Bachelor of Science in Computer Science & Engineering

Electricity Demand Forecasting Using Machine Learning Models

by

Sharmin Ara

ID: 1604044

Department of Computer Science & Engineering

Chittagong University of Engineering & Technology (CUET)

Chattogram-4349, Bangladesh.

August, 2022

Electricity Demand Forecasting Using Machine Learning Models

Submitted in partial fulfilment of the requirements for
Degree of Bachelor of Science
in Computer Science & Engineering

by

Sharmin Ara
ID: 1604044

Supervised by

Dr. Md. Mokammel Haque
Professor
Department of Computer Science & Engineering

Chittagong University of Engineering & Technology (CUET)
Chattogram-4349, Bangladesh.

The thesis titled '**Electricity Demand Forecasting Using Machine Learning Models'** submitted by ID: 1604044, Session 2019-2020 has been accepted as satisfactory in fulfilment of the requirement for the degree of Bachelor of Science in Computer Science & Engineering to be awarded by the Chittagong University of Engineering & Technology (CUET).

Board of Examiners

Chairman

Dr. Md. Mokammel Haque
Professor
Department of Computer Science & Engineering
Chittagong University of Engineering & Technology (CUET)

Member (Ex-Officio)

Dr. Md. Mokammel Haque
Professor & Head
Department of Computer Science & Engineering
Chittagong University of Engineering & Technology (CUET)

Member (External)

Dr. Mohammad Moshiul Hoque
Professor
Department of Computer Science & Engineering
Chittagong University of Engineering & Technology (CUET)

Declaration of Originality

This is to certify that I am the sole author of this thesis and that neither any part of this thesis nor the whole of the thesis has been submitted for a degree to any other institution.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. I am also aware that if any infringement of anyone's copyright is found, whether intentional or otherwise, I may be subject to legal and disciplinary action determined by Dept. of CSE, CUET.

I hereby assign every rights in the copyright of this thesis work to Dept. of CSE, CUET, who shall be the owner of the copyright of this work and any reproduction or use in any form or by any means whatsoever is prohibited without the consent of Dept. of CSE, CUET.

Signature of the candidate

Date:

Acknowledgements

Without the help of the Almighty and the guidance, support, and direction of so many people and institutions, it is very difficult to complete a task. In the initial phase, I would like to express my unending adoration, obligations, and gratitude to God Almighty, who is infinitely merciful, gracious, and beneficent, and whose favoritism made it possible for me to successfully complete this thesis. Then I would like to express my heartfelt gratitude to my supervisor Dr. Md. Mokammel Haque, Professor, Department of Computer Science and Engineering of Chittagong University of Engineering and Technology for his continuous guidance, support, and suggestions of this work. I want to thank my family for supporting me emotionally and financially. I also want to thank all the other teachers in my department from whom I learned so many things that helped me during this work.I owe my gratitude to Prof. Dr. Mohammad Shamsul Arefin,Department of Computer Science and Engineering, Chittagong University of Engineering and Technology Finally, my appreciation and heartfelt gratitude are extended to each and every person who somehow helped us in this regard.

Abstract

Electricity demand forecasting is one of the essential ways to manage the power system of a country efficiently and effectively. By facilitating the decision-making process for electricity generation and consumption, electricity demand forecasting serves a useful purpose in decision-making for power system companies. In this work, we develop an electricity demand forecasting model that predicts electricity demand using our own dataset. Our dataset includes numerous parameters relevant to Chittagong's electricity covering more than ten years of data (January 2012 to June 2022). The demand data was taken from the Bangladesh Power Development Board website (BPDB) using web scapping and other features via manual insertion. Several models like Linear Regression(LR), Decision Tree(DT), Random Forest(RF), Support Vector Regression(SVR), and K Nearest Neighbors(KNN) based on machine learning (ML) methods have been applied to observe how these algorithms perform in forecasting electricity. The performance of the discussed ML methods has been analyzed based on various evaluation metrics such as accuracy, MAE, MSE, and RMSE. The outcomes obtained from the analysis show that the Random Forest ML approach outperforms the other conventional ML approaches in terms of accuracy, MAE, MSE, and RMSE since it has the highest accuracy (86.24%) and the lowest MAE, MSE, and RMSE score (0.0450, 0.0042, and 0.0652, respectively) compared to the other discussed conventional ML approaches. This Random Forest method has the capability to forecast the electricity demand more precisely and can plan ahead for maintenance and load distribution, which can aid electricity generation and distribution companies.

Keywords: Electricity Demand Forecast; Machine Learning; Prediction; Linear Regression; Decision Tree; Random Forest; Support Vector Regression; K Nearest Neighbours.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

Since electricity has become such an integral element of everyday life in the modern world, every nation's economic progress is closely linked to its power infrastructure, network, and availability. As a result, the worldwide demand for electricity for commercial and residential purposes has increased. On the other hand, electricity rates have been shifting over the years, hiding the fact that electricity generation is insufficient to fulfill world demand. As a result, a number of studies have been undertaken to estimate future electrical energy usage for commercial and residential purposes, allowing power generators, distributors, and suppliers to plan ahead and encourage customers to conserve energy.

In the entire operation and planning of power systems, electricity demand forecasting is critical. To operate the power system effectively and efficiently, precise electricity demand forecasting is important. Electricity demand forecasting can be beneficial in areas such as power outages, demand-supply coordination, maintenance, operating costs, and structural construction. As a result, power demand forecasting has been a prominent research topic in recent decades [1]. The natures of these projections are also distinct. Forecasts for the short term range from one hour to a week. The term "hourly load forecast" refers to a prediction of short-term power consumption. Forecasts for the medium term range from a few weeks to a few months, and even a few years. The monthly load forecast is a term used to describe a medium-term forecast. Long-term electricity demand predictions are used in distribution system planning, cost regulation, and energy trading. A long-term forecast must be valid for at least 5 years and up to 25 years. This forecast is used to determine the system's production and distribution expansion plans. An annual peak load is a term used to describe a long-term forecast [2].

Despite the fact that there has been a lot of new research in the subject of electricity demand forecasting, more robust and accurate electricity demand forecast models are still needed. Because it is used in decision making, an accurate prediction of variance in future electricity demand is critical for both electric customers and utilities. The various influencing elements such as changeable climate, humidity, temperature, calendar indications, occupancy patterns, and societal norms, however, are the major obstacles in future electricity demand forecasting. Ordinary forecasting methods are insufficient in such a dynamic context, and more complex methodologies are necessary.

We will propose a system that will predict the future electricity demand using machine learning approach.

1.2 Categories of Load Forecasting

The three types of load forecasting are as follows:

1. Short term forecasts
2. Medium term forecasts
3. Long term forecasts

However, predicting load can be challenging, particularly in the short term. The foundation for the safe and efficient functioning of electricity systems is short-term load forecasting. The future short-term load can be predicted in large part by taking into consideration influential factors like the weather, population density, etc. and how these relate to the load. Both network managers and utility companies need to be aware of how much energy has to be produced the following day as well as other crucial details like the peak demand in order to operate the plants and plan effectively. This is where short-term forecasting is useful because it can predict the amount of electricity consumed up to a week in advance by analyzing demand. Since it addresses the demand for the next day or the following week, the forecasting must produce results with a low margin of error to prevent shortages and waste. Our task is to compare the forecast load data outcomes of several models in order to identify the most promising result using the machine learning approach.

1.3 Difficulties

Every work in the scientific sector is difficult and demands a lot of work to be successful. While completing this thesis, we have experienced a few difficulties. The following are some of the significant challenges we experienced:

1. Building the dataset: Our work's most difficult part was creating our own dataset. Although there isn't any publicly available dataset for Bangladesh's electricity demand, it is simpler to utilize one from another country. Even though we have considered a variety of aspects for our dataset, it can be challenging to find all of these features in a single source. Additionally, there is no single source of information that provides data for the entire city of Chittagong.

2. Building the models: The most important factors that influence the demand for electricity were taken into consideration to improve prediction. We have trained the model using the well-liked machine learning regression algorithms in an effort to discover the optimum model. The evaluation matrices were compared in order to choose the best model.

1.4 Applications

1. The ability to plan ahead for maintenance and load distribution is one benefit of electricity demand forecasting for electricity generation companies.

2. By observing this demand forecasting pattern, generation companies may readily predict the price of the electricity because the electricity market price is the intersection of the supply curve and the demand curve.

3. Actions to reduce load shedding can be done by estimating the future electricity demand.

1.5 Motivation

1. Electricity demand forecasting is an effective way to balance the requirement and distribution of electricity.
2. With a better prediction of electricity demand, the system operator will be able to reduce the gap between demand and supply ensuring the power system's stability.

1.6 Contribution of the thesis

Even though this work has been done using historical data from various countries, there has only been a very limited amount of research done on historical demand data from Bangladesh to forecast load on electricity demand using machine learning based algorithms. In our thesis work, we have tried to investigate whether machine learning methods help discover hidden patterns in our dataset and how these patterns impact prediction. The main contribution of our thesis are as follows:

1. To build our own dataset combining energy data, weather data and population density data.
2. To predict the electricity demand accurately using well defined machine learning methods.
3. To compare the accuracy of the applied algorithms for achieving the best approach for our system.

1.7 Thesis Organization

Organization of the rest of the thesis report is as follows -

- **Chapter - 2** provides the a brief summary of the previous works in the related fields and it's background.
- **Chapter - 3** represents a detailed information and general procedure of our methodology.
- **Chapter - 4** shows the experimental results and related information about the system.
- **Chapter - 5** brings a conclusion to the findings and discusses the future scopes of the work.

1.8 Conclusion

The basics of our work have been introduced in this chapter. It is not at all simple to complete a task like this. The fundamental overview of electricity demand has been thoroughly discussed. The motivation for the thesis and the challenges I have experienced have both been described. Our work's contribution has also been discussed. The problem's current state and historical background are discussed in the next section.

Chapter 2

Literature Review

2.1 Introduction

Many academic publications on this topic have been published in recent years. This might be because the study of energy economy is seeing increased interest in forecasting electricity demand. In light of the major objective of this study.Forecasting electricity demand is critical to the management and planning of electrical power and energy systems. Researchers use a variety of approaches to anticipate power consumption. So, it is a very demanding research area.

2.2 Related Work

Çamurdan et al. [3], developed a method to predict electricity demand using three different machine learning models. They use different evaluation matrices to test the accuracy of the results. Here the highest accuracy of 98% was obtained from the random forest model.

Shabbir et al. [4], applied three different ML algorithms: linear regression, tree-based regression, and Support Vector Machine (SVM)-based regression on a large dataset of an Estonian household.The results show that SVM is the most effective method since it provides the lowest Root Mean Square Error(RMSE).

Anik et al. [5], a simultaneous equation framework was applied to an annual database over a 47-year period (1972-2018) to forecast future energy demand.

Liu et al. [6], compared single (ANN and SVM) and hybrid machine learning methods which have different faetures and they are applicable in various situations and they also have their own strengths and weaknesses.

Gajowniczek et al. [7], developed an algorithmic technique for load modeling through peak detection and used the Polish Power System based on data between 1 January 2008 and 31 December to feed the forecasting system for this purpose. SVM gives the most promising results for peak classification and ANN for forecasting.

Jain et al. [8], proposed a method to forecast the electricity consumption using the ARIMA model. For the years 2004–2008, the seasonal ARIMA model was determined to be the best model for predicting energy usage in IIT(ISM). The limitation of this paper is that ARIMA model can be limited in forecasting extreme values.

Mati et al. [9], proposed a multiple regression time series modeling approach to predict Nigeria's energy demand using previous data of 1970 to 2005. The main regressors in the model are power usage and percentage connectivity to the national grid. One drawback of the time series models is that they cannot clearly represent technologies and rely on heavily aggregated data, as well as ignoring the technically most efficient solutions, resulting in an underestimation of energy improvement potential.

Behm et al. [10], presented a method for forecasting long-term climate hourly power consumption using ann. Their main contribution is to present a new method to forecast electricity loads as necessary input data for energy system modeling. The limitation is that they didn't use the correct scaling method for loads.

Islam et al. [11], presented an LSTM-based electrical load forecasting approach, which was deployed for a proposed smart grid infrastructure in Bangladesh's Chattogram city, and the findings demonstrate that the LSTM outperforms the SVM. One of the LSTM's significant flaws is that important parameters such the number of time steps, neuron size, batch size, and others must be chosen manually and are dependent on the researcher's expertise.

Eseye et al. [12], suggested a ml-based hybrid feature selection technique for extracting the most relevant and nonredundant information for improved short-term forecasting of power demand in distributed energy systems.

Usha et al. [13], introduced to numerous prognosis and forecasting methodologies that were established in the electrical domain. This project's main goal is to use a seasonal data method to anticipate power consumption. The limitations of this paper is that different consumers use different meters, either standard or smart meters, and their electricity usage varies depending on the meter.

Deng et al. [14], proposed a time series analysis for short term Singapore electricity demand forecasting. The multiplicative decomposition model and the seasonal ARIMA model are presented as time series models. The drawbacks of this research include that load demands often fluctuate greatly between seasons and temperatures, making it difficult to fit a trend line.

Patel et al. [15], introduced a time series model for predicting the future load demand. With the aid of projected power system or future consumption, an electric utility or energy management system may plan effectively. The paper's drawbacks include that consumption may fluctuate as a result of price changes, making it difficult to obtain reliable data for predicting.

Rabbi et al. [16], proposed a model to forecast Bangladesh's annual electricity demand using a multivariate time series model. They anticipate Bangladesh's power consumption using a Multivariate Time Series since it provides more data.

In [17], using univariate historical data on the country's power usage in megawatt-hour from 2003 to 2017, the authors utilized the well-known ARIMA(p,d,q) model to estimate Philippines electricity use for the years 2018 to 2022. The purpose of this research is to forecast electric consumption in the Philippines for household, institutional, and commercial applications. The ARIMA method, which is a type of time series analysis model, was utilized in this work.

In [18], authors presented a new Grey-based prediction method for forecasting very-short-term electric power consumption in order to manage electricity demand. Electric consumption waste can be prevented using this approach.

2.3 Conclusion

This chapter provides an in-depth analysis of the literature, highlighting the various algorithms and methods that have been tested in numerous studies and experiments to forecast electricity demand.The methods that were applied in the study to forecast load demand are covered in full in the following chapter, together with a detailed analysis of the data and important forecasting factors, as well as an experimental setup that includes an explanation of how the algorithm was fully implemented.

Chapter 3

Methodology

3.1 Introduction

The systematic use of algorithms and methods to conduct research using a variety of theoretical approaches is known as methodology. It also talks about the implementation strategies we used to make the application a fully functional solution with no significant usage obstacles. A work's methodology is created by careful preparation, research, and analysis. After conducting extensive research, we developed the technique for our work in a series of steps. The methodology's overall structure is divided into several steps, much like a roadmap. The sections that follow provide descriptions of them.

3.2 Overview of framework

The overall design of the framework is depicted in Figure 3.1.Initially, the dataset is read from the CSV file. The dataset consists of three types of data, such as energy data, weather data, and population density data. Data preprocessing is the combination of three following steps: data cleaning, data normalization, and data structuring. After going through the preprocessing stage, we get the final dataset for further implementation. The final dataset is split into two portions: the training set and the testing set. Machine learning algorithms are applied to the training sets. The prediction models are built and they make predictions on the test set, yielding the accuracy in percentage and comparing their performances.

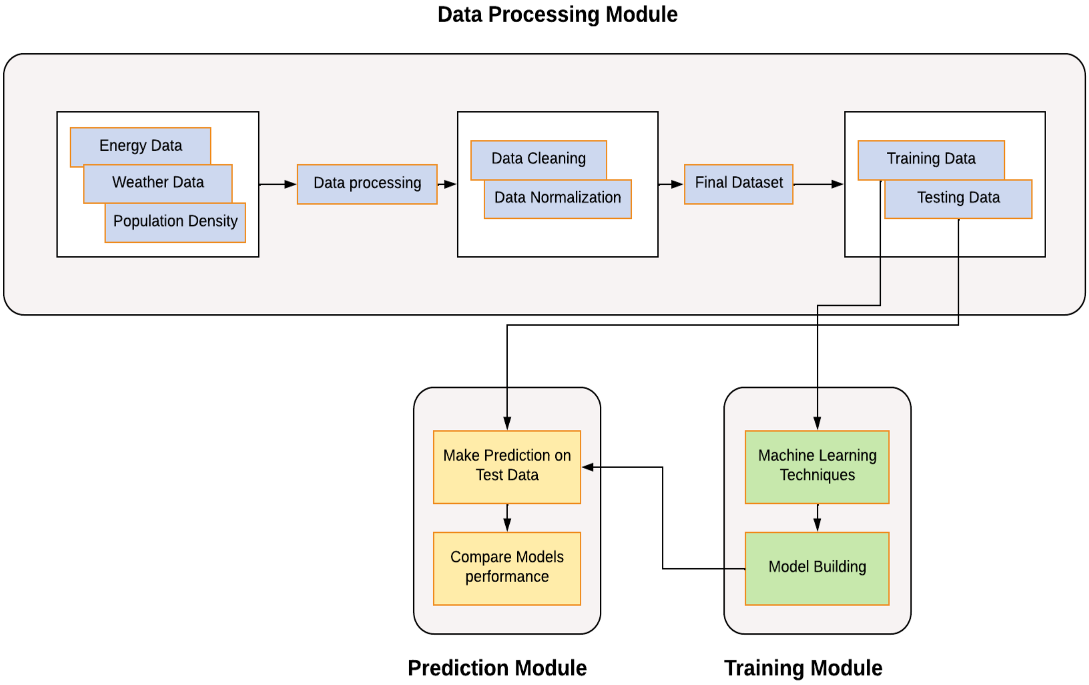


Figure 3.1: Overview of framework

3.3 Explanation of the Proposed Method

The proposed method has been explained in this section. Numerous techniques and algorithms have been implemented in support of the main implementation.

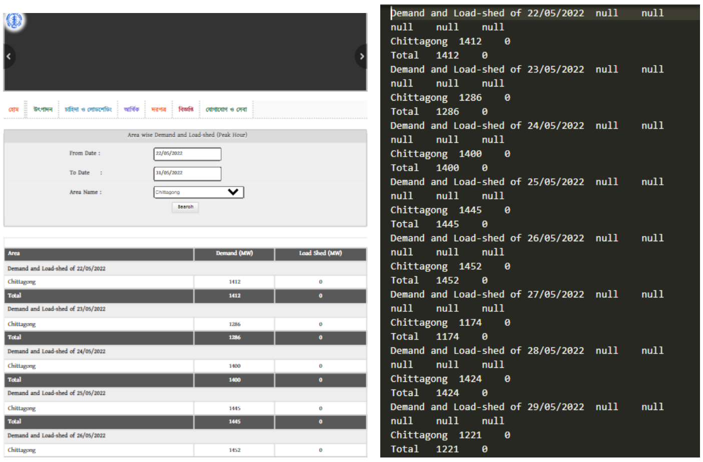
3.3.1 Data Preprocessing

Data preprocessing is the process of preparing the raw data into an understandable format. It is an important step in machine learning models that improves the quality of the data to encourage the extraction of valuable insights from the data.Data preprocessing is also an data mining approach that can't work with raw data.

3.3.1.1 Data Collection

Data collection is the fundamental process in the machine learning pipeline for training the ML models. On the basis of the evidence gathered, a researcher might assess their hypothesis. So, to forecast demand in every situation always requires a meaningful dataset with the necessary features. Initially, the data was collected from various sources for the Chittagong district to build the dataset.The dataset contains data for a period of more than 10 years (01-01-2012 to 31-05-2022). We have considered some features as the main features for our research, and these are: Date, Demand (MW), Temp_max, Temp_min, Humidity, Pressure, Wind_speed, and Population Density. Studying some research papers and articles, ,we have found that these features have a high correlation with electricity demand.

1. The data for demand (mw) was obtained from the Bangladesh Power and Development Board (BPDB), which hosts the areawise electricity demand. The data was extracted using the web scrapper chrome extension.By installing the web scrapper extension from the chrome web store to extract the data. In this technique, the scrapper traverses the whole website and extracts the data. After stop the scraping, using the sitemap tab the data was converted to a text file.Then splitting the file into lines and remove the unnecessary information and extract only date and datewise demand information from the text file and then converted it into a csv file.[19]



Area wise Demand and Load-shed (Peak Hour)

From Date : 22/05/2022

To Date : 31/05/2022

Area Name : Chittagong

Area	Demand (MW)	Load Shed (MW)
Demand and Load-shed of 22/05/2022		
Chittagong	1412	0
Total	1412	0
Demand and Load-shed of 23/05/2022		
Chittagong	1286	0
Total	1286	0
Demand and Load-shed of 24/05/2022		
Chittagong	1400	0
Total	1400	0
Demand and Load-shed of 25/05/2022		
Chittagong	1445	0
Total	1445	0
Demand and Load-shed of 26/05/2022		
Chittagong	1452	0

```
Demand and Load-shed of 22/05/2022  null    null
null    null    null
Chittagong  1412    0
Total   1412    0
Demand and Load-shed of 23/05/2022  null    null
null    null    null
Chittagong  1286    0
Total   1286    0
Demand and Load-shed of 24/05/2022  null    null
null    null    null
Chittagong  1400    0
Total   1400    0
Demand and Load-shed of 25/05/2022  null    null
null    null    null
Chittagong  1445    0
Total   1445    0
Demand and Load-shed of 26/05/2022  null    null
null    null    null
Chittagong  1452    0
Total   1452    0
Demand and Load-shed of 27/05/2022  null    null
null    null    null
Chittagong  1174    0
Total   1174    0
Demand and Load-shed of 28/05/2022  null    null
null    null    null
Chittagong  1424    0
Total   1424    0
Demand and Load-shed of 29/05/2022  null    null
null    null    null
Chittagong  1221    0
Total   1221    0
```

Figure 3.2: Extracting Demand data from BPDB website

2. The weather information (temp_max, temp_min, humidity, pressure, and wind_speed) was collected from the timeanddate.com website.[20]

3. The population density data was collected from the worldpopulationreview.com website.[21]

3.3.1.2 Data cleaning

The dataset includes data during a more than ten-year period (01-06-2015 to 10-06-2018). The dataset ultimately contained 3800 rows as a result of this. One of the important steps in machine learning is data cleaning. There are numerous methods for cleaning data. The handling of missing values is one of them. In our dataset, there are some missing values when collecting the data from the website. Therefore, each row containing a date was checked for any null or zero values before applying ML models. The mean value of all the other values from earlier observations was used to replace any null or zero values that were found.

3.3.1.3 Data normalization

One of the most popular methods for preparing data is normalization, which enables us to convert the values of the dataset's numerical columns to a common scale. Every dataset does not require normalization for machine learning. It is applied whenever a dataset's features have different ranges.It helps to improve a machine learning model's performance and reliability.Even though there are numerous feature normalization methods in machine learning, we have used the Min-Max Scaling method in our dataset because our dataset's features have a different range of values. Min-Max Scaling: Normalization, often known as the Min-Max scaling approach, is a scaling technique where values are shifted and rescaled to conserve their ranges between 0 and 1. Here's the formula for normalization:

$$X' = \frac{X - Xmin}{Xmax - Xmin}$$

Where:

- **X**' is our normalized value
- **X** is the original value
- **X**min is the minimum value of the column
- **X**max is the maximum value of the column

3.3.1.4 Data Analysis

Figures 3.3 show date wise electrical demand over more than 10 years (2012-2022) and the normalized electrical demand over more than 10 years(2012-2022) is shown in figure 3.4.

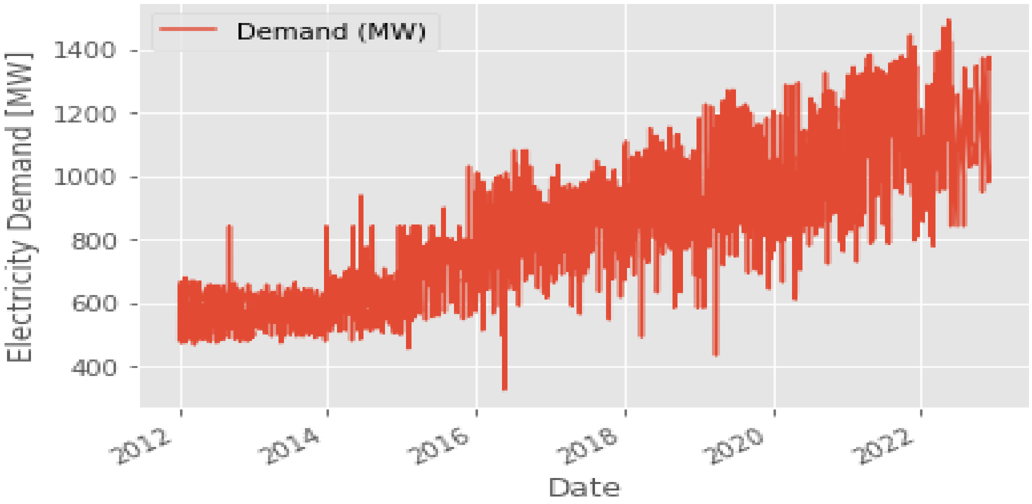


Figure 3.3: Electricity demand of 10 years(2012-2022)

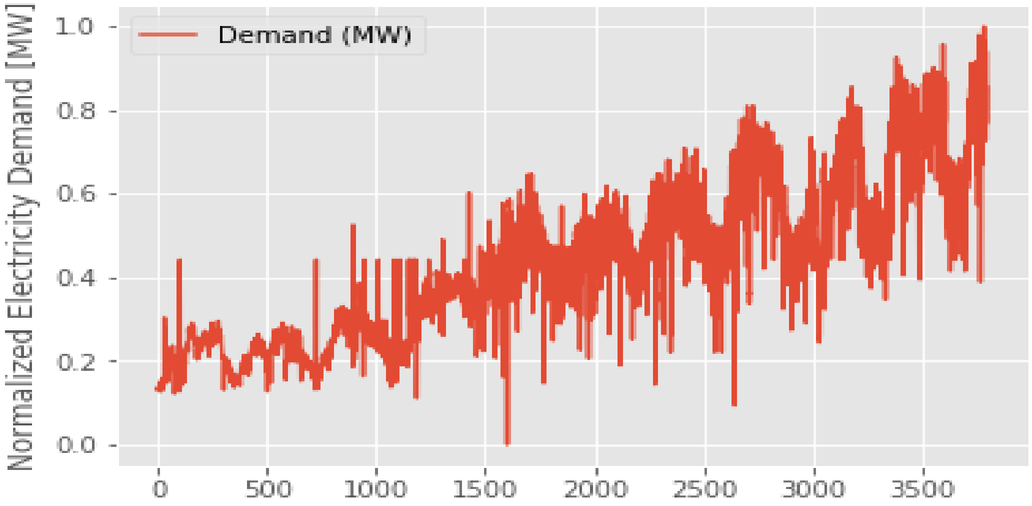


Figure 3.4: Normalized electricity demand of 10 years(2012-2022)

Figure 3.5 depicts the histogram digram of demand distribution.

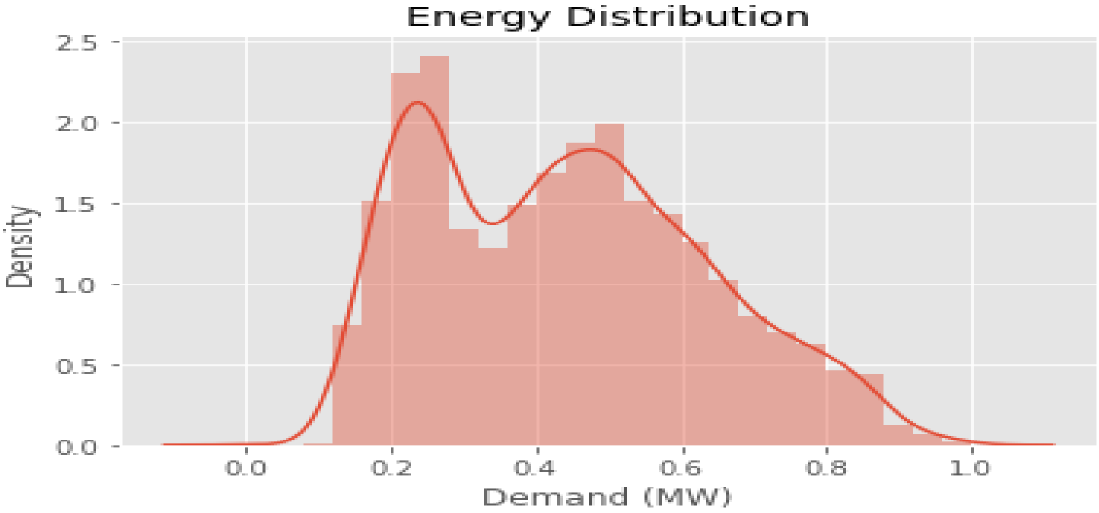


Figure 3.5: Electricity demand distribution

3.4 Machine Learning Algorithm

With the aid of training data and without explicit guidance, machine learning is the process of learning and adapting. Machine learning approaches can be broadly divided into two categories.

1. Supervised Learning:The method of learning from level data when the output is provided for the input data is known as supervised learning. And the algorithms perform in accordance.

2. Unsupervised Learning:Unsupervised learning is a type of learning where the training or operation is carried out on a dataset that has not been labeled. The method puts comparable data types together in order to identify aspects that are similar.

Regardless, the labeled data serve as the foundation for our model. It is possible to formulate the issue as a supervised learning problem. Two types of supervised learning algorithms exist.

1. Classification :The categorical output values used by classification algorithms allow each output to be allocated to a class. For discrete output values, the methods are effective.

2. Regression :Continuous regressive output values are what regression algorithms operate on. The output in this case is a continuous value.

The continuous output is predicted by our model. Therefore, it can be claimed that the issue is a regression one. We'll discuss about various machine learning regression techniques in the next subsections.

3.4.1 Linear Regression

Regression is carried out through the supervised machine learning method known as the linear regression algorithm. It merely determines whether the input variable and the output variable have a linear connection. Multivariable regression is the term used when their connection is not linear. By assuming a linear relationship between a dependent variable (y) and an independent variable (x), linear regression is the modeling technique that is most frequently utilized.The equation for linear regression is defined as follows [4]:

$$y = a_0 + a_1 x$$

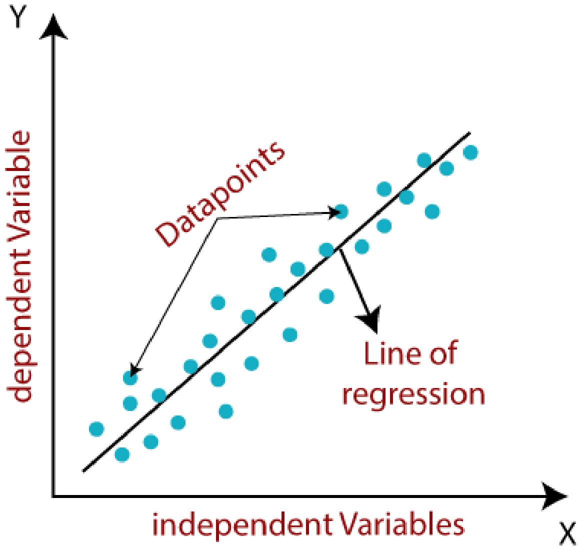


Figure 3.6: Linear Regression

3.4.2 Decision tree

The supervised learning algorithms group includes the decision-tree algorithm. It works with output variables that are categorized and continuous. Decision tree regression trains a model in the form of a tree to predict data in the future and generate useful continuous output by observing the properties of an item. Continuous output denotes the absence of discrete output, i.e., output that is not only represented by a discrete, well-known set of numbers or values.

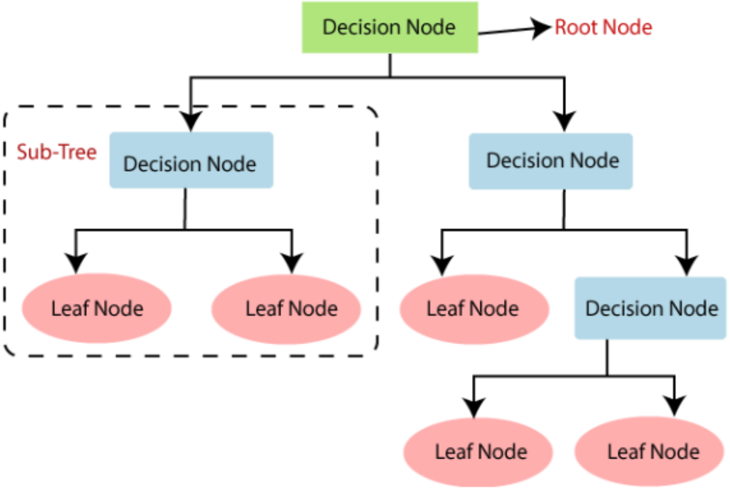


Figure 3.7: Decision Tree

3.4.3 Random Forest

One of the most effective supervised learning algorithms is random forest, which is capable of both classification and regression tasks. The Random Forest regression is an ensemble learning technique that mixes various decision trees and predicts the final result based on the average of each tree's output. The base is a set of merged decision trees. Aggregated decision trees are executed in parallel by Random Forest using the Bagging or Bootstrap Aggregation ensemble learning technique. By generating random subsets of the dataset, Random Forest regression allows us to prevent overfitting in the model.

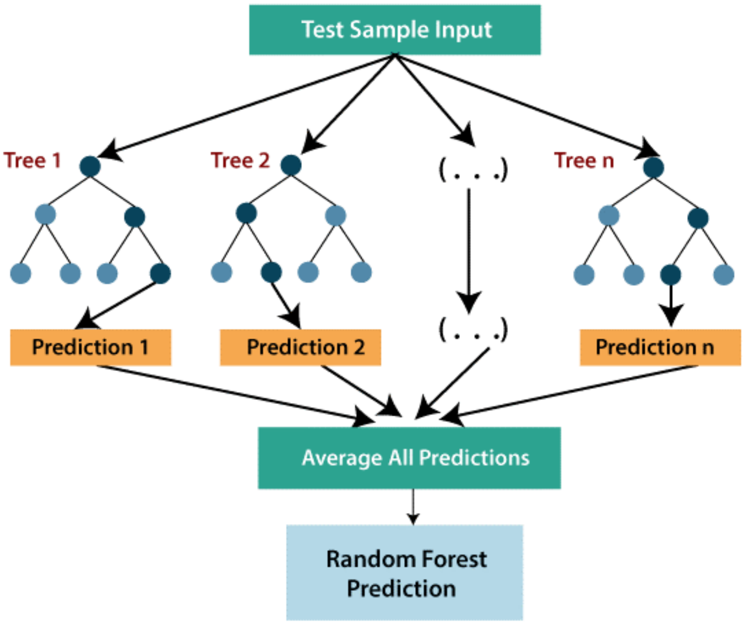


Figure 3.8: Random Forest

3.4.4 Support Vector Regression

Regression techniques include Support Vector Machine. It maintains all crucial characteristics, characterizing the algorithm with the greatest margin. With only minor modifications, the Support Vector Regression Method applies the SVM's classification assumptions. The result is a real number. As a result, it is challenging to forecast the information given all the potential outcomes. Regression uses tolerance or epsilon to approximate the SVM that was requested to solve the problem. There is also a more complex explanation. The algorithm is much difficult to understand. The main aim is to personalize the hyper plane and reduce the mistake. By accepting some mistake, the hyper plane exceeds the margin.[22]

3.4.5 K Nearest Neighbours

A straightforward supervised machine learning approach is the K-Nearest Neighbors algorithm. Both classification and regression issues are resolved by it. Understanding and applying it are simple. Its main flaw is that as data sizes increase, it gets slower. The KNN algorithm believes that similar things can be found nearby. It implies that objects that are similar are located close to one another. KNN

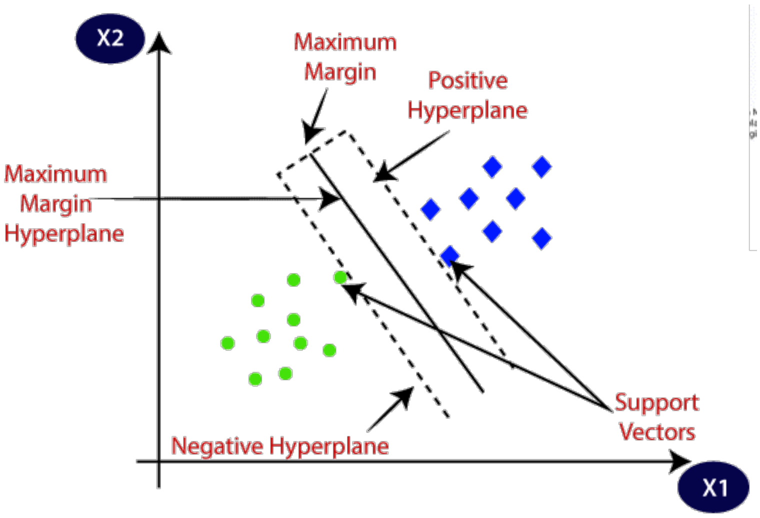


Figure 3.9: Support Vector Regression

chooses the label with the highest frequency in the case of classification or averages the labels in the case of regression after calculating the distances between a query and all the examples in the data that are closest to it.[22]

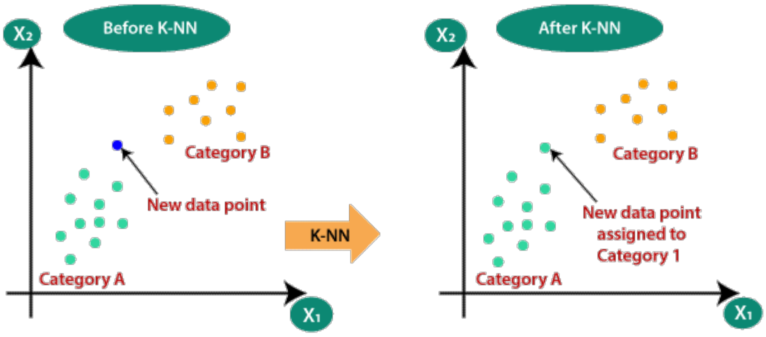


Figure 3.10: K Nearest Neighbours

3.5 Implementation

1. Loading the daraset: The dataset is initially read from the CSV file. Before being used for a further step, the data entries from the dataset are analyzed based on their features.

2. Working with datetime: In our dataset, the date was in a string format. In order to work with it, we need to convert the dates into datetime format. To accurately train the machine learning models, we changed the column type from a string to a datetime format using the pd.to datetime() function.

3. Feature Engineering: In this step, we need to preprocess the raw data to make it suitable for a machine learning model.In our dataset, some missing values were found; we replaced them by taking the mean value.We have also applied the normalization technique to our dataset because our dataset's features have a different range of values. And finally, the data are visualized to see the correlation of different features in our dataset.

4. Data split: In this step, data are split into train(80%) and test(20%) set. we used the cross validation method to assess how well our machine learning models perform on unseen data. In the 5-fold cross-validation (CV) procedure, typically involves dividing the data into k folds at random; in our case, k = 5. After that, k-1 folds are used to train the model, and only one fold is used to test the model.This process is repeated k times.However, all data in this work are first divided into training and testing datasets, with a training dataset being used for cross-validation.

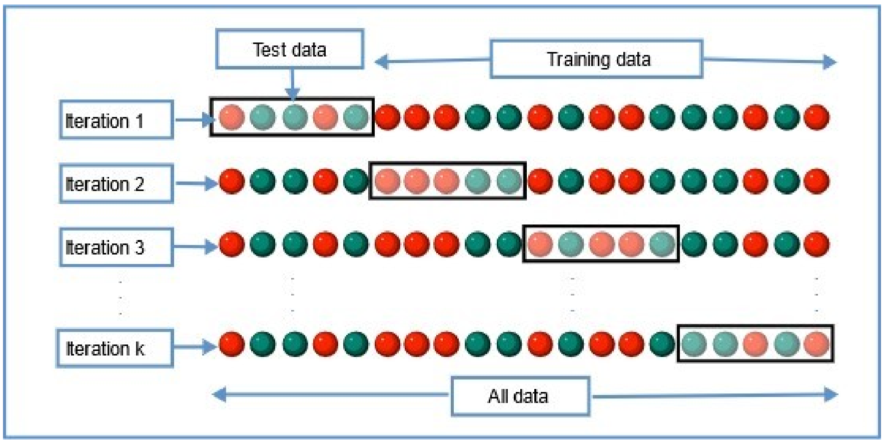


Figure 3.11: Cross Validation

5. Build the model: We will allow our model to observe the target values Demand (MW), along with other features, during the training session. By doing so, we can make sure that our model is developing the knowledge necessary to predict the target value (Demand (MW)) from features other than the target value. The model gradually discovers any correlations between the features and the target value during the training phase. For it comes time to evaluate the model, it no longer has access to the target values and instead relies entirely on the characteristics of the test set when making

predictions. The accuracy of the model is then calculated by comparing the predictions to the actual values of the target from the test set.

6. Model Evaluation: In this step, we will discuss how our machine learning algorithms performed when compared to common regression evaluation measures. Regression analysis is used to predict the intensity of electricity demand within a particular range. The regression metrics MAE, MSE, and RMSE are employed to assess the predicted outcomes, which are elaborately discussed in the result section.

7. Compare Performance: In this step compare the performance of the models based on their accuracy for selecting the best one

3.6 Conclusion

The theoretical analysis of several machine learning models and machine learning in general has been outlined in this chapter. Then, we discussed how we would implement our proposed approach.In the next section, we will go over the results and findings of our work.

Chapter 4

Results and Discussions

4.1 Introduction

The comprehensive explanation for our approach has been given in the previous chapter.One of the most important components in any ML model is performance. The more useful a model is, the better it performs.This chapter aims to demonstrate the comparative evaluation of our proposed model. We'll also mention more discussion.

4.2 Dataset Description

We have collected the dataset for our analysis from several websites. There are 3800 data in the dataset, and there are 8 feature values. The dataset sample is shown in Figure. The description of the features are given below:

Time factor play a major role in forecasting electricity demand. Different weekdays exhibit different patterns in the electricity distribution. Noticeable variances exist between the load on weekdays and weekends, and these differences are significant for load forecasting.During the weekend demand declines, especially on Saturday, the second day of the weekend, and then sharply increases on Sunday.According to the monthly historical analysis of load demand, the power consumption rate behaves in a certain way where it is prominently greater in summer than in winter. If temperature sensitivity is considered, load forecasting during the summer and winter seasons can be significantly improved.[23]

- **Demand data:** Daily electrical demand.

- **Weather Data:** One of the most crucial independent variables for forecasting power is the weather.[24]

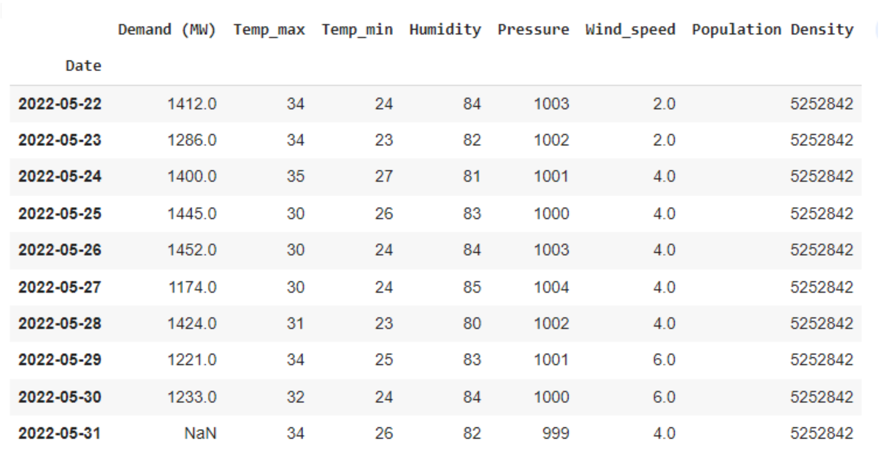
  Temperature: During the summer, the electricity demand is about twice as high as the winter electricity demand. In Chittagong, the coldest part of the year begins in early November. Then the temperature drops, and mild weather conditions lower the amount of electricity on the power network.

  Humidity: Humidity is another weather element that affects the total load curve. It is known as relative humidity in real life and is expressed in percentage numbers. A typical finding is that while humidity has no effect on the actual temperature, it can affect how people perceive it to be. However, humidity can make a temperature of 33°C appear harsher than, say, 40°C.

  Pressure: The temperature drops when air pressure drops. Additionally, it explains why air becomes colder at higher elevations due to decreased pressure. A low-pressure area with stormy, rainy weather is headed your way, according to falling pressure. Then the temperature drops, also decreasing the electricity demand.

  Wind Speed: A windless day with an air temperature of 15 degrees has no wind effect. 15 degrees is how it feels. A 20-mph wind, on the other hand, makes a 15-degree air temperature feel like only 2! The procedure grows more effective the faster the wind blows, making it appear colder and colder.Then the demand for electricity decreased.

- **Population Density:** Rural to urban migration is the main cause of this population growth. Population growth has been named as the main cause of the temperature rise and subsequent increase in power use. This essay seeks to explain and examine the rise in electricity use in this metropolis brought on by rising temperatures. Intensely hot, muggy, and rusty describe Dhaka. So, for cooling purposes, it uses a lot of electricity. In spite of this, a significant portion of its population is growing every day.[24]



	Demand (MW)	Temp_max	Temp_min	Humidity	Pressure	Wind_speed	Population Density
Date							
2022-05-22	1412.0	34	24	84	1003	2.0	5252842
2022-05-23	1286.0	34	23	82	1002	2.0	5252842
2022-05-24	1400.0	35	27	81	1001	4.0	5252842
2022-05-25	1445.0	30	26	83	1000	4.0	5252842
2022-05-26	1452.0	30	24	84	1003	4.0	5252842
2022-05-27	1174.0	30	24	85	1004	4.0	5252842
2022-05-28	1424.0	31	23	80	1002	4.0	5252842
2022-05-29	1221.0	34	25	83	1001	6.0	5252842
2022-05-30	1233.0	32	24	84	1000	6.0	5252842
2022-05-31	NaN	34	26	82	999	4.0	5252842

Figure 4.1: Dataset Sample

4.3 Feature Selection

An illustration of a correlation matrix that shows the relationship between various features is called a correlation heatmap. Correlation can have any value between -1 and 1. A weaker linear relationship between the two variables is indicated by values that are closer to zero. When the values are close to 1 the variables are said to be more positively correlated whereas values close to -1 are said to be more negatively correlated. The darkest shade of blue in the heatmap typically represents the positive correlation between two or more variables, whereas the pure white shade typically represents the negative correlation.

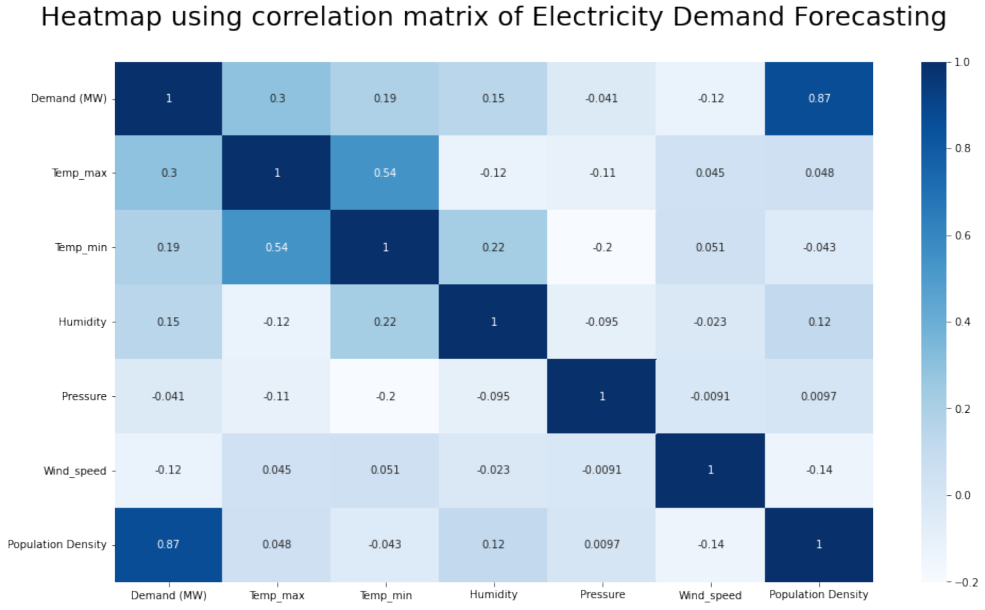


Figure 4.2: Correlation Heatmap

4.4 Impact Analysis

Our work also has some impacts. They are as follows:

4.4.1 Social Impact

Long-term planning and decision-making on future generation and transmission predictions are made easier when organizations are aware of the future load, which reduces risks for service providers. Electricity utilities benefit from the ability to skillfully build power plants because they have an understanding of future consumption or load demand.

4.4.2 Ethical Impact

The world energy system also brings up significant problems with justice. There, the unfair distribution of the effects of our energy systems is emphasized. The majority of the harm is caused by developed nations, while the adverse impacts are more likely to affect developing nations.The energy was divided fairly based on advanced electricity demand prediction.

4.5 Evaluation of Performance

The findings from the models are determined together with common evaluation metrics like MAE, MSE, and RMSE in the chapter on evaluation. The best model is picked based on the results received.[25]

- **Mean Absolute Error (MAE)** The mean absolute error(MAE) is the most fundamental regression error statistic to understand . In order to prevent negative and positive residuals from canceling out, we shall calculate the residual for each data point using just their absolute values. Then, we average all of these residuals. The typical residual magnitude is effectively described by MAE. The formula for MAE is as follows:

$$MAE = \frac{1}{n} \sum |y - y'|$$

Where,

n = number of data points

y= actual output value

y'= predicted output value

|y-y'| = absolute value of residual

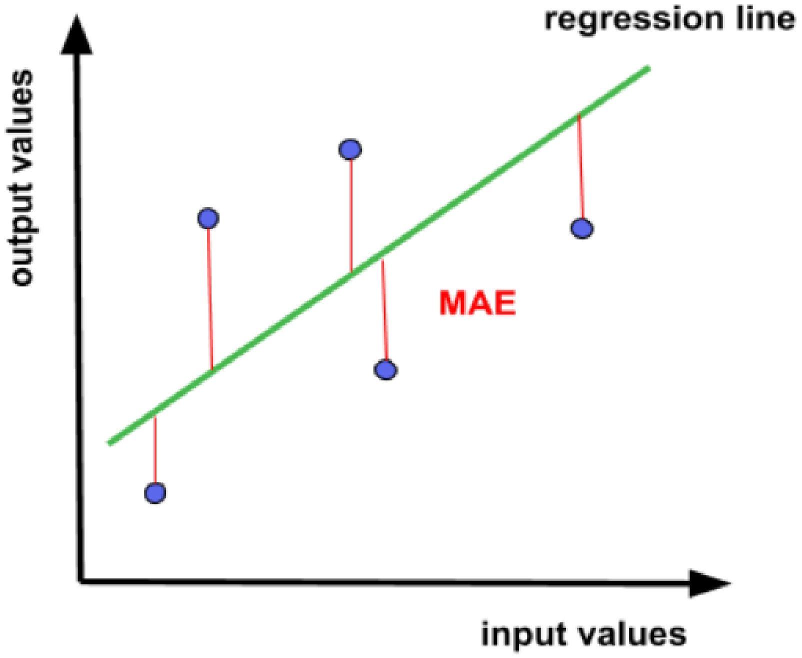


Figure 4.3: Graphical representation of MAE

The data are represented by the blue points, while the green line is the best-fitting prediction model. Utilizing the absolute value of the residual, underperformance or overperformance can be decreased in MAE. When the MAE is low, the model is thought to be stable, whereas when it is high, the model can have problems in some areas. A performance prediction model with an MAE of zero will be the most accurate.

- **Mean Square Error (MSE)** MSE is determined by averaging the squared differences between the actual value and the projected or estimated value provided by the regression model (line or plane). Another name for it is mean squared deviation (MSD). Mathematically, it is expressed as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y - y')^2$$

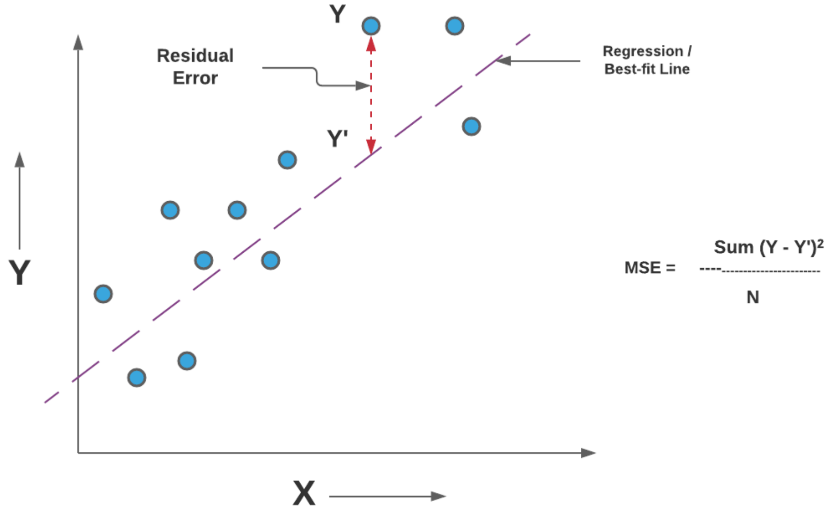


Figure 4.4: Graphical representation of MSE

The value of MSE is always positive. A value close to zero will represent better quality of the estimator/predictor (regression model). An MSE of zero (0) represents the fact that the predictor is a perfect predictor.

- **Root Mean Squared Error (RMSE)** The square root of MSE is used to calculate RMSE. The Root Mean Square Deviation is another name for RMSE. It measures the average error magnitude and is interested in the variations from the true value. When the RMSE number is zero, the model fits the data perfectly. The model's performance and predictions improve with decreasing RMSE. A greater RMSE denotes a significant departure from the ground truth in the residual. RMSE can be applied to a variety of features because it aids in determining whether a feature enhances model prediction or not. Mathematically, it is expressed as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y - y')^2}$$

4.5.1 Evaluation of Linear Regression model and results

Figure 4.5 shows a comparison between the actual and expected electricity usage based on our model. We can see that an LR virtually predicts data points that match the pattern of electricity usage. In addition, as Table 4.1 illustrates, it produces lower MAE, MSE, and RMSE scores.It was observed that linear regression models give 82.53% accuracy. The Evaluation results of Linear Regression model are as follows:

Table 4.1: Prediction Errors and accuracy score of LR

Prediction Errors and accuracy score of LR	
Mean Absolute Error	0.0582
Mean Square Error	0.0057
Root Mean Square Error	0.0761
Accuracy(%)	82.53%

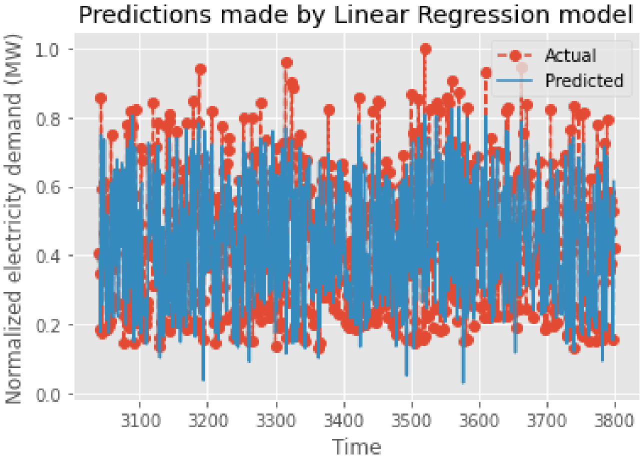


Figure 4.5: Graphical representation of Actual vs predicted results

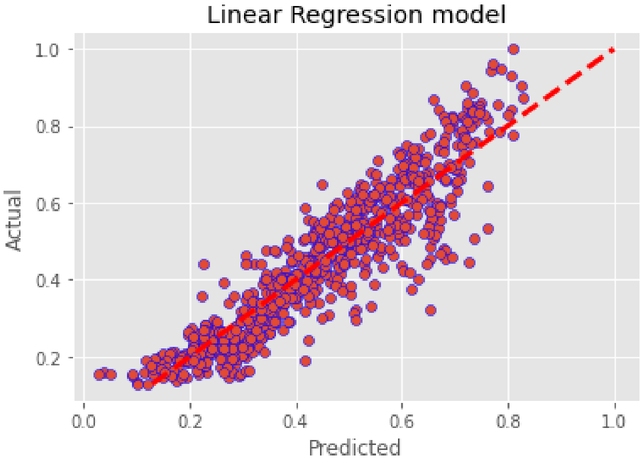


Figure 4.6: Actual Vs Predicted observations for Linear Regresion

4.5.2 Evaluation of Decision Tree Regression model and results

Figure 4.7 shows a comparison between the actual and expected electricity usage based on our model. We can see that an DT virtually predicts data points that match the pattern of electricity usage. In addition, as Table 4.2 illustrates, it produces lower MAE, MSE, and RMSE scores.It was observed that Decision Tree regression models give 74.16% accuracy. The Evaluation results of Decision Tree Regression model are as follows:

Table 4.2: Prediction Errors and accuracy score of DT

Prediction Errors and accuracy score of DT	
Mean Absolute Error	0.0615
Mean Square Error	0.0081
Root Mean Square Error	0.0903
Accuracy(%)	74.16%

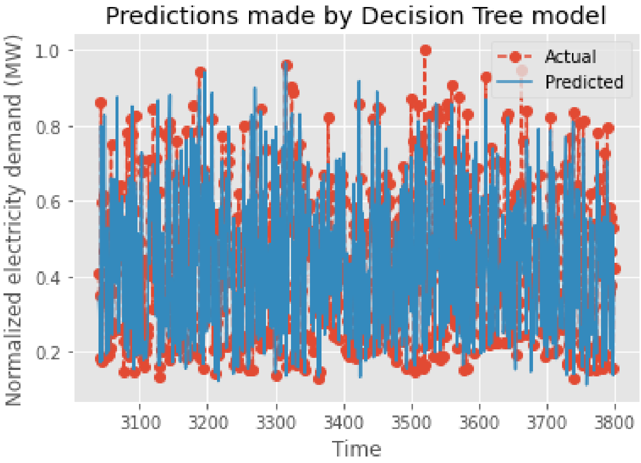


Figure 4.7: Graphical representation of Actual vs predicted results

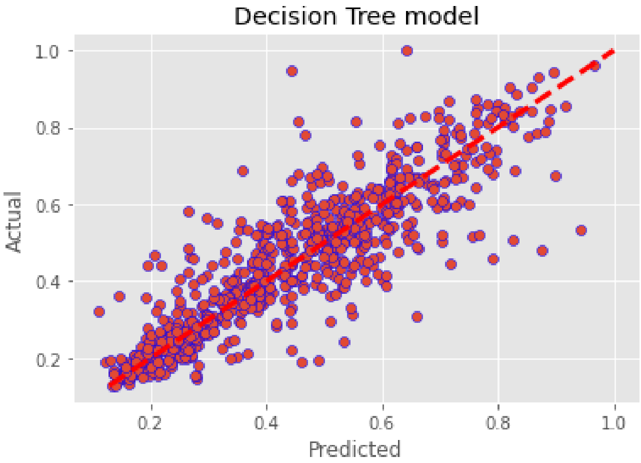


Figure 4.8: Actual Vs Predicted observations for Decision Tree Regression

4.5.3 Evaluation of Random Forest Regression model and results

Figure 4.9 shows a comparison between the actual and expected electricity usage based on our model. We can see that an RF virtually predicts data points that match the pattern of electricity usage. In addition, as Table 4.3 illustrates, it produces lower MAE, MSE, and RMSE scores.It was observed that Random forest regression models give 86.24% accuracy. The Evaluation results of Random Forest Regression model are as follows:

Table 4.3: Prediction Errors and accuracy score of RF

Prediction Errors and accuracy score of RT	
Mean Absolute Error	0.0450
Mean Square Error	0.0042
Root Mean Square Error	0.0652
Accuracy(%)	86.24%

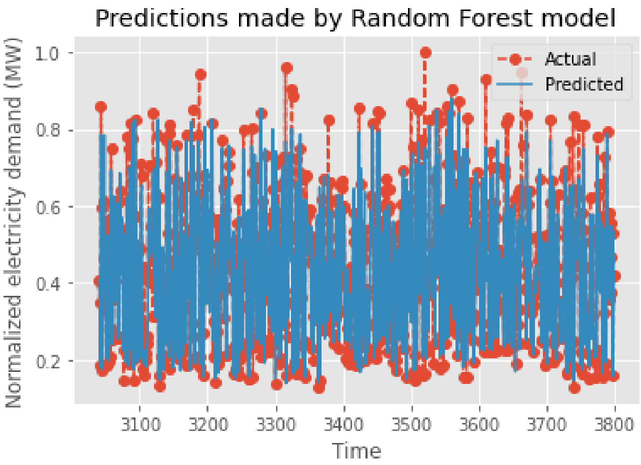


Figure 4.9: Graphical representation of Actual vs predicted results

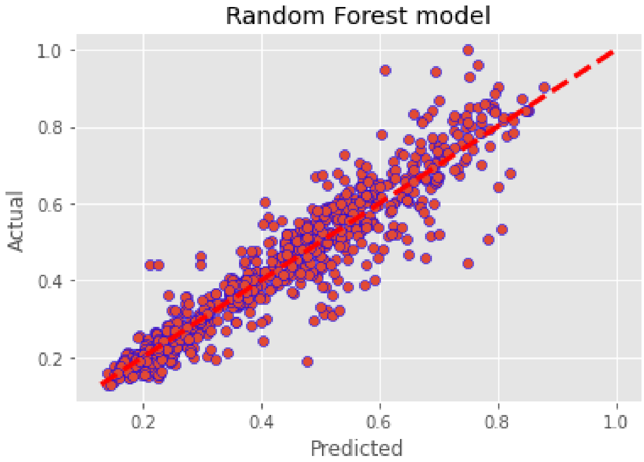


Figure 4.10: Actual Vs Predicted observations for Random Forest Regression

4.5.4 Evaluation of Support Vector Regression model and results

Figure 4.11 shows a comparison between the actual and expected electricity usage based on our model. We can see that an SVR virtually predicts data points that match the pattern of electricity usage. In addition, as Table 4.4 illustrates, it produces lower MAE, MSE, and RMSE scores.It was observed that Support vector regression models give 84.52% accuracy. The Evaluation results of Support Vector Regression model are as follows:

Table 4.4: Prediction Errors and accuracy score of SVR

Prediction Errors and accuracy score of SVR	
Mean Absolute Error	0.0540
Mean Square Error	0.0052
Root Mean Square Error	0.0724
Accuracy(%)	84.52%

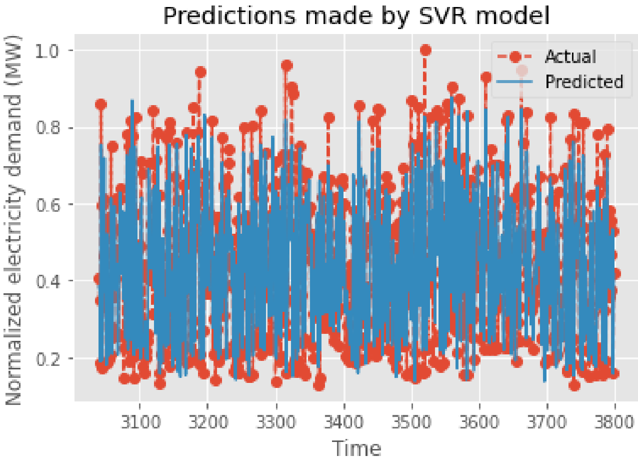


Figure 4.11: Graphical representation of Actual vs predicted results

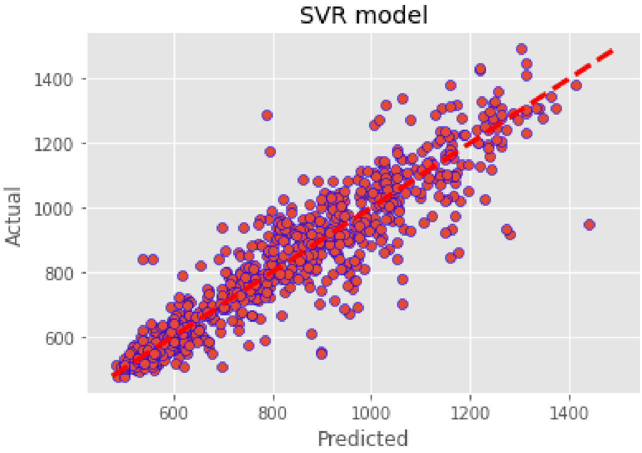


Figure 4.12: Actual Vs Predicted observations for Support Vector Regression

4.5.5 Evaluation of K Nearest Neighbours model and results

Figure 4.13 shows a comparison between the actual and expected electricity usage based on our model. We can see that an KNN virtually predicts data points that match the pattern of electricity usage. In addition, as Table 4.5 illustrates, it produces lower MAE, MSE, and RMSE scores.It was observed that K Nearest Neighbours models give 74.16% accuracy.

The Evaluation results of K Nearest Neighbours model are as follows:

Table 4.5: Prediction Errors and accuracy score of KNN

Prediction Errors and accuracy score of KNN	
Mean Absolute Error	0.0534
Mean Square Error	0.00612
Root Mean Square Error	0.0782
Accuracy(%)	81.15%

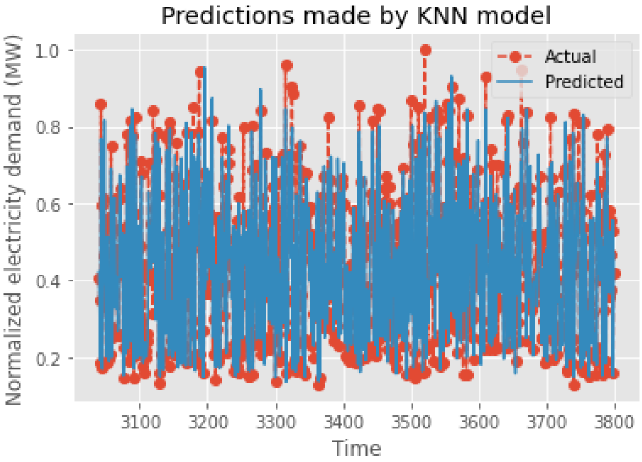


Figure 4.13: Graphical representation of Actual vs predicted results

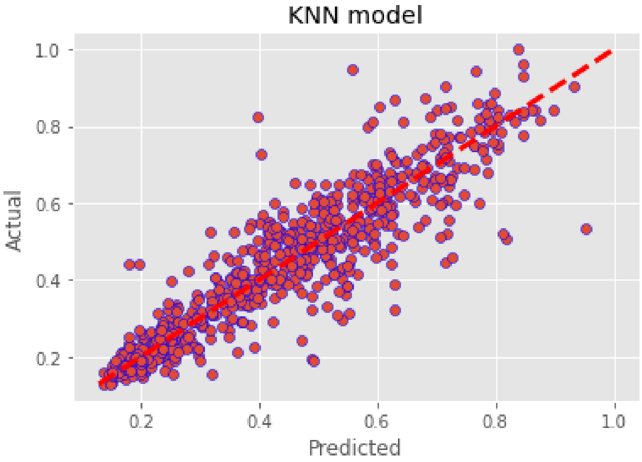


Figure 4.14: Actual Vs Predicted observations for K Nearest Neighbours

4.6 Comparison and Result Summary

A comparison of various machine learning-based demand forecasting models is shown in Table 4.6. Future electricity demand can be predicted using these models. The findings show that, in comparison to earlier investigations, our approach provides significantly better predictions. In this work, many machine learning methods are employed. Table 4.6 displays a thorough performance comparison of different algorithms. The outcomes obtained from the analysis represent that the Random Forest ML approach outperforms the other conventional ML approaches in terms of accuracy, MAE, MSE, and RMSE error rates since it has the highest accuracy (86.24%) and lowest MAE, MSE, and RMSE error rates (0.0450, 0.0042, and 0.0652, respectively) compared to other discussed conventional ML approaches.

The result presented in Table 4.6 demonstrates the obtained accuracy and error values generated in the applied algorithms.

Table 4.6: Comparison among Various Algorithms

Algorithms	**Accuracy**	**MAE**	**MSE**	**RMSE**
Linear Regression	82.53%	0.0582	0.0057	0.0761
Decision Tree	74.16%	0.0615	0.0081	0.0903
Random Forest	86.24%	0.0450	0.0042	0.0652
Support Vector Regression	84.52%	0.0540	0.0052	0.0724
K Nearest Neighbours	81.84%	0.0534	0.0061	0.0782

We plotted all the accuracy and the error values of the used algorithms in a bar plot to better visualize the outcome comparison. As seen in figures 4.15 and 4.16, Random Forest outperforms every other regression analysis technique.

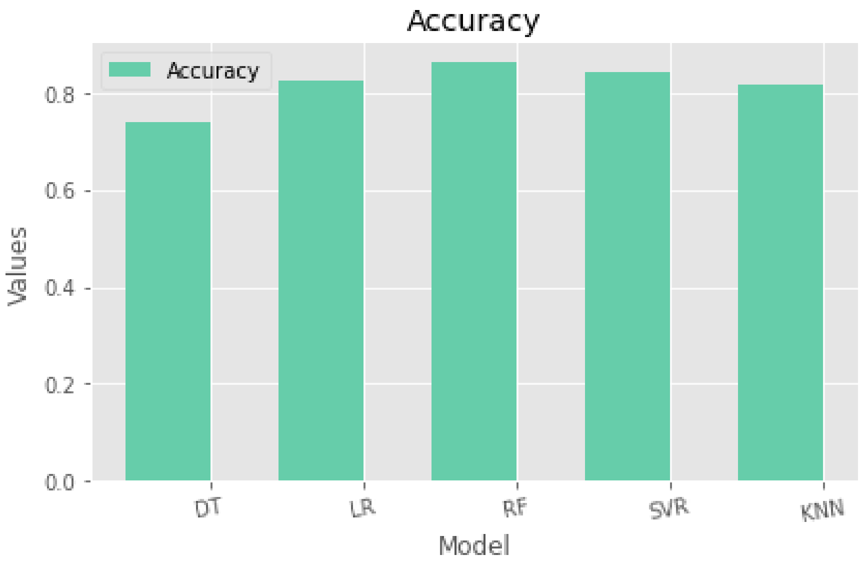


Figure 4.15: Accuracy

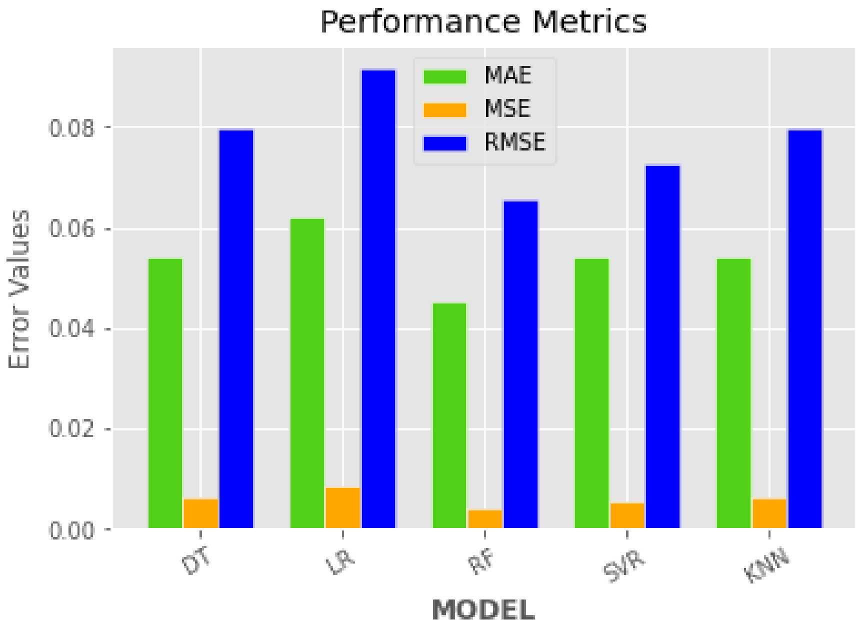


Figure 4.16: Error comparision of the models

4.7 Conclusion

The overall effectiveness of the proposed method has been analyzed in this section. The dataset's detailed explanation has been provided. Also demonstrated is the feature selection procedure. The model has also been assessed using the evaluation metrics MAE, MSE, and RMSE.Here, the best model has been chosen as Random Forest Regression.

Chapter 5

Conclusion

5.1 Conclusion

Energy shortages could seriously impede Bangladesh's economic development given its rapidly expanding economy. Although electricity demand has been a hot topic for some time, there haven't been many in-depth studies done on the state of our energy production. Artificial intelligence and other contemporary computation techniques can undoubtedly produce useful results that can reduce the gap between supply and demand.

In our thesis, we successfully prepared a dataset of historical demand data for Chittagong, Bangladesh, covering more than ten years (January 2012 to June 2022). The data was taken from the Bangladesh Power Development Board website. Additionally, other important features were collected manually from different websites. To demonstrate the system's dependability, we tested it using testing data sets and accurate error computation, with the Random Forest Regression Model outperforming competing models.

Overall, outcomes show that the proposed ML approach outperforms the conventional ML approaches in terms of accuracy, MAE, MSE, and RMSE error rates since it has the highest accuracy (86.24%) and the lowest MAE, MSE, and RMSE error rates (0.0450, 0.0042, and 0.0652, respectively) compared to other discussed conventional ML approaches. The RF method provides better performance due to the bagging algorithm, which is the foundation of Random Forest, which employs ensemble learning. The output of all the trees is combined once as many trees as possible have been created on the subset of the data. In doing so, it lessens the issue of overfitting in decision trees, as well as reduces variation and improves accuracy. Therefore, the RF method has the ability to predict the future demand for electricity more accurately and also be able to plan ahead for maintenance and load distribution using the obtained results, which can help electricity generation companies.

5.2 Future Work

Bangladesh’s economy is exhibiting aspirational growth in many areas, and this expansion requires a reliable electricity supply. We intend to continue working on our work soon with this viewpoint in mind. Our goal is to create a hybrid system that can recognize significant features to see if we can more accurately predict load. We intend to gain relevant data and forecast Bangladesh’s electricity demand on a region basis.

References

[1] S. Masum, Y. Liu and J. Chiverton, 'Multi-step time series forecasting of electric load using machine learning models,' in *International conference on artificial intelligence and soft computing*, Springer, 2018, pp. 148–159 (cit. on p. 1).

[2] N. Phuangpornpitak and W. Prommee, 'A study of load demand forecasting models in electric power system operation and planning,' *GMSARN International Journal*, vol. 10, no. 2016, pp. 19–24, 2016 (cit. on p. 2).

[3] Z. Çamurdan and M. C. Ganiz, 'Machine learning based electricity demand forecasting,' in *2017 International Conference on Computer Science and Engineering (UBMK)*, 2017, pp. 412–417. DOI: `10.1109/UBMK.2017.8093428` (cit. on p. 6).

[4] N. Shabbir, R. Ahmadiahangar, L. Kütt and A. Rosin, 'Comparison of machine learning based methods for residential load forecasting,' in *2019 Electric power quality and supply reliability conference (PQ) & 2019 symposium on electrical engineering and mechatronics (SEEM)*, IEEE, 2019, pp. 1–4 (cit. on pp. 6, 16).

[5] A. R. Anik and S. Rahman, 'Commercial energy demand forecasting in bangladesh,' *Energies*, vol. 14, no. 19, p. 6394, 2021 (cit. on p. 6).

[6] Z. Liu *et al.*, 'Accuracy analyses and model comparison of machine learning adopted in building energy consumption prediction,' *Energy Exploration & Exploitation*, vol. 37, no. 4, pp. 1426–1451, 2019 (cit. on p. 6).

[7] K. Gajowniczek and T. Ząbkowski, 'Two-stage electricity demand modeling using machine learning algorithms,' *Energies*, vol. 10, no. 10, p. 1547, 2017 (cit. on p. 7).

[8] P. K. Jain, W. Quamer and R. Pamula, 'Electricity consumption forecasting using time series analysis,' in *International Conference on Advances in Computing and Data Sciences*, Springer, 2018, pp. 327–335 (cit. on p. 7).

[9] A. Mati, B. Gajoga, B. Jimoh, A. Adegobye and D. Dajab, 'Electricity demand forecasting in nigeria using time series model,' *The Pacific Journal of Science and Technology*, vol. 10, no. 2, pp. 479–485, 2009 (cit. on p. 7).

[10] C. Behm, L. Nolting and A. Praktiknjo, 'Forecasting long-term electricity demand time series using artificial neural networks,' 2020 (cit. on p. 7).

[11] M. R. Islam, A. Al Mamun, M. Sohel, M. L. Hossain and M. M. Uddin, 'Lstm-based electrical load forecasting for chattogram city of bangladesh,' in *2020 International Conference on Emerging Smart Computing and Informatics (ESCI)*, IEEE, 2020, pp. 188–192 (cit. on p. 7).

[12] A. T. Eseye, M. Lehtonen, T. Tukia, S. Uimonen and R. J. Millar, ‘Machine learning based integrated feature selection approach for improved electricity demand forecasting in decentralized energy systems,’ *IEEE Access*, vol. 7, pp. 91 463–91 475, 2019 (cit. on p. 7).

[13] T. Usha and S. A. A. Balamurugan, ‘Seasonal based electricity demand forecasting using time series analysis,’ *Circuits and Systems*, vol. 7, no. 10, pp. 3320–3328, 2016 (cit. on p. 8).

[14] J. Deng and P. Jirutitijaroen, ‘Short-term load forecasting using time series analysis: A case study for singapore,’ in *2010 IEEE Conference on Cybernetics and Intelligent Systems*, IEEE, 2010, pp. 231–236 (cit. on p. 8).

[15] M. R. Patel, R. B. Patel and N. A. Patel, ‘Electrical energy demand forecasting using time series approach,’ 2020 (cit. on p. 8).

[16] F. Rabbi, S. U. Tareq, M. M. Islam, M. A. Chowdhury and M. A. Kashem, ‘A multivariate time series approach for forecasting of electricity demand in bangladesh using arimax model,’ in *2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI)*, IEEE, 2020, pp. 1–5 (cit. on p. 8).

[17] A. J. P. Delima, ‘Application of time series analysis in projecting philippines’ electric consumption,’ *International Journal of Machine Learning and Computing*, vol. 9, no. 5, pp. 694–699, 2019 (cit. on p. 8).

[18] A. W. Yao, S. Chi and J. Chen, ‘An improved grey-based approach for electricity demand forecasting,’ *Electric Power Systems Research*, vol. 67, no. 3, pp. 217–224, 2003 (cit. on p. 8).

[19] *Electricity demand of chittagong*, `http://119.40.95.168/bpdb/area_wise_demand?fbclid=IwAR3DQP2OXz8eOnJlRoLsX_fSx6y1tG8D4_M7RhTyBf4EMteykYGx0X59FOs` (cit. on p. 12).

[20] *Past weather history of chittagong*, `https://www.timeanddate.com/weather/bangladesh/chittagong/historic?month=1&year=2012` (cit. on p. 12).

[21] *Chittagong population*, `https://worldpopulationreview.com/world-cities/chittagong-population` (cit. on p. 12).

[22] T. R. Jena, S. S. Barik and S. K. Nayak, ‘Electricity consumption & prediction using machine learning models,’ *Muktshabd*, vol. 9, no. 6, pp. 2804–2818, 2020 (cit. on pp. 18, 19).

[23] B. N. Mahmud, Z. Ferdoush and L. T. Mim, ‘Modelling and forecasting energy demand of bangladesh using ai based algorithms,’ Ph.D. dissertation, Brac University, 2019 (cit. on p. 22).

[24] A. Istiaque, S. I. Khan *et al.*, ‘Impact of ambient temperature on electricity demand of dhaka city of bangladesh,’ *Energy and Power Engineering*, vol. 10, no. 07, p. 319, 2018 (cit. on p. 23).

[25] S. A. Shaik, ‘Forecasting the electricity demand using machine learning algorithms,’ Ph.D. dissertation, Dublin Business School, 2020 (cit. on p. 25).