UNIVERSITY OF ENERGY AND NATURAL RESOURCES, SUNYANI

SCHOOL OF ENGINEERING

DEPARTMENT OF COMPUTER AND ELECTRICAL ENGINEERING

BSC. COMPUTER ENGINEERING

REPORT ON ELECTRICAL LOAD FORECASTING USING MACHINE LEARNING

PRESENTED BY:

DAMBA ERIC	**(UEB0904119)**
GYIMAH GIDEON	**(UEB0913419)**
ANTWI AUGUSTINE	**(UEB0902619)**

SUPERVISOR: ING. ISAAC OTCHERE

REPORT ON ELECTRICAL LOAD FORECASTING USING MACHINE LEARNING

A Thesis Presented to the University of Energy and Natural Resources in Partial

Fulfillment of the Requirements for the Degree of
Bachelor of Science in Electrical/Electronic
Engineering

SEPTEMBER, 2023

AUTHORS' DECLARATION

We understand that copyright in our thesis is transferred to University of Energy and Natural Resources.

DAMBA ERIC ____________________ ____________________

GYIMAH GIDEON ____________________ ____________________

ANTWI AUGUSTINE ____________________ ____________________

RESEARCH SUPERVISOR

SUPERVISOR: ING. ISAAC OTCHERE ____________________ ____________________

,

DECLARATION

We, the undersigned members of the research team, declare that the work presented in this project thesis, titled "[Electrical Load Forecasting Using Machine]," is the result of our collective efforts and is original, unless otherwise acknowledged. We have appropriately cited all sources and references used in this thesis. The data and information used in this project were obtained from legitimate sources, publicly available datasets, and academic publications. We have provided proper attribution to these sources, and any direct quotations or paraphrased content have been cited accordingly. This thesis has not been submitted for any other academic award, degree, or qualification at any other institution. It does not infringe upon the intellectual property or copyright of any third party. We take full responsibility for the content and conclusions presented in this thesis. Any opinions, findings, or recommendations expressed herein are our own and do not necessarily reflect the views of University of Energy and Natural Resources or any other organization. We acknowledge and adhere the academic integrity policies and regulations of University of Energy and Natural Resources.

DEDICATION

This project thesis is dedicated to those who have been a constant source of inspiration and support throughout this journey. To our families, whose unwavering encouragement and understanding made it possible for us to pursue our academic goals. Your love and patience have been our rock.

To our supervisor, Ing. Isaac Otchere who guided us with wisdom and expertise, nurturing our intellectual growth and shaping our perspectives. To our friends and colleagues, for the countless hours of discussion, collaboration, and shared experiences that enriched our learning and made this project a reality. And to all those who believe in the power of knowledge and innovation to make a positive impact on our world, this work is dedicated to you.

Your belief in us has been the driving force behind our efforts, and for that, we are profoundly grateful.

ACKNOWLEGEMENT

We, the members of the research team behind this project thesis on "Electrical Load Forecasting Using Machine Learning," wish to express our sincere gratitude to all those who contributed to the successful completion of this endeavor.

First and foremost, we extend our heartfelt appreciation to our academic advisor, Ing. Isaac Otchere, for their invaluable guidance, mentorship, and unwavering support throughout this research journey. Their expertise and feedback significantly shaped the direction of this work.

We are also thankful to the faculty members of the Computer and Electrical Department at University of Energy and Natural Resources for their wisdom and the knowledge they shared, which enriched our academic experience.

We wish to acknowledge the dedication and collaborative spirit of our fellow research team members. Our collective efforts, shared insights, and teamwork were fundamental to achieving the goals of this project.

Special thanks go to Ing. Isaac Otchere who provided technical expertise and played a vital role in data collection and analysis.

Furthermore, we express our deep gratitude to our families for their continuous support, encouragement, and understanding throughout our academic pursuits.

Lastly, we appreciate the cooperation of research participants and all individuals, as well as organizations, who contributed data and resources crucial to this project.

This work stands as a testament to the collective efforts, support, and contributions of all these individuals and entities. We are profoundly thankful for their indispensable role in the completion of this project.

ABSTRACT

Effective electrical load forecasting is paramount for operational efficiency, especially in energyintensive industries such as aluminum production. Accurate forecasting models aid in optimizing energy consumption, thus, contributing to economic operations. This project aims to develop and compare multiple machine learning algorithms for predicting the electrical load demand of the Volta Aluminium Company in Ghana. We employed various algorithms including Linear Regression, Ridge, Decision Tree, Lasso, Random Tree, Gradient Boosting, AdaBoost, Support Vector Machine, XGBoost, LightGBM, and CatBoost to forecast the daily load demand. A historical dataset comprising load demand, weather conditions, and operational hours was used to train and evaluate the performance of the developed models. The best algorithm was fine-tuned to achieve the best predictive accuracy. The models' performances were assessed using metrics such as Mean Absolute Error (MAE) and Coefficient of Determination (R^2). Our findings reveal that ensemble methods like Gradient Boosting and XGBoost performed exceptionally well in capturing the non-linear relationship between the features and the load demand. This comparative analysis provides insights into the strengths and weaknesses of each algorithm, and sheds light on the potential of machine learning in enhancing load forecasting accuracy for industrial setups in emerging economies like Ghana. Our work paves the way for developing robust load forecasting systems, essential for strategic energy management and sustainability in the aluminum production sector.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABREVIATIONS

LF- Load Forecasting

SVMs- Support Vector Machines.

GPR- Gaussian Process Regression

KNN- K-Nearest Neighbors

RMSE- Root Mean Square Error

MSE- Mean Square Error

MAE- Mean Absolute Error

LTF- Long Term Forecasting

MTF- Medium Term Forecasting

STF- Short Term Forecasting ELD-

Economic Load Dispatch ML-

Machine Learning.

SLT- Statistical Learning Theory. GA-

Genetic Algorithms

ANN- Artificial Neural Networks

CNN- Convulsion Neural Networks

RNNs- Recurrent Neural Networks

LSTM- Long-Short Term Memory Unit

PSO- Particle Swarm Optimization

ERNN- Efficient Recurrent Neural Network.

DNN- Deep Neural Network.

XGBoost- eXtreme Gradient Boosting.

VALCO-Volta Aluminium Company Limited

CNN- Convolutional Neural Network

LSTM – Long Short-Term Memory

CHAPTER I

1.0 INTRODUCTION

1.1 Overview / Introduction

As the population and industrializing are increasing, the corresponding energy demand is increasing year by year exponentially. Electricity is the central nervous system of our modern existence commodities. It is also critical to any country's national security, social and economic prosperity. As a result, the security and sustainability of energy sources, as well as the continuity of generation, are critical problems, and electrical load forecasting may be quite useful in this regard. Load forecasting (LF) is the anticipation and prediction of demand load that is computed using a systematic approach to alter future expectations based on existing characteristics and information to anticipate future demands and system requirements. Precise load prediction or forecasting can help energy and power providers ensure the security and continuity of the power supply with no (or minor) disruptions. In addition, because electrical power is difficult to store, scheduling and limiting energy waste are important. Electrical load forecasting may be used to manage generation capacity, schedule maintenance and outages, grid and transmission management, peak reduction, and reserve capacity management. Furthermore, it is critical for market evaluation and assessment of power trade capacity and interconnection capabilities, as well as preparing reserve to overcome any power deficits for any VALCO, or the Volta Aluminium Company Limited, holds a storied place in Ghana's industrial history. Established in the early 1960s, VALCO emerged as a symbol of the nation's aspirations for industrialization and economic diversification. Founded through a partnership between the Ghanaian government and American aluminum giants Kaiser Aluminum and Reynolds Metals Company. The Volta Aluminum Company (VALCO) in Ghana relies on hydroelectric power from the Akosombo Dam for its electricity. This dam, which is located on the Volta River, is the largest hydroelectric power plant in West Africa and one of the largest in Africa. The Akosombo Dam supplies VALCO with about 80% of its power needs. The remaining 20% of its power needs are met by the Aboadze Thermal Power Plant, which is located in Takoradi. Together, these two power sources ensure that VALCO has a reliable and sustainable source of energy

For decades, VALCO played a pivotal role in Ghana's economy, generating substantial revenue and providing employment opportunities for local communities. However, this journey has been marked by intermittent challenges and periodic shutdowns.

One of the primary reasons for these shutdowns was the company's heavy reliance on electricity, an energy-intensive necessity for aluminum smelting. Ghana's energy supply infrastructure faced periodic disruptions, leading to production halts at VALCO. Additionally, the global aluminum market's price volatility posed economic challenges, prompting temporary closures during periods of unfavorable pricing. VALCO also dealt with environmental pressures, ownership changes, and financial complexities.

Despite these challenges, VALCO remains an enduring symbol of Ghana's industrial ambition and resilience, reflecting both the opportunities and obstacles encountered on the path to aluminum production and economic development

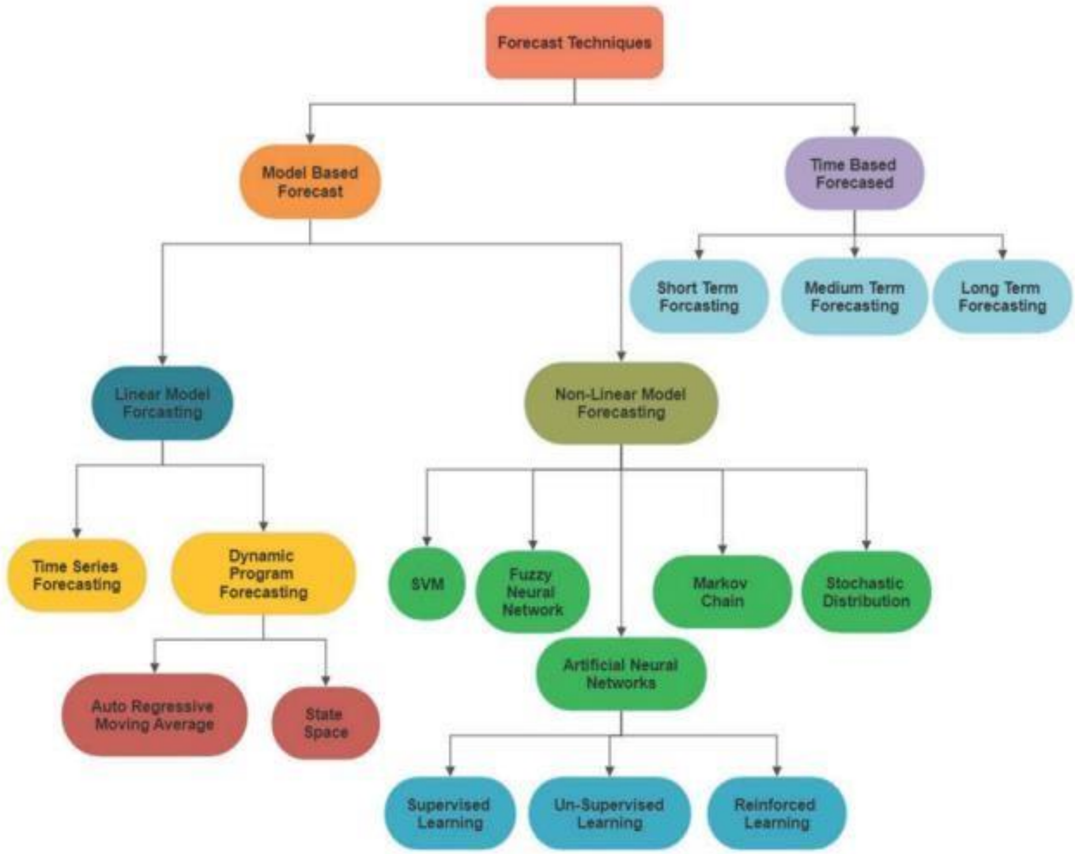
Consequently, gaining precise insights into VALCO's energy requirements is essential to safeguard against power interruptions within Ghana. Our proposal to undertake a comprehensive load forecasting study for VALCO exemplifies a proactive strategy aimed at securing a dependable and consistent energy supply. This not only ensures the stability of industrial operations but also fosters sustainable economic growth for the nation linked utilities.

Machine Learning in Power System

Electrical load is classified into two types: base load and peak load. The baseload is nearly constant, but the peak load varies naturally. The electrical grid's stability may be improved by forecasting peak load using machine learning techniques. As a result, the study is based on machine learning and better comprehending Existing approaches are discussed in light of the model. Linear and non-linear classes are two sorts of classes. Linear models are straightforward models. It uses the least-squares approximation approach to find the best line that adequately describes the data. Least-squares approach is a frequently used linear class strategy. There might be a when data points are thinly separated, there might be significant inaccuracy. Nonlinear models can be used to improve accuracy and reduce errors. Because the electrical demand is mostly highly nonlinear, this technique may not be appropriate. Non-linear models are complicated, timeconsuming, and need a large number of computations. However, the results provided are superior to linear models.

The Prediction variable and the Response variable are the two most utilized terminology in this study. They need prediction variables, which are traits or factors on which the response variable is dependent. The prediction variable's value is used to forecast the response variable's value. Non-linear models are utilized when there is non-linear dependency between prediction variables. Non-linear models include Support Vector Machines (SVMs), Gaussian Process Regression (GPR), and others. SVMs combined with ant colony (AC) optimization seem to be an accurate load prediction tool. Non-linear models are utilized where there is non-linear dependency between prediction variables. Nonlinear models include Support Vector Machines (SVMs), Gaussian Process Regression (GPR), and others. SVMs combined with ant colony (AC) optimization seem to be an accurate strategy for load.

Prediction is a difficult task. Correct prediction necessitates correct information and elements on which the forecast will be based. When all factors are included, the amount of the data grows, as do the risks of data over-fitting. The characteristics must be chosen in such a way that they are not redundant. The load is projected using past load data. As a result, a large amount of historical load data from the location will be included for reliable load forecast. As illustrated in the diagram, prediction approaches are categorized.



1.2 Aim

The main aim of this project is to develop a machine learning model to accurately predict future electricity / power demand, enabling efficient resource planning, cost optimization, grid stability, demand response management, renewable energy integration, maintenance scheduling, improved customer service and effectively capturing the impact of dynamic factors on our model building.

1.3 Objectives

- Accurate load prediction: The primary objective is to develop machine learning models that can accurately forecast future electrical load demand. The models should capture the underlying patterns and dependencies in the historical load data to make precise predictions**.**
- Improved operational efficiency: Utility companies and grid operators optimize resource allocation, plan maintenance schedules, and manage electricity generation and distribution more efficiently. By accurately forecasting load demand, they can make informed decisions to balance supply and demand, minimize energy wastage, and reduce operational costs.
- Enhanced grid reliability: It plays a crucial role in maintaining the stability and reliability of the electrical grid. By anticipating peak load periods or sudden load fluctuations, grid operators can take proactive measures such as load shedding, load shifting, or deploying additional resources to ensure uninterrupted power supply and prevent system failures.
- Demand-side management: Design a model to enable better demand-side management by providing insights into consumer behavior and load patterns. It helps utilities and energy providers develop demand response programs, incentivize energy conservation during peak periods, and encourage load shifting strategies to balance electricity consumption.
- Planning and investment decisions: Assists in long-term planning and investment decisions related to infrastructure development. By predicting future load growth accurately, utilities and policymakers can plan for capacity expansions, grid reinforcements, and optimize investments in new generation facilities or transmission infrastructure.
- Customer satisfaction: Allowing utilities to provide better customer service. By accurately predicting load demand, utilities can anticipate and mitigate potential outages, minimize service disruptions, and ensure a stable and reliable power supply, leading to higher customer satisfaction.

- Energy efficiency and conservation: To promote energy efficiency and conservation initiatives by identifying energy usage patterns and highlighting opportunities for optimization. It enables consumers to make informed decisions about energy consumption, encourages load reduction during peak hours, and supports the implementation of energy-saving measures.
- Deployment: Deployment is the method by which you integrate a machine learning model into an existing production environment to make practical business decisions based on data. In order to start using our model for practical decision-making, it needs to be effectively deployed into production using streamlit.

1.4 Problem statement

The electrical demand in sub-Saharan/. West Africa is increasing due to rapid growth in population, which results in the highest increase in electrical demand worldwide. VALCO, a vital component of Ghana's industrial sector, faces recurrent operational disruptions due to electricity supply issues. These interruptions pose a substantial threat not only to VALCO's productivity but also to the nation's broader goals of achieving sustainable industrialization. The challenge at hand is to develop a reliable and adaptive energy management solution that not only addresses VALCO's immediate concerns but also serves as a blueprint for ensuring uninterrupted energy supply to support industrial growth across Ghana.

Therefore, a good predictive model trained with all possible parameters to handle future load demands would be needed.

1.5 Scope of work

In this project, we successfully predicted electric demand loads at specific hours of the day based on various prediction variables. This was framed as a regression problem, where we aimed to predict a response variable using the values of prediction variables and their relationship with the response variable. Our project encompassed various regression scenarios: Linear Regression:

Predicting the response variable with a single prediction variable. Multiple Regression: Predicting the response variable using multiple prediction variables. Multivariate Regression: Handling many response and prediction variables.

Our choice of the best model was driven by extensive modeling and simulation results. We explored various regression algorithms, including Support Vector Machines (SVMs), K-Nearest Neighbors (KNN), "Extreme Gradient Boosting (XGBoost), Neural Networks, Gaussian Process Regression (GPR), among others. Each algorithm had its strengths and weaknesses, and we made a data-driven selection based on factors such as computation time, complexity, and hyperparameter requirements.

Ultimately, we adopted a Short-Term Load Forecasting Technique, blending both Linear and NonLinear models for comparison. To assess model performance, we used essential metrics such as Root Mean Square Error (RMSE)/Mean Square Error (MSE) and Mean Absolute Error (MAE).

Once our selected model (XGBoost) was fine-tuned and validated, we seamlessly deployed it for practical use. Deployment involved integrating the machine learning model into our existing production environment, enabling us to make informed business decisions based on real-time data. This ensured that practical insights were readily available for decision-makers. We achieved this by deploying the model within the Grid production environment and implementing Streamlit for user-friendly access, facilitating practical decision-making based on our machine learning predictions.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview

Electrical load forecasting may be used to manage generation capacity, schedule maintenance and outages, grid and transmission management, peak reduction, and reserve capacity. Furthermore, it is critical for market evaluation and assessment of power trade capacity and interconnection capabilities, as well as preparing reserve to overcome any power deficits for any linked utilities [1], [2]. Electrical load forecasting may be divided into three categories:

1. Long term (1-20 years)

2. Medium term (1 week – 1 year)

3. Short term (1 hour – 1 week)

Each of these categories can have different methods of compiling as well as different benefits and different constrains. Therefore, would require different machine learning methods in order to obtain the forecast. The different benefits for each category are organized in Table 1.

Time scope	Application
LTF (1-20 yrs)	Planning for the power system involves load forecasting, network planning, and generation planning.
MTF (1w-1y)	arranging repairs and maintenance
STF (1min-1w)	Analysis of unit commitment
	Economic load dispatch (ELD) and OPF analysis
	Automatic generation control
Milliseconds to seconds	Power system dynamic analysis
Nanoseconds to micro seconds	Power system transient analysis

Table 2.1 time based applications of load forecasting

Long term forecasting (LTF), as detailed in Table 1, is critical to the economic planning of forthcoming future additions to the generating system, as well as extra transmission planning. Other applications of medium-term forecasting (MTF) include tariff setting, maintenance and repair scheduling, fuel supply scheduling, and financial management. Meanwhile, short term forecasting (STF) is used to provide the essential data for scheduling generator start-up and shutdown, preparing spinning reserves, and conducting a detailed examination of transmission limitations. The STF is also used to assess system and ELD security [1]. Forecasting is timedependent; the longer the timeframe, the less accurate the forecasting.

2.2 Key theories or concepts from the previous work and their methods

Their methods are;

A. Classical methods

1. Qualitative method
2. Quantitative method

B. Machine Learning methods

1. Artificial Neural Networks
2. Support Vector Machine
3. Fuzzy Logic
4. Genetic Algorithms

The below are the explanations of their methods and concepts

A. Classical Methods

Classical methods approach load forecasting in several ways and as mentioned in [3] – [6] could be divided in 2 categories which have been listed in Figure 1.

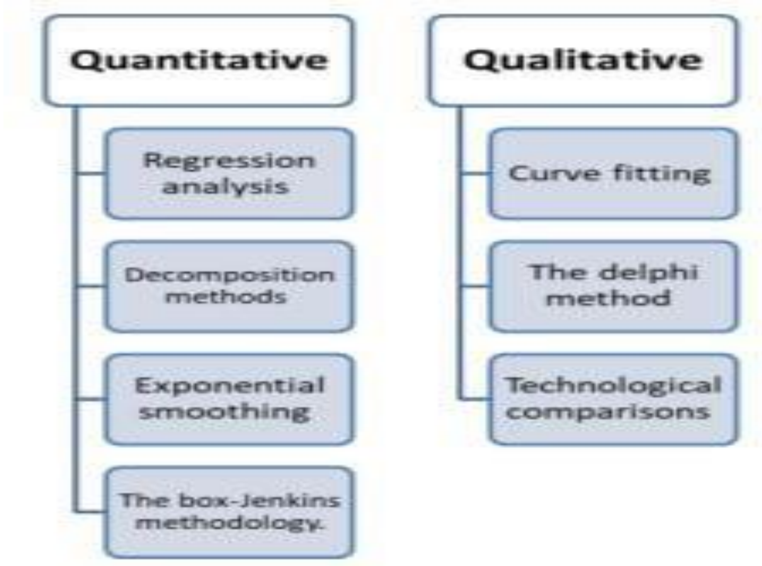


Figure 2.1 the base categories of classical forecasting methods

Figure 1 depicts the two primary groupings for classical approaches; selecting between these categories is primarily a matter of preference relies on the information and data provided. In qualitative approaches, no historical data are utilized to create a prediction; instead, the expert opinion in a structurally defined methodology is used to anticipate loads, some of which are depicted in figure 1.

On the other hand, quantitative techniques are based on objective forecasting methods that use mathematical and statistical approaches on existing historical data while assuming the continuation of specific numerical previous attributes (such as temperature). Quantitative approaches abound, and each has a unique set of costs, faults, and characteristics; which one to employ relies on data availability, desired application, and output, among other factors [3]. Many new methods and processes have recently been researched and published to address the inadequacies of some of these preceding LF techniques, as many of the characteristics required by these approaches are difficult to obtain

B. Methods of Machine Learning

Classical methodologies have been obsolete in current times owing to their limits, and machine learning has arisen as a subject, evolved in application diversity, and become well recognized in the power analysis and forecasting sector for its numerous virtues. Machine Learning is a subfield of artificial intelligence that focuses on using data and algorithmic principles to gradually improve the accuracy of machines and models through pattern analysis and recognition, almost replicating the human learning process.

Machine learning methods are numerous, each of them serving a certain purpose or better fitting for certain circumstances, input features, and type of data. Theses machine learning approaches include the following:

1) ANN (Artificial Neural Networks)

It is a model architecture that attempts to mimic the way the human brain processes information. This seeks to intelligently interpret, recognize, and categorize data. Data is pumped into neural network layers that may give weights after processing data, which is subsequently transmitted to the network's next layer of nodes [7].

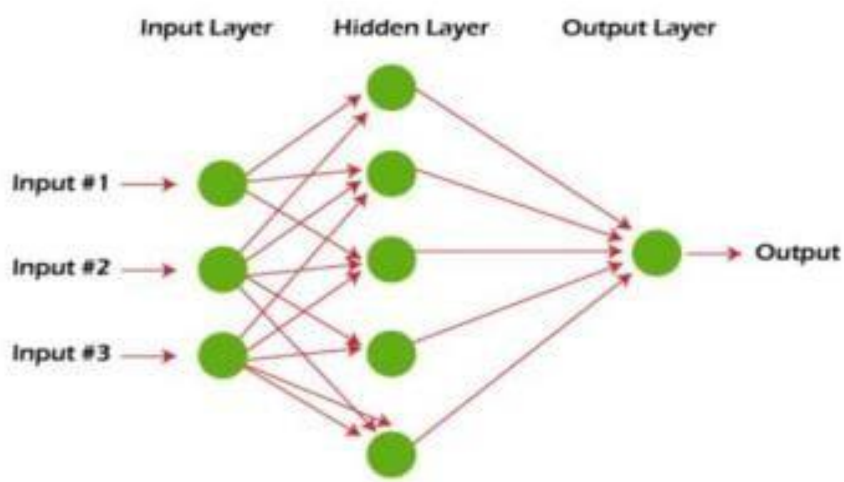


Figure 2.2 Simple representation of neural network

Neural networks can adapt to changes in input and still give the best possible output without the need to alter the model or output properties, making it a reasonably accurate machine learning model [9]. This method is widely used in anticipating load patterns since it is incredibly versatile and allows for accurate forecasting even with few inputs. At its most basic, standard artificial neural networks (see Figure 2) consist of an input layer, a hidden layer, and an output layer. ANN are used for load forecasting in a variety of studies, including [10] - [13], some of which date back to the early days of machine learning as a subject in the 1990s.

2) SVM (Support Vector Machine)

It is a machine learning (ML) technique that uses statistical learning theory (SLT), and it is commonly utilized for categorization applications and regression analysis due to SLT's skills in identifying patterns and evaluating data [14]. The goal of this approach is to find a discriminant function based on the independent training data set that can predict labels for newly obtained occurrences. It needs less training data and computer resources than generative techniques. Because SVM takes an analytical approach to optimization issues, it always provides the same optimal parameters of solution, unlike other ML approaches such as genetic algorithms. SVM has strong generalization ability and is resilient, making it one of the most widely recognized supervised ML algorithms while also being one of the simplest. Furthermore, SVMs have a quick convergence time as well as a strong nonlinear processing capabilities. However, they perform badly with large amounts of data [15]. Figure 3 depicts a generic framework for an SVM technique.

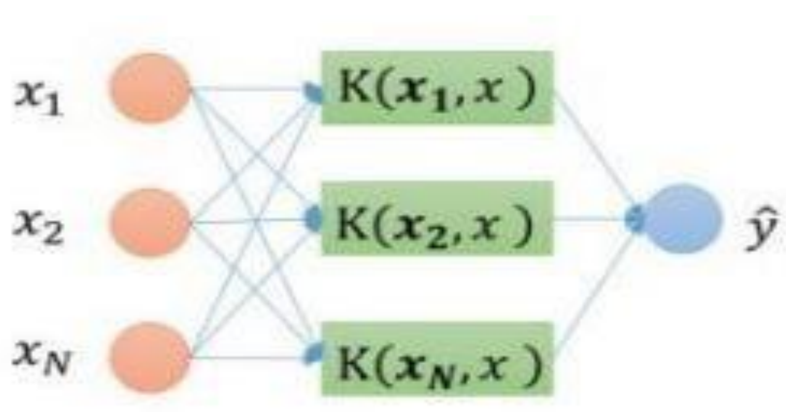


Figure 2.3 general structure for SVM model

3) Fuzzy Logic

Fuzzy logic is a way of describing or viewing information and data that is not based on the usual classical logic which is typically binary (either 0 or 1). Fuzzy logic acknowledges the grey area between the black and white. As in, objects can have "degrees" or "grades" of truths to them rather than the "yes" to "no" range of computer reasoning. This concept was introduced initially in the 1960s by Zadeh in his paper "Fuzzy sets," [16] for the purpose of better understanding and analyzing data that doesn't fall in a classical logic frame.

Generally, most objects in the real world don't fall into a "yes" or "no" frame. Uncertainties and partial degrees of truths play a big rule in human reasoning which is what artificial intelligence and consequentially machine learning is aiming to replicate. This is where fuzzy sets have emerged as a concept which Zadeh suggested in [16] to deal with class criteria that aren't Sharpley defined. These grades of relations of objects or data in a fuzzy set have a range that extends from 0 to 1 (in between values).

This logic is defined by a membership function that designates the "grades" to each fuzzy set of objects. This means that any "object" or "data point" in a set has a value relating it to the membership function of the said "set" or "class", the value of this membership function at the calculated point denotes exactly the strength of this point at this set (ea. the closer the number is to 1, the stronger the grade of membership of this particular point is to the set).

In the field of electrical load forecasting, this can be very useful in representing the vagueness in some of the parameters contributing to the load pattern where the effect of each parameter can be defined as its own membership function, such application can be observed in [17], [18] where a combination of fuzzy sets logic and convolution neural networks is used to develop a forecasting model.

4) Genetic Algorithms (GA)

Genetic algorithm is a machine learning concept that focuses on mimicking the same principle as biological genetic evolution and natural selection. It is very effective for search optimization that includes huge and unorganized data. It is generally used to solve constrained and unconstrained complex issues. Genetic algorithms are a subset of the field of evolutionary computation, and they have proven to be very adaptable [19]

In GA, a variety of solutions are provided to a problem. These solutions would encounter processes of “mutation” and recombination to produce new offspring and the previous process is then repeated for several “generations” of solutions.

This concept is similar to genetic reproduction in humans, and likewise, the new solutions would “mate” and “combine” based on the fittest match to produce the optimum “offspring”. This procedure aims to guarantee the creation of fitter or better solutions in subsequent generations and it continues until it reaches the desired outcome criterion [19]. Some of these pre-mentioned operations, which get performed on the solutions to new generations, are:

(1) Selection

(2) Crossover

(3) Mutation

Therefore, it can be considered as the analog version of chromosomic biology. Moreover, due to the flexibility of this method it has been used also to generate load forecasts, in [20]a function for the long-term load forecasting problem utilizing genetic algorithm has been designed and optimized.

Plentiful other methods are available in literature, however, these are the most popular. Some methods merge classical and modern methods to come up with hybrid approaches to load forecasting. In addition, a combination of different machine learning methods could be used together in one forecasting model like [21] that suggests an ANN model associated with genetic algorithm for optimization of weights and biases. All these different methods have their own perks and flaws and most of them can be used to predict loads at varying efficiencies and computation time. Some of these applications will be discussed in details in the following sections of this paper. Such as a deep neural networks approach, recurrent neural network with Elman’s and particle swarm optimization along with an SVM application employing a kernel function and fuzzy logic.

Methodological and Other Issues

1. LOAD FORECASTING WITH DEEP NEURAL NETWORKS.

Deep neural networks have been employed in predicting applications in few research applications. Similarly, [2] presents a deep neural network architecture for hourly dayahead estimates of electric demands. It employs several NN components, models, and layers to represent various elements that may influence load consumption and behavior.

A. Factors affecting Load forecasting

Each factor influences the load forecasting in different proportions and each model may include and consider different factors, some of which are mentioned below.

1) The weather

Weather is one of the most important elements influencing load fluctuations throughout the day and year. As severely low temperatures can raise the use of heaters, significantly high temperatures can increase the use of air conditioning. Furthermore, inclement weather increases the amount of time spent indoors. This shows how weather has a significant impact on power use patterns. Alternative weather parameters such as wind speed and humidity may also be useful for load predictions; however, we lack the necessary information about such factors in the available data set, and their effect is less pronounced than that of temperature, except in extreme cases.

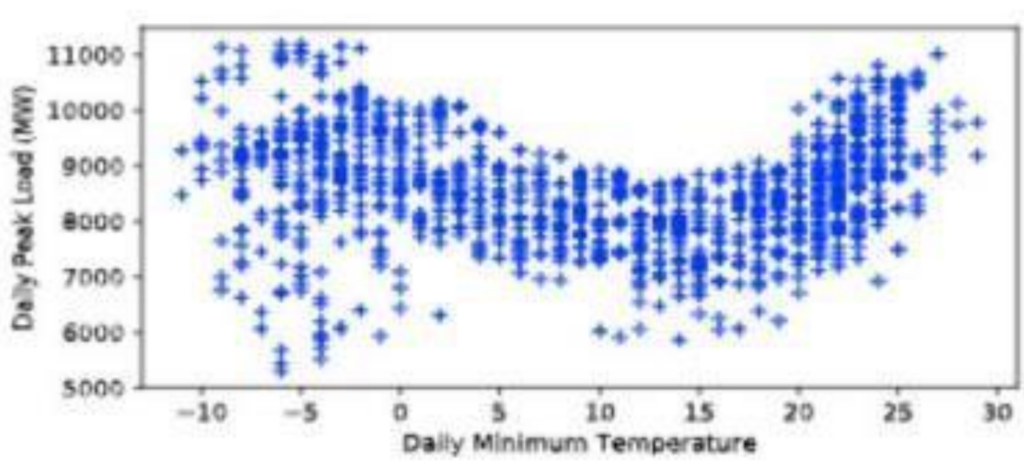


Figure 2.4 the effect of temperature in daily load patterns

Additionally, the time itself and the week day of the target time are clearly useful indicators for predicting hourly load. The variation that time of day and temperature can cause are displayed in Figure 4.

2) Holidays On long holidays and public/national vacations, people's consumption patterns vary as it is increased due to home gatherings or decreased due to travel. It is variable depending on the location as well.

3) Electricity price and energy policies some countries have energy regulations that can cause the load to follow a certain pattern. For example, in some countries the price of electricity can be higher during specific times of day to encourage people to conserve energy at usually high/peak load hours.

The patterns stated above are some of the characteristics that were used as inputs to the NN, especially the deep feed forward component of the model. This model is found in [2]. Seeks to forecast hourly loads for a certain day and hour based on weather, time of day, and holiday for this specific area and time utilizing previous data spanning 24 hours

The use of multiple Convulsion Neural Networks (CNN) layers to adapt and analyze historical data for load pattern is proposed in reference [2]. CNNs have demonstrated their efficiency in tasks like as extracting and learning features from input data, according to [22]. As shown in Figure 5, three parallel CNN components with varying filter widths were used to turn the historical load pattern into a collection of variable features. Due to their efficacy in representing sequential dynamic data, these time-series characteristics and input components are subsequently processed by recurrent neural networks (RNNs).

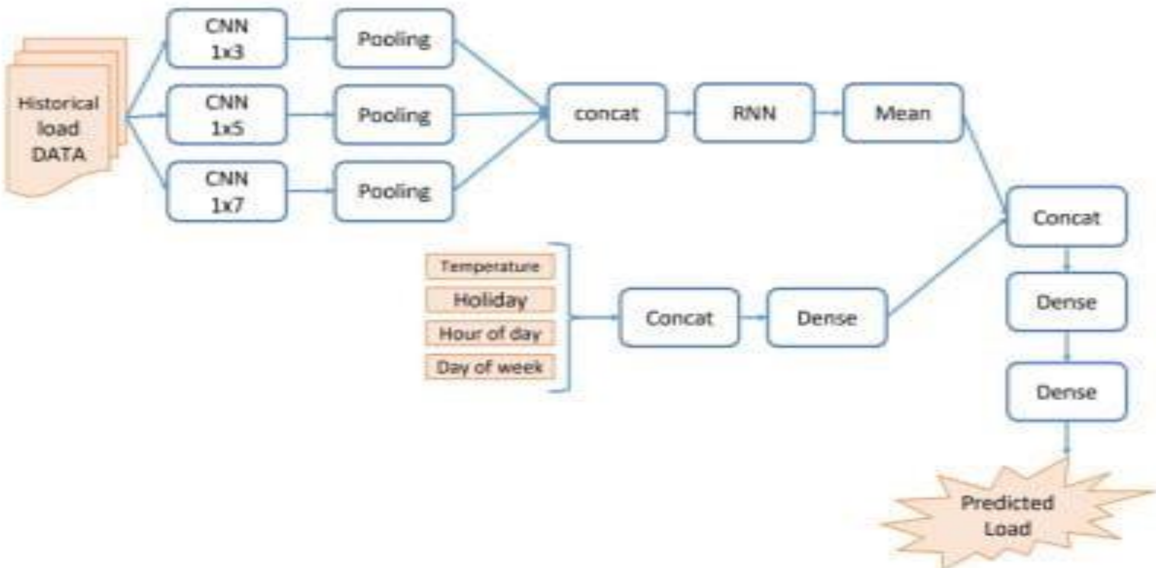


Figure 2.5 the layers and model for the proposed CNN method

To eliminate gradient difficulties that arise with basic RNNs, RNNs are linked with a long-short term memory unit (LSTM), and the neural component of LTSM is used to mimic variability in previous load data.

Other elements, such as temperature and national vacation days, are also incorporated in the vector representation via feed-forward components. All of these characteristics are then concatenated and sent as inputs to the DNN element of this model, which computes and learns from the raw data on its own. The data set is organized into three sections: training, validation, and testing. The evaluation is based on three years of hourly load data, and the strategy has been shown successful based on testing findings.

The suggested approach is compared to SVM, linear regression, and a third method (similar to the proposed but with differences in layers). The suggested in load forecasting, this strategy appears to outperform the others.

2. RECURRENT NEURAL NETWORKS FOR LOAD FORECASTING

A recurrent neural network model for load forecasting is suggested in reference [23]. A recurrent neural network (RNN) is a type of neural network in which components are linked in a direct cycle. RNNs may sort arbitrary input patterns using their internal memory. For model training, simple RNNs employ the idea of back propagation (BP). It is a supervised learning neural network model in which the network constantly adjusts weights and biases until it fits the historical data, which is then assigned specific numerical values that match the input characteristics with the output predictions using training data.

A RNN is made up of at least one feedback circle. It might be made up of a single layer of neurons, with the output of each neuron feeding back into the input of all the others. The output of neurons in the output and hidden layers is determined by taking the weighted total quantity of inputs and applying the tangent sigmoid activation function and pure linear transfer functions, respectively. Arithmetically, the weights are revised using a learning rate and the variable's partial derivation, indicating the direction of change with respect to a local or global minimum, with respect to the output error, which are updated every loop to adjust the best fitted solution of weights to the current problem.

Elman's recurrent neural network was chosen to be modeled in [23] to analyze the load forecasting data since it has shown to be an efficient RNN design. As illustrated in Figure 6, ERNN is made up of recurrent linkages in the hidden layer of neurons that connect to a delay unit that holds the output of these neurons for a single time period before re-feeding them to the hidden layer's input.

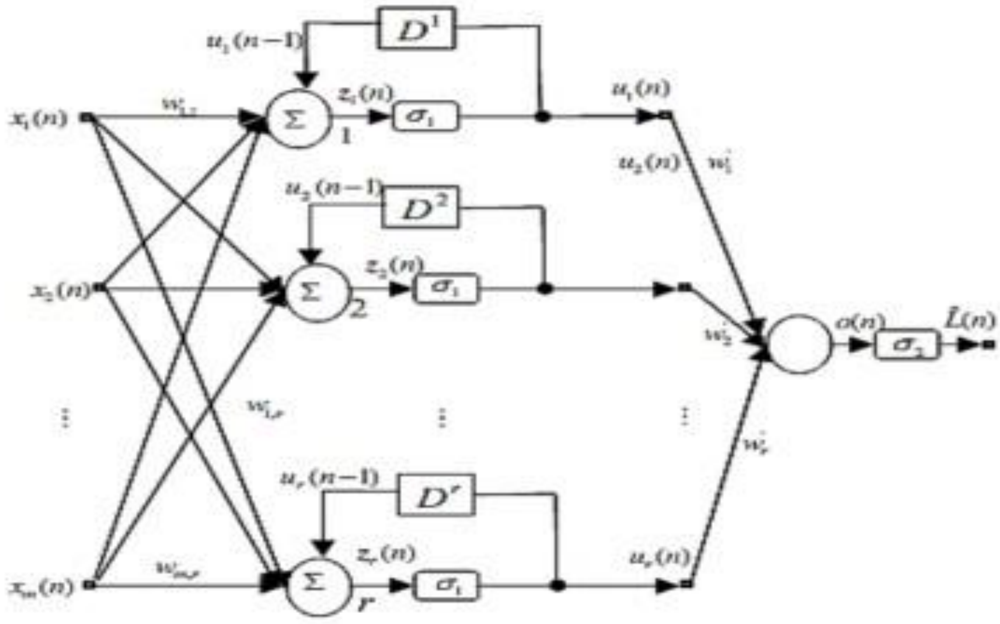


Figure 2.6 the topology of an Elman's RNN

Particle Swarm Optimization (PSO) is a type of optimization algorithm. Particle swarm optimization (PSO) is a method for conducting population-based searches. Particles in a PSO system (Input data) move across a multidimensional search space. During their motion, these particles undergo their own tunings or changes (position modifications) based on their own experience in multidimensional space or the experience of a neighboring particle. In this manner, the particles modify themselves by employing their optimal position and the assistance of a neighbor for better suited outcomes. As a result, PSO utilizes and integrates both local and global search methodologies to achieve a balance of exploration and exploitation

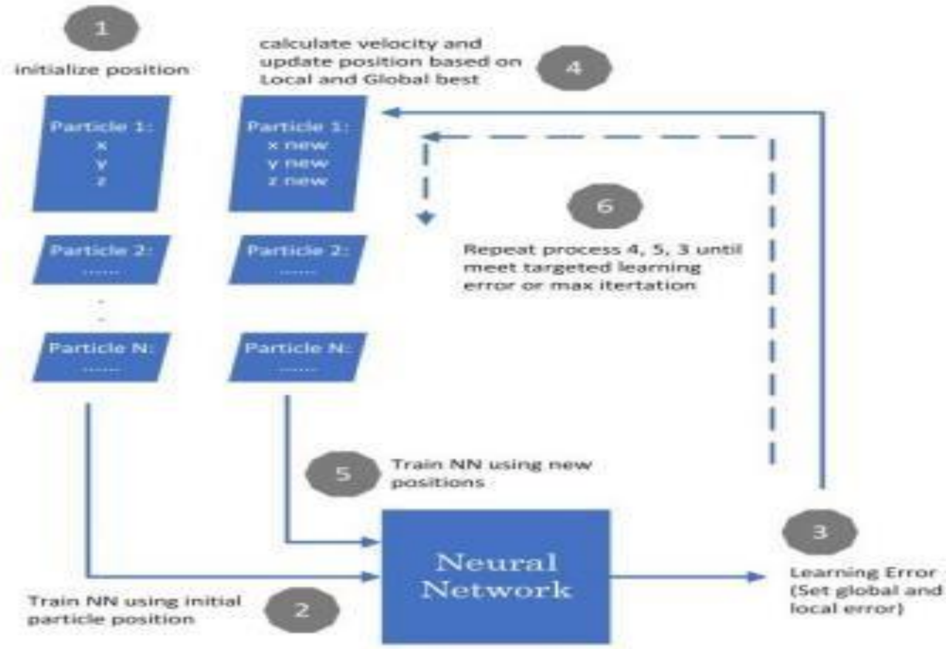


Figure 2.7 the training process for this PSO-ERNN network

PSO was employed as a tool to alter the best fitting weights and locations in the Elman's RNN in the application outlined in [23], and then the optimal answer to this learning process is utilized to anticipate and forecast loads. Figure 7 depicts the network's learning process. [23] Determined that a particle swarm optimized Elman recurrent neural network (PSO-ERNN) is the best prototype for predicting 168 hours ahead of time. This type of network can be highly effective in terms of projecting future demands. Though the simulations appeared to be incredibly successful, data from other sources must still be used to evaluate the produced models in order to ensure their consistency for additional data sets. This model may be enhanced by taking into account other elements that influence the daily load forecast

3. SUPPORT VECTOR MACHINE BASED LOAD FORECASTING

Support vector machine is a machine learning (ML) learning load forecasting technique that incorporates statistical learning theory (SLT) [14]. It collaborates with the kernel function to generate a linear map from the input data space to the high-dimensional feature area. Nonlinearity, short sample sizes, huge dimensions, and local optimization may all be handled using SVM, which gives a more global answer. Traditional SVM algorithms, on the other hand, are incapable of filtering out worthless data worthless, redundant data, and unable to reduce dimensionality. In [24], an SVM-based load prediction model is presented to anticipate loads using data mining technologies to address the inadequacies of traditional SVM models by overcoming challenges with huge amounts of data processing and eliminating information redundancy.

This plan considers a number of elements and metrics, including typical load range, weather conditions, and special date features. This new data sequence will serve as the SVM model's training data. This lowers the dimensions of the intake and computations while also making the previous data more regular and hence more applicable.

This model's data mining method chooses comparable day samples while accounting for all external and internal aspects (as previously described, which standard SVM implementations fail to consider). This method combines similarities mining, fuzzy grouping, and grey relational theory. The suggested technique succeeds in making the input data and calculations simpler. Furthermore, it enhances the criteria for the training samples. This suggested model produces a decent prediction by picking similar-days from the initial data; samples that considerably differ from the load attributes of the projected day in the original set of data are discarded. As a result, the model's training time is obviously reduced, and the predictions made are more accurate.

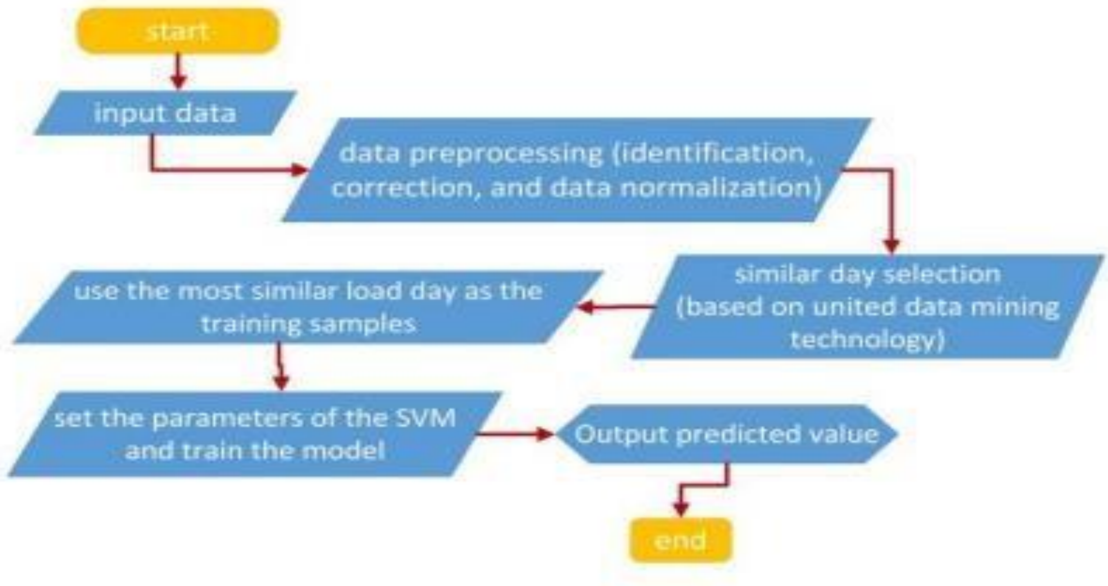


Figure 2.8 Flow chat of the load forecasting SVM similar days based application

The gaussian radial basis function is used as the SVM's kernel function. When the projected results are compared to the results of the basic SVM technique, the average error is 2.06%, which is smaller than the average error for the classical SVM, which came out to be approximately 2.54%. This is owing to the suggested method's capacity to efficiently remove. Data redundancy reduces computation complexity. It is also efficient, quick, and capable of meeting criteria.

Gaps or Controversies in the Literature and Some Deductions.

Classical method (Qualitative Methods)

In classical approaches, such as qualitative methods, historical data are not utilized for predictions. Instead, experts rely on structured methodologies to forecast electrical loads. However, this approach often leads to significant inaccuracies due to its reliance on subjective expert opinions.

Artificial Neural Networks (ANNs)

Artificial Neural Networks (ANNs) possess the unique ability to adapt to changing input data without requiring extensive model modifications. This adaptability contributes to their effectiveness as a machine learning model, allowing them to consistently produce accurate predictions.

Support Vector Machines (SVMs)

Support Vector Machines (SVMs) are characterized by their swift convergence and robust nonlinear data processing capabilities. Nevertheless, their performance tends to deteriorate when dealing with massive datasets.

Genetic Algorithms (GAs)

Genetic Algorithms (GAs) excel in optimizing searches, particularly in scenarios involving extensive and unstructured data. They are highly effective in such contexts

The substantial gap is that in previous approaches, there was a lack of consideration for factors such as holidays, weather conditions, date or timestamp, and whether it was the dry or rainy season.

CHAPTER 3

SYSTEM DESIGN

3.1 The Concept

- Data Collection: Gathering of historical load data from utility companies or relevant sources. This data would include electricity consumption values along with associated timestamps and any other relevant factors like weather conditions

- Preprocessing: Clean and preprocess the collected data. Handle missing values, outliers, and data inconsistencies. Convert timestamps into appropriate formats and adjust for time zones if necessary. Perform data normalization or scaling to ensure consistent ranges for features.

- Feature Engineering: Extracting meaningful features from the data to capture important patterns and relationships. Some common features for load forecasting may include time of day, day of the week, month, holidays, weather conditions (temperature, humidity), and historical load values (lagged load).

- Data Split: Splitting the preprocessed dataset into training and testing sets. The training set will be used to train the machine learning model, while the testing set will be used to evaluate the model's performance

- Model Selection: Choosing a suitable machine learning algorithm for load forecasting. Depending on the specific requirements and characteristics of the data, algorithms such as linear regression, decision trees, random forests, support vector machines (SVM), or neural networks (e.g., LSTM) can be considered.

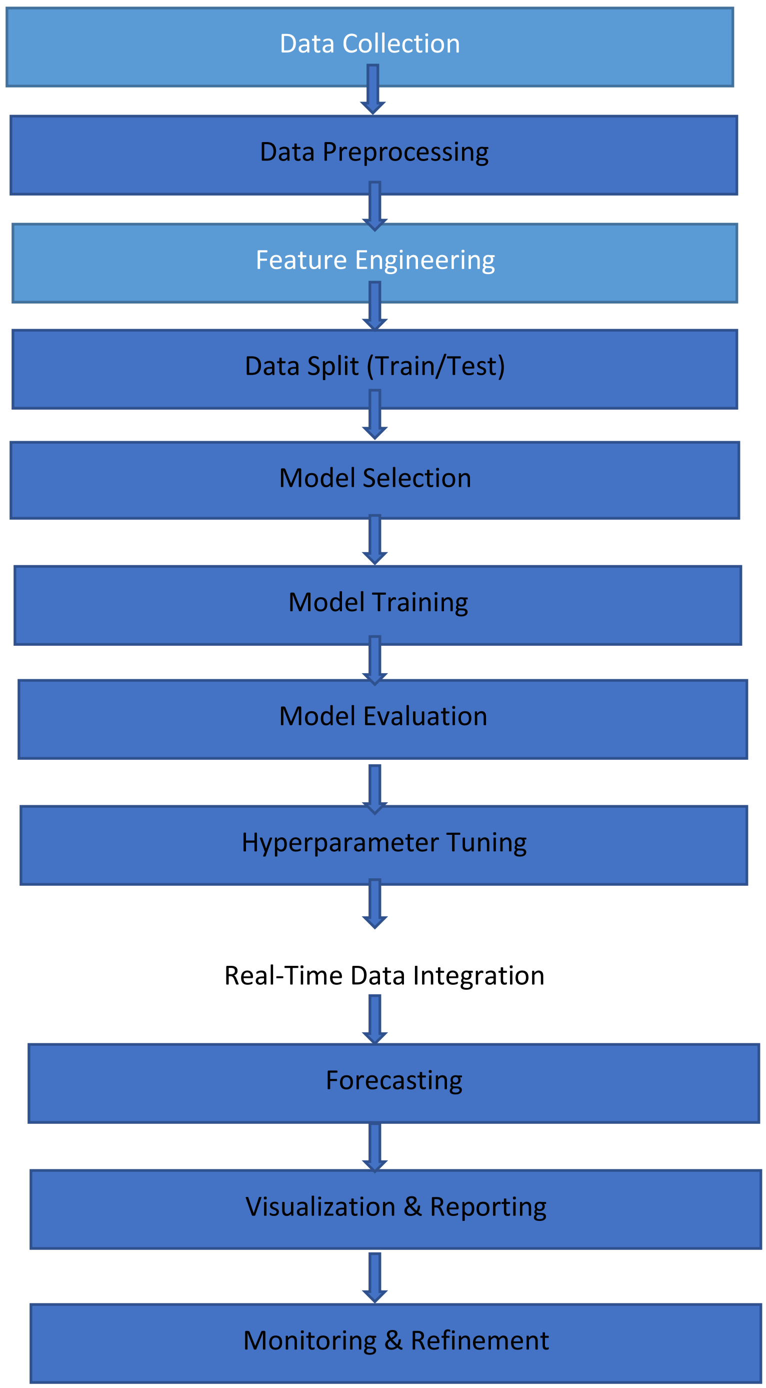
- Model Training: Training the selected machine learning model using the training dataset. The model learns the underlying patterns and relationships between the load and the selected features

- Model Evaluation: Evaluating the trained model using the testing dataset. Calculate appropriate performance metrics such as mean absolute percentage error (MAPE), root mean square error (RMSE), or coefficient of determination (R-squared) to assess the accuracy and reliability of the load forecasts.
- Hyperparameter Tuning: Fine-tune the model by adjusting hyperparameters to optimize its performance. This can be done using techniques like grid search or random search, which explore different combinations of hyperparameters to find the optimal configuration

- Real-time Data Integration: Implement a mechanism to integrate real-time data, such as live load measurements and up-to-date weather information, into the forecasting model. This allows the model to adapt and generate more accurate forecasts based on the most recent data.
- Forecasting: Once the model is trained and validated, it can be used to generate load forecasts for future time periods. Incorporate relevant input data (e.g., weather forecasts) to generate accurate predictions.
- Visualization and Reporting: Visualizing the forecasted load values along with actual load data to provide insights and aid decision-making. Creating informative plots, charts, and graphs to communicate the forecasted results effectively.
- Monitoring and Refinement: Continuously monitoring the performance of the load forecasting system and refining the models as new data becomes available. Implementing feedback mechanisms to ensure the system adapts to changes in load patterns and improves over time.

Note: It's important to note that the system design may vary depending on the specific requirements, available data, and the machine learning algorithms we get as time goes on. This concept provides a general framework for developing an electrical load forecasting system using machine learning.

3.2 BLOCK DIAGRAM

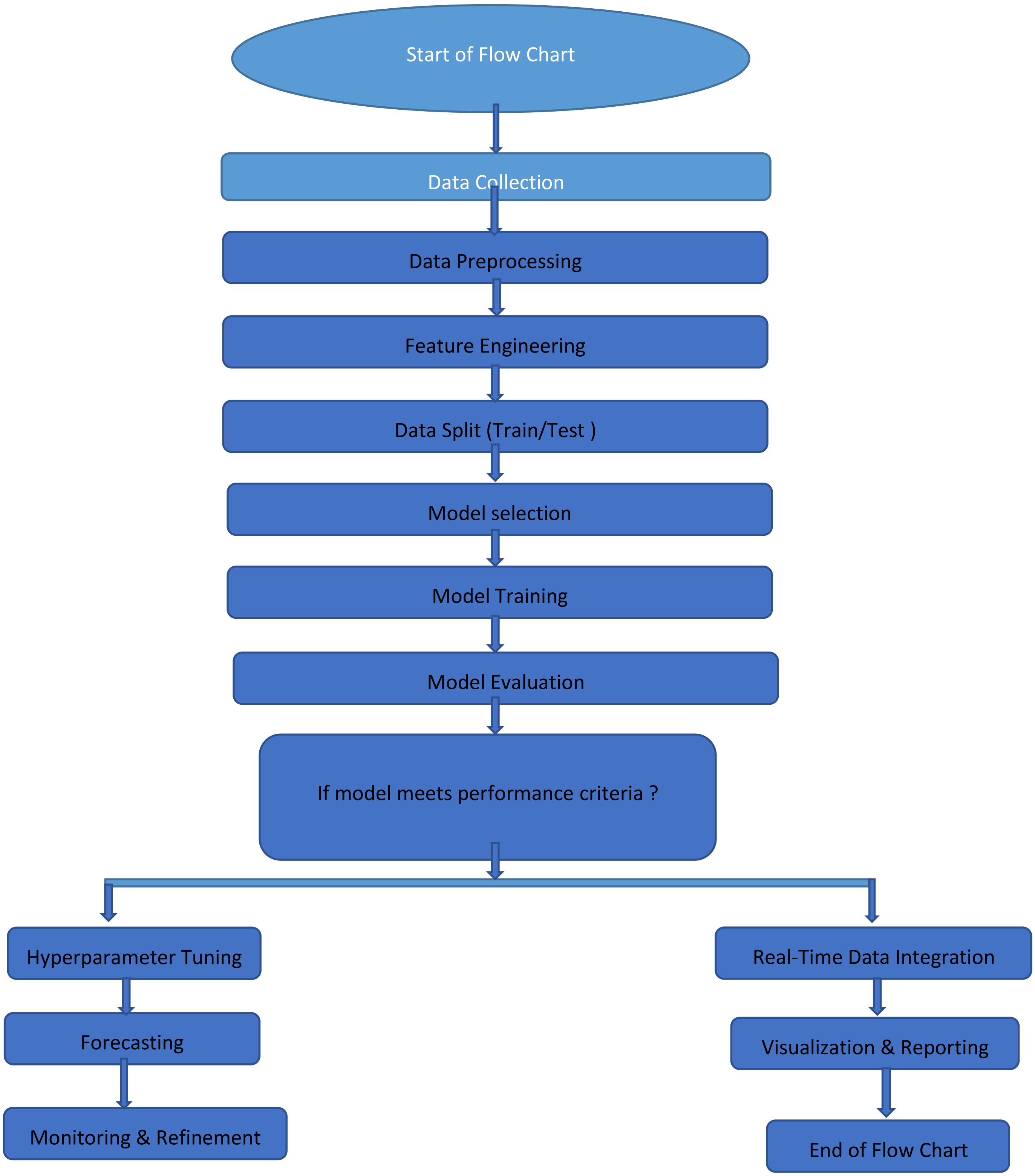


Description:

- Data Collection: Gather historical load data from utility companies or relevant sources. This data includes electricity consumption values along with timestamps and other relevant factors like weather conditions.
- Data Preprocessing: Clean and preprocess the collected data. Handle missing values, outliers, and data inconsistencies. Convert timestamps into appropriate formats and adjust for time zones if necessary. Perform data normalization or scaling.
- Feature Engineering: Extract meaningful features from the data to capture important patterns and relationships. Features may include time of day, day of the week, month, holidays, weather conditions, and historical load values (lagged load).
- Data Split (Train/Test): Split the preprocessed dataset into training and testing sets. The training set is used to train the machine learning model, while the testing set is used to evaluate its performance.
- Model Selection: Choose a suitable machine learning algorithm for load forecasting based on the specific requirements and characteristics of the data. Algorithms can include linear regression, decision trees, random forests, support vector machines (SVM), or neural networks (e.g., LSTM).
- Model Training: Train the selected machine learning model using the training dataset. The model learns the patterns and relationships between the load and selected features.
- Model Evaluation: Evaluate the trained model using the testing dataset. Calculate performance metrics such as MAPE, RMSE, or R-squared to assess the accuracy and reliability of the load forecasts.
- Hyperparameter Tuning: Fine-tune the model by adjusting hyperparameters to optimize its performance. Techniques like grid search or random search explore different combinations of hyperparameters to find the optimal configuration.
- Real-time Data Integration: Implement a mechanism to integrate real-time data, such as live load measurements and up-to-date weather information, into the forecasting model. This allows the model to adapt and generate more accurate forecasts.

- Forecasting: Once the model is trained and validated, it is used to generate load forecasts for future time periods. Relevant input data, such as weather forecasts, can be incorporated for accurate predictions.
- Visualization & Reporting: Visualize the forecasted load values alongside actual load data to provide insights and aid decision-making. Create informative plots, charts, and graphs to communicate the forecasted results effectively.
- Monitoring & Refinement: Continuously monitor the performance of the load forecasting system and refine the models as new data becomes available. Implementing feedback mechanisms to ensure the system adapts to changes in load patterns and improves over time

FLOW CHART


Start of Flow Chart
Data Collection
Data Preprocessing
Feature Engineering
Data Split (Train/Test)
Model selection
Model Training
Model Evaluation
If model meets performance criteria ?
Hyperparameter Tuning
Forecasting
Monitoring & Refinement
Real-Time Data Integration
Visualization & Reporting
End of Flow Chart

Explanation

- The flowchart starts with data collection, where historical load data is gathered from various sources.
- The collected data then undergoes preprocessing, where cleaning, handling missing values, and normalization take place.
- Feature engineering is performed to extract relevant features from the data.
- The preprocessed data is split into training and testing sets.
- Model selection involves choosing an appropriate machine learning algorithm for load forecasting.
- The selected model is trained using the training data.
- The trained model is evaluated using the testing data to assess its performance.
- If the model meets the predefined performance criteria, hyperparameter tuning and realtime data integration are performed.
- Hyperparameter tuning involves optimizing the model's hyperparameters to improve its performance.
- Real-time data integration involves incorporating real-time data, such as live load measurements and weather information, into the model.
- The forecasting stage utilizes the trained and tuned model to generate load forecasts.
- Visualization and reporting are carried out to present the forecasted load values and insights derived from the model.
- The system is monitored, and refinement is performed as needed to enhance the load forecasting system.
- The flowchart ends, indicating the completion of the process.

Please note that the flowchart provides a high-level overview of the system design concept. The specific steps and details may vary depending on the project's requirements and implementation choices as time goes on.

CHAPTER FOUR

ANALYSIS AND DISCUSSIONS

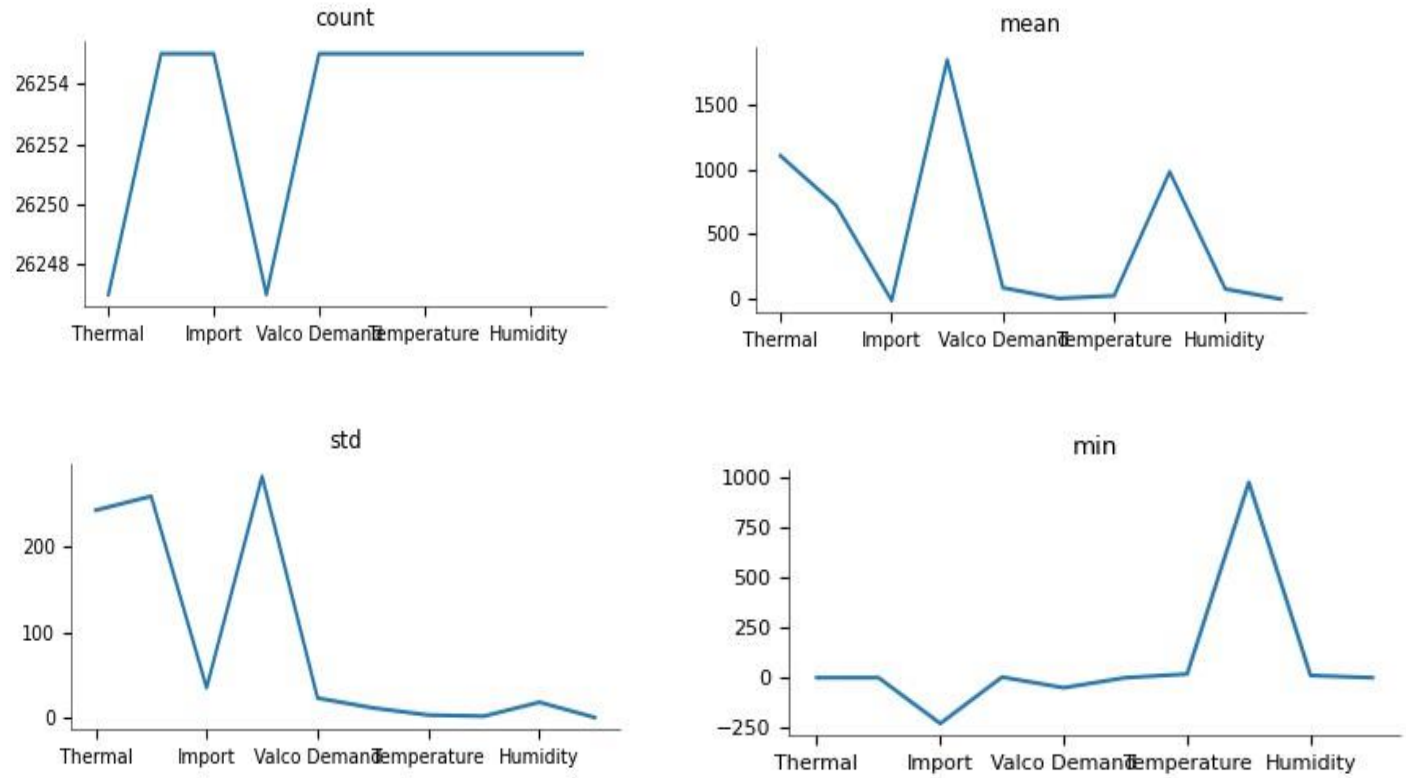
In this chapter, the focus is on employing descriptive analysis to investigate the dataset meticulously to garner insightful findings, which are paramount in understanding the dynamics of load demand at the Volta Aluminium Company (VALCO) in Ghana. Descriptive analysis is the first step in analyzing the raw data to find trends, patterns, and anomalies that can significantly affect load forecasting. This step serves as a precursor to the application of machine learning algorithms for predictive analysis

4.1: Data Description

Before delving into the machine learning algorithms' application, a thorough examination and understanding of the data are essential. The dataset comprises historical load demand records, weather conditions, and operational hours of VAC over a specified period. A summary statistics table is presented to give a preliminary understanding of the data distribution, including mean, median, standard deviation, and range values for each variable .

	count	mean	std	min	25%	50%	75%	max
Thermal	26247.0	1104.143824	241.625078	0.0	964.00	1128.00	1269.00	2527.00
Hydro	26255.0	722.185262	257.880964	0.0	501.00	692.00	910.00	4356.00
Import	26255.0	-10.815616	35.287878	-228.0	-21.00	0.00	5.00	172.00
Total	26247.0	1842.772564	281.073509	3.0	1646.40	1811.50	2019.10	5294.80
Valco Demand	26255.0	87.992420	23.354500	-50.0	72.00	74.00	109.00	1221.00
Export	26255.0	5.636603	11.880852	0.0	0.00	0.00	5.00	172.00
Temperature	26255.0	26.140675	3.602987	18.2	23.26	25.21	28.64	37.84
Pressure	26255.0	979.966243	2.323356	972.3	978.40	980.10	981.60	989.80
Humidity	26255.0	78.374702	18.503669	11.1	68.90	83.70	93.10	100.00
Wind_Speed	26255.0	0.723420	0.755362	0.0	0.00	0.40	1.10	6.90

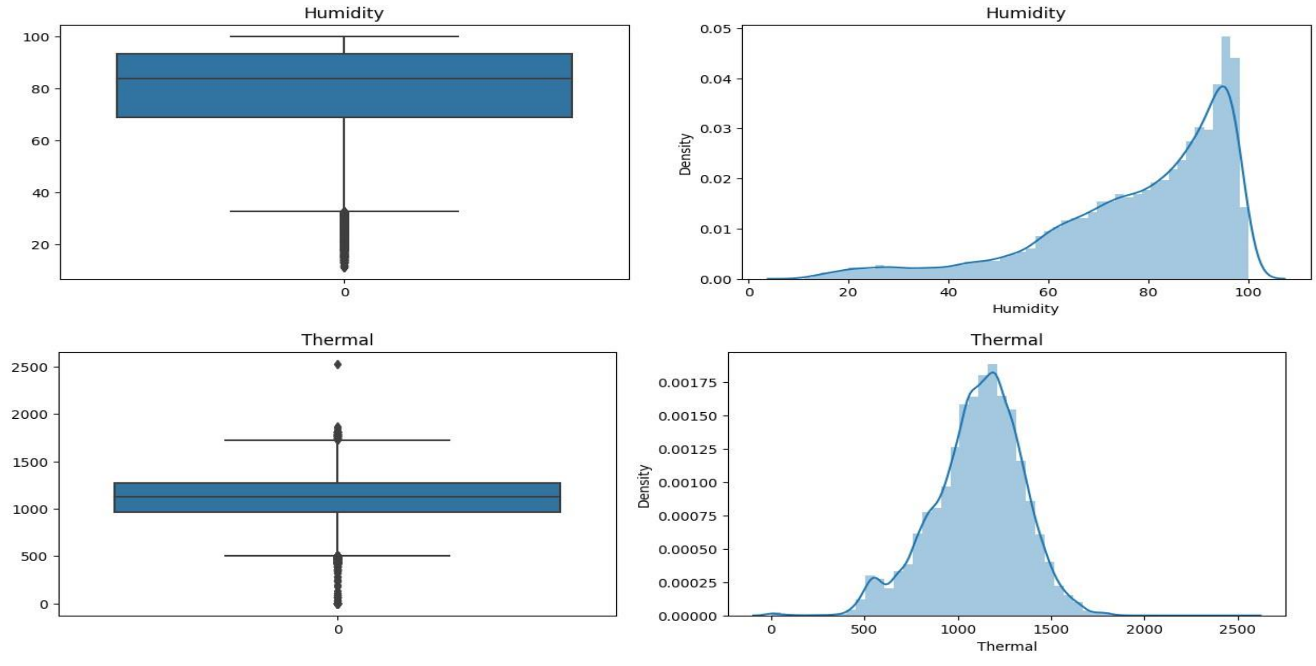
Visualization

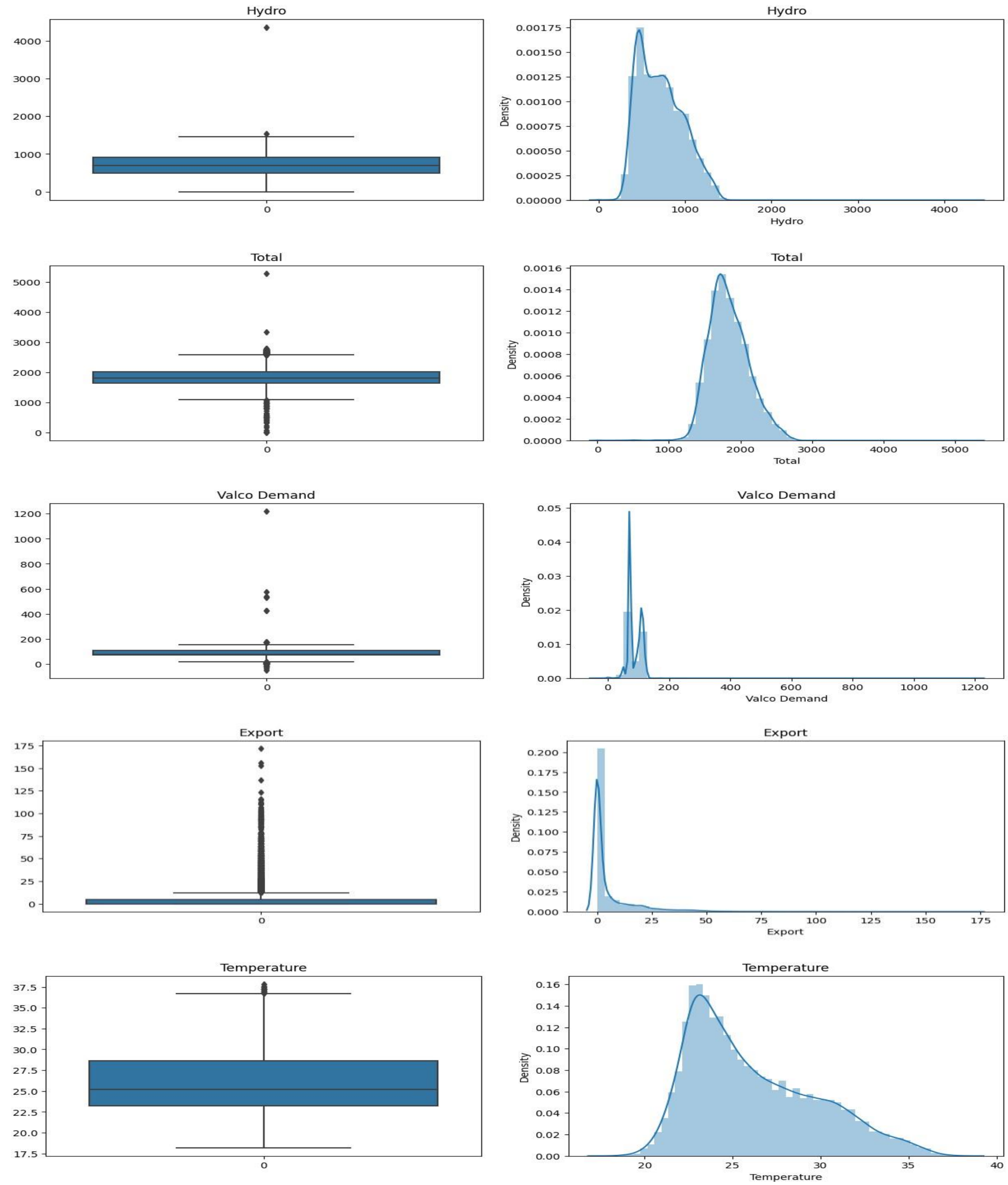


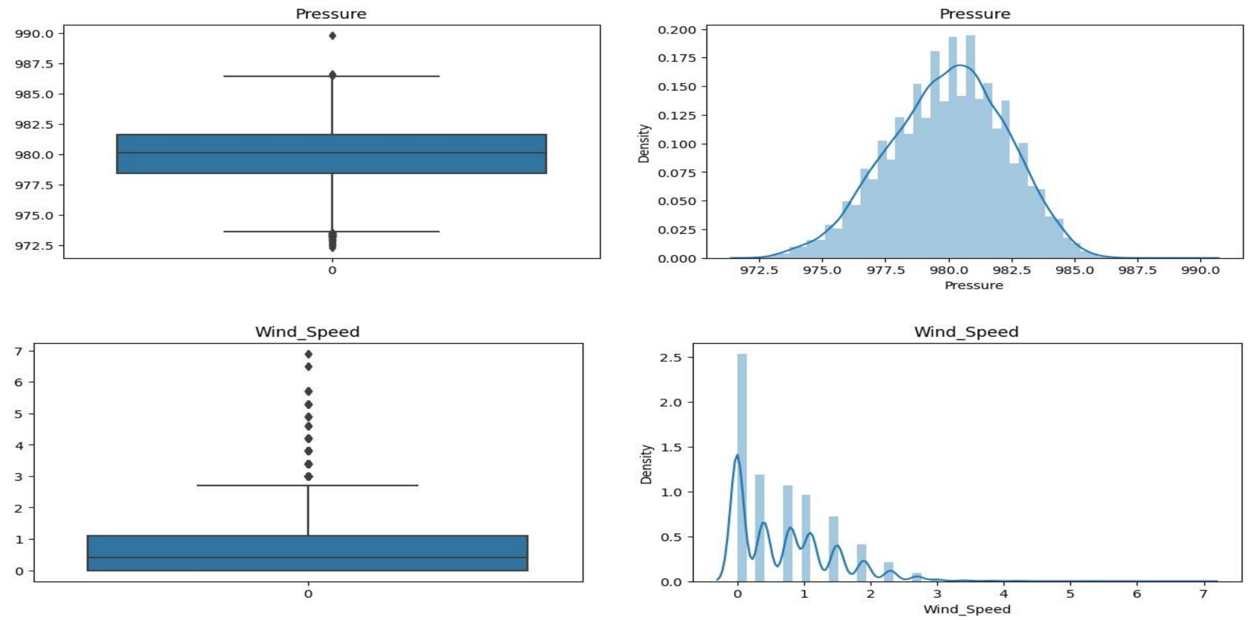
Understanding the data distribution through metrics like mean, median, standard deviation, and range is critical in this project as it lays the groundwork for predictive model development, helps in identifying and addressing outliers, and assesses data normality which many machine learning algorithms assume. These metrics are essential for feature scaling to ensure the input variables are on a similar scale, which in turn aids in better performance of the machine learning models. They also enhance data comparability, interpretability, and provide insights into relationships among variables, assisting in effective data cleaning and preprocessing. Furthermore, they serve as simple benchmarks for model performance comparison, and aid in effectively communicating data characteristics to stakeholders, ensuring the development of robust, accurate, and reliable machine learning models for load forecasting in this project.

Outlier Detection/ Imputation Strategy

The use of Box Plots and Symmetric Distribution Plots was instrumental in identifying outliers and understanding skewness in the dataset, respectively. The Box Plots helped in visualizing the distribution of data and easily spotting outliers through the plotting of data points outside the whiskers, while also providing a visual representation of data dispersion through the interquartile range. On the other hand, Symmetric Distribution Plots aided in assessing the skewness of data distribution which is essential in understanding how data deviates from normality. These visualizations were pivotal in devising an imputation strategy for handling missing values, should they arise. For normally distributed data or data with slight skewness, mean imputation was considered, while for highly skewed data, or in the presence of outliers, median imputation was deemed more suitable due to its robustness against extreme values. Through this detailed analysis, a sound imputation strategy was established, ensuring data integrity and accuracy for further analysis and machine learning modeling in the project. ***Visualization***




Hydro
Hydro
Density
Hydro
Total
Total
Density
Total
Valco Demand
Valco Demand
Density
Valco Demand
Export
Export
Density
Export
Temperature
Temperature
Density
Temperature

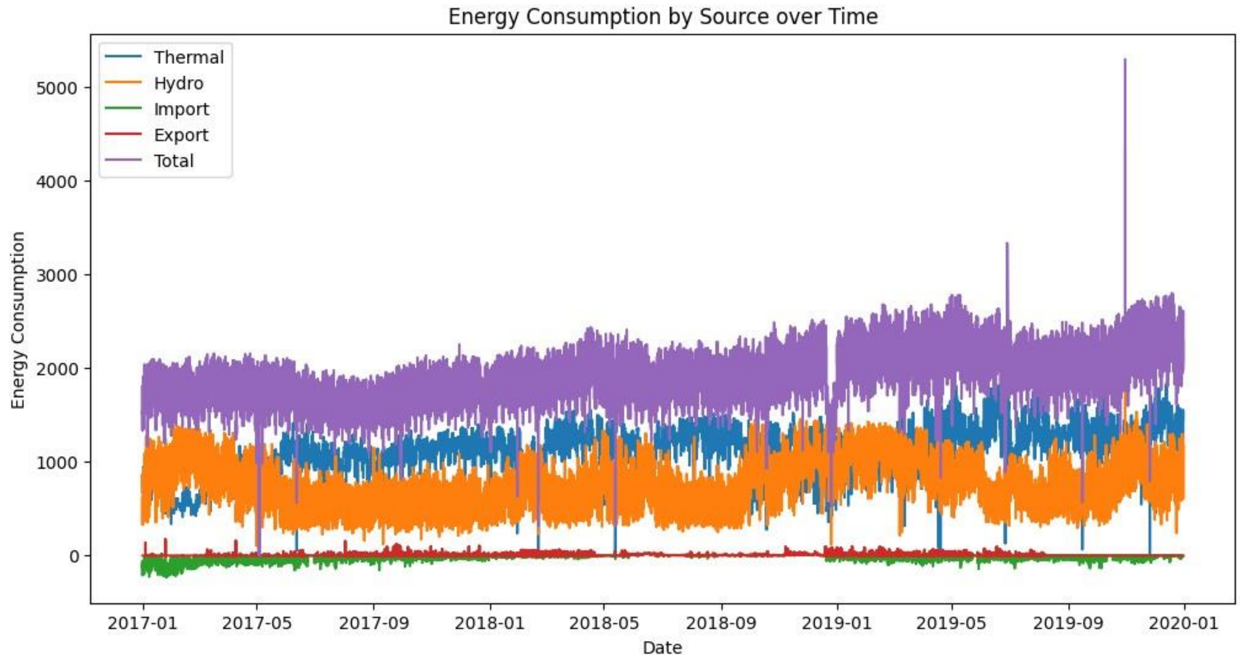


Handling missing values properly is critical for preserving the integrity and quality of the dataset, which in turn significantly impacts the performance and predictive accuracy of machine learning models. Addressing missing values adequately helps avoid biased or misleading results, ensuring that the analysis remains representative of the underlying phenomena. This is essential for models to comply with certain statistical assumptions and for ensuring valid, interpretable outputs. The proper handling of missing values also contributes to more robust models, informed decisionmaking, and an increased understanding of data patterns. In scenarios where data-driven decisions are paramount, having accurate and reliable information is crucial.

4.2 Trend Analysis

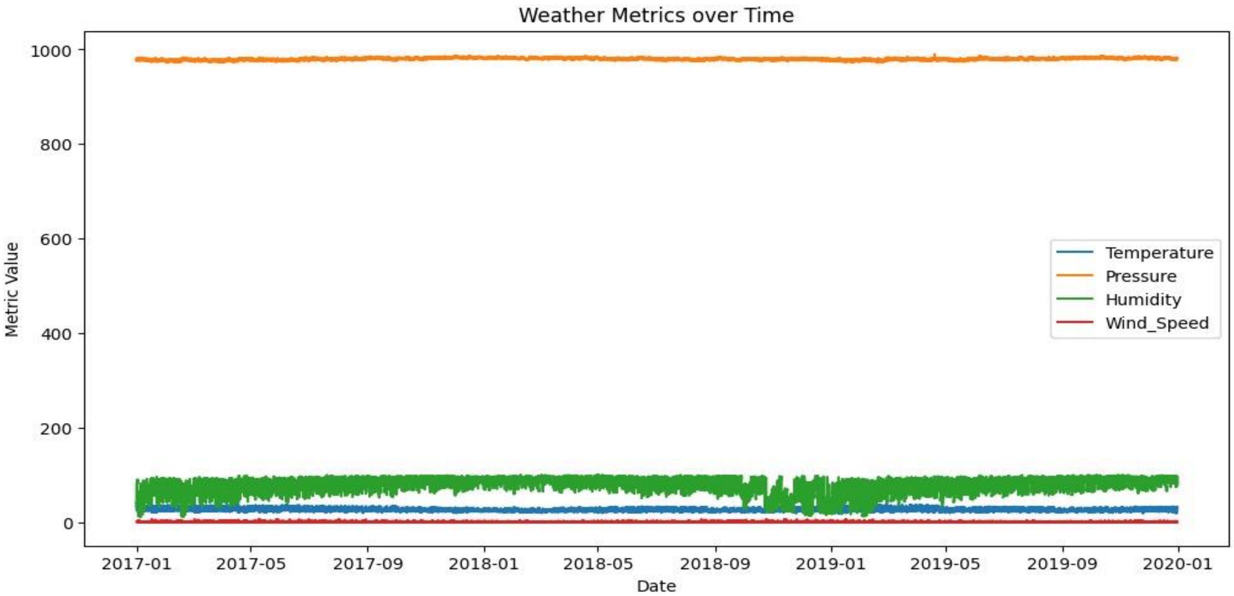
Trend analysis unveils the general direction in which the load demand is moving over time. By plotting the time series data, we observe daily and seasonal trends in electricity consumption. This section elaborates on the identified trends and discusses possible correlations with operational and environmental factors such as temperature and humidity.

Energy Consumption by Source over Time



The insights derived from this time series analysis were instrumental in understanding the dynamics of energy consumption within the Volta Aluminium Company. Observing the historical trends and patterns helped to make more informed decisions regarding future energy consumption forecasting. This analysis set a solid foundation for the predictive modeling endeavors in subsequent sections of the project, assisting in anticipating how the trends seen in the historical data may continue or change in the future, and how these trends relate to the broader objectives of load forecasting for the Volta Aluminium Company.

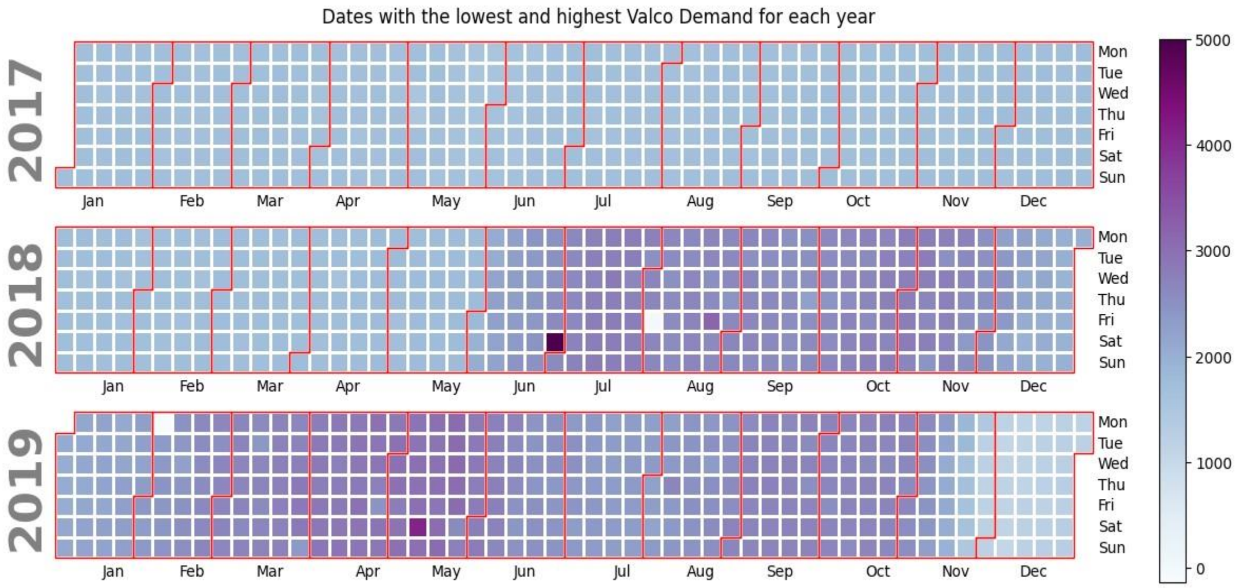
Weather Metrics over Time



Objective was to visualize and analyze weather metrics such as Temperature, Pressure, Humidity, and Wind Speed over time to identify trends and patterns that could potentially impact energy consumption. A multi-line plot was generated to showcase the variations in these metrics on a unified time axis, thereby facilitating a visual comparison amongst them.

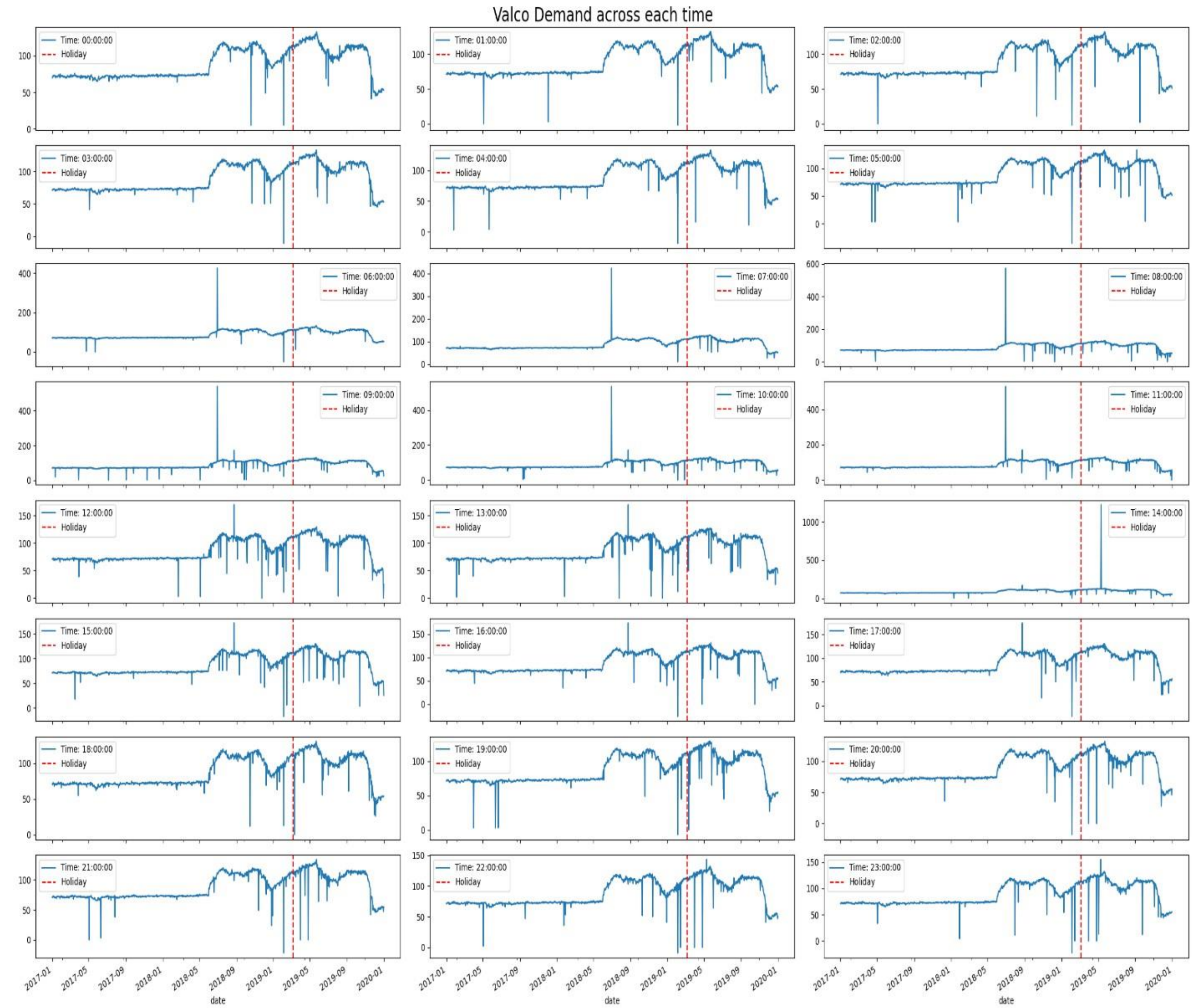
This graphical analysis was pivotal in unraveling the dynamics between weather conditions and energy consumption within the Volta Aluminium Company. It aided in understanding how weather variables could potentially influence energy usage patterns, thus providing an informed basis for incorporating weather metrics into the subsequent predictive modeling for more accurate load forecasting. The visual comparison allowed for a deeper understanding of the multifaceted factors impacting energy consumption, setting a comprehensive groundwork for the ensuing predictive analysis aimed at forecasting energy demand.

Dates with the lowest and highest Valco Demand for each year



This section utilized a specialized visualization to depict Valco's energy demand trends across 2017, 2018, and 2019. The analysis highlighted a low demand phase in 2017, possibly linked to reduced industrial activity or external economic factors at the Ghana Volta Aluminium Company. A significant demand uptick was observed between June and November 2018, and February to October 2019, potentially reflecting increased production levels, operational expansions, or favorable economic conditions. Weather conditions could also have contributed to the demand variations. The insights from this analysis are pivotal for understanding the multifaceted factors influencing energy demand, thereby enriching the foundation for the ensuing predictive modeling and load forecasting tasks of this project.

Valco Demand across time



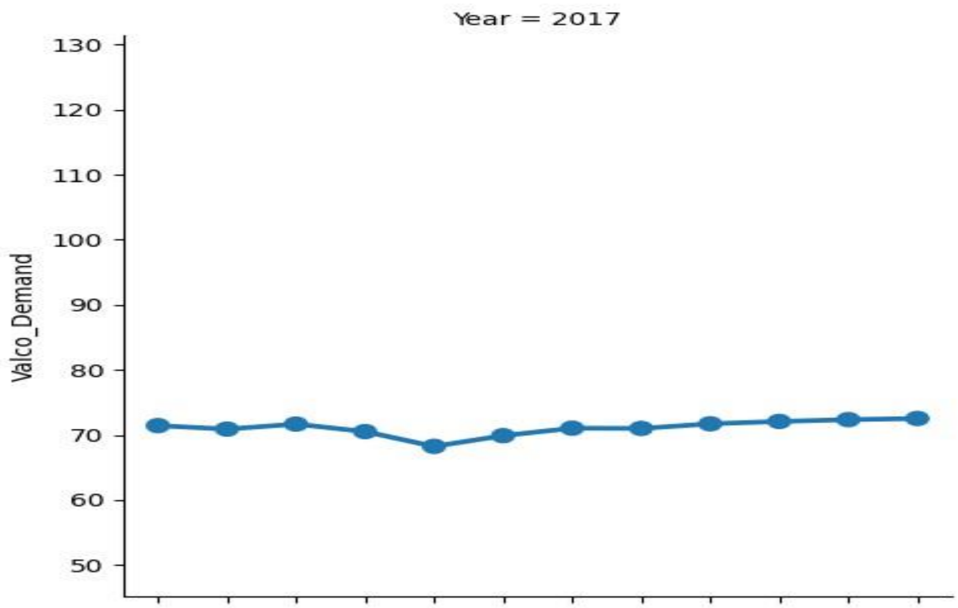
In this section, a comprehensive analysis was carried out to understand the hourly demand trends over the years and to inspect the demand behavior on a specific holiday, 6th March, which is celebrated in Ghana. A series of plots were generated to represent the demand across each hour of the day, over the years, with a red vertical line indicating the demand on 6th March 2019 at each respective hour.

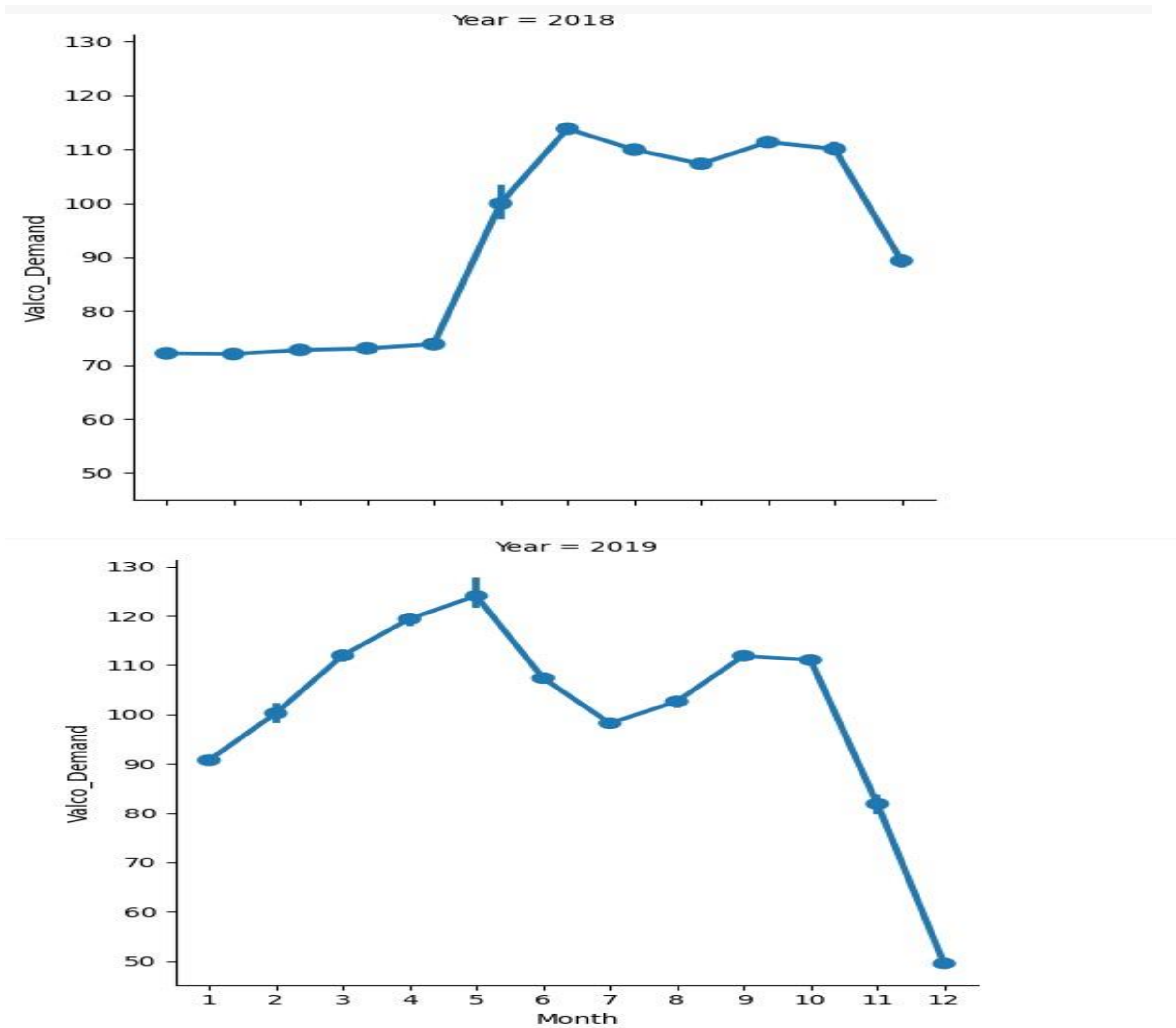
The resultant visualizations revealed a notable increase in demand around the red line across each time slot. This surge in demand on 6th March 2019 could be attributed to several factors, which will captured in our future work. The analysis provides crucial insights into how special days or holidays might influence energy demand at different hours of the day within the Ghana Volta

Aluminium Company. Understanding such demand behavior is instrumental in devising more accurate and nuanced load forecasting models. It accentuates the necessity to incorporate temporal and special event considerations into the predictive modeling framework to better capture the dynamics of energy demand.

Moreover, this analysis underscores the importance of considering external factors, such as holidays and other special events, in load forecasting endeavors, thus enriching the analytical perspective for better-informed decision-making and more accurate demand prediction in subsequent project phases.

Valco Demand trend over the months



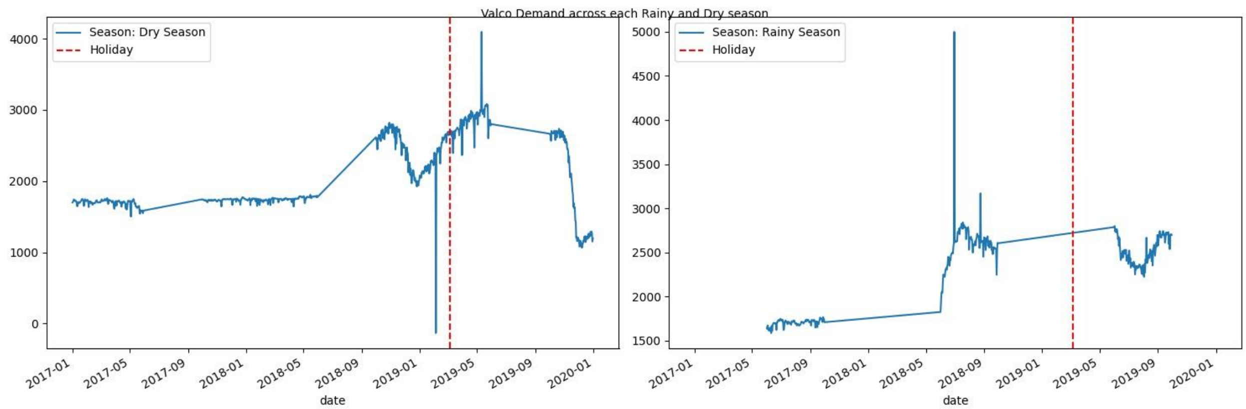


Categorical plotting technique was utilized to delineate the monthly demand trends for the years 2017, 2018, and 2019 at the Volta Aluminium Company. The objective was to visually represent and compare the monthly demand patterns across these years, thereby providing a comprehensive perspective on how demand dynamics have evolved.

The analysis yielded distinct plots for each year, plotting the monthly demand against each month of the year. The trends observed were clear and discernible, making it straightforward to understand the demand variations over the months for each respective year. This analysis is pivotal as it unveils the seasonal and monthly demand fluctuations, providing critical insights that will be invaluable in tailoring the predictive models to accurately forecast the load demand based on time variables. Understanding such temporal demand variations is crucial for devising robust forecasting models that can cater to the dynamic demand patterns observed in the data.

Analysis of Seasonal Variations on Demand Over Time

In this section, the focus was on analyzing the impact of Ghana's distinctive seasonal variations - the dry season and rainy season - on the energy demand at the Ghana Volta Aluminium Company. These seasons here in Ghana are known for their contrasting weather characteristics which could potentially influence energy consumption patterns in significant ways.

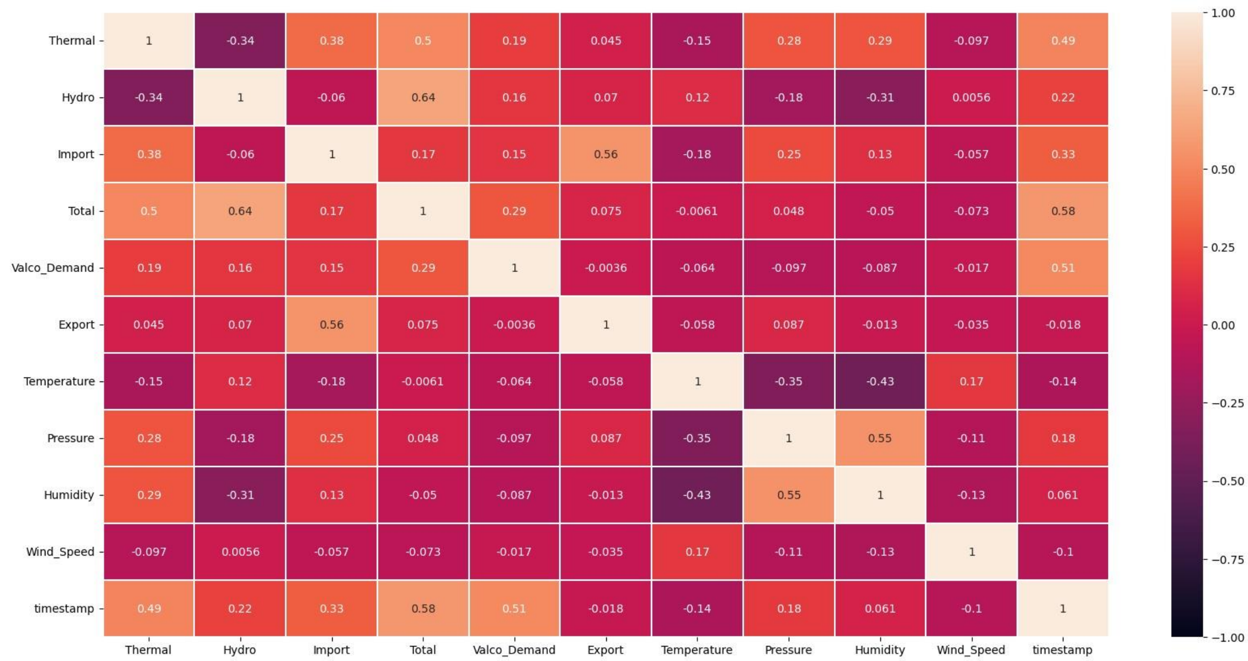


Again**,** this unique visualization strategy was employed to analyze the impact of seasonal variations and national holidays on energy demand at the Ghana Volta Aluminium Company. A red line, representing 6th May 2019, a national holiday, was integrated into the visualization, offering a clear demarcation for comparing demand on this day against other days across different seasons.

When readers delve into this visualization, they will find it engaging and easy to comprehend how the demand shifts during different seasons, and how a national holiday could potentially interrupt the typical demand pattern. The clarity and simplicity of this visualization make the trends easily interpretable, allowing readers and power plant operators to understand Valco Demand for each season across these years.

4.3 Correlation Analysis

A correlation matrix is constructed to investigate the relationships between load demand and other explanatory variables. This section discusses the identified strong correlations and their implications for load forecasting.



In this, the correlation matrix analysis is carried out to elucidate the linear relationships between various variables and the energy demand at Valco, which is a critical step for enhancing the effectiveness of the forecasting models. The correlation matrix serves as a robust tool for identifying and quantifying the interactions between these variables and Valco demand, establishing a foundational basis for effective predictive modeling. This analysis significantly aids in the process of feature selection, allowing for the identification of key variables that have a substantial impact on Valco demand, thereby improving the model's performance while reducing the necessity for computational resources. Furthermore, the correlation matrix is instrumental in diagnosing issues of multicollinearity which, if unaddressed, could distort the predictive models. By identifying highly correlated independent variables, necessary adjustments can be made to mitigate the issue, thereby enhancing the reliability and accuracy of the forecasting models.

4.4 Discussion

Model Training

Multiple machine learning models were trained to forecast the energy demand for the Ghana Volta Aluminium Company. This diverse approach, using models like Linear Regression, XGBoost, and Catboost, among others, was chosen to increase the chance of finding the best-fitting model, as each captures different data patterns. The data was first pre-processed, and then each model was trained. This meticulous process aimed to enhance prediction reliability, crucial for the company's energy management and planning. The methodology underscores the importance of evaluating multiple models to ensure optimal energy forecasting results.

Model Performance and Selection

In the process of predicting the energy demand for the Volta Aluminium Company in Ghana, a range of machine learning models were employed as discussed earlier and their performance was evaluated based on the Mean Square Error (MSE). The MSE is a significant metric as it quantifies the average of the squares of the errors, thereby indicating the magnitude of error a model is likely to make in its predictions. Below is the discussion on the performance of each model, ordered from the highest to the lowest MSE.

- Support Vector Machine (MSE: 437.61): The Support Vector Machine model exhibited the highest MSE among all models, indicating its struggle to model the underlying patterns in the data accurately.
- Lasso (MSE: 340.91): Lasso regression, known for its feature selection properties, came next with a relatively high MSE, suggesting that the regularization term might have omitted essential features or the linear assumptions were violated.
- Ridge (MSE: 326.71) and Linear Regression (MSE: 326.05): Ridge regression and Linear Regression displayed similar performance, reflecting the challenge linear models faced in capturing the intricacies of the energy demand data.
- AdaBoost (MSE: 283.04): The AdaBoost model performed better than the aforementioned models, showcasing the benefits of boosting; however, it still had a considerably high MSE.
- Decision Tree (MSE: 95.35): The Decision Tree model significantly improved the MSE, underlining its capacity to capture non-linear relationships in the data.
- LightGBM (MSE: 69.64), Gradient Boosting (MSE: 67.27), and Catboost (MSE: 65.43): These gradient boosting frameworks demonstrated a marked improvement in reducing the MSE, emphasizing the strengths of ensemble learning and gradient boosting techniques in demand forecasting.

- Random Tree (MSE: 62.15): Random Tree further improved the MSE, reflecting the effectiveness of bagging and ensemble learning in minimizing prediction errors.
- XGBoost (MSE: 57.18): Outshining all the other models, XGBoost achieved the lowest MSE, showcasing its robustness and efficacy in forecasting energy demand. The optimized gradient boosting framework of XGBoost effectively harnessed the power of ensemble learning to provide the most accurate predictions among all models evaluated.

	Model	Mean squared Error
0	Linear Regression	326.046089
1	Ridge	326.711675
2	Decision Tree	95.354215
3	Lasso	340.913570
4	Random Tree	62.147024
5	Gradient Boosting	67.266743
6	AdaBoost	283.035974
7	Support Vector Machine	437.610565
8	XGBoost	57.181542
9	LghtGBM	69.640861
10	Catboost	65.431207

Table 4.1 model summary

The elucidation of the models' performance based on MSE affords a meticulous understanding of their predictive prowess. It also underscores the paramountcy of selecting the most accurate model, in this case, XGBoost, to ensure precise and reliable energy demand forecasts.

How XGBoost Works

XGBoost is a widely-used machine learning algorithm for regression and classification tasks. It builds decision trees in sequence, with each tree correcting its predecessor's errors through a method called boosting. Key features of XGBoost include its in-built regularization to prevent overfitting, ability to handle missing data without imputation, and its design for efficient hardware utilization, supporting parallel and distributed computing. By continuously learning from prior errors and refining its predictions, XGBoost offers high performance and flexibility across various data types and tasks, making it a top choice for practitioners seeking accurate predictive models.

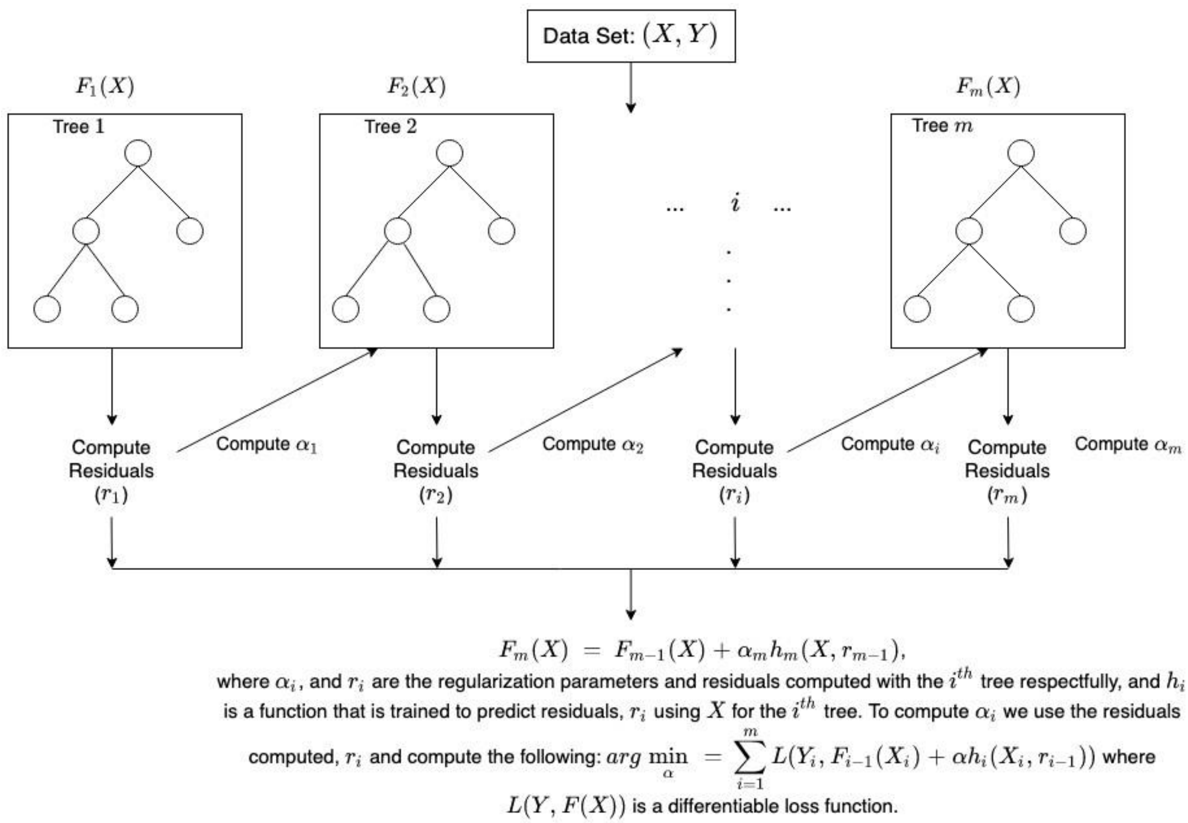


Figure 4.1 XGBoost tree

Live performance of XGBoost on our train data

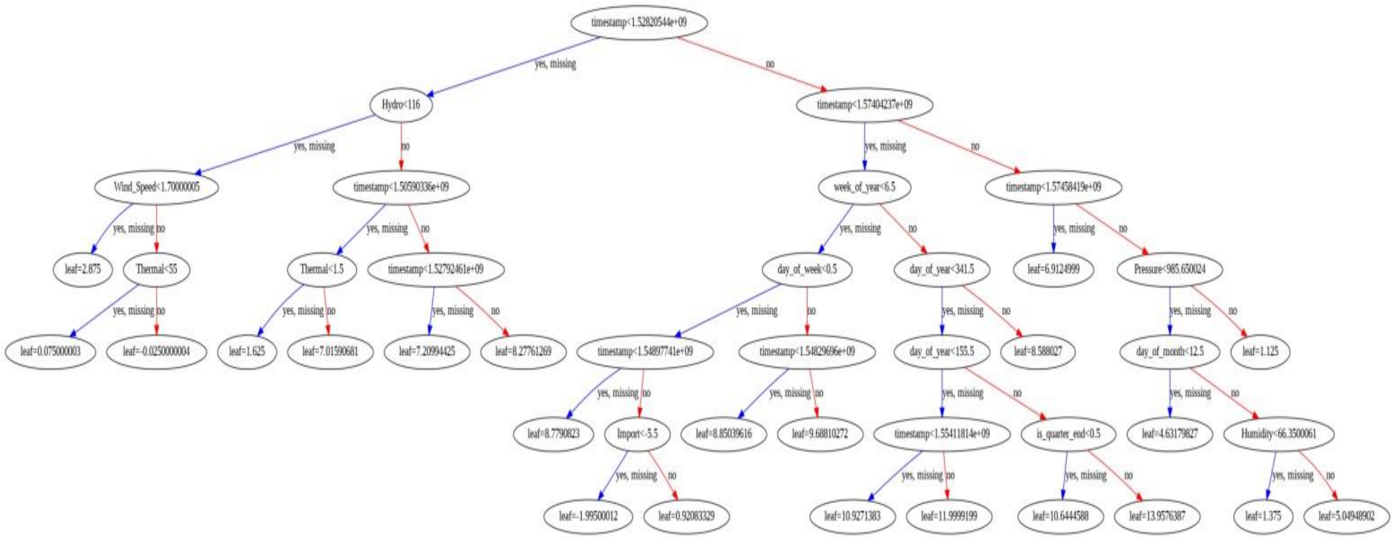


Figure 4.2 XGBoost Live performance

Hyper Tuning of XGBoost

Hyperparameter tuning was a crucial step to optimize the model's performance. For our XGBoost model concerning the load forecasting for Volta Aluminium Company in Ghana, a systematic approach was employed to fine-tune the hyperparameters. Specifically, Randomized Search was utilized due to its efficiency and effectiveness in navigating the hyperparameter space.

The following hyperparameters were considered in the tuning process:

n_estimators: This parameter defines the number of boosting rounds or trees to be built, which directly affects the model's complexity and accuracy.

learning_rate: It is the step size shrinkage used to prevent overfitting, and it's crucial for the model's performance and speed.

max_depth: This parameter specifies the maximum depth of the trees, controlling the model's complexity.

Subsample: It's the fraction of the training data sampled for each boosting round, which helps in preventing overfitting.

colsample_bytree: This parameter defines the fraction of features to choose for each boosting round, also helping to prevent overfitting.

Hyperparameters	Range of values to search from
n_estimators	range(50, 200, 10)
Learning_rate	[0.01, 0.1, 0.2, 0.3]
Max_depth	range(3, 10)
Subsample	[0.8, 0.9, 1.0]
Colsample_bytree	[0.6, 0.7, 0.8, 0.9, 1.0]

Table 4.2 Hyperparameters

Utilizing Randomized Search for hyperparameter tuning, different combinations of these parameters were randomly selected and tested. The performance of the model for each combination was evaluated using Mean Absolute Error (MAE) as the scoring metric as explained earlier due to its interpretability and robustness against outliers. This systematic exploration helped in identifying the optimal set of hyperparameters that minimized the MAE, leading to a more accurate and robust model for forecasting the load demand.

These were the selected parameters after the Random search;

Best Parameters: {'subsample': 0.9, 'n_estimators': 110, 'max_depth': 4, 'learning_rate': 0.2, 'colsample_bytree': 1.0}

Model Evaluation and Prediction

Actual vs Predicted value on Perfect Fit

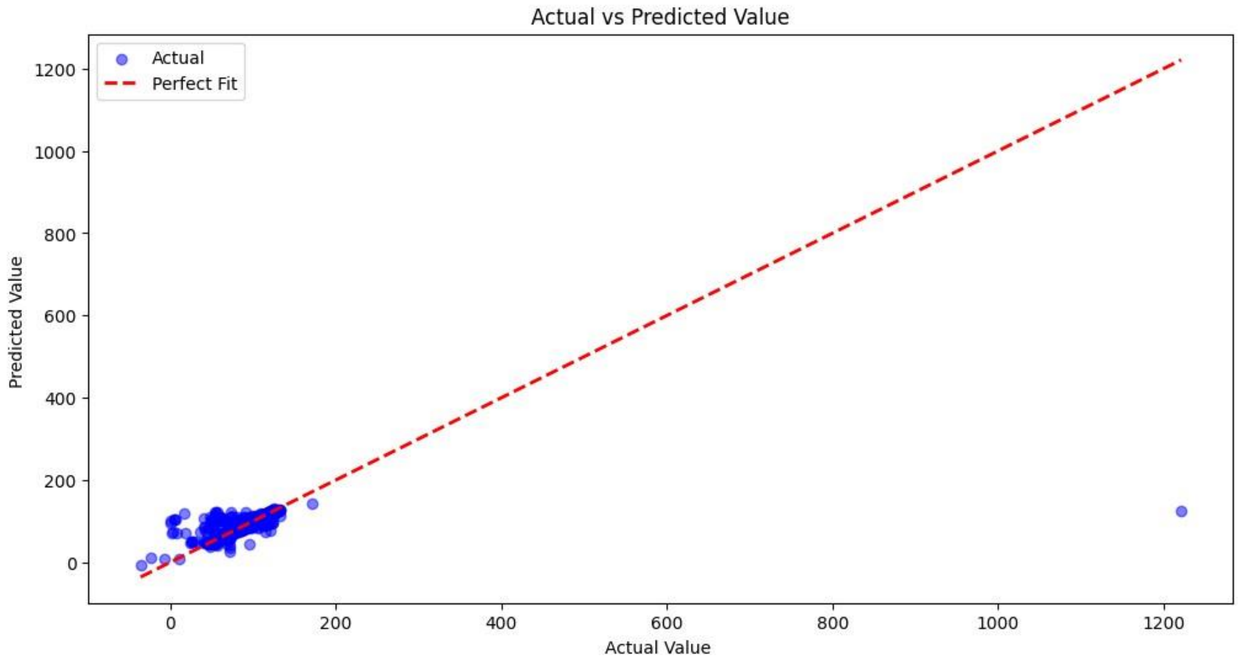


Figure 4.3 Model Evaluation best fit

In assessing the XGBoostRegressor model's performance, a scatter plot was used, plotting actual values on the x-axis against predicted values on the y-axis. A red dashed diagonal line, termed 'Perfect Fit,' was added as a benchmark indicating where predictions would match actual values perfectly. The closer the data points are to this line, the better the model's accuracy. This visual method provides an intuitive grasp of the model's effectiveness in predicting the demand for the Volta Aluminium Company in Ghana, emphasizing the importance of visual tools in understanding complex data results.

Actual vs predicted value on Data Point

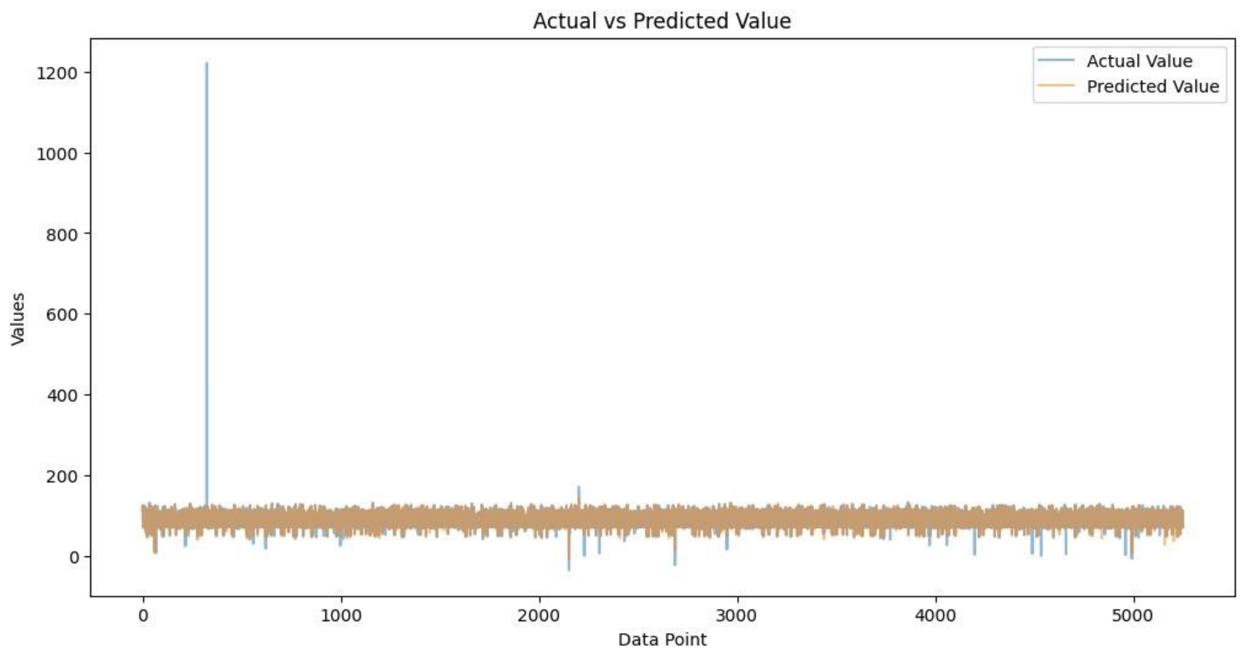


Figure 4.4 Model Evaluation and Prediction

This line plot was used to visually compare the actual and predicted values of the XGBoostRegressor model. This graph represent the true demand and 'Predicted Value' illustrating the model's forecasts. The x-axis lists the data points, and the y-axis shows the demand values. Using this translucent lines helps discern overlaps and differences between the two lines. This graph just aids in easily understanding the model's performance. Such visual representation is crucial for intuitively assessing discrepancies and alignments, highlighting the model's accuracy, and communicating the results to a wider audience.

Testing Model with Date

A thorough backtesting strategy was used to validate the XGBoost model's accuracy over various time periods. The approach involves training the model on certain data segments and testing it on withheld segments, mimicking real-world forecasting scenarios. This was done over three 15-day periods, with each testing phase rolling into the next training phase. The backtests started from January 9, 2017, and extended to August 15, 2019. This method assessed the model's adaptability to new data and its response to potential shifts in data trends. Such rigorous validation bolsters the model's credibility, confirming the XGBoost's consistent performance across the tests.

Performance After Backtesting

Evaluation on eval data
RMSE: 10.00
R2(R-squared score): 0.85

Feature Importance

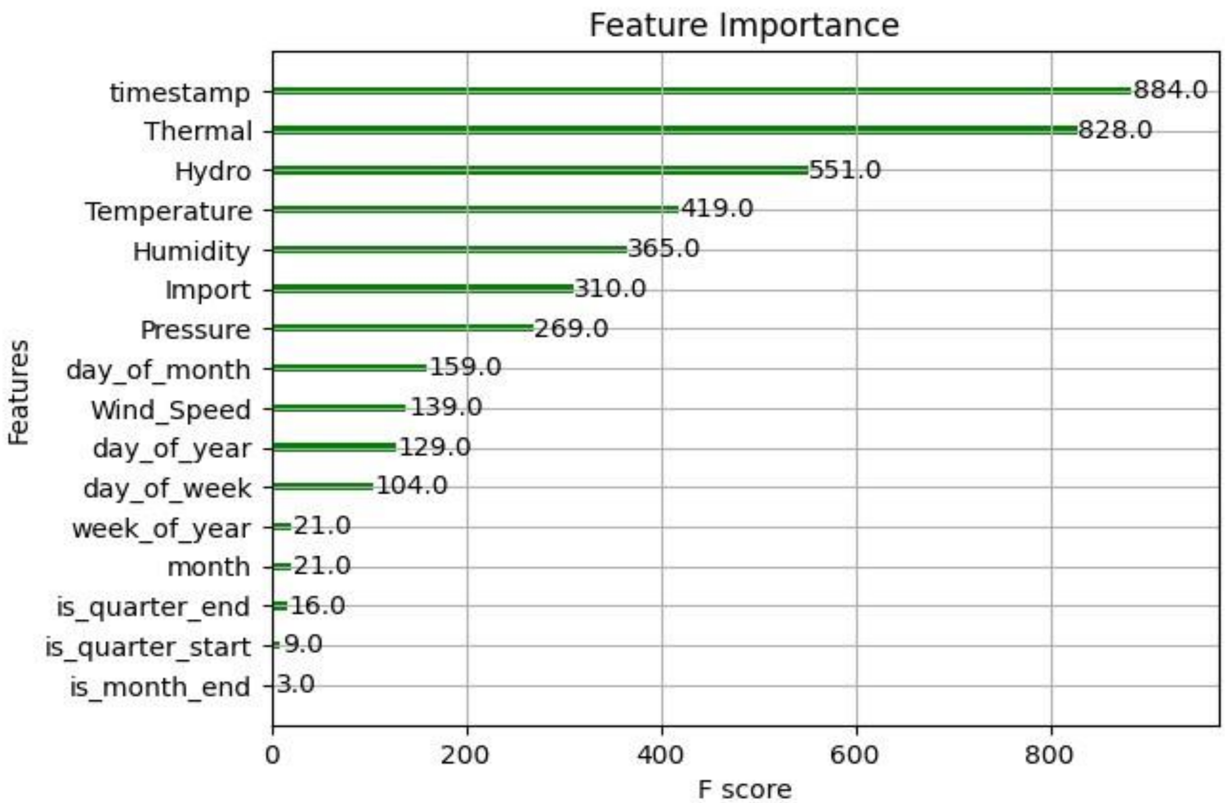


Figure 4.5 Feature Importance

We wanted to see which features had the most impact on our prediction which will aids in explaining the model's decision, especially in regulated problems. Feature importannce helps in feature selection, ensuring only the most influential features are used, which can lead to more efficient and accurate model. Understanding which variables heavily impact the target variable is essential for our decision-making.

Model Interpretabilty

SHAP (SHapley Additive exPlanations) is a powerful tool for interpreting machine learning models, and it provides a measure of the impact of each feature on the prediction for individual data points. When utilizing SHAP for model interpretability, several visualizations and metrics are introduced, including the base value, which serves as a reference point for the explanations.

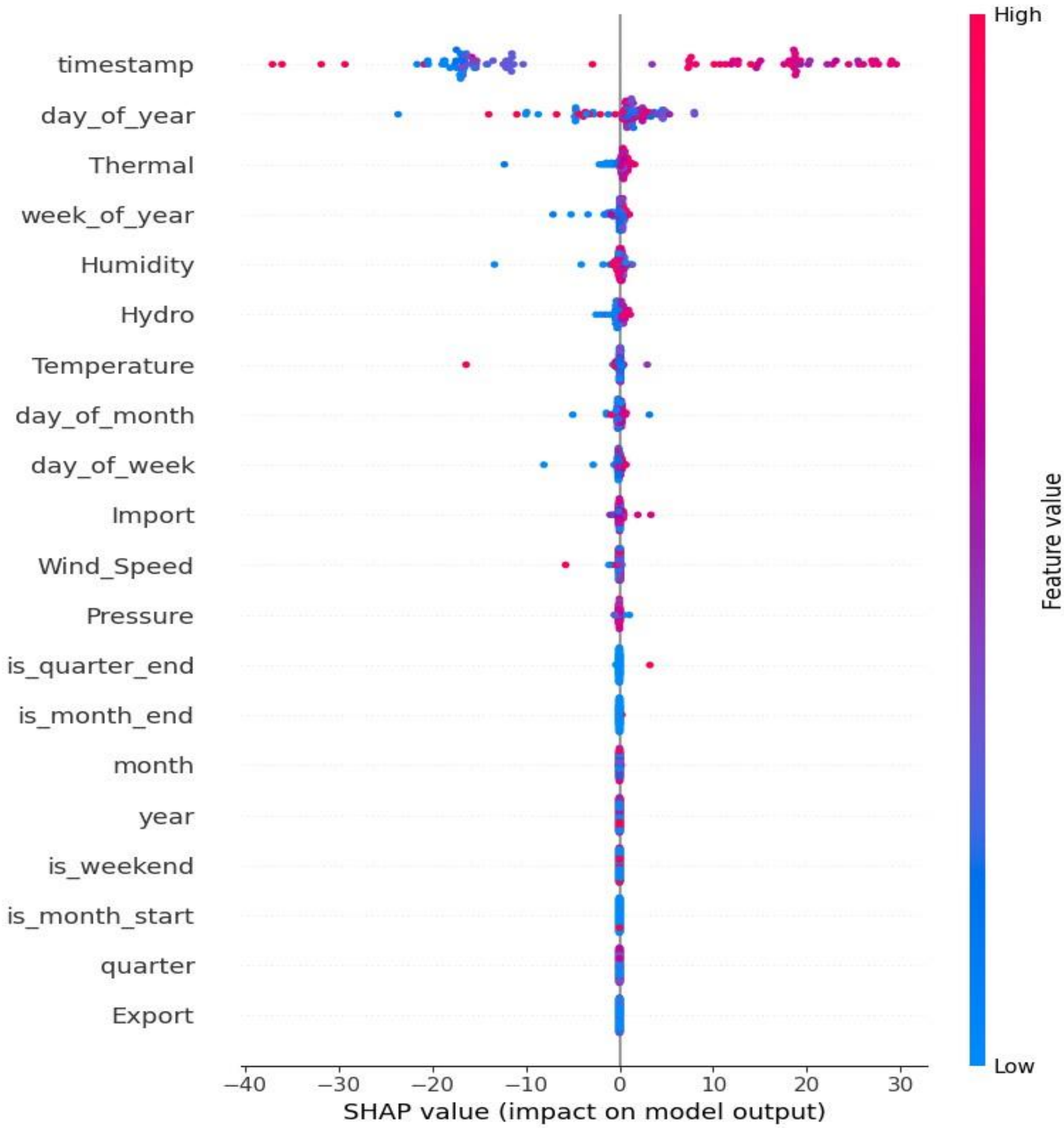
Base Value

The base value in SHAP is the expected value of the model prediction over the entire dataset. In simple terms, it's the value that would be predicted if we did not know any features for the current output. It's the point where the SHAP value explanations start, representing the average prediction of the model across all data points

.

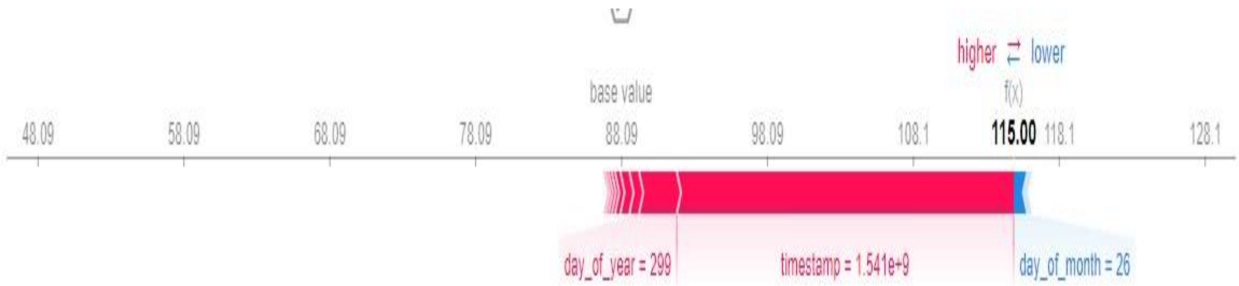
1. Summary Plot:

The Summary Plot encapsulates feature importance and feature effects in one view. The plot displays all the features ranked by their importance (the total absolute SHAP value) across all samples. It's an amalgam of information where each dot represents a SHAP value for a feature and a sample, thus showing us the distribution of the impacts each feature has on the model output.



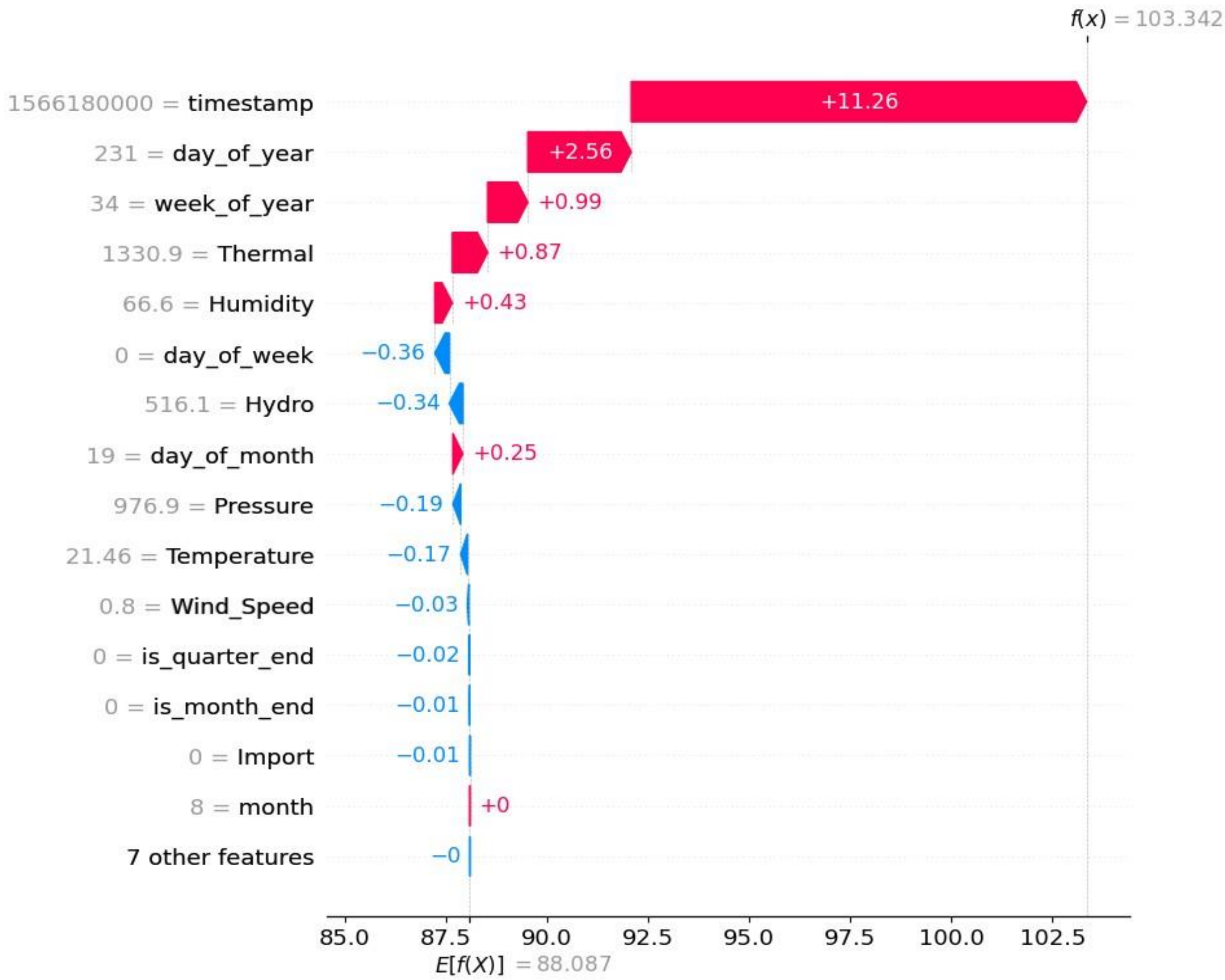
2. Force Plot:

The Force Plot illustrates the influence of different features on a single prediction. Each feature value pushes the prediction higher or lower, away or towards the base value. The visualization enabled us to understand the driving forces behind a particular prediction by showing the positive or negative contribution of each feature.



3. Waterfall Plot:

Similar to the Force Plot but represented in a waterfall format, this plot visually breaks down individual predictions to understand the contribution of each feature, starting from the base value. It sequentially adds the effect of each feature, showing how the cumulative sum changes from the base value to the model output.



These visualizations help peel layers of model complexity, making it easier to understand and explain the model's predictions and behaviors. Understanding the contributions of different features towards predictions can provide vital insights, enabling better decision-making and trust in the model's outcomes. Furthermore, these plots help in identifying any unwanted biases or anomalies in the model, which could be critical in certain applications.

Model Deployment with Streamlit App

Deploying ouur machine learning model into a production environment, making it available to others to make predictions, is a critical phase of the project lifecycle. In our case, we utilized Streamlit, a fast and efficient way to create interactive web applications around machine learning models, to provide a user interface for plant operators to input data and receive predictions on demand for a particular date. Here's a structured narration of the deployment and integration process:

Model Export:

After training and validating our XGBoost model, we exported it into a format suitable for deployment, such as a Pickle file , ensuring that it could be loaded into the Streamlit app with all its trained parameters intact.

Streamlit App Development:

Leveraging Streamlit, we created an interactive web application. we designed an intuitive user interface where plant operators could easily input the required data, such as temperature, date, and other relevant parameters. Streamlit's simplistic scripting nature allowed for a seamless integration between the UI elements and the backend logic of the application.

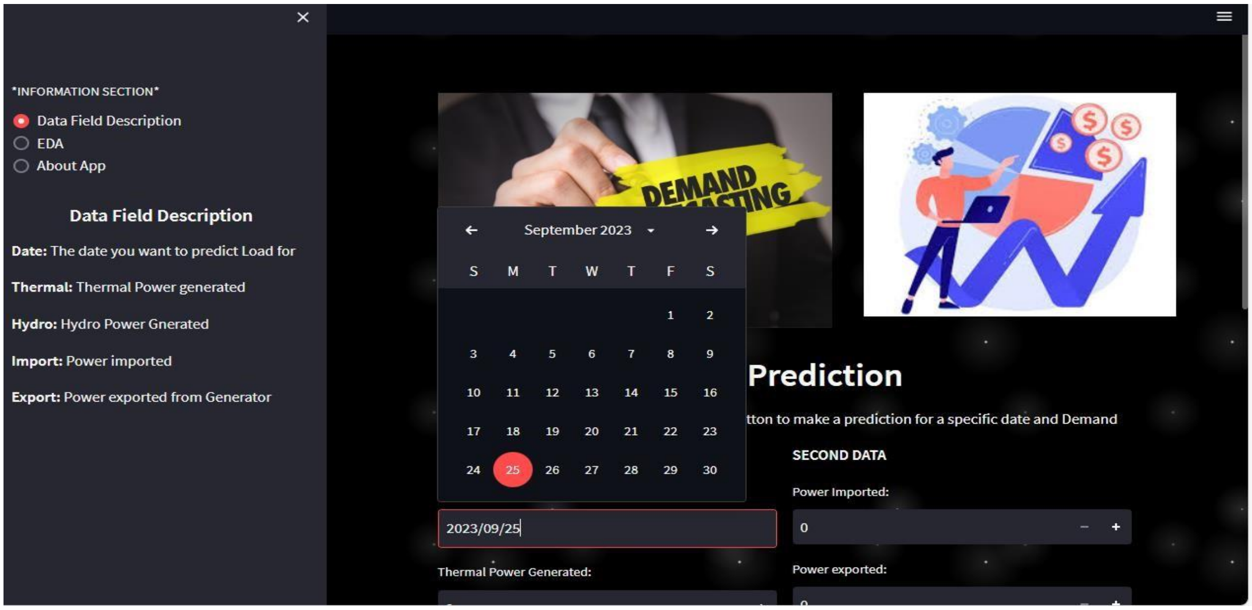
Model Integration:

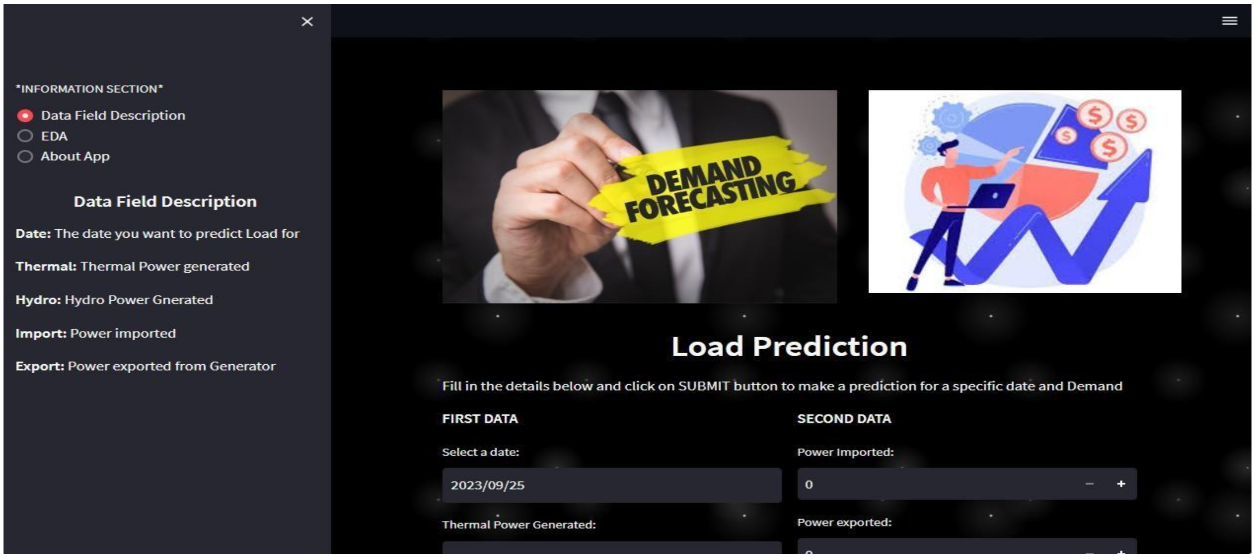
Within the Streamlit app, we integrated the exported model, ensuring that it could access the input data provided by the user. The app was set up to load the model into memory upon initialization, ready to make predictions on demand.

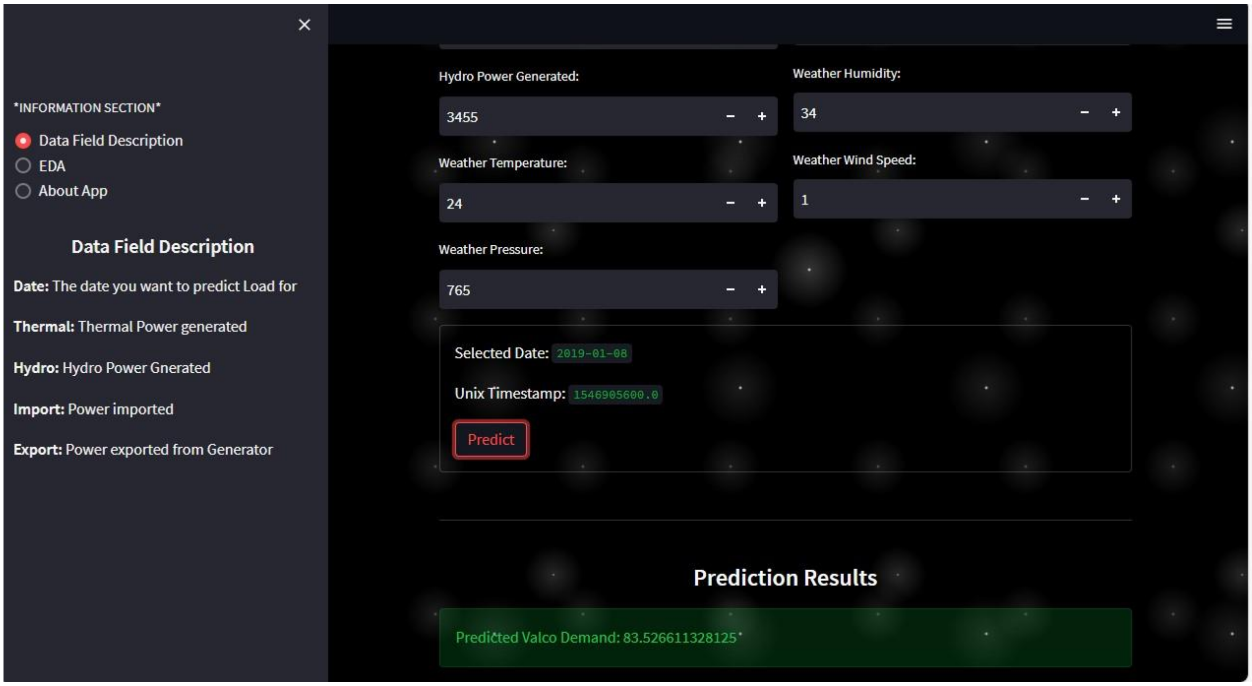
Prediction and Response:

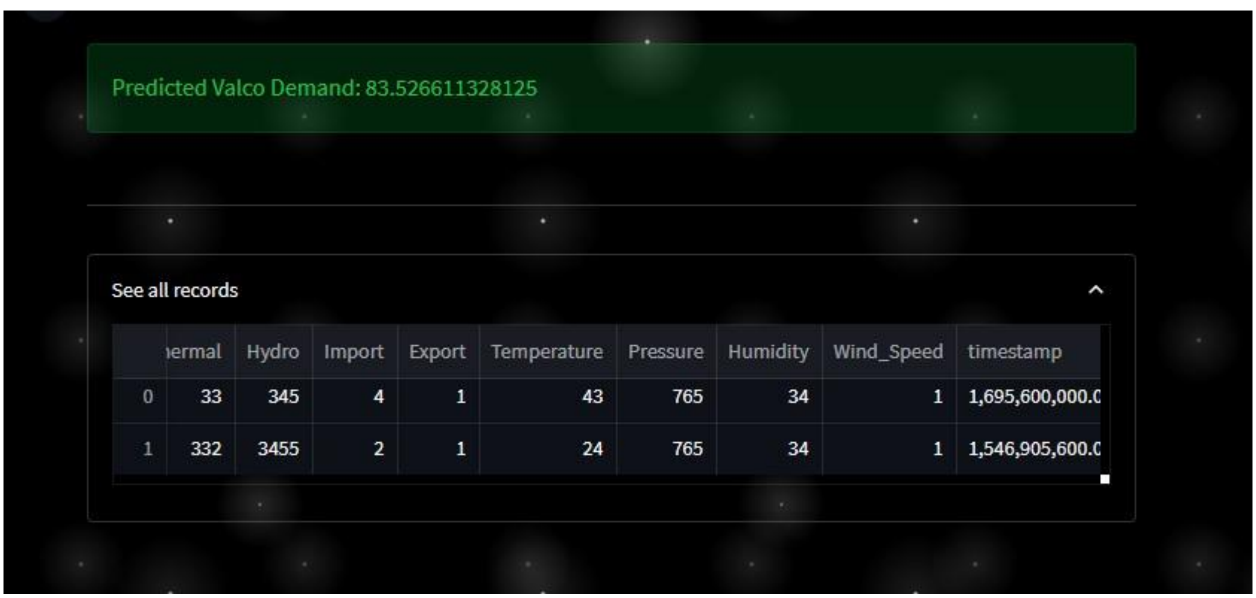
Upon receiving input from the operators via the UI, the app processes the data, feeding it to the loaded model, and obtaining the predicted demand. The predictions were then beautifully displayed on the Streamlit app, giving immediate feedback to the plant operators on the expected demand for the specified date.

Interfaces


INFORMATION SECTION
Data Field Description
EDA
About App
Data Field Description
Date: The date you want to predict Load for
Thermal: Thermal Power generated
Hydro: Hydro Power Gnerated
Import: Power imported
Export: Power exported from Generator
DEMAND
September 2023
S M T W T F S
1 2
3 4 5 6 7 8 9
10 11 12 13 14 15 16
17 18 19 20 21 22 23
24 25 26 27 28 29 30
Prediction
tton to make a prediction for a specific date and Demand
SECOND DATA
Power Imported:
2023/09/25
0
Thermal Power Generated:
Power exported:


INFORMATION SECTION
Data Field Description
EDA
About App
Data Field Description
Date: The date you want to predict Load for
Thermal: Thermal Power generated
Hydro: Hydro Power Gnerated
Import: Power imported
Export: Power exported from Generator
DEMAND FORECASTING
Load Prediction
Fill in the details below and click on SUBMIT button to make a prediction for a specific date and Demand
FIRST DATA
SECOND DATA
Select a date:
Power Imported:
2023/09/25
0
Thermal Power Generated:
Power exported:


INFORMATION SECTION
Data Field Description
EDA
About App
Data Field Description
Date: The date you want to predict Load for
Thermal: Thermal Power generated
Hydro: Hydro Power Gnerated
Import: Power imported
Export: Power exported from Generator
Hydro Power Generated:
3455
Weather Humidity:
34
Weather Temperature:
24
Weather Wind Speed:
1
Weather Pressure:
765
Selected Date: 2019-01-08
Unix Timestamp: 1546905600.0
Predict
Prediction Results
Predicted Valco Demand: 83.526611328125


Predicted Valco Demand: 83.526611328125
See all records
ıermal Hydro Import Export Temperature Pressure Humidity Wind_Speed timestamp
0 33 345 4 1 43 765 34 1 1,695,600,000.0
1 332 3455 2 1 24 765 34 1 1,546,905,600.0

CHAPTER FIVE

CONCLUSION AND RECOMMENDATION

5.1 Conclusion

In this project, we used the XGBoost algorithm to accurately forecast Ghana VALCO's electrical demand and integrated the model into a user-friendly app for real-time decision support. This initiative will empower VALCO to efficiently adapt to changing energy needs, optimize operations, and save costs. The system's adaptability ensures resilience in fluctuating operational environments. This endeavor not only boosts VALCO's current operational prowess but also opens doors for future enhancements, reinforcing the blend of machine learning and practical application. Ultimately, this project underscores the transformative role of machine learning in industrial load forecasting, positioning VALCO for continued excellence in a dynamic energy sector.

5.2 Recommendation

Based on the findings and insights obtained from this research on electrical load forecasting using machine learning and the analysis of VALCO data, some key recommendations are proposed for further research and practical applications:

Enhanced Data Collection: To improve the accuracy of load forecasting models, future research should focus on enhancing data collection efforts. This includes obtaining more granular data related to weather conditions, holidays, and other external factors that influence electrical load patterns. Additionally, exploring real-time data integration can further enhance forecasting precision.

Advanced Machine Learning Techniques: Investigate the utilization of advanced machine learning techniques, such as deep learning models (e.g., LSTM or CNN), ensemble methods, and hybrid models, to potentially achieve even more accurate load forecasting results. Continuously staying abreast of advancements in the machine learning field is crucial.

Interdisciplinary Research: Encourage interdisciplinary research collaborations that involve experts from fields such as meteorology, economics, and policy analysis. Combining expertise from various domains can lead to more holistic and accurate load forecasting models

REFERENCES

[1] H. Malik, N. Fatema, and A. Iqbal, “Intelligent Data Analytics for TimeSeries Load Forecasting Using Fuzzy Reinforcement Learning (FRL),” in Intelligent data-analytics for condition monitoring: Smart Grid Applications, London, United Kingdom: Academic Press, 2021, pp. 193–213.

[2] W. He, “Load forecasting via Deep Neural Networks,” Procedia Computer Science, vol. 122, pp. 308–314, 2017.

[3] M. A. Hammad, B. Jereb, B. Rosi, and D. Dragan, “Methods and models for Electric Load Forecasting: A comprehensive review,” Logistics & Sustainable Transport, vol. 11, no. 1, pp. 51– 76, Feb. 2020.

[4] S. A.-h. Soliman and A. M. Al-Kandari, Electrical Load Forecasting: Modeling and Model Construction, 1st ed. Butterworth–Heineman, 2010.

[5] R. J. Hyndman and G. Athanasopoulos, Forecasting: Principles and Practice, 2nd ed. OTexts: Melbourne, Australia, 2018.

[6] N. Abu-Shikhah and F. Elkarmi, "Medium-Term Electric Load Forecasting Using Singular Value Decomposition," Energy Conversion and Management, vol. 36, no. 7, pp. 4259-4271, 2011.

[7] "What Is Machine Learning (ML)?", UCB-UMT, 2020. [Online]. Available: https://ischoolonline.berkeley.edu/blog/what-is-machinelearning/. [Accessed: 15- Sep- 2022].

[8] "Machine Learning Techniques", www.javatpoint.com, 2022. [Online]. Available: https://www.javatpoint.com/machine-learning-techniques. [Accessed: 14- Sep- 2022].

[9] B. Mahesh, "Machine Learning Algorithms -A Review", International Journal of Science and Research (IJSR), vol. 9, no. 1, pp. 381-386, 2020. Available: DOI: 10.21275/ART20203995 [Accessed 15 September 2022]

[10] M. Hayati and Y. Shirvany, “Artificial neural network ap-proach forshort term load forecasting for illam region,” World Academy of Science, Engineering and Technology, vol. 28, pp.280–284, 2007.

[11] G. Zhang, B. E. Patuwo, and M. Y. Hu, “Forecasting with artificial neural networks: The state of the art,” International journal of forecasting, vol. 14, no. 1, pp. 35–62, 1998.

[12] N. Kandil, R. Wamkeue, M. Saad, and S. Georges, “An efficient approach for short term load forecasting using artificial neural networks,” International Journal of Electrical Power& Energy Systems, vol. 28, no. 8, pp. 525–530, 2006.

[13] D. C. Park, M. El-Sharkawi, R. Marks, L. Atlas, and M. Damborg, “Electric load forecasting using an artificial neural network,” IEEE transactions on Power Systems, vol. 6,no. 2, pp. 442– 449, 1991.

[14] kumar singh, Ibraheem, S. Khatoon, M. Muazzam, and D. Chaturvedi, “Load forecasting techniques and methodologies: A review,” in 2012 2nd International Conference on Power, Control and Embedded Systems, 2013.

[15] Z. Wang, J. Li, S. Zhu, J. Zhao, S. Deng, S. Zhong, H. Yin, H. Li, Y. Qi, and Z. Gan, “A review of load forecasting of the Distributed Energy System,” IOP Conference Series: Earth and Environmental Science, vol. 237, p. 042019, 2019.

[16] L. A. Zadeh, “Fuzzy sets,” Information and Control, vol. 8, no. 3, pp. 338–353, 1965.

[17] P. Mukhopadhyay, G. Mitra, S. Banerjee, and G. Mukherjee, “Electricity load forecasting using fuzzy logic: Short term load forecasting factoring weather parameter,” 2017 7th International Conference on Power Systems (ICPS), Jun. 2018.

[18] H. J. Sadaei, P. C. de Lima e Silva, F. G. Guimarães, and M. H. Lee, “Short-term load forecasting by using a combined method of convolutional neural networks and Fuzzy Time Series,” Energy, vol. 175, pp. 365–377, 2019.

[19] A. Lambora, K. Gupta, and K. Chopra, “Genetic algorithm- A literature review,” 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), Oct. 2019.

[20] “Applications of genetic algorithms to load forecasting problem,” Studies in Computational Intelligence, pp. 383–402, 2008.

[21] P. Ray, S. K. Panda, and D. P. Mishra, “Short-term load forecasting using genetic algorithm,” Advances in Intelligent Systems and Computing, pp. 863–872, 2018.

[22] C. Szegedy et al., "Going deeper with convolutions," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1-9, doi: 10.1109/CVPR.2015.7298594.

[23] S. Quaiyum, Y. I. Khan, S. Rahman, and P. Barman, “Artificial Neural Network based Short Term Load Forecasting of Power System,” International Journal of Computer Applications, vol. 30, no. 4, pp. 1–7, Sep. 2011.

[24] X. Liu, W. Zhang, S. Ye, and Y. Han, “Short-term load forecasting approach with SVM and similar days based on United Data Mining Technology,” 2019 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia), Oct. 2019

[25] https://arxiv.org/pdf/1603.02754.pdf XGBoost: A Scalable Tree Boosting System