Short-Term Electric Load Forecasting Using Machine Learning

by

Zihadul Haque Talukdar

Roll: 1713023

A thesis report submitted in partial fulfillment of the requirements for the degree of Bachelor of Energy Science and Engineering

Department of Energy Science and Engineering

Khulna University of Engineering & Technology

Khulna-9203, Bangladesh

February, 2023

Short-Term Electric Load Forecasting Using Machine Learning

by

Zihadul Haque Talukdar

Roll: 1713023

Supervisor: **Dr. Naruttam Kumar Roy**

A thesis report submitted in partial fulfillment of the requirements for the degree of Bachelor of Energy Science and Engineering

Department of Energy Science and Engineering

Khulna University of Engineering & Technology

Khulna-9203, Bangladesh

February, 2023

Declaration

This is to certify that the project and thesis work entitled *"**Short-Term Electric Load Forecasting Using Machine Learning**"* has been carried out by *Zihadul Haque Talukdar, Roll No. 1713023* in the Department of Energy Science and Engineering, Khulna University of Engineering & Technology, Khulna, Bangladesh. The above thesis work or any part of this work has not been submitted anywhere for the award of any degree or diploma.

Dr. Naruttam Kumar Roy
Professor
Department of Electrical and Electronic Engineering

Zihadul Haque Talukdar
Roll: 1713023

Approval

This is to certify that the project and thesis work submitted by *Zihadul Haque Talukdar*, *Roll No. 1713023* entitled *"**Short-Term Electric Load Forecasting Using Machine Learning**"* has been approved by the board of examiners for the partial fulfillment of the requirements for the degree of B.Sc. in the Department of Energy Science and Engineering, Khulna University of Engineering & Technology, Khulna, Bangladesh in February, 2023.

BOARD OF EXAMINERS

1. ____________________

Dr. Naruttam Kumar Roy (Supervisor)
Professor
Khulna University of Engineering & Technology
Khulna-9203, Bangladesh

2. __________________________ External Member

Dr. Md. Rafiqul Islam
Professor
Khulna University of Engineering & Technology
Khulna-9203, Bangladesh

Acknowledgment

First and foremost, I would want to express my heartfelt thanks to Allah for the opportunity to finish this thesis. I also would like to thank my supervisor, **Dr. Naruttam Kumar Roy**, Professor, Department of Electrical and Electronic Engineering, Khulna University of Engineering and Technology (KUET), Khulna-9203, Bangladesh, for his inspiration, support, and valuable suggestions during my thesis work. His academic advice, constructive comments, and cooperative attitude have always encouraged me to continue working on "***Short-Term Electric Load Forecasting Using Machine Learning*** ". I also express gratitude and profound respect to **Dr. Md. Rafiqul Islam,** Professor, Department of Electrical and Electronic Engineering, Khulna University of Engineering & Technology, Khulna.

Lastly, I want to express my deepest gratitude to my parents, elder brothers, family members, and all of my friends for their counsel, financial, and spiritual support. This study article would not be feasible without their contributions.

Zihadul Haque Talukdar

Abstract

One of the most important inputs for the planning committee of power plant units is accurate short-term load forecasting (STLF). STLF helps to lower the costs associated with producing energy in a power system by controlling excess power production. However, choosing accurate time series models to develop and use is a difficult task because it necessitates training a variety of models to determine which one is the best, as well as extensive feature extraction to derive important features and find the best time delays, which are frequently used input variables for time series forecasting. This study presents a collection of machine learning (ML) models to increase STLF accuracy. Linear regressioin model performance is not up to the mark. Random forest (RF) and Extreme Gradient Boosting (XGBoost) models performance better than the Artificial Neural Network (ANN) and the RMSE scores of these three models are 169.27, 173.47 and 176.93 respectively. The accuracy of these three models was improved by hyperparameter tuning and the optimized RMSE score is 163.09, 159.72 and 139. 62 for RF and XGBoost and ANN model respectively. ANN model surpasses all other model performance.

Contents

page

Title page……………………………………………………………………………i

Declaration ii

Approval iii

Acknowledgment..iv

Abstract..v

Contents vi

List of Tables viii

List of Figures..ix

Nomenclature xi

CHAPTER I 1

1 Introduction..1

1.1 Background and Motivation for load forecasting 1

1.2 Problem Statement and Research Question 4

1.3 Research objectives and contribution...5

1.4 Overview of thesis organization...5

CHAPTER II...6

2 Literature Review...6

Chapter III...12

3 Proposed Methodology 12

3.1 Data Collection and Dataset Analogy 12

3.2 Data Preprocessing..13

3.2.1 Filling of Missing Value 13

3.2.2 Handling Outliers...14

3.2.3 Dataset Scaling...16

3.3 Exploratory Data Analysis 17

3.3.1 Time Series Analysis 17

3.3.2 Trend Analysis 19

3.4 Feature Engineering 21

3.4.1 Feature Extraction and Creation 21

3.4.2 Feature Selection 23

3.5 Model Implementation and Evaluation 24

3.5.1 Data Splitting 24

3.6 Applied Algorithms 25

3.6.1 Linear Regression (LR) 25

3.6.2 Random Forest (RF) 26

3.6.3 Extreme Gradient Boosting (XGBoost) 27

3.6.4 Artificial Neural Network (ANN) 27

3.7 Performance Metrics for Evaluation 29

CHAPTER IV 30

4 Result & Discussion 30

4.1 Comparative analysis on different machine learning models 30

4.2 Evaluation of prediction accuracy 34

4.2.1 Hyper parameter tuning of selected model 36

Chapter V 39

5 Conclusion 39

5.1 Conclusion 39

5.2 Recommendation for Future Work 39

References 41

List of Tables

Table 1: Hyperparameter of machine learning model 35

Table 2: Performance metrics of different model 35

Table 3: Hyperparameter tuning result 36

Table 4: RMSE value comparison before and after hyperparameter tuning 37

List of Figures

Figure 3.1 Flowchart of forecasting methodology 12

Figure 3.2. Dataset Representation 13

Figure 3.3 Energy generation distribution before outlier handling 15

Figure 3.4 Energy generation distribution after outlier handling 15

Figure 3.5: Energy generation distribution on hourly basis throughout the day 18

Figure 3.6: Energy generation on monthly throughout the year 19

Figure 3.7 Annual Energy Distribution 20

Figure 3.8 Quarterly Electric load Distribution 20

Figure 3.9 Electric load demand by weekday vs weekend 21

Figure 3.10: Correlation matrix for (a) lag-related features (b) rolling window related feature 23

Figure 3.11: Data splitting into training set and testing set 25

Figure 3.12 Random Forest regression model. 26

Figure 3.13 Artificial Neural Network (ANN) model [39] 28

Figure 4.1: Two day ahead load prediction using linear regression 31

Figure 4.2: one week ahead load prediction using linear regression model 31

Figure 4.3: Two day ahead load prediction using Random Forest Regression model 32

Figure 4.4: one week ahead load prediction by using random forest regression model 32

Figure 4.5:two day ahead load prediction using Extreme Gradient Boosting (XGBoost) model 33

Figure 4.6 one week ahead load prediction by using Extreme Gradient Boosting (XGBoost) model 33

Figure 4.7: two day ahead load prediction using Artificial Neural Network (ANN) model 34

Figure 4.8: one week ahead load prediction using Artificial Neural Network (ANN) model......34

Figure 4.10 Mean Absolute Error (MAE) score for different models......36

Nomenclature

Symbol or Acronym	Description of Symbol or Acronym
LR	Linear Regression
STLF	Short-Term Load Forecasting
XGBoost	Extreme Gradient Boosting
RF	Random Forest
MAPE	Mean Absolute Percentage Error
SVR	Support Vector Regression
MSE	Mean Squared Error
RMSE	Root Mean Squared Error
MAE	Mean Absolute Error
NLDC	National Load Dispatch Center
PGCB	Power Grid Company of Bangladesh
STLF	Short-term Load Forecasting
MTLF	Medium-term Load Forecasting
LTLF	Long-term Load Forecasting
ANN	Artificial Neural Network
SVM	Support Vector Machine
MW	Mega Watt
SD	Standard Deviation
CV	Coefficient of Variation
MA	Moving Average

CHAPTER I

Introduction

1.1 Background and Motivation for load forecasting

Utility companies use load forecasting to simulate and predict power loads in order to balance market forces, reduce production costs, estimate realistic energy prices, manage scheduling, and plan for the future. The main factor used to separate prediction model is the forecasting horizon. Power demand forecasting was divided into three categories by the author of the reference [1]: short-term predictions that range from one hour to one week, medium-term forecasts that range from one week to one year, and long-term forecasts that span more than one year. The forecasting of short-term demand has received a lot of attention in the literature. Predictive forecasting is essential for managing the electricity system, unit commitment, economic dispatch, and the energy markets. On the other hand, although being a crucial component of budget allocation and power system planning, mid- and long-term forecasting has gotten less attention [2].

This research will look at several predicting horizons for the short future. In order to deal with the uncertainty of forecasting parameters, which always have a probability of occurring, long-term and mid-term forecasting need stochastic models. The expected horizons encounter two problems. In the short term, accuracy is crucial for the best daily operating effectiveness of electrical power distribution; in the long run, prediction stability is required for precise fuel supply scheduling and on-time maintenance operations. For medium-term prediction stability, a minimum forecast error should be maintained. The forecasting model should thus continue to function successfully in the longer term or, at the very least, should not be too sensitive to the passage of time.

Every economy relies on electricity to maintain highly technologically sophisticated industrialization [1], [2]. Practically every contemporary activity is dependent on power. Electric energy demand and use have increased substantially in recent years [3] but the process of creating, transferring, and transporting electrical energy remains difficult and expensive. As a result, effective grid management is critical in lowering the cost of energy generation and increasing capacity to satisfy the rising demand for electric energy [4].

As a consequence, successful grid management comprises appropriate load demand planning, an efficient control program for generating, transmission, and distribution system, and efficient load distribution along distribution system. As a consequence, accurate load forecasting will assist to increase the efficiency of the strategic planning in the power production sectors [4], [5]. Several computational and statistical methodologies have been utilized to construct prediction models it boost the accuracy of Power systems Energy Demand (EED) projections [6].

Correlation, extrapolation, and a combination of the two are three (3) separate EED forecasting techniques that may be combined together. Extrapolation techniques (Trend analysis) are used to simulate the growing trend by applying trend curves to the major historical data on electrical energy consumption [6], [7]. The future value of power demand is determined in this case using the estimated trend curve function for the chosen future point. Its results may sometimes be very accurate despite how simple it is [7].

In contrast, correlation methods (end-use and market processes) connect the system load to a variety of monetary and social variables [6], [7]. The processes, however, guarantee that the analysts can identify any connections between the patterns of load rise and other quantifiable factors. Nevertheless, the disadvantage is in the forecasting of economic and demographic difficulties, which is more difficult than the load estimate itself [6], [7]. Economic and demographic parameters, such as population, building permits, heating, employment, ventilation, knowledge of the air conditioning system, weather data, building structure, and firm, are often used in correlation strategies [6], [8]. Nevertheless, some academics divide engineering approaches—also referred to as correlation techniques—and data-driven (artificial intelligence) methods—which are associated with extrapolation techniques—into two groups (which are analogous to extrapolation techniques). Yet, no one approach is consistently recognized as being better in terms of science.

For many years, load forecasting has been a major subject of study in the electric power business. As electricity became more commonly utilized in the early twentieth century, electric utilities needed to estimate demand for energy in order to properly manage power supply and distribution. Originally, load forecasting was a straightforward procedure that relied on statistical approaches to assess past power demand data.

Load forecasting is a procedure used by power or energy-providing organizations to predict the power/energy necessary at a certain moment in the future to meet the market and economic balance. This is a crucial duty for power system operators and energy dealers as they attempt to offer a consistent and steady supply of electricity to clients while controlling energy production costs.

At present days, Load forecasting is appeared as the most challenging task due to the increased power demand and integration of renewable energy sources. Renewable energy sources are very unpredictable and irregular which brings uncertainties and fluctuation in the power system. The accuracy of load forecasting is reduced because of it only depends on historical load patterns that may not represent the current and future dynamics of the power system.

As a consequence, load forecasting has become a flourishing area of research and development, with a growing interest in the application of deep learning and machine learning techniques technologies. These strategies have shown promising results in improving the accuracy and durability of load forecasting models. They make it easier to integrate large and diverse information, such as real-time and near-real-time data that may represent the dynamic character of the power system. Moreover, machine learning and deep learning models may reveal complex and non-linear relationships between input and output variables, allowing for more reliable and accurate load estimates.

To recap, growing power consumption, the incorporation of renewable energy sources, and the necessity for dependable and precise load forecasting have prompted the use of machine learning and deep learning approaches in load forecasting advancement and research. These technologies have the capability to overcome the difficulties and complexities of load forecasting, allowing for more efficient and efficient power system operations.

1.2 Problem Statement and Research Question

The problem statement for load forecasting is to improve the accuracy of the prediction and reliability of handling uncertainties and variations in the power system. Widely used statistical models for load forecasting such as ARIMA and exponential smoothing can't capture non-linear and dynamic relationships between input features and target variables. Machine learning models can identify hidden patterns, complex and non-linear relationships between input features and output variables.

The following are the research inquiries that form the basis of this thesis:

- Which machine learning and deep learning models are best for load forecasting?
- How can machine learning and deep learning models take historical data, seasonal patterns and other relevant factors into account?
- How can performance and resilience of machine learning and deep learning models be evaluated and compared?
- What are the real-world impacts and prospective uses of ML and DL models in the power sector?

Short-term load forecasting (STLF) is a vital part of electric power system design and operation. The major goal of STLF is to guarantee that energy needs are supplied effectively and reliably while ensuring power grid stability. STLF models give estimates of the predicted load demand for a short-term period, often up to one week in advance. These predictions are crucial for the scheduling and dispatch of power producing units, which assist limit the cost of energy production and decrease the danger of power outages. Moreover, STLF enables grid operators anticipate and react to changes in load demand and generating capacity to maintain the balance between supply and demand and avoid blackouts and brownouts. Furthermore, STLF is an intriguing study field for data scientists and engineers since it needs the development of advanced algorithms and approaches to build precise and accurate forecasting models for load demand.

1.3 Research objectives and contribution

The following are the research objectives for load forecasting using machine learning and deep learning:

i. To examine the effect of various feature engineering and selection strategies on the performance of machine learning (ML) models for load forecasting.
ii. To implement different machine learning model
iii. To assess the accuracy and reliability of load forecasting using machine learning models.
iv. To evaluate the ML model by different metrics to find out the best model for forecasting load demand.

1.4 Overview of thesis organization

The thesis on load forecasting by machine learning and deep learning organized as follows:

Chapter 1: Introduction

The background, motivation, objective of the study, problem statement and research question of this topic are discussed in this section.

Chapter 2: Literature review

A brief history of literature studied on load forecasting, different method such statistical method, machine model and advance deep learning algorithm are discussed here.

Chapter 3: Methodology

Background of different machine learning and deep learning models will be described

Chapter 4: Result and Discussion

This section analyze different machine learning model performance and compare model performance to other model.

Chapter 5: Conclusion

Draws a conclusion from overall methodology, result and discussion section.

CHAPTER II

Literature Review

The use of short-term power demand forecasting meets a variety of requirements and has a broad range of applications. The load size varies significantly across study investigations, ranging from a single transformer to whole cities, regions, and nations. The forecasting horizon is another important difference in the study area, which may range from short-term projections of the next 900 seconds to larger horizons such as weekly forecasts. Forecast granularity varies between research projects, with some concentrating on 15 or 30-minute granularities and others on hourly predictions. Despite the variety of forecasting applications, the present study will focus on the technique, variables, algorithms, and assessment criteria employed since the forecast's effectiveness is dependent on the judgments made throughout these development phases.

A wide variety of approaches and algorithms have been used in the area of short-term load forecasting (STLF). The persistence technique, described in Reference [9], is the simplest strategy since it simply expects that today's load will be the same as tomorrow's. Modern deep learning methods are at the opposite extreme of the spectrum, as discussed in Reference [10].

The Box-Jenkins and Holt-Winters processes are heavily used in the time-series analysis, a method for forecasting that is often discussed. For instance, the reference [11] authors utilized these techniques to estimate the weekly demand for the Riyadh Electricity System in Saudi Arabia and discovered that they provide a mechanism to break down the anticipated electric load. For predicting the next 24 hours in Iran, the autoregressive integrated moving average (ARIMA) model was suggested in another research [12]. The conventional ARIMA model is enhanced by this modified ARIMA model, which incorporates temperature and load data with estimate. Nevertheless, utilizing the ARIMA model alone is computationally expensive and does not considerably improve prediction accuracy, demonstrating the necessity to include outside inputs to improve the outcomes.

Although machine learning techniques have been demonstrated to generate superior outcomes in recent research, there has recently been a trend away from utilizing traditional models for estimating power usage[13], [14]. In instance, the previous referenced work [15] indicated that the ARIMA model has a number of constraints for addressing the load forecasting issue. This study compared the performance of 6 traditional regression models and 2 deep learning models for day-ahead forecasting in Jiangsu province, China. This included the model's sole reliance on time-series data for forecasting, the challenge of figuring out the model's order through computation or trial and error, and the requirement for numerous iterations and analysis of autocorrelation and partial autocorrelation functions in order to fine-tune the model.

Machine learning methods offer the benefit over conventional statistical time-series models in that they can take a larger variety of important aspects into account, such the state of the weather, to increase the precision of short-term load forecasting (STLF). As seen in studies like Reference [16], where it was used to forecast the hourly weekly load in Thailand and achieved an average mean absolute percentage error (MAPE) of 7.71% for 250 testing weeks, multiple linear regression (MLR) has been a popular choice for STLF. This study highlights the importance of temperature for load prediction. Similar to how Reference [17] utilized MLR to anticipate power use, this study took into consideration meteorological factors including temperature, humidity, and daylight hours to predict electricity usage 24 hours in advance for 14 West African nations. MLR has been observed to perform badly for irregular load profiles, despite the speed of training and interpretability it provides.

On the other hand, artificial neural networks (ANNs) have also been widely used for STLF in recent decades due to their flexible algorithms. For example, Reference [18] proposed an ANN with a Levenberg-Marquardt training algorithm to forecast hourly, daily, and weekly loads in Ontario, Canada and achieved good results without comparing with other methods. Additionally, Reference [19] used ANN to predict the load of a single transformer using quarter-hour load records and weather data with hourly records and achieved a MAPE performance below 1% for both summer and winter seasons. In more recent research, Reference [20] applied STLF for urban smart grid systems in Australia and commented on the good generalization ability of ANN for the task. However, this approach also has its limitations, such as the tendency to quickly fall into local optimums, overfitting, and a relatively slow convergence rate. Despite these challenges, forecasting smart grid loads with an increasing number of renewable energy sources requires complex solutions and a robust approach in order to obtain good results.

The Support Vector Regression (SVR) model is a well-regarded method for short-term load forecasting. It is often preferred due to the linear relationship between the inputs and the forecast, as evidenced by research in references [15] and [21]. In reference [15], the authors used a linear kernel and achieved a Mean Absolute Percentage Error (MAPE) of under 2.6% for day-ahead predictions, outperforming other models such as multiple linear regression and multivariate adaptive regression splines. The authors of reference [21] found that SVR was a more efficient option for forecasting 48 hours of Portuguese electricity consumption, compared to the previously used Artificial Neural Network (ANN) as mentioned in reference [22]. The main advantage of SVR was its efficiency in hyperparameter tuning for daily online forecasting, which resulted in a MAPE between 1.9% and 4% for the first and second-day predictions. In reference [23], the authors compared SVR with ANN for forecasting South-Iranian day-ahead hourly load, and found that the nu-SVR variant improved upon SVR by changing the optimization problem and allowing the epsilon tube width to adapt to the data. The average MAPE was 2.95% for nu-SVR and 3.24% for ANN across all seasons.

The random forest (RF) ensemble technique is a method of combining multiple independent models to improve the overall forecasting ability. In a study referenced in [24], this technique was used to predict the daily hourly energy consumption in office buildings, leveraging environmental variables like temperature and humidity, as well as lagged load records. The use of RF resulted in a 6.11% mean absolute percentage error (MAPE) in the study.

Another research cited in [25] compared the performance of RF, ARIMA, Seasonal ARIMA (SARIMA), and extreme gradient boosting models for predicting the electrical usage in smart buildings (XGB). Although RF functioned well, the findings revealed that XGB surpassed the other approaches in terms of accuracy and execution time. Lastly, a research cited in [26] evaluated RF with XGB for next day load forecasting and came to the conclusion that XGB, an ensemble learning method, offers superior prediction accuracy, with a root mean squared error (RMSE) of 2.01 compared to 3.31 for RF.

To forecast the hourly weekly demand of a power plant, the authors of Reference [27] suggest utilizing XGB, which considers environmental factors and historical load. Nonetheless, they admit that the XGB's hyperparameter tweaking phase might be challenging. To solve this, they suggest using the Fireworks Algorithm to identify the global minimum in the hyperparameter space, which would increase the load forecast's precision.

Nevertheless, because to the unpredictable load profile of the holiday load (STLF), forecasting it might be difficult. The authors of Reference [28] claim that SVR, ANN, and deep learning are only a few examples of sophisticated STLF algorithms. These techniques do have certain drawbacks, however, such as SVR's sensitive to outliers, ANN's difficulties in determining the ideal number of hidden layers, and deep learning's reliance on huge datasets for efficient performance. The author reference [27]contend that since XGB is not constrained by these issues, it performs better than other approaches in solving STLF. These findings are supported by XGB, which beats Random Forest, SVR, ANN, and even XGB alone, in addition to averaging daily profile curves for comparable vacations.

The authors in references [29] and [30]suggest that training the XGB model on similar days can enhance the forecast. A comparison was done between the traditional XGB and the similar days XGB, which showed a noticeable improvement. This highlights that the accurate selection of similar days plays a significant role in short-term load forecasting (STLF).

According to reference [31], the authors suggested a hybrid approach that used K-means clustering with input from XGB feature significance findings to categorize related days. After the categorization, the data from the comparable days was divided into numerous intrinsic mode functions using the empirical mode approach. These intrinsic mode functions were then used to train several long short-term memory (LSTM) models. Subsequently, a time-series reconstruction using each LSTM model prediction was carried out. When compared to ARIMA, SVR, or a back-propagation neural network using the same comparable day strategy as the original input, this hybrid model incorporating LSTM performed better for STLF across 24 and 168 hour horizons.

A multi-step-ahead forecasting method employing XGB and SVR was suggested in reference [32] for predicting hourly heat demand. The terms "direct" and "recursive" were contrasted by the writers. The recursive technique employs a single model that iterates across the forecasting horizon, utilizing the previous forecasts as input for the next phase, as opposed to the direct method, which uses a separate model to anticipate each period on the forecasting horizon. The direct method's drawback is that it requires the training of several models, one for each interval on the horizon. On the other side, the recursive approach is vulnerable to prediction mistakes, which may mount as the forecasting horizon lengthens.

A decomposition-based technique was used in a research to assess the performance of XGB and SVR in forecasting the 10-day streamflow for a hydroelectric facility. Using the Fourier Transform, the streamflow time-series was divided into seven frequency components. The SVR or XGB model was then used to independently predict each component. According to assessment criteria, the study's findings showed that when utilizing the Fourier decomposition approach, the SVR outperformed the XGB model.

Artificial neural networks (ANNs) may be made more generic by integrating them with ensemble techniques, notably bagging and boosting, according to the authors of Reference [33]. The approach minimizes the inaccuracy in the short-term load forecasting (STLF) but increases computation time owing to numerous training methods. Ensembles of ANNs are trained in parallel utilizing bootstrapped samples of the training data. A different method, suggested in Reference [34], suggests an evolutionary optimization process for fine-tuning a single ANN in order to prevent problems like overfitting and choose the optimum design. This approach produced findings with a mean absolute percentage error of 4.86%. (MAPE). According to the findings from References [18], [31], robust hyperparameter optimization may help ANNs overcome the difficulties associated with ANN tuning and outperform alternative forecasting techniques for STLF.

In the area of short-term load forecasting, recurrent neural networks, particularly Long Short-Term Memory (LSTM) networks, are becoming more significant (STLF). LSTM networks include feedback connections that are advantageous for tackling time-series forecasting issues, in contrast to conventional feedforward neural networks. The results of a study that compared the forecasting abilities of LSTM, back-propagation Artificial Neural Network (ANN), and Support Vector Regression (SVR) for predicting the upcoming 24 hours of load in a smart grid showed that LSTM had a Mean Absolute Percentage Error (MAPE) of 1.9%, while ANN had a MAPE of 3.3%, and SVR had a MAPE of 4.8% [35]. Another study that forecasted the load for a furniture company using a multi-layer LSTM approach and compared it to models like ARIMA, exponential smoothing, k-nearest neighbors regressor, and ANN discovered that LSTM performed the best in terms of both Root Mean Squared Error (RMSE) and MAPE, followed by Support Vector Machine (SVM), and ANN[36]. Another research indicated that deep learning techniques outperform other approaches in electrical STLF, but that they have not yet reached their full potential, particularly in terms of hidden layer architectures. In order to solve this, the research assessed deep-stacked LSTM with many layers, including unidirectional LSTM (Uni-LSTM), bidirectional LSTM (Bi-LSTM), and SVR, and discovered that Bi-LSTM had a MAPE of 0.22%, but Uni-LSTM and SVR had MAPE scores over 2% [37].

Chapter III

Proposed Methodology

Forecasting methodology of thesis given in Figure 3.1, details process depicted in this flowchart.

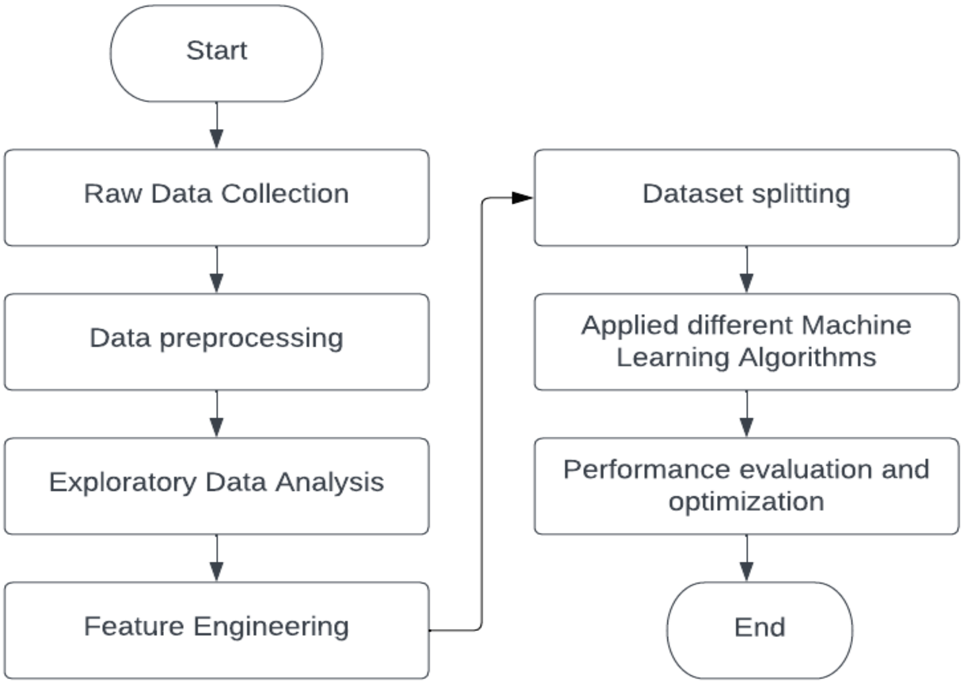


Figure 3.1 Flowchart of forecasting methodology

3.1 Data Collection and Dataset Analogy

The dataset used in this thesis was collected from National Load Dispatch Centre (NLDC), Power Grid Company of Bangladesh contain hourly basis generation of

electric energy. The dataset containing time series data from 11th January, 2016 to 15th March, 2021 total 44518 records with two columns, one has datetime information and other has the energy generation on hourly basis.The dataset is represented on Figure *3.2*
.

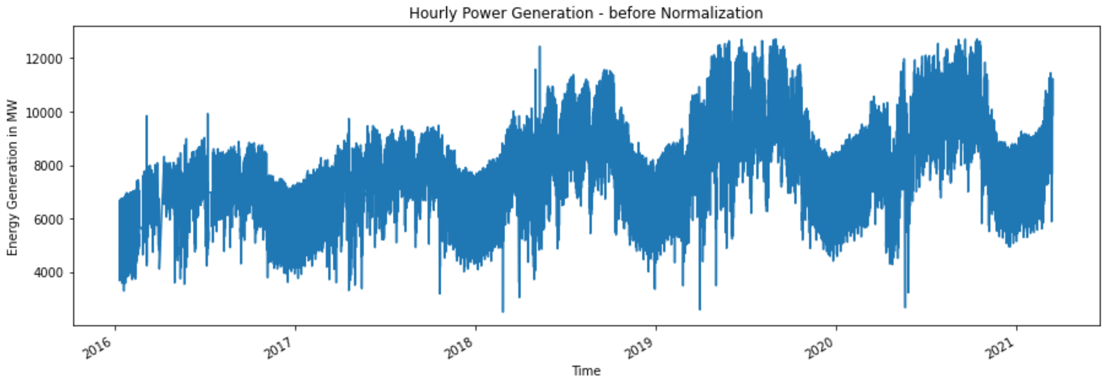


Figure 3.2. Dataset Representation

3.2 Data Preprocessing

Data preprocessing involves several steps to ensure that it is ready for analysis. This includes checking for missing values and outliers, scaling the data to a specific range, and dividing the time series data into training and testing subsets while maintaining the temporal order. The goal of this process is to clean and prepare the data for further analysis.

3.2.1 Filling of Missing Value

Clean dataset is required for training a machine learning Algorithms. Dataset with missing data lead the some problem. It is a great idea that fill the missing values of existing dataset. There are many methods to fill missing values in pandas a python library used for data analysis. Some of the filling methods are given below:

- Fill missing values with the mean value of the column
- Fill missing values with median value of the column
- Fill missing values with a specified value of the column

- Forward fill missing value
- Backward fill missing value
- Interpolate missing value

In this thesis, Fill the missing values with mean used.

3.2.2 Handling Outliers

Outlier handling is important in machine learning (ML) because outliers can have a significant impact on the performance of a model. Outliers are data points that are significantly different from the other data points in a dataset. In ML, outliers can lead to the following issues:

Biased model: Outliers can skew the model's parameters, leading to a biased model that may not generalize well to new data.

Overfitting: Outliers can also increase the risk of overfitting, as the model may try to fit the outliers instead of the underlying patterns in the data.

Reduced accuracy: Outliers can reduce the accuracy of a model, especially if the model is sensitive to the presence of outliers.

There are several ways to handle outliers in ML, including:

Removing outliers: This is the simplest approach, but it may not always be appropriate, especially if the outliers contain important information.

Replacing outliers: This involves replacing the outliers with more representative values, such as the mean or median.

Robust models: Some ML algorithms, such as decision trees and random forests, are more robust to outliers and do not require explicit outlier handling.

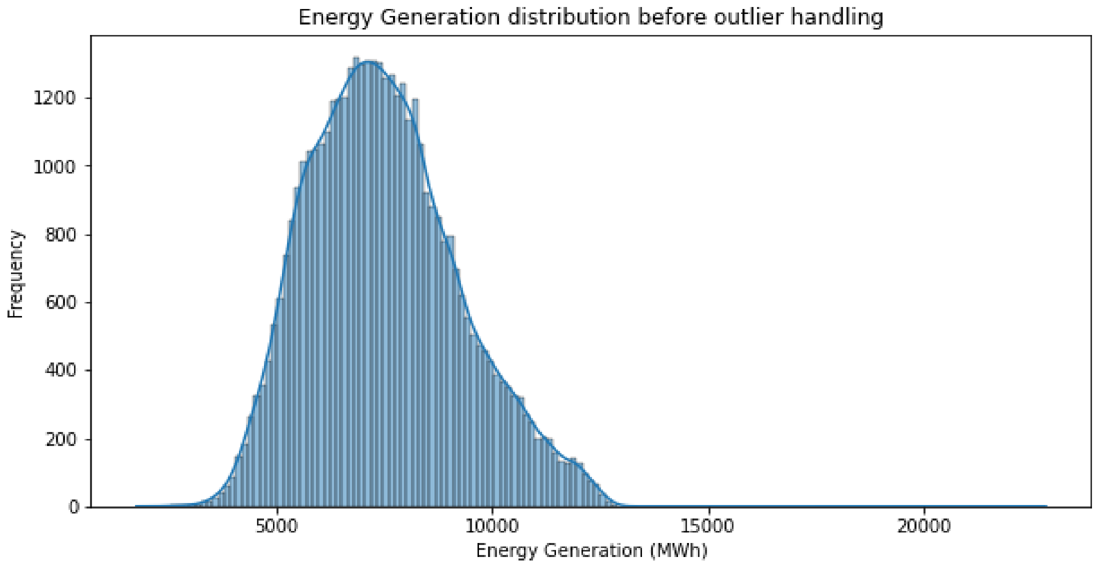


Figure 3.3 Energy generation distribution before outlier handling

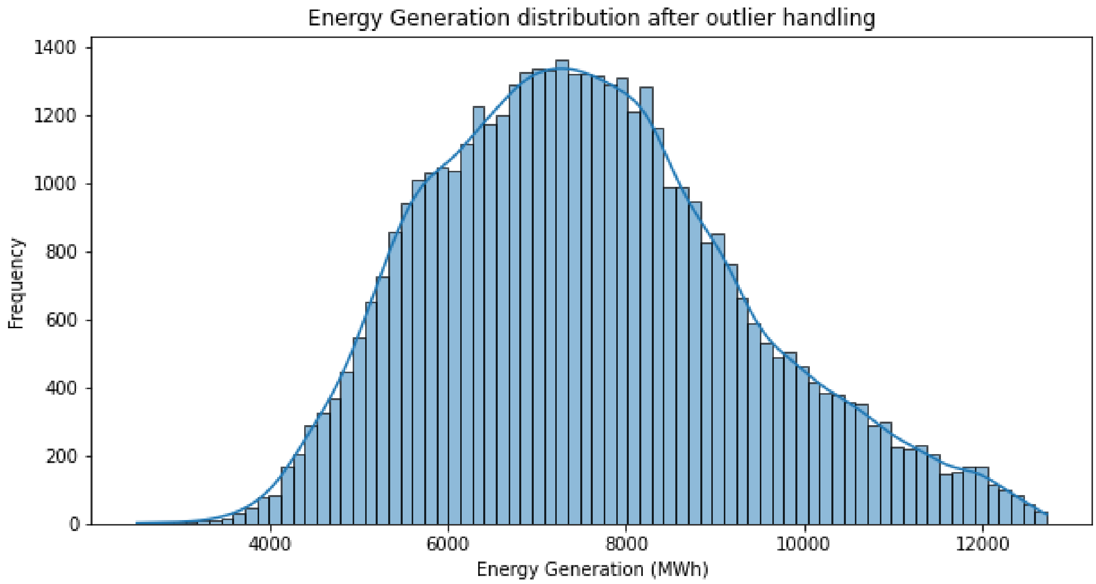


Figure 3.4 Energy generation distribution after outlier handling

3.2.3 Dataset Scaling

Dataset scaling is important in machine learning (ML) because many ML algorithms are sensitive to the scale of the input features. This means that the results of an ML model can be significantly impacted by the scale of the features used to train the model.

Here are some reasons why dataset scaling is important:

Improved convergence: Some ML algorithms, such as gradient-based optimization algorithms, are sensitive to the scale of the input features. Scaling the features to have similar ranges can help these algorithms converge faster and more reliably.

Reduced sensitivity to outliers: Scaling the features can also reduce the sensitivity of an ML model to outliers. Outliers can have a disproportionate impact on the results of an ML model when the features have different scales, but scaling the features to have similar ranges can reduce the impact of outliers.

Improved interpretability: Scaling the features can also make it easier to interpret the results of an ML model. For example, if features have different units (e.g., some features are measured in meters and others in kilometers), scaling the features to have similar ranges can make it easier to understand the relative importance of each feature in the model.

There are several methods for scaling datasets, including:

Min-Max Scaling: This method scales the features to have values between 0 and 1, by subtracting the minimum value of each feature and dividing by the range of each feature.

Standardization: This method scales the features to have a mean of 0 and a standard deviation of 1, by subtracting the mean of each feature and dividing by the standard deviation of each feature.

Normalization: This method scales the features to have a norm (i.e., length) of 1, by dividing each feature by its Euclidean norm.

Standardization is the best-performing technique for dataset scaling in forecasting. So, in this thesis standardization technique is used for dataset scaling.

3.3 Exploratory Data Analysis

Exploratory data analysis (EDA) is a process to study and explore datasets and describe their essential properties, frequently using data visualization approaches. It aids in determining how to effectively modify data sources to get the answers required, making it simpler for data scientists to uncover patterns, detect anomalies, test hypotheses, and validate assumptions.

EDA is mainly used to discover what data may disclose beyond the formal modeling or hypothesis testing tasks, and it offers a deeper knowledge of data set variables and their interactions. It also assist to assess if the statistical approaches thinking about using for data analysis are acceptable.

3.3.1 Time Series Analysis

A particular method of studying a collection of data points over a period of time is called a time series analysis. In time series analysis, data points are recorded at regular intervals throughout a predetermined length of time rather than merely occasionally or arbitrarily. But gathering data over time for this kind of research is not all that it entails.

Decomposing a time series entails seeing it as a collection of level, trend, seasonality, and noise components. Decomposition is an effective abstract approach for considering time series more broadly and for better comprehending issues that arise during time series research and forecasting.

Moving average or Rolling average is most popular techniques for time series analysis. It calculates the average over a moving window of a specified number where normal average is calculate from all the data points. This window moves through the dataset and each time it calculate new average for specified number. It helps to find out trend from the dataset and smooth the small fluctuation in datasets.

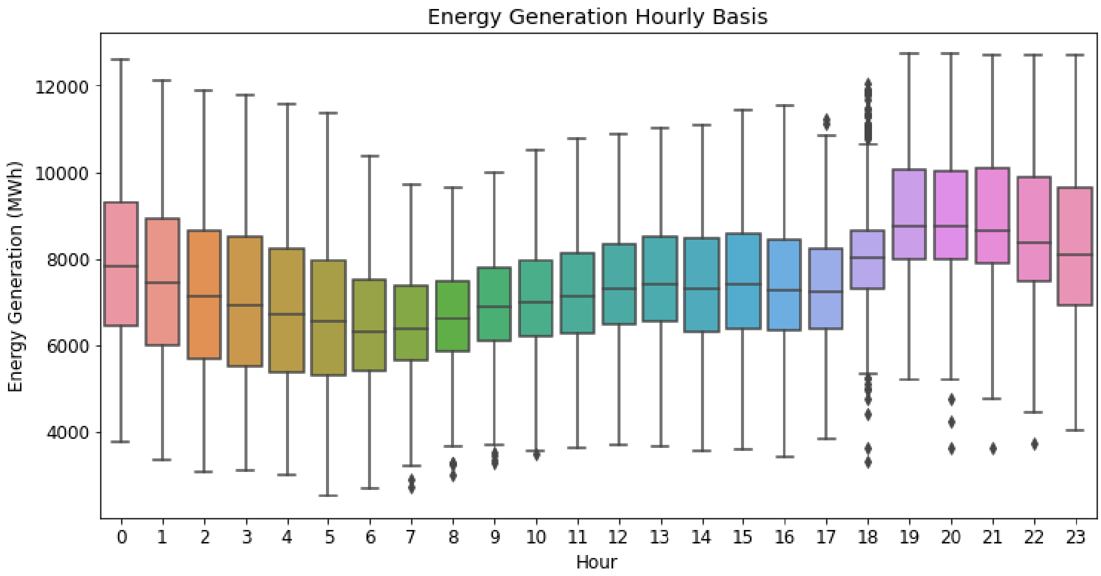


Figure 3.5: Energy generation distribution on hourly basis throughout the day

Figure 3.5 highlight the hourly load pattern by a boxplot representation, at the evening the electricity demand highest and Figure 3.6 shows energy generation monthly basis over the year. At the summer season electric load demand is higher. Starting of the year and ending of the are in the winter season for that reason electricity demand is lower comparatively to summer season.

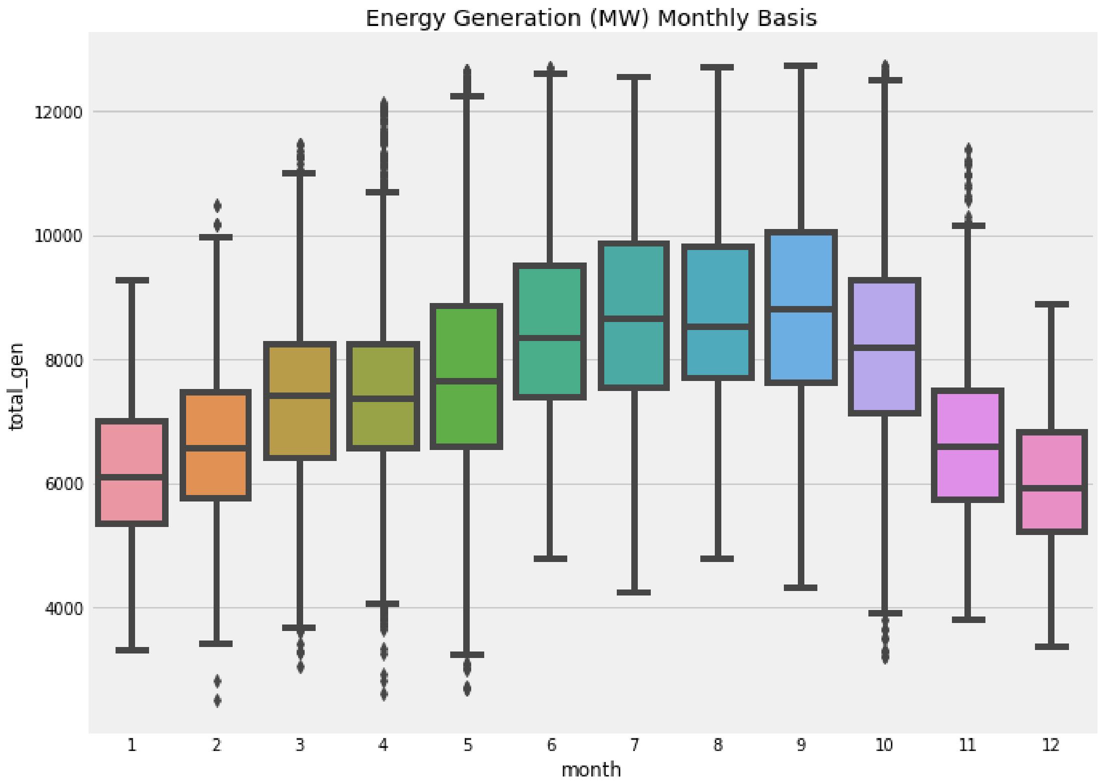


Figure 3.6: Energy generation on monthly throughout the year

3.3.2 Trend Analysis

Energy demand increases day by day. Figure *3.7* boxplot energy demand represent the energy generation trend which in the upward direction this is because population growth, economic development technological advancements and climate control. Year by year, energy consumption pattern would be in upward direction in developing country like Bangladesh because industrial development running this type of country. This pattern will be very important feature for long term load forecasting.

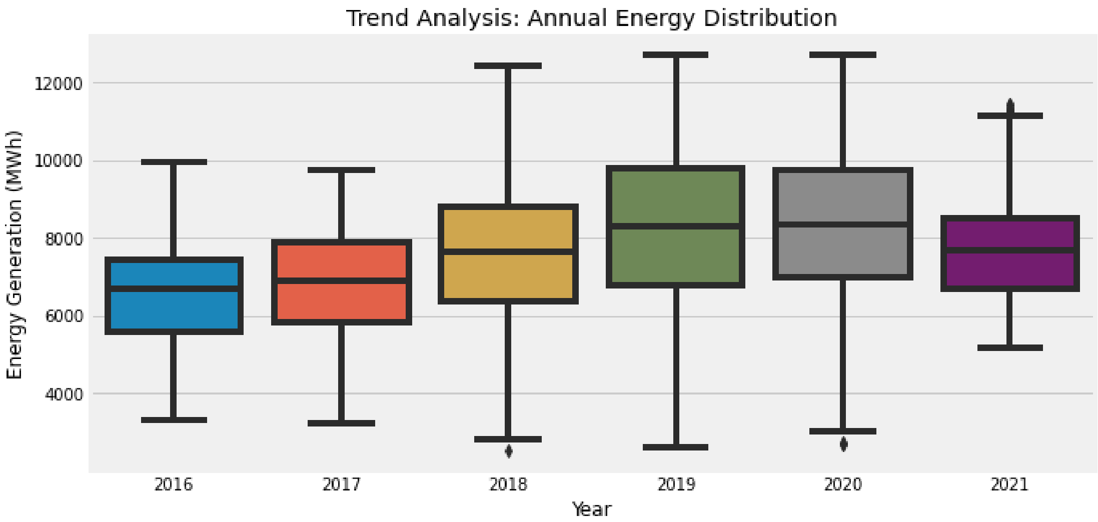


Figure 3.7 Annual Energy Distribution

Figure *3.8* depicts the quarterly energy distribution plot shows first 3 quarter power demand increases due to first three quarter cooling load increase quarter by quarter. Last quarter power demand reduced due to less cooling load in winter months.

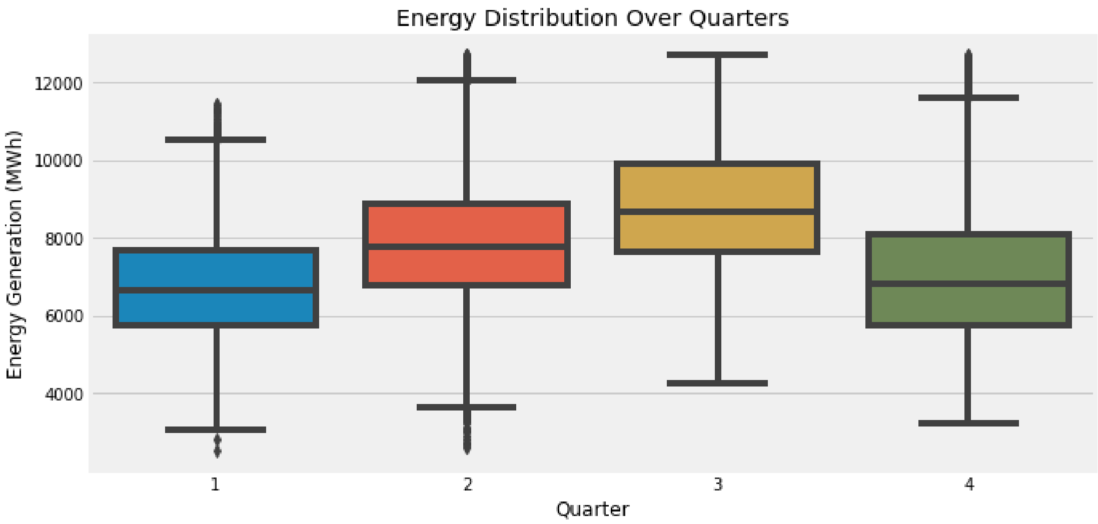


Figure 3.8 Quarterly Electric load Distribution

Holidays and weekends often have an impact on power consumption since use patterns vary dramatically from ordinary days. This feature may be used to increase the accuracy of our model. Figure *3.9* depicts a box plot of electricity use for weekdays and weekends, revealing that weekend consumption is usually lesser than weekday consumption throughout all years.

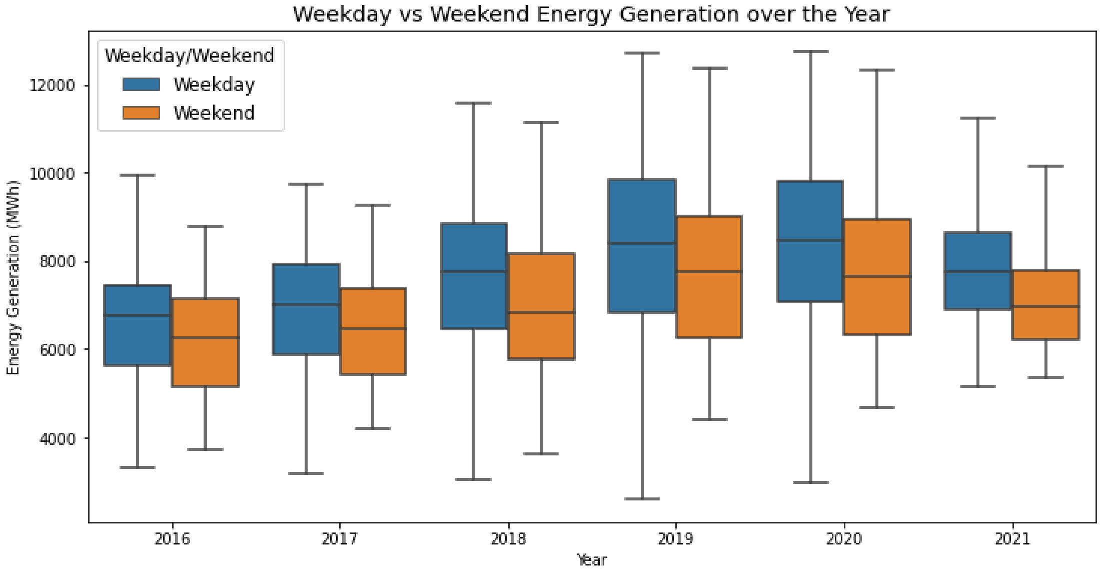


Figure 3.9 Electric load demand by weekday vs weekend

3.4 Feature Engineering

Feature engineering entails selecting, developing, and transforming the most useful and relevant input feature for the algorithm to trained-up.

3.4.1 Feature Extraction and Creation

The practice of extracting features from data collected in order to expose meaningful information is known as feature extraction. This minimizes the quantity of data into manageable numbers for algorithms to handle without affecting the original connections or vital information. Adding features includes providing extra variables that will be advantageous to our model. This might mean adding or deleting features. There are many sorts of characteristics are created and extracted in this study as follows:

Date-time Related features:

- Year
- Quarter
- Month
- Date
- Time – hourly in a day
- Week- number of weeks in year
- Day type

Lag-related features:

Capturing temporal relationship amongst observations in time series data required lag-related features. The value of past data point use as a feature for a current data point is considered a lag-related feature. Lag-related features highlight the interdependence that occurs throughout the period.

Lag-related features derived from past observation:

- Energy generation before 1 hour
- Energy generation before 3 hours
- Energy generation before 5 hours
- Energy generation before 8 hours
- Energy generation before 12 hours
- Energy generation before 1 day for the same hour
- Energy generation before 1 week for the same hour
- Energy generation before 3 quarters (3 months) for the same hour
- Energy generation before 1 year for the same hour

Moving average features:

Moving averages calculate the average value over a specific time of the window and this average calculate several times. The window moves through the dataset, moving average smooth the small fluctuation of data.

Rolling window features derived from past values:

- Moving average of generation for last 3 hours

- Moving average of generation for last 5 hours
- Moving average of generation for last 8 hours
- Moving average of generation for last 12 hours
- Moving average of generation for last 24 hours (1 day)
- Moving average of generation for last 1 week (7 days)
- Moving average of generation for last 1 month
- Moving average of generation for last 3 months
- Moving average of generation for last 3 months

3.4.2 Feature Selection

Too many features were created from time series data. The number features should be reduced to lower the computational cost of developing machine learning model and sometime increases performance of the model. Feature selection helps to increase the accuracy and efficiency of the machine learning model. There are different approaches to select the appropriate feature. Filter method, Wrapper method, and Analyzing correlation matrix method are some of them. Analyzing correlation matrix method was used here to select important feature.

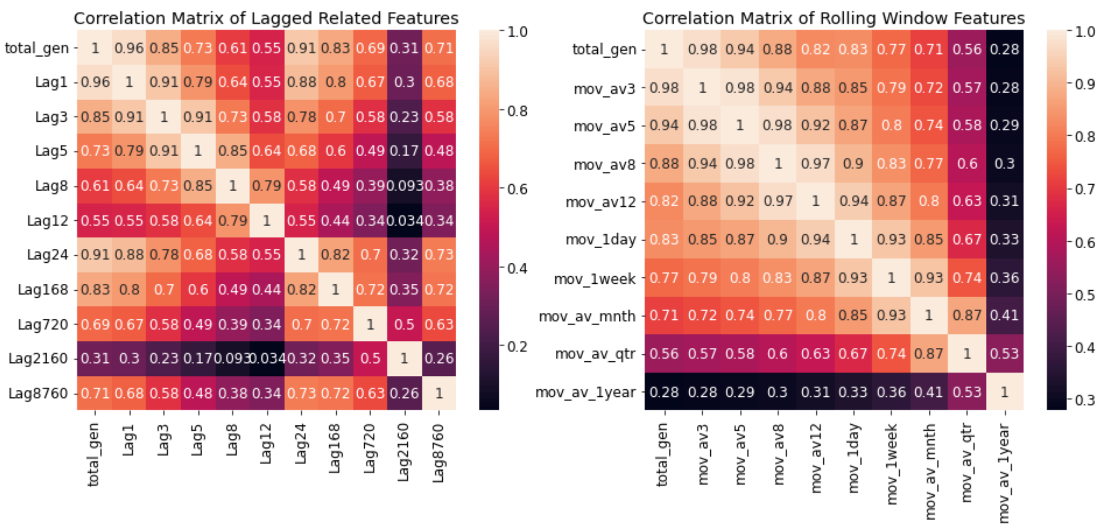


Figure 3.10: Correlation matrix for (a) lag-related features (b) rolling window related feature

There are many methods for analyzing correlation features, Pearson correlation matrix or correlation matrix with heat map used here. This matrix shows the relationship between feature. **Error! Reference source not found.** (a, b) shows the correlation or r elationships between the features. Generally correlation coefficient ranges between -1 to +1, where values between -1 and 0 indicate inverse relationship, 0 indicate no correlation and between 0 to 1 indicate positive relationship.

Pearson correlation coefficient which analyzes how the value of two independent variables varies with regard to one other. Equation (3.1 calculate Pearson correlation coefficient r

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad (3.1)$$

Where, x_i and y_i are two input features, n is the sample size, $\bar{x}$ and $\bar{y}$ are the sample means of the x and y .

From Pearson correlation matrix shows on **Error! Reference source not found.** highly c orrelated feature to the target feature were selected. Here 17 features which correlation coefficient was greater than 0.5 has a great impact on target were selected from 21 features.

3.5 Model Implementation and Evaluation

This is main part of this thesis. After Exploratory Data Analysis (EDA), dataset is ready to implement on machine learning model. In this section, Data splitting into training and testing set, multiple used model discussed here.

3.5.1 Data Splitting

There is a technique before training a machine learning model, a dataset need to be separated into training set and testing set. Train set would be used for training different machine model, test set would be used for evaluate machine learning model.

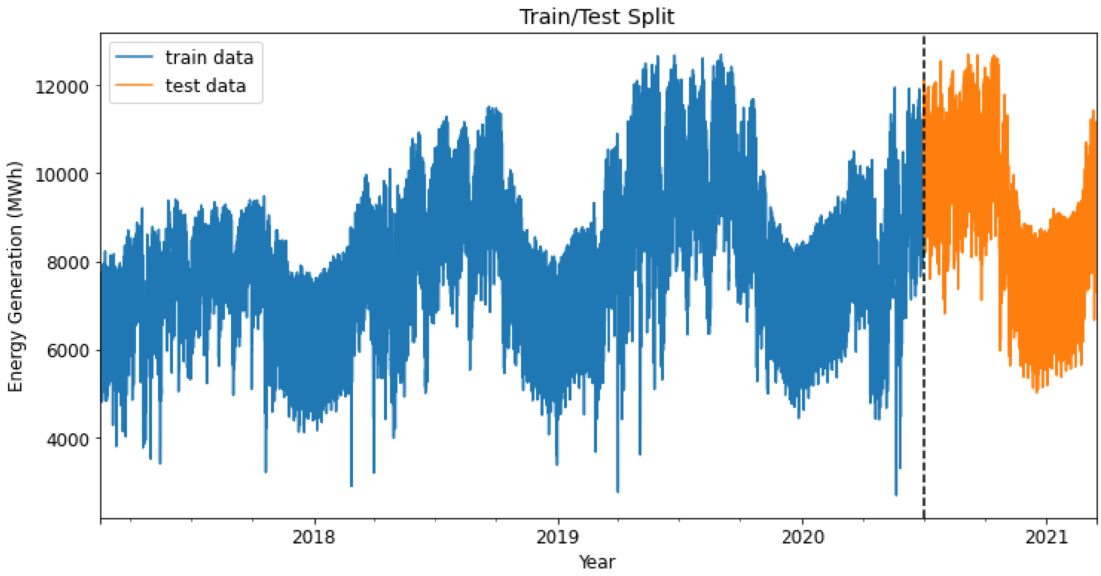


Figure 3.11: Data splitting into training set and testing set

3.6 Applied Algorithms

In this thesis 3 machine learning model and 2 deep learning model total 5 model were used. The five model are given below:

- Linear Regression (LR)
- Random Forest (RF)
- Extreme Gradient Boosting (XGBoost)
- Artificial Neural Network (ANN)

3.6.1 Linear Regression (LR)

Multiple linear regression is a statistical technique for analyzing the relationship between a dependent variable and a number of independent factors. Creating a linear equation that best captures the dependence between the dependent variable and the independent variables is the goal.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon \quad (3.2)$$

Where, y is the dependent variable, x_1, x_2, x_p independent variable and β_1, β_2, β_p are regression coefficient, β_0 is the y intercept and ε is the error term.

Multiple linear regression may be used by the researcher to include each of these conceivably important factors into a single model. The advantages of this approach include a more accurate knowledge of the connection between each distinct factor and the result. [38]

3.6.2 Random Forest (RF)

A random forest (RF) model is a collection of several weak learners, sometimes referred to as decision trees. Classification and regression problems may be solved with it. Moreover, regression analysis makes extensive use of this model. The following steps may be taken to complete the regression technique utilizing random forest:

Data splitting: The procedure involves dividing the characteristics into rows, with each row producing a decision tree.

Making decisions: Each tree decides for itself depending on the data.

Decision aggregation: The outcome of this stage is the average value forecast made by the trees.

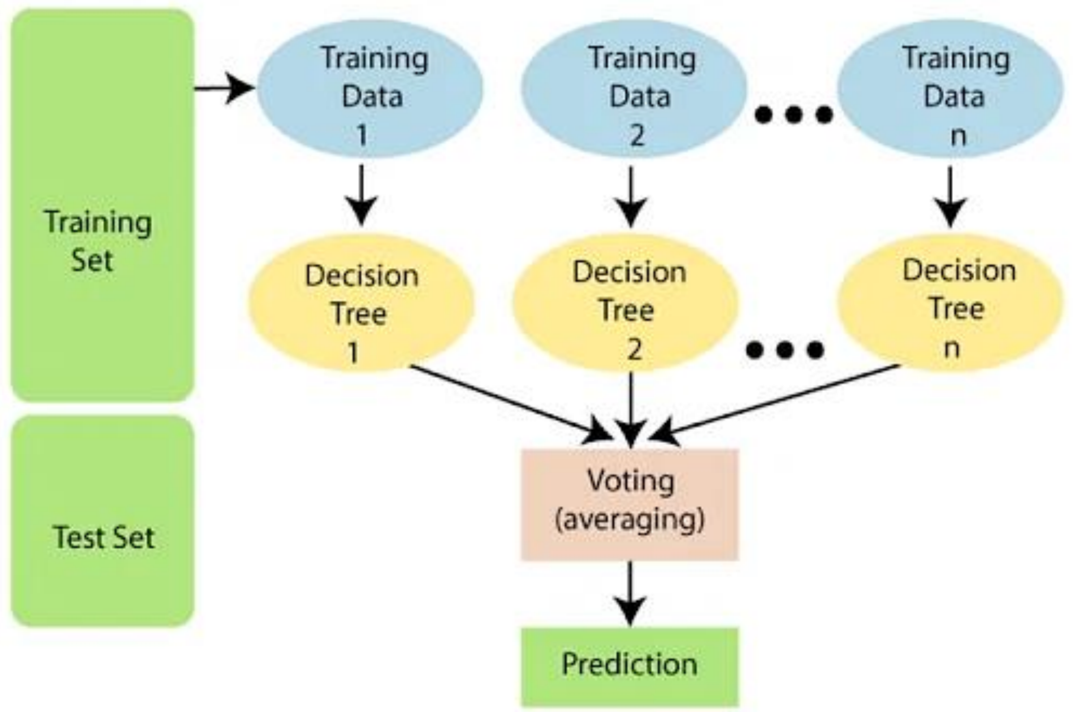


Figure 3.12: Random Forest regression model.

3.6.3 Extreme Gradient Boosting (XGBoost)

For supervised learning issues like classification and regression, XGBoost is a well-liked machine learning method. Decision trees serve as the algorithm's base learners in this particular kind of gradient boosting.

A number of decision trees are constructed by XGBoost throughout the training phase and are subsequently added to the ensemble model. To increase the prediction ability of the model, it first builds a single decision tree and then progressively adds additional trees. The loss function, which calculates the difference between the anticipated and actual values, is minimized by the algorithm using a technique called gradient descent.

With regard to the anticipated values, XGBoost calculates the gradients and hessians of the loss function in each iteration. The residuals from the previous iteration are then predicted using a new decision tree that is built using these data. After the required number of trees has been built, the new tree is then added to the ensemble model and the procedure is repeated.

The L1 and L2 regularization algorithms are used by XGBoost to the tree weights and leaf scores to minimize overfitting. Moreover, it imposes a minimum amount of samples in each leaf and restricts the maximum depth of the decision trees.

The model calculates the weighted total of all the ensemble's decision trees' predictions during the prediction step to provide the final forecast. A highly precise and reliable model that can manage intricate and nonlinear interactions between the input and output variables is produced using this method.

3.6.4 Artificial Neural Network (ANN)

Artificial Neural Networks (ANNs), a kind of machine learning method, are loosely modeled on the composition and functionality of the human brain. Artificial neural networks (ANNs) are used to tackle a variety of difficulties, including classification, regression, and image recognition.

Figure 3.13 illustrate the working principle of ANN. Several interconnected layers of artificial neurons make form an artificial neural network (ANN). For the layer below it, each layer is responsible for transforming the raw data into a more useful representation. Each hidden layer that comes after the input layer applies a transformation to the output of the layer before it. The input layer receives the raw data. The last output layer creates the network's forecast.

To lessen the difference between predictions and actual values in the training set, the network alters its parameters (weights and biases) throughout training. This is done by determining the output difference between predicted and actual output, then using backpropagation to adjust the network's weights in order to reduce this difference.

By calculating the gradient of the loss function with respect to the network weights, backpropagation is the process of updating the weights in the network in the direction that is opposite to that gradient. The process is repeated until there is just a little discrepancy between the expected and actual figures.

ANNs may be used to learn the complex and nonlinear relationships between the input and output variables. They are also beneficial for a range of machine learning applications since they can generalize to new datasets. Nevertheless, they could need careful network architecture and hyperparameter selection during training, which might be computationally expensive.

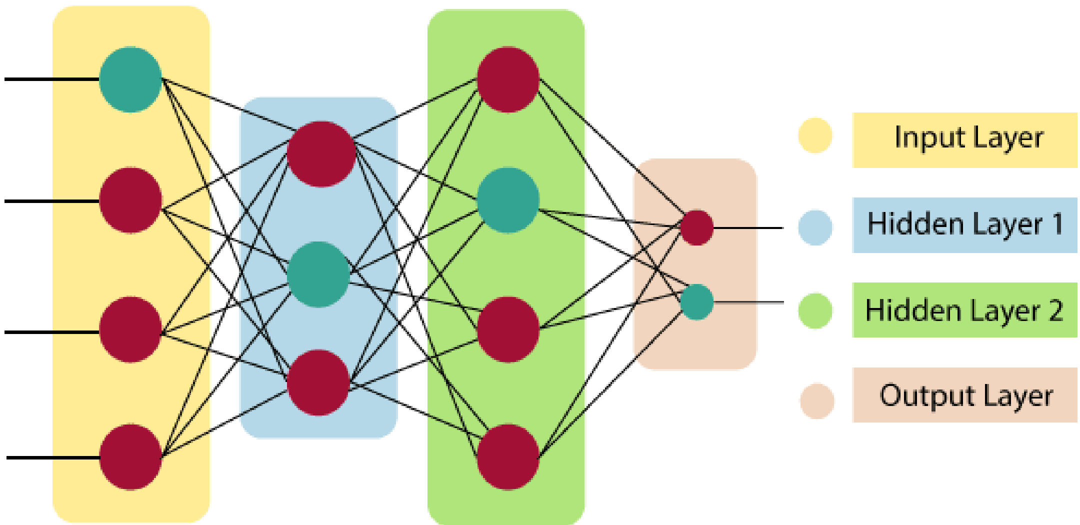


Figure 3.13: (a) Artificial Neural Network (ANN) architecture [39]

3.7 Performance Metrics for Evaluation

A number of measures, including as the root mean squared error (RMSE) , mean absolute error (MAE) and mean absolute percentage error (MAPE) are often used to evaluate the precision of load forecasting models [40] . Another often used metric is MAE, which is the mean of the total absolute discrepancies between actual and predicted values. While RMSE is also often used, it penalizes bigger error terms more severely, producing steadily greater results for outliers than MAE. The formulae for these error measurements are as follows:

Mean Absolute Error:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \mathrm{y_p}\right| \tag{3.3}$$

Mean Absolute Percentage Error:

$$\mathrm{MAPE} = \frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{\mathrm{y_i} - \mathrm{y_p}}{\mathrm{y_i}}\right| \tag{3.4}$$

Root Mean Square Error:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \mathrm{y_p}\right)^2} \tag{3.5}$$

Here ,

y_i = actual value

y_p = predicted value

n = number of observations

CHAPTER IV

Result & Discussion

The time frame of load forecasting is broadly classified into 3 categories, Short Term Load Forecasting (STLF- Several hours to several days), Mid Term Load Forecasting (MTLF-Several days to several months) and Long-Term Load Forecasting (LTLF – Several months to several year). In this research, all model predict the short term electric load demand.

4.1 Comparative analysis on different machine learning models

Comparing different machine learning is a crucial step in assessing their effectiveness and choosing the model that is best suited for the purpose of load forecasting. This part will provide a thorough evaluation of the performance of various models as well as the findings of our research.

The following models were compared:

- Linear Regression (LR)
- Random Forest Regression (RFR)
- Extreme Gradient Boosting (XGBoost)
- Artificial Neural Network (ANN)

A common dataset and experimental setup used here to provide fair comparisons across the models. The datasets divided into training and testing sets by 8:2 ratio.

Popular data-driven methods for time series data include linear regression, which minimizes the sum of the squared vertical distance to estimate the parameters, and utilizing ordinary least square technique to estimate the parameters by fitting the best straight line across the training data. Figure 4.1 and Figure 4.2 shows the actual load demand to the linear regression (LR) model predicted load demand for two days and one week. In case of similar load pattern, in example morning time, the model perform well but at the fluctuation condition the forecast error is high.

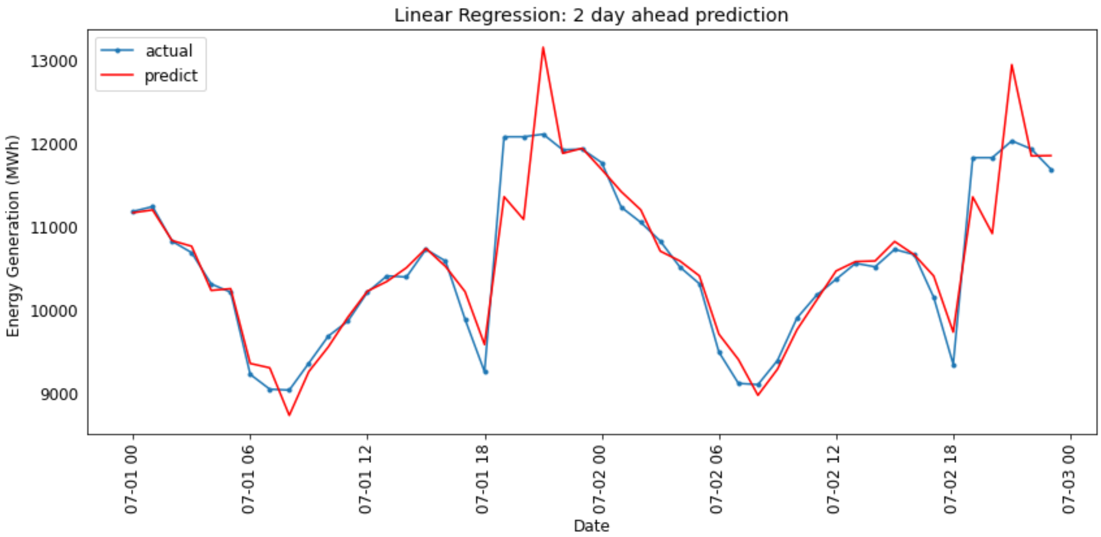


Figure 4.1: Two day ahead load prediction using linear regression

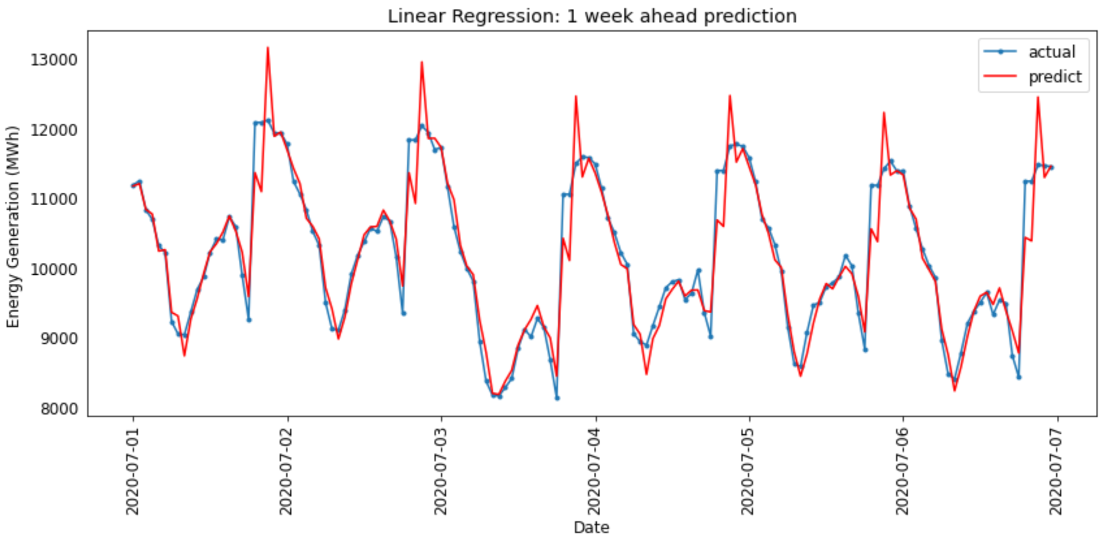


Figure 4.2: one week ahead load prediction using linear regression model

Random Forest regressor is use ensemble technique to predict the load, and it is a tree based algorithm which can capture hidden non-linear complex relationship. Figure 4.3 and Figure 4.4 shows the two days and one week predicted load in comparison to actual load. This model perform well in the peak load condition.

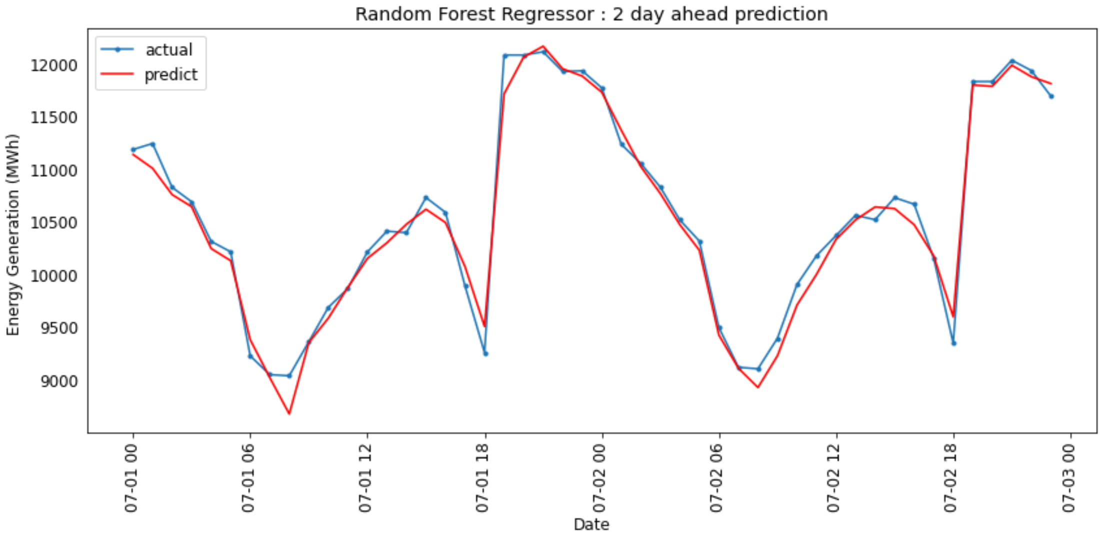


Figure 4.3: Two day ahead load prediction using Random Forest Regression model

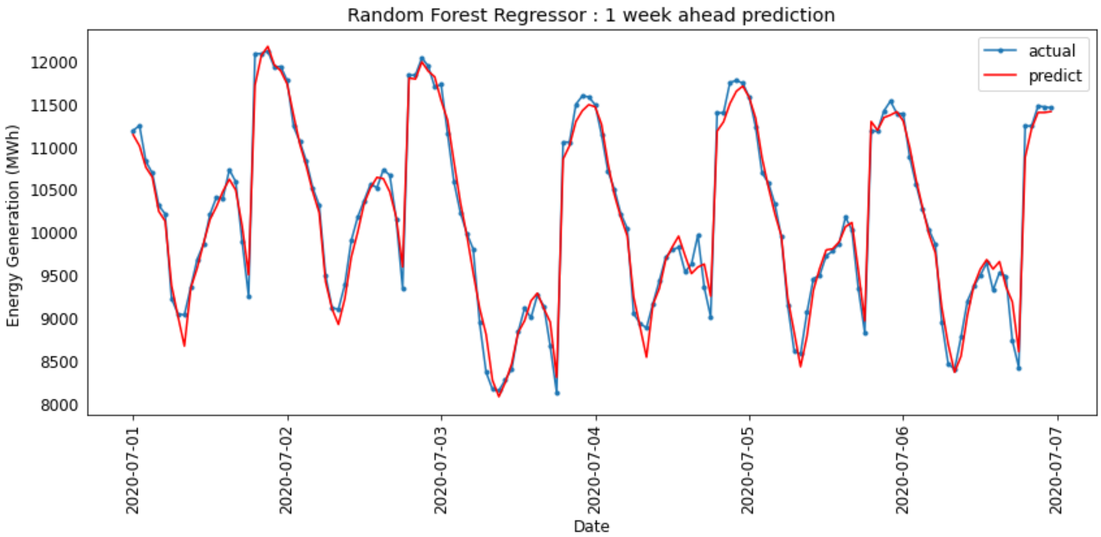


Figure 4.4: one week ahead load prediction by using random forest regression model

Figure 4.5 Figure 4.6 shows the comparisons with actual two day load and one week load to predicted load of two day ahead and one week ahead by using XGBoost Regressor model. This model may outperform linear regression model. It can capture the hidden complex relation from the data. This model forecasted value almost follow the real load.

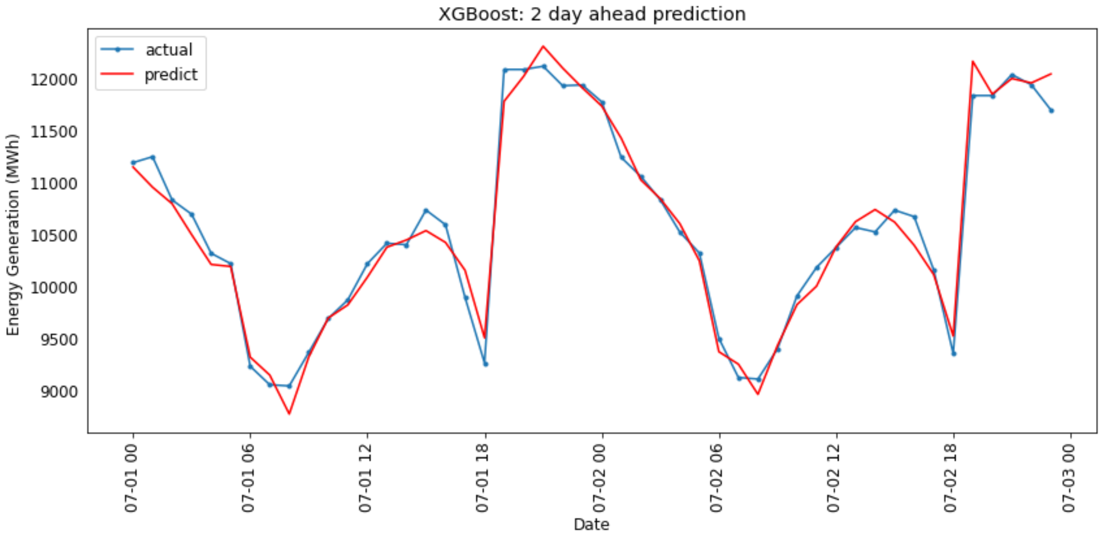


Figure 4.5:two day ahead load prediction using Extreme Gradient Boosting (XGBoost) model

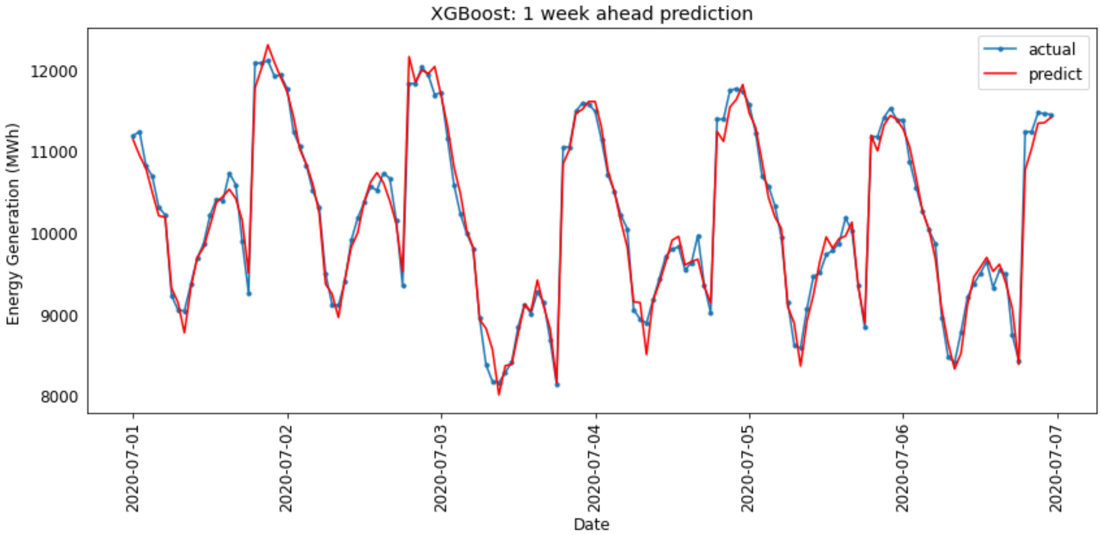


Figure 4.6 one week ahead load prediction by using Extreme Gradient Boosting (XGBoost) model

Artificial Neural Network (ANN) works based human neural system ideology. ANN also perform well to predict two day ahed load and one week ahead load. Figure 4.7 and Figure 4.8 shows the comparison graph actual load demand to predicted load two day ahead deman and one week ahead demand by ANN. ANN perform well LR model but low performance than XGBoost.

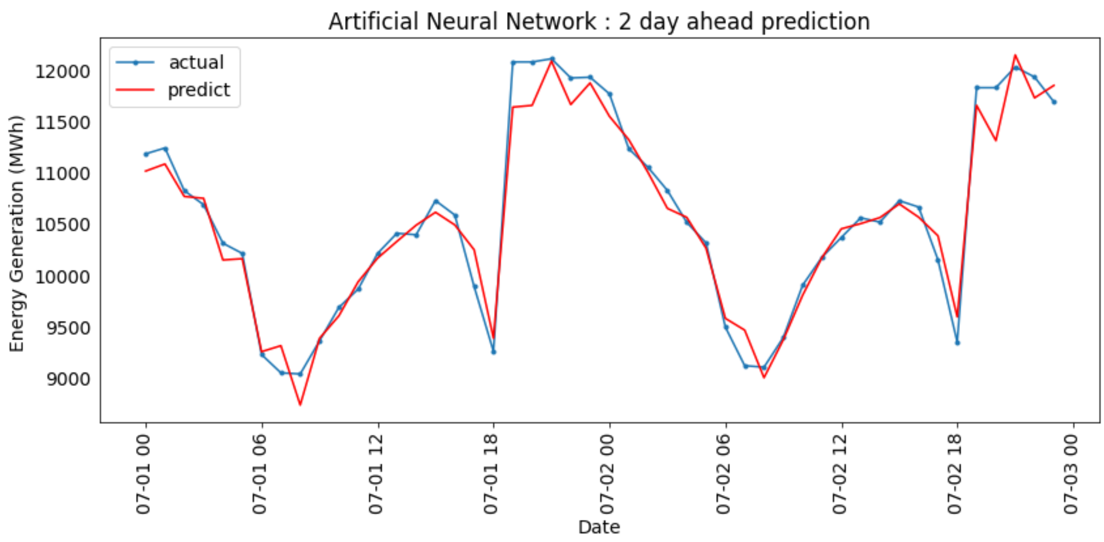


Figure 4.7: two day ahead load prediction using Artificial Neural Network (ANN) model

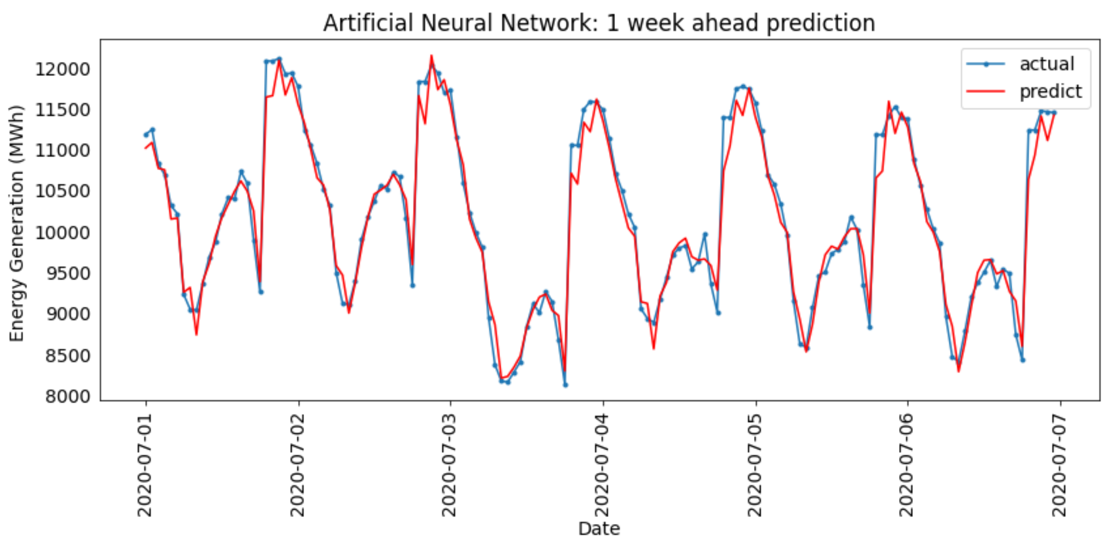


Figure 4.8: one week ahead load prediction using Artificial Neural Network (ANN) model

4.2 Evaluation of prediction accuracy

Machine learning model performance evaluated by different performance metrics (RMSE, MAE, MAPE). Hyperparameter greatly affect performance of the ML model.

Table 1: Hyperparameter of machine learning model

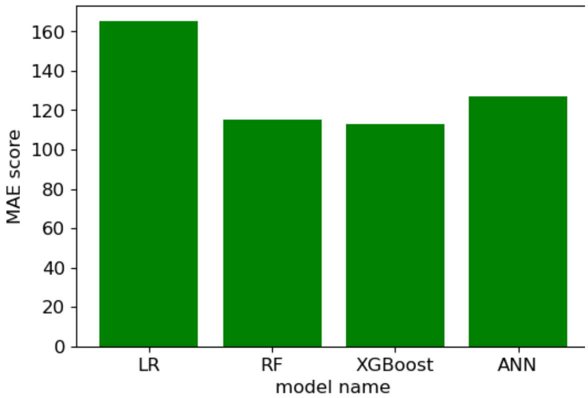
Model	Hyperparameter
Linear Regression	It has no hyperparameter
Random Forest	No of estimator = 200, max depth of tree=40, min sample split = 4, Min sample leaf=4
XGBoost	No of estimators =100, maximum depth=10 min samples split= 2, min sample leaf=1, learning rate=0.1
Artificial Neural Network (ANN)	Number of input layer=3, optimizer= adam Activation function= relu, epochs=200, loss function= mse

Table 1 shows the hyperparameter associated with each model. Linear Regression model has simple structure it has no hyperparameter.

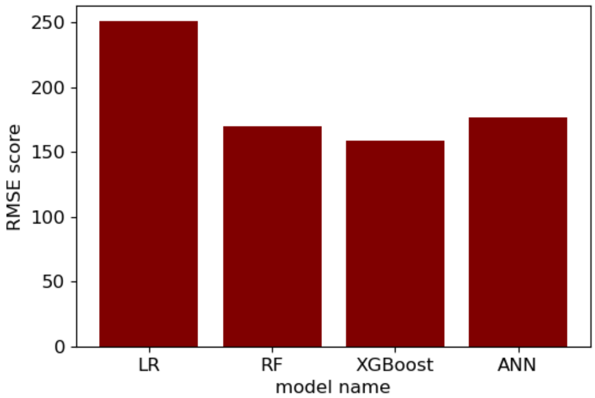
Table 2: Performance metrics of different model

Model	RMSE	MAE	MAPE
Linear Regression (LR)	250.51	165.80	1.84%
Random Forest (RF)	169.27	115.89	1.34%
XGBoost	173.47	122.9	1.41%
Artificial Neural Network (ANN)	176.93	143.66	1.62%

Table 2 shows three performance metrics (Root Mean Square Error, Mean Absolute Error, Mean Absolute Percentage Error) scores, Random forest and XGBoost perform best on this dataset, both of them outperform the Linear regressor model and ANN model. Random forest and XGBoost models MAPE scores 1.34% and 1.41% and RMSE scores 169.27 and 173.47 respectively.



(a)



(b)

Figure 4.9: different models performance scores (a) MAE Scores and (b) RMSE Scores

Figure 4.9 (a) shows the different model performance in MAE score. It clearly seen that Random Random Forest (RF) and Extreme Gradient Boosting (XGBoost) performing well and of them outperform the other LR and ANN model.

4.2.1 Hyper parameter tuning of selected model

Random forest and XGBoost model perform well on two days and one week prediction. Both of the model consist of several hyperparameter. Optimized hyperparameter increase the model performance. Optuna library use for hyperparameter tuning result shown in .

Table 3.

Table 3: Hyperparameter tuning result

Model	hyperparametr list	Optimized hyperparameter
Random Forest	n_estimators= 10:200	n_estimators= 133

	max_depth=5:50 min_samples_split=1:20 min_sample_leaf=1:20 max_feature=0.1:1.0	max_depth=29 min_samples_split=2 min_sample_leaf=1 max_feature=0.65
XGBoost	n_estimators= 100:500 max_depth=5:20 learning_rate=0.001:0.2 reg_alpha=0:1 reg_lamda=0:1	n_estimators= 100 max_depth=9 learning_rate=0.05 reg_alpha=0 reg_lamda=0
ANN	n_hidden_layers=1:5 n_nodes=10:128 learning_rate=0.0001	n_hidden_layers=3 n_nodes=115 learning_rate=0.0003

Model performance can be increased by Hyperparameter Tunig. Table 4 shows the RMSE value before and after hyperparameter tuning. After hyperparameter tuning, ANN has optimum RMSE value of 139.62 where Random forest and XGBoost score are 163.09 and 159.62 respectively.

Table 4: RMSE value comparison before and after hyperparameter tuning

Model	RMSE value before	RMSE value after
Random Forest	169.27	163.09
XGBoost	173.47	159.72
ANN	176.93	139.62

Chapter V

Conclusion

5.1 Conclusion

This study offers a trustworthy machine learning method for forecasting Bangladesh's short-term power load demand. The data collection is narrowed down to just the elements that are crucial to the forecasting goal using an advanced feature engineering method. The feature selection approach drastically decreases dimensionality while keeping important features. To take into consideration the cyclical nature of the production process and add value to the forecasting assignment, moving average and time lag characteristics are developed. In this situation, an useful method for identifying seasonality in the data set is to use the Pearson correlation matrix to find strongly connected temporal delays.

It can seen that, Linear regression model performance is not up to the mark, it can’t capture nonlinear complex relation between the feature and target variable. Random forest and XGBoost models can capture complex the non-linear relationship between the feature and load demand. ANN model perform better than linear regression model but cannot outperform the Random forest and XGBoost model performance. Primarily well performed three model are XGBoost, Random Forest and ANN, and those model performance can be improved by hyperparameter tuning. The optimized models RMSE score are 163.09, 159.29 and 139.62 for Random Forest and XGBoost and ANN model respectively. ANN model shows the best result and surpassing the other model performances.

5.2 Recommendation for Future Work

The fundamental flaw in this study is a decline in short-term forecasting precision based on long-term load forecasting. This is due to a lack of features that are suitable for short-term forecasting. More historical load data may be added to future work in this area of study so that models can be trained for specific scenarios, such vacations. A further strategy to improve the accuracy of the holiday forecast is to create a separate model only for holidays, integrate the predictions with those from a normal days model, and employ stacking methods [28] even if this lengthens the training and forecasting processes [41].

References

[1] A. Hussain, M. Rahman, and J. A. Memon, "Forecasting electricity consumption in Pakistan: the way forward," *Energy Policy*, vol. 90, pp. 73–80, Mar. 2016, doi: 10.1016/j.enpol.2015.11.028.

[2] A. Dedinec, S. Filiposka, A. Dedinec, and L. Kocarev, "Deep belief network based electricity load forecasting: An analysis of Macedonian case," *Energy*, vol. 115, pp. 1688–1700, Nov. 2016, doi: 10.1016/j.energy.2016.07.090.

[3] "Browse Live Wire by Title." https://openknowledge.worldbank.org/handle/10986/17135/browse?type=title (accessed Feb. 19, 2023).

[4] M. Uz Zaman, A. Islam, and N. Sultana, "Short term load forecasting based on Internet of Things (IoT)," Thesis, BRAC University, 2018. Accessed: Feb. 19, 2023. [Online]. Available: http://dspace.bracu.ac.bd/xmlui/handle/10361/10170

[5] "Load Forecasting of Andhra Pradesh Grid using PSO, DE Algorithms | Semantic Scholar." https://www.semanticscholar.org/paper/Load-Forecasting-of-Andhra-Pradesh-Grid-using-PSO%2C-Kumar-Veerakumari/2066842b2b6551b902ccd706dd4774eb0a355d2c (accessed Feb. 19, 2023).

[6] I. Nti, A.-A. Samuel, and A. Michael, "Predicting Monthly Electricity Demand Using Soft-Computing Technique," Jun. 2019.

[7] "Types of Forecasting Methods - EEEGUIDE.COM." https://www.eeeguide.com/types-of-forecasting-methods/ (accessed Feb. 19, 2023).

[8] S. Bouktif, A. Fiaz, A. Ouni, and M. A. Serhani, "Optimal Deep Learning LSTM Model for Electric Load Forecasting using Feature Selection and Genetic Algorithm: Comparison with Machine Learning Approaches †," *Energies*, vol. 11, no. 7, Art. no. 7, Jul. 2018, doi: 10.3390/en11071636.

[9] S. Dutta *et al.*, "Load and Renewable Energy Forecasting for a Microgrid using Persistence Technique," *Energy Procedia*, vol. 143, pp. 617–622, Dec. 2017, doi: 10.1016/j.egypro.2017.12.736.

[10] N. G. Paterakis, E. Mocanu, M. Gibescu, B. Stappers, and W. van Alst, "Deep learning versus traditional machine learning methods for aggregated energy demand prediction," in *2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, Sep. 2017, pp. 1–6. doi: 10.1109/ISGTEurope.2017.8260289.

[11] E. H. Barakat and J. M. Al-Qasem, "Methodology for weekly load forecasting," *IEEE Trans. Power Syst.*, vol. 13, no. 4, pp. 1548–1555, Nov. 1998, doi: 10.1109/59.736304.

[12] "Short-term hourly load forecasting using time-series modeling with peak load estimation capability | IEEE Journals & Magazine | IEEE Xplore." https://ieeexplore.ieee.org/document/962429 (accessed Feb. 18, 2023).

[13] M. S. Al-Musaylh, R. C. Deo, J. F. Adamowski, and Y. Li, "Short-term electricity demand forecasting with MARS, SVR and ARIMA models using aggregated demand data in Queensland, Australia," *Adv. Eng. Inform.*, vol. 35, pp. 1–16, Jan. 2018, doi: 10.1016/j.aei.2017.11.002.

[14] M. A. A. Amin and Md. A. Hoque, "Comparison of ARIMA and SVM for Short-term Load Forecasting," in *2019 9th Annual Information Technology, Electromechanical Engineering and Microelectronics Conference (IEMECON)*, Mar. 2019, pp. 1–6. doi: 10.1109/IEMECONX.2019.8877077.

[15] K. Chapagain and S. Kittipiyakul, "Short-Term Electricity Demand Forecasting with Seasonal and Interactions of Variables for Thailand," in *2018 International Electrical Engineering Congress (iEECON)*, Mar. 2018, pp. 1–4. doi: 10.1109/IEECON.2018.8712189.

[16] "A Neural Network Based Short Term Electric Load Forecasting in Ontario Canada | IEEE Conference Publication | IEEE Xplore." https://ieeexplore.ieee.org/document/4052750 (accessed Feb. 18, 2023).

[17] O. Adeoye and C. Spataru, "Modelling and forecasting hourly electricity demand in West African countries," *Appl. Energy*, vol. 242, pp. 311–333, May 2019, doi: 10.1016/j.apenergy.2019.03.057.

[18] "Designing a short-term load forecasting model in the urban smart grid system - ScienceDirect." https://www.sciencedirect.com/science/article/pii/S0306261920303627?via%3Dihub (accessed Feb. 18, 2023).

[19] E. Bećirović and M. Ćosović, "Machine learning techniques for short-term load forecasting," in *2016 4th International Symposium on Environmental Friendly Energies and Applications (EFEA)*, Sep. 2016, pp. 1–4. doi: 10.1109/EFEA.2016.7748789.

[20] P. M. Ferreira, A. E. Ruano, and R. Pestana, "Improving the Identification of RBF Predictive Models to Forecast the Portuguese Electricity Consumption*," *IFAC Proc. Vol.*, vol. 43, no. 1, pp. 208–213, Jan. 2010, doi: 10.3182/20100329-3-PT-3006.00039.

[21] P. M. Ferreira, I. D. Cuambe, A. E. Ruano, and R. Pestana, "Forecasting the Portuguese Electricity Consumption using Least-Squares Support Vector Machines," *IFAC Proc. Vol.*, vol. 46, no. 20, pp. 411–416, Jan. 2013, doi: 10.3182/20130902-3-CN-3020.00138.

[22] M. Zou, D. Fang, G. Harrison, and S. Djokic, "Weather Based Day-Ahead and Week-Ahead Load Forecasting using Deep Recurrent Neural Network," in *2019 IEEE 5th International forum on Research and Technology for Society and Industry (RTSI)*, Sep. 2019, pp. 341–346. doi: 10.1109/RTSI.2019.8895580.

[23] T. Pinto, I. Praça, Z. Vale, and J. Silva, "Ensemble learning for electricity consumption forecasting in office buildings," *Neurocomputing*, vol. 423, pp. 747–755, Jan. 2021, doi: 10.1016/j.neucom.2020.02.124.

[24] S. Hadri, Y. Naitmalek, M. Najib, M. Bakhouya, Y. Fakhri, and M. Elaroussi, "A Comparative Study of Predictive Approaches for Load Forecasting in Smart Buildings," *Procedia Comput. Sci.*, vol. 160, pp. 173–180, 2019, doi: 10.1016/j.procs.2019.09.458.

[25] J. Cai, H. Cai, Y. Cai, L. Wu, and Y. Shen, “Short-term Forecasting of User Power Load in China Based on XGBoost,” in *2020 12th IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, Sep. 2020, pp. 1–5. doi: 10.1109/APPEEC48164.2020.9220335.

[26] G. Suo, L. Song, Y. Dou, and Z. Cui, “Multi-dimensional Short-Term Load Forecasting Based on XGBoost and Fireworks Algorithm,” in *2019 18th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES)*, Nov. 2019, pp. 245–248. doi: 10.1109/DCABES48411.2019.00068.

[27] X. Liao, N. Cao, M. Li, and X. Kang, “Research on Short-Term Load Forecasting Using XGBoost Based on Similar Days,” in *2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)*, Jan. 2019, pp. 675–678. doi: 10.1109/ICITBS.2019.00167.

[28] K. Zhu, J. Geng, and K. Wang, “A hybrid prediction model based on pattern sequence-based matching method and extreme gradient boosting for holiday load forecasting,” *Electr. Power Syst. Res.*, vol. 190, p. 106841, Jan. 2021, doi: 10.1016/j.epsr.2020.106841.

[29] Y. Liu, H. Luo, B. Zhao, X. Zhao, and Z. Han, “Short-Term Power Load Forecasting Based on Clustering and XGBoost Method,” in *2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS)*, Beijing, China, Nov. 2018, pp. 536–539. doi: 10.1109/ICSESS.2018.8663907.

[30] H. Zheng, J. Yuan, and L. Chen, “Short-Term Load Forecasting Using EMD-LSTM Neural Networks with a Xgboost Algorithm for Feature Importance Evaluation,” *Energies*, vol. 10, no. 8, p. 1168, Aug. 2017, doi: 10.3390/en10081168.

[31] P. Xue, Y. Jiang, Z. Zhou, X. Chen, X. Fang, and J. Liu, “Multi-step ahead forecasting of heat load in district heating systems using machine learning algorithms,” *Energy*, vol. 188, p. 116085, Dec. 2019, doi: 10.1016/j.energy.2019.116085.

[32] X. Yu, Y. Wang, L. Wu, G. Chen, L. Wang, and H. Qin, “Comparison of support vector regression and extreme gradient boosting for decomposition-based data-driven 10-day streamflow forecasting,” *J. Hydrol.*, vol. 582, p. 124293, Mar. 2020, doi: 10.1016/j.jhydrol.2019.124293.

[33] P. Singh and P. Dwivedi, “Integration of new evolutionary approach with artificial neural network for solving short term load forecast problem,” *Appl. Energy*, vol. 217, pp. 537–549, May 2018, doi: 10.1016/j.apenergy.2018.02.131.

[34] G. Yan, T. Han, W. Zhang, and S. Zhao, “Short-Term Load Forecasting of Smart Grid Based on Load Spatial-Temporal Distribution,” in *2019 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia)*, May 2019, pp. 781–785. doi: 10.1109/ISGT-Asia.2019.8881259.

[35] H. Abbasimehr, M. Shabani, and M. Yousefi, “An optimized model using LSTM network for demand forecasting,” *Comput. Ind. Eng.*, vol. 143, p. 106435, May 2020, doi: 10.1016/j.cie.2020.106435.

[36] S. Atef and A. B. Eltawil, “Assessment of stacked unidirectional and bidirectional long short-term memory networks for electricity load forecasting,” *Electr. Power Syst. Res.*, vol. 187, p. 106489, Oct. 2020, doi: 10.1016/j.epsr.2020.106489.

[37] “Hybridization of the SGTM Neural-Like Structure Through Inputs Polynomial Extension | IEEE Conference Publication | IEEE Xplore.” https://ieeexplore.ieee.org/document/8478456 (accessed Feb. 18, 2023).

[38] “Advanced statistics: linear regression, part II: multiple linear regression - PubMed.” https://pubmed.ncbi.nlm.nih.gov/14709437/ (accessed Feb. 18, 2023).

[39] “Artificial Neural Network Tutorial - Javatpoint,” *www.javatpoint.com*. https://www.javatpoint.com/artificial-neural-network (accessed Mar. 23, 2023).

[40] B. Yildiz, J. I. Bilbao, and A. B. Sproul, “A review and analysis of regression and machine learning models on commercial building electricity load forecasting,” *Renew. Sustain. Energy Rev.*, vol. 73, pp. 1104–1122, Jun. 2017, doi: 10.1016/j.rser.2017.02.023.

[41] M. Massaoudi, S. S. Refaat, I. Chihi, M. Trabelsi, F. S. Oueslati, and H. Abu-Rub, “A novel stacked generalization ensemble-based hybrid LGBM-XGB-MLP model for Short-Term Load Forecasting,” *Energy*, vol. 214, p. 118874, Jan. 2021, doi: 10.1016/j.energy.2020.118874.