Leiden University

ICT in Business

Smart energy management as a means

towards improved energy efficiency

Name:	Dylan te Lindert
Date:	30/08/2017
1st supervisor:	Dr. Knobbe, A.J.
2nd supervisor:	Dr. Pinho Rebelo de Sá, C.F.
3rd supervisor:	Dr. Soares, C.M.

MASTER'S THESIS

Leiden Institute of Advanced Computer Science (LIACS)
Leiden University
Niels Bohrweg 1
2333 CA Leiden
The Netherlands

Abstract

Refrigeration is recognized as being one of the major energy consumers in supermarkets. The costs associated with refrigerator equipment often represent more than half of the total energy costs. This presents a good motivation for running these systems efficiently. Therefore, a Portuguese Retailer challenged us to define baselines, so they can examine the energy performance of their refrigeration departments. In this study, we investigate different ways to construct a reference behavior, which can serve as a baseline for judging the performance of energy consumption. We used 3 distinct learning models: Multiple Linear Regression, Random Forests, and Artificial Neural Networks. During our experiments we used a variation of the sliding window method in combination with learning curves. We applied this approach on five different supermarkets, owned by the Retailer, across Portugal. In the main section of this study, we describe solutions for the business questions laid down by the Retailer. We are able to create baselines using off-the-shelf data mining techniques. Moreover, we found a way to create them based on short term historical data. With our findings, the Retailer is able to create baselines to examine the energy performance. Furthermore, these baselines enable the Retailer to compare the energy savings with the investments made for the measure. This way, the Retailer is able to determine if energy efficiency measures are effective. We believe that our research will serve as a base for future studies, for which we provide interesting directions.

Acknowledgements

First of all, I want to take this occasion to display my gratitude to a number of people. I would like to express my special thankfulness to my supervisor Claudio for his close supervision, knowledge, and friendship during this challenging project. I have learned a lot from your expertise and insightful discussions. Next, I would also like to thank Carlos and Arno for their intelligent comments and instructions plus enabling me to execute this project at INESC-TEC Porto in Portugal. I am pleased to all of those with whom I have had the joy to work during this project at the Centro de Engenharia de Sistemas Empresariais (CESE), and I want to especially thank Elisa for her help and assistance. Furthermore, I want to thank my close friends Hugo, Kaylesh, and Sebastian for their support from the beginning of the research. Their discussions and brainstorm sessions have helped me a lot during this period. Finally, I need to express my very profound appreciation to Cristiana for believing in me and providing me with endless support and encouragement so I could conduct this research as best as possible.

Contents

Introduction

Portugal has been developing and transforming their energy policy since 2014, with an emphasis on energy efficiency, towards a more sustainable economy. In 2016, Portugal created the National Renewable Energy Action Plan (NREAP) which established an energy saving target of 25% by 2020 [1]. The program intends to generate awareness about the energy usage of companies. Moreover, it aims to stimulate businesses to perform energy management in a more sustainable way. To achieve this goal, the Portuguese government, in cooperation with the EU, set up a fund named 'Portugal 2020 Strategy'. This fund invests in research in order to create and transfer the scientific outputs to production sites [2]. Besides research, the Portuguese government is also actively supporting and encouraging companies to develop strategies aligned with their sustainability. Since the start of the fund, in 2014, more than €1 billion of applications have been approved [3].

Research shows that when the government takes the role of a driver behind energy savings, by providing financial incentives, this encourages businesses the most to investigate better ways of utilizing energy [4]. It is beneficial to have a solid energy management program since, first of all, this is the core of saving energy in an organization. From a business point of view, greater energy efficiency is of importance because it provides a number of direct economic benefits such as increased competitiveness, e.g., decreasing energy consumption reduces production costs, and higher productivity, e.g., by decreasing lead times [5]. Besides direct benefits, the indirect economic consequences are, e.g., an improved image, contributing to climate protection or resource conservation [5]. Additional research shows that there is a major untapped potential for improved industrial energy efficiency [6, 7]. There are several explanations that reveal the causes behind this efficiency gap. For example, when inadequate information is available about the results of investments in energy efficient products, this can limit companies to invest in them. Furthermore, liquidity constraints, when upfront costs are high and access to credit is limited, can prevent companies to make energy efficient improvements. Hence, Energy management should focus on utilizing as much of this potential as possible to create major benefits by addressing factors that prevent businesses to make investments.

One particular business area where energy management is particularly important is retail [8]. Retailers operate in an industry that is characterized as a competitive and low-margin [9]. Therefore, these companies are continually searching for ways to become more efficient and, thus, more competitive. Supermarkets, e.g., sell mainly food products and a substantial amount of these products need to be stored in fridges and freezers to slow down the deterioration of the food. Hence, refrigeration plays a significant role in retaining these items at a planned temperature. These fridges and freezers utilize energy day and night and, because of that, they are the largest energy consumers in supermarkets [10]. This is, therefore, a good place to search for efficiency gains.

1.1 Problem Statement

Studies on energy efficiency have been conducted for residential [11], historical [12], and state owned buildings [13]. Commercial buildings, like supermarkets, have received less attention because the availability of data is limited. Supermarkets, in particular, are one of the most energy intensive buildings and that is why energy management is vital in this business [8].

The last couple of years there has been a growing focus on energy efficiency [1], mostly because, for supermarkets, energy costs are the second-highest operating cost after labor [14]. Compared to other buildings, supermarkets consume proportionately more energy [10]. This is caused by the refrigeration which conserves the chilled and frozen products [10]. In this type of stores, the costs associated with refrigerator equipment represent more than 50% of the total energy costs [8, 9, 15, 16]. This outlines the importance of operating the system at its optimum performance level by reducing the associated energy costs. Therefore, it is essential to implement energy management policies to find potential improvement opportunities, because, in this area, a small improvement could yield significant results.

As soon as energy management policies are in place, one can analyze the energy consumption or the effectiveness of improvement measures, by monitoring energy usage over time. In order to evaluate the performance of a store, the actual energy usage should be compared to a *reference behavior*. There are two ways to establish this reference behavior [17]. The first is called *energy benchmarking* which consist of the comparison of the energy usage in a store with the current or past performance of another store with a similar configuration. The energy usage of this 'similar' store will be transferred and appear as the reference behavior. The second method is called *energy baselining*, in which the reference behavior is based on the previous, or optimal energy performance, of the current store.

Since supermarkets tend to have different configurations, it becomes challenging to find a store with similar characteristics to form a benchmark. Therefore, we will focus on the second method and create our own baseline which will be used as reference behavior. This baseline will be based on the previous energy consumption of the store.

1.2 Research Questions

A giant Portuguese retailer, hereafter referred to as 'The Retailer', challenged us to define baselines for the energy consumption of the refrigerator sections in their stores. To establish this, we decided to use standard data mining techniques. The approach will be tested with data derived from several stores that belong to the Retailer. The Retailer laid down several practical challenges. These are translated into research questions and used to answer the main question of this study.

Main question

- Can off-the-shelf data science technologies be used to create an energy baseline that supports improved energy management?

Multiple sessions with the Retailer have been organized to gather business requirements. From these sessions, three issues emerged, each of these will be presented here.

Retailers frequently open new stores. When a new store is opened, no data has been collected about the energy performance of *this* specific store. To create a baseline as soon as possible, it is essential to know how many days it takes to collect sufficient data. Therefore, we have to study the minimum amount of days needed to create a solid baseline. This information is also suitable for updating the baseline when the configuration of the store changes, e.g., due to an upgrade of the refrigeration equipment.

It is important to determine how reliable the baseline is and if it needs updating. The assumption is that the prediction error will grow over time, and therefore will behave differently for short and long term predictions. We will have to test how the reliability of the predictions develop over time. With this information, the life-cycle of a model can be determined, which defines how often the model needs to be updated.

When a new energy saving policy is implemented, the Retailer wants to estimate how much energy is saved. Therefore, a model has to be developed which is able to make long term predictions based on the old configuration of the store. With this baseline, the Retailer can see what the estimated energy consumption would be if they did not change the layout. By comparing this baseline with the observed energy consumption or the new baseline, the difference can be estimated. We will examine the behavior of the model for long term predictions because the Retailer needs to know for how long he can estimate, with a reasonable accuracy, the energy gains from a certain energy policy.

Business questions

1. When I open a new store, how many days do I need before I am able to train my model for a baseline, and which model should I use?
2. For how long does this baseline stay reliable?
3. Can we estimate the energy gains after a new energy efficiency measure is implemented?

Research approach

1. Study the prediction accuracy of a model based on the amount of training data.
2. Discover how often the model needs to be updated.
3. Empirically test the model behavior for long term predictions.

1.3 Organizational Relevance

The Retailer can benefit from this study in several ways

- The Retailer is currently monitoring their day-to-day energy consumption with an advanced energy monitoring system. However, as there is no reference behavior, they can not assess the performance of their energy policies. By creating an energy baseline they will be able to analyze their performance.
- This baseline helps to identify potential improvements because, in order to calculate the energy consumption, it is often necessary to identify variables that affect this consumption. The Retailer will get insight into factors that influence the energy usage and this information can be useful for decision-makers when deciding to invest in measures that lead to improved energy efficiency.

- Improved energy efficiency reduces risk exposure to energy price fluctuations. Being energy efficient can decrease the electricity costs. Therefore, it becomes an interesting subject to study for a commercial company. When the energy consumption can be estimated in advance, this information can be used in the Retailers‘ energy price strategy. It can be applied to determine if a fixed or variable price would suit best, which can be useful in energy price negotiations.

1.4 Thesis Overview

This thesis consists of 5 chapters, following this introductory chapter.

Chapter 2, Literature Review on Energy Management, presents a literature review based on Energy Management. It describes areas related to EM and presents several examples.

Chapter 3, Materials & Methods, contains an extensive overview of the learning models and methods that are used for energy baselining.

Chapter 4, Experimental Setup, outlines the approach used in this study. Next, it includes an overview of the used data and explains how it is prepared.

Chapter 5, Results, displays the results obtained from the executed experiments. It answers the research questions and also includes a discussion of the results.

Finally, in Chapter 6, Conclusion, each research question will be individually answered. Further, it also consists the main conclusion of this thesis, and directions for future work are proposed.

Literature Review on Energy Management

In this chapter, we make an extensive overview on Energy Management and describe some best practices.

Energy Management (EM) has been the subject of numerous studies throughout the years, and, because the field of EM is wide, it can be described in many different ways [18]. To define the field of focus, it is important to start with a description of Energy Management that is going to be used in this study. The following description captures the essence of this project; *"Energy management can be seen as a means to overcome barriers to improved energy efficiency"* [4].

The purpose of EM is to search for improved strategies to consume energy in a more efficient way. From a business point of view, greater energy efficiency is of importance because it provides a number of direct, and indirect, economic benefits. Examples of direct economic benefits can be increased competitiveness or higher productivity, where indirect economic benefits can be an improved image, climate protection or resource conservation [5]. Other research shows that there is a major untapped potential for improved industrial energy efficiency [6, 7]. Energy management is focused on utilizing as much of this potential as possible to create benefits for the company and their environment.

The way companies perform EM has been evolving throughout the years but the basic framework of the EM process is still the same. In 2001, [19] described successful energy management as a continuous process that consists of three steps, Figure 2.1.

1. This process starts with measurement, in this step, the data about the energy consumption is collected, e.g., by directly measuring the energy usage.

2. During the analysis, data is collected, prepared and presented to a decision maker that will evaluate the information and decide what action to take.

3. After analysis, action might be taken, e.g., implementing new energy policies. After this step, it is important to start measuring again to determine the effect.

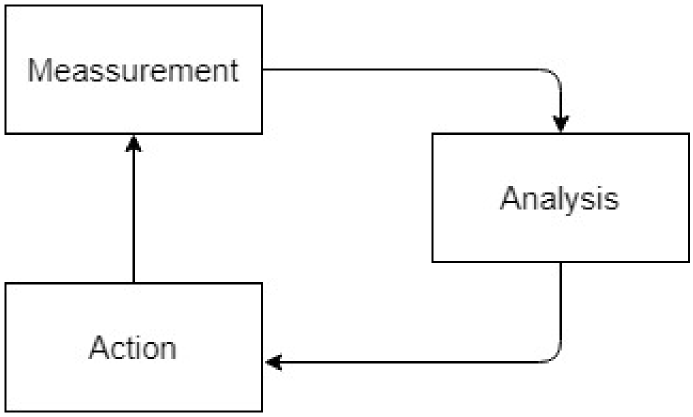

FIGURE 2.1: Example of the Energy Management Process

The characteristics of the food-retail industry, such as fierce competition and low margins, makes retailers continually search for ways to operate more efficiently [9].Since energy costs are the second highest costs for a retailer [14], a decent energy management process is vital for improving efficiency [18]. Electricity costs associated with refrigeration accounts for a large part of the operating costs because these machines are continually utilizing energy, day and night. In an office building, one could simply turn off the machines, air conditioning or lights, when no one is around but it is not possible to simply shut down the refrigerators in supermarkets during the night. Refrigerators are used to retain items on a planned temperature to keep them *sellable*. Keeping products sellable means that they can be sold over a longer period of time and this improves the store economics [14]. Thus, savings in energy costs can be undone by the depreciation of products. This implies that a trade-off has to be found between the length of time a product is sellable and the energy consumption. For retailers and researchers, this presents an attractive opportunity to explore different ways to obtain energy efficiency gains.

A lot of studies found in Energy management literature is focused on aspects to consume energy in a more efficient manner [18]. The scope and results of the studies performed are quite diverse. Several studies discovered and developed best practices [14, 20, 21]. Other studies developed simulation models to estimate the energy consumption or savings [22]. Another part of the studies identified factors that influence the usage of energy [22, 23]. Finally, there is a part that tries to predict energy consumption using data mining methods [24].

2.1 Best Practices

As described before, a great number of studies have developed best practices to reduce energy consumption in supermarkets. These 'proven ways of success' have a lot of different applications. One study [20], focused on pre-packed food that is stored in freezers and fridges. It proposed best practices on how the packaging of these products can reduce energy usage. Another study [14], developed best practices based on the use of low energy doors, equipped with anti-sweat heaters, that consume one-third less energy. The same study also suggests that implementing an energy monitoring system is key to spot new areas, of the supermarket, to improve energy efficiency. To conclude, another study [22], proposes several measures to optimize efficiency, like installing specialized compressors and heat exchangers.

Generally, best practices are documented well which makes it possible to implement them. This enables others to test if they can obtain similar results. When you operate multiple stores, it is advisable to first test a best practice in a single store before executing them in others. Given the circumstances, it can be that a certain best practice is not profitable, possible of appropriate. When a best practice proves to be a success, this will lead to replication. It also simplifies decisions to reserve funding for the required investments in other stores.

2.2 Simulation Models

A simulation model is a representation of the construction and working of a system of interest [25]. The purpose of simulation is to enable decision makers to predict the effect of changes in the physical environment, e.g., to evaluate potential benefits of best practices. Therefore, a virtual model needs to be developed where all components and aspects of the system require examination. One study [24], used six different simulation models to validate if future changes, e.g., in technology, had a positive impact on energy efficiency. Each of these models consisted of a diversified set of variables, collected by sensors throughout the store. Another study [10], used the supermarket energy simulation software from Energyplus to examine opportunities for energy efficiency in a specific case study.

Simulation models can be an adequate and suitable way to quickly evaluate the impact of measures. Off-the-shelf simulation models are available and can be immediately used for this practice. However, to have a simulation model that accurately reflects the impacts in the real world, it is essential to have a wide range of information available about the specific store layout and configuration. Since it can be hard to find supermarkets with similar layouts, it is a time intensive process to create a simulation model that is able to precisely assess the effect of measures for each independent store. Moreover, since supermarkets continually implement new equipment and best practices, the models need frequent updating in order to reflect the reality.

2.3 Influential Factors

Collecting data and studying their relevance to the energy consumption is a way of smart energy management. One study [26], names a couple of challenges that have to be addressed in order to realize the full potential out of the data. a) how to effectively collect, store and manage energy data, b) how to efficiently analyze and mine energy data, c) how to use energy data to support effective and efficient decision making, d) how to create insights and get understanding from energy data and e) how to effectively prevent risks and protect privacy while utilizing energy data.

Several studies [27], identified and studied driving factors of energy use in buildings. It is valuable information to know which factors influence the most the consumption of energy if you want to work towards a more efficient way of using energy. The International Energy Agency's Energy in the Buildings and Communities Program (IEA EBC) [27], indicated that there are six main factors that impact energy use in buildings, e.g., climate, building design and occupant behavior. Another study [28], showed that other factors like the indoor humidity, occupancy behavior, operating hours, or the building size can be influential factors.

The influence of factors is different in each situation [23, 24], therefore, the impact on energy consumption is also a topic of study. One study [23], reviewed literature about the relationship between climate change and energy consumption of consumers and technology. They conclude that the changing climate will influence the technological change and that some technologies that are not economic today may become so in the future. Another study [24], describes how the energy consumption of a supermarket is expected to change based on the indoor- and outdoor temperature and relative humidity.

Food quality has also been addressed in energy management since this is essential for food retail stores like supermarkets. In [14], the authors researched the trade-off between food quality loss and the energy consumption. In a second study [21], traditional defrost schemes were used and optimized with food quality as a new variable. Another study [10], used operational data like the store layout, opening hours and the type of products as variables to determine the optimal energy consumption.

Several studies showed that a number of people that visit a store, occupant behavior also has a significant impact on the energy usage. However, one of the biggest challenges associated with the collection of this data is the lack of a standardized format and the regulation of privacy issues [27]. Therefore, supermarkets that want to study occupant behavior, need to design a data collection process that protects peoples privacy. This makes the usefulness of this data questionable, e.g., people need to be aware that they are monitored which can result in them altering their behavior. Because of these issues, it happens that this data is typically unavailable for analysis.

2.4 Energy Prediction

Numerous studies focused on energy prediction because forecasting the energy consumption is an important component of any energy management system [29]. Traditional data mining techniques like Multiple Linear Regression (MLR), Artificial Neural Networks (ANN) or ensemble methods like Random Forests (RF) have been used for this.

In New Zealand [30], researchers used MLR to calculate the optimal energy usage level for office buildings, based on monthly outside temperatures and numbers of full-time employees. With this knowledge, they could build an energy monitoring and auditing system for the optimization and reduction of energy consumption. In the UK [24], researchers used an MLR to forecast the expected effect of climate change on the energy consumption of a specific supermarket. They estimated that, until 2040, the gas consumption will increase 28%, which is more, compared to the electricity usage, which will increase 5,5%.

In the UK, most supermarkets negotiate energy prices and, when they exceed their predicted demand, they have to pay a penalty.Therefore, their ability to accurately predict energy consumption will facilitate their negotiations on electricity tariffs with suppliers. One supermarket in the UK used ANN's to analyze the Stores Total Electricity Consumption as well as their individual systems, such as Refrigeration and Lighting [15]. For each of these systems, they developed a model to provide an energy baseline. This baseline is used for performance monitoring which is vital to ensure systems to perform adequately and guarantee operating costs and energy use are kept to a minimum. Finally, ANN's have been used for energy prediction with the final goal of estimating the supermarkets future CO_2 emissions [31].

A recent paper [28], provides a detailed literature review on the state-of-the-art developments of Artificial Intelligent (AI) based models for building energy use prediction. It provides insight into ensemble learning, which combines multiple AI-based models to improve prediction accuracy. The paper concludes that ensemble methods have the best prediction accuracy but that a high level of technical knowledge and computational resources is required to develop them. Consequently, this has hindered their application in real practice. An advantage of high prediction accuracy is that this can allow early detection of equipment faults that could disrupt store operations [15].

These studies show that predicting energy consumption is possible with data mining techniques and that they can predict energy usage within acceptable errors. Compared to other engineering methods, ensemble methods require less detailed information of the physical building parameters [28]. This saves money and time in conducting predictions compared to simulation tools. Hence, they could replace them in the future. Because studies use different types and volumes of input data, there is no unified input data format. Therefore, knowledge of the methods and a variety of data is needed to create meaningful and accurate predictions.

2.5 Energy Baselining

Creating energy forecasts is an important aspect of the energy management and maintenance of buildings [28]. It helps to evaluate energy efficiency, conduct building commissioning, and detect and diagnose building system faults. In order to evaluate the performance of a store or a system as well as to detect any faults, the operation of the store/system must be compared to a *reference behavior* [17].

This reference behavior can be used to determine the performance of the store and can be established in two different ways. One way is called *energy benchmarking* and it includes the comparison of the stores' current energy usage with the current *or* previous performance of a similar store. The energy consumption of another store, with a similar configuration, size or equipment, will be transferred to create the reference behavior. The second method is called *energy baselining*, here the reference behavior is defined as the previous, historically best, or ideal, theoretical performance of the given store [15]. Since supermarkets differ in a lot of ways, a baseline for the energy consumption should be created for each store individually. This baseline is usually created on the analysis of historical data [17], from a selected period of time, and can be developed using traditional data mining techniques.

Two different approaches are described to implement a baseline [17]. The first way includes creating the baseline once, without further updating. When the configuration of a building, either does not change often or a sufficient amount of data is available, this could be a good approach. When the amount of data is limited or the baseline is expected to reflect changes in equipment or building layout, the second way is more suitable. This baseline approach is more dynamic and includes frequent updates, e.g., with the usage of new input variables or data. When a building is regularly renewing their equipment or the volume of data is small, this approach fits best.

A first step to estimate a baseline involves determining the dependent and independent variables [15]. Consequently, the dependent variable will be the energy consumption for the store or system. The independent variables will be factors that affect the energy consumption like temperature, occupancy behavior or floor size. When a baseline is established, it can serve as a starting point for setting energy efficiency improvement goals as well as a comparison point for evaluating the effects of changes or spotting trends in energy performance [15]. The baseline becomes part of the Energy Management process as is shown in Figure 2.1.

Materials & Methods

In this chapter, we start with a description of the time series nature of our data in Section 3.1. We will also introduce time series modeling and some specific time series algorithms that can be used. Section 3.2 describes the machine learning algorithms that are used. Finally, Section 3.3 and Section 3.4 report the techniques that are used to train and evaluate the algorithms' performance.

3.1 Time Series

Time series data are defined as the sort of data that is captured over a period of time [32].

$$X_1, X_2, \ldots X_{t-1}, X_t \ldots \tag{3.1}$$

Where X is the value measured at time t. A time series variable can be observed on a regular basis, e.g., daily temperature measures, or at irregular intervals, e.g., tweets from your favorite American president. The main characteristic is that these observations are dependent in time, so that previous observations are related with the subsequent ones [33].

Time series data is ordered in a chronological order, often referred to as "time", which makes it possible to analyze how a given variable changes over time. The time order can either be based on a sequence of observations or on the timestamp, which indicates the specific moment of measurement [33]. This time order is important for time series data because it enables the analysis of trends, seasonal events or other changes over time. However, seasonality, irregular effects or other developments can cause time series data to deviate drastically throughout a period of time, in time series this is called non-stationary [34].

3.1.1 Stationarity

The biggest part of literature concerning time series analysis is concentrated on stationary time series data [35]. A stationary time series is a set of data, of which, the statistical properties such as the mean and variance do not change over time [33]. Classical time series algorithms work on the assumption that a dataset is stationary [33]. Therefore, it is important to know if a time series data set is stationary or not, because this influences, e.g., the choice of prediction algorithms. A typical approach is to convert non-stationary time series into stationary time series to enable this implementation of classical time series algorithms. Converting the time series can be done by different kinds of techniques, e.g., seasonal adjusting, where the seasonal component of the time series is removed [36].

To discover if a time series is stationary, one can decide to plot the output variable [36]. Figure 3.1, shows the plot of the output variable of a stationary time series. We notice the graph evolving around a mean, which is similar throughout time, and the graph does not show trends or periodic fluctuations. However, Figure 3.2 presents a rising line which exhibits a strong trend and visualizes that the mean is changing over time. Therefore, this is a non-stationary time series dataset.

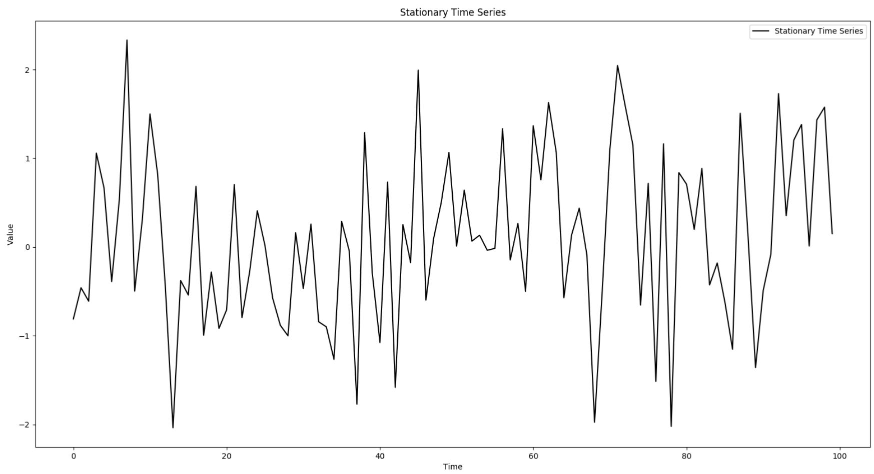

FIGURE 3.1: Example of a Stationary Time Series

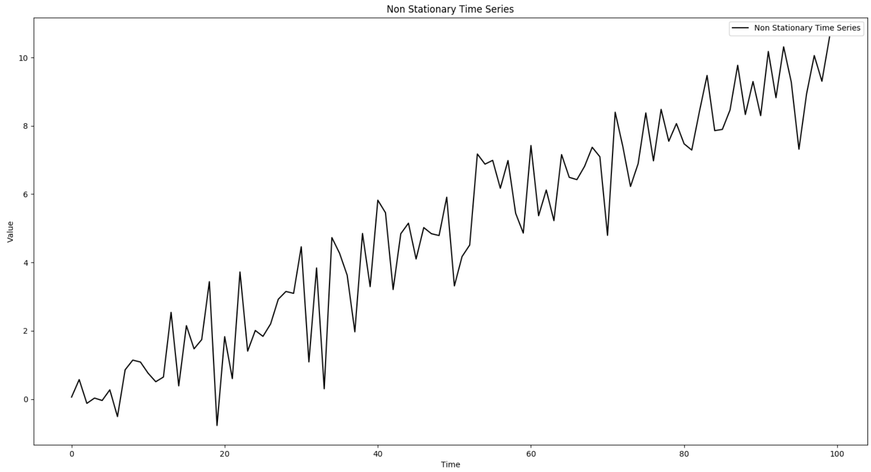

Figure 3.2: Example of a Non Stationary Time Series

It is possible to study the non stationary components of time series data [37]. An example of a non-stationary component is, irregular effects, which can be caused by equipment upgrades, replacements or breakdowns. Seasonality is another component of time series data in which it undergoes consistent and expected variations that are repeated every calendar year.

3.1.2 Time Series Forecasting

In time series forecasting, forecasts are made on the basis of data comprising one or more time series [38]. Making forecasts can also help in model evaluation when testing different time series algorithms [38]. Based on the analysis of historical data the forecast will provide new future values. One way to test the accuracy of forecasts is to compare these future values with the actual observed values.

It is very hard to forecast a numeric value correctly but the difference can be larger or smaller. Every forecast $\hat{Y}_i$ of an observed value Y_i will have a forecast error E, which describes the deviation among them. A lot of studies are focused to minimize the prediction error, i.e., to get it as close to zero as possible [39, 40]. Reducing errors on prior values can be a good practice for building a reliable forecast model. The difference in focusing on prior *or* future values distinguishes two types of forecasts, *Ex-post forecasts* and *Ex-ante forecasts* [41]. When using a regression model in combination with time series data, it is possible to produce these two forecasts. Where ex-ante forecasts are forward looking, ex-post forecasts are backward looking. I will demonstrate their difference using two examples.

- *Ex-ante forecasts* are entirely based on available information at the time the forecast is produced. For example, when we produce an ex-ante forecast of the number of tourists that are expected to visit Portugal next year (2018), we can *only* use data up to this year (2017).
- *Ex-post forecasts* use information beyond the time at which the forecast is made. For example, when we create an ex-post forecast of the number of tourists that visited Portugal in the past year (2016), we can use the observed values, of the independent and dependent variables, that are known after this forecast period.

3.1.3 Forecast Algorithms

Forecasts can be produced using a time series algorithm which learns from historical data to predict future values. For time series forecasting we can use several statistical techniques to learn from past data to build such a model. We can divide them into methods where *all* data points hold equal relevance, *or* methods where all data points carry a weight that strengths their relevance [42]. Here we explain several classical methods for time series forecasting, Regression, Moving average, and Exponential smoothing.

Regression Regression is not a time series specific method for forecasting, however, it can be applied to make time series forecasts. Regression can be useful when we want to use multiple input variables to make a prediction, e.g., domain knowledge can be applied to create these variables. In a regression analysis, we search for a linear relationship between these independent variables. Based on this relationship the algorithm will be able to predict a value for the dependent variable.

In multiple regression models, we forecast the dependent variable using a linear combination of the independent variables. This is slightly different than what happens with **Auto Regression**, where this is done using a linear combination of the prior values of the dependent variable. Regression assumes that prior values have an effect on the subsequent ones. This method makes use of linear relationships and, therefore, works best when the time series data is stationary. By defining the number of past values, that are included in the analysis, we can realize variations in the prediction performance.

Moving Average The word *moving* symbolizes that the period over which an average is computed changes over time. Every moment a new observation becomes available this one replaces the oldest one. Values can vary widely per observation, yet, depending on the size of the sample used for averaging and the existence of outliers, the average across the entire period resides steady. As a result, the average moves slowly when additional observations are attached. This method is applied in time series analysis to smooth out short-term fluctuations and highlight long-term trends or cycles [33].

If a time series dataset is stationary one can obtain good results with the moving average method. This method assumes that the average, computed over a specified interval, is more representative of the forecast period than using only recent observations [42]. To use this method, a period of historical data needs to be selected. If recent values are considered relevant, one can select a small number of observations, when more historical data is preferred a larger quantity can be selected. Determining the optimal amount can be done based on a mixture of domain knowledge, understanding of the particular time series data and behavior of the model.

Exponential Smoothing The name *exponential* smoothing reflects the point that weights exponentially decrease when observations grow older [42]. Methods based on exponential smoothing have the characteristic that forecasts are weighted aggregates of prior observations. In these aggregates, recent observations carry relatively more weight than older ones [42]. The concept is that the closer the data remains to the forecast period, the more suitable it is to be predictive, opposed to older data. A method based on exponential smoothing can be applied to forecast values for non-stationary datasets.

3.2 Data Mining and Machine Learning Methods

In this section, we will describe the Data mining and Machine learning methods that are used in this study. We will start with the introduction of a descriptive data mining method, Subgroup discovery. Next, the predictive machine learning algorithms will be introduced. Because the nature of this preliminary study is to use off-the-shelf techniques, we will briefly introduce these predictive algorithms. The literature described in Section 2.4 gives examples of successful applications of these methods for the prediction of energy consumption.

Descriptive data mining is a knowledge discovery approach [43]. It consists of searching for patterns in data to explain what is happening. Predictive techniques try to forecast what might happen in the future based on the dependency amongst this historical data [44].

3.2.1 Subgroup Discovery

Subgroup discovery (SD) is acknowledged as a general technique to perform descriptive and exploratory data mining [45]. This method can be used to, e.g., perform knowledge engineering, feature creation, or to identify relations between dependent and independent variables [43, 46]. It aims to identify interesting subsets of a dataset that show an unusual behavior. The differences in behavior between the subgroup and the complete dataset these methods search for are not due to statistical variations but to external factors [43]. Identifying these factors helps to produce a general understanding of the data and provides insights for the analyst [43].

3.2.2 Multiple Linear Regression

Linear regression is a simple and widely used statistical technique for predictive modeling [24]. When more than one input variable is used, we talk about Multiple Linear Regression (MLR). The easiness of using this method is viewed as the main advantage of linear regression because there are no parameters that have to be tuned. However, one major limitation, linear regression is not able to handle nonlinear problems well [28].

The Multiple Linear Regression model can be represented with this formula [24]

$$Y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \ldots + \beta_k x_{k,i} \tag{3.2}$$

Where Y_i is the dependent variable, i.e., the value we want to predict. $x_1, \ldots x_k$ represent all the independent variables that are used as input for the prediction. $\beta_1, \ldots \beta_k$ are the coefficients of the corresponding independent variables. The coefficients show the change in the value of Y_i per unit of change in the independent variable when all other factors stay equal. For example, when the coefficient β_1 is 100, and the corresponding input variable x_1 changes from 29 to 30, the output variable will be adjusted with 100.

MLR is able to explain the relationships between the dependent and the independent variables. The coefficients can also be used to determine variable importance, to see which variable is the most important for making the prediction [47]. However, variable importance can only be determined when the input variables have the same scale, i.e., are normalized.

To determine the coefficients, MLR can use multiple methods of which the least-squares method is the most popular. The goal of this method is to minimize the error between the observed and predicted values [24, 48]. The result of the least squares method is the "best fit", which can be represented by a line.

In Figure 3.3, the red dots represent the real values. The blue line is the linear model, represented by the formula shown above. The green lines are the differences between the predicted and the observed values, i.e., the errors. The coefficients are obtained by finding the minimum sum of squared errors. The total error is calculated with the following formula:

$$E = (e_1)^2 + (e_2)^2 + (e_3)^2 + \ldots + (e_6)^2 \tag{3.3}$$

Where $e1, \ldots, e6$ are the individual squared errors.This is done so that the positive and negative errors won't cancel each other out, the error will always have a positive value. By adjusting the coefficients the model can minimize the differences and create the best fit, where E is the smallest. The best fit, e.g., linear model, is found when the value for E is the lowest. This is described as "least squares" estimation because it gives the least value for the sum of the squared errors.

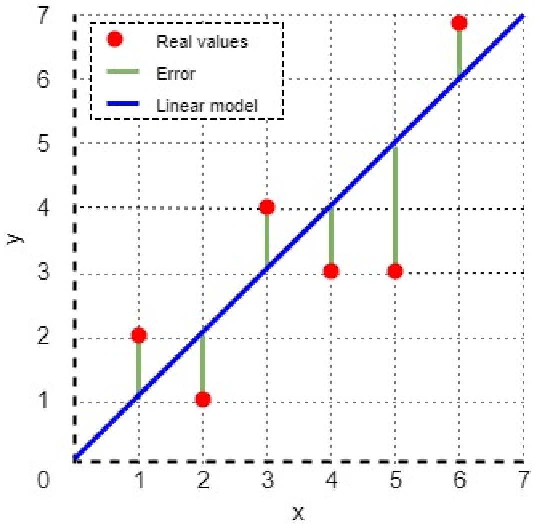

FIGURE 3.3: Example of linear regression, which has one independent variable

3.2.3 Random Forest

Random Forests (RF) are an ensemble method originally introduced for classification and regression problems [49]. It essentially consists of the generation of multiple decision trees obtained using different randomization techniques (Figure 3.4). The set of predictions made by each of these trees is aggregated to obtain the prediction of the RF [50].

The Random Forest method has specific characteristics that make it interesting for this research. First, RF is considered to be one of the most accurate general-purpose learning techniques available and is popular because of its good off-the-shelf performance [51, 52]. It generally produces highly accurate predictions and is able to handle a large number of input variables [28].

A state-of-the-art review of artificial intelligence methods used for energy prediction concludes that ensemble methods offer several advantages over single-model methods [28, 53]. The advantage of the ensemble methods lies in its remarkably improved prediction accuracy and stability [28]. Where a single-model method uses one type of model for prediction, ensemble methods combine multiple models. When a single method is used, the user needs to select a suitable learning algorithm for their problem because no particular model has dominated others in predicting [28]. Since the ensemble method uses multiple models for prediction, the combination of these models will reduce the overall prediction error of the model [28]. The eventual prediction of the ensemble method is made on the integration of all these different models. And because it is unlikely that all the models will fail, at the same time, there is a little chance that the ensemble method causes great errors [28].

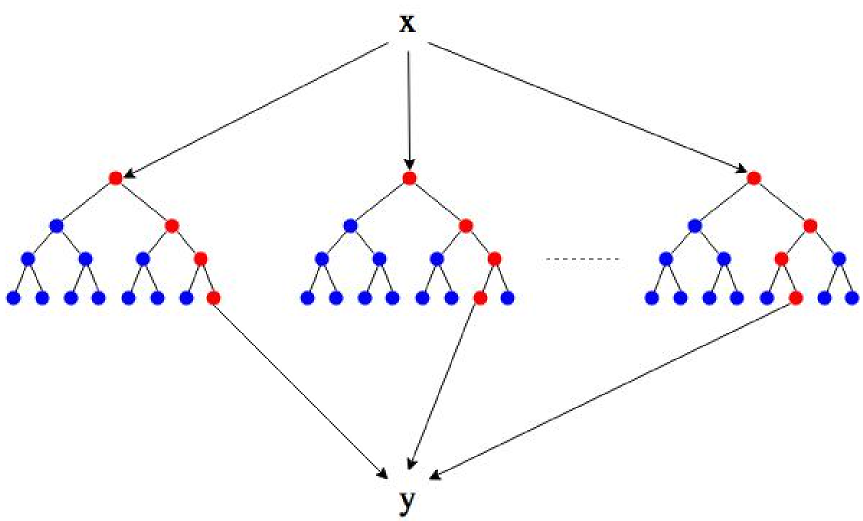

FIGURE 3.4: Example of a Random Forest

3.2.4 Artificial Neural Network

Artificial Neural Networks (ANN) have successfully been used in numerous studies to predict energy consumption [15, 28, 29, 31, 54, 55]. The success can partly be explained because of their ability to make time-series predictions [56]. Artificial Neural Networks are also used in the area of deep learning [56].

In a Neural network, each neuron is connected to the next layer of neurons, through synaptic weights [56]. If this connection is unidirectional (one way), the network is characterized as feedforward (Figure 3.5). The connections between the neurons receive a certain weight, we can compare this to the coefficients in the MLR [15]. The number of input and output neurons are equal to the amount of independent and dependent variables.

Models developed using Neural networks can achieve very high predictive performance. ANN's try to imitate the learning process of the human brain and learn from examples [56]. The model is able to identify relationships between the input and output data using previously recorded data [15]. A limitation is that they can only do this for data that belongs within the boundary of the training set [57]. This means that to be able to make full advantage of using a Neural network it is essential to use a substantial amount of training data, that is representative of the whole range of the variables [57]. For example, because one of the variables is Temperature, samples of different seasons or multiple months are needed to enable the network to generalize and predict better [55]. When multiple seasons are included, a lot of different temperature values can be fed to the model.

Artificial Neural Networks can be harder to tune in comparison to MLR and RF since there are more parameters that can be tuned [29, 56]. For example, one study was devoted to using various configurations, several training algorithms, and multiple sets of input variables to create the most efficient predictions with the ANN [54].

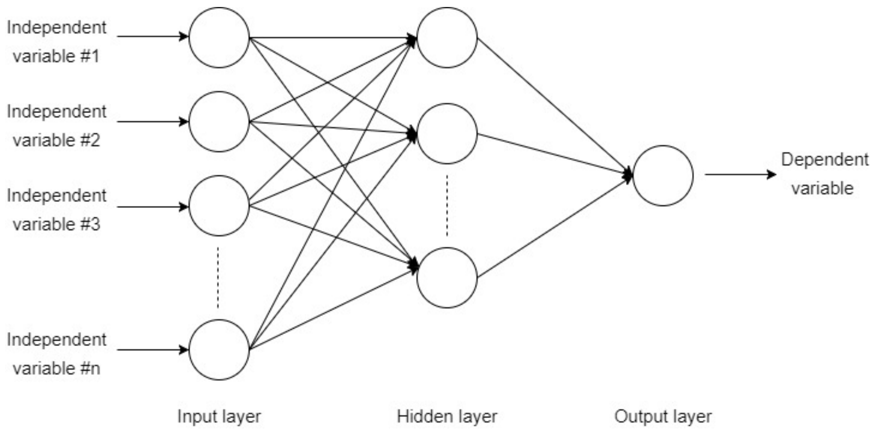

FIGURE 3.5: Example of a Feedforward Artificial Neural Network

3.3 Experimental Methods

During this study, several experimental methods have been applied to research the questions described in Section 1.2. In this chapter, we will discuss each of the methods that we used, Learning Curves and Sliding Window. In the experimental setup, Chapter 4, a description of the chosen approach is outlined which explains how they are used in combination.

3.3.1 Learning Curves

In Machine Learning, learning curves are used to reflect the predictive performance as a function of the number of training examples [58]. Figure 3.6 reveals the developing learning ability of a model when the number of training examples increases. The curve indicates how much better the model gets in predicting when more training examples are used. The general idea is to find out how good the model can become in predicting and what the subsequent number of training examples is [58]. Since we are searching for the minimum number of training days to create a baseline, we can use the learning curves to identify this number.

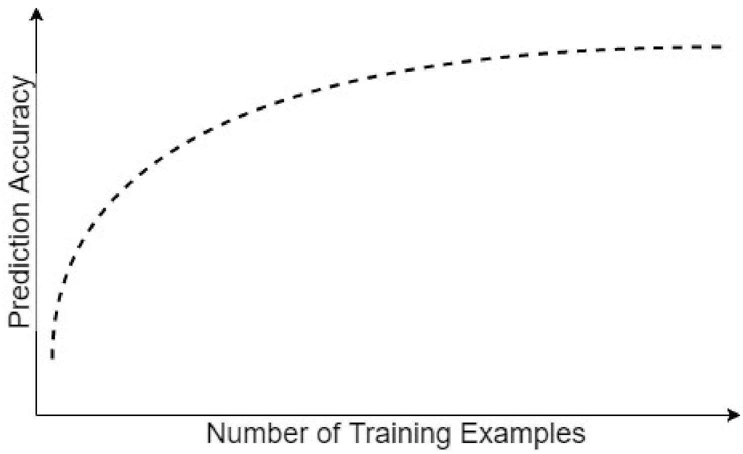

FIGURE 3.6: Example of a graphical representation of a learning curve

To test the learning ability of a model one can create several training sets of data and evaluate their performance on a test set [59]. These training sets can differ in, e.g., volume. It is preferred that the data for these sets are randomly selected from the available data [59]. The purpose is to train the model multiple times, and after every training, the model performance should be tested. The results of these tests can be plotted to draw a learning curve which shows the evolution in the performance of the model. These curves can be clarifying, especially when the performance of multiple models is compared. Besides for model selection, also the performance of a model can be compared in relation to the number of training examples used [58]. Such a learning curve will tell how the model behaves when it is constructed with varying volumes of training data.

3.3.2 Sliding Window

The Sliding window method uses historical data to predict future values [60]. In this method, a window of specified length slides over the data [34], (Figure 3.7). This makes it possible to train and test the algorithm in multiple periods of time. The window defines the length of the data over which the predictions are computed, the window size can be fixed or variable. The Sliding window model can be either sequence or time-stamp based. If it is sequence based, the window sizes are defined by the number of observations, time-stamp based means that it is defined by a duration in time [33]. The window slide can be triggered due to different reasons, e.g., when new data becomes available or when a specific point in time is reached [61].

The sliding window is a popular method in the field of statistics, where it is used to calculate statistical information over a fixed period of time or instances, with the necessity to forget old observations [33].

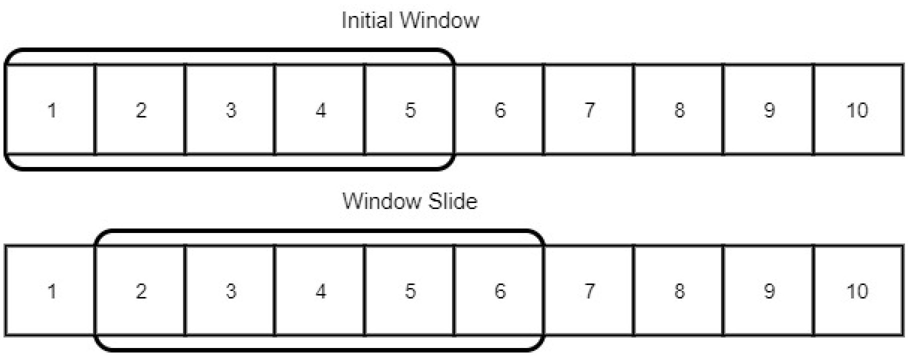

FIGURE 3.7: Example of a Sliding Window Process

3.3.3 Updating the Forecast Model

Having a reliable model that makes accurate predictions is a good starting point. The accuracy of a model influences the effectiveness and actionability of the outcomes. Therefore, it is important to observe the performance of a learning model, e.g., with learning curves. To prevent that errors in predictions become too large, and keep the model reliable, one can choose to perform regular updates, with fresh data, to keep the model accuracy stable.

A simple way to determine the life cycle of a forecast model is to analyze how prediction errors emerge over time. When they fall outside an acceptable range one can choose to update the model. The most widely used updating procedure, for time series models, is renewing output variables [62]. A general assumption is that forecasts strongly depend on the most recent behavior of the input data [62]. Therefore, the model needs to be filled with fresh data to make the most accurate predictions. During the updating process, following the first in, first out principle, up-to-date input data will be added, some input data will stay in place and the oldest part will be removed.

3.4 Evaluation

Nowadays, Machine Learning models and methods are applied in various areas and are used to make important decisions which can have far-reaching consequences [63]. Therefore, it is important to evaluate their performance. Evaluating is also done to choose the best configuration out of different methods, parameters, features or datasets [63]. Currently, Cross-Validation (CV) is the widely accepted and most used evaluating technique in data analysis and machine learning [63, 64] and serves as a standard procedure to estimate model performance.

Cross-Validation is used to test the performance of a predictive algorithm that is trained with cross sectional data [64]. Cross sectional data is data which is collected at the same point of time *or* is without consideration to differences in time [65]. Normally, this data is collected to study a population at a current point in time, instead of studying a development over a period of time.

When used for evaluating prediction performance, the algorithm is given a dataset on which it will be constructed, *training data*, and a dataset on which this model will be tested, *test data*. The goal of Cross-Validation is to repeat this process multiple times to make full use of the total data set [63]. Therefore, the dataset will be divided into equal parts, called folds. Each fold will once represent the testing set, when all the other parts are used for training, Figure 3.8. Since time is not a factor of importance, folds from each point in time can be used for training and testing. Once all the folds are used for testing, the average of the error terms is used to calculate the prediction accuracy.

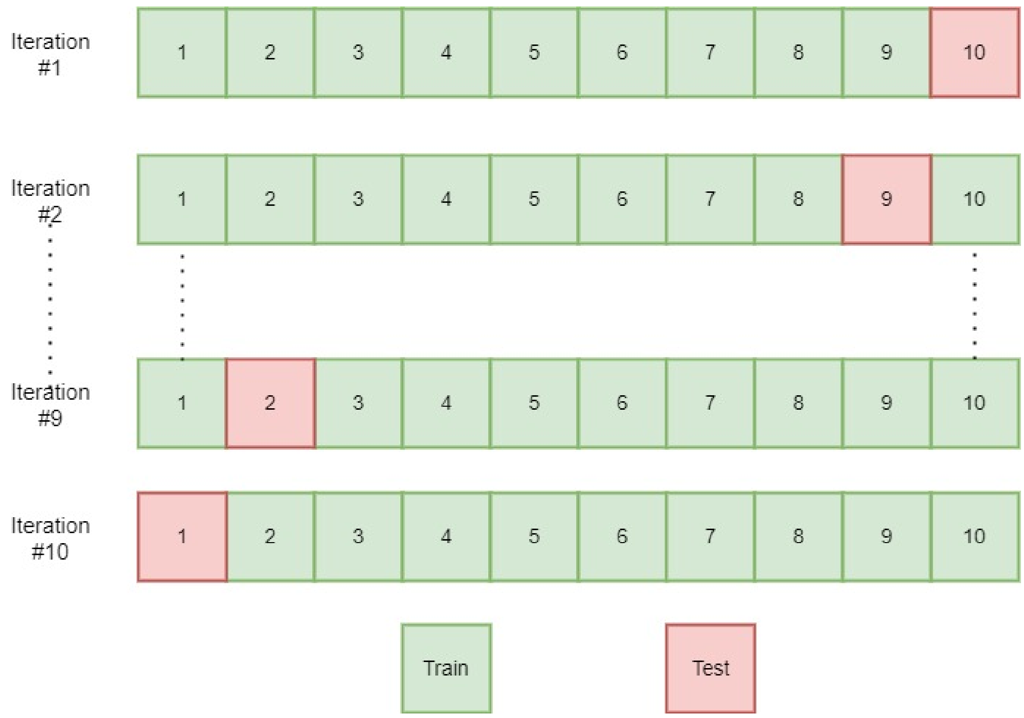

Figure 3.8: Example of a Cross Validation

However, Cross Validation does not work well in evaluating the predictive performance of time series [64]. Time-series prediction is a method of forecasting future values based on historical data [61]. In contrast to cross-sectional data, the time order and period of data collection are important. However, a consensus regarding standard procedures to evaluate time series data does not exist [63]. This is not only due to the different characteristics that time series can have, but also the different evaluation techniques that exist like forward validation, modified cross-validation or blocked cross-validation [63].

One way to validate the prediction performance of a time series model is to make use of a Sliding Window design [66], (Figure 3.9). In this method, the training and test data can not be randomly selected throughout time for training and testing. This is where this method differs with the Cross-Validation technique, (Figure 3.8).

Figure 3.9 shows the process of a Sliding Window approach with an initial window size 5. Each number represents an observation of the time series data $(1, 2, 3, \ldots, 10)$. The initial window at $t = 1$ shows that the first until the third observation is used to predict the value for the fourth. Next, the sliding window will slide one observation to the right and the first until the fourth are used to predict the value of the fifth observation. This process will be continued until the sliding window reaches the end of the time series. Once all the examples are used for testing, the average of the error terms is used to calculate the prediction accuracy.

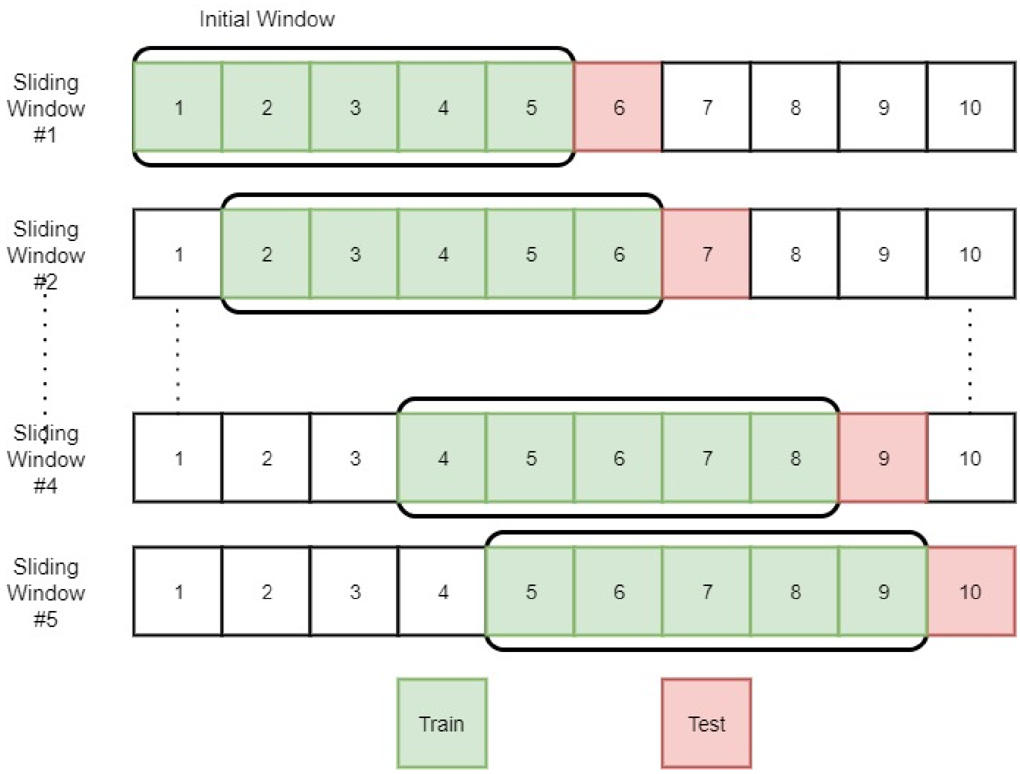

FIGURE 3.9: Example of a Sliding Window Validation

3.4.1 Metrics

Error metrics are used to evaluate the prediction performance of a model. Error metrics show the difference between the observed and the predicted value. The Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are the most frequently used error metrics in Machine Learning for regression [67].

The following formula shows how the Mean Absolute Error is calculated:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|\hat{Y}_i - Y_i\right| \tag{3.4}$$

Here $\hat{Y}_i$ is the predicted value and Y_i is the real value. The calculation of the MAE is relatively simple. First, the distance between $\hat{Y}_i$ and Y_i is calculated, this represents the absolute error. After all the tests are executed, all the absolute errors are summed and divided by the number of tests. This gives the Mean Absolute Error [68]. This error represents the average absolute error of the model, not to confuse with the RMSE.

The following formula shows how the Root Mean Squared Error is calculated;

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{Y}_i - Y_i)^2} \tag{3.5}$$

Here $\hat{Y}_i$ is the predicted value and Y_i is the real value. The calculation of this error is done in three steps, First, the individual absolute errors are squared and summed to form the total squared error. By squaring the absolute errors, bigger errors are penalized heavier than smaller ones. This means that large errors have a higher influence on the total squared error. Secondly, the total square error is divided by the number of tests, this produces the Mean Squared Error (MSE). Finally, the RMSE is calculated by taking the square root of the MSE [68].

Pearson's correlation can be used to describe the dependence between the independent and the dependent variables [69]. The squared Pearson correlation coefficient, (R^2), can also be used to assess model performance [70].

Mean Absolute Error is the most natural measure of the average prediction error [68, 71]. The RMSE describes more than just the average prediction error alone, e.g., since the errors are squared, the RMSE also describes the magnitude of the errors [71].

Experimental Setup

This chapter starts with describing the characteristics of the data used (Section 4.1). Next, Section 4.2 describes the approach used and Section 4.3.1 illustrates it with a practical example.

4.1 The Data

This section provides insight into the data collection process and the actions taken to prepare the data for training the models. Furthermore, an overview of the datasets will be provided.

For this study, we obtained data from five stores across Portugal. The stores are located in the regions of Aveiro, Fatima, Macedo de Cavaleiros, Regua, and Mangualde, Figure 4.1. These supermarkets sell mainly food products and some general merchandise. Part of the food products is stored in fridges and freezers, of which the electricity consumption is measured in 15-minute intervals.

The datasets provided are mainly based on the energy consumption and weather data for the whole year of 2016 and the first half of 2017 (Table 4.1). The data for each store is available from the moment the store opened or started to collect the data. Hence, for each store, the maximum amount of data is available. The energy consumption is measured in kilowatt hour (kWh) from the Retailer's energy monitoring system. The weather data consists of the outside temperature derived from a sensor placed on the roof of the store and is measured in degrees Celsius (C°). Both the energy and weather data are delivered in 15-minute intervals. Because we focus on daily forecasts, they were aggregated by day.

FIGURE 4.1: Example of the Store Locations

	First day	Last day	Observations
Aveiro	04/12/2015	26/04/2017	510 days
Fatima	07/01/2016	26/04/2017	476 days
Macedo de Cavaleiros	13/11/2015	26/04/2017	531 days
Mangualde	16/05/2016	16/05/2017	366 days
Regua	16/05/2016	16/05/2017	366 days

TABLE 4.1: Overview datasets

4.1.1 Data Standardization

In order to apply a similar approach to the data of each store, we decided to work separately with datasets that have a similar structure. Before creating these datasets, it is important to identify the dependent and the independent variables [15, 28]. In this study, an energy baseline will be created that reflects the estimated refrigeration energy consumption. Consequently, this will be the dependent variable, and the independent variables are the ones influencing this consumption. As can be read in Section 2.3, there are a lot of factors that can influence the energy consumption, but only the factors that are measured, by all stores, can be used here as an independent variable. The effect of the independent variables, in the refrigerations' energy usage has to be examined. By examining their importance, on the final predictions, we can create a dataset that gives us good prediction performance. The following variables were available for each store; *Timestamp*, *Refrigeration Energy Consumption* and *Outside temperature*. Based on the two available variables, *Timestamp* and *Outside temperature*, we tried to create new features with additional information that the algorithm can use.

To determine which features to create, knowledge about the behavior of the store is important [15]. The process of designing new features, based on domain knowledge, is called Feature engineering [72]. Designing appropriate features is one of the most important steps to create good predictions because they can highly influence the results that will be achieved with the learning model [73]. The domain knowledge required for this process can be acquired through conversations with experts, reviewing similar studies and using descriptive data mining techniques, e.g., Subgroup discovery (SD). SD is a method to identify, unusual, behaviors between dependent and independent variables in the data [45, 46]. In this study, SD will be used to improve our understanding of the behavior of the energy consumption.

Similar studies [15, 24, 30, 31, 54, 74, 75], have been reviewed to analyze what kind of variables, based on the temperature and timestamp, were used to predict energy consumption. Based on the two available variables, a large set of variables were created, such as *Week*, *Date* and *Maximum Temperature*.

It is important to limit the number of variables in the input dataset to avoid over-parameterization, i.e., using an excessive number of features. This makes the model complex and can lead to overfitting [15]. When a model is overfitting, instead of learning from the input-output patterns the model will start memorizing them, and it will lose the ability to generalize well. This leads to poor predictive performance, hence, the model will not be able to make accurate predictions for the baseline. Therefore, an optimal subset of features needs to be selected from this large set of variables. The process of selecting the optimal subset of variables is called Feature selection [76]. During this process, redundant or irrelevant features are removed from the input dataset. In this study, the set of features was selected using a combination of domain knowledge and trying different sets of variables for making predictions. The features that are selected for the datasets will be described in the following section.

4.1.2 Features

The original time series data was provided, in sometimes irregular, 15-minute intervals. After this restructuring, the data is converted into hourly values and eventually, transformed to daily formats. Table 4.2 gives an overview of the final features that are derived from the Timestamp variable. As used in other studies, we also use the Workday variable [15], because supermarkets tend to have a different energy consumption during weekdays and weekends due to different opening hours and shopping habits [15]. The Retailer stores also have different opening hours during weekends and workdays and therefore this feature can be useful.

Name	Type	Description	Derived from
Weekday	Categorical (1-7)	Day of the week	Timestamp
Week of the Month	Categorical (1-4)	Week of the Month	Timestamp
Workday	Binary (0-1)	Workday or Weekend	Timestamp

Table 4.2: Variables derived from the Timestamp

Table 4.3 gives an overview of the final features that are derived from the Temperature variable. Variables related to temperature are used in a lot of similar studies [15, 24, 30, 31, 54]. One study used this amplitude variable, for the indoor temperature, to improve their regression analysis [74]. The variable Temperature Amplitude is the maximum change of a temperature in a given period. In this dataset, it represents the maximum change of outdoor temperature for one day.

Name	Type	Description	Derived from
Max Temperature	Numerical	Max Temperature of the Day	Temperature
Mean Temperature	Numerical	Mean Temperature of the Day	Temperature
Min Temperature	Numerical	Min Temperature of the Day	Temperature
Temperature Amplitude	Numerical	Absolute Difference Min and Max	Temperature

Table 4.3: Variables derived from Temperature

During meetings, to collect requirements and acquire domain knowledge, the experts explained that thermal inertia of the building plays a role in the energy consumption of refrigeration. Thermal inertia means that the indoor conditions are affected by the outside temperature [75]. During days with a high outside temperature, the building warms up and this effect even lasts the following day. Therefore, Table 4.4 contains variables that are created based on the temperature of the day before. Another effect of including this variable is that the observation concerning the first day of each dataset had to be omitted because this row could not contain temperature values of the day before.

Name	Type	Description	Derived from
Max Temperature Y..	Numerical	Max Temperature of Yesterday	Temperature
Mean Temperature Y..	Numerical	Mean Temperature of Yesterday	Temperature
Min Temperature Y..	Numerical	Min Temperature of Yesterday	Temperature
Temperature Amplitude Y..	Numerical	Absolute Difference Min and Max	Temperature

Table 4.4: Variables derived from Temperature

4.1.3 Indicator Variables

The final dataset, which is the combination of Table 4.2, 4.3 and 4.4, contains several categorical variables. Because not all methods, that we are going to test, deal well with categorical variables we will transform these into indicator variables. Designing indicator variables is a commonly used method to 'trick' the regression analysis to accurately examine these variables [77].

For our experiments, we transformed the variables *Weekday* and *Week of the Month* into indicator variables. We decided to do this because all the values that these features hold are evenly meaningful. The values of these features are used to identify the level of the variable, e.g., 1 = Monday or 5 = Friday. The following example illustrates how we want to avoid that these variables are treated. For the variable Weekday, the number 6 represents Saturday, but Saturday is not twice as much as number 3, representing Wednesday. Or number 3, Wednesday, is not the same as 1 plus 2, i.e., Monday plus Tuesday. So, for the variable *Weekday* and *Week of the Month*, 6 and 3 indicator variables are created.

Indicator variables hold binary values, 0 or 1, which indicate the presence or absence of their particular category, (Table 4.5). Using an equal number of indicator variables to represent an equal number of categories is called the indicator *variable trap*. This variable trap can distort the regression analysis [78]. This can happen when, e.g., instead of six, seven indicator variables would be used to represent Weekday. The general rule to bypass this trap is to use one fewer indicator variable then there are categories [78]. As can be noticed in Table 4.5, instead of one variable representing Weekday, 1-7, we will now have six indicator variables to code the seven categories of Weekday. That is because when all the values are set to zero, the category Sunday is specified.

	D1	D2	D3	D4	D5	D6
Monday	1	0	0	0	0	0
Tuesday	0	1	0	0	0	0
Wednesday	0	0	1	0	0	0
Thursday	0	0	0	1	0	0
Friday	0	0	0	0	1	0
Saturday	0	0	0	0	0	1
Sunday	0	0	0	0	0	0

Table 4.5: Example of Indicator variables for Weekday

4.1.4 Normalization

After creating the indicator variables, the next step was to perform normalization on the data. Normalization is a general method which is performed to improve predictive accuracy and can be useful for some algorithms [79]. For this study, normalization has a couple of advantages. First, normalization makes it possible, depending on the algorithm, to calculate variable importance, because the weights of the variables can be compared. Next, because the values of all features lie in a similar range, also, the error metrics can be calculated in this similar range. For example, since the absolute values of all features are transformed into a similar range, 0-1, it does not matter if the max energy consumption is around 1200 kWh or 2500 kWh. They will both receive a value 1 if the scale is 0-1. This makes it possible to compare the learning models' and the stores' performance. Finally, the normalized data has a new scale and therefore serves to protect data privacy, since the original values are not displayed.

The objective of normalization is, to transform the values of the variables into a similar value range. By doing this, one can prevent that particular features have a higher impact on the prediction due to their greater numeric value range [80]. Because our variables have different scales, one variable might overpower another one [81].

For the normalization in this study, the Z-score method is used on all the variables except the indicator variables. Using the Z-score method, the values for the attributes x are normalized by subtracting the population mean from the value x and next, dividing this by their standard deviation [81]. This method is used because it is particularly useful when the actual minimum and maximum of attribute x are unknown [81].

4.2 Experimental Setup

The data is obtained from five stores across Portugal, as is mentioned in Section 4.1 We used commercially available software from Python (3.6) and **R** [82] to perform the experiments. The Python libraries from Scikit-Learn and Keras [83] were used to implement the machine learning algorithms. Because our goal was to use off-the-shelf approaches, no changes have been made to the default configuration of the used machine learning models.

In order to study the questions described in Section 1.2, we designed an approach based on Learning curves in combination with a Sliding window. Our experimental setup is a variation of the Time series approach used by [84, 85]. The method we propose is visualized in Figure 4.2. We decided to use this particular method because we want to train machine learning models with different sizes of historical training data. The learning curves enable us to visualize and evaluate their performance.

The method we designed, is able to keep the time series data in order. This is an important challenge since we do not use a time series specific model [61]. Specific time series learning models like SARIMA and ARIMA require more technical knowledge before they can be applied [61]. Furthermore, we want to be able to use domain-specific knowledge to engineer new features, therefore, we decided to follow a regression approach. We selected off-the-shelf machine learning algorithms like MLR, RF and ANN.

4.2.1 Approach

The approach, that is described here, is based on the business questions and their corresponding research approaches that are described in Section 1.2.

We start our approach with data exploration. In this study, we use the Subgroup Discovery (SD) method to perform this task. This method has the ability to search for interesting behavior in the data (Section 3.2.1). As soon as we find interesting insights, we share them with the business. For the SD analysis, we use a dataset that contains the original values, therefore, we excluded the indicator variables. These variables can make it hard for SD to find interesting patterns. For example, for the variable Weekday, Sunday can only be recognized if the six other variables, Monday-Saturday are tested for 0. The result is that SD has to combine six different rules, which makes the patterns too deep. Excluding the indicator variables improves the explorative power of SD, and using the original values makes the results easier to interpret.

Because we are interested in changes in the energy consumption, we decided to not transform our time series dataset as described in Section 3.1.1. When we remove the irregular changes in the energy consumption, to make the dataset stationary, we lose the ability to study them. Our non-stationary time series dataset enables us to detect these irregular changes in the energy consumption.

The first goal of this research is to define the minimal set of training examples needed to build a solid energy baseline. To do this, we train the machine learning models with different sizes of training examples. Each iteration we increased the number of training examples and evaluated the models prediction accuracy using the approach described in Section 3.4. When all iterations have been completed, we are ready to plot the error metrics in the learning curves. Because this approach is replicated for the three models, this also reveals which one performs the best.

After we selected the model which is able to create the baseline with the least amount of data, we define the update frequency of this setup. We expect the prediction error to grow over time, and therefore the baseline will become unreliable at some point when the prediction error becomes too high. To find the point of which we recommend updating, we use the previously defined setup, to make predictions for the remaining dataset. As soon as the predictions are made, we compute a MAE for each of 10 subsequent predictions. Once all the errors are computed, we can plot them in the learning curves. This enables us to analyze how the prediction accuracy develops along the prediction horizon, and define the update frequency.

Finally, the third part of this research is to analyze the long term prediction performance. This was done by training each model with various sizes of training data and let it predict for the remaining dataset. After the predictions were made, we then calculated a MAE for every 10 subsequent predictions. Having plotted the error metrics in the learning curves, meant that we could study their performance over time.

The Retailer requested us to use the Mean Absolute Error (MAE) as the error metric because its simplicity makes it easier to explain the results to the business (Section 3.4.1). Therefore, we decided to use the MAE to describe the difference between the observed and the predicted values. Furthermore, the Retailer required us to calculate the Pearson correlation.

4.3 Learning Curves in Sliding Windows

Before we can execute the method, Figure 4.2, we need to define three parameters.

1. First, we define the number of training examples for the *initial window*. This number represents the maximum size the training set can reach.
2. Next, we determine the size of the *step*, symbolized by the green square in iteration 1. This iteration starts with this single step, and after each iteration, one more step is added. Each iteration the initial window increases with historical data, and therefore expands backward, until the maximum size is reached.
3. Finally, we set the volume of the *prediction window*, which is symbolized by the red square. The prediction window is used to test the prediction performance of the models. As soon as the initial window is filled with green steps, i.e., reached its maximum size, it slides to the right. The size of the prediction window is also used for this *window slide*.

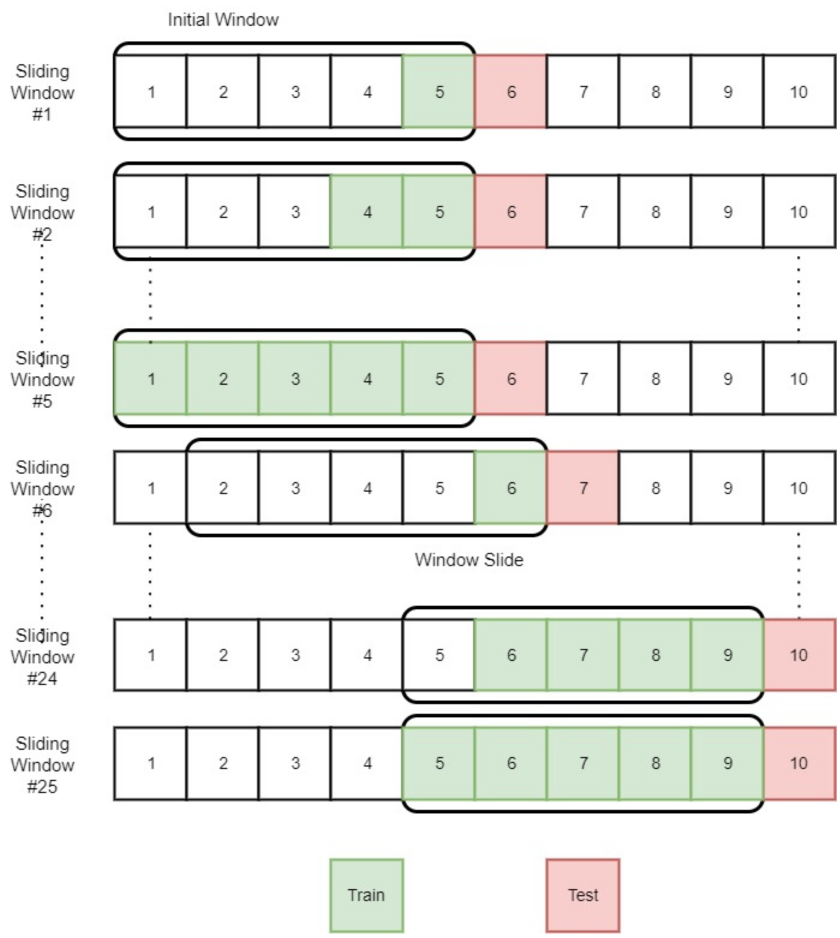

FIGURE 4.2: Example of our Sliding Window Process

4.3.1 Example

For example, the parameters were set as *Initial window* = 180, *Step* = 10, *Prediction window* and *Window slide* = 50. By choosing these values, we can make full use of the available dataset. With this setup, we can train and test the model in six different periods of time.

In the first iteration, 10 subsequent training examples, from the period November until May, are used to train the model. The examples cover a period of 10 consecutive days prior to the prediction window, which is colored red and covers a period of 50 days. When in the next iteration 10 more days are added, the training set now covers a period of 20 days preceding the prediction window. For each iteration that follows, the steps add 10 more training examples until the test set holds 180 days. As soon as the initial window reached its maximum size, the whole window slides 50 days to the right. This process replicates itself until it arrives at the end of the dataset, where the period of July until January predicts the values for February and March. After each test, the error metric is computed and plotted in the learning curve.

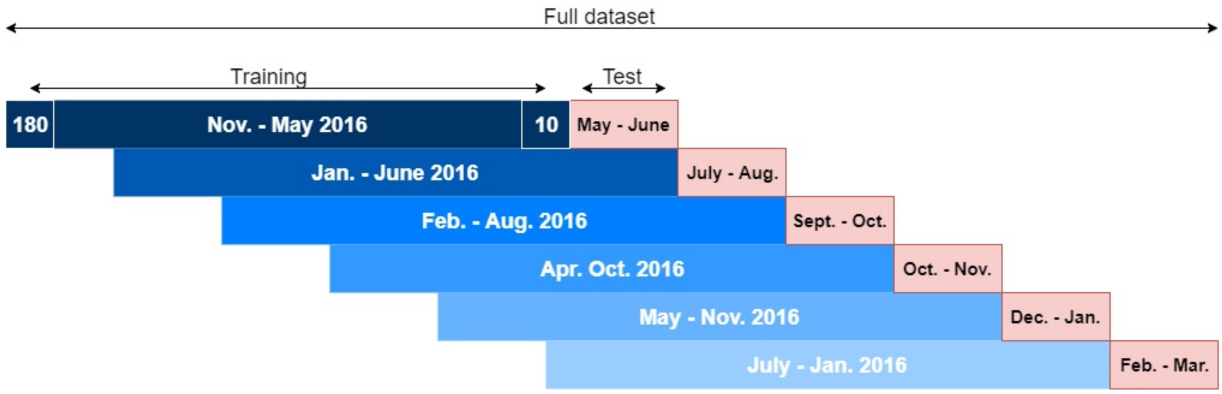

FIGURE 4.3: Example of the used Sliding Window Process

Results

This chapter is organized around the three research approaches described in Section 1.2. We will first describe the results, that we derived from data exploration in Section 5.1. Next, each research approach will be subsequently answered in Sections 5.2, 5.3 and 5.4. Each section also contains a discussion of the results.

5.1 Data Analysis with Subgroup Discovery

We used the open-source software *Cortana* to perform Subgroup Discovery (SD) [86]. Cortana has the ability to identify interesting patterns in datasets. We will use observations, retrieved from the store located in Macedo Cavaleiros, to show insights retrieved from SD.

We used the default values, and we selected the Average as the quality measure. With this setup, Cortana found 26 interesting subgroups in the dataset from Macedo Cavaleiros. The subgroup with the highest score is *Mean Temperature >= 24*. Figure 5.1, shows that this subgroup, (the black line), has a different distribution than the whole dataset, (gray line). This subgroup reveals, that when the Mean Temperature surpasses this threshold of 24 degrees, the energy consumption follows a different energy profile.

In Figure 5.1, the gray line shows us how the energy consumption is distributed for the whole dataset of Macedo Cavaleiros, the whole dataset contains 531 observations. The black line shows this distribution for days where the Mean Temperature was 24 degrees *or* higher, and is based on 68 observations.

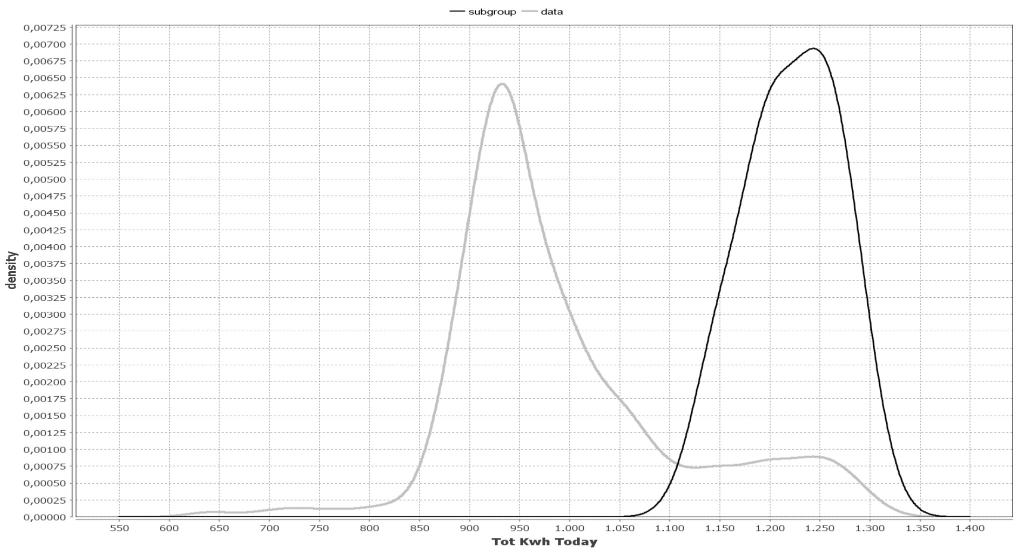

FIGURE 5.1: Distribution of the energy consumption of the subgroup *Mean Temperature* ≥ 24, Macedo Cavaleiros, based on 68 observations

The algorithm noticed that there is a different energy profile when the mean temperature is ≥ 24. Therefore, we decided to test the correlation between Mean Temperature and Energy Consumption. We plotted these variables to analyze their relation. Figure 5.2 shows, that there is a different relation when Mean Temperatures reach values of 24 degrees or above. This supports our previous finding with Cortana, that there are two energy profiles, one when temperature is < 24, and another one when the temperature is ≥ 24. Finally, we produced the same analysis for the other two stores in Fatima, Figure 5.4, and Aveiro, Figure 5.3, and we notice a similar behavior when temperatures reach higher values.

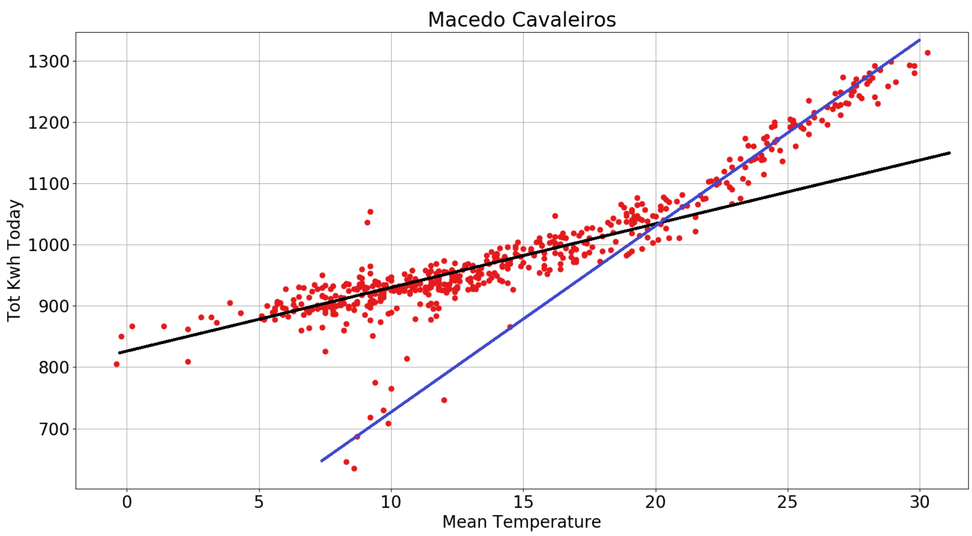

FIGURE 5.2: Distribution of the Total Kwh Today vs Mean Temperature, Macedo Cavaleiros

The patterns detected by subgroup discovery algorithms are not always explained by statistical variation but, rather by external factors [43]. Since we believed that the higher energy consumption is not solely due to mean temperatures of 24 degrees or above, We presented our findings to the Retailer.

The Retailer explained, that the refrigeration machines are programmed to work on a different mode above a user defined temperature, which is mostly above 20 degrees. This threshold varies for different stores, but with SD we are able to detect this threshold from data.

Table 5.1, shows the average temperature per month and store. We can see that the days where the average temperatures are ≥ 20 (highlighted in bold) are mainly during the months of June, July, August, and September.

Store	Jan	Feb	Mar	Apr	May	June	July	Aug	Sep	Oct	Nov	Dec
Aveiro	12	13	14	16	17	**20**	**21**	**21**	19	18	14	14
Fatima	9	11	11	14	15	19	**22**	**22**	**20**	17	12	10
Macedo Cavaleiros	8	10	12	15	16	**22**	**26**	**25**	**22**	16	10	8

TABLE 5.1: Average Temperature per Month and Store

In this study, we used SD for retrieving useful insights for the Retailer. We were able to detect that every store has a different threshold, after which the energy consumption follows a different profile. We believe that this information can be useful to improve our machine learning models in two ways. First, we could decide to split the data, based on this threshold, and develop a different model for when temperature reach higher values. A second option is to create an indicator variable, which can be included in the input dataset for each specific store. It will be used to indicate when the mean temperature is above the threshold for that specific store, e.g., < 24 or ≥ 24.

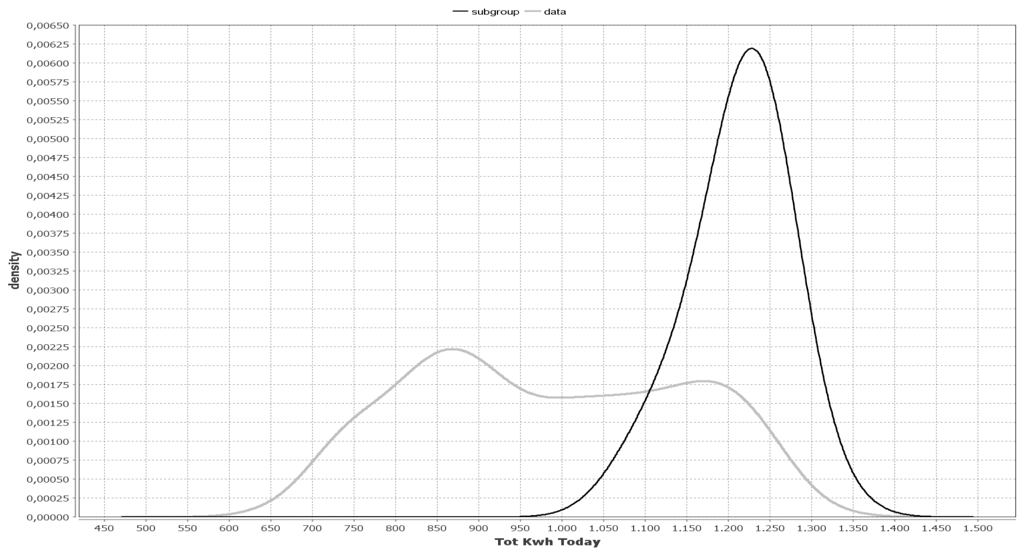

FIGURE 5.3: Distribution of the energy consumption of the subgroup *Mean Temperature* ≥ 20.6, Aveiro, based on 66 observations

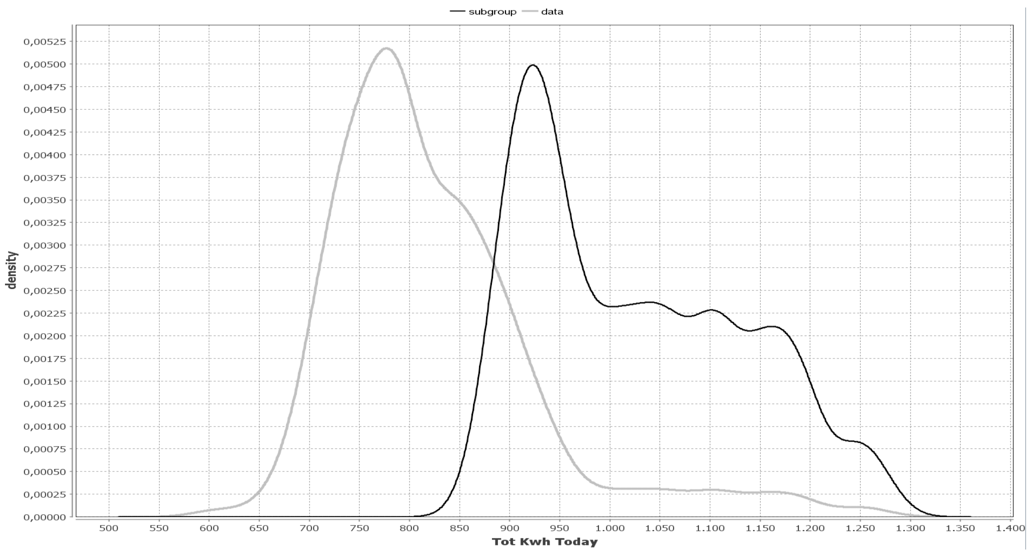

FIGURE 5.4: Distribution of the energy consumption of the subgroup *Mean Temperature* ≥ 20.3, Fatima, based on 62 observations

5.2 Minimum number of days to create a Baseline

In this section, we empirically test the minimum number of days needed to create a reliable baseline. We also test different machine learning algorithms, ANN, MLR, and RF to find out which one can create this baseline with the least amount of training days.

The plots in Figure 5.5, 5.6 and 5.7, show the learning curves of the associated machine learning methods for the store Macedo Cavaleiros. When we applied this method, on the data from the other stores, it showed similar learning curves, for simplicity we decided to show plots from only one store. The approach that we use here, is described in Section 4.3.1 the blue colors, in Figures 5.5, 5.6 and 5.7, are related to the ones in Figure 4.3.

Our analysis of the learning curves in Figure 5.5 shows, that the error at 10 days is quite high. We further observe, that in the first 30 days, the error decreases dramatically when more data is used to train the machine learning model. The learning curves also reveal, that there is a local minimum around the 30 training days. From this point on, the average MAE increases for almost all training sets, alongside the number of training days. Therefore, we decide that 30 days are enough for the MLR to make its best predictions.

When we look at Figure 5.6, we notice that the learning curves show a quite stable behavior. Given our findings, it is hard to determine a point in time where this RF performs best.

An interesting development of learning curves can be seen in Figure 5.7. The overall trend is that the error, for the ANN, decreases when more training examples are included in the training set. We believe that this trend will continue when more training examples are included.

Taken together, our experiments show that the MLR has a clear advantage over the other two methods for creating a baseline with a minimum amount of days. We repeated this approach on all the other stores and the results are discussed next.

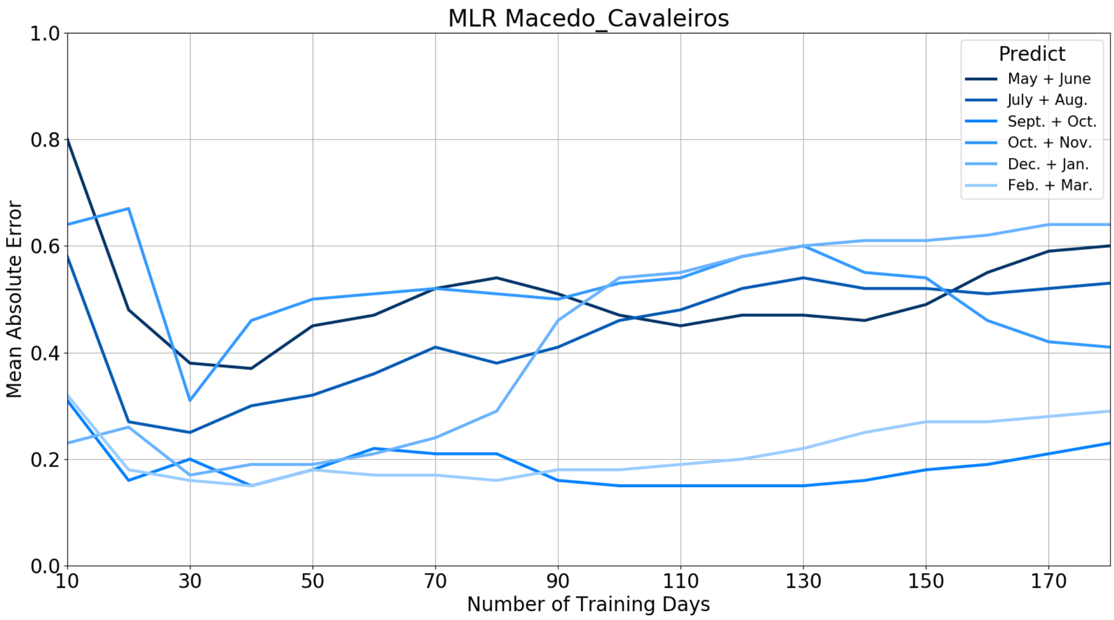

Figure 5.5: Example of MLR prediction performance for Macedo Cavaleiros

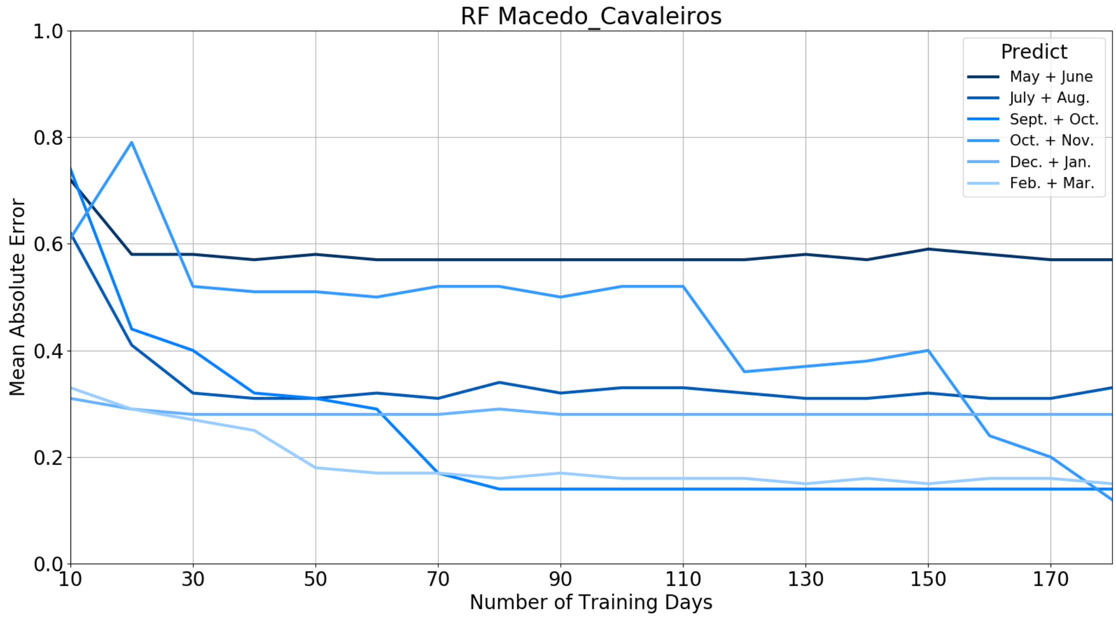

FIGURE 5.6: Example of RF prediction performance for Macedo Cavaleiros

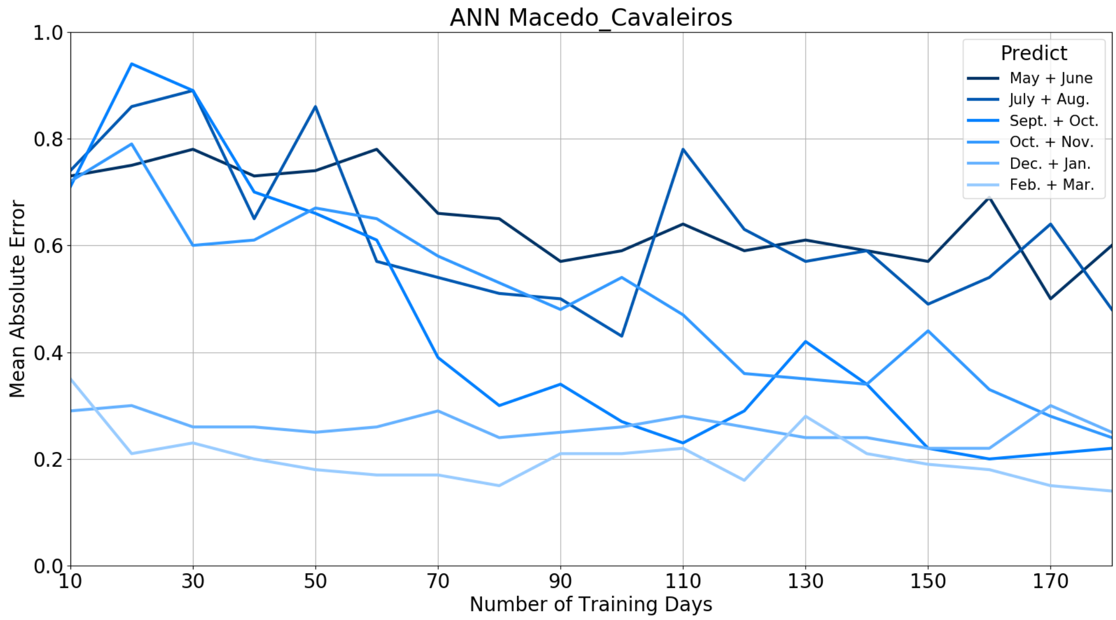

FIGURE 5.7: Example of ANN prediction performance for Macedo Cavaleiros

In Figure 5.8, we see how the error evolves as we train the model with more and more days. This plot displays the learning curves obtained for each of the trained models, MLR, ANN, and RF. The number of training examples ranged from 10 up to 180 days, with threads of 10, and have been tested for a period of 50 days. Each line represents the mean of 18 iterations, for all stores, Aveiro, Fatima, and Macedo Cavaleiros, we performed six iterations regarding the method described in 4.3.1.

In Figure 5.8, we observe that the MLR is the most reliable by a number of 30 days with a MAE of 0.25. Besides, we observe that using the MLR, as we expand the size of training examples, there is an increase in the MAE. Furthermore, we perceive a different behavior for the other two learning models. We see that the performance of the RF stabilizes when we increase the training data following 70 training examples up to 180. Moreover, we remark that the ANN exhibits a continuous reduction in the MAE when more training examples, up to 180, are attached to the training set.

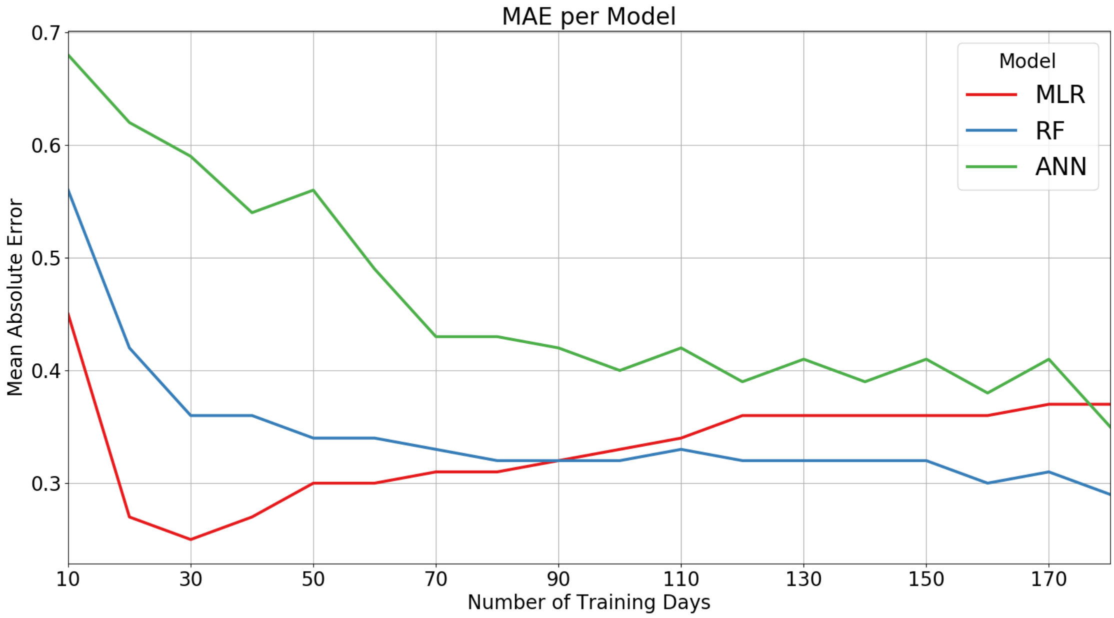

Figure 5.8: Learning curves, based on an average for all stores and methods

The learning curves in Figure 5.8, reveal that each of the learning models is affected differently by the change in the training set size. We notice that the MLR outperforms the other two methods, for making a reliable baseline using the least amount of days. Furthermore, we see that the performance of the MLR worsens when we increase the number of training examples. This can be explained by the nonstationary nature of the datasets. This non-stationarity is a problem for the MLR since it has difficulties with nonlinear relationships. Because the MLR works well with a smaller number of training examples, we assume that the dataset contains periods of local stationarity. One study [87], shows that it is possible that nonstationary time series appear stationary when examined close up. In this local period, the statistical properties change slowly over time. As a consequence, the data that lies close to the forecast period is more likely to be predictive for this forecast period.

For the ANN and RF, stationarity is irrelevant since they are able to handle more complex, nonlinear relations. We see evidence for this in our results, there is a promising development over time in the associated learning curve. We believe that with more diverse data, the ANN could be able to predict a baseline with less number of training days than the MLR. Unfortunately, we were not able to investigate this further.

5.3 Study of the Update Frequency

This section studies, how frequent the model has to be updated when its trained based on the recommendations of Section 5.2. With that, we are able to propose an update frequency required to keep the baseline reliable.

As shown in Section 5.2, we are able to create a reliable model with the MLR trained on 30 days. Therefore, we trained the MLR for each of the stores during the same period of the year, March 2016, and we estimated the energy consumption for the period of one year, from April 2016, until February 2017.

Figure 5.9, shows the evolution of the MAE throughout this period. We observe that during the first 30 days of predictions, the MAE remains quite low, under 0.5. As mentioned in Section 3.1.3, this is somewhat expected, because the prediction period is close to the period used for training.

In Figure 5.9, we see that during the period between 50 till 180 days, the MAE is higher for all the stores. As a matter of fact, this period represents the months June, July, August, and September. Table 5.1 shows, that throughout these months, temperature levels reach higher values than in March, the period that was used for training the model. Our findings in Section 5.1, show that the energy consumption follows a different profile when average temperatures are higher than 20 degrees. This explains why the MAE is higher. To avoid this problem, we could train a different model for each of the two energy profiles identified in Section 5.1. Because our dataset is limited, we were not able to test this in practice.

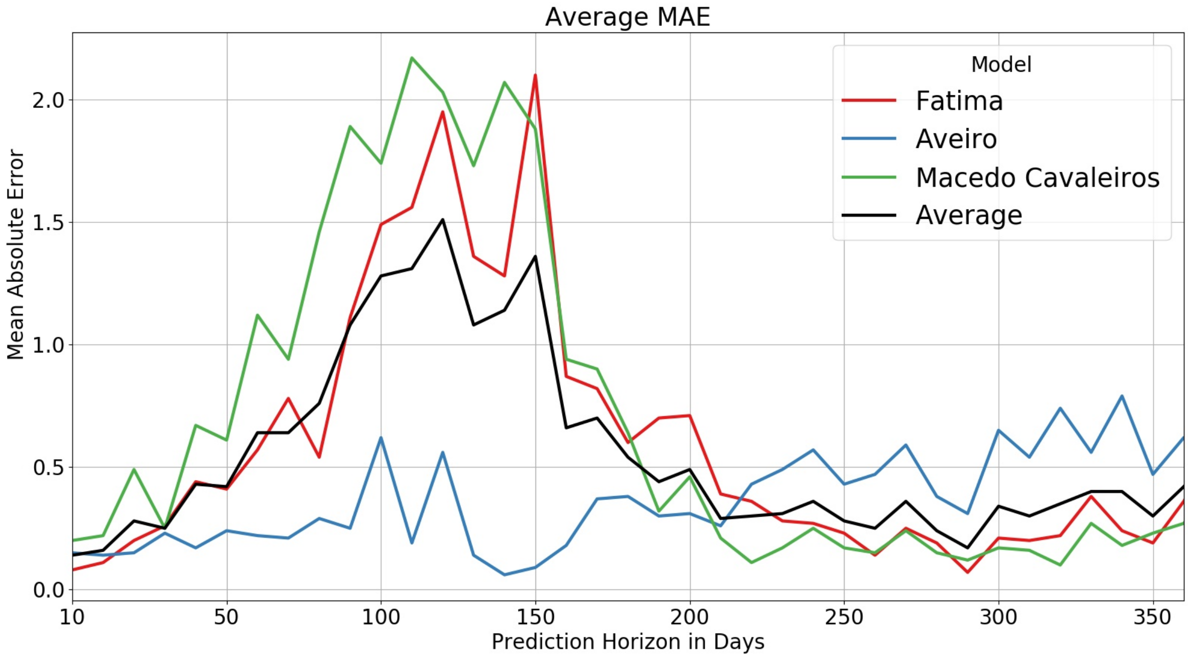

Figure 5.9: MAE over time using MLR

We observe, in Figure 5.9, that in Aveiro the influence of seasonality is less evident than for the supermarkets in Fatima and Macedo Cavaleiros. Since all stores are trained and tested with the same model and in the same period of time, the most plausible factor, for this, are the variables that are related to Temperature. The average temperatures of the three stores follow a similar pattern, higher in the summer and lower in the winter. However, if we focus on the amplitudes of the average temperatures per month, (Table 5.1), we observe that Aveiro registered the smallest amplitude, with a difference of $9C^{\circ}$. The other stores, Fatima and Macedo Cavaleiros, noted an amplitude of $13C^{\circ}$ and $18C^{\circ}$ respectively. This seems to explain why the model trained for the store of Aveiro, is less affected by seasonality.

In Figure 5.9, we notice that after 220 days the accuracy of the model increases again. When we look at Table 5.1, we see that the temperature values from November on, are comparable to the ones in March. Nevertheless, the error is still higher than in the period of the first 30 days.

We applied this method in different periods of time, and we perceived similar behavior.

In conclusion, we base our decision on the average prediction. Figure 5.9 shows that the average prediction remains stable until 30 days, therefore, we recommend updating the model up to 30 days.

5.4 Study of Prediction Accuracy over time

This section describes how the learning models perform for making long term predictions when we use more than 180 days of data to train. Finally, in Section 5.5, two other stores will be used to test if these predictions are accurate enough to estimate energy savings.

Each store has a different number of observations, and they are also collected in different periods of time. We will train the MLR, RF, and ANN with the first 180 and 360 days of data, and test for the remaining days. We will do this for the stores located in Aveiro, Fatima, and Macedo Cavaleiros. Therefore, we train each store in different periods, and not within the same period.

In Section 5.3, we noticed that 30 training days were not enough to make accurate long term predictions. Therefore, we decide to include more training days into our training set. Each of the following plots, in Figures 5.10, 5.11, 5.12, 5.13, 5.14, and 5.15, show how the prediction error evolves over time, per store, per model and number of training days. Each point shows the average error for 10 subsequent predictions.

Figures 5.10, 5.11, and 5.12 show the evolution of the prediction error when the models are trained on the first 180 days of data. We observe, that each store shows a similar behavior as described in Section 5.3. This is more evident when we compare the error of the MLR (red line) with the error in Figure 5.9. Overall, the MAE is lower for the stores of Fatima and Macedo Cavaleiros, if we use 180 days instead of 30 days. These results also show, that the effect of the different consumption modes is still visible, but less dramatically.

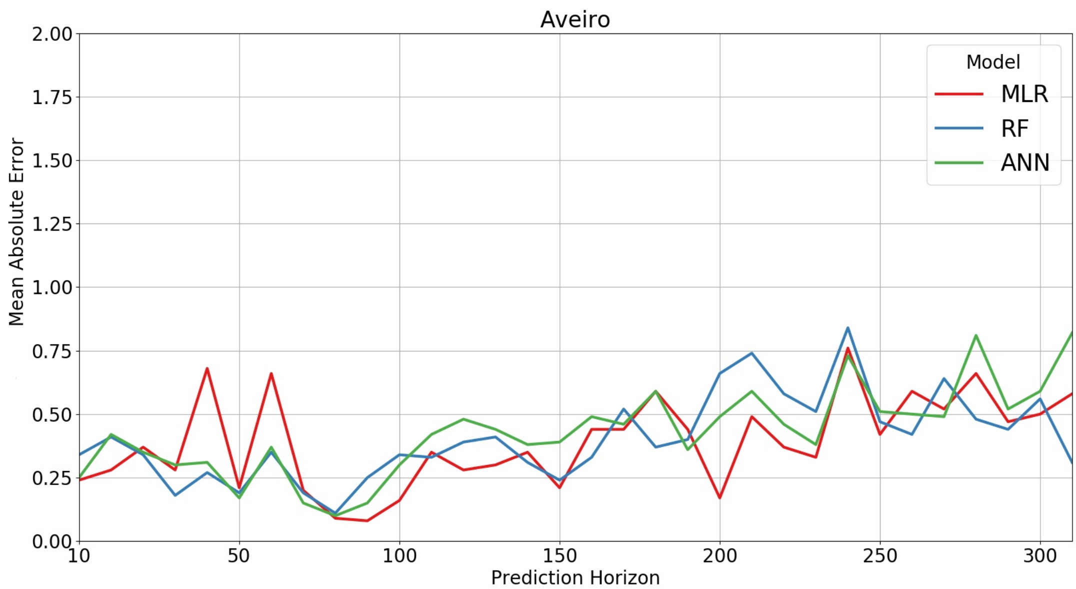

FIGURE 5.10: MAE over time using 180 training days, Aveiro

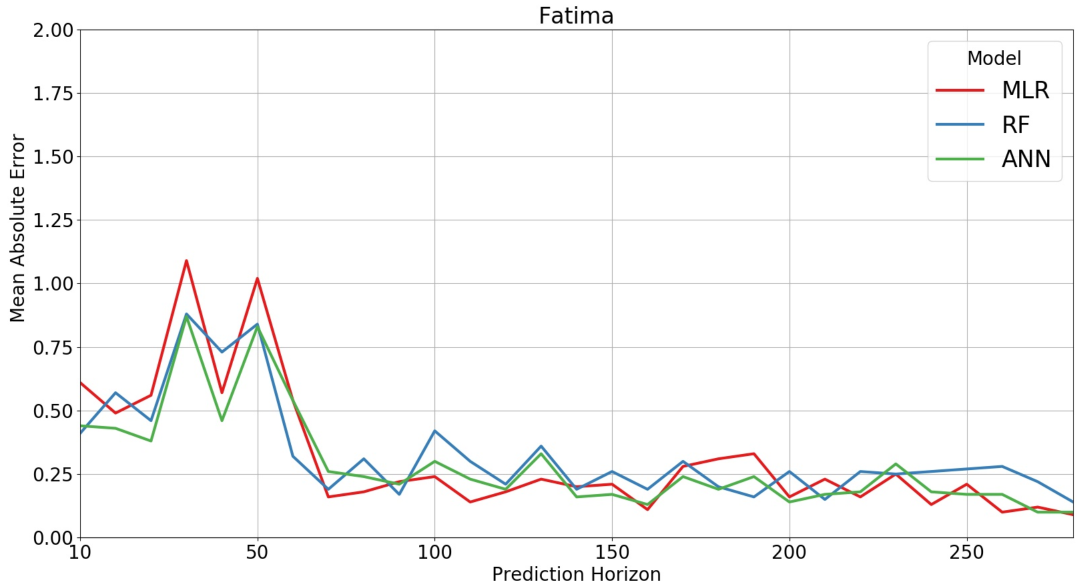

FIGURE 5.11: MAE over time using 180 training days, Fatima

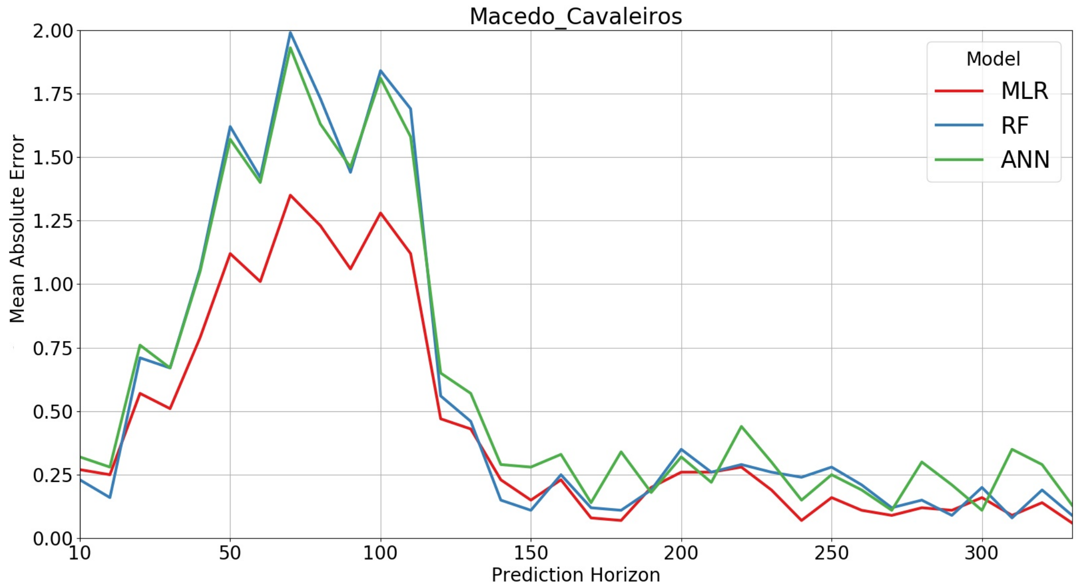

FIGURE 5.12: MAE over time using 180 training days, Macedo Cavaleiros

We expect that long term predictions become more accurate when we use 360 training days to train the model because the model is trained with data from all periods of the year. Because we use this number of training days, a bigger variation of temperature values is included in the training set. Therefore, we decided to train the models, for all stores, on the first 360 training days and study the predictions on the remaining days. Figures 5.13, 5.14, and 5.15 show us how the MAE error evolves for this period of time. We observe, that the for the corresponding period of time, the MAE is a bit lower than for the models trained on 180 days.

In contrast to Section 5.2, that MLR has the worst performance, while the RF and ANN perform somewhat similar. The results of this experimental part supports the general idea that when we train the models with more data, our predictions will improve.

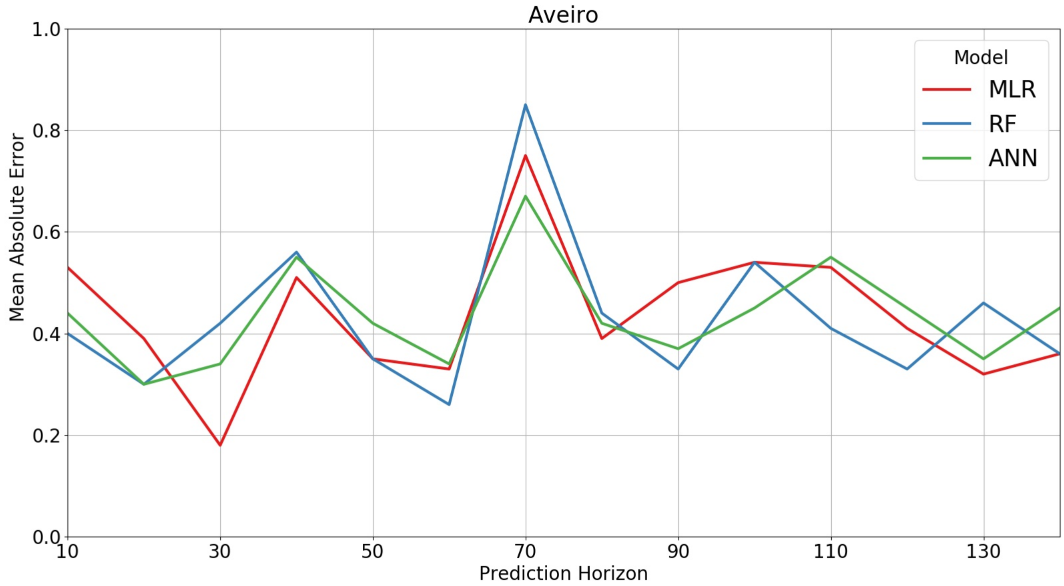

FIGURE 5.13: MAE over time using 360 training days, Aveiro

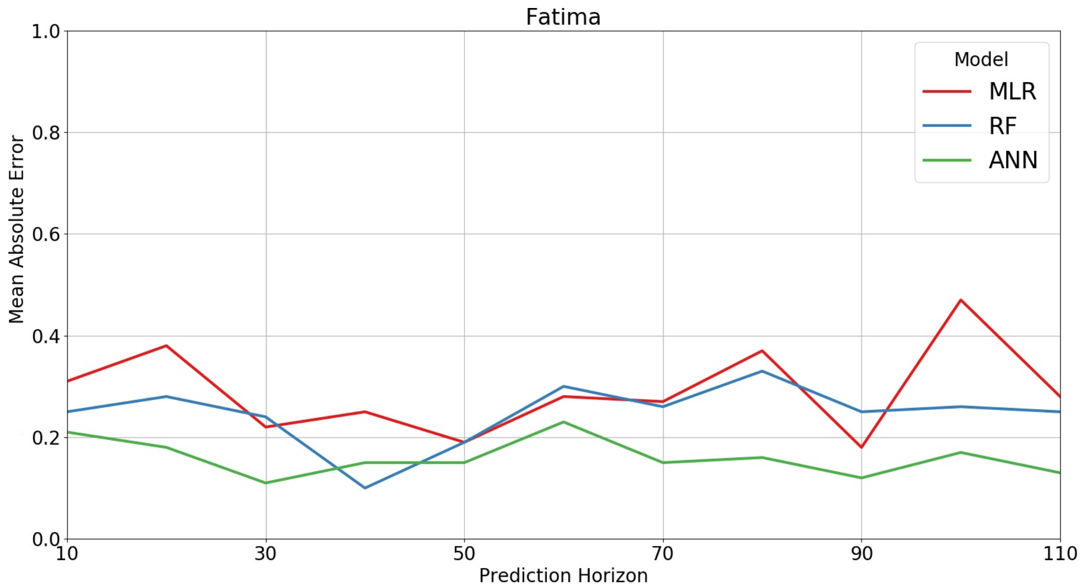

FIGURE 5.14: MAE over time using 360 training days, Fatima

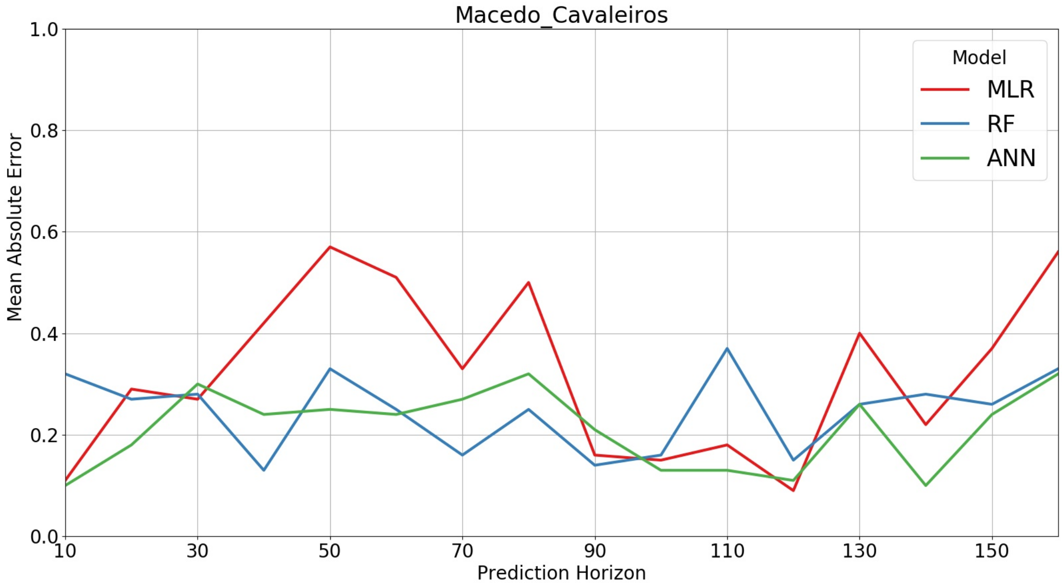

FIGURE 5.15: MAE over time using 360 training days, Macedo Cavaleiros

When the algorithms are trained with 180 training days, the effect of the different energy consumption modes is still visible. When we use 360 training days, we observe that the predictions become more accurate. Therefore, we advice to train algorithms on 360 training days to create long term predictions.

5.5 Estimate Energy Savings

The Retailer wants to estimate, with reasonable accuracy, the energy savings resulting from its energy policies. Changes in energy policies, such as the retrofitting an equipment, require high investments. This makes it important for the Retailer to know if the investments are truly effective, in the reduction of energy consumption. If we use a baseline trained with data before some measure is implemented, we can estimate the energy savings by comparing its estimates with the observed consumption.

We selected two stores that have undergone a retrofitting of the equipment. From these stores exactly one year of data is available. Mangualde and Regua had, respectively, 170 and 200 training days available before the Retrofit. Because we have less than a year of data available, we decide to use the MLR, trained on 30 days, as described in Section 5.2.

Figures 5.16 and 5.17 show the observed consumption (orange lines) versus the baseline estimates (blue lines) for these two stores. We trained the MLR for both stores, on 30 training days, between 50 and 20 days before the Retrofit and we predicted for 50 days.

This makes it easier to visualize how the baseline compares with the energy consumption before and after the Retrofit.

The deviations, between the baseline and the energy consumption, can result from poor prediction performance or energy savings/losses. We chose a setup that gives us a reliable baseline, therefore, we believe that the deviations are caused by energy savings. In both Figures 5.16 and 5.17, we observe that, before the Retrofit, the baseline and the real energy consumption intertwine in several points. This behavior, which was also seen before, shows that the predictions are close to the real consumption. After the Retrofit, however, the observed consumption is always lower than the prediction, which offers strong evidence that the implemented measure was effective.

Hence, if we assume that the baseline is accurate enough, we can estimate the energy savings using the difference between the predicted and observed energy consumption.

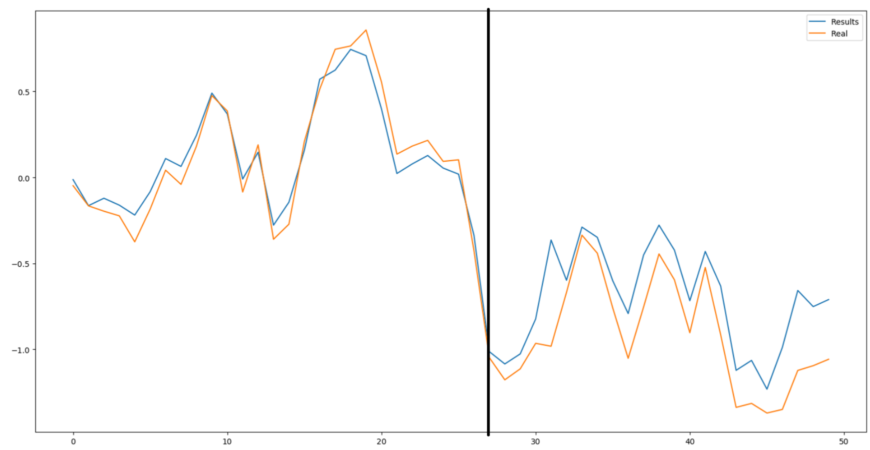

FIGURE 5.16: Example of the predicted and observed Energy Consumption, Mangualde

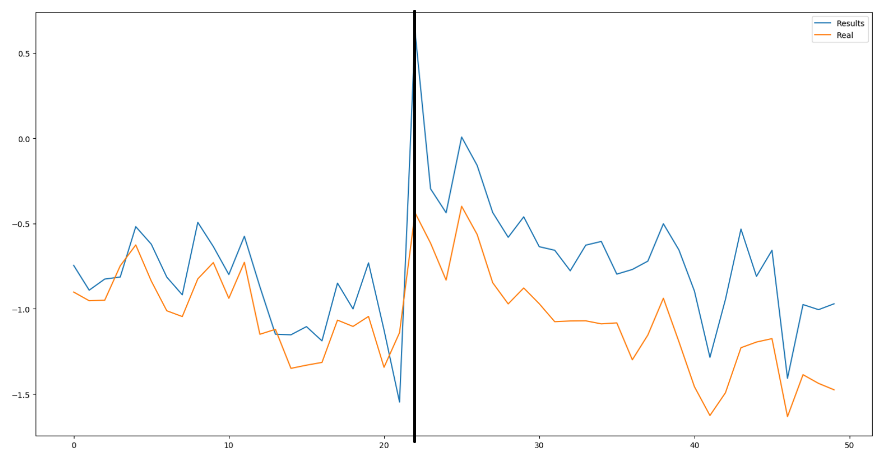

FIGURE 5.17: Example of the predicted and observed Energy Consumption, Regua

Conclusions & Future Work

Several reasons can keep a company from investing in energy efficiency measures. For example, when inadequate information is available about the results of investments in energy efficiency measures. Energy management can focus on addressing these factors that prevent a company to make investments.

Energy efficiency measures can require high investments. This makes it important for the Retailer to know if the investments are truly effective, in reducing energy consumption. Energy baselines can be used to study the effectiveness of energy efficiency measures. The results can simplify decisions to reserve funding for the required investments in other stores.

In this study, we researched if off-the-shelf data science technologies can be used to create energy baselines that support improved energy management. Before that, we also performed some exploratory analysis to better understand the data.

With our exploratory analysis, we were able to detect that every store has a distinct threshold for the activation of a different energy consumption mode. As temperatures pass this threshold, the energy consumption of the store follows a different profile. The Retailer was able to explain that this is due to the refrigeration machines that are programmed to work on a different mode above a user defined temperature, which is mostly above 20 degrees.

Our first goal, was to determine the minimum amount of training days needed to create a reliable baseline, and which model performs best. For that, we studied the prediction accuracy of three machine learning models, ANN, RF, and MLR, based on various datasets. For the experiments, we proposed a sliding window approach in which we systematically expanded the size of the training set with historical data. Our experiments show, that the MLR has a clear advantage over the other two methods for creating a baseline with a minimum amount of days. This model needs 30 training days to estimate a reliable baseline.

The second goal was to determine how often the algorithm needs to be updated when trained with a MLR on 30 training days. We trained our algorithm multiple times, on all stores, and in different time periods. Our analysis shows that the MAE stays low for a period of 30 days, after this the MAE dramatically increases. Moreover, we observed that the energy consumption follows a different profile when average temperatures are higher than 20 degrees. These findings are in line with our insights derived from Subgroup Discovery. Our analysis shows, that the amplitude of the average temperature affects the prediction performance. Hence, we advise updating the model up to 30 days.

Our third goal, was to determine if we can estimate energy savings after implementing an energy efficiency measure. To answer this question, we trained our models with 180 and 360 training days and predicted for the remaining days. Our findings show, that the predictions become the most accurate when trained with 360 training days. Because we use 360 training days, a bigger variation of temperature values is included in the training set. This supports the general idea that when we train the models with more data, our predictions will improve. With a baseline, trained on 360 training days, the Retailer is able to estimate, with reasonable accuracy, the energy savings resulting from its energy policies. Moreover, he can compare the energy savings to the investment made for the measure. This has obvious advantages for the retailer.

In summary, the results of this study show that we have been able to create reliable energy baselines using off-the-shelf data science technologies. Moreover, we found a way to create them based on short term historical data.

6.1 Future Work

Further work needs to be done to establish whether *transfer learning* can be applied, to enable baselines to be calculated quicker. In traditional machine learning, a predictive model is constructed from scratch. Transfer learning attempts to transfer learning from a previous task to a new task [88]. Thus, it would be interesting to examine to what extent a model could be trained on other stores, and implemented in a new store. In our experiments we observed that the ANN needs more data than the MLR, to construct a good model. Therefore, transfer learning could enable us, by making use of the powerful capability of Neural networks in handling nonlinear relations, to establish baselines quicker than the MLR can do now.

The second direction of research includes fine tuning Neural networks to optimize their performance. Our experiments show, that recent data is more likely to be predictive for future predictions. Therefore, we suggest examining the utilization of time series algorithms that make use of exponential smoothing. Furthermore, we are also curious how a Long Short Term Memory (LSTM) algorithm could be applied in this context [89].

With SD, we were able to detect that every store has a different threshold, after which the energy consumption follows a different profile. We believe that this information can be useful to improve our machine learning models in two ways. First, we could decide to split the data, based on this threshold, and develop a different model for when temperature reach higher values. A second option is to create an indicator variable, which can be included in the input dataset, it will indicate when the mean temperature is, e.g., < 20 degrees or ≥ 24. We think that an important issue for future research, is to determine how one of these options can improve the predictive performance.

A unified set of input data needs to be generated to construct a model that can be transferred towards different stores [10, 88]. We believe that including store specific variables could lead to achieving better prediction performance. Therefore, we are interested in researching which supplementary data can be collected, and how.

Also, we view that analyzing the significance of each variable, and how this is affected by energy measures, creates valuable information for the Retailer. This is an important aspect for future research. This knowledge can be used to perform a 'Retrofit' analysis, this specifies which stores benefit most from specific upgrades, i.e., saves the most energy.

Finally, we recommend that further research should be undertaken to study if the baselines can become solid enough to detect equipment faults. With our current models, we can only detect equipment failures when the energy consumption shows extremely unusual behavior. In [15], a baseline is designed that can detect system faults based on energy consumption. Deviations from their baselines are interpreted as prediction errors or faults in the equipment [15]. They designed a method, where, if five consecutive data points deviate from the baseline, according to a pre-determined percentage, the system warns for a possible failure. Furthermore, this information serves for setting up a maintenance plan to prevent equipment faults or help to react quickly when faults occur.

Bibliography

[1] IEA. Energy policies of iea countries; portugal 2016, 2016. URL https://www.iea.org/publications/freepublications/publication/energy-policies-of-iea-countries---portugal-2016-review.html. [Online; accessed 14-April-2017].

[2] Portugal 2020. Portugal 2020 what is it?, n.d. URL https://poseur.portugal2020.pt/en/portugal-2020/. [Online; accessed 16-July-2017].

[3] Portugal 2020. Approved applications, n.d. URL https://poseur.portugal2020.pt/en/applications/approved-applications/. [Online; accessed 16-July-2017].

[4] Patrik Thollander, Sandra Backlund, Andrea Trianni, and Enrico Cagno. Beyond barriers a case study on driving forces for improved energy efficiency in the foundry industries in finland, france, germany, italy, poland, spain, and sweden. *Applied Energy*, 111:636 – 643, 2013. ISSN 0306-2619. doi: http://dx.doi.org/10.1016/j.apenergy.2013.05.036. URL http://www.sciencedirect.com/science/article/pii/S0306261913004376.

[5] Ernst Worrell, John A Laitner, Michael Ruth, and Hodayah Finman. Productivity benefits of industrial energy efficiency measures. *Energy*, 28(11):1081 – 1098, 2003. ISSN 0360-5442. doi: http://dx.doi.org/10.1016/S0360-5442(03)00091-4. URL http://www.sciencedirect.com/science/article/pii/S0360544203000914.

[6] Patrik Thollander and Jenny Palm. Industrial energy management decision making for improved energy efficiencystrategic system perspectives and situated action in combination. *Energies*, 8 (6):5694–5703, 2015. ISSN 1996-1073. doi: 10.3390/en8065694. URL http://www.mdpi.com/1996-1073/8/6/5694.

[7] Kenneth Gillingham and Karen Palmer. Bridging the energy efficiency gap: Policy insights from economic theory and empirical evidence. *Review of Environmental Economics and Policy*, 8(1):18, 2014. doi: 10.1093/reep/ret021. URL http://dx.doi.org/10.1093/reep/ret021.

[8] Lelde Timma, Roberts Skudritis, and Dagnija Blumberga. Benchmarking analysis of energy consumption in supermarkets. *Energy Procedia*, 95:435 – 438, 2016. ISSN 1876-6102. doi: http://dx.doi.org/10.1016/j.egypro.2016.09.056. URL http://www.sciencedirect.com/science/article/pii/S1876610216306956. International Scientific Conference Environmental and Climate Technologies, CONECT 2015.

[9] Garcia and Coelho. Energy efficiency strategies in refrigeration systems of large supermarkets. *International Journal of Energy, Environment and Economics*, 4(3):6470, 2010.

[10] S. A. Tassou Z. Mylona, M. Kolokotroni. Frozen food retail: Measuring and modelling energy use and space environmental systems in an operational supermarket. *Energy and Buildings*, 144: 129 – 143, 2017. ISSN 0378-7788. doi: http://doi.org/10.1016/j.enbuild.2017.03.049. URL http://www.sciencedirect.com/science/article/pii/S0378778817309544.

[11] Edite Kamendere, Gatis Zogla, Agris Kamenders, Janis Ikaunieks, and Claudio Rochas. Analysis of mechanical ventilation system with heat recovery in renovated apartment buildings. *Energy Procedia*, 72:27 – 33, 2015. ISSN 1876-6102. doi: http://dx.doi.org/10.1016/j.egypro.2015.06.005. URL http://www.sciencedirect.com/science/article/pii/S1876610215006955. International Scientific Conference Environmental and Climate Technologies, CONECT 2014.

[12] Kristaps Kass, Andra Blumberga, and Gatis Zogla. Energy performance of historical brick buildings in northern climate zone. *Energy Procedia*, 72:238 – 244, 2015. ISSN 1876-6102. doi: http://dx.doi.org/10.1016/j.egypro.2015.06.034. URL http://www.sciencedirect.com/science/article/pii/S1876610215007249. International Scientific Conference Environmental and Climate Technologies, CONECT 2014.

[13] Liva Asere and Andra Blumberga. Government and municipality owned building energy efficiency system dynamics modelling. *Energy Procedia*, 72:180 – 187, 2015. ISSN 1876-6102. doi: http://dx.doi.org/10.1016/j.egypro.2015.06.026. URL http://www.sciencedirect.com/science/article/pii/S187661021500716X. International Scientific Conference Environmental and Climate Technologies, CONECT 2014.

[14] Henrik B. Christensen. The low-energy supermarket project. *Heating/Piping/Air Conditioning Engineering*, 75(12):48 – 51, 2003. ISSN 15274055. URL http://search.ebscohost.com.ezproxy.leidenuniv.nl:2048/login.aspx?direct=true&db=a2h&AN=11657670&site=ehost-live.

[15] N. Shah G. Mavromatidis, S.Acha. Diagnostic tools of energy performance for supermarkets using artificial neural network algorithms. *Energy and Buildings*, 62:304 – 314, 2013. ISSN 0378-7788. doi: https://doi.org/10.1016/j.enbuild.2013.03.020. URL http://www.sciencedirect.com/science/article/pii/S0378778813001886.

[16] S.A. Tassou, Y. Ge, A. Hadawey, D. Marriott. Energy consumption and conservation in food retailing. *Elsevier*, (31), 2011.

[17] P. Stluka M. Hrnčár. Performance monitoring strategies for effective running of commercial refrigeration systems. In *Proceedings of the 12th WSEAS International Conference on Automatic Control, Modelling & Simulation*, ACMOS'10, pages 177–180, Stevens Point, Wisconsin, USA, 2010. World Scientific and Engineering Academy and Society (WSEAS). ISBN 978-954-92600-1-4. URL http://dl.acm.org/citation.cfm?id=1844174.1844209.

[18] Mike Schulze, Henrik Nehler, Mikael Ottosson, and Patrik Thollander. Energy management in industry a systematic review of previous findings and an integrative conceptual framework. *Journal of Cleaner Production*, 112:3692 – 3708, 2016. ISSN 0959-6526. doi: http://dx.doi.org/10.1016/j.jclepro.2015.06.060. URL http://www.sciencedirect.com/science/article/pii/S0959652615007891.

[19] Thomas C. Mort CEM. The energy management process. *Strategic Planning for Energy and the Environment*, 21(1):7 – 11, 2001. doi: https://doi.org/10.1080/10485230109509566. URL http://www-tandfonline-com.ezproxy.leidenuniv.nl:2048/doi/citedby/10.1080/10485230109509566?scroll=top&needAccess=true.

[20] Luca Falasconi Marco Pagani, Matteo Vittuari. Does packaging matter? energy consumption of pre-packed salads. *British Food Journal*, 117(7):1961–1980, 2015. doi: 10.1108/BFJ-10-2014-0348. URL http://dx.doi.org/10.1108/BFJ-10-2014-0348.

[21] J. Cai, J. Stoustrup, and B.D. Rasmussen. An active defrost scheme with a balanced energy consumption and food quality loss in supermarket refrigeration systems. *IFAC Proceedings Volumes*, 41(2):9374 – 9379, 2008. ISSN 1474-6670. doi: http://dx.doi.org/10.3182/20080706-5-KR-1001.01584. URL http://www.sciencedirect.com/science/article/pii/S1474667016404611.

[22] S.A. Tassou Y.T. Ge. Performance evaluation and optimal design of supermarket refrigeration systems with supermarket model supersim, part i: Model description and validation. *International Journal of Refrigeration*, 34(2):527 – 539, 2011. ISSN 0140-7007. doi: http://doi.org/10.1016/j.ijrefrig.2010.11.010. URL http://www.sciencedirect.com/science/article/pii/S0140700710002719.

[23] E. T. Mansur M. Auffhammer. Measuring climatic impacts on energy consumption: A review of the empirical literature. *Energy Economics*, 46:522 – 530, 2014. ISSN 0140-9883. doi: http://doi.org/10.1016/j.eneco.2014.04.017. URL http://www.sciencedirect.com/science/article/pii/S0140988314001017.

[24] S. Beck M. Braun, H. Altan. Using regression analysis to predict the future energy consumption of a supermarket in the uk. *WIT Transactions on Ecology and the Environment*, 176:313, 2013. doi: 10.2495/ESUS130101.

[25] Anu Maria. Introduction to modeling and simulation. In *Proceedings of the 29th conference on Winter simulation, WSC 1997, Atlanta, GA, USA, December 7-10, 1997*, pages 7–13, 1997. doi: 10.1109/WSC.1997.640371. URL http://doi.ieeecomputersociety.org/10.1109/WSC.1997.640371.

[26] Kaile Zhou, Chao Fu, and Shanlin Yang. Big data driven smart energy management: From big data to big insights. *Renewable and Sustainable Energy Reviews*, 56:215 – 225, 2016. ISSN 1364-0321. doi: http://dx.doi.org/10.1016/j.rser.2015.11.050. URL http://www.sciencedirect.com/science/article/pii/S1364032115013179.

[27] Tianzhen Hong, Sarah C. Taylor-Lange, Simona DOca, Da Yan, and Stefano P. Corgnati. Advances in research and applications of energy-related occupant behavior in buildings. *Energy and Buildings*, 116:694 – 702, 2016. ISSN 0378-7788. doi: http://dx.doi.org/10.1016/j.enbuild.2015.11.052. URL http://www.sciencedirect.com/science/article/pii/S0378778815005307.

[28] Zeyu Wang and Ravi S. Srinivasan. A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models. *Renewable and Sustainable Energy Reviews*, 75:796 – 808, 2017. ISSN 1364-0321. doi: http://dx.doi.org/10.1016/j.rser.2016.10.079. URL http://www.sciencedirect.com/science/article/pii/S1364032116307420.

[29] G. E. Nasr, E. A. Badr, and M. R. Younes. Neural networks in forecasting electrical energy consumption. pages 489–492, 2001. URL http://dl.acm.org/citation.cfm?id=646814.708150.

[30] Majeed Safa, Mahdi Safa, Jeremy Allen, Arash Shahi, and Carl T. Haas. Improving sustainable office building operation by using historical data and linear models to predict energy usage. *Sustainable Cities and Society*, 29:107 – 117, 2017. ISSN 2210-6707. doi: http://dx.doi.org/10.1016/j.scs.2016.12.001. URL http://www.sciencedirect.com/science/article/pii/S2210670716305212.

[31] Andreas Chari and Symeon Christodoulou. Building energy performance prediction using neural networks. 2017. doi: DOI:10.1007/s12053-017-9524-5. URL https://www.researchgate.net/publication/316440622_Building_energy_performance_prediction_using_neural_networks.

[32] James Douglas Hamilton. *Time series analysis*, volume 2. Princeton university press Princeton, 1994.

[33] João Gama. *Knowledge Discovery from Data Streams*. Chapman and Hall / CRC Data Mining and Knowledge Discovery Series. CRC Press, 2010. ISBN 978-1-4398-2611-9. URL http://www.crcpress.com/product/isbn/9781439826119.

[34] Richa Handa H.S. Hota and A.K. Shrivas. Time series data prediction using sliding window based rbf neural network. *International Journal of Computational Intelligence Research*, 13(5):1145 – 1156, 2017. ISSN 0973-1873. URL https://www.ripublication.com/ijcir17/ijcirv13n5_47.pdf.

[35] Chris Chatfield. *The Analysis of Time Series: An Introduction, Sixth Edition*, pages 1–352. Chapman and Hall/CRC, 2003. ISBN 9781584883173. URL https://www.crcpress.com/The-Analysis-of-Time-Series-An-Introduction-Sixth-Edition/Chatfield/p/book/9781584883173.

[36] Carsten Jentsch and Suhasini Subba Rao. A test for second order stationarity of a multivariate time series. *Journal of Econometrics*, 185(1):124 – 161, 2015. ISSN 0304-4076. doi: http://dx.doi.org/10.1016/j.jeconom.2014.09.010. URL http://www.sciencedirect.com/science/article/pii/S030440761400195X.

[37] Sarah J. Halliday, Andrew J. Wade, Richard A. Skeffington, Colin Neal, Brian Reynolds, Philip Rowland, Margaret Neal, and Dave Norris. An analysis of long-term trends, seasonality and short-term dynamics in water quality data from plynlimon, wales. *Science of The Total Environment*, 434:186 – 200, 2012. ISSN 0048-9697. doi: http://dx.doi.org/10.1016/j.scitotenv.2011.10.052. URL http://www.sciencedirect.com/science/article/pii/S0048969711012460. Climate Change and Macronutrient Cycling along the Atmospheric, Terrestrial, Freshwater and Estuarine Continuum - A Special Issue dedicated to Professor Colin Neal.

[38] Chris Chatfield. *Time-series forecasting*. CRC Press, 2000.

[39] Leon de Bruin and John Michael. Prediction error minimization: Implications for embodied cognition and the extended mind hypothesis. *Brain and Cognition*, 112:58 – 63, 2017. ISSN 0278-2626. doi: http://dx.doi.org/10.1016/j.bandc.2016.01.009. URL http://www.sciencedirect.com/science/article/pii/S0278262616300094. Perspectives on Human Probabilistic Inferences and the 'Bayesian Brain'.

[40] Thomas D. Sambrook and Jeremy Goslin. Mediofrontal event-related potentials in response to positive, negative and unsigned prediction errors. *Neuropsychologia*, 61:1 – 10, 2014. ISSN 0028-3932. doi: http://dx.doi.org/10.1016/j.neuropsychologia.2014.06.004. URL http://www.sciencedirect.com/science/article/pii/S0028393214001857.

[41] Michael P. Clements. Forecast uncertaintyex ante and ex post: U.s. inflation and output growth. *Journal of Business & Economic Statistics*, 32(2):206–216, 2014. doi: 10.1080/07350015.2013.859618. URL http://dx.doi.org/10.1080/07350015.2013.859618.

[42] R. Hyndman, A.B. Koehler, J.K. Ord, and R.D. Snyder. *Forecasting with Exponential Smoothing*, pages 1–362. Springer-Verlag Berlin Heidelberg, 2008. ISBN 978-3-540-71918-2. doi: 10.1007/978-3-540-71918-2. URL https://10.1007/978-3-540-71918-2.

[43] Martin Atzmueller. Subgroup discovery. volume 5, pages 35–49, 2015. doi: 10.1002/widm.1144. URL https://doi.org/10.1002/widm.1144.

[44] Fida Hussain Muzammil Khan and Imran Khan. Single level drill down interactive visualization technique for descriptive data mining results. *International Journal of Grid and Distributed Computing*, 7(4):33 – 40, 2014. URL http://www.sersc.org/journals/IJGDC/vol7_no4/4.pdf.

[45] Martin Atzmueller. Subgroup discovery. *Wiley Interdisc. Rew.: Data Mining and Knowledge Discovery*, 5(1):35–49, 2015. doi: 10.1002/widm.1144. URL https://doi.org/10.1002/widm.1144.

[46] Francisco Herrera, Cristobal Jose Carmona, Pedro Gonzalez, and Maria Jose del Jesus. An overview on subgroup discovery: foundations and applications. *Knowledge and Information Systems*, 29(3): 495–525, Dec 2011. ISSN 0219-3116. doi: 10.1007/s10115-010-0356-2. URL https://doi.org/10.1007/s10115-010-0356-2.

[47] D. Roland Thomas, Edward Hughes, and Bruno D. Zumbo. On variable importance in linear regression. *Social Indicators Research*, 45(1):253–275, Nov 1998. doi: 10.1023/A:1006954016433. URL https://doi.org/10.1023/A:1006954016433.

[48] William Menke. Review of the generalized least squares method. *Surveys in Geophysics*, 36(1): 1–25, Jan 2015. ISSN 1573-0956. doi: 10.1007/s10712-014-9303-1. URL https://doi.org/10.1007/s10712-014-9303-1.

[49] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, Oct 2001. ISSN 1573-0565. doi: 10.1023/A:1010933404324. URL https://doi.org/10.1023/A:1010933404324.

[50] C.F. Pinho Rebelo de S. Pattern mining for label ranking. 2016. URL https://doi.org/10.4121/uuid:21b1959d-9196-423e-94d0-53883fb0ff21.

[51] Manuel Fernández-Delgado, Eva Cernadas, Senén Barro, and Dinani Amorim. Do we need hundreds of classifiers to solve real world classification problems? *J. Mach. Learn. Res.*, 15(1):3133–3181, January 2014. ISSN 1532-4435. URL http://dl.acm.org/citation.cfm?id=2627435.2697065.

[52] Gérard Biau. Analysis of a random forests model. *Journal of Machine Learning Research*, 13: 1063–1095, 2012. URL http://dl.acm.org/citation.cfm?id=2343682.

[53] R. Maclin and D. Opitz. Popular Ensemble Methods: An Empirical Study. *ArXiv e-prints*, June 2011.

[54] S. Karatasou, M. Santamouris, and V. Geros. Modeling and predicting building's energy use with artificial neural networks: Methods and results. *Energy and Buildings*, 38(8):949 – 958, 2006. ISSN 0378-7788. doi: http://dx.doi.org/10.1016/j.enbuild.2005.11.005. URL http://www.sciencedirect.com/science/article/pii/S0378778805002161.

[55] Aurlie Foucquier, Sylvain Robert, Frdric Suard, Louis Stphan, and Arnaud Jay. State of the art in building modelling and energy performances prediction: A review. *Renewable and Sustainable Energy Reviews*, 23:272 – 288, 2013. ISSN 1364-0321. doi: http://dx.doi.org/10.1016/j.rser.2013.03.004. URL http://www.sciencedirect.com/science/article/pii/S1364032113001536.

[56] Florian Neukart. Reverse engineering the mind. In *Consciously Acting Machines and Accelerated Evolution*, volume 94. Springer, 2017. doi: 10.1007/978-3-658-16176-7. URL http://dx.doi.org/10.4271/2016-01-0114.

[57] Guoqiang Zhang, B. Eddy Patuwo, and Michael Y. Hu. Forecasting with artificial neural networks:. *International Journal of Forecasting*, 14(1):35 – 62, 1998. ISSN 0169-2070. doi: http://dx.doi.org/10.1016/S0169-2070(97)00044-7. URL http://www.sciencedirect.com/science/article/pii/S0169207097000447.

[58] Claudia Perlich. *Learning Curves in Machine Learning*, pages 577–580. Springer US, Boston, MA, 2010. ISBN 978-0-387-30164-8. doi: 10.1007/978-0-387-30164-8_452. URL https://doi.org/10.1007/978-0-387-30164-8_452.

[59] Pat Langley. Machine learning as an experimental science. *Machine Learning*, 3(1):5–8, Aug 1988. ISSN 1573-0565. doi: 10.1007/BF00115008. URL https://doi.org/10.1007/BF00115008.

[60] Ali K. M. Al-Qurabat and Ali Kadhum Idrees. Distributed data aggregation protocol for improving lifetime of wireless sensor networks. *Qalaai Zanist Scientific Journal*, 04 2017.

[61] Jui-Sheng Chou and Ngoc-Tri Ngo. Time series analytics using sliding window metaheuristic optimization-based machine learning system for identifying building energy consumption patterns. *Applied Energy*, 177:751 – 770, 2016. ISSN 0306-2619. doi: http://dx.doi.org/10.1016/j.apenergy.2016.05.074. URL http://www.sciencedirect.com/science/article/pii/S0306261916306717.

[62] Kailei Liu, Cheng Yao, Ji Chen, Zhijia Li, Qiaoling Li, and Leqiang Sun. Comparison of three updating models for real time forecasting: a case study of flood forecasting at the middle reaches of the huai river in east china. *Stochastic Environmental Research and Risk Assessment*, 31(6): 1471–1484, Aug 2017. ISSN 1436-3259. doi: 10.1007/s00477-016-1267-x. URL https://doi.org/10.1007/s00477-016-1267-x.

[63] Christoph Bergmeir and Jos M. Bentez. On the use of cross-validation for time series predictor evaluation. *Information Sciences*, 191:192 – 213, 2012. ISSN 0020-0255. doi: http://dx.doi.org/10.1016/j.ins.2011.12.028. URL http://www.sciencedirect.com/science/article/pii/S0020025511006773. Data Mining for Software Trustworthiness.

[64] Gaoxia Jiang and Wenjian Wang. Markov cross-validation for time series model evaluations. *Inf. Sci.*, 375:219–233, 2017. doi: 10.1016/j.ins.2016.09.061. URL https://doi.org/10.1016/j.ins.2016.09.061.

[65] Manuel Jaen-Garcia. Empirical Analysis of Wagners Law for the Spains Regions. *International Journal of Academic Research in Accounting, Finance and Management Sciences*, 1(1):1–17, November 2011. URL https://ideas.repec.org/a/hur/ijaraf/v1y2011i1p1-17.html.

[66] Nathan R. Hoot, Larry J. LeBlanc, Ian Jones, Scott R. Levin, Chuan Zhou, Cynthia S. Gadd, and Dominik Aronsky. Forecasting emergency department crowding: A discrete event simulation. *Annals of Emergency Medicine*, 52(2):116 – 125, 2008. ISSN 0196-0644. doi: http://dx.doi.org/10.1016/j.annemergmed.2007.12.011. URL http://www.sciencedirect.com/science/article/pii/S0196064407018604.

[67] Feng Xie, Zhen Chen, Jiaxing Shang, and Geoffrey C. Fox. Grey forecast model for accurate recommendation in presence of data sparsity and correlation. *Knowledge-Based Systems*, 69:179 – 190, 2014. ISSN 0950-7051. doi: http://dx.doi.org/10.1016/j.knosys.2014.04.011. URL http://www.sciencedirect.com/science/article/pii/S0950705114001336.

[68] C.J. Willmott and K. Matsuura. Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. *Climate Research*, 30:79 – 82, 2005.

doi: http://dx.doi.org/10.3354/cr030079. URL http://www.int-res.com/articles/cr2005/30/c030p079.pdf.

[69] Vladik Kreinovich, Nguyen Hung T., and Wu Berlin. Towards a localized version of pearson's correlation coefficient. *International Journal of Intelligent Technologies & Applied Statistics*, 6(3):215 – 224, 2013. ISSN 19985010. URL https://login.ezproxy.leidenuniv.nl:2443/login?URL=http://search.ebscohost.com.ezproxy.leidenuniv.nl:2048/login.aspx?direct=true&db=aph&AN=90656030&site=ehost-live.

[70] Guocheng Zhu, Peng Zhang, Tiroyaone Tshukudu, Jun Yin, Gongduan Fan, and Xuxu Zheng. Forecasting traffic-related nitrogen oxides within a street canyon by combining a genetic algorithm-back propagation artificial neural network and parametric models. *Atmospheric Pollution Research*, 6(6):1087 – 1097, 2015. ISSN 1309-1042. doi: http://dx.doi.org/10.1016/j.apr.2015.06.006. URL http://www.sciencedirect.com/science/article/pii/S1309104215000227.

[71] C. J. Willmott and K. Matsuura. On the use of dimensioned measures of error to evaluate the performance of spatial interpolators. *International Journal of Geographical Information Science*, 20(1):89–102, 2006. doi: 10.1080/13658810500286976. URL http://dx.doi.org/10.1080/13658810500286976.

[72] Zheng Li, Xianfeng Ma, and Hongliang Xin. Feature engineering of machine-learning chemisorption models for catalyst design. *Catalysis Today*, 280:232 – 238, 2017. ISSN 0920-5861. doi: http://dx.doi.org/10.1016/j.cattod.2016.04.013. URL http://www.sciencedirect.com/science/article/pii/S0920586116302760. A Decade of Effort in Addressing the Grand Challenges in Catalysis.

[73] Lucas Silva. A feature engineering approach to wind power forecasting. *International Journal of Forecasting*, 30(2):395 – 401, 2014. ISSN 0169-2070. doi: http://dx.doi.org/10.1016/j.ijforecast.2013.07.007. URL http://www.sciencedirect.com/science/article/pii/S0169207013000836.

[74] Dirk Jacob, Sebastian Dietz, Susanne Komhard, Christian Neumann, and Sebastian Herkel. Black-box models for fault detection and performance monitoring of buildings. *Journal of Building Performance Simulation*, 3(1):53–62, 2010. doi: 10.1080/19401490903414454. URL http://dx.doi.org/10.1080/19401490903414454.

[75] Jos A. Orosa and Armando C. Oliveira. A field study on building inertia and its effects on indoor thermal environment. *Renewable Energy*, 37(1):89 – 96, 2012. ISSN 0960-1481. doi: http://dx.doi.org/10.1016/j.renene.2011.06.009. URL http://www.sciencedirect.com/science/article/pii/S0960148111003016.

[76] Jie Zhao, Xueya Wang, and Peiquan Jin. Feature selection for event discovery in social media: A comparative study. *Computers in Human Behavior*, 51:903 – 909, 2015. ISSN 0747-5632. doi: http://dx.doi.org/10.1016/j.chb.2014.11.007. URL http://www.sciencedirect.com/science/article/pii/S0747563214005834. Computing for Human Learning, Behaviour and Collaboration in the Social and Mobile Networks Era.

[77] H.E.T. Holgersson, L. Nordström, and Ö. Öner. Dummy variables vs. category-wise models. *Journal of Applied Statistics*, 41(2):233–241, 2014. doi: 10.1080/02664763.2013.838665. URL http://dx.doi.org/10.1080/02664763.2013.838665.

[78] Daniel B. Suits. Use of dummy variables in regression equations. *Journal of the American Statistical Association*, 52(280):548–551, 1957. ISSN 01621459. URL http://www.jstor.org/stable/2281705.

[79] B. Misra S. Nayak and H. Behera. Impact of data normalization on stock index forecasting. *International Journal of Computer Information Systems and Industrial Management Applications.*, 6: 257–269, 2014. ISSN 2150-7988. URL https://www.researchgate.net/publication/291962265_Impact_of_Data_Normalization_on_Stock_Index_Forecasting.

[80] Guan-De Wu and Shang-Lien Lo. Effects of data normalization and inherent-factor on decision of optimal coagulant dosage in water treatment by artificial neural network. *Expert Systems with Applications*, 37(7):4974 – 4983, 2010. ISSN 0957-4174. doi: http://dx.doi.org/10.1016/j.eswa.2009.12.016. URL http://www.sciencedirect.com/science/article/pii/S0957417409010628.

[81] Ismail Bin Mohamad and Dauda Usman. Standardization and its effects on k-means clustering algorithm. *Research Journal of Applied Sciences, Engineering and Technology*, 6(17):3299–3303, 9 2013. ISSN 2040-7459.

[82] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016. URL https://www.R-project.org/.

[83] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[84] Enzo Busseti, Ian Osb, and Scott Wong. Deep learning for time series modeling, 2012.

[85] Jan N. van Rijn, Salisu Mamman Abdulrahman, Pavel Brazdil, and Joaquin Vanschoren. *Fast Algorithm Selection Using Learning Curves*, pages 298–309. Springer International Publishing, Cham, 2015. ISBN 978-3-319-24465-5. doi: 10.1007/978-3-319-24465-5_26. URL https://doi.org/10.1007/978-3-319-24465-5_26.

[86] Marvin Meeng and Arno Knobbe. Flexible enrichment with cortana software demo.

[87] Alessandro Cardinali and Guy Nason. Costationarity of locally stationary time series using costat. *Journal of Statistical Software, Articles*, 55(1):1–22, 2013. ISSN 1548-7660. doi: 10.18637/jss.v055.i01. URL https://www.jstatsoft.org/v055/i01.

[88] Bar Koer and Ahmet Arslan. Genetic transfer learning. *Expert Systems with Applications*, 37 (10):6997 – 7002, 2010. ISSN 0957-4174. doi: http://dx.doi.org/10.1016/j.eswa.2010.03.019. URL http://www.sciencedirect.com/science/article/pii/S0957417410002009.

[89] Francois Rivest, John F. Kalaska, and Yoshua Bengio. Conditioning and time representation in long short-term memory networks. *Biological Cybernetics*, 108(1):23–48, Feb 2014. ISSN 1432-0770. doi: 10.1007/s00422-013-0575-1. URL https://doi.org/10.1007/s00422-013-0575-1.