PREDICTION OF ENERGY PERFORMANCE GAP IN BUILDINGS WITH MACHINE LEARNING ALGORITHMS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

DERYA YILMAZ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
BUILDING SCIENCE IN ARCHITECTURE

JULY 2022

Approval of the thesis:

PREDICTION OF ENERGY PERFORMANCE GAP IN BUILDINGS WITH MACHINE LEARNING ALGORITHMS

Submitted by **DERYA YILMAZ** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy** i**n Building Science in Architecture, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. F. Cânâ Bilsel
Head of Department, **Architecture** _____

Prof. Dr. Ali Murat Tanyer
Supervisor, **Architecture Dept., METU** _____

Prof. Dr. İrem Dikmen Toker
Co-Supervisor, **Civil Engineering Dept., METU** _____

Examining Committee Members:

Prof. Dr. Arzu Gönenç Sorguç
Architecture Dept., METU _____

Prof. Dr. Ali Murat Tanyer
Architecture Dept., METU _____

Prof. Dr. Gülser Çelebi
Interior Architecture Dept., Çankaya University _____

Assoc. Prof. Dr. Başak Güçyeter
Architecture Dept., Eskişehir Osmangazi University _____

Assist. Prof. Dr. Mehmet Koray Pekeriçli
Architecture Dept., METU _____

Date: 08.07.2022

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name Last name: Derya Yılmaz

Signature:

ABSTRACT

PREDICTION OF ENERGY PERFORMANCE GAP IN BUILDINGS WITH MACHINE LEARNING ALGORITHMS

Yılmaz, Derya
Doctor of Philosophy, Building Science in Architecture
Supervisor: Prof. Dr. Ali Murat Tanyer
Co-Supervisor: Prof. Dr. İrem Dikmen Toker

July 2022, 274 pages

The energy performance gap presents obstacles for governments to reduce dependency on foreign energy sources, as well as for policymakers to be successful in future ambitions about reducing the environmental impacts of buildings. Although considerable research focused on the reasons, reduction, or quantification of the energy performance gap in buildings, little attention has been given in the past to methods that would inform decision makers about a likely performance gap. Such studies could change the actions and strategies of decision makers to achieve expected energy performance levels in buildings. This research supports the idea that performance gap occurs due to project risks occurring during the building life cycle. Therefore, the research examined the energy performance gap in buildings through the lens of project risks and demonstrated the potential of the machine learning classification technique in the prediction of performance gaps.

To achieve the aim of the research, first, a conceptual performance gap framework was proposed after a literature review and conducting semi-structured interviews with 11 experts working on six energy-efficient buildings. Later, the framework was designed as a web-based survey for data collection. Information about the

performance gap and project risks of a total of 77 buildings was collected using the web-based survey. Lastly, using this dataset, the performance of the four machine learning classification algorithms (1) Naïve Bayes, (2) KNN, (3) SVM, and (4) Random Forest were compared to find the best prediction model based on six performance metrics (Accuracy, kappa statistic, precision, recall, F-measure, and AUC). Energy performance gap prediction problems were studied focusing on the heating demand and electricity demand of buildings. First, datasets were studied to distinguish cases that might show a positive and negative performance gap (binary classification problem). Following that, smaller datasets were studied to predict the level of performance gap in buildings using the following categories, low (0-15%), medium (15.1-40%), and high (40.1-90%), (multi-class classification problem).

The results indicate that *Naïve Bayes* gives the best classification accuracy on the majority of the classification problems. While *KNN* shows the highest accuracy for electricity performance gap prediction (79.00%), *Naïve Bayes* is the suggested algorithm for heating performance gap prediction (72.50% accuracy) to solve binary classification. Moreover, in multi-class classification, while Naïve Bayes achieved an accuracy of 77.08% for positive electricity performance gap, SVM achieved an accuracy of 76.04% for positive heating performance gap prediction. Lastly, in multi-class classification, Naïve Bayes provided an accuracy of 83.85% and 71.81% to solve the negative electricity performance gap and heating performance gap, respectively.

Keywords: Energy Performance Gap, Machine Learning, Classification, Algorithms, Risk Identification

ÖZ

MAKİNE ÖĞRENME ALGORİTMALARIYLA BİNALARDA ENERJİ PERFORMANS FARKININ TAHMİNİ

Yılmaz, Derya
Doktora, Yapı Bilimleri, Mimarlık
Tez Yöneticisi: Prof. Dr. Ali Murat Tanyer
Ortak Tez Yöneticisi: Prof. Dr. İrem Dikmen Toker

Temmuz 2022, 274

Enerji performansı farkı, hükümetlerin yabancı enerji kaynaklarına bağımlılığı azaltmalarının yanı sıra, politika yapıcıların binaların çevresel etkilerini azaltma konusundaki gelecekteki hedeflerinde başarılı olmaları için engeller sunuyor. Binalardaki enerji performans farkının nedenleri, azaltılması veya nicelleştirilmesine odaklanan önemli araştırmalara rağmen, geçmişte karar vericileri olası bir performans boşluğu hakkında bilgilendirecek yöntemlere çok az ilgi gösterilmiştir. Bu tür çalışmalar, binalarda beklenen enerji performans seviyelerine ulaşmak için karar vericilerin eylemlerini ve stratejilerini değiştirebilir. Bu araştırma, bina yaşam döngüsü boyunca meydana gelen proje riskleri nedeniyle performans farkı oluştuğu fikrini desteklemektedir. Bu nedenle araştırma, binalardaki enerji performans farkını proje riskleri merceğinden incelemiş ve performans farklarının tahmininde makine öğrenimi sınıflandırma tekniğinin potansiyelini ortaya koymuştur.

Araştırmanın amacına ulaşmak için, ilk olarak, literatür taraması ve altı enerji verimli bina üzerinde çalışan 11 uzmanla yarı yapılandırılmış görüşmeler yapıldıktan sonra kavramsal bir performans farkı çerçevesi önerilmiştir. Daha sonra çerçeve, veri toplama için web tabanlı bir anket olarak tasarlanmıştır. Web tabanlı anket kullanılarak toplam 77 binanın performans farkı ve proje riskleri hakkında bilgi

toplanmıştır. Son olarak, bu veri seti kullanarak, dört makine öğrenimi sınıflandırma algoritmasının (1) Naïve Bayes, (2) KNN, (3) SVM, (4) Random Forest performansı, altı performans metriğine dayalı olarak en iyi tahmin modelini bulmak için karşılaştırılmıştır (Doğruluk, kappa istatistiği, kesinlik, geri çağırma, F-skoru ve AUC). Enerji performans farkı tahmin problemi binaların ısıtma talebi ve elektrik talebine odaklanarak incelenmiştir. İlk olarak, pozitif ve negatif performans farkı gösterebilecek vakaları ayırt etmek için veri kümeleri incelenmiştir (ikili sınıflandırma problemi). Bunu takiben, düşük (% 0-15), orta (% 15,1-40) ve yüksek (% 40,1-90), kategorileri kullanılarak binalardaki performans farkı seviyesini tahmin etmek için daha küçük veri kümeleri incelenmiştir (çok sınıflı sınıflandırma problemi).

Sonuçlar, Naive Bayes'in sınıflandırma problemlerinin çoğunda en iyi sınıflandırma doğruluğunu verdiğini göstermektedir. KNN, elektrik performans farkı tahmini için en yüksek doğruluğu (%79,00) gösterirken, Naïve Bayes, ikili sınıflandırmada ısıtma performans farkı tahmini için önerilen (%72,50 doğruluk) algoritmadır. Ayrıca çok sınıflı sınıflandırmada Naïve Bayes pozitif elektrik performans farkı için %77,08 doğruluk elde ederken, SVM pozitif ısıtma performans farkı tahmini için %76,04 doğruluk elde etmiştir. Son olarak, çok sınıflı sınıflandırmada Naïve Bayes, negatif elektrik performans farkı ve ısıtma performans farkını çözmek için sırasıyla %83,85 ve %71,81 doğruluk sağlamıştır.

Anahtar Kelimeler: Enerji Performans Farkı, Makine Öğrenimi, Sınıflandırma, Algoritmalar, Risk Tanımlama

To my dear Aren, Arden

and

all children

ACKNOWLEDGMENTS

First of all, I would like to express my deepest gratitude to my supervisor, Prof. Dr. Ali Murat Tanyer, and co-supervisor Prof.Dr.İrem Dikmen Toker, for their guidance, advice, and criticism throughout the research. Prof. Dr. Ali Murat Tanyer accompanied me tirelessly for years when I needed his guidance most. I am very grateful to Prof.Dr.İrem Dikmen Toker for agreeing to join us and guiding me at very critical times. After, I would like to express my appreciation to the dissertation committee members who contributed to my work with their valuable comments and feedback.

Next, I would like to thank Prof. Glenn Kukkola, who attaches great importance to transferring knowledge and experience to his students. It is a big chance for me to have the opportunity to work with you and learn from you.

Afterward, I would like to thank Prof.Dr.Conrad Völker. If he had not accepted my application two years ago, I would not have been able to meet Building Physics team at Bauhaus University Weimar, where I spent one of the best periods of my life. It was a unique opportunity to see the importance they attach to collaboration and to experience their research culture. Many thanks to Anja Willmann for her valuable time, my friends Ammar who accompanied me with his sincerity and moral support, Mara with her cheerful mood, and Alex with his generosity. Many thanks to Chance for his friendship, and of course family Heinrich for our countless wonderful days in Weimar.

My dear Aren and Arden, your presence and laughter, even though we are far away, the games we played with you, and the daily lives that we shared from mobile screens

shortly, gave me a breath when I was very tired. I am very happy about your presence.

My dear parents, Yücel Yılmaz and Bedriye Kalın thank you so much for having the patience to witness me overcome yet another challenge.

I want to thank my friends who accompanied me from time to time on this journey, Bircan, Melike, Gülşah, Raziye, Forró Ankara family, Damla, and Serhat, for the motivation and advice you have given.

Lastly, I would like to thank Yalduz Family, Kalın Family, Murat Ayhan and İlknur Baygün. I think she has a great passion for life and helping others. I am really pleased we met.

This work is partially funded by the Scientific and Technological Research Council of Turkey under grant number TUBİTAK 1059B142000267.

TABLE OF CONTENTS

ABSTRACT ..... **v**

ÖZ…… ..... **vii**

ACKNOWLEDGMENTS ..... **xi**

LIST OF TABLES ..... **xix**

LIST OF FIGURES ..... **xxi**

LIST OF ABBREVIATIONS ..... **xxv**

LIST OF SYMBOLS ..... **xxvii**

CHAPTERS

1. INTRODUCTION **29**

1.1. Background of the Research 29

1.2. Problem Definition 32

1.3. Aim and Objectives of the Research 34

1.4. Disposition 34

2. LITERATURE REVIEW **37**

2.1. Building Performance 37

2.1.1. Evaluating Energy Performance of Buildings 42

2.1.2. Energy Performance Simulation Tools 45

2.1.3. Energy Performance Gap in Buildings 49

2.1.3.1. Disadvantages of Energy Performance Gap 54
2.1.3.2. Reasons of Energy Performance Gap 55
2.1.3.3. How to Decrease the Energy Performance Gap 63
2.2. Risk Assessment in Buildings 68
2.2.1. Risk and its Definition 68
2.2.2. Risk Assessment in Construction Sector 70
2.2.3. Risk Studies in Energy Efficient Buildings 74
2.3. Machine Learning Concept 77
2.3.1. Classification in Machine Learning 79
2.3.2. Machine Learning in Energy Performance of Buildings 80
2.3.3. Machine Learning Algorithms 82
2.3.3.1. Naïve Bayes Algorithm 83
2.3.3.2. Nearest Neighbor Algorithm 84
2.3.3.3. Support Vector Machines 85
2.3.3.4. Random Forests 88
2.3.4. Performance Measures in Machine Learning 90
2.3.5. Feature Selection in Machine Learning 94
2.3.6. Synthetic Minority Over-sampling Technique 96
2.4. Inferences Drawn from Literature Review 98
3. METHOD AND MATERIAL 101

3.1. Method 101
3.1.1. Energy Performance Gap Framework 103
3.1.1.1. Semi Structured Interviews 103
3.1.1.2. Case Studies 105
3.1.1.3. Evaluation of the Project Process 108
3.1.1.4. Results of Semi-Structured Interviews 115
3.1.1.5. Final Conceptual Framework 121
3.1.1.6. Web-based Survey 126
3.1.2. Machine Learning Application 128
3.1.2.1. Binary Classification to Predict Electricity EPG 132
3.1.2.2. Multi Class Classification to Predict Positive EPG 138
3.1.2.3. Multi Class Classification to Predict Negative EPG 142
3.1.2.4. Binary Classification to Predict Heating EPG 143
3.1.2.5. Multi Class Classification to Predict Positive EPG 144
3.1.2.6. Multi Class Classification to Predict Negative EPG 145
3.1.3. The Weka Workbench 145
3.2. Material 149
3.2.1. Dataset used in the ML Classification Models 149
3.2.2. Analysis of Data in the Dataset 152
4. RESULTS 157

4.1. Electricity Performance Gap Prediction.....157
4.1.1. Binary Classification of Electricity Performance Gap 157
4.1.1.1. Selected Attributes with Wrapper Attribute Evaluator.....157
4.1.1.2. Performance Evaluation for Electricity Gap Prediction 159
4.1.2. Multi Class Classification of Electricity Performance Gap 163
4.1.2.1. Selected Attributes for Positive EPG.....163
4.1.2.2. Performance Evaluation for Positive Electricity Gap Prediction….....164
4.1.2.3. Selected Attributes for Negative EPG 166
4.1.2.4. Performance Evaluation for Negative Electricity Gap Prediction.....167
4.2. Heating Performance Gap Prediction.....168
4.2.1. Binary Classification of Heating Performance Gap 169
4.2.1.1. Selected Attributes with Wrapper Attribute Evaluator.....169
4.2.1.2. Performance Evaluation for Heating Gap Prediction 171
4.2.2. Multi Class Classification of Heating Gap.....173
4.2.2.1. Selected Attributes for Positive EPG.....173
4.2.2.2. Performance Evaluation for Positive Heating Gap Prediction ……..174
4.2.2.3. Selected Attributes for Negative EPG 176

4.2.2.4. Performance Evaluation for Negative Heating Gap Prediction……..177

4.3. Comparison of the Results 179

5. CONCLUSION 181

5.1. Discussion of Findings 181

5.1.1. Discussion of Findings from Semi-Structured Interviews 182

5.1.2. Discussion of Findings related to Electricity Gap Database 184

5.1.3. Discussion of Findings related to Heating Gap Database 189

5.1.4. Discussion of Findings related to Machine Learning Models 194

5.2. Research Limitations 206

5.3. Recommendations for Future Work 208

REFERENCES 211

6. APPENDICES 239

A. How to Reduce EPG 239

B. Approval of Applied Ethics Research Center 243

C. Interview Questions 244

D. Cognitive Maps of the Case Studies 245

E. Risk-Vulnerability EPG Framework 248

F. Final Conceptual Framework: Unexpected Events and Changes, Project Management, Assumptions 253

G. Web-based Survey: Questions 258

H. Some Performance Measures for a Classfier (Bramer, 2020).....262

İ. List of attributes considered in ML models 263

J. Dataset used in ML models.....264

K. Analysis of Electricity Gap Dataset 268

L. Analysis of Heating Gap Dataset 269

M. Conceptual Web-Interface to Predict the Negative Electricity Performance Gap.....270

N. Conceptual Web-Interface to Predict the Positive Heating Performance Gap……..270

O. Conceptual Web-Interface to Predict the Negative Heating Performance Gap…….……271

CURRICULUM VITAE 273

LIST OF TABLES

TABLES

Table 2.1 Reasons of Energy Performance Gap in Buildings 62

Table 2.2 Risks in Construction Projects 69

Table 2.3 Example of a 2x2 Confusion Matrix 92

Table 3.1 Interviewee Profile 105

Table 3.2 Project Profile 106

Table 3.3 Problems/Limitations: Design Phase Factors 122

Table 3.4 Problems/Limitations: Construction Phase Factors 123

Table 3.5 Problems/Limitations: Use and Operational Phase Factors 124

Table 3.6 Types of Data 152

Table 4.1 Attributes Selected with Naïve Bayes Classifier for Electricity Gap .. 158

Table 4.2 Attributes Selected with KNN Classifier for Electricity Gap 158

Table 4.3 Attributes Selected with SVM Classifier for Electricity Gap 158

Table 4.4 Attributes Selected with Random Classifier for Electricity Gap 159

Table 4.5 Performance Evaluation of the Classifiers 160

Table 4.6 Attributes Selected using ChiSquaredAttributeEval for Positive Electricity Gap 163

Table 4.7 Performance Evaluation of the Classifiers 165

Table 4.8 Attributes Selected using ChiSquaredAttributeEval for Negative Electricity Gap 166

Table 4.9 Performance Evaluation of the Classifiers 167

Table 4.10 Attributes Selected with Naïve Bayes Classifier for Heating Gap 169

Table 4.11 Attributes Selected with KNN Classifier for Heating Gap.....170

Table 4.12 Attributes Selected with SVM Classifier for Heating Gap.....170

Table 4.13 Attributes Selected with Random Forest Classifier for Heating Gap.170

Table 4.14 Performance Evaluation of the Classifiers 171

Table 4.15 Attributes Selected using ChiSquaredAttributeEval for Positive Heating Gap.....174

Table 4. 16 Performance Evaluation of the Classifiers 175

Table 4.17 Attributes Selected using ChiSquaredAttributeEval for Negative Heating Gap.....176

Table 4.18 Performance Evaluation of the Classifiers 178

Table 5.1 Analyses of Electricity Gap Dataset.....189

Table 5.2 Analyses of Heating Gap Dataset.....193

Table 5.3 Top Two Performing Classifiers for Classification Problems.....205

LIST OF FIGURES

FIGURES

Figure 1.1 Outline of the Research 35

Figure 2.1 The BRPU Conceptual Model 38

Figure 2.2 Building Performance Context. 42

Figure 2.3 Selected History of Building Performance Tools, 1960-present 46

Figure 2.4 Number of Publications and the Highest Cited Authors per Year..... 50

Figure 2.5 Main Focus of the Most Cited Publications 51

Figure 2.6 The Risk Management Process based on ISO 31000:2018 71

Figure 2.7 Techniques used in Risk Identification..... 72

Figure 2.8 The Process of Developing ML Algorithms..... 78

Figure 2.9 The Structure of the Naïve Bayes Network. 83

Figure 2.10 The SVM Algorithm..... 86

Figure 3.1 Steps Applied in the Research 102

Figure 3.2 Risk and Vulnerability Paths 116

Figure 3.3 Cognitive Map of Building A 117

Figure 3.4 Renewed Conceptual Model..... 119

Figure 3.5 Finalized EPG Framework 120

Figure 3.6 Goal of the Classification Problems Applied in the Research..... 129

Figure 3.7 Flowchart Explaining the Method of the Research 131

Figure 3.8 Weka GUI Chooser 146

Figure 3.9 Weka Experimenter Interface 147

Figure 3.10 Countries and Construction Period of Cases in Dataset..... 151

Figure 3.11 Heated Floor Area and Construction Type 151

Figure 4.1 Performance of the Final Model: KNN Classifier 162

Figure 4.2 Performance of the Final Model: Naïve Bayes 166

Figure 4.3 Performance of the Final Model: Naïve Bayes 168

Figure 4.4 Performance of the Final Model: Naïve Bayes 173

Figure 4.5 Performance of the Final Model: SVM..... 176

Figure 4.6 Performance of the Final Model: Naïve Bayes 179

Figure 5.1 Some of the Selected Attributes and their Values in the Electricity Gap Dataset 185

Figure 5.2 Some of the Selected Attributes and their Values in the Heating Performance Gap Dataset 190

Figure 5.3 Conceptual Web-Interface to Predict a Positive or Negative Gap 195

Figure 5.4 Selected Attributes and their Values in the First Classification Problem 196

Figure 5.5 Conceptual Web-Interface to Predict the Positive Gap for Electricity Demand..... 197

Figure 5.6 Selected Attributes and their Values in the Second Classification Problem..... 198

Figure 5.7 Selected Attributes and their Values in the Third Classification Problem 200

Figure 5.8 Conceptual Web-Interface to Predict a Positive or Negative Heating Gap 201

Figure 5.9 Selected Attributes and their Values in the Fourth Classification Problem 202

Figure 5.10 Selected Attributes and their Values in the Fifth Classification Problem 203

Figure 5.11 Selected Attributes and their Values in the Sixth Classification Problem 204

LIST OF ABBREVIATIONS

ABBREVIATIONS

AI	Artificial Intelligence
AC	Air Conditioner
AUROC	Area under ROC Curve
ARFF	Attribute Relation File Format
CSV	Comma Separated Values
ECC	Exhaustive Correction Code
EPC	Energy Performance Contracting
ESCO	Energy Service Companies
FN	False Negative
FP	False Positive
IAQ	Indoor Air Quality
KNN	K-Nearest Neighbor
ML	Machine Learning
OVA	One-Vs-All

OVO	One-Vs-One
RBF	Radial Basis Function
RCC	Random Correction Cod
RF	Random Forest
SVM	Support Vector Machines
SMOTE	Synthetic Minority Over-sampling Technique
TP	True Positive
TN	True Negative
WEKA	Waikato Environment for Knowledge Analysis

LIST OF SYMBOLS

SYMBOLS

α	Predetermined level of significance
C	Penalty parameter in SVM
diff	Difference between X_i and its nearest neighbors
Gap	Random number
IDT	Decision tree inducer
K	Number of nearest neighbors
N	Number of attributes used in each node
n_i	Synthetic samples
S	Training set
T	Number of iterations
μ	Subsample size
$\bar{x}$	Mean of a set of samples
X_i	Considered sample
$\bar{Y}$	Mean of another set of samples

CHAPTER 1

INTRODUCTION

This chapter presents an overview of the thesis by providing information on research background, definition of the problem, aim and objectives of the research, contribution of the thesis, and disposition.

1.1 Background of the Research

Intergovernmental organizations and governments around the world implement policies to promote energy reduction and carbon mitigation because of the rapid growth of global energy consumption (Lu & Lai, 2019). A wide range of methodologies has been developed for predicting the energy performance of buildings since the introduction of the first building regulations aimed at energy conservation in the 1970s (Borgstein *et al.*, 2016). In addition to the buildings' widespread negative impacts on the environment, a large amount of energy is being consumed or wasted in the buildings' lifecycle.

As a result, many rating systems, benchmarking, and standards have been developed like the US's Leadership in Energy and Environment Design (LEED), the UK's Building Research Establishment Environment Assessment Method (BREEAM), Australia's Green Star, and the European Union's Passive House (Zou *et al.*, 2018) to support resource efficiency, and the spread of buildings that are more environmentally friendly.

However, many publications in the literature confirm that there is a difference between a building's planned and actual performance (Galvin, 2014; Fedoruk *et al.*, 2015; Figueiredo *et al.*, 2018; Liang *et al.*, 2019; Cuerda *et al.*, 2020), whether they are new or retrofitted (Calì *et al.*, 2016; Niu *et al.*, 2016). This phenomenon has been defined as "the energy performance gap." It is universally accepted that "gap" means "difference" or "discrepancy", and "the performance gap" is usually referred to as the difference between the performance predicted (or anticipated, calculated, designed, *etc*.) in the design stage, and that measured (or real, actual, achieved, *etc*.) in the occupation stage (Shi *et al.*, 2019).

According to Gram-Hanssen & Georg (2018), the energy performance gap is not an entirely new problem. The term emerged in the mid-1990s (Niu *et al.*, 2016), and it has been widely studied for well over two decades (Shi *et al.*, 2019). Moreover, conducting a literature survey using the Web of Science database indicated that 157 publications addressed the energy performance gap phenomenon in the fields of Architecture and Civil Engineering between 2011 and 2021. The average number of publications per year was 15.7, and the number of publications has increased sharply since 2014.

Although the performance gap is usually connected with increased energy consumption, it can also mean reduced consumption (Mojic *et al.*, 2019). However, measured energy use is higher than predicted in many cases. Moreover, although the performance might be better than expected in some areas, it might be worse in others (De Wilde, 2018). For instance, the measured energy output of an installed PV system can be higher than the expectations. Even so, it could cause the unreasonable design of the system and improper use of capital investment (Harrison & Jiang, 2018).

Janser *et al.* (2020) observed that studies on this issue often cope with one or more of the following topics: how to define, quantify, explain, and overcome energy performance gaps. Agreeing with Janser *et al.* (2020)'s inference, in the reviewed energy performance gap literature, a common point is that some researchers demonstrated the importance of using a life-cycle perspective. Brejnrod *et al.* (2017) state that the focus on the whole life cycle perspective of buildings has become more common in recent years. As anticipated, previous researchers underlined the causes of the energy performance gap (De Wilde, 2014; Zou *et al.*, 2018; Eon *et al.*, 2020), and identified the ways to reduce the gap (Fedoruk *et al.*, 2015), or outlined a methodology which produces accurate building performance (Horrigan *et al.*, 2015) considering a life-cycle perspective.

Moreover, studies quantifying the performance gap in the literature showed that the results could be almost double the target energy use intensity (Fedoruk *et al.*, 2015), could present an average deviation of 30% (Herrando *et al.*, 2016), or could show values up to 287% performance gap (Calì *et al.*, 2016). As a result, several researchers suggested different approaches because of the importance of reducing the discrepancy. While Demanuele (2008) highlights the significance of the quality of data and the use of a probabilistic and statistical approach to simulation, De Wilde (2018) suggests reexamining the tools used in the sector regarding this issue.

Consequently, the literature review indicated that reducing the performance gap can be addressed in many different ways. Some researchers conducted pre-occupancy (Niu *et al.*, 2016), post-occupancy evaluations (Menezes *et al.*, 2012), or suggested dealing with uncertainty in energy simulations (Li *et al.*, 2019b). Some others simulated the occupant behavior through exhaustive modeling (Ali *et al.*, 2020), or used monitoring data to calibrate simulation models (Cuerda *et al.*, 2020). Besides, some researchers dealt with the problem from a stakeholder perspective like designers (Grover, 2019), contractors, policymakers (Dollard, 2018), industry, and universities (Imam *et al.*, 2017). As Gram-Hanssen & Georg (2018) highlighted,

there are many stakeholders apart from the occupants whose actions must be taken into account to "close the gap".

1.2 Problem Definition

The energy performance gap can cause negative impacts on a small and large scale. Shrubsole *et al.* (2019) state that the performance gap risks weaken design, technology, sustainability, economic, health, and well-being benefits. Under-performance can be a burden not only on the energy performance contractors but also on the building owners or occupants because of added costs, longer payback periods, and the potential for contractual disputes (Mohareb *et al.*, 2017). Moreover, it can damage the reputation of the person behind this failure like the designer or contractor of the project (Mohammadi, 2015), and people can be hesitant to adopt the energy performance certification system for buildings (Abela *et al.*, 2016).

Additionally, the performance gap also introduces reputational dangers to the industry (Dollard, 2018). Since the general EU policy is focused on restricting overall energy consumption in buildings and reducing harmful emissions into the environment, the investment potential in buildings is high (Nižetić & Papadopoulos, 2017). However, the performance gap affects the implementation of different policy programs (Goulden *et al.*, 2020), making it difficult for policymakers to successfully make plans for future strategies (Li *et al.*, 2019a) like the zero-carbon housing policy (Gupta *et al.*, 2013). Furthermore, significant increases in energy cost and environmental impact can be seen due to the discrepancies (Birchall, 2011).

A review of the literature reveals that the factors that can contribute to the energy performance gap (EPG) in buildings have been studied in various publications (Menezes *et al.*, 2012; De Wilde, 2014; Cali *et al.*, 2014; Herrando *et al.*, 2016; Zou *et al.*, 2018; Eon *et al.*,2020) sometimes by examining risks of the green building sector in different countries (Zhao *et al.*, 2016; Qin *et al.*, 2016; Yang *et al.*, 2016a;

Yang *et al.*, 2014). Doylend (2015), Topouzi *et al.* (2019), and Alam *et al.* (2017) addressed that the energy performance gaps were the result of risk factors occurring during different building life-cycle stages. However, several researchers have addressed the significance of studying risk routes instead of studying risk factors separately to have a better understanding of their impact and how to deal with them (Yildiz *et al.*, 2014; Liu *et al.*, 2016b; Qazi *et al.*, 2016).

Besides, Nižetić & Papadopoulos (2017) mentioned the importance of having a tool for estimating a possible gap between real and expected energy savings and proposed a novel but just a conceptual method for energy performance gap prediction. However, aside from this study, no approach has been found aiming to predict the energy performance gap in buildings, especially by using a group of projects from the past with energy performance gap information and the experienced project risks. Machine Learning (ML), as a part of a growing field of data science, proposes to predict future events and scenarios that are not well-known to the computer by using example data or former experience. Moreover, ML methods appear frequently in energy performance prediction studies (Paudel *et al.*, 2017; Seyedzadeh *et al.*, 2020b; Arjunan *et al.*, 2020; Walker *et al.*, 2020). However, to the best of the knowledge of the researcher, studies have not been performed to predict the energy performance gap in buildings by utilizing ML methods. By identifying the patterns in the data that would be usually difficult to be noticed by humans, ML could allow project stakeholders to understand the risk of an energy performance gap with its direction and level. In this way, decision-makers can take action to change their project decisions and suggest new strategies to control the gap.

Therefore, this research, through the lens of project risks, proposes energy performance gap prediction models for buildings by utilizing ML algorithms. This research contributes to the knowledge base by (1) developing several ML models by using past experience to predict future scenarios of energy performance gaps in buildings, and (2) determining risk routes of energy-efficient buildings instead of

studying the project risks separately, (3) creating a risk-energy performance gap database of buildings that represents risk histories of various energy-efficient buildings.

1.3 Aim and Objectives of the Research

This research aims to answer the question of 'Can we know the risk of energy performance gap in buildings beforehand?' and propose a model by which these discrepancies can be predicted. These aims are reflected in the following objectives:

- to identify risks leading to performance gaps in buildings in the project life cycle.
- to map major causal factors resulting in discrepancies between design and actual building energy performance.
- to establish a risk database of buildings where an energy performance gap exists.
- to compare the performance of various ML algorithms to propose a model that offers the highest prediction performance.

1.4 Disposition

This thesis consists of five chapters. *The first chapter* starts with providing information on the research background, the definition of the problem, the aim and objectives of the research, the contribution of the thesis, and disposition.

The second chapter presents the concepts related to the research context about the energy performance gap, risk assessment concerning the construction and energy-efficient building projects, and machine learning. Relevant studies in the literature are briefly explained, and the chapter is finalized with a discussion of the literature.

The third chapter aims to explain the method and the material of the study. First, the steps followed to achieve the aims of the study were presented briefly. Subsequently,

semi-structured interviews focusing on the energy performance gap context are discussed. Later, the proposed conceptual EPG framework is presented. The chapter continues with describing a web-based survey founded on a conceptual EPG framework to gather data. After that, it explains ML application as it is the suggested method to solve performance gap prediction problems. Then, the software, Weka Version 3.8.5, used to solve ML classification problems is introduced. Finally, the buildings in the dataset are presented as the material of the research, and an analysis of data in the dataset is given. Figure 1.1 represents the stages of the research.

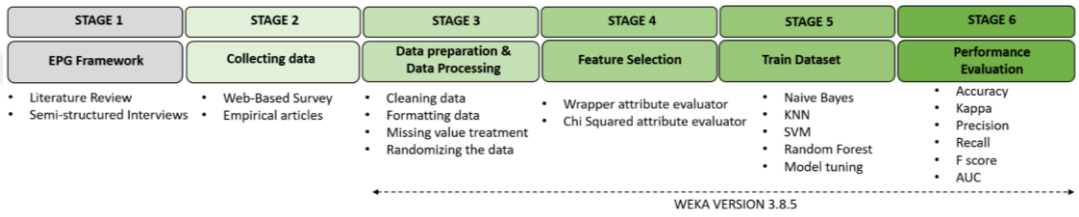


Figure 1.1 Outline of the Research

Chapter four presents the results of performance comparison of six different classification problems for EPG prediction. In the last chapter, *Chapter five*, the major findings of the research, its contributions to the theory, expected benefits for practice, research limitations, and recommendations for further studies are stated.

CHAPTER 2

LITERATURE REVIEW

This chapter overviews the theoretical background of building energy performance analysis, risk assessment, and machine learning subjects. To do this, first, an introduction to building performance context is given. Then, evaluation methods of building energy performance are explained by introducing energy performance simulation tools. Next, energy performance gap literature is reviewed considering its disadvantages, the reasons, and mitigation methods. After that, risk assessment concepts are described by giving various definitions of risk and introducing methods applied in the construction sector and energy-efficient buildings. Later, the machine learning concept is reviewed focusing on classification techniques, studies related to building energy performance, machine learning algorithms used in the research, performance metrics, feature selection, and oversampling techniques.

2.1 Building Performance

Performance levels emerge from human needs because of the interaction of users with various settings in the built environment (Preiser & Vischer, 2006). The term "building performance" indicates that buildings can be defined, and their performance can be assessed (Baird *et al.*, 2018). Building performance is usually expressed by key performance indicators which are measurable variables to analyze and assess the performance of the building such as electricity consumption, gas consumption, CO_2 emissions, and heating and cooling balance (van Heijst, 2017).

Baird *et al.* (2018) explain that systematic studies on building performance provided by the Building Performance Research Unit (BRPU) at Strathclyde University were pioneering in the 1970s. This unit suggested a model which is composed of five principal systems including the objectives system, the activity system, the environmental system, the building system, and the resources system (Figure 2.1). The model not only provided a general framework for the analysis of other sides of building performance but also a different aspect by being interested in the building and all the people involved.

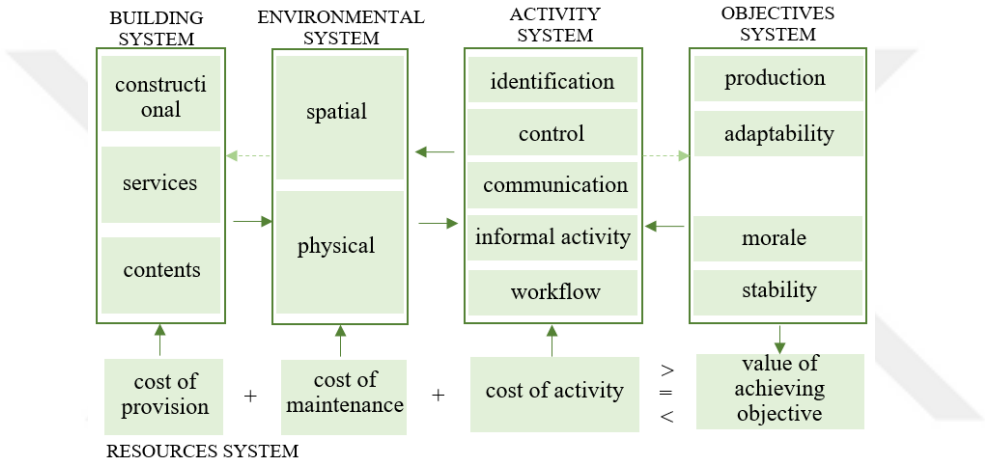


Figure 2.1 The BRPU Conceptual Model (modified from Baird *et al.*, 2018).

While Hartkopf *et al.* (1986) identify health, safety, and welfare as the traditional building demands, Preiser &Vischer (2006) depicted three levels of priority in building performance:

- health, safety, and security performance.
- functional, efficiency, and workflow performance.
- aesthetic, cultural, social, and psychological performance.

Each category as a goal has a sub-category or sub-goals. Performance levels interact with several sub-goals and may also conflict with each other and require a solution

to be effective. According to De Wilde (2018), different disciplines have their own way to performance, and different stakeholders involved in building production define performance differently. The author identifies three main viewpoints of the building performance concept: an engineering view, a process view, and an aesthetic view. While the engineering view is focused on how well a building performs its functions and tasks, the process view is focused on how well the construction process provides buildings. Lastly, the aesthetic view is focused on the success of buildings as a form of appreciation or presentation.

- **Energy Consideration on Building Performance**

Energy considerations started to impact the building design in other forms with the approach of the industrial age and the energy-intensive buildings lead to an era of energy growth (Baird *et al.*, 2018). Hartkopf *et al.* (1986) point out that the international prominence given especially on energy and economical resources those years puts a stress on the building industry. Nowadays, because of the raising requirements for quality and building performance, numerous building stakeholders start to give attention to policy, energy, economic, environmental, and comfort aspects (Zheng *et al.*, 2018).

Energy performance is a complex physical process founded on building characteristics (e.g., orientation, envelope, etc.), environmental effects (e.g., shading, climate, etc.), occupancy patterns (Güçyeter & Günaydın, 2012; Delzendeh *et al.* 2017), installed equipment (Amasyali & El-Gohary, 2018), lifestyle and culture (Van Someren, 2017). Energy efficiency becomes an especially critical design principle when fossil fuels are unavailable (Baird *et al.*, 2018). Although the building sector offers many innovative building solutions for energy efficiency such as building integrated electricity production, switchable glazing, super insulation, energy storage systems, adaptive building skins, and demand response (Hensen & Lamberts, 2019),

a review of building performance literature has shown some criticisms about sustainability ambitions and the green buildings approach. For instance, Hamza (2013) criticizes that sustainability ambitions have a reductive performative notion, which considers buildings as simply optimized machines. Besides, Yudelson (2016) claims that the green building approach has comprehensive and excessive technical criteria, many complex rating systems, large and cumbersome bureaucracies, excessive costs, and limited emphasis on real long-term building performance. Additionally, Hensen & Lamberts (2019) argue that indoor environmental quality for well-being and health will gain more importance than energy-related performance in close future. Similarly, some researchers highlight the importance of improving the employee's well-being and productivity (Hamza, 2013), considering the complex relationship between energy, indoor air quality (IAQ), and productivity holistically (Burman *et al.*, 2018). Although understanding the critical balance to ensure all building performance demands (air quality, visual comfort, spatial comfort, thermal comfort, building integrity, and acoustic comfort) is defined as the challenge of the 1980s and 1990s by Hartkopf *et al.* (1986), it seems that it will continue to be a challenge of the future.

- **Complexity of Building Performance**

Building performance context has numerous limitations and complexities. The challenges of the built environment such as the increasing complexity of buildings and integrated building systems (Hensen & Lamberts, 2019) also bring challenges to the building performance. Primarily, buildings consist of many products which can be combined in an infinite number of possibilities, and this requires understanding specific functions provided by them (De Wilde, 2018). Related to this topic, Hartkopf *et al.* (1986) address a limitation of the understanding of the building performance. The authors explain that designing an assembly (e.g., structural, exterior enclosure, etc.) only to the respective component performance requirement leads to the inadequacy of two assemblies to sustain different performance mandates.

Besides, focusing on building retrofit, Regnier *et al.* (2018) state that building systems are often treated as isolated individual systems. As a result, different stakeholders frequently prefer a staged approach or single equipment replacements which ignores the interactive effects between various building systems and prevents extra savings. On the other hand, the difference between the life expectancy of buildings and the service life of building systems necessitates paying special attention to maintenance, renovation, and replacement (De Wilde, 2018).

Furthermore, Hensen & Lamberts (2019) think that some of the challenges of the built environment are unique challenges in the design, building, and operational phases. However, the focus on the whole life cycle perspective of buildings has become more common in recent years (Brejnrod *et al.*, 2017). In addition, the wide variety of stakeholders in the built environment with varying and often conflicting interests (Hensen & Lamberts, 2019) and the tension between some fields such as architecture and engineering bring another complexity to it (De Wilde, 2018).

Lastly, several authors state that instead of oversimplifying the performance of a building to a record of the percentage of final energy use, or the related greenhouse gas emissions, there is a necessity to study issues with IAQ, occupant health, comfort, and productivity in it (Hamza, 2013; Burman *et al.*, 2018, De Wilde, 2018; Hensen & Lamberts, 2019). Figure 2.2 summarizes the context of building performance and depicts the complexity involved in it considering the aforementioned readings on building performance.

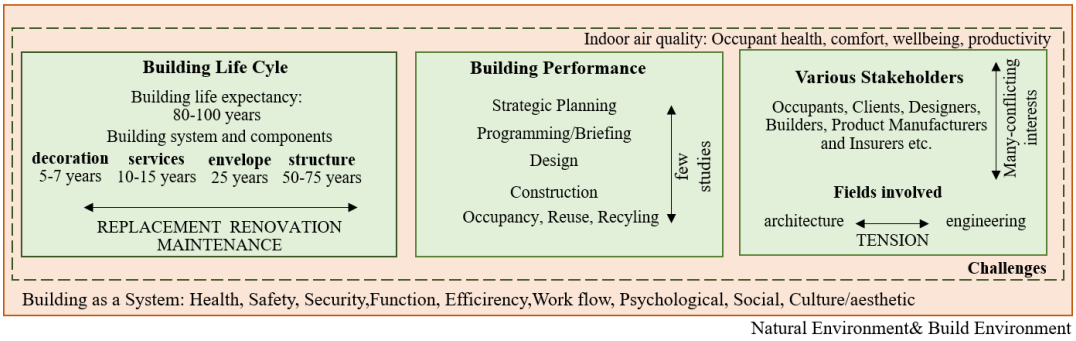


Figure 2.2 Building Performance Context (De Wilde, 2018; Hamza, 2013; Burman *et al.*, 2018; Hensen & Lamberts, 2019).

2.1.1 Evaluating Energy Performance of Buildings

A large range of methodologies has been developed to predict, analyze, and evaluate the building energy performance since the introduction of the first building regulations intended for energy conservation in the 1970s, the increased emphasis on this area in the late 1990s and early 2000s (Borgstein *et al.*, 2016).

While De Wilde (2018) addresses the building performance analysis through four primary methods: physical testing and measurement, calculation and simulation, stakeholder assessment, and expert judgment, Borgstein *et al.* (2016) divide energy performance evaluation methodologies into simulation, engineering calculations, statistical methods, and machine learning for non-domestic buildings. Besides, Mhalas *et al.*, (2013) explain that in the residential sector, there are “top-down” and “bottom-up” approaches to model the energy consumption and these approaches vary regarding their complexity, the data input required by the user, level of detail, their geographical coverage and the time periods covered.

CIBSE (2013) lists five principles to evaluate energy use:

- using dynamic simulation models to assess energy use for heating, cooling, fans and pumps, to use more realistic inputs and profiles
- conversation between the designers and all the future occupants about energy management systems and potential operating hours
- to include all the energy uses in the building, rather instead of the central plant and equipment
- a comparison against existing energy use benchmarks to assure that the results are into different possible outcomes
- to present the results as a range with a low and high energy use scenario; this helps to indicate that there is uncertainty about energy use prediction.

As one of the suggested principles for energy use evaluation, benchmarking is referred to by different terminologies such as energy labeling, energy certification, energy rating, asset rating, operational rating, and operations and maintenance rating (Arjunan *et al.*, 2020). Energy benchmarking is explained as a macroscopic level of performance assessment and it uses metrics to measure buildings' performance with relative to other buildings or its previous performance (Li *et al.*, 2014). Originally, the word benchmark was used particularly in topography to precisely describe a reference point in the geologic analysis. The term started to be used for the comparison of energy use in buildings of similar characteristics in the 1990s (Pe´rez-Lombard *et al.*, 2009). Li *et al.* (2014) categorized existing energy benchmarking methods into white box method, gray box method, black box method.

- White box method requires a large amount of design documentations (e.g., detailed energy simulation method)
- Grey box method integrates both data fitting techniques and physical knowledge of the system to obtain a useful energy model. (e.g., degree day method)

- Black box method demands training data and a pre-selected statistical model and (e.g., artificial neural network method)

As one of the primary methods of building performance analysis, calculation and simulation are generally encountered in design projects, allowing analysis without physical access to a building (De Wilde, 2018) and calculating energy use at the design stage can provide the following outcomes (CIBSE, 2013):

- Future occupants can have clearer evidence of the potential range of carbon dioxide emissions, energy use, and costs of running the building
- Designers and future occupants can recognize how and where energy is expected to be used in the building
- Designers can understand which actions have the biggest impact on energy use
- Future occupants do not have imaginary expectations about the building's performance

Although the energy use of a building would seem to be one of the more easily quantifiable measures of building performance in terms of measurement, many practical challenges originate when one tries to develop a consistent measure of energy performance. These challenges vary from the choice of units to concerns with the use of indigenous and nonrenewable energy resources. Issues of the weighting to be given to several types of delivered energy are yet another difficulty for those who would aim to measure energy performance accurately (Baird *et al.*, 2018). Besides, Buratti *et al.* (2014) state that energy performance certificates are presented as a tool to estimate the energy performance of buildings; nevertheless, costly, and time-consuming controls are needed to prove the accuracy of the set and claimed data. Hong *et al.* (2014) claim that traditional energy auditing or analysis methods are not adequate to examine more closely the energy use of high-performance buildings. Therefore, the authors proposed a new holistic method, powered by measured

building performance data and analytics to guide energy retrofit of high performance buildings.

2.1.2 Energy Performance Simulation Tools

The 1960s were the beginning of research, development, and application of building energy simulation, and the discipline is continually developing and maturing (Hensen & Lamberts, 2019). “Expected” building performance is calculated using energy simulation programs (Cuerda *et al.*, 2020). Computer simulations’ results are helpful, to the design team, and the client, in assessing alternative design or retrofit options and making informed decisions about optimizing building operation (Demanuele, 2008); in testing performance compliance with various codes and regulations (Peters & Peters, 2018). Countless architects believe that energy modeling is the domain of engineers because of the generally non-integrated design process in the building industry, the reluctance of professionals to communicate across disciplines, the difficulty of using energy simulation tools, and the perceived cost (Peters & Peters, 2018).

Many tools explore and estimate building performance, especially energy performance (Rezaee *et al.*, 2015). Some of the well-known building performance modeling and analysis software tools include IES <VE>, DOE-2, TAS, Energy Plus, Ecotect, ESP-R (Birchall, 2011), and eQuest (Amasyali & El-Gohary, 2018). Computational tools vary from reduced-order models to higher-order dynamic simulation models. While the main goal of the former is to compare the normative performance of design alternatives accurately (e.g., MIT Design Advisor), the latter is more detailed and is used to numerically solve sophisticated sets of simultaneous differential-algebraic equations to calculate the transportation of air, heat, and moisture (e.g., Energy Plus) (Rezaee *et al.*, 2015). Figure 2.3 represents the selected history of building performance tools from 1960 onwards.

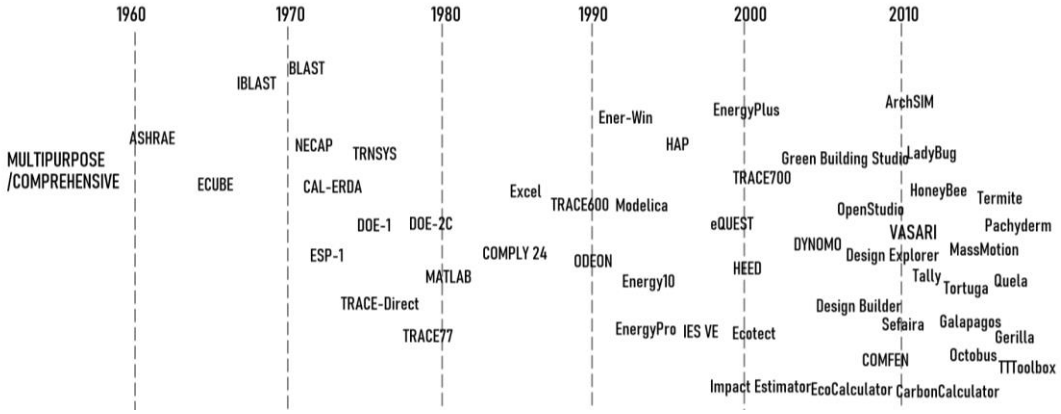


Figure 2.3 Selected History of Building Performance Tools, 1960-present (modified from Peters & Peters, 2018)

Various input parameters are defined in the software to perform the building performance simulation (Birchall, 2011), and the final output is sensitive to the input values (Nižetić & Papadopoulos, 2017). These inputs are composed of constant inputs such as building size and orientation, building location, construction materials, and variable inputs referred to as "schedules" and "profiles" which may change during the day, such as lighting, occupancy, and small power use (Birchall, 2011). Information about occupancy patterns, HVAC, construction, and boundary conditions such as climate information is involved within the building model. Consequently, many assumptions are made when the model is to present a sophisticated energy flow path and interactions (Cuerda *et al.*, 2020).

Van Heijst (2017) divided the modeling methods into three:

- Compliance modeling: It is beneficial in assessing the energy efficiency of buildings under standard operating conditions (operating schedules, set-points, occupant density, etc.) that are determined in national calculation methodologies. In this way, these buildings can be checked whether they meet the minimum

performance requirements or not. Anyhow, comparing these calculations with measured energy use may cause a wrong view of the performance gap.

- Performance modeling: Since the actual operating conditions are simulated with respect to the building context and the actual operating conditions of the building, the performance gap could be reduced with this method.
- Calibration techniques: Using this method, the impact of underlying reasons for the discrepancy and design assumptions on the energy performance can be quantified.

Moreover, in addition to model calibration, sensitivity analysis and uncertainty analysis are some of the approaches taken to get more reliable results and to adapt the presentation of the results to the needs of specific users of the simulation tools (International Energy Agency, 2016). Van Heijst (2017) mentioned two distinct types of uncertainty: Design parameter uncertainty and scenario uncertainty. While the nominal distribution around a thermal resistance value, due to on-site workmanship or wrong specifications of the manufacturer is an example of the former, the boundary conditions and scenario of use, such as weather conditions and occupancy are examples of the latter. The author state that scenario uncertainty has a wider distribution in comparison with the design parameter uncertainty.

The complexity and power of building performance modeling and simulation come from the use of several underlying theories from various disciplines (Hensen & Lamberts, 2019). Although energy simulation provides a wide variety of outputs and detailed information, it may require experienced users, a vast number of inputs, and quite a lot of time to collect and input the necessary data, all of which can make the procedure expensive (Pérez-Lombard *et al.*, 2009). Hensen & Lamberts (2019) state that (1) Simulation does not immediately produce answers and (2) usually it is difficult to guarantee the quality of simulation results. Moreover, most building energy modeling tools are inadequate at the design stage, and they are not user-friendly (Imam *et al.*, 2017).

Although energy simulation is becoming a standard practice, there are still big concerns about the accuracy of predictions (Demanuele, 2008). Turner& Frankel (2008) state that the inherent complexity of buildings, additionally the change in operational factors such as occupancy and building schedule, internal plug loads, and weather limit the accuracy of modeling. Moreover, it is challenging to repeat the real context consisting of different factors (Güçyeter & Günaydın, 2012), due to the unavailability of some detailed data to users during simulation (Amasyali & El-Gohary, 2018; Lee *et al.*, 2007; Rezaee *et al.*, 2015), and high level of uncertainty (Rezaee *et al.*, 2015; Garwood, 2019; Sun *et al.*, 2019). Therefore, it is possible to claim that evaluation tools predict actual state with a deviation.

However, Shi *et al.* (2019) state that blaming the performance gap entirely on the modeling programs is not fair. A more reasonable and useful view is to agree on the constraints and keep developing these programs to better capture the complicated essences of building performance. For instance, lack of training for practitioners is also one of the key reasons for the performance gap (Garwood, 2019). The personal performance of thermal modelers might deviate over time. Additionally, a comparison of their results with the performance of the finished building is rare, and the feedback is limited (Imam *et al.*, 2017). Besides, occupants' energy behaviors are considered fully neither within energy certification methods nor in energy simulation software (Delzendeh *et al.*, 2017). A deviation from design conditions in comparison to real use conditions can be seen especially for factors that are controlled by tenants' preference, e.g., indoor set-point condition, or by to a certain level random in nature, e.g., the use patterns of lighting and equipment and the occupancy pattern Lee *et al.*, 2007).

Attia *et al.* (2013) interviewed twenty-eight experts to report potential challenges and opportunities to integrate optimization tools in net zero energy buildings design. According to the results, optimization tools do not reflect the needs of the industry (user-friendly interfaces, ease of use, fast turn-around, high return on invested time,

shallow learning curve). The experts desire improved visualization of results, faster computing (e.g., cloud computing and real-time feedback of results), improved methodologies (e.g., automated error-checking, validation, and uncertainty analysis), and standardized costs and performance databases in the future. Imam *et al.* (2017) asked 108 building modelers to comment on the importance of using accurate values for twenty-one common modeling input variables to estimate annual energy demand when using dynamic simulation. The results indicated little correlation between which variables were thought to be important by the modelers and which assured to be objectively important. Moreover, it was seen that having many years of experience in modeling did not improve the accuracy of people's predictions. Attia *et al.* (2012) interviewed architects and engineers to outline the major criteria for building performance simulation tool selection. The results indicated that the most prominent features of a simulation tool are different for engineers and architects. While engineers think that accuracy and ability to simulate detailed and complex components is the most important feature among four other features; architects think that integration of an intelligent design knowledge base to support decision-making is the most important feature.

2.1.3 Energy Performance Gap in Buildings

The problem of the energy performance gap is hardly new (Gram-Hanssen & Georg, 2018). The term "building energy performance gap" appeared in the mid-1990s (Niu *et al.*, 2016) and it has been largely studied for well over two decades (Shi *et al.*, 2019). It is universally accepted that 'gap' means 'difference' or 'discrepancy,' and 'the performance gap' is usually referred to as the difference between the performance predicted in the design stage and that measured in the occupation stage (Shi *et al.*, 2019).

Using the Web of Science database, the number of publications published between 2011 and 2021 that includes the “energy performance gap” keyword was searched in the Topic (including title, abstract, author keywords, and keywords plus) in the fields of Architecture and Civil Engineering. Firstly, 933 publications were found. After reading the title and abstract of the publications, irrelevant studies were excluded, and 157 publications were selected that address the energy performance gap phenomenon. While some of these studies focus directly on the energy performance gap, others state that the focus of the article will also benefit performance gap studies. As Figure 2.4 represents, the number of publications increased sharply after 2014, with an average of 15.7 publications per year.

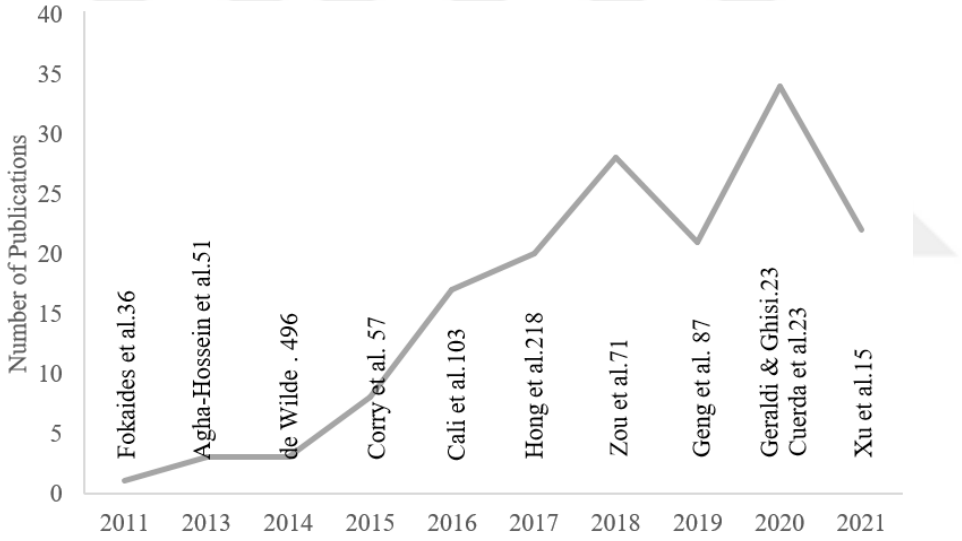


Figure 2.4 Number of Publications and the Highest Cited Authors per Year

This chart also includes the highest cited authors per year. Three of these studies provided a review of the literature on the energy performance of buildings (Geraldi & Ghisi, 2020); post-occupancy performance of green buildings (Geng *et al.*, 2019); performance gap (De Wilde, 2014). Moreover, life cycle building information modeling (Xu *et al.*, 2021); methods to reduce the gap by using simulation tools

(Cuerda *et al.*, 2020); how occupant behavior influences the energy performance of buildings (Hong *et al.*, 2017); employee satisfaction (Agha-Hossein *et al.*, 2013) was the focus of some others. While Corry *et al.* (2015) developed a performance assessment ontology and software tool, Zou *et al.* (2018) studied the root causes of the performance gap. Additionally, Fokaides *et al.* (2011) compared the measured and calculated performance of ten dwellings and Calì *et al.* (2016) provided the results from the monitoring activity of three demonstration buildings by analyzing the reasons for the gap. Figure 2.5 represents the main focus of the most cited articles between 2011 and 2021.

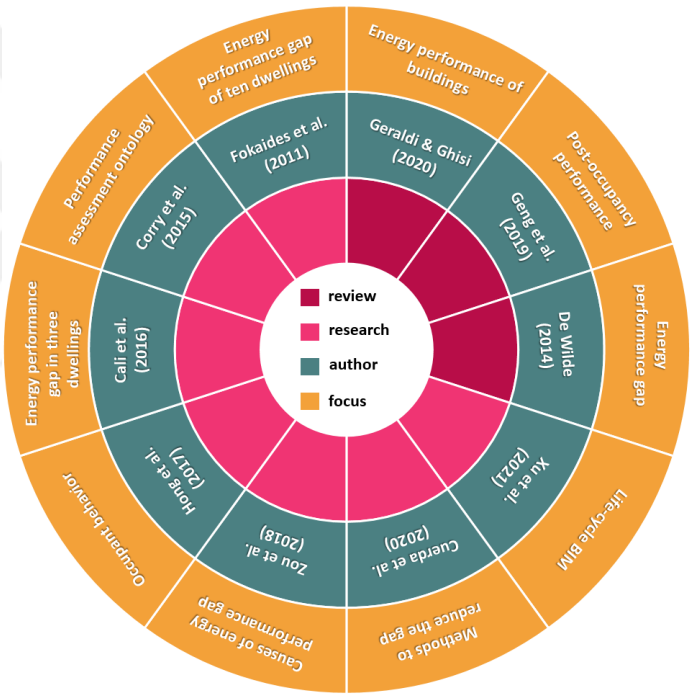


Figure 2.5 Main Focus of the Most Cited Publications

Basically, many publications in the literature confirm that there is a difference between a building's planned and actual performance (Galvin, 2014; Fedoruk *et al.*, 2015; Figueiredo *et al.*, 2018; Liang *et al.*, 2019; Cuerda *et al.*, 2020). A small number of buildings in the UK perform in line with their design intent (Mashford &

Reason, 2014). The authors exemplify a BREEAM Excellent building which is intended to be 'ultra-low energy,' demonstrated Display Energy Certificate rating of G. Performance gap is widespread and complex in origin and spreading across the industry. The Innovative UK and the Zero Carbon Hub study indicated that the performance gap is on average 2.6 times worse than the design predictions (Dollard, 2018). Sunnika-Blank & Galvin (2012) studied existing data on 3400 German homes and compared energy performance ratings with actual measured consumption. The results showed that while occupants consume on average 30% less heating energy than the calculated rating, occupants consume more than the rating for low-energy dwellings. The analysis indicated four aspects of the data that can be identified in similar ways in Belgium, France, the Netherlands, and the UK. Galvin (2014) focused on domestic heating and analyzed three retrofitted apartment buildings in Germany. The results indicated that the energy performance gap ranged from 2% for one building, 56.8% for the second, and 272.9% for the third. Calì *et al.* (2016) presented that values up to 287% could be seen for the performance gap. Herrando *et al.* (2016) provided a detailed analysis of the Energy Performance Certification of 21 Faculty Buildings in Spain. They found an average deviation of 30% between actual vs. estimated energy consumption. Moreover, discrepancies for research buildings were significantly higher than academic buildings (45% vs. 23% on average). Fedoruk *et al.* (2015) analyzed an innovative university building that aimed for net-positive energy performance in Canada. The 'performance goal' was to target an energy-use intensity (EUI) of 70 kWh/m^2/year, however, the result was almost double the target EUI, 130 kWh/m^2/year. Harrison & Jiang (2018) state that the use of solar energy is a popular path to decarbonization in the UK. The authors focused on the performance gap of photovoltaic systems. The results through a case study showed an 8.6% higher measured energy output from the installed PV system.

Although the performance gap is usually associated with increased energy consumption, it can also mean reduced consumption (Mojic *et al.*, 2019; De Wilde,

2018). However, actual measured energy use is higher than predicted in many cases. While the performance might be better than expected in some areas, it might be worse in others. Moreover, the performance gap can not only be identified in energy use, but also other areas such as daylighting, acoustics, and indoor air quality (De Wilde, 2018).

Expectedly, attempts were made to reduce the gap using diverse ways. Conducting pre-occupancy (Niu *et al.*, 2016) and post-occupancy evaluations (Menezes *et al.*, 2012), or dealing with uncertainty in energy simulations (Li *et al.*, 2019b) are some of those. In more detail, Mhalas *et al.* (2013) proposed a methodology that integrates a GIS platform, aerial and terrestrial imagery, digital maps, and information from various national statistics and databases. The results show that the predicted energy performance is within a maximum range of ±8% of the actual energy performance provided by the social housing provider. Majcen *et al.* (2015) used descriptive statistics and multiple regression to recommend improvements to the theoretical heating consumption provided in the European Performance of Buildings Directive. Niu *et al.* (2016) developed a virtual reality integrated method and used pre-occupancy evaluation to close the performance gap. Marshall *et al.* (2017) modeled a typical pre-1920s UK house in Design builder and performed electric co-heating tests to calculate the building's heat transfer coefficient. Moreover, in-situ air permeability and U-value tests were made. The modified model resulted in a decrease in the performance gap to 2.4%. Liu *et al.* (2019) developed a system dynamics model to simulate the interactions between facility managers and occupants and to discover targeted operations management tactics to minimize energy performance gaps. Sun *et al.* (2019) conducted a longitudinal questionnaire survey for one year in ten offices located in China. Window-opening behavior profiles which are generated through the results were linked to simulation software to further study the discrepancies between simulation and monitored results. Carpino *et al.* (2020) compared the impact of three different occupancy profiles in

performance prediction: i) the ‘Compliance profile’, specified based on regulations, ii) the ‘Standard profile’, based on average data gathered from surveys, and iii) the ‘Actual profile’, customized on measured data from the actual building case.

2.1.3.1 Disadvantages of Energy Performance Gap

The performance gap’s existence creates a gap between reality and the policies legislated by governments for the reduction of energy use and greenhouse gas emissions (Imam *et al.*, 2017). The performance gap has affected the implementation of different policy programs (Goulden *et al.*, 2020), and made it difficult for policymakers to successfully make plans for future strategies (Li *et al.*, 2019a). There is concern that the performance gap has the potential to weaken the zero carbon housing policy and present significant commercial risk for the wider industrial sector (Gupta *et al.*, 2013). Shrubsole *et al.* (2019) state that the performance gap risks weaken design, technology, sustainability, economic, health, and well-being benefits. Significant increases in energy cost and environmental impact can be seen due to this difference (Birchall, 2011). Moreover, in case the measured energy output is higher than the expectations in an installed PV system, it could result in the unreasonable design of the PV system and unsuitable use of capital investment (Harrison & Jiang, 2018).

A better understanding of the performance gap, along with more accurate predictions, would enhance designer confidence, facilitate risk assessment, and benefit all stakeholders involved (Demanuele, 2008). The underperformance can be a burden not only on the energy performance contractors but also on the building owner or occupant. Where the realized savings do not match those that had been budgeted for, added costs, longer payback periods, and the potential for contractual disputes can be seen (Mohareb *et al.*, 2017).

Additionally, the energy performance gap introduces reputational dangers to the industry (Dollard, 2018) including the company or person behind this failure such as the designer or contractor of the project (Mohammadi, 2015). The performance gap is a potential barrier to widespread acceptance of energy retrofit (Loucari *et al.*, 2016). It can reduce the confidence of building owners and third-party investors and result in economic loss to the ESCOs (Alam *et al.*, 2017; Loucari *et al.*, 2016). Besides, it can also be another reason for general hesitance to adopt the energy performance certification system for buildings (Abela *et al.*, 2016). Buratti *et al.* (2014) state that the revision of certificates represents just another major problem since it is highly priced and requires considerable time. Moreover, it is extremely hard to determine if the declared data is correct without performing an inspection of the building (Buratti *et al.*, 2014).

2.1.3.2 Reasons of Energy Performance Gap

De Wilde (2014) grouped the causes for the mismatch between prediction and measurements in three main categories: causes belonging to the design stage, the construction stage, and the operational stage.

Considering the design stage of buildings, modeling methods and simulation programs is one of the most frequently cited reasons for the performance gap by many researchers. Peters & Peters (2018) state that the phase at which the modelling was undertaken have a key role to identify the accuracy of the output. It is well accepted that a model is just an abstract and simplified version of a complex reality (Zou *et al.*, 2018). Menezes *et al.* (2012) address shortcomings of current modeling programs and add that current modeling methods are not able to represent the realistic use and operation of buildings. Different calculation methods may lead to quite different results (Calì *et al.*, 2016) and there is a misusage of calculation methodologies and tools or input parameters (Kampelis *et al.*, 2017). Moreover,

some assign ‘the gap’ to building regulations and their connected energy calculations’ unrealistic behavioral assumptions (Gram-Hanssen & Georg, 2018).

If we dive deeper into the assumptions of energy calculations, researchers mention fundamental uncertainties inherent in simulation inputs. They are oversimplified and-or unrealistic (Menezes *et al.*, 2012). One of the main reasons for these discrepancies is that standard operating conditions are considered in the simulation software rather than the real ones (Herrando *et al.*, 2016). If enough data can be obtained from the operation stage, the accuracy of design would increase (Zou *et al.*, 2018). Moreover, variations are seen in real life with respect to occupancy schedule, actual weather conditions, plug loads, and internal heat gains (De Wilde, 2014). The most frequently files used in digital models, typical meteorological year files are not designed to present current or future realities (Peters &Peters, 2018). Since how buildings are designed to perform and how they are built-in reality could also be different (Sunikka-blank & Galvin, 2012), uncertainties in built quality and fabric performance (Menezes *et al.*, 2012) which also involve insulation, thermal leaks as well as to building services are present (Sunikka-blank & Galvin, 2012).

Furthermore, the design itself is often the first cause of performance discrepancies within the design stage (De Wilde, 2014). Several authors point out that design complexity (Jradi *et al.*, 2018; Bros-Williamson *et al.*, 2016; Alam *et al.*, 2017; Alencastro *et al.*, 2018) or lack of attention to simplicity, buildability, sequencing of the construction process, or of appropriate detail can be the reason of performance gap (Zou *et al.*, 2018; De Wilde, 2014).

Although knowledge and experience had a significant impact on building energy consumption, they are largely out of the attention of existing research. (Zou *et al.*, 2018). “Performance Gap” is a “Knowledge Gap” between off-site designers and on-site constructors and within both the policymakers and administrators (Littlewood & Smallwood, 2015). Lack of knowledge and experience present in the design stage

(Zou *et al.*, 2018; Zou *et al*, 2019), tool modelers knowledge (De Wilde, 2014; Jradi *et al.*, 2018; Li *et al.*, 2019b), and experience (Demanuele, 2008) might impose uncertainty for the building energy analysis. Imam *et al.* (2017) call universities to increase their efforts about building physics education to reduce a part of the performance gap.

More effective communication between key stakeholders accompanying the design, construction, and operation of buildings is considered essential to achieve successful performance (Fedoruk *et al*, 2015). Miscommunication about performance targets for the building between design team and client (De Wilde, 2014; Zou *et al.*, 2018; Jradi *et al.*, 2018; Alencastro *et al.*, 2018) might be a reason for the discrepancy. Moreover, owners' concepts, decisions, and behavior would have an impact on the design, for instance wrong ideas about building energy-saving and changing an order (Zou *et al.*, 2018).

At the construction and commissioning stage, limited experience, and knowledge of the designer (Zou *et al.*, 2018), inexperienced installers (Mohareb *et al.*, 2017), workers and project managers (Qi *et al.*, 2019), inexperienced contractors, subcontractors (Alam *et al.*, 2017) are considered to lead to quality failures (Qi *et al.*, 2019) and create risks potentially affecting energy performance of buildings (Alam *et al.*, 2017).

Construction quality has been identified as a significant contributor between all other factors to the energy performance gap. The properties of building fabric may differ from the initial values set in the simulation due to the changes, cost-saving exercises (Demanuele, 2008), poor workmanship (Delzendeh *et al.*, 2017; Zou *et al.*, 2018; Alencastro *et al.*, 2018), and improper construction technique (Zou *et al.*, 2018). The quality of the building is generally not in line with the specification, with inadequate attention to both airtightness and insulation (De Wilde, 2014; Zou *et al.*, 2018).

Gaps in the insulation and thermal bridges are usual (Menezes *et al.*, 2012; Herrando *et al.*, 2016) and building envelope components (*e.g.*, glazing *etc*.) are installed improperly (Kampelis *et al.*, 2017; Jradi *et al.*, 2018). Besides, a contractor with limited experience and knowledge usually decides details that are left unspecified by designers (Zou *et al.*, 2018). Additionally, the performance features of the components of the building can be endangered due to the changing of orders by clients or material specifications by value engineering (De Wilde, 2014; Menezes *et al.*, 2012; Alencastro *et al.*, 2018, Jradi *et al.*, 2018) which can be defined as a systematic method used by a project team to provide cost optimization while considering functional performance and project quality.

Many researchers state that the equipment which is not performing as well as specified by the manufacturer can cause a performance gap (De Wilde, 2014; Calì *et al.*, 2016; Fedoruk *et al.*, 2015; Bros-Williamson *et al.*, 2016; Zou *et al.*, 2019; Van Heijst, 2017; Salimi & Hammad, 2019; Heo, 2011). Zou *et al.* (2019) suggest using products from trusted manufacturers who could offer sufficient and accurate technical details and specifications of their products.

Another critical contributor is occupant behavior. It is challenging to predict occupants' behavior (Salehi *et al.*, 2015) and it is usually different from the assumptions (De Wilde, 2014). Building occupants do not every time have direct control over building services such as heating, still, in highly automated buildings, occupants can influence the energy consumption by affecting the indoor conditions (*e.g.*, window opening behavior *etc*.) (Menezes *et al.*, 2012). Moreover, diverse heating patterns of people might be a part of the cause of this discrepancy (Sunikka-Blank& Galvin, 2012).

Occupants’ knowledge and skills are necessary for the operational stage (Zou *et al.*, 2018). The field needs professional expertise and new ways to integrate user involvement in building performance to completely succeed (Janda, 2011).

Baborska-Narozny& Stevenson (2015) carried building performance evaluation of 40 households in two UK developments using a case study approach. Lack of previous experience with continuous mechanical ventilation and airtight dwellings were the similarities between case buildings' gap factors.

The quality of commissioning and its coordination is vital for how the building will run once occupied both in terms of energy efficiency and the comfort wellbeing and productivity of users (Alam *et al.*, 2017). Problems with commissioning (Menezes *et al.*, 2012), the inadequacy of commissioning (Jradi *et al.*, 2018), or lack of continuity of it (Jradi *et al.*, 2018) might cause the gap. Moreover, De Wilde (2014); Herrando *et al.* (2016); Jradi *et al.* (2018) address the problems with soft landings which is a method to provide a smooth transition from construction to occupation and an optimized operational performance. Besides, good management and controls can bring an efficient operation of the building services whereas improper strategies can give rise to unwanted waste of energy (Menezes *et al.*, 2012). Several researchers mention the role of management when considering the performance gap (Menezes *et al.*, 2012; Kampelis *et al.*, 2017; Zou *et al.*, 2018). In more detail, suboptimum performance in systems' integration, misuse of energy systems' operation, or lack of maintenance and service (Kampelis *et al.*, 2017) the energy managers' management behavior (Zou *et al.*, 2018) might also be some of the reasons for the gap.

Some think that the most important reason for the energy performance gap is a necessity for application of all energy use in a building like plug loads, ICT systems, special functions etc. Others notice problems with software, software users, building commissioning, maintenance and recording. Yet others support that a better understanding and modelling of the energy-related occupant behaviour in buildings is a key to improvement (De Wilde, 2018).

- **Analyzing the reasons of the Gap**

Table 2.1 below presents the reasons for the energy performance gap in three building life cycle stages based on a detailed analysis of 20 journal articles. Seven factors are presented in each project life cycle stage including the design phase, the construction phase, and the use or operational phase. The factors are listed considering the number of times they occur in the articles.

Analyzing the factors further shows that the factors can be subcategorized into five groups *i.e.*, (i) human-related factors, (ii) material or equipment-related factors, (iii) tools and data-related factors, (iv) design-related factors, and (v) maintenance-related factors. Such subcategorization can generate a deeper understanding of the essence of the reasons for the gap and help find simpler mitigation techniques for the discrepancy. It can be observed that most of the reasons for the energy performance gap are human-related. Subcategories are represented in the table with distinct colors. All the factors in the subcategories are listed below:

Human-related factors:

- lack of knowledge and experience of the project stakeholders such as design, construction team, and users (D4, C4, U7),
- miscommunication or lack of communication among clients, design team, and builders (D3, C2),
- changing requests or orders from clients (C3),
- low-quality construction due to low-quality workmanship and poor site management (C1),
- Uncertainties related to building occupants, their behavior, and preferences (U1).

Material- and equipment-related factors:

- poor quality of materials and equipment (C5),
- unavailability of the specific products (C6),
- Deterioration or aging of materials (C7).

Tool- and data-related factors:

- limitations and shortcomings present in modeling and software (D1),
- unrealistic simulation inputs (D2),
- Poor quality of the measured data due to faulty sensors and meters (U4).

Design related factors:

- poorly designed buildings or systems (D5),
- the complexity of the design (D6),
- incomplete design (D7)

Maintenance-related factors:

- lack of commissioning (U2),
- lack of monitoring (U5)
- poor building management (U3)

Table 2.1 Reasons of Energy Performance Gap in Buildings

DESIGN PHASE	CONSTRUCTION PHASE	USE OR OPERATION PHASE
D1 LIMITATIONS / SHORTCOMINGS RELATED TO MODELING AND SOFTWARE (3, 8, 9, 11, 12, 13, 14, 16, 17, 20)	**C1** POOR QUALITY OF WORKMANSHIP / INSUFFICIENT BUILDING QUALITY (2, 3, 4, 5, 7, 9, 12, 13, 14, 16, 18, 19, 20)	**U1** UNCERTAINTY ON OCCUPANTS BEHAVIOR AND ACTIVITIES (1, 3, 5, 6, 7, 8, 9, 11, 12, 13, 15, 16, 17, 18, 19, 20)
D2 UNREALISTIC / OVERSIMPLIFIED SIMULATION INPUTS: WEATHER, OCCUPANCY, ETC. (5, 6, 9, 12, 13, 14, 16, 18, 20)	**C2** MISCOMMUNICATION / LACK OF COMMUNICATION BETWEEN PROJECT PARTIES (2, 3, 5, 10)	**U2** LACK OF CONTINUOUS COMMISSIONING / POOR COMMISSIONING / PROBLEMS WITH SOFT LANDING (2, 7, 9, 12, 13, 16)
D3 MISCOMMUNICATION / LACK OF COMMUNICATION BETWEEN PROJECT PARTIES (3, 5, 9, 10, 13, 20)	**C3** CHANGING REQUESTS / VALUE ENGINEERING (2, 5, 9, 13,16)	**U3** UNSUITABLE / POOR BUILDING MANAGEMENT (7, 12, 14, 16, 19, 20)
D4 LACK OF KNOWLEDGE, SKILL OR EXPERIENCE OF PROJECT PARTIES (2, 9, 13, 15, 19, 20)	**C5** POOR QUALITY OF MATERIALS, EQUIPMENT OR TECHNOLOGIES (6, 8, 9, 10, 19, 20)	**U4** POOR QUALITY OF THE MEASURED DATA DUE TO FAULTY SENSORS AND METERS (7 , 9, 13,19)
D5 DESIGN DEFICIENCIES / POORLY DESIGNED BUILDING AND BUILDING SYSTEMS (3, 9, 13, 15, 17)	**C4** LACK OF KNOWLEDGE, SKILL OR EXPERIENCE OF PROJECT PARTIES (2, 5, 20)	**U5** LACK OF MONITORING / CONTINUITY OF MONITORING (9, 13,16, 20)
D6 THE COMPLEXITY OF THE DESIGN / LACK OF SIMPLICITY I LACK OF ATTENTION TO BUILDABILITY (3, 9, 13, 20)	**C6** LACK OF GREEN MATERIAL IN THE MARKET / NON-AVAILABILITY OF THE SPECIFIC PRODUCT (3, 19)	**U6** THE COMPLEXITY OF CONTROL STRATEGIES (5, 12)
D7 INCOMPLETE DESIGN / LACK OF PROPER DETAILS (2, 3)	**C7** DETERIOTION / AGING / LIFETIME DEFECTS OF MATERİALS (2, 4)	**U7** LACK OF KNOWLEDGE OR EXPERIENCE OF PROJECT PARTIES (13, 20)

HUMAN | MATERIAL & EQUIPMENT | TOOLS & DATA | DESIGN | MAINTENANCE

References: 1. Ahn et al.(2017), 2. Alam, et al. (2017), 3. Ali et al.(2020), 4.Aïssani et al.(2016), 5.Alencastro et al. (2018), 6. Ascione et al.(2016), 7. Birchall (2011), 8.Cali et al.(2016), 9.De Wilde, P. (2014), 10.Fedoruk et al.(2015), 11.Gram-Hanssen & Georg (2018), 12.Herrando et al.(2016), 13.Jradi et al.(2018), 14.Kampelis et al.(2017), 15.Menezes et al.(2012), 16.Mojic et al.(2019), 17.Salehi et al.(2015), 18.Sunnika-Blank & Galvin (2012), 19.Zou et al.(2019), 20.Zou et al.(2018)

2.1.3.3 How to Decrease the Energy Performance Gap

To answer the energy performance gap, the Zero Carbon Hub identified three common issues throughout all stages: Unclear distribution of responsibility, lack of understanding, knowledge, and skills, and ineffective communication of information (Dollard, 2018). Shi *et al.* (2019) group effective energy performance gap reduction measures into three categories, technical measures, non-technical (managerial) measures, and hybrid measures.

According to Demanuele (2008), the following points are necessary to reduce the performance gap:

- Better and more reliable data: (weather data, building envelope components, scheduled occupant loads)
- Probabilistic and statistical approach to simulation (for the prediction of variables such as use of windows, blinds, and heating systems)
- Sensitivity analysis/ Uncertainty analysis (to assess the relationship between input parameters and output values, how well the model represents reality)

According to Hong *et al.* (2015), to overcome the performance gap, a new method is necessary in the design phase, for example, using the operational rating which comes from the estimation method that considers the actual energy performances of similar buildings with the data mining or machine learning approach. Secondly, the stochastic approach should be used to consider the uncertainty to estimate the operating rating in the design phase. De Wilde (2018) suggests that some of the prediction and analysis tools used in the sector should be reexamined thoroughly to bridge the gap. Moreover, how well the default parameters (*e.g.*, daily occupancy, equipment use, and lighting use patterns) can represent the typical situation in buildings is necessary to avoid wrong performance arrangements (Lee *et al.*, 2007). The predictions can be improved by considering the effect of occupants with

different energy usage rates. The simulation of behavior through exhaustive modeling can be a way to do that (Ali *et al.*, 2020). For designers, there is a need to evaluate user experience and that can be considered while designing and conceptualization building performance (Grover, 2019). Additionally, practitioners' focus should be on the sequencing of work, the distribution of design responsibility, and the location of expertise to reduce the 'performance gap' (Boyd & Schweber, 2018).

Suggestions of Dollard (2018) to designers are as follows:

- designing a simple and compact building envelope
- increasing simplicity of detailing
- specifying simple building services that are easy to operate and perform as intended
- reducing thermal bridging
- ensuring drawings and specifications are clear
- considering construction sequencing, buildability, and as-built performance

Greater understanding and attention to detail regarding thermal bridges during the design and construction phases is necessary to reduce heat losses through the building fabric (Gupta *et al.*, 2013). Poor energy performance is a result of poor design and construction quality. Therefore good thermal performance necessitates high levels of quality assurance performance. Besides, the products' performance goes side by side with the installation quality. Therefore, increasing the frequency and quality of inspection and testing is an answer to improve quality (Dollard, 2018). A gap between designed and actual air tightness might be related to construction quality or the built form. Better air tightness can be achieved because of the higher quality of detailing at key junctions, skirtings, and service penetrations. Air leakages can be found in several places such as penetrations in ground floor bathrooms, solar

and MVHR cupboards, electrical cupboards, below-skirting boards and around external doors on the ground floor, and roof gables (Gupta *et al.*, 2013). Thermal performance defects (*e.g*., discontinuity of insulation layer or ruptures in the vapor barrier) need specialist tests (*e.g*., air tightness tests or thermographic imaging) during the handover and post-handover stage. Therefore, these hidden defects (e.g., missing cavity closer) must be quickly spotted during the construction stage before new works start or are covered up (Alencastro *et al.*, 2018).

Dollard (2018) proposes that contractors can improve quality assurance on-site with additional building control visits. With better communication between the designer and the site team, the role and responsibility of the designer on site can be increased. One of the project teams can be in charge of as-built energy performance and check the airtightness, the continuity of insulation, wind tightness, correct building services, and commissioning. Moreover, products should be installed in line with specifications and SAP. They should make sure that the building services are commissioned well, particularly ventilation, and provide information and training to the occupants on use and maintenance. Lastly, they should make sure that subcontractors have the right training and experience for their job (Dollard, 2018). Besides, to solve problems of accessibility and efficiency loss, a detailed and coordinated services layout plan presenting the location of controls will be effective. The handover review, initial group training sessions, and home user guides with comprehensive, simple, and visual guidance with diagrams for occupants are suggested to ensure effective control of MVHR, heating system, electric panel, etc. The performance gap can be further closed by resolving the issues of comfort and control efficiently (Gupta *et al.*, 2013). According to CIBSE (2013), when the building is complete, the designers/contractors should collaborate with the building occupants and operators to determine the differences between the design calculations and the real operational energy use. This collaboration should aim to fine-tune the building and optimize performance, instead of apportioning blame.

Different studies use monitoring data for the calibration of simulation models or the comparison of predicted vs. actual building energy performance to reduce the performance gap (Cuerda *et al.*, 2020). Moreover, Turner & Frankel (2008) suggest that Advanced Commissioning and Measurement and Verification could be reworked to contribute to better energy performance right away and provide more directly useful information to owners and operators related to LEED credits. Additionally, Gupta & Gregg (2016) suggest soft landings approach to close the performance gap. The soft landing approach differs from the conventional brief, design, build and occupy system by providing expectations management during design and construction and a better handover, followed by period aftercare and post occupancy evaluation.

A more detailed analysis of some of the best and worst performing buildings could identify ways to omit the poorer outcomes and communicate lessons from the best results (Turner & Frankel, 2008). Besides, policymakers can award good practices with incentives, increase building control for better energy performance and construction quality, make a construction guidance web resource for standard details and sequences of installation, and train design and site teams to reduce the energy performance gap (Dollard, 2018). To improve as-built performance, building regulations should be clarified with a special focus on compliance and site checks (Dollard, 2018). Moreover, regulations could take measures to provide more operational guidance to occupants and limit increases in energy demand. This may involve the pre-occupancy calculations (or pre-retrofit data) and post-occupancy energy measurements in the regulations (Hansen *et al.*, 2018). A part of the performance gap could be closed quickly by both industry and universities by rising their efforts with respect to building physics education (Imam *et al.*, 2017). For researchers, there is a need to develop methodological understanding to highlight the significance of humanistic factors in closing the building energy "performance gap" (Grover, 2019).

According to Khoury *et al.* (2017), particular attention is paid during the design and execution phase, and conducting a series of optimization measures in HVAC systems after the retrofit process can help buildings perform better. A significant reduction of the performance gap can be achieved through building optimization and responsible behavior (both occupant and energy operator). The impacts of occupant behavior on building energy performance have been largely oversimplified or ignored during the entire process of building operation (Hong *et al.*, 2017). For building operators and managers, there is a need to develop a regular feedback loop to explore green building feature awareness of occupants, and their response to it (Grover, 2019). For occupants of the building, there need to be more action-oriented reforms through their awareness and behaviors within the space that have implications on the building's performance (Grover, 2019). Continuous monitoring and obtaining feedback about how buildings perform is necessary at every stage of a project (Birchall, 2011).

Furthermore, there are many stakeholders apart from the occupants whose actions must be taken into consideration to "close the gap" which necessitates considering how buildings are designed, constructed, and operated by different professionals (Gram-Hanssen & Georg, 2018). A better understanding of the practices of the professionals involved in the process could potentially lead to more effective suggestions for supporting more sustainable home lives in their work (Wade *et al.,* 2018). Besides, greater integration, and construction project management operations, coordination between project partners enables building performance improvements. The authors also state that communication of partner ambitions, design intent, and responsibilities may increase project performance and reduce the performance gap. (Papachristos *et al.*, 2020). Response strategies to reduce the energy performance gap are given in a tabular form in Appendix A.

The organizational and technological complexity of construction projects causes enormous risks (Zou *et al.*, 2007) and project risks may generate positive or negative

consequences or both. Doylend (2015) used a commercial office building as a case study and found that the energy performance gaps were the result of whole-system lifecycle risk factors ranging from the design phase to the use and operational phase. Moreover, Topouzi *et al.* (2019) focused on deep renovation projects and Alam *et al.* (2017) studied retrofitted buildings and both confirmed that project risks in different life cycle stages cause poor outcomes and the energy performance gap. Besides, Alam *et al.* (2017) highlighted that considering risk paths will lessen the possibility of missing some crucial risks. Therefore, in the next section, the risk assessment concept is studied to be able to propose an energy performance gap framework with different project risks.

2.2 Risk Assessment in Buildings

This part focuses on the risk assessment concept firstly by giving various definitions of risk and then introducing methods applied in the construction sector and energy-efficient buildings.

2.2.1 Risk and its Definition

Risk is inevitable and present in every human situation (Ramanathan & Rathinakumar, 2017). International Electrotechnical Commission (IEC, 2019) explains that risk involves the impacts of any of the forms of uncertainty that may generate positive or negative consequences or both. While a positive risk is defined as an opportunity, a negative risk is defined as a threat. Therefore, something that is well defined as risky does not inevitably mean that it is always a bad thing, only that it is an uncertain thing (Cretu *et al.*, 2011). Additionally, the risk is generally defined in terms of risk sources, potential events, their consequences, and their likelihoods (IEC, 2019).

Besides, risk can have different definitions based on the domain of interest (Thomé *et al.*, 2016; Pruvost & Scherer, 2017). For instance, risks are uncertain situations or events that can have positive or negative consequences on the project goals in construction projects (Hatefi & Tamošaitienė, 2019).

Moreover, different risks can be structured by risk classification (Zou *et al.*, 2007) in several ways (Jayasudha & Vidivelli, 2016). However, the motivation for choosing a method to classify the risks must serve the purpose of the research (Zou *et al.*, 2007). For example, construction project risks can be analyzed from the perspectives of different stakeholders such as from the point of view of the client and from the point of view of the contractor (Ramanathan & Rathinakumar, 2017). According to Jayasudha & Vidivelli (2016), the typology of the risks appears to depend mostly on whether the project is local or international. Therefore, the authors classified risks into internal risks and external risks in construction projects. Table 2.2 represents the various ways suggested by different authors to classify risks in construction projects.

Table 2.2 Risks in Construction Projects

No	Author	Risks for Construction
1	Ramanathan & Rathinakumar (2017)	(1) Project, (2) Design, (3) Contracts, (4) Subcontractors, (5) Labors, (6) Material Suppliers, (7) Equipment, (8) Site Staff, (9) Execution, (10) System, (11) Owner, (12) Finance, (13) Environment
2	Jayasudha & Vidivelli (2016)	(1) External Risks, (2) Internal Risks
3	Plebankiewicz *et al.* (2016)	(1) Technological, (2) Construction Process Related, (3) Financial, (4) Political, (5) Environmental, (6) Legal

Jayasudha & Vidivelli (2016) state that risks play a significant role in decision-making and may impact the performance of a project. Risk elements related to construction projects influence the cost, time, and quality performance of the project (Akintoye & MacLeod, 1997). Moreover, many of the risks inherent in construction projects influence each other. For instance, the risk associated with the cost of a construction project may increase the project time (Hatefi & Tamošaitienė, 2019).

Although the organizational and technological complexity of construction projects causes enormous risks (Zou *et al.*, 2007), risk had not appeared in construction literature until the 1960s (Taroun, 2014). Cost overrun risk in public construction projects (Jung *et al.*, 2016); stakeholder-related risks in construction projects (Xia *et al.*, 2017); thermal discomfort risk in a renovated single-family house (Psomas *et al.*, 2017); occupational health and safety risk for sustainable projects (Karakhan & Gambatese, 2017); political risks in international construction projects (Chang *et al.*, 2018) are some of the examples of recently studied risks in construction projects in literature.

2.2.2 Risk Assessment in Construction Sector

As it is shown in the diagram used in the ISO 31000:2018 standard, risk assessment is at the core of the risk management process and it comprises three steps risk identification, risk analysis, and risk evaluation (Figure 2.6).

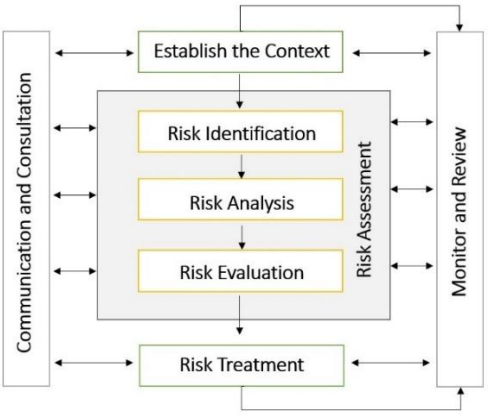


Figure 2.6 The Risk Management Process based on ISO 31000:2018

Risk identification is significantly important since the process of risk analysis and response management may only be performed on determining potential risks (Al-Bahar *et al.*, 1991; Jung *et al.*, 2016). The information should be appropriate, relevant, and up to date in risk identification (ISO 31000: 2018). According to Szymański (2017), the following steps are necessary to identify risks in construction projects:

- deciding what types of risks may influence the project
- identification of their distinctive parameters
- estimation of the probability of their occurrence in the project

Hanna *et al.* (2013) conducted an initial questionnaire, web-based survey, and phone interviews to identify construction risks with high potential for conflict and to support assessing and distributing these risks to suitable parties. Xia *et al.* (2017) studied stakeholder-related risks in construction projects. The authors reviewed the literature first to establish a list of general stakeholders encountered by construction projects. After that, experts used this list as a reference to identify a final list of stakeholders. Figure 2.7 indicates the methods used in the risk identification process, as well as the data needed, and the required effort to apply the method.

B.1 Techniques to elicit views	Data needs	Effort to apply
• Brainstorming	None	Low
• Delphi technique	None	Medium
• Nominal group technique	None	Medium
• Interviews	None	High
• Surveys	Low	High
B.2 Risk identification Techniques to identify risk		
• Check lists	High to develop,low to use	Low/Medium
• FMEA/FMECA	Depends on application	Low/High
• HAZOP	Medium	Medium/High
• Scenario analysis	Low/Medium	Low/Medium
• SWIFT	Medium	Low/Medium
B.3 Determining sources, causes and drivers of risk		
• Cindynic approach	Low	High
• Ishikawa method	Low	Low
• Root cause analysis	-	-

Figure 2.7 Techniques used in Risk Identification (modified from ISO 31000:2018).

Risk analysis is several analytical methods used to evaluate the level of hazardous occurrences (Nolan, 2019). Risk analysis incorporates basically: (1) knowledge about risk-related situations, events, etc., and (2) concepts, theories, frameworks, approaches, principles, and models to understand, assess, represent, communicate, and manage risk, overall and for particular applications (Hansson & Aven, 2014). Risk analysis methods can be qualitative, quantitative, or a combination of these, dependent on the conditions and intended use (ISO 31000: 2018). Hansson (2010) mentions about two contradictory concepts of risk: an objectivist and a subjectivist view. According to these concepts, although some researchers take risk as objectively given and defined by physical facts, others support the idea of independence of risk from physical facts. The author concludes that risk includes both subjective and objective components and suggests understanding how factual and valuational components are combined in risk statements. Gondia *et al.*

(2020) explain that conventional risk analysis methods, which are strongly relying on subjective data sources and/or do not consider variable interdependencies within the data, are used mostly by construction firms. The following four steps are essential in risk analysis (Nolan, 2019):

- Specify incident occurrences or scenarios
- Evaluate the frequency of the occurrences
- Identify the consequences of each occurrence
- Determine the consequences of each occurrence

A hybrid approach that uses the bootstrap technique and interval computations (Hashemi *et al.*, 2011); a soft-margin support vector machines scheme (Kifokeris & Xenidis, 2019); a decision tree, and Naïve Bayesian classification algorithms (Gondia *et al.*, 2020) are some of the methods applied in literature for analyzing the risks.

Risk analysis is followed by risk evaluation in Figure 2.6. Risk evaluation aims to support decisions and enables to define where additional action is needed (ISO 31000:2018). Risk evaluation involves some rankings relating to the importance which suggests measuring some aspects of the factors (Fung *et al.*, 2012). Monte Carlo Simulation, Fault Trees, Event Trees, Fuzzy set, Failure mode and Effective analysis, Game Theory, Grey Systems, and Multicriteria Verbal analysis, are several methods for risk evaluation (Fouladgar *et al.*, 2012). Additionally, meta-network analysis (Wang *et al.*, 2018); deep neural network (Xu, 2020) and multi-criteria decision making, and fuzzy logic (Fouladgar *et al.*, 2012) are some other techniques applied for risk evaluation in the literature.

2.2.3 Risk Studies in Energy Efficient Buildings

Several researchers presented the risks present in energy-efficient buildings such as Krechowicz (2020); Topouzi *et al.* (2019); Mills *et al.* (2006); Qin *et al.* (2016); Zhao *et al.* (2016); Yang & Zou (2014); Yang *et al.* (2016a). Krechowicz (2020) identified the risks in the design and construction of passive houses using Failure Mode and Effect Analysis. Topouzis *et al.* (2019) analyzed the risk of different deep retrofit approaches including the whole house; fabric first; room-by-room, step-by-step, and measure-by-measure building renovation by using evidence-based data and reports, and publications. The authors identified three distinct types of risk at ten retrofit stages: assessment, sequence, and communication which contribute to the performance gap. The authors state that new management systems, documents (e.g., Building passports), and new practices of the profession, with significance for education and training, are crucial to minimizing the performance gap in deep retrofits. Mills *et al.* (2006) determined ten categories for energy-efficient project risks including economic, technology, operational, contextual, and measurement and verification risk where each of these categories has both uncontrollable and controllable dimensions. The authors provided two risk analysis methods to analyze energy-efficient projects: Coefficient of variation and Monte Carlo Simulation. Qin *et al.* (2016) examined the risk factors of green building in China considering a life cycle perspective. The authors presented thirty-six key risk factors impacting the successful construction of green buildings. Moreover, clear differences between owners and contractors, owners and residents, engineers, and designers and contractors were seen in the ranking of risk importance. Zhao *et al.* (2016) analyzed the risks in green building projects in Singapore and developed a risk assessment model using the fuzzy synthetic evaluation approach. The risk factors were categorized into eleven groups. Yang & Zou (2014) studied stakeholder-associated risk in complex green building projects and developed a Social Network Analysis method to assess and analyze the risks and their interactions in projects. Yang *et*

al. (2016a) compared the differences between critical risks and stakeholders in green building projects in China and Australia. According to the results, while reputation risk is critical in both countries, the ethical risk of "assessment experience and fairness" is important in the Chinese context. Zou *et al.* (2016) categorized risk factors in building retrofit projects into eleven groups including financial risk, market risk, economic risk, social risk, legislative risk, industry risk, project design risk, technological risk, installation risk, operational risk, measurement, and verification risk.

Some authors proposed new methods to use in risk management including Heo *et al.* (2013); Sun *et al.* (2016); Zou *et al.* (2016); Zheng *et al.* (2019). Heo *et al.* (2013) proposed a method based on Bayesian calibration of building energy models for risk analysis. Sun *et al.* (2016) developed a probabilistic method of risk assessment for the calculation of energy use intensity and total utility cost in energy performance using the Monte Carlo method. Zou *et al.* (2016) reviewed the qualitative and quantitative methods used for identifying, analyzing, and responding to the risks linked to building energy/water retrofitting projects. Afterward, they proposed a risk management framework. Zheng *et al.* (2019) proposed a techno-economic-risk decision-making methodology based on Monte Carlo simulation for large-scale building energy efficiency retrofit projects. Nižetić & Papadopoulos (2017) state that tools to predict performance gaps can be useful in the phase where energy efficiency projects are rated and evaluated for possible investments. The authors proposed a conceptual novel method to determine gaps between designed and achieved energy savings beforehand.

Several authors focused on the risks in Energy Performance Contracting (EPC) industry. EPC promises energy savings or energy-cost savings directly related to the total cost of the improvement contracted by the Energy Service Companies (ESCOs) (Heo *et al.*, 2013) and it has been viewed as an alternative way to increase energy efficiency in existing buildings (Lee *et al.*, 2015). Depending on the contract clauses

of a performance contract, if the savings are not achieved during the contract period or are overestimated, ESCOs may have to compensate building owners for the deficit (Heo *et al.*, 2013). Lee *et al.* (2015) identified the main risks inherent in EPC projects. The results show that the main risks to ESCOs are uncertainty of baseline measurement, possible payment default of hosts after installation, and increase in installation costs. Wang *et al.* (2019) identified the risk factors in China's EPC industry. The authors developed a method that is based on the traditional Analytic Network Process and the Best Worst method for ranking the identified risk factors. Garbuzova-Schlifter & Madlener (2016) studied typical risk factors and reasons of risk associated with EPC projects implemented in three Russian sectors: (1) public; (2) housing and communal services; and (3) industrial.

Regarding energy-efficient buildings topic, numerous researchers emphasize overheating risk in buildings such as Liu *et al.* (2016a); Silvero *et al.* (2019); Fosas *et al.* (2018); Chvatal & Corvacho (2009); Gupta & Gregg (2018). Liu *et al.* (2016a) presented two alternative methods to develop current and future weather files. The authors suggested a method based on Weighted Cooling Degree Hours to use for assessing the severity and occurrence of overheating. Silvero *et al.* (2019) evaluated the original state and energy-efficient version of a historical residential building located in Paraguay using future climate scenarios. According to the results, the expected temperature rise will cause an increase in discomfort rates. However, energy refurbishment measures could mitigate it. Fosas *et al.* (2018) studied the impact of insulation on overheating considering a range of relevant variables. The authors created 576,000 building variants and implemented data mining techniques. The results indicated that insulation only answers for up to 5% of the overall overheating response. Chvatal & Corvacho (2009) focused on the impact of insulation on the overheating risk. The authors studied three types of buildings in three different climates with different combinations of building parameters. The results indicate that it is essential to control solar and internal gains to avoid

overheating in summer with added insulation. Gupta & Gregg (2018) made a simulation study of two case study dwellings in the UK to achieve a net zero energy target. Summertime overheating was tested in three different future climates. It was seen that summertime overheating is expected to be a risk for dwellings by the 2050s.

2.3 Machine Learning Concept

This part of the dissertation explains machine learning concepts as it is the suggested method to solve performance gap prediction problems. After giving a brief description of machine learning, and a summary of classification applications, it introduces four machine learning algorithms utilized in the research. Then, metrics used to evaluate the performance of these ML algorithms are given. The topic continues by introducing feature selection and an oversampling technique applied in the process.

ML is a branch of artificial intelligence. It uses example data or past experience to optimize a performance criterion (Alpaydın, 2010) and aims to predict future events and scenarios that are unfamiliar to the computer (Awad & Khanna, 2015). ML is a multidisciplinary field and uses results from various research fields such as artificial intelligence, computational complexity theory, Bayesian methods, control and information theory, psychology, philosophy, and neurobiology (Pantic, 2006). Since the central task of ML is to make assumptions from a sample, the theory of statistics is used in building mathematical models in ML (Alpaydın, 2010) and due to its multidisciplinary nature, techniques from a variety of disciplines, involving statistics, medicine, and information retrieval have influenced different developments of the performance measures (Japkowicz & Shah, 2011).

Various studies introduce ML algorithms as the most contemporary and best method for prediction (Olu-Ajayi *et al.*, 2022). However, Borgstein *et al.* (2016) underline

that ML is a powerful computational technique when sufficient data is available. Therefore, ML is dependent on the collection of this information. According to Awad & Khanna (2015), ML learning application is generally convenient when there are the following conditions: (a) very high accuracy is not needed; (b) large volumes of data involve undiscovered patterns to be synthesized; (c) the problem itself is not pretty well understood due to lack of knowledge of historical information as a reason to develop suitable algorithms, and (d) the problem requires to adapt to changing environmental conditions.

The authors divide the process of developing ML algorithms into seven steps as shown in Figure 2.8. The first step is to collect data that might be effective in solving the problem. The second step is preprocessing the data that is composed of formatting, cleaning by removing, replacing, or fixing missing data, and sampling to minimize repetition. The third step is transforming the data by feature scaling, decomposition, or aggregation. The fourth step is training the algorithm by selecting the training and testing datasets from the transformed data and training the algorithm on the training data set. In the fifth step, the algorithm is evaluated to determine quickly if any learnable structure can be identified in the data. In the sixth step, reinforcement learning is applied in case supervised learning may not be practical. In the last step, a validated model is applied to perform an actual prediction task.

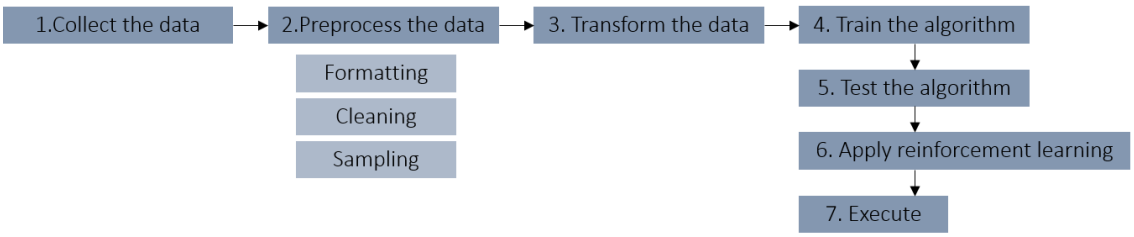


Figure 2.8 The Process of Developing ML Algorithms (modified from Awad & Khanna, 2015).

2.3.1 Classification in Machine Learning

One of the most widespread applications for data mining is classification (Bramer, 2020). Classification learning is at times called supervised since the method operates under supervision by being arranged with the actual outcome for each of the training examples (Witten *et al*., 2011). Regression problems are also categorized under supervised learning. However, when the output variable is a real value in a regression problem, it is a category in a classification problem (Brownlee, 2018b). Spam email detection, medical diagnosis, image classification, musical genre identification (Joshi, 2020); performance prediction, fraud detection, and manufacturing are some examples of the application of ML classification (Han *et al*., 2012).

Classification aims to learn a system adequate for the prediction of the unknown output class of a previously unseen instance with a good generalization ability (Galar *et al.*, 2011) and requires the construction of a classifier which is a function that gives a class label to instances defined by a set of attributes (Friedman *et al.*, 1997). Synthesizing a model, using a learning algorithm, and applying the model to categorize new data are the primary steps engaged in classification (Awad & Khanna, 2015).

When there are only two values as labels to predict future unseen examples, the problem is named a binary classification problem (Japkowicz & Shah, 2011). When there are more than two classes that instances can be assigned, this kind of classification is named "multiclass data classification" (Ayhan, 2019).

Binary classification problems are much easier than others. However, lots of problems are multiclass problems in the real-world (Li *et al.*, 2004). The authors divide multiclass classification methods into two types. In the first method, the binary classification algorithms which can be naturally extended to manage multiclass problems directly, are utilized. In the second method, multiclass problems

are decomposed into binary ones. The second method is a simple, but accurate way to reduce their complexity (García *et al.*, 2015). Moreover, Witten *et al*. (2011) highlight that the success of classification learning can be evaluated by trying out the concept description that is learned on an independent set of test data in which the real classifications are known but not presented to the machine.

2.3.2 Machine Learning in Energy Performance of Buildings

The interest in using big data methods like data mining, artificial intelligence, and machine learning is increasing due to the fast growth in the availability and quantity of data on buildings (Borgstein *et al.*, 2016). While modeling is the first path to predict building energy performance, machine learning, which develops a correlation between inputs and outputs without explicitly modeling the systems and physical processes, is another and usually neglected path for energy performance prediction (De Wilde, 2014).

Amasyali & El-Gohary (2018) presented a review of the research in data-driven building energy simulation prediction. According to the authors, two major constraints of data-driven energy consumption prediction are: Their performance may be reduced aside from their training range and since data-driven models are black-box models, although their prediction accuracy may be adequate, to provide a detailed understanding of the different parameters and its behavior concerning energy consumption may be limited.

Wang *et al.* (2021) state that although more than 128 ML algorithms have been used in building energy efficiency studies, no optimal algorithm has been found. Because no single ML algorithm has been shown to outperform other ML algorithms for all conditions when several different ML solutions have been crafted to solve the same problem. Since each algorithm has its own advantages and disadvantages for different cases, the algorithms should be selected according to the specific

application (Amasyali & El-Gohary, 2018). Besides, artificial neural networks (Borgstein *et al.*, 2016; Tardioli *et al.*, 2015), SVM (Tardioli *et al.*, 2015; Seyedzadeh *et al.*, 2020b), and decision tree (Tardioli *et al.*, 2015; Salimi & Hammad, 2019) are some of the most common algorithms used in the context of building energy performance. This study also indicated that random forest (Ye *et al.*, 2021; Anand *et al.*, 2021; Walker *et al.*, 2020; Amasyali & El-Gohary, 2018), and SVM (Ye *et al.*, 2021; Zhou and Chang, 2021; Walker *et al.*, 2020; Amasyali & El-Gohary, 2018; Paudel *et al.*, 2017) were commonly used in the context of energy performance at the building or neighborhood scale. However, Olu-Ajayi *et al.* (2022) add that the classification method has not been examined in many previous studies to predict energy performance.

Ye *et al.* (2021) compared the performance of different algorithms for city-block-level energy consumption classification. Zhou and Chang (2021) studied the automated classification of building structures in a complex urban environment for contributing to the progress of sustainable urban studies and evaluated twelve popular ML algorithms. Anand (2021) studied the predictability of electrical energy consumption with occupancy density information using ML. Arjunan *et al.* (2020) proposed a methodology that improves the existing Energy Star calculation method. Seyedzadeh *et al.* (2020b) introduced an energy performance prediction model through a machine learning model using a gradient-boosted regression tree for non-domestic buildings. Grillone *et al.* (2020) analyzed the measurement and verification process, and the prediction and recommendation process to improve building energy performance. Guyot *et al.* (2019) reviewed studies using neural networks for energy-related applications in the building sector. Moreover, a study on predicting the electricity demand of buildings individually and at the group level, showed that boosted-tree, random forest, and ANN models outperformed the other considered regression algorithms considering computational power, time, and error accuracy (Walker *et al.*, 2020). Amasyali & El-Gohary (2018) focused on predicting the

cooling energy consumption of a building by comparing ANN and other ML models. Paudel *et al.* (2017) introduced two kinds of prediction modeling approaches for heating consumption based on support vector machine.

2.3.3 Machine Learning Algorithms

Brownlee (2018b) divides machine learning algorithms into two: *Parametric* and *Non-parametric* machine learning algorithms. *Parametric algorithms* are usually also named linear machine learning algorithms. They simplify the function to a known form. Arjunan *et al.* (2020) state that linear algorithms are easy to interpret, simple and intuitive. Moreover, they have advantages such as being very fast to learn from data and requiring fewer data. Their limitations are being highly constrained to the specified form, more suitable for easier problems, and uncertain to match the underlying mapping function (Brownlee, 2018b). Besides, they are not adequate to model complex relationships that are essential to many real-world systems. Hence, they are usually found to be inaccurate in many energy management applications (Arjunan *et al.*, 2020). Logistic Regression, Linear Discriminant Analysis, and perceptron are some examples of this group (Brownlee, 2018b).

Nonparametric algorithms outperform linear models relating to accuracy in many application domains (Arjunan *et al.*, 2020). They are advantageous in terms of flexibility, power, and performance (Brownlee, 2018b). However, they lack interpretability (Arjunan *et al.*, 2020) and do not make strong assumptions about the form of the mapping function. Additionally, they have limitations such as requiring more data, being slower to train, and carrying more of a risk to overfit the training data. Decision Trees such as C4.5, Naïve Bayes, Neural Networks, and Support Vector Machines are some examples of this group (Brownlee, 2018b).

Any learning algorithm may be perfectly accurate for one dataset and very poor for another. A good learning algorithm means how well its inductive bias matches the properties of the data (Alpaydın, 2010). Pantic (2006) explains that it is beneficial to analyze several factors to decide which machine learning algorithm will perform better for the given problem and the given target function. These are the associations between the size of the hypotheses space, the completeness of it, the previous knowledge held by the learner, the number of training examples accessible, and the confidence we can have that a hypothesis that is compatible with the training data will accurately generalize to unseen examples.

2.3.3.1 Naïve Bayes Algorithm

Naïve Bayes Classifier uses probability theory to find the most likely of the possible classifications (Bramer, 2020). It is suitable for binary and multiclass classification problems (Brownlee, 2018b) and is described as one of the most effective classifiers in the way that it's predictive performance despite its assumptions about independence (Friedman *et al.*, 1997). Figure 2.9 represents a Naïve Bayesian classifier structure. This network captures that every attribute is independent of the other attributes, considering the state of the class variable (Friedman *et al.*, 1997).

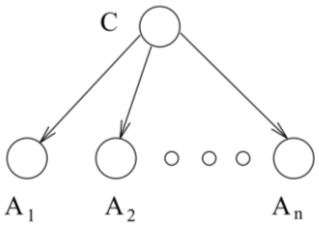


Figure 2.9 The Structure of the Naïve Bayes Network (Friedman *et al.*, 1997).

Assuming that the effect of each attribute on a given class is independent of the other attributes simplifies the computation. Therefore, Naïve Bayes Classifier is defined

as “naïve” (Yang *et al.*, 2016b). A Naïve Bayes model learns from the training data quickly (Brownlee, 2018b). Additionally, Naïve Bayes Classifier has exhibited high accuracy and efficiency when applied to a large database (Yang *et al.*, 2016b). It is robust, quite easy to construct and use (Awad & Khanna, 2015), and excellent at handling irrelevant attributes. Moreover, the Naïve Bayes algorithm can be naturally extended to solve multiclass data (Ayhan, 2019). On the other hand, it is very vigorous to violations of its independence assumptions (Bauer *et al.*, 1999). Additionally, the Naïve Bayes approach depends on all categorical attributes. Moreover, if the number of instances is small, estimating probabilities by relative frequencies can deliver a poor evaluation. The continuous attributes can be transformed into categorical ones to overcome the first problem and a more complicated formula can be used to estimate probabilities to overcome the second problem (Bramer, 2020). If the normality assumptions are clearly not correct, kernel density estimators can be used to improve performance (Witten *et al*., 2011).

> “This classifier learns from training data the conditional probability of each attribute A_i given the class label C. Classification is then done by applying Bayes rule to compute the probability of C given the particular instance of $A_1,...,A_n$, and then predicting the class with the highest posterior probability. This computation is rendered feasible by making a strong independence assumption: all the attributes A_i are conditionally independent given the value of the class C. By independence we mean probabilistic independence, that is, A is independent of B given C whenever Pr (A|B, C) = Pr (A|C) for all possible values of A, B and C, whenever Pr(C) > 0 (Friedman *et al.*, 1997)”.

2.3.3.2 Nearest Neighbor Algorithm

K-Nearest Neighbor (KNN) is a ML algorithm that can be used for regression and classification problems, and it has names such as instance-based learning, and lazy

learning in different disciplines (Brownlee, 2018b). Moreover, it is yet to receive a lot of attention in the field of building energy prediction (Olu-Ajayi *et al.*, 2022). The KNN classification specifies a group of k objects in the training set that is closest to the test object and gives a label according to the most predominant class in this neighborhood (Awad & Khanna, 2015). The three main components of this method are (Awad & Khanna, 2015):

- A current set of labeled objects
- A distance metric to calculate distance between objects
- The number of nearest neighbors (k)

The training dataset is used directly to make predictions in KNN (Brownlee, 2018b). The KNN algorithm can manage multiclass data classification problems apart from binary classification problems (Ayhan, 2019). Additionally, the nonlinear relationships can be managed well by KNN (Joshi, 2020). The nearest neighbor's weakness is its unsatisfactory way of dealing with categorical attributes. However, modifications can be made to deal with them (Bramer, 2020). The KNN algorithm has low bias and high variance. However, increasing the value of k contributes to the prediction and in turn increases the bias of the model (Brownlee, 2018b). If there are many unobserved instances, the lazy learning approach can be computationally expensive. Moreover, it does not give any idea of the fundamental causality of the task domain. Besides, when the number of attributes increases, it quickly becomes impossible to visualize them (Bramer, 2020). Therefore, it can benefit from an effective feature selection process (Zhang, 2020).

2.3.3.3 Support Vector Machines

Support Vector Machines (SVM) which were developed in the 1990s are one of the most popular machine learning algorithms (Brownlee, 2018b). SVMs are supervised learning methods that are mainly used for regression, classification analysis, feature

reduction, and novelty detection (Awad & Khanna, 2015). SVMs are a combination of linear modeling and instance-based modeling (Witten *et al*., 2011).

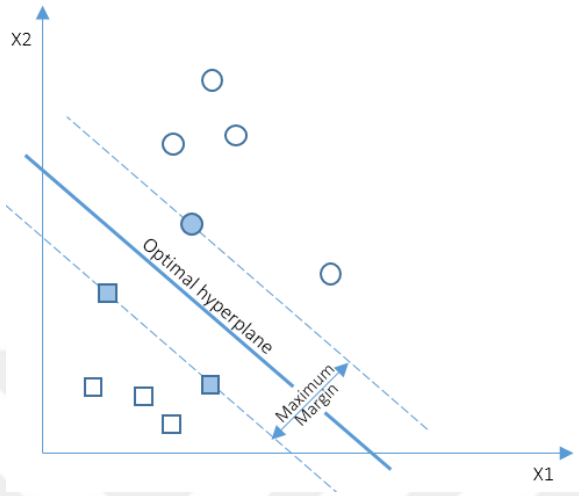


Figure 2.10 The SVM Algorithm (Awad & Khanna, 2015).

As Figure 2.10 shows, a SVM model maps the observations as points in space, and they are classified into a separate partition which is divided by the largest distance to the closest observation data point of any class. Later, new observations are predicted to belong to a class depending on which side of the partition they fall. Support vectors are the data points closer to the hyperplane that separates the classes (Awad & Khanna, 2015).

Although their principle is quite simple, SVMs are famous for their good generalization capability, their robustness, and their unique global optimum solutions (Awad & Khanna, 2015). Even though the amount of training data is relatively small, the algorithm is good at solving non-linear problems (Amasyalı & El-Gohary, 2018). They have shown excellent performance in binary classification tasks (Li *et al.*, 2004). However, methods such as One-Vs-One (OVO) and One-Vs-All (OVA), *etc.* should be externally used for multiclass classification problems

(Ayhan, 2019). Overfitting is not likely to occur with support vectors. However, compared to other methods like decision trees, the computational complexity of SVMs is still a problem and SVMs are slower (Witten *et al*., 2011).

Since SVM assumes that the inputs are numeric, categorical inputs may need to be converted to binary dummy variables (Brownlee, 2018b). Least squares SVM and parallel SVM were also implemented to increase the computational efficiency of SVM in building energy consumption prediction (Amasyali & El-Gohary, 2018). The SVM algorithm is applied in practice using a kernel. The kernel specifies a distance measure or the similarity between the support vectors and new data (Brownlee, 2018b). In binary SVMs, kernels can be understood as dissimilarity measures of pairs of objects in the training set. Linear, Polynomial, and Gaussian are popular positive kernels (Sidaoui & Sadouni, 2013). The learning performance of the SVM depends greatly on the choice of kernel function to transform data from input space to a higher dimensional feature space (Chamasemani & Singh, 2011). Therefore, the applied kernel function should be selected carefully.

Kadyrova & Pavlova (2018) state that machine tuning and determination of the optimal parameters of the tuned machine are the two principal stages of constructing SVMs. Machine tuning involves searching for values of hyperparameters, regularization parameters, and the kernel that provide the highest quality of the machine. Seyedzadeh *et al.* (2020b) state that selecting inappropriate hyper-parameters will cause poor accuracy which may wrongly be translated as a model failure. Moreover, the authors mention that although it is common to use the default values for hyperparameters in the building energy performance field, recently, researchers have started to tune the ML models for more accurate predictions of energy metrics.

Gaspar *et al.* (2012) state that controlling the slack-variables penalty weight (C) is the most basic approach to SVM classification improvement. C parameter is

independent of kernel choice, and can every time be adjusted. The authors explain that optimizing its sigma parameter jointly with the C parameter is one of the most popular strategies while using Radial Basis Function (RBF) kernel, polynomial, or sigmoid kernels. In their study, Arjunan *et al.* (2020) proposed a methodology that enhanced the existing Energy Star calculation method and used the following range for hyperparameters of an SVM-poly model: degree: 1-3, scale: 0.001-1, c: 0.25-1. Chamasemani & Singh (2011) compared the accuracy and computation time of multi-class SVM with three kernel functions, namely, linear, polynomial, and RBF using the following four C parameters, respectively 10,40,60,80.

2.3.3.4 Random Forests

A classification tree or a decision tree is one way of generating classification rules. Representation of data in the form of a decision tree is more meaningful and easier to interpret (Bramer, 2020). A decision tree creates a binary classification tree (Li *et al.*, 2004). While a tree illustrates the learned function, each node in the decision tree defines a test of some attribute of the query instance. Moreover, each branch going down from that node refers to one of the possible values for this attribute (Pantic, 2006). Decision tree theory does not depend on numerical data (Joshi, 2020). Several algorithms can be used to construct a tree out of some data set. Random forests are among the most powerful decision tree techniques available (Brownlee, 2018b) and it is one of the popular decision tree methods in the field of building energy consumption prediction (Amasyali & El-Gohary, 2018).

"Weak learners" work together to form "strong learners" through a diverse collection of decorrelated decision trees in Random Forest (Awad & Khanna, 2015). It combines the output from several shallow trees, rather than developing a solution founded on the output of a single deep tree. While single decision trees experience

high variance and high bias, random forest averages several decision trees to increase estimation performance (Awad & Khanna, 2015).

The individual trees are constructed using the algorithm presented below (Rokach & Maimon, 2008):

Require: *IDT* (a decision tree inducer), *T* (the number of iterations), *S* (the training set), μ (the subsample size). *N* (number of attributes used in each node)

Ensure: M_t; t =1,..., T

1: t ← 1

2: repeat

3: S_t ← Sample μ instances from S with replacement.

4: Build classifier M_t using IDT (N) on S_t

5:t ++

6: **until** t>T

Random Forest needs a few numbers of hyperparameters to be set. The number of independent trees is the main parameter in the forest. Other parameters cover the number of features to consider when looking for the best split, whether bootstrap samples are used when creating trees, and a minimum number of data samples to split a node and required in each node (Seyedzadeh *et al.*, 2020a). Besides, it gives good accuracy even if a substantial portion of data is missing (Awad & Khanna, 2015). It is fast (Rokach & Maimon, 2008). RF can evaluate the relative importance of a variable for classification (Awad & Khanna, 2015). Even though not all features

are equally important to the outcome and specific features can have some interdependencies, Random Forest tree architecture enhances model performance in similar situations (Joshi, 2020). Additionally, the author explains that the features are sampled randomly with replacement and this method makes the model more robust and resilient to noisy data.

2.3.4 Performance Measures in Machine Learning

Machine learning algorithms are evaluated using several criteria. This part explains the used performance evaluation metrics in the research (accuracy, precision, recall, F-measure, AUC) in addition to some of the most widely used ones (confusion matrix). Japkowicz & Shah (2011) explain that performance measures and statistical significance testing are some of the principal factors that are necessary to evaluate learning algorithms. Therefore, statistical significance testing is briefly introduced since the performance of the algorithms was also analyzed using a Paired T-Tester statistical test.

- **Predictive Accuracy**

Predictive accuracy is the most evident measure to use for evaluating the performance of a classifier (Bramer, 2020). Predictive accuracy is often presented as a percentage and is the ratio of the number of correct predictions out of all predictions made (Brownlee, 2018a). Besides, the author adds that, regarding evaluating the predictive accuracy of a classifier, three main methods are often used: Splitting the data into a training set and a test set, k-fold cross-validation, and N-fold (or leave-one-out) cross-validation (Bramer, 2020). As Brownlee (2018b) explains, using a resampling method such as k-fold cross-validation and keeping back a validation dataset to evaluate the accuracy of the model are two vital methods that can be used to limit overfitting. In an overfitting situation, the induction algorithm generates a classifier that fits the training data very well but has lost the capability of generalizing

to instances not introduced during training (Rokach & Maimon, 2008). Overfitting is one of the reasons for inferior performance in machine learning and it is very likely with nonparametric and nonlinear models (Brownlee, 2018b).

In this research, k-fold cross validation is utilized method to limit overfitting and reduce the inferior performance of ML models. In K-fold cross-validation, the dataset X is splitted randomly with K equal- cross-validation sized pieces, Xi,i = 1,...,K. To generate each pair, one of the K parts is kept out as the validation set, and the remaining K-1 parts are combined to form the training set (Alpaydın, 2010). K-fold cross-validation enables us to achieve statistically valid results when the size of the original sample is not large (Kadyrova & Pavlova, 2018). The standard procedure is to repeat the cross-validation process 10 times and average the results (Witten *et al*., 2011). By doing that, the limitations such as the small size of training and validation sets, sources of randomness in the learning method, or availability of noise and outliers in the dataset can be overcomed (Alpaydın, 2010). Moreover, Witten *et al*. (2011) add that given a single, fixed sample of data, using stratified 10-fold cross-validation is the standard way to predict the error rate of a learning method. Stratification is ensuring that each class is adequately represented in both training and test sets by random sampling.

- **Confusion Matrix**

In addition to the predictive accuracy, it is useful to see an analysis of the classifier's performance in a confusion matrix (Bramer, 2020) which is more commonly called a contingency table (Bouckaert *et al.*, 2013). A Confusion matrix is a table that indicates the number of predictions for each class in comparison to the number of instances that really belong to each class (Brownlee, 2018a). An example of a 2x2 confusion matrix can be seen in Table 2.3.

Table 2.3 Example of a 2x2 Confusion Matrix

2x2 confusion matrix		predicted condition	
		positive	negative
actual condition	positive	True Positive	False Negative
	negative	False Positive	True Negative

In a confusion matrix, if the prediction is positive for a positive example, this is called a true positive; if the prediction is negative for a positive example, this is called a false negative. If the prediction is also negative for a negative example, we have a true negative, and if we predict a negative example as positive, we have a false positive (Alpaydın, 2010). It is important to note that the ideal combination of false positives and false negatives may be changing for different applications (Bramer, 2020).

- **Precision, Recall, F-Score**

Using the confusion matrix above, various performance metrics including precision, recall, and F score are calculated. While precision is defined as the proportion of instances classified as positive that are really positive, recall is defined as the proportion of positive instances that are correctly classified as positive (Bramer, 2020). There is a trade-off between precision and recall. Therefore, the F score which is the harmonic mean of the precision and recall can be quite useful as it is a combination of them (Mortaz, 2020). More detailed information on these performance measures is presented in Appendix H.

- **Kappa Statistic**

Moreover, the kappa statistic is a measurement that shows the agreement of prediction with the true class, and 1.0 means complete agreement (Bouckaert *et al.*, 2013). The Kappa metric presents the predictive accuracy information taking the class balance into account. Therefore, this metric can be reviewed in case the data has unbalanced classes (Brownlee, 2018a). In this research, the interpretation given by Mun *et al.* (2019) was applied. A value below 0.2 was viewed as poor, 0.2–0.4 was fair, 0.4–0.6 moderate, 0.6–0.8 substantial, and over 0.8 were defined as good.

- **Area Under the Curve**

Bramer (2020) states that when the classes are seriously unbalanced, predictive accuracy alone is an unreliable indicator of a classifier's effectiveness. For instance, ROC curves, which stand for receiver operating characteristic, are used in situations in which the learner is aiming to select samples of test instances that have a high proportion of positives (Witten *et al*., 2011). Rokach & Maimon (2008) state that ROC curves can help to find the best model as long as one model dominates. If there is not a dominating model, the area under the curve (AUC) comparison can create a dominance relationship between classifiers. Additionally, the imbalance of the training set does not influence AUC (Rokach & Maimon, 2008). Witten *et al*. (2011) explain that when the area under the curve is larger, the model is better.

- **Statistical Significance Testing**

Lastly, as Japkowicz & Shah (2011) state statistical significance testing is one of the principal factors that is necessary to evaluate learning algorithms. A statistical significance test can be used to judge whether the accuracy between two classifiers is different due to chance (Han *et al*., 2012). A student's t-test or paired t-test that is a more sensitive version of the former can be used to determine whether the mean of a set of samples ($\bar{x}$) is significantly greater or less than the mean of another ($\bar{Y}$).

When there are enough samples, the mean of ($\bar{x}$) of a set of independent samples has a normal (*e.g.*, Gaussian) distribution (Witten *et al.*, 2011). The variance can be estimated from the set of samples to obtain confidence limits on (the true value of the mean) given the mean of the samples. To decide whether the means $\bar{x}$ and $\bar{Y}$, are the same or not, the differences between corresponding observations are considered. If they're significantly different, the difference will be significantly different from zero. For a given confidence level, usually, 5% or 1% is used in practice, it is checked if the actual difference exceeds the confidence limit. If the value of t is greater than z, or less than -z, the null hypothesis is rejected. It means that there is a significant difference between the two learning methods (Witten *et al.*, 2011). Joshi (2020) explains the four main parts of hypothesis testing: (1) Defining the null hypothesis, (2) Gathering observational or experimental data, (3) Computation of the probability (also called as P value), (4) Comparison of P value with the predetermined level of significance, e, symbolized as α.

2.3.5 Feature Selection in Machine Learning

Feature selection also named attribute selection, or variable subset selection has been an effective technique to reduce the complexity of machine learning and data mining applications (Zhang, 2020), evaluate the informative features, and reduction of the dimension of data (Trivedi, 2020). Joshi (2020) explains that the density of the data decreases greatly with an increase in dimensions. Data with a higher density will be more likely to find a good model and have higher confidence in the accuracy of the model. Moreover, feature selection can be beneficial to solve the problems of data overfitting and poor prediction accuracy on test samples (Zhang, 2020).

A feature selection method can be categorized into three groups: Filter methods, Wrapper methods, and embedded methods (Ma & Guo, 2014). Rachman & Ratnayake (2019) add that domain knowledge can also be used as a feature selection

method to explain the relationship among features and to avoid data overfitting if the nature of a problem is clearly understood.

Feature selection methods offered by Weka have two main parts, namely, "Attribute Evaluator" and "Search Method." Each part provides several techniques. Witten *et al*. (2011) explain that search methods aim to find a good subset of attributes. Some attribute evaluator methods should be used with specific search methods (Brownlee, 2018a).

In this research, although the impact of the features selected by the researcher on the performance of the algorithms was reviewed, a better performance was not achieved. This might be due to the complexity of studying building performance and the difficulty to depict the underlying reasons for the performance gap without monitoring the buildings. While a wrapper attribute evaluator was used in the first part to select the relevant features to solve binary classification problems, a Chi-squared attribute evaluator was utilized for choosing the relevant features to solve multi-class classification problems. The following parts introduce these two methods.

- **Selection with Wrapper Attribute Evaluator**

While a machine learning algorithm is used in "Wrapper methods" to select the best subset of features (Rachman & Ratnayake, 2019), an independent assessment is made based on general characteristics of the data in Filter methods (Witten *et al*., 2011). Wrapper methods usually create better predictive models than filter methods whereas, filter methods are usually much faster (Magoulès & Zhao, 2016) and more practical than wrapper methods, particularly to use on data of high dimensionality (Rokach & Maimon, 2008). Moreover, Brownlee (2018a) states that the WrapperSubsetEval method must be used with a GreedyStepwise or BestFirst

Search Method in Weka and clarifies that the algorithm used to evaluate the subsets does not have to be the algorithm that you will use to model your problem.

- **Selection with Chi-Squared Attribute Evaluator**

Chi-squared attribute evaluator is classified as a "Filter method." Witten *et al*. (2011) divide attribute evaluation methods for attribute selection in Weka into two: attribute subset evaluator and single attribute evaluator. A chi-squared attribute evaluator is a single attribute evaluator which computes the chi-squared statistic of each attribute with reference to the class. The ranker method is used with single attribute evaluators, and it implements attribute selection by removing the lower-ranking ones. Trivedi (2020) compared three feature selection techniques (such as Information-gain, Gain-Ratio, and Chi-Square) and five ML classifiers (Bayesian, Naïve Bayes, Random Forest, Decision Tree (C5.0), and SVM). The results indicated that the Chi-square feature selection method was suitable with the most informative predictors for all machine learning models.

2.3.6 Synthetic Minority Over-sampling Technique

In this research, the Synthetic Minority Over-sampling Technique (SMOTE) were utilized in classification problems when there was class imbalance in datasets. If the classes are not almost equally represented, a dataset is imbalanced (Chawla *et al.*, 2002). Therefore, oversampling techniques are applied to create a balanced training set by adding new data to the minority class of an imbalanced training set (Hong *et al.*, 2020). Because standard classifier learning algorithms have a bias toward the classes with greater number of instances in imbalanced datasets. Because rules that correctly predict those instances are positively weighted in favor of the accuracy metric, while specific rules that predict examples from the minority class are often

ignored. Thus, minority class instances are more frequently misclassified than those from the other classes (Galar *et al.*, 2011).

Some alternative methods to SMOTE include, randomly under-sampling the majority class, randomly oversampling the minority class, and a combination of oversampling and under-sampling (Liu & Zio, 2019). Generally, oversampling is preferred over undersampling (Somu *et al.*, 2021). Because removing the instances by the under-sampling may cause a loss of information and give an undersized dataset (Somu *et al.*, 2021; Chen & Zhang, 2022).

Somu *et al.* (2021) describe the procedure for synthetic sample generation as follow:

Computing the difference between the considered sample and its nearest neighbors. Multiplying the difference with a random number in the range of [0,1] and adding it to the considered sample. In this way, synthetic samples (n_i) that fall in the line segment joining the nearest neighbors of the minority class are generated (Eqn. (3)).

$$n_i = X_i + gap*diff \quad (3)$$

Where,

X_i is the considered sample, gap is the random number, and diff is the difference between X_i and its nearest neighbors.

Many researchers utilized SMOTE to overcome class imbalance problems in their studies such as Ye *et al.* (2021), Tian *et al.* (2021), Hong *et al.* (2020), and Mun *et al.* (2019). Tian *et al.* (2021) proposed an agent-based modeling framework for the simulation of household energy consumption and conducted SMOTE to improve the deep learning model results for small family samples. Ye *et al.* (2021) developed a data-driven random forest-based framework to model the relationship between city-block-level building-oriented features and building energy consumption. Hong *et*

al. (2020) compared several oversampling methods to optimally predict complaints because of environmental pollutants from construction sites using imbalanced empirical data. Mun *et al.* (2019) focused on residential Air Conditioner (AC) use behavior and conducted the prediction and validation of AC on-off state.

2.4 Inferences Drawn from Literature Review

Studies show that the notion of building performance is complex due to several reasons including; (1) the interactions between buildings and the natural and built environment which is constantly changing, and (2) the diversity of materials, components, and systems that constitute the buildings, (3) the necessity to design these systems holistically so that they can serve several goals (4) sometimes conflicting natures of these goals (5) the redundancy and diversity of the stakeholders involved in the building production stages, (6) the different goals of these stakeholders.

The importance of energy conserving measures in buildings has increased especially after the energy crisis of the 1970s, and its applications have gained momentum globally due to the energy dependency of countries, environmental concerns, and climate change. Nowadays, there is a broad range of building energy standards, and certification systems developed in different countries which are considered a way to energy dependency, climate change mitigation, and carbon dioxide emissions reduction.

Several methods have been presented in the literature for building performance analysis, such as engineering calculations, statistical methods, simulation, and machine learning. Among these methods, despite its limitations such as requiring many inputs, a considerable amount of time, or experienced users, energy simulation is frequently used to understand the performance of a design option and to achieve certification.

Nevertheless, many scholars address a widespread phenomenon, defined as the energy performance gap, the difference between a building's planned and actual performance. This phenomenon addresses not only increased energy consumption but also reduced consumption. Moreover, the gap is observed in both new buildings and energy retrofits. It is represented in the literature that there could be an average deviation of 30% between actual vs. estimated consumption (Herrando *et al.*, 2016), actual building performance could almost double the target (Fedoruk *et al.*, 2015), or could be up to 287% (Calì *et al.*, 2016).

As a result of the importance of the energy performance gap in buildings, several researchers proposed distinctive methods to reduce the performance gap in buildings. Some of them are applying the probabilistic and statistical approaches to simulation (Demanuele, 2008; Hong *et al.*, 2015), reexamining the tools used in the sector (De Wilde, 2018), simulating the user behavior through exhaustive modeling (Ali *et al.*, 2020), using monitoring data to calibrate simulation models (Cuerda *et al.*, 2020). Another way of approach to mitigate the gap is integrating risk management into energy-efficient buildings. Topouzi *et al.* (2019) used case studies of deep retrofits across the UK and EU in their study to estimate risks associated with renovation processes and technologies. Zou *et al.* (2016) developed a risk management framework for managing the risks associated with building retrofit projects. Moreover, several researchers studied the risks present in energy-efficient buildings. Stakeholder-associated risk in complex green building projects (Yang & Zou, 2014), risks in green building projects in Singapore (Zhao *et al.*, 2016), risk factors of green building in China (Qin *et al.*, 2016), overheating risk in buildings (Liu *et al.*, 2016a; Silvero *et al.*, 2019; Fosas *et al.*, 2018; Chvatal & Corvacho, 2009) are some of these studies.

The energy performance gap can cause many negative effects both on a small and large scale. It may not only bring reputational dangers to designers, contractors, or companies; but also bring disadvantages such as economic loss to a variety of

stakeholders including occupants, building owners, and the ESCOs. Moreover, it might present obstacles for governments to reduce dependency on foreign energy sources, as well as for policymakers to be successful in future ambitions about reducing the environmental impacts of buildings and greenhouse gas emissions.

Nižetić & Papadopoulos (2017) mentioned the importance of having a tool for estimating a possible gap between real and expected energy savings and proposed a novel but only conceptual approach for energy performance gap prediction. Apart from this study, no approach has been found pointing to predicting the energy performance gap in buildings, primarily by using a group of projects from the past with the information on energy performance gap and the project risks encountered. As a part of the growing field of data science, machine learning methods appear often in energy performance prediction studies (Paudel *et al.*, 2017; Seyedzadeh *et al.*, 2020b; Arjunan *et al.*, 2020; Walker *et al.*, 2020). However, to the best of the knowledge of the researcher, there are no studies applying the machine learning concept and identifying project risks ***to predict energy performance gaps in buildings***. The availability of such a decision support tool can enable project stakeholders to prevent unnecessary investments and take action to control the performance gap by changing their project decisions. Therefore, in this research, an energy performance gap prediction model applying ML through the lens of project risks is suggested.

CHAPTER 3

METHOD AND MATERIAL

To accomplish the objectives of the thesis, two main research methods were adopted. *First*, an energy performance gap framework was proposed after a comprehensive literature review, and semi-structured interviews with experts working on different energy-efficient buildings. The framework was designed as a web-based survey for data collection. *Later*, the machine learning classification concept was utilized to solve EPG prediction problems. This chapter first introduces the method of the study. *Following this par*t, the material of the research which is a dataset containing energy performance gaps and project risk information of different buildings is presented. *After that*, brief information about the data types in the dataset is given.

3.1 Method

The research was carried out in two main steps.

EPG Framework Development: After a comprehensive literature review, semi-structured interviews were conducted with 11 participants working on 6 energy-efficient buildings. At the end of this process, an energy performance gap conceptual framework that demonstrates how the risks encountered during the life cycle of buildings work together and cause EPG was proposed. Afterward, this framework was designed as a web-based survey that would be used for data collection.

Machine Learning Classification: The machine learning classification technique was utilized and the performance of four machine learning classification algorithms (1) Naïve Bayes, (2) KNN, (3) SVM, and (4) Random Forest was compared to propose energy performance gap prediction models. To decide a group of algorithms for the prediction of given problems, the dataset was first trained by a larger number of ML classification algorithms. Since the aforementioned algorithms provided higher accuracy, they were included in the study. At this stage, the Weka workbench, open-source software that provides methods for standard data mining problems, was used to apply the ML classification techniques. Using several performance metrics, the algorithm that achieves the highest performance was presented as the final classifier. Figure 3.1 indicates a detailed representation of the applied stages in the research.

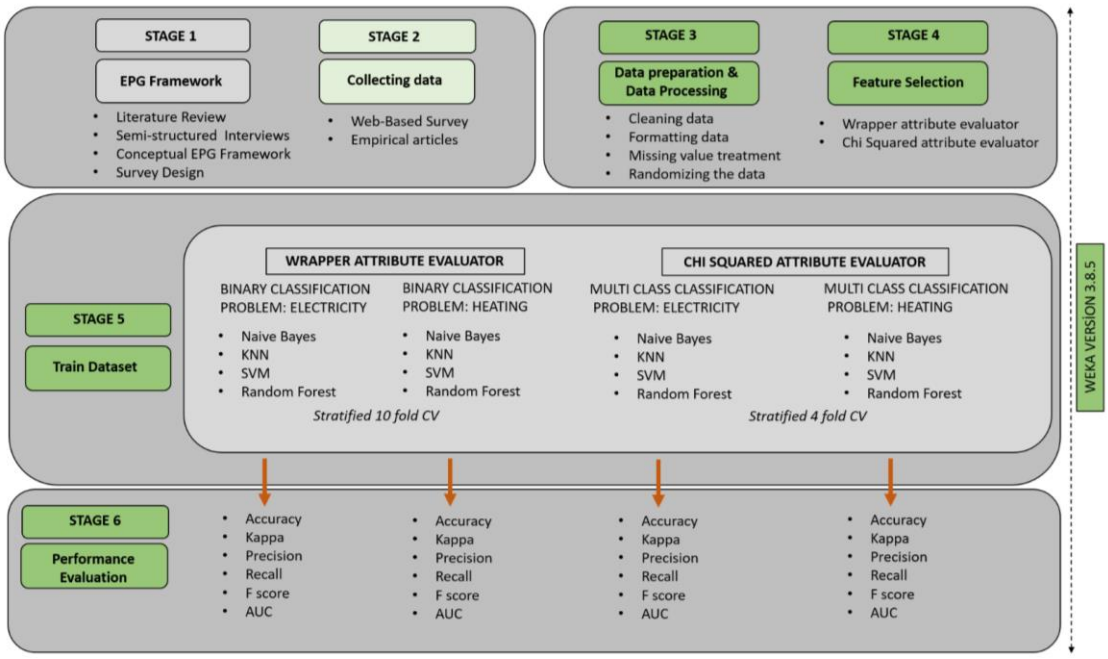


Figure 3.1 Steps Applied in the Research

3.1.1 Energy Performance Gap Framework

The energy performance gap framework is designed after a comprehensive literature review, and semi-structured interviews conducted with the experts working on energy-efficient buildings. Each step to reach the final framework is introduced in the following stages.

3.1.1.1 Semi Structured Interviews

Interviewing is a data collection method for qualitative research, and it is used to obtain a vivid picture of the participant's perspective on the research topic (Liu, 2015). There are three types of interviews: unstructured, semi-structured, and structured. Semi-structured interviews are conducted by a list of questions in a flexible way and give essential information on individual perspectives and experiences through direct discussion (Liu, 2015).

In this research, following the literature review, semi-structured interviews were conducted in two rounds with experts working on the design, construction, and operation of energy-efficient buildings. The cases were selected so that different risk paths of a variety of buildings, and different company and country conditions could be observed.

The goal of the semi-structured interviews was (i) to verify the factors that were drawn from the literature, (ii) to identify the project risks, and (iii) to see how project risks work together. Based on the experiences of different stakeholders, a causal relationship between these factors was established and the dynamics of the performance gap were better understood. As a result, a conceptual framework that explains the energy performance gap phenomenon in buildings was proposed.

Semi-structured interviews were conducted face to face or on an online meeting platform after the approval of the METU Human Subjects Ethics Committee between December 2020 and March 2021. The approval of the Applied Ethics Research Center can be found in Appendix B.

Questions were sent to the experts over the web before, so they could have prior knowledge and preparation. One-hour interviews were recorded using a tape recorder or web platform. During the interviews, the experts were first asked to give information about themselves and their professional experience and company. Next, they were asked to explain a specific building project that they were working on and to describe the project process in general. Following that, they were asked to explain the problems/challenges they encountered during the project process, the results of these problems, how they solved them and if these problems caused some other problems. Interview questions can be found in Appendix C.

After the first round of the semi-structured interviews, the problems described by experts were represented using the cognitive mapping method to review the risk paths of the projects. Cognitive mapping is a method that is used to show and actively shape the mental models or belief systems that people use to perceive, simplify, contextualize, and interpret problems that are complex otherwise (Tanrıverdi, 2019). According to Fidan (2008), this method is used in the field of risk management to discover risks in projects, explore the process of risk identification, and model complex projects. In the second round, experts were asked about the interrelations in the cognitive maps and to indicate whether they agreed with them or wanted to revise any part. Information about case studies, interviewees, the findings of the interviews, and cognitive maps of each case are described in the following sections.

The experts consisted of eleven professionals with on average 12 years of experience in energy-efficient buildings. One CEO, three project managers, three mechanical engineers, one site manager, one electrical technician, one quality manager, and one

commissioner participated in the semi-structured interviews. Since one of the research goals is to provide a life cycle perspective to the study, it is aimed that the experts will be the ones who participated in the design, construction, and use or operational phases of the buildings. While nine of the participants are from Turkey, two of them are from Germany. Moreover, one of the participants had a doctoral degree, one participant had an associate degree, and the rest of the participants had a master's degree or bachelor's degree. The interviewee profile can be found in Table 3.1.

Table 3.1 Interviewee Profile

Case	No	Profession	Job Position	Country	Experience (years)		Education Level
					Constr. sector	Energy Efficient Buildings	
A	I1	Architect	Project Manager	Turkey	21	8	Master's
	I2	Civil Engineer	Site Manager		39	8	Bachelors
	I3	Mechanical Engineer	Mechanical Engineer		23	8	Bachelors
	I4	Mechanical Engineer	Commissioner		36	12	Ph.D
B	I5	Mechanical Engineer	Mechanical Engineer		5	2.5	Bachelors
	I6	Electrical Technician	Electrical Technician		37	10	Associate degree
	I7	Printing Educator	Quality Manager		8	8	Bachelors
C, D	I8	Architect I City Planer	Project Manager		11	9	Master's
	I9	Mechanical Engineer	Mechanical Engineer		9	8	Master's
E	I10	Architect	Project Manager	Germany	28	21	Master's
F	I11	Architect I Civil Engineer	CEO		34	34	Master's
*4th interviewee participated in the interviews as the Commissioner of Building A and B.							

3.1.1.2 Case Studies

The buildings used for the case studies had different project types (e.g., educational, office, residential), certificates (e.g., LEED, Passive House), and locations (Turkey, Germany). Table 3.2 represents detailed information about the project profile.

Table 3.2 Project Profile

Case Study	City	Certificate	Building Type	Construction Year	Area
A	Ankara	LEED Platinum	Head Quarters	2014	9.538 m²
B	Ankara	LEED Gold	Head Quarters	2020	45.782 m²
C	Ankara	x	Educational	2017	17.030 m²
D	Ankara	x	Head Quarters	2019	8.955 m²
E	Munich	Passive House	Residential	2019	4.009 m²
F	Heidelberg	Passive House	Residential	2018	15.150 m²

Building A is a Headquarters office constructed in 2014, in Ankara which has two basement floors, a ground floor, 5 upper floors, and 9.538m^2 air-conditioned area. The building has earned LEED Platinium certification. High-level insulated walls and window frames are some of the energy-saving measures of the building (10cm for external walls, 13cm+6cm for the roof). Heating and cooling are provided with water-sourced variable refrigerant volume systems in the buildings. The heating system includes a cogeneration unit, thermal storage tanks, solar collectors, boiler, and solar wall, the cooling system consists of thermal storage, absorption chiller, and cooling tower. Moreover, the building uses solar collectors and PV systems as renewable energy sources.

Having a building area of 45.782 m^2, Building B is used as the Headquarters office of a Public Institution and it was constructed in 2020 in Ankara. High levels of insulation (15cm for external walls, 10cm for the roof) and triple glazing are some of the energy-saving measures of the building. It is a LEED certification candidate and is expected to earn Gold certification. Air handling units are used for ventilation and 4 pipe fan coils are used for heating and cooling. In addition, the building has solar collectors, photovoltaics, condensing boilers, and an automation system.

Building C and Building D were carried out in collaboration with various firms from Turkey and abroad. The goal was to design exemplary buildings with an Integrated Design approach that had improved energy efficiency, and reduced greenhouse gas emissions. While Building C is an educational facility constructed in 2017, Building D is a Headquarters office of a Public Institution constructed in 2019. The project team proposed solar chimneys and solar walls to support heating passively, and a pre-cooling / labyrinth system to support heating and cooling. Both projects have environmental and mechanical features like underfloor heating and cooling, energy-efficient ventilation, chilled beam units for cooling, building automation, heat recovery systems, environmentally friendly materials, green roofs, and economical water usage.

Building E is the largest wooden housing project in Germany and has a Passive House certificate. The project which was constructed in 2019 in Munich has 4.009 m^2. It includes facilities such as a swimming pool, kindergarten, and common work areas in addition to thirty-six dwellings. A German and an Austrian architecture firm collaborated on the design of the project. Some of the energy-saving measures of the project are aluminum-clad wood frame windows, triple glazing, heat recovery ventilation units, district heating, green roof, and solar panels for the generation of electricity and hot water.

Building F, which was constructed in 2018 in Heidelberg, is a Passive House certified building. The project has 15.150m^2 usable space and consists of residential and commercial areas. The investment firm is also the project developer, architect, and property owner of the project. The project includes 162 apartments. Some of the energy-saving measures of the project are a highly insulated building envelope (28 cm thick insulation on the exterior walls), a vertical garden, a green roof, and solar power.

3.1.1.3 Evaluation of the Project Process

In this section, the strengths and the difficulties encountered during the project process of each case will be summarized based on the comments of the experts.

- **Building A**

The goal of Building A was to construct an exemplary building that is going to be truly green, therefore the project team planned the process carefully [I4]. The project firm has more than 40 years of experience in the construction industry and provides engineering, construction, manufacturing, and specialized software development services. This has enabled the accessibility of knowledgeable and experienced personnel within the company, the ability to quickly adapt the desired changes in the project to the construction site environment, and the accessibility to the professional, technical staff at the construction site [I2]. Interviewee number 4 believed that since the building was a private-sector project, the project process had been easier. There was effective communication with the energy consultant. The systems proposed by the energy consultant have been taken into account as much as possible. The building was commissioned and the performance of the system was evaluated. According to Interviewee number 2, the firm had sufficient financial resources which facilitated the selection of higher-quality products. Since they created a work plan using a project schedule, potential delays in the construction process were prevented as much as possible. They gathered feedback from occupants during the operation phase of the project. Based on the results, building occupants were informed about the building systems. Additionally, they organized training for the building occupants when necessary.

Besides, since the firm constructed a building for their use, there were some difficulties while making decisions regarding material selections [I2]. Green or energy-efficient building practices are new in Turkey. Therefore, the team faced

difficulties to find certified products or companies [I3]. Lack of local and high quality mechanical products caused imports for material supply, which resulted in a 10-15% increase in costs and delayed material supplies [I2]. The issues that the design is considered to be insufficient were divided into three: According to the project manager [I1], transparency is a significant aspect of contemporary architecture. The project manager thought that the ratio of transparent surfaces on the facade exceeded a certain percentage and that was in contradiction with the energy efficiency concept. This has increased heat losses and gains in the building. Moreover, since triple glazing did not seem economically feasible, it was changed with a budget-friendly option. The project manager stated that the importance of some design elements could have been explained better to other teams. For instance, the design team suggested sunshades on the south and west facades against the disturbing effect of light. However, the shading devices were canceled afterward and there were complaints from some users on those facades after moving to the building. Interviewee number 2 addresses that there are inadequacies in the simulation programs or the way the buildings are simulated. Moreover, according to Interviewee number 1, the inability to prevent air circulation in large gallery spaces appeared to be a design deficiency and resulted in complaints about comfort conditions. The experts agree that since laboratory or the ideal measurements in the factory environment are different from the conditions in-situ, values given in some catalogs did not match the reality. Furthermore, Interviewee number 3 explained that operating and maintenance services of trigeneration could not be found in Turkey at those times. Therefore, these services had to be brought from abroad. Due to the high cost of these services, trigeneration systems were not in operation in the building. However, the PV system integrated afterward became a more efficient alternative to trigeneration. Lastly, Interviewee number 2 explained that since some users had insufficient knowledge of the building systems, there were complaints about comfort.

- **Building B**

The positive factors experienced in Building B were as follows: Interviewee number 6 stated that the client was knowledgeable and experienced in energy efficiency. The project team aimed to construct a building that truly fulfilled the green building goal. In this project, building user guides have been prepared. According to Interviewee number 4, the communication with the project team was effective. Their suggestions were considered as much as possible. The building was commissioned, and the mechanical equipments were checked if they operated properly. Interviewee number 6 explained that the project owner had sufficient financial resources which enabled the selection of higher-quality products. Moreover, the payments could be made on time based on the availability of financial resources. Interviewee number 4 stated that the availability of a technician who had a good knowledge of the project process helped a more successful follow-up of the systems in the building.

Difficulties encountered during the project process were as follows: Since some projects were drawn by people who did not have construction site experience [I6] or project teams preferred 2D programs instead of BIM applications [I5], there were several revisions due to drawing mistakes during the implementation phase. Besides, the energy efficiency concepts were integrated later in the project [I6]. Therefore, there were some problems with integrated design. The interviewees pointed out inadequacies in regulations. According to them, regulations require solutions that increase the cost too much. Moreover, some stakeholders were in favor of aesthetic and form-oriented design approaches and due to a lack of knowledge, the application of more energy-efficient solutions was not always possible. Due to the excessive cost of material testing processes and lack of sufficient material test labs, stakeholders had to trust catalog values in material specifications [I6]. Early delivery of some materials to the construction site can cause variations in material performance [I5]. Due to the unavailability of workers with sufficient certification and the shortage of field control staff, technical personnel worked overtime in the supervision part.

However, this might have increased the mistakes and impacted building quality [I6]. Besides, lack of knowledge of technical personnel in the field, impacted building quality. The subcontractor went bankrupt in this project and there were troubles in the orientation of the new contractor and changes in the work plan. The lack of local high-quality mechanical equipment and changes in material decisions caused delays in the project schedule. There were problems with supply due to importing the products from abroad during the pandemic [I6]. There are inconsistencies between public tenders and specifications. Policies are not supporting local production, therefore there is a lack of suitable local products. Besides, excessive taxes on labor causes a preference for cheap workmanship that leads to lower quality of workmanship. Additionally, when there is economic inconsistency, it is difficult to find some raw materials in the market and this impacts the delivery of some materials or products [I5]. Furthermore, hours of use in the building were different than planned during the pandemic. Since occupants were complaining about the comfort, heating setpoints were adjusted differently than in the simulations. This was a time that there were extreme weather conditions [I5]. Since construction work did not finish in some parts of the building, heat loss was expected to be higher than assumed levels [I6]. Lastly, poor detailing in some parts of the building and using window frames without thermal breaks increased heat losses due to thermal bridges [I4].

- **Building C**

Building C and Building D had not only the aim of improving energy efficiency in Buildings but also introducing the Integrated Building Design Approach. Therefore there was very frequent communication between the project teams. This process had been an educational tool for different stakeholders. Since some teams did not have experience with sustainable buildings, their lack of experience was tried to be compensated by working in collaboration with partners from abroad. Users of Building C were qualified and had technical knowledge about mechanical systems.

Therefore, this enabled them to provide solutions to some technical problems of the building.

Since experts in Turkey are not familiar with Integrated Building Design Approach, some problems were experienced with the adaptation to the process and embracing the project. Moreover, public sector projects bring some challenges like unit price work, more restrictions to design, and not being able to special items. Budget limitations affected the selection of the product and not being able to hire a commissioner. Besides, there were inadequacies in the tender process, and this impacted the selection of a contractor. Some team members left the team, and changes occurred in the expectations of new team members. The project manager explained that the material they proposed for the roof cover showed different properties than those given in the catalog. As a result, comfort conditions deteriorated, and overheating occurred in the courtyard. Additionally, the experts added that they could not find local high-standard products in mechanical equipment. Cost performance trade-offs affected the systems used. Either foreign products are bought or lower performance products were selected. Besides, due to changes in regulations or bureaucratic obstacles, some mechanical systems could not be operated. In Building C, the contractor was not experienced in sustainability. The project manager expressed that the quality of workmanship was problematic and there were problems with material storage.

- **Building D**

Building D had similar difficulties to Building C related to the public tendering process and the conventional work habits of some stakeholders. However, a lower number of stakeholders made the project process easier in comparison to Building C. There were some problems with the administrative processes. The project location changed several times and this caused loss of budget and time. The first contractor of this project was not experienced in sustainable buildings and he refused to

implement some design decisions. Moreover, the first contractor went bankrupt, and to get progress payment, the contractor made wrong applications and damaged some building components. The new contractor had one energy-efficient building project experience. However, there were construction delays due to budget limitations and the damage created by the previous contractor. Besides, after the building is complete, the project team could not receive thermal images of the building or any feedback related to the airtightness test results.

- **Building E**

The firm which was responsible for the design and energy consultancy of Building E has a long experience with wood construction and Passive Houses. This enabled them to act quickly on construction details and reach knowledgeable staff readily about the energy efficiency subject. Budget limitations constrained material decisions. However, the availability of incentives from the municipality encouraged choosing more environmentally friendly materials. The project manager believed that bureaucracy while making the project difficult to complete at the intended time, also encouraged the project team to be careful with their schedule. Due to budget limitations, technical staff selection in the construction site was always made based on cost instead of experience. The language was seen as a chance and barrier as well. At the construction site, the necessity of effective communication is very critical. A multinational atmosphere increases conflicts on the construction site and the lack of knowledge and motivation of the workers' results in reduced construction quality and building performance. The project manager thought that the software used for calculation was quite well. However, the interface could be redesigned to be more explanatory for new users. He also added that the software focused more on winter comfort than summer comfort. Although passive design strategies such as exterior, or interior shading devices, and natural ventilation are used, it might still be difficult to achieve the summer comfort expectations due to climate change. Therefore, it is vital to communicate everything clearly with the clients/users such as the necessity

for active strategies. Time constraints to complete the project was overcomed by changing the design process or if possible, pre-ordering the materials. The expert observed that usually very few users were interested in the technical systems of the building in their projects.

- **Building F**

The firm is experienced in sustainable architecture and urban planning for more than 50 years and they are not only the investor of Building F, but also the project developer, architect, and landlord of it. Therefore, they followed the use or operational phase more closely. Building F consumed more energy in the summertime than the team's expectations. The expert stated that energy laws and low energy building standards should change their approach and pay more attention to thermal comfort conditions in summer. In this building, they applied vertical gardens to reduce energy consumption in summer. The expert explained that airtightness and the quality of workmanship had vital importance in Passive Houses. As detecting all air leakages is very difficult in a very large building that has thousands of windows, construction workers must know what they are doing and are motivated to bring high-quality applications. Additionally, the expert stated that it was very difficult to predict the behavior of the building users and it is not sufficient to look only at the user behavior. Necessary measurements must be made in the building and the building must be monitored. It was added that users consumed much more than their predictions in another project and they did not know the causes of excessive consumption since they did not monitor the building due to the high cost of the measuring devices. Lastly, the expert mentioned that the contractor of this building quit. Although it was a very big and experienced construction firm, it was their first experience with Passive Houses. Since they underestimated the demand for constructing a Passive House, they realized that they would not complete the project on time and within budget. For this reason, the project developer took over

the construction, but they had a lot of trouble since they did not have enough workers on the construction site at that time.

3.1.1.4 Results of Semi-Structured Interviews

The project processes described above provide an understanding of the risks encountered in the process for each case. According to Fidan *et al.* (2011), risk models that mimic project performance should be founded on risk paths rather than individual risk sources. The authors proposed a risk and vulnerability ontology to collect risk-related project data and understand the reasons for cost overruns in construction projects. This ontology integrated not only risk paths but also vulnerabilities into risk models. Dikmen *et al.* (2008) explain that while vulnerability is the concept of being susceptible to a risky situation, risk is used to define the severity of consequences within a scenario. It was seen that the ontology proposed by Fidan *et al.* (2011) was also applicable to understanding the reasons for the energy performance gap in buildings. In this ontology, risk events occur due to a risk source and lead to a risk consequence.

Moreover, while some vulnerability parameters (V1) influence the probability of risk occurrence, some (V2) are about the manageability of risk, and some others (V3) influence the impact of risk events on project success. Additionally, vulnerability sources were categorized into three types: Robustness, Resilience, and Sensitivity source. The authors defined the robustness source as factors that indicate the "weakness" within a project system that affects the probability of occurrence of risks. Resilience source is defined as factors that affect the manageability of risk and sensitivity source is defined as factors that affect the magnitude of risk consequences. Figure 3.2 illustrates risk and vulnerability paths.

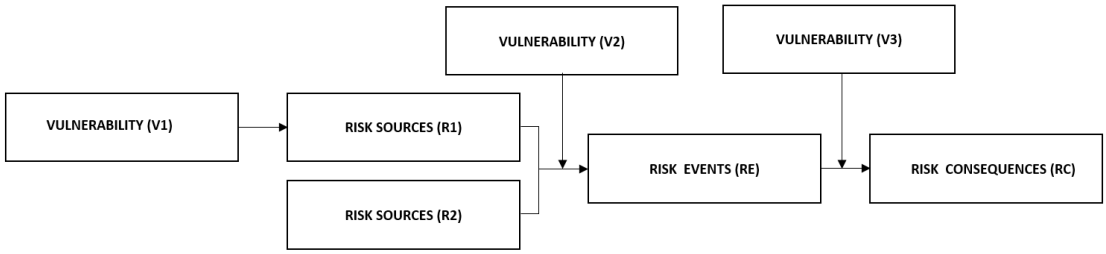


Figure 3.2 Risk and Vulnerability Paths (Fidan *et al.*, 2011)

For the second round of semi-structured interviews, each project process was presented as cognitive maps inspired by the risk and vulnerability ontology of Fidan *et al.* (2011). Suggested paths were discussed with project participants during the second meeting so that cognitive maps could be verified. The cognitive map of Building A can be found in Figure 3.3. While the UP abbreviation stands for use or operational phase, DP, CP, and MP stand for the design phase, construction phase, and manufacturing phase, respectively. These factors indicate the "weakness" within the project system and influence the probability of occurrence of risk sources. Risk events and risk sources are shown with RE and RS, respectively. S is an abbreviation for sensitivity and these factors influence the magnitude of risk consequences (energy performance gap). Additionally, factors represented with R influence the manageability of risks. The cognitive maps of other projects can be found in Appendix D. Considering the cognitive maps of each project and the reasons for the gap in the literature, a conceptual risk-vulnerability framework was prepared to understand EPG, as seen in Appendix E.

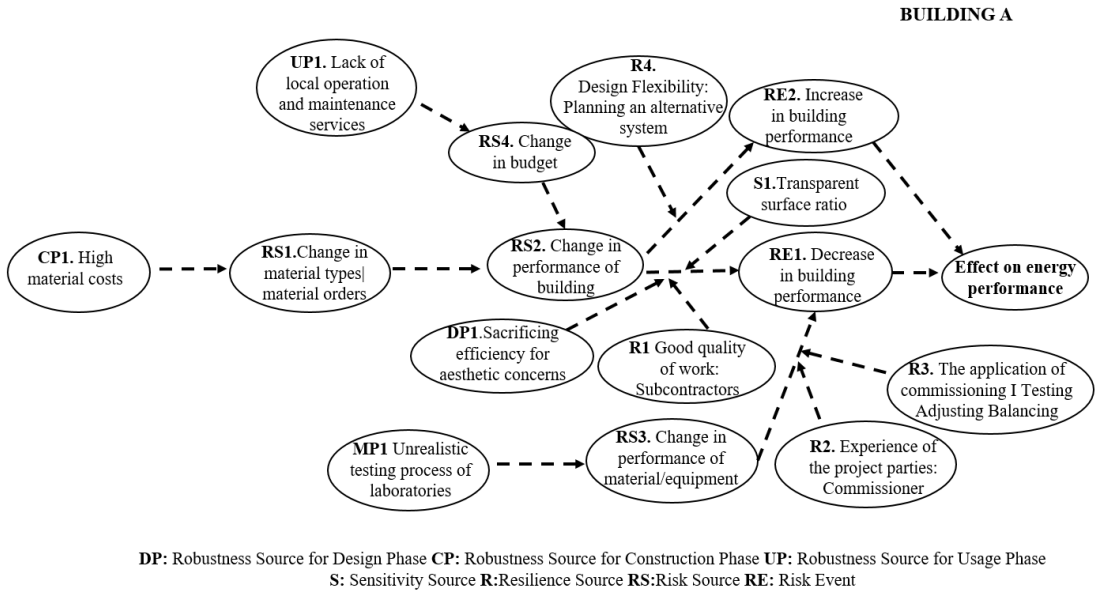


Figure 3.3 Cognitive Map of Building A

It is important to point out that, the experts not only mentioned the problems that can lead to a gap in energy consumption but also in the budget, project time, and user comfort during the meetings. Hatefi & Tamošaitienė (2019) state that many of the risks inherent in construction projects influence each other. For example, the risk associated with the cost of a construction project may increase the project time. Since the EPG framework was more detailed in this case, risks associated directly with project budget, project time, and user comfort were excluded from the conceptual framework.

The first conceptual framework which is based on the study of Fidan *et al.* (2011) became a survey to gather risk and performance gap-related data about buildings. The survey was pretested by two research assistants who work on energy-efficient building research projects and one energy-efficient building consultant. The participants were encouraged to comment on the design of the questionnaire, involving unclear questions, completeness, terminology, clarity, *etc*. Based on the

feedback of the participants and observations of the researcher, the framework and the survey were revised and restructured due to the following shortcomings:

- The proposed survey was excessively long
- Some factors were repeated in more than one phase in the questionnaire with the aim of reflecting the life-cycle perspective. As is shown in Table 3.1, some factors cause performance gaps to occur in more than one phase of the building life cycle (*e.g.*, Miscommunication/ lack of communication between project parties, Lack of knowledge, skill, or experience of project parties).
- Some factors were listed separately with the aim of giving a detailed view of the performance gap issue. (*e.g.*, Unrealistic/oversimplified simulation inputs: Weather, Unrealistic/oversimplified simulation inputs: Occupancy).
- A few factors were also listed about the manufacturing phase of buildings for the evaluation of participants. However, later, it was thought that the target group of the survey might not be able to answer those questions.
- The participants did not find the difference between the requirements of two questions clear (related to risk sources)
- Having two contradictory statements in the same question caused confusion and participants found it difficult to answer (*e.g.*, decrease in building quality, increase in building quality)

As a result of the pretesting of the online survey, it was seen that some factors represent similar conditions generically so they could be defined as a broader concept. Moreover, the factors repeated several times in the questionnaire could be recategorized, and the "Manufacturing phase" could be excluded from the survey questions since the target group of the survey might not be able to answer them. Unclear questions could be reevaluated to see if other parameters in the framework could replace them. Consequently, a new framework was suggested as it is presented in Figure 3.4.

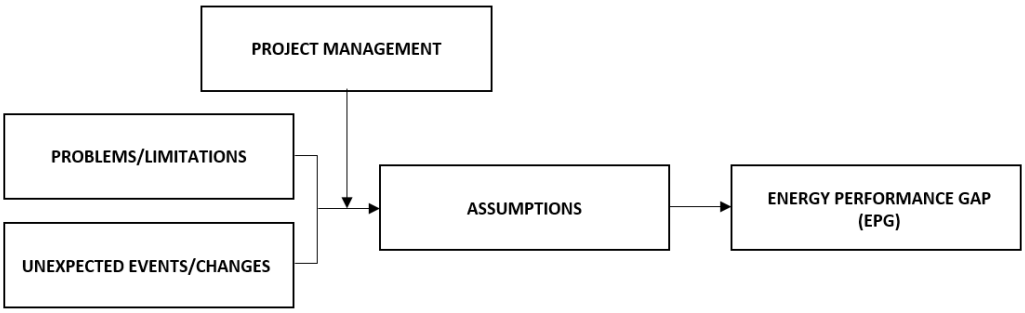


Figure 3.4 Renewed Conceptual Model

There are five categories in this model. (1) Problems, Limitations, (2) Unexpected Events and Changes, (3) Project Management, (4) Assumptions, (5) Energy Performance Gap. Problems, Limitations include factors that occur in the three different life cycle stages of the project and create weakness in the system. These factors may influence the occurrence of Unexpected Events and Changes. However, Unexpected Events and Changes may occur independently as well. Factors that influence the manageability of the groups are categorized as Project Management. According to the framework, factors listed in the first three categories may trigger a change in the Assumptions. As a result, the energy performance gap can be observed. The factors that appeared most in the cognitive maps were selected for this framework. Moreover, the factors were supported by a literature review to be able to cover a broader range of energy-efficient buildings. Many of the factors listed under Assumptions are extracted from the sensitivity analysis literature since sensitivity analysis can be used to specify the key variables impacting building thermal performance. The next section introduces the final conceptual model. A detailed representation of the final conceptual model can be found in Figure 3.5.

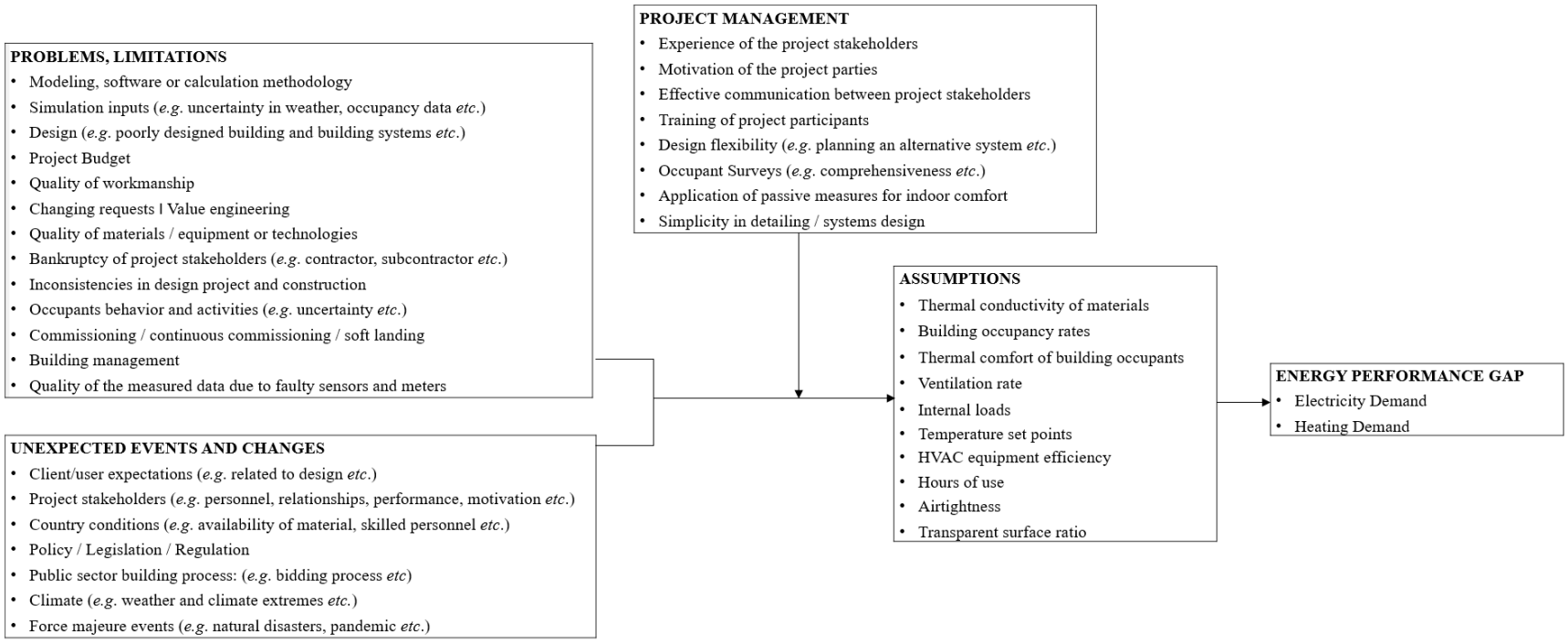


Figure 3.5 Finalized EPG Framework

3.1.1.5 Final Conceptual Framework

"Problems-Limitations" is the first category that contains several factors related to the design, construction, use, or operational phase of buildings. The design phase sub-category includes four factors:

- (D1) Modeling, software, or calculation methodology,
- (D2) Simulation inputs,
- (D3) Design,
- (D4) Project Budget.

Cases in which similar factors became a concern for the performance goals of the projects are given in Table 3.3 with references from the literature. The construction phase sub-category involves five factors:

- (C1) Quality of workmanship,
- (C2) Changing Requests-Value engineering,
- (C3) Quality of materials, equipment, or technologies,
- (C4) Bankruptcy of project stakeholders,
- (C5) Inconsistencies between design project and construction.

Case studies faced such problems and articles mentioning these factors in literature can be seen in Table 3.4. There are four factors in the use or operational phase sub-category. These are:

- (U1) Occupants' behavior and activities,
- (U2) Commissioning-continuous commissioning-soft landing,
- (U3) Building management,
- (U4) Quality of measured data due to faulty sensors and meters.

Table 3.5 illustrates case studies and articles referring to these factors.

Table 3.3 Problems/Limitations: Design Phase Factors

Phase	No	Code	Parameter	References	Case	Comment of Interviewee
Design	1	D1	**Modeling, software or calculation methodology**	De Wilde (2014), Cali et al. (2016), Ali et al. (2020), Gram-Hanssen & Georg (2018), Herrando et al.(2016), Jradi et al. (2018), Kampelis et al.(2017), Mojic et al.(2019), Salehi et al.(2015), Zou et al.(2018), van Heijst (2017), Mills et al.(2006), Xu et al.(2021), Doylend (2015), Ipeec (2019)	Building A Building B Building E	A1. It is assumed that the building is closed on all sides in simulations. For example, it is never taken into account that the entrance doors will be opened and closed. B1. We see some mistakes in drawings since the 3rd dimension is not considered. Preparing drawings in programs such as Revit, instead of using 2d programs such as Autocad can help to solve such problems. E1.I think the software is quite well. However, it must be easier to handle...It must also have some parts that is more explanatory for the people who does not have knowledge on this topic so architect can understand what to do with it. E2. Software puts more focus on winter comfort than summer comfort.
	2	D2	**Simulation inputs**	Alencastro et al. (2018), Ascione et al.(2016), De Wilde (2014), Herrando et al. (2016), Jradi et al. (2018), Kampelis et al. (2017), Mojic et al. (2019), Sunikka-Blank & Galvin (2012), Zou et al. (2018), van Heijst(2017), Xu et al.(2021)	x	x
	3	D3	**Design**	De Wilde (2014), Jradi et al. (2018), Ali et al. (2020), Menezes et al.(2012), Salehi et al.(2015), Zou et al. (2018), Alam et al. (2017), Mills et al.(2006), Coşkun (2019), Doylend (2015)	Building A Building B	A1.The air flow in the atrium spaces could not be prevented after the building construction is finished. This was a design problem.B1. The architect rejected the application of some technical approaches due to their concern for aesthetics and form. For this reason, applications that will provide more efficiency cannot be always made. B1. There are thermal bridges due to poor detailing at several locations, resulting in excessive heat gain in summer and heat losses in winter. B3. There are design deficiencies in some building systems as well, which result in reduced comfort levels and increased energy consumption.
	4	D4	**Project Budget**	Delgado (2016), Eon et al.(2020), Dasgupta et al.(2012), Bhuiyan et al.(2015), Brady& Hanmer-Dwight (2016), Mirzaie& Menzie (2017), Doylend (2015)	Building C	C1.The budget is lower in the context of Turkey since one of the targets is to design repeatable public buildings . When the project's budget is not too big...

Table 3.4 Problems/Limitations: Construction Phase Factors

Phase	No	Code	Parameter	References	Case	Comment of Interviewee
Construction	5	C1	**Quality of workmanship**	Alam et al. (2017), Ali et al. (2020), Aïssani et al.(2016), Alencastro et al. (2018), Birchall (2011),De Wilde (2014), Herrando et al. (2016), Jradi et al. (2018), Kampelis et al. (2017), Mojic et al. (2019), Sunikka-Blank & Galvin (2012), Zou et al. (2018), Zou et al. (2016),Topouzi et al. (2019), van Heijst(2017), Coşkun (2019), Xu et al.(2021),Ipeec (2019)	Building C Building F	C1.We have seen several cracks in the building envelope. There were places we warned them during the construction phase. For instance we told that they did not apply the plaster mesh on the insulation correctly... Besides, there were problems with the storage of the material.F1. You have to make sure that everything is very airtight in Passive Houses .You have to make sure the one, who is applying insulation is doing it very well, he has to do this work very precisely. Because of this, it needs simply high motivation and high qualification...you need to have high quality of workmanship from the beginning.
	6	C2	**Changing requests / Value engineering**	Alam et al.(2017), Alencastro et al. (2018), De Wilde (2014), Jradi et al. (2018), Mojic et al. (2019),Xu et al.(2021), Zou et al.(2018), Doylend (2015)	x	x
	7	C3	**Quality of materials / equipment or technologies**	Calì et al. (2016), De Wilde (2014), Fedoruk et al. (2015), Zou et al. (2018), Zou et al. (2019), Yang & Zou (2014), van Heijst(2017), Qin et al.(2016), Zou et al.(2016) Mills et al.(2006), Mohammadi (2015), Coşkun (2019),Xu et al.(2021), Doylend (2015), Ipeec (2019)	Building A	A1. We saw that material specifications in the catalogs was inaccurate. This may be due to the differences between measurements in the laboratory or factory environment and the conditions in the natural environment.
	8	C4	**Bankruptcy of project stakeholders**	Baas et al.(2016), Garbuzova-Schlifter&Madlener(2016), Wang et al.(2018)	Building B Building D	B1.The subcontractor went bankrupt. D1. The first contractor of this project went bankrupt.
	9	C5	**Inconsistencies in design project and construction**	Mahdavi& Berger(2019), van Dronkelaar et al.(2016) , Zou et al.(2019), Menezes et al.(2012), Shi et al.(2019)	Building C Building D	C1.There is an ETFE roof covering an internal courtyard space. We feel that light transmission is more than given in the catalog. One of the users already measured that it is more than planned. This is not the amount we wrote in the specification. There is over-heating at the moment. However, I do not know if the contractor bought a different product or the performance of the product is different. D1. There was a place where we wanted adobe to be used in this project. Although we explained to the first contractor its purpose and how it should be, he did not apply it. The new contractor took over that job, but it wasn't what we planned. At the end something was done with smaller bricks.

Table 3.5 Problems/Limitations: Use and Operational Phase Factors

Phase	No	Code	Parameter	References	Case	Comment of Interviewee
Use and Operational	10	U1	**Occupants behavior and activities (e.g. uncertainty etc.)**	Ahn et al. (2017), Ali et al. (2020), Alencastro et al. (2018), Ascione et al.(2016) , Birchall (2011), Cali et al. (2016), De Wilde (2014), Gram-Hanssen & Georg (2018), Herrando et al. (2016), Jradi et al. (2018), Menezes et al.(2012), Mojic et al. (2019), Salehi et al. (2015), Sunikka-Blank & Galvin (2012), Zou et al. (2016), Zou et al. (2018), Doylend (2015), Ipeec (2019), Bros-Williamson et al.(2016)	Building F	F1.Some tenants open the windows and open the heating system at the same moment, so this means you have an increased energy consumption, other tenants are much more careful with it. They do not waste energy.
	11	U2	**Commissioning / continuous commissioning / soft landing**	Alam et al. (2017), Birchall (2011), De Wilde (2014), Herrando et al. (2016), Jradi et al. (2018), Mojic et al. (2019), Qin et al.(2016), Gultekin (2015), Xu et al.(2021), Doylend (2015), Ipeec (2019), Salehi et al.(2015)	Building B	B1.Commissioning runs smoothly in this building. Testing and balancing are made. They are still in contact with us, they share their problems with us.
	12	U3	**Building management**	Birchall (2011), Herrando et al. (2016), Kampelis et al. (2017), Mojic et al. (2019), Zou et al. (2016), Zou et al. (2018), Topouzi et al. (2019), Xu et al.(2021), Doylend (2015), Ipeec (2019)	Building F	F1. What we learned is if you only look at the energy consumption of the tenants we are not right, we have to look on the physical situations of our buildings.Therefore we are doing all of these measurements and monitoring.
	13	U4	**Quality of the measured data due to faulty sensors and meters**	Birchall (2011), De Wilde (2014), Jradi et al. (2018), Zou et al. (2016), Mills et al.(2006), Xu et al.(2021)	x	x

The second category is the “Unexpected events and changes” category which consists of seven factors:

- (UE1) Client, user expectations,
- (UE2) Project stakeholders,
- (UE3) Country conditions,
- (UE4) Policy, legislation, regulation,
- (UE5) Public sector building process,
- (UE6) Climate,
- (UE7) Force major events.

Appendix F represents factors belonging to “Unexpected events and changes”, “Project Management” and “Assumptions” categories, as well as cases in which these factors could be observed and articles mentioning them. “Project Management” is the third category. The following eight factors constitute this category:

- (PM1) Experience of the project stakeholders,
- (PM2) Motivation of the project parties,
- (PM3) Effective communication between project stakeholders,
- (PM4) Training of project participants,
- (PM5) Design flexibility,
- (PM6) Occupant Surveys,
- (PM7) Application of passive measures,
- (PM8) Simplicity in detailing/systems/design.

The last category is named “Assumptions.” There are ten factors in this category. These are:

- (A1) Thermal conductivity of materials,
- (A2) Building occupancy rates,

- (A3) Thermal comfort of building occupants,
- (A4) Ventilation rate,
- (A5) Internal loads,
- (A6) Temperature setpoints,
- (A7) HVAC equipment efficiency,
- (A8) Hours of use,
- (A9) Airtightness,
- (A10) Transparent surface ratio.

3.1.1.6 Web-based Survey

A questionnaire that can be analyzed using qualitative and quantitative methods is a written, online, or verbal tool to collect data from individuals or groups. Questionnaires can have open, closed, or mixed questions responses (Wilson, 2013). While respondents can answer in their own words in open-ended questions, they can either choose from one of the unordered response categories (*e.g.*, a list of building types) or ordered response categories (*e.g.*, rating scales) in closed-ended questions. Mixed questions contain both particular responses and an open-ended "Other" response (Wilson, 2013).

A Web-based survey consisting of six pages and forty-two questions is designed based on the final conceptual framework for this research. While the first page required information on project characteristics, the rest of the survey was designed by listing all the factors of the final conceptual framework with the given categories. Next, the participants were asked to evaluate the magnitude of the listed factors. The content of the survey is summarized below:

- The first page involves general information such as project type, location, and year of construction, as well as the building's electricity demand given in

simulation software and real electricity consumption, and the building's heating demand given in simulation software and real heating consumption.

- The second page is designed to collect information about the problems/limitations encountered in three stages, namely, design, construction, and use or operational stage.
- Participants are asked to evaluate factors related to project management on the third page.
- Ten factors are listed below the assumptions category for participants' evaluation on the fourth page.
- Fifth-page lists factors categorized as unexpected events and changes, and the participants were asked to evaluate these factors.
- Participant information such as education level, profession, etc. is collected on the last page. Moreover, a space is left for the participants to make any further comments at the end of the survey.

The survey instrument firstly was developed in English. Later, it was translated into German and Turkish. The participants were required to score the main survey questions on a 5-point Likert scale from very low to very high including N/A. Darko & Chan (2018) state that five point Likert scale can be interpreted without difficulty since the results provided by it are unambiguous. In this survey, open, closed, and mixed questions were used. Several methods were used to administer the survey to experts working on energy-efficient buildings like online survey tools, e-mail, face-to-face interviews, and phone interviews.

Professionals responsible for project management, energy consultancy, and commissioning in energy-efficient building projects were the selected participants for this research. The survey was sent to 1000 contacts based on cluster sampling. The contact list of the respondents was gathered through LinkedIn, Passive House Database, USGBC, and BREEAM websites. Esurveypro which allows users to create online surveys is used in this study to build the survey. The survey was carried

out between the 2nd of August and the 15th of November 2021. Survey questions can be found in Appendix G.

3.1.2 Machine Learning Application

In this research, building data collected through the web-based survey introduced in the previous section were used for developing prediction models with the machine learning classification technique. *Firstly*, the energy performance gap prediction problem was considered a binary classification problem in this study. This part deals with whether a project may observe a positive or negative performance gap regarding the electricity demand and heating demand. A positive performance gap stands for the cases in which the actual performance is more than the predicted performance and a negative performance gap stands for the cases in which the actual performance is less than the predicted performance. *After* studying binary classification problems, various ML models were created to predict the level of performance gap in percentages as a multi-class classification problems. In this part, each model was studied as a multi-class classification problem to forecast the performance gap as a percentage. Ram *et al*. (2022) state that the increasing number of classes may reduce the classification accuracy. Moreover, the problem was narrowed down due to unavailability of a large dataset in this study. Therefore, multi-class classification problems were divided into two steps: positive performance gap and negative performance gap for each category (considering electricity demand and heating demand). As a result, ML prediction models were developed with the following classes: low (0-15%), medium (15.1-40%), high (40.1-90%). Figure 3.6 represents the process described above for electricity demand. The same procedure was applied afterward for heating demand.

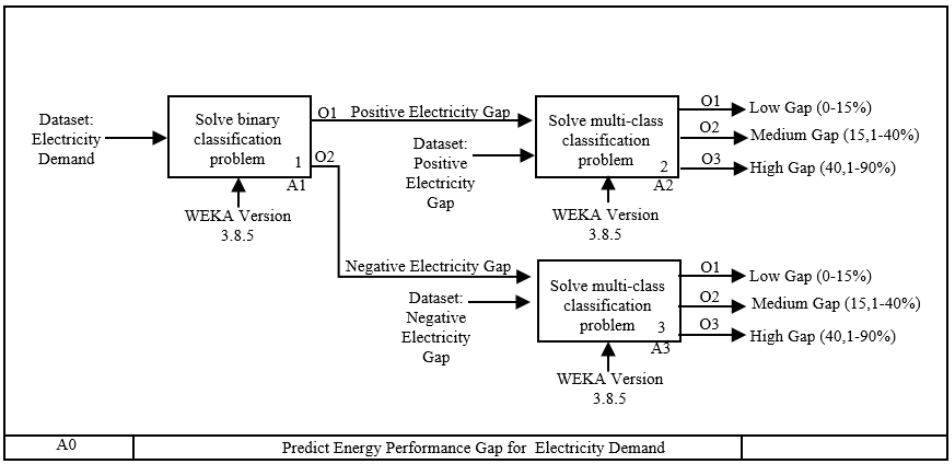


Figure 3.6 Goal of the Classification Problems Applied in the Research

A flowchart explaining the method in detail can be seen in Figure 3.7. As explained in previous sections, based on a comprehensive literature review, reasons behind the gap were identified. After conducting semi-structured interviews with the experts working on different energy-efficient buildings, and proposing a conceptual energy performance gap framework, a web-based survey was designed for data collection. After data collection, the process of developing ML prediction models requires several steps such as (i) data preprocessing, (ii) feature selection, (iii) training, and (iv) testing the ML algorithms. During data preprocessing, steps such as data cleaning, transforming the raw data into a machine-readable form, and missing value treatment were applied.

Following that, unnecessary inputs were eliminated by feature selection which is a critical task to develop a good ML model. This procedure is used to reduce the dimension of data. Because, data with reduced dimensions will provide a higher confidence in the accuracy of the model. In this research, two different feature selection methods were applied for each classification problem: (1) the Wrapper method for binary classification problems, and (2) the ChiSquared Attribute evaluator for multi-class classification problems. Moreover, Chawla *et al.* (2002)

explain that if the classes are not almost equally represented, a dataset is imbalanced. If there is a class imbalance problem, an oversampling technique which is called SMOTE was applied to create a balanced training set. Additionally, optimization of model parameters which is known as tuning is essential for the accuracy of ML model predictions (Seyedzadeh *et al.*, 2020b). Since these parameters control the learning and prediction processes and affect the performance of ML models, the models were trained by tuning the algorithms. Finally, the performance of the algorithms was tested on unseen data (test data) using several performance metrics including accuracy, kappa, precision, recall, F-measure, and AUC. The algorithm that provides the highest performance was selected as the final classifier. The following sections explain each step in more detail.

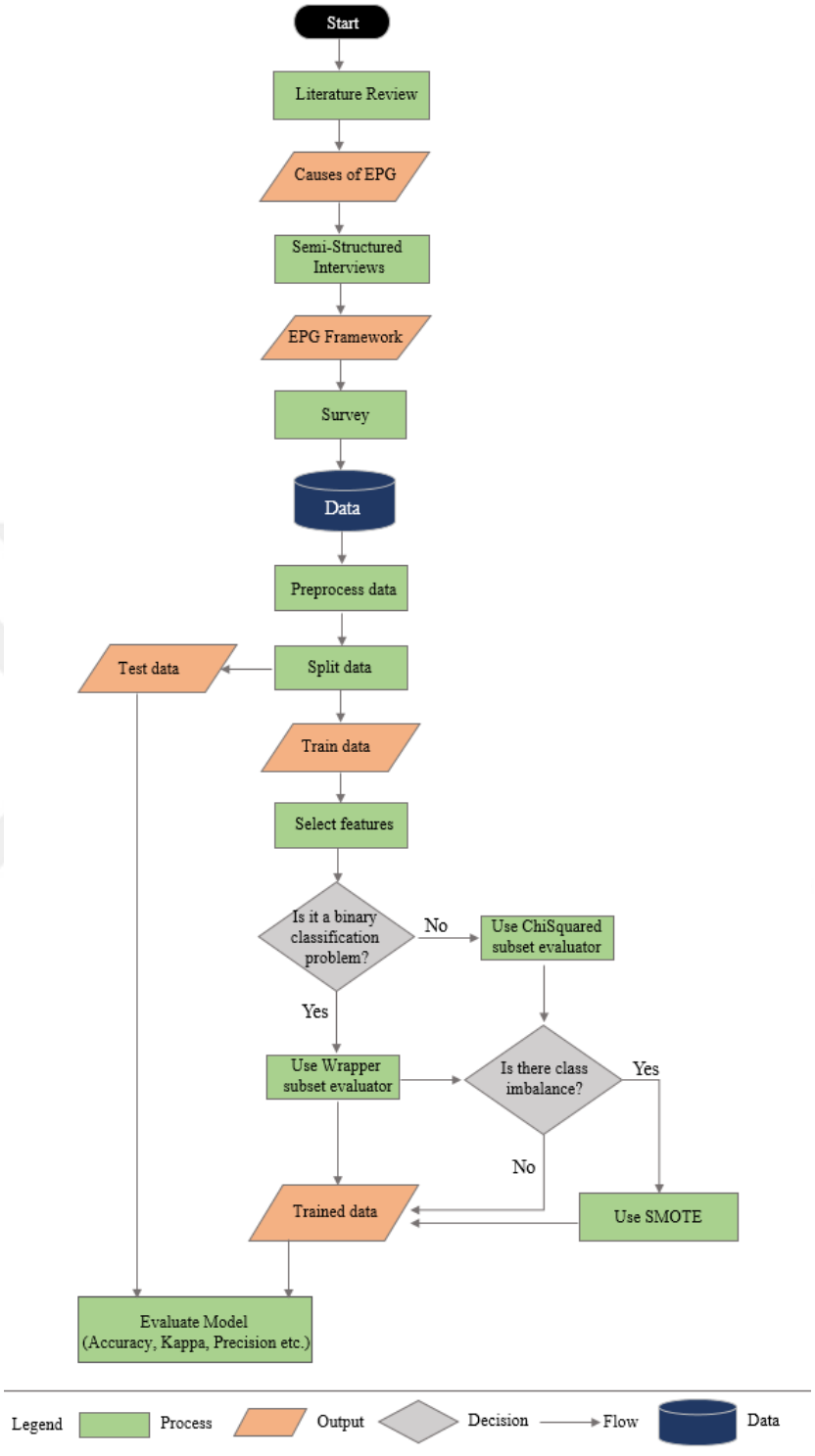


Figure 3.7 Flowchart Explaining the Method of the Research

3.1.2.1 Binary Classification to Predict Electricity EPG

The dataset for energy performance gap prediction is composed of 77 instances. However, since 2 instances which does not show a performance gap for electricity demand, they were excluded from the analysis. The final dataset regarding electricity is composed of 75 instances and 34 attributes. The input attributes are the survey responses excluding the "Assumptions" section. A list of attributes included in the analysis is presented in Appendix I. The output attribute is "Electricity Gap." Firstly, this dataset was classified into two classes. The instances whose actual performance is more than the predicted performance are classified as "positive" and the instances whose actual performance is less than the predicted performance are classified as "negative." The analysis in this part aims:

1) to predict whether the actual performance regarding electricity demand will be more than the intended performance in a building or not

2) to compare and find the algorithm which performs best among the four other classification algorithms.

After loading the ARFF file to Weka, missing values were identified. There were 1% of data in 1 attribute and more than 5% of data in 8 attributes that were missing. One expert who participated in the survey with a group of buildings as part of a research project conducted at their university did not have information about some questions such as problems with the budget, or bankruptcy of project stakeholders. Therefore, the expert did not answer related questions. Most of the missing values come from those cases. All missing values were replaced with the field mode not to lose any data. Next, data was randomized using Randomize filter so that learning algorithms would not be susceptible to the order the data arrives in.

After missing value treatment and randomizing the data, the dataset was split into a training set (80%) and a testing set (20%). The training set includes 60 instances with 31 of them being labeled as positive, and the test set includes 15 instances with 9 of them being labeled as positive. While the training set was used for feature selection and parameter tuning, the testing set was used to verify the generalization capacity of the model which was obtained after the training process. Using the testing set, the performance of the classifiers was compared based on six different performance metrics including i) Accuracy, ii) kappa statistics, iii) precision, iv) recall, v) F measure and vi) AUC. Stratified 10-fold CV were applied in 10 repeats in this section.

Stratified 10-fold cross-validation is generally suggested to estimate accuracy because of its relatively low bias and variance (Han et al., 2012). Repeating the cross-validation process 10 times and averaging the results is a standard procedure (Witten *et al*., 2011). This procedure helps to overcome the limitations such as the small size of training and validation sets, sources of randomness in the learning method, or availability of noise and outliers in the dataset (Alpaydın, 2010).

- **Configuration of Wrapper Attribute Evaluator**

Attribute selection is necessary mainly (i) to reduce the complexity of machine learning applications, and (ii) to solve the problems of poor prediction accuracy on the test sample and data overfitting.

To remove attributes using the Wrapper method, weka.filters.supervised. attribute.AttributeSelection. was used. The Wrapper method involves an *evaluation* method and a *search* method. The method starts with selecting a classifier. Classifiers map the input space into predefined classes (Rokach & Maimon, 2008). The wrapper method evaluates a specific ML algorithm for finding optimal attributes. In this research, (1) Naïve Bayes, (2) KNN, (3) SVM (4) Random

Forest algorithms were selected as the classifier one by one. “Folds” requires the number of xval folds to use when estimating subset accuracy. “Threshold” requires repeating xval if stdev of mean exceeds this value. For each algorithm, “Folds” was set to 10, and “threshold” was set to -1. The configuration method was the same for all binary classification problems.

Search methods aim to find a good subset of attributes. The “Best First” option was used as the search method. There are three direction options in Weka, including backward, forward, and bi-dimensional search. Clicking on the “Best First” search method, different attribute subsets were created choosing each direction one by one. “Search Termination” was set to values between 1 to 10 for each classifier. This setting specifies the number of consecutive non-improving nodes to allow before terminating the search to select attributes. Several experimentations were made until changing the search termination does not yield a different attribute subset. Models providing the highest classifier performance were saved in ARFF format. The selected models were reserved to compare them further using the Weka Experimenter interface. The attributes selected to train the dataset is given in Section 4.

- **Configuration of the Naïve Bayes Algorithm**

Optimization of model parameters which is known as tuning plays an essential role in the accuracy of ML model predictions (Seyedzadeh *et al.*, 2020b). Therefore, the configuration of the algorithms is necessary to tune the hyperparameters as they control the learning and prediction processes and affect the performance of ML models.

The Naïve Bayes algorithm can be found in “weka.classifiers.bayes.NaïveBayes”. The algorithm can manage binary, nominal and numeric attributes. Since all attributes are categorical in the dataset and a kernel estimator is applied on numeric

attributes to estimate class-conditional marginal densities of data, the "useKernelEstimator" setting was disabled. Supervised discretization is used to convert numeric attributes to nominal ones; therefore, this setting was also disabled. The same configuration method was applied in the binary classification of heating performance gap prediction.

- **Configuration of the KNN Algorithm**

The KNN Algorithm can be found in "weka.classifiers.lazy.IBk". The algorithm can be used on binary, nominal, and numeric data. The k value and the distance metric are two important parameters to tune. Moreover, there are three different weighting options in the settings: no distance weighting, weight by 1/ distance, and weight by 1-distance.

The "meanSquared" configuration is used when doing cross-validation for regression problems. Therefore, it was set as "False" during the analysis. There were five different distance functions used to find the neighbor: Chebyshev, Euclidean, Filtered, Manhattan, and Minkowski. In this part, different models were created by setting the K value from 1 to 10. "crossValidate" option was set to true. "No distance weighting" and "Weight by 1/distance" options were tried for "distanceWeighting". Clicking on LinearNNSearch, Manhattan Distance was set as "distanceFunction". The optimum k value was identified as 3. The same configuration method was applied in the binary classification of heating performance gap prediction.

- **Configuration of the SVM Algorithm**

The SVM Algorithm can be found in "weka.classifiers.functions.SMO". The skills of the algorithm can be seen by clicking on "Capabilities." The SVM Algorithm can be used on binary, nominal, and numeric data. Although the SVM is developed for

binary classification, extensions to the method support multi-class classification as well.

The C parameter and the type of the Kernel are two critical parameters in SVM. While high values of C will mostly penalize misclassified examples, low values only slightly penalize misclassifications, and the result might be a wrong separation (Gaspar *et al.*, 2012). The kernel is a similarity function that enables the creation of more dimensions to the dataset to work better with non-linear data. Some of the popular kernels are linear, polynomial, and gaussian. The default setting for SVM is a Polynomial kernel in Weka. It is suggested to use different kernels on the dataset to see which one works best (Brownlee, 2018a).

In this part, "weka.classifiers.meta.CVParameterSelection" which performs parameter selection by cross-validation was used to identify an optimum C parameter. This Meta classifier can be used with any classifiers. Firstly, the classifier option was set to "SMO" and "numFolds" was set to 10. Clicking on the "CVParameters" option, the search range was specified as "C 1 1000 1" in which the numbers indicate the lower bound, the upper bound, and the number of steps, respectively.

After the configuration of "CVParameterSelection", the SVM algorithm was configured. Clicking on the SMO, if "buildCalibrationModels" was set to "True," the probability of an observation belonging to each possible class label may be predicted and the number of folds used to generate training data for calibration models should be entered into the "numFolds" field. In this study, "buildCalibrationModels" was set to False.

SVM calibration window includes normalize and standardize filters. With data normalization, the attributes which have values on different scales can be brought on a scale between 0 and 1. Data standardization is another method to rescale data. A

standardized dataset will have a mean of 0 and a standard deviation of 1 without having a specific upper or lower bound. In this study, “filterType” was set to “No normalization/standardization”. “Polykernel”, “RBF kernel” and “NormalizedPolyKernel” options were tried and Polykernel provided a higher accuracy. The configuration method was the same for the binary classification of heating performance gap prediction.

- **Configuration of the Random Forest Algorithm**

NumFeatures which considers the number of attributes in each split point is an important parameter for Random Forest. “bagSizePercent” and “numIterations” are the two other important parameters to tune (Brownlee, 2018a). Sun & Chen (2020) mention that grid search is one of the most popular techniques in the field of parameter optimization. The grid search method tests each set of possible combinations of the parameters and selects the one with the highest accuracy as the final value (Ye *et al.*, 2021). In this part, the grid search method was applied to test each set of possible combinations of the “numFeatures” and “numIterations” parameters. Finally, the method selects the one with the highest accuracy as the final value. The grid search algorithm can be found in “weka.classifiers.meta.GridSearch”. Random Forest was selected as the classifier. numFeatures was written to “XProperty.” XBase, XMax, XMin, and XStep are set to 2, 3, -3, and 1, respectively. numIterations was written to “YProperty”. YBase, YMax, YMin, and YStep were set to 10, 2, 0, and 1, respectively. “Evaluation” was set to Accuracy. After the configuration of the Grid Search, Random Forest configuration default settings was used in Weka. The configuration method was the same for the binary classification of heating performance gap prediction.

3.1.2.2 Multi Class Classification to Predict Positive EPG

The number of instances in which the energy performance of buildings regarding the electricity demand was more than predictions is 40 in this dataset. After replacing all missing values with the field mode and randomizing the data, the dataset was split into a training set and a testing set. The training set includes 31 instances and the test set includes 9 instances. In the training set, 10 instances show a percentage gap between 0-15% (low), 13 instances show a percentage gap between 15.1-40% (medium), and 8 instances show a percentage gap between 40.1-90% (high). In the testing set, each class has 3 instances. He (2011) state that if the most common class is less than twice as common as the rarest class in a dataset, it would only be slightly unbalanced, that datasets with an imbalance ratio of about 10: 1 would be modestly imbalanced, and datasets with imbalance ratios above 1000: 1 would be greatly imbalanced.

When there was a class imbalance problem in a dataset, even if it was slightly unbalanced, SMOTE technique was employed to increase the number of instances in minority classes in the training set after feature selection. The second class (high) was oversampled 70% by using 5 nearest neighbors and the third class (low) was oversampled 30% by using the 5 nearest neighbors. Similarly, to the binary classification, the training set was used for feature selection and parameter tuning, and the testing set was used to verify the generalization capacity of the model which was obtained after the training process.

Using the testing set, the performance of the classifiers was compared based on six different performance metrics including accuracy, kappa statistics, precision, recall, F measure, and AUC. Due to the limited number of instances, stratified 4-fold CV was applied in 4 repeats in this section. The following sections introduce the feature selection, and configuration of the algorithms, respectively.

- **Configuration of Attribute Selection**

A supervised attribute filter, "the Attribute Selection filter" was used to apply the Chi Squared Attribute evaluator in the training set. The "Evaluator" option was set to ChiSquaredAttributeEval and the "Search" option was set to Ranker. The users can define how many attributes to retain. Certain attributes can be retained despite their rank using "Ranker" as well. Clicking on Ranker "numToSelect" option was set to values between 3 to 10 one by one. Each selection was saved in ARFF format to load to the Experimenter interface for hyperparameter tuning. The performance of each model was compared based on the pre-selected performance metrics and the group of attributes providing the best performance is presented in Section 4. The configuration method was the same for all multi-class classification problems.

- **Configuration of the Naïve Bayes Algorithm**

The Naïve Bayes algorithm is suitable both for binary and multiclass classification problems. Therefore, it can be naturally extended to solve multi-class problems. Another way to manage a multi-class problem is to decompose the problem into binary problems. García *et al.* (2015) state that OVO and OVA are the most studied binary decomposition strategies in the literature. While OVO trains a classifier to distinguish between each pair of classes, OVA trains a classifier to distinguish each class from all other classes. The author adds that the OVO strategy often out-stands OVA. Witten *et al*. (2011) mention that error-correcting output codes are another method to improve the performance of classification algorithms in multiclass learning problems. Several independent two-class problems are created in the dataset, the algorithm is run on each one, and the outputs of the resulting classifiers are combined. The authors suggest applying it even if the learning algorithm can manage multi-class problems directly. Ayhan (2019) explains that there are two extensions of the original error-correcting output codes in the WEKA workbench, random correction code (RCC) and the exhaustive correction code (ECC) methods

namely. It is stated that the ECC method becomes infeasible as the number of classes increases. Since too many classifiers must be generated when the number of classes is large, the RCC method is preferred.

In this study, in addition to the configuration represented for binary classification problems, the Naïve Bayes algorithm was applied using "weka.classifiers. meta.MultiClassClassifier". The classifier was set to "Naïve Bayes" and the decomposition method was selected by clicking on the dropdown menu of "method." The default settings were used for the "Naïve Bayes" classifier. OVA, OVO, RCC, and ECC methods were tried, respectively. "usePairwiseCoupling" was set to "True" only for the OVO technique, for the other methods the option was set to "False". The method providing better performance is presented in the results. The configuration method was the same for the multi-class classification of heating performance gap prediction.

- **Configuration of the KNN Algorithm**

The KNN algorithm can manage binary and multi-class classification problems. Therefore, in addition to the configuration of the KNN algorithm described in binary classification problem, "weka.classifiers.meta.MultiClassClassifier" was used to apply binary decomposition methods. OVA, OVO, RCC, and ECC methods were tried one by one. Additionally, in this part, to optimize the k parameter, "CVParameterSelection" was used. Firstly, "CVParameterSelection" was chosen as the classifier. After clicking on "CVParameterSelection", "weka.classifiers. lazy.IBk" was chosen as the classifier. "numFolds" was set to 4. Clicking on the "CVParameters" option, "K 1 35 1" was entered to define a search range between 1 and 35 with 1 increment. After that, clicking on IBk, the "crossValidate" option was set to true. "No distance weighting" and "Weight by 1/distance" options were tried as "distanceWeighting". Clicking on LinearNNSearch, "distance Function" was set

to Manhattan Distance. The same configuration method was applied for the multi-class classification of heating performance gap prediction.

- **Configuration of the SVM Algorithm**

Although the SVM is developed for binary classification, extensions to the method support multi-class classification as well. Similar to the previous sections, "weka.classifiers.meta.MultiClassClassifier" was used to apply binary decomposition methods. In this window, OVA, OVO, RCC, and ECC methods were selected from the drop-down menu. "usePairwiseCoupling" was set to "True" for the OVO method. For other methods, it was set to "False."

Following that, "weka.classifiers.meta.CVParameterSelection" which performs parameter selection by cross-validation was selected as a "classifier" to optimize the C parameter. Firstly, the classifier option was set to "SMO" and "numFolds" was set to 4. Clicking on the "CVParameters" option, the search range was specified as "C 1 1000 1" in which the numbers indicate the lower bound, the upper bound, and the number of steps, respectively. After the configuration of "CVParameterSelection", the SVM algorithm was configured. "buildCalibrationModels" was set to False and "filterType" was set to "No normalization/standardization." "Polykernel," "RBF kernel" and "NormalizedPolyKernel" options were tried and Polykernel provided a higher accuracy. The same configuration method was applied for the multi-class classification of heating performance gap prediction.

- **Configuration of the RF Algorithm**

"weka.classifiers.meta.MultiClassClassifier" was used to apply binary decomposition methods to configure Random Forest for multi-class classification. OVA, OVO, RCC, and ECC methods were selected as "method" one by one. "usePairwiseCoupling" was set to "True" for the OVO method. For other methods,

it was set to “False.” The “Grid Search” method was selected as the classifier to test each set of possible combinations of the “numFeatures” and “numIterations” parameters. numFeatures was written to “XProperty”. XBase, XMax, XMin, and XStep were set to 2, 3, -3, and 1, respectively. numIterations was written to “YProperty.” YBase, YMax, YMin, and YStep were set to 10, 2, 1, and 1, respectively. “Evaluation” was set to Accuracy. After the configuration of the Grid Search, default settings were used for the configuration of the Random Forest algorithm in Weka. The configuration method was the same for the multi-class classification of heating performance gap prediction.

3.1.2.3 Multi Class Classification to Predict Negative EPG

The number of instances in which the performance gap of buildings regarding the electricity demand was less than expected, is 33 in this dataset. After replacing all missing values with the field mode and randomizing the data, the dataset was split into a training and testing set. The training set includes 27 instances and the test set includes 6 instances. In the training set, 8 instances show a percentage gap between 0-15% (low), 5 instances show a percentage gap between 15.1-40% (medium), and 14 instances show a percentage gap between 40.1-90% (high). In the testing set, the number of instances in the aforementioned classes are 2, 1, and 3, respectively. Since the dataset is modestly unbalanced, after feature selection, SMOTE technique was employed to increase the number of instances in minority classes in the training set. Minority classes were oversampled using “weka.filters.supervised.instance. SMOTE”. The first class (medium) was oversampled firstly 100% using 4 nearest neighbors, after that the same class was oversampled 40% using 5 nearest neighbors. The second class (low) was oversampled 80% using 5 nearest neighbors. Similarly, to the binary classification, the training set was used for feature selection and parameter tuning, and the testing set was used to verify the generalization capacity of the model which was obtained after the training process.

Using the testing set, the performance of the classifiers was compared based on six different performance metrics including accuracy, kappa statistics, precision, recall, F measure, and AUC. Due to the limited number of instances, stratified 4-fold CV was applied in 4 repeats in this section. The feature selection, configuration of the algorithms, and results are given in the following sections, respectively.

3.1.2.4 Binary Classification to Predict Heating EPG

The final dataset regarding heating is composed of 71 instances and 34 attributes. The output attribute is "Heating Gap." Firstly, this dataset was classified into two classes. Similarly, the instances whose actual performance is more than the intended performance are classified as "positive" and the instances whose actual performance is less than the intended performance are classified as "negative." The analysis in this part aims:

1) To predict whether the actual performance regarding heating demand will be more than the predicted performance in a building or not

2) To compare and find the algorithm which performs best among the four other classification algorithms.

Like the previous section, all missing values were replaced with the field mode since it was not preferred to lose data. Later, data was randomized using Randomize filter so that learning algorithms would not be susceptible to the order the data arrives in. After missing value treatment and randomizing the data, the dataset was split into a training set and a testing set. The training set includes 57 instances with 33 of them being labeled as positive, and the test set includes 14 instances with 8 of them being labeled as positive. While the training set is used for feature selection and parameter tuning, the testing set is used to verify the generalization capacity of the model which is obtained after the training process. Using the testing set, the performance of the

classifiers was compared based on six different performance metrics including *Accuracy*, *kappa statistics*, *precision*, *recall*, *F measure*, and *AUC*. Stratified 10-fold CV were applied in 10 repeats in this section.

3.1.2.5 Multi Class Classification to Predict Positive EPG

The number of instances in which the performance gap of buildings regarding the heating demand was more than expected is 36 in this dataset. After replacing all missing values with the field mode and randomizing the data, the dataset was split into a training set and a testing set. The training set includes 27 instances and the test set includes 9 instances. In the training set, 8 instances show a percentage gap between 0-15% (low), 13 instances show a percentage gap between 15.1-40% (medium), and 6 instances show a percentage gap between 40.1-90% (high). In the testing set, each class has 3 instances. Since the training dataset is modestly imbalanced, SMOTE technique is employed to increase the number of instances in minority classes after feature selection. The first minority class was oversampled 100% and 20% using 5 nearest neighbors, second minority class was oversampled 70% using 5 nearest neighbors. Similarly, to the binary classification, the training set is used for feature selection and parameter tuning, and the test set is used to verify the generalization capacity of the model which is obtained after the training process.

Using the testing set, the performance of the classifiers was compared based on six different performance metrics including *accuracy*, *kappa statistics*, *precision*, *recall*, *F measure*, and *AUC*. Due to a smaller number of samples in the dataset, stratified 4-fold CV was applied in 4 repeats in this section. The following sections introduce the feature selection, configuration of the algorithms, and results, respectively.

3.1.2.6 Multi Class Classification to Predict Negative EPG

The number of instances in which the performance gap of buildings regarding the heating demand was more than expected is 32 in this dataset. After replacing all missing values with the field mode and randomizing the data, the dataset was split into a training set and a testing set. The training set includes 23 instances and the test set includes 9 instances. In the training set, 10 instances show a percentage gap between 0-15% (low), 5 instances show a percentage gap between 15.1-40% (medium), and 8 instances show a percentage gap between 40.1-90% (high). In the testing set, each class has 3 instances. Since the training set was modestly imbalanced, SMOTE technique is employed to increase the number of instances in minority classes after feature selection to the training set. The first minority class (high) was oversampled 30% using 5 nearest neighbors and the second minority class was oversampled 100% using 4 nearest neighbors. Similarly, to the binary classification, the training set is used for feature selection and parameter tuning, and the testing set is used to verify the generalization capacity of the model which is obtained after the training process.

Using the testing set, the performance of the classifiers was compared based on six different performance metrics including *accuracy*, *kappa statistics*, *precision*, *recall*, *F measure*, and *AUC*. Due to the limited number of instances, stratified 4-fold CV was applied in 4 repeats in this section. The following sections introduce the feature selection, configuration of the algorithms, and results, respectively.

3.1.3 The Weka Workbench

Weka which stands for the Waikato Environment for Knowledge Analysis was developed at the University of Waikato in New Zealand. The open-source software provides methods for standard data mining problems such as regression,

classification, clustering, association rule mining, and attribute selection. Using Weka, a dataset can be preprocessed, entered into a learning plan, and the resulting classifier and its performance can be analyzed. The system is written in the Java language. It is an easy-to-use collection of up-to-date machine learning algorithms and data preprocessing tools (Witten *et al.*, 2011).

The Explorer interface helps the users access all its facilities including data preprocessing, classification, clustering, association, attribute selection, and data visualization. The experimenter interface allows users to compare a variety of classification and regression techniques. Users can perform experiments and conduct statistical tests between different classifiers. The Knowledge Flow interface enables to design of configurations by dragging boxes that represent learning algorithms and data sources around the screen. Alternatively, users can write commands to implement the main method of classes in Simple CLI. Weka GUI Chooser which is presented in Figure 3.8 is the entry point into the Weka interface.

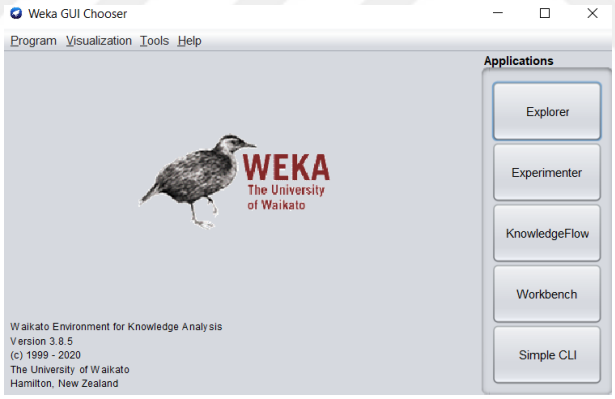


Figure 3.8 Weka GUI Chooser

Data is loaded in the ARFF format to Weka. ARFF which is an extension of the Comma Separated Values (CSV) file format, stands for Attribute Relation File Format. CSV files can be converted to ARFF format with the Weka workbench (Brownlee, 2018a). Moreover, users can install a list of official Weka packages using

Package Manager. In this study, using the Experimenter interface, several experiments were performed to tune ML algorithms, evaluate the performance of the models, and conduct Paired T-Tester statistical test. The Experimenter Interface can be seen in Figure 3.9.

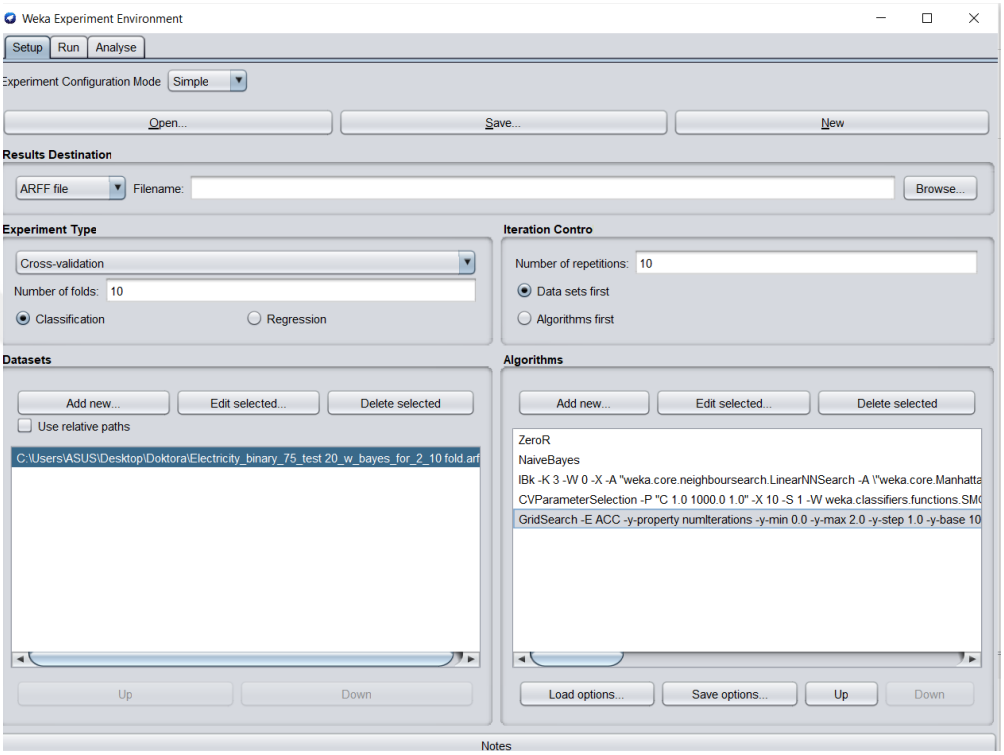


Figure 3.9 Weka Experimenter Interface

3.2 Material

One of the critical bottlenecks among the many challenges in machine learning is data collection. Data collection techniques come from different disciplines (Roh *et al.*, 2021). Moreover, understanding the data is the first step to building an application of artificial intelligence (Joshi, 2020). In this research, the main data source is a web-based survey that is designed after a comprehensive literature review and a range of semi-structured interviews on diverse types of energy-efficient buildings. This survey is designed to explore a broad range of risks encountered during the project process of various buildings and energy performance gaps related to electricity and heating demand. Therefore, in this part, firstly the examples in the dataset are introduced. Afterward, data types in the dataset were analyzed in detail.

3.2.1 Dataset used in the ML Classification Models

As explained earlier, the survey responses became the main source of the dataset to be used to train and test machine learning models. A total of 72 responses were received to the survey. The survey response rate is 7.2%. However, it was observed that 9 of these responses did not respond to either or both of the electricity and heating performance gaps and 5 of these responses showed a gap between 200-410%. Since the variance was quite large, and it was difficult to group them in one class, they were excluded from the study. Moreover, one case contained the same information as another case in terms of responses to the survey. This was due to a participant who entered two different projects into the survey. Both projects were public sector projects having the same project type. Since the projects had remarkably close performance gaps to classify them in the same group, one of them was considered a duplicate and it was removed. Since duplicate records cause an overweighting of the data values in those records, if the records are truly duplicates, only one set of them should be kept (Larose & Larose, 2014).

Furthermore, to increase the number of cases in the dataset, cases analyzed in empirical articles were included in the survey based on certain rules. In this part, studies by Herrando *et al.* (2016); Korjenic & Bednar (2012); Pegg *et al.* (2007) were reviewed. 20 cases from the literature were included in the survey. In order to avoid researcher bias, (1) Cases were entered into the survey after an in-depth review of the articles twice considering the survey questions, and (2) the survey response strategy was reviewed by two academicians. If there were no statements about a question in the survey, the response was rated as very low. If the article described a problem with the building using certain adverbs or pronouns, it was rated accordingly. If a problem was described similarly with small, short, modest, etc., it was rated as low. If a context is described with moderate, some, or average, it was rated as medium. If the reason was described using important, significant, or highly, it was rated as high. If a problem was described similarly with very much, extremely, so much, etc., it was rated as very high. Moreover, if the article analyzes more than one building, a linear relationship was assumed between the energy performance gap and the reason of the gap. Dataset used in the research is given in a tabular form in Appendix J. As a result, a total of 77 projects with both electricity and heating performance gaps were considered for machine learning models.

Figure 3.10 indicates the countries where the projects in the dataset were constructed and the construction period of the projects. Most of the cases are from Germany. Following that, there are 18 cases from Spain, 8 cases from Turkey, and 6 cases from the UK in the dataset. While 38 out of 77 cases are from the 21st century, 29 cases are from the 20th century and 10 cases are from the 18th-19th century. Moreover, 39 cases are refurbishment projects and 38 of them are newly built.

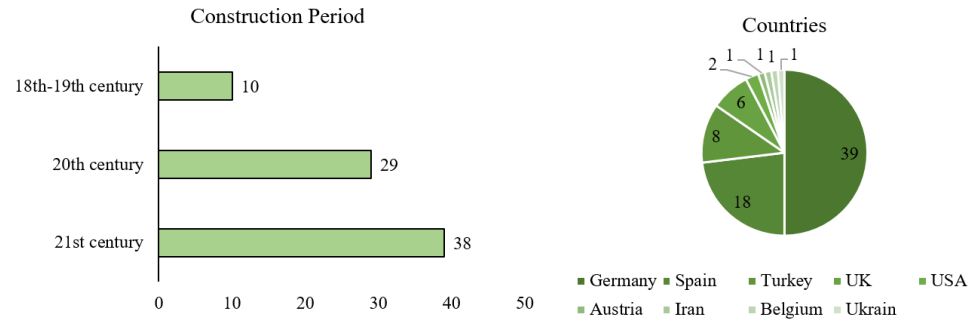


Figure 3.10 Countries and Construction Period of Cases in Dataset

Additionally, an analysis of projects based on construction year and heated floor area can be seen in Figure 3.11. There are 43 reinforced concrete buildings, 20 mixed construction, 10 masonry construction, and 4 timber frame building in the dataset. Besides, 11 of the buildings have a heated floor area of up to 500m^2. Most of the buildings have a heated floor area between 501m^2-2500m^2 (33). While 11 of the buildings have a heated floor area between 2501-6000m^2, 12 of the buildings have a heated floor area between 6001-10000m^2. Lastly, 10 of the buildings have more than 10001m^2 heated floor area.



Figure 3.11 Heated Floor Area and Construction Type

3.2.2 Analysis of Data in the Dataset

This part introduces some of the common terminologies used in machine learning applications and provides preliminary information about the data types present in the dataset. Understanding the inputs and outputs is much more essential than knowing what goes on in between in ML. The inputs are in form of concepts, instances, and attributes in data mining (Witten *et al*., 2011). Bramer (2020) explains that in general there is a dataset of examples which are also called instances. Each instance comprises the values of several variables, which are often called attributes. Larose & Larose (2014) define an attribute (or a variable) as a characteristic of an element, which undertakes different values for different elements. Each attribute can be viewed as a column in the table (Joshi, 2020). There are lots of distinct types of attributes. However, typical data mining methods are only applicable to numeric and nominal, or categorical, ones (Witten *et al*., 2011). There are 33 attributes used in ML classification problems in this study, as they can be seen in Appendix I.

Table 3.6 Types of Data (summarized from Larose & Larose, 2014)

Variables	**Qualitative**	**Categorical**	Nominal data	e.g., Marital status
			Ordinal data	e.g., Income rank
	Quantitative	**Numerical**	Interval data	e.g., Year
			Ratio data	e.g., Income

Larose & Larose (2014) explain that a variable can be either qualitative or quantitative. A qualitative variable which is also called a categorical variable, categorizes, or classifies an element based on some characteristics. Categorical variables are divided into two, namely nominal and ordinal variables. While there is no natural ordering on a nominal variable, a particular order can be given to an ordinal one (*e.g*., Likert scale). A quantitative variable is also called a numerical variable, and it has a quantifiable character that enables arithmetic to be performed

on it. Interval and ratio are two types of quantitative variables. While interval data is defined on intervals and holds no true zero, ratio data has a defined zero point. Table 3.6 summarizes different data types with some examples.

Classification is one kind of prediction, where the value to be predicted is a label or category such as “very good,” “good” or “poor” (Bramer, 2020). In this study, the dataset contains survey responses about various building characteristics, performance gaps, and risk evaluation of their project process. The main goal of the study is to predict the class of an unseen instance based on several labeled attributes.

If we examine the data types of the survey in more detail, survey responses mainly consist of categorical data. Data indicating project type (e.g., Single-family house, multi-family dwelling, etc.), construction type (e.g., Reinforced concrete, steel frame, etc.), and new building or refurbishment are nominal, whereas all survey questions that provide a Likert scale for evaluation are ordinal. However, the year of construction and heated floor area are numerical, and they are examples of interval and ratio data, respectively. The performance gap which is expressed as percentages in this study is also an example of ratio data. However, some algorithms prefer categorical rather than continuous predictors, in this case, any numerical predictors need to be partitioned into bins or bands (Larose & Larose, 2014). Moreover, using binning techniques is also useful to reduce the dimensionality and complexity of the data set and to improve the predictive power of a variable (García *et al.*, 2015). Therefore, these three numerical predictors were partitioned into bins. While the year of construction is partitioned into three bins as 18^{th} -19^{th} century, 20^{th} century, and 21^{st}-century buildings, the heated floor area is partitioned into five bins very small, small, medium, large, and very large. Moreover, the performance gap is partitioned into two bins namely positive (energy consumption is more than design prediction), and negative (energy consumption is less than design prediction) for binary classification problems. Lastly, the performance gap is partitioned into three bins namely low (0-15%), medium (15.1-40%), and high (40.1-90%) to solve multi-

classification problems. Since the accuracy of the classifier often decreases with the increase in the number of class labels in multi-class classification, the number of classes was limited to three. Moreover, since a homogenous separation of the classes with equal intervals such as 30% increased the imbalance ratio greatly, a random separation with narrower intervals is preferred due to the lack of a large amount of data.

Additionally, Joshi (2020) mentions the concept of linearity and nonlinearity for the data. If the data has a linear relationship between its input and the output, the data is called linear. For instance, when the input value decreases, the output value also decreases and vice versa. Analyzing the scatter plots provided by Weka, it was seen that the data was not linearly separable. In this study, the algorithms were selected after testing the performance of 20 classification algorithms on the training set in Weka. The selected algorithms were Naïve Bayes, KNN, Support Vector Machines, and Random Forest because of their better performance on the given dataset. Better performance of four algorithms on the given dataset can be explained with their ability to perform well with non-linearly separable data.

- **Missing values in the dataset**

It is critical to identify, mark and manage missing data when developing machine learning models to get the best performance (Brownlee, 2018a). In real-world datasets, missing values are usual (Rokach & Maimon, 2008). Reasons for missing values may be because of unanswered questions on a survey, privacy concerns of survey participants, problems with the measuring mechanism, or values that were obviously wrong and had to be removed (Rivera, 2020). Missing values can be handled during data preprocessing which involves dimension reduction (*i.e.*, feature selection and sampling); data cleansing (i.e., removing noise or outliers, as well as dealing with missing values); and data transformation (*i.e.*, discretization of numerical attributes and attribute extraction) (Rokach & Maimon, 2008).

Missing data keeps plaguing data analysis methods (Larose & Larose, 2014). Moreover, this situation can complicate not only induction (a training set where some of its values are missing) but also the classification of a new instance that has certain missing values (Rokach & Maimon, 2008).

Although missing observations can be discarded altogether, since usually there are not big enough samples to be able to afford this, and the non-missing entries contain information, losing data is not preferred (Bastanlar & Ozuysal, 2014). Besides, since the pattern of missing values may be systematic, simply deleting the records with missing values would result in a biased subset of the data (Larose & Larose, 2014).

The authors list some of the common guidelines to choose replacement values for missing values as follows:

- Replacing the missing value with some constant, specified by the researcher
- Replacing the missing value with the field mode (for categorical variables) or mean (for numeric variables)
- Replacing the missing values with a value generated randomly from the observed distribution of the variable
- Replacing the missing values with imputed values with reference to the other features of the record

In this study, all missing observations were replaced with the field mode. Additionally, since one instance contained the same information as another instance in terms of the inputs, and the outputs in the dataset, it was removed.

CHAPTER 4

RESULTS

In this chapter, the results of the research in electricity and heating performance gap prediction are given, respectively.

4.1 Electricity Performance Gap Prediction

This part starts with explaining the result of binary classification of the electricity performance gap and continues with presenting the results of multi-class classification of the electricity performance gap.

4.1.1 Binary Classification of Electricity Performance Gap

Binary classification aims to predict whether the actual performance regarding electricity demand will be more than the predicted performance in a building or not. In this part, first, selected features using Weka Version 3.8.5 are presented. Following that, a performance evaluation of the algorithms is given.

4.1.1.1 Selected Attributes with Wrapper Attribute Evaluator

Table 4.1 indicates the attributes chosen with Wrapper subset evaluator using Naïve Bayes Classifier. When the number of folds was set to 10, the direction of the search was set to forward, and search termination was set to 2, Weka selected the following attributes. A more detailed information about the configuration of the Wrapper subset evaluator is given in Section 3.1.2.1.

Table 4.1 Attributes Selected with Naïve Bayes Classifier for Electricity Gap

No	Wrapper Attribute Selection: Classifier: Naïve Bayes, folds:10
1	Uncertainty related to simulation inputs
2	Simplicity of construction details /systems design

The attributes chosen with Wrapper subset evaluator using KNN Classifier are shown in Table 4.2. The following attributes were given by Weka when the number of folds was set to 10, the direction of the search was set to forward, and search termination was set to 2.

Table 4.2 Attributes Selected with KNN Classifier for Electricity Gap

No	Wrapper Attribute Selection: Classifier: KNN, folds:10
1	Construction type
2	Shortcomings of modeling, software, or calculation methodology
3	Quality of workmanship
4	Quality of commissioning/continuous commissioning/soft landing
5	Occupant surveys (*e.g.*, comprehensiveness *etc*.)
6	Availability of passive design strategies for indoor comfort

Table 4.3 lists the attributes selected with Wrapper subset evaluator using SVM Classifier. When the number of folds was set to 10, the direction of the search was set to backward, and search termination was increased step by step from 1 to 4, attributes were provided by the software as follows.

Table 4.3 Attributes Selected with SVM Classifier for Electricity Gap

No	Wrapper Attribute Selection: Classifier: SVM, folds:10
1	Heated floor area
2	Shortcomings of modeling, software, or calculation methodology
3	Quality of workmanship
4	Inconsistencies between design project and construction
5	Occupant related uncertainty
6	Quality of commissioning/continuous commissioning/soft landing
7	Simplicity of construction details /systems design
8	Problems in country conditions (*e.g.*, Availability of material *etc*.)
9	Uncertainty related to Policy/Legislation/Regulation

The attributes chosen with Wrapper subset evaluator using Random Forest Classifier are shown in Table 4.4. The following attributes were given when the number of folds was set to 10, the direction of the search was set to forward, and search termination was set to 1. Moreover, Appendix I presents a list of all attributes selected with different feature selection methods in different classification models.

Table 4.4 Attributes Selected with Random Classifier for Electricity Gap

No	Wrapper Attribute Selection: Classifier: Random Forest, folds:10
1	Shortcomings of modeling, software, or calculation methodology
2	Uncertainty related to Policy/Legislation/Regulation
3	Uncertainty related to climate (*e.g.*, weather and climate extremes)

4.1.1.2 Performance Evaluation for Electricity Gap Prediction

The results for all classification models are given in Table 4.5. 10-fold cross-validation is repeated 10 times and the average value of the performance metrics including *accuracy*, *kappa statistic*, *precision*, *recall*, *F measure*, and *AUC*, is reported. Looking at the testing set, KNN, SVM, and Random Forest have the top three accuracies (79.00%, 75.00%, and 73.50%, respectively) among all the

classifiers. Naïve Bayes has the lowest accuracy (66.00%) of all the classifiers. The agreement between predicted and observed categorizations of a dataset is measured by the Kappa statistic (Witten *et al.*, 2011). The interpretation given by Mun *et al.* (2019) was applied in this research. A value below 0.2 was viewed as poor, 0.2–0.4 was Fair, 0.4–0.6 Moderate, 0.6–0.8 Substantial, and over 0.8 were defined as Good. While KNN and SVM show a substantial agreement between the predicted and actual values in the dataset, Naïve Bayes and Random Forest represent a moderate agreement.

Table 4.5 Performance Evaluation of the Classifiers

Performance Gap Prediction regarding Electricity Demand								
Binary classification	**Test set results**				**Training set results**			
	Models				**Models**			
Performance Metric	**Naïve Bayes**	**KNN**	**SVM**	**Random Forest**	**Naïve Bayes**	**KNN**	**SVM**	**Random Forest**
Accuracy (%)	66.00	**79.00**	**75.00**	73.50	**70.83**	**80.33**	**77.67**	**73.50**
Kappa	0.53	**0.65**	**0.61**	**0.59**	**0.42**	**0.61**	**0.55**	**0.47**
Precision	**0.81**	**0.82**	**0.83**	**0.81**	**0.74**	**0.82**	**0.80**	**0.76**
Recall	*0.68*	*0.86*	*0.79*	*0.81*	*0.73*	*0.85*	*0.77*	*0.78*
F-Measure	**0.90**	**0.89**	**0.91**	**0.87**	0.71	**0.81**	**0.77**	**0.74**
AUC	**0.80**	0.62	**0.72**	**0.67**	**0.67**	**0.87**	**0.78**	**0.81**

Accuracy: The ratio of the number of correct predictions out of all predictions
Kappa: The agreement of prediction with the true class
Precision: The proportion of instances classified as positive that are really positive
Recall: The proportion of positive instances that are correctly classified as positive
F-Measure: The harmonic mean of the precision and recall
AUC: How well a classifier can separate positive and negative instances
Bold font: The classifiers which are significantly better than the base classifier
Italic font: The classifiers which are significantly worse than the base classifier

Precision represents the proportion of instances classified as positive that are really positive. The average precision value that indicates the positive predicting power of SVM classifiers is “0.83” and it is the highest among all the classifiers. KNN,

Random Forest, and Naïve Bayes provide an average precision value of “0.82,” “0.81,” and “0.81”, respectively. Besides, recall represents how often a positive class instance in the dataset was predicted as a positive class instance by the classifier (He & Ma, 2011). A recall value of 1 means that every positive instance is classified correctly. The average recall value shows that the success of the KNN algorithm to identify projects with a positive class instance is “0.86.” Random Forest, SVM, and Naïve Bayes provide lower average recall values than the KNN algorithm, “0.81,” “0.79,” and “0.68”, respectively. Since there is a tradeoff between precision and recall, the F score is quite useful as it is a combination of recall and precision. F score is the harmonic mean of the precision and recall (Mortaz, 2020). An F score of 1 indicates perfect precision and recall. On the testing set, the SVM algorithm has the highest F score of 0.91. Slightly lower F scores are seen for Naïve Bayes (0.90), KNN (0.89), and Random Forest (0.87) classifier.

AUC corresponds to the probability that a randomly drawn positive instance will be ranked higher than a randomly drawn negative instance (Kontokosta & Tull, 2017). While an AUC equal to 0.5 represents a model that does not predict any better than random guessing, an AUC of 1 indicates a perfect prediction. In this research, the following rules is used to interpret AUC: value >0.9 excellent, 0.8-0.9 very good, 0.7-0.8 good, 0.6-0.7 average, 0.5-0.6 poor. The Naïve Bayes algorithm has the highest AUC of 0.80 on the testing set, which is a good result. Larger the AUC value, better the performance. SVM provided an AUC of 0.72. Other classifiers showed average AUC values. While Random Forest provided an AUC of 0.67, KNN provided an AUC of 0.62.

Finally, the results of the algorithm performances were compared with the baseline classifier. ZeroR is selected as the baseline classifier. It is a simple classifier that always classifies to the largest class in Weka. ZeroR provided an accuracy of 65.00%, a Kappa value of 0.40, a precision of 0.65, a recall of 1.00, an F measure of 0.81, and an AUC of 0.50. While using Paired T-Tester statistical test, the

“Significance level” was set to 0.05. In this way, the differences between any of the pairwise algorithm performance comparisons are analyzed to see if there is a statistical significance with a confidence of 95%. Statistical significance comparison is reflected in the results in Table 4.5. The classifiers who are significantly better than the base classifier are written in bold and the classifiers who are significantly worse are written in italic. Otherwise, we cannot be sure whether there is a statistically significant difference or not. Considering testing results, while SVM is statistically significant in five performance metrics, Random Forest and KNN are statistically significant in four performance metrics and Naïve Bayes is statistically significant in three performance metrics. Moreover, the results indicate that all classifiers are statistically insignificant in the recall metric.

While SVM outperforms KNN in three performance metrics (Precision, F measure, and AUC), KNN outperforms SVM in Accuracy, Kappa statistic, and Recall. Since KNN exhibits the best accuracy among the four classification algorithms, and it has also a simpler model, KNN is the suggested model due to its highest accuracy. Figure 4.1 represents the predictive accuracy of the KNN classifier on the unseen (test) data and the final model.

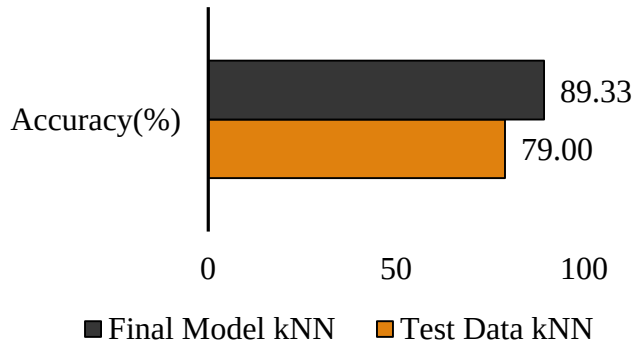


Figure 4.1 Performance of the Final Model: KNN Classifier

4.1.2 Multi Class Classification of Electricity Performance Gap

The models represented in this part aim to predict the energy performance gap in buildings regarding electricity demand as percentages. Three classes were created in two directions (positive and negative): low (0-15%), medium (15.1-40%), high (40.1-90%). The following parts introduce the results of the positive performance gap and negative performance gap prediction problems, respectively. In each part, first, selected features using Weka Version 3.8.5 are presented. Following that, a performance evaluation of the algorithms is given.

4.1.2.1 Selected Attributes for Positive EPG

Table 4.6 lists the attributes selected using ChiSquaredAttributeEval in WEKA. 3 to 10 attributes that were ranked highest, were retained one by one. Each subset of features was evaluated using the Experimenter interface. The performance of each model was compared based on the performance metrics and the subset of features that offered higher performance for most classifiers is represented. Appendix I presents a list of all attributes selected by the Chi Squared method in different classification models. A more detailed information about the configuration of the method is given in Section 3.1.2.2.

Table 4.6 Attributes Selected using ChiSquaredAttributeEval for Positive Electricity Gap

No	ChiSquaredAttributeEval, Ranker search method
1	Shortcomings of modeling, software, or calculation methodology
2	Building management problems
3	Quality of materials/ equipment/technologies
4	Motivation of the project stakeholders
5	Experience of the project stakeholders
6	Design problems (*e.g.*, Poorly designed building and building systems *etc*.)
7	Inconsistencies between design project and construction

4.1.2.2 Performance Evaluation for Positive Electricity Gap Prediction

Considering the testing set results of the models using ChiSquaredEval features, Naïve Bayes (OVO) provided the highest accuracy of 77.08% and Random Forest (ECC) achieved the second highest accuracy (73.96%) of all the classifiers as can be seen in Table 4.7. SVM (OVO) and KNN (OVA) achieved 65.63% and 60.42% average classification accuracy, respectively. While Naïve Bayes and Random Forest show a substantial agreement between the predicted and actual values in the dataset, other classifiers show a moderate agreement.

The average precision value that indicates the positive predicting power of Naïve Bayes classifiers is 100%. This means that all positive instances are classified correctly. Random Forest, SVM, and KNN show the much lower precision value of "0.67," "0.67," and "0.43", respectively. The average recall value shows that the success of the Random Forest algorithm to identify projects with a positive class instance is highest (0.75). SVM, Naïve Bayes, and KNN show a lower average recall value (0.67) than the Random Forest algorithm. Naïve Bayes, SVM, and Random Forest have an F score of 1 which indicates perfect precision and recall. A lower F score is seen for the KNN classifier (0.83). On the testing set, Random Forest and SVM have an excellent AUC value of 0.92 and 0.90, respectively. While Naïve Bayes shows a very good AUC value (0.83), KNN has a good value of AUC (0.75).

Table 4.7 Performance Evaluation of the Classifiers

Positive Performance Gap Prediction regarding Electricity Demand: ChiSquaredEval								
Multi Class classification	**Test set results**				**Training set results**			
	Models				**Models**			
Performance Metric	**Naïve Bayes**	**KNN**	**SVM**	**Random Forest**	**Naïve Bayes**	**KNN**	**SVM**	**Random Forest**
Accuracy (%)	**77.08**	**60.42**	**65.63**	**73.96**	**81.18**	**82.92**	**69.38**	**78.47**
Kappa	**0.65**	0.42	0.49	**0.62**	**0.72**	**0.74**	**0.54**	**0.68**
Precision	**1.00**	0.43	0.67	0.67	**0.69**	**0.86**	**0.62**	**0.76**
Recall	**0.67**	**0.67**	**0.67**	**0.75**	**0.83**	**0.82**	**0.66**	**0.72**
F-Measure	**1.00**	0.83	**1.00**	**1.00**	0.73	**0.84**	0.62	0.76
AUC	0.83	0.75	0.90	**0.92**	**0.87**	**0.93**	**0.80**	**0.90**

Table 4.7 reflects the statistical significance comparison of the classifiers with the baseline algorithm based on Paired T-Tester statistical test with a confidence of 95%. Zero R provided an accuracy of 8.33%, a kappa value of 0.00, a precision of 0.17, a recall of 0.33, an F measure of 0.50, and an AUC of 0.50. As described in the previous section, if a classifier's performance was significantly better than the baseline algorithm, the result was written in bold in the table. If the result is written in italic, the classifier is significantly worse than the baseline algorithm. Otherwise, we cannot be sure whether there is a statistically significant difference or not. Considering the performance of the models on the unseen data using ChiSquaredEval features, both Naïve Bayes and Random Forest are significantly better than the baseline algorithm in five performance metrics. While Random Forest outperforms Naïve Bayes in Recall and AUC, Naïve Bayes outperforms Random Forest in other performance metrics (Accuracy, Kappa statistic, Precision). Since there is no class imbalance in the dataset, the best performance is selected based on the best accuracy. Therefore, Naïve Bayes is the suggested algorithm. The predictive accuracy of the Naïve Bayes classifier on the unseen data and the final model can be found in Figure 4.2.

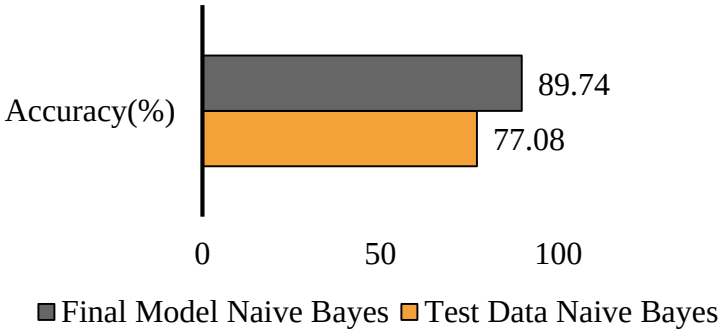


Figure 4.2 Performance of the Final Model: Naïve Bayes

4.1.2.3 Selected Attributes for Negative EPG

Table 4.8 lists the attributes selected using ChiSquaredAttributeEval in WEKA. Like the other sections, 3 to 10 of the attributes that were ranked highest, were retained one by one. Each subset of features was evaluated using the Experimenter interface. The performance of each model was compared based on the performance metrics and the subset of features that offered higher performance for most classifiers is represented. A more detailed information about the configuration of the method is given in Section 3.1.2.2.

Table 4.8 Attributes Selected using ChiSquaredAttributeEval for Negative Electricity Gap

No	ChiSquaredAttributeEval, Ranker search method
1	Shortcomings of modeling, software or calculation methodology
2	Uncertainty related to simulation inputs
3	Motivation of the project stakeholders
4	Quality of materials/ equipment/ technologies
5	Effective communication between project stakeholders
6	Occupant related uncertainty
7	Quality of workmanship

4.1.2.4 Performance Evaluation for Negative Electricity Gap Prediction

In this part, precision and F-Measure were indeterminable when evaluating the model on unseen data. Therefore, the results are based on the whole data set that also includes unseen data. Considering the whole dataset results of the models created using ChiSquaredEval features, the prediction accuracy of Naïve Bayes (OVO) is 83.85% and it is the highest of all the algorithms. While KNN (OVA) and Random Forest (RCC) achieved 82.29% and 79.17% average classification accuracy, SVM (OVA) has the least accuracy (78.13%). Considering kappa statistics, all classifiers show a substantial agreement between the predicted and actual values in the dataset.

Naïve Bayes classifiers can achieve a precision of 100%. While Random Forest and SVM have a precision value of 0.92, KNN has a precision value of 0.90. The average recall value shows that the success of the KNN algorithm to identify projects with a positive class instance is 0.94. While Naïve Bayes has an average recall value of 0.90, Random Forest and SVM have an average recall value of 0.84. The KNN has the highest F score of 0.94 amongst all the other classifiers. SVM and Random Forest have an F value of 0.92 and Naïve Bayes has an F value of 0.90. All algorithms have an excellent AUC value. Table 4.9 represents the comparison of the performance metrics for all classification models.

Table 4.9 Performance Evaluation of the Classifiers

Negative Gap	**ChiSquaredEval**			
Multi Class classification	**Whole data set**			
	Models			
Performance Metric	**Naïve Bayes**	**KNN**	**SVM**	**Random Forest**
Accuracy (%)	**83.85**	**82.29**	**78.13**	**79.17**
Kappa	**0.76**	**0.73**	**0.67**	**0.69**
Precision	**1.00**	**0.90**	**0.92**	**0.92**
Recall	0.90	**0.94**	0.84	**0.84**
F-Measure	**0.90**	**0.94**	**0.92**	**0.92**
AUC	**0.98**	**0.98**	**0.97**	**0.97**

A comparison of the statistical significance of the models with the baseline algorithm indicates that KNN and Random Forest are statistically significant in all metrics. Zero R provided an accuracy of 31.33%, a kappa value of 0.00, a precision of 0.25, a recall of 0.25, an F measure of 0.40, and an AUC of 0.50. According to the results in Weka, we cannot say whether Naïve Bayes is statistically significant or not in the recall metric. However, it outperforms KNN in Accuracy, Kappa statistic, and Precision. KNN outperforms Naïve Bayes in Recall and F measures. Since there is no class imbalance problem, the best model is selected based on the highest accuracy. Therefore, as it is presented in Figure 4.3, Naïve Bayes is the suggested classifier for the negative electricity gap prediction.

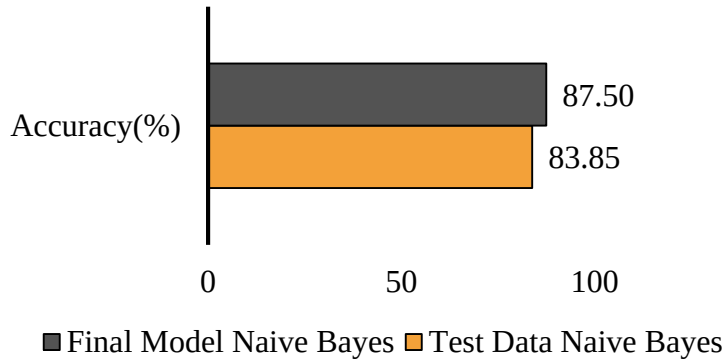


Figure 4.3 Performance of the Final Model: Naïve Bayes

4.2 Heating Performance Gap Prediction

This part starts with explaining the result of binary classification of the heating performance gap and continues with presenting the results of multi-class classification of the heating performance gap.

4.2.1 Binary Classification of Heating Performance Gap

Binary classification aims to predict whether the actual performance regarding heating demand will be more than the predicted performance in a building or not. In this part, first, selected features using Weka Version 3.8.5 are presented. Following that, a performance evaluation of the algorithms is given.

4.2.1.1 Selected Attributes with Wrapper Attribute Evaluator

Table 4.10 lists the attributes selected with Wrapper subset evaluator using Naïve Bayes Classifier. When the number of folds was set to 10, the direction of the search was set to forward, and search termination was set to 1, attributes were provided by Weka as follows. A more detailed information about the configuration of the method is given in Section 3.1.2.1.

Table 4.10 Attributes Selected with Naïve Bayes Classifier for Heating Gap

No	Wrapper Attribute Selection: Classifier: Naïve Bayes, folds:10
1	Design problems (*e.g.* poorly designed building and building systems *etc*.)
2	Quality of workmanship
3	Design flexibility (*e.g.* planning an alternative system)

The attributes chosen with the Wrapper subset evaluator using KNN are shown in Table 4.11. The following two attributes were given by the software when the number of folds was set to 10, the direction of the search was set to forward, and search termination was set to 1.

Table 4.11 Attributes Selected with KNN Classifier for Heating Gap

No	Wrapper Attribute Selection: Classifier: KNN, folds:10
1	New or Refurbishment
2	Quality of workmanship

Table 4.12 indicates the attributes chosen with the Wrapper subset evaluator using SVM Classifier. When the number of folds was set to 10, the direction of the search was set to forward, and search termination was set to 1, Weka selected the following attributes.

Table 4.12 Attributes Selected with SVM Classifier for Heating Gap

No	Wrapper Attribute Selection: Classifier: SVM, folds:10
1	Construction type
2	Design problems (*e.g.*, Poorly designed building and building systems *etc*.)
3	Quality of workmanship
4	Motivation of project stakeholders

The attributes chosen with Wrapper subset evaluator using Random Forest Classifier are shown in Table 4.13. The following attributes were given when the number of folds was set to 10, the direction of the search was set to forward, and search termination was set to 7.

Table 4.13 Attributes Selected with Random Forest Classifier for Heating Gap

No	Wrapper Attribute Selection: Classifier: Random Forest, folds:10
1	Heated floor area
2	Construction type
3	Design problems (*e.g.*, poorly designed building and building systems *etc*.)
4	Quality of workmanship
5	Occupant surveys (*e.g.*, comprehensiveness *etc*.)

4.2.1.2 Performance Evaluation for Heating Gap Prediction

Performance comparison of the classifiers is given in Table 4.14. Since features selected with Naïve Bayes provided higher performance for all classifiers, the results are based on the selected features using the Wrapper method with the Naïve Bayes classifier. Moreover, higher performance was achieved when applying the distance weight method using Weight by 1/distance during the hyperparameter tuning of the KNN. 10-fold cross-validation is repeated 10 times and the average of the performance is reported.

Looking at the testing set, Naïve Bayes has the highest accuracy of 72.50%. SVM shows an average accuracy of 68.50%. KNN and Random Forest achieve an accuracy of 68.00% and 67.50%, respectively. On the testing set, while Naïve Bayes shows a substantial agreement (0.62) between the predicted and actual values in the dataset, Random Forest (0.56), SVM (0.55), and KNN (0.54) represent a moderate agreement.

Table 4.14 Performance Evaluation of the Classifiers

Performance Gap Prediction regarding Heating Demand								
Binary classification	**Test set results**				**Training set results**			
	Models				**Models**			
Performance Metric	**Naïve Bayes**	**KNN**	**SVM**	**Random Forest**	**Naïve Bayes**	**KNN**	**SVM**	**Random Forest**
Accuracy (%)	**72.50**	68.00	**68.50**	67.50	**78.63**	**71.63**	**76.80**	**71.60**
Kappa	**0.62**	**0.54**	**0.55**	**0.56**	**0.53**	**0.37**	**0.49**	**0.38**
Precision	**0.74**	**0.70**	**0.70**	**0.72**	**0.77**	**0.72**	**0.75**	**0.72**
Recall	*0.90*	*0.87*	*0.87*	*0.84*	*0.94*	*0.89*	*0.94*	*0.87*
F-Measure	**0.90**	0.86	0.86	**0.89**	**0.84**	**0.78**	**0.83**	**0.78**
AUC	0.64	**0.67**	**0.66**	**0.69**	**0.65**	**0.66**	**0.74**	**0.66**

The average precision value that indicates the positive predicting power of Naïve Bayes classifiers is “0.74” and it is the highest among other classifiers. Random Forest, KNN, and SVM show an average precision value of “0.72,” “0.70,” and “0.70”, respectively. The average recall value shows that the success of the Naïve Bayes algorithm to identify projects with a positive class instance is “0.90.” SVM, KNN, and Random Forest show high recall values as well, “0.87,” “0.87,” and “0.84”, respectively. Naïve Bayes classifier has the highest F score of 0.90. Random Forest (0.89) has a slightly lower F score than Naïve Bayes. KNN and SVM achieve an F score of 0.86. All algorithms show average AUC values.

A comparison of the algorithm performances with the baseline classifier (Zero R) using Paired T-Tester statistical test in Weka can be seen in Table 4.14. The classifiers who were significantly better than the baseline classifier were written in bold and the classifiers who were significantly worse than the baseline classifier are written in italic. Otherwise, we cannot be sure whether there is a statistically significant difference or not. According to the results, Naïve Bayes, SVM, and Random Forest are statistically significant in four metrics. All classifiers are not statistically significant in the recall metric. While Naïve Bayes is statistically significant in accuracy, kappa statistics, precision, and F measure, Random Forest is statistically significant in kappa statistics, precision, F measure, and AUC. Since Naïve Bayes outperforms Random Forest in five performance metrics including accuracy, it is the suggested model for the heating performance gap prediction in binary classification problems. Figure 4.4 represents the predictive accuracy of the Naïve Bayes model on unseen data and the final model.

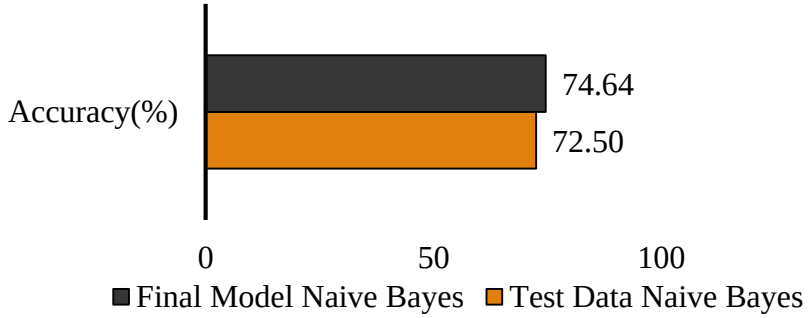


Figure 4.4 Performance of the Final Model: Naïve Bayes

4.2.2 Multi Class Classification of Heating Gap

The models represented in this section aims to predict the energy performance gap in buildings regarding heating demand as percentages. The following sections firstly introduces the models to predict the performance gap in buildings for the cases that might consume more than the design predictions. After that, the models to predict the performance gap for the cases that might consume less than the design predictions are introduced.

4.2.2.1 Selected Attributes for Positive EPG

Table 4.15 lists the attributes selected using ChiSquaredAttributeEval in WEKA. 3 to 8 attributes that were ranked highest, were retained one by one. Each subset of features was evaluated using the Experimenter interface. The performance of each model was compared based on the performance metrics and the subset of features that offered higher performance for most classifiers is represented. A more detailed information about the configuration of the method is given in Section 3.1.2.2.

Table 4.15 Attributes Selected using ChiSquaredAttributeEval for Positive Heating Gap

No	ChiSquaredAttributeEval, Ranker search method
1	Shortcomings of modeling, software, or calculation methodology
2	Heated Floor Area
3	Uncertainty related to simulation inputs
4	Effective communication between project stakeholders
5	Project type

4.2.2.2 Performance Evaluation for Positive Heating Gap Prediction

Considering the performance metrics of the models on unseen data, SVM (ECC) provided the highest accuracy of 76.04%. Random Forest (OVA), Naïve Bayes (RCC), and KNN (OVO) achieved 69.79%, 59.38%, and 55.21% average classification accuracy, respectively. On the testing set, while SVM shows a substantial agreement between the predicted and actual values in the dataset, other classifiers show a moderate agreement. The average precision value that indicates the positive predicting power of SVM is 50%. Naïve Bayes and KNN show a precision value of "0.44" and Random Forest has a precision value of 0.43. The average recall value shows that the success of the SVM classifier to identify projects with a positive class instance is highest (0.92). Other classifiers show a lower average recall value (0.83) than the SVM classifier. Random Forest has an F score of 1 which indicates perfect precision and recall on the testing set. Other classifiers achieve an F score of 0.92. Random Forest, KNN, and Naïve Bayes have an excellent AUC value (1.00) on the testing set. SVM shows a lower, however still an excellent AUC value (0.96). Performance evaluation of the classifiers based on ChiSquaredEval features can be seen in Table 4.16.

Table 4. 16 Performance Evaluation of the Classifiers

Positive Performance Gap Prediction regarding Heating Demand: ChiSquaredEval								
Multi Class classification	**Test set results**				**Training set results**			
	Models				**Models**			
Performance Metric	**Naïve Bayes**	**KNN**	**SVM**	**Random Forest**	**Naïve Bayes**	**KNN**	**SVM**	**Random Forest**
Accuracy (%)	**59.38**	**55.21**	**76.04**	**69.79**	**62.71**	**58.12**	**65.35**	**68.40**
Kappa	**0.45**	**0.42**	**0.68**	**0.57**	**0.44**	**0.37**	**0.48**	**0.53**
Precision	0.44	0.44	0.50	0.43	**0.57**	**0.50**	**0.55**	**0.71**
Recall	**0.83**	**0.83**	**0.92**	**0.83**	0.75	0.67	0.69	0.69
F-Measure	**0.92**	**0.92**	**0.92**	**1.00**	**0.63**	**0.56**	**0.59**	**0.67**
AUC	**1.00**	**1.00**	**0.96**	**1.00**	**0.80**	**0.76**	**0.73**	**0.80**

A comparison of the classifier performances with the baseline classifier (Zero R) using Paired T-Tester statistical test is reflected in Table 4.16. Zero R provided an accuracy of 8.33%, a kappa value of 0.00, a precision of 0.17, a recall of 0.33, an F measure of 0.50, and an AUC of 0.50. The classifiers that are significantly better than the baseline classifier are written in bold. In other cases, we cannot be sure whether there is a statistically significant difference or not. All classifiers are significantly better than the baseline classifier in five metrics as can be seen in Table 4.16. Since SVM outperforms Random Forest in accuracy, kappa statistics, precision, and recall on unseen data, it is the suggested classifier for the final model. Figure 4.5 indicates the predictive accuracy of the SVM classifier on unseen data and the final model.

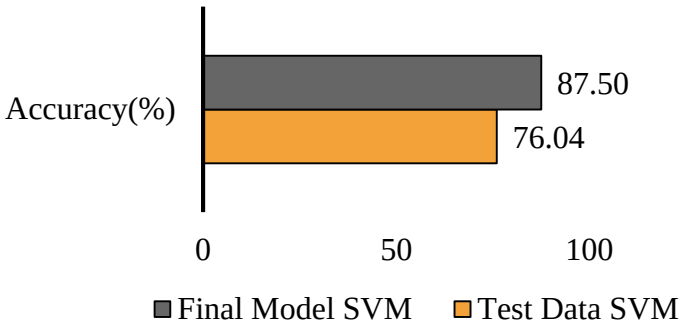


Figure 4.5 Performance of the Final Model: SVM

4.2.2.3 Selected Attributes for Negative EPG

Table 4.17 lists the attributes selected using ChiSquaredAttributeEval in WEKA. 3 to 8 attributes that were ranked highest, were retained one by one. Each subset of features was evaluated using the Experimenter interface. The performance of each model was compared based on the performance metrics and the subset of features that offered higher performance for most classifiers is represented. A more detailed information about the configuration of the method is given in Section 3.1.2.2.

Table 4.17 Attributes Selected using ChiSquaredAttributeEval for Negative Heating Gap

No	ChiSquaredAttributeEval, Ranker search method
1	Uncertainty related to simulation inputs
2	Shortcomings of modeling, software, or calculation methodology
3	Simplicity of construction details /systems design
4	Quality of commissioning/continuous commissioning/soft landing
5	Problems in country conditions (*e.g.*, Availability of material, skilled personnel *etc*.)
6	Availability of passive design strategies for indoor comfort

4.2.2.4 Performance Evaluation for Negative Heating Gap Prediction

In this part, precision and F-Measure were indeterminable when evaluating the model on unseen data. Therefore, the results are based on the whole data set that also includes unseen data. Considering the results of the models using ChiSquaredEval features, Naïve Bayes (OVA) has the highest accuracy of 71.81%. Random Forest (OVO), SVM (OVO), and KNN (OVA) achieved 65.14%, 64.58%, and 60.69 average classification accuracy, respectively. Table 4.18 represents the results. Considering kappa statistics, all algorithms show a moderate agreement between the predicted and actual values in the dataset. The average precision value that indicates the positive predicting power of Naïve Bayes classifiers is 0.77 and it is the highest among other classifiers. Random Forest, SVM, and KNN show a lower precision value of "0.64", "0.63", and "0.59", respectively. The average recall value shows that the success of the Random Forest algorithm to identify projects with a positive class instance is highest (0.60). SVM, Naïve Bayes, and KNN show lower average recall values than the Random Forest algorithm, 0.59, 0.56, and 0.45, respectively. Naïve Bayes has an F score of 0.61. Lower F scores are seen for SVM (0.57), Random Forest (0.56), and KNN classifier (0.58). Naïve Bayes has a very good AUC value (0.81). Random Forest (0.76), SVM (0.76), and KNN (0.76) indicate a lower but very good AUC value.

Table 4.18 Performance Evaluation of the Classifiers

Negative Heating Gap	**ChiSquaredEval**			
Multi Class classification	**Whole data set results**			
	Models			
Performance Metric	**Naïve Bayes**	**KNN**	**SVM**	**Random Forest**
Accuracy (%)	**71.81**	**60.69**	**64.58**	**65.14**
Kappa	**0.58**	**0.41**	**0.47**	**0.47**
Precision	**0.77**	**0.59**	**0.63**	**0.64**
Recall	*0.45*	*0.56*	*0.59*	*0.60*
F-Measure	0.61	0.58	0.57	0.56
AUC	**0.81**	**0.76**	**0.76**	**0.76**

A comparison of the results of the algorithm performances with the baseline classifier (Zero R) using Paired T-Tester statistical test can be seen in Table 4.18. If a classifier is significantly better than the baseline classifier, the result is written in bold. If a classifier is significantly worse than the baseline classifier, the result is written in italic. In other cases, we cannot be sure whether there is a statistically significant difference or not. Zero R provided an accuracy of 8.33%, a kappa value of 0.00, a precision of 0.17, a recall of 0.33, an F measure of 0.50, and an AUC of 0.50. Considering that, all classifiers are statistically better than the baseline classifier in four performance metrics, namely, accuracy, kappa statistics, precision, and AUC. All classifiers are statistically worse than the baseline classifier in the recall. Since the Naïve Bayes classifier performs better than other classifiers in five performance metrics (accuracy, kappa statistics, precision, F-measure, AUC), it is the suggested classifier as can be seen in Figure 4.6.

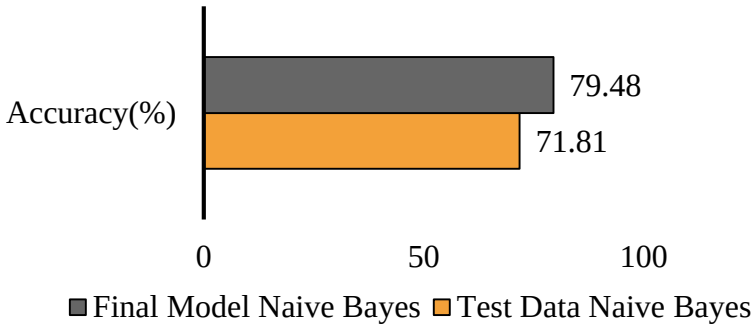


Figure 4.6 Performance of the Final Model: Naïve Bayes

4.3 Comparison of the Results

Considering the prediction of the electricity performance gap in buildings, three different classification problems were studied. While KNN is the top performer in the binary classification of the electricity performance gap based on predictive accuracy, Naïve Bayes performs better in the multi-class classification of the positive and negative electricity performance gap. As the top performer of the first classification problem in terms of accuracy, KNN indicates high precision and high recall. The model can manage the class quite well. Besides, the model shows an average performance to separate positive and negative instances. Naïve Bayes, as the top performer of the positive electricity performance gap prediction in terms of accuracy, indicates perfect precision and average recall. This can be interpreted as the prediction is very reliable however the model shows an average performance to spot the class. Besides, this model has a very good performance to separate the positive and negative instances. Naïve Bayes indicates again the highest accuracy in multi-class classification of negative electricity performance gap. In this case, the model shows perfect precision and very high recall. This indicates that the class is managed very successfully by the model and the model has an excellent performance to separate positive and negative instances.

A comparison of the performance of the algorithms shows that Naïve Bayes is the top performer to predict the heating performance gap in terms of accuracy. While SVM performs better to predict positive heating performance gap, Naïve Bayes performs better to predict negative electricity performance gap in terms of accuracy. As the top performer of the first classification problem, Naïve Bayes shows very high recall and good precision. This indicates that although the model can manage the class quite well, there are also instances from other classes. The model has an average performance to separate positive and negative instances. SVM, as the top performer in the prediction of the positive heating performance gap, shows a very high recall, but low precision. This can be interpreted as the class being detected very well; however, the prediction is correct 50% of the time. This model performs excellently to separate positive and negative instances. Naïve Bayes indicates the highest accuracy in predicting negative heating performance gap. This time, the model shows a high precision, but very low recall. This indicates that the class is managed poorly by the model, however, the prediction is trustable. This model performs very well to separate positive and negative instances.

Considering all classification models, the results indicate that Naïve Bayes gives the best classification accuracy on the majority of the classification problems. Van de Schoot & Milica (2020) explain that if researchers have access to a rather small number of participants, Bayesian methods are attractive to them. Because, Bayesian statistics may provide good results despite small to moderate sample sizes since they are not founded on large samples (i.e., the central limit theorem). Moreover, Stribos (2020) studied how data noise impacts the accuracy of different classification algorithms. This study showed that Naïve Bayes is less sensitive to data noise in comparison to the Random Forest algorithm. Additionally, since the Naïve Bayes method depends on categorical attributes, this might be another reason for the success of the Naïve Bayes algorithm in the given datasets which are mainly based on categorical data.

CHAPTER 5

CONCLUSION

5.1 Discussion of Findings

This research examined the energy performance gap in buildings through the lens of project risks. First, semi-structured interviews were conducted with 11 experts to identify risks that may cause energy performance gaps in buildings. Using a survey listing various project risks, the performance gap observed related to electricity and heating demand in buildings is collected via a web-based survey. The data gathered through the survey is then studied as a classification problem utilizing machine learning algorithms. The first classification problem aims to predict whether the buildings might consume less (negative) or more (positive) than design expectations. The second classification problem aims to predict the level of performance gap as percentages. To the best of the knowledge of the researcher, there is not any study that uses the experiences of the past to estimate a gap between the designed and observed performance of buildings. Prediction of the energy performance gap is critical for deciding on possible investment in buildings. The energy performance gap in buildings causes significant increases in energy cost and environmental impact (Birchall, 2011), improper use of capital investment, and unreasonable design of the system (Harrison & Jiang, 2018). Moreover, it introduces reputational dangers to the industry (Dollard, 2018). Therefore, if project stakeholders can predict the direction (positive or negative) and the level of a performance gap (low, medium, high), their strategies and project decisions might change. The following sections explain the main findings of the semi-structured interviews, the dataset for electricity gap prediction, the dataset for heating gap prediction, machine learning analysis, limitations of the study, and recommendations for the future work, respectively.

5.1.1 Discussion of Findings from Semi-Structured Interviews

Semi-structured interviews were conducted with 11 professionals about the project process of six different energy-efficient buildings to understand how risk factors work together and cause performance gaps. These cases were selected in a way that risk paths of different energy-efficient buildings, building types, and company and country conditions could be observed. The cases include a Leed Platinium certified office building, one office building that is expected to earn a Leed Gold certificate, two Passive House certified residential buildings, and two non-certified exemplary buildings aiming at improved energy efficiency (one office, one educational building). The general conclusions drawn from the semi-structured interviews are as follows:

- Project risks that may cause the energy performance gap may be different in number and type considering each case.
- Availability of material or certified equipment appears to be a country-specific risk. All cases faced difficulties to find local high-quality mechanical equipment in Turkey. As a result, project teams felt an obligation to import those products. However, since imports are more expensive, project stakeholders may lower the quality expectations in that or other products or change the level of skill they hire.
- Several participants addressed variations in the properties of some equipment. The information given in some catalogs was not correct. One participant stated that they had to rely on the catalog values due to the shortage of material testing laboratories and the expense of each test. Another participant claimed that based on their calculations, they proved that the value given in the catalog was wrong. In another case, the project manager was not sure if the contractor ordered another product or if the quality of the product was lower than claimed.

- Various participants talked about the modeling or software impact on the project. According to one participant, building modeling does not realistically reflect the building. Another participant suggested integrating BIM into the project process to avoid any mistakes causing revisions in the implementation stage. Two other participants mentioned that summer comfort should be paid more attention to software or calculation methodologies due to climate change.
- Changes or revisions are common in construction, and they can be seen in material decisions, systems, space design, etc. Modifications can cause improved indoor comfort for occupants and less energy consumption or reduced comfort and more energy consumption.
- However, in all cases, a significant majority of participants highlighted the importance of experience and effective communication to reach project goals as much as possible. In addition to the availability of experienced stakeholders, their numbers, and the time they joined the team is critical. Ideally, they should be introduced at an early project stage.
- Moreover, it was obvious that project stakeholders were constantly faced with several trade-offs during the project process such as cost-energy performance trade-off, and cost-comfort trade-off. As Hatefi & Tamošaitienė (2019) state, many of the risks inherent in construction projects influence each other. Although experts also mentioned the problems that can lead to a gap in the budget, project time, and user comfort during the meetings, these parts were not included in the study. The application of ML classification is not only used in performance prediction, but also in medical diagnosis, fraud detection, image classification, etc. Likewise, the method suggested in this study could not only be applied to the prediction of the energy performance gap but also cost or time overrun of energy-efficient buildings.

5.1.2 Discussion of Findings related to Electricity Gap Database

In this section, the survey responses regarding the energy performance gap are evaluated considering the electricity demand of the buildings. Questions having missing responses were not considered. There are 55 buildings in the electricity gap database coming from survey responses. While 29 of them are educational buildings, 13 of them are office buildings. The number of buildings showing a negative gap is slightly more than the buildings showing a positive gap both for educational and office buildings. When the buildings are analyzed based on the year of construction, a large majority of buildings from the 18^{th}-20^{th} century have a negative performance gap. However, a large majority of buildings from the 21^{st} century show a positive performance gap. Moreover, there are 21 new buildings and 34 refurbishments in this dataset. Similarly, while a large majority of the new buildings consume more electricity than expected, most refurbishments consume less electricity than expected. As Mojic *et al.* (2019) explain, several studies indicate that in older, not renovated buildings up to the year of construction 1995 and with a design heating demand of more than 100-130 kWh/m^2a, the energy consumption is often lower than the calculated energy demand. Analyzing the performance gap in terms of the heated floor area of buildings, if the buildings have a heated floor area up to 2500m^2, it is more likely that there will be a negative performance gap. 23 out of 55 buildings are reinforced concrete and most of them consume more electricity than expected. While most of the buildings consume less than expected in masonry buildings, the number of buildings having a positive or negative gap is almost the same in mixed construction. Appendix K visualizes the analysis of the dataset in bar charts. Moreover, the spider chart given in Figure 5.1 indicates the responses to the survey questions in case the buildings have a positive performance gap or a negative performance gap.

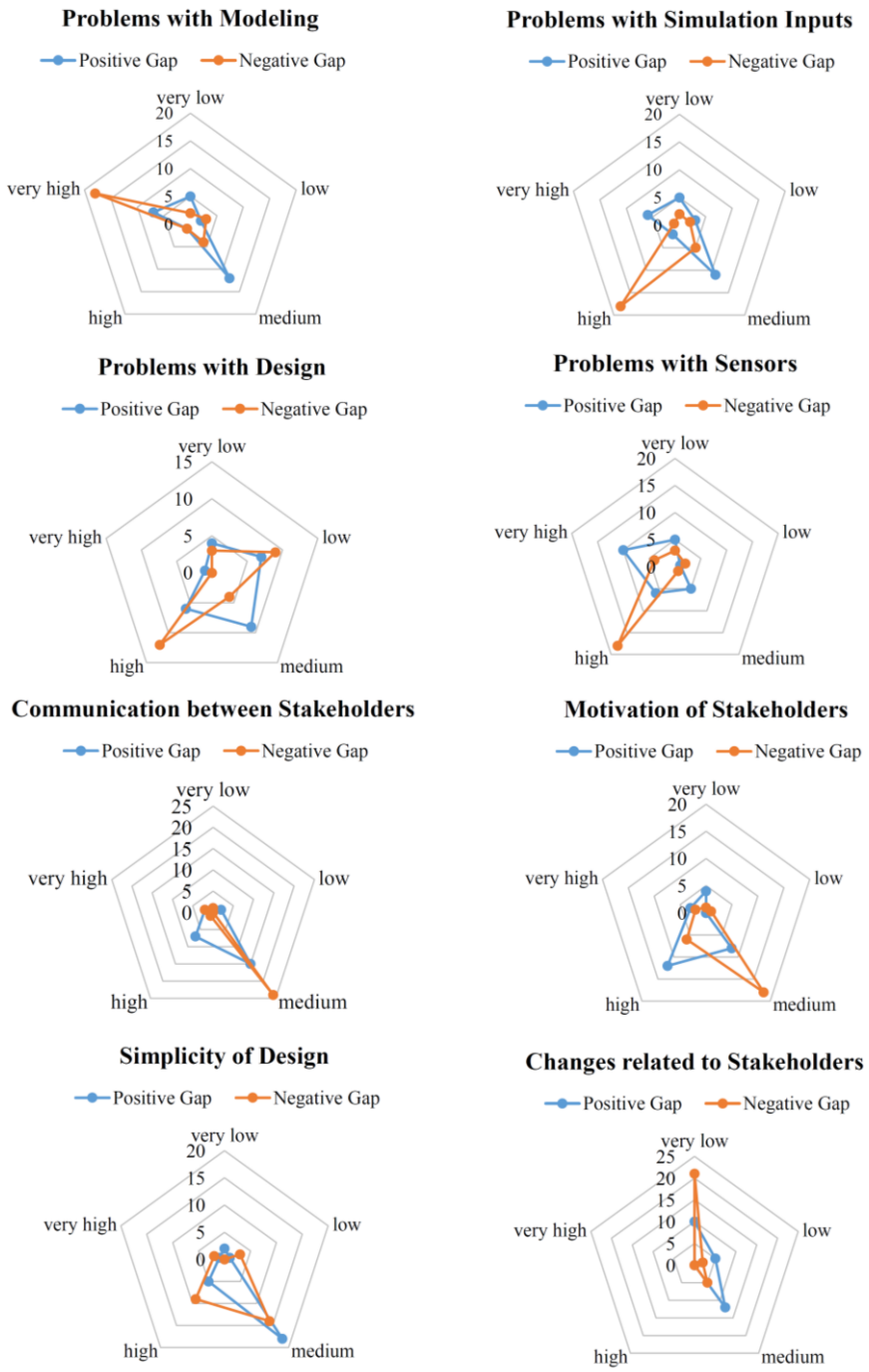


Figure 5.1 Some of the Selected Attributes and their Values in the Electricity Gap Dataset

While the survey questions constitute the attributes taken into account in ML models, answers to these questions were used as values of the given attribute. Figure 5.1 shows that respondents most often chose from medium, medium, medium, and very high while answering the survey questions about the problems with modeling/software, simulation inputs, design, and sensors/meters, respectively. In this case, the building has a positive performance gap. Moreover, most often, communication between stakeholders, their motivation, simplicity of design, and changes related to stakeholders were chosen as medium, high, medium, and medium, respectively. If the building has a negative performance gap, responses to the given questions were most often very high, high, high, high, medium, medium, medium, and very low, respectively.

Considering the survey responses about modeling and software problems, approximately 70 percent of cases with high and very high problems related to modeling or software have a negative performance gap. Given the problems encountered with simulation inputs, problems were rated as high most of the time (20 cases). If problems with simulation inputs were rated as high and very high, around 30 percent of buildings showed a positive performance gap. In the case of design-related problems, problems were mostly high (18 cases). If the problems with design were rated as high and very high, around 37% of the cases showed a positive gap.

The problems with workmanship were rated very low in 7 cases, low in 13 cases, medium in 17 cases, and high in 17 cases. Around 50 percent of cases with high and very high ratings related to the problems with the quality of workmanship have a positive performance gap. Regarding the problems related to the quality of materials/equipment, problems were mostly rated medium (30 cases). In these cases, electricity consumption is less than expected majority of the time. About 44 percent of buildings that had medium and high ratings related to the problems with the quality of materials/equipment show a positive performance gap.

If problems related to occupants were categorized as high and very high, around 45% of the time the performance gap was positive. Participants were asked to evaluate the magnitude of the problems related to sensors/meters as well. The problems were rated as very low in 8 cases, high in 24 cases, and very high in 14 cases. Cases in which problems are very low, medium, and very high, often have higher electricity consumption than expected. Besides, cases in which problems were considered to be high, had less electricity consumption than the expected large majority of the time.

Effective communication between project stakeholders was suggested as a way to mitigate the performance gap both in articles and semi-structured interviews. While a large majority of cases have a negative gap when they are rated as having a medium level of effective communication, they have a positive performance gap a large majority of the time even though the level of effective communication was rated as high. Approximately 33 percent of buildings with a positive performance gap indicate a high and very high level of effective communication. The training of project stakeholders is also given as an effective factor to reduce the performance gap. The large majority of the cases had a negative gap in cases where the level of training of project stakeholders was rated as low and this might seem contradictory. The large majority of the cases had a positive gap in cases where the level of training was rated as medium. The simplicity of design is presented as a factor to achieve the targeted building performance. The survey responses support a similar view. In cases where the design simplicity is very low, low, or medium, the majority of the time the performance gap is positive. However, the gap is negative the large majority of the time when the cases were rated with a high and very high level of simplicity. Around 26 percent of the positive performance gap cases have been rated with a high and very high level of simplicity.

Participants were asked to evaluate the level of changes in client expectations/requirements (related to design) and the change was rated as very low in 43 cases. Buildings consume less electricity than the predictions large majority of

the time. Next, participants were asked to rate changes related to project stakeholders. In 31 cases, it was rated as very low, and similarly, the large majority of the cases have a negative performance gap. Moreover, the change was rated as medium in 17 cases and this time a large majority of the time, the gap is positive. Following that, the participants were asked to evaluate the unexpected events and changes in country conditions such as unavailability of materials. In 30 cases, the unexpected events in country conditions were rated as very low, and a large majority of the time, these cases show a negative performance gap. Cases that have a high level of change indicate a positive gap most of the time. Additionally, the participants were asked to evaluate unexpected events and changes in policy/ legislation/ regulation and public project processes such as tendering. A similar conclusion to country conditions can be drawn from these evaluations as well. The last two questions were asked for the evaluation of unexpected events and changes (if happened) such as extreme weather conditions or pandemics. The distribution of the responses was similar to the previous evaluations. The percentage of buildings showing a positive and negative gap is given in case the listed survey questions were answered as it is given in the Table 5.1.

Table 5.1 Analyses of Electricity Gap Dataset

Questions	Survey Responses	Positive Performance Gap	Negative Performance Gap
problems with modeling	high-very high	~30%	~70%
problems with inputs	high-very high	~30%	~70%
problems with design	high-very high	~37%	~63%
problems with workmanship	high-very high	~50%	~50%
problems with materials	medium-high	~44%	~56%
problems with occupants	high-very high	~45%	~55%
problems with sensors, meters *etc*.	high-very high	~40%	~60%
experience of stakeholders	very low-low	~67%	~33%
motivation of stakeholders	high-very high	~70%	~30%
communication of stakeholders	high-very high	~75%	~25%
training of stakeholders	very low-low	~29%	~71%
simplicity of design	high-very high	~35%	~65%
changes in client expectations	very low-low	~45%	~55%
change in stakeholders	very low-low	~39%	~61%
electricity gap		49%	51%

5.1.3 Discussion of Findings related to Heating Gap Database

In this section, the survey responses regarding the energy performance gap are evaluated considering the heating demand of the buildings. There are 56 buildings in the heating gap database out of which 29 are educational buildings and 14 are office buildings. While cases having a positive gap in educational buildings is slightly more than cases having a negative gap, cases having a negative and positive gap are the same in office buildings. When considering the year of construction, if a building is constructed in the 21st century, it is more likely that it has a positive performance gap. In the database, 21 buildings are newly built, and 35 buildings are refurbishments. The probability that the building has a positive or negative gap is almost equal for both categories. When examining the performance gap considering

the heated floor area of buildings, if the heated floor area is up to 2500 m^2, it is more likely that the direction of the gap will be negative. Regardless of its construction type, the probability of observing a positive or negative performance gap is equal for all construction types. Appendix L visualizes the analysis of the dataset in bar charts. Moreover, the spider charts in Figure 5.2 illustrate some of the selected attributes and their values. While the survey questions constitute the attributes taken into account in ML models, answers to these questions were used as values of the given attribute.

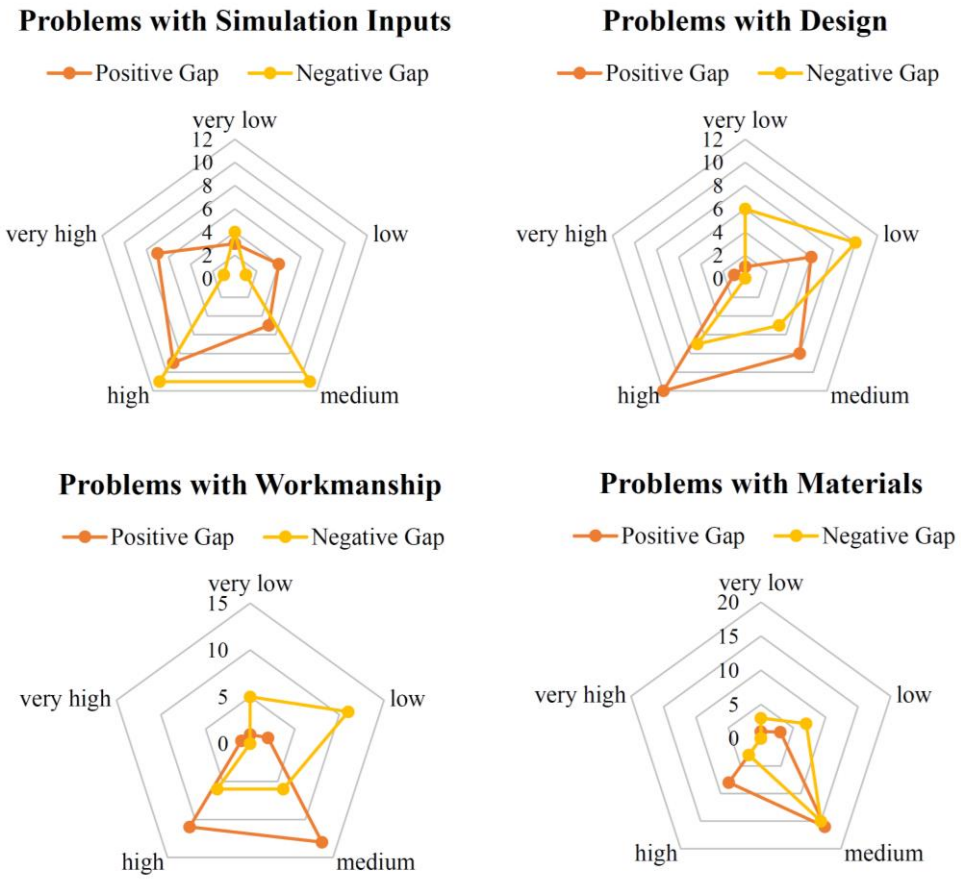


Figure 5.2 Some of the Selected Attributes and their Values in the Heating Performance Gap Dataset

As it can be seen, when the building has a positive performance gap, the respondents most often choose from high, high, medium, and medium while answering the survey questions about the problems with simulation inputs, design, workmanship, and materials, respectively. When the building has a negative performance gap, they most often choose from high/medium, low, low, and medium.

Given the survey responses about the problems related to modeling and software, most of the time, the problems were identified as very high (26 cases). Around 46 percent of buildings showing a positive performance gap have very high problems related to modeling and software. Given the problems encountered with simulation inputs, problems were rated most of the time as high (20 cases). In very high-ranked cases, the consumption is more than expected large majority of the time. Around 57 percent of cases with a positive performance gap indicate high and very high level of problems with simulation inputs. In the case of design-related problems, problems were rated as very low in 7 cases, low in 16 cases, medium in 13 cases, and high in 19 cases. The large majority of the cases showed a negative performance gap if the problems were rated to be very low or low. Other groups often indicate a positive performance gap. Problems related to design were ranked high and very high around 46 percent of buildings with a positive performance gap. Table 5.2 gives a summary of the findings, representing the percentage of buildings showing a positive and negative gap in case the listed survey questions were answered as it is given in the table.

Considering the problems with the quality of workmanship, around 43 percent of buildings with a positive performance gap indicate a high and very high level of quality of workmanship problems. Following that, an evaluation of problems related to the quality of materials and equipment indicates that problems were rated medium most of the time (31 cases). Around 86 percent of buildings with a positive performance gap indicate medium and high problems related to material/equipment quality.

Most of the time, problems related to occupants (*e.g.*, uncertainty in occupant behavior, *etc*.) were ranked very high (23 cases). Around 79% of cases with a positive performance gap indicate a high and very high level of problems with the occupants. Moreover, problems with sensors/meters were ranked high in 24 cases, and very high in 15 cases. Approximately 71 percent of buildings with a positive performance gap have a high and very high level of problems with the sensors or meters.

Evaluation of the experience of project stakeholders indicates that the level of experience is rated low in 14 cases, medium in 27 cases, and high in 9 cases. Around 25 percent of cases with a positive performance gap have a rank of high and very high regarding the experience of the project parties. An evaluation of the motivation of the project stakeholders indicates that around 43 percent of cases with a positive performance gap have high and very high levels of motivation of project participants. Considering effective communication, around 79 percent of buildings with a positive performance gap indicate a low and medium level of effective communication. Moreover, in the survey, the training of project stakeholders was ranked as low in 27 cases and medium in 21 cases. Low-ranked cases show a negative performance gap majority of the time. Around 46 percent of buildings with a positive performance gap, the training level of stakeholders was ranked as very low and low. Moreover, only around 14 percent of buildings with a positive performance gap indicated a very low and low level of design simplicity. Table 5.2 represents a summary of the findings, the percentage of buildings showing a positive and negative gap is given in case the listed survey questions were answered as it is given in the table.

Table 5.2 Analyses of Heating Gap Dataset

Questions	Survey Response	Positive Performance Gap	Negative Performance Gap
problems with modeling	high-very high	~46%	~54%
problems with inputs	high-very high	~57%	~43%
problems with design	high-very high	~65%	~35%
problems with workmanship	high-very high	~67%	~33%
problems with materials	high	~73%	~27%
problems with occupants	high-very high	~51%	~49%
problems with sensors, meters *etc*.	high-very high	~51%	~49%
experience of stakeholders	very low-low	~44%	~56%
motivation of stakeholders	high-very high	~52%	~48%
communication between stakeholders	high-very high	~55%	~45%
training of stakeholders	very low-low	~45%	~55%
simplicity of design	high-very high	~47%	~53%
changes in client expectations	very low-low	~48%	~52%
change in stakeholders	very low-low	~49%	~51%
heating gap		50%	50%

Considering the level of changes in client expectations/requirements (e.g., related to design, etc.), the change was rated as very low in 42 cases. A large majority of the time, the performance gap was negative for these cases. Next, participants were asked to rate changes in project stakeholders. In 31 cases, it was rated as very low and the change was rated as medium in 17 cases. In both groups, the direction of the gap is negative majority of the time. Following that, the participants were asked to evaluate the unexpected events and changes in country conditions such as unavailability of materials. In 29 cases, the unexpected events in country conditions are rated as very low and the direction of the gap is negative a large majority of the time. Moreover, a similar conclusion to country conditions can be drawn from the evaluations related to changes in policy/implementation/regulations and public project process. The last two questions were to evaluate the unexpected events and changes (if happened) such as extreme weather conditions and pandemics. The

responses were similar. Most of the cases had a very low level of change. The majority of the time, the gap is negative in these cases.

5.1.4 Discussion of Findings related to Machine Learning Models

In this research, to find the best-performing prediction model, several feature selection techniques and four different ML algorithms were tested. While Wrapper methods use a machine learning algorithm to select the best subset of features (Rachman & Ratnayake, 2019), an independent assessment is made based on general characteristics of the data using a Filter method (Witten et al., 2011). Wrapper methods are known for creating better predictive models compared to filter methods. Therefore, WrapperSubsetEval was applied for binary classification problems since it is the starting point for the whole classification procedure. Since the number of models increased in multi-class classification problems, the preference was a Filter method as they are much faster and more practical than Wrapper methods. In this part, models created with ChiSquareAttributeEval were used for each multi-classification problem. Although InfoGainAttributeEval was also used to suggest models for multi-classification problems, the results showed that InfoGainAttributeEval and ChiSquareAttributeEval list very similar attributes in some datasets. However, better predictive performance was achieved with the features selected with ChiSquareAttributeEval. Besides, an overall analysis of the selected features based on ChiSquareAttributeEval indicates that the following features appeared more than once in the selected subsets in different prediction problems:

- Shortcomings of modeling, software, or calculation methodology,
- Uncertainty related to simulation inputs,
- Effective communication between project stakeholders.

Six different classification problems in two distinct categories were studied in this research. The first category deals with the prediction of the performance gap for electricity demand in buildings and the second category deals with the performance gap prediction for heating demand in buildings. In each category, first, datasets were studied to solve binary classification problems. The main goal was to distinguish cases that might show a positive and negative performance gap in this part for each category. Considering electricity gap prediction for binary classification with 10-fold CV, ***KNN*** and ***SVM*** are the top two performers regarding the accuracy, kappa statistic, precision, and F-measure. In this part, KNN and SVM achieved an accuracy of 79% and 75% on unseen data, respectively. ***KNN*** is the suggested classifier to solve binary classification for ***electricity gap prediction*** due to its higher prediction accuracy. A conceptual web interface that introduces a performance gap prediction tool for electricity demand can be seen in Figure 5.3.

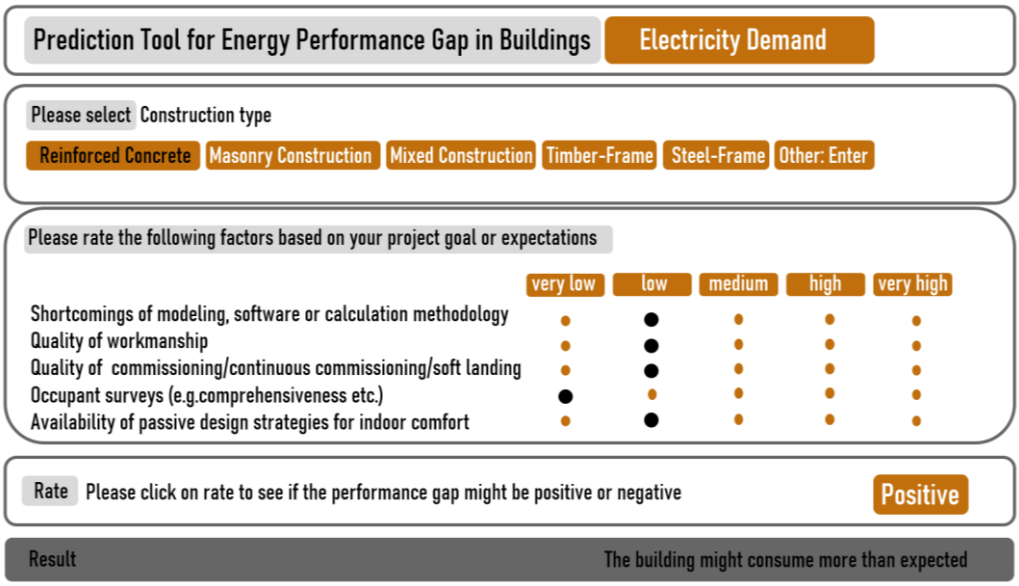


Figure 5.3 Conceptual Web-Interface to Predict a Positive or Negative Gap

The listed factors in the interface (which are selected by feature selection methods) will be evaluated by a project stakeholder using a ranking between “very low” to “very high.” Based on this evaluation, the tool will predict whether the gap might be

positive or negative in this building. Figure 5.4 indicates how the answers to the related survey questions are distributed considering the binary classification of electricity demand in the dataset. Survey questions constitute the attributes utilized in ML and the answers given to them are values of these attributes.

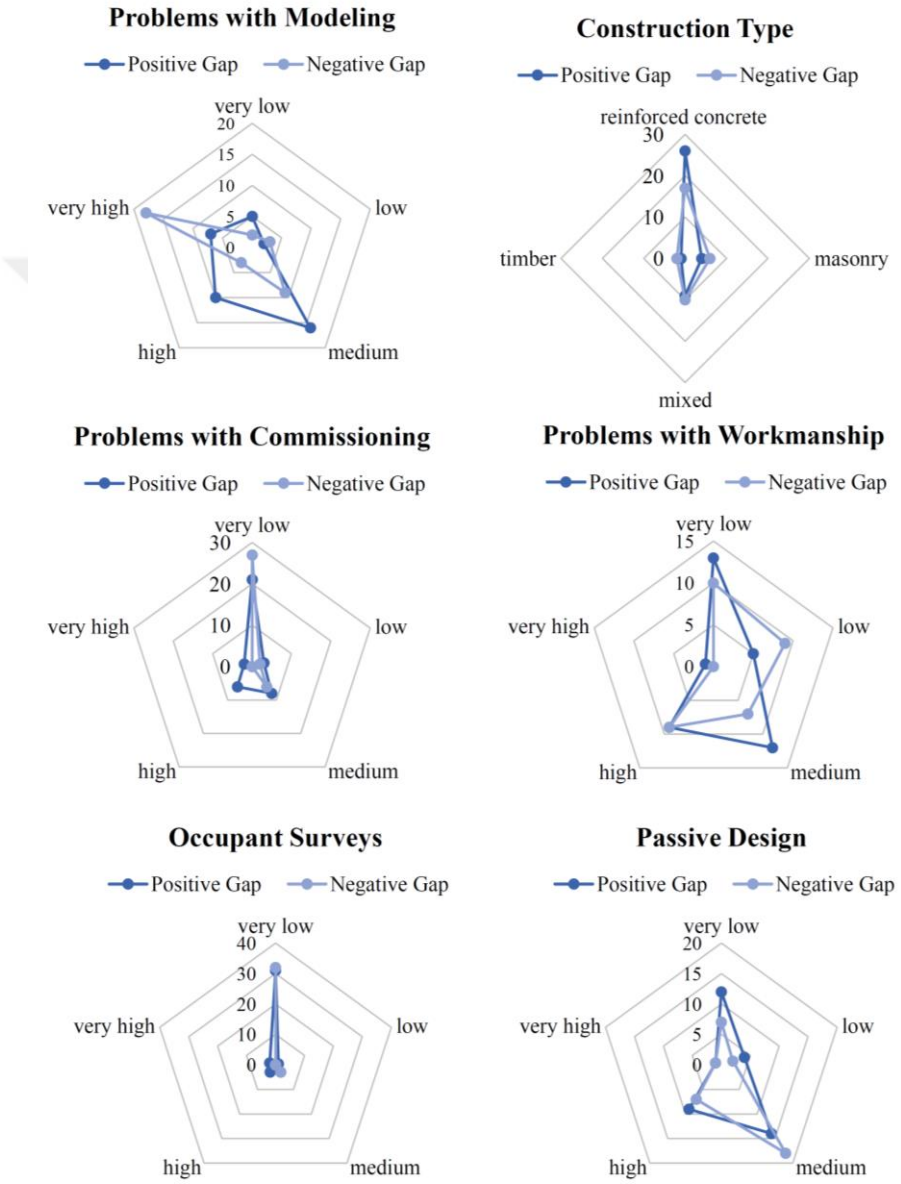


Figure 5.4 Selected Attributes and their Values in the First Classification Problem

Following binary classification, smaller datasets were studied for multi-class classification problems to predict the level of performance gap in buildings based on the following classes, low (0-15%), medium (15.1-40%), and high (40.1-90%). Repeating the 4-fold CV four times, ***Naïve Bayes*** and ***Random Forest*** are the top two performers regarding accuracy and kappa statistic in predicting a positive performance gap for electricity demand in multi-class classification. Naïve Bayes (OVO) and Random Forest (ECC) provided an accuracy of 77.08% and 73.96% on unseen data (test data), respectively. ***Naïve Bayes*** is the suggested classifier to solve multi-class classification for ***positive electricity gap prediction*** due to its higher prediction accuracy. The conceptual web interface that introduces a positive performance gap prediction tool for electricity demand can be seen in Figure 5.5. Similarly, the listed factors in the interface (which are selected by a feature selection method) will be evaluated by a project stakeholder using a ranking between "very low" to "very high." Based on this evaluation, the tool will predict the level of the gap between the classes of low (0-15%), medium (15.1-40%), and large (40.1-90%).

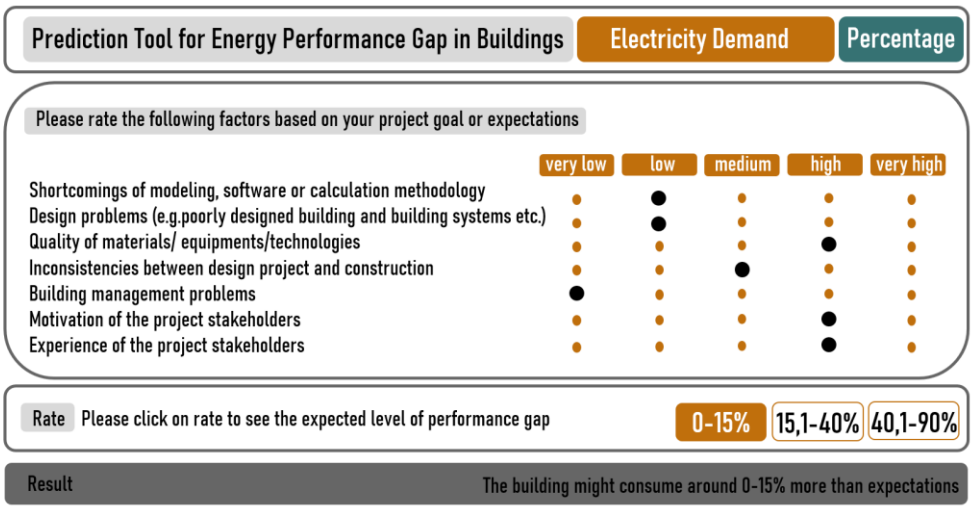


Figure 5.5 Conceptual Web-Interface to Predict the Positive Gap for Electricity Demand

Figure 5.6 indicates how the answers to the related survey questions are distributed considering the multi-class classification of electricity demand when the buildings have a positive performance gap.

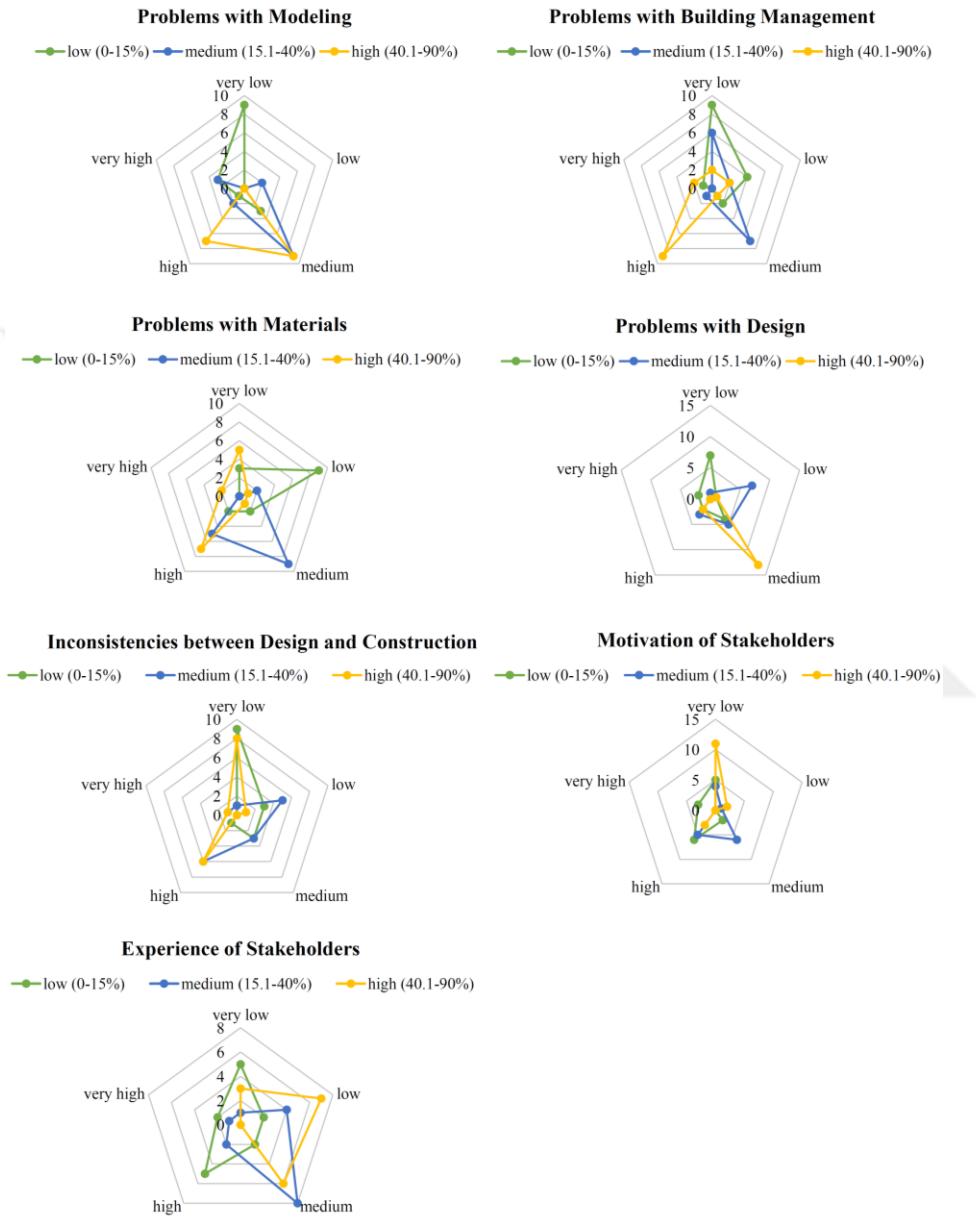


Figure 5.6 Selected Attributes and their Values in the Second Classification Problem

Considering multi-class classification of the negative performance gap prediction for electricity demand, ***Naïve Bayes*** (RCC) and ***KNN*** (OVA) are the top two performers regarding accuracy, kappa statistic, recall, and AUC. In this part, precision and F-Measure were indeterminable when evaluating the model on unseen data. Therefore, the results are based on repeating 4-fold CV 4 times on the whole data set that also includes unseen test data. Naïve Bayes (RCC) and KNN (OVA) achieved an accuracy of 83.85% and 82.29%, respectively. The conceptual web interface that introduces a negative performance gap prediction tool for electricity demand can be seen in Appendix M. ***Naïve Bayes*** is the suggested classifier to solve multi-class classification for ***negative electricity gap prediction*** due to its higher prediction accuracy. Figure 5.7 indicates how the answers to the related survey questions are distributed considering the multi-class classification of electricity demand when the buildings have a negative performance gap.

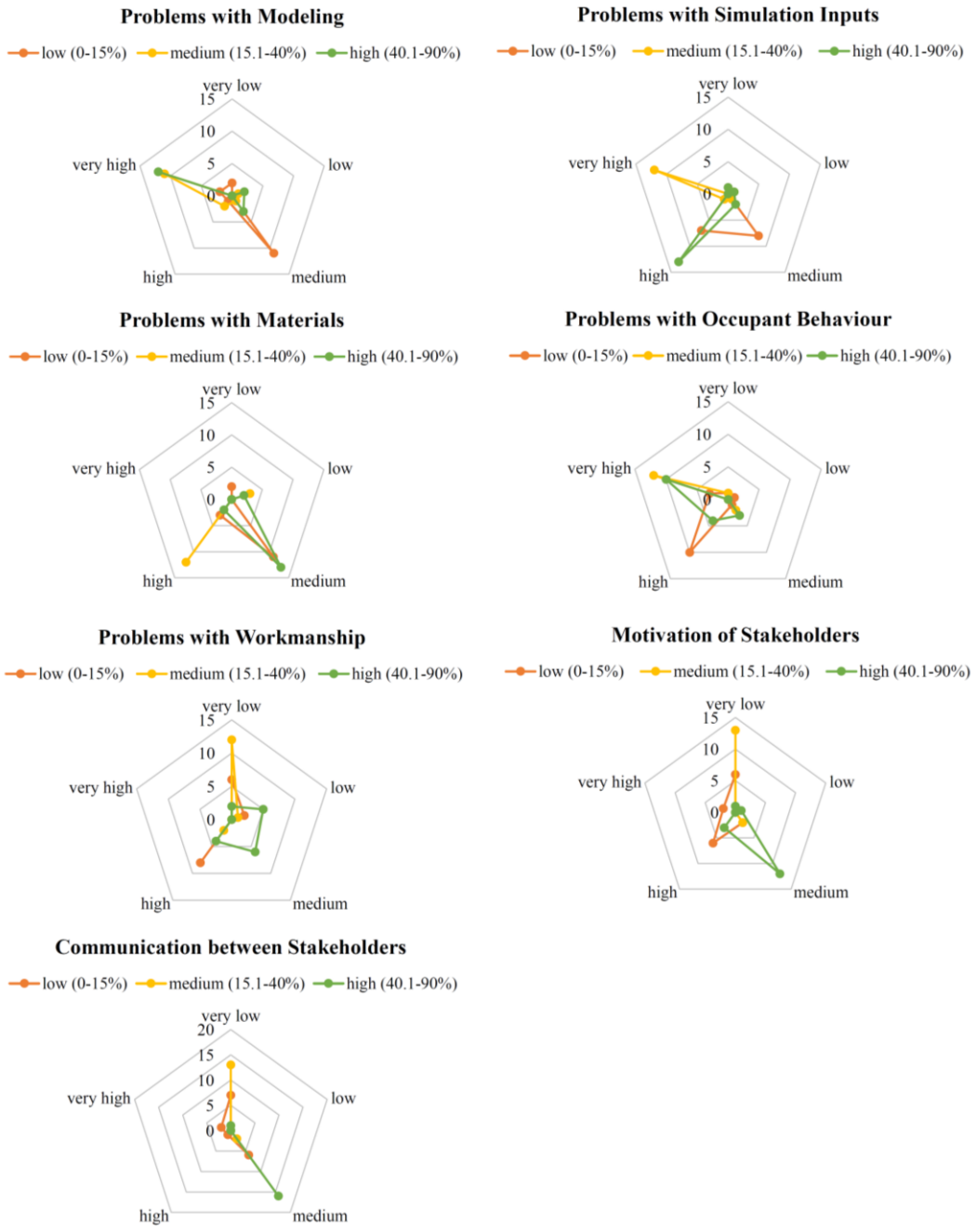


Figure 5.7 Selected Attributes and their Values in the Third Classification Problem

Naïve Bayes and ***SVM*** are the top two performers for binary classification of the heating gap with a 10-fold CV. In this part, Naïve Bayes and SVM achieved an

accuracy of 72.5% and 68.50% on unseen data, respectively. While Naïve Bayes outperforms other classifiers in five performance metrics except for AUC, SVM outperforms the other two classifiers (KNN and Random Forest) in two performance metrics. ***Naïve Bayes*** is the suggested classifier to solve binary classification for ***heating gap prediction*** due to its higher prediction accuracy. A conceptual web interface that is suggested to use as a prediction tool for the heating gap can be found in Figure 5.8. Only three attributes are required for the prediction of the direction of the heating demand.

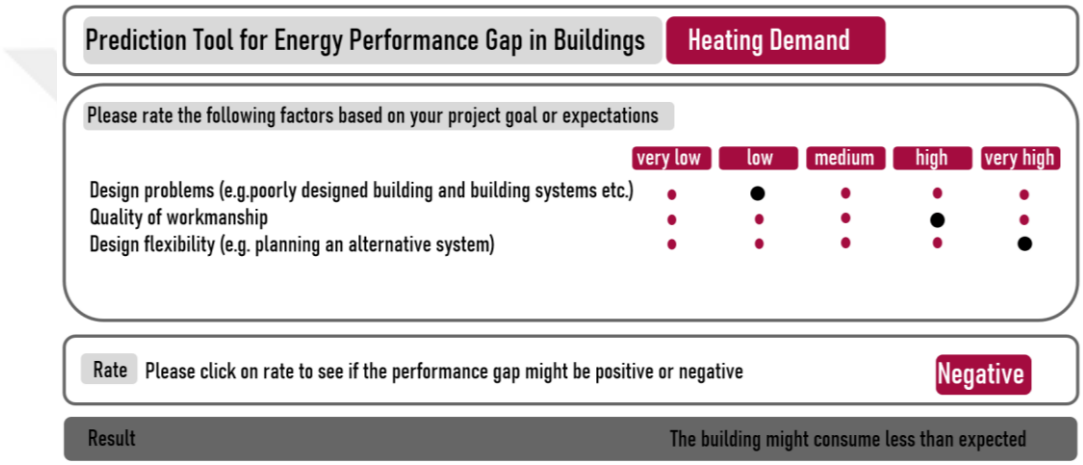


Figure 5.8 Conceptual Web-Interface to Predict a Positive or Negative Heating Gap

Selected attributes to solve the classification problem and the distribution of the values in the dataset are given in Figure 5.9. According to the spider chart, responses to the survey questions are most often low, medium, and low to the problems with design, workmanship, and design flexibility, respectively when there is a positive performance gap. The responses to the survey questions are most often low, very low, and low, respectively when there is a negative performance gap.

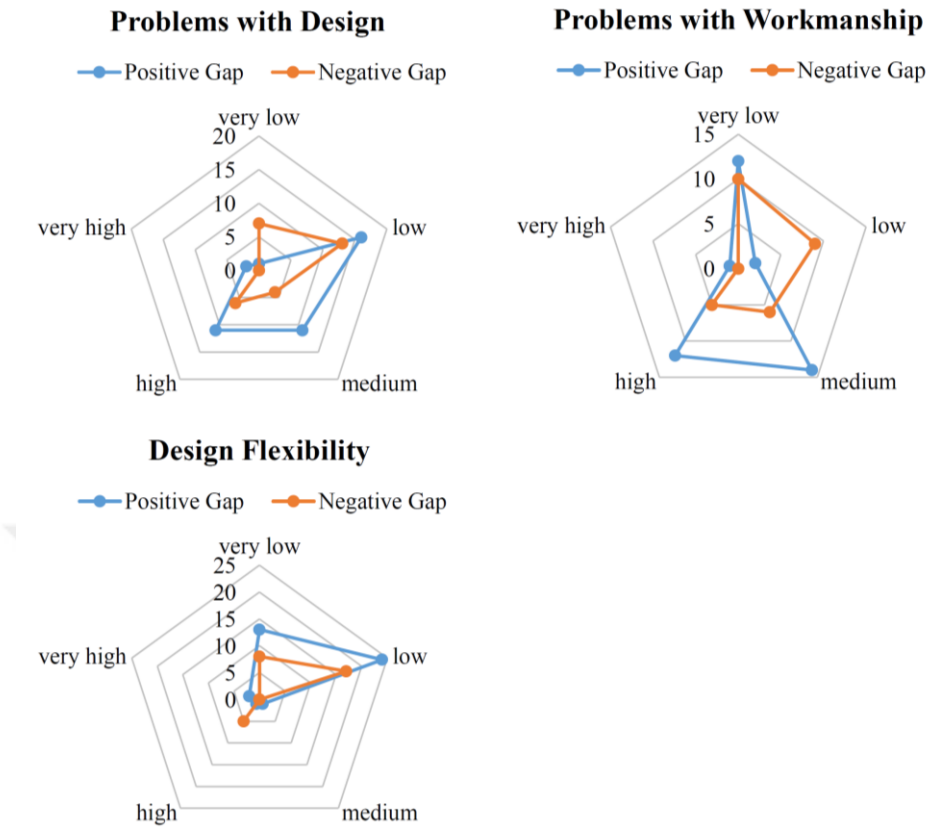


Figure 5.9 Selected Attributes and their Values in the Fourth Classification Problem

SVM (ECC) and Random Forest (OVA) are the top two performers regarding accuracy and kappa statistics for the positive heating performance gap prediction with a 4-fold CV in multi-class classification. SVM (ECC) and Random Forest (OVA) provided an accuracy of 76.04 % and 69.79% on unseen data, respectively. SVM outperforms Random Forest in accuracy, kappa statistics, precision, and recall. ***SVM*** is the suggested classifier in this classification problem. The conceptual web interface that introduces a positive performance gap prediction tool for heating demand can be found in Appendix N. Selected attributes to solve the multi-class classification problem and the distribution of the values in the dataset are given in Figure 5.10.

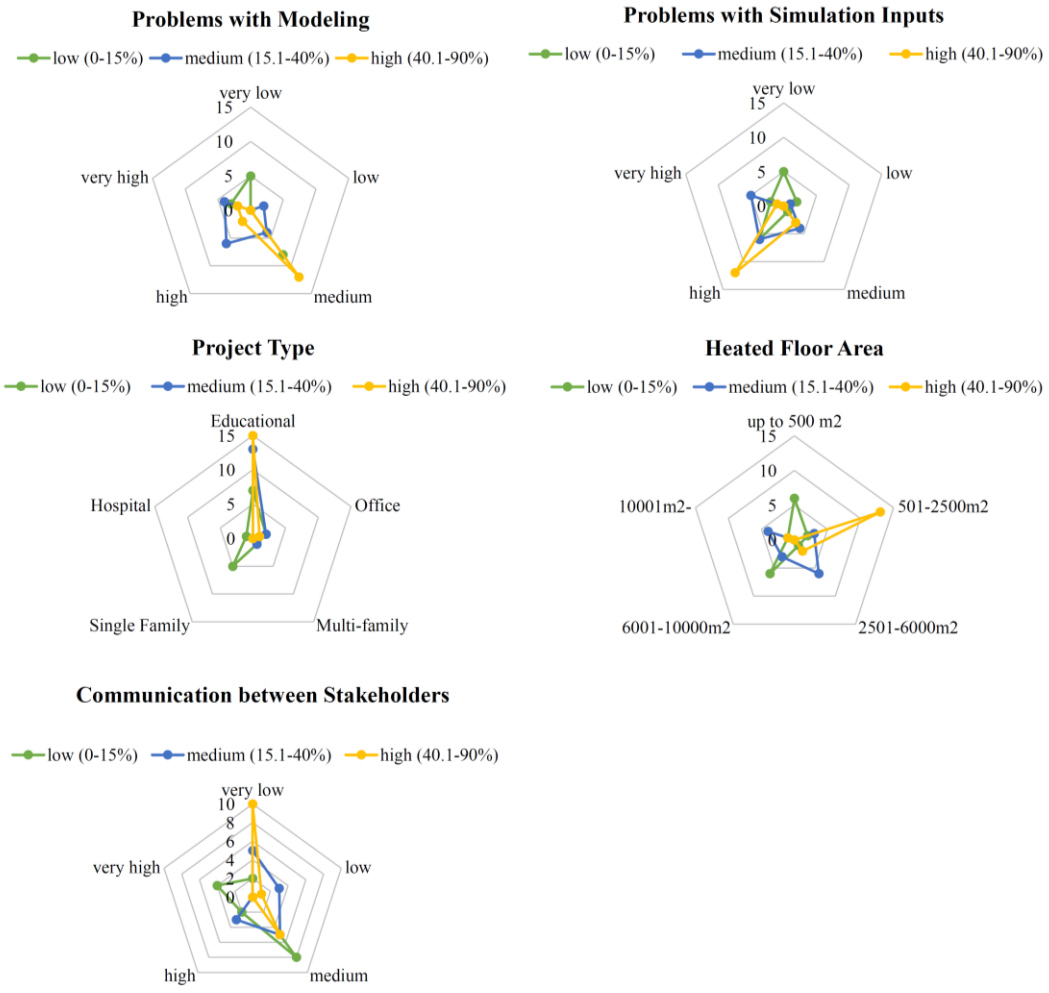


Figure 5.10 Selected Attributes and their Values in the Fifth Classification Problem

Lastly, Naïve Bayes (OVA) and Random Forest (OVO) are the top two performers regarding accuracy and precision to predict the negative heating gap in multi-class classification. Naïve Bayes (OVA) and Random Forest (OVO) achieved an accuracy of 71.81% and 65.14%, respectively. However, like the multi-classification problem of negative performance gap prediction for electricity demand, precision and F-Measure were indeterminable when evaluating the model on unseen data. Therefore, the results are based on repeating 4-fold CV 4 times on the whole data set that also

includes unseen test data. Due to its higher prediction accuracy, ***Naïve Bayes*** is the suggested classifier to solve multi-class classification for ***negative heating gap prediction***. The conceptual web interface that introduces a negative performance gap prediction tool for heating demand can be seen in Appendix O. Selected attributes to solve the multi-class classification problem and the distribution of the values in the dataset are given in Figure 5.11.

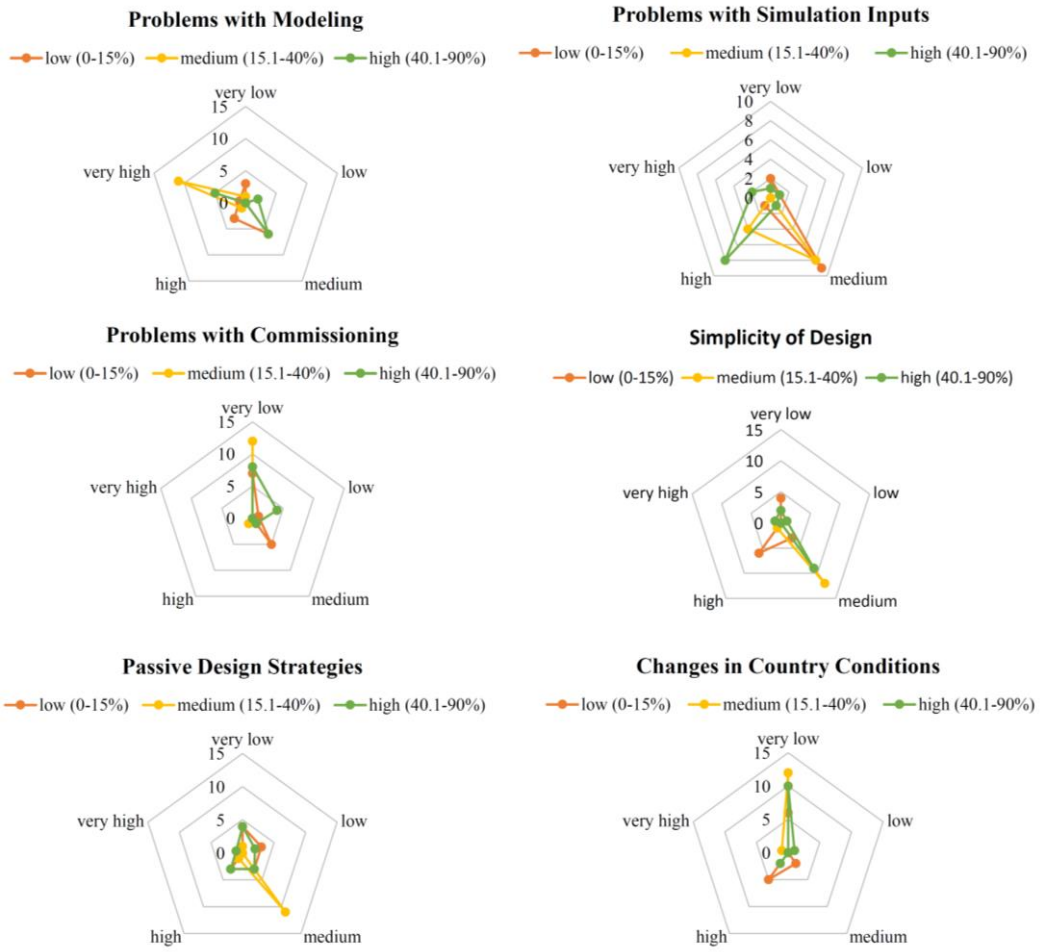


Figure 5.11 Selected Attributes and their Values in the Sixth Classification Problem

Moreover, the top two performing algorithms for each classification problem can be seen in Table 5.3 as a summary of the results.

Table 5.3 Top Two Performing Classifiers for Classification Problems

Results	Accuracy of the algorithm (%)	
	Rank: 1	**Rank:2**
Binary classification: Electricity Gap	KNN (79.00%)	SVM (75.00%)
Multi class classification: Positive Electricity Gap	Naïve Bayes (77.08%)	Random Forest (73.96%)
Multi class classification: Negative Electricity Gap	Naïve Bayes (83.85%)	KNN (82.29%)
Binary classification: Heating Gap	Naïve Bayes (72.50%)	SVM (68.50%)
Multi class classification: Positive Heating Gap	SVM (76.04%)	Random Forest (69.79%)
Multi class classification: Negative Heating Gap	Naïve Bayes (71.81%)	Random Forest (65.14%)

An overall comparison of the predictive accuracy of the classifiers shows that there was an obvious reduction in the performance of the classifiers when dealing with the gap related to heating demand in buildings. This tendency may be connected to the noise in the dataset mainly due to the following reasons:

- Having contradictory examples in the dataset. Analyzing the survey responses showed that although the response was overly pessimistic, the dataset contained cases in which the performance gap could still be negative or very low. Alternatively, while some cases showed a negative performance gap related to electricity demand, they showed a positive performance gap related to heating demand or vice versa. It might be that the complexity of building performance in some cases, cannot be explained with the current survey parameters. Therefore, several open-ended questions could be asked to participants to gather project-specific factors that cause the gap. Although literature review and semi-structured interviews showed that the causes of the performance gap were much

greater than those listed in the survey, many factors were eliminated to keep the survey short and simple.

- Incorrect answers due to misunderstanding of survey questions. This tendency can be categorized as attribute noise. Around 40 percent of the survey responses were gathered via a website. While conducting surveys, it was observed that two participants misread and interpreted some questions, especially those related to project management. This misunderstanding could be resolved over the phone. However, if similar misunderstandings occurred among respondents who completed the survey via the website, this would further increase the complexity of the classification problem due to their erroneous responses. Therefore, it is suggested to conduct the survey on a platform that allows the participants to easily ask questions to the researcher.

5.2 Research Limitations

This study cannot answer the need for ML models to deploy into production. To benefit more from ML classification, more data should be obtainable. This study indicated that the proposed method could work well in the future, and a good level of performance can be achieved especially for the prediction of the electricity performance gap in buildings. However, the models suggested for heating gap prediction provided poorer results than the other category. ML is suggested to be a powerful computational method when sufficient data is available. The main limitation of this study was trying to solve classification problems with a small number of data which is gathered through surveys and empirical articles. Although ML literature confirms that higher predictive accuracies can be also achieved in smaller datasets, it is critical to have access to quality data and a representative sample for the target population.

In this study, due to the length, and details of the first survey, the survey was designed twice to reduce lower response rates. However, the response rate was still limited during the field period (three and a half months). The reasons for the low response rate might be similar to the reasons of some of the target participants:

1) a nondisclosure agreement signed with clients,

2) not having access to the data,

3) not having the opportunity to obtain in-use metered data for buildings,

4)only having a couple of projects with information on operational energy consumption as well as energy modeling,

5) time limitations to obtain the client's permission to use them as a case study.

To increase the number of cases in the dataset, twenty cases from the literature (26% of the dataset) were included in the study. Moreover, to avoid researcher bias, (1) Cases were entered into the survey after an in-depth review of the articles twice considering the survey questions, and (2) the research strategy was reviewed by two academicians. Although collecting data from several resources enrich the dataset, since the responses collected from empirical articles were based on the interpretation of the researcher, this method can generate attribute noise in the dataset. However, this study demonstrates a generic framework can be developed with ML classification in energy performance gap prediction even with a small sample size. To achieve more responses to such a survey, the survey process could be kept longer, and surveys could be conducted face-to-face with the participants to increase the accuracy of the responses.

5.3 Recommendations for Future Work

This study contributes to building energy performance by demonstrating the potential of machine learning in the prediction of energy performance gaps in buildings. Although the prediction of building energy performance utilizing ML methods is common, there is room for the application of the classification methods in energy performance gap prediction. In this sense, this study opens a new perspective by combining project risks and the ML classification method.

Since the dataset used in this research has a great influence on the main findings, different datasets would provide different results. The dataset used in ML classification problems is derived mainly from a web-based survey that lists several risks encountered during the project process and requires evaluation of these risks by professionals who worked on these buildings. The evaluations can be questionable as they depend on their subjective judgment and perceptions. A combination of qualitative and quantitative methods can be suggested in future studies to increase the reliability and objectivity of the risk evaluation. The web-based survey suggested in this study asks questions that should be answered based on the access of the professionals to quantitative data as well (data gathered through detailed building monitoring, etc.). Since it was not clear if participants have access to the data, these factors (assumptions category in the survey) were excluded from the study while training ML algorithms. Future studies may try to solve this problem by collecting data from a larger number of buildings in which the experts conducted quantitative analysis on building energy performance (e.g., sensitivity analysis, building monitoring, etc.). In this way, it can be ensured that the qualitative judgments are based on quantitative assessments. Another alternative can be to apply fuzzy set theory to handle subjective, imprecise, ambiguous judgments of the respondents.

Moreover, a low level of performance gap might be considered negligible by some project stakeholders, whereas medium and high levels of gap might be more of a concern. Although this research suggests a tool to detect energy performance gaps in buildings at various levels and provided some suggestions to reduce energy performance gaps, it does not provide comprehensive mitigation techniques to reduce them. To do that, other methods are necessary to depict the relationships between input inputs and inputs with outputs. Therefore, future studies can suggest decision support tools that both provide a prediction for the energy performance gap and mitigation techniques to reduce the gap.

During semi-structured interviews, it was observed that participants generally were more self-confident and willing to explain the building performance when the actual energy consumption was less than their design predictions in a building (a negative performance gap). The reputation dangers of the energy performance gap make it a sensitive topic although its existence has been studied for decades all around the world. Therefore, project stakeholders might be hesitant to be involved in such studies. To better understand the problems in this area, instead of seeing a positive performance gap as a failure and a negative performance gap as a success, future studies can objectively compare the two situations and examine the advantages and disadvantages of both from a life-cycle perspective.

REFERENCES

Abela, A., Hoxley, M., McGrath, P., & Goodhew, S. (2016). An investigation of the appropriateness of current methodologies for energy certification of Mediterranean housing. *Energy and Buildings*, 130, 210–218. https://doi.org/10.1016/j.enbuild.2016.07.056

Agha-Hossein, M. M., El-Jouzi, S., Elmualim, A. A., Ellis, J., & Williams, M. (2013). Post-occupancy studies of an office environment: Energy performance and occupants' satisfaction. *Building and Environment*, *69*, 121–130. https://doi.org/10.1016/j.buildenv.2013.08.003

Ahn, K. U., Kim, D. W., Park, C. S., & de Wilde, P. (2017). Predictability of occupant presence and performance gap in building energy simulation. *Applied Energy*, *208*, 1639–1652. https://doi.org/10.1016/j.apenergy.2017.04.083

Aïssani, A., Chateauneuf, A., Fontaine, J. P., & Audebert, P. (2016). Quantification of workmanship insulation defects and their impact on the thermal performance of building facades. *Applied Energy*, *165*, 272–284. https:// doi.org/ 10.1016/j.apenergy.2015.12.040

Akintoye, A. S., & MacLeod, M. J. (1997). Risk analysis and management in construction. *International Journal of Project Management*, *15*(1), 31–38. https://doi.org/10.1016/S0263-7863(96)00035-X

Alam, M., Phung, V. M., Zou, P. X. W., & Sanjayan, J. (2017). Risk identification and assessment for construction and commissioning stages of building energy retrofit projects. *International Conference of Advancement of Construction Management and Real Estate*, (November).

Al-Bahar, J. F., & Crandall, K. C. (1991). Systematic risk management approach for construction projects. ASCE. *Engineering*, *116*(3), 533–546. https://doi.org/ 10.1061/(ASCE)0733-9364(1990)116

Alencastro, J., Fuertes, A., & de Wilde, P. (2018). The relationship between quality defects and the thermal performance of buildings. *Renewable and Sustainable Energy Reviews*, *81*(August 2017), 883–894. https://doi.org/ 10.1016/ j.rser. 2017.08.029

Ali, Q., Thaheem, M. J., Ullah, F., & Sepasgozar, S. M. E. (2020). The performance gap in energy-efficient office buildings: How the occupants can help? *Energies*, *13*(6), 1–27. https://doi.org/10.3390/en13061480

Alpaydın, E. (2010). *Introduction to machine learning*. (2nded.). The MIT Press. https://mitpress.mit.edu/books/introduction-machine-learning

Amasyali, K., & El-Gohary, N. M. (2018). A review of data-driven building energy consumption prediction studies. *Renewable and Sustainable Energy Reviews*, *81*(September 2017), 1192–1205. https://doi.org/10.1016/j.rser.2017.04.095

Anand, P., Deb, C., Yan, K., Yang, J., Cheong, D., & Sekhar, C. (2021). Occupancy-based energy consumption modelling using machine learning algorithms for institutional buildings. *Energy and Buildings*, *252*, 111478. https:// doi.org/ 10.1016/j.enbuild.2021.111478

Arjunan, P., Poolla, K., & Miller, C. (2020). EnergyStar++: Towards more accurate and explanatory building energy benchmarking. *Applied Energy*, *276*(January), 115413. https://doi.org/10.1016/j.apenergy.2020.115413

Ascione, F., Bianco, N., Böttcher, O., Kaltenbrunner, R., & Vanoli, G. P. (2016). Net zero-energy buildings in Germany: Design, model calibration and lessons learned from a case-study in Berlin. *Energy and Buildings*, *133*, 688–710. https://doi.org/10.1016/j.enbuild.2016.10.019

Attia, S., Hensen, J. L. M., Beltrán, L., & De Herde, A. (2012). Selection criteria for building performance simulation tools: Contrasting architects' and engineers' needs. *Journal of Building Performance Simulation*, *5*(3), 155–169. https://doi.org/10.1080/19401493.2010.549573

Attia, S., Hamdy, M., O'Brien, W., & Carlucci, S. (2013). Assessing gaps and needs for integrating building performance optimization tools in net zero energy buildings design. *Energy and Buildings*, *60*, 110–124. https://doi.org /10.1016/j.enbuild.2013.01.016

Ayhan, M. (2019). *Development of dispute prediction and resolution method selection models for construction disputes*. [Doctoral dissertation, Middle East Technical University]. Institutional Repository at the Middle East Technical University. https://open.metu.edu.tr/handle/11511/45067

Awad, M., & Khanna, R. (2015). *Efficient learning machines: Theories, concepts, and applications for engineers and system designers*. (1sted.). Apress. https://doi.org/10.1007/978-1-4302-5990-9

Baas, D.,Maaijen, R.,Maasen,W., Zeiler,W. (2016). Predicting energy savings for energy performance contracting : the impact of the energy performance gap.*Rehva Journal*, October 2017, 59–62.

Baborska-Narozny, M., & Stevenson, F. (2015). Continuous mechanical ventilation in housing - understanding the gap between intended and actual performance and use. *Energy Procedia*, *83*, 167–176. https://doi.org/ 10.1016/ j.egypro. 2015.12.207

Baird, G., Donn, M. R., Pool, F., Brander, W. D. S., & Seong Aun, C. (2018). Energy performance of buildings. *Energy Performance Buildings*. https://doi.org/ 10.1201/9781351071697

Bastanlar, Y., & Ozuysal, M. (2014). Introduction to machine learning. *Methods in molecular biology (Clifton, N.J.)* (Vol. 1107). Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/24272434

Bauer, E., & Kohavi, R. (1999). Empirical comparison of voting classification algorithms: bagging, boosting, and variants. *Machine Learning*, *36*(1), 105–139. https://doi.org/10.1023/a:1007515423169

Bhuiyan, S. I., Jones, K., & Wanigarathna, N. (2015). An approach to sustainable refurbishment. *31st Annual ARCOM Conference*, (September), 1093–1102.

Birchall, S. J. (2011). *An appraisal of the performance of a "green" office building*. [Doctoral dissertation, University of Leeds]. White Rose eTheses Online. Retrieved from http://etheses.whiterose.ac.uk/2620/

Brady, L., & Hanmer-Dwight, R. (2016). The management and control of energy at the design stage of buildings. *Proceedings of the 32nd Annual ARCOM Conference, ARCOM 2016*, *2* (September), 1171–1180.

Bramer, M. (2020). *Principles of data mining*. (4thed.). Springer. https:// doi.org/10.1007/978-1-4471-7493-6

Bros-Williamson, J., Garnier, C., & Currie, J. I. (2016). A longitudinal building fabric and energy performance analysis of two homes built to different energy principles. *Energy and Buildings*, *130*, 578–591. https://doi.org/ 10.1016/j.enbuild.2016.08.052

Borgstein, E. H., Lamberts, R., & Hensen, J. L. M. (2016). Evaluating energy performance in non-domestic buildings: A review. *Energy and Buildings*, *128*, 734–755. https://doi.org/10.1016/j.enbuild.2016.07.018

Bouckaert, R. R., Frank, E., Kirkby, R., Reutemann, P., Seewald, A., & Scuse, D. (2013). WEKA manual for version 3-7-2. Retrieved from http://netcologne.dl.sourceforge.net/project/weka/documentation/3.7.x/Weka Manual-3-7-12.pdf

Boyd, P., & Schweber, L. (2018). Unintended consequences: institutional artefacts, closure mechanisms and the performance gap. *Building Research and Information*, *46*(1), 10–22. https://doi.org/10.1080/09613218.2017.1331096

Brejnrod, K. N., Kalbar, P., Petersen, S., & Birkved, M. (2017). The absolute environmental performance of buildings. *Building and Environment*, *119*, 87–98. https://doi.org/10.1016/j.buildenv.2017.04.003

Bros-Williamson, J., Garnier, C., & Currie, J. I. (2016). A longitudinal building fabric and energy performance analysis of two homes built to different energy principles. *Energy and Buildings*, *130*, 578–591. https://doi.org/ 10.1016/j.enbuild.2016.08.052

Brownlee, J. (2018a). *Machine learning mastery with WEKA: Analyze data, develop models and work through projects*. (v1.5). https:// machinelearningmastery. com/machine-learning-mastery-weka/

Brownlee, J. (2018b). *Master machine learning algorithms: Discover how they work and implement them from scratch.* (v1.13). https://machinelearningmastery. com/master-machine-learning-algorithms/

Buratti, C., Barbanera, M., & Palladino, D. (2014). An original tool for checking energy performance and certification of buildings by means of Artificial Neural Networks. *Applied Energy*, *120*, 125–132. https://doi.org/ 10.1016/ j.apenergy.2014.01.053

Burman, E., Kimpian, J., & Mumovic, D. (2018). Building schools for the future: Lessons learned from performance evaluations of five secondary schools and academies in england. *Frontiers in Built Environment*, *4*(April), 1–16. https://doi.org/10.3389/fbuil.2018.00022

Cai, X., Zhou, C., Zhou, L., & Xu, Q. (2019). A bibliometric analysis of energy performance contracting research from 2008 to 2018. *Sustainability 2019 (11-3548)*, 1–23. https://doi.org/10.7717/peerj.7992

Calì, D., Osterhage, T., Streblow, R., & Müller, D. (2016). Energy performance gap in refurbished German dwellings: Lesson learned from a field test. *Energy and Buildings*, *127*, 1146–1158. https://doi.org/10.1016/j.enbuild.2016.05.020

Carpino, C., Loukou, E., Heiselberg, P., & Arcuri, N. (2020). Energy performance gap of a nearly Zero Energy Building (nZEB) in Denmark: the influence of occupancy modelling. *Building Research and Information*, *0*(0), 1–23. https://doi.org/10.1080/09613218.2019.1707639

Chamasemani, F. F., & Singh, Y. P. (2011). Multi-class Support Vector Machine (SVM) classifiers - An application in hypothyroid detection and classification. *Proceedings - 2011 6th International Conference on Bio-Inspired Computing: Theories and Applications, BIC-TA 2011*, 351–356. https://doi.org/ 10.1109/BIC-TA.2011.51

Chang, T., Deng, X., Zuo, J., & Yuan, J. (2018). Political risks in Central Asian Countries: Factors and strategies. *Journal of Management in Engineering*, *34*(2), 04017059. https://doi.org/10.1061/(asce)me.1943-5479.0000588

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, *16*(Sept. 28), 321–357. Retrieved from https://www3.nd.edu/~dial/ publications/chawla2002smote.pdf

Chen, W., & Zhang, L. (2022). Building vulnerability assessment in seismic areas using ensemble learning: A Nepal case study. *Journal of Cleaner Production*, *350*(February), 131418. https://doi.org/10.1016/j.jclepro.2022.131418

Chen, R., & Tsay, Y. S. (2021). An integrated sensitivity analysis method for energy and comfort performance of an office building along the Chinese coastline. *Buildings*, *11*(8). https://doi.org/10.3390/buildings11080371

CIBSE. (2013). Evaluating operational energy performance of buildings at the design stage, 1–30. https://doi.org/978-1-906846-38-1

Chvatal, K. M. S., & Corvacho, H. (2009). The impact of increasing the building envelope insulation upon the risk of overheating in summer and an increased energy consumption. *Journal of Building Performance Simulation*, *2*(4), 267–282. https://doi.org/10.1080/19401490903095865

Coakley, D., Raftery, P., & Keane, M. (2014). A review of methods to match building energy simulation models to measured data. *Renewable and Sustainable Energy Reviews*, *37*, 123–141. https://doi.org /10.1016/ j.rser. 2014.05.007

Corry, E., Pauwels, P., Hu, S., Keane, M., & O'Donnell, J. (2015). A performance assessment ontology for the environmental and energy management of

buildings. *Automation in Construction*, *57*, 249–259. https://doi.org/10.1016/j.autcon.2015.05.002

Coşkun, C. (2019). *Development of a risk assessment method for sustainable construction of megaprojects*. [Master's thesis, Middle East Technical University]. Institutional Repository at the Middle East Technical University. https://open.metu.edu.tr/handle/11511/44237

Cretu, O., Stewart, B.R., & Berends, T. (2011). *Risk management for design and construction*. (1st ed.). John Wiley & Sons. https://onlinelibrary. wiley.com/ doi/book/10.1002/9781118984017

Cuerda, E., Guerra-Santin, O., Sendra, J. J., & Neila, F. J. (2020). Understanding the performance gap in energy retrofitting: Measured input data for adjusting building simulation models. *Energy and Buildings*, *209*, 109688. https://doi.org/10.1016/j.enbuild.2019.109688

Darko, A., & Chan, A. P. C. (2018). Strategies to promote green building technologies adoption in developing countries: The case of Ghana. *Building and Environment*, *130* (October 2017), 74–84. https://doi.org /10.1016/ j.buildenv.2017.12.022

Dasgupta, A., Prodromou, A., & Mumovic, D. (2012). Operational versus designed performance of low carbon schools in England: Bridging a credibility gap. *HVAC and R Research*, *18*(1–2), 37–50.https://doi.org/ 10.1080/ 10789669. 2011.614318

Day, J. K., & Gunderson, D. E. (2015). Understanding high performance buildings : The link between occupant knowledge of passive design systems, corresponding behaviors, occupant comfort and environmental satisfaction. *Building and Environment*, *84*, 114–124. https://doi.org/10.1016/ j.buildenv. 2014.11.003

Delgado, J. M. P. Q. (2016). Indoor environmental quality of school buildings. *Defect and Diffusion Forum*, *369*, 24–29. https://doi.org/ 10.4028/ www.scientific.net/DDF.369.24

Delgarm, N., Sajadi, B., Azarbad, K., & Delgarm, S. (2018). Sensitivity analysis of building energy performance: A simulation-based approach using OFAT and variance-based sensitivity analysis methods. *Journal of Building Engineering*, *15*(November 2017), 181–193. https://doi.org/10.1016/j.jobe.2017.11.020

Delzendeh, E., Wu, S., Lee, A., & Zhou, Y. (2017). The impact of occupants' behaviours on building energy analysis: A research review. *Renewable and Sustainable Energy Reviews*, *80*(May), 1061–1071. https://doi.org/10.1016/j.rser.2017.05.264

Demanuele, C. (2008). *Investigating the discrepancy between the predicted and actual energy performance of buildings*. [Master's Thesis, University College London]. ProQuest. https://www.proquest.com/docview/1432125085

De Wilde, P. (2014). The gap between predicted and measured energy performance of buildings: A framework for investigation. *Automation in Construction*, *41*, 40–49. https://doi.org/10.1016/j.autcon.2014.02.009

De Wilde, P. (2018). *Building performance analysis*. John Wiley& Sons Ltd. https://onlinelibrary.wiley.com/doi/book/10.1002/9781119341901

Dikmen, I., Talat Birgonul, M., & Fidan, G. (2008). Assessment of project vulnerability as a part of risk management in construction. *Procs. of Joint 2008 CIB W065/W055 Symposium*. Retrieved from https://www.irbnet.de/daten/iconda/CIB17595.pdf

Dollard, T. (2018). *Designed to perform: An illustrated guide to delivering energy efficient homes*. (1st ed.). RIBA Publishing. https://doi.org/ 10.4324/9780429345838

Doylend, N. (2015). *Evaluating building energy performance : A lifecycle risk management methodology*. [Doctoral dissertation, Loughbourgh University]. Institutional Repository at the Loughbourgh University. https://hdl. handle.net/2134/18022

Eon, C., Breadsell, J. K., Byrne, J., & Morrison, G. M. (2020). The discrepancy between as-built and as-designed in energy efficient buildings: A rapid review. *Sustainability (Switzerland)*, *12*(16). https://doi.org/10.3390/SU12166372

Fedoruk, L. E., Cole, R. J., Robinson, J. B., & Cayuela, A. (2015). Learning from failure: Understanding the anticipated-achieved building energy performance gap. *Building Research and Information*, *43*(6), 750–763. https://doi.org/10.1080/09613218.2015.1036227

Fidan, G. (2008). *A risk and vulnerability ontology for construction projects*. [Master's thesis, Middle East Technical University]. Institutional Repository at the Middle East Technical University. https://etd.lib.metu.edu.tr/upload/12609793/index.pdf

Fidan, G., Dikmen, I., Tanyer, A. M., & Birgonul, M. T. (2011). Ontology for relating risk and vulnerability to cost overrun in international projects. *Journal of Computing in Civil Engineering*, *25*(4), 302–315. https://doi.org/ 10.1061/(ASCE)CP.1943-5487.0000090

Figueiredo, A., Kämpf, J., Vicente, R., Oliveira, R., & Silva, T. (2018). Comparison between monitored and simulated data using evolutionary algorithms: Reducing the performance gap in dynamic building simulation. *Journal of Building Engineering*, *17*(February), 96–106. https://doi.org/ 10.1016/ j.jobe.2018.02.003

Fosas, D., Coley, D. A., Natarajan, S., Herrera, M., Fosas de Pando, M., & Ramallo-Gonzalez, A. (2018). Mitigation versus adaptation: Does insulating dwellings increase overheating risk? *Building and Environment*, *143*(May), 740–759. https://doi.org/10.1016/j.buildenv.2018.07.033

Fouladgar, M. M., Yazdani-Chamzini, A., & Zavadskas, E. K. (2012). Risk evaluation of tunneling projects. *Archives of Civil and Mechanical Engineering*, *12*(1), 1–12. https://doi.org/10.1016/j.acme.2012.03.008

Fokaides, P. A., Maxoulis, C. N., Panayiotou, G. P., Neophytou, M. K. A., & Kalogirou, S. A. (2011). Comparison between measured and calculated energy performance for dwellings in a summer dominant environment. *Energy and Buildings*, *43*(11), 3099–3105. https://doi.org/10.1016/j.enbuild.2011.08.005

Friedman, N., D. Geiger, and M. Goldszmidt. (1997). Bayesian network classifiers. *Machine Learning*, 29 (2):131–163. https://doi.org/10.1023/ A:1007465528199

Fung, I. W. H., Lo, T. Y., & Tung, K. C. F. (2012). Towards a better reliability of risk assessment: Development of a qualitative & quantitative risk evaluation model (Q 2 REM) for different trades of construction works in Hong Kong. *Accident Analysis and Prevention*, *48*, 167–184. https://doi.org/ 10.1016/ j.aap.2011.05.011

Galar, M., Fernandez, A., Barrenechea, E., & Bustince, H. (2011). A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches. *IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications and Reviews publication information*, 41(4), 1–22. https://doi.org/10.1109/TSMCC.2011.2157553.

Galvin, R. (2014). Making the 'rebound effect' more useful for performance evaluation of thermal retrofits of existing homes: Defining the 'energy savings

deficit' and the 'energy performance gap.' *Energy and Buildings*, *69*, 515–524. https://doi.org/10.1016/J.ENBUILD.2013.11.004

Garbuzova-Schlifter, M., & Madlener, R. (2016). AHP-based risk analysis of energy performance contracting projects in Russia. *Energy Policy*, *97*, 559–581. https://doi.org/10.1016/j.enpol.2016.07.024

García, S., Luengo, J., Herrera, F. (2015). *Data preprocessing in data mining*. Springer. https://doi.org/10.1007/978-3-319-10247-4

Garwood, T. (2019). *Closing the performance gap in building energy modelling through digital survey methods and automated reconstruction*. [Doctoral dissertation, The University of Sheffield]. White Rose eTheses Online. https://etheses.whiterose.ac.uk/25583/

Gaspar, P., Carbonell, J., & Oliveira, J. L. (2012). On the parameter optimization of Support Vector Machines for binary classification. *Journal of Integrative Bioinformatics*, *9*(3), 201. https://doi.org/10.1515/jib-2012-201

Geng, Y., Ji, W., Wang, Z., Lin, B., & Zhu, Y. (2019). A review of operating performance in green buildings: Energy use, indoor environmental quality and occupant satisfaction. *Energy and Buildings*, *183*, 500–514. https://doi.org/10.1016/j.enbuild.2018.11.017

Geraldi, M. S., & Ghisi, E. (2020). Building-level and stock-level in contrast: A literature review of the energy performance of buildings during the operational stage. *Energy and Buildings*, *211*, 109810. https://doi.org/10.1016/j.enbuild.2020.109810

Goulden, S., Erell, E., Pearlmutter, D., & Garb, Y. (2020). Embracing uncertainty in building energy efficiency policy: A case study of a building energy standard. *Energy Policy*, *139* (December 2019), 111303. https://doi.org/ 10.1016 /j.enpol.2020.111303

Gondia, A., Asce, S. M., Siam, A., El-dakhakhni, W., Asce, F., & Nassar, A. H. (2020). Machine Learning Algorithms for Construction Projects Delay Risk Prediction. *American Society of Civil Engineers. 146*(1), 1–16. https://doi.org/ 10.1061/(ASCE)CO.1943-7862.0001736

Gram-Hanssen, K., & Georg, S. (2018). Energy performance gaps: promises, people, practices. *Building Research and Information*, *46*(1), 1–9. https://doi.org/ 10.1080/09613218.2017.1356127

Grover, A. C. (2019). Understanding the role of Occupants in green building energy performance. *Proceedings of the International Conference of Architectural Science Association*, *2019-Novem*, 547–556. https://anzasca.net/wp-content/uploads/2020/03/57-Understanding-the-role-of-Occupants-in-Green-Building-Energy-Performance.pdf

Grillone, B., Danov, S., Sumper, A., Cipriano, J., & Mor, G. (2020). A review of deterministic and data-driven methods to quantify energy efficiency savings and to predict retrofitting scenarios in buildings. *Renewable and Sustainable Energy Reviews*, *131*(March), 110027. https://doi.org/10. 1016/j.rser. 2020.110027

Gupta, R., Gregg, M., & Cherian, R. (2013). Tackling the performance gap between design intent and actual outcomes of new low / zero carbon housing. *ECEEE Summer Study Proceedings*, 1315–1328.

Gupta, R., & Gregg, M. (2016). Empirical evaluation of the energy and environmental performance of a sustainably-designed but under-utilized institutional building in the UK. *Energy and Buildings*, *128*, 68–80. https://doi.org/10.1016/j.enbuild.2016.06.081

Gupta, R., & Gregg, M. (2018). Assessing energy use and overheating risk in net zero energy dwellings in UK. *Energy and Buildings*, *158*(2018), 897–905. https://doi.org/10.1016/j.enbuild.2017.10.061

Güçyeter, B., & Günaydin, H. M. (2012). Optimization of an envelope retrofit strategy for an existing office building. *Energy and Buildings*, *55*(2012), 647–659. https://doi.org/10.1016/j.enbuild.2012.09.031

Guyot, D., Giraud, F., Simon, F., Corgier, D., Marvillet, C., & Tremeac, B. (2019). Overview of the use of artificial neural networks for energy-related applications in the building sector. *International Journal of Energy Research*, *43*(13), 6680–6720. https://doi.org/10.1002/er.4706

Hamza, N. (2013). *The sustainable corporate image and renewables. From technique to the sensory experience. Sustainability, Energy and Architecture: Case Studies in Realizing Green Buildings*. Elsevier. https://doi.org/ 10.1016/B978-0-12-397269-9.00004-9

Han,J., Kamber, M., Pei, J.(2012). *Data mining concepts and techniques. (3rd ed.).* Morgan Kaufmann. https://doi.org/10.1016/C-2009-0-61819-5

Hanna, A. S., Thomas, G., & Swanson, J. R. (2013). Construction risk identification and allocation: Cooperative approach. *Journal of Construction Engineering and Management*, *139*(9), 1098–1107. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000703

Hansen, A. R., Gram-Hanssen, K., & Knudsen, H. N. (2018). How building design and technologies influence heat-related habits. *Building Research and Information*, *46*(1), 83–98. https://doi.org/10.1080/09613218.2017.1335477

Hansson, S. O. (2010). Risk: Objective or subjective, facts or values. *Journal of Risk Research*, *13*(2), 231–238. https://doi.org/10.1080/13669870903126226

Hansson, S. O., & Aven, T. (2014). Is risk analysis scientific? *Risk Analysis, 34*(7), 1173–1183. https://doi.org/10.1111/risa.12230

Harrison, S., & Jiang, L. (2018). An investigation into the energy performance gap between the predicted and measured output of photovoltaic systems using dynamic simulation modelling software-A case study. *International Journal of Low-Carbon Technologies, 13*(1), 23–29. https://doi.org/10.1093/ijlct/ctx016

Hartkopf, V.H., Loftness, V.E., and Mill, P. A.D. (1986). "*The concept of total building performance and building diagnostics*," Building performance: Function, preservation, and rehabilitation, ASTM STP 901,G. Davis, Ed., American Society for Testing and Materials, Philadephia, pp.5-22.

Hashemi, H., Mousavi, S. M., & Mojtahedi, S. M. H. (2011). Bootstrap technique for risk analysis with interval numbers in bridge construction projects. *Journal of Construction Engineering and Management*, *137*(8), 600–608. https://doi.org/10.1061/ (ASCE)CO.1943-7862.0000344

Hatefi, S. M., Tamošaitienė,J. (2019). An integrated fuzzy DEMATEL-Fuzzy ANP model for evaluating construction projects by considering. *Journal of Civil Engineering and Management, 25*(2), 114–131. https://doi.org/ 10.3846/jcem. 2019.8280

He, H., & Ma, Y. (2011). *Imbalanced learning: Foundations, algorithms, and applications*. Wiley. https://doi.org/10.1002/9781118646106

Hensen, L.M.J., & Lamberts, R. (2019). *Building performance simulation for design and operation. (2nd ed.).* Routledge. https://doi.org/10.1201/ 9780429402296

Heo, Y. (2011). *Bayesian calibration of building energy models for energy retrofit decision-making under uncertainty*. [Doctoral Dissertation, Georgia Institute of

Technology]. Institutional Repository at the Georgia Institute of Technology. https://smartech.gatech.edu/handle/1853/42878

Heo, Y., Augenbroe, G., & Choudhary, R. (2013). Quantitative risk management for energy retrofit projects. *Journal of Building Performance Simulation*, *6*(4), 257–268. https://doi.org/10.1080/19401493.2012.706388

Herrando, M., Cambra, D., Navarro, M., de la Cruz, L., Millán, G., & Zabalza, I. (2016). Energy performance certification of faculty buildings in Spain: The gap between estimated and real energy consumption. *Energy Conversion and Management*, *125*(January 2014), 141–153. https://doi.org/ 10.1016/ j.enconman.2016.04.037

Hong, T., Yang, L., Hill, D., & Feng, W. (2014). Data and analytics to inform energy retrofit of high performance buildings. *Applied Energy*, *126*, 90–106. https:// doi.org/10.1016/j.apenergy.2014.03.052

Hong, T., Koo, C., Kim, J., Lee, M., & Jeong, K. (2015). A review on sustainable construction management strategies for monitoring, diagnosing, and retrofitting the building's dynamic energy performance: Focused on the operation and maintenance phase. *Applied Energy*, *155*, 671–707. https://doi.org /10.1016/ j.apenergy.2015.06.043

Hong, T., Yan, D., D'Oca, S., & Chen, C. fei. (2017).Ten questions concerning occupant behavior in buildings: The big picture. *Building and Environment*, *114*, 518–530. https://doi.org/10.1016/j.buildenv.2016.12.006

Hong, J., Kang, H., & Hong, T. (2020). Oversampling-based prediction of environmental complaints related to construction projects with imbalanced empirical-data learning. *Renewable and Sustainable Energy Reviews*, *134*(September), 110402. https://doi.org/10.1016/j.rser.2020.110402

Horrigan, M., O'Donnell, J., Corry, E., & Curry, A. E. (2015). A holistic life-cycle data analysis approach to bridge the energy performance gap in buildings. *14th International Conference of IBPSA - Building Simulation 2015, BS 2015, Conference Proceedings*, 2051–2058.

Imam, S., Coley, D. A., & Walker, I. (2017). The building performance gap: Are modellers literate? *Building Services Engineering Research and Technology*, *38*(3), 351–375. https://doi.org/10.1177/0143624416684641

International Energy Agency (2016). Total energy use in buildings: Analysis and evaluation methods (Annex 53). https://www.iea-ebc.org/Data/ publications /EBC_PSR_Annex53.pdf

International Electrotechnical Commission. (2019). *Risk management-Risk assessment techniques* (Standard No.31010). Retrived from https://www.iso.org/standard/72140.html

International Organization for Standardization. (2018). *Risk management guidelines.* (Standard No.31000). https://www.iso.org/standard/65694.html

IPEEC. (2019). Building energy performance gap issues: An international review. *International Partnership for Energy Efficiency Cooperation*, 1–36. Retrieved from https://ipeec.org/upload/publication_related_language/pdf/1626.pdf

Joshi,V.A. (2020). *Machine learning and artificial intelligence.* Springer. https://doi.org/10.1007/978-3-030-26622-6

Janda, K. B. (2011). Buildings don't use energy: People do. *Architectural Science Review*, *54*(1), 15–22. https://doi.org/10.3763/asre.2009.0050

Janser, M., Hubbuch, M., & Windlinger, L. (2020). Call for a definition and paradigm shift in energy performance gap research. *IOP Conference Series: Earth and Environmental Science*, *588*(5). https://doi.org/10.1088/1755-1315/588/5/052052

Japkowicz, N., & Shah, M. (2011). *Evaluating learning algorithms: a classification perspective.* Cambridge University Press. https://doi.org/10.1017/ CBO9780511921803

Jayasudha, K., & Vidivelli, B. (2016). Analysis of major risks in construction projects. *ARPN Journal of Engineering and Applied Sciences*, *11*(11), 6943–6950.

Jradi, M., Arendt, K., Sangogboye, F. C., Mattera, C. G., Markoska, E., Kjærgaard, M. B., … Jørgensen, B. N. (2018). ObepME: An online building energy performance monitoring and evaluation tool to reduce energy performance gaps. *Energy and Buildings*, *166*, 196–209. https://doi.org/10.1016/ j.enbuild.2018.02.005

Jung, J. H., Kim, D. Y., & Lee, H. K. (2016). The computer-based contingency estimation through analysis cost overrun risk of public construction project.

KSCE Journal of Civil Engineering, *20*, 1119–1130. https://doi.org/10.1007/s12205-015-0184-8

Kadyrova, N. O., & Pavlova, L. V. (2018). An analysis of methods for tuning a Support-Vector Machine for binary classification. *Biophysics (Russian Federation)*, *63*(6), 994–1003. https://doi.org/10.1134/S0006350918060131

Kampelis, N., Gobakis, K., Vagias, V., Kolokotsa, D., Standardi, L., Isidori, D., … Santamouris, M. (2017). Evaluation of the performance gap in industrial, residential & tertiary near-Zero energy buildings. *Energy and Buildings*, *148*, 58–73. https://doi.org/10.1016/j.enbuild.2017.03.057

Karakhan, A. A., & Gambatese, J. A. (2017). Identification, quantification, and classification of potential safety risk for sustainable construction in the United States, *143*(7). https://doi.org/10.1061/ (ASCE)CO.1943-7862.0001302

Khoury, J., Alameddine, Z., & Hollmuller, P. (2017). Understanding and bridging the energy performance gap in building retrofit. *Energy Procedia*, *122*, 217–222. https://doi.org/10.1016/J.EGYPRO.2017.07.348

Kifokeris, D., & Xenidis, Y. (2019). Risk source-based constructability appraisal using supervised machine learning. *Automation in Construction*, *104*(April), 341–359. https://doi.org/10.1016/j.autcon.2019.04.012

Kontokosta, C. E., & Tull, C. (2017). A data-driven predictive model of city-scale energy use in buildings. *Applied Energy*, *197*, 303–317. https://doi.org/10.1016/j.apenergy.2017.04.005

Korjenic, A., & Bednar, T. (2012). Validation and evaluation of total energy use in office buildings: A case study. *Automation in Construction*, 23, 64–70. https://doi.org/10.1016/j.autcon.2012.01.001

Krechowicz, M. (2020). Qualitative risk assessment of passive house design and construction processes. *IOP Conference Series: Materials Science and Engineering*, *960*(4). https://doi.org/10.1088/1757-899X/960/4/042068

Larose T.D., Larose, D.C. (2014). *Discovering knowledge in data: An introduction to data mining*. Wiley. https://doi.org/10.1002/9781118874059

Lee, W. L., Yik, F. W. H., & Burnett, J. (2007). Assessing energy performance in the latest versions of Hong Kong Building Environmental Assessment Method (HK-BEAM). *Energy and Buildings*, *39*(3), 343–354. https://doi.org/10.1016/j.enbuild.2006.08.003

Lee, P., Lam, P. T. I., & Lee, W. L. (2015). Risks in Energy Performance Contracting (EPC) projects. *Energy and Buildings*, *92*, 116–127. https://doi.org/ 10.1016/j.enbuild.2015.01.054

Liang, J., Qiu, Y., & Hu, M. (2019). Mind the energy performance gap: Evidence from green commercial buildings. *Resources, Conservation and Recycling*, *141*(June 2018), 364–377. https://doi.org/10.1016/j.resconrec.2018.10.021

Li, Z., Han, Y., & Xu, P. (2014). Methods for benchmarking building energy consumption against its past or intended performance: An overview. *Applied Energy*, *124*, 325–334. https://doi.org/10.1016/j.apenergy.2014.03.020

Li, Y., Kubicki, S., Guerriero, A., & Rezgui, Y. (2019a). Review of building energy performance certification schemes towards future improvement. *Renewable and Sustainable Energy Reviews*, *113*(February). https://doi.org/ 10.1016/ j.rser.2019.109244

Li, J., Yu, Z. (Jerry), Haghighat, F., & Zhang, G. (2019b). Development and improvement of occupant behavior models towards realistic building performance simulation: A review. *Sustainable Cities and Society*, *50*(February), 101685. https://doi.org/10.1016/j.scs.2019.101685

Li, T., Zhang, C., & Ogihara, M. (2004). A comparative study of feature selection and multiclass classification methods for tissue classification based on gene expression. *Bioinformatics*, *20*(15), 2429–2437. https://doi.org/ 10.1093/ bioinformatics/bth267

Liu, F. (2015). *Integrated change and knowledge management for energy efficient retrofit projects*. [Doctoral Dissertation, the Pennsylvania State University]. Institutional Repository at the Pennsylvania State University. https://etda.libraries.psu.edu/files/final_submissions/11328

Liu, C., Kershaw, T., Eames, M. E., & Coley, D. A. (2016a). Future probabilistic hot summer years for overheating risk assessments. *Building and Environment*, *105*, 56–68. https://doi.org/10.1016/j.buildenv.2016.05.028

Liu, J., Zhao, X., & Yan, P. (2016b). Risk paths in international construction projects: Case study from Chinese contractors. *Journal of Construction Engineering and Management*, *142*(6), 1–11. https://doi.org/ 10.1061/ (ASCE)CO.1943-7862.0001116

Liu, J., & Zio, E. (2019). Integration of feature vector selection and support vector machine for classification of imbalanced data. *Applied Soft Computing Journal*, *75*, 702–711. https://doi.org/10.1016/j.asoc.2018.11.045

Liu, P., Lin, B., Wu, X., & Zhou, H. (2019). Bridging energy performance gaps of green office buildings via more targeted operations management: A system dynamics approach. *Journal of Environmental Management*, *238* (September 2018), 64–71. https://doi.org/10.1016/j.jenvman.2019.02.111

Littlewood, J. R., & Smallwood, I. (2015). Testing building fabric performance and the impacts upon occupant safety, energy use and carbon inefficiencies in dwellings. *Energy Procedia*, *83*, 454–463. https://doi.org/10.1016/ j.egypro.2015.12.165

Love, P., & Bullen, P. A. (2009). Toward the sustainable adaptation of existing facilities. *Facilities*, *27*(9–10), 357–367. https://doi.org/ 10.1108/ 02632770910969603

Loucari, C., Taylor, J., Raslan, R., Oikonomou, E., & Mavrogianni, A. (2016). Retrofit solutions for solid wall dwellings in England: The impact of uncertainty upon the energy performance gap. *Building Services Engineering Research and Technology*, *37*(5), 614–634. https://doi.org/10.1177/0143624416647758

Lu, M., & Lai, J. H. K. (2019). Building energy: A review on consumptions, policies, rating schemes and standards. *Energy Procedia*, *158*, 3633–3638. https://doi.org/10.1016/j.egypro.2019.01.899

Ma, Y., & Guo, G. (2014). *Support vector machines applications*. (1st ed.). Springer. https://doi.org/10.1007/978-3-319-02300-7

Magoulès, F., & Zhao, H. X. (2016). *Machine learning in building energy analysis*. Wiley. https://doi.org/10.1002/9781118577691.ch1

Mahdavi, A., & Berger, C. (2019). Predicting buildings' energy use: Is the occupant-centric "Performance Gap" research program Ill-advised? *Frontiers in Energy Research*, *7*(November), 1–11. https://doi.org/ 10.3389/ fenrg.2019.00124

Majcen, D., Itard, L., & Visscher, H. (2015). Statistical model of the heating prediction gap in Dutch dwellings: Relative importance of building, household and behavioral characteristics. *Energy and Buildings*, *105*(2015), 43–59. https://doi.org/10.1016/j.enbuild.2015.07.009

Marshall, A., Fitton, R., Swan, W., Farmer, D., Johnston, D., Benjaber, M., & Ji, Y. (2017). Domestic building fabric performance: Closing the gap between the in situ measured and modelled performance. *Energy and Buildings*, *150*, 307–317. https://doi.org/10.1016/j.enbuild.2017.06.028

Mashford, K., & Reason, L. (2014). *How much energy does your building use*? (1st ed.). Routledge. https://doi.org/10.4324/9781351274760

Mastrucci, A., Pérez-López, P., Benetto, E., Leopold, U., & Blanc, I. (2017). Global sensitivity analysis as a support for the generation of simplified building stock energy models. *Energy and Buildings*, *149*, 368–383. https://doi.org/10.1016/j.enbuild.2017.05.022

Menezes, A. C., Cripps, A., Bouchlaghem, D., & Buswell, R. (2012). Predicted vs. actual energy performance of non-domestic buildings: Using post-occupancy evaluation data to reduce the performance gap. *Applied Energy*, *97*, 355–364. https://doi.org/10.1016/j.apenergy.2011.11.075

Mhalas, A., Kassem, M., Crosbie, T., & Dawood, N. (2013). A visual energy performance assessment and decision support tool for dwellings. *Visualization in Engineering*, *1*(1). https://doi.org/10.1186/2213-7459-1-7

Mills, E., Kromer, S., Weiss, G., & Mathew, P. A. (2006). From volatility to value: Analysing and managing financial and performance risk in energy savings projects. *Energy Policy*, *34*(2 SPEC. ISS.), 188–199. https://doi.org/10.1016/j.enpol.2004.08.042

Mirzaie, S., & Menzies, G. (2017). How can the combination of BREEAM and soft landings successfully deliver a low energy, comfortable building? *Proceedings of 33rd PLEA International Conference: Design to Thrive, PLEA 2017*, *1*, 511–518.

Mohammadi, S. (2015). *The future of green building contracts in Turkey*. [Master's thesis, Middle East Technical University]. Institutional Repository at the Middle East Technical University. https://etd.lib. metu.edu.tr/upload /12619263/index.pdf

Mohareb, E., Hashemi, A., Shahrestani, M., & Sunikka-Blank, M. (2017). Retrofit planning for the performance gap: Results of a workshop on addressing energy, health and comfort needs in a protected building. *Energies*, *10*(8). https://doi.org/10.3390/en10081177

Mojic, I., Lehmann, M., Van Velsen, S., & Haller, M. (2019). ImmoGap - Analysis of the performance gap of apartment buildings. *E3S Web of Conferences*, *111*(2019). https://doi.org/10.1051/e3sconf/201911104016

Mortaz, E. (2020). Imbalance accuracy metric for model selection in multi-class imbalance classification problems. *Knowledge-Based Systems*, *210*, 106490. https://doi.org/10.1016/j.knosys.2020.106490

Montazami, A., Gaterell, M., & Nicol, F. (2015). A comprehensive review of environmental design in UK schools: History, conflicts and solutions. *Renewable and Sustainable Energy Reviews*, *46*, 249–264. https://doi.org/10.1016/j.rser.2015.02.012

Mun, S. H., Kwak, Y., & Huh, J. H. (2019). A case-centered behavior analysis and operation prediction of AC use in residential buildings. *Energy and Buildings*, *188–189*, 137–148. https://doi.org/10.1016/j.enbuild.2019.02.012

Niu, S., Pan, W., & Zhao, Y. (2016). A virtual reality integrated design approach to improving occupancy information integrity for closing the building energy performance gap. *Sustainable Cities and Society*, *27*, 275–286. https://doi.org/10.1016/j.scs.2016.03.010

Nižetić, S., & Papadopoulos, A. M. (2017). Concept of building evaluation methodology for gap estimation between designed and achieved energy savings. *Procedia Environmental Sciences*, *38*, 538–545. https://doi.org/10.1016/j.proenv.2017.03.118

Nolan, P. D. (2019). *Handbook of fire and explosion protection engineering principles for oil, gas, chemical, and related facilities*. (2nd ed.). Gulf Professional Publishing. https://www.elsevier.com/books/handbook-of-fire-and-explosion-protection-engineering-principles/nolan/978-1-4377-7857-1

Olu-Ajayi, R., Alaka, H., Sulaimon, I., Sunmola, F., & Ajayi, S. (2022). Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques. *Journal of Building Engineering*, *45*(September 2021), 103406. https://doi.org/10.1016/j.jobe.2021.103406

Ortiz, M. A., Kurvers, S. R., & Bluyssen, P. M. (2017). A review of comfort, health, and energy use: Understanding daily energy use and wellbeing for the development of a new approach to study comfort. *Energy and Buildings*, *152*, 323–335. https://doi.org/10.1016/j.enbuild.2017.07.060

Pantic, M. (2006). *Introduction to machine learning & case-based reasoning*. [Lecture notes]. Imperial College London. https://ibug.doc.ic.ac.uk/media/uploads/documents/courses/syllabus-CBR.pdf

Papachristos, G., Jain, N., Burman, E., Zimmermann, N., Mumovic, D., Davies, M., & Edkins, A. (2020). Low carbon building performance in the construction industry: A multi-method approach of project management operations and building energy use applied in a UK public office building. *Energy and Buildings*, *206*. https://doi.org/10.1016/j.enbuild.2019.109609

Pappaccogli, G., Giovannini, L., Zardi, D., & Martilli, A. (2020). Sensitivity analysis of urban microclimatic conditions and building energy consumption on urban parameters by means of idealized numerical simulations. *Urban Climate*, *34*(August), 100677. https://doi.org/10.1016/j.uclim.2020.100677

Paudel, S., Elmitri, M., Couturier, S., Nguyen, P. H., Kamphuis, R., Lacarrière, B., & Le Corre, O. (2017). A relevant data selection method for energy consumption prediction of low energy building based on support vector machine. *Energy and Buildings*, *138*, 240–256. https://doi.org/ 10.1016/ j.enbuild.2016.11.009

Pegg, I. M., Cripps, A., & Kolokotroni, M. (2007). Post-occupancy performance of five low-energy schools in the UK. *ASHRAE Transactions*, *113 PART 2*, 3–13.

Pérez-Lombard, L., Ortiz, J., González, R., & Maestre, I. R. (2009). A review of benchmarking, rating and labelling concepts within the framework of building energy certification schemes. *Energy and Buildings*, *41*(3), 272–278. https://doi.org/10.1016/j.enbuild.2008.10.004

Preiser, W. F. E., & Vischer, J. C. (2006). The evolution of building performance evaluation: An introduction. *Assessing Building Performance*, 3–14. https://doi.org/10.4324/9780080455228

Peters, B., Peters, T. (2018). *Computing the environment: Digital design tools for simulation and visualization of sustainable* Peters *architecture*. Wiley. Doi: 10.1002/9781119097921

Plebankiewicz, E., Zima, K., & Wieczorek, D. (2016). Life cycle cost modelling of buildings with consideration of the risk. *Archives of Civil Engineering*, 62(12), 149-166. https://doi.org/ 10.1515/ace-2015-0071

Pruvost, H., & Scherer, R. J. (2017). Analysis of risk in building life cycle coupling BIM-based energy simulation and semantic modeling. *Procedia Engineering*, *196*(June), 1106–1113. https://doi.org/ 10.1016/ j.proeng. 2017.08.068

Psomas, T., Fiorentini, M., Kokogiannakis, G., & Heiselberg, P. (2017). Ventilative cooling through automated window opening control systems to address thermal discomfort risk during the summer period: Framework, simulation and parametric analysis. *Energy and Buildings*, *153*, 18–30. https://doi.org/10.1016/j.enbuild.2017.07.088

Qazi, A., Quigley, J., Dickson, A., & Kirytopoulos, K. (2016) Project Complexity and Risk Management (ProCRiM): Towards modelling project complexity driven risk paths in construction projects. *International Journal of Project Management*, 34(7), 1183–1198. https://doi.org/ 10.1016/ j.ijproman.2016.05.008

Qi, Y., Qian, Q. K., Meijer, F. M., & Visscher, H. J. (2019). Identification of quality failures in building energy renovation projects in Northern China. *Sustainability (Switzerland)*, *11*(15). https://doi.org/10.3390/su11154203

Qin, X., Mo, Y., & Jing, L. (2016). Risk perceptions of the life-cycle of green buildings in China. *Journal of Cleaner Production*, *126*, 148–158. https://doi.org/10.1016/j.jclepro.2016.03.103

Rachman, A., & Ratnayake, R. M. C. (2019). Machine learning approach for risk-based inspection screening assessment. *Reliability Engineering and System Safety*, *185*(February), 518–532. https://doi.org/10.1016/j.ress.2019.02.008

Rafsanjani, H. N., Ahn, C. R., & Alahmad, M. (2015). A review of approaches for sensing, understanding, and improving occupancy-related energy-use behaviors in commercial buildings. *Energies* (Vol. 8). https://doi.org/ 10.3390/ en81010996

Ram, V. S. S., Kayastha, N., & Sha, K. (2022). OFES: Optimal feature evaluation and selection for multi-class classification. *Data & Knowledge Engineering*, *139*(March), 102007. https://doi.org/10.1016/j.datak.2022.102007

Ramanathan, S., & Rathinakumar, V. (2017). Analysis of risk factors in small, medium & large construction projects. *International Journal of Civil Engineering and Technology*, *8*(4), 1977–1984.

Regnier, C., Sun, K., Hong, T., & Piette, M. A. (2018). Quantifying the benefits of a building retrofit using an integrated system approach: A case study. *Energy*

and Buildings, *159*, 332–345. https://doi.org/ 10.1016/ J.ENBUILD. 2017.10.090

Rezaee, R., Brown, J., Augenbroe, G., & Kim, J. (2015). Assessment of uncertainty and confidence in building design exploration. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing: AIEDAM*, *29*(4), 429–441. https://doi.org/10.1017/S0890060415000426

Rivera, R. (2020). *Principles of managerial statistics and data science*. John Wiley & Sons, Inc. https://doi.org/ 10.1002/ 9781119486473

Roh, Y., Heo, G., & Whang, S. E. (2021). A survey on data collection for machine learning: A big data-AI integration perspective. *IEEE Transactions on Knowledge and Data Engineering*, *33*(4), 1328–1347. https://doi.org/10.1109/ TKDE.2019.2946162

Rokach, L., &Maimon, Z. O. (2008). *Data mining with decision trees: Theory and applications*. World Scientific.

Salehi, M. M., Terim Cavka, B., Frisque, A., Whitehead, D., & Bushe, W. K. (2015). A case study: The energy performance gap of the Center for Interactive Research on Sustainability at the University of British Columbia. *Journal of Building Engineering*, *4*, 127–139. https://doi.org/10.1016/j.jobe.2015.09.002

Salimi, S., & Hammad, A. (2019). Critical review and research roadmap of office building energy management based on occupancy monitoring. *Energy and Buildings*, *182*, 214–241. https://doi.org/10.1016/j.enbuild.2018.10.007

Seyedzadeh, S., Rahimian, F. P., Oliver, S., Glesk, I., & Kumar, B. (2020a). Data driven model improved by multi-objective optimisation for prediction of building energy loads. *Automation in Construction*, *116*(January), 103188. https://doi.org/10.1016/j.autcon.2020.103188

Seyedzadeh, S., Pour Rahimian, F., Oliver, S., Rodriguez, S., & Glesk, I. (2020b). Machine learning modelling for predicting non-domestic buildings energy performance: A model to support deep energy retrofit decision-making. *Applied Energy*, *279*(September), 115908. https://doi.org/ 10.1016/ j.apenergy. 2020.115908

Shi, X., Si, B., Zhao, J., Tian, Z., Wang, C., Jin, X., & Zhou, X. (2019). Magnitude, causes, and solutions of the performance gap of buildings: A review. *Sustainability (Switzerland)*, *11*(3), 1–21. https://doi.org/10.3390/su11030937

Shrestha, M.P. (2018). *Impacts of energy-efficiency retrofits on ventilation and indoor air quality in low-income households*. [Doctoral Dissertation, University of Colorado]. Institutional Repository at the University of Colorado. https://scholar.colorado.edu/

Shrubsole, C., Hamilton,I.G., Zimmermann, N., Papachristos, G., Broyd, T., Burman, E., … Davies, M. (2019). Bridging the gap: The need for a systems thinking approach in understanding and addressing energy and environmental performance in buildings. *Indoor and Built Environment*, *28*(1), 100–117. https://doi.org/10.1177/1420326X17753513

Stribos, R. (2020). The impact of data noise on a Naive Bayes classifier. *34th Twente Student Conference on IT*, *4*(3), 1–9. Retrieved from http://www.mitpressjournals.org/doi/pdf/10.1162/neco.1992.4.3.366

Sidaoui, B., & Sadouni, K. (2013). OVA tree multiclass framework for support vector machine. *Lecture Notes in Engineering and Computer Science*, *2202*(March), 36–41.

Silvero, F., Lops, C., Montelpare, S., & Rodrigues, F. (2019). Impact assessment of climate change on buildings in Paraguay—Overheating risk under different future climate scenarios. *Building Simulation*, *2050*, 943–960. https://doi.org/ 10.1007/s12273-019-0532-6

Somu, N., Sriram, A., Kowli, A., & Ramamritham, K. (2021). A hybrid deep transfer learning strategy for thermal comfort prediction in buildings. *Building and Environment*, *204*(April), 108133. https://doi.org/10.1016/ j.buildenv. 2021. 108133

Sun, S., & Chen, H. (2020). Data-driven sensitivity analysis and electricity consumption prediction for water source heat pump system using limited information. *Building Simulation*. https://doi.org/10.1007/s12273-020-0721-3

Sun, S., Kensek, K., Noble, D., & Schiler, M. (2016). A method of probabilistic risk assessment for energy performance and cost using building energy simulation. *Energy and Buildings*, *110*, 1–12. https://doi.org/ 10.1016/ j.enbuild. 2015. 09.070

Sun, C., Zhang, R., Sharples, S., Han, Y., & Zhang, H. (2019). Thermal comfort, occupant control behaviour and performance gap – A study of office buildings in north-east China using data mining. *Building and Environment*, *149* (September 2018), 305–321. https://doi.org/10.1016/ j.buildenv.2018. 12.036

Sunikka-Blank, M., & Galvin, R. (2012). Introducing the prebound effect- the gap between performance and actual energy consumption. *Building Research &Information, 40* (3), 37–41. https://doi.org/10.1080/09613218.2012.690952

Szymański, P. (2017). Risk management in construction projects. *Procedia Engineering, 208*, 174–182. https://doi.org/10.1016/j.proeng.2017.11.036

Tanrıverdi, C. (2019). *A causal mapping approach for the analysis of construction disputes*. [Master's thesis, Middle East Technical University]. Institutional Repository at the Middle East Technical University. https://open.metu.edu.tr/handle/11511/44587

Tardioli, G., Kerrigan, R., Oates, M., O'Donnell, J., & Finn, D. (2015). Data driven approaches for prediction of building energy consumption at urban level. *Energy Procedia, 78*, 3378–3383. https://doi.org/10.1016/ j.egypro. 2015. 11.754

Taroun, A. (2014). Towards a better modelling and assessment of construction risk: Insights from a literature review. *International Journal of Project Management*, *32*(1), 101–115. https://doi.org/10.1016/j.ijproman.2013.03.004

Thomé, A. M. T., Scavarda, L. F., Scavarda, A., & Thomé, F. E. S. de S. (2016). Similarities and contrasts of complexity, uncertainty, risks, and resilience in supply chains and temporary multi-organization projects. *International Journal of Project Management*, *34*(7), 1328–1346. https://doi.org/ 10.1016/ j.ijproman. 2015.10.012

Tian, S., Lu, Y., Ge, X., & Zheng, Y. (2021). An agent-based modeling approach combined with deep learning method in simulating household energy consumption. *Journal of Building Engineering*, *43*(September). https:// doi.org/ 10.1016/j.jobe.2021.103210

Topouzi, M., Killip, G., Fawcett, T., & Owen, A. (2019). Deep retrofit approaches: Managing risks to minimise the energy performance gap. *Eceee Summer Study Proceedings, 2019-June*, 1345–1354.

Trivedi, S. K. (2020). A study on credit scoring modeling with different feature selection and machine learning approaches. *Technology in Society*, *63*(September), 101413. https://doi.org/10.1016/j.techsoc.2020.101413

Turner, C., & Frankel, M. (2008). *Energy performance of LEED ® for new construction buildings*. (Final Report). New Buildings Institute.

https://newbuildings.org/resource/energy-performance-leed-new-construction-buildings/

Van de Schoot, R., & Milica, M. (2020). *Small sample size solutions: A Guide for applied researchers and practitioners*. Routledge. https://doi.org/10.4324/9780429273872-6

Van Heijst, A. (2017). *Analyzing the performance gap by identifying its causal factors and quantifying their individual impacts for Energy Performance Contracts*. [Master's thesis, Eindhoven University of Technology]. Institutional Repository at the Eindhoven University of Technology. https://pure.tue.nl/ws/portalfiles/portal/72496420/Heijst_v_0768016.pdf

Van Someren, L.K. (2017).*User behaviour and perception as drivers for lightşng energy efficiency and performance gap reduction in higher education*. [Doctoral Disstertation, University of Reading]. Institutional Repository at the University of Reading. https://centaur.reading.ac.uk/77156/

Wade, F., Murtagh, N., & Hitchings, R. (2018). Managing professional jurisdiction and domestic energy use. *Building Research and Information*, *46*(1), 42–53. https://doi.org/10.1080/09613218.2017.1324698

Walker, S., Khan, W., Katic, K., Maassen, W., & Zeiler, W. (2020). Accuracy of different machine learning algorithms and added-value of predicting aggregated-level energy performance of commercial buildings. *Energy and Buildings*, *209*(February). https://doi.org/10.1016/j.enbuild.2019.109705

Wang, T., Gao, S., Li, X., & Ning, X. (2018). A meta-network-based risk evaluation and control method for industrialized building construction projects. *Journal of Cleaner Production*, *205*, 552–564. https://doi.org/ 10.1016/ j.jclepro.2018. 09.127

Wang, Z., Xu, G., Lin, R., Wang, H., & Ren, J. (2019). Energy performance contracting, risk factors, and policy implications: Identification and analysis of risks based on the best-worst network method. *Energy*, *170*, 1–13. https://doi.org/10.1016/j.energy.2018.12.140

Wang, Z., Liu, J., Zhang, Y., Yuan, H., Zhang, R., & Srinivasan, R. S. (2021). Practical issues in implementing machine-learning models for building energy efficiency: Moving beyond obstacles. *Renewable and Sustainable Energy Reviews*, *143*(February). https://doi.org/10.1016/j.rser.2021.110929

Wei, T. (2013). A review of sensitivity analysis methods in building energy analysis. *Renewable and Sustainable Energy Reviews*, *20*, 411–419. https://doi.org/ 10.1016/j.rser.2012.12.014

Wilson, C. (2013). Questionnaires and surveys. *Credible checklists and quality questionnaires A user centered design method*. Morgan Kaufmann. 29–79. https://doi.org/10.1016/b978-0-12-410392-4.00002-7

Witten, H.I., Frank, E., Hall, M. A. (2011). *Data mining: Practical machine learning tools and techniques*. (3rded.). Morgan Kaufmann Publishers. https:// catalog.library.metu.edu.tr/

Xia, N., Zhong, R., Wu, C., Ph, D., Asce, S. M., Wang, X., & Wang, S. (2017). Assessment of stakeholder-related risks in construction projects : Integrated analyses of risk attributes and stakeholder influences. *Journal of Construction Engineering and Management*, *143*(8), 1–11. https://doi.org/ 10.1061/(ASCE)CO.1943-7862.0001322

Xu, X. (2020). Risk factor analysis combined with deep learning in the risk assessment of overseas investment of enterprises. *PLoS ONE*, *15*(10 October), 1–14. https://doi.org/10.1371/journal.pone.0239635

Xu, X., Li, C. Z., Wang, J., & Huang, W. (2019). Collaboration between designers and contractors to improve building energy performance. *Journal of Cleaner Production*, *219*, 20–32. https://doi.org/10.1016/j.jclepro.2019.02.036

Xu, X., Mumford, T., & Zou, P. X. W. (2021). Life-cycle building information modelling (BIM) engaged framework for improving building energy performance. *Energy and Buildings*, *231*, 110496. https://doi.org/ 10.1016/ j.enbuild.2020.110496

Yang, R. J., & Zou, P. X. W. (2014). Stakeholder-associated risks and their interactions in complex green building projects: A social network model. *Building and Environment*, *73*, 208–222. https://doi.org/ 10.1016/ j.buildenv. 2013.12.014

Yang, R. J., Zou, P. X. W., & Wang, J. (2016a). Modelling stakeholder-associated risk networks in green building projects. *International Journal of Project Management*, *34*(1), 66–81. https://doi.org/10.1016/j.ijproman.2015.09.010

Yang, T., Qian, K., Lo, D. C. T., Xie, Y., Shi, Y., & Tao, L. (2016b). Improve the prediction accuracy of Naïve Bayes classifier with association rule mining. *Proceedings - 2nd IEEE International Conference on Big Data Security on*

Cloud, IEEE BigDataSecurity 2016, 2nd IEEE International Conference on High Performance and Smart Computing, IEEE HPSC 2016 and IEEE International Conference on Intelligent Data and S, 129–133. https://doi.org/10.1109/BigDataSecurity-HPSC-IDS.2016.38

Ye, Z., Cheng, K., Hsu, S. C., Wei, H. H., & Cheung, C. M. (2021). Identifying critical building-oriented features in city-block-level building energy consumption: A data-driven machine learning approach. *Applied Energy*, *301*(April), 117453. https://doi.org/10.1016/j.apenergy.2021.117453

Yildiz, A. E., Dikmen, I., Birgonul, M. T., Ercoskun, K., & Alten, S. (2014). A knowledge-based risk mapping tool for cost estimation of international construction projects. *Automation in Construction*, *43*, 144–155. https://doi.org/10.1016/j.autcon.2014.03.010

Yudelson, J. (2016). *Reinventing green building : Why certification systems aren't working and what we can do about it*. New Society Publishers. https:// www.nlb.gov.sg/biblio/202978564

Zhang, S. (2020). Cost-sensitive KNN classification. *Neurocomputing*, *391*, 234–242. https://doi.org/10.1016/j.neucom.2018.11.101

Zhao, X., Hwang, B. G., & Gao, Y. (2016). A fuzzy synthetic evaluation approach for risk assessment: A case of Singapore's green projects. *Journal of Cleaner Production*, *115*(January 2005), 203–213. https://doi.org/10.1016/ j.jclepro. 2015.11.042

Zhou, P., & Chang, Y. (2021). Automated classification of building structures for urban built environment identification using machine learning. *Journal of Building Engineering*, *43*(July), 103008. https://doi.org/10.1016/ j.jobe. 2021. 103008

Zheng, D., Yu, L., Wang, L., & Tao, J. (2018). A screening methodology for building multiple energy retrofit measures package considering economic and risk aspects. *Journal of Cleaner Production*. https://doi.org/10.1016/J. JCLEPRO. 2018.10.196

Zheng, D., Yu, L., & Wang, L. (2019). A techno-economic-risk decision-making methodology for large-scale building energy efficiency retrofit using Monte Carlo simulation. *Energy*, *189*, 116169. https://doi.org/10.1016/j. energy. 2019.116169

Zhu, J., & Xu, J. (2019). Driving factors of green supply chain management in building materials enterprises. *IOP Conference Series: Earth and Environmental Science*, *295*(2). https://doi.org/10.1088/1755-1315/295/2/012063

Zou, P. X. W., Zhang, G., & Wang, J. (2007). Understanding the key risks in construction projects in China. *International Journal of Project Management*, *25*(6), 601–614. https://doi.org/10.1016/j.ijproman.2007.03.001

Zou, P. X. W., Xu, X., Sanjayan, J., & Wang, J. (2018). Review of 10 years research on building energy performance gap: Life-cycle and stakeholder perspectives. *Energy and Buildings*, *178*, 165–181. https://doi.org/10.1016/j.enbuild.2018.08.040

Zou, P. X. W., Alam, M., Sanjayan, J., Wilson, J., Stewart, R., Sahin, O., … Ellis-Jones, D. (2016). Managing risks in complex building retrofit projects for energy and water efficiency. *International Conference on Innovative Production and Construction (IPC 2016), 29-30th Sep 2016, Perth, Australia*, (October), 1–16.

Zou, P. X. W., Wagle, D., & Alam, M. (2019). Strategies for minimizing building energy performance gaps between the design intend and the reality. *Energy and Buildings*, *191*, 31–41. https://doi.org/10.1016/j.enbuild.2019.03.013

APPENDICES

A. How to Reduce EPG

Design Phase	
Problems	**Solutions**
1. Limitations-shortcomings related to modeling and software	(1) reexamine the tools (De Wilde, 2018), (2) improve the simulation software including improvement in the surface area graphic implementation, default usage profiles, data implementation, and the possibility of implementing the energy consumption of specific equipment, (3) include the stochastic aspects of occupants' behavior into building energy performance simulation software (Shi *et al.*, 2019)
2. Unrealistic-oversimplified simulation inputs: Weather, occupancy, etc.	(1) use on-site weather data to increase the accuracy of building energy simulation, (2) use pre-occupancy evaluation, (3) use post-occupancy evaluation (Shi *et al.*, 2019), (4) have better and more reliable data, (5) use sensitivity analysis/ uncertainty analysis (Demanuele, 2008), (6) check how well the default parameters (e.g. daily occupancy etc.) can represent typical situation in buildings (Lee *et al.*, 2007), (7) consider the effect of occupants with different energy usage rates (Ali *et al.*, 2020)
3.Design deficiencies-poorly designed building and building systems	(1) evaluate user experience as designers (Grover, 2019), (2) focus on the distribution of design responsibility, sequencing of work (Boyd & Schweber, 2018)
4.The complexity of the design- lack of simplicity-lack of attention to buildability	(1) design a simple and compact building envelope, (2) increase simplicity of detailing, (3) consider construction sequencing, buildability and as-built performance (Dollard, 2018)
5.Incomplete design-lack of proper details	(1) reduce thermal bridging, (2) ensure drawings and specifications are clear (Dollard, 2018), (3) provide higher quality of detailing at key junctions, skirtings, and service penetrations (Gupta *et al.*, 2013), (4) policy makers can make a construction guidance web resource for standard details and sequences of installation (Dollard, 2018)

Construction Phase	
Problems	**Solutions**
1.Poor quality of workmanship-insufficient building quality	(1) increase the frequency and quality of inspection and testing (Dollard, 2018), (2) use specialist tests (e.g., thermographic imaging or air tightness tests) during the construction (Alencastro *et al.*, 2018)
2.Changing requests-Value engineering	(1) provide a robust design change management system (Alencastro *et al.*, 2018), (2) re-check the changes carefully so that the changes may not impact on performance criteria (Alam *et al.*, 2019)
3.Poor quality of materials, equipment or technologies	(1) use products from trusted manufacturers who could offer sufficient and accurate technical details and specifications of their products (Zou *et al.*, 2019)
4.Lack of green material in the market- non-availability of the specific product	(1) more publicity using media, (2) financial and better market-based incentives for green building technologies adoption, (3) an improved green building technologies research and development (Darko & Chan, 2018), (4) since consumers' green environmental trends have a critical impact on green supply chain management, consumers should improve their environmental awareness and actively communicate with the company about product information (Zhu & Xu, 2019)
5.Deteriotion-aging-lifetime defects of materials	(1) pay special attention to maintenance, renovation, and replacement due to the difference between the life expectancy of buildings and the service life of building systems (De Wilde, 2018)

Use and Operational Phase	
Problems	**Solutions**
1.Uncertainty on occupants' behavior and activities	(1) provide more operational guidance to occupants (Hansen *et al.*, 2018), (2) develop a regular feedback loop to explore green building feature awareness of occupants, and their response to it (Grover, 2019), (3) seek ways of integrating user involvement in building performance (Janda, 2011)
2.Lack of continuous commissioning-poor commissioning-problems with soft landing	(1) provide advanced commissioning and measurement and verification (Turner& Frankel, 2008), (2) use soft landing approach (Gupta & Gregg, 2016)
3.Unsuitable-poor building management	(1) ensure good management and controls (Menezes *et al.*, 2012)
4.Poor quality of the measured data due to faulty sensors and meters	(1) replace the faulty products (Doylend, 2015), (2) assess the possibility of having faulty or underperforming components (Jradi *et al.*, 2018)
5.Lack of monitoring-continuity of monitoring	(1) use monitoring data to calibrate simulation models (Cuerda *et al.*, 2020), (2) use continuous monitoring and obtain feedback about how buildings perform at every stage of a project (Birchall, 2011).
6.Complexity of control strategies	(3) specify simple building services that are easy to operate and perform as intended (Dollard, 2018)

Critical for all Project Phases	
Problems	**Solutions**
1.Miscommunication-lack of communication between project parties	(1) provide better management to the building design, construction, and operation process (Shi *et al.*, 2019), (2) ensure better communication between designer and site team (Dollard, 2018), (3) communicate partner ambitions, design intent and responsibilities (Papachristos *et al.*, 2020), (4) when the building is completed, the designers/contractors should collaborate with the building users and operators to determine differences between the design estimates and the actual operational energy use (CIBSE, 2013).
2.Lack of knowledge, skill or experience of project parties	(1) make sure that sub-contractors have right training and experience for their job (2) train design and site teams (Dollard, 2018), (3) industry and universities can rise their efforts with respect to building physics education (Imam *et al.*, 2017), (4) develop professional expertise in the field (Janda, 2011)

B. Approval of Applied Ethics Research Center

UYGULAMALI ETİK ARAŞTIRMA MERKEZİ
APPLIED ETHICS RESEARCH CENTER

ORTA DOĞU TEKNİK ÜNİVERSİTESİ
MIDDLE EAST TECHNICAL UNIVERSITY

DUMLUPINAR BULVARI 06800
ÇANKAYA ANKARA/TURKEY
T: +90 312 210 22 91
F: +90 312 210 79 59
ueam@metu.edu.tr
www.ueam.metu.edu.tr

No: 28620816 /

29 January 2021

Ali Murat TANYER, Phd
Professor

Re: Your application to Human Participants Research Committee

Dear Ali Murat TANYER and Derya YILMAZ

The protocol titled "A DECISION SUPPORT TOOL USING RISK MANAGEMENT TO REDUCE ENERGY PERFORMANCE GAP IN NZEB REFURBISHMENTS" has been reviewed and approved by the METU Human Subjects Ethics Committee. The protocol number is **353-ODTU-2020.**

Sincreley,

Prof. Dr. Mine MISIRLISOY
Chair

C. Interview Questions

Participant Preliminary Information Form

In this research, *the problems encountered during the design, construction, use or operation stages will be examined* to understand the causes of the energy performance gap in buildings. In the literature, "Performance gap" is defined as the mismatch between the predicted energy performance of buildings and the actual measured performance.

We will ask the questions below to understand *the factors that weakened or strengthened the project process, the adverse changes, and unpredictable events*. With your answers, we aim *to establish a causal relationship* between these factors.

Interviews can be made on an online meeting platform on the web, or in person at a time that suits you. These one-hour interviews will be recorded using a tape recorder or web platform. Afterward, you will be asked for another meeting.

Research Title: A Decision Support Tool Using Risk Management to Reduce Energy Performance Gaps in Buildings

Researcher: Derya Yılmaz

Advisor at Bauhaus University Weimar: Prof. Dr.-Ing. Conrad Voelker

Name Surname	Profession	Education Level	Years of Experience	Current Job Position	Firm/Company Name	Years of Experience in Energy Efficient Buildings

First Session:

1. Please tell me about yourself and your professional experience highlighting the information above.

2. Could you give me brief information about your company?

3. Please explain briefly an energy-efficient building project you were working on.

4. In which processes did you participate in this project?

5. Could you tell me about your responsibilities in the project?

6. Please tell me about the project process in general.

7. What were the problems you encountered or the challenges you faced during the project process?

8. How did these problems result?

9. Did these problems cause some other problems?

10. If not, do you think they were solved with managerial abilities, financial or technical resources? (Please state if you think the problem is solved in some other ways)

11. What can be done to prevent such problems during the project process?

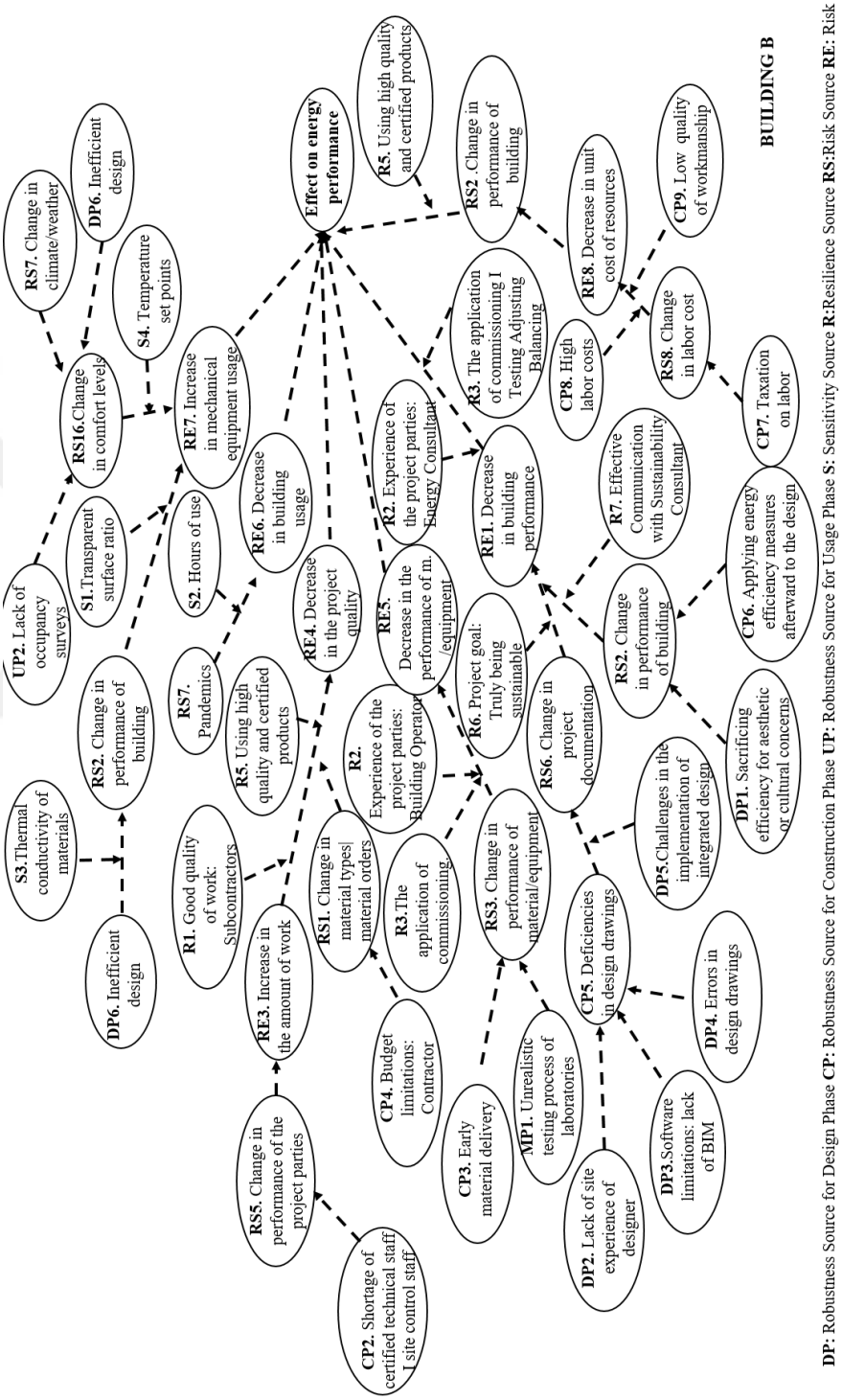
12. In your opinion, what are the most important points to be taken into consideration to achieve energy performance targets in buildings? (Please answer this question considering the construction sector in your country).

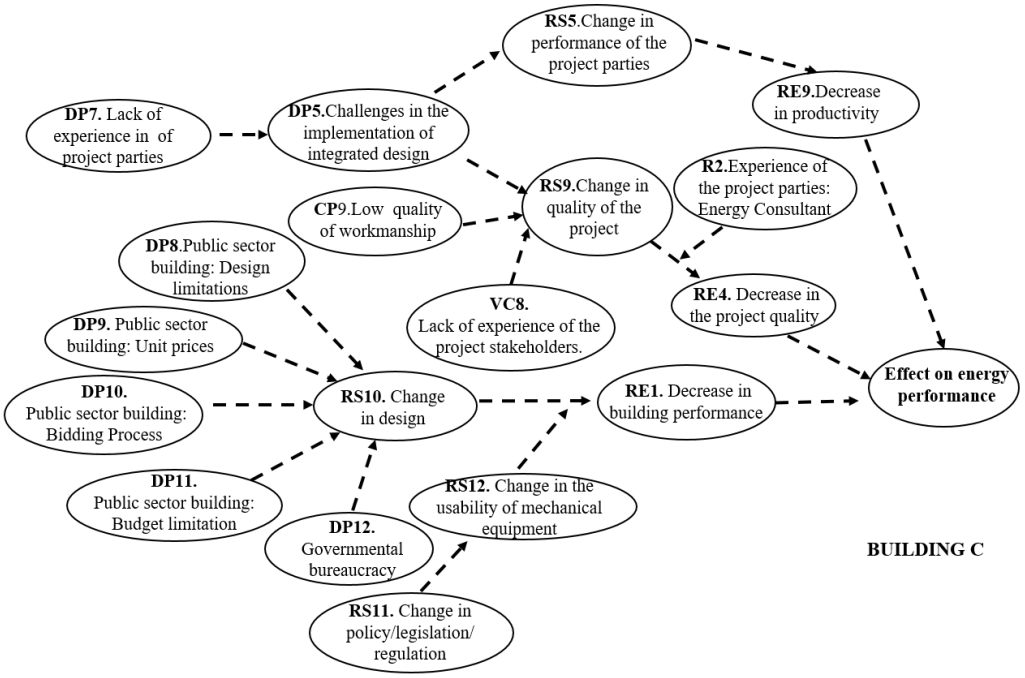
Second Session:

The second session can be held a few hours after the first session or another day.

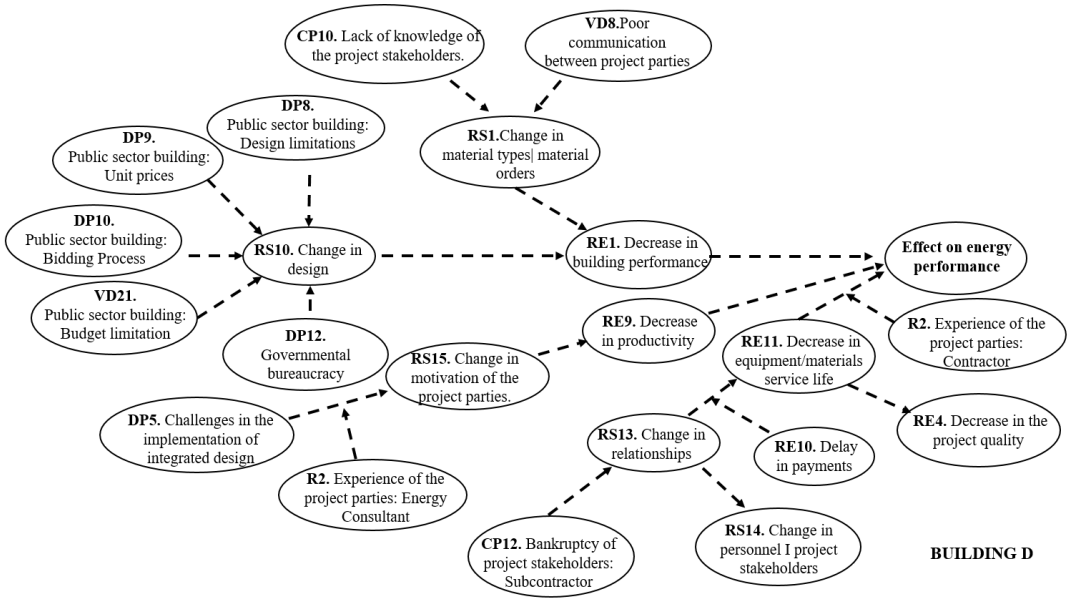
13. Please check the cognitive maps and explain if there is anything you want to add or change.

D. Cognitive Maps of the Case Studies

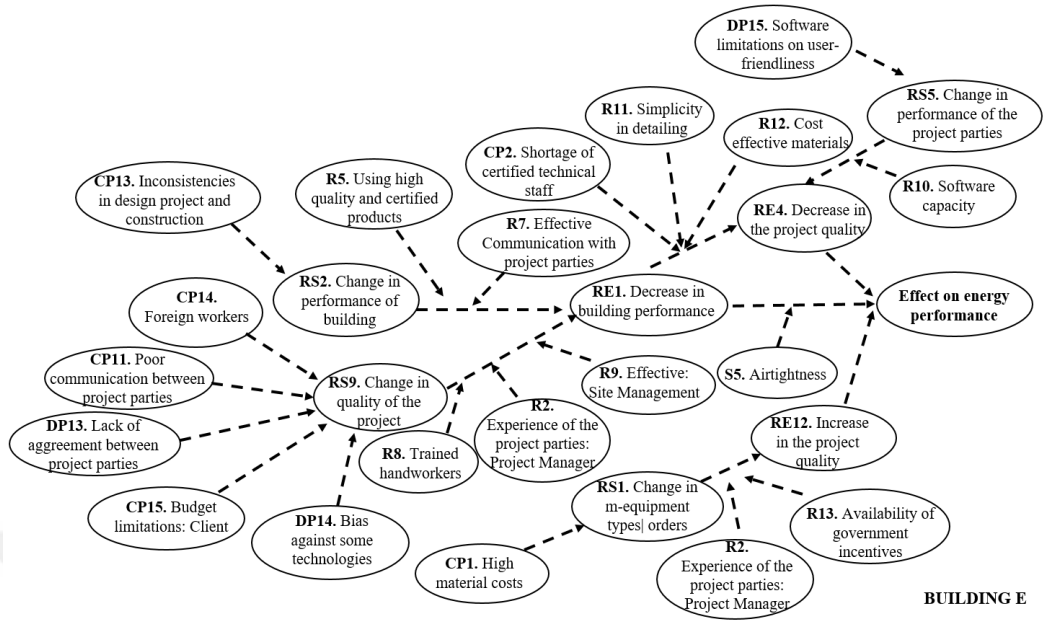




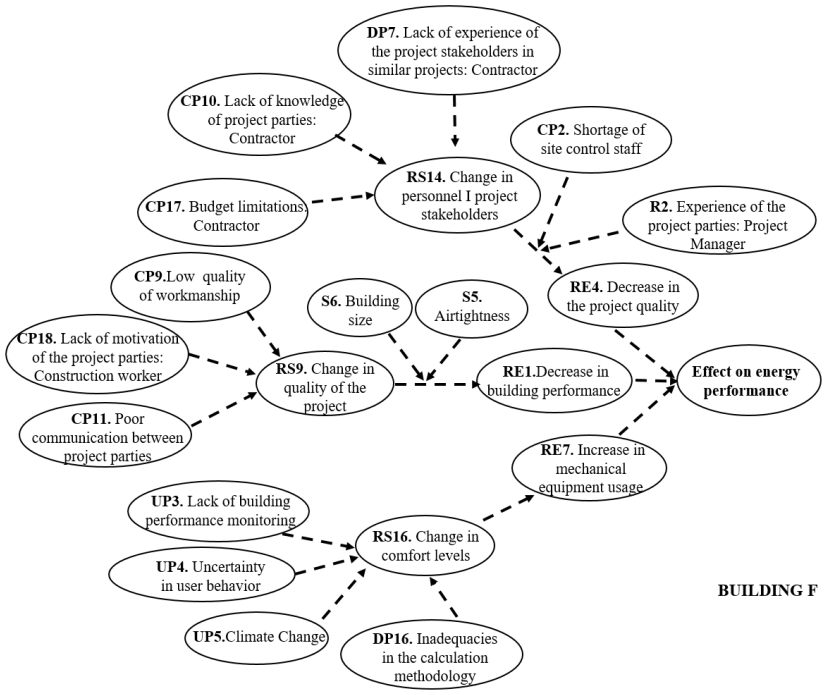
DP: Robustness Source for Design Phase **CP:** Robustness Source for Construction Phase **UP:** Robustness Source for Usage Phase
S: Sensitivity Source **R:**Resilience Source **RS:**Risk Source **RE:** Risk Event



DP: Robustness Source for Design Phase **CP:** Robustness Source for Construction Phase **UP:** Robustness Source for Usage Phase
S: Sensitivity Source **R:**Resilience Source **RS:**Risk Source **RE:** Risk Event



DP: Robustness Source for Design Phase **CP:** Robustness Source for Construction Phase **UP:** Robustness Source for Usage Phase
S: Sensitivity Source **R:**Resilience Source **RS:**Risk Source **RE:** Risk Event



DP: Robustness Source for Design Phase **CP:** Robustness Source for Construction Phase **UP:** Robustness Source for Usage Phase
S: Sensitivity Source **R:**Resilience Source **RS:**Risk Source **RE:** Risk Event

E. Risk-Vulnerability EPG Framework

Project Life Cycle	No	Factor: Vulnerability	Case
Design Phase	DP1	Sacrificing efficiency for aesthetic or cultural concerns	Building C, Building F
	DP2	Lack of site experience of designer	Building C, Building D
	DP3	Software limitations: lack of BIM	Building C, Building D
	DP4	Errors in design drawings	Building C, Building D
	DP5	Challenges in the implementation of integrated design	Building C, Building D
	DP6	Inefficient design	Building C, Building D
	DP7	Lack of experience in of project parties about sustainable buildings I Contractor	Building D, Building F
	DP8	Public sector building: Design limitations	Building D, Building E,Building F
	DP9	Public sector building: Unit prices	Building D
	DP10	Public sector building: Bidding process	Building E
	DP11	Public sector building: Budget limitation	Building E
	DP12	Governmental bureaucracy	Building E
	DP13	Lack of aggreement between project parties	Building E
	DP14	Bias against some technologies	Building E
	DP15	Software limitations on user-friendliness	Building E
	DP16	Inadequacies in the calculation methodology	Building E
Usage Phase	UP1	Lack of local operation and maintenance services	Building F
	UP2	Lack of occupancy surveys	Building F
	UP3	Lack of building performance monitoring	Building F
	UP4	Uncertainty in user behavior	Building F
	UP5	Climate Change	Building F
Manufacturing Phase	MP1	Unrealistic testing process of laboratories	Building F

Project Life Cycle	No	Factor: Vulnerability	Case
Construction Phase	CP1	High material costs	Building A, Building E
	CP2	Shortage of certified technical staff I site control staff	Building A, Building B
	CP3	Early material delivery	Building A, Building B
	CP4	Budget limitations: Contractor	Building A
	CP5	Deficiencies in design drawings	Building B, Building F
	CP6	Applying energy efficiency measures afterwards to the design	Building B
	CP7	Taxation on labor	Building B, Building F
	CP8	High labor costs	Building B
	CP9	Low quality of workmanship	Building B
	CP10	Lack of knowledge of the project stakeholders on sustainable buildings: Contractor	Building B
	CP11	Poor communication between project parties	Building B
	CP12	Bankruptcy of project stakeholders: Subcontractor	Building B,Building C, Building D
	CP13	Inconsistencies in design project and construction	Building B
	CP14	Foreign workers	Building B
	CP15	Budget limitations: Client	Building B
	CP16	Shortage of certified technical staff	Building B
	CP17	Budget limitations: Contractor	Building B
	CP18	Lack of motivation of the project parties: Construction worker	Building B, Building C, Building F

No	Factor: Risk Sources	Case
RS1	Change in Material Types I Material Orders	Building A, Building C
RS2	Change in Performance of Building	Building C, Building D
RS3	Change in Performance of Material/Equipment	Building A
RS4	Change in Budget	Building B
RS5	Change in Performance of the Project Parties	Building A, Building B, Building D, Building E
RS6	Change in Project Documentation	Building B
RS7	Change in Climate/Weather	Building B, Building F
RS8	Change in Labor Cost	Building B
RS9	Change in Quality of the Project	Building C
RS10	Change in Design	Building C, Building E, Building F
RS11	Change in Policy/Legislation/Regulation	Building A, Building B, Building E
RS12	Change in the Usability of Material/Equipment	Building A, Building B
RS13	Change in Relationships	Building B, Building C, Building E
RS14	Change in Personnel I Project stakeholders	Building D, Building F
RS15	Change in Motivation of the Project Participants	Building D
RS16	Change in Comfort Levels	Building D

No	Factor: Sensitivity	Case
S1	Transparent Surface Ratio	Building A, Building B
S2	Hours of use	Building B
S3	Thermal conductivity of materials	Building B
S4	Temperature set points	Building B
S5	Airtightness	Building E, Building F
S6	Building size	Building F
No	**Factor: Robustness**	**Case**
R1	Good quality of work: Subcontractors	Building A, Building B
R2	Experience of the project parties: Commissioner I Building Operator I	Building A, Building B, Building C, Building D, Building E
R3	The application of commissioning I Testing Adjusting Balancing	Building A, Building B
R4	Design Flexibility: Planning an alternative system	Building A
R5	Using high quality and certified products	Building B, Building E
R6	Project goal: Truly being sustainable	Building B
R7	Effective Communication with Sustainability Consultant I Project Parties	Building B, Building E
R8	Trained handworkers	Building E
R9	Effective: Site Management	Building E
R10	Software capacity	Building E
R11	Simplicity in detailing	Building E
R12	Cost effective materials	Building E
R13	Availability of government incentives	Building E

No	Factor: Risk Events	Case
RE1	Decrease in building performance	Building A, Building B, Building C, Building D, Building E, Building F
RE2	Increase in building performance	Building A, Building B, Building F
RE3	Increase in the amount of work	Building B
RE4	Decrease in the project quality	Building B
RE5	Decrease in the performance of materials/ equipment	Building B,Building F
RE6	Decrease in building usage	Building B
RE7	Increase in mechanical equipment usage	Building D
RE8	Decrease in unit cost of resources	Building D
RE9	Decrease in productivity	Building D
RE10	Delay in payments	Building B,Building C,Building D,Building E, Building F
RE11	Decrease in equipment/materials service life	Building E
RE12	Increase in the project quality	Building B
No	**Factor: Force Major**	**Case**
RS7	Pandemics	Building B

F. Final Conceptual Framework: Unexpected Events and Changes, Project Management, Assumptions

Group	No	Code	Parameter	References	Case	Comment of Interviewee
Unexpected Events & Changes	14	UE1	Client/user expectations	Coşkun (2019), Montazami et al.(2015), Alam et al.(2017), Jradi et al.(2018), Doylend (2014), Ipeec(2019)	Building A	A1.We did not find triple glazing very feasible. Therefore, we preferred high performance glazing. B1. Due to design changes in the project, there were changes in orders. C1. We had too many revisions in this project. C2. There were many stakeholders and decision makers in this project, and there were many different requests. F1.Right now we try to green the facade, and by greening facade, it might help a lot but it will need few more years…
					Building B	
					Building C	
					Building F	
	15	UE2	Project stakeholders	Qin et al.(2016), Mohammadi (2015), De Wilde(2018), Dasgupta et al.(2012)	Building B	B1.There were difficulties with the new contractor's orientation. B1. Some of construction workers who were undisciplined were changed. C1. Some team members leave the team, because people do not work in the same position in the public sector for a long time. For example, we do something according to a team member's request. However, after the building is complete, the person who brings that discussion may not be there and the newcomers can ask why we did that. F1. We had a big problem with this particular building because we hired a general contractor and suddenly he realized that he mistook this building and will not be able to complete this building on time and it will exceed the overall budget. Therefore he gave up.
					Building C	
	16	UE3	Country conditions	Topouzi et al. (2019), Yang & Zou (2014), Zou et al.(2016), Mohammadi (2015), Coşkun (2019)	Building A	A1. There is already a typical problem in Turkey regarding the supply of materials, what I mean is: 70-80% of the materials used in this building are imported.B1. Another important issue is the insufficient number of staff controlling the project on-site and lack of experience. B2. We had difficulties with product suppy. There is a problem with local production of technical equipment. Therefore they are imported from abroad. C1. We depend on importing foreign technology on a number of matters. There are many non-local products. Either we do not use them in Turkey or we need to choose another spec product that is both affordable and available. C2. We are not used to a high level of insulation in Turkey. Contractors are not used to it either. As the thickness increases, the products on the market become more limited and more expensive. We even had difficulties in finding suitable dowels. C3. There may not always be a personnel who will operate the buildings in the targeted way, who will make the adjustments and maintenance correctly. E1. For windows,we do not have so many companies to work with.
					Building B	
					Building C	
					Building E	

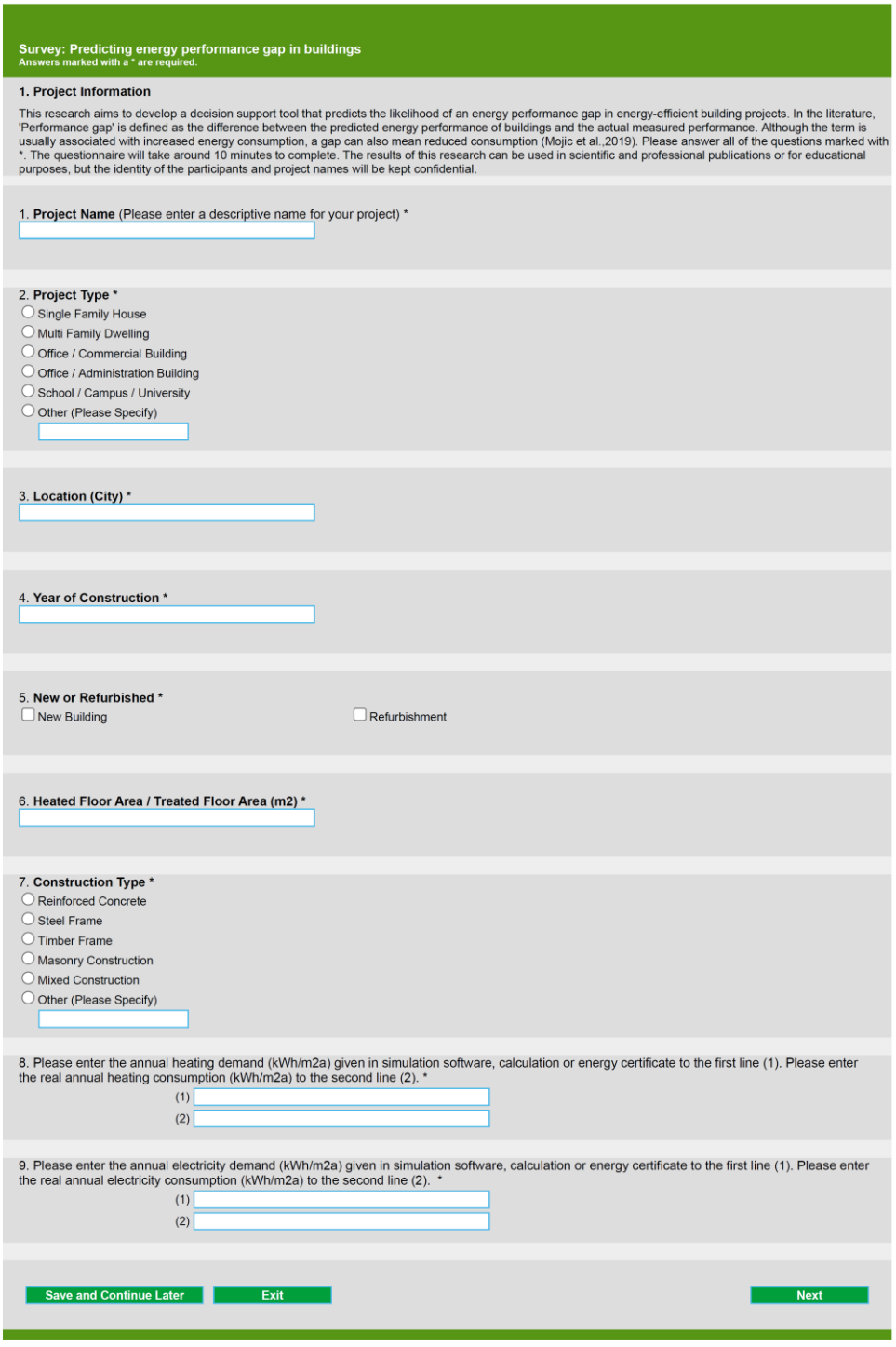
Group	No	Code	Parameter	References	Case	Comment of Interviewee
Unexpected Events & Changes	17	UE4	**Policy / Legislation / Regulation**	Yang & Zou (2014), Qin et al.(2016), Mohammadi (2015), Coşkun (2019)	Building C	C1.Regulations can change a lot in Turkey. The regulation on cogeneration devices has changed after the design in this project.Thus some decisions were not be implemented due to such changes. C2. There are barriers in terms of bureaucracy and regulations in Turkish conditions. Not all of them are predictable.
	18	UE5	**Public sector building process**	Mohammadi (2015), Coşkun (2019)	Building B Building D	C1. This is a public project which poses a number of challenges to the process. C2.Public procurement processes need to be transformed, some changes are needed in tendering both for the project design phase and for the delivery of qualified buildings to contractors. C3. In public projects, assigning a commissioning specialist to the project in Turkey creates a conflicting authorization situation with control teams. C4. They are already trying to improve the tendering process. Maybe they need to make not only economic but also technical evaluations during the tendering process. D1.This project progressed very parallel to the other, so I cannot distinguish the two projects. We had similar problems with the other such as conventional work and public tendering.
	19	UE6	**Climate**	Zou et al.(2016), Coşkun (2019), Brady& Hanmer-Dwight (2016), De Wilde (2014), Xu et al.(2019), Love &Bullen (2009), Salehi et al. (2015), Bros-Williamson et al.(2016), Harrison &Yiang (2018)	Building B Building F	B1. Since weather and climate extremes were experienced this year, the occupants reported reduced comfort levels; therefore, temperature set points and HVAC operating hours were applied differently from the design stage assumptions. F1.What I want to say is you have to change the way of approach. Especially, because there is climate change.
	20	UE7	**Force majeure events**	Cai et al.(2019),De Wilde (2018), Wang et al.(2019), Shrestha (2015)	Building B	B1.Flexible work schedules for the employees during the pandemic caused a change in building usage and the number of users decreased, which impacted electricity consumption.

Group	No	Code	Parameter	References	Case	Comment of Interviewee
Project Management	21	PM1	**Experience of the project stakeholders**	De Wilde (2014), Jradi et al. (2018), Zou et al. (2018), Zou et al. (2019), Alam et al. (2017), Menezes et al.(2012), Alencastro et al. (2018), Jradi et al. (2018), Yang & Zou (2014), Qin et al. (2016), Zou et al.(2016), Gultekin (2015), Mohammadi (2015), Coşkun (2019), Ipeec (2019)	Building A Building B Building C Building D Building E Building F	A1.In this project, the client had its own architectural and mechanical staff. Since they are a project developer and consulting firm, this is a well-known issue for them. B1. The client was knowledgeable in this project. C1. In that regard, they are actually lucky in this project, because they have skilled labor. C2. It was a little bit difficult to develop a project with people who did not work on sustainability. C3. We had foreign partners who were more experienced in this regard. We tried to lessen the lack of experience by working with experts from abroad. D1. For example, we did not have a contractor that we could choose in the other project. However, we have now. We know that the contractor who worked on the previous project has at least some experience. D2. This project was much easier than the other, we both worked with experienced people and less number of stakeholders got involved. E1.The buildings completely built out of wood. This was the major focus of this project. We have a lot of experience in wood construction so we do not have a problem at least with detailing. E2. I think the important point is everybody have to know about it including the project owners, the planners and also the manufacturer... In our company, everybody has to know Passive House topic.F1. The contractor indeed had lots of experience with conventional buildings but not with Passive Houses. It was their first Passive House. F2. We learned that if workers do not understand what the construction is all about, then the best construction does not worth anything. So we need the involvement of high quality workers very much.
	22	PM2	**Motivation of the project parties**	Yang & Zou (2014), Janser et al.(2020), Gilani et al.(2020), Rafsanjani et al. (2015), Tuohy & Murphy (2015), Fernandez-Antolin et al.(2021), Ortiz et al.(2017), Delzendeh et al.(2017), Alencastro et al.(2018)	Building A Building F	A1.We aimed to construct an exemplary building. Rather than earning points for the project, our main goal was to make a truly green building. F1.You have to make sure that everything is very airtight in Passive Houses. You have to make sure the one, who is applying insulation is doing it very well, he has to do this work very precisely. Because of this, it needs simply high motivation and high qualification.
	23	PM3	**Effective communication between project stakeholders**	De Wilde (2014), Jradi et al. (2018), Zou et al. (2018), Ali et al. (2020), Alencastro et al. (2018), Fedoruk et al. (2015), Alam et al. (2017), Ali et al. (2020), Topouzi et al. (2019), Yang & Zou (2014), Xu et al.(2021), Doylend (2015), Ipeec (2019)	Building A Building B Building C Building D Building E	A1. They tried to take into account the things we talked about in every subject, they applied what we wanted during the construction phase, it was a more integrated process. A2.The importance of some of the design elements could have been explained to other teams more effectively. B1.Here the project developers were open for cooperation. They took into account our recommendations for the project. C1. We were in frequent communication with all teams...this sounds too perfect. However, we had problems with it. D1.We did not receive much feedback about the operation of this project, we do not know how it is operated at the moment. Airtightness test results are not available. We asked for commissioning information but we do not have them. E1. In this project there is also another architectural firm that is a part of our planning team. It was very nice to work with him. E2. It's a small family...Firstly, the important thing is you should have a good relationship with them. You must be fair to them. E3. The most important thing is how you talk to them because no one looks at the plans at the construction site.

Group	No	Code	Parameter	References	Case	Comment of Interviewee
Project Management	24	PM4	**Training of project participants**	Zou et al.(2016), Fernandez-Antolin et al.(2020), Janser et al.(2020), Rafsanjani et al.(2015), Brady&Hanmer-Dwight(2016), Alencastro et al.(2018), Fernandez-Antolin et al.(2021), Eon et al.(2020), Gupta &Kapsali (2016), Doylend (2015),Ipeec (2019)	Building A	A1. There were minor difficulties in using some systems in accordance with their intended purpose, and these were tried to be solved with the announcements or the trainings.
	25	PM5	**Design flexibility**	Delgado (2016), Al Dakheel et al.(2020), Zou et al.(2018)	Building A Building C	A1. We could not directly have the operation and maintenance services for trigeneration systems in Turkey. We had to procure it from abroad. Since trigeneration systems are very expensive to maintain and operate, we use photovoltaics, which are much more energy efficient than trigeneration. C1. Lessons we learned from the project: Design should be robust,and flexible. When a system is shut down, the rest of the system should not be affected much. There should be different operation methods...
	26	PM6	**Occupant Surveys**	Mills et al.(2006), Li et al.(2019), Gupta& Chandiwala (2010), Salimi & Hammad (2019), Grover (2019), Rafsanjani et al.(2015), Tuohy &Murphy(2015), Gupta & Kapsali (2016)	Building A	A1. Occupant surveys were conducted as a part of a Master's Thesis. The satisfaction rate was very high except a very few places.
	27	PM7	**Application of passive measures for indoor comfort**	Day & Gunderson (2015), Ramesh et al.(2010), Burman et al.(2018)	Building A Building F	A1. We suggested sunshading devices on the south and west facades against the disturbing effect of light. However, they were canceled afterward and there were complaints from some users in south and west facing spaces after moving to the building. F1.Right now we try to green the facade, and by greening facade, it might help a lot but it will need few more years…
	28	PM8	**Simplicity in detailing / systems design**	Alencastro et al. (2018), Herrando et al. (2016), Qin et al.(2016), Coşkun (2019), Zou et al.(2019), Ipeec (2019), Bros-Williamson et al.(2016)	Building E	E1.Our detailing is not so sophisticated as well. When we do it, we try to do the detailing very simple. E2. Sophisticated things are difficult to realize and cost lots of money.Moreover, you can not find the people who will construct it in that way.

Group	No	Code	Parameter	References	Case	Comment of Interviewee
Assumptions	29	A1	**Thermal conductivity of materials**	Jumabekova et al.(2020), de Rubeis et al. (2020), Gelesz et al. (2020), Mastrucci et al. (2017), Tian et al. (2017), Gagnon et al.(2018), Chen & Tsay (2021)	Building A	A1. Product catalogs gave inaccurate information about some materials such as glazing and some mechanical equipment.
	30	A2	**Building occupancy rates**	Kim et al. (2020)	Building B	B1.Flexible work schedules for the employees during the pandemic caused a change in building usage and the number of users decreased, which impacted electricity consumption.
	31	A3	**Thermal comfort of building occupants**	Coakley et al. (2014), Burman et al.(2018)	Building B Building C	B1. Since weather and climate extremes were experienced this year, the occupants reported reduced comfort levels; therefore, temperature set points and HVAC operating hours were applied differently from the design stage assumptions. C1.We could not achieve the comfort conditions we aimed for in this building.Some systems could not be operated due to operational problems or changes in regulations.
	32	A4	**Ventilation rate**	Wei (2013), Gunay et al. (2019), Pappaccogli et al. (2020)	x	x
	33	A5	**Internal loads**	Wei (2013),Zou et al.(2016), De Wilde (2014)	x	x
	34	A6	**Temperature set points**	Wei (2013), Mastrucci et al. (2017), Chen & Tsay (2021), Gagnon et al.(2018), Tian et al. (2017),	Building B Building C	B1. Since weather and climate extremes were experienced this year, the occupants reported reduced comfort levels; therefore, temperature set points and HVAC operating hours were applied differently from the design stage assumptions. C1.The preferences of the user may be different from the target. They may want to adjust the temperature set points differently.
	35	A7	**HVAC equipment efficiency**	Wei (2013), Tian et al. (2017), Kim et al. (2020)	Building A	A1. Product catalogs gave inaccurate information for some materials such as glazing and some mechanical equipment.
	36	A8	**Hours of use**	Wei (2013),Zou et al.(2016), De Wilde (2014)	Building B	B1. Since weather and climate extremes were experienced this year, the occupants reported reduced comfort levels; therefore, temperature set points and HVAC operating hours were applied differently from the design stage assumptions. B1.Flexible work schedules for the employees during the pandemic caused a change in building usage and the number of users decreased, which impacted electricity consumption.
	37	A9	**Airtightness**	Wei (2013), Pappaccogli et al. (2020), Chen & Tsay (2021), Gagnon et al. (2018)	Building E	E1.In the past we did not have many problems on airtightness, but we have more and more problems nowadays and the reason is that it's not so easy to get at this high level of understanding .
	38	A10	**Transparent surface ratio**	Wei (2013), Delgarm et al. (2018)	Building A	A1.The ratio of the transparent surfaces on the facade should not exceed a certain percentage, but we have passed that figure.

G. Web-based Survey: Questions



Survey: Predicting energy performance gap in buildings
Answers marked with a * are required.

1. Project Information

This research aims to develop a decision support tool that predicts the likelihood of an energy performance gap in energy-efficient building projects. In the literature, 'Performance gap' is defined as the difference between the predicted energy performance of buildings and the actual measured performance. Although the term is usually associated with increased energy consumption, a gap can also mean reduced consumption (Mojic et al.,2019). Please answer all of the questions marked with *. The questionnaire will take around 10 minutes to complete. The results of this research can be used in scientific and professional publications or for educational purposes, but the identity of the participants and project names will be kept confidential.

1. **Project Name** (Please enter a descriptive name for your project) *

2. **Project Type** *
- Single Family House
- Multi Family Dwelling
- Office / Commercial Building
- Office / Administration Building
- School / Campus / University
- Other (Please Specify)

3. **Location (City)** *

4. **Year of Construction** *

5. **New or Refurbished** *
- New Building
- Refurbishment

6. **Heated Floor Area / Treated Floor Area (m2)** *

7. **Construction Type** *
- Reinforced Concrete
- Steel Frame
- Timber Frame
- Masonry Construction
- Mixed Construction
- Other (Please Specify)

8. Please enter the annual heating demand (kWh/m2a) given in simulation software, calculation or energy certificate to the first line (1). Please enter the real annual heating consumption (kWh/m2a) to the second line (2). *
(1)
(2)

9. Please enter the annual electricity demand (kWh/m2a) given in simulation software, calculation or energy certificate to the first line (1). Please enter the real annual electricity consumption (kWh/m2a) to the second line (2). *
(1)
(2)

Save and Continue Later | Exit | Next

Survey: Predicting energy performance gap in buildings
Answers marked with a * are required.

2. Problems/Limitations

In this section, a number of problems/limitations that can be faced during the design, construction, use, and operation stage of buildings will be presented.

10. **Design Phase Factors** *

Please consider the design phase of your specific project. What was the magnitude of the problems/limitations you faced for the following factors? (Please select "N/A" if the factor was not considered during the design phase.)

	N/A	Very Low	Low	Medium	High	Very High
Modeling, software or calculation methodology	○	○	○	○	○	○
Simulation inputs (e.g. uncertainty in weather, building envelope, occupancy, mechanical equipment data etc.)	○	○	○	○	○	○
Design (e.g. poorly designed building and building systems, incomplete design, lack of proper details, complexity of design etc.)	○	○	○	○	○	○
Project Budget	○	○	○	○	○	○

11. **Construction Phase Factors** *

Please consider the construction phase of your specific project. What was the magnitude of the problems/limitations you faced for the following factors? (Please select "N/A" if the factor was not considered during the construction phase.)

	N/A	Very Low	Low	Medium	High	Very High
Quality of workmanship	○	○	○	○	○	○
Quality of materials / equipment or technologies	○	○	○	○	○	○
Inconsistencies in design project and construction	○	○	○	○	○	○
Bankruptcy of project stakeholders (e.g. contractor, subcontractor etc.)	○	○	○	○	○	○

12. **Use and Operation Phase Factors** *

Please consider the use and operation phase of your specific project. What was the magnitude of the problems/limitations you faced for the following factors? (Please select "N/A" if the factor was not considered during the use and operation phase.)

	N/A	Very Low	Low	Medium	High	Very High
Occupants behavior and activities (e.g. uncertainty etc.)	○	○	○	○	○	○
Commissioning / continuous commissioning / soft landing	○	○	○	○	○	○
Building management	○	○	○	○	○	○
Regular maintenance of the equipment	○	○	○	○	○	○
Quality of the measured data due to faulty sensors and meters	○	○	○	○	○	○

Save and Continue Later | Exit | Back | Next

Survey: Predicting energy performance gap in buildings
Answers marked with a * are required.

3. Project Management *

Rate the magnitude of the following factors considering your specific project. (Please select "N/A" if the factor was not considered.)

13. **Project Management Factors** *

	N/A	Very Low	Low	Medium	High	Very High
Experience of the project stakeholders (e.g. energy-sustainability consultant, project manager, site manager, building operator etc.)	○	○	○	○	○	○
Motivation of the project parties (e.g. energy-sustainability consultant, project manager, building operator etc.)	○	○	○	○	○	○
Effective communication between project stakeholders (e.g. sustainability consultant etc.)	○	○	○	○	○	○
Training of project participants	○	○	○	○	○	○
Design flexibility (e.g. planning an alternative system etc.)	○	○	○	○	○	○
Occupant Surveys (e.g. comprehensiveness etc.)	○	○	○	○	○	○
Application of passive measures for indoor comfort	○	○	○	○	○	○
Simplicity in detailing / systems design	○	○	○	○	○	○

Save and Continue Later | Exit | Back | Next

Survey: Predicting energy performance gap in buildings
Answers marked with a * are required.

4. Assumptions *

Please consider the conditions of your specific project. How much change did you see in the following factors during the use/operation phase in comparison to the design phase of your project? (If it is not available, or there is no answer please select N/A).

14. **Assumptions Factors** *

	N/A	Very Low	Low	Medium	High	Very High
Thermal conductivity of materials	○	○	○	○	○	○
Building occupancy rates	○	○	○	○	○	○
Thermal comfort of building occupants	○	○	○	○	○	○
Ventilation rate	○	○	○	○	○	○
Internal loads	○	○	○	○	○	○
Temperature set points	○	○	○	○	○	○
HVAC equipment efficiency	○	○	○	○	○	○
Hours of use	○	○	○	○	○	○
Airtightness	○	○	○	○	○	○
Transparent surface ratio	○	○	○	○	○	○

Save and Continue Later | Exit | Back | Next

Survey: Predicting energy performance gap in buildings
Answers marked with a * are required.

5. Unexpected events and changes *

Please consider your specific project. Rate the magnitude of the unexpected events and changes that occurred related to the following factors.

For example, were there unexpected events and changes related to client/user expectations? If yes, please rate the magnitude. (If there is no change, please indicate it as N/A).

15. **Unexpected events and changes in...***

	N/A	Very Low	Low	Medium	High	Very High
Client/user expectations (e.g. related to design, material-equipment types/orders etc.)	○	○	○	○	○	○
Project stakeholders (e.g. personnel, relationships, performance, motivation of the project parties etc.)	○	○	○	○	○	○
Country conditions (e.g. availability of material, certified equipment, skilled personnel etc.)	○	○	○	○	○	○
Policy / Legislation / Regulation	○	○	○	○	○	○
Public sector building process: (e.g. bidding process)	○	○	○	○	○	○
Climate (e.g. weather and climate extremes)	○	○	○	○	○	○
Force majeure events (e.g. natural disasters, pandemic etc.)	○	○	○	○	○	○

Save and Continue Later | Exit | Back | Next

H. Some Performance Measures for a Classifier (Bramer, 2020)

True Positive Rate or Hit Rate or Recall or Sensitivity or TP Rate	TP/P	The proportion of positive instances that are correctly classified as positive.
False Positive Rate or False Alarm Rate or FP Rate	FP/N	The proportion of negative instances that are erroneously classified as positive
False Negative Rate or FN Rate	FN/P	The proportion of positive instances that are erroneously classified as negative = 1-True Positive Rate
True Negative Rate or Specificity or TN Rate	TN/N	The proportion of negative instances that are correctly classified as negative
Precision or Positive Predictive Value	TP/(TP+FP)	Proportion of instances classified as positive that are really positive
F1 Score	(2 × Precision × Recall) /(Precision + Recall)	A measure that combines Precision and Recall
Accuracy or Predictive Accuracy	(TP + TN)/(P + N)	The proportion of instances that are correctly classified
Error Rate	(FP + FN)/(P + N)	The proportion of instances that are incorrectly classified

İ. List of attributes considered in ML models

			BEG				PEG	NEG	BHG				PHG	NHG
			Wrapper				Filter	Filter	Wrapper				Filter	Filter
Category	No	Attributes	NB	KNN	SVM	RF	Chi	Chi	NB	KNN	SVM	RF	Chi	Chi
General Info	1	Project Type												
	2	Year of Construction												
	3	New or Refurbished												
	4	Heated Floor Area												
	5	Construction Type												
Problems/Limitations	6	Modeling												
	7	Inputs												
	8	Design												
	9	Project Budget												
	10	Quality of Workmanship												
	11	Quality of Materials												
	12	Inconsistencies in project												
	13	Bankruptacy												
	14	Occupants behaviour												
	15	Commissioning												
	16	Building management												
	17	Regular maintenance												
	18	Quality of measured data												
Project Management	19	Experience of the project stakeholders												
	20	Motivation of the project parties												
	21	Effective communication												
	22	Training of project stakholders												
	23	Design flexibility												
	24	Occupant Surveys												
	25	Application of passive measures												
	26	Simplicity in detailing												
Adverse events/changes	27	Client/ user expectations												
	28	Project stakeholders												
	29	County conditions												
	30	Policiy/Legislation/Regulations												
	31	Public sector buidling process												
	32	Climate												
	33	Force majeure events (e.g. Pandemics)												

BEG: Binary electricity gap, **PEG:** Positive Electricity Gap, **NEG:** Negative Electricity Gap, **BHG:** Binary Heating gap, **PHG:** Positive Heating Gap, **NHG:** Negative Heating Gap, **Wrapper:** Wrapper Attribute Evaluator, **Chi:** Chi-Squared Attribute Evaluator

J. Dataset used in ML models

Part 1: Instances

Part 1: Attributes

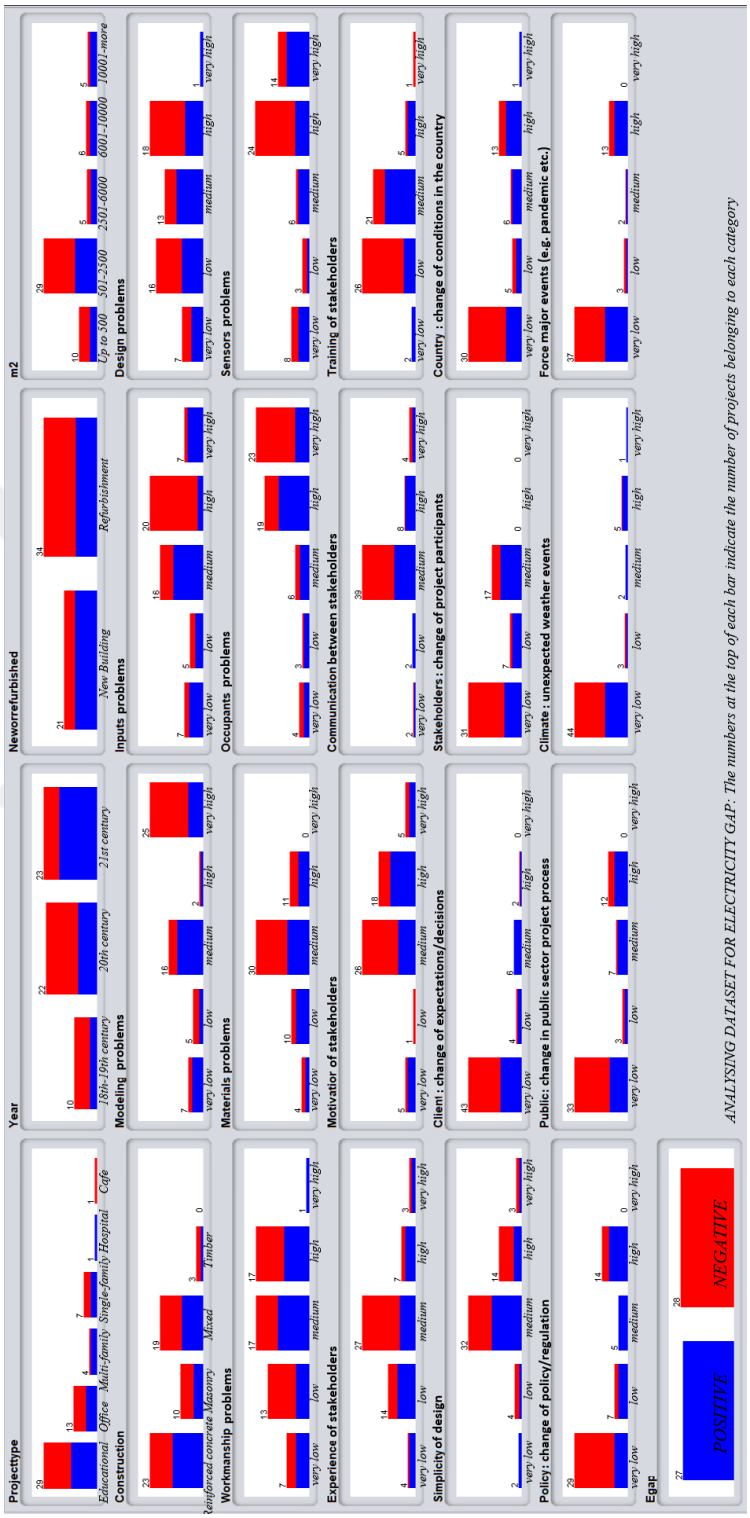
ID	Projecttype	Year	Newornrefurbished	m2	Construction	Modeling	Inputs	Design	Budget	Workmanship	Materials	Inconsistencies	Bankruptcy	Occupants	Commissioning	Management	Maintenance
1	SchoolCampus/University	1996	Refurbishment	8600	Reinforced Concrete	low	high	high	medium	medium	medium	high	very low	very high	high	medium	high
2	Single Family House	2020	New Building	213	Masonry Construction	medium	low	high	high	high	high	high	high	medium	high	high	medium
3	Single Family House	1974	Refurbishment	164	Masonry Construction	low	low	low	low	low	low	very low	very low	very low	low	medium	low
4	Multi Family Dwelling	1894	Refurbishment	400	Masonry Construction	very low	very low	medium	medium	high	low	medium	very low	very high	medium	low	low
5	Single Family House	2017	New Building	208	Timber Frame	very low	very low	very low	very low	very low	very low	very low	very low	very low	very low	very low	very low
6	SchoolCampus/University	2016	New Building	12000	Mixed Construction	very low	low	medium	medium	medium	high	medium	medium	high	low	low	low
7	Multi Family Dwelling	2018	New Building	3000	Timber Frame	medium	medium	very high	medium	medium	medium	medium	medium	medium	high	very high	very high
8	Single Family House	1959	Refurbishment	127	Masonry Construction	very low	very low	very low	very low	very low	very low	very low	very low	medium	very low	very low	very low
9	Office / Administration Building	2019	New Building	35000	Reinforced Concrete	low	low	medium	low	medium	high	medium	medium	medium	medium	medium	medium
10	Office / Administration Building	2015	New Building	8000	Reinforced Concrete	very high	very high	very high	high	very low	very low	low	very low	high	very low	very high	high
11	Single Family House	2013	New Building	366	Timber Frame	very low	low	very low	medium	low	low	very low	very low	high	medium	high	medium
12	Office / Administration Building	2020	New Building	2596	Reinforced Concrete	medium	very high	high	low	medium	low	very high	very low	high	very low	high	very low
13	SchoolCampus/University	2017	New Building	21.940	Mixed Construction	low	low	medium	high	medium	medium	high	high	low	medium	medium	medium
14	Multi Family Dwelling	2008	Refurbishment	535	Reinforced Concrete	very low	very low	very low	?	very low	low	very low	very low	very low	very low	very low	very low
15	Office / Administration Building	2014	New Building	10000	Mixed Construction	very low	very low	very low	very low	very low	very low	very low	very low	very low	low	very low	very low
16	Single Family House	2016	New Building	182	Masonry Construction	very low	very low	very low	low	medium	very low	very low	low	very low	very low	very low	low
17	Office / Administration Building	2011	New Building	6978	Reinforced Concrete	high	medium	low	low	high	high	low	very low	high	high	medium	?
18	Single Family House	1930	Refurbishment	174	Mixed Construction	high	medium	low	very low	very low	very low	very low	very low	very high	medium	high	medium
19	Hospital	2015	Refurbishment	10568	Reinforced Concrete	very high	very high	high	medium	medium	medium	high	very low	high	very high	high	very high
20	Office / Administration Building	2011	New Building	50000	Reinforced Concrete	medium	medium	very low	low	medium	medium	low	very low	low	very low	?	medium
21	Multi Family Dwelling	2019	New Building	7367	Reinforced Concrete	low	very low	very low	high	high	medium	medium	very low	medium	low	low	low
22	Single Family House	2016	New Building	120	Timber Frame	very low	medium	very low	high	low	medium	?	very low	low	very low	medium	very low
23	SchoolCampus/University	2001	New Building	921	Mixed Construction	very high	medium	low	?	low	low	medium	?	high	?	?	low
24	Office / Administration Building	2001	New Building	890	Mixed Construction	very high	high	low	?	low	low	medium	?	high	?	?	low
25	SchoolCampus/University	2003	New Building	1303.83	Mixed Construction	medium	medium	low	?	high	medium	high	?	high	low	medium	?
26	SchoolCampus/University	2007	New Building	6368	Mixed Construction	medium	very high	medium	?	low	low	medium	?	high	low	medium	?
27	SchoolCampus/University	2008	New Building	782.53	Mixed Construction	medium	high	medium	?	high	medium	high	?	high	medium	medium	?
28	SchoolCampus/University	2011	New Building	985.3	Mixed Construction	very high	high	low	?	very low	low	very low	?	high	?	?	low
29	SchoolCampus/University	2014	New Building	671	Reinforced Concrete	very high	high	low	?	very low	low	very low	?	very high	?	?	low
30	Office / Administration Building	2014	New Building	1705	Mixed Construction	very high	very high	medium	?	medium	medium	very high	?	very high	?	?	low
31	SchoolCampus/University	1905	Refurbishment	1704	Masonry Construction	very high	high	high	?	low	medium	very high	?	very high	?	?	medium
32	SchoolCampus/University	1911	Refurbishment	6250	Masonry Construction	very high	high	high	?	low	medium	very low	?	very high	?	?	medium
33	SchoolCampus/University	1925	Refurbishment	1251.72	Reinforced Concrete	medium	medium	high	?	low	medium	low	?	high	medium	medium	?
34	SchoolCampus/University	1925	Refurbishment	1054.09	Reinforced Concrete	medium	medium	medium	?	high	high	low	?	high	medium	medium	?
35	SchoolCampus/University	1929	Refurbishment	2604	Reinforced Concrete	medium	medium	high	?	high	high	low	?	high	medium	medium	?
36	SchoolCampus/University	1952	Refurbishment	2019.3	Reinforced Concrete	medium	medium	medium	?	high	medium	low	?	high	medium	medium	?
37	SchoolCampus/University	1956	Refurbishment	835.72	Reinforced Concrete	medium	medium	low	?	high	medium	low	?	high	medium	medium	?
38	Office / Administration Building	1956	Refurbishment	1043	Reinforced Concrete	very high	high	high	?	medium	medium	low	?	very high	?	?	medium

		Part 2: Attributes																	
Part 1: Instances	ID	Sensors	Experience	Motivation	Communication	Training	Flexibility	Surveys	Passive	Simplicity	Client	Stakeholders	Country	Policy	Public	Climate	Force	Heatingap (%)	Electricity gap (%)
	1	low	high	medium	low	very low	very low	very low	very low	low	medium	low	low	low	low	high	high	16,6	17,9
	2	high	medium	high	high	medium	?	?	?	medium	very low	very low	very low	medium	low	very low	very low	6,3	56,5
	3	medium	very low	very low	very low	low	?	?	?	low	very low	very low	very low	very low	very low	very low	very low	-52,0	-20,0
	4	very low	very high	very high	very high	high	?	?	?	very high	medium	medium	medium	very low	very low	high	very low	2,6	4,9
	5	very low	very high	very high	very high	very high	?	?	?	very high	very low	very low	very low	very low	very low	very low	very low	0,0	-3,9
	6	very low	very low	very low	low	medium	?	?	?	medium	low	low	medium	medium	medium	very low	very low	13,3	8,0
	7	medium	high	high	high	high	?	?	?	medium	medium	medium	medium	low	medium	medium	medium	33,3	25,0
	8	very low	very low	very low	high	medium	very low	very low	very low	high	very low	very low	very low	medium	very low	very low	very low	-22,4	8,6
	9	low	medium	low	medium	medium	low	very low	?	high	low	low	low	low	medium	low	low	118,3	-60,0
	10	very high	very high	very high	high	medium	very high	high	high	medium	low	low	low	low	very low	high	very low	33,3	11,7
	11	very low	very high	high	high	medium	medium	very low	high	very high	low	medium	low	very low	low	very low	very low	0,0	0,0
	12	medium	low	very low	high	high	very high	very high	high	medium	medium	very low	high	high	very low	very low	low	27,2	45,8
	13	medium	medium	high	medium	medium	low	low	high	high	low	low	low	medium	medium	very low	very low	40,0	20,3
	14	very low	very low	very low	very low	very low	very low	very low	very low	very low	very low	very low	very low	low	very low	very low	very low	-8,4	13,0
	15	very low	high	high	medium	high	high	high	high	high	medium	very low	high	high	medium	high	high	-4,3	0,0
	16	medium	high	very high	very high	high	high	very high	very high	very low	very low	very low	very low	very low	very low	very high	very low	5,0	7,0
	17	very low	high	high	high	medium	high	high	medium	medium	medium	medium	very high	high	medium	low	very low	-29,4	119,0
	18	low	high	high	high	low	very low	very low	medium	high	high	medium	very low	very low	very low	high	very low	-7,1	-8,1
	19	very high	high	medium	medium	low	medium	very low	high	medium	medium	low	medium	low	medium	high	very low	11,6	31,5
	20	medium	medium	medium	high	medium	high	medium	low	medium	high	medium	medium	medium	medium	medium	very low	-11,8	16,7
	21	very low	low	high	medium	medium	high	medium	very high	very high	very low	low	low	low	low	low	low	-57,1	-49,1
	22	very low	high	very high	very high	high	high	very low	high	high	very low	very low	medium	very low	very low	very low	medium	-7,9	-1,6
	23	high	medium	medium	medium	low	low	very low	medium	medium	very low	very low	very low	very low	very low	very low	very low	-35,9	34,4
	24	high	medium	medium	medium	low	low	very low	high	medium	very low	very low	very low	very low	very low	very low	very low	-38,3	-86,2
	25	very high	low	high	medium	medium	low	very low	high	medium	very low	medium	high	high	high	very low	high	155,7	38,7
	26	very high	low	high	medium	medium	low	very low	high	medium	very low	medium	high	high	high	very low	high	-47,7	39,1
	27	very high	low	high	medium	medium	low	very low	medium	medium	very low	medium	high	high	high	very low	high	77,0	-62,5
	28	high	medium	medium	medium	low	low	very low	medium	medium	very low	very low	very low	very low	very low	very low	very low	-83,6	-67,9
	29	high	medium	medium	medium	low	low	very low	high	medium	very low	very low	very low	very low	very low	very low	very low	-63,6	2,4
	30	high	medium	medium	medium	low	low	very low	medium	medium	very low	very low	very low	very low	very low	very low	very low	11,4	164,3
	31	high	medium	medium	medium	low	low	very low	high	medium	very low	very low	very low	very low	very low	very low	very low	-61,3	-10,1
	32	high	medium	medium	medium	low	low	very low	high	low	very low	very low	very low	very low	very low	very low	very low	38,0	-66,3
	33	very high	low	high	medium	medium	low	medium	low	medium	very low	medium	high	high	high	very low	high	-44,6	9,5
	34	very high	low	high	medium	medium	low	very low	medium	medium	very low	medium	high	high	high	very low	high	-11,8	16,0
	35	very high	low	high	medium	medium	low	medium	high	low	very low	medium	high	high	high	very low	high	19,8	-1,9
	36	very high	low	high	medium	medium	low	very low	low	medium	very low	medium	high	high	high	very low	high	-0,7	67,9
	37	very high	low	high	medium	medium	low	medium	high	high	very low	medium	high	high	high	very low	high	-1,6	-48,0
	38	high	medium	medium	medium	low	low	very low	medium	medium	very low	very low	very low	very low	very low	very low	very low	0,2	-86,2

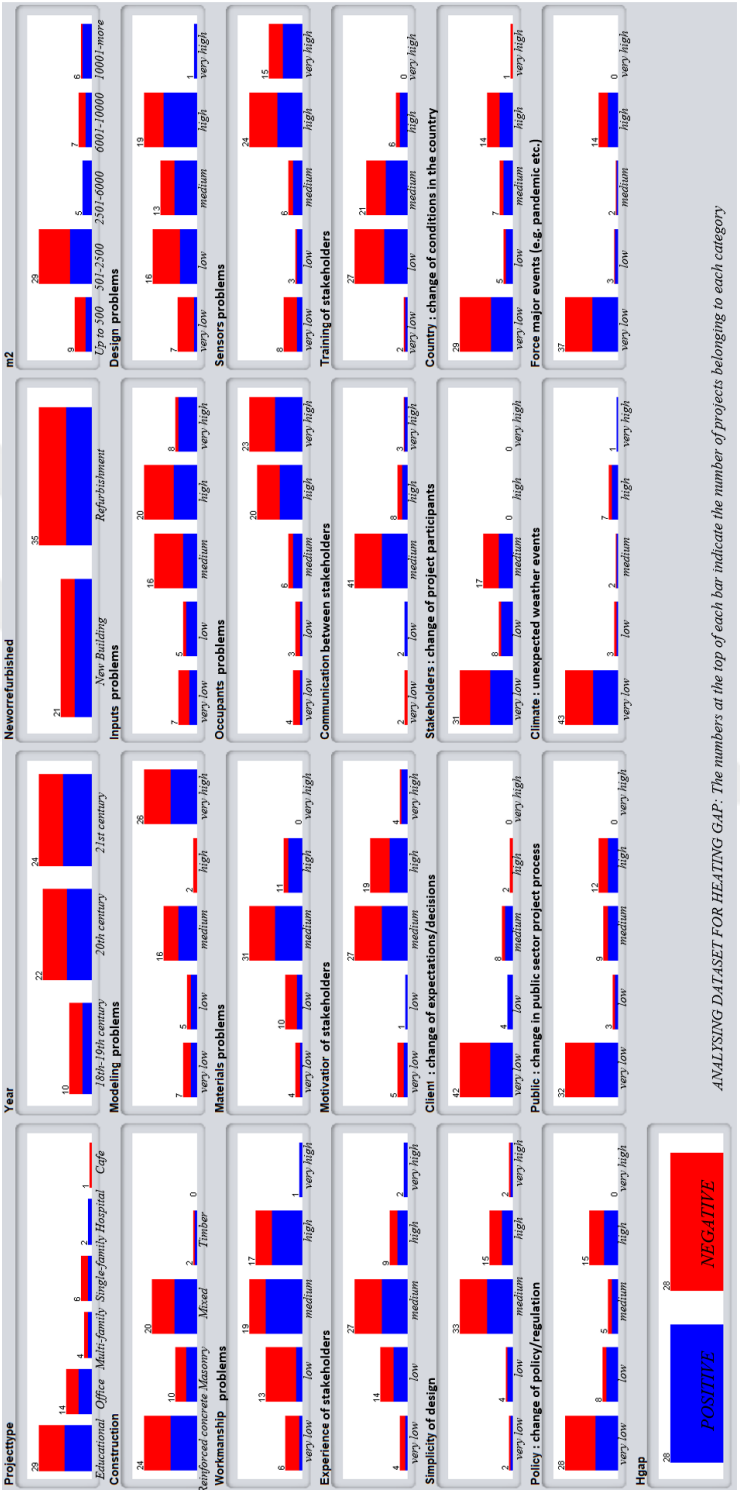
	Part 1: Attributes																	
	ID	Projecttype	Year	Neworrefurbished	m2	Construction	Modeling	Inputs	Design	Budget	Workmanship	Materials	Inconsistencies	Bankruptcy	Occupants	Commissioning	Management	Maintenance
Part 2: Instances	39	School/Campus/University	1957	Refurbishment	3100	Reinforced Concrete	very high	high	high	?	low	medium	very low	?	very high	?	?	medium
	40	School/Campus/University	1960	Refurbishment	2235,16	Reinforced Concrete	medium	medium	high	?	high	medium	low	?	high	medium	medium	?
	41	School/Campus/University	1962	Refurbishment	323,2	Reinforced Concrete	very high	high	low	?	low	medium	very low	?	very high	?	?	medium
	42	School/Campus/University	1968	Refurbishment	603	Reinforced Concrete	medium	medium	low	?	low	medium	medium	?	high	medium	medium	?
	43	Office / Administration Building	1968	Refurbishment	824,5	Reinforced Concrete	very high	high	low	?	high	high	very low	?	very high	?	?	medium
	44	School/Campus/University	1970	Refurbishment	2162	Mixed Construction	very high	very high	medium	?	high	high	high	?	very high	?	?	high
	45	School/Campus/University	1974	Refurbishment	3231,8	Mixed Construction	medium	very low	low	?	high	medium	very low	?	medium	medium	medium	?
	46	School/Campus/University	1974	Refurbishment	596,79	Mixed Construction	medium	medium	low	?	high	medium	low	?	high	medium	medium	?
	47	School/Campus/University	1975	Refurbishment	1279	Reinforced Concrete	very high	high	medium	?	medium	medium	low	?	very high	?	?	high
	48	Office / Administration Building	1983	Refurbishment	2280	Reinforced Concrete	very high	high	high	?	medium	medium	very high	?	very high	?	?	medium
	49	School/Campus/University	1818	Refurbishment	2162	Mixed Construction	very high	high	high	?	high	medium	high	?	very high	?	?	medium
	50	Office / Administration Building	1829	Refurbishment	1304	Mixed Construction	very high	high	high	?	medium	medium	very high	?	very high	?	?	high
	51	School/Campus/University	1830	Refurbishment	1000	Mixed Construction	very high	high	high	?	medium	medium	low	?	very high	?	?	medium
	52	Office / Administration Building	1854	Refurbishment	1157	Mixed Construction	very high	medium	medium	?	medium	high	very high	?	very high	?	?	medium
	53	Office / Administration Building	1870	Refurbishment	669	Masonry Construction	very high	high	high	?	low	medium	low	?	very high	?	?	medium
	54	School/Campus/University	1881	Refurbishment	1079	Masonry Construction	very high	high	high	?	medium	medium	medium	?	very high	?	?	medium
	55	Cafe/Restaurant	1886	Refurbishment	86,2	Masonry Construction	very high	high	high	?	low	medium	low	?	very high	?	?	low
	56	School/Campus/University	1720	Refurbishment	542	Mixed Construction	very high	high	low	?	high	high	very low	?	very high	?	?	medium
	57	School/Campus/University	1720	Refurbishment	651,5	Mixed Construction	very high	very high	low	?	very high	high	medium	?	very high	?	?	medium
	58	School/Campus/University	2003	New Building	10627	Reinforced Concrete	high	high	medium	very low	high	very low	very low	very low	high	high	high	medium
	59	School/Campus/University	2003	New Building	10529	Reinforced Concrete	high	very high	medium	very low	medium	very low	very low	very low	very high	very high	very high	very low
	60	School/Campus/University	2005	New Building	83043	Reinforced Concrete	high	very high	medium	very low	medium	very low	very low	very low	high	high	high	very low
	61	Office / Administration Building	2007	New Building	4811	Reinforced Concrete	high	low	very low	very low	low	low	very low	very low	high	very low	very low	very low
	62	School/Campus/University	1946	Refurbishment	9000	Reinforced Concrete	medium	very high	low	very low	very low	high	low	very low	high	very low	very low	very low
	63	School/Campus/University	1975	Refurbishment	7600	Reinforced Concrete	medium	high	low	very low	very low	high	low	very low	medium	very low	very low	very low
	64	School/Campus/University	1991	New Building	3750	Reinforced Concrete	high	very high	low	very low	very low	high	low	very low	medium	very low	very low	very low
	65	School/Campus/University	1996	New Building	6000	Reinforced Concrete	medium	high	low	very low	very low	medium	low	very low	high	very low	very low	very low
	66	School/Campus/University	2002	New Building	8000	Reinforced Concrete	medium	medium	low	very low	very low	medium	low	very low	medium	very low	very low	very low
	67	School/Campus/University	2005	New Building	3750	Reinforced Concrete	medium	medium	low	very low	very low	high	low	very low	medium	very low	very low	very low
	68	School/Campus/University	1949	New Building	6300	Reinforced Concrete	medium	very high	low	very low	very low	low	low	very low	very high	very low	very low	very low
	69	School/Campus/University	2010	New Building	5700	Reinforced Concrete	medium	high	low	very low	very low	very high	low	very low	very high	very low	very low	very low
	70	School/Campus/University	2005	New Building	1000	Reinforced Concrete	medium	medium	low	very low	very low	high	low	very low	high	very low	very low	very low
	71	School/Campus/University	1996	Refurbishment	6500	Reinforced Concrete	medium	high	low	very low	very low	medium	low	very low	medium	very low	very low	very low
	72	School/Campus/University	2006	New Building	12000	Reinforced Concrete	high	medium	medium	very low	very low	very high	high	very low	medium	very low	very low	very low
	73	School/Campus/University	2013	New Building	5500	Reinforced Concrete	high	medium	low	very low	very low	high	high	very low	medium	very low	very low	very low
	74	School/Campus/University	2000	New Building	800	Reinforced Concrete	high	medium	low	very low	very low	low	low	very low	medium	very low	very low	very low
	75	School/Campus/University	2009	New Building	1000	Reinforced Concrete	high	medium	low	very low	very low	high	low	very low	medium	very low	very low	very low
	76	School/Campus/University	2009	Refurbishment	2000	Reinforced Concrete	high	very high	very high	very low	very low	high	high	very low	high	very low	very low	very low
	77	School/Campus/University	1998	New Building	13750	Reinforced Concrete	high	high	low	very low	very low	high	low	very low	medium	very low	very low	very low

		Part 2: Attributes																	
	ID	Sensors	Experience	Motivation	Communication	Training	Flexibility	Surveys	Passive	Simplicity	Client	Stakeholders	Country	Policy	Public	Climate	Force	Heatingap (%)	Electricity gap (%)
Part 2: Instances	39	high	medium	medium	medium	low	low	very low	high	medium	very low	very low	very low	very low	very low	very low	very low	38,1	-70,2
	40	very high	low	high	medium	medium	low	medium	medium	medium	very low	medium	high	high	high	very low	high	42,5	8,6
	41	high	medium	medium	medium	low	low	very low	medium	high	very low	very low	very low	very low	very low	very low	very low	-6,5	-65,7
	42	very high	low	high	medium	medium	low	very low	low	high	very low	medium	high	high	high	very low	high	-0,7	17,5
	43	high	medium	medium	medium	low	low	very low	medium	high	very low	very low	very low	very low	very low	very low	very low	96,2	-43,6
	44	high	medium	medium	medium	low	low	very low	low	medium	very low	very low	very low	very low	very low	very low	very low	149,6	-37,6
	45	very high	low	high	medium	medium	low	very low	medium	high	very low	medium	high	high	high	very low	high	14,2	15,5
	46	very high	low	high	medium	medium	low	very low	medium	high	very low	medium	high	high	high	very low	high	42,8	-7,2
	47	high	medium	medium	medium	low	low	very low	medium	medium	very low	very low	very low	very low	very low	very low	very low	-23,3	-58,0
	48	high	medium	medium	medium	low	low	very low	medium	medium	very low	very low	very low	very low	very low	very low	very low	76,6	-47,2
	49	high	medium	medium	medium	low	low	very low	medium	medium	very low	very low	very low	very low	very low	very low	very low	-35,3	-67,3
	50	high	medium	medium	medium	low	low	very low	low	medium	very low	very low	very low	very low	very low	very low	very low	-40,2	-78,7
	51	high	medium	medium	medium	low	low	very low	medium	medium	very low	very low	very low	very low	very low	very low	very low	-46,6	-91,2
	52	high	medium	medium	medium	low	low	very low	medium	medium	very low	very low	very low	very low	very low	very low	very low	-19,4	37,9
	53	high	medium	medium	medium	low	low	very low	medium	medium	very low	very low	very low	very low	very low	very low	very low	-38,2	-61,8
	54	high	medium	medium	medium	low	low	very low	medium	medium	very low	very low	very low	very low	very low	very low	very low	35,6	-77,7
	55	high	medium	medium	medium	low	low	very low	medium	high	very low	very low	very low	very low	very low	very low	very low	-94,7	-97,7
	56	high	medium	medium	medium	low	low	very low	medium	high	very low	very low	very low	very low	very low	very low	very low	58,4	-13,8
	57	high	medium	medium	medium	low	low	very low	medium	high	very low	very low	very low	very low	very low	very low	very low	127,4	0,4
	58	very low	low	very low	low	low	low	very low	high	low	very low	very low	very low	very low	very low	very low	very low	45	65
	59	very low	low	very low	low	low	low	very low	medium	low	very low	very low	very low	very low	very low	very low	very low	30	83
	60	very low	low	very low	low	low	low	very low	medium	low	very low	very low	very low	very low	very low	very low	very low	32	73
	61	very low	medium	very low	very low	very low	very low	high	very low	very low	very low	very low	very low	very low	very low	medium	very low	-7,3	13,2
	62	very low	medium	very low	very low	medium	very low	very low	very low	very low	very low	very low	very low	very low	very low	very low	very low	-62,0	69,0
	63	very low	medium	very low	very low	medium	very low	very low	very low	very low	very low	very low	very low	very low	very low	very low	very low	-47,0	29,0
	64	very low	medium	very low	very low	medium	very low	very low	very low	low	very low	very low	very low	very low	very low	very low	very low	43,0	-17,0
	65	very low	medium	very low	very low	medium	very low	very low	very low	very low	very low	very low	very low	very low	very low	very low	very low	54,0	-14,0
	66	very low	medium	very low	very low	medium	very low	very low	very low	very low	very low	very low	very low	very low	very low	very low	very low	2,0	23,0
	67	very low	medium	very low	very low	medium	very low	very low	very low	very low	very low	very low	very low	very low	very low	very low	very low	-9,0	-53,0
	68	very low	medium	very low	very low	medium	very low	very low	very low	very low	very low	very low	very low	very low	very low	very low	very low	-111,0	-19,0
	69	very low	medium	very low	very low	medium	very low	very low	very low	low	very low	very low	very low	very low	very low	very low	very low	28,0	53,0
	70	very low	medium	very low	very low	medium	very low	very low	very low	very low	very low	very low	very low	very low	very low	very low	very low	45,0	-8,0
	71	very low	medium	very low	very low	medium	very low	very low	very low	very low	very low	very low	very low	very low	very low	very low	very low	12,0	-14,0
	72	very low	medium	very low	very low	medium	very low	very low	very low	very low	very low	very low	very low	very low	very low	very low	very low	-15,0	86,0
	73	very low	medium	very low	very low	medium	very low	very low	very low	very low	very low	very low	very low	very low	very low	very low	very low	38,0	61,0
	74	very low	medium	very low	very low	medium	very low	very low	very low	very low	very low	very low	very low	very low	very low	very low	very low	100	-26
	75	very low	medium	very low	very low	medium	very low	very low	very low	very low	very low	very low	very low	very low	very low	very low	very low	24	36
	76	very low	medium	very low	very low	medium	very low	very low	high	low	very low	very low	very low	very low	very low	very low	very low	39	83
	77	very low	medium	very low	very low	medium	very low	very low	very low	very low	very low	very low	very low	very low	very low	very low	very low	35	37

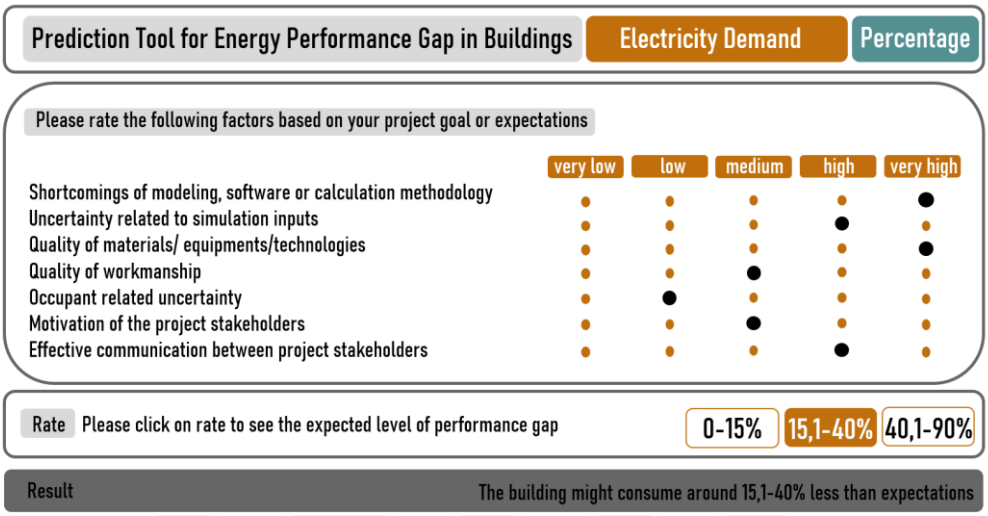
K. Analysis of Electricity Gap Dataset



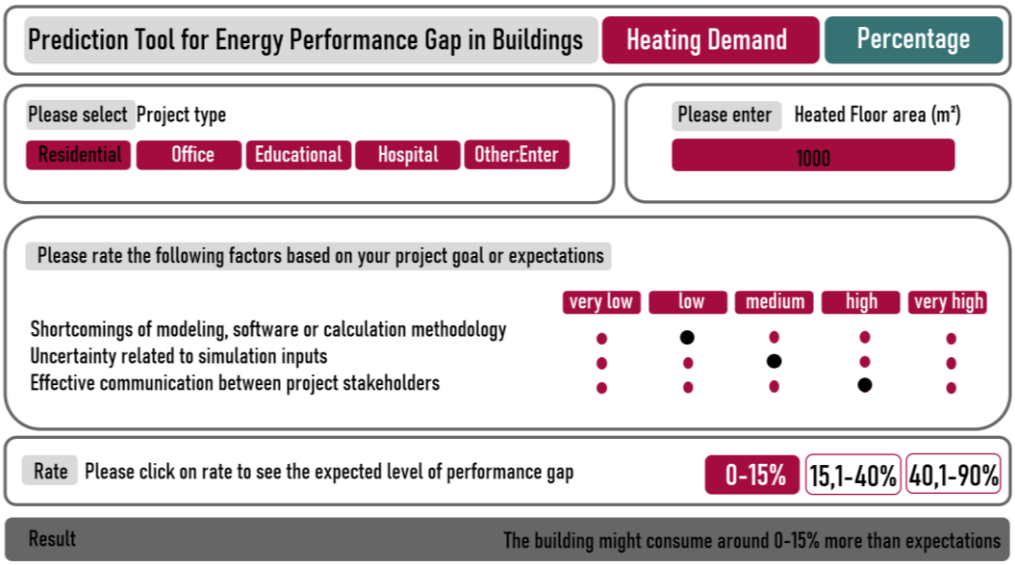
L. Analysis of Heating Gap Dataset



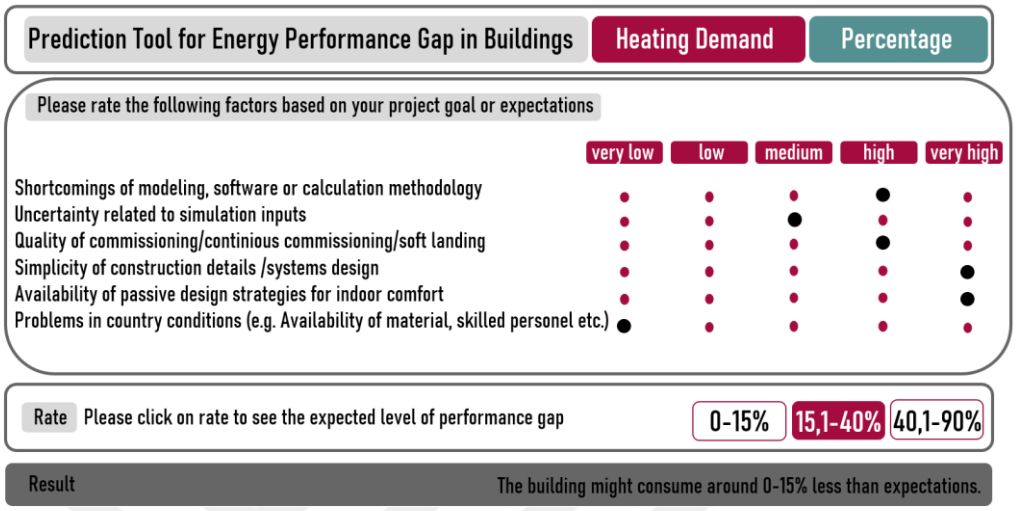
M. Conceptual Web-Interface to Predict the Negative Electricity Performance Gap



N. Conceptual Web-Interface to Predict the Positive Heating Performance Gap



O. Conceptual Web-Interface to Predict the Negative Heating Performance Gap



CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Yılmaz, Derya

Nationality: Turkish (TC)

Date and Place of Birth: 27 Eylül 1985, Zonguldak

Marital Status: Single

Phone: +90 506671415

Email: yilmaz.derya@metu.edu.tr

EDUCATION

Degree	Institution	Year of Graduation
MS	METU Architecture	2014
BS	YTU Architecture	2009
High School	Batıkent High School, Ankara	2003

WORK EXPERIENCE

Year	Place	Enrollment
2021-2022	Bauhaus University Weimar	Visiting Researcher
2018-2020	Bilkent University	Part-time Instructor
2014-2016	Gökhan Aksoy Architecture	Architect
2012-2013	Passive House Institute, Darmstadt	Intern Architect
2011-2011	Ercan Çoban Architecture	Architect
2009-2010	Şahinbaş-Fikirlier Architects	Architect
2009 May	Emre Arolat Architects	Intern Arc. Student
2007 June	Yapı Merkezi Construction and Industry Inc.	Intern Arc. Student
2006 June	Mimarlar and Han Tümertekin Project Consultancy Services Ltd.	Intern Arc. Student

FOREIGN LANGUAGES

Advanced English, Fluent German, Beginner Portuguese

PUBLICATIONS

1. A Study on Multi Objective Optimization Applications Used to Optimize the Shape of Buildings for Energy Efficiency, Central Europe towards Sustainable Building Conference, Prag, Czech Republic, 2-4 July 2019

2. A Review on Sustainability in School Facilities through the Perspective of Architects, 5th International Project and Construction Management Conference, Kyrenia, Northern Cyprus, 16-18 November 2018

3. Impacts of Detailing on Annual Heating Demand in a Passive House Building,

2nd International Sustainable Buildings Symposium, Ankara, Turkey, 28-30 May 2015

4. Evaluation of a Tunnel Form Social Housing Unit based on Passive Design Principles, 2nd International Sustainable Buildings Symposium, Ankara, Turkey, 28-30 May 2015

Travelling, Tennis, Dance, Documentaries