AMERICAN UNIVERSITY OF BEIRUT

AN APPLICATION OF NEURAL NETWORKS TO FORECAST BUILDING OCCUPANT COMPLAINTS

by
SENA ALI ASSAF

A thesis
submitted in partial fulfillment of the requirements
for the degree of Master of Engineering
to the Department of Civil & Environmental Engineering
of the Maroun Semaan Faculty of Engineering and Architecture
at the American University of Beirut

Beirut, Lebanon
September 2020

AMERICAN UNIVERSITY OF BEIRUT

AN APPLICATION OF NEURAL NETWORKS TO FORCAST BUILDING OCCUPANT OCMPLAINTS

by

SENA ALI ASSAF

Approved by:

Dr. Issam Srour, Associate Professor — Advisor
Dept. of Civil and Environmental Engineering

Dr. Ibrahim Alameddine, Assistant Professor — Member of Committee
Dept. of Civil and Environmental Engineering

Dr. Jordan Ludders Srour, Assistant Professor — Member of Committee
Dept. of Information Technology and Operation Management

Date of thesis/dissertation defense: September 7th, 2020

AMERICAN UNIVERSITY OF BEIRUT

THESIS, DISSERTATION, PROJECT RELEASE FORM

Student Name:
_____Assaf_________________ Sena____________________ Ali__
Last First Middle

● Master's Thesis ○ Master's Project ○ Doctoral Dissertation

☐ I authorize the American University of Beirut to: (a) reproduce hard or electronic copies of my thesis, dissertation, or project; (b) include such copies in the archives and digital repositories of the University; and (c) make freely available such copies to third parties for research or educational purposes.

☑ I authorize the American University of Beirut, to: (a) reproduce hard or electronic copies of it; (b) include such copies in the archives and digital repositories of the University; and (c) make freely available such copies to third parties for research or educational purposes after :

One ---- year from the date of submission of my thesis, dissertation, or project.

Two ---- years from the date of submission of my thesis, dissertation, or project.

Three ---- years from the date of submission of my thesis, dissertation, or project.

Sena. A September 11 2020

Signature Date

ACKNOWLEDGMENTS

I would like to provide many thanks for my thesis advisor Dr. Issam Srour for his continuous support throughout this two-year journey. Dr. Issam was a great mentor who showed exceptional support, assistance, and guidance during both regular follow-ups and challenging times. Being part of Dr. Issam's research team has grown my research skills, analytical thinking skills, and professionalism skills.

My deep gratitude also goes to the committee members Dr. Ibrahim Alameddine and Dr. Jordan Srour for providing their time to be part of this thesis work, and for their insightful feedback during the comprehensive exam.

Also, I would like to thank Mohamed Awada, who was part of Dr. Issam Srour research team, for assisting me to get started with this research work and for providing proper guidance and follow up at the early stages.

I would also like to thank my parents, friends, and family for being a resilient support system throughout this journey.

AN ABSTRACT OF THE THESIS OF

Sena Ali Assaf for Master of Engineering
Major: Civil and Environmental Engineering

Title: An Application of Neural Networks to Forecast Building Occupant Complaints

Occupant complaints are a reflection of poor building performance and an unsatisfactory indoor environment. One way to mitigate those complaints and to ensure occupants' satisfaction with regards to building performance is through a well-performing facility management that is capable of planning for and addressing maintenance services. This thesis proposes a machine learning-based multistep generic framework to analyze occupant complaint data and to forecast the number of thermal complaints in particular for the upcoming week as part of the facility management's predictive maintenance approach. Moreover, the developed forecasting model is benchmarked against a traditional statistical model to ensure proper performance. The proposed methodology was tested for a period of three years on a highly unstructured and unsolicited occupants' complaints data recorded by facility management operators in a residential complex composed of 16 buildings. Text mining results of more than 6,000 occupant complaints showed that thermal related complaints are among the most common ones thus require further attention of facility managers. The developed Multi-Layer Perceptron (MLP) models to forecast the number of thermal complaints for the upcoming week showed proper performance with improvements over the traditional Autoregressive Integrated Moving Average (ARIMA) model with a higher ability to generalize to new data. It is also evident that the developed MLP forecasting models could assist facility managers in planning for the staffing resources required to handle these complaints thus enhancing occupant satisfaction and building performance.

TABLE OF CONTENTS

ACKNOWLEDGMENTS v

TABLE OF CONTENTS vii

ILLUSTRATIONS ix

TABLES xi

ABBREVIATIONS xii

1. INTRODUCTION 1

2. LITERATURE REVIEW 4

A. Facility Management and Building Occupant Complaints 4

B. Building Maintenance and Resource Management 6

C. Data Analytics and Machine Learning 8

D. Previous Works: Potential of Building Occupant Complaints to Assist FM 10

3. RESEARCH OBJECTIVE AND METHODS 15

A. Research Objective 15

B. Materials and Methods 16

1. Model Selection 17

2. ML- Based Generic Framework to Forecast Thermal Occupant Complaints 18

3. Benchmark Model: ARIMA 37

4. RESULTS AND DISCUSSION: CASE STUDY 44

A. An application of the Developed ML-Based Generic Framework 44

1. Data Collection and Preprocessing 44

2. Text Cleaning and Mining 52

3. Data Split 60

4. Time Series Smoothing 61

5. MLP Input Data Preparation 63

6. MLP model training, validation, and testing 69

B. Benchmark Models Development 78

1. Time Series Preparation 78

2. ARIMA Input Data Preparation 78

3. Splitting Design Data into Training and Validation 79

4. ARIMA (p, d, q) (P, D, Q) s Modeling 80

C. MLP and ARIMA Models Comparison and Discussion 89

D. Work Significance: Resource Staffing Plan 90

E. The Need for Proper Data Management 100

5. CONCLUSIONS AND FUTURE WORKS 102

APPENDIX A 105

REFERENCES 109

ILLUSTRATIONS

Figure Page

1. Proposed ML-based generic framework to forecast thermal complaints 19
2. ANN structure 28
3. Structure of a neuron 28
4. Walk forward validation method (Hyndman & Athanasopoulos, 2018, p.84) 35
5. ARIMA methodology 43
6. Maintenance data high-level dictionary 47
7. Time series plots for potential weather features 51
8. TDM example 52
9. Word cloud for year one 54
10. Word cloud for year two 54
11. word cloud for year three 55
12. Most frequent terms for year one 55
13. Most frequent terms for year two 56
14. Most frequent terms for year three 56
15. Developed AC time series 58
16. Developed heater time series 59
17. AC time series split between design and holdout sets 60
18. Heater time series split between design and holdout sets 61
19. AC time series smoothing for design set 62
20. Heater time series smoothing for design set 63
21. Scatter plot matrix 65
22. Density plots 67

23. AC time series split into training and validation 69

24. Heater time series split into training and validation 69

25. AC MLP model actual and predicted values in train set 73

26. AC MLP model actual and predicted values in validation set.... 74

27. AC MLP model actual and predicted values in test set 75

28. Heater MLP model actual and predicted values in train set 76

29. Heater MLP model actual and predicted values in validation set.... 76

30. Heater MLP model actual and predicted values in test set 78

31. AC time series ACF and PACF plots 81

32. AC ARIMA model actual and predicted values in train set 82

33. AC ARIMA model actual and predicted values in validation set.... 83

34. AC ARIMA model residual diagnostics 83

35. AC ARIMA model actual and predicted values in test set 84

36. Heater time series ACF and PACF plots 85

37. Heater ARIMA model actual and predicted values in train set 86

38. Heater ARIMA model actual and predicted values in validation set 87

39. Heater ARIMA model diagnostics 88

40. Heater ARIMA model actual and predicted values in test set 88

41. Number of predicted thermal complaints 91

42. Planned and predicted weekly man hours.... 93

43. The variation of the number of understaffed man hours 97

44. The variation of the number of overstaffed man hours 98

45. The variation of the number of technicians per month 106

46. The variation of the percentage of thermal complaints per month 108

TABLES

Table Page

1. Summary of previous works addressing the potential of building occupant complaints to assist FM ... 13
2. Summary statistics of weather features ... 50
3. TDM summary per year ... 53
4. AC and heater time series summary statistics ... 59
5. Time series Spearman correlation ... 66
6. AC and heater MLP models hyperparameters and characteristics ... 72
7. ARIMA AC model iterations for estimation step ... 81
8. ARIMA heater model iterations for estimation step ... 86
9. MLP and ARIMA models comparison ... 89
10. Assumptions for each staffing option ... 106
11. Allocation of technicians per day per shift ... 107

ABBREVIATIONS

FM: Facility Management

AC: Air Conditioner

OC: Occupant Complaint

ML: Machine Learning

ANN: Artificial Neural Networks

MLP: Multi-Layer Perceptron

ARIMA: Autoregressive Integrated Moving Average

MinT: Minimum Temperature

AvT: Average Temperature

MaxT: Maximum Temperature

MinRH: Minimum Relative Humidity

AvRH: Average Relative Humidity

MaxRH: Maximum Relative Humidity

MinW: Minimum Wind

AvW: Average Wind

MaxW: Maximum Wind

CHAPTER 1

INTRODUCTION

A Facility Management (FM) unit in a building is responsible for running the operation phase of a built and occupied facility. Its role is to ensure that the users are satisfied with the different aspects of the building (Shin, Lee, Park, & Lee, 2018). Moreover, it aims at ensuring that the building is running at its optimum performance (Nutt, 1999) which can only be guaranteed through a continuous long-term maintenance strategy (Ismail, Ali, Othman, & Jaffar, 2017).

Building maintenance is one important service under the scope of work of the FM for two main reasons. The first is that a high number of maintenance requests is a reflection of unsatisfied occupants and underperforming building systems and equipment, and the second is that maintenance activities represent a large portion of the operational costs of a facility (Higgins, Mobley, & Wikoff, 2008). Reactive maintenance, although very common among facility managers, often incurs additional unnecessary costs on different levels since actions are taken once the equipment or system failure had already occurred (Akcamete & Akinci, 2004; Higgins et al., 2008). Preventive maintenance could also incur additional costs since it is based on routine and prescheduled maintenance which might lead to unnecessary maintenance (Wahab, Samad, & Basari, 2013). With the development of data analytics and machine learning (ML) tools, it has become feasible for the FM to adopt a predictive maintenance strategy. By predicting occupant complaints, the FM is able to prepare the necessary resources needed to resolve the issue before failure occurs thus preventing unexpected equipment breakdown (Sipos, Fradkin, Moerchen, & Wang, 2014). However, records of occupant complaint logs are large and highly unstructured datasets that often include a textual description of the complaint. The use of text mining techniques makes it possible to restructure and analyze such data in order to extract further useful information (Bortolini & Forcada, 2019; Gunay, Shen, & Yang, 2019).

It is crucial for the FM to adopt a predictive maintenance strategy for the maintenance services in order to mitigate the overwhelming expenses associated with a building's operation phase and to ensure that the occupants are satisfied with the building performance and the indoor environment. This however heavily relies on historical records of maintenance requests obtained from unsolicited occupant complaints. Moreover, it requires proper staffing planning to avoid extra costs of over or under staffing and to ensure the failure is resolved at the right time to minimize equipment downtime (Elazouni & Shaikh, 2008). To this end, this research work presents a generic framework based on data analytics and advanced ML tools to assist FM first to better understand building occupant complaints (type, frequency, and correlations), and second to forecast the frequency of thermal complaints. This framework acts as a sound decision making tool to properly allocate staffing resources among the corresponding maintenance activities to ensure occupants' satisfaction, maintenance cost savings, and an adequate building performance.

Chapter two of this thesis presents an overview of the FM's scope of work in addition to its role in handling building occupant complaints and maintaining the building systems. It also discusses previous works in the literature that have used occupant complaint data to assist the FM's decision making. Chapter three discusses the objectives and significance of this research work. It also describes the methodology followed to develop and evaluate the ML-based generic framework to develop a machine learning model to forecast thermal complaints and the corresponding benchmark model. The results of an application of the proposed methodology on a selected case study are presented and discussed in chapter four. Chapter five presents conclusions for this research work along with future work.

CHAPTER 2

LITERATURE REVIEW

This chapter is divided into four subsections as follows: facility management and building occupant complaints, building maintenance and resource management, data analytics and machine learning, and previous works: the potential of building occupant complaints.

A. Facility Management and Building Occupant Complaints

The role of the FM in a building is to ensure that the users feel safe and satisfied in a friendly environment (Shin et al., 2018). Its scope of work covers several activities ranging from handling of physical issues (services, maintenance, adaptation, built space etc.) to human and business concerns (use and function, comfort, safety, security etc.) and even to financial concerns (occupancy cost and benefits, property investment etc.). Most importantly, it aims at integrating the decisions among the three latter areas of concern through proper management to enhance the productivity, usage and performance of the facility (Nutt, 1999). In this context, the FM's strategic role requires planning, designing, and ensuring continuous improvement of the service quality (Alexander, 2003) to ensure that the delivered services meet the expectations of the occupants (S. Y. Lee, 2002). Unfortunately, occupants are likely to complain when the facility's performance fails to meet their needs. It is the responsibility of the FM to address and resolve those complaints (Goins & Moezzi, 2013) while ensuring optimum resource allocation among all its services (Kral & Bartosova, 2016). Occupant complaints can be of two types: either volunteered or solicited. Volunteered complaints are raised voluntarily by occupants when they are not feeling comfortable in their environment physically or psychologically. On the other hand, collecting solicited complaints is done through a specific request such as a survey addressing occupants' satisfaction. In this case, the survey questions guide the arising complaints thus might miss on some other complaints that could be more important, disturbing, or easy to deal with (Goins & Moezzi, 2013).

In a first attempt to understand unsolicited building occupant complaints, Federspiel (1998) studied a large set of unsolicited occupant complaints data obtained from commercial buildings. A statistical analysis showed that the most frequent types of unsolicited complaints were those associated with thermal sensation (too hot or too cold complaints) accounting for 77 percent of the total unsolicited environmental complaints. A more recent study conducted by the International Facility Management Association (2009) also showed that thermal complaints are the most common among office employees. Moreover, a survey conducted by Goins and Moezzi (2013) among the occupants of 575 buildings in a university campus showed that 43 percent of the respondents were dissatisfied with the indoor temperature which does not meet the threshold of the 80 percent occupant satisfaction stated by ASHRAE 55-2010 standard (ASHRAE, 2010). Thermal complaints appear to be the most common among building occupants regardless of the building use and thus must be given wide attention of the FM.

B. Building Maintenance and Resource Management

According to Weinstein (1997), maintenance can be defined as the set of activities that aim at preserving or restoring systems and components to ensure that they provide the optimal performance level and function they were designed for. Building maintenance is one major aspect of the FM's scope of work that could tremendously affect the building's performance (Horner, El-Haram, & Munns, 1997). Costs associated with maintenance activities represent a large portion of the total costs of the operation phase of a building (Madureira, Flores-Colen, de Brito, & Pereira, 2017). A well-performing FM necessitates continuous maintenance to ensure that the facility will remain in good conditions (Nawi, Baharum, Ibrahim, & Riazi, 2017) and suitable for its intended use. Maintaining buildings on a regular basis allows owners to maximize both their profit and the building performance at the lowest possible cost and ensures a comfortable and suitable environment for the occupants (Ismail et al., 2017). One way to assist the FM to evaluate the facility's performance and to mitigate the high costs of maintenance activities would be through analyzing maintenance request data obtained from records of unsolicited building occupant complaints (Bortolini & Forcada, 2019) and taking relevant managerial and operational actions.

Practitioners in the FM industry often adopt a reactive strategy to maintain the building systems and equipment (Akcamete & Akinci, 2004) through which corrective actions are taken when the breakdown or damage had already occurred (Higgins et al., 2008). Despite the convenience of such strategy, it is often considered costly and myopic. On the other hand, predictive maintenance is a proactive maintenance strategy that makes use of past equipment failure data as a main source to predict when a breakdown will occur. This allows to plan and schedule maintenance in advance to the breakdown thus preventing unforeseen equipment downtime and saving on the long-term cost of equipment maintenance (Akcamete & Akinci, 2004). A practical way to obtain this past failure data is often achieved through the use of sensors that monitor the equipment's performance one of which is equipment-failure's time and frequency of occurrence. However, this is not always a feasible solution due to the high cost and effort sensors require to be installed and to remain functional (Sipos et al., 2014). When there are no records of equipment logs that record failure data, an alternative solution would be using maintenance request data obtained from unsolicited occupant complaints' description as a source for equipment failure data. This is because occupants are likely to complain when they are not satisfied with the building performance and thus would issue a complaint when a certain equipment fails to perform its function.

Maintenance management strategies require proper planning at both strategic and operational levels of an organization (Lee & Scott, 2009). At an operational level, a maintenance activity entails the presence of sufficient resources depending on the size of the facility. These include human resources (technical crews, engineers, supervisors etc.), their required skill levels, suitable tools, maintenance time requirement, in addition to a maintenance schedule or plan (Márquez, León, Fernández, Márquez, & Campos, 2009). Maintenance personnel often argue that most of the times the budget and resources allocated for maintenance activities are inadequate. As a result, it becomes crucial for the top management at a strategic level to ensure optimal resource allocation among the different services it provides one of which is the maintenance services (Lee & Scott, 2009). This allows to avoid significant problems such as crew over or under staffing where both cases would eventually lead to additional operation and maintenance costs and building users' dissatisfaction (Elazouni & Shaikh, 2008).

Given the benefits associated with predictive maintenance strategies and the corresponding need for structured historical records of building occupant complaints, this research work entails the use of data analytics and ML tool.

C. Data Analytics and Machine Learning

When used in the context of building occupant complaints, data analytics, text mining techniques, machine learning tools, and time series analysis tools could act ought to be useful in assisting the FM's decision-making.

Data analytics is the process of analyzing a set of collected data to extract insights that are of added value to the decision makers (Kelleher, Mac Namee, & Aoife, 2015). Occupant complaint descriptions are often recorded in a textual form. What is special about this type of data is that the information to be inferred is clearly stated in the text; however, it is not in a form that makes it suitable for further use by humans or computers (Witten & Frank, 2005). Thus, text mining becomes useful to study such large amount of written text to transform it into structured data for future analysis (Aggarwa & Zhai, 2012). ML can be defined as the process of training a computer model on a training dataset to perform a certain task so that it will be able to perform that exact task when given new data it had not encountered before (Panos & Christof, 2016). It is often used to build predictive models. Predictive data analytics is a subcategory of data analytics that makes use of historical data to extract patterns in order to develop predictive models (Kelleher et al., 2015).

Considering that the use and thus the frequency of various types of occupant complaints vary depending on the time of the year, it is important to study the variation of those complaints as a function of time. The sequence of the observations in successive order over a certain period of time is known as a “time series”. Time series analysis allows to identify and analyze trends, patterns, and seasonality effects in the sequence understudy (Tabachnick, Fidell, & Ullman, 2007). Forecasting time series data aims to estimate how the sequence of the data under study will continue in the future (Hyndman & Athanasopoulos, 2018). Among the most common time series forecasting methods are the Autoregressive Integrated Moving Average (ARIMA) model and the Multi-Layer Perceptron (MLP) which is a type of feedforward Artificial Neural Network (ANN) whereby both models have been employed in several time series forecasting applications (Babu & Reddy, 2014). The autoregressive part of the ARIMA model ensures its ability to regress the variable of interest against itself, which means that it uses a linear combination of past values of this variable as a predictor to forecast future values of the variable of interest The moving average part of the ARIMA on the other hand ensures its ability to linearly model past errors that are assumed to be independently distributed with a normal distribution. As for the integration part of the ARIMA model, it is employed to ensure that the data is made stationary before fitting the model (Hyndman & Athanasopoulos, 2018). In addition to that, it has the ability to take as an input additional exogenous data that might influence the prediction other than lagged values of the variable of interest. It can also be extended to include seasonal effects for example if seasonality is observed on a weekly, monthly, or yearly basis (Box & Tiao, 1975). ANNs on the other hand are ML algorithms that have the ability to learn from a training data set, store the information, and recall it when needed. They are capable of representing highly complex and non-linear problems. (Yu, Zhu, & Zhang, 2012). ANN are used in a wide variety of applications including optimization, classification, clustering, recognition, and prediction (Mohammed, Hamdan, Abdelhafez, & Shaheen, 2013). ANNs are advantageous over ARIMA models for their ability to model any form of non-linear mapping without any prior assumptions regarding the data being studied (G. P. Zhang, 2003).

D. Previous Works: Potential of Building Occupant Complaints to Assist FM

Previous works in the literature have studied and analyzed building occupant complaints to understand them and to show their potential in supporting the FM's work using different techniques ranging from statistical modelling to advanced machine learning.

Federspiel (1998) investigated unsolicited occupant complaint data obtained from computerized logs in commercial buildings. A statistical analysis showed that thermal complaints, being the most common, were triggered by an unsatisfactory performance of the HVAC systems, whether fault detection or inadequate control. A mathematical model was then developed in a follow-up paper to predict the frequency of thermal complaints based on past complaint data as a function of certain properties of the indoor temperature. The model's purpose was to assist the FM in making sound economic decision such as determining the optimum indoor temperature and evaluating the benefits of upgrading the temperature control system on the cost of both occupant complaints and energy consumption (Federspiel, 2000). The prediction model was recalibrated to enhance its accuracy. It showed good performance when used to design for the optimum building indoor temperature with the objective of minimizing the thermal complaints' frequency or the combined cost associated with energy, operations, and maintenance (Federspiel, Rodney, & Hannah, 2003).

In order to mitigate the unstructured, variable, and static nature of the collected occupant-generated work orders, Mcarthur et al. (2018) developed an automatic work order classification system to better organize and handle work orders. The first level of the system classifies the work order based on the trade of the complaint such as "plumbing" and the second level identifies the subcategory of the issue such as "shower". A set of follow-up questions were also developed for each subcategory. Several classification models for each of the two levels were trained, tested and cross validated based on historical data of work orders collected from a university's set of buildings. Integrating this classification system with BIM visualization allows for real-time follow-up with the occupants regarding the generated work order instead of conducting multiple trips to acquire more information to better define the problem and thus resolve it. This acts as a tool to prioritize the work orders and identify how urgent each is and thus the response.

Gunay et al. (2018) studied the impact of the indoor and outdoor climate settings on the generated unsolicited thermal complaints by analyzing temperature setpoint change data. They demonstrated their work on data obtained from a Building's Automation System (BAS) from a set of office buildings. Building operators are responsible for changing the temperature setpoint upon receiving a complaint call from the occupants since they do not have direct access to do such modification. They developed predictive models using Markov logistic regression to forecast the chance of observing a temperature setpoint change. Forecasting the frequency of thermal complaints by predicting the number of thermostat setpoint change requests assists the FM to benchmark occupants satisfaction and to design for the optimum indoor temperature to minimize the complaints' frequency.

In another study, Gunay et al. (2019) introduced a systematic method using the process of text-mining to extract and analyze useful information from unstructured work order logs obtained from a CMMS from a set of university buildings and a heating and cooling plant. The work order logs were first pre-processed to obtain the frequency of each term per each work order. Top terms included the names of the systems or components, adjectives describing their characteristics, and verbs describing actions taken to address the request. Clustering the work orders allowed to isolate the terms that address HVAC failures. Association rule-mining was then conducted within this cluster to discover certain relationships and patterns among the different terms such as linking an equipment's name to the adjective describing it. A probabilistic model was then put together to predict the probability of not observing a certain failure or warning during a certain period of time. This method of text mining and statistical modeling provides valuable insights for the FM to benchmark the facility's maintenance performance.

Text mining was also used by Bortolini and Forcada (2019) to analyze maintenance request data extracted from a CMMS in a set of laboratories office buildings, and academic buildings. The obtained data was classified into categories depending on the problem type they address. Each category was characterized by a set of keywords defined based on the most frequent terms. Then they were also classified into three levels based on how severe the problem is. This allows to better understand the number, type, and severity of the maintenance requests and the level of performance of the building systems. The FM can then develop preventive maintenance strategies to ensure that occupants are satisfied and that the building systems are functioning at the required performance levels.

Table 1 summarizes the collected studies from the literature that have used building occupant complaints to support the FM's scope of work including: source of the occupants' complaint data, the methods of analysis and the contribution of this work to the FM.

Table 1 Summary of previous works addressing the potential of building occupant complaints to assist FM

Reference	Data Source	Methods	Contribution to FM
(Federspiel, 1998)	Computerized logs form commercial buildings	Statistical analysis	To investigate the common types of unsolicited building occupant complaints
(Federspiel, 2000)	Computerized logs form commercial buildings	Mathematical modelling	To predict the frequency of thermal complaints To make sound economic decisions
(Federspiel et al., 2003)	Computerized logs form commercial buildings	Mathematical modelling	To design for the optimum building indoor temperature
(Mcarthur et al., 2018)	Work order logs from a university's set of buildings	Supervised ML (classification) BIM	To prioritize the work orders To identify how urgent each work order is and thus the response
(Gunay et al., 2018)	BAS from a set of office buildings	Markov logistic regression	To predict the frequency of thermal complaints To benchmark occupants satisfaction To design for the optimum indoor temperature

(Gunay et al., 2019)	CMMS from a set of university buildings and a heating and cooling plant	Text mining Statistical modelling	To predict the probability of not observing a certain equipment failure or warning To benchmark the facility's maintenance performance
(Bortolini and Forcada, 2019)	CMMS from a set of laboratories, office buildings, and academic buildings	Text mining	To better understand the maintenance requests and the performance level of the building systems To develop preventive maintenance strategies

CHAPTER 3

RESEARCH OBJECTIVE AND METHODS

A. Research Objective

The FM's research literature has shown how important it is for the FM to study and analyze historical records of unsolicited building occupant complaints. It is also evident that using certain data analytics and ML tools could significantly assist the FM in several aspects of its scope of work to eventually ensure that the occupants are satisfied with the building operations and that the building is running at its optimum performance.

Previous works have assisted the FM to make sound economic decisions when it comes to identifying the optimum building indoor temperature for example, to develop an automated work order classification system, to benchmark occupants' satisfaction and the facility's maintenance performance, in addition to developing preventive maintenance strategies. Despite the potential of the previous works to assist facility managers, none has addressed forecasting unsolicited building occupant complaints for the purpose of developing a specific staffing plan for a building's maintenance activities.

The significance of this research work relies on presenting a sound decision making tool for FM to plan for staffing resources with regards to maintenance activities related to AC and heater use. Its aims to guide FM on how it could adopt a predictive maintenance strategy based on previous records of occupant complaints and how to plan ahead of time for the corresponding maintenance activities. The overall aim is to mitigate the unnecessary maintenance costs by better planning the staffing resources and to ensure that occupants are satisfied with the building's performance.

This research work presents a multi-step generic framework to collect, clean, restructure and analyze highly unstructured historical data of building occupant complaints with the help of data analytics and text mining techniques. The framework also incorporates the use of machine learning tools for time series forecasting in order to develop and evaluate a neural network model to forecast building occupant complaints for thermal complaints in particular. Moreover, this research work presents in great details how the developed forecasting models can be benchmarked against a traditional statistical model to ensure proper performance. The proposed methodology is then tested on an occupant complaint dataset obtained from a facility management unit of a residential complex comprised of 16 buildings for a period of three years. Moreover, the significance of this work is reflected upon implementing the forecast results in a staffing model that will be used to dynamically plan for staffing resources needed to address the corresponding maintenance activities.

B. Materials and Methods

The methodology of this research work is divided into three parts. The first part justifies the reasoning behind the model selection for the time series forecasting problem of thermal complaints. The second part proposes an ML-based generic framework that the FM could adopt to analyze building occupant complaints and to forecast the frequency of thermal complaints in particular. As for the third part, it provides detailed steps on developing the statistical based time series forecasting model that will be used as a benchmark for the results from the ML-based model.

1. Model Selection

There are several models from traditional statistical analysis and advanced machine learning that are used for time series forecasting which is at the core of this research work. As mentioned in section C of chapter 2, ARIMA and MLP models are among the most common ones. ARIMA is a linear model which assumes a linear relationship between historical data and between past errors. It also requires the time series to be made stationary before fitting the model (Babu & Reddy, 2014). Linear models have the advantage of being easily understood, implemented, and analyzed. However, real-life problems are often complex and thus might not satisfy the linearity assumption (G. Zhang, Patuwo, & Hu, 1998). The Multi-Layer Perceptron (MLP) ANN is a powerful feedforward network that is able to learn any form of continuous non-linear mapping (Behrang, Assareh, Ghanbarzadeh, & Noghrehabadi, 2010). ANNs could be employed in time series forecasting since they do not assume linearity of the data and are capable of non-linear modeling. They also do not require the time series data to be stationary nor do they require any specific model form (G. P. Zhang, 2003). As such, ANNs are data driven nonparametric models that requires only few assumptions regarding the process by which the data was collected from. Another property of ANNs is their ability to adapt and ensure accurate generalization capabilities when certain characteristics of the time series change over time. Similar to the ARIMA model, ANNs can incorporate in addition to time-lagged values of the variable of interest other predictors that might influence the output (G. Zhang et al., 1998).

Moreover, both MLP and ARIMA models could incorporate exogenous data as predictors in addition to the lagged variables of the time series. In fact, including lagged variables of the time series as predictors is vital when it comes to predicting occupant complaints because the frequency of complaints is not going to be the same for each day. As a matter of fact, if on a Monday the FM received an extensive amount of complaints, there is a much less chance to get a high number of complaints on the day after, for example. The chance of receiving fewer complaints the next day is justified by the fact that the complaints from the previous day would have been addressed and resolved, thus making it less likely for the occupants who initially filed a complaint, to file it again the next day.

Thus, the MLP model shows great potential to model non-linear dynamic systems one of which is time series modeling of thermal complaints with the ability to account for additional exogenous data. The developed MLP models to forecast thermal complaints will be benchmarked against the corresponding traditional state of the art ARIMA model.

2. ML- Based Generic Framework to Forecast Thermal Occupant Complaints

This section describes the developed ML- based multi step generic framework which includes data collection and preprocessing, text cleaning and mining, time series smoothing, data splitting, in addition to MLP data preparation, training, validating, and testing. The flowchart in figure 1 summarizes the proposed multi-step generic framework.

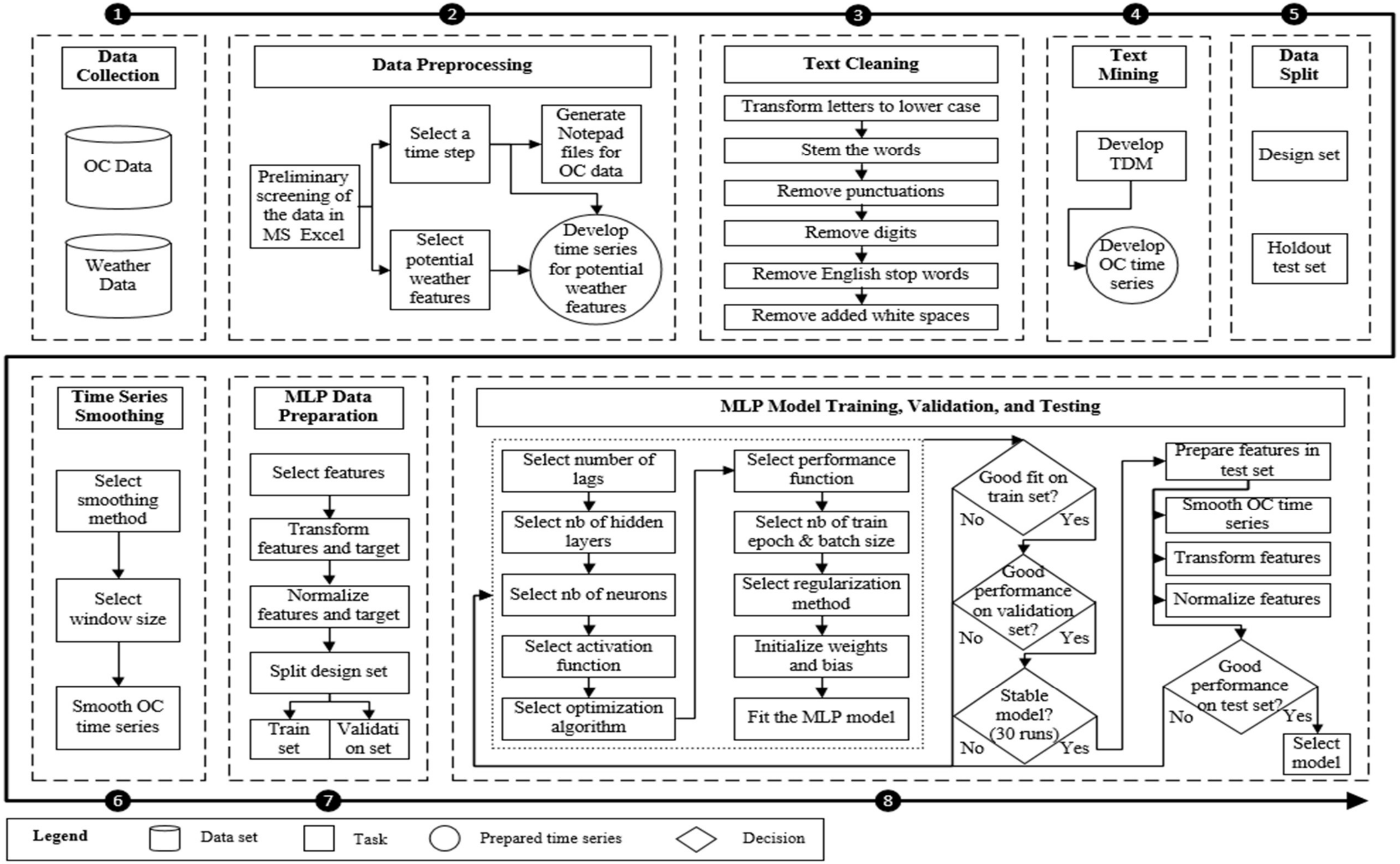

Figure 1 Proposed ML-based generic framework to forecast thermal complaints

a. Data Collection and Preprocessing

Historical records of occupant complaint (OC) data could be obtained from the facility management unit with a description of the complaints issued by occupants and recorded by building operators. An OC is a representative of a maintenance request since the complaint was issued due to a certain disruption in the building that requires actions to be taken by building operators. Since this research work addresses forecasting thermal complaints, weather data could act as an important predictor for the reason that the use of the heater and AC heavily depends on the outdoor weather conditions. Weather-related data could be obtained from the meteorological department. Both obtained data sets are exported to Microsoft Excel where they are thoroughly investigated to better understand each set and then to go through a process of preliminary screening in order to remove irrelevant features, to remove duplicate observations, to handle missing data, and to deal with any discrepancies.

At this stage, a time step for prediction (ex: day, week, month etc.) is selected based on data availability and on the ultimate application of the forecasting tool in the FM's scope of work. It is important to include the time factor since the use of the AC and the heater and thus their complaints' frequencies depend on the time of the year.

After that, the screened data will need some preprocessing as described by the following steps:

- Import the text description of the complaints of the OC data on Microsoft Excel to Notepad files based on the selected time step (ex: week1, week2, etc.).
- Identify the most relevant features of the weather data that could potentially be used as inputs for the prediction model to be developed.

- Develop time series for the selected features from the weather data based on the selected time step. Summary statistics of each of these weather features could be used as an input (ex: minimum, average, maximum).

b. Text Cleaning and Mining

The OC descriptions text data is cleaned as described below after inputting the Notepad files per the selected time step into the software R:

- Transform all letters to lower case to ensure that the model for example does not treat the terms "Apartment" and "apartment" as distinct words.
- Stem the words to ensure that the model for example does not treat the terms "plumber" and "plumbing" as distinct words.
- Remove punctations by replacing them with white spaces since they are of no interest in this study.
- Remove digits by replacing them with white spaces since they are of no interest in this study.
- Remove English stop words by replacing them with white spaces since they are of no interest in this study (ex: "and", "or", "the" etc.).
- Remove all the added white spaces.

Now that the OC data is cleaned, it is used in the text mining process in R using the "tm" package. Text mining allows to develop a Term Document Matrix (TDM) which is a mathematical matrix that permits quantitative analysis of the data. Rows in a TDM represent all the terms found in all the Notepad files, columns represent each

Notepad file (each represents a single time step), and numbers in cells represent how many times each term was repeated in each Notepad file. Given the obtained frequencies of the terms per time step, the time series representing the frequencies of the terms of interest ("ac" and "heater") are thus developed.

c. Data Split

At this stage, there will be a set of weather time series that could potentially be used as features to the MLP model. Also, the target time series of what the model is trying to predict (AC or heater occupant complaints) are developed as well. This data is structured in a table format where each column represents one input feature and the last column represents the target, and each row represents an instance meaning one-time step.

As such, the total data instances as per the selected time step should be divided into two parts. The first part will be referred to as the "design set" which will be used develop the MLP model and to ensure that it is performing well on data it had been trained on. The second part is known as the "holdout test set" that will be used to test the final model's ability to generalize, meaning to test its performance on data it had not encountered before (James, Witten, Hastie, & Tibshirani, 2013).

d. Time Series Smoothing

The AC and heater time series of the design set that are obtained from the process of cleaning and mining will have some embedded noise that is treated using smoothing techniques to ensure that their characteristics stand out and to reduce random fluctuations (Guiñón, Ortega, García-Antón, & Pérez-Herranz, 2007). The generated smoothed time series for both AC and heater complaints are then used as an input in the prediction model of each.

The most common time series smoothing techniques are the Moving Average (MA) filters that replace an observation by the average of a number of neighboring points depending on the selected window size (Guiñón et al., 2007). An arithmetic MA considers equal weights of the points unlike a weighted MA (Hyndman & Athanasopoulos, 2018). A Gaussian weighted MA for example assigns a normal distribution for the weights (Shumway & Stoffer, 2017) and an exponentially weighted MA puts more weight on recent data points with an exponential decay of the weights (Achelis, 2001).

e. MLP Data Preparation

The MLP model for each the AC or heater will have two types of inputs. The first is the AC or heater time series obtained after smoothing the corresponding time series obtained from the text mining process. The OC time series will be used as an input to account for the occupant complaints in the previous time steps (lagged variable). The second corresponds to additional exogeneous data that, as stated earlier, includes time series of the potentially relevant features of the weather data.

i. Feature Selection

In order to identify which of these weather features will be used as an input, a good practice is first to plot all the potential input features and target from the design set against one another as a scatter plot matrix. One could visually assess which input features seem to be correlated with the target, and could act as good predictors, and which features seem highly correlated with each other and thus explain one another and perhaps one could be eliminated. Then, it is recommended to use statistical methods to quantify this correlation between different variables and the target and between the variables themselves such as Spearman method. Spearman method is a non-parametric test that is used to measure the correlation between two variables without enforcing any assumption on the distribution of the variables. Once this correlation is quantified by the rho parameter and is tested for significance, the MLP model input features will be selected to include the relevant weather time series along with a lagged time series for AC or heater OC.

ii. Data Transformation and Scaling

In order to better understand each variable and the target, several visuals could be analyzed such as line plots, histograms, and density plots. Line plots allow to visualize how each variable is changing over time and to check if there is any obvious trend or seasonality. Histograms and density plots allow to visualise the distribution each variable is most likely to be following. For example, if the histogram of a certain variable shows a heavy tail, it might be difficult for the ML model to detect trends. Thus, it is recommended to transform the data into a distribution to is closer to a bell-shaped one (Géron, 2019). Examples of data transformations include logarithmic, square root and cube root transformations (Hyndman & Athanasopoulos, 2018). In addition to that, often the variables will have different ranges that could be very far from one another. ML algorithms also often do not perform well when the scales of the input variables vary widely. As such, it is recommended to transform the input variables (and sometimes even the target) into the same scale. This can be done by normalizing the variables into the range of [0,1] whereby the minimum value of the instances of a certain variable is subtracted from the value of interest and then divided by the difference of the maximum and minimum. Another method for feature scaling is through standardization, whereby the mean of the instances of the variable is subtracted from the value of interest and is then divided by the variance of the instances of the variable. So, the instances of the variable will be standardized to have a mean of zero and a unit variance but will not necessarily be bound to a certain range which might still be problematic for certain ML algorithms (Géron, 2019).

iii. Splitting Design Data into Training and Validation

Now that the target times series (AC or heater) is smooth, and both input and target times are transformed and scaled as needed, developing the MLP model could start. But prior to that, it is important to divide the design dataset into two parts. The first is the training set that is used to fit the model to identify its different coefficients. The second part is validation set that will be used assess to the model and to evaluate the performance of different scenarios by tuning the different hyperparameters that will be identified in the upcoming section. But this evaluation becomes biased since the validation dataset is now part of the model's configuration. For that reason, a testing set was held out earlier to evaluate the final model and provide a test error (James et al., 2013). The need to split the data is crucial since it allows to evaluate the model prior to testing it on the holdout set thus overcoming both underfitting and overfitting potential problems. Overfitting happens when the model has a high complexity where it performs well on the training data set but does not have the capacity to perform well on data points outside the this set. Underfitting on the other hand occurs when the model has a low complexity where it does not learn the training data well in the first place (Goodfellow, Bengio, & Courville, 2016).

f. MLP Model Training, Validation, and Testing

Two machine learning-based prediction models will be built: one for predicting the number of AC complaints (AC model) and one for predicting the number heater complaints (heater model). Building both MLP models will follow a similar procedure using the Keras library in Python which is an open source for developing neural networks.

i. Number of Lags

The number of lags can be defined as the number of previous time steps that will be used in the prediction. This will then define the number of input features that will be used as predictors in the model including previous values of the input features and previous values of the target since the problem at hand is an autoregressive one with exogenous predictors. For example, suppose the MLP model has 5 input features and 1 target to be predicted. As per the tabular structure of the data, suppose a number of 2 lags was selected, this means that the features of instances 1 and 2 will be used to predict the target at instance 3. The model will now require 12 input features; the 1st 5 represent the input features of instance 1, feature number 6 will represent the target of instance 1, the next 5 represent the input features of instance 2, feature number 12 represents the target of instance 2. It should be noted that since a lag of 2 was selected, the first 2 instances will be used for initialization, and the first prediction will start at instance 3. Thus, restructuring the dataset as such allows to transform this time series forecasting problem into a typical supervised ML problem. Selecting the number of lags is crucial to the problem, because the higher the number of lags the higher is the total number of predictors used to fit the model, and the more is the historical data required to make each prediction. So, the number of lags can be increased starting with a lag of one until the model starts overfitting.

ii. Network Architecture and Activation Function

Now that the dataset is structured properly, the MLP network architecture can be defined. This includes the number of hidden layers and the number of neurons in each layer. Those are referred to as hyperparameters since they are variables that the training algorithm does not learn by itself, rather they are selected by the user prior to the learning process (Bergstra, Yamins, & Cox, 2013).

A typical MLP structure consists of a minimum of three layers: an input layer, at least one hidden layer, and an output layer as shown in figure 2. X_j represents the MLP's input vector including both the target time series (AC or heater) and the time series of the exogenous features (weather features) of the previous time steps with j being the index of the input, a_i represents the vector of neurons used with i being the index of the neuron in the corresponding layer, f represents the activation function, and $\hat{y}\,(t)$ represents the predicted times series (AC or heater). Figure 3 shows a more detailed structure of a single neuron In each layer, every neuron sums the elements of the input vector X_j after weighting each by its respective weight given by the weight vector w_{ij} which represents the weights vector needed to carry out the mapping. It then performs an activation function to the obtained sum as stated by to obtain an output a_i (Behrang et al., 2010).

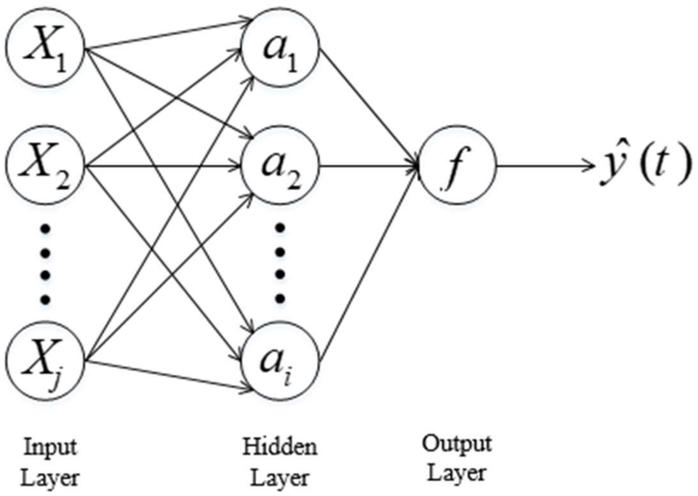

Figure 2 ANN structure

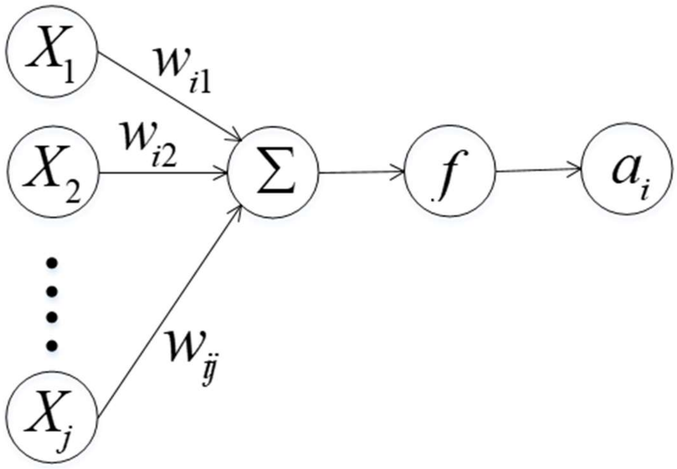

Figure 3 Structure of a neuron

According to Géron (2019) often using one hidden layer in MLPs is enough to provide reasonable results given that it has a sufficient number of neurons. However, using a deeper network (one with number of hidden layers greater than two), could exponentially decrease the number of neurons in each hidden layer as compared to a shallower network. Decreasing the number of neurons in the network makes the training less computationally expensive. So, using one or two hidden layers could be a good starting point when fitting the MLP model. The number of hidden layers should be selected prior to the training process. It can be increased until the model starts overfitting the training dataset (Géron, 2019).

As for the neurons, their number in the input layer is defined by the number of input variables, one for each. In the output layer, their number depends on the type of output the model is trying to predict for example, one neuron is used if the model is trying to predict one target for a regression problem. Whereas in the hidden layers, a common practice is to select their number in a funnel form; to use a smaller number of neurons in each layer. However, this increases the number of hyperparameters one needs to be selecting and tuning, so a more common practice is to use the same number of neurons in each hidden layer thus having one hyperparameter to select instead of one for each layer. As with the number of hidden layers, the number of neurons should be selected prior to the training process and could be increased until the model starts overfitting the training dataset (Géron, 2019).

Several types of activation functions $f\ (X)$ presented below can be used in the mapping process where X represents the output from the previous layer. The most common types are the linear (equation 1), sigmoid (equation 2), hyperbolic tangent (equation 3) and the rectified linear unit (ReLU) activation functions (equation 4) (Nwankpa, Ijomah, Gachagan, & Marshall, 2018).

$$f\ (X) = X \qquad (1)$$

$$f(X) = \frac{1}{1+e^{-X}} \qquad (2)$$

$$f(X) = \frac{e^{X}-e^{-X}}{e^{X}+e^{-X}} \qquad (3)$$

$$f(X) = max\ (0, X) = \begin{cases} X,\ if\ X\ \geq\ 0 \\ 0,\ if\ X\ < 0 \end{cases} \qquad (4)$$

The ReLU activation function, or one of its variants, are often used in the hidden layers of a neural network since they require less computational effort and offer a better generalization performance as compared to the sigmoid and hyperbolic tangent functions (Géron, 2019; Nwankpa et al., 2018). As for the output layer, the linear activation function can be used since the problem at hand is a regression problem (Géron, 2019).

iii. Optimization Algorithm

The learning process of the MLP model is carried out by a training algorithm as an optimization problem to determine the weights and bias of the model with the objective of minimizing the model error, also known as a cost function, defined by a performance function. The weights and bias are tuned in a way that permits the model to produce an output that is as close as possible to the actual value (James et al., 2013). Common training algorithms that solve non-linear optimization problems include: gradient descent and its variants, Gauss-Newton (Battiti, 1992), Levenberg-Marquardt

(Gavin, 2011), Momentum optimization, RMSprop , and Adaptive Moment Estimation (Adam) (Géron, 2019).

For a long time, gradient descent algorithm was the most commonly used optimization algorithm in neural networks. This optimization is done by updating the model weights and bias after every iteration throughout the training process. The direction of this update should be opposite to that of the gradient of the cost function with respect to the weights. As for the size of the step of this update, it is determined by the learning rate that is a hyperparameter of the gradient descent that should be configured (Ruder, 2016). If a very high learning rate is used, the training might diverge or might never converge where it keeps going around the minimum. Whereas if a very low rate is used, the training will eventually converge but will take a lot of time. So, selecting the learning rate is a very critical process to ensure convergence and a reasonable computational time (Géron, 2019). Current practices recommend using Adam optimizer to train an MLP model. Adam emerged from two other optimizers: Momentum optimization and RMSprop. Adam has the property of accelerating the optimization problem towards the minimum leading to faster convergence as compared to the traditional gradient descent. This requires keeping track of the average gradients of past iterations in an exponentially decaying manner. As such, a new hyperparameter known as "momentum decay hyperparameter β_1" should be configured. β_1 ranges from 0 to 1 with a value of 0 reflecting a high friction, and a value of 1 reflecting no friction. Often, a value of 0.9 is used. Also, another new hyperparameter known as "scaling decay hyperparameter β_2" ranging from 0 to 1 should be configured as well. β_2 is usually initialized to 0.999. Other hyperparameters of Adam include a smoothing term ϵ that mitigates division by zero with a typical value of 10^{-9}, in addition to the learning rate η that is usually set to 0.001. Often those default parameters provide a good performance (Géron, 2019). The main advantage of Adam over Gradient Descent algorithm, is that it is an adaptive learning rate algorithm, meaning that the learning rate is updated throughout the learning process in a way that allows it first to learn quickly and then to converge slowly towards a good solution, and this requires less tuning of the learning rate (Géron, 2019).

Two more hyperparameters should also be configured when training the MLP model, the number of epoch and the batch size. The batch size represents the number of samples from the training dataset that are passed to the model at once during the training process. The model then is evaluated at the end of each epoch. As such, one epoch represents one complete pass over all the samples in the training dataset. The number of epochs is usually high for example 500, 1000 or even higher (Géron, 2019).

iv. Performance Function

Several statistical error measurements can often be used when evaluating the model's performance. The Mean Absolute Error (MAE) for example measures the average absolute error e which is defined as the difference between the predicted and the actual outputs. The Mean Squared Error (MSE) is similar but the values are squared to obtain positive values (Cadenas, Rivera, Campos-amezcua, & Heard, 2016) Equations 5 and 6 represent the MAE and the MSE respectively.

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |\hat{y}_i - y_i| \quad (5)$$

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \quad (6)$$

Where, n is the number of instances in the data set, y_i is the actual value, and $\hat{y}_i$ is the predicted value.

v. Regularization

As mentioned earlier, one very common problem of neural networks is overfitting. Regularization is a method that can be used to mitigate this problem by constraining the model thus decreasing its complexity (Géron, 2019). Large model weights are not recommended since they are a reflection that the model has a high complexity and had learned the train set so well and will most likely not perform as good on data it had not seen before (overfit). Moreover, with larger weights, any small change in the input features will have a much larger impact on the output (Smithing, 1999). As such, regularization can be applied during the learning process of the model and is controlled by a hyperparameter of the training algorithm that should be set prior to the training process. Two common regularization methods include: Ridge and Lasso. They both add a regularization term (ranging from 0 to 1) to the cost function that the training algorithm is trying to minimize. Ridge regularization forces the learning algorithm to produce a good model fit while keeping the weights as small as possible. On the other hand, Lasso regularization completely eliminates weights of features throughout the training that are not of much importance thus automatically performing some kind of feature selection. Thus, the method and the corresponding hyperparameters should be carefully chosen and a combination of both could be used (Géron, 2019) . Although the weights are calculated per layer of a MLP, it is common practice to select the same regularization hyperparameter (s) for all layers in order to avoid having more hyperparameters to tune (Goodfellow et al., 2016).

vi. MLP Model Fitting and Evaluation

Now that the model inputs are prepared and structured, the number of lags is selected, the hyperparameters are selected (including: number of hidden layers, number of neurons, batch size, number of epoch,), activation functions, optimization algorithm and its corresponding hyperparameters, performance function, in addition to regularization method (s) and corresponding hyperparameter (s) that are selected as well, the model weights are initialized and the model can now be trained.

A model is said to perform well on train set if it has a low error as per the defined performance function and as per the context of the problem, along with a good model fit (R^2). If this is not the case, steps in section "MLP Model Training, Validation, and Testing" be repeated while changing certain hyperparameters and properties, and the model is fit again. After experimenting with different possible options to train the model, and reaching an option that has a good performance on the training set (low train error, and high R^2), the model should be tested on the validation set to ensure proper selection of the hyperparameters and that the model is unlikely to overfit new data. As such, walk forward validation can be adopted.

Walk forward validation is a suggested method to cross- validate the MLP model that is specific for time series data. A certain number of data samples should first be selected to train the model (train set). Then, the fit model will make a one-step prediction on the validation set. This prediction is evaluated against its actual value, and the train set used to fit the model is then expanded to include this true observation. This process repeats throughout the number of samples in the validation set. This allows the model to make predictions using the most recent data (Hyndman & Athanasopoulos, 2018). Simple models are often refit every time a true observation of the validation set is added to the train set; however, this becomes computationally expensive for neural networks.

The figure below adapted from Hyndman & Athanasopoulos (2018) illustrates the walk forward validation for one step ahead prediction. The blue dots represent the samples used to fit the model in the train set, and the red ones show samples in the validation set. The model's forecasting accuracy is then computed by averaging the errors in the validation set for each one step prediction. Each row represents one iteration throughout the validation set (Hyndman & Athanasopoulos, 2018).

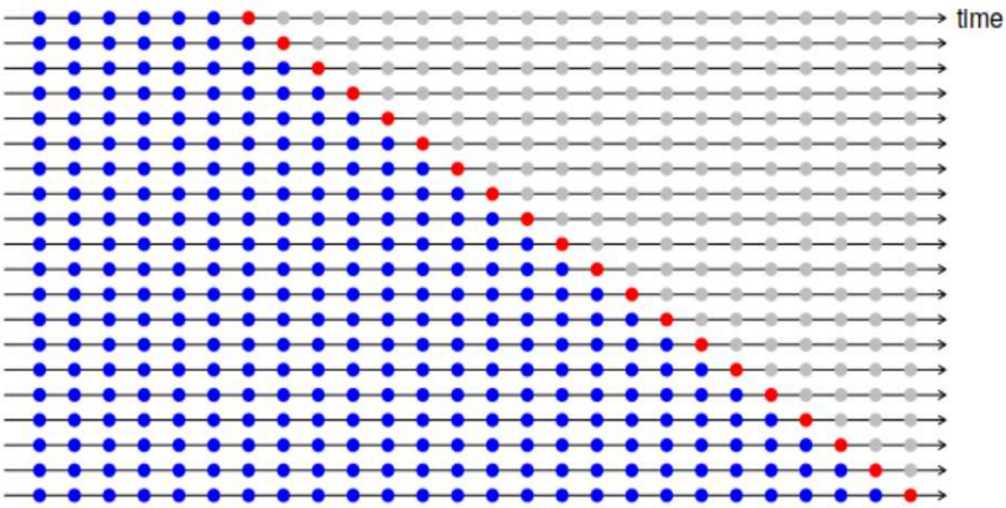

Figure 4 Walk forward validation method (Hyndman & Athanasopoulos, 2018, p.84)

If the fit model does not show good performance on the validation set, steps in section " MLP Model Training, Validation, and Testing". should be repeated while changing certain hyperparameters and properties where the mode is fit again. Whereas if the model shows good performance on the validation set, it will be evaluated by conducting 30 model fit runs while keeping everything the same and the only things changing are the initial model weights due to the stochastic nature of the optimization algorithm. This is to ensure that the model is stable which can be evaluated by obtaining the average error and the corresponding standard deviation for both training and validation sets for those 30 model fit runs. If the average train error and the average validation error are low, are close to each other (model is unlikely to overfit), and the standard deviation of each is low as compared to the corresponding average of the error, this means that the model is stable and the model can now be evaluated on the holdout test set. If this is not the case, steps in section "MLP Model Training, Validation, and Testing" . should be repeated while changing certain hyperparameters and properties where the mode is fit again. In order to evaluate the model on the holdout test set, the features in the test set should first be prepared following the exact process used to prepare features in the design set. As such, the OC data in the test set should be smoothened, and features should be scaled and transformed in equivalence to how the corresponding feature was treated on the design set**.** If the model does not show good performance on the test set, section "MLP Model Training, Validation, and Testing". should be repeated while changing certain hyperparameters and properties where the mode is fit again. Whereas if the model shows good performance on the test set, it will be selected and saved as a suggested model. A good performance on the test set means that the error on the test set is low as per the context of the problem at hand, and the test error is very close or slightly higher than the train error to ensure that the model is not overfitting. It should also be noted that every time the error and the R^2 are calculated at any evaluation level, It is important to inverse any kind of scaling and then to inverse any kind of transformation of each predicted value so that it is comparable to the actual value allowing to obtain the errors and R^2 that are in the same range as the actual context of the problem

3. Benchmark Model: ARIMA

The state of the art ARIMA model for time series forecasting will be developed for each of the AC and the heater related number of complaints to act as benchmark for the developed corresponding MLP models. This includes time series preparation, ARIMA input data preparation, splitting the data into training and validation, and modelling the ARIMA process as described in the sections below.

a. Time Series Preparation

The ARIMA benchmark model will follow the same steps as the ML-generic framework for data collection and preprocessing, text cleaning and mining, data split, and time series smoothing.

b. ARIMA Input Data Preparation

As with the MLP model, the ARIMA model for each the AC or heater models will have two types of inputs. The first is the AC or heater time series obtained after smoothing the corresponding time series obtained from the text mining process and known as the endogenous series. The second corresponds to additional exogeneous data that, as stated earlier, and includes time series of the relevant potential features of the weather data

i. Feature Selection

As with the MLP model, the weather features that will be selected as potential exogenous variables will be based on visualizing the scatter plot matrix of all these variables and the target (endogenous series) and based on Spearman correlations. This allows to determine which exogenous input features seem correlated with the target, and could act as good predictors, and which exogeneous features seem highly correlated with each other and thus explain one another and perhaps one could be eliminated

ii. Data Transformation and Scaling

As with the MLP, several visualizations could be used to better understand each exogeneous variable and the endogenous time series.

As mentioned earlier, ARIMA model requires the endogenous series to be stationary whereby the characteristics do not change over time. This can first be tested by the statistical Augmented Dicky-Fuller (ADF) test where the null hypothesis states that the series has a unit root and thus is not stationary, and the alternative hypothesis states the series does not have a unit root and thus is stationary. The null hypothesis can be rejected or not based on the selected significance level (Goh & Law, 2002). If the data is not stationary, several transformations could be used to transform the series into a stationary one. This is done by removing the trend and/or seasonality of the time series which in turn requires transformation such as differencing or log transformation (Hyndman & Athanasopoulos, 2018). Certain power transformations can assist in stabilizing the variance of the endogenous time series, while differencing can assist in stabilizing the mean of the time series (Hyndman & Athanasopoulos, 2018).

Differencing can be used to remove trend by replacing each value of the endogenous series with this value subtracted from it the value of the previous time step, while differencing to remove seasonality can be done by subtracting from the previous season. If needed, differencing can be done more than once.

Data scaling methods mentioned in section B.e.2 could also be adopted here if the range of the endogenous and exogenous series differ drastically.

c. Splitting Design Data into Training and Validation

Splitting the design data into training and validation will follow the same procedure as with the MLP model to avoid both underfitting and overfitting.

d. ARIMA (p, d, q) (P, D, Q) s modelling

The methodology followed to develop and diagnose an ARIMA model is based on that suggested by Box and Jenkins (Box, Jenkins, & Reinsel, 2011).

The first step to develop an ARIMA model is known as "identification" and aims to select the model parameters by analyzing the design data set. The Autocorrelation Function (ACF) represents correlations between residuals and thus can be used to determine an estimate of the "q" parameter representing the order of the moving average. The Partial Autocorrelation Function (PACF) represents the correlations between the lagged values of the endogenous time series without including intermediate lags, and thus can be used to determine an estimate of the "p" parameter representing the order of autoregression. The "p" and "q" parameters are selected in a way that no significant correlation is found after the selected number of lags for each of the PACF and ACF plots respectively based on the selected confidence level. As for the "d" parameter, it refers to the degree of integration required to make the time series stationary before fitting the ARIMA model (Box et al., 2011). The equation of the ARIMA model then becomes:

$$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + \ldots + \beta_p y_{t-p} + \epsilon t + \Theta_1 \epsilon_{t-1} + \Theta_2 \epsilon_{t-2} + \ldots + \Theta_q \epsilon_{t-q}$$

(equation 7)

where, y_t represents the value of the endogenous series at time t, $\beta_1 y_{t-1} + \beta_2 y_{t-2} + \ldots + \beta_p y_{t-p}$ represents the autoregressive part which includes the previous p lagged values of the endogenous time series, and $\epsilon t + \Theta_1 \epsilon_{t-1} + \Theta_2 \epsilon_{t-2} + \ldots + \Theta_q \epsilon_{t-q}$ represents the moving average part which includes the error terms at time t and past times till $t - q$. p and q as mentioned earlier represent the autoregression and the moving average orders respectively. As for β_1 to β_p, they represent the autoregression (AR) coefficients, and Θ_1 to Θ_q represent the moving average (MA) coefficients.

There is also a tendency for the time series to vary periodically if it is long enough, this is known as seasonal variation, and should be considered in the modeling process (Tabachnick et al., 2007). As such, the ARIMA model is now referred to as seasonal ARIMA, and three additional parameters should be identified "P", "D", and "Q" following a similar process to "p", "d", and "q" identification but now to model the seasonal variation instead of the trend of the time series. As for the parameter "s", it represents the period of the time series For example, "s" is 12 for a monthly recorded time series that varies yearly, and "s" is four for a quarterly recorded time series that varies yearly (Box et al., 2011). The seasonal ARIMA model then ensures that the variations between consecutive observations and between consecutive observations of consecutive seasons are both modeled (Hamzaçebi, 2008). When there are additional exogenous features that are to be added in the forecasting process, the ARIMA model is referred to as seasonal ARIMAX.

Once tentative parameters are selected, the second step of "estimation" can proceed. The model coefficients are estimated using the train set provided that they minimize the model errors (Box et al., 2011). A model is said to perform well on train set if it has a low error as per the defined performance function and as per the context of the problem, along with a good model fit (R^2). The AIC (Akaike Information Criterion) is another measure that can be used for model selection upon fitting different ARIMA models. It accounts for both the model fit and the simplicity for the model. The lower the AIC values the better the model is (Babu & Reddy, 2014). It should be noted that the AIC measure should only be used to compare hierarchical ARIMA models (Tabachnick et al., 2007). At this stage, the model coefficients including those – as applicable- for autoregression, moving average, and exogeneous features should be statistically significant from zero as per the selected significance level. If not significant, the corresponding variable (s) are to be removed one at a time (if more than one is insignificant), another tentative model is selected, and the coefficients are estimated again (Tabachnick et al., 2007). The model is then tested on the validation set, using the walk forward validation method discussed in section B. 2. f. vi, to ensure it is not overfitting.

By now, the coefficients of the selected model parameters are significant, and the model does not seem prone to overfitting. The third step of "diagnosis" can proceed. This step verifies that the assumptions of the residuals have been met ensuring that the model has captured all the information that could be modeled (Box et al., 2011). The residuals of a good model should not be correlated otherwise there is some information that was not modeled, and this can be identified by plotting the corresponding ACF plot. If this is not the case, the model parameters should be identified again as in step one. They also should have a mean of zero otherwise they are considered biased, but this problem can be solved by adding the mean of the residuals to every forecast. Moreover, the residuals should have a constant variance and should be normally distributed (Hyndman & Athanasopoulos, 2018). Once a good model is selected with proper residual diagnostics, it will be tested on the holdout set in a walk forward manner. If the test results are not satisfactory, the selected model parameters could be updated, and the steps repeat again. The obtained generalization error is used to compare this model to the MLP model. The ARIMA model procedure could be summarized by figure 5.

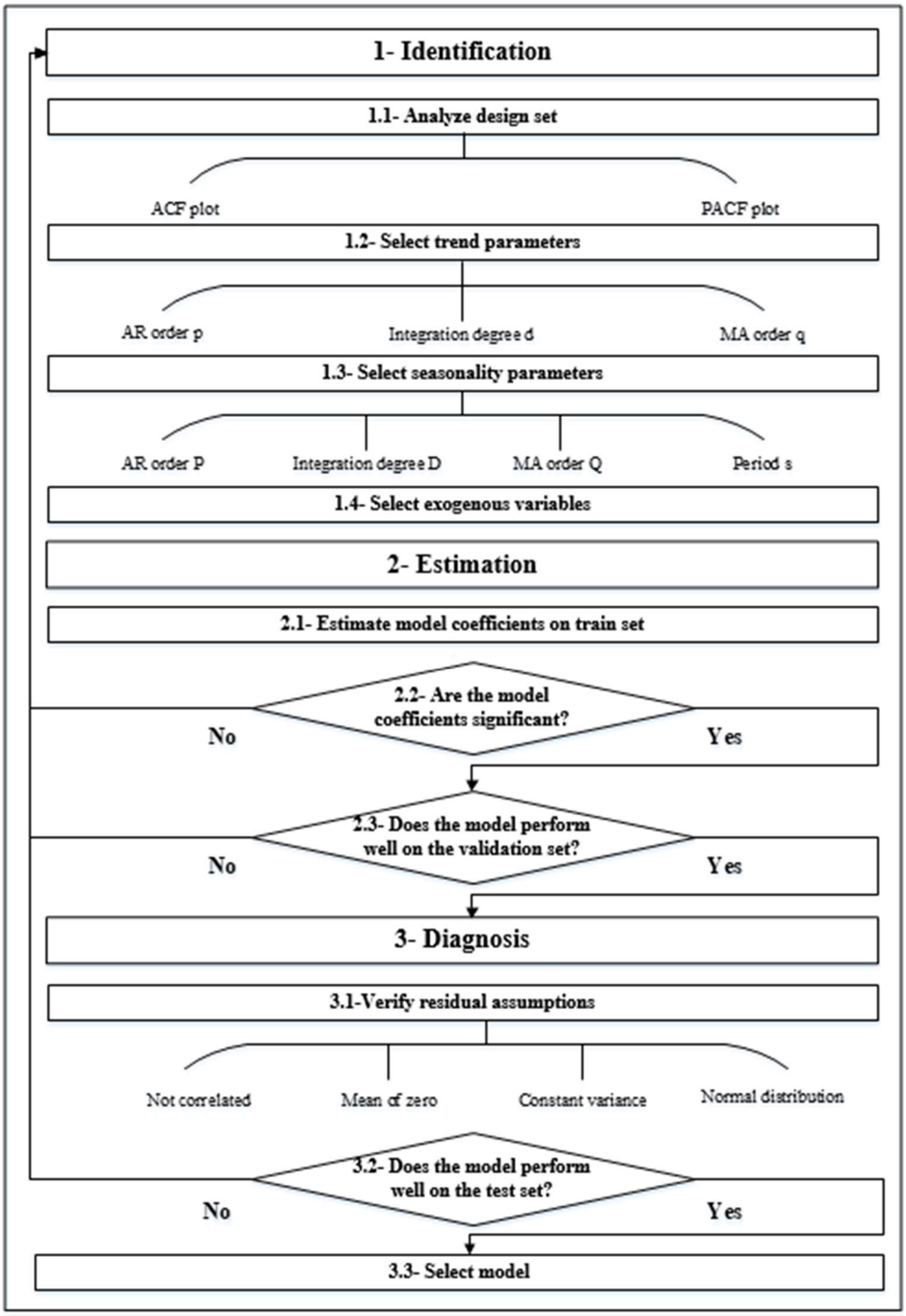

Figure 5 ARIMA methodology

CHAPTER 4

RESULTS AND DISCUSSION: CASE STUDY

This chapter provides a detailed analysis of the results obtained upon implementing the proposed methodology on a selected case study. Section A tests and evaluates the application of the developed ML-based generic framework to forecast thermal complaints, section B provides a detailed analysis of the developed ARIMA benchmark models, section C compares and evaluates both the ML-based models and the benchmark models, and section D presents a staffing application of the developed ML-based models. The need for proper data management is addressed in section E.

A. An application of the Developed ML-Based Generic Framework

The developed generic framework to forecast thermal complaints was tested on an occupant complaint dataset obtained from a facility management unit of a residential complex in Beirut, Lebanon for a period of three years to forecast the number of AC and heater related complaints in the upcoming week.

1. Data Collection and Preprocessing

The residential complex comprises of 16 new buildings with a total of 145 apartments. It is mainly served by two separate centralized systems for heating and air conditioning. The air conditioning system comprises of air cool chillers. The water side of the system utilizes primary and secondary chilled water pumps that supply chilled water to terminal fan coil units at each apartment. Similarly, the heating system includes two boilers. The same pumping strategy is used to serve radiators for space heating as well as hot water bottles for domestic water heating (during the winter season).

The maintenance related data was obtained in Microsoft Excel documents comprised of three different files: one for each of the three years. Each of these Excel files had 12 sheets: one for each month. The data for each year included daily routine maintenance records, Building Management System (BMS) notes, and recorded occupant complaint calls. The data was recorded on a daily basis in each sheet, where a row might represent an occupant complaint, a maintenance record, or a BMS note. As for the columns, they included for each observation: the date, the building's name, apartment's number, name of the operator who received the complaint call if applicable, time of the call if applicable, time of the response to the call if applicable, name of the technician put in charge to handle the complaint if applicable, action taken by the technician if applicable, and a thorough description in text format of the complaint issued by the occupant if applicable. After a thorough preliminary screening and investigation of the maintenance related data for the three years the following conclusions can be drawn:

- The date of each observation is important, because as stated earlier this is a time series forecasting problem, so time is an important factor.
- Considering that the observations were recorded on a daily basis, this gives flexibility to study the data with a time step of days, weeks, month or years.
- Although the name of the building and the apartment number could assist to better understand where the complaints are coming from, they are out of the scope of this study and thus will be excluded.

- The name of the operator who received the call is of no relevance to this study and thus will be excluded.
- The time of the call and the time of the response to the call were only recorded for few observations, mostly for those where the time of the call and the time of the response are very close in time. Such calls are likely to be urgent matters, or during which a technician was readily available at the time. This shows a lack of follow-up on updating this log when the complaint or maintenance requests were addressed. So, these two columns had missing data for almost all observations and thus will be excluded.
- The name of the technician in charge and the action taken by the technician were not filled for most of the observation so, they were not included in any further analysis.
- The description of the complaint issued by the occupant was of great use to this study. It was the major source of information on the type and number of complaints issued, and sometimes even to a more detailed description on what the problem exactly is. Yet, this was written in a text format which made it challenging to extract further information: perhaps there is some information stated explicitly in the text but cannot be easily spotted or searched for. As such, this text data could be transformed into a more structured format using the process of text mining which will be employed in the next section.

Considering that this study focuses on the occupants' complaint calls and that several columns encountered missing data points for several observations, the columns that will be used in this study include the date and the description of the occupant complaint. The latter was first filtered to ensure it only includes description of occupant complaints and no BMS notes or routine maintenance follow ups. The complaint descriptions were further investigated, and it was deduced that a good number of those - especially in year three- was recorded in the Arabic language while the primary language of the documents was English. As such, those were scanned and translated manually to English. While translating, it was critical to ensure as much as possible that the same terms in English are used consistently, for example being consistent in using the term "AC" instead of "air conditioning", and for that purpose, this task was conducted by one person only. However, a term could always be replaced by another if they refer to the same meaning such as "ac" and "air conditioner". After cleaning the data, a total of 6, 577 complaints were analyzed with 2, 692 in the first year, 1, 924 in the second year, and 1, 961 in the third year.

The chart below in figure 6 provides a high-level data dictionary of the obtained data set including which columns of the Excel files were used in this study and which were not.

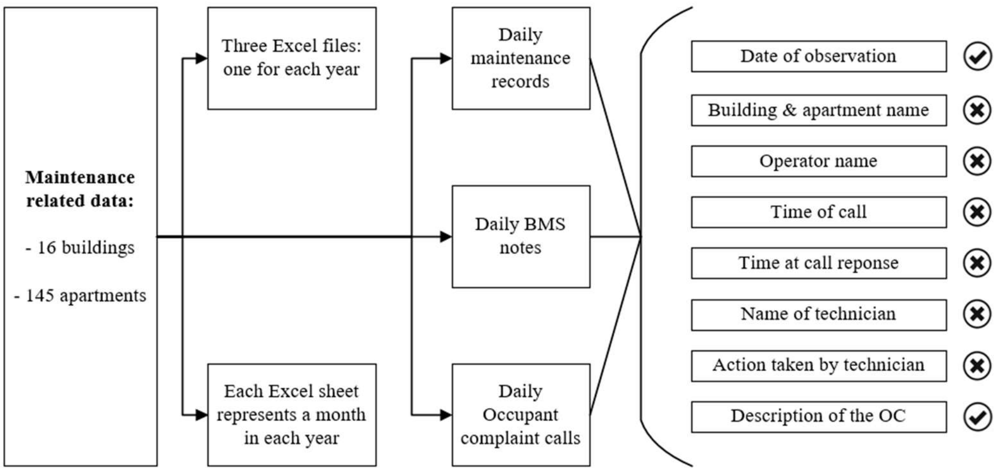

Figure 6 Maintenance data high-level dictionary

A time step of weeks was selected since staffing decisions are typically made on a weekly basis which is the ultimate application of the proposed methodology. So, a total of 156 weeks, 52 in each year, with an average of 42 complaints was obtained per week. As such, the occupant complaint descriptions were exported from Excel into Notepad files to be used in the next section in the text mining process. So, a total of 156 Notepad files was obtained.

As for the weather-related data, they were also obtained for the corresponding three years, and was clean without any missing data. The data was recorded on a minute by minute basis which provides some flexibility for the time step to be selected. It included several features: temperature, relative humidity, mean sea level pressure, total precipitation, snowfall amount, cloud dover, in addition to wind speed and direction. The temperature and the relative humidity seem to be good predictors of thermal complaints because they direct the use of air conditioning and heating systems which in turn drives the corresponding number of complaints. Moreover, the wind speed was added to ensure the real-feel temperature was included in the prediction.

Since a time-step of weeks was selected, the weather observations recorded on a minute basis were aggregated to obtain the daily average which can be used to obtain three features for each of the selected variables based on the summary statistics varying on a weekly basis: minimum, average, and maximum to ensure that those variables are fully represented. For example, the observations for week one for the temperature features are obtained as follows: the minimum temperature feature is obtained by taking the minimum temperature of the first seven days, the average temperature variable is obtained by averaging the temperature of the first seven days, and the maximum temperature is obtained by taking the maximum temperature of the first seven days. The same process was applied for the following weeks till week 156, and for the relative humidity and wind speed variables. As such, a total of nine features is obtained and the following abbreviations will be used throughout this thesis: MinT, AvT, MaxT, MinRH, AvRH, MaxRH, MinW, AvW, and MaxW. The temperature features were measured in degrees Celsius, the relative humidity features were measured in percentage, and the wind features were measured in kilometres per hour. Each of these features represents a time series. The corresponding line plots are shown in figure 7, and table 2 shows the summary statistics of each. These time series line plots show obvious repetitive variations from one year to another, in addition to seasonal variations within the same year. In particular, temperature features show low values at the beginning and end of each year since this period represents the winter season, and mid year they show higher values whereby this period represents the summer season. There is no clear seasonal variation for the relative humidity features, but they appear to have a wide range. As for the wind features, the minimum wind feature does not show obvious seasonal variation, but this is clear for the average and maximum wind features that show seasonal variation opposite to that of the temperature features but not as strong.

Table 2 Summary statistics of weather features

	Min T	Av T	Max T	Min RH	Av RH	Max RH	Min W	Av W	Max W
Count	156	156	156	156	156	156	156	156	156
Mean	17.55	20.2	23.37	42.99	59.95	76.01	7.39	11.06	16.06
Standard deviation	6.12	5.58	5.44	13.49	9.83	8.74	1.59	2.86	6.25
Minimum	3.51	7.24	11.74	11.42	26.65	53.79	3.48	6.06	8.04
Maximum	28.08	30.60	33.57	73.46	81.55	92.63	12.68	22.55	42.65

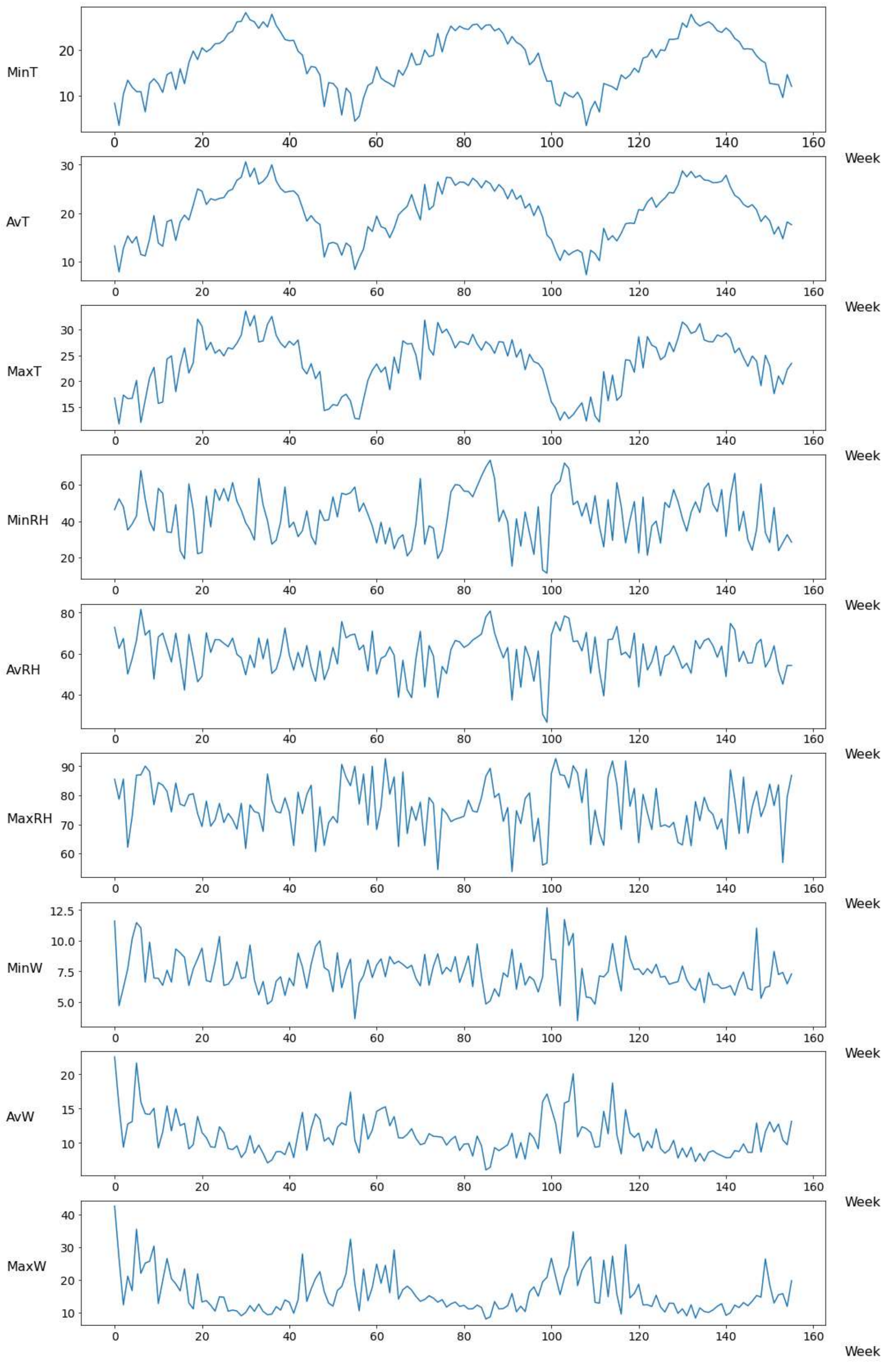

Figure 7 Time series plots for potential weather features

2. Text Cleaning and Mining

The Notepad files containing the OC descriptions for each of the 156 weeks are now imported into the software R. These files are cleaned by transforming all the letters to lower case, stemming the words, replacing punctuations with white spaces, replacing digits with white spaces, replacing the English stop words with white spaces, since non of these are of relevance to the study, and finally deleting all the additional white spaces.

Once cleaned, these Notepad files were used in the text mining process to obtain the TDM. Figure 8 shows an example of an obtained TDM of two Notepad files each for one week with one occupant complaint only

Occupant complaints
W 1: Can you send someone to fix my heater?
W 2: The power went off.

DTM

	fix	heater	off	power	send	someone	went
W 1:	1	1	0	1	1	0	0
W 2:	0	0	1	0	0	1	1

Figure 8 TDM example

The number of occupant complaints, their average number per week, along with the total number of distinct terms, the sum of the total number of terms, and the average number of terms per occupant complaint derived form the TDM are summarized in table 3 for each year. It is noted that on average only four terms were used to describe an occupant complaint, so these terms ought to be useful. However, it might be a sign that the recorded description is not much detailed and might be missing on some important information, but this can be explained by the fact that an occupant is issuing the complaint and not an experienced technician for example. The results also show how sparse a TDM can be: out of more than 1,000 distinct terms in each year, only 4 on average were used to describe an occupant complaint

Table 3 TDM summary per year

	Year 1	Year 2	Year 3
Number of OC	2,692	1,924	1,961
Average number of OC per week	52	37	38
Number of distinct terms	1,389	1,043	1,288
Total number of terms	11,894	7,485	8,066
Average number of terms per OC	4.42	3.89	4.11

Figures 9, 10, and 11 show the word clouds for each year for the top 200 most frequent terms as per the obtained TDM. A word cloud is a visualization of the terms that are most frequently repeated. The size of a word is proportional to its frequency—the larger the font size, the more frequent it appears and vice versa. Figures 12, 13, and 14 show bar charts for the top 20 most frequent terms for each year. It can be inferred that for year one the word with the highest frequency was “water” followed by “ac”. Other words that appeared frequently include: “leak”, “heater”, “alarm”, “edl” (EDL refers to electricity of Lebanon company), and “lamps”. As for year two, the most common term was “edl” followed by “water” and “ac”. For year three, the most common term was “water” followed by “ac” and “problem”.

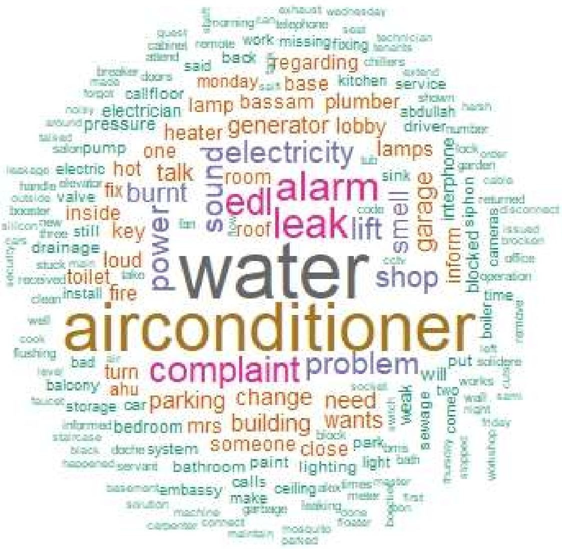

Figure 9 Word cloud for year one

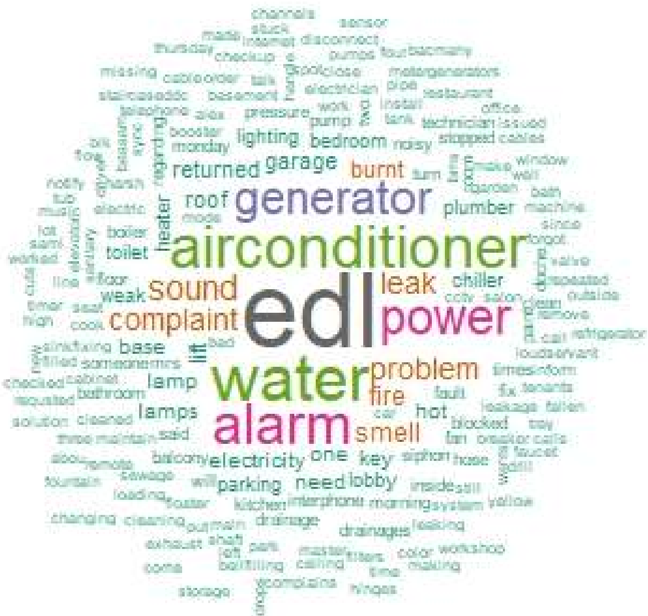

Figure 10 Word cloud for year two

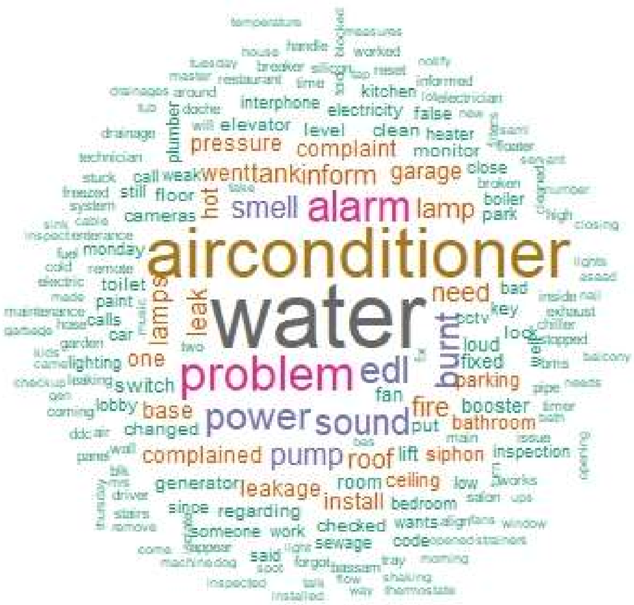

Figure 11 word cloud for year three

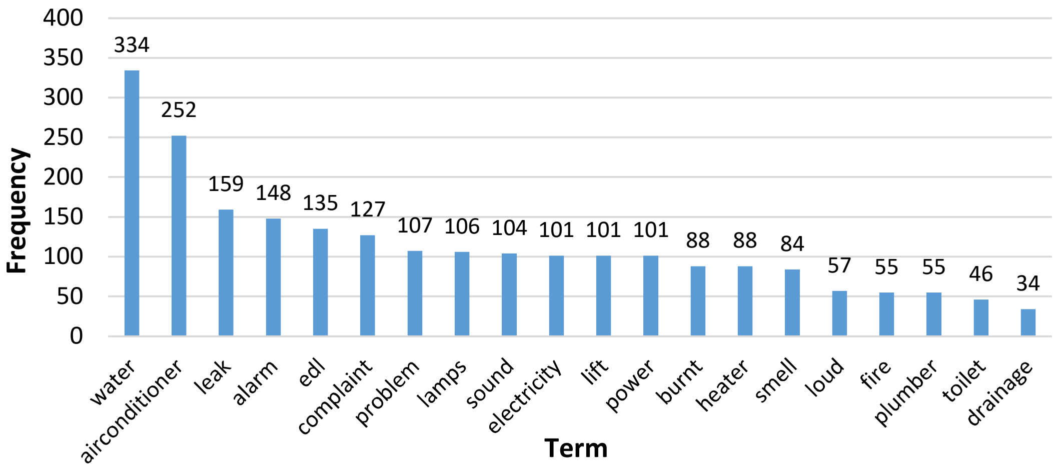

Figure 12 Most frequent terms for year one

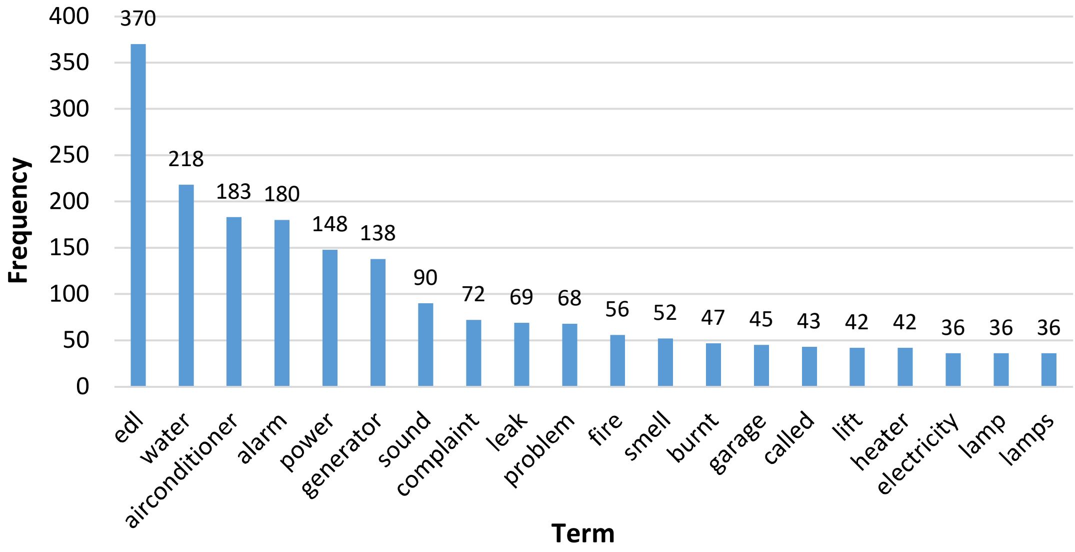

Figure 13 Most frequent terms for year two

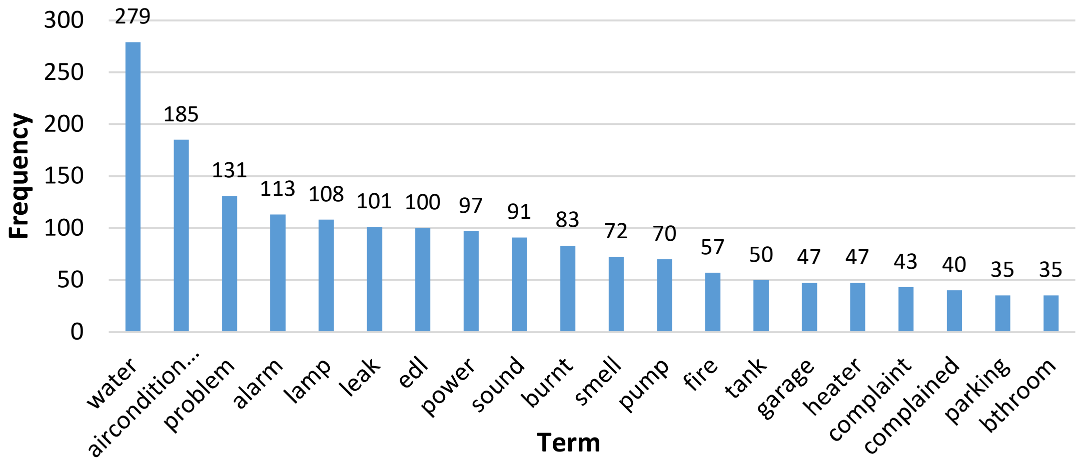

Figure 14 Most frequent terms for year three

Examining the variation of thermal complaints in particular, it can be noted that the number AC related complaints in each year is higher than that of heater related complaints. This is expected since the number of summer months is higher than that of winter months. Another reason is that the maximum weekly temperature for example could go as high as 35 degrees Celsius, and the corresponding feel temperature is expected to be even higher due to the effect for the humidity and wind, whereas the minimum weekly temperature does not go any lower than about 3 degrees Celsius. So, the cooling system seems to be used more than the heating system and even more intensively reflecting the gap between the number of the corresponding complaints obtained. Moreover, it can be noted that the total number of thermal complaints in year one was 340, which showed a great drop in year two to 225 complaints and then very slight increase in year three to 232. This can be explained by upgrading the heating and cooling systems of the complex before the start of year 2. When examining the number of thermal complaints per each month, obtained from the TDM for each year, a certain discrepancy was noted. For some of the winter months for example January, February, and December, the term "ac" appeared several times and this did not make sense since those are winter months, thus the heater would be used not the air conditioning system, and thus it is not expected to receive complaints related the AC. These specific complaints were further examined to come to the root cause of this issue by reviewing the OC description in the Notepad files for the weeks related to those months in every year. This investigation came to the conclusion that the operators seem to use the term "ac" to describe a heater related complaint, because for them the air conditioning system of the building was used in the heating mode, and as call center operators they are not expected to have the technical knowledge about the heating and cooling systems for the buildings. As such, the count of the "ac" terms for the winter months was manually shifted to the term "heater" to describe a heater complaint. Another important point is that for some of the thermal related complaints they were described using the term "hvac", thus those complaints were examined to check if each is AC or heater related complaint, and those were shifted manually. This was doable since the term “hvac” appeared only about 3 times in each year. The bar graphs and word clouds represented above were developed after the numbers were shifted.

Given the obtained number of the “ac” and “heater” terms per week, the time series representing the numbers of these terms versus time in weeks are thus developed and represented respectively by figures 15 and 16 below. The vertical lines show the boundaries between the years. For the AC complaints, they are null at the beginning and end of each year, and peak mid year during summer months. Whereas the heater complaints are null mid year and peak at the beginning or end of each year. Such variation is repetitive from one year to another. It is clear as well that the months without heater complaints (summer season) are less than the months without AC complaints (winter season). Moreover, the average number of AC complaints per week is much more than that of heater complaints. The reasoning behind the latter two points was discussed earlier in this section.

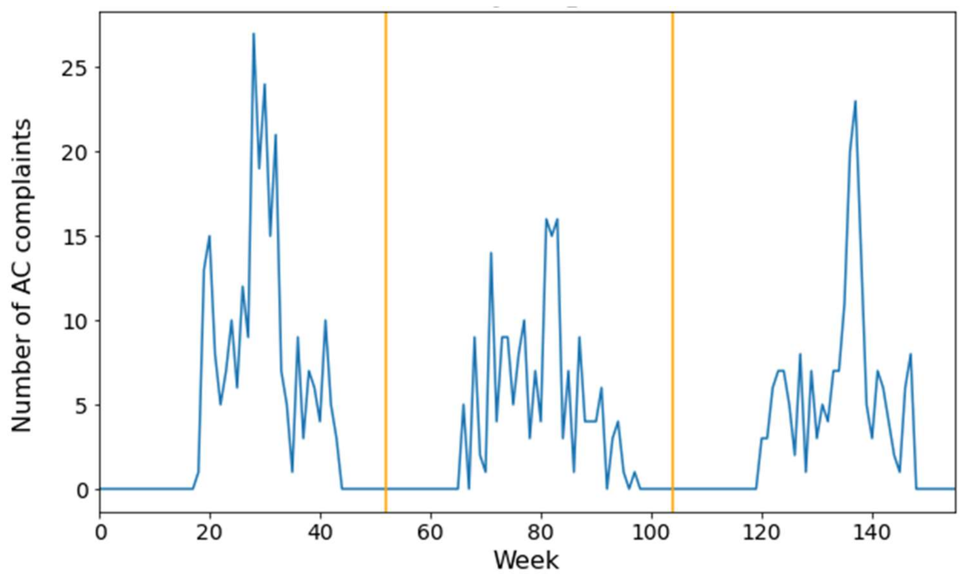

Figure 15 Developed AC time series

The summary statistics for the AC and heater time series are shown in the table 4. It can be noted that each has a very wide range, and a low mean compared to the range.

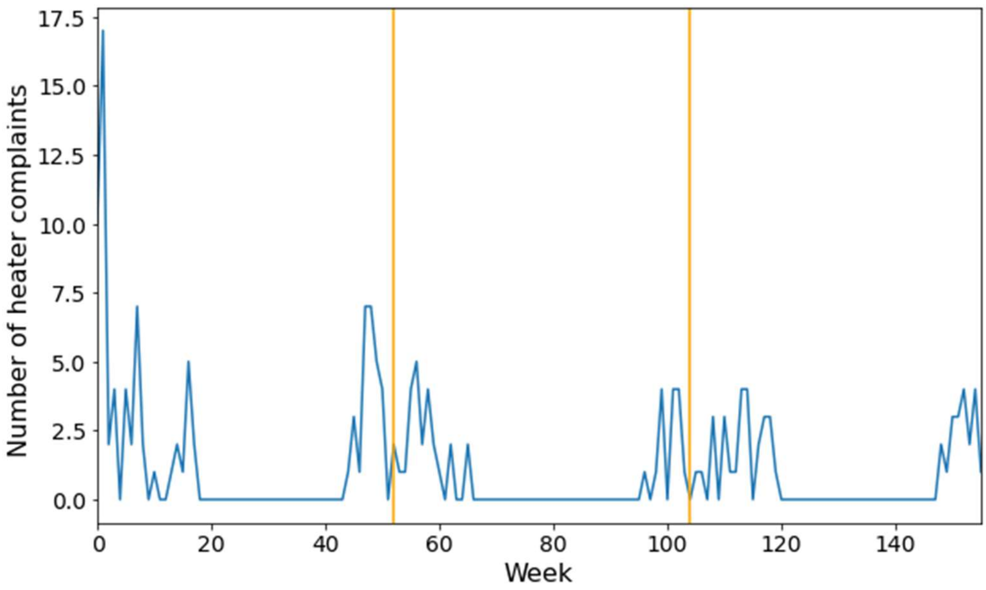

Figure 16 Developed heater time series

Table 4 AC and heater time series summary statistics4

	AC complaints	Heater complaints
Count	156	156
Mean	3.981	1.128
Standard deviation	5.576	2.181
Minimum	0.000	0.000
Maximum	27.000	17.000

3. Data Split

At this stage, there is a total of 156 instances, a total of nine potential weather features, and the target of what the corresponding model is trying to predict (AC or heater complaints). This data set will be divided into two parts sequentially. For each of the AC and heater models, the first 90% of the data (140 instances) –"design set"- will used to design each MLP model, and the following 10% (16 instances) will be used to test the model to obtain the generalization error- "holdout set". The 10% holdout set might seem small; however, the data split was conducted in an iterative way for this data split step and upcoming steps of data smoothing and further division of the design set. This was to ensure that each sub-part of the data is representative -as much as possible- of the whole data set while respecting the sequential variation of the data. Figures 17 and 18 below is a visualization of the data split into "design set" and "holdout set" for each the AC and heater time series, respectively. All other variables will be split the same way.

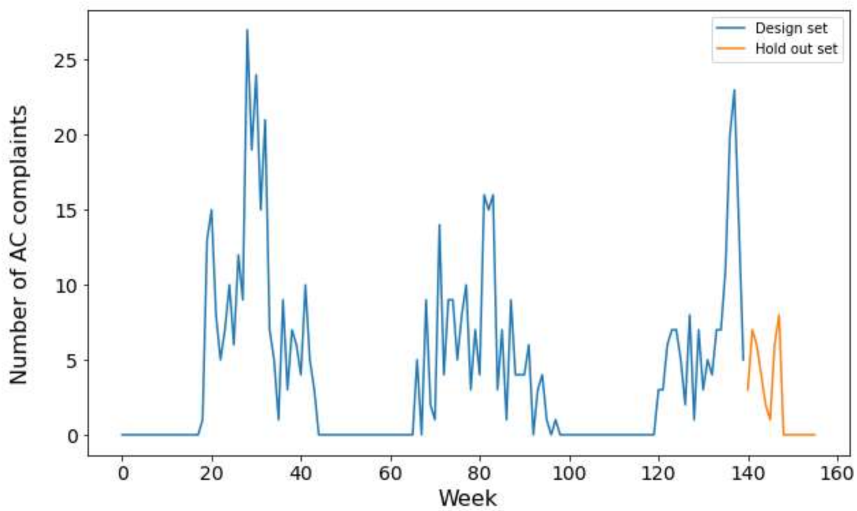

Figure 17 AC time series split between design and holdout sets

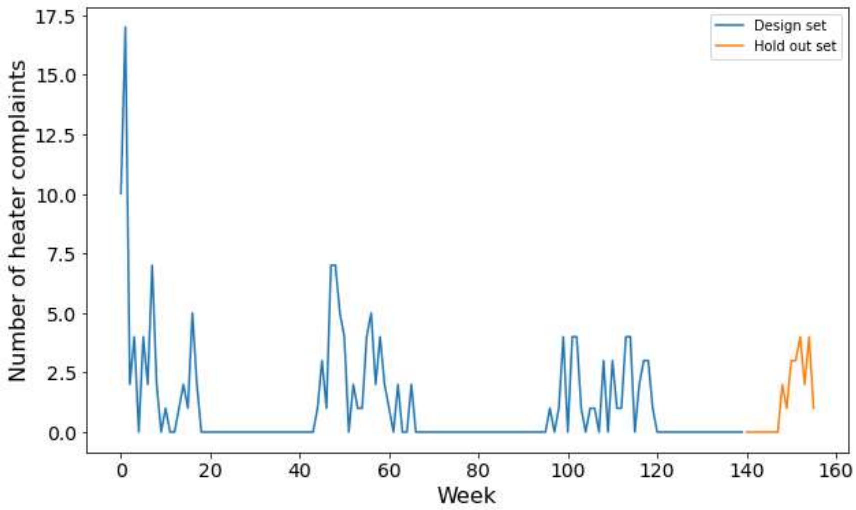

Figure 18 Heater time series split between design and holdout sets

4. Time Series Smoothing

Upon inspecting the AC and heater times series, it can be noted they have several spikes and some fluctuations that seem to be random. As such, time series smoothing was carried on. Several smoothing techniques were examined, for both AC and heater, a Gaussian weighted moving average filter was selected because it was able to reduce those spikes and random fluctuations yet preserve the overall morphology of each time series. As for the window size, a size of six was selected for the AC time series, and a size of four was selected for the heater time series. The larger the window size, the smoother is the time series, thus the easier it is for the model to predict later on; however, extreme smoothing changes the characteristics of the time series and should be avoided. For the AC model the first five values of the time series will be omitted since they are used for smoothing the sixth value, and same for the heater time series where the first three values will be omitted since they are used to smooth the fourth value. As such, the length of the observations in the design set becomes 135 for the AC model, and 137 for the heater model. After examining both tailored and central smoothing for the same window size, central smoothing provided better results, however tailored smoothing was selected because the aim of the MLP model - that will use these smooth time series as an input- is for forecasting purposes. So, whenever the model is used for prediction, applying central smoothing for a certain observation will require future values of the time series, which is not applicable, while tailored smoothing will only require past values of the time series. Figures 19 and 20 below represent the AC and heater time series before and after smoothing respectively.

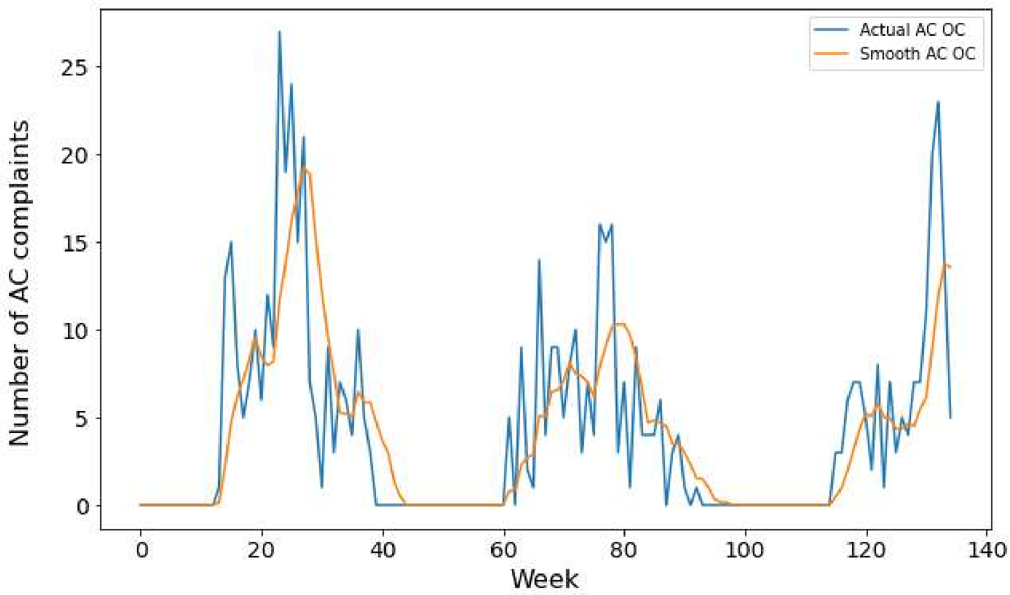

Figure 19 AC time series smoothing for design set

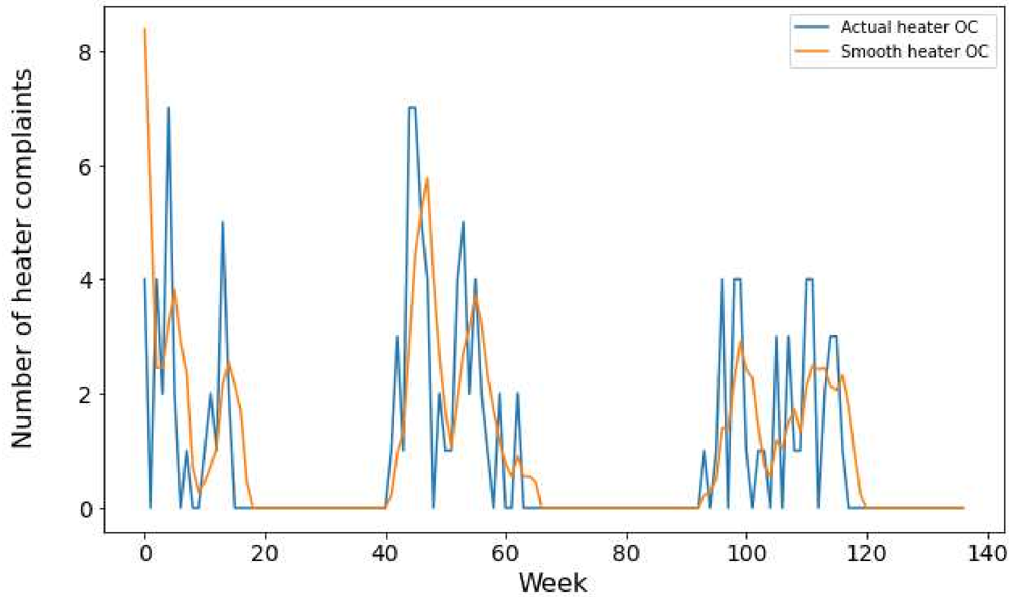

Figure 20 Heater time series smoothing for design set

5. MLP Input Data Preparation

Each of the AC and heater models will have two types of inputs. The first is the developed corresponding occupant complaints smooth series. As for the second type, it represents the exogeneous weather features. Nine time series of the weather features were extracted: MinT, AvT, MaxT, MinRH, AvRH, MaxRH, MinW, AvW, MaxW. These weather features were further investigated to select relevant ones. As such, all these features along with the AC and heater time series were plot against one another (from the design set), and the Spearman correlation coefficients were calculated. Figure 21 shows the scatter plot, and table 5 shows the rho value for the Spearman correlations. The number of AC complaints is highly correlated with the MinT, AvT, Max T, AvW, and MaxW, so these seem to be good predictors for the number of AC complaints. These same weather variables are highly correlated with the number of heater complaints but in the opposite direction. It should also be noted that the MinT, AvT, and

MaxT are highly correlated with one another, and the AvW and MaxW are highly correlated with one another as well. The MinRH, AvRH, MaxRH, and MinW do not show high correlations with the number of AC complaints nor with the number of heater complaints, so they do not seem to have any major effect on the complaint predictions. However, all the weather variables will be used in the MLP prediction models for the AC and the heater because they all could contribute to the prediction even if some had much less contribution than others. It is feasible to input nine variables every time the model (AC or heater) is to be used. Later, when validating the model, there might be a need to drop some variables if the model seems to be overfitting and no other feasible solution were able to resolve the issue. Also, another binary variable "season" was added to distinguish between the weeks where the AC was used and those where the heater was used. A value of 0 refers to the use of the AC, whereas a value of 1 refers to the use of the heater. Figure 22 shows the density plots for each of the weather variables and the AC and heater time series to check if any transformation is required. It can be noted the AC and heater time series are skewed to the right, and this is due to the fact that the use of each depends on the change of the winter and summer seasons, and so a lot of zero observations are expected in the season the heating or cooling system is not being used. Thus, a log(x+1) transformation was used for the number of AC and heater complaints to make their distribution closer to a Gaussian one. Also, the AvW and MaxW distributions are right skewed, so a logarithmic transformation was used for both to them closer to a Gaussian distribution. The other variables show density plots close to bell shaped one.

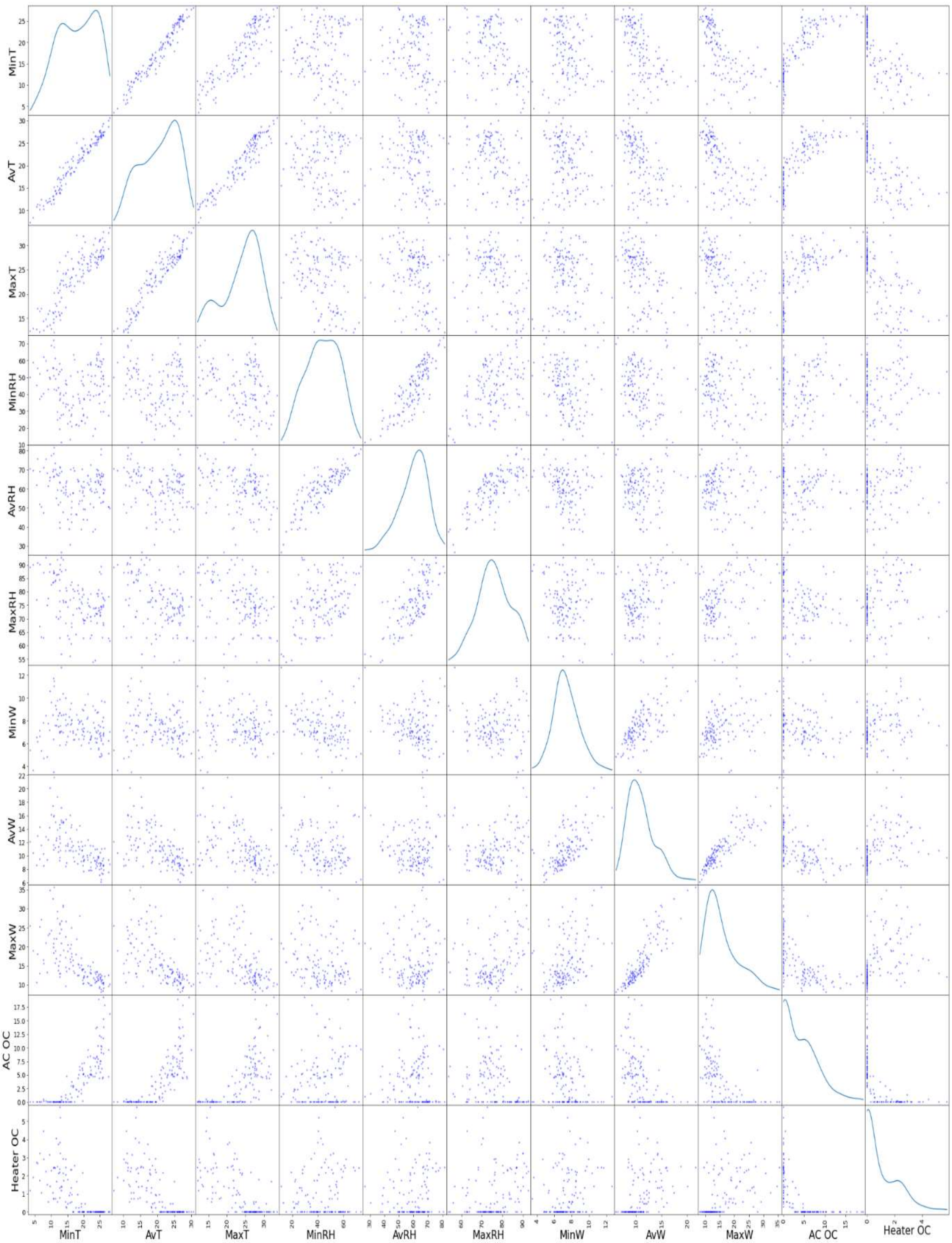

Figure 21 Scatter plot matrix

Table 5 Time series Spearman correlation

	Min T	Av T	Max T	Min RH	Av RH	Max RH	Min W	Av W	Max W	AC OC	Heater OC
MinT	**1.00**	0.97	0.86	0.04	-0.15	-0.38	-0.21	-0.64	-0.75	0.89	-0.84
AvT	0.97	**1.00**	0.94	-0.06	-0.24	-0.39	-0.17	-0.61	-0.71	0.87	-0.83
MaxT	0.86	0.94	**1.00**	-0.26	-0.35	-0.38	-0.10	-0.49	-0.56	0.77	-0.75
MinRH	0.04	-0.06	-0.26	**1.00**	0.82	0.34	-0.31	-0.26	-0.22	0.12	-0.02
AvRH	-0.15	-0.24	-0.35	0.82	**1.00**	0.70	-0.22	-0.03	0.02	-0.05	0.09
MaxRH	-0.38	-0.39	-0.38	0.34	0.70	**1.00**	-0.01	0.26	0.32	-0.32	0.32
MinW	-0.21	-0.17	-0.10	-0.31	-0.22	-0.01	**1.00**	0.65	0.43	-0.23	0.26
AvW	-0.64	-0.61	-0.49	-0.26	-0.03	0.26	0.65	**1.00**	0.91	-0.60	0.56
Max W	-0.75	-0.71	-0.56	-0.22	0.02	0.32	0.43	0.91	**1.00**	-0.68	0.62
AC OC	0.89	0.87	0.77	0.12	-0.05	-0.32	-0.23	-0.60	-0.68	**1.00**	-0.86
Heater OC	-0.84	-0.83	-0.75	-0.02	0.09	0.32	0.26	0.56	0.62	-0.86	**1.00**

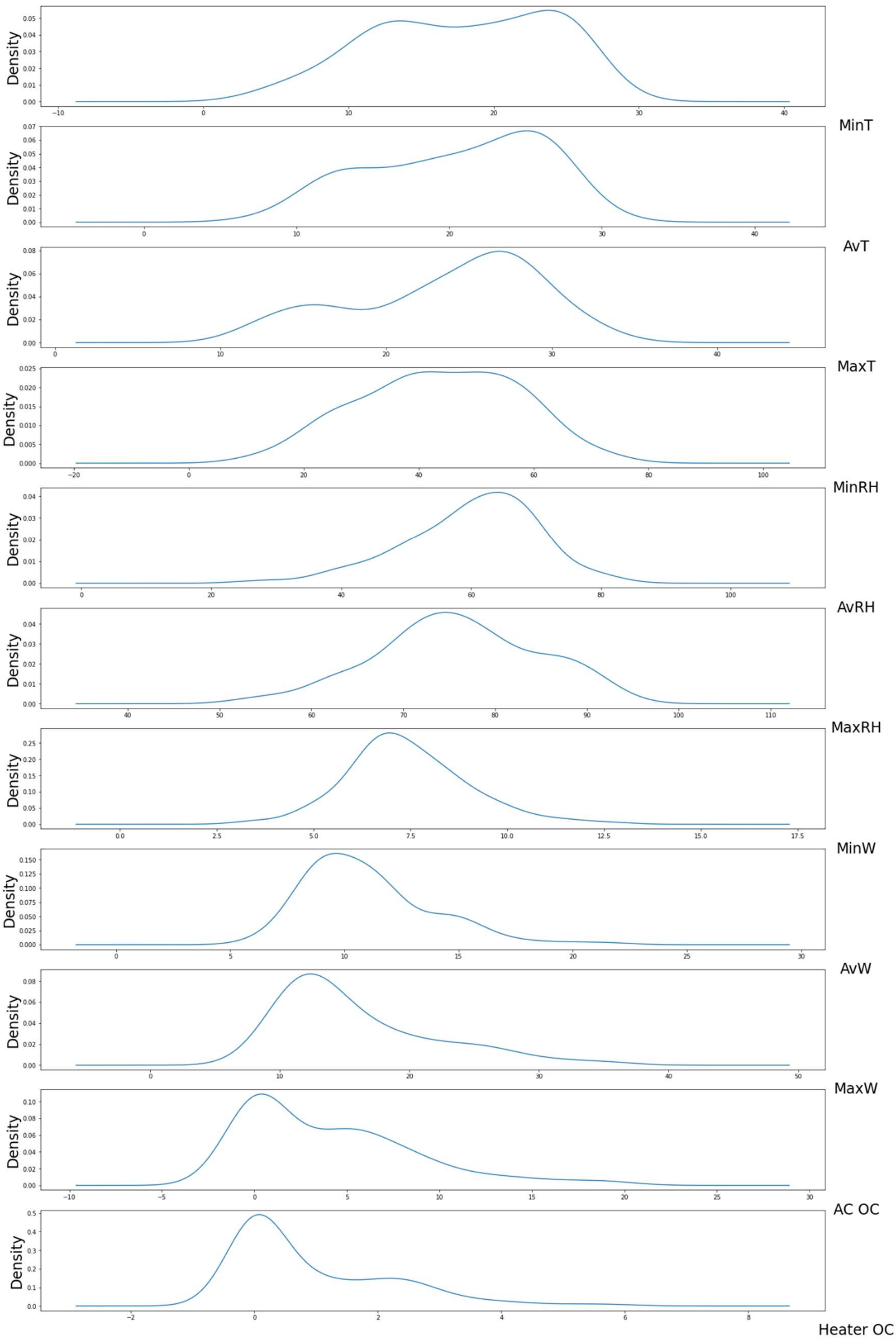

Figure 22 Density plots

Observing tables 2 and 4, it can be noted that the weather variables and the number of AC or heater complaints have ranges that are very different from one another. As such all the variables and AC or heater target were normalized into the range of [0 ,1].

At this stage, the design set is prepared to be used in the MLP models, where the target time series (AC or heater) is smooth, and both input weather features, and the target times are transformed and normalized as applicable. The design set for each of the AC and heater models is then divided into training and validation. The first 70% of the entire data set (109 instances) will constitute the training set, and the following 20% of the entire dataset (31 instances) will constitute the validation set. These two for each model add up to 90% of the observations of the entire data forming the first 140 instances which is the length of the design dataset. It should be noted the data the training set actually has 104 observations not 109 for the AC model because five points were used for initialization in the smoothing process of the AC time series of the design set carried out in the previous sections. Same applies for the heater model where the training data set is formed of 106 instances not 109 where three were used for initialization in the smoothing process of the heater time series of the design set carried out in the previous sections. Figures 23 and 24 present a visualization of the design set of the AC and heater time series respectively split into training and validation sets.

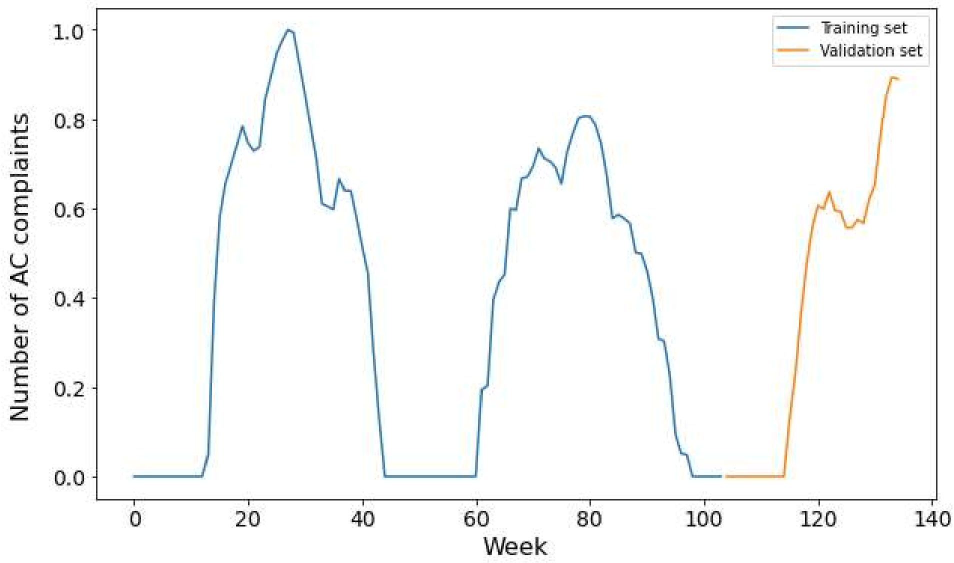

Figure 23 AC time series split into training and validation

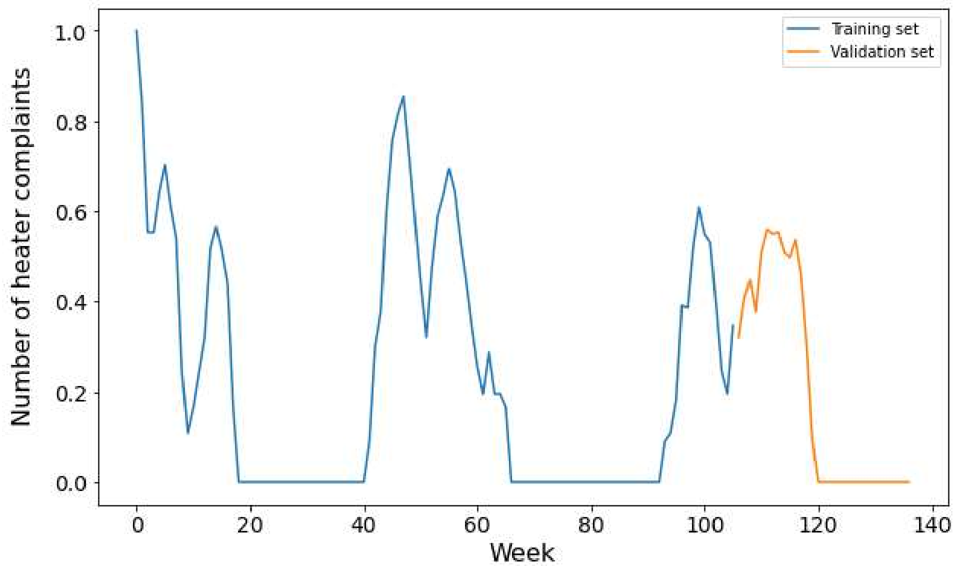

Figure 24 Heater time series split into training and validation

6. MLP model training, validation, and testing

Developing both the AC and heater MLP ANN was conducted in Python using the Keras library. During the development of each, several potential models were evaluated and failed certain evaluations levels (training, validation, stability or testing), so another model had to be tested by changing the different hyperparameters and characteristics as systematically as possible. A suggested model for each is discussed in this section.

For the AC model, a lag of three weeks was selected meaning that every time the model is to make a prediction, it will use the input (the nine weather features, season feature, and the corresponding number of occupant complaints) of each of the previous three weeks. As such, the total number of input features becomes 33; 11 features for each of the previous 3 weeks. And the target is the number of AC occupant complaints for the upcoming (fourth) week. So, the first three instances will be used for -initialization, and the total number of instances in the train set becomes 101 instead of 104. The same applies for the heater model, where a lag of two weeks was selected. So, the model will use the features of the past two weeks to predict the target in the third week. As such, the total number of input features becomes 22; 11 for each of the previous two weeks. Also, the first two instances will be used for initialization, so the number of instances in the train set becomes 104 instead of 106.

The network architecture of the AC MLP model consists of one input layer, two hidden layers and an output layer. The number of neurons in the input layer is equivalent of the number of input features which is 33, whereas in each hidden layer 150 neurons were used, and one neuron was used in the output layer since this is a regression problem and only one target (number of AC complaints) is being predicted. As for the heater MLP model, the network architecture consists also of one input layer, two hidden layers, and an output layer as well. The number of neurons in the input layer is equivalent to the number of input features which is 22, whereas 100 neurons were used in each hidden layer, and one neurons was used in the output layer since this is a regression problem as well and only on target is being predicted which is the number of heater complaints.

For both the AC and heater MLP models, the activation function ReLu was used in the hidden layers in all model updates because it showed good performance and several studies from the literature proved that it outperforms other activation functions. Also, a linear activation function was always selected in the output layer since this is a regression problem and one target being predicted (the number of complaints).

The Adam optimizer was selected as the training algorithm for both AC and heater MLP models because it showed good performance and several studies from the literature proved that it outperforms other traditional optimization algorithm. The default values were used for the parameters β_1 (0.9), for β_2 (0.999), and for ϵ (10^{-9}). A learning rate η of 0.0025 was selected for both models. Moreover, A training batch size of 34 was selected for the AC MLP model and 26 for the heater MLP model. A number of training epoch of 1000 was selected for both models.

The performance function MSE was selected for both and will be used to evaluate the model performance.

Both Ridge $(\ell 2)$ and Lasso $(\ell 1)$ regularization methods were used in both AC and heater MLP models to ensure that the model is not prone to overfitting. This eliminated the need to remove any input features of the weather data, thus all nine were used in each model. The selected Ridge and Lasso hyperparameters were respectively 0.0003 and 0.0001 for the AC MLP model, and both had a value of 0.00035 for the heater MLP model.

What was being updated between one model iteration and another included: the number of lags and the corresponding number of input features, the number of hidden layers, number of neurons in the input layer and each hidden layer, the learning rate of Adam optimizer, training batch size, number of training epoch, and regularization method and the corresponding hyperparameter, while the rest remained the same.

Table 6 below summarizes the selected hyperparameters and characteristics for both the AC and heater MLP models.

Table 6 AC and heater MLP models hyperparameters and characteristics

	AC MLP model	**Heater MLP model**
Number of lags	3	2
Network architecture:		
Number of hidden layers	2	2
Number of neurons in input layer	33	22
Number of neurons in each hidden layer	150	100
Number of neurons in output layer	1	1
Network activation function:		
Activation function in hidden layer (s)	ReLu	ReLu
Activation function in output layer	Linear	Linear
Optimization algorithm:	Adam	Adam
Momentum decay hyperparameter: β_1	0.9	0.9
Scaling decay hyperparameter: β_2	0.999	0.999
Smoothing term: ϵ	10^-9	10^-9
Learning rate: η	0.0025	0.0025
Training batch size	34	26
Number of training epoch	1000	1000
Performance function	MSE	MSE
Regularization:		
Ridge regularization $\ell 1$	0.0003	0.00035
Lasso regularization $\ell 2$	0.0001	0.00035

Using the selected hyperparameters and characteristics, each of the AC and heater MLP models was fit and evaluated.

The AC MLP model had a RMSE on the training data set of 0.801 and a model for (R^2) of 0.971. The error seems to be low as compared to the range of AC complaints and seems to be an acceptable error, and the model is able to explain about 97.1% of the training dataset. Figure 25 shows the actual and predicted values for the number of AC complaints for the training set. It can be noted that the maximum error was obtained at the peak of year one

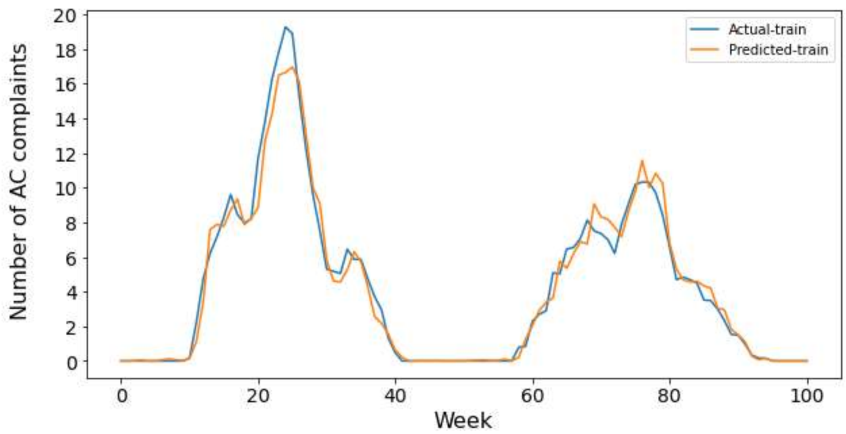

Figure 25 AC MLP model actual and predicted values in train set

The AC MLP model was further tested on the validation set using walk forward validation to obtain an RMSE of 0.836. The validation RMSE of 0.836 is very close to the training RMSE of 0.801, so it seems that the model is not overfitting. It is also reasonable that the validation error is slightly higher than the training error because the model is expected to perform worse on data that was not used to fit it. Figure 26 shows the actual and predicted number of the AC complaints for the validation set.

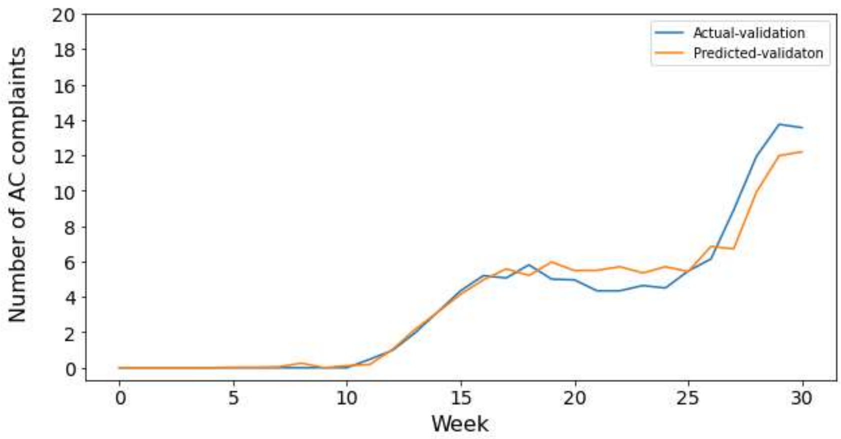

Figure 26 AC MLP model actual and predicted values in validation set

In order to ensure that the AC MLP model was stable, it was fit 30 times using the same hyperparameters and characteristics, but the only thing changing is the initial model weights due to the stochastic nature of the optimization algorithm. An average train RMSE of 0.826 with standard deviation of 0.069 was obtained along with an average validation RMSE of 0.88 and a standard deviation of 0.08. The model is stable in the sense that average train and validation errors are close to each other with the validation one being a bit higher, and each had a low standard deviation. As a result, the selected model hyperparameters and characteristics seem adequate.

Since the AC MLP model is performing well on the training and validation sets, and it seems stable, it was further tested on the holdout test set. The features of the holdout test set should first be prepared in a similar way to the design set. As such, the AC time series was first smoothened using a Gaussian weighted MA and a window size of six. The first five points for initialization were used from the end of the validation set. The AvW and MaxW were transformed using a log transformation, and the target AC time series was transformed using a log(x+1) transformation. All the input features and target were normalized into the range [0,1]. Upon testing the developed model on the holdout set, a test RMSE of 0.878 was obtained which is considered low, close to the train and validation RMSEs, and slightly higher than the train RMSE ensuring that the model is not overfitting. Figure 27 shows the actual and predicted values for the occupant complaints in the test set.

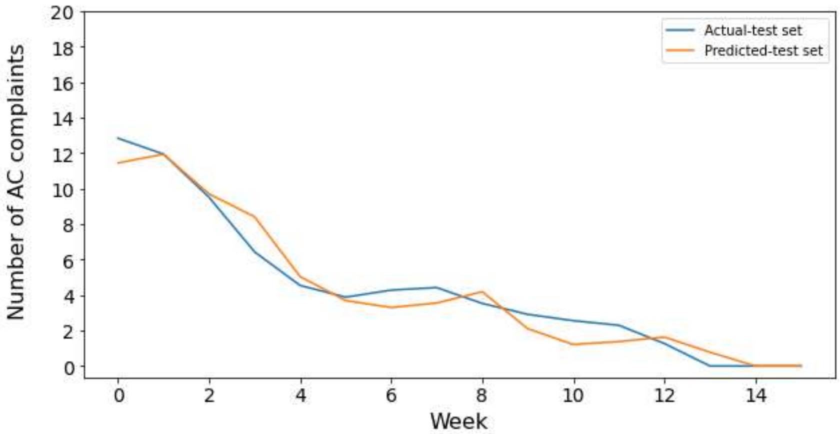

Figure 27 AC MLP model actual and predicted values in test set

The heater MLP model had a RMSE on the training data set of 0.246 and a model for (R^2) of 0.965. The error seems to be low as compared to the range of heater complaints and seems to be an acceptable error, and the model is able to explain about 96.5% of the training dataset. Figure 28 shows the actual and predicted values for the number of heater complaints for the training set. The highest error was observed at the peak at the end of year one.

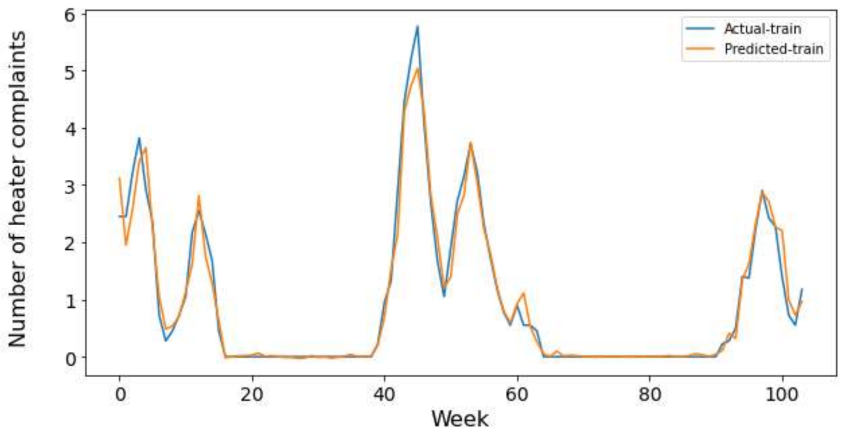

Figure 28 Heater MLP model actual and predicted values in train set

The heater MLP model was further tested on the validation set using walk forward validation to obtain an RMSE of 0.352. The validation RMSE of 0.352 is very close to the training RMSE of 0.246, so it seems that the model is not overfitting. Figure 29 shows the actual and predicted values for the number of heater complaints for the validation set. The maximum error was for the first instance and is considered a bit high.

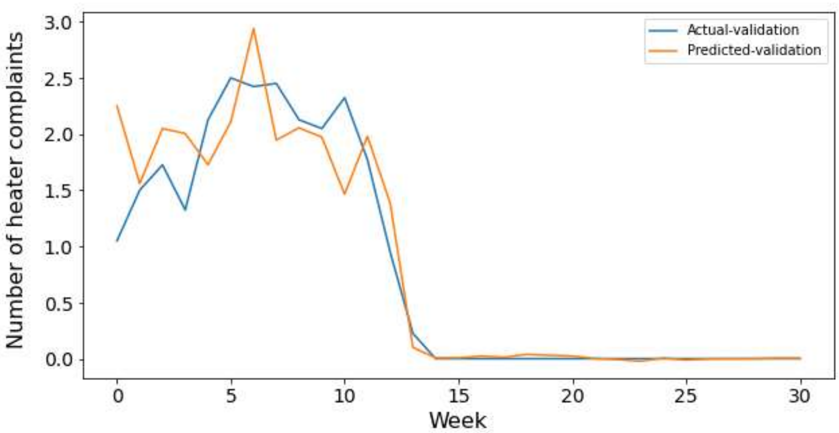

Figure 29 Heater MLP model actual and predicted values in validation set

In order to ensure that the heater MLP model was stable, it was fit 30 times using the same hyperparameters and characteristics, but the only thing changing is the initial model weights due to the stochastic nature of the optimization algorithm. An average train RMSE of 0.326 with standard deviation of 0.032 was obtained along with an average validation RMSE of 0.395 and a standard deviation of 0.037. The model is stable in the sense that average train and validation errors are close to each other with the validation one being a bit higher, and each had a low standard deviation. As a result, the selected model hyperparameters and characteristics seem adequate.

Since the heater MLP model is performing well on the training and validation sets, and it seems stable, it was further tested on the holdout test set. The features of the holdout test set should first be prepared in a similar way to the design set. As such, the heater time series was first smoothened using a Gaussian weighted MA and a window size of 4. The first 3 points for initialization were used from the end of the validation set. The AvW and MaxW were transformed using a log transformation, and the target heater time series was transformed using a log(x+1) transformation. All the input features and target were normalized into the range [0,1]. Upon testing the developed model on the holdout set, a test RMSE of 0.372 was obtained which is considered low, close to the train and validation RMSEs, and slightly higher than the train RMSE ensuring that the model is not overfitting. Figure 30 shows the actual and predicted values for the occupant complaints in the test set.

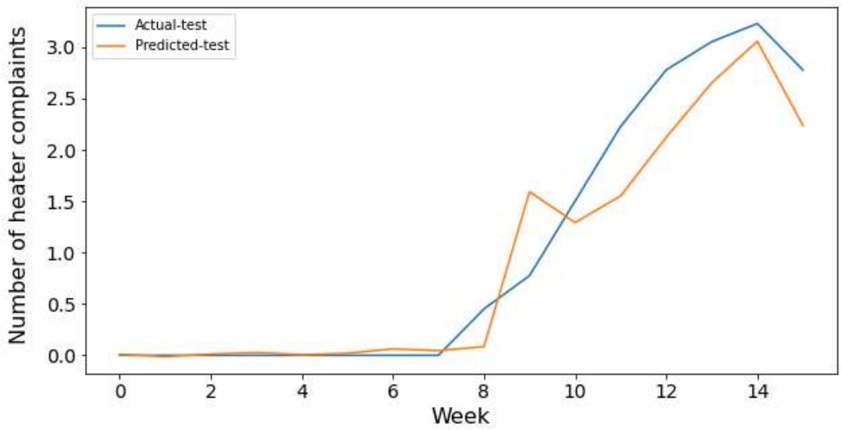

Figure 30 Heater MLP model actual and predicted values in test set

B. Benchmark Models Development

The developed AC and heater MLP models were benchmarked against the state of the art ARIMA model. This section shows how these ARIMA models were developed.

1. Time Series Preparation

For both the AC and the heater ARIMA models, the same steps and results provided in the previous section of developing the MLP models apply for: data collection and preprocessing, text cleaning and mining, data split, and time series smoothing.

2. ARIMA Input Data Preparation

Each of the AC and heater ARIMA models will have two types of input as with the MLP models including the corresponding smooth AC or heater times series, the season, and the additional potential weather features.

The weather features that will be used to develop the ARIMA models are selected based on the scatter matrix plot and Spearman correlation provided earlier in figure 21 and table 5 for the design set. MinT, AvT, and MaxT are highly correlated thus the MinT and MaxT will be removed. Also, the MinRH and AvRH are highly correlated and thus the MinRH will be removed. The MaxW and AvW are also highly correlated, so the AvW will be removed. These are removed for both the AC and heater ARIMA models. So, the remaining input weather features are: AvT, AvRH, MaxRH, MinW, and MaxW. Also, similar to the MLP model a binary variable "season" was added to the dataset. The target AC and heater time series will be transformed using the log(x+1) so that the distribution of each is closer to a bell-shaped one. The ADF test was conducted for both time series where the AC time series turned out to be stationary at the significance level of 1% where an ADF statistic of -4.95 was obtained which is smaller than the critical value of -3.49. As for the heater time series, it turned out to be non-stationary at the 1% significance level where the corresponding ADF statistic turned out to be -2.50 which is larger than the critical ADF value of -3.48 at the 1% significance level. As such, transforming the heater time series into a stationery one will be taken care of in the upcoming section when selecting the parameters of the ARIMA model. Moreover, all the input features and the target for each of the AC and heater models were normalized into the range of the [0,1] because these had a widely varying range compared to one another.

3. Splitting Design Data into Training and Validation

Splitting the design data into training and validation will follow the same procedure as with the MLP model. So, the training set of the AC ARIMA model will include 106 instances and that of the heater ARIMA model will include 104 instances. The data split for the target time series was visualized in figures 23 and 24.

4. ARIMA (p, d, q) (P, D, Q) s Modeling

The ACF and PACF plots for the AC time series shown in figure 31 were used to identify the initial parameters for the AC ARIMA model. An AR order "p" of 2 was selected, and a MA order "q" of 4 was selected. As stated earlier, the AC time series is stationary, so an integration order "d" of 0 was selected. From the ACF plot, there is an obvious seasonality, and it seems seasonal integration order "D" of 1 is required with a seasonal period "s" of 52 since the weekly data varies yearly. The ACF and PACF plots do not show any correlation at any season for example at weeks 52 and 104. So, a seasonal AR order "P" of 0 and a seasonal MA order "Q" of 0 were selected. Based on that, an ARIMA (2, 0, 4) (0, 1, 0) 52 was selected as the initial model, including the previously selected exogenous variables **:** AvT, AvRH, MaxRH, MinW, MaxW and the season, to start the next step of coefficient estimation. During this step, the model fit R^2, the RMSE, and the AIC value were recorded in every iteration during which insignificant variables are dropped and the model is fit again. The results are summarized in table 7. Model 11 was selected because it has the lowest AIC value and the model fit R^2 and the model RMSE both represent good values, and all the AR and MA coefficients ended up being significant except for one AR coefficient, and none of the exogenous variables was significant.

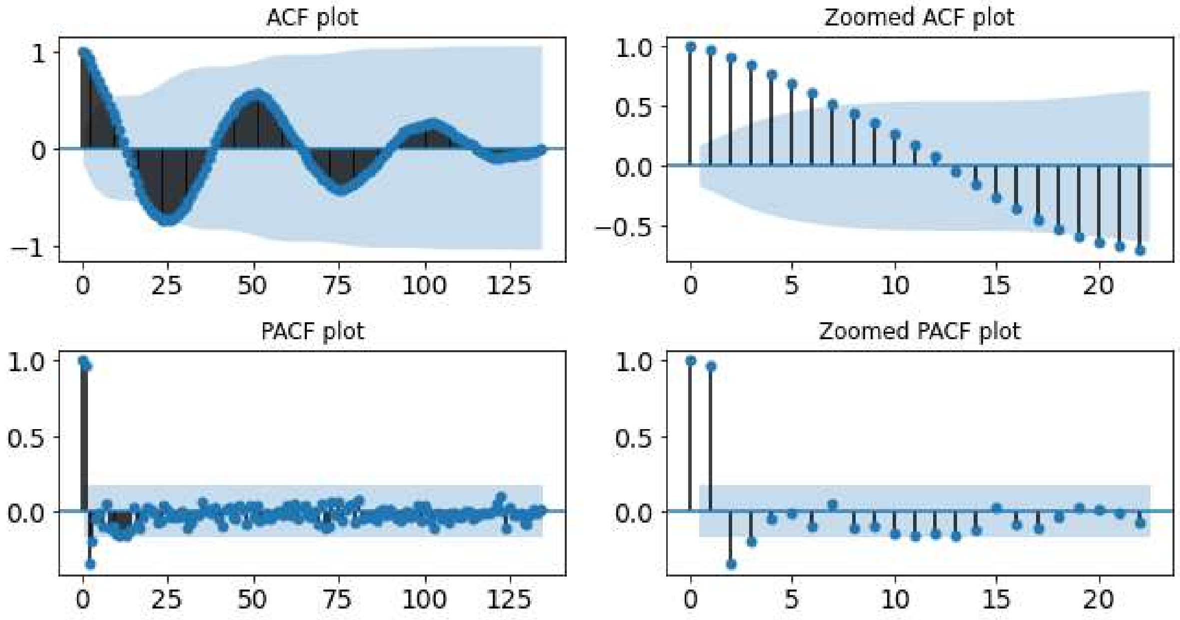

Figure 31 AC time series ACF and PACF plots

Table 7 ARIMA AC model iterations for estimation step

Model number	(p, d, q)	(P, D, Q, s)	R2 train	RMSE train	AIC	Exogeneous variables
1	(2, 0, 4)	(0, 1, 0, 52)	0.952	0.942	-116.254	AvT, AvRH, MaxRH, MinW, MaxW, season
2	(2, 0, 3)	(0, 1, 0, 52)	0.922	0.957	-113.273	AvT, AvRH, MaxRH, MinW, MaxW, season
3	(2, 0, 2)	(0, 1, 0, 52)	0.918	0.984	-114.293	AvT, AvRH, MaxRH, MinW, MaxW, season
4	(2, 0, 1)	(0, 1, 0, 52)	0.906	1.058	-111.712	AvT, AvRH, MaxRH, MinW, MaxW, season
5	(2, 0, 0)	(0, 1, 0, 52)	0.906	1.054	-112.795	AvT, AvRH, MaxRH, MinW, MaxW, season
6	(2, 0, 0)	(0, 1, 0, 52)	0.906	1.056	-114.77	AvT, AvRH, MaxRH, MaxW, season
7	(2, 0, 0)	(0, 1, 0, 52)	0.907	1.048	-116.694	AvT, MaxRH, MaxW, season
8	(2, 0, 0)	(0, 1, 0, 52)	0.908	1.048	-118.646	AvT, MaxW, season
9	(2, 0, 0)	(0, 1, 0, 52)	0.908	1.043	-120.376	AvT, season
10	(2, 0, 0)	(0, 1, 0, 52)	0.913	1.015	-120.44	AvT
11	(2, 0, 0)	(0, 1, 0, 52)	0.913	1.015	-122.442	None

12	(1, 0, 0)	(0, 1, 0, 52)	0.893	1.126	-119.605	None

Figure 32 shows the actual and predicted values for the number of AC complaints in the train set using the AC ARIMA model. The maximum error is shown at the peak. It should be noted that the model used out of the train set 52 points to account for the seasonal differencing "D" of 1.

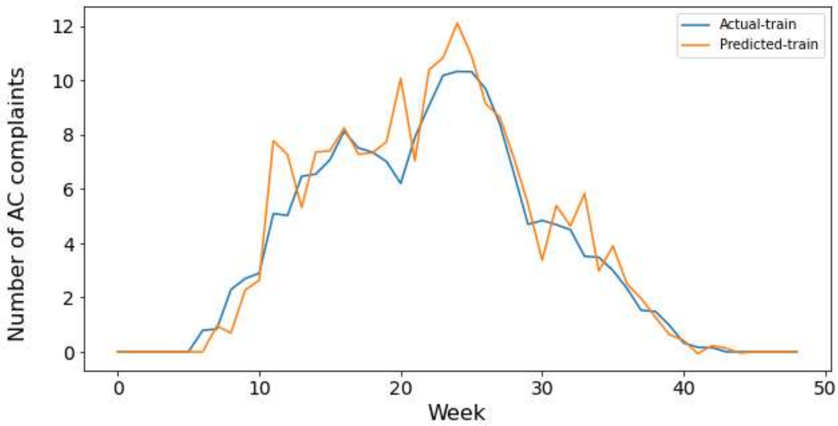

Figure 32 AC ARIMA model actual and predicted values in train set

The selected model was then tested on the validation set using the walk forward validation method. An RMSE of 0.998 was obtained which is slightly lower than the RMSE on the train set, so the model seems to be slightly overfitting. Further attempts to improve the selected to eliminate the overfitting issue have failed. Figure 33 shows the actual and predicted values for the number of AC complaints in the validation set using the selected AC ARIMA model.

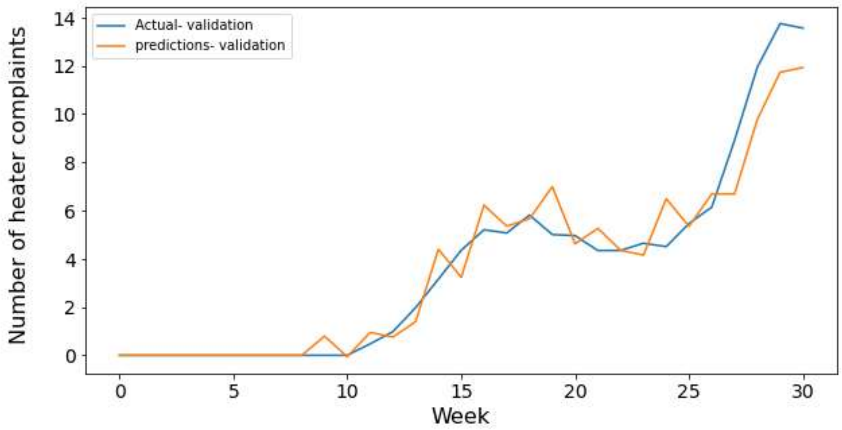

Figure 33 AC ARIMA model actual and predicted values in validation set

Moving to step 3 of diagnosis, the residual assumptions upon fitting the model were tested. Figure 34 represents the residual plot, histogram, normal Q-Q plot, and the ACF correlogram of the residuals.

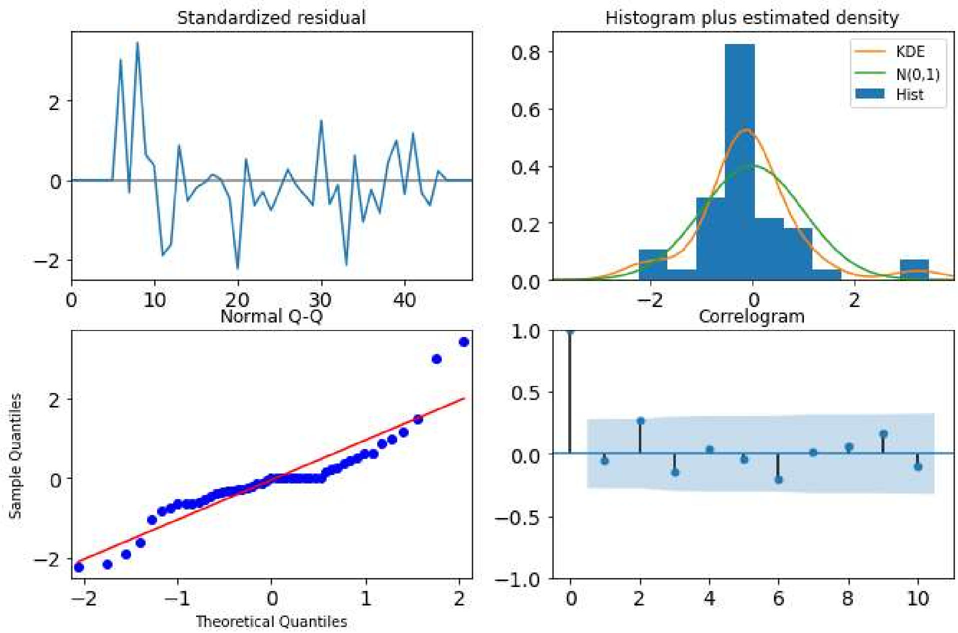

Figure 34 AC ARIMA model residual diagnostics

It can be inferred that the residuals are not correlated, seem to have a mean very close to zero, a constant variance and seem to be normally distributed but slightly skewed.

Since the residual assumptions are verified, the selected AC model was tested on the holdout test set. A RMSE of 1.117 was obtained which is very close to the RMSE on the train set ensuring that the model is unlikely to overfit. Figure 35 shows the actual and predicted values for the number of AC complaints in the holdout test set using the selected AC ARIMA model.

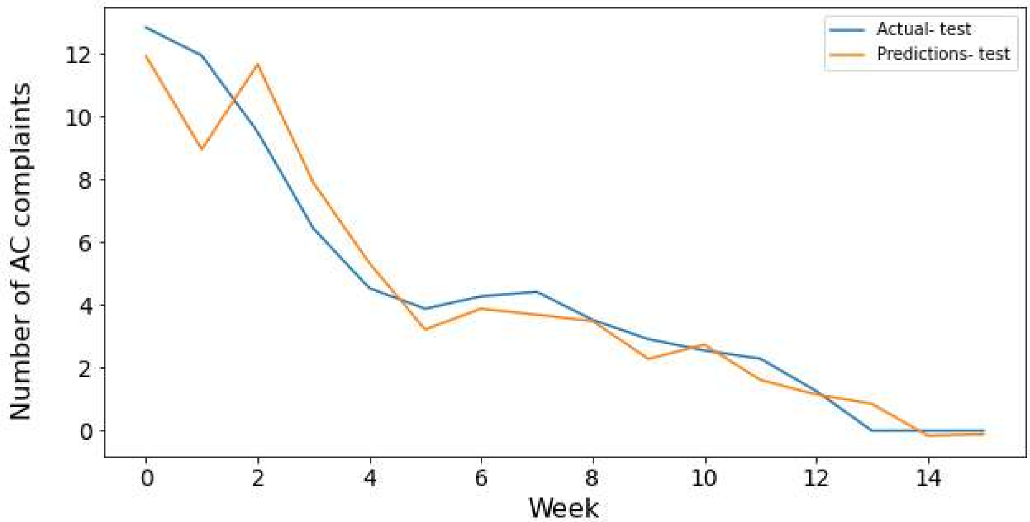

Figure 35 AC ARIMA model actual and predicted values in test set

Developing the heater ARIMA model followed a similar process to that of the AC ARIMA model. The ACF and PACF plots of the heater time series are shown in figure 36. These were used to identify the initial ARIMA model. An AR order “p” of 2 was selected, and a MA order “q” of 4 as selected. As stated earlier, the heater time series is not stationary, so an integration order “d” of 1 was selected. From the ACF plot, there is an obvious seasonality, and it seems that a seasonal integration order “D” of 1 is required with a seasonal period "s" of 52 since the weekly data varies yearly. The ACF and PACF plots do not show any correlation at any season for example week 52 and 104. So, a seasonal AR order "P" of 0 and a seasonal MA order "Q" of 0 were selected. Based on that, an ARIMA (2, 1, 4) (0, 1, 0) 52 was selected as the initial model, including the previously selected exogenous variables **:** AvT, AvRH, MaxRH, MinW, MaxW and the season, to start the next step of coefficient estimation. During this step, the model fit R^2, the RMSE, the AIC value were recorded in every iteration during which insignificant variables are dropped and the model is fit again. The results are summarized in table 8. Model 7 is selected because it has the lowest AIC value and the model fit R^2 and the model RMSE both represent good values.

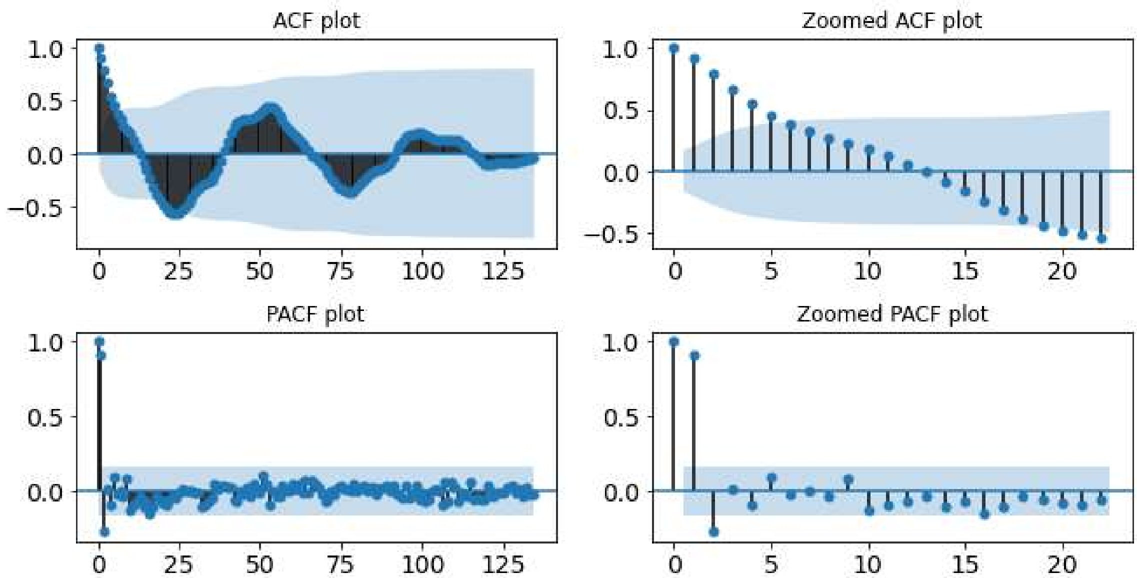

Figure 36 Heater time series ACF and PACF plots

Table 8 ARIMA heater model iterations for estimation step

Model number	**(p, d, q)**	**(P, D, Q, s)**	**R2 train**	**RMSE train**	**AIC**	**Exogeneous variables**
1	(2, 1, 4)	(0, 1, 0, 52)	0.643	0.606	-116.094	AvT, AvRH, MaxRH, MinW, MaxW, season
2	(2, 1, 3)	(0, 1, 0, 52)	0.663	0.59	-115.852	AvT, AvRH, MaxRH, MinW, MaxW, season
3	(2, 1, 2)	(0, 1, 0, 52)	0.671	0.582	-116.123	AvT, AvRH, MaxRH, MinW, MaxW, season
4	(2, 1, 1)	(0, 1, 0, 52)	0.661	0.591	-116.599	AvT, AvRH, MaxRH, MinW, MaxW, season
5	(2, 0, 1)	(0, 1, 0, 52)	0.854	0.388	-117.582	AvT, AvRH, MaxRH, MinW, MaxW, season
6	(2, 0, 1)	(0, 1, 0, 52)	0.853	0.388	-119.558	AvT, AvRH, MinW, MaxW, season
7	(2, 0, 1)	(0, 1, 0, 52)	0.852	0.391	-121.264	AvT, AvRH, MaxW, season
8	(2, 0, 0)	(0, 1, 0, 52)	0.847	0.398	-119.13	AvT, AvRH, MaxW, season

Figure 37 shows the actual and predicted values for the number of heater complaints in the train set using the heater ARIMA model. The maximum error is shown at the peak at the beginning of the year.

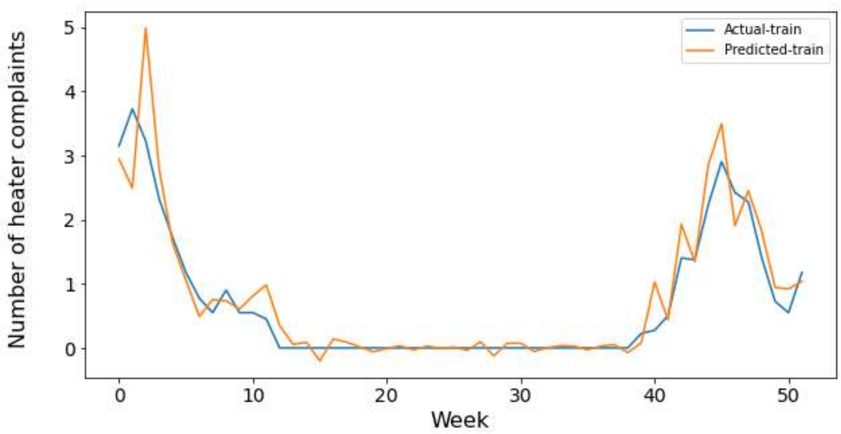

Figure 37 Heater ARIMA model actual and predicted values in train set

It should be noted that the model used out of the train set 52 points to account for the seasonal differencing “D” of 1.

The selected model was then tested on the validation set using the walk forward validation method. An RMSE of 0.384 was obtained which is very close to the RMSE on the train set, so the model does not seem to overfit. Figure 38 shows the actual and predicted values for the number of heater complaints in the validation set using the selected heater ARIMA model.

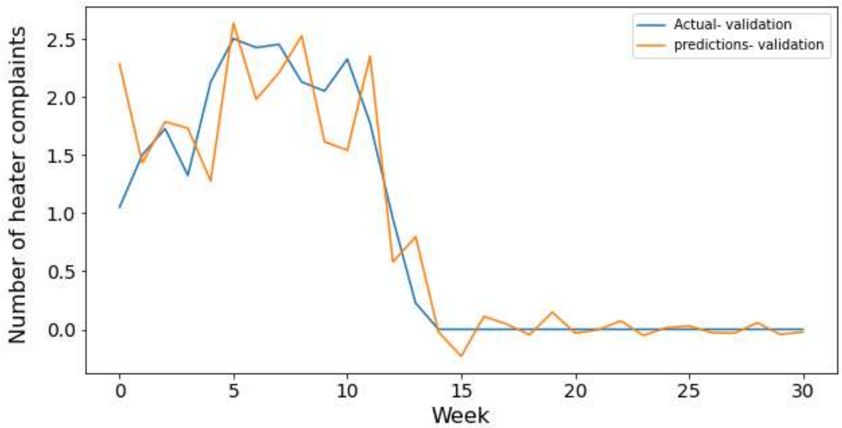

Figure 38 Heater ARIMA model actual and predicted values in validation set

Moving to step 3 of diagnosis, the residual assumptions upon fitting the model were tested. Figure 39 presents the residual plot, histogram, normal Q-Q plot, and the ACF correlogram. It can be inferred that the residuals are not correlated, seem to have a mean very close to zero, a constant variance and seem to be normally distributed.

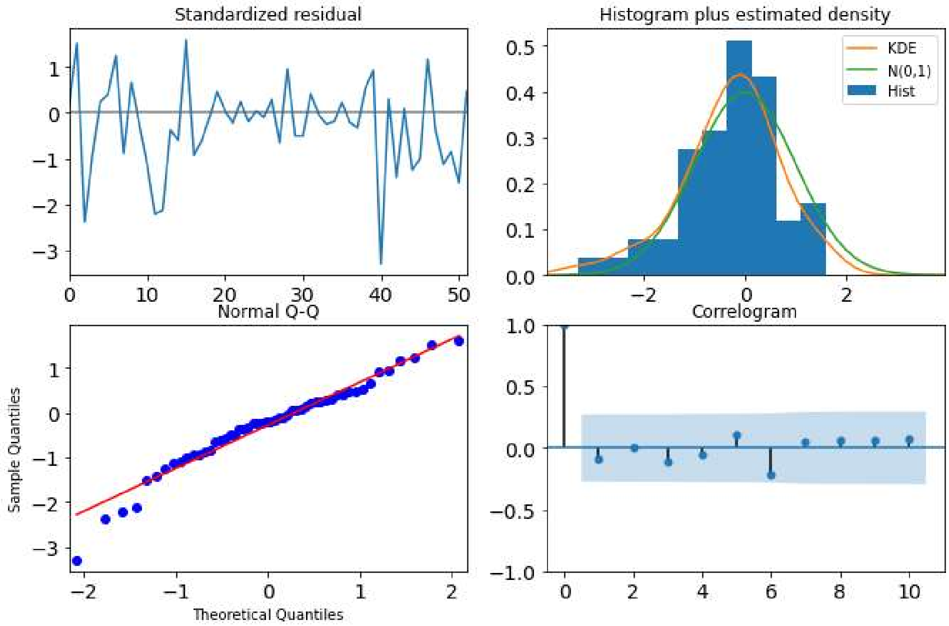

Figure 39 Heater ARIMA model diagnostics

Since the residual assumptions are verified, the selected heater model was tested on the test set. A RMSE of 0.265 was obtained which is smaller than the RMSE on the train set (0.391) meaning that the model is likely to overfit. Further attempts to improve the model have failed. Figure 40 shows the actual and predicted values of the number of heater complaints in the holdout test set using the selected heater ARIMA model.

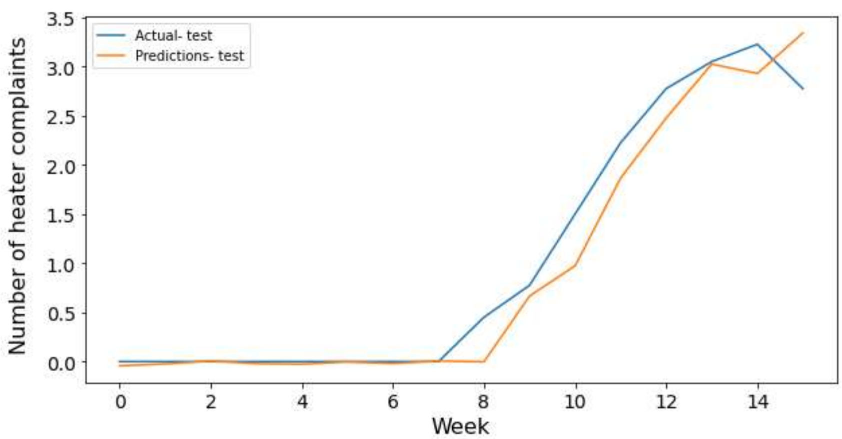

Figure 40 Heater ARIMA model actual and predicted values in test set

C. MLP and ARIMA Models Comparison and Discussion

Table 9 summarizes the obtained model fit R^2, RMSEs on the train, validation, and test sets of both MLP and ARIMA AC and heater models.

Table 9 MLP and ARIMA models comparison

	Train R^2	Train RMSE	Validation RMSE	Test RMSE	Exogenous variables
AC MLP	0.971	0.801	0.836	0.878	MinT, AvT, MaxT, MinRH, AvRH, MaxRH, MinW, AvW, MaxW, season
AC ARIMA	0.913	1.015	0.998	1.117	None
Heater MLP	0.965	0.246	0.352	0.372	MinT, AvT, MaxT, MinRH, AvRH, MaxRH, MinW, AvW, MaxW, season
Heater ARIMA	0.852	0.391	0.384	0.265	AvT, AvRH, MaxW, season

The R^2 of the AC MLP (0.971) model is slightly higher than that of the AC ARIMA model (0.913). However, such slight increase at such an already high value (> 0.9) is not easy to obtain. The RMSEs of the AC MLP model on the test set (0.878) is lower than that of the AC ARIMA model (1.117) meaning that the MLP model has a higher ability to generalize to new data it had not encountered before.

As for the heater models, the R^2 of the heater MLP (0.965) model is higher than that of the heater ARIMA model (0.852) which shows that the MLP model was able to explain the training data set better. The RMSEs of the heater MLP model on the test set (0.372) is higher than that of the heater ARIMA model (0.265) , but this does not mean that that the ARIMA model has a higher ability to generalize to new data because if compared to the train RMSE (0.391), the ARIMA model is highly overfitting, whereas the MLP model is not.

It should be noted that the heater models RMSE is always smaller than that of the AC models, this is because the range of the heater complaints is much lower than that of the AC complaints, and thus a lower error is expected lowering the threshold of an acceptable error. Moreover, it is notable that the MLP models require all 10 exogenous variables where the ARIMA models require 4 or none. Having a lower number of variables does not give any superiority for the ARIMA models because the weather data is a very clean dataset and creating the weather time series is a very smooth task that is not time-consuming.

Based on the above, the MLP models for both the AC and heater showed good performance and provided some improvements over the traditional corresponding ARIMA model and are valid to forecast the number of thermal complaints for the upcoming week.

D. Work Significance: Resource Staffing Plan

The developed AC and heater MLP models were employed in a staffing problem for thermal complaints to show the significance of this research work. Two scenarios were developed and compared. The first represents the base case that the FM unit would have been adopting corresponding to the discussed case study. Appendix A provides details on how this hypothetical base case was developed. The second one represents the updated case upon virtually employing the developed MLP models of this research work.

The scenarios were based on a span of four months, equivalent to 16 weeks that represent the test set that was used earlier to obtain the generalization error of each of the MLP models. These represent the last four months of year three: September, October, November, and December. The problem addresses staffing for thermal complaints, which is a combination of both AC and heater related complaints. So, each MLP model was used to predict the number of the corresponding complaints per week, and these were summed up for each week to obtain the predicted number of weekly thermal complaints. The variation of the predicted thermal complaints along the 16-week span is shown in figure 41. It should be noted that, as per the MLP models development for one-step ahead forecasting, the number of complaints was predicted one week at a time. Common thermal problems at the apartment level based on the occupant complaints included but were not limited to: dirty AC filters, refrigerant leakage or shortage, AC water drainage, the need for radiator bleeding, radiator leakage, electric circuit issues, thermostat related issues, and others.

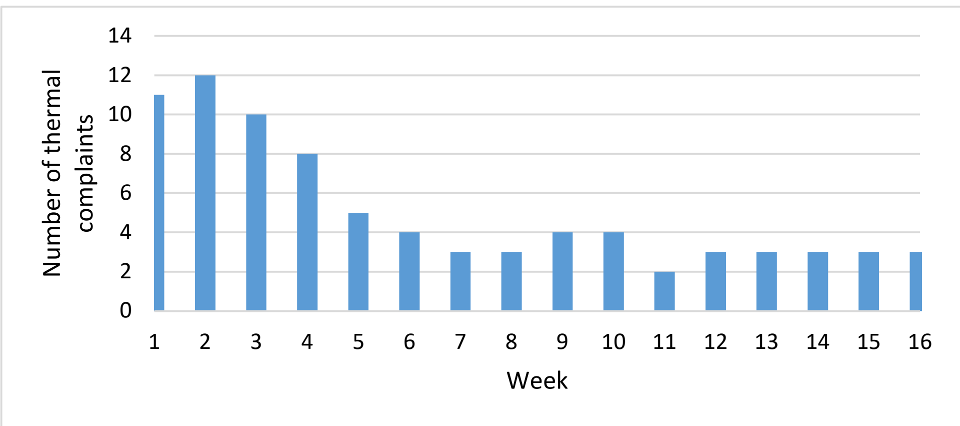

Figure 41 Number of predicted thermal complaints

As for the allocation of maintenance technicians to handle the predicted thermal complaints, the maintenance department of the FM unit had a constant number of 21 hired technicians that are present throughout each month. Considering a 24- hour and seven days a week service, each technician is required to work five shifts per week, with each shift being eight hours long. This sums up to a total of three shifts per day for seven days a week, with each shift having five technicians. As mentioned earlier, this department is responsible for handling three main tasks: Monitoring and handling BMS related issues, conducting routine check-ups for regular maintenance activities, and handling occupant complaints. It is assumed that the number of man hours of each technician is split equally between these three domains, 33.33% for each. Moreover, thermal complaints often constitute 13% of the overall complaints received and handled by the maintenance department as inferred from historical data. So, it is assumed that out of the total number of man hours of each technician allocated to handle occupant complaints, 13% will be allocated for thermal complaints in particular. As such, the number of man hours each technician is responsible to allocate for handling thermal complaints becomes: 40-man hours / week * 33.33% * 13% = 1.733-man hours per week. So, the total man hours available per week becomes 1.733* 21 = 36.396-man hours per week. This is shown by the constant straight line in figure 42. It represents the number of man hours that would have been allocated by the maintenance department to handle the weekly thermal complaints had they not been using the developed tool of this research work.

As for the MLP- based case, the predicted number of thermal complaints was translated to equivalent man hours after obtaining the estimated time taken to address each thermal complaint. According an interviewed expert in staffing for HVAC maintenance activities, most frequently, it takes about four hours to handle a common thermal complaint and the best-case scenario, if it is a very basic problem, it would take about an hour. The worst case scenario is when a complaint takes a longer time, around 12 hours, to be handled for example if the technician has to assess the issue, purchase a certain part of the equipment that needs to be replaced, and come back and actually replace it. As such, the time to handle a thermal complaint can be modeled by a triangular distribution with a minimum of one hour, a mode of four hours and a maximum of 12 hours. As such, the estimated time to handle each complaint was sampled from the corresponding triangular distribution. The resource histogram in figure 42 shows the number of man hours required to handle the predicted weekly thermal complaints.

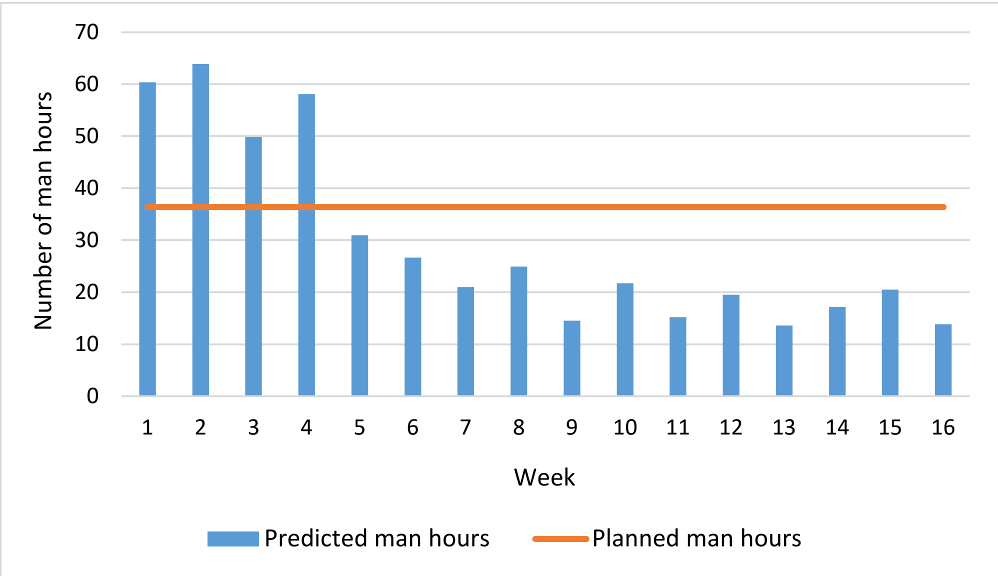

Figure 42 Planned and predicted weekly man hours

Overall, the developed man hour resource histogram shows a total of 582 available man hours throughout the 16-week period and a required 472-man hours upon predicting the number of weekly thermal complaints. This shows that about 19% of the total available man hours are overstaffed.

Considering the first four weeks, it can be inferred that the maintenance department would have been understaffing with a total of 86.57-man hour-shortage if they were to allocate 36.396-man hours per week. These understaffing man hours represent 37.29% out of the total required hours over the span of these four weeks. As such, the thermal complaints will not be handled properly. The occupants will not be satisfied to have their complaints being resolved few weeks later maybe for the required technicians to be available, or maybe not resolved in the first place and they are likely to be issuing even more complaints. Moreover, such delay to address thermal complaints might cause the problem to be amplified and thus will require more recourses to handle. In addition to that, some thermal issues at the level of the apartments might be a sign of an even more major problem related to the central heating or cooling systems which would require a long time to be handled and a large amount of resources. This understaffing situation could have been handled by two scenarios: either by working overtime or by postponing handling the complaints until technicians are available. For the first scenario, it is assumed that, on average, a maintenance technician's wage in Lebanon is 10$/hour and increases to 15$/hour during overtime hours. So, if the additional staffing hours required (86.57-man hours) were to be covered by working overtime, this would lead to an incurred additional cost of 433$ over the span of the first four weeks. This cost is equivalent to 15.7% of the total cost of the required man hours during these four weeks due to working overtime. Although the complaints are resolved and the occupants are relatively satisfied, this has incurred an additional cost for the maintenance department and must have hindered the technicians' productivity. While in the second scenario, it would have taken until mid of week 11 to handle all the thermal complaints issued during the first four weeks considering the available man hours. This reflects a delay of 1 to 10.5 weeks to handle these complaints depending on when each was received and when exactly it was handled which in turns depends on the available man hours and the newly received complaints in each week.

As for the remaining 12 weeks, the maintenance department would have been overstaffing with a total of 197-man hours. As such, this also represents a loss of resources that could have been allocated to other tasks in the maintenance department. These total lost hours due to overstaffing represent 45.14% of the available man hours over the 12-week span. This percentage is expected to be slightly lower in reality because the maintenance department would most likely be allocating, on the spot, some of these man hours to other tasks in the department that are understaffed, or to other thermal complaints that were deferred from previous weeks that were not resolved earlier due to understaffing. Moreover, idle 197-man hours could be translated to an incurred cost of 1,971$ without getting any thermal complaint handled in return. This is equivalent to 45% of the cost of the available man hours during these 12 weeks.

This example shows that staffing based on a static number of technicians per month would incur additional unnecessary costs, promote occupants' dissatisfaction along with a deteriorated building performance. As such, a good practice would be to allocate the number of technicians and the corresponding man hours based on the predicted number of thermal complaints for the upcoming week to be able to handle such complaints. This ensures mitigating the issues of overstaffing and understaffing along with their accompanied drawbacks and ensures occupant satisfaction and proper performance of the heating and cooling related systems at the level of the apartment and even at the level of the building.

Given that the developed example encountered several assumptions and uncertainties, other possible scenarios were assessed ana analyzed to ensure that the significance of this research work still holds. This analysis is conducted considering that some factors remain constant in each scenario because their corresponding uncertainties were mitigated otherwise. These factors include: the number of predicted weekly thermal complaints because they were obtained upon employing the developed MLP models, the estimated time required to handle each complaint because it was sampled from a triangular distribution based on input provided by the interviewed expert, and thus the corresponding number of predicted man hours required to handle the weekly thermal complaints. However, the uncertainty in the developed problem comes from lack of staffing related data corresponding to the developed case study including: the constant number of technicians hired per month in the maintenance department and the actual percentage of the number of man hours of each technician allocated to thermal complaints in particular. These two factors reflect on the number of planned man hours available every week to handle thermal complaint. As such, nine scenarios were developed, compared and analyzed to assess how the number of overstaffed man hours and the number of understaffed man hours vary over the span of 16 weeks to study the impact of the two mentioned factors. For the number of technicians per month, three options were considered: 16, 21, and 25, and for the percentage of thermal complaints three options were considered as well: 6.95%, 13%, and 23.9%. The reasons for selecting these options in particular is justified in Appendix A considering the humble availability of staffing related data.

Figure 43 shows for each percentage of thermal complaints, the variation of the total number of understaffed man hours as a function of variation of the number of technicians hired per month over the span of 16 weeks, and figure 44 shows a similar graph but for the variation of the number of total overstaffed man hours.

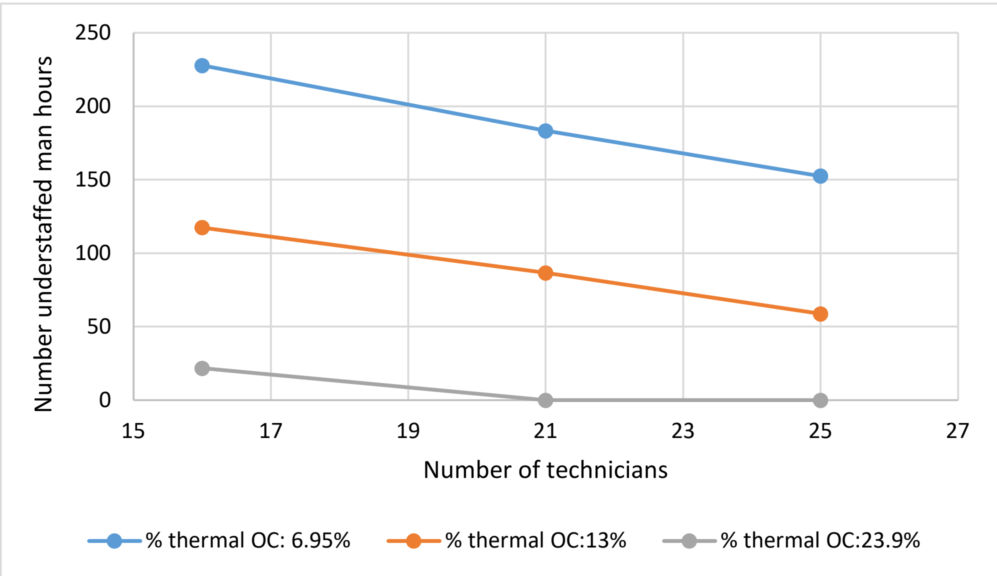

Figure 43 The variation of the number of understaffed man hours

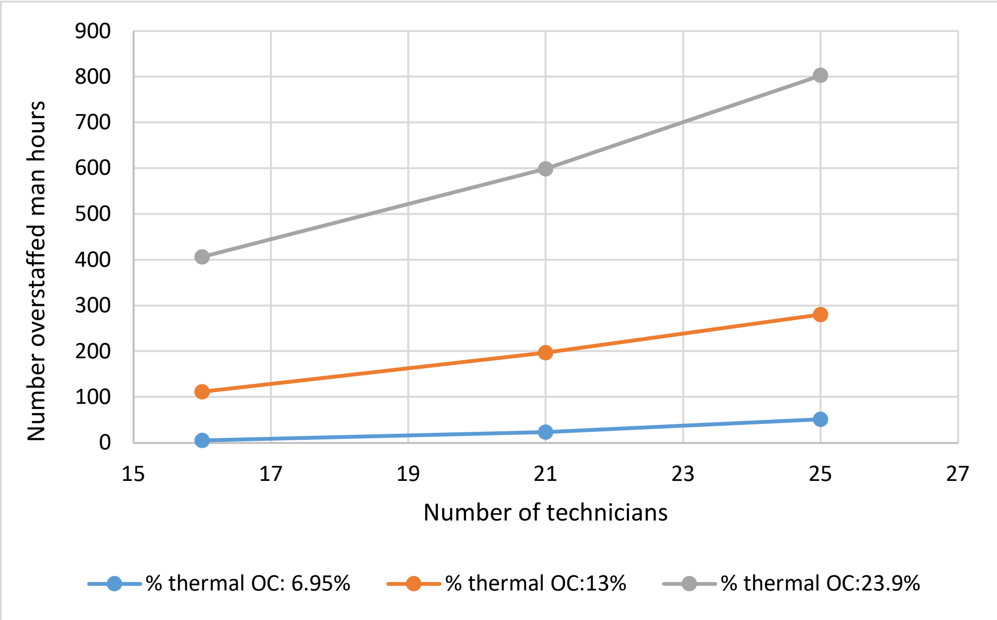

Figure 44 The variation of the number of overstaffed man hours

Based on the analysis of the two developed graphs the following can be inferred:

- For a certain number of available technicians per month (16, 21, or 25), the more thermal complaints received (6.23% to 13% to 23.9%), the more man hours of each technician will be allocated to handle thermal complaints, and thus this will decrease the total number of the understaffed man hours span, but will increase the total number of the overstaffed man hours beyond the available limit over the 16-week span.

- For a certain percentage of thermal complaints (6.23%, or 13%, or 23.9%), if more technicians are hired per month (16 to 21 to 25), then a larger number of thermal complaints can be handled, and thus there would be a decrease in the total number of understaffed man hours, but an increase in the total number of overstaffed man hours.

- For an increase in the number of technicians hired coupled with an increase in the percentage of received thermal complaints and thus the number of man hours allocated to thermal complaints (16 and 6.95% to 21 and 13% to 25 and 23.9%) , there is a decrease in the total number of understaffed man hours form one case to the following one, and an increase in the number of overstaffed man hours as expected.
- It should also be noted that, the analyzed scenarios are clear evidence on the need to allocate the number of man hours to handle thermal complaints dynamically. Using a static number of workers will not balance between the need of additional man hours during weeks where the number of thermal complaints peaks, and between the extra available man hours during weeks where there is a minimal number of thermal complaints. The gap between the two cases, in terms of thermal complaints, is high, and thus this show the importance of planning for a better estimate of the number of man hours to handle thermal complaints based on the precited number of thermal complaints for the upcoming week.

The above analysis shows similar expected results from one case to another while varying the number of technicians hired per month and the percentage of thermal complaints received out of all the received complaints. This also shows that the concept of the problem holds in each analyzed scenario that tackles certain assumptions and uncertainties. Thus, the developed base case is feasible to show the significance of this research work to staff in accordance with the predicted number of thermal complaints for the upcoming week.

E. The Need for Proper Data Management

The suggested data driven methodology and the significance of this research work in terms of understanding building occupant complaints, forecasting thermal complaints, and staffing accordingly are but a solid proof on the importance of the data itself, thus stressing on the need to properly collect, store, and sustain the occupant complaint data. More specifically, the main driver of the proposed methodology was the textual description of the complaints issued by occupants and recorded by operators in the call center along with the corresponding date. However, the unstructured format of the dataset and several encountered data-related issues, highlighted throughout the sections of chapter 4, are at the root of limiting proper, accurate, and further use of the occupant complaint data. Such issues narrow down to:

- Unavailable records for several required data entries (for example not recording the name of the technicians that were responsible to handle the complaint)
- Inaccurate records of some data entries (for example using the term AC to describe a heater related complaint)
- Inconsistent records of some data entries (for example writing the name of the same technician differently between different complains, and using for example more than one terms to describe an AC- related complaint: AC, air conditioner, and HVAC)

Looking at the bigger picture while considering the obtained maintenance logs, no further use was possible for both the BMS notes and the routine maintenance checkups due to several unavailable entries and having the same recording being copied from one Excel cell to another. Had such data been available, it could have further assisted facility managers in making strategic decisions that eventually bring back benefits in term of ensuring occupants satisfaction, enhancing the building's performance, and mitigating unnecessary costs. The facility management literature is rich with such examples.

Facility managers are recommended to sustain proper data management strategies given the importance of data driven decision making, and thus the need for reliable data. Such strategies ought to define a road map on how to collect, store, maintain, and share the obtained data. The selected data management strategy is case-specific depending on the user and on the corresponding business objectives and strategic goals. Insights about such modern systems, and those that specifically tackle maintenance management coupled with modern IT solutions, could be further assessed to select the optimum system for better management of the maintenance related data used in this research work.

CHAPTER 5

CONCLUSIONS AND FUTURE WORKS

Occupant complaints in buildings reflect an underperforming building and unsatisfied occupants. The facility management's role in buildings is to ensure those complaints are addressed and resolved to ensure that the occupants are satisfied and that the building is performing properly. More specifically, they aim to plan the for required staffing resources for this task of handling occupant complaints. As such, this work has presented a sound decision making tool that facility managers can adopt when planning these staffing resources to handle thermal complaints in particular. By forecasting the number of these complaints for the upcoming week, facility managers can plan their staffing resources accordingly.

The proposed methodology of this thesis was divided into three main parts. The first part included a theoretical selection of the time series forecasting model to predict the number of thermal complaints. The second part presented a ML- multistep generic framework to analyze building occupant complaints and forecast the number of thermal complaints. It adopted text mining techniques to transform highly unstructured occupant complaints into a structured format that could be used in further analysis. It also adopted the development of neural networks, MLPs in particular, for the forecasting purposes that undergo a systematic evaluation process before being selected. The third part of the methodology then suggested a traditional statistical model for time series forecasting that was used as a benchmark for the developed ML- based forecasting.

The proposed methodology was tested on a selected case study of highly unstructured real-world building occupant complaints data. Text mining results showed that thermal complaints are in fact among the most common types and thus require attention form facility managers. Two ML-models were developed: an AC MLP model to forecast the number of AC-related complaints in the upcoming week and a heater MLP model to forecast the number of heater-related complaints in the upcoming week. These two models were benchmarked against the state-of-the-art traditional statistical model ARIMA, where one was also built to forecast AC complaints and another to forecast heater complaints. Comparing the model errors showed that the MLP models outperform the ARIMA models. These models were tested on a test set composed of 16 weeks. The AC MLP model had a lower RMSE (0.878) than the AC ARIMA model (1.117) and was not overfitting. As for the heater MLP model it had a RMSE (0.265) slightly lower than that of the heater MLP model (0.372), however the heater MLP model was highly overfitting. Thus, the MLP models were selected for the forecasting purpose. Also, it is evident how the selected forecasting models could assist facility managers to plan accordingly for the staffing resources required to handle these complaints.

The success of obtaining a well performing ML-based forecasting model for the selected case study provides evidence on the applicability of the developed ML- based generic framework. The AC and heater MLP models are specific to the selected case study of forecasting thermal complaints for the upcoming week for the 16-building residential complex with its corresponding weather conditions and properties. Had there been another residential complex with similar weather conditions and similar trend and seasonal variation of thermal complaints, these developed MLP models could have been used in this case as well. However, the developed ML- based generic framework provides facility mangers with a road map on how they can developed their own AC and heater MLP models to forecast the number of thermal complaints based on the recorded occupant complaint calls in a text format. This gives facility managers the flexibility to custom tailor the models to reflect the actual case as per the data availability and the ultimate use of the forecast.

It is worth noting that upon working with the data obtained from the selected case study, constant occupancy levels were assumed constant because no further information was available from the data owners. If such data was available, it could have been incorporated in the forecasting models. Moreover, one main limitation for this specific case study is that the complaints were issued by the occupants and recoded by operators where both do not have the technical expertise to further describe the complaint using technical terms and including more details on the problem that could have been used in the analysis.

Future works aim at further investigating the occupant complaints to assist facility managers in making strategic decisions other than these related to maintenance resource staffing. Moreover, other predictors could be incorporated in the prediction model to better represent the actual situation such as the occupancy levels along with the performance of the building.

APPENDIX A

In order to select reasonable numbers for the options of the available technicians per month, the maintenance log sheets for occupant complaints were visited again. As mentioned previously, the name of the technician responsible to handle each complaint was not filled consistently throughout the log sheets. So, in order to get a better idea onto the number of technicians hired per month, the distinct names of technicians were counted in each month for years one and three. The numbers of distinct names obtained do not necessarily reflect the numbers actually available because there might be technicians who are hired and handling thermal complaints, but their names were never mentioned in the log sheets due to poor data entry strategies. Figure 45 shows the number of technicians available in each month for the years one and three. This, although not complete due to lack of available data, provides an idea about the range of the number of available technicians, their variability from one month to another, and from one year to another.

Investigating the number of technicians, it varies from 20 to 25 in year three, and from 19 to 23 in year one during the last four months. Looking over the span of the two years, the number can get as low as 13 and as high as 26. This explains the selection of the options for the available number of technicians per month: 16, 21, and 23.

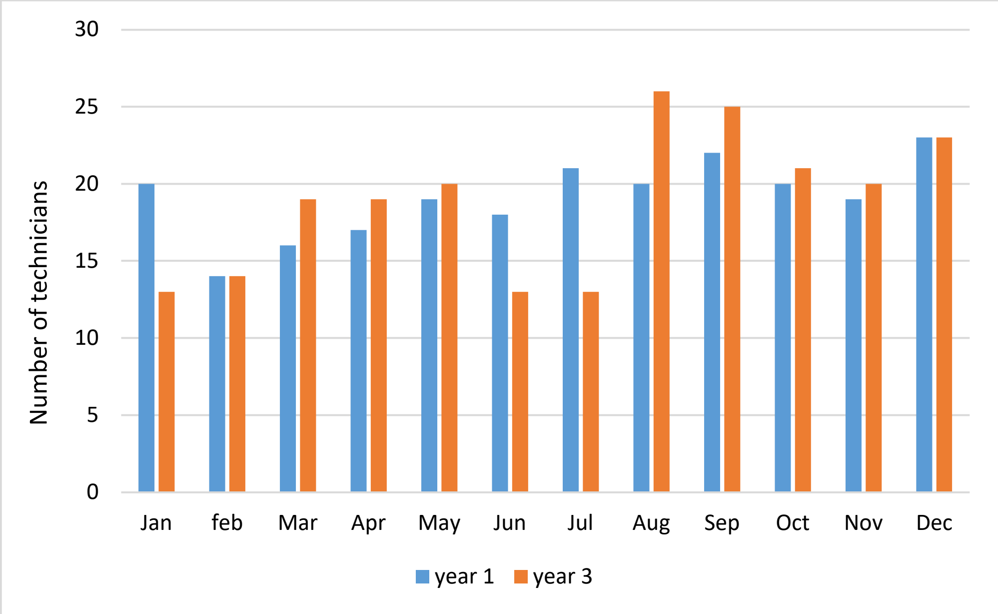

Figure 45 The variation of the number of technicians per month

Moreover, for each selected option, it was ensured that the number of available technicians could be allocated to ensure a 24-hour service, 7 days a week, while each technician works about 40 hours per week. Table 10 shows the assumptions taken for each option.

Table 10 Assumptions for each staffing option

	option 1	**option 2**	**option 3**
Number of technicians per month	16	21	25
Number of shifts per day	4	3	3
Number of hours per shift	6	8	8
Number of technicians per shift	4	5	6
Number of shifts per technician	7	5	5
Number of hours per technician per week	42	40	40

Table 11 shows how the technicians were assumed to be allocated in each day of the week (D), and in each shift (s). Each number in the cells represents a technician's ID.

Table 11 Allocation of technicians per day per shift

		D1	D2	D3	D4	D5	D6	D7
Option 1	**s1**	1,2,3,4	1,2,3,4	1,2,3,4	1,2,3,4	1,2,3,4	1,2,3,4	1,2,3,4
	s2	5,6,7,8	5,6,7,8	5,6,7,8	5,6,7, 8	5,6,7,8	5,6,7, 8	5,6,7,8
	s3	9,10,11, 12	9,10,11, 12	9,10,11, 12	9,10,11, 12	9,10,11, 12	9,10,11, 12	9,10,11, 12
	s4	13,14,15, 16	13,14,15, 16	13,14,15, 16	13,14,15, 16	13,14,15, 16	13,14,15, 16	13,14,15, 16
Option 2		**D1**	**D2**	**D3**	**D4**	**D5**	**D6**	**D7**
	s1	1,2,3, 4,5	16,17,18, 19,20	10,11,12, 13,14	4,5,6, 7,8	19,20,21, 1,2,	13,14,15, 16,17	7,8,9, 10,11
	s2	6,7,8, 9,10	21,1,2, 3,4	15,16,17, 18,19	9,10,11, 12,13	3,4,5, 6,7	18,19,20, 21,1	12,13,14, 15,16
	s3	11,12,13, 14,15	5,6,7,8,9	20,21,1, 2,3	14,15,16, 17,18	8,9,10, 11,12	2,3,4,5, 6	17,18,19, 20,21
Option 3		**D1**	**D2**	**D3**	**D4**	**D5**	**D6**	**D7**
	s1	1,2,3, 4,5,6	19,20,21, 22,23,24	12,13,14, 15,16,17	5,6,7, 8,9,10	23,24,25, 1,2,3	16,17,18, 19,20,21	8,9,10, 11,12,13
	s2	7,8,9,10, 11,12	25,1,2, 3,4,5	18,19,20, 21,22,23	11,12,13, 14,15,16	4,5,6, 7,8,9	22,23,24, 25,1,2	14,15,16, 17,18,19
	s3	13,14,15, 16,17,18	6,7,8, 9,10,11	24,25,1, 2,3,4	17,18,19, 20,21,22	10,11,12, 13,14,15	3,4,5, 5,6,7	20,21,22, 23,24,25

As for the percentage of thermal complaints out of all the received complaints, three options were investigated: 6.95%, 13%, and 23.9%. The number of the total complaints was obtained from the maintenance logs for each month of the three years.

As for the number of thermal complaints per month, this was obtained after conducting the text mining process earlier. As such, figure 46 shows the percentage of thermal complaints out of all complaints per month for the three years.

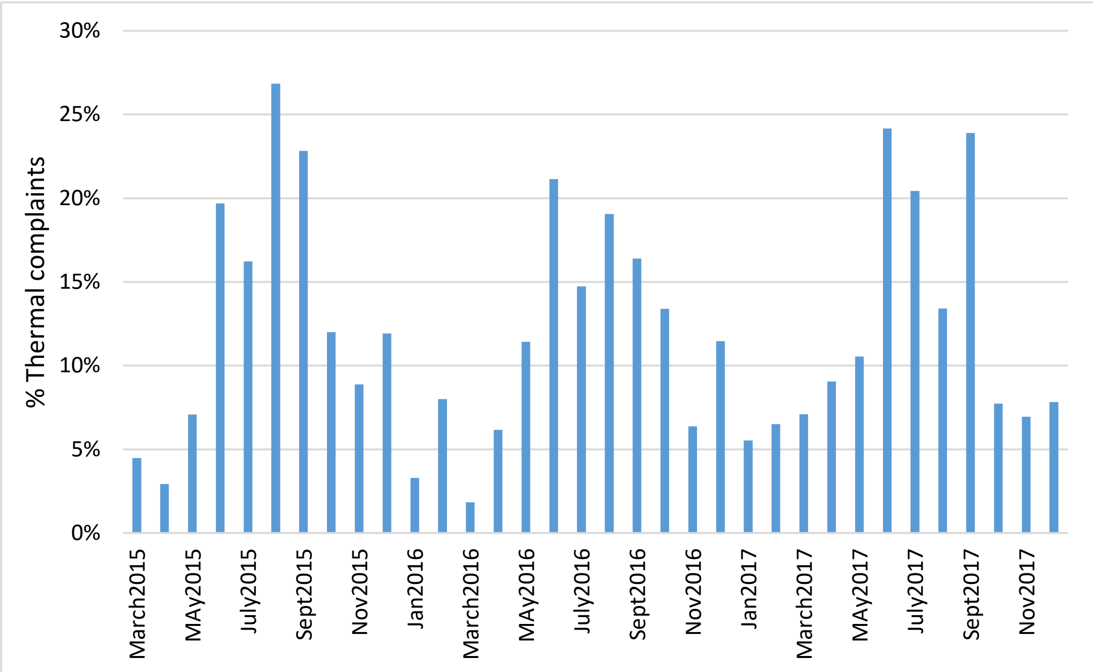

Figure 46 The variation of the percentage of thermal complaints per month

It can be noted that there is a repetitive pattern in each year, with the percentage of thermal complaints peaking mid-year during summer months and decreasing at the beginning and end of each year. For the last four months of 2017 in particular, the percentage of thermal complaints varies between 6.95% to 23.9%. This explains the selection of these two options in the developed scenarios. As for the 13% option, it was selected by calculating the percentage of thermal complaints out of all complaints for each of the three years, and then averaging these three

REFERENCES

Achelis, S. B. (2001). *Technical Analysis from A to Z*. McGraw Hill New York.

Aggarwa, C., & Zhai, C. (2012). *Mining Text Data Textbook*.

Akcamete, A., & Akinci, B. (2004). Potential utilization of building information models for planning maintenance activities. *Proceedings of the International Conference on Computing in Civil and Building Engineering*. Nottingham University Press.

Alexander, K. (2003). A strategy for facilities management. *Facilities*, *21*(11), 269–274. https://doi.org/10.1108/02632779410070200

ASHRAE, A. S. of H. R. A. A. C. E. (2010). *Thermal environmental conditions for human occupancy*. ASHRA.

Babu, C. N., & Reddy, B. E. (2014). A moving-average filter based hybrid ARIMA–ANN model for forecasting time series data. *Applied Soft Computing*, *23*, 27–38.

Battiti, R. (1992). First-and second-order methods for learning: between steepest descent and Newton's method. *Neural Computation*, *4*(2), 141–166.

Behrang, M. A., Assareh, E., Ghanbarzadeh, A., & Noghrehabadi, A. R. (2010). The potential of different artificial neural network (ANN) techniques in daily global solar radiation modeling based on meteorological data. *Solar Energy*, *84*(8), 1468–1480.

Bergstra, J., Yamins, D., & Cox, D. D. (2013). *Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures*.

Bortolini, R., & Forcada, N. (2019). Analysis of building maintenance requests using a text mining approach : building services evaluation services evaluation. *Building Research & Information*, 1–11. https://doi.org/10.1080/09613218.2019.1609291

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2011). *Time series analysis: forecasting and control* (Vol. 734). John Wiley & Sons.

Box, G. E. P., & Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association*, *70*(349), 70–79.

Cadenas, E., Rivera, W., Campos-amezcua, R., & Heard, C. (2016). Wind speed prediction using a univariate ARIMA model and a multivariate NARX Model. *Energies*, *9*(2).

Elazouni, A. M., & Shaikh, A. (2008). Simulation model for determining maintenance staffing in higher education facilities. *JOURNAL OF PERFORMANCE OF CONSTRUCTED FACILITIES*, *22*(1), 54–61.

Federspiel, C. C. (2000). Predicting the frequency and cost of hot and cold complaints in buildings. *HVAC and R Research*, *6*(4), 289–305.

Federspiel, C. (1998). Statistical analysis of unsolicited thermal sensation complaints in commercial buildings / Discussion. In Anon (Ed.), *Proceedings of the 1998 ASHRAE Winter Meeting. Part 1 (of 2)* (pp. 912–923). San Francisco, CA, USA: ASHRAE,Atlanta, GA, United States.

Federspiel, Clifford, Rodney, M., & Hannah, Y. (2003). *Thermal comfort models and complaint frequencies*.

Gavin, H. (2011). The Levenberg-Marquardt method for nonlinear least squares curvefitting problems. *Department of Civil and Environmental Engineering, Duke University*, 1–15.

Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media.

Goh, C., & Law, R. (2002). Modeling and forecasting tourism demand for arrivals with stochastic nonstationary seasonality and intervention. *Tourism Management*, *23*(5), 499–510.

Goins, J., & Moezzi, M. (2013). Linking occupant complaints to building performance. *Building Research & Information*, *41*(3), 361–372.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.

Guiñón, J. L., Ortega, E., García-Antón, J., & Pérez-Herranz, V. (2007). Moving average and Savitzki-Golay smoothing filters using Mathcad. *Papers ICEE*, *2007*.

Gunay, H. B., Shen, W., Newsham, G., & Ashouri, A. (2018). Modelling and analysis of unsolicited temperature setpoint change requests in o ffice buildings. *Building and Environment*, *133*(December 2017), 203–212.

Gunay, H. B., Shen, W., & Yang, C. (2019). Text-mining building maintenance work orders for component fault frequency. *Building Research and Information*, *47*(5), 518–533.

Hamzaçebi, C. (2008). Improving artificial neural networks' performance in seasonal time series forecasting. *Information Sciences*, *178*(23), 4550–4559.

Higgins, L. R., Mobley, R. K., & Wikoff, D. J. (2008). *Maintenance engineering* *handbook* (7th ed.). New York: McGraw Hill, McGraw-Hill, McGraw-Hill Professional Publishing.

Horner, R. M. W., El-Haram, M. A., & Munns, A. K. (1997). Building maintenance strategy: a new management approach. *Journal of Quality in Maintenance Engineering*, *3*(4), 273–280.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.

International Facility Management Association. (2009). *Temperature wars : savings vs. comfort*.

Ismail, N., Ali, S. N. A. M., Othman, N. A. A., & Jaffar, N. (2017). Occupants' satisfaction on building maintenance of government quarters. *AIP Conference Proceedings*, *1891*, 020066.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). Springer.

Kelleher, J. D., Mac Namee, B., & Aoife, D. (2015). *Fundementals of machine learning for predictive data analytics*. The MIT Press.

Kral, P., & Bartosova, V. (2016). Optimal resource allocation in facility management. *Proceedings of the 20th International Scientific Conference. Transport Means*, 1036–1039.

Lee, H. H. Y., & Scott, D. (2009). Overview of maintenance strategy, acceptable maintenance standard and resources from a building maintenance operation perspective. *Journal of Building Appraisal*, *4*(4), 269–278.

Lee, S. Y. (2002). A facility manager's role to provide quality service through reflecting user needs. *Proceedings of the CIB W*, *70*, 2002.

Madureira, S., Flores-Colen, I., de Brito, J., & Pereira, C. (2017). Maintenance planning of facades in current buildings. *Construction and Building Materials*, *147*, 790–802.

Márquez, A. C., León, P., Fernández, J. F., Márquez, C. P., & Campos, M. L. (2009). The maintenance management framework: a practical view to maintenance management. *Journal of Quality in Maintenance Engineering*, *15*(2), 167–178.

Mcarthur, J. J., Shahbazi, N., Fok, R., Raghubar, C., & Bortoluzzi, B. (2018). Machine learning and BIM visualization for maintenance issue classi fi cation and enhanced data collection. *Advanced Engineering Informatics*, *38*(June), 101–112.

Mohammed, L. B., Hamdan, M. A., Abdelhafez, E. A., & Shaheen, W. (2013). Hourly solar radiation prediction based on nonlinear autoregressive exogenous (Narx) neural network. *JJMIE*, *7*(1).

Nawi, M. N. M., Baharum, F., Ibrahim, S. H., & Riazi, S. R. M. (2017). A review study of maintenance and management issues in Malaysian commercial building towards sustainable future practice. *AIP Conference Proceedings*, *1891*, 020100. https://doi.org/10.1063/1.5005433

Nutt, B. (1999). Linking FM practice and research. *Facilities*, *17*(1/2), 11–17.

Nwankpa, C., Ijomah, W., Gachagan, A., & Marshall, S. (2018). Activation functions: Comparison of trends in practice and research for deep learning. *ArXiv Preprint ArXiv:1811.03378*.

Panos, L., & Christof, E. (2016). Machine learning. In *Vector Consulting Services*. IEEE Computer Society.

Ruder, S. (2016). An overview of gradient descent optimization algorithms. *ArXiv Preprint ArXiv:1609.04747*.

Shin, H., Lee, H.-S., Park, M., & Lee, J. G. (2018). Facility management process of an office building. *Journal of Infrastructure Systems*, *24*(3), 04018017.

Shumway, R. H., & Stoffer, D. S. (2017). *Time series analysis and its applications: with R examples*. Springer.

Sipos, R., Fradkin, D., Moerchen, F., & Wang, Z. (2014). Log-based predictive maintenance. *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1867–1876. ACM.

Smithing, N. (1999). *Supervised Learning in Feedforward Artificial Neural Networks*. Cambridge, MA: MIT Press.

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using multivariate statistics* (Vol. 5). Pearson Boston, MA.

Wahab, Y., Samad, A., & Basari, H. (2013). Building maintenance management preliminary finding of a case study in Icym. *Middle-East Journal of Scientific Research*, *17*(9), 1260–1268. https://doi.org/10.5829/idosi.mejsr.2013.17.09.12286

Weinstein, L. B. (1997). *Decision support for the integration of maintenance activities with front-end production planning activities.*

Witten, I., & Frank, E. (2005). *Data mining practical machine learning tools and techniques*. Morgan Kaufmann Publishers.

Yu, S., Zhu, K., & Zhang, X. (2012). Energy demand projection of China using a path-coefficient analysis and PSO–GA approach. *Energy Conversion and Management*, *53*(1), 142–153.

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, *50*, 159–175.

Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks:: The state of the art. *International Journal of Forecasting*, *14*(1), 35–62.