BAŞKENT UNIVERSITY
SOCIAL SCIENCES INSTITUTE
DEPARTMENT OF TECHNOLOGY AND KNOWLEDGE MANAGEMENT
TECHNOLOGY AND KNOWLEDGE MANAGEMENT MASTER'S PROGRAM WITH THESIS

ESTIMATION OF ELECTRIC CONSUMPTION OF THE FIRMS OPERATING IN DIFFERENT SECTORS WITH MACHINE LEARNING ALGORITHMS

PREPARERD BY
MERT ENSARİ

MASTER'S THESIS

THESIS ADVISOR
DOÇ. DR. MEHMET GÜRAY ÜNSAL

ANKARA – 2023

BAŞKENT ÜNİVERSİTESİ

SOSYAL BİLİMLER ENSTİTÜSÜ

YÜKSEK LİSANS ÇALIŞMASI ORİJİNALLİK RAPORU

Tarih: 26/12/2022

Öğrencinin Adı, Soyadı: Mert ENSARİ

Öğrencinin Numarası: 22110461

Anabilim Dalı: Teknoloji ve Bilgi Yönetimi

Programı: Teknoloji ve Bilgi Yönetimi Tezli Yüksek Lisans

Danışmanın Unvanı/Adı, Soyadı: Doç. Dr. Mehmet Güray ÜNSAL

Tez Başlığı: Estimation of electric consumption of the firms operating in different sectors with machine learning algorithms

Yukarıda başlığı belirtilen Yüksek Lisans/Doktora tez çalışmamın; Giriş, Ana Bölümler ve Sonuç Bölümünden oluşan, toplam 62 sayfalık kısmına ilişkin, 26/12/2022 tarihinde şahsım/ tez danışmanım tarafından Turnitin adlı intihal tespit programından aşağıda belirtilen filtrelemeler uygulanarak alınmış olan orijinallik raporuna göre, tezimin benzerlik oranı % 11'dir. Uygulanan filtrelemeler:

1. Kaynakça hariç

2. Alıntılar hariç

3. Beş (5) kelimeden daha az örtüşme içeren metin kısımları hariç

"Başkent Üniversitesi Enstitüleri Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Usul ve Esaslarını" inceledim ve bu uygulama esaslarında belirtilen azami benzerlik oranlarına tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Öğrenci İmzası:…….

ONAY

Tarih: 26/12/2022

Öğrenci Danışmanı Unvan, Ad, Soyad, İmza:

Doç. Dr Mehmet Güray Ünsal

ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisor Assoc. Prof. Mehmet Güray ÜNSAL even though it is not enough to show my gratitude to him. I am very thankful to his patience, innovativeness, open-mindnesses and mostly to him great knowledge. I am very grateful that he gave me the chance to work with him.

I would like to thank also to Prof. Dr. Hakkı Okan Yeloğlu for his contribution to me and my master's degree education. I would like to thank to Prof. Dr. Hulusi ÖĞÜT for his valuable contribution, especially in Python coding process. I would also like to thank jury members, Asst. Prof. Nurcan ALKIŞ BAYHAN and Assoc. Prof. Deniz ÖZONUR for their valuable comments and evaluations in the jury.

I would like to thank my family, especially to my brother Tolga ENSARİ because of his great contributions and support to me and this work.

Finally, I would like to thank my friends; Tenin ÇOLAK and A. Oğulcan DALKILIÇ for their kind friendship and support.

ÖZET

MERT ENSARİ, FARKLI SEKTÖRLERDE FAALİYET GÖSTEREN FİRMALARIN MAKİNE ÖĞRENME ALGORİTMALARI İLE ELEKTRİK TÜKETİMLERİNİN TAHMİNİ, BAŞKENT ÜNİVERSİTESİ, SOSYAL BİLİMLER ENSTİTÜSÜ, TEKNOLOJİ VE BİLGİ YÖNETİMİ BÖLÜMÜ, 2023

Firmaların etkin bir yol haritası çizebilmesi için elektrik tüketim miktarlarını en iyi seviyede tahmin edebilmesi gerekmektedir. Elektrik tüketim tahmini ne kadar başarılı olursa firmaların maliyeti de o kadar düşük olmaktadır. Bu yüzden elektrik tüketim tahminini en başarılı şekilde tahmin edebilmek günümüz rekabetçi piyasalarda bütük önem arz etmektedir. Gerçek zamanlı dengesizlikleri tahmin eden doğru tüketim tahmini, diğer firmalara karşı rekabet avantajı sağlayacaktır. Bu çalışmada, kısa vadeli tüketim tahmini yapmak ve yöntemlerin doğruluğunu test etmek için Makine Öğrenmesi (ML) yöntemleri önerilmekte, Türkiye'de çeşitli sektörlerden firmaların (şirketlerin) saatlik tüketim tahminleri yapılmakta ve sonuçlar değerlendirilmektedir. Farklı yöntemler farklı zamanlarda iyi performans gösterseler bile, tutarlı bir şekilde aynı performansı gösteremezler. Bu tez çalışmada, Aşırı Gradyant Destekleme Regresyonu ve Gradyant Destekeleme Regresyonu yöntemlerini kullanarak ortalama mutlak hata (performans ölçüm göstergesi olarak) sonuçları değerlendirilmekte ve sonuçları yorumlanmaktadır. Programlama kodları Python programlama dilinde yazılmıştır. Gelecek dönemler için elektrik tüketimi tahmini ve benzeri çalışmalarda makine öğrenmesi yöntemlerinin kullanılmasının uygun olduğu gözlemlenmektedir.

Anahtar Kelimeler: Tahmin, Elektrik Tüketimi, Makine Öğrenmesi, Destekleme

ABSTRACT

MERT ENSARİ, ESTIMATION OF ELECTRIC CONSUMPTION OF THE FIRMS OPERATING IN DIFFERENT SECTORS WITH MACHINE LEARNING ALGORITHMS, BAŞKENT UNIVERSITY, INSTITUTE OF SOCIAL SCIENCES DEPARTMENT OF TECHNOLOGY AND KNOWLEDGE MANAGEMENT, 2023

Firms or companies need to be able to draw a sustainable and effective roadmap and to estimate the electricity consumption amounts at the best level. The more successful the electricity consumption estimation is, the lower the cost of the companies. That's why it's so important in today's world. Accurate consumption estimating real-time imbalances will provide a competitive advantage with other retail sales companies. In this study, Machine Learning (ML) methods are proposed to make short-term consumption estimation and to test the accuracy of the methods, hourly consumption estimations of firms (companies) from various sectors in Turkey are made and the results are evaluated. Even if different methods perform well at different times, they cannot consistently perform the same. In this thesis study, we evaluate the results of mean absolute error (as performance measurement indicator) using Extreme Gradient Boosting Regression (XGBR) and Gradient Boosting Regression (GBR) methods and interpreted the results. The programming codes are written in Python programming language. It is observed that it is a suitable to use ML methods for the estimation of electricity consumption for future periods and similar studies.

Keywords: Estimation, Electric Consumption, Machine Learning, Boosting

TABLE OF CONTENTS

TABLES

Page

FIGURES

Page

1. INTRODUCTION

“Can a machine think and how can it think?”, this question is the title of the first official Turkish resource explaining the basis of Artificial Intelligence and Machine Learning concepts by our world-renowned Ordinaryus Professor Cahit Arf. For this reason, it is quite appropriate to start this thesis with this sentence and reference, and it is important to take the subject from the ground up. After the Second World War, three main issues that emerged with the effect of science during the war process were emphasized. These are; obtaining atomic energy, rockets launched into the air, and thinking machines, in other words electronic brains, that make decisions according to the situation, do things in accordance with these decisions (Arf, 1959). We can make the general definition that we describe as a machine here, in terms of computers. It was emphasized in this study that computers do not have the ability to think aesthetically as much as the human brain, but it is hoped that the ability to make decisions can be transferred to machines in case of uncertainty in the decision-making process of humans. In the circumstances, we encounter a process in which machines can be a figure in human history.

Humanity has listed the historical chronology as the First Age, the Middle Ages, the New Age and the Modern Age, depending on the occurrence of certain events or the emergence of some inventions. Undoubtedly, the period we are in shows that we are in a different era in terms of consumption habits and production models. We can call this age the Technology and Information Age or the "Digital Age" in general (Gürsakal, 2018). What this means is that we are living in a period in which there are developments so important that we can talk about a new era in historical chronology. In this age we live in, the technological revolution that has emerged due to new consumption habits reveals the need to develop different production models. The technological revolution here should be explained in more detail. In this day and age, machine-human interactive smart factories are the result of this process. In this process, which is called Industry 4.0, machines have started to gain the ability to perform some actions with human intelligence. These; It means transferring the skills of thinking like a human, establishing relationships between subjects, solving problems and making decisions to machines. In such a production process, people are still inside or outside the factory in some way, the next stage is a process dominated by machines, ignoring people. Stores and factories managed only by machines present Industry 5.0, which is essentially the 5th industrial revolution, as a reality waiting for us in the future.

The main debate about this reality is based on two main issues. First, "intelligent" machines can only be used in production, health, service, etc. It is the possibility that it can be used in bad situations in human history, such as war, in addition to its well-intentioned contributions to humanity in sectors. Secondly, machines can completely replace human beings in this process, making some areas of business history. For instance, if we consider that self-driving cars will be a completely autonomous process independent of people, professions such as truck and taxi driver will disappear, there will be no need for driver's licenses, perhaps there will be no business line related to this because the occurrence of traffic accidents will disappear (Gürsakal, 2018). However, it can be said that the zero accident situation is not guaranteed for now. Although some lines of business will disappear completely with machines with human-specific intelligence, it is a fact that new business areas will emerge within the framework of many new terminologies such as the intersection and subset of Artificial Intelligence. Although the idea of transferring human intelligence-based behaviors to machines was first brought to the literature by John McCarthy with the use of the term "Artificial Intelligence" in 1956, the descriptions of artificial intelligence extend to the depiction of mechanical servants serving the gods in the remains of Homer (Gürsakal, 2018). Considering that the dividing line between human and machine may become blurred in this process, it is not difficult to think that studies on this subject are also being carried out. For example, Captcha, which represents a system of questions asked to us when registering on some websites or while recording information, is a security measure for human and computer separation. With the awareness of small-capitalization and sign separation, these are used in some applications on computers for the purpose of separating humans from machines (Gürsakal, 2018). In addition, the Turing test, developed by Alan Turing in 1950, draws attention as another application that distinguishes humans from machines (Gürsakal, 2018; Balaban & Kartal, 2015):

- Must be able to communicate successfully in the language used,
- Should be able to answer some questions and make inferences using given information
- Must be able to adapt to new conditions.

The transition to machines that behave in a human-like manner is of course the product of a certain process and did not appear out of nowhere. Until the Artificial Intelligence process,

the process of incorporating machines into production has passed from infancy to running by passing different steps. From the use of steam machines (Industry 1.0), to the use of electrically powered machines and these machines, which have become widespread with wireless electricity transmission (Industry 2.0), from computer and automated production (Industry 3.0) to smart factory systems (Industry 4.0), and finally to produce only within the machine-machine interaction. This development process continues at an accelerated pace until Industry 5.0, which aims at.

In other words, the process of revealing machines that can replace the human brain with Artificial Intelligence science to be used in production processes continues. In this case, artificial intelligence can be briefly called 'the work of making machines intelligent'. In summary, the subject of Artificial Intelligence is to develop computers and machines that think and act rationally like humans (Gürsakal, 2018; Balaban & Kartal, 2015). Artificial Intelligence is mentioned together with two methods called Machine Learning and Deep Learning in the literature. Alternative definitions of these three terms, which will be used frequently in this study, are followed by Ghatak (2017) as follows:

- "Artificial Intelligence (AI) is the broadest term applied to any technique that enables computers to mimic human intelligence using logic, if-then rules, decision trees, and ML (including deep learning)."

- "ML is the subset of AI that includes abstruse statistical techniques that enable machines to improve at tasks with the experience gained while executing the tasks."

- "Deep Learning is a scalable version of ML. It tries to expand the possible range of estimated functions. If ML can learn, say 1000 models, deep learning allows us to learn, say 10000 models. Although both have infinite spaces, deep learning has a larger viable space due to the math, by exposing multilayered neural networks to vast amounts of data."

The concepts of Artificial Intelligence (AI), Machine Learning (ML) and Deep Learning (Ramasubramanian & Singh, 2019) described above are concepts that can be graphically represented as subsets of each other. This depiction is given in Figure 1 below.

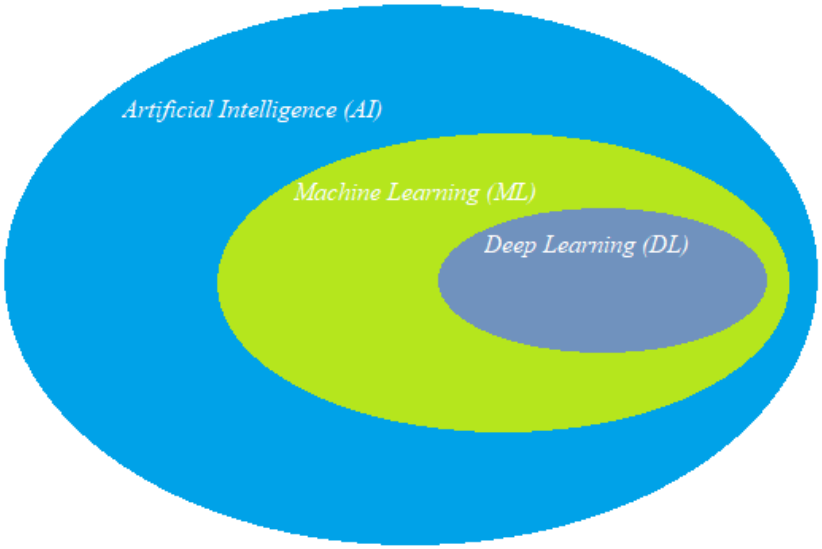


Figure 1. Artificial intelligence, machine learning and deep learning

The concept of ML, which we can examine as a subset of AI methods, is another terminology that has been brought to the literature with AI. Mitchell's general definition of ML: 'If a computer program increases the performance measured by P in G tasks, by experience D, it can be said that this computer program has learned (Kubat, 2017; Balaban & Kartal, 2015).

If we focus more on the term ML; as mentioned, in this day and time, the age of management of machines has come. ML used to be dull under different disciplines in the past, but this situation has now changed. A group of researchers conducted research on where data comes from in data-based frameworks in the 1970s. By courtesy of studies developed, these issues were also clarified (Kubat, 2017). It is very difficult to describe machines how to approach any problem, but it can be taught indirectly how to approach a problem. At this point, ML algortihms and model comes into play.

Nowadays, human thinking and solution ability are being tried to be integrated into machines. Forwhy, in the growing and developing World in Digital Era, activities such as decision making, estimation and optimization are getting harder and harder due to the proliferation of data. Therefore, ML gets busy in all fields all over the world. All industrial companies in the world are experiencing some problems in energy consumption. Due to the increasing world population, the products and services produced for people are also increasing. This causes more energy consumption.

One of the services most needed by the sectors is to determine the best electricity consumption forecasts during the day or the day before. Because the industries should make the

best estimates of electricity consumption so that there is no energy deficit in the energy market. If the electricity consumption forecasts are not estimated in the best way, price differences occur and it is necessary to buy electricity at a high price. In this case, an imbalance occurs and all companies's cost expense increase. Therefore, estimating the electricity consumption data by the company in advance and creating the demand has an important place in minimizing the cost of the production process. Therefore, the electricity consumption forecast is very valuable. For this reason, this subject is studied within the scope of this thesis. The scope of the study is to determine how the boosting-based algorithms perform from this estimation process. For this reason, it is desired to make these estimations over two algorithms based on boosting in ML.

Our aim in this study is to estimate the electricity consumption of industrial companies in the best way with ML algorithms. Algorithms and models described in the study are only a part of ML. The remaining parts of this thesis are structered as follows: Section 2 describes the general framework of ML by discussing the strategies based on ML, performance measurement & evaluation, parameter tuning concepts of ML and a brief literature survey. In Section 3, the methodologies used in this study are investigated. All the results and findings are given in sub-sections of Section 4. The study is finalized in Section 5.

2. MACHINE LEARNING

Machine Learning (ML) refers to coding programs whose activities are formed according to their exposure to the information in the data (Igual & Segui, 2017). ML can be expressed as the stages of discovering the trends of complex and big data beyond simple analysis methods (Ghatak, 2017). In ML most significant digit, taking the right steps in identifying the problem.

ML is a subfield of artificial intelligence (AI). ML is in the subject of artificial intelligence. In ML, the goal is to learn from experiences and optimize their performance. The purpose of artificial intelligence is to integrate human intelligence into machines (Gürsakal, 2018).

We can summarize ML as follows (Gürsakal, 2018):

$$Y = f(x) + e$$

The machine learns the f function from the labeled training data. Machine predicts Y from unlabeled test data. In any classification based problem, Y represents the labels, in other words, the class or group of the units. This is also called target attribute. The target attribute has to be categorical data (nominal and ordinal variable in Statistics) such as defected or non-defected units or customer segments as 1st class, 2nd class and 3rd class. In these kind of cases, it can be defined a binary variable for identification of defected and non-defected ones as 0 and 1, respectively. For custumer segments, it is suitable to use 1, 2 and 3 for representing of 1st, 2nd and 3rd class customers. Here, x represents the attributes which are helpful in prediction process of target attribute for any unit. For any classification problem, these attributes or variables can be, in the form of nominal, ordinal, interval or ratio scale of measurement. The nominal scale of measurement defines the identity property of data and there is no numerical or hierarchical dominance within the levels of attribute or variable. The colors of eyes (Brown, hazel, blue, green, others), gender (male, female) or profession (doctor, engineer, teacher, others) can be given as examples for attributes or variables scaled in nominal scale of measurement. The ordinal scale defines data that is placed in a ranking or a specific descending or ascending order. There has to be a numerical or hierarchical dominance within the levels of attribute or variable. Group of ages (18-30, 20-50, 50+) or category of employees in a factory (clerical, custodial, manager) can be given as examples for attributes or variables scaled in ordinal scale of measurement. Interval scale of measurement can take a value within a specific

interval and it is characterised by the fact that the number zero is an existing variable. In other words, an absolute starting point must not be defined. For example, if you measure degrees, zero has a temperature, it is not a starting point fort his measure. Ratio scale of measurement differ from interval scales. Difference between these two scales is based on a defined stariting or true zero. For example, someone can not be zero centimetres tall. His or her weight can not be measured as zero kilos.

f(x) represents a function that models the effect of other attributes on target attribute. In this definition, the model indicates a methodology or algorithm of ML. These models, algorithms or methodologies are based on Multivariate Statistical Methods. Here, e is an error term including the noise, in other words, unexplained variability in target variable. ML use Multivariate Statistical Methods by considering its own logic.

ML brings a new approach to software production. It learns from examples instead of programming for a determined purpose and it is the most important factor (Mcafee & Brynjolfsson, 2017). As mentioned above, Statistics and ML concepts are intertwined. Algorithms within ML are linked to statistical tools. These statistical tools and models often derive meaningful presumptions from the data, and based on these presumptions, they make robust statements about the outcomes (Ghatak, 2017) On the other hand, if the presumptions in the learning model are inaccurate, the effectiveness of the model becomes arguable (Ghatak, 2017). The data inserted into the algorithms transforms the input data into a large amount of output information. Therefore, there is a direct proportionality between the input data and the output data (Ghatak, 2017). If the missing values in the data we process are more than we expect, the algorithm will assign these values itself. For this reason, the possibility of inaccuracies in the output results will increase. The model created will unfortunately give wrong results (Ghatak, 2017).

ML is used in many fields today. Especially in the field of industry, also in academic processes, estimating and optimizing are among the important operations today. For instance, automotive industy, marketing, medicine, entertainment industry, science, web search, spam filters, credit scoring etc.

Four important factors affecting ML are listed as follows (Balaban & Kartal, 2015; Gürsakal, 2018):

- Dataset presented to the machine as an experience

- Determination of the variables thought to have an effect on the result

- Selected learning strategy

- Algorithm used and algorithm parameters, if any.

Dataset presented to the machine as experience: They will learn better with more experience with machines or computers, just like humans. In other words, the more and different possible situations they encounter as experience, the learning will be better. In addition, the quality of the data collected about the problem in ML is as important as the quantity (Balaban & Kartal, 2015; Gürsakal, 2018). For example, suppose we have a classification problem. Let's wait for the machine to distinguish whether the photo on any paper to be displayed after the apple and pear pictures shown to the machine is apple or pear. When there is a problem with classification into two groups like here, if the data we show and introduce the machine in the learning process includes only examples of apples, an accurate and reliable classification will not be made, since there is no information about pear. It is clear that we cannot expect high performance in the discrimination power of the machine after such a learning process. Performance measurement will be discussed in more detail in the following sections.

Determination of the variables that are thought to have an effect on the result: One of the factors that significantly affect ML process is the selected features in the established model. The features to be used in the ML process are called attributes. These are called variables in statistics. Although different terminologies seem to be used, they basically mean the same thing. Being able to distinguish between apples and pears in the previous example also depends on the qualities we will use in the process of distinguishing photographs. For instance, it is clear that the diameter, shape and color of the object in the photograph should be used as attributes. Thanks to these qualities, the machine will try to distinguish between apples and pears. As another example, in the classification of a bank's customers, features such as the amount of income, whether it has existing credit, whether it makes regular credit card payments can be determined as problem-specific features. In some problems, the features that are found to be highly related to each other or have little contribution to the solution are eliminated (Balaban & Kartal, 2015; Gürsakal, 2018).

Selected learning strategy: This part will be explored in more detail in the next section. What we mean by the term learning strategy is closely related to both the task for which machine learning is desired and the dataset required for it. Because while ML needs an output value (target attribute) in the data set to solve some problems, it does not need these values in some cases. In the process of grouping the object in the photographs as apple (E) or pear (A) in

relation to the photographic experiment we gave earlier, being a target attribute means labeling the objects. Secondly, the risk groups of the customers (Low/Medium/High) can be given as an example of target qualification while the bank classifies its customers according to risk groups in terms of credit. When the examples are examined, it can be seen that the problems that need the target attribute are mostly on estimation and classification problems. Therefore, it is important to determine the learning strategy according to the problem to be addressed in ML and to provide the appropriate data set for this strategy (Gürsakal, 2018).

Algorithm used and algorithm parameters, if any: What we mean by algorithm in this definition is the method or methodology that will be used in the ML process. For this reason, we see that the terms method, methodology or algorithm are used in the same sense in different sources. The algorithm can be expressed as the step-by-step path to be followed to solve a given problem. In general, these algorithms are based on Multivariate Statistical Analysis methods or Optimization methods. Also some algorithms can even use probability theory. In these algorithms, there are a number of steps required for the machine to learn by considering the dataset it has. In some algorithms, no parameter is needed for analysis, while in some algorithms, some parameter values must be determined and the method must be applied (Balaban & Kartal, 2015; Gürsakal, 2018). Determining the most appropriate parameter values for the data structure we have and the selected algorithm is an important issue that will affect the learning performance. This subject will be discussed in more detail in the following sections.

Since algorithms developed for ML can be transferred to computers with various programming languages, machine expression can simply be thought of as computers we use in daily life (Gürsakal, 2018).

2.1. Machine Learning Strategies

An approach frequently used in the literature for grouping machine learning techniques is to classify ML systems according to the learning strategy used in the learning process. According to this approach, we can divide the subject of machine learning into these 4 classes. These are:

- Supervised Learning

- Unsupervised Learning

- Ensemble Learning

- Reinforcement Learning

2.1.1. Supervised Learning

Supervised learning can be defined as using labeled data (in other words, target attribute) in the training phase of algorithms. As the input data (other attributes which try to estimate the target attribute) is adapted to the model, the model continues to adjust its weights until the machine deems it appropriate (Igual & Segui, 2017) The first step in supervised learning problems starts with the training set containing the training examples associated with the correct labels. The steps of supervised learning are summarized in Figure 2.

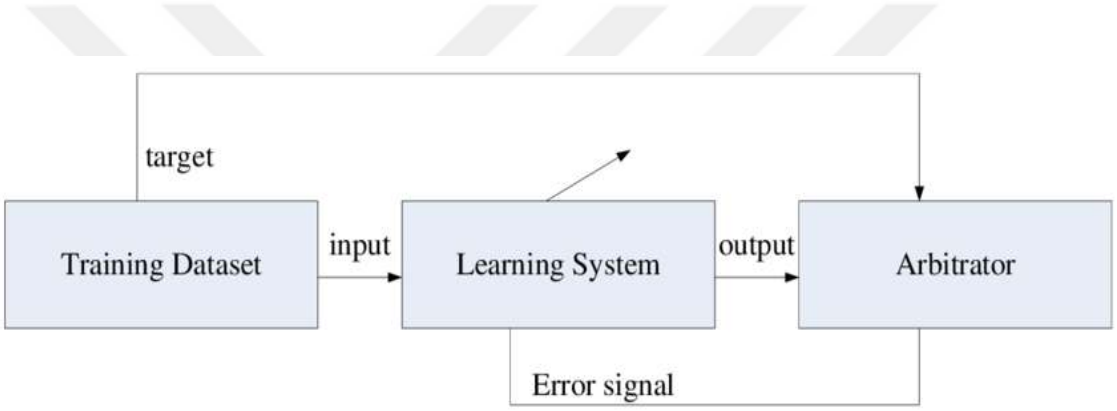


Figure 2. Block Diagram of Supervised Learning (Sapon, Ismail, Zainudin, & Ping, 2011)

Supervised learning is carried out by the following two methods: A continuous numerical value can be estimated by regression or a label can be assigned to an object by classification (Gürsakal, 2018). An appropriate algorithm has to be choosen according to the structure of target attribute. This algorithm is a model in which the difference between the estimated output values and the known output values is minimized. For example, a bank wants to examine the creditworthiness of customers. It decides whether to give a loan according to the risk situation. Given that, a bank has so many customers, it is impossible to manually infer from lists. At this point, a decision can be made using the k-NN (k nearest neighbourhood) method or algorithm, which is a supervised learning technique (Balaban & Kartal, 2015).

In supervised learning, what the model will predict, in other words, its target quality is clear. There is labeled data. As stated above, the purpose of supervised learning algorithms is classification and regression. Depending on whether the target attribute is discrete or

continuous, the algorithm that can be used may change. Support vector machines (SVM), decision trees, logistic regression, naive bayes classifier, k-NN (k nearest neighborhood) algorithms can be given as an example of the most common supervised learning algorithms or methods in the literature in classification problems where the target attribute is a discrete variable. Multiple linear regression, Ridge regression, LASSO, support vector regression can be given as examples of the most common supervised learning algorithms or methods in the literature in classification problems where the target attribute is continuously variable. Depending on whether the target quality is discrete or continuous, the performance evaluation criteria that will be mentioned in the future vary.

2.1.2. Unsupervised Learning

It would not be right to think that ML can only learn by examples. Convenient information can also be gathered from examples without classes (Kubat, 2017). Unsupervised learning is gaining ground day by day. Some examples are; a shopping site can see the most searched products and customers have purchased, so shopping site can highlight products that customers can buy. These applications are realized with statistical learning methods and unsupervised learning systems (James, Witten, Hastie, & Tibshrani, 2017).

In unsupervised learning, the system does not know about the correct output. Networks that can be trained unsupervised adjust their weight values according to the characteristics of the input information without the target output. Networks that can be trained unsupervised adjust their weight values according to the characteristics of the input information without the target output (Elmas, 2018).

In unsupervised learning, the network works with the entered information, not the requested external data. The network must find a way to organize themselves without help. Since there are no output examples, the network learns by improving itself (Elmas, 2018).

Unsupervised learning includes methods that attempt to explain the properties or structures of data. To illustrate the distinction between supervised and unsupervised learning, we need to show a simple example in Figure 3:

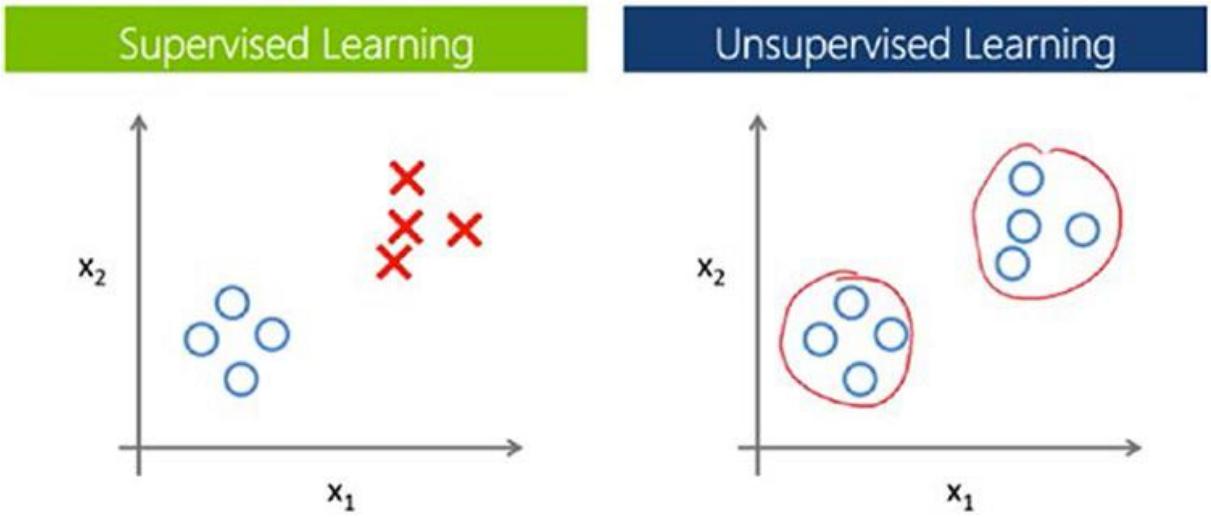


Figure 3. Supervised versus unsupervised learning (Ramasubramanian & Singh, 2019)

The graph on the left, the data are labeled as two different types, therefore the algorithm realizes that objects are dissimilar. On the right side has the same objects, but the algorithm doesn't tell us which one is (Ramasubramanian & Singh, 2019).

2.1.3. Ensemble Learning

The purpose of ensemble learning is to make a better resolution with a holistic intelligence. This concept makes use of multiple algorithms to improve the performance of the system in ML (Ramasubramanian & Singh, 2019). As with all research problems, we may not find the desired result in some constrained iterations. In situations as this, ensembles can reunite multiple theroies to create a better theory (Ramasubramanian & Singh, 2019).

Ensemble learning has multiple models. Therefore, the calculations take up time longer. Mostly, it is recommended to use faster algorithms. On the other hand, solution time may take time longer than you expect. In other words we can say “learning together in harmony”. (Gürsakal, 2018) .Because multiple algorithms are used in the model. For example, When buying a car, parameters such as whether the car’s wide, mile per galon, and brand value should be considered. Then you decide whether to buy the car or not. This is how ensemble learning works, too. There are many variables and parameters.

In the process of ensemle learning, three main technique are used to perform a better estimation process for a target attribute. These are bagging, boosting and stacking.

2.1.3.1. Bagging

Bagging is an ensemble learning algorithm. For simplicity, we will assume that the samples are either positive or negative. As usual, the classifier is based on a pre-trained training sample. The strategy, called bagging, induces a certain group of classifiers. Classifiers are used concurrently and offer insight into which class to label. The base classifier then collects and analyzes all the information and selects the most selected label.

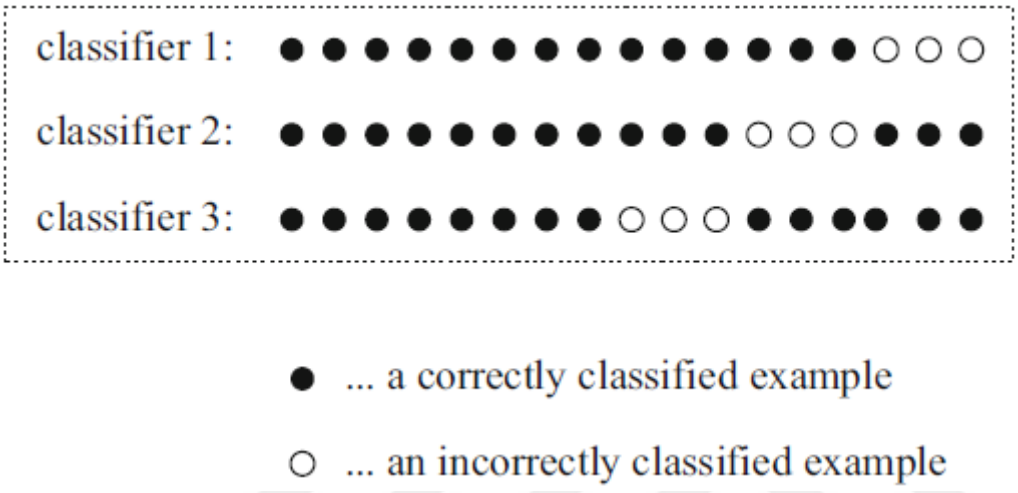


Figure 4. Bagging (Kubat, 2017)

In the figüre 4 above shows how the voting process effectively reduces the error rate. In this example, three classifiers were expected to classify 17 samples. Here each made an error in the examples for each classifier. These errors are corrected by the voting process. Each sample dataset is misclassified at most once. Other tags are correct. Thus, the mistake made will be corrected by others. (Kubat, 2017)

This method can be used in many applications. These cases take the weighted average of the outputs of the generated models in the case of continuous functions. (Ramasubramanian & Singh, 2019)

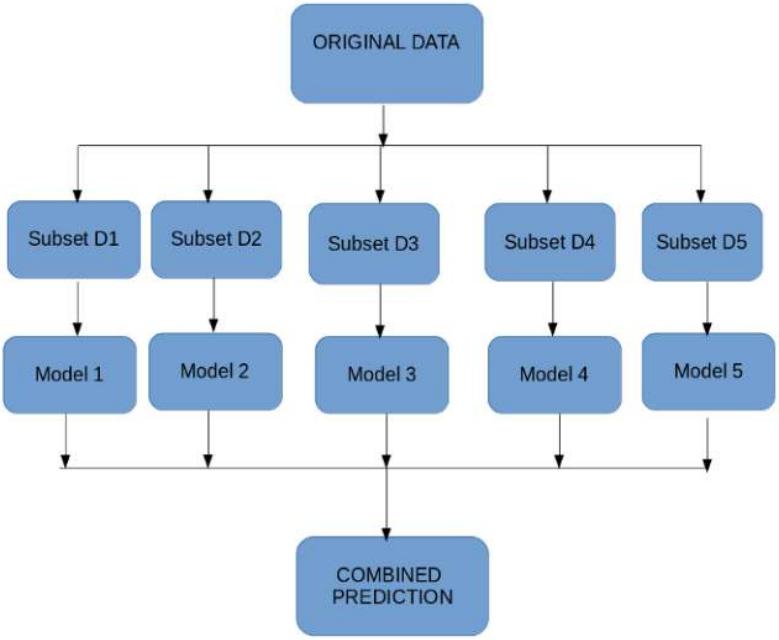


Figure 5. Bagging ensemble flow

There are 4 steps as you can see in Figure 5. These are:

1. Generating samples from training datasets
2. The model is trained on examples.
3. Classifiers are created from each model
4. Results are combined using weighted averages and voting.

2.1.3.2. Boosting

Boosting method, it is a statistical learning algorithm like the bagging method and also it is an ensemble learning's model. Boosting models is a sequential build process based on minimizing errors from previous models while enhancing the impact of high-performance models. These models are techniques by which the bugs detected in previous models are resolved.

The boosting algorithm seeks to answer the following question, suppose we have a model that generates less than 50 classifications. Can we build a regular model to get the minimum error? Typically, successive models constitute a very mixed relationships in the final models that are formed last. Therefore, models can become inexplicable. The boosting method can occasionally generate a black box. (Ramasubramanian & Singh, 2019)

The basic idea in the Boosting method is to make inferences from the collection of trees obtained as a result of giving different weights to the dataset. Initially, all observations have equal weight. As the tree community begins to grow, weightings are adjusted based on problem information. The weight of misclassified observations is increased, while the weight of rarely misclassified observations is decreased. In this way, trees gain the ability to regulate themselves in the face of difficult situations.

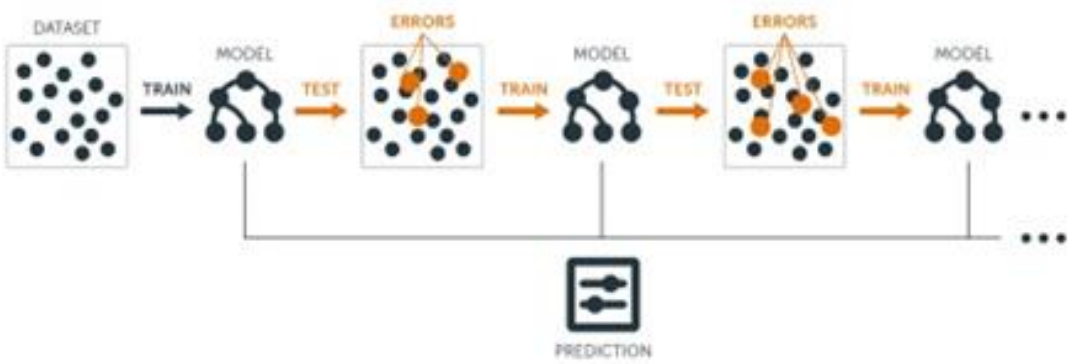


Figure 6. Boosting (Boehmke & Greenwell, 2020)

If we explain how the algorithm works based on the above model as given Figure 6, firstly a subset is constitute from the original dataset. All data points are given equal weight. A base model is created in this subset. This model is used to make predictions for the entire dataset. Errors are calculated using actual values and predicted values. More weight is given to incorrectly predicted observations(More weight will be given here to the three unclassified blue plus points). Another model is created and predictions are made on the dataset (This model attempts to fix the bugs of the previous model). Similarly, multiple models are created, each correcting the errors of the previous model. The final model (strong learner) is the weighted average of all models.

2.1.3.3. Stacking

Stacking is a method to ensemble multiple classification problems or regression models in ML. Stacking (or Stacked Generalization) has a different idea from boosting and bagging. In stacking, ML models are stacked on top of each other in layers. In this respect, it is another argument used by ensemble learning method. Each model transfers its own estimations or forecasts to the next model, and the final result is obtained with the last model (Gürsakal, 2018).

In other words, stacking is a way of ensemble ML used to evaluate multiple models to build a new model which improves model performance and gives better estimates or forecasts.

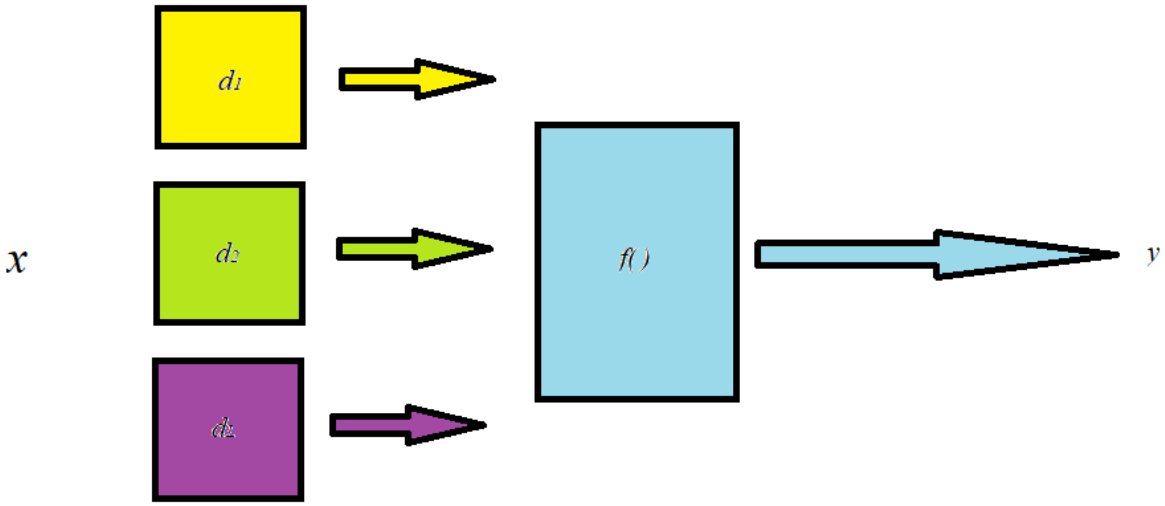


Figure 7. Stacking with two layers in ensemble learning

As can be seen in the Figure 7, which shows a two-layer stacking technique, models such as d_1, d_2 and d_L take input x as data and give the estimations which they produce to the f( ) function, and the f( ) function uses these inputs to produce the final outputs (forecasts or estimations). It converts these inputs to y value. The number of layers can be more than two (Gürsakal, 2018). Increasing the number of layers gives more detailed solutions, but also increases the cpu time.

2.1.4. Reinforcement Learning

This type of learning, the model not only experiences the data, but also affects each other with the surrounding and receives feedbacks on its experimentation (Ghatak, 2017). Supervised and unsupervised learning models require tidy and correct data to achieve optimum outcomes. Besides, to work with non visible samples, the data requires extensive. Driverless cars, robotic coding are connected with these modeling classes. Reinforcement learning programmed to constantly learn from the surrounding. It gains experience by communicating with the environment. In this way, all situations are tried to be explored (Ramasubramanian & Singh, 2019).

Reinforcement learning is learning by interacting with trial and error in a dynamic environment. There is no need for true value label information and the learning process is started with temporary labels. The necessary is tutorial feedback indicating whether the

temporary labels are true or false. Feedback indicates something is wrong or right. But, the model does not specify the cause of the error. Instead of presenting true value, a signal, a score or a rating is reported by the consultant indicating how accurate the output value is (Balaban & Kartal, 2015).

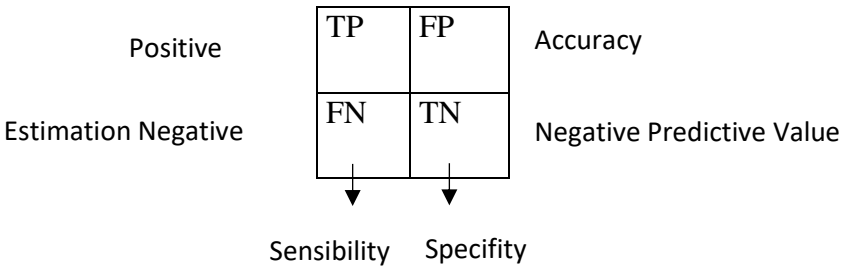
2.2. Performance Measurement

2.2.1. Confusion Matrix

In classification operations, confusion matrix is used to compare the actual data with the predicted value. The confusion matrix is tabulated and easily analyzed. In binary result examples, the matrix have four results (Ghatak, 2017):

1. True Positives (TP): are positive components directly classified as positive.
2. True Negatives (TN): are negative components directly defined as negative.
3. False Positives (FP): are negative components classified as positive.
4. False Negative (FN): are positives components classified as negative.

Specified topics can be demonstrated by matrix and formulas, as follows:



The conjoining of these items authorizes to describe many metrics: (Igual & Segui, 2017)

- Accuracy:

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

- Metrics are obtained by column-wise:

-Sensitivity:

$$\text{sensitivity} = \frac{TP}{\text{Real Positives}} = \frac{TP}{TP + FN}$$

-Specificity:

$$\text{specificity} = \frac{TN}{\text{Real Negatives}} = \frac{TN}{TN + FP}$$

- Metrics are obtained by row-wise:

-Accuracy or Positive Predictive Value:

$$\text{Accuracy} = \frac{TP}{\text{Predicted Positive}} = \frac{TP}{TP + FP}$$

-Negative Predictive Value:

$$\text{NPV} = \frac{TN}{\text{Predicted Negative}} = \frac{TN}{TN + FN}$$

These fractional performance measures respond questions about how often a classifier estimate a confidental class (Igual & Segui, 2017). If the TP and TN values are high, the trueness of the model is at the highest level. However, it is not always desirable to maximize trueness. Nevertheless, we may not always want to maximize trueness. In this case, FP and FN need to be minimized (Ghatak, 2017). Therefore, truness is not the most regular metrical for interpreting classifiers, Confusion matrix takes into account such events and provides new definitions of measurement and evaluation (Igual & Segui, 2017).

2.2.2. Mean Absolute Error (MAE):

In Applied Statistics, mean absolute error (MAE) is one of the popular measures to evaluate the model performance. It is a measure based on errors. These errors are calculated as the differences between the observed values and estimated values. The estimated values can be calculated by using any regression based model. For each observation in data, the error value (difference between observed and estimated value) is obtained and the absolute value of these error are considered. Then, the mean value of these absolute errors gives MAE value of the relavant model. MAE derives from the unaltered magnitude (absolute value) of each difference. In other words, MAE is obtained as the summation of absolute errors divided by the sample size as follows (Williams & Matsuura, 2005):

$$MAE = \frac{\sum_{i=1}^{n} |\hat{y}_i - y_i|}{n} = \frac{\sum_{i=1}^{n} |e_i|}{n} \qquad i = 1, 2, ..., n$$

As mentioned, MAE is an arithmetic average of the absolute errors and here, $|e_i|$ is equal to $|\hat{y}_i - y_i|$, where $\hat{y}_i$ is estimations of units in the sample in regression models and y_i is observed values. Here, n is sample size and $i = 1, 2, ..., n$ (Williams & Matsuura, 2005).

The mean absolute error (MAE) is one of the most popular ways of comparing estimations or forecasts with their eventual outcomes. Two of the other alternatives are the mean absolute scaled error (MASE) which summarize performance of models by disregarding the direction of over or under prediction. Another measure can be defined as the mean squared error (MSE) based on the least squares approach in regression. The root mean squared error (RMSE) is the square root of MSE. MAE is not identical to root-mean square error (RMSE). MAE is simpler than RMSE in terms of calculation process, and also easier to interpret than RMSE (Pontius, Thontteh, & Chen, 2008). Thus, MAE is choosed as a performance measure in the concept of this study.

2.3. Performance Evaluation Methods (K-fold Cross Validation)

Cross-validation is the most used verification process in these days. It is used more in the computer world than residual-based metrics. The problem with residue-based methods is that you also need to keep a test set. It does not say how to interfere with the confidential data

contained in the model. Therefore, while training, testing, and validation are successful, simulation and illustration give us more opportunities to test (Ramasubramanian & Singh, 2019).

K-Fold Cross Validation is one of the methods of splitting the dataset for evaluating classification models and training the model. Sometimes, some problems may occur in separating the train and test sets. Some problems may occur at this stage. We may not have been able to make the dataset separation randomly. Only certain professions, from certain cities, can be chosen as men or women and a model can be built on them. As a result, we may encounter an overfitting problem. This problem can be solved with cross validation.

In K-Folds Cross Validation, we divide our data into k different subsets. We use k-1 subsets to train our data and use the last subset as test data. The average error value obtained as a result of k experiments indicates the validity of our model.

Here we will talk about the steps of cross validation (Ramasubramanian & Singh, 2019):

1. Split the data into k subsets
2. Train a model on k-1 subsets.
3. Test the model on the remaining one subset and estimate the error.
4. Recap Steps 1-3 until all subsets are used completely once for testing
5. It is necessary to find the cross validation error. For this, the errors are averaged using a simulation exercise.

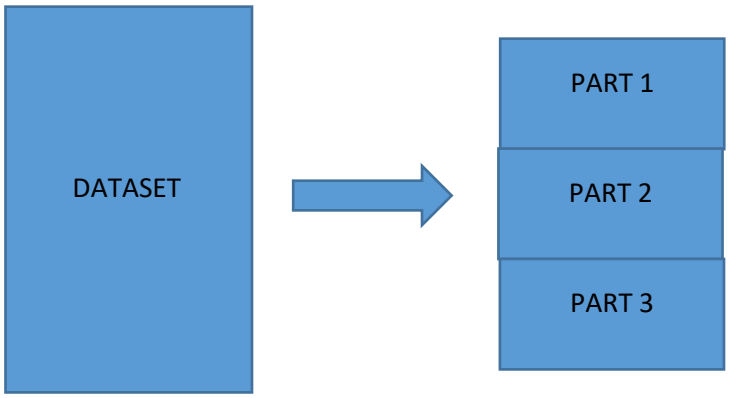
The advantage of this method is that the method of k-subsets generated is insignificant compared to the case in the training/test method. This method provides that all data points must be in the test set once and enter the training set k-1 times. The variance of the resulting forecast also decreases as k increases (Ramasubramanian & Singh, 2019).

The disadvantage of cross validation is that the created model has to be forecasted k times and tested k times. Unfortunately, this creates a high computational cost (the computational cost is proportional to the coefficient). A variant splits the data randomly. It then takes control of all folding dimensions. The advantage of this is the convenience of determining the size of test sets and which trials to average (Ramasubramanian & Singh, 2019).

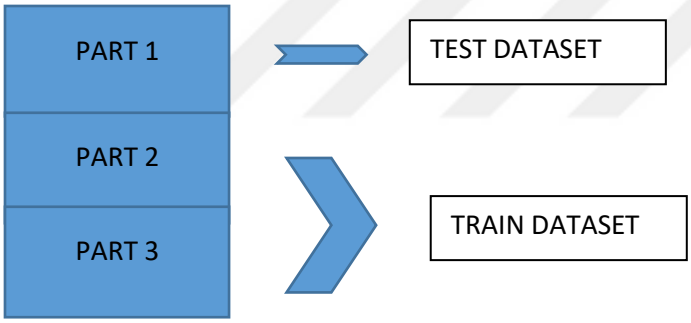
A training set is needed to train a model used for very large datasets. It can be divided into a validation set to validate the model and a test set to evaluate the trained model. Cross-validation is useful when models do not have large data samples. Cross-validation helps us with this when choosing a model and estimating the error in the model (Ghatak, 2017).

In the k-fold cross-validation method, a k value is first selected. In this example, the value of k will be taken as 3. As additional information, the most preferred k value in the

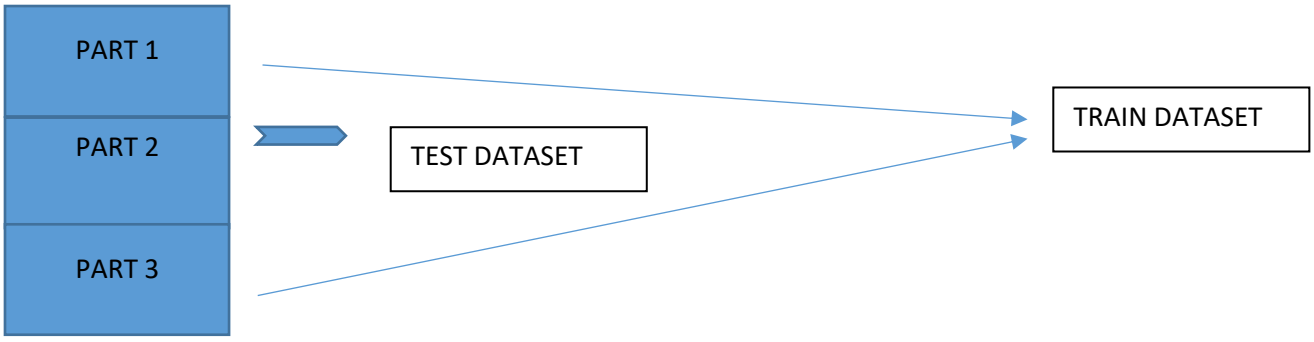
literature is 10. For K= 3, first divide the dataset into 3 equal parts. For example, suppose there are 30 records in the dataset. It is divided into 10 equal parts.



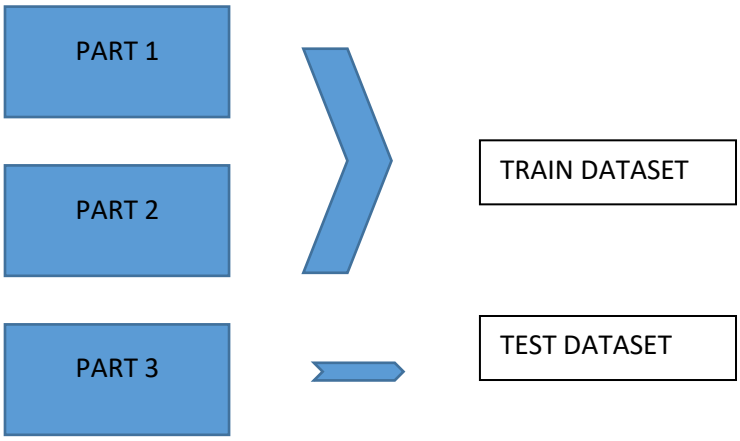
As can be seen in the figure, 3 datasets are divided into 3 since k=3 is taken. After that, k-fold cross-validation starts working. One of the pieces will be used for testing and the remaining 2 will be used for training.



When the system starts to work, if we go from the classification model example, the result is called A1. Then the process continues by selecting the 2nd part.



The classifier is run again and a result appears. The value of this result is also assigned as A2. Finally, section 3 is treated as the test set.



The final result is called A3. As you can see, the system is run in 3 different training and test sets to determine successes and errors. The overall success and error are also calculated as the average of 3 results. The 1/k part of the system dataset, which has not been used for testing before, is used for testing, while the rest is used for training.

The formula below shows the operations performed mathematically:

$$t_i \in VK, \qquad Result = \frac{\sum_{i=0}^{k} CF(t_i, DB - t_i)}{k}$$

In the above equation, CF(test, train) is the classification function, DB is the dataset, k is how many folds are used and t is each test set selected from the dataset. The result is the sum of the performances of all the classification functions divided by k and averaged.

2.4. Parameter Tuning (Grid Search)

The multi-layered architectural structures that emerged with deep learning also revealed a set of hyperparameter groups waiting to be decided by the controller. Some of these parameters are for the selection of the basic algorithm to be applied in the model among a certain number of algorithm groups, as in the examples such as the optimization algorithm and the activation function. Dealing with the selection of such hyperparameters is usually not difficult, as the number of algorithms is limited. However, the number of layers, the number of neurons, the number of learning coefficients, etc. There are also hyperparameter types that wait for us to choose from a set whose range extends to infinity within certain limits or on the number line. The selection of such hyperparameters is a tedious and time-consuming process (Ramasubramanian & Singh, 2019).

The first choices we make for the hyperparameters while designing the model usually do not lead us to the right results. By iteratively changing the hyperparameters one after the other, the performance of the model is observed and the most suitable hyperparameter group for the model is tried to be selected.

Some of the hyperparameters are in a position to take an infinite number of values. Using the preliminary information we have about the problem, we can determine the ranges for the values that the hyperparameters can take. Value lists are created for hyperparameters by selecting certain main points from these ranges we have determined. In hyper parameter selection process with Grid search; The network is trained for combinations of all values in the specified range. Afterwards, the results are checked and the best combination is chosen as the hyperparameter group. An image of grid search is given Figure 8.

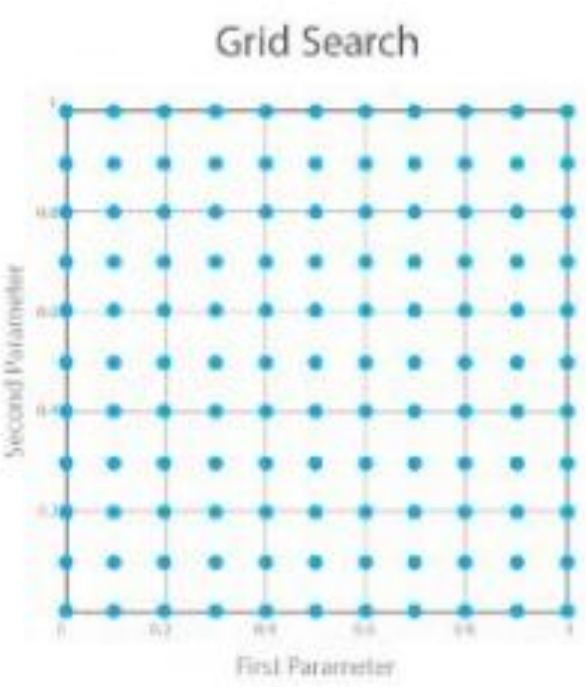


Figure 8. Grid Search (Liashchynskyi & Liashchynskyi, 2019)

It should not be focused only on successful results. However, other results should be considered. So all results should be examined in detail. Fixed parameters or ranges should be observed by observing the relationships between the parameters.

Training deep networks is a long process. Therefore, to save time, a subset of the dataset can be studied to select the hyperparameters. The aim here is to form a general opinion for the parameters. It is determined which ranges hyperparameters will focus on. The values of hyperparameters can be kept large initially. Afterwards, the value ranges can be reduced to the desired level. This will save time. For example, a range such as the number of neurons in the

network [100, 1000, 2000] can be selected and then the range coverage can be changed according to the order [50, 100, 150, 200, 250]. In this way, saves time.

2.5. Literature Survey on ML & Electric Consumption

In this section, a literature survey is conducted on some other studies based on ML and electric consumption in the last one decade. Burger and Moura (2015) suggested to use of an ensemble learning algorithm for short-term electricity consumption estimation. GBR algorithm was more successful than other ML methods in estimation of electric consumption (Touzani, Granderson, & Fernandes, 2018). Divina et al. (2018) proposed that tecniques based on enseble learning has to be used in estimation of electric consumption. Chen et al. (2018) also investigated that enseble learning models or algorithms gave better results than classical regression in estimation of electric consumption.

Khan and Byun (2020) used supervised learning and enseble learning algorithms and Genetic Algorithm for electric consumption estimation. Banik et al. (2021) suggested to use of Random Forest and XGBR algorithms in estimation of electric consumption. Şen et al. (2021) worked on electric consumption estimates in OECD countries by using machine learning algorithms. Opera and Bara (2021) used machine learning methods to estimate electric consumption for Tunisia and found that the best result was GBR. In an estimation study on the electricity consumption of residential buildings, GBR algorithm gave better results than a hybrid model made with ARIMA (Nie, Roccotelli, Fanti, Ming, & Li, 2021) . In another study, three ensemble learning techniques were applied to estimate the electricity consumption of offices as GBR, Random Forests and AdaBoost. In the results, it was seen that the AdaBoost algorithm was more successful (Pinto, Praça, Vale, & Silva, 2021). Bassi et al. (2021) used CatBoost, GBR and XGBR. In this study, XGBR gave the most successful result. Sepulveda et al. (2021) estimated electric consuption using GBR and other machine learning techniques. Porteiro et al. (2022) discussed that ensemble learning is successful in estimating electricity consumption for industrial factories and residences.

According to the results in Sujan Reddy et al. (2022), XGBR model gave more accurate results in the electricity consumption estimation made with the data obtained periodically for 9 years. Albuquerque et al. (2022) investigated that ML models perform better than conventional ones in estimation of electric consumption. Mariano-Hernandez et al. (2022) used different ML methods to estimate electric consiption in the campus of University of Valladolid. Reddick et al. (2022) used GBR to estimate electric consumption for Monash University. Random Forest,

GBR and XGBR models were used to make predictions about the electricity consumption of buildings and successful results were obtained (Moon, Park, Rho, & Hwang, 2022) . In the study of Tyagi and Singh (2022), XGBR was found to be the best method within the different ML and AI methods for estimation of electric consumption. Shin and Woo (2022) also used XGBR method in estimation process.

By considering the last decade studies investigated in this literature survey, it can be seen that ensemble learning methods have been proposed in estimation of electric consumption by many studies. Among the ensemble learning algorithms, the most used and proven methods are the gradient booosting methods, such as GBR and XGBR. Therefore, the motivation of this thesis is to focus on the estimation of electricity consumption data through these two ensemble learning methods.

3. METHODOLOGIES BASED ON GRADIENT BOOSTING

In the concept of Section 3, Both Gradient Boosting (GB) and Extreme Gradient Boosting (XGB) will be discussed. GB and XGB follow the same principle. The main differences between them lie in the details of the implementation. XGB achieves more efficient and effective performance by controlling the complexity of trees using different editing technique (Yıldırım, 2022). An important factor for good results when training an XGB model is parameter tuning. It consists of two steps: first a base model is developed to observe how well XGB performs overall, and then a second model is trained using parameter tuning by comparing the results to the base model. First, it uses a more efficient learning rate to accelerate training. Then, after the tree-specific parameters are set, and finally, the learning rate is reduced. It is then trained to obtain the optimal set of classifiers (Yıldırım, 2022).

XGB helps reduce over-learning or over-fitting as it has regularization. In addition, thanks to its high flexibility, this algorithm allows users to evaluate criteria and define the most appropriate objective function, which adds a new dimension to the model (Chen, Wang, & Pan, 2019). Another advantage of XGB is deep priority search. It is a type of algorithm that has been used so that we can search for tree structures in this algorithm. From the moment it starts searching, it goes to the deepest node it can reach in nodes. If there is no deeper knot, it rewinds and continues with deep knots prioritizing. In simpler terms, these algorithms look for low-level neighbors first. The scope of the study is to determine how the boosting-based algorithms perform in the electricity consumption estimation process. For this reason, it is desired to make these predictions on two algorithms based on bossting in ML.

3.1. Gradient Boosting Regression (GBR)

Gradient Boosting (GB) is an ensemble learning algorithm working based on boosting and it can be used for classification or regression predictive modeling problems. If the target attribute is categorical variable measured in nominal or ordinal scale, the classification ability is considered for GB. If the target variable is a continious variable such as regression modeling or time series, Gradient Boosting Regression (GBR) can help us in estimation or forecasting processes. Due to being a boosting method, GB and GBR are set of algorithms used to solve ML problems. They are an efficient algorithm that can solve millions of examples with a small amount of resources. In 2014, it was added to the library called DMLC (Distributed Machine

Learning Community), which was developed by Tianqi Chen and contains open source machine learning algorithms (Chen, Wang, & Pan, 2019).

Friedman's work in (2001) and (2002) enhanced the work of Friedman, Hastie and Tibshirani (2000) and laid the groundwork for a new generation method.

Below is the algorithm created by Freidman (Ridgeway, 2005):

Initialise $\hat{f}(x)$ to bea steady, $\hat{f}(x) = \arg min_{\rho} \sum_{i=1}^{N} \Psi(y_i, \rho)$.

For t in 1,….., T do

1. Calculate the negative gradient as the running reply

$$z_i = -\frac{\partial}{\partial f(x_i)} \Psi\big(y_i, f(x_i)\big)\big|_{f(x_i)=\hat{f}(x_i)} \quad (1)$$

2. Fit a regression model, g(x), estimating z_i from the covariates x_i.

3. Select a gradient descent step size as

$$\rho = \arg min_{\rho} \sum_{i=1}^{N} \Psi\left(y_i, \hat{f}(x_i) + \rho g(x_i)\right) \quad (2)$$

4. Update the forecast of $f(x)$

$$\hat{f}(x) \leftarrow \hat{f}(x) + \rho g(x) \quad (3)$$

$\Psi\big(y_i, f(x_i)\big)$ function is considered as loss function in the literatere. It can be constructed for continuous response, y. The basic framework shown in Figure 1 can be improved. They have various ways. Friedman's work, which reduces computation times and improves prediction performance (Friedman J. H., 2001).

$$\hat{f}(x) = arg \min_{f(x)} E_{y,x}\, \Psi\big(\mathrm{y}, \mathrm{f(x)}\big)$$

$$= arg \min_{f(x)} E_{\propto}\left[E_{y|x} \Psi\big(\mathrm{y}, \mathrm{f(x)}\big)\big|\mathrm{x}\right] \quad (4)$$

We will concentrate on obtainment calculations of f(x) such that:

$$\hat{f}(x) = arg \min_{f(x)} E_{y|x}\left[\Psi\big(\mathrm{y}, \mathrm{f(x)}\big)\big|\mathrm{x}\right] \quad (5)$$

Assuming that f(x) has a limited number of parameters, β, parametric regression models estimate them by choosing parameter values that minimize a loss function (such as squared error loss) over a training sample of N observations on (y, x) pairs as in (6).

$$\hat{\beta} = arg \min_{\beta} \sum_{i}^{N} \Psi\,(y_i, f(x_i; \beta)) \quad (6)$$

The task gets harder when we want to estimate f(x) non-parametrically. We can see that we want to reduce the N-dimensional function by setting $f_i = f(x_i)$

$$J(\boldsymbol{f}) = \sum_{i=1}^{N} \Psi\left(y_i, f(x_i)\right)$$
$$= \sum_{i=1}^{N} \Psi\,(y_i, F_i) \quad (7)$$

The direction of the locally greatest reduction in $J(\boldsymbol{f})$ is indicated by the negative gradient of $J(\boldsymbol{f})$. We would then need to change f in order to account for gradient descent, where ρ denotes the magnitude of the step along the direction of greatest fall.

$$\hat{\boldsymbol{f}} \leftarrow \hat{\boldsymbol{f}} - \rho \Delta J(\boldsymbol{f}) \quad (8)$$

This action alone obviously falls short of our desired outcome. First, it only accounts for f for x values for which we have data. Second, it ignores the likelihood that observations with similar x values will have comparable values of f(x). Both of these issues would drastically increase generalization error. Friedman advises picking a group of functions, typically a regression tree, that use the covariate data to approximate the gradient. At each iteration, the algorithm chooses a specific model from the list of permissible functions that is most consistent with the direction in which it wants to enhance the fit to the data. Regarding squared-error loss $\Psi\left(y_i, f(x_i)\right) = \sum_{i}^{N}(y_i - f(x_i))^2$ with residual fitting, this algorithm correlates exactly. (Ridgeway, 2005)

3.2. Extreme Gradient Boosting Regression (XGBR)

Extreme Gradient Boosting (XGB or XGBoost) method is a new application method for Gradient Boosting. XGB purposes to avoid overfitting and at the same time optimize computational sources. Besides, parallel calculations are automatically made for functions in XGB during the training phase (Fan, et al., 2018).

XGB and its regression version (XGBR) are popular and well-known algorithms. They are frequently used in the energy, health and finance sectors. One of its advantages is that it is more performant and faster than other algorithms. Estimation studies have high predictive power.

The function at step t is as given below: (Fan, et al., 2018)

$$f_i^{(t)} = \sum_{k=1}^{t} f_k(x_i) = f_i^{(t-1)} + f_t(x_i) \quad (9)$$

$f_t(x_i)$ seen in the formula means the learner at step t. Estimates obtained at t and t−1 steps are also shown as $f_i^{(t)}$ and $f_i^{(t-1)}$. x_i indicates the outputs.

To avoid the problem of overfitting, the XGBR model derives the following analytical expression from the above-mentioned function to evaluate the "goodness" of the model (Fan et al., 2018):

$$Obj^{(t)} = \sum_{k=1}^{n} |l(\overline{y_i}, y_i)| + \sum_{k=1}^{t} \Omega(f_i) \quad (10)$$

l denotes the loss function. n is the number of observations. Ω is the regularisation term and is as below (Fan, et al., 2018):

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \|\omega^2\| \quad (11)$$

Where ω denotes the vector of points on the leaves. λ is the regularization parameter and γ is the minimum loss required to uttermore division the leaf node (Fan, et al., 2018). The detailed information and computation procedures can be found in Chen and Guestrin (2016) .

4. ESTIMATION OF TOTAL ELECTRIC CONSUMPTIONS OF FIRMS: A REAL WORLD DATA APPLICATION

Our aim in this study is to estimate the electricity consumption of industrial companies in the best way with machine learning algorithms. Algorithms and models described in the study are only a part of ML. In our study, datasets of companies in the real world were used. Since these companies are profit making companies and are in competition with each other, their names will not be given. Companies will be shown with letters and the sector they are in. Our estimates were obtained by using the aggregate consumption data in the dataframe.

Section 4 will be explained in 3 sub-sections as “Data Description” and “Models and Analysis” and “Results”. Data is collected from 1st of January 2021 to 31st of December 2022. The data were obtained as a result of the negotiations with the customers (firm A, firm B, firm C, etc.) of a commercial firm engaged in energy management, and demanding the amount of electricity consumed on an hourly basis within 1 year from the firms. These data were obtained with the permission of each company. Furthermore, real names of the companies are not used in the study to ensure confidentiality. These companies consist of production factories. Since the electricity consumption of the factories is at very high levels, energy management and electricity consumption forecast data are of great importance for the factories. Below are the codes of the companies whose datasets are used and information about which sector they are in. There are 11 companies (firms) in total.

FIRM A ➔ Wire Sector

FIRM B ➔ Ceramic Sector

FIRM C ➔ Food Sector

FIRM D ➔ Chemical Sector

FIRM E ➔ Ceramic Sector

FIRM F ➔ Textile Sector

FIRM G ➔ Wire Sector

FIRM H ➔ Ceramic Sector

FIRM I ➔ Aluminium Sector

FIRM J ➔ White Appliance Sector

FIRM K ➔ Textile Sector

4.1. Data Description

The total consumption data of these companies were taken hourly. In other words, there is data on how much electricity consumption will be at the beginning of every hour. Since the electricity consumption consumptions of the companies on weekends and weekdays are not the same, “weekdays” and “weekend days” are also considered in our study. The explanations of the variables in the dataframe are given below in Table 1:

Hourly Data	Data between 00.00 – 23.00 hours were used.
Whether it’s the weekend	*If the day is weekday assigned “0” *If the day is weekend assigned “1”
Total Consumption	Continous Variable (Megawatt-hour)

Table 1. The Variables in Data Frame

The descriptive statistics values of 11 companies are shown in the table below. In total, 8760 hours, ie exactly 1 year of data, were used and this is the same for all companies.

FIRM A (Wire Sector)

For Firm A, there is 8760 aggregate consumption data collected during the year on an hourly basis. The descriptive statistics for the consumption data of Firm A is shown in Table 2 and the sequence graph of the electricity consumption realized in 1 year for hourly data is shown in Figure 9.

Mean	14.6517 MW
Std	2.6968 MW
Min	0.0000 MW
1st Quartile	13.5223 MW
Median	15.2010 MW
3rd Quartile	16.3528 MW
Max	25.4864 MW

Table 2. Descriptive Statistics for Total Electric Consumptions of Firm A

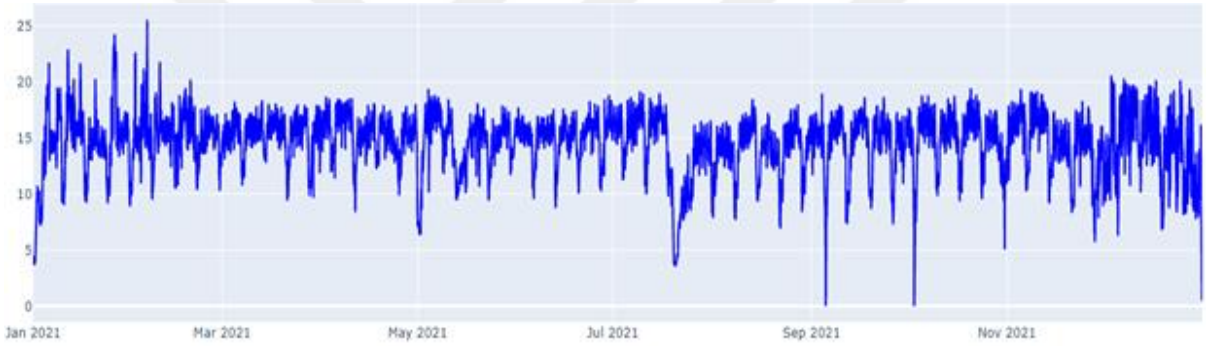


Figure 9. Sequence Graph of Total Electric Consumption of Firm A

FIRM B (Ceramic Sector)

For Firm B, there is 8760 aggregate consumption data collected during the year on an hourly basis. The descriptive statistics for the consumption data of Firm B is shown in Table 3 and the sequence graph of the electricity consumption realized in 1 year for hourly data is shown in Figure 10.

Mean	8.4213MW
Std	0.6633 MW
Min	0.0033 MW
1st Quartile	8.0231 MW
Median	8.4436 MW
3rd Quartile	8.8508 MW
Max	18.2348 MW

Table 3. Descriptive Statistics for Total Electric Consumptions of Firm B

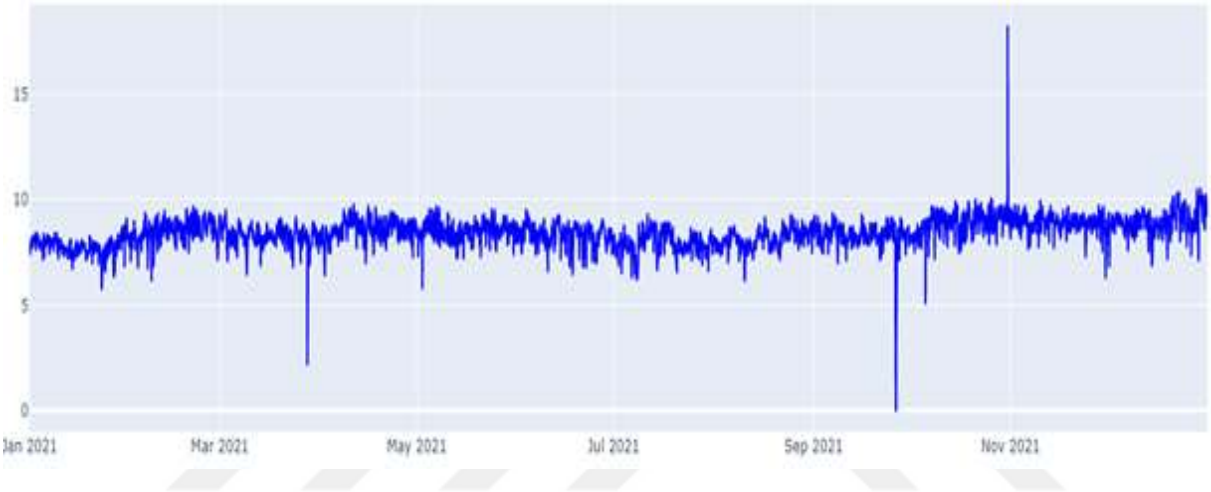


Figure 10. Sequence Graph of Total Electric Consumption of Firm B

FIRM C (Food Sector)

For Firm C, there is 8760 aggregate consumption data collected during the year on an hourly basis. The descriptive statistics for the consumption data of Firm C is shown in Table 4 and the sequence graph of the electricity consumption realized in 1 year for hourly data is shown in Figure 11.

Mean	10.3949 MW
Std	2.5303 MW
Min	0.1465 MW
1st Quartile	9.2845 MW
Median	10.5623 MW
3rd Quartile	11.3513 MW
Max	19.2585 MW

Table 4. Descriptive Statistics for Total Electric Consumptions of Firm C

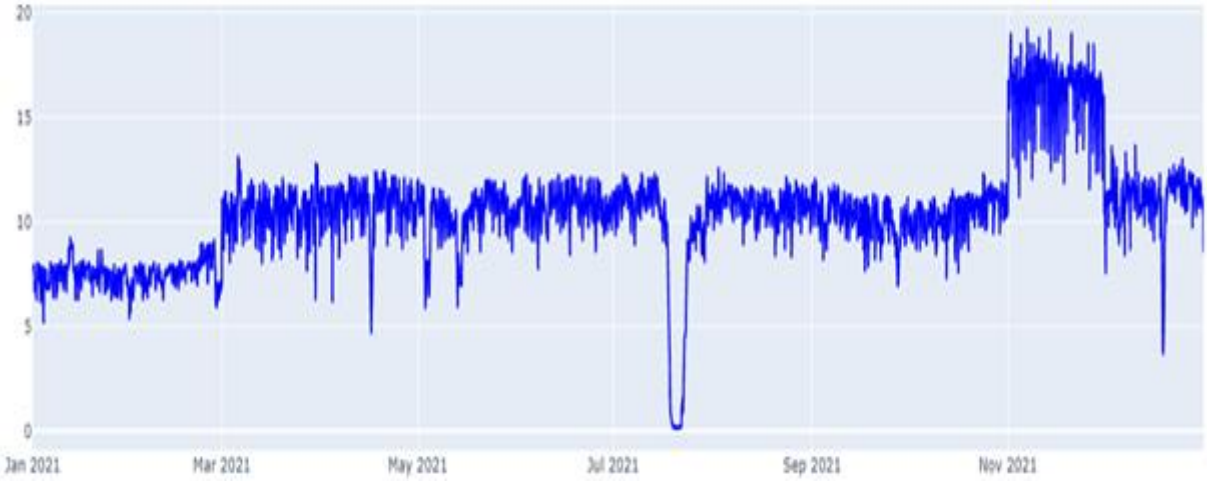


Figure 11. Sequence Graph of Total Electric Consumption of Firm C

FIRM D (Chemical Sector)

For Firm D, there is 8760 aggregate consumption data collected during the year on an hourly basis. The descriptive statistics for the consumption data of Firm C is shown in Table 5 and the sequence graph of the electricity consumption realized in 1 year for hourly data is shown in Figure 12.

Mean	42.5407 MW
Std	6.2925 MW
Min	12.4155 MW
1st Quartile	38.6413 MW
Median	43.4490 MW
3rd Quartile	47.5700 MW
Max	58.5225 MW

Table 5. Descriptive Statistics for Total Electric Consumptions of Firm D

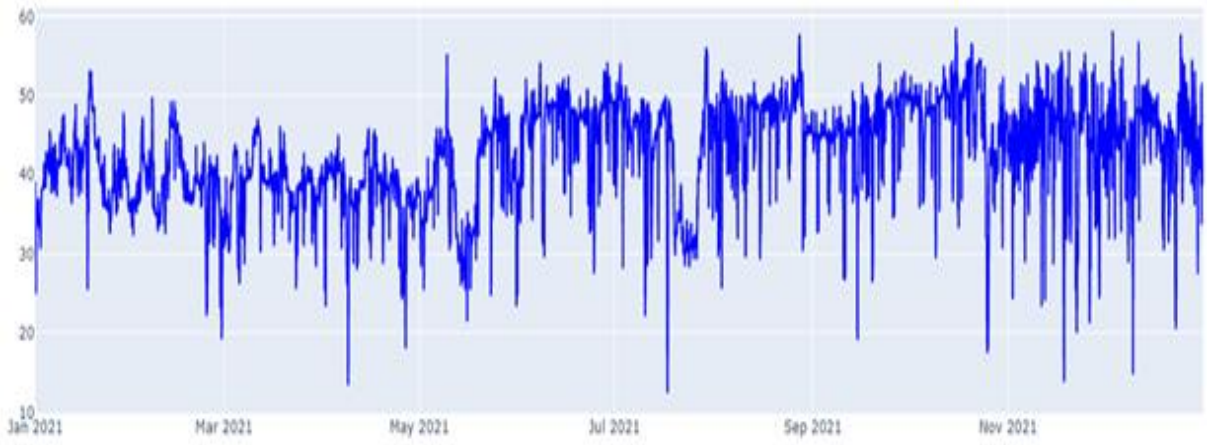


Figure 12. Sequence Graph of Total Electric Consumption of Firm D

FIRM E (Ceramic Sector)

For Firm B, there is 8760 aggregate consumption data collected during the year on an hourly basis. The descriptive statistics for the consumption data of Firm E is shown in Table 6 and the sequence graph of the electricity consumption realized in 1 year for hourly data is shown in Figure 13.

Mean	17.8512 MW
Std	1.3665 MW
Min	7.0484 MW
1st Quartile	16.9947 MW
Median	17.9711 MW
3rd Quartile	18.8060 MW
Max	27.5103 MW

Table 6. Descriptive Statistics for Total Electric Consumptions of Firm E

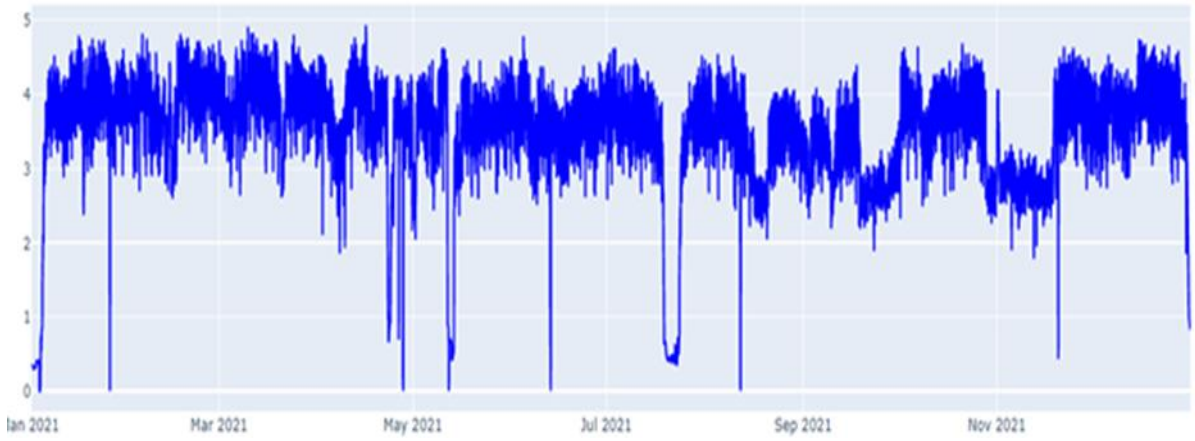


Figure 13. Sequence Graph of Total Electric Consumption of Firm E

FIRM F (Textile Sector)

For Firm F, there is 8760 aggregate consumption data collected during the year on an hourly basis. The descriptive statistics for the consumption data of Firm F is shown in Table 7 and the sequence graph of the electricity consumption realized in 1 year for hourly data is shown in Figure 14.

Mean	17.8183 MW
Std	2.1832 MW
Min	0.2178 MW
1st Quartile	17.2260 MW
Median	18.2099 MW
3rd Quartile	18.8742 MW
Max	24.5898 MW

Table 7. Descriptive Statistics for Total Electric Consumptions of Firm F

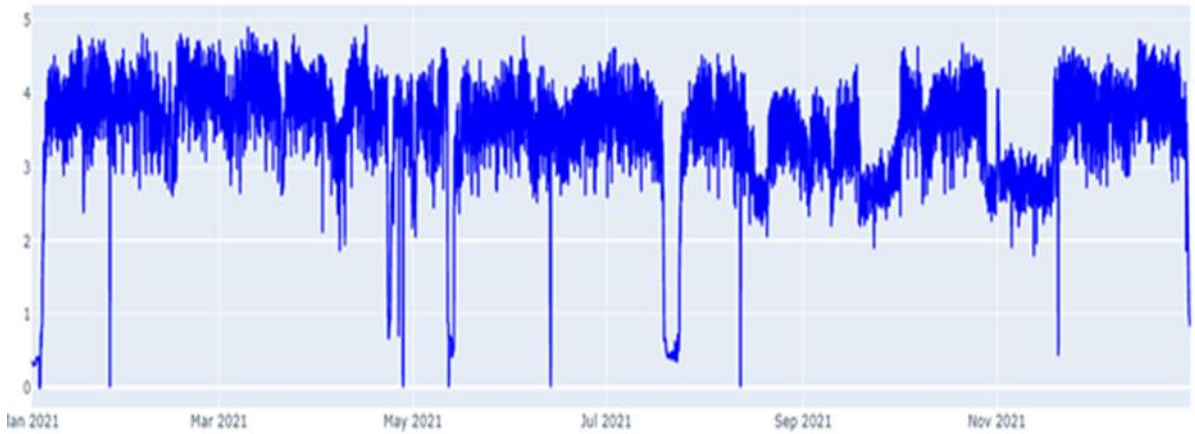


Figure 14. Sequence Graph of Total Electric Consumption of Firm F

FIRM G (Wire Sector)

For Firm G, there is 8760 aggregate consumption data collected during the year on an hourly basis. The descriptive statistics for the consumption data of Firm G is shown in Table 8 and the sequence graph of the electricity consumption realized in 1 year for hourly data is shown in Figure 15.

Mean	3.5099 MW
Std	0.7744 MW
Min	0.0000 MW
1st Quartile	3.2091 MW
Median	3.6592 MW
3rd Quartile	4.0038 MW
Max	4.9340 MW

Table 8. Descriptive Statistics for Total Electric Consumptions of Firm G

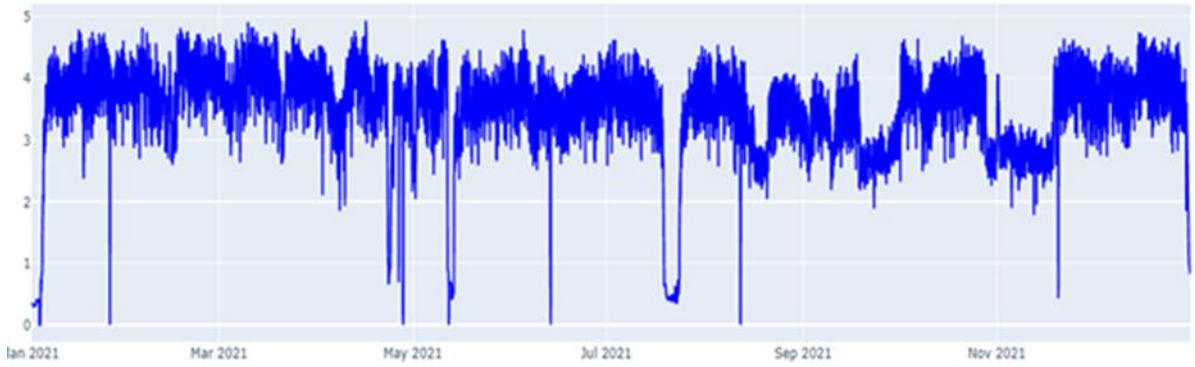


Figure 15. Sequence Graph of Total Electric Consumption of Firm G

FIRM H (Ceramic Sector)

For Firm H, there is 8760 aggregate consumption data collected during the year on an hourly basis. The descriptive statistics for the consumption data of Firm H is shown in Table 9 and the sequence graph of the electricity consumption realized in 1 year for hourly data is shown in Figure 16.

Mean	20.4492 MW
Std	1.4715 MW
Min	4.0600 MW
1st Quartile	19.9800 MW
Median	20.5750 MW
3rd Quartile	21.1500 MW
Max	27.5500 MW

Table 9. Descriptive Statistics for Total Electric Consumptions of Firm H

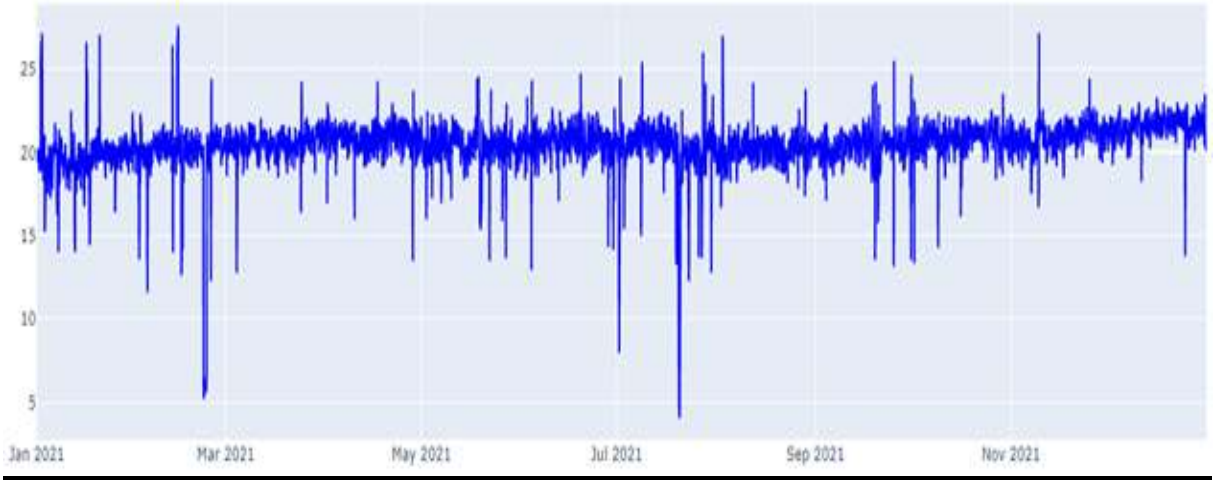


Figure 16. Sequence Graph of Total Electric Consumption of Firm H

FIRM I (Alumunium Sector)

For Firm I, there is 8760 aggregate consumption data collected during the year on an hourly basis. The descriptive statistics for the consumption data of Firm I is shown in Table 10 and the sequence graph of the electricity consumption realized in 1 year for hourly data is shown in Figure 16.

Mean	21.5921 MW
Std	2.8344 MW
Min	0.0264 MW
1st Quartile	20.6482 MW
Median	22.1383 MW
3rd Quartile	23.2913W
Max	28.5157 MW

Table 10. Descriptive Statistics for Total Electric Consumptions of Firm I

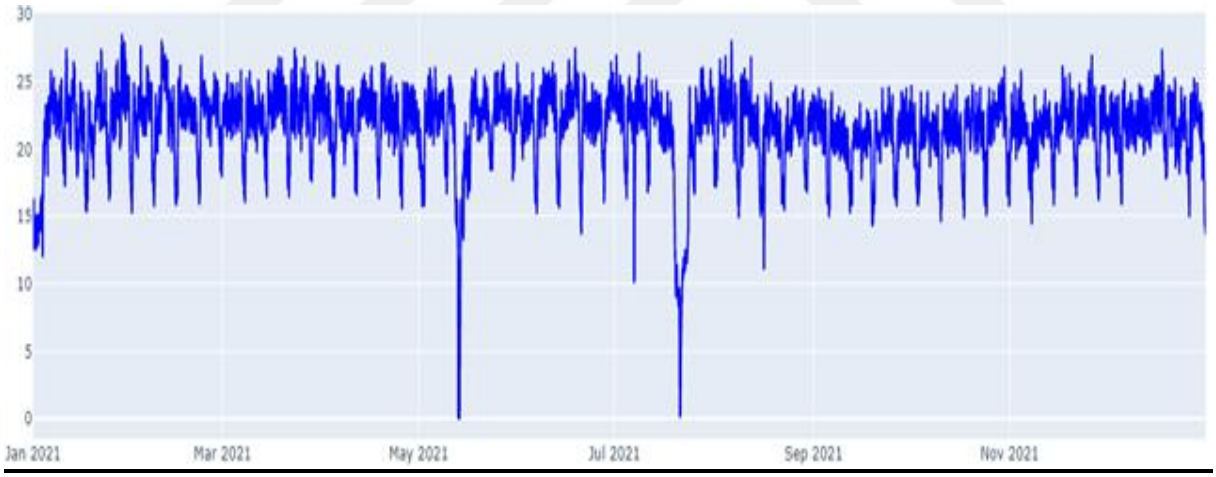


Figure 17. Sequence Graph of Total Electric Consumption of Firm I

FIRM J (White Appliance Sector)

For Firm J, there is 8760 aggregate consumption data collected during the year on an hourly basis. The descriptive statistics for the consumption data of Firm J is shown in Table 11 and the sequence graph of the electricity consumption realized in 1 year for hourly data is shown in Figure 18.

Mean	23.8275 MW
Std	6.0923 MW
Min	4.2050 MW
1st Quartile	21.9021 MW
Median	25.6207 MW
3rd Quartile	27.9539 MW
Max	33.0328 MW

Table 11. Descriptive Statistics for Total Electric Consumptions of Firm J

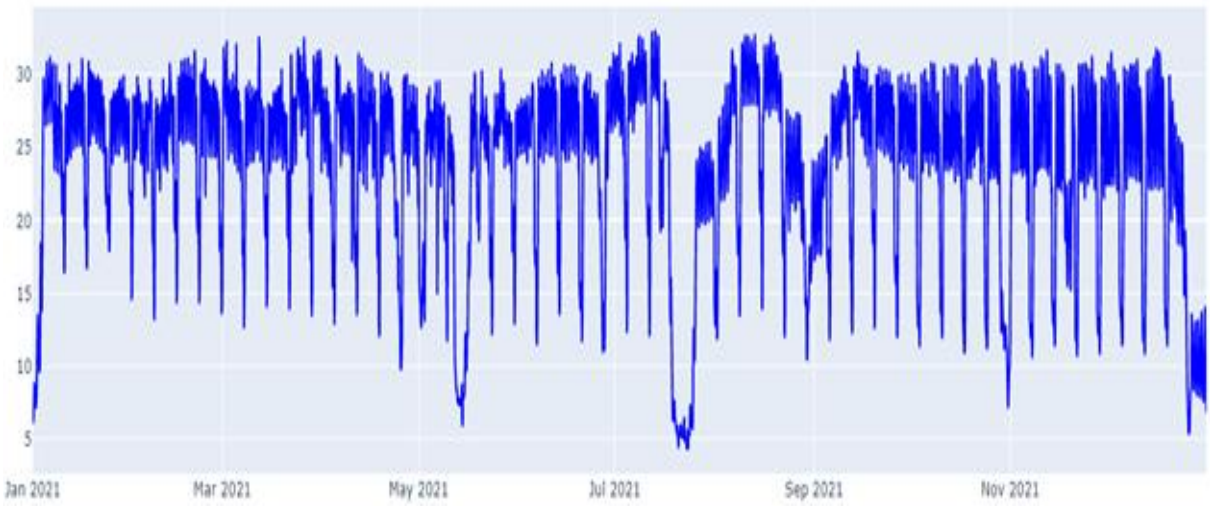


Figure 18. Sequence Graph of Total Electric Consumption of Firm J

FIRM K (Textile Sector)

For Firm K, there is 8760 aggregate consumption data collected during the year on an hourly basis. The descriptive statistics for the consumption data of Firm K is shown in Table 12 and the sequence graph of the electricity consumption realized in 1 year for hourly data is shown in Figure 19.

Mean	5.1889 MW
Std	1.1964 MW
Min	0.0000 MW
1st Quartile	4.5260 MW
Median	5.1689 MW
3rd Quartile	6.1046 MW
Max	8.4854 MW

Table 12. Descriptive Statistics for Total Electric Consumptions of Firm K

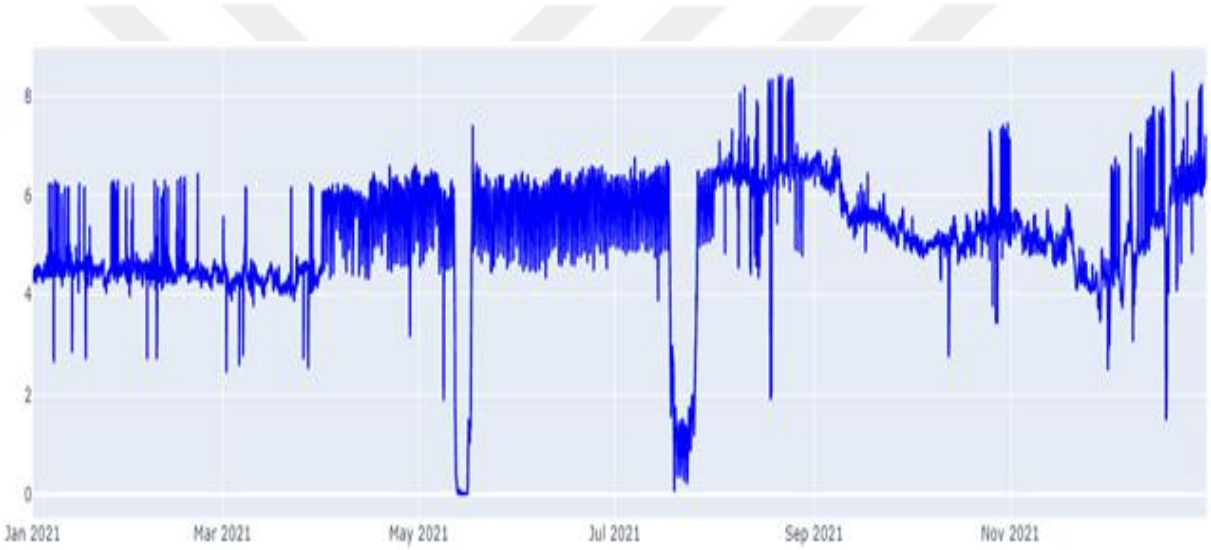


Figure 19. Sequence Graph of Total Electric Consumption of Firm K

4.2. Models and Analysis

Total consumption data of each firm is estimated over 2 different algorithms described in Section 3. In model building process, Hour (0-23), being week da yor weekend (0 or 1) and different lags of the electric consumptions are considered for each different day. The lags are determinated by significant autocorrelations in time series. By this way, it is tried to obtain more accurate estimations. In these estimations, the training set is taken as the first 16 weeks and the test set as the 17th week immediately following the training set. In the next iteration, the training set is shifted from Week 2 to Week 17, and the test set is shifted to Week 18 by 1 week. In this way, this iteration is repeated 36 times (for 1 year data), and the absolute value

averages of the differences between the predicted values and the actual values, in other words absolute errors are taken in each iteration. Finally, the arithmetic mean corresponding to these absolute values (Mean Absolute Error, MAE) is calculated for each method. These procedures are valid for the analysis of each firm.

4.2.1. Models for Firm A (Wire Sector)

In Firm A, the following models are established as estimation model for days.

Monday:

$$\hat{y}_{t_{Firm\ A}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-72_{Firm\ A}}, y_{t-96_{Firm\ A}}, y_{t-168_{Firm\ A}}\}$$

Tuesday:

$$\hat{y}_{t_{Firm\ A}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-96_{Firm\ A}}, y_{t-120_{Firm\ A}}, y_{t-168_{Firm\ A}}\}$$

Wednesday:

$$\hat{y}_{t_{Firm\ A}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-48_{Firm\ A}}, y_{t-120_{Firm\ A}}, y_{t-168_{Firm\ A}}\}$$

Thursday:

$$\hat{y}_{t_{Firm\ A}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-48_{Firm\ A}}, y_{t-72_{Firm\ A}}, y_{t-168_{Firm\ A}}\}$$

Friday:

$$\hat{y}_{t_{Firm\ A}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-48_{Firm\ A}}, y_{t-72_{Firm\ A}}, y_{t-168_{Firm\ A}}\}$$

Saturday:

$$\hat{y}_{t_{Firm\ A}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-168_{Firm\ A}}, y_{t-168*2_{Firm\ A}}, y_{t-168*3_{Firm\ A}}\}$$

Sunday:

$$\hat{y}_{t_{Firm\ A}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-168_{Firm\ A}}, y_{t-168*2_{Firm\ A}}, y_{t-168*3_{Firm\ A}}\}$$

4.2.2. Models For Firm B (Ceramic Sector)

In Firm B, the following models are established as estimation model for days.

Monday:

$$\hat{y}_{t_{Firm\ B}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-72_{Firm\ B}}, y_{t-96_{Firm\ B}}, y_{t-168_{Firm\ B}}\}$$

Tuesday:

$$\hat{y}_{t_{Firm\ B}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-96_{Firm\ B}}, y_{t-120_{Firm\ B}}, y_{t-168_{Firm\ B}}\}$$

Wednesday:

$$\hat{y}_{t_{Firm\ B}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-48_{Firm\ B}}, y_{t-120_{Firm\ B}}, y_{t-168_{Firm\ B}}\}$$

Thursday:

$$\hat{y}_{t_{Firm\,B}} = gbr\ or\ xgbr\ \{ y_{t_{hour}}, y_{t_{HoW}}, y_{t-48_{Firm\,B}}, y_{t-72_{Firm\,B}}, y_{t-168_{Firm\,B}} \}$$

Friday:

$$\hat{y}_{t_{Firm\,B}} = gbr\ or\ xgbr\ \{ y_{t_{hour}}, y_{t_{HoW}}, y_{t-48_{Firm\,B}}, y_{t-72_{Firm\,B}}, y_{t-168_{Firm\,B}} \}$$

Saturday:

$$\hat{y}_{t_{Firm\,B}} = gbr\ or\ xgbr\ \{ y_{t_{hour}}, y_{t_{HoW}}, y_{t-168_{Firm\,B}}, y_{t-168*2_{Firm\,B}}, y_{t-168*3_{Firm\,B}} \}$$

Sunday:

$$\hat{y}_{t_{Firm\,B}} = gbr\ or\ xgbr\ \{ y_{t_{hour}}, y_{t_{HoW}}, y_{t-168_{Firm\,B}}, y_{t-168*2_{Firm\,B}}, y_{t-168*3_{Firm\,B}} \}$$

4.2.3. Models For Firm C (Food Sector)

In Firm C, the following models are established as estimation model for days.

Monday:

$$\hat{y}_{t_{Firm\,C}} = gbr\ or\ xgbr\ \{ y_{t_{hour}}, y_{t_{HoW}}, y_{t-72_{Firm\,C}}, y_{t-96_{Firm\,C}}, y_{t-168_{Firm\,C}} \}$$

Tuesday:

$$\hat{y}_{t_{Firm\,C}} = gbr\ or\ xgbr\ \{ y_{t_{hour}}, y_{t_{HoW}}, y_{t-96_{Firm\,C}}, y_{t-120_{Firm\,C}}, y_{t-168_{Firm\,C}} \}$$

Wednesday:

$$\hat{y}_{t_{Firm\,C}} = gbr\ or\ xgbr\ \{ y_{t_{hour}}, y_{t_{HoW}}, y_{t-48_{Firm\,C}}, y_{t-120_{Firm\,C}}, y_{t-168_{Firm\,C}} \}$$

Thursday:

$$\hat{y}_{t_{Firm\,C}} = gbr\ or\ xgbr\ \{ y_{t_{hour}}, y_{t_{HoW}}, y_{t-48_{Firm\,C}}, y_{t-72_{Firm\,C}}, y_{t-168_{Firm\,C}} \}$$

Friday:

$$\hat{y}_{t_{Firm\,C}} = gbr\ or\ xgbr\ \{ y_{t_{hour}}, y_{t_{HoW}}, y_{t-48_{Firm\,C}}, y_{t-72_{Firm\,C}}, y_{t-168_{Firm\,C}} \}$$

Saturday:

$$\hat{y}_{t_{Firm\,C}} = gbr\ or\ xgbr\ \{ y_{t_{hour}}, y_{t_{HoW}}, y_{t-168_{Firm\,C}}, y_{t-168*2_{Firm\,C}}, y_{t-168*3_{Firm\,C}} \}$$

Sunday:

$$\hat{y}_{t_{Firm\,C}} = gbr\ or\ xgbr\ \{ y_{t_{hour}}, y_{t_{HoW}}, y_{t-168_{Firm\,C}}, y_{t-168*2_{Firm\,C}}, y_{t-168*3_{Firm\,C}} \}$$

4.2.4. Models For Firm D (Chemical Sector)

In Firm D, the following models are established as estimation model for days.

Monday:

$$\hat{y}_{t_{Firm\,D}} = gbr\ or\ xgbr\ \{ y_{t_{hour}}, y_{t_{HoW}}, y_{t-72_{Firm\,D}}, y_{t-96_{Firm\,D}}, y_{t-168_{Firm\,D}} \}$$

Tuesday:

$$\hat{y}_{t_{Firm\,D}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-96_{Firm\,D}}, y_{t-120_{Firm\,D}}, y_{t-168_{Firm\,D}}\}$$

Wednesday:

$$\hat{y}_{t_{Firm\,D}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-48_{Firm\,D}}, y_{t-120_{Firm\,D}}, y_{t-168_{Firm\,D}}\}$$

Thursday:

$$\hat{y}_{t_{Firm\,D}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-48_{Firm\,D}}, y_{t-72_{Firm\,D}}, y_{t-168_{Firm\,D}}\}$$

Friday:

$$\hat{y}_{t_{Firm\,D}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-48_{Firm\,D}}, y_{t-72_{Firm\,D}}, y_{t-168_{Firm\,D}}\}$$

Saturday:

$$\hat{y}_{t_{Firm\,D}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-168_{Firm\,D}}, y_{t-168*2_{Firm\,D}}, y_{t-168*3_{Firm\,D}}\}$$

Sunday:

$$\hat{y}_{t_{Firm\,D}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-168_{Firm\,D}}, y_{t-168*2_{Firm\,D}}, y_{t-168*3_{Firm\,D}}\}$$

4.2.5. Models For Firm E (Ceramic Sector)

In Firm E, the following models are established as estimation model for days.

Monday:

$$\hat{y}_{t_{Firm\,E}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-72_{Firm\,E}}, y_{t-96_{Firm\,E}}, y_{t-168_{Firm\,E}}\}$$

Tuesday:

$$\hat{y}_{t_{Firm\,E}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-96_{Firm\,E}}, y_{t-120_{Firm\,E}}, y_{t-168_{Firm\,E}}\}$$

Wednesday:

$$\hat{y}_{t_{Firm\,E}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-48_{Firm\,E}}, y_{t-120_{Firm\,E}}, y_{t-168_{Firm\,E}}\}$$

Thursday:

$$\hat{y}_{t_{Firm\,E}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-48_{Firm\,E}}, y_{t-72_{Firm\,E}}, y_{t-168_{Firm\,E}}\}$$

Friday:

$$\hat{y}_{t_{Firm\,E}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-48_{Firm\,E}}, y_{t-72_{Firm\,E}}, y_{t-168_{Firm\,E}}\}$$

Saturday:

$$\hat{y}_{t_{Firm\,E}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-168_{Firm\,E}}, y_{t-168*2_{Firm\,E}}, y_{t-168*3_{Firm\,E}}\}$$

Sunday:

$$\hat{y}_{t_{Firm\,E}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-168_{Firm\,E}}, y_{t-168*2_{Firm\,E}}, y_{t-168*3_{Firm\,E}}\}$$

4.2.6. Models For Firm F (Textile Sector)

In Firm F, the following models are established as estimation model for days.

Monday:

$$\hat{y}_{t_{Firm\,F}} = gbr\ or\ xgbr\ \{ y_{t_{hour}}, y_{t_{HoW}}, y_{t-72_{Firm\,F}}, y_{t-96_{Firm\,F}}, y_{t-168_{Firm\,F}} \}$$

Tuesday:

$$\hat{y}_{t_{Firm\,F}} = gbr\ or\ xgbr\ \{ y_{t_{hour}}, y_{t_{HoW}}, y_{t-96_{Firm\,F}}, y_{t-120_{Firm\,F}}, y_{t-168_{Firm\,F}} \}$$

Wednesday:

$$\hat{y}_{t_{Firm\,F}} = gbr\ or\ xgbr\ \{ y_{t_{hour}}, y_{t_{HoW}}, y_{t-48_{Firm\,F}}, y_{t-120_{Firm\,F}}, y_{t-168_{Firm\,F}} \}$$

Thursday:

$$\hat{y}_{t_{Firm\,F}} = gbr\ or\ xgbr\ \{ y_{t_{hour}}, y_{t_{HoW}}, y_{t-48_{Firm\,F}}, y_{t-72_{Firm\,F}}, y_{t-168_{Firm\,F}} \}$$

Friday:

$$\hat{y}_{t_{Firm\,F}} = gbr\ or\ xgbr\ \{ y_{t_{hour}}, y_{t_{HoW}}, y_{t-48_{Firm\,F}}, y_{t-72_{Firm\,F}}, y_{t-168_{Firm\,F}} \}$$

Saturday:

$$\hat{y}_{t_{Firm\,F}} = gbr\ or\ xgbr\ \{ y_{t_{hour}}, y_{t_{HoW}}, y_{t-168_{Firm\,F}}, y_{t-168*2_{Firm\,F}}, y_{t-168*3_{Firm\,F}} \}$$

Sunday:

$$\hat{y}_{t_{Firm\,F}} = gbr\ or\ xgbr\ \{ y_{t_{hour}}, y_{t_{HoW}}, y_{t-168_{Firm\,F}}, y_{t-168*2_{Firm\,F}}, y_{t-168*3_{Firm\,F}} \}$$

4.2.7. Models For Firm G (Wire Sector)

In Firm G, the following models are established as estimation model for days.

Monday:

$$\hat{y}_{t_{Firm\,G}} = gbr\ or\ xgbr\ \{ y_{t_{hour}}, y_{t_{HoW}}, y_{t-72_{Firm\,G}}, y_{t-96_{Firm\,G}}, y_{t-168_{Firm\,G}} \}$$

Tuesday:

$$\hat{y}_{t_{Firm\,G}} = gbr\ or\ xgbr\ \{ y_{t_{hour}}, y_{t_{HoW}}, y_{t-96_{Firm\,G}}, y_{t-120_{Firm\,G}}, y_{t-168_{Firm\,G}} \}$$

Wednesday:

$$\hat{y}_{t_{Firm\,G}} = gbr\ or\ xgbr\ \{ y_{t_{hour}}, y_{t_{HoW}}, y_{t-48_{Firm\,G}}, y_{t-120_{Firm\,G}}, y_{t-168_{Firm\,G}} \}$$

Thursday:

$$\hat{y}_{t_{Firm\,G}} = gbr\ or\ xgbr\ \{ y_{t_{hour}}, y_{t_{HoW}}, y_{t-48_{Firm\,G}}, y_{t-72_{Firm\,G}}, y_{t-168_{Firm\,G}} \}$$

Friday:

$$\hat{y}_{t_{Firm\,G}} = gbr\ or\ xgbr\ \{ y_{t_{hour}}, y_{t_{HoW}}, y_{t-48_{Firm\,G}}, y_{t-72_{Firm\,G}}, y_{t-168_{Firm\,G}} \}$$

Saturday:

$$\hat{y}_{t_{Firm\,G}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-168_{Firm\,G}}, y_{t-168*2_{Firm\,G}}, y_{t-168*3_{Firm\,G}}\}$$

Sunday:

$$\hat{y}_{t_{Firm\,G}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-168_{Firm\,G}}, y_{t-168*2_{Firm\,G}}, y_{t-168*3_{Firm\,G}}\}$$

4.2.8. Models For Firm H (Ceramic Sector)

In Firm H, the following models are established as estimation model for days.

Monday:

$$\hat{y}_{t_{Firm\,H}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-72_{Firm\,H}}, y_{t-96_{Firm\,H}}, y_{t-168_{Firm\,H}}\}$$

Tuesday:

$$\hat{y}_{t_{Firm\,H}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-96_{Firm\,H}}, y_{t-120_{Firm\,H}}, y_{t-168_{Firm\,H}}\}$$

Wednesday:

$$\hat{y}_{t_{Firm\,H}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-48_{Firm\,H}}, y_{t-120_{Firm\,H}}, y_{t-168_{Firm\,H}}\}$$

Thursday:

$$\hat{y}_{t_{Firm\,H}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-48_{Firm\,H}}, y_{t-72_{Firm\,H}}, y_{t-168_{Firm\,H}}\}$$

Friday:

$$\hat{y}_{t_{Firm\,H}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-48_{Firm\,H}}, y_{t-72_{Firm\,H}}, y_{t-168_{Firm\,H}}\}$$

Saturday:

$$\hat{y}_{t_{Firm\,H}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-168_{Firm\,H}}, y_{t-168*2_{Firm\,H}}, y_{t-168*3_{Firm\,H}}\}$$

Sunday:

$$\hat{y}_{t_{Firm\,H}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-168_{Firm\,H}}, y_{t-168*2_{Firm\,H}}, y_{t-168*3_{Firm\,H}}\}$$

4.2.9. Models For Firm I (Aluminium Sector)

In Firm I, the following models are established as estimation model for days.

Monday:

$$\hat{y}_{t_{Firm\,I}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-72_{Firm\,I}}, y_{t-96_{Firm\,I}}, y_{t-168_{Firm\,I}}\}$$

Tuesday:

$$\hat{y}_{t_{Firm\,I}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-96_{Firm\,I}}, y_{t-120_{Firm\,I}}, y_{t-168_{Firm\,I}}\}$$

Wednesday:

$$\hat{y}_{t_{Firm\,I}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-48_{Firm\,I}}, y_{t-120_{Firm\,I}}, y_{t-168_{Firm\,I}}\}$$

Thursday:

$$\hat{y}_{t_{Firm\ I}} = gbr\ or\ xgbr\ \{ y_{t_{hour}}, y_{t_{HoW}}, y_{t-48_{Firm\ I}}, y_{t-72_{Firm\ I}}, y_{t-168_{Firm\ I}} \}$$

Friday:

$$\hat{y}_{t_{Firm\ I}} = gbr\ or\ xgbr\ \{ y_{t_{hour}}, y_{t_{HoW}}, y_{t-48_{Firm\ I}}, y_{t-72_{Firm\ I}}, y_{t-168_{Firm\ I}} \}$$

Saturday:

$$\hat{y}_{t_{Firm\ I}} = gbr\ or\ xgbr\ \{ y_{t_{hour}}, y_{t_{HoW}}, y_{t-168_{Firm\ I}}, y_{t-168*2_{Firm\ I}}, y_{t-168*3_{Firm\ I}} \}$$

Sunday:

$$\hat{y}_{t_{Firm\ I}} = gbr\ or\ xgbr\ \{ y_{t_{hour}}, y_{t_{HoW}}, y_{t-168_{Firm\ I}}, y_{t-168*2_{Firm\ I}}, y_{t-168*3_{Firm\ I}} \}$$

4.2.10. Models For Firm J (White Appliance Sector)

In Firm J, the following models are established as estimation model for days.

Monday:

$$\hat{y}_{t_{Firm\ J}} = gbr\ or\ xgbr\ \{ y_{t_{hour}}, y_{t_{HoW}}, y_{t-72_{Firm\ J}}, y_{t-96_{Firm\ J}}, y_{t-168_{Firm\ J}} \}$$

Tuesday:

$$\hat{y}_{t_{Firm\ J}} = gbr\ or\ xgbr\ \{ y_{t_{hour}}, y_{t_{HoW}}, y_{t-96_{Firm\ J}}, y_{t-120_{Firm\ J}}, y_{t-168_{Firm\ J}} \}$$

Wednesday:

$$\hat{y}_{t_{Firm\ J}} = gbr\ or\ xgbr\ \{ y_{t_{hour}}, y_{t_{HoW}}, y_{t-48_{Firm\ J}}, y_{t-120_{Firm\ J}}, y_{t-168_{Firm\ J}} \}$$

Thursday:

$$\hat{y}_{t_{Firm\ J}} = gbr\ or\ xgbr\ \{ y_{t_{hour}}, y_{t_{HoW}}, y_{t-48_{Firm\ J}}, y_{t-72_{Firm\ J}}, y_{t-168_{Firm\ J}} \}$$

Friday:

$$\hat{y}_{t_{Firm\ J}} = gbr\ or\ xgbr\ \{ y_{t_{hour}}, y_{t_{HoW}}, y_{t-48_{Firm\ J}}, y_{t-72_{Firm\ J}}, y_{t-168_{Firm\ J}} \}$$

Saturday:

$$\hat{y}_{t_{Firm\ J}} = gbr\ or\ xgbr\ \{ y_{t_{hour}}, y_{t_{HoW}}, y_{t-168_{Firm\ J}}, y_{t-168*2_{Firm\ J}}, y_{t-168*3_{Firm\ J}} \}$$

Sunday:

$$\hat{y}_{t_{Firm\ J}} = gbr\ or\ xgbr\ \{ y_{t_{hour}}, y_{t_{HoW}}, y_{t-168_{Firm\ J}}, y_{t-168*2_{Firm\ J}}, y_{t-168*3_{Firm\ J}} \}$$

4.2.11. Models for Firm K (Textile Sector)

In Firm K, the following models are established as estimation model for days.

Monday:

$$\hat{y}_{t_{Firm\ K}} = gbr\ or\ xgbr\ \{ y_{t_{hour}}, y_{t_{HoW}}, y_{t-72_{Firm\ K}}, y_{t-96_{Firm\ K}}, y_{t-168_{Firm\ K}} \}$$

Tuesday:

$$\hat{y}_{t_{Firm\,K}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-96_{Firm\,K}}, y_{t-120_{Firm\,K}}, y_{t-168_{Firm\,K}}\}$$

Wednesday:

$$\hat{y}_{t_{Firm\,K}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-48_{Firm\,K}}, y_{t-120_{Firm\,K}}, y_{t-168_{Firm\,K}}\}$$

Thursday:

$$\hat{y}_{t_{Firm\,K}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-48_{Firm\,K}}, y_{t-72_{Firm\,K}}, y_{t-168_{Firm\,K}}\}$$

Friday:

$$\hat{y}_{t_{Firm\,K}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-48_{Firm\,K}}, y_{t-72_{Firm\,K}}, y_{t-168_{Firm\,K}}\}$$

Saturday:

$$\hat{y}_{t_{Firm\,K}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-168_{Firm\,K}}, y_{t-168*2_{Firm\,K}}, y_{t-168*3_{Firm\,K}}\}$$

Sunday:

$$\hat{y}_{t_{Firm\,K}} = gbr\ or\ xgbr\ \{y_{t_{hour}}, y_{t_{HoW}}, y_{t-168_{Firm\,K}}, y_{t-168*2_{Firm\,K}}, y_{t-168*3_{Firm\,K}}\}$$

4.3.Results

Accordingly, although the parameter estimation is made, the significance tests of papameter estimations are not considered due to the logic of ML. ML deals with how the model that learns from the training dataset performs with the test dataset. In the concept of this study, as mentioned in Section 2, MAE is used as a performance measure. The MAE values for the methods for each firm are given in Table 13-24 and sequence graphs of estimations for the methods are visualized in Figure 20-41 below:

FIRM A	*Mean Absolute Error*
GBR	1.2653267483124182
XGBR	1.261672658274571

Table 13. Mean Absolute Error (MAE) for Firm A

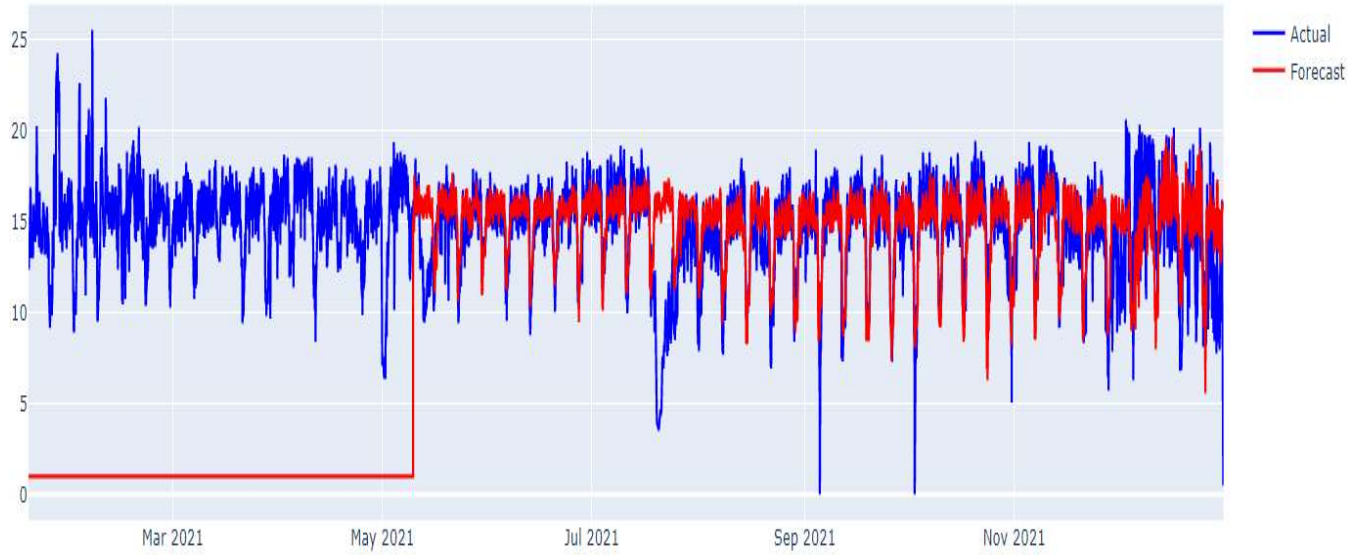


Figure 20.Sequence Graph of GBR Estimations and Total Electric Consumption of Firm A

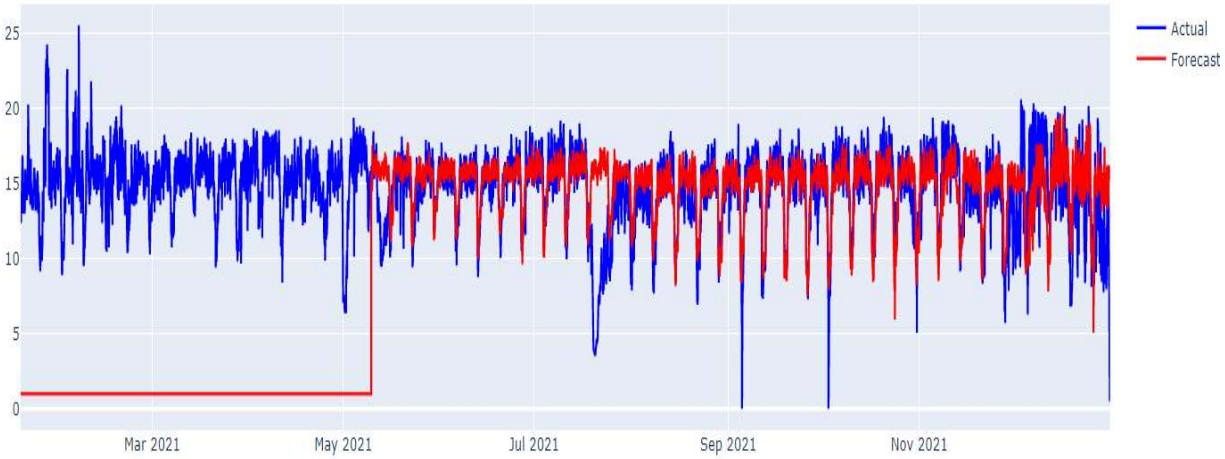


Figure 21. Sequence Graph of XGBR Estimations and Total Electric Consumption of Firm A

FIRM B	*Mean Absolute Error*
GBR	0.44388306164427427
XGBR	0.4409897354516245

Table 14. Mean Absolute Error (MAE) for Firm B

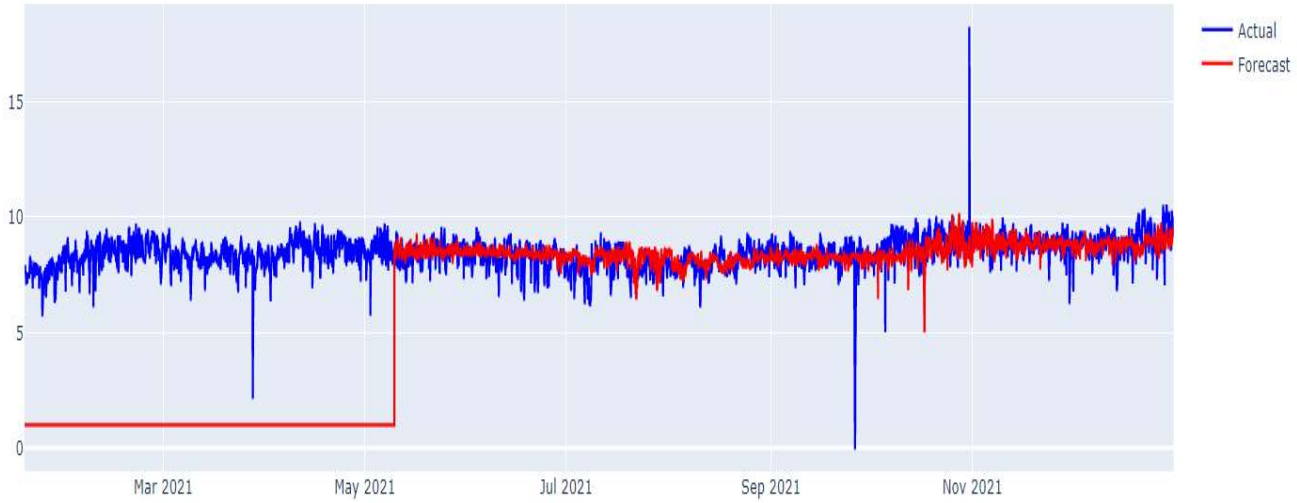


Figure 22. Sequence Graph of GBR Estimations and Total Electric Consumption of Firm B

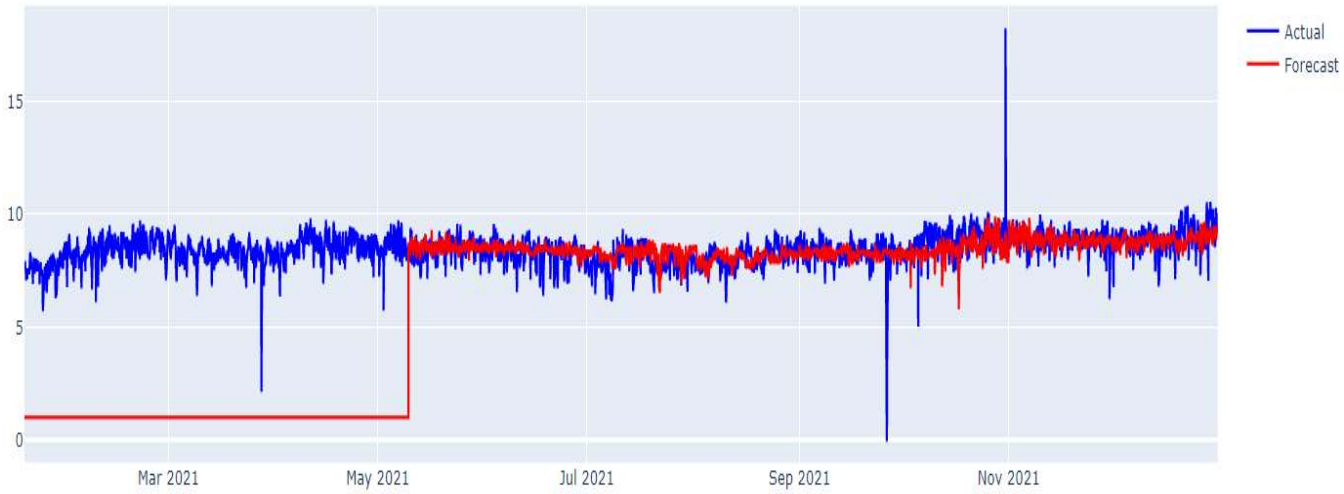


Figure 23. Sequence Graph of XGBR Estimations and Total Electric Consumption of Firm B

FIRM C	*Mean Absolute Error*
GBR	1.071974065092944
XGBR	1.0728962020918016

Table 15. Mean Absolute Error (MAE) for Firm C

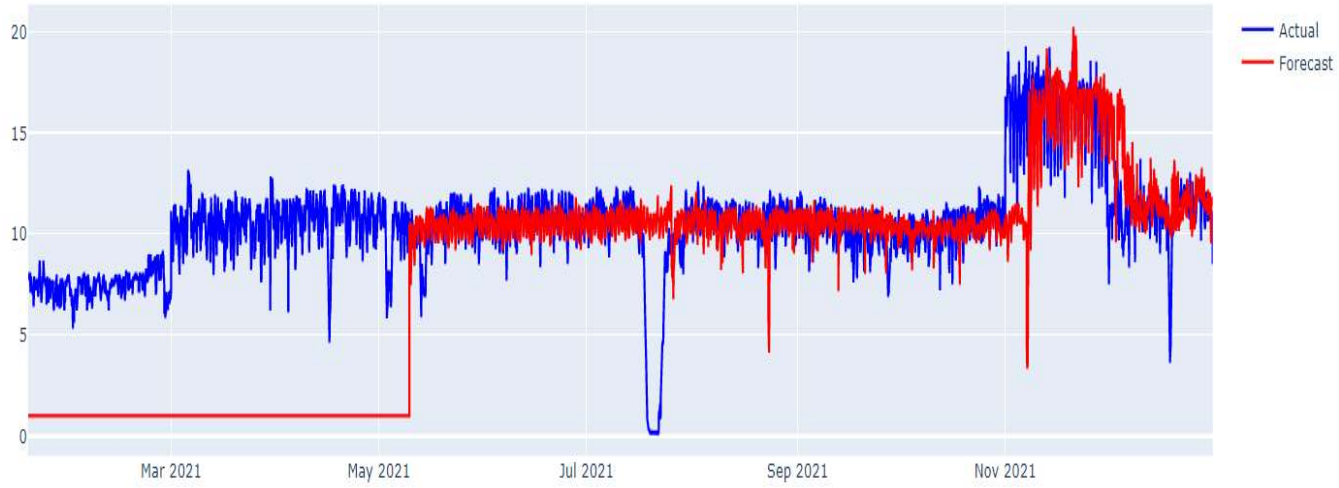


Figure 24. Sequence Graph of GBR Estimations and Total Electric Consumption of Firm C

Figure 25. Sequence Graph of XGBR Estimations and Total Electric Consumption of Firm C

FIRM D	*Mean Absolute Error*
GBR	4.751989541882788
XGBR	4.736551897891931

Table 16. Mean Absolute Error (MAE) for Firm D

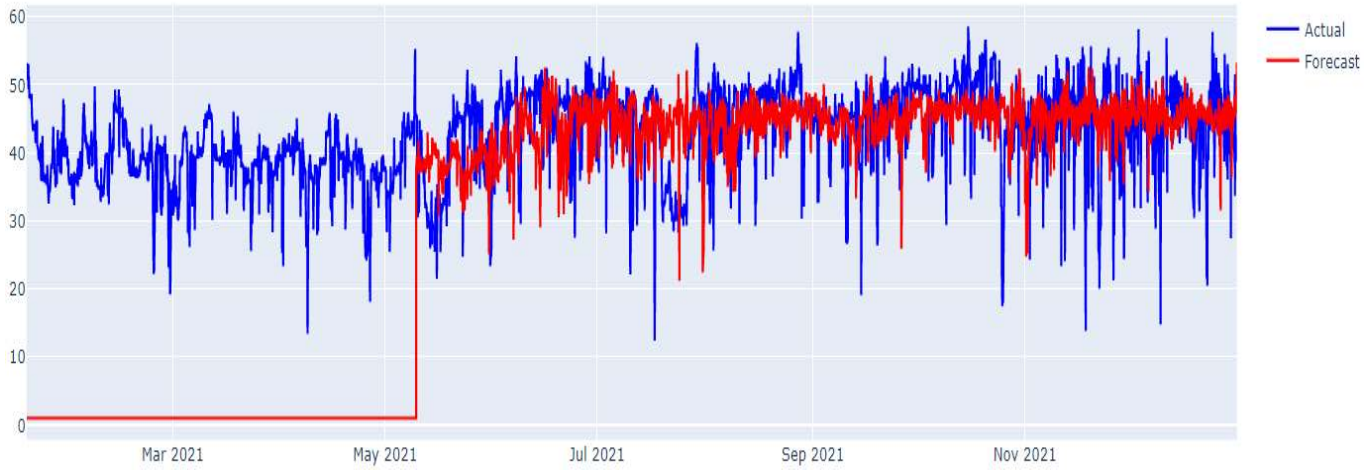


Figure 26. Sequence Graph of GBR Estimations and Total Electric Consumption of Firm D

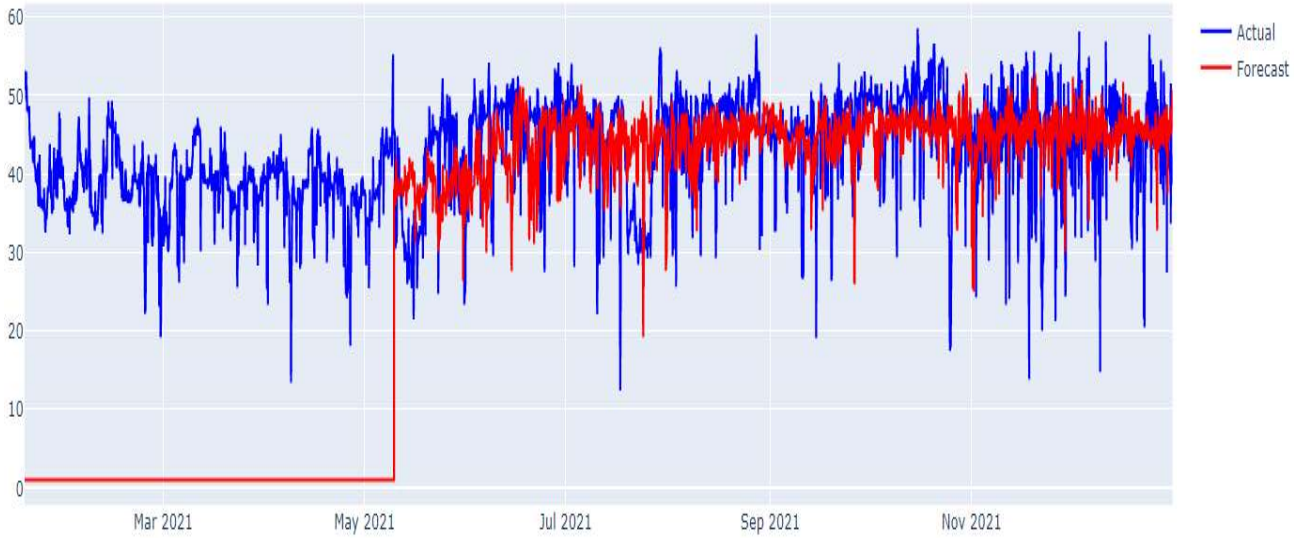


Figure 27. Sequence Graph of XGBR Estimations and Total Electric Consumption of Firm D

FIRM E	*Mean Absolute Error*
GBR	0.6761688573979254
XGBR	0.6751756112893422

Table 17. Mean Absolute Error (MAE) for Firm E

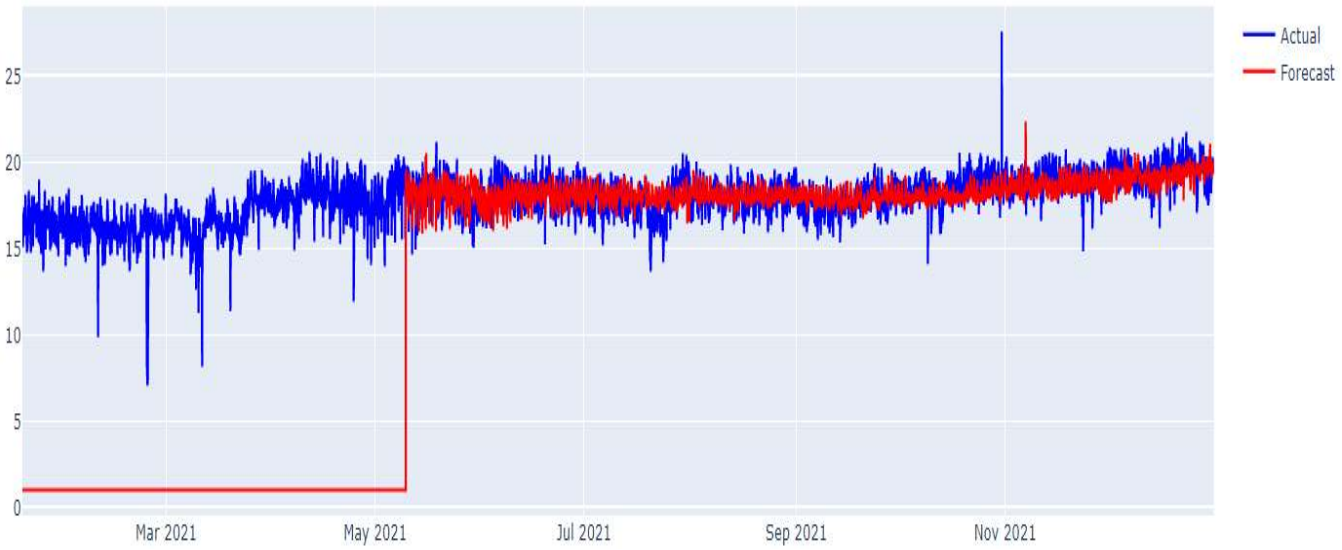


Figure 28. Sequence Graph of GBR Estimations and Total Electric Consumption of Firm E

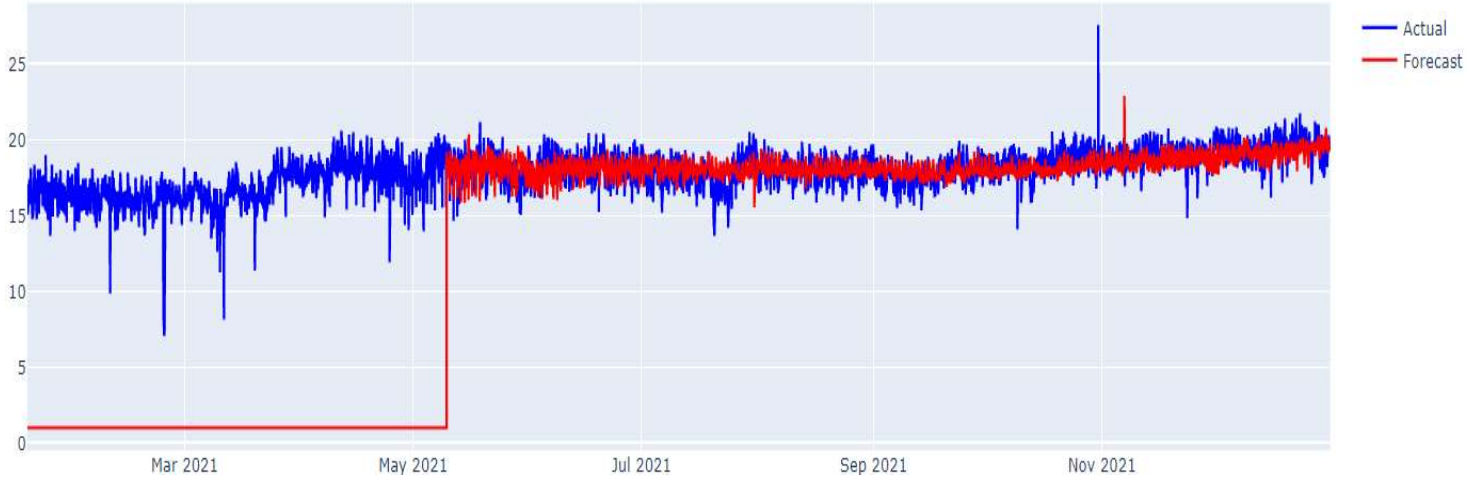


Figure 29. Sequence Graph of XGBR Estimations and Total Electric Consumption of Firm E

FIRM F	*Mean Absolute Error*
GBR	0.9078347264753566
XGBR	0.9044053499877099

Table 18. Mean Absolute Error (MAE) for Firm F

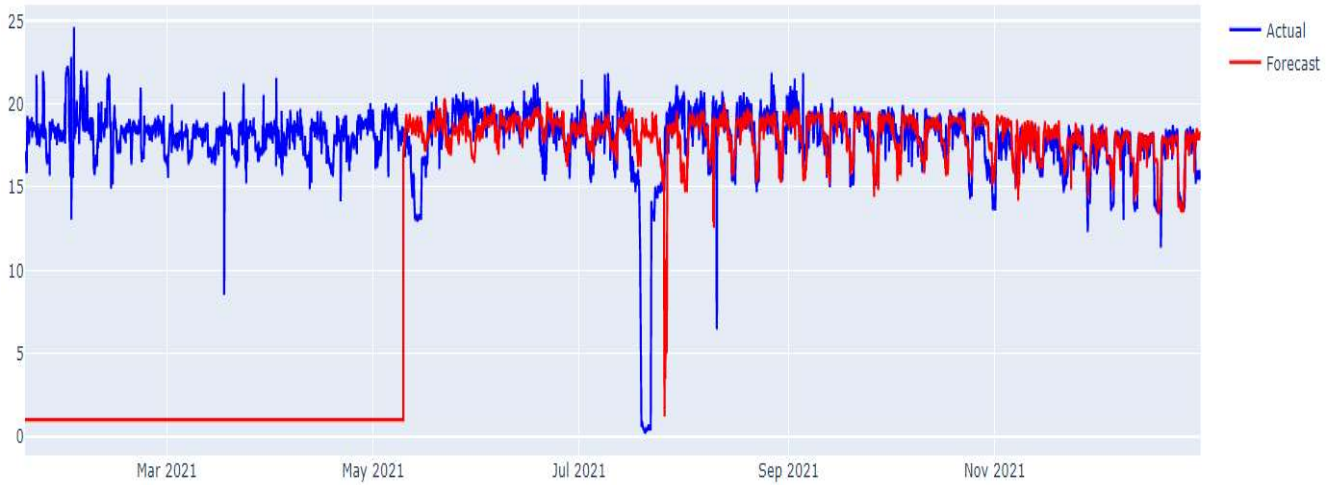


Figure 30. Sequence Graph of GBR Estimations and Total Electric Consumption of Firm F

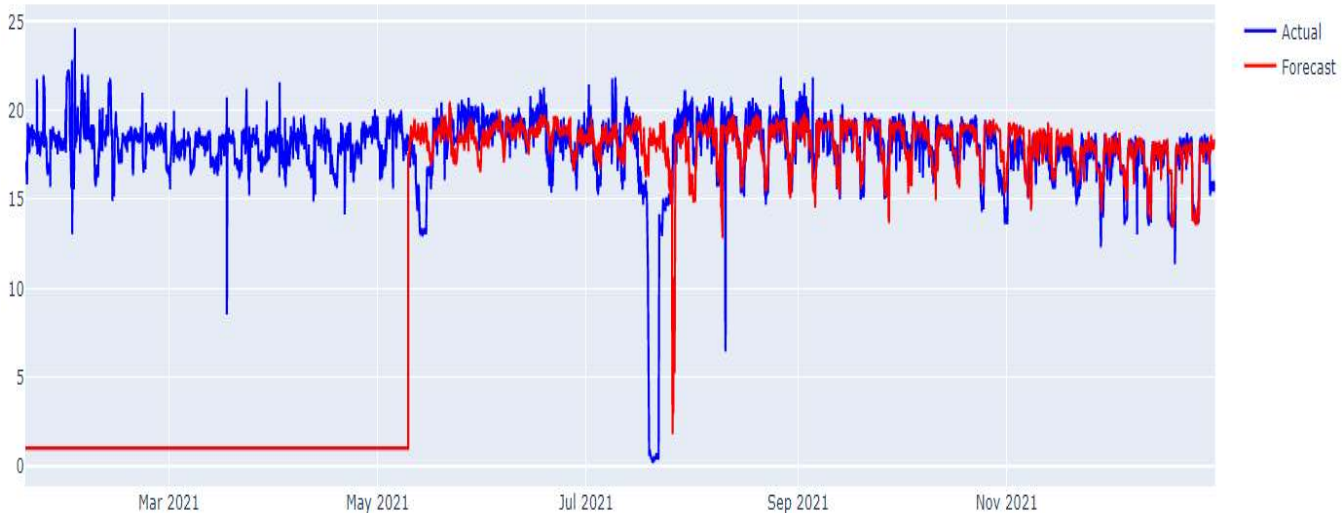


Figure 31. Sequence Graph of XGBR Estimations and Total Electric Consumption of Firm F

FIRM G	*Mean Absolute Error*
GBR	0.421076861156264
XGBR	0.4208396688886342

Table 19. Mean Absolute Error (MAE) for Firm G

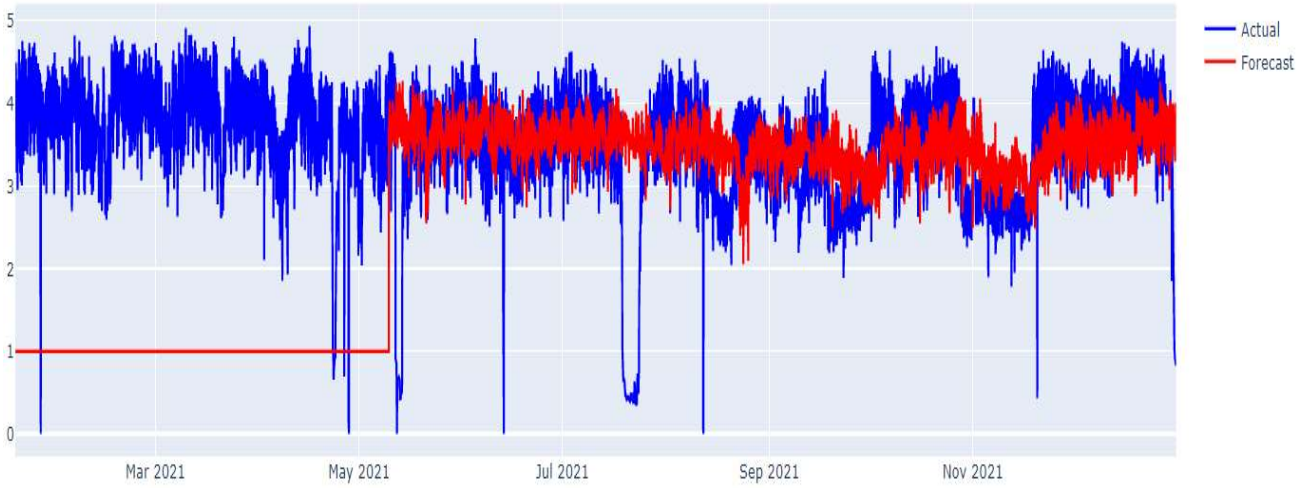


Figure 32. Sequence Graph of GBR Estimations and Total Electric Consumption of Firm G

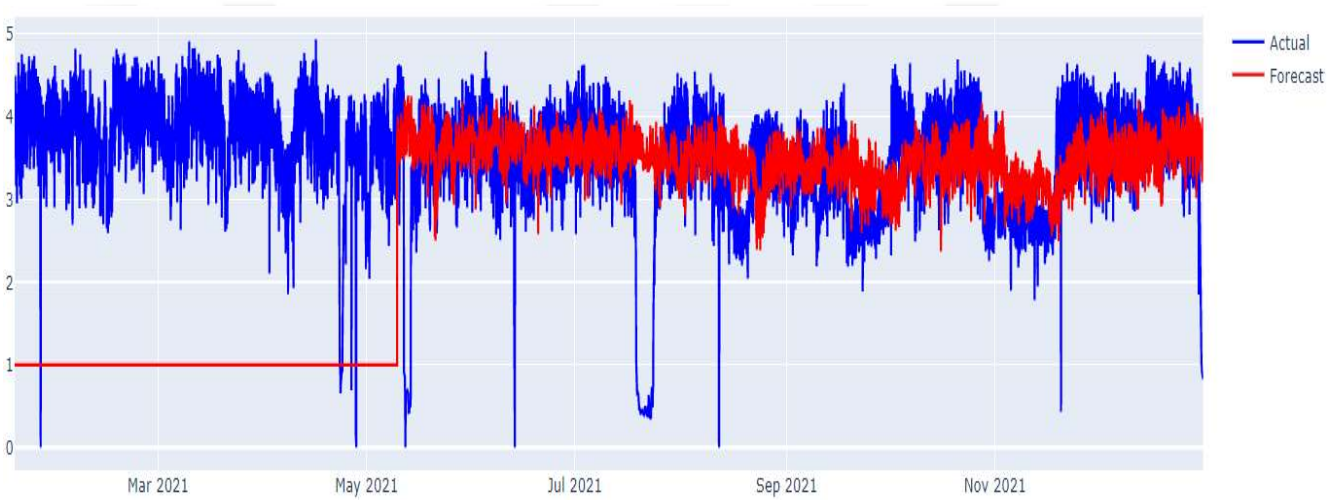


Figure 33. Sequence Graph of XGBR Estimations and Total Electric Consumption of Firm G

FIRM H	*Mean Absolute Error*
GBR	0.6687708525532885
XGBR	0.670595225500209

Table 20. Mean Absolute Error (MAE) for Firm H

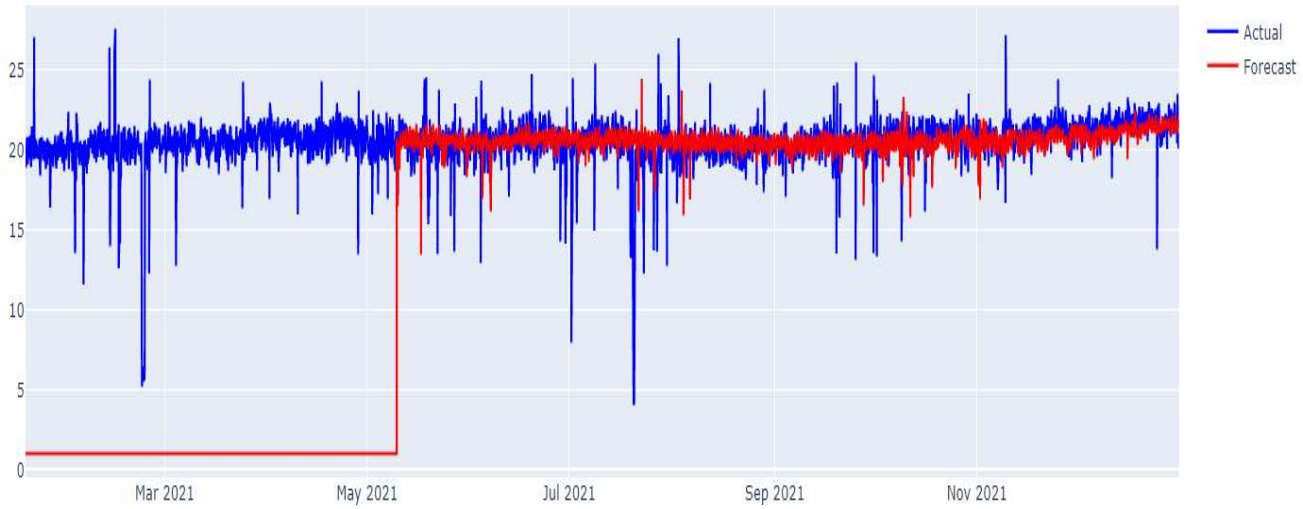


Figure 34. Sequence Graph of GBR Estimations and Total Electric Consumption of Firm H

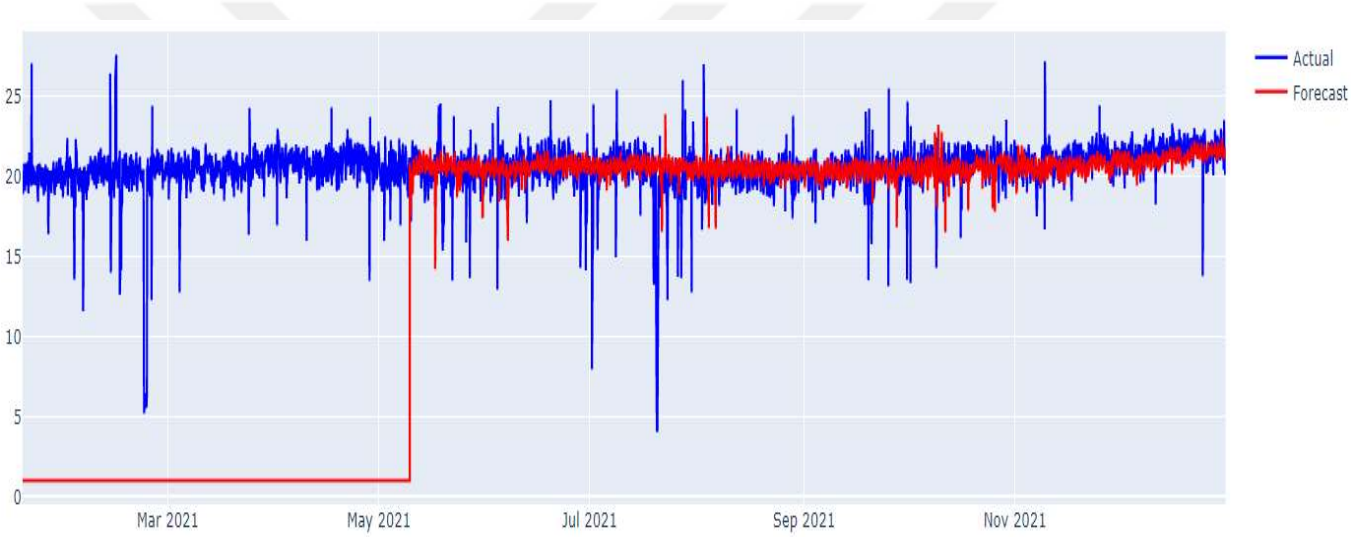


Figure 35. Sequence Graph of XGBR Estimations and Total Electric Consumption of Firm H

FIRM I	*Mean Absolute Error*
GBR	1.175271976780397
XGBR	1.171272103932358

Table 21. Mean Absolute Error (MAE) for Firm I

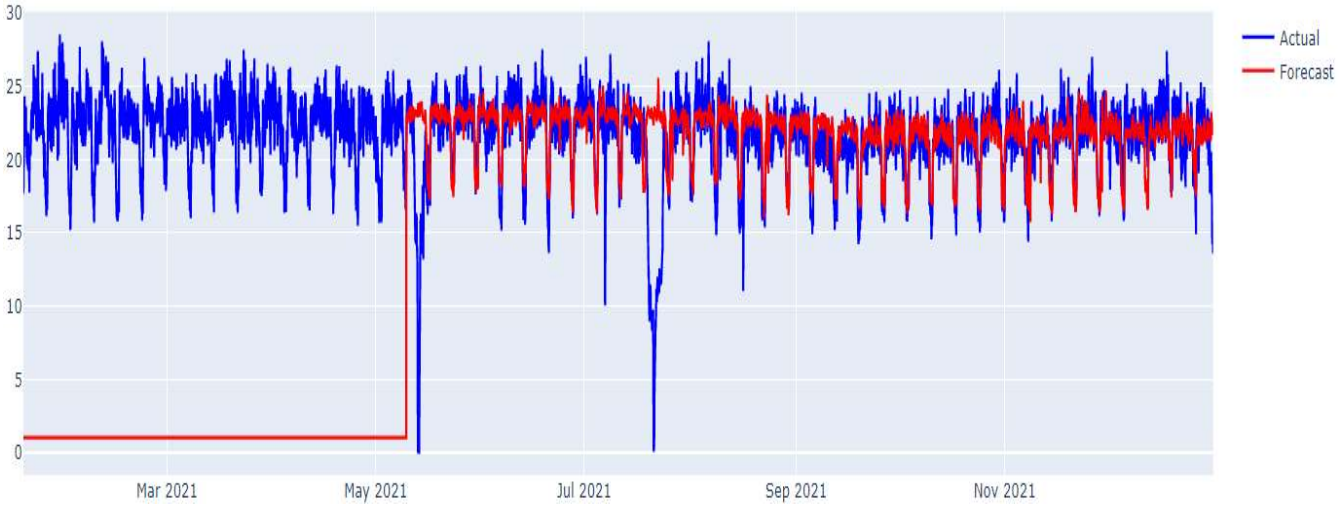


Figure 36. Sequence Graph of GBR Estimations and Total Electric Consumption of Firm I

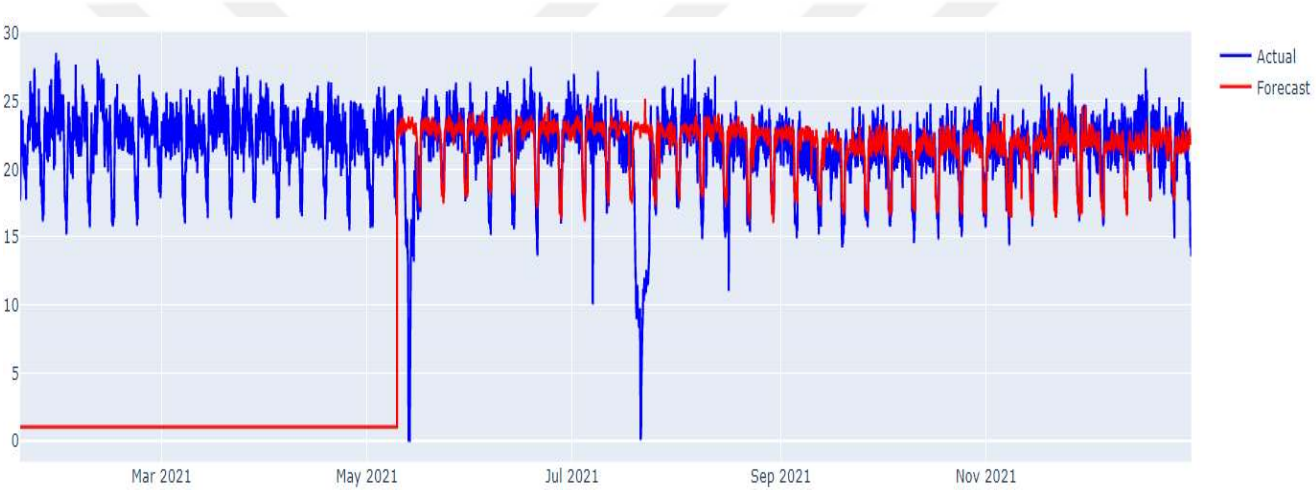


Figure 37. Sequence Graph of XGBR Estimations and Total Electric Consumption of Firm I

FIRM J	*Mean Absolute Error*
GBR	2.4798617711074322
XGBR	2.4683623024591377

Table 22. Mean Absolute Error (MAE) for Firm J

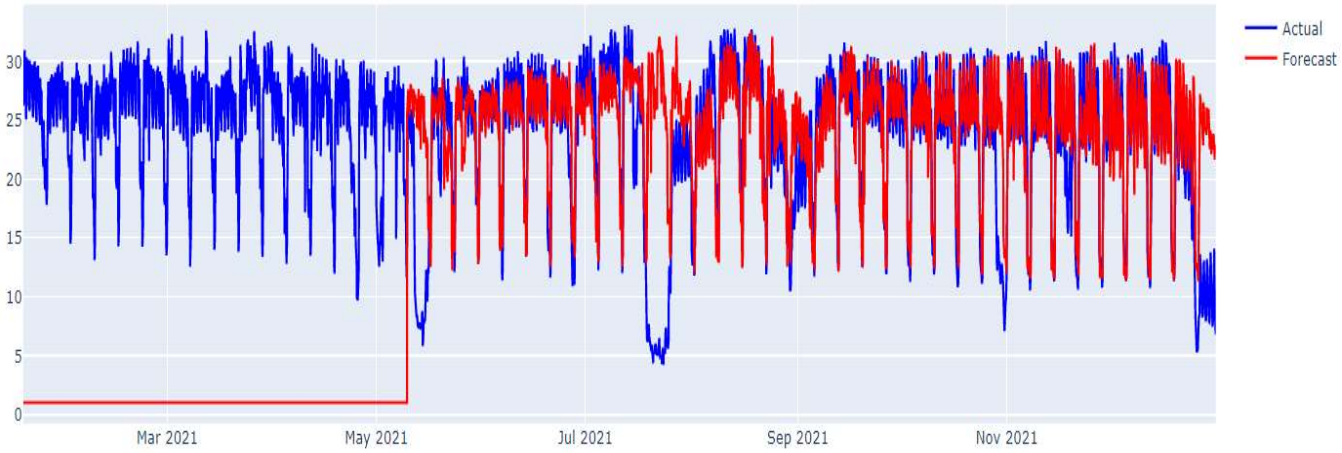


Figure 38. Sequence Graph of GBR Estimations and Total Electric Consumption of Firm J

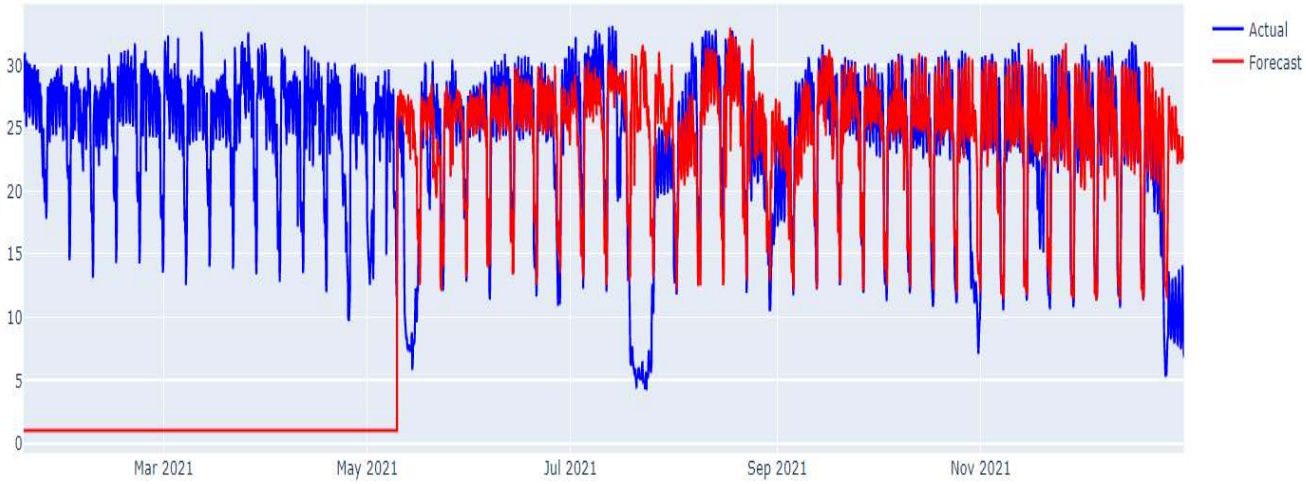


Figure 39. Sequence Graph of XGBR Estimations and Total Electric Consumption of Firm J

FIRM K	*Mean Absolute Error*
GBR	0.6103127865813966
XGBR	0.6125562043070616

Table 23. Mean Absolute Error (MAE) for Firm K

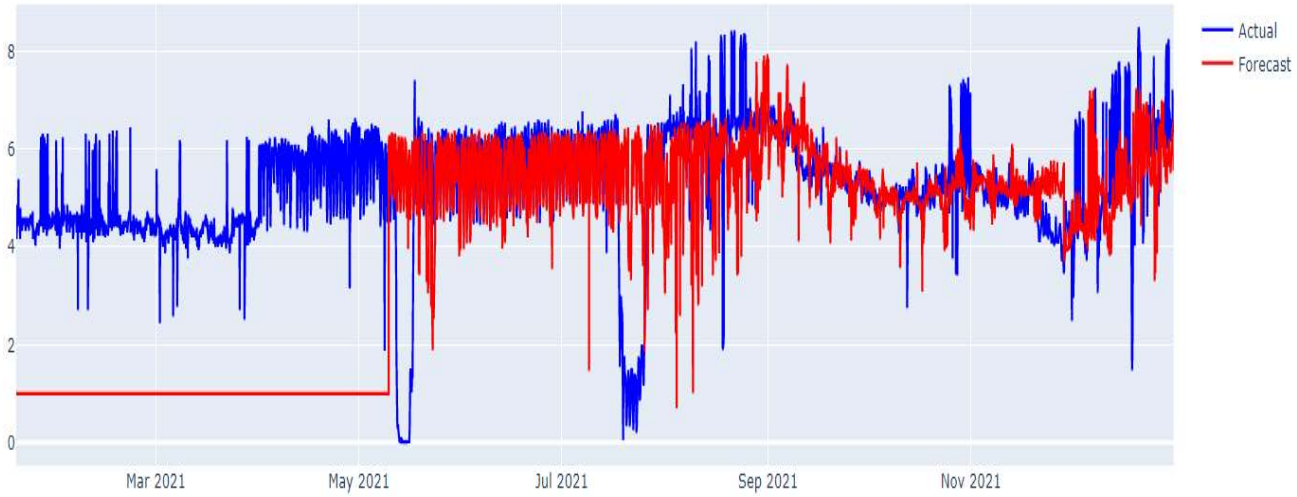


Figure 40. Sequence Graph of GBR Estimations and Total Electric Consumption of Firm K

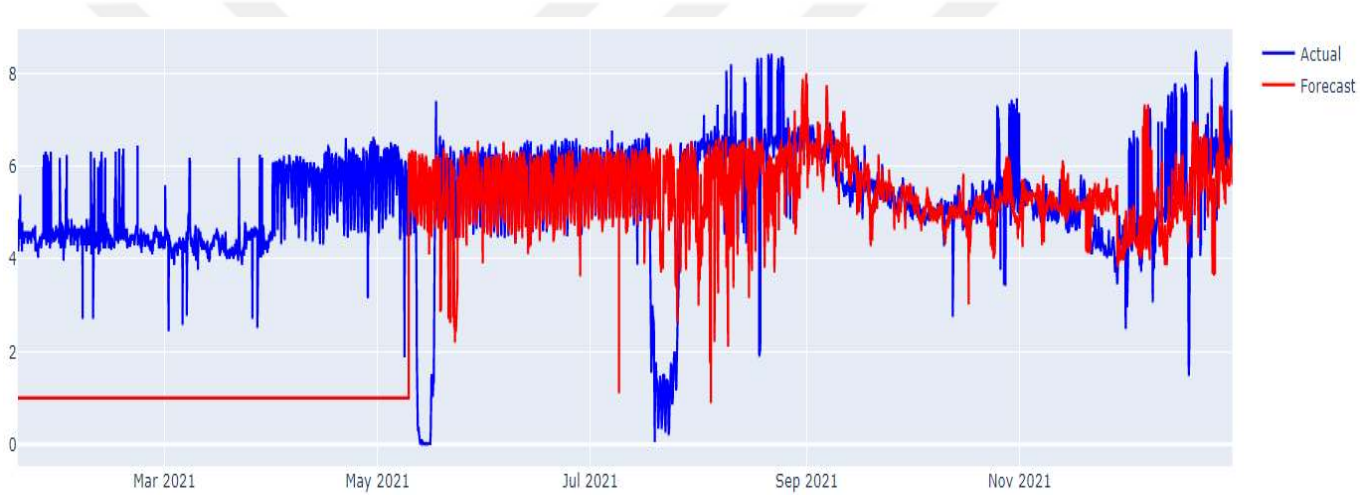


Figure 41. Sequence Graph of XGBR Estimations and Total Electric Consumption of Firm K

The MAE results given separately for all firms are summarized in following table:

FIRMS	MAE values of GBR	MAE values of XGBR
Firm A (Wire Sector)	1,26532675	1,26167266
Firm B (Ceramic Sector)	0,44388306	0,44098974
Firm C (Food Sector)	1,07197407	1,07289620
Firm D (Chemical Sector)	4,75198954	4,73655190
Firm E (Ceramic Sector)	0,67616886	0,67517561
Firm F (Textile Sector)	0,90783473	0,90440535
Firm G (Wire Sector)	0,42107686	0,42083967
Firm H (Ceramic Sector)	0,66877085	0,67059523
Firm I (Aluminium Sector)	1,17527198	1,17127210
Firm J (White Appliances Sector)	2,47986177	2,46836230
Firm K (Textile Sector)	0,61031279	0,61255620
Mean of MAE values	1,31567920	1,31230154

Table 24. MAE values of GBR and XGBR for the firms

As seen in the Table 24 above the Gradient Boosting Regression (GBR) method gave more successful results for

- Firm C (Food Sector)
- Firm H (Ceramic Sector)
- Firm K (Textile Sector)

The Extreme Gradient Boosting Regression (XGBR) method gave better results for

- Firm A (Wire Sector)
- Firm B (Ceramic Sector)
- Firm D (Chemical Sector)
- Firm E (Ceramic Sector)
- Firm F (Textile Sector)
- Firm G (Wire Sector)

- Firm I (Aluminium Sector)
- Firm J (White Appliance Sector)

In general, it can be said that XGBR performs better in estimating of electricity consumptions of firms operating in different sectors.

5. DISCUSSION AND CONCLUSION

It is known that it plays an active role in determining the direction of institutions with long or short-term forecasts for future periods by using estimation or forecasting methods and past period data, and in meeting the assets, visions and needs of customers. For this reason, it can be said that making healthy estimations or estimating uncertain future conditions is very important for the accuracy of the decisions to be taken. The companies in question need to be able to draw a sustainable and effective roadmap and to estimate the electricity consumption amounts at the best level. This depends on the success of the estimations and forecasts they make in-house or outsource. Successfully estimating the energy that will be needed in this market, where market participants share their electricity generation and consumption plans for the next day, will prevent the emergence of bottlenecks or energy surpluses. This will reflect positively on the costs of the companies.

The aim of this research is to make estimations by taking into account the previous period electricity consumption data, taking into account the researches on consumption estimations so far. The importance of the estimations here is that electricity cannot be stored in large quantities and it provides the opportunity to take action in advance according to the need, based on the future consumption estimations. Especially, ML algorithms have been started to use for electric comsumption of companies, cities, regions and/or universities in the last decade. ML algorithms based on ensemble learning are frequently used methods in these issues. GBR and XGBR are some of the popularly used ensemble learning methods, and their performance in estimation of electric consumption has been proved by many studies in the literature. Thus, gradient boosting based algorithms are used to estimate electric consuption of the firms mentioned in this study. As a result of the study, the Gradient Boosting Regression (GBR) model gave better results for 3 companies (Firm C, Firm H, Firm K). These results comfirm the results of some studies, such as Touzani et al. (2018) and Opera and Bara (2021). For 8 companies (Firm A, Firm B, Firm D, Firm E, Firm F, Firm G, Firm I, Firm J) Extreme Gradient Boosting Regression (XGBR) model gave better results similar to some other studies, such as Banik et al. (2021), Bassi et al. (2021) and Tyagi and Singh (2022). Despite the achievement of its purpose in both models, it can be said that the XGBR model is more successful because the XGBR model gives better results in 8 companies.

If we need to make suggestions based on our consumption estimation results, the consistency of the estimations should be checked by comparing the consumption estimations

made with the data realized in the following periods. Better results can be obtained by applying more techniques. It is recommended by us to continue the necessary research on different ML and AI techniques.

REFERENCES

Albuquerque, P. C., Cajueeiro, D. O., & Rossi, M. D. (2022). Machine learning models for forecasting power electricity consumption using a high dimensional dataset. *Expert Systems with Applications*, 115917.

Arf, C. (1959). Makine Düşünebilir mi ve Nasıl Düşünebilir? *Atatürk Üniversitesi 1958-1959 Öğretim Yılı.* Erzurum: Atatürk Üniversitesi.

Balaban, M. E., & Kartal, E. (2015). *Veri Madenciliği ve Makine Öğrenmesi Temel Algoritmaları ve R Dili ile Uygulamaları.* Çağlayan Kitabevi.

Banik, R., Das, P., Ray, S., & Biswas, A. (2021). Prediction of electrical energy consumption based on machine learning technique. *Electrical Engineering*, 909-920.

Bassi, A., Shenoy, A., Sharma, A., Sigurdson, H., Glossop, C., & Chan, J. H. (2021). Building Energy Consumption Forecasting: A Comparison of Gradient Boosting Models. *IAIT2021: The 12th International Conference on Advances in Information Technology* (s. 1-9). New York: Association for Computing Machinery.

Boehmke, B., & Greenwell, B. (2020). *Hands-On Machine Learning with R.* CRC Press.

Burger, E. M., & Moura, S. J. (2015). Gated ensemble learning method for demand-side electricity load forecasting. *Energy and Buildings*, 23-34.

Chen, K., Jiang, J., Zheng, F., & Chen, K. (2018). A novel data-driven approach for residential electricity consumption prediction based on ensemble learning. *Energy*, 49-60.

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (s. 785-794). New York: Association for Computing Machinery.

Chen, X., Wang, Z.-X., & Pan, X.-M. (2019). HIV-1 tropism prediction by the XGboost and HMM methods. *Scientific Reports*. doi:10.1038/s41598-019-46420-4

Divina, F., Gilson, A., Gomez-Vela, F., Torres, M. G., & F.Torres, J. (2018). Stacking Ensemble Learning for Short-Term Electricity Consumption Forecasting. *Energies, 11*(4), 11(4). doi:10.3390/en11040949

Elmas, Ç. (2018). *YApay Zeka Uygulamaları.* Ankara: Seçkin Yayıncılık.

Fan, J., Wang, X., Wu, L., Zhou, H., Zhang, F., Yu, X., . . . Xiang, Y. (2018). Comparison of Support Vector Machine and Extreme Gradient Boosting for. *Energy Conversion and Management*, 102-111.

Freidman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 367-378.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 1189-1232.

Friedman, J., Hastie, T., & Tibshirani, R. (2000). Additive logistic regression: a statistical view of boosting. *Annals of Statistics*, 337-374.

Ghatak, A. (2017). *Machine Learning with R.* Singapore: Springer.

Gürsakal, N. (2018). *Makine Öğrenmesi.* Dora Yayınları.

Igual, L., & Segui, S. (2017). *Introduction to Data SCience A Python Approach to Concepts, Techniques and Applications.* Switzerland: Springer.

James, G., Witten, D., Hastie, T., & Tibshrani, R. (2017). *An Introduction to Statistical Learning with Applications R.* New York: Springer.

Khan, P. W., & Byun, Y.-C. (2020). Genetic Algorithm Based Optimized Feature Engineering and Hybrid Machine Learning for Effective Energy Consumption Prediction. *IEEE Access, 8*, 196274-196286.

Kubat, M. (2017). *An Introduction To Machine Learning (Second Edition).* Switzerland: Springer.

Liashchynskyi, P., & Liashchynskyi, P. (2019). Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS. *arXiv preprint arXiv:1912.06059*.

Mariano-Hernandez, D., Callejo, L. H., Solis, M., A. Zorita-Lamadrid, O. D.-P., Gonzalez-Morales, L., Alonso-Gomez, V., . . . Garcia, F. S. (2022). Comparative study of continuous hourly energy consumption forecasting strategies with small data sets to support demand management decisions in buildings. *Energy Science & Engineering* , 4694-4707.

Mcafee, A., & Brynjolfsson, E. (2017). *Harnessing Our Digital Future, Machine Platform Crowd.* Norton & Company.

Moon, J., Park, S., Rho, S., & Hwang, E. (2022). Robust building energy consumption forecasting using an online learning approach with R ranger. *Journal of Building Engineering, 47*, 103851.

Nie, P., Roccotelli, M., Fanti, M. P., Ming, Z., & Li, Z. (2021). Prediction of home energy consumption based on gradient boosting regression tree. *Energy Reports*, 1246-1255.

Opera, S.-V., & Bara, A. (2021). Machine learning classification algorithms and anomaly detection in conventional meters and Tunisian electricity consumption large datasets. *Computers & Electrical Engineering, 94*, 107329.

Pinto, T., Praça, I., Vale, Z., & Silva, J. (2021). Ensemble learning for electricity consumption forecasting in office buildings. *Neurocomputing*, 747-755.

Pontius, R. G., Thontteh, O. E., & Chen, H. (2008). Components of information for multiple resolution comparison between maps that share a real variable. *Environmental and Ecological Statistics* , 111-142.

Porteiro, R., Hernandez-Callejo, L., & Nesmachnow, S. (2022). Electricity demand forecasting in industrialand residential facilities using ensemblemachine learning. *Revista Facultad de Ingeniería, Universidad de Antioquia*, 9-25.

Ramasubramanian, K., & Singh, A. (2019). *Machine Learning Using R With Time Series and Industry-Based Use Cases in R (Second Edition).* New York: APress.

Reddick, J., Genov, E., Camargo, L. R., Coosemans, T., & Messagie, M. (2022). Evolutionary Scheduling of University Activities Based on Consumption Forecasts to Minimise Electricity Costs. *2022 IEEE Congress on Evolutionary Computation (CEC)* (s. 1-8). Padua/Italy: IEEE.

Ridgeway, G. (2005). *Generalized Boosted Models: A Guide to the GBM Package.*

Sapon, M. A., Ismail, K., Zainudin, S., & Ping, C. S. (2011). Diabetes Prediction with Supervised Learning Algorithms of Artifical Neural Network. *2011 International Conference on Software and Computer Applications. 9.* Singapore: IACSIT Press.

Sepulveda, L. F., Diniz, P. S., Diniz, J. O., Netto, S. M., Cipriano, C. L., Araujo, A. C., . . . C.Paiva, A. (2021). Forecasting of individual electricity consumption using Optimized Gradient Boosting Regression with Modified Particle Swarm Optimization. *Engineering Applications of Artificial Intelligence, 105*, 104440.

Shin, S.-Y., & Woo, H.-G. (2022). Energy Consumption Forecasting in Korea Using Machine Learning Algorithms. *Energies, 15*(13). doi:doi.org/10.3390/en15134880

Sujan Reddy, A., Akashdeep, S., Harshvardhan, R., & Sowmya Kamath, S. (2022). Stacking Deep learning and Machine learning models for short-term energy consumption forecasting. *Advanced Engineering Informatics, 52*, 101542.

Şen, D., Tunç, K. M., & Günay, M. E. (2021). Forecasting electricity consumption of OECD countries: A global machine learning modeling approach. *Utilities Policy, 70*, 101222.

Touzani, S., Granderson, J., & Fernandes, S. (2018). Gradient boosting machine for modeling the energy consumption of commercial buildings. *Energy and Buildings, 158*, 1533-1543.

Tyagi, S., & Singh, P. (2022). Short Term and Long term Building Electricity Consumption Prediction Using Extreme Gradient Boosting. *Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science), 15*(8), 1082-1095.

Williams, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research, 30*, 79-82.

Yıldırım, E. (2022, Ağustos). Hızlandırılmış Makine Öğrenmesi Algoritmaları ile Türkçe Sahte Haber Tespiti. Karabük, Türkiye: Karabük Üniversitesi.