Thesis for M.SC in SWE

A Hybrid Approach for Improved Electric Short Term Load Forecasting

Submitted By

Md.Sakib Rahman
Student ID: 221-44-236

Supervisor

Dr. Md. Fazla Elahe
Assistant Professor & Associate Head
Department of Software Engineering

A thesis submitted in partial fulfillment of the requirement for the degree of

Master of Science in Software Engineering

Department of Software Engineering
Daffodil International University
Dhaka, Bangladesh

Semester-Spring-2023

APPROVAL

This thesis/project/internship titled on "A Hybrid Approach for Short Term Load Forecasting", submitted by Md. Sakib Rahman, ID: 221-44-236 to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Masters of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

-- Chairman

Dr. Imran Mahmud
Associate Professor and Head
Department of Software Engineering
Daffodil International University

Fazla Elahe 18.08.23
-- Internal Examiner 1

Dr. Md. Fazla Elahe
Assistant Professor and Associate Head
Department of Software Engineering
Daffodil International University

-- Internal Examiner 2

Afsana Begum
Assistant Professor
Department of Software Engineering
Daffodil International University

-- External Examiner

Dr. Md. Sazzadur Rahman,
Associate Professor
Institute of Information Technology
Jahangirnagar University

Declaration

It hereby declare that this thesis has been done by Sakib Rahamn, ID: 221-44-236 under the supervision of Dr. Md. Fazla Elahe, Assistant Professor & Associate Head ,Department of Software Engineering, Daffodil International University. It also declare that neither this thesis nor any part of this has been submitted elsewhere for award of any degree.

....

MD.Sakib Rahman
ID:221-44-236
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Certified by:

Fazla Elahe
....

Dr. Md. Fazla Elahe
Assistant Professor & Associate Head
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Dedicated to My Parents

Abstract

For the daily management and planning of power grids, short-term load forecasting (STLF) is a crucial task in power systems. In STLF, the electrical load demand is forecasted from a few hours to several days in advance. For the power grid to remain stable and reliable, to avoid overloading or underutilizing power plants, and to maximize energy management tactics, accurate STLF is important.

Various STLF methods, including those based on machine learning (ML) and artificial intelligence (AI), have been developed over the years, in addition to more conventional statistical techniques. Artificial neural networks (ANNs), support vector machines (SVMs), autoregressive integrated moving averages (ARIMA), and deep learning (DL) models are a few well-liked methods. Each technique has advantages and disadvantages, and the best one to use depends on a number of variables, including the availability of data, the time horizon for forecasting, and the level of accuracy required. The complexity and dynamic nature of power systems, the inherent uncertainty and variability in load demand, and the influence of outside factors like weather and human behavior mean that STLF remains a difficult task despite recent advancements. To increase STLF techniques' precision, robustness, and suitability for use in various power system scenarios, more research is required.

Overall, STLF is a crucial task for ensuring the dependable and effective operation of power grids, and the development of precise and dependable STLF techniques is an ongoing research topic in the field of power systems engineering.

Key Words: STLF: Short Term Load Forecasting, NN: Neural Network, RF: Random Forest, LR: Linear Regression,SVR: Support Vector Machine, Hybrid Meth

Acknowledgment

In this very special moment, first and foremost I would like to express my heartiest gratitude to the almighty Allah for allowing me to accomplish this M.SC study successfully. I am really thankful for the enormous blessings that the Almighty has bestowed upon me not only during my study period but also throughout my life. In achieving the gigantic goal, I have gone through the interactions with and help from other people, and would like to extend my deepest appreciation to those who have contributed to this dissertation itself in an essential way.

I would like to take the opportunity to express my gratitude to my respected supervisor Dr. Md. Fazla Elahe sir. He has been a great influence throughout and I am grateful to him for not only his intellectual support, providing me with vast knowledge, but also for the moral support and proper guidelines that he gave me, encouraging me always to work harder, raise the bar and his honest quest to extract the best out of us. It is needless to say without his influence and support this dissertation would never be successful.

In the finishing note, I am thankful to my family and friends for their help and immense support in different aspects to make this dissertation a success.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter presents an overview of Short Term Load Forecasting including the proposed model in this area. Based on the discussion, motivations behind the proposed method are explained clearly in this thesis.

1.1 Overview

Short-term load forecasting (STLF) is a technique used to predict the amount of electricity required to meet demand in the near future, usually within the next 24 to 48 hours. STLF is important for companies and grid operators as it helps them anticipate changes in demand and plan accordingly to ensure robust and reliable electricity. There are three types of forecasts - short term, medium term and long term [1]. STLF is frequently based on statistical models that utilize historical data on electricity consumption, weather, and other variables that may affect demand. These models are frequently created using machine learning algorithms, such as neural networks and decision trees [2]. The reliability of STLF is dependent on the quality and availability of data, the sophistication of the statistical models employed, and the capacity to take into account unforeseen events like sudden weather changes or equipment failures. Despite these difficulties, STLF has grown in significance as a tool for managing the electrical grid and ensuring a consistent supply of electricity.

1.2 Contribution

Short-term load forecasting (STLF) can handle huge data volumes and complex patterns in the data that may be challenging to model using conventional statistical methods, machine learning which is used for short-term load forecasting[3]. The analysis of historical data on energy use and weather conditions by machine learning algorithms, such as neural networks and decision trees, can reveal patterns and relationships that are challenging to spot by hand. These algorithms can use this data to forecast future electricity demand based on the weather at the time and other variables. Utilities and grid operators can create more precise short-term load forecasting models using machine learning algorithms[4], which will enable them to better manage their electricity supply and prevent blackouts or other disruptions. Additionally, as more data becomes available, machine learning models can adapt and advance, making them suitable for forecasting applications that call for regular updates and modifications.

1.3 Motivation

The machine learning has been a well- known research point among specialist and researcher's of computer science .We are too much interest to conduct to identify problem or finds answers to uncertainties. Research is conducted because there is uncertainty about a phenomenon that either has, or has not occurred[5]. Research also aims to use the best method to solve problem, weather or not experiments are conduct the main purpose of writhing and publishing is to disseminate research finding and to share new knowledge with other researchers in the their machine learning or artificial intelligence fields[6]. We are analysis and try to attempt to give some calculation which is better then recent result. Generally this recent research motivated us to doing new research.

1.4 Organization of the Thesis

The dissertation is organized as follows:

Chapter 1 Introduction. In this chapter an introduction to the semantic segmentation is presented. The definition and importance are clearly introduced. After that, the dissertation focuses the contribution.

Chapter 2 Literature Review. This chapter first shows the related works of semantic segmentation around the world. Limitations of these methods and works are clearly addressed.

Chapter 3 Methodology. This chapter presents the main contribution of the thesis: the design and develop- ment of proposed methodology.

Chapter 4 Experimental Evaluation In this chapter, We are going to show all the performances and result analysis.

Chapter 5 Conclusion and Future Work. Finally, this chapter concludes the dissertation indicating the limitations and future works.

Chapter 2

Literature Review

In this chapter, before we start we have to clear some concept and first thing we need to know about is Short term load forcasting. What is short term load forcasting? The prediction of electrical power demand over a brief period, typically lasting from a few minutes to a few weeks, is known as short-term load forecasting (STLF). Given that it enables utilities to efficiently manage resources and prevent system instability, STLF is a crucial part of power system operations and planning.

Dealing with load data uncertainly and variability due to things like weather, holidays, and events is one of the main challenge in short term load forcasting (STLF)[7]. Researchers have proposed a number of hybrid models that combine statistical and AI/ ML techniques to address this issue, or they have used ensemble methods to combine multiple modles.

Overall, STLF is a highly active field of study, and there is a wealth of literature on it. Depending on the data characteristics, accuracy standards, and computational resources available, a method for STLF is chosen.

2.1 Background Study

When the need arose to precisely forecast the demand for electricity over a brief period, the background study of short-term load forecasting (STLF) was born. Early STLF models primarily relied on regression analysis, exponential smoothing, and moving averages as their foundation. These models relied on historical data and basic time series analysis to make predictions. These conventional statistical models started to have trouble correctly forecasting load demand as electricity systems grew more complicated and variable[8], [9]. As a result, more complex statistical models were created, including the seasonal decomposition of time series (STL), autoregressive integrated moving average (ARIMA), and other time series techniques.

The development of machine learning (ML) and artificial intelligence (AI) techniques in the 1990s provided STLF with new opportunities. The modeling of more intricate connections between load data and outside variables like the weather, holidays, and events was made possible by AI and ML techniques. Since they can model complex relationships and change in response to changing conditions, deep learning techniques like convolutional neural networks (CNNs) and Artificial neural networks (ANN) have become more and more popular for STLF[10].

The need for more precise and reliable load forecasting in an increasingly complex and dynamic electricity system has led to an overall evolution of the background study of STLF away from traditional statistical methods and toward more cutting-edge AI/ML techniques.

2.1.1 Machine Learning

Machine learning is the art of inspiring computers to act without being expressly customized. In the previous decade, machine learning has given us self-driving autos, handy discourse acknowledgment, viable web look, and a listlessly enhanced comprehension of the human genome. Machine learning is so inescapable today that you likely utilize it many times each day without knowing it[11]. Numerous specialists likewise think it is the most ideal approach to gain ground towards human-level AI . In this class, you will find out about the best machine learning systems, and pick up work on executing them and motivating them to work for yourself. Allthe more critically, you'll find out about the hypothetical underpinnings of learning, as well as pick up the commonsense know-how expected to rapidly and intensely apply these methods to new issues.The Python programming language is establish- ing itself as one of the most popular languages for scientific computing. Thanks toits high-level interactive nature and its maturing co system of scientific libraries, itis an appealing choice for algorithmic development and exploratory data analysis[1]. Yet, as a general-purpose language, itis increasingly used not only in academic settings but also in industry.Scikit-learnharnesses this rich environment to provide state-of-the-art implementations of many well known machine learning algorithms, while maintaining an easy-to-use interface tightly integrated with the Python language. This answers the growing need for statistical data analysis by non-specialists in the software and web industries, as well as in fields outside of computer-science,such as biology or physics. Scikit-learn differs from other machine learning toolboxes in Python for various reasons: i) it is distributed under the BSD license ii) it incorporates compiled code for efficiency, unlike MDP and pybrain

iii) it depends only on numpy and scipy to facilitate easy distribution, unlike pymvpa that has optional dependencies such as R and shogun, and iv) it focuses on imperative programming, unlike pybrain which uses a data-flow frame-work. While the package is mostly written in Python, it incorporates the C++ libraries LibSVM and LibLinear that provide reference implementations of SVMs and generalized linear models with compatible licenses. Binary packages are available on a rich set of platforms including Windows and any POSIX platforms. MACHINE LEARNING IN PYTHON Furthermore, thanks to its liberal license, it has been widely distributed as part of major free software distributions such as Ubuntu, Debian, Mandriva, Net BSD and Mac ports and in commercial distributions such as the En thought Python Distribution.

2.1.2 Types of Load Forecasting

There are several methods for load forecasting. Long term load forecasting, Medium term load forecastin, short term load forecasting[12].

1. **Long term load forecasting:** Long-term load forecasting (LTLF) is the process of estimating the future demand for electricity over a period of several years, typically spanning several decades. To make sure they have enough resources to meet the future demand for electricity, utilities and energy companies must plan and develop their infrastructure in accordance with LTLF. In LTLF models, a variety of variables that may have an impact on the demand for electricity are typically taken into account, including population growth, economic development, modifications to technology and way of life, climatic patterns, and governmental regulations[13]. Emerging technologies like distributed generation and electric vehicles may also be taken into account in these models to determine how they will affect the demand for electricity. LTLF models can be developed using a range of statistical, econometric, and machine learning

techniques. These models can be complex, and the accuracy of the forecasts can depend on the quality of the data and the assumptions made about the future trends and factors that influence the electricity demand.

2. **Medium term load forecasting:** The process of estimating the future demand for electricity over a time period of a few weeks to a few months is known as medium-term load forecasting (MTLF). In order to meet the anticipated demand for electricity, utilities and energy firms must plan and optimize the operation of their power systems. This is done while reducing costs, preserving reliability, and stabilizing the system[14]. A variety of variables, such as weather patterns, seasonal variations, holidays, events, and economic indicators are frequently taken into account by MTLF models when predicting how much electricity will be consumed. These models may also include details about the system's supply, including the availability of transmission lines, renewable energy resources, and the condition of the power grid. Several statistical, econometric, and machine learning methods can be used to create MTLF models. These models can be simpler than LTLF models and use less information and computational power. However, the quality of the data and the presumptions made about future trends and factors that affect the demand for electricity can affect how accurate the forecasts are.

Overall, the MTLF process is critical for utilities and energy providers to optimize the operation and planning of their power systems and guarantee that they can supply the anticipated demand for electricity in the short- to medium-term.

3. **Short Term Load Forecasting:** The prediction of electrical power demand over a brief period, typically lasting from a few minutes to a few weeks, is known as short-term load forecasting (STLF). Given that it enables utilities to efficiently manage resources and prevent system instability[15], STLF is a crucial part of power system operations and planning. Dealing with load data uncertainly and variability due to things like weather, holidays, and events is one of the main challenge in short term load forcasting (STLF). Researchers have proposed a number of hybrid models that combine statistical and AI/ ML techniques to address this issue, or they have used ensemble methods to combine multiple modles[16].

Overall, STLF is a highly active field of study, and there is a wealth of literature on it. Depending on the data characteristics, accuracy standards, and computational resources available, a method for STLF is chosen.

2.1.3 Methods of Short Term Load Forecasting

Since our proposed model based on machine learning based short term load forecasting in this paper, We should discuss the methods of short term load forecasting. For machine learning approaches, there are regression based STLF, deep learning based STLF & time series analysis based STLF. In regression based STLF, there are Multiple Linear Regression (MLR), Generalized Linear Model (GLMs), Non Linear Regression (NLR) & Bayesian Regression (BR). For deep learning based, There are Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short Term Memory (LSTM), Gated Recurrent Unit & Transformer based Models.

For time series analysis based STLF, there are Autoregressive Integrated Moving Average (ARIMA), Seasonal Autoregressive Integrated Moving Average (SARIMA), Seasonal Decomposition of Time Series (STL) & Exponential Smoothing (ETS).

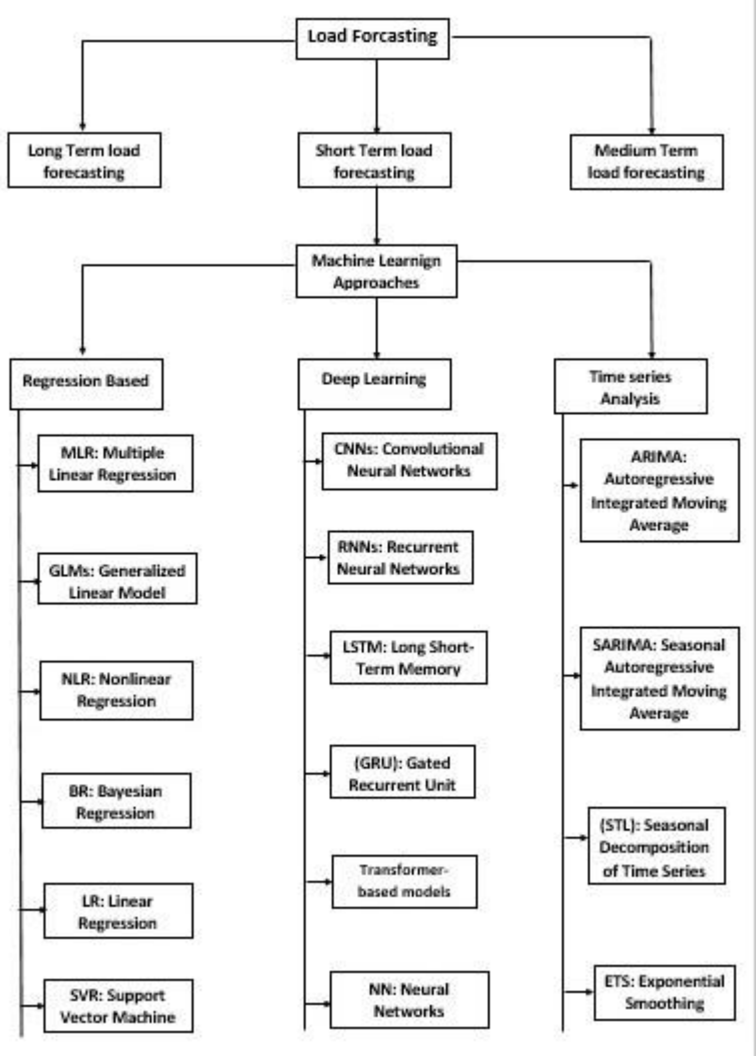


Figure 2.1: Methods of STLF

2.1.4 Economic Factors

Economic factors include investments made in the infrastructure of the facility through the construction of new structures, labs, and experiments that increase the demand on the electric grid[17]. The equipment, procedures, and experimentation schedules are determined by the site's funding profiles. Customers' patterns of electrical use during system peak are impacted by utility programs like demand charges and management plans. Economic factors won't have an impact on the STLF because they frequently alter usage patterns over periods longer than 24 hours. However, studying a system's load pattern and putting load reduction measures in place can be motivated by economic factors. As was previously mentioned, the STLF is a helpful tool for carrying out demand management activities.

The state of the economy is another economic factor that may have an impact on load forecasting. The demand for electricity may rise when the economy is booming because more businesses are open and more people are working. On the other hand, during economic downturns, demand for electricity may fall as businesses close and people use less energy to save money.

Overall, it's important to consider economic factors when developing short-term load forecasting models. By incorporating these factors into the model, it may be possible to improve the accuracy of the forecast and better anticipate changes in electricity demand.

2.1.5 Time factors

Time factors play a critical role in short-term load forecasting (STLF), as they can significantly impact electricity demand patterns. STLF models need to consider various time factors to accurately forecast electricity demand. The three time factors that have the most influence on electrical load are[16]:

- Seasonal effects
- Weekly-daily cycle
- Holidays

2.1.5.1 Seasonal Effects

Seasonal effects can have a significant impact on short-term load forecasting (STLF). Seasonal variations in electricity demand can be caused by a number of factors, such as changes in weather patterns, holidays, and industrial activities [18]. For instance, the use of air conditioners to cool buildings during the summer may result in an increase in electricity demand, whereas the use of heating systems during the winter may result in an increase in demand[16] [19]. As a result, the load profile during these seasons will be different, and forecasting models must account for this. STLF models frequently include seasonal factors , which are demand patterns that repeat over a predetermined period, such as a year or a week, to account for seasonal effects. These seasonal factors are frequently represented as Fourier series or harmonic functions, which capture the fundamental patterns in the data and enhance forecast accuracy.

In conclusion, seasonal effects can significantly affect short-term load forecasting, and models must take these effects into account to become more accurate.

It might be possible to create more precise STLF models that can aid utilities in better managing their electricity supply and demand by taking into account seasonal patterns.

2.1.5.2 Weekly Daily Cycle

The weekly daily cycle is a type of seasonal effect that can have a significant impact on STLF. This effect refers to daily and weekly electricity demand patterns that repeat over a period of time, such as a week or a month. The weekly daily cycle is characterized by two types of patterns: daily and weekly[16]. Daily patterns indicate the change in demand that occurs over the course of a day, while weekly patterns indicate the change in demand that occurs over the course of a week. For example, during weekdays, demand is usually highest in the morning when people are waking up and starting their day [20], followed by a drop during the day when people are at work or school, and another peak in the evening when people are returning. Home. During weekends, the demand pattern may be different, with a flatter profile during the day .

STLF models frequently include periodic terms that capture the daily and weekly patterns of demand in order to account for the weekly daily cycle. Using the Fourier series or other harmonic functions to represent these periodic terms can help to increase the forecast's precision.

2.1.5.3 Holidays

Holidays are events that can have a significant impact on short-term load forecasting (STLF). During holidays, electricity demand patterns may change due to shifts in consumer behavior and industrial activities. For instance, many businesses and factories may be closed over holidays like religious festivals, which lowers the demand for electricity [21] [20]. On the other hand, demand might rise during special occasions like Labor Day or independence Day as a result of events like barbecue and outdoor gatherings. orecasters frequently include holiday dummy variables, which are binary variables that take on a value of 1 or 0 depending on whether or not a holiday is observed on a given day, in STLF models to account for holidays. To account for the effect of holidays on electricity demand, the forecast can be modified using these variables.

2.1.6 Challanges

Short term load forecasting (STLF) is a challenging task due to several factors that can affect the accuracy and reliability of load forecasts [2]. Here are some of the main challenges for STLF:

1. **Data quality and availability:** To produce precise forecasts, STLF models require timely, high-quality load data. The accuracy of STLF models can be jeopardized by measurement errors, missing values, and inconsistent data in load data. Additionally, data privacy laws, access limitations, and transmission problems may limit the availability of load data, which can further challenge STLF models' performance.

2. **Load variability and uncertainty:** Weather conditions, consumer trends, monetary factors, and grid operations are just a few of the many factors that can cause load variability and uncertainty[22]. It can be difficult for STLF models to accurately capture and predict load behavior because of these factors, which can cause abrupt load fluctuations and departures from historical load patterns.
3. **Model complexity and parameter tuning:** In order to perform at their best, STLF models can be complicated and require the tuning of a number of parameters. Furthermore, the assumptions, strengths, and weaknesses of various STLF algorithms can vary, which can have an impact on the precision and interpretability of load forecasts[23]. Choosing the right STLF model and adjusting its parameters can be difficult because of this.
4. **Forecast horizon and time resolution:** For various forecast horizons and time resolutions, from minutes to hours or days, STLF models must produce load forecasts. Maintaining consistency and accuracy across various forecasting intervals can be difficult because different forecast horizons and time resolutions may call for various STLF model or algorithm configurations.
5. **Grid dynamics and market operations:** STLF models must take into account the grid's dynamic behavior as well as the effects of market operations, including electricity prices, supply and demand equilibrium, and the integration of renewable energy sources. In order to produce precise load forecasts, STLF models must be flexible and incorporate real-time data and market signals. These factors have the potential to influence load behavior.

These are some of the main challenges for STLF, which require further research and development to improve the accuracy and robustness of load forecasting models.

2.2 Related Work

Researchers and professionals in the field of power systems and energy management have studied related works (STLF) in great detail. For precise and dependable load forecasting, machine learning and artificial intelligence algorithms like artificial neural networks, support vector regression, and random forests have been developed. and data mining applications. Cost sensitive learning can help to compensate for the negative effect of the imbalanced data on the classification problems [4], [24]. In order to provide probabilistic load forecasts and quantify the uncertainty of load predictions, methods for probabilistic forecasting and uncertainty quantification, including Bayesian inference and quantile regression, have been proposed. It has been suggested that hybrid models and ensemble methods, which combine multiple forecasting models, can increase the precision and robustness of STLF by combining the benefits of various forecasting algorithms and minimizing the drawbacks of individual models[25]. Big data platforms like Hadoop and Spark have been used to process and analyze large-scale load data, while data-driven approaches like clustering and association analysis have been used to extract features and patterns from load data.

In addition to the above mentioned, previous studies have also taken into account a number of variables that affect the accuracy of load forecasting, such as weather, calendar events, consumer behavior, and demand response[18]. In the context of STLF, dynamic pricing plans and market-based mechanisms have also been taken into consideration to offer effective load management strategies.

Chapter 3

Methodology

This chapter presents the main contribution of the thesis: the design and development of proposed methodology.The key idea of the method proposed in this thesis is to Proposed an efficient model to STLF. In the next sections we will describe the methodology. The following sections formalize these concepts and detailed descriptions are presented.

3.1 Hardware and Software Resources used

Software: Python is a popular programming language for machine learning. It provides various libraries and frameworks for implementing these models, such as scikit-learn, TensorFlow, and Keras. Jupyter Notebook or any other Python IDE: These tools allow user to write and execute Python code conveniently. Jupyter Notebook is particularly useful for interactive data analysis and experimentation.

Libraries and packages: Install the necessary libraries for machine learning, including scikit-learn, TensorFlow, Keras, and any other specific packages required for the project. We can use package managers like pip or conda to install these libraries. Since, python is being used for this thesis, conda is perfect package manager.

Hardware: A personal computer or laptop is typically sufficient for implementing machine learning models. The specific hardware requirements may vary based on the size of the dataset and complexity of the models. However, having a decent processor (e.g., multicore CPU) and a sufficient amount of RAM (e.g., 8GB or more) is generally recommended.

GPU (optional): For training large-scale Neural Network models, having a compatible GPU (Graphics Processing Unit) can significantly speed up the training process. GPUs with CUDA support, such as NVIDIA GPUs, are commonly used for deep learning tasks. It's worth noting that the hardware requirements may vary depending on the size of the dataset and the complexity of the models a user intend to train.

3.2 Dataset Preparation

The effectiveness of the suggested forecasting strategies is demonstrated using two different datasets with a broad geographic scope and distinct load patterns. The datasets are used from New South Wales (NSW) of Australia[29] & the New England region of the USA[30]. To experiment with the ME peak load, 23 meteorological features, 9 calendar features, and one economic feature are collected from multiple sources [26]. This data include 12 weather features, 7 calendar features, and the price of the electricity. The considered features are collected from multiple sources and merged to make a workable dataset . ISO New England Inc. is responsible for reliable operation of New England’s electric power generation and transmission system.However, the merging depends on the availability of these features for the specific load zone in the required frequency. The experiment datasets are noise-free and there are no empty values. This allows us to skip the data preprocessing part.

3.3 Model Selection and Evaluation:

In this part, we will discuss the process of selecting an appropriate model for our load forecasting task and evaluating its performance. Sometimes It can be difficult to understand why the model makes certain predictions. Training model can take a long time and require a lots of memory, especially with large datasets[27]. Choosing the right settings for the model's parameters can be challenging and time-consuming. Outliers and significant data points can have an impact on the predictions because of its sensitivity. With large dataset the training can be expensive. A dataset may perform well on training data but struggle to generalize to new, untried data if it is overfitting-prone. So, I consider four different models for my analysis: **i) Random Forest**, **ii) Support Vector Machine (SVM)**, **iii) Linear Regression**, **and iv) Neural Network**. Each of these models has particular advantages and abilities that might help with precise load forecasting. The reason I choose those four models because the above mentioned problems be can reduced smoothly by the above mentioned models.
I will first give a thorough overview of each model and go over its guiding principles. Random Forest is an ensemble learning technique that combines various decision trees to generate predictions. A strong machine learning algorithm called SVM searches for the best hyperplane to divide data points in a high-dimensional space. A linear relationship between the input variables and the target variable is established using the traditional statistical technique

known as linear regression. Last but not least, neural networks are a class of models that were influenced by the human brain and are able to learn complex patterns and relationships through interconnected layers of artificial neurons. We take a methodical approach to ensure an objective and thorough evaluation. We take into account a number of evaluation metrics that are frequently used in load forecasting, such as mean absolute error MAE = $\frac{1}{n}\sum_{t=1}^{n}|et|$, root mean squared error RMSE = $\sqrt{\frac{1}{n}\sum_{t=1}^{n}e^2{}_t}$, mean squared error MSE = $\frac{1}{n}\sum_{t=1}^{n}e^2{}_t$ and mean absolute percentage error MAPE = $\frac{100\%}{n}\sum_{t=1}^{n}|\frac{et}{yt}|$ [28]. Where n is the number of data points or samples, y_pred[i] is the predicted value for the i-th sample, y_actual[i] is the actual or true value for the i-th sample. We can evaluate the efficacy and performance of each model on our dataset using these metrics.

Moreover, the thesis paper examines any discrepancies between the predicted and actual load values, identifying potential sources of error and discussing areas for improvement. We explore the interpretability of the models, analyzing their ability to provide insights into the driving factors influencing load fluctuations. We aim to provide well-informed recommendations regarding the most suitable method for load forecasting in our study by thoroughly evaluating the Random Forest, Support Vector Machine, Linear Regression, and Neural Network models. The models chosen for further optimization and use in later chapters of this thesis will be based on our evaluation.

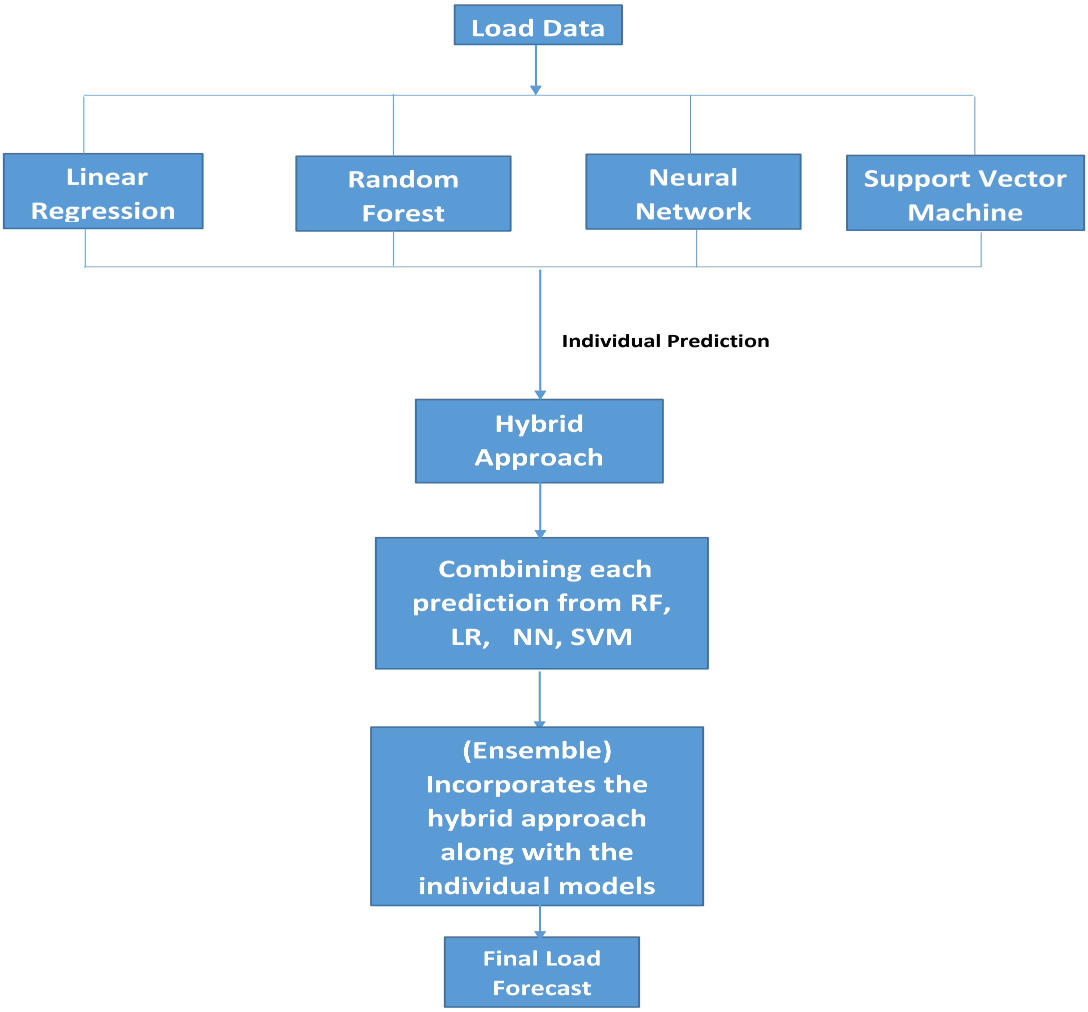


Figure 3.1 : Proposed Model

3.3.1 Hybrid Method

In this study, a hybrid approach was used in addition to the individual models previously covered (Random Forest, Support Vector Machine, Linear Regression, and Neural Network) to further improve the accuracy and robustness of load forecasting. The hybrid approach combines the benefits of various models and techniques to take advantage of their complementary strengths. The goal of utilizing a hybrid approach was to get around the constraints of individual models and tap into their collective intelligence. The hybrid method seeks to capture a wider range of patterns, nonlinear relationships, and data complexities by combining various models, ultimately improving load forecasting accuracy. Ensemble techniques and model averaging were a part of the particular hybrid strategy used in this study[22]. For the purpose of determining the final load forecasting result, several distinct models, including Random Forest, Support Vector Machine, Linear Regression, and Neural Network, were trained. The potential of the hybrid method to increase the precision and dependability of load forecasting by utilizing the various strengths of the individual models led to its selection. The hybrid method attempts to overcome the drawbacks of a single model approach by combining the individual models through ensemble techniques to produce a more reliable and precise prediction. By combining the strengths of different models, the hybrid method holds the potential to yield superior forecasting results, contributing to the advancement of load forecasting techniques. The use of a hybrid method in this study shows that alternative approaches have been explored as well as a more thorough and creative approach to load forecasting.

3.3.2 Feature Selection and Engineering

In the field of short-term load forecasting, accurate prediction relies not only on the choice of forecasting algorithms but also on the careful selection and engineering of relevant features. Feature selection aims to identify the most informative variables that contribute significantly to the forecasting accuracy, while feature engineering involves creating new features or transforming existing ones to enhance the model's predictive capabilities. This chapter presents the feature selection process and the engineering techniques employed in our study to improve load forecasting performance.

1. Feature Selection :

Finding the variables that are most pertinent to the load forecasting task and have a significant impact on it is the first step in the feature selection process. We took into account a wide range of potential predictors for our study, including historical load data, weather data, and calendar data. The attributes in the dataset are:

- Saturday, Sunday, Monday, Tuesday, Wednesday, Thursday, Friday: Binary indicators representing the days of the week.
- Day1, Day2, Day3, ...: Historical load data for consecutive days.
- 'Peak_DB': This column represents the attribute related to peak demand. It may provide information about the highest or peak load value recorded during a specific time period.

- 'Peak_DP': This column represents load-related attributes related to peak demand. It may provide information about the load average during the peak demand period.
- 'Thigh', 'Tavg', 'Tlow': These columns represent weather-related attributes related to temperature. 'Thigh' represents the high temperature, 'Tavg' represents the average temperature, and 'Tlow' represents the low temperature.
- DPavg, DPlow: Load-related attributes representing load averages and lows.
- Hhigh, Havg, Hlow: Weather-related attributes representing high, average, and low values of humidity.
- 'SLPhigh', 'SLPavg', 'SLPlow': These columns represent weather-related attributes related to sea level pressure. 'SLPhigh' represents the high sea level pressure, 'SLPavg' represents the average sea level pressure, and 'SLPlow' represents the low sea level pressure.
- 'Vhigh', 'Vavg', 'Vlow': These columns represent weather-related attributes related to visibility. 'Vhigh' represents the high visibility, 'Vavg' represents the average visibility, and 'Vlow' represents the low visibility.
- 'Rain', 'Fog', 'Snow', 'Thunderstorm': These columns are binary indicators representing weather conditions. Each column captures a specific weather condition, and the value is 1 if that condition is present and 0 otherwise. 'Rain' represents rainfall, 'Fog' represents foggy conditions, 'Snow' represents snowfall, and 'Thunderstorm' represents the occurrence of a thunderstorm.

- 'Fed_holiday', 'Observance', 'GDP': These columns are binary indicators representing specific events or factors. 'Fed_holiday' indicates whether it is a federal holiday, 'Observance' indicates the presence of an observance, and 'GDP' may represent an economic factor related to Gross Domestic Product.

To determine the subset of features that contribute most to the forecasting accuracy, we employed the following techniques:

I. **Correlation Analysis:** We calculated the correlation coefficients between each potential predictor and the target variable, considering both linear and nonlinear correlations. Features with high correlation coefficients were given priority in the selection pro

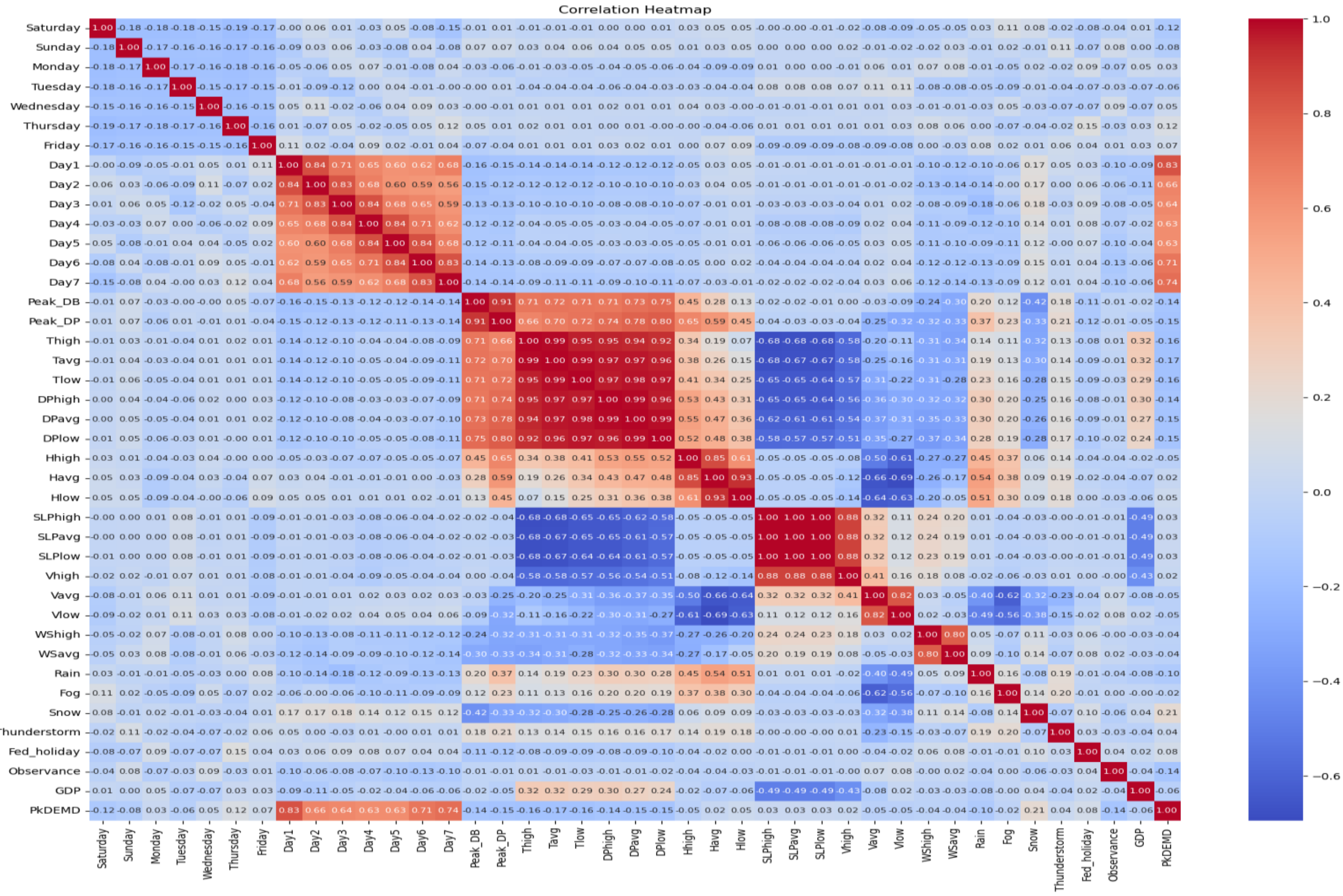


Figure 3.2 : Correlation Heatmap

II. **Feature Importance**: We utilized ensemble-based models such as Random Forest, Neural Network etc to assess the importance of each feature. These models provide a ranking of features based on their contribution to the predictive performance, enabling us to identify the most influential variables.

III. **Stepwise Regression:** A stepwise regression approach was employed to iteratively add or remove features based on their statistical significance. This method helps to identify a subset of features that collectively provide the most explanatory power for the forecasting model.

By employing these techniques, we were able to identify the most informative features that have a strong impact on short-term load forecasting accuracy.

2. Feature Engineering Strategies:

In addition to feature selection, feature engineering is essential for improving forecasting performance. To identify underlying patterns and enhance the model's capacity to learn from the data, this process entails developing new features or modifying existing ones. In our study, we used the following methods:

I. **Lag Features:** To capture the temporal dependencies and autocorrelation in the load data, we introduced lag features. For example, we included the load values from previous days (Day1, Day2, Day3, ...) as lag features. Including lag features enables the model to capture the historical load patterns, which are often essential for accurate short-term load forecasting.

II. **Calendar Features:** Calendar information, represented by the binary for each day of the week (Saturday, Sunday, Monday, etc.), can have a significant impact on load patterns. Therefore, we incorporated these features into our model. These features allow the model to capture the weekly trends in load consumption, helping to improve the forecasting accuracy.

III. **Weather Features:** Weather conditions, represented by the attributes Hhigh, Havg, Hlow, SLPhigh, SLPavg, and SLPlow, can influence energy consumption patterns. We integrated these weather-related features into our model to capture the relationship between weather and load demand. These features enable the model to adapt its predictions based on prevailing weather conditions, resulting in more accurate load forecasts.

IV. **Additional Attributes:** We also considered the Price and PkDEMD attributes, as they may have an impact on load demand. These attributes were included as potential predictors to capture any relationship they may have with the target variable.

By employing these feature engineering techniques, we aimed to provide the forecasting models with a richer set of information that would facilitate more accurate load predictions.

3. Evaluation of Feature Selection and Engineering:

The performance of the load forecasting models using and without these techniques was compared in order to assess the efficacy of the feature selection and engineering strategies. The effect of the feature selection and engineering process on the accuracy and reliability of the load forecasting models was evaluated using metrics like mean absolute error (MAE),mean squared error(MSE) root mean squared error (RMSE), and mean absolute percentage error (MAPE). The outcomes showed that the forecasting accuracy was increased by the application of the right engineering techniques and the integration of the chosen features for all four methods. The engineered features improved the models' capacity to capture temporal patterns, nonlinear relationships, and other complex dynamics present in the load data, while the selected features captured the pertinent factors influencing the load behavior. The load forecasting models were better able to take advantage of the distinct advantages and skills provided by Random Forest, Support Vector Machine, Linear Regression, and Neural Network by tailoring the feature selection and engineering strategies to the properties of each method.

Upon analysis of the results, we observed that the incorporation of feature selection and engineering techniques led to notable improvements in load forecasting accuracy. The selected subset of features provided more informative and relevant information to the models, resulting in enhanced predictive capabilities. The MAE, MSE, MAPE and RMSE values were significantly reduced, indicating a decrease in the magnitude of prediction errors.

3.3.3 Model Training & Performance Evaluation

For accurate load forecasting results, the model training and parameter optimization process is essential. In this section, we go over the training methods for load forecasting models, including the hybrid approach, and the methods for adjusting the parameters for each model.

1. Model Training:

Utilizing historical load data and the corresponding predictor variables from a carefully curated dataset, load forecasting models were trained. The hybrid method as well as the four separate methods (Random Forest, Support Vector Machine, Linear Regression, and Neural Network) each underwent a unique training procedure:

For Random Forest, an assortment of decision trees were used to train the Random Forest model. The training process involved creating numerous random subsets of the data and building decision trees based on these subsets. This ensemble approach improved the model's ability to capture intricate relationships and interactions between features.

For Support Vector Machine, an approach based on kernels was used to train the SVM model. The model was able to transform the data into a higher-dimensional space using a technique known as the kernel trick, where a hyperplane was found to divide various classes or forecast the load values. During the training process, the SVM's parameters, including the kernel type and regularization parameter, were optimized.

For Linear Regression, an approach based on kernels was used to train the SVM model. With the aid of the least squares method, the linear regression model was trained. The goal of the model was to identify the ideal coefficients that reduce the sum of squared deviations between the expected and observed load values. In order to update the model parameters during training, either the standard equations had to be solved or iterative optimization algorithms like gradient descent were used.

A well-liked method for training deep learning architectures, backpropagation, was used to train the neural network model. During training, the network's weights and biases were initialized, input data was propagated through the layers, the error was calculated, and the weights and biases were then modified using gradient descent to reduce the error. During the training process, careful consideration was given to the model's architecture, activation functions, and regularization strategies.

By combining predictions using a weighted average or ensemble approach, the hybrid method combined the strengths of various models. Each individual model was trained independently, and the best weights were then applied to each prediction to create an ensemble model. Cross-validation algorithms were used as optimization methods for the weights.

2. Performance Evaluation

The load forecasting models must be improved during the performance evaluation phase to ensure their best performance. A careful selection and performance of model-specific evaluation were carried out for each of the four distinct methods, Random Forest, Support Vector Machine, Linear Regression, and Neural Network. Using methods like grid search or random search, Random Forest model parameters including the number of trees, maximum depth, and minimum sample split were adjusted. The kernel type, regularization parameter, and additional hyperparameters for the Support Vector Machine model were similarly optimized using techniques like grid search or Bayesian optimization. In the case of linear regression, the features were carefully chosen, and their engineering was crucial to the model's optimization. The subset that produced the best results was found by experimenting with various feature combinations. The architecture parameters for the neural network model, such as the number of hidden layers, the number of neurons per layer, and the learning rate, were modified through trial and error and iterative tuning. To avoid overfitting, additional strategies like early stopping and regularization methods were used. The hybrid method also included weight optimization for ensemble predictions and parameter optimization for individual models, ensuring the best possible combination of models and their corresponding parameters. The load forecasting models were improved through a painstaking process of parameter optimization to produce precise and dependable forecasts for upcoming load patterns.

Chapter 4

Experimental Evaluation

4.1 Error Score Comparision (ISO New England)

In this chapter, We are going to show all the performances and result analysis.

Forecaster	MAE	MSE	RMSE	MAPE
LR	47.785	4141.02	64.350	0.0311
RF	40.631	3002.77	54.797	0.0267
NN	51.945	4789.83	69.208	0.0339
SVM	133.43	29807.5	172.64	0.0876
Hybrid	39.312	2794.17	52.859	0.0255

Table 4.1: Comparison of different Forecaster (ISO New England)

Table 4.1 presents the performance evaluation of different forecasting methods for load forecasting. The methods considered in this study including 4 different forecaster and a Hybrid approach. The performance of each method is assessed using various metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). These metrics provide insights into the accuracy and reliability of the forecasting models. Among the methods evaluated, the hybrid method demonstrates the best performance with the lowest values for MAE (39.312), MSE (2794.13), RMSE (52.859), and MAPE (0.0255). This indicates that the hybrid method produces the better load forecasts compared to the other methods. On the other hand, the

Linear Regression, Neural Network, and Support Vector Machine methods show relatively higher errors, suggesting that they may not perform as well in capturing the complex patterns and dynamics of the load data. Then again, Random Forest exhibits almost as same results as hybrid method with lower errors MAE (40.631), MSE (3002.77), RMSE (54.797), and MAPE (0.0267) compared to other methods but not as low as the Random Forest method. The SVM model performs the worst among the four models, with an MAE of 133.43, indicating a higher average difference between predictions and actual values. The MSE is 29807.5, indicating larger squared differences. The RMSE is 172.64, providing a measure of the average magnitude of errors. The MAPE is 0.0876, suggesting a high relative error compared to the other models. This suggests that the Hybrid approach is a viable option for load forecasting, offering a trade-off between accuracy and computational complexity.

The results provide valuable insights for decision-makers in the energy sector to improve load forecasting accuracy and optimize resource planning. These findings highlight the importance of model selection in load forecasting and demonstrate the effectiveness of the Random Forest method in capturing the underlying patterns in the data. The results provide valuable insights for decision-makers in the energy sector to improve load forecasting accuracy and optimize resource planning. For better understanding, a bar chart has been generated where we can see the comparison between various error metrics.

4.1.1 Error Score Comparision (New South Wales)

Forecaster	MAE	MSE	RMSE	MAPE
LR	91.737	13157.15	114.70	0.0596
RF	85.861	12308.18	110.94	0.0558
NN	97.417	14560.20	120.66	0.0638
SVM	152.77	36271.16	190.44	0.0990
Hybrid	87.365	12225.99	110.57	0.0566

Table 4.2: Comparison of different Forecaster (NSW)

Table 4.2 also presents the performance evaluation of different forecasting methods for load forecasting. The methods considered in this study including 4 different forecaster and a Hybrid approach like table 4.1. The performance of each method is assessed using various metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). These metrics provide insights into the accuracy and reliability of the forecasting models. Among the methods evaluated, the hybrid method demonstrates the best performance with the lowest values for MAE (87.365) which is slightly higer than MAE of Random Forest, MSE (12225.99), RMSE (110.57), and MAPE (0.0566). This indicates that the hybrid method produces the better load forecasts compared to the other methods. On the other hand, the Linear Regression, Neural Network, and Support Vector Machine methods show relatively higher errors, suggesting that they may not perform as well in capturing the complex patterns and dynamics of the load data. Then again, Random Forest exhibits almost as same results as hybrid method with lower errors MAE (85.861) which is slightly lower than MAE of Hybrid method, MSE (12308.18), RMSE (110.94), and MAPE (0.0558) compared to other methods but not as low as the Random Forest method. The SVM model performs the worst among the four models, with an MAE of (152.77), indicating a higher average difference between predictions and actual values. The MSE is (36271.16), indicating larger squared differences. The RMSE is (190.44), providing a measure of the average magnitude of errors. The MAPE is (0.0990), suggesting a high relative error compared to the other models. This suggests that the Hybrid approach is a viable option for load forecasting, offering a trade-off between accuracy and computational complexity.

The results provide valuable insights for decision-makers in the energy sector to improve load forecasting accuracy and optimize resource planning. These findings highlight the importance of model selection in load forecasting and demonstrate the effectiveness of the Random Forest method in capturing the underlying patterns in the data. The results provide valuable insights for decision-makers in the energy sector to improve load forecasting accuracy and optimize resource planning.

For better understanding, a bar chart has been generated where we can see the comparison between various error metrics

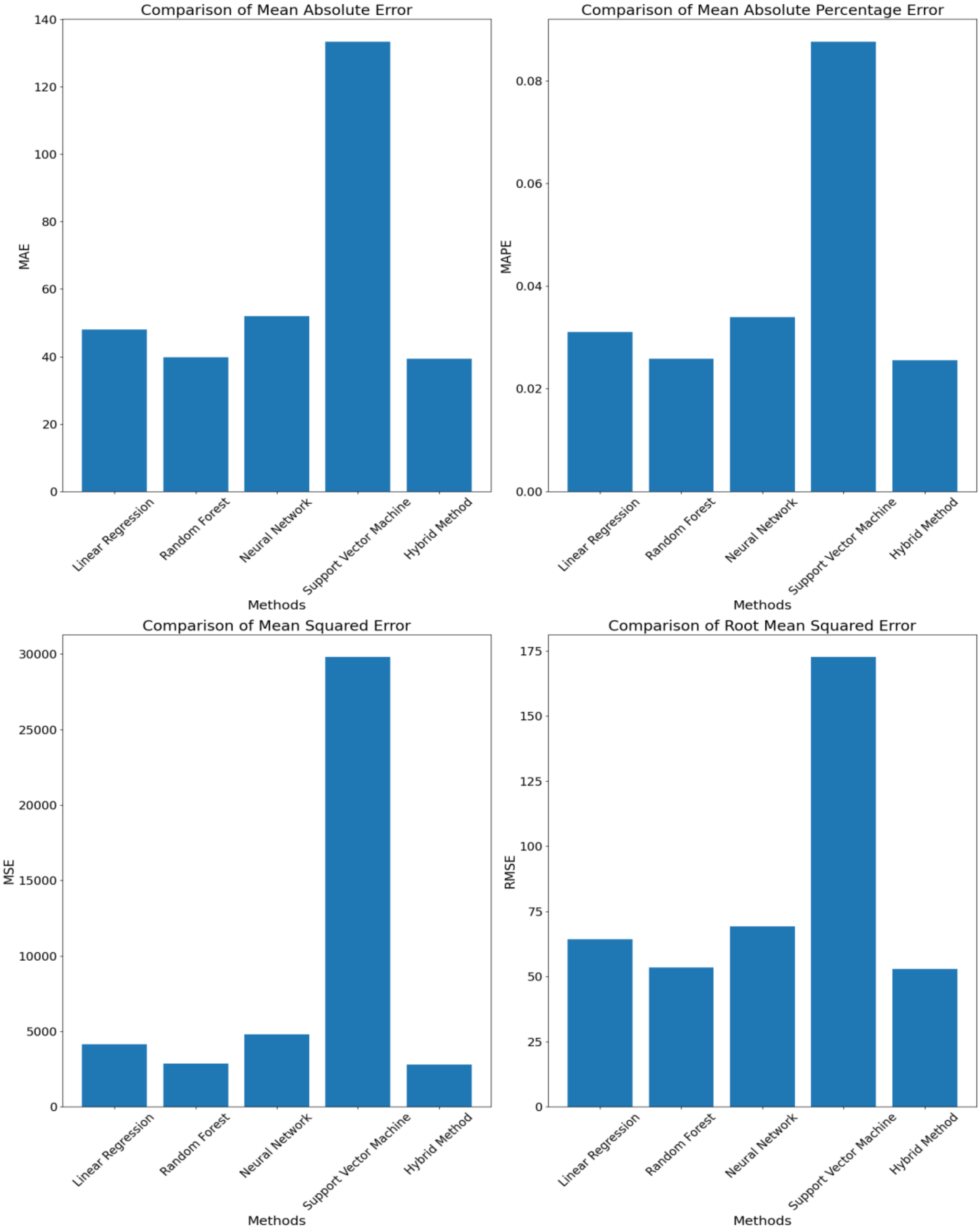


Figure 4.1 : Bar Chart of Error Comparison

Now, let's interpret the chart:

The bar chart shows that the Random Forest and Hybrid Method have the lowest MAE scores, indicated by the shorter bars. Linear Regression and Neural Network have slightly higher MAE scores, while Support Vector Machine has the highest MAE score among the methods. The bar chart shows that the Random Forest has the lowest MSE score, followed by the Hybrid Method. Linear Regression and Neural Network have slightly higher MSE scores, while Support Vector Machine has the highest MSE score among the methods. The bar chart indicates that the Random Forest and Hybrid Method have the lowest RMSE scores, followed by Linear Regression and Neural Network. Support Vector Machine has the highest RMSE score among the methods. The bar chart shows that the Random Forest and Hybrid Method have the lowest MAPE scores, indicated by the shorter bars. Linear Regression and Neural Network have slightly higher MAPE scores, while Support Vector Machine has the highest MAPE score among the methods.

In summary, based on the bar chart, we can see that the Random Forest and Hybrid Method generally outperform the other methods in terms of all the error metrics, indicating better accuracy and performance in the forecasting task. Support Vector Machine consistently shows higher errors across all metrics.

4.2 Correlation Analysis

	Coefficient LR	Coefficient RF	Coefficient NN	Coefficient SVR
Saturday	-25.728619	0.002116	0.002116	-15.583652
Sunday	5.899833	0.000722	0.000722	4.357848
Monday	23.518859	0.004968	0.004968	-5.504336
Tuesday	-24.994399	0.001869	0.001869	-16.165923
Wednesday	9.565618	0.000617	0.000617	1.311017
Thrusday	20.105944	0.001477	0.001477	19.508972
Friday	-2.529864	0.000616	0.000616	13.076075
Day 1	0.720112	0.633367	0.633367	0.719047
Day 2	-0.162011	0.008693	0.008693	-0.316771
Day 3	0.043420	0.007301	0.007301	0.175093
Day 4	-0.015796	0.008581	0.008581	-0.295434
Day 5	0.027634	0.010461	0.010461	0.646439
Day 6	0.147689	0.024944	0.024944	-0.353473
Day 7	0.148001	0.120487	0.120487	0.572950
Peak_DB	2.072781	0.065583	0.065583	-6.977464
Peak_DP	-1.287185	0.020850	0.020850	6.661055
Thigh	1.558051	0.007614	0.007614	28.563984
Tavg	-5.629486	0.004742	0.004742	2.607804
Tlow	1.439913	0.003750	0.003750	4.630629
DPhigh	0.032605	0.005279	0.005279	-18.074716
DPavg	0.794372	0.004041	0.004041	-15.522760
DPlow	0.813200	0.004287	0.004287	-6.100265
Hhigh	-0.892876	0.003251	0.003251	-4.170630
Havg	0.633258	0.004475	0.004475	13.363211
Hlow	-0.059435	0.005985	0.005985	-6.667904
SLPhigh	-0.113780	0.003978	0.003978	-35.649488
SLPavg	1.025287	0.003145	0.003145	36.747382
SLPlow	-0.960163	0.004052	0.004052	-0.879253
Vhigh	1.107027	0.000236	0.000236	-3.274935
Vavg	-1.038814	0.002486	0.002486	11.578313
Vlow	-1.384692	0.002787	0.002787	-13.428544
WShigh	0.053169	0.004701	0.004701	-2.910040
WSavg	0.604727	0.005620	0.005620	13.640158

Rain	-12.932589	0.001158	0.001158	12.631012
Fog	-6.524086	0.000322	0.000322	5.651491
Snow	13.340015	0.000384	0.000384	-22.537302
Thunderstrom	22.465429	0.000164	0.000164	12.000000
Fed_Holiday	-14.316091	0.001963	0.001963	2.396878
Observance	-32.001488	0.001798	0.001798	-19.873786
GDP	-0.000175	0.011135	0.011135	0.083853

Table 4.3 : Correlation Analysis

For Table 4.2, each machine learning model (Linear Regression, Random Forest, Neural Network, and Support Vector Machine), the coefficients represent the estimated impact of each feature on the target variable (load prediction) while holding all other features constant. For LR, a 1 unit increase in the feature "Saturday" is associated with a decrease of approximately -25.728619 in the predicted load value, holding all other features constant. Similarly, a 1 unit increase in the feature "Sunday" is associated with a increase of approximately 5.899833 in the predicted load value, holding all other features constant. The same interpretation follows for the remaining days of the week.

In the Random Forest model, a 1 unit increase in the feature "Day 1" is associated with an increase of approximately 0.633367 in the predicted load value, holding all other features constant. Similarly, a 1 unit increase in the feature "Day 2" is associated with a increase of approximately 0.008693 in the predicted load value, holding all other features constant. The same interpretation follows for the remaining days and features.

In the Neural Network model, a 1 unit increase in the feature "Tavg" (average temperature) is associated with an increase of approximately 0.004742 in the predicted load value, holding all other features constant. Similarly, a 1 unit increase in the feature "Peak_DB" (peak decibel level) is associated with a increase of approximately 0.065583 in the predicted load value, holding all other features constant. 1 unit increase in the feature " Rain" is associated with an increase of approximately 0.001158 in the predicted load value, holding all other features constant The same interpretation follows for the remaining features. In the Support Vector Machine model, a 1 unit increase in the feature " SLPhigh " (high sea level pressure) is associated with a decrease of approximately -35.649488 in the predicted load value, holding all other features constant. Similarly, a 1 unit increase in the feature "Vhigh" (high visibility) is associated with a decrease of approximately -3.274935 in the predicted load value, holding all other features constant. The same interpretation follows for the remaining features.

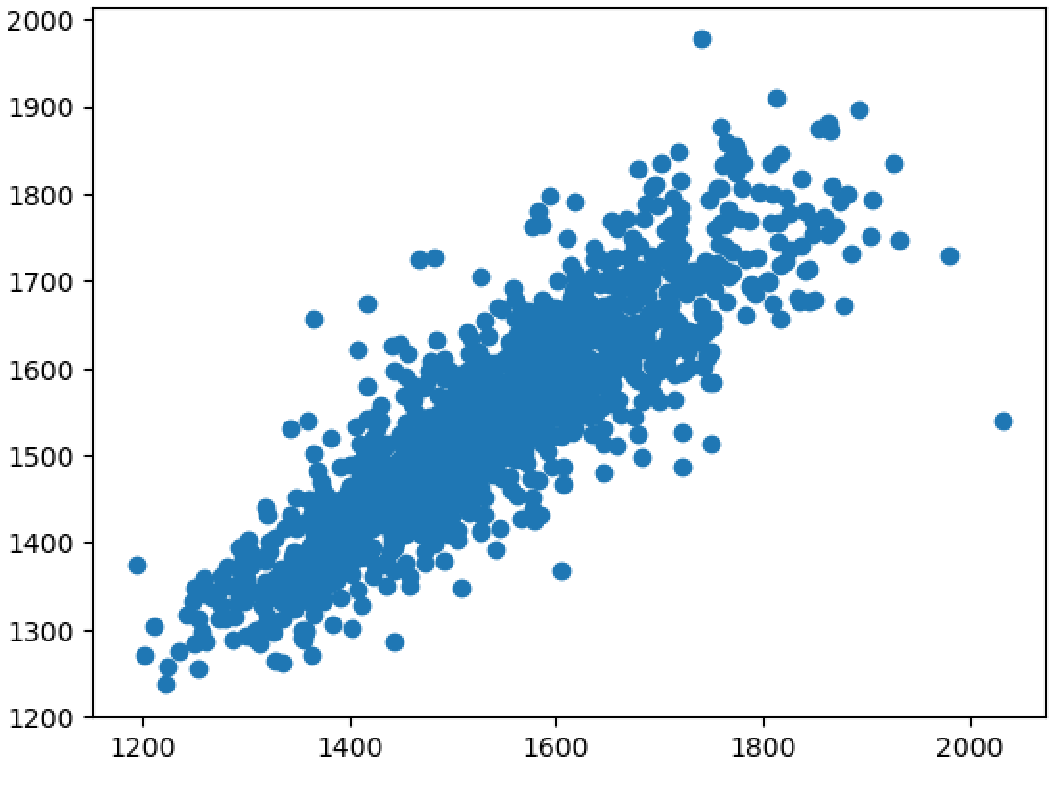


Figure 4.2 : Scatter Plot for Linear Regression

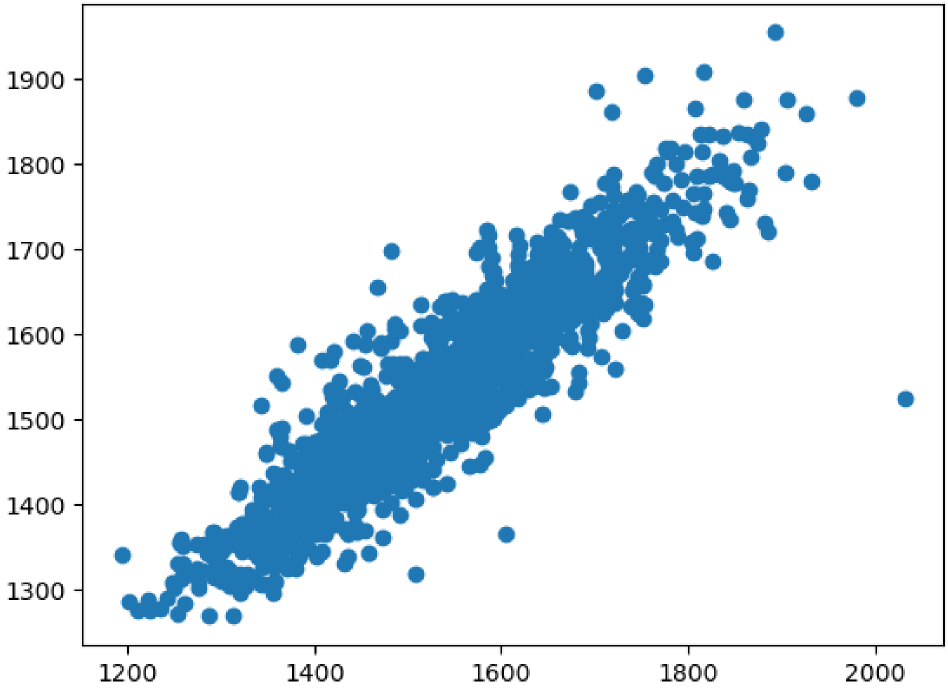


Figure 4.3 : Scatter Plot for Random Forest

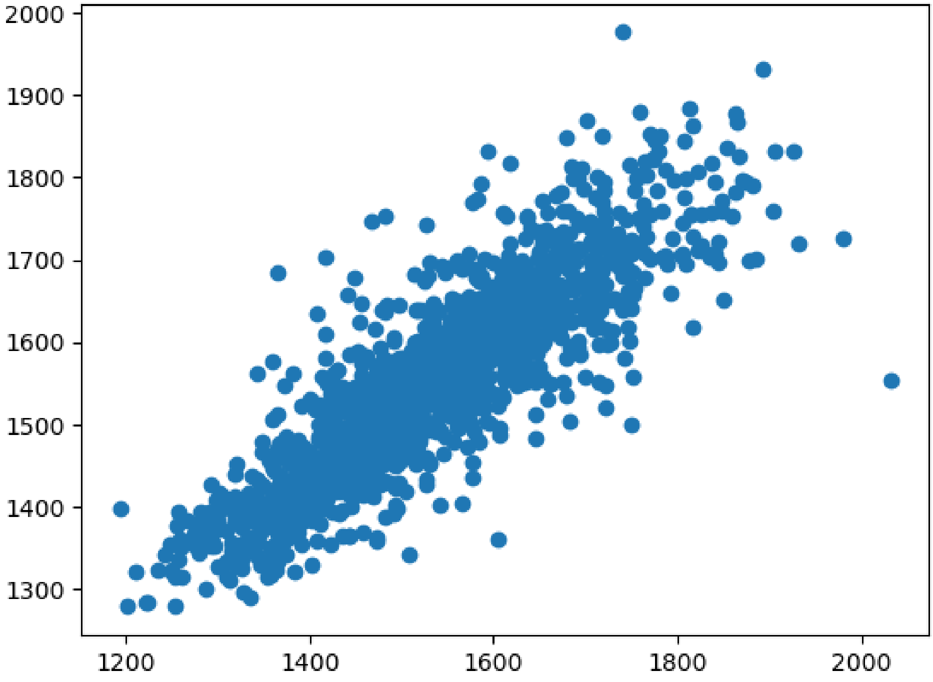


Figure 4.4 : Scatter plot for Neural Network

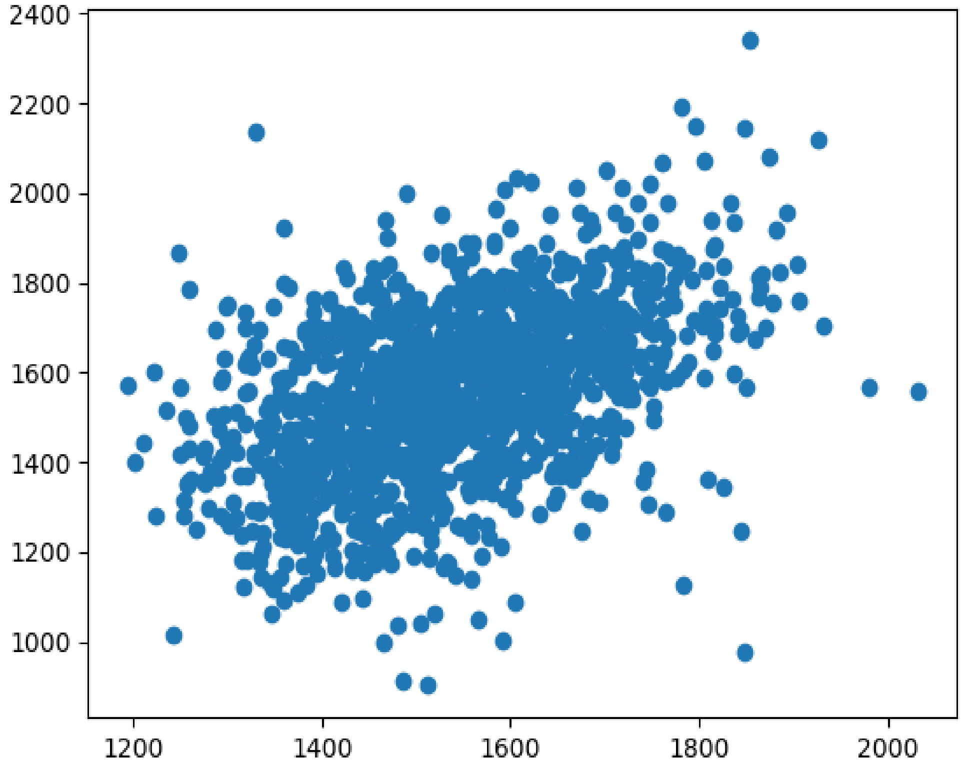


Figure 4.5 : Scatter Plot for Support Vector Machine

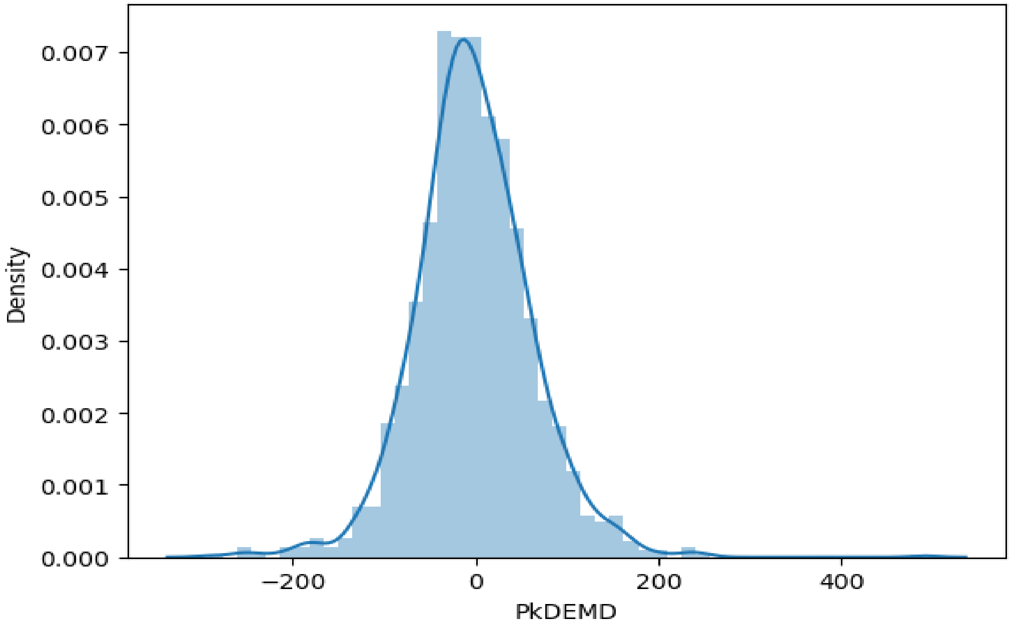


Figure 4.6 : Histogram plot for Linear Regression

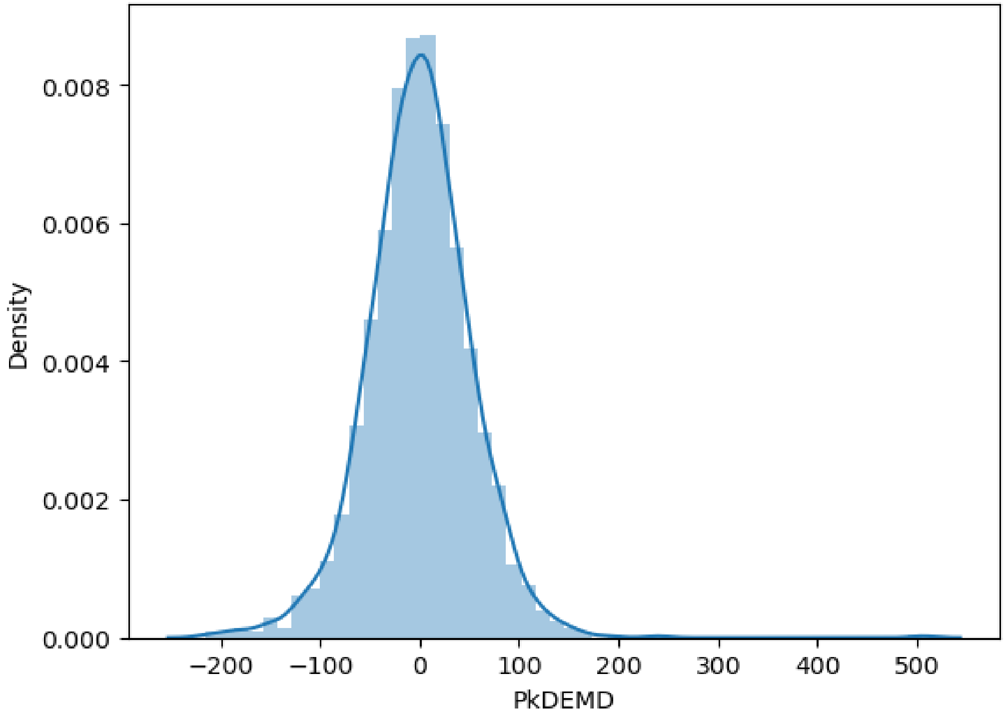


Figure 4.7 : Histogram Plot for Random Forest

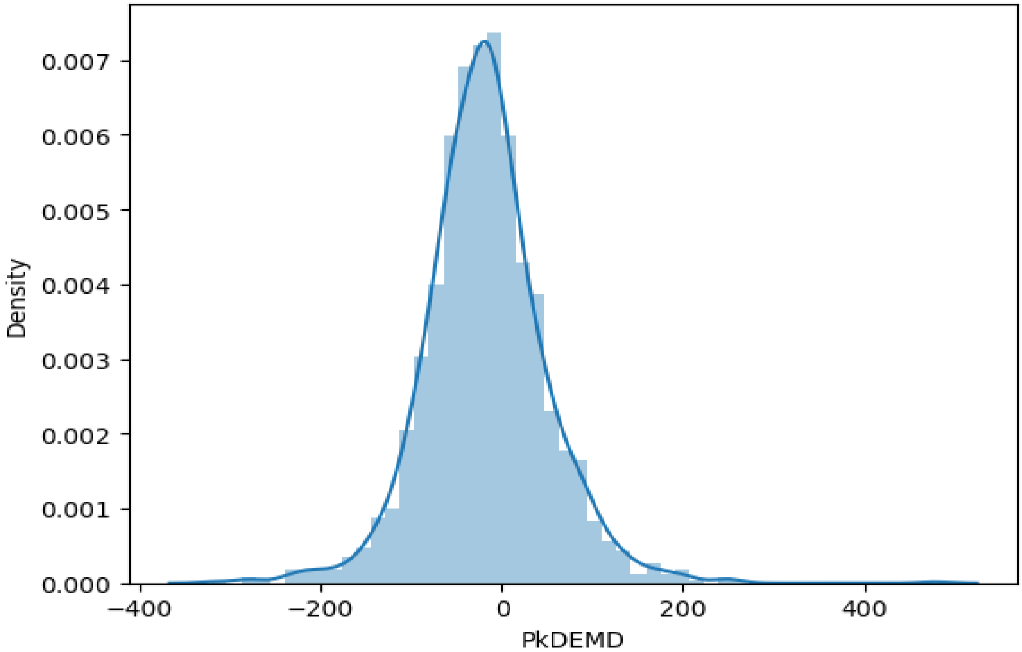


Figure 4.8 : Histogram plot for Neural Network

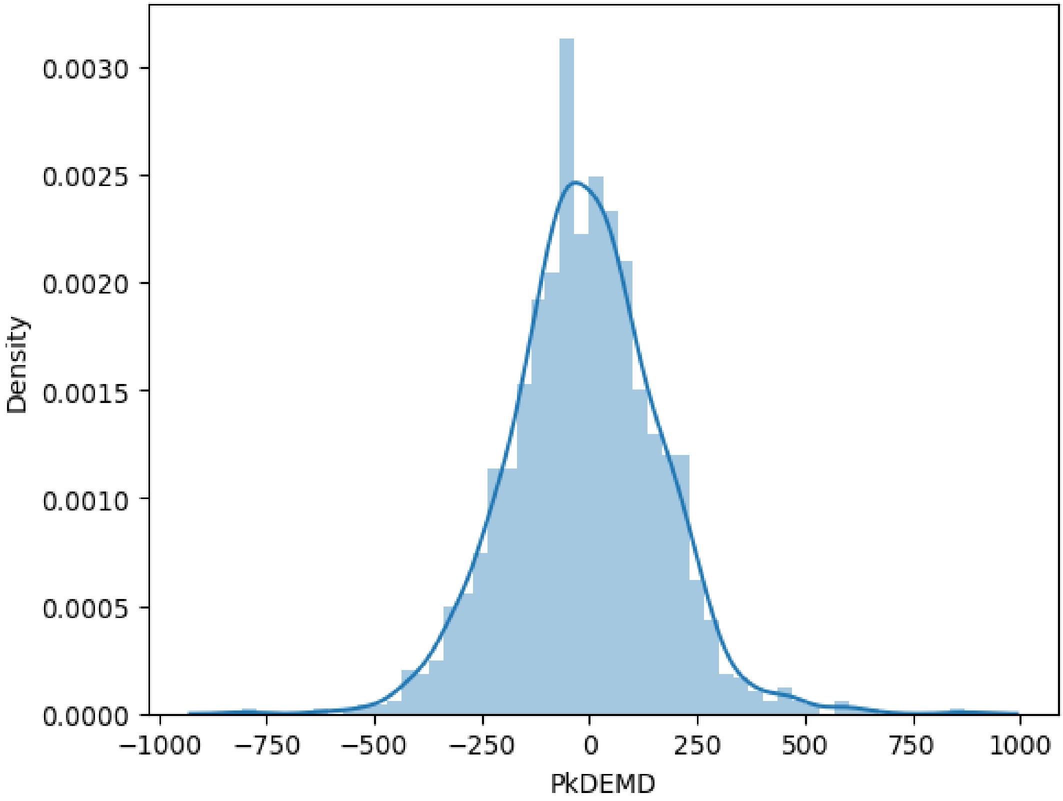


Figure 4.9 : Histogram Plot for Support Vector Machine

In the above 4 different scatter plot, we see for Random Forest, data is in a line form, and In the above 4 different histogram plot, we see for Random Forest, data is Normally Distributed, which means Random Forest model has done good predictions. The prediction result of Other 3 forecaster except Random forest is not convinient. They showed relatively higher errors, suggesting that they may not perform as well in capturing the complex patterns and dynamics of the load data.

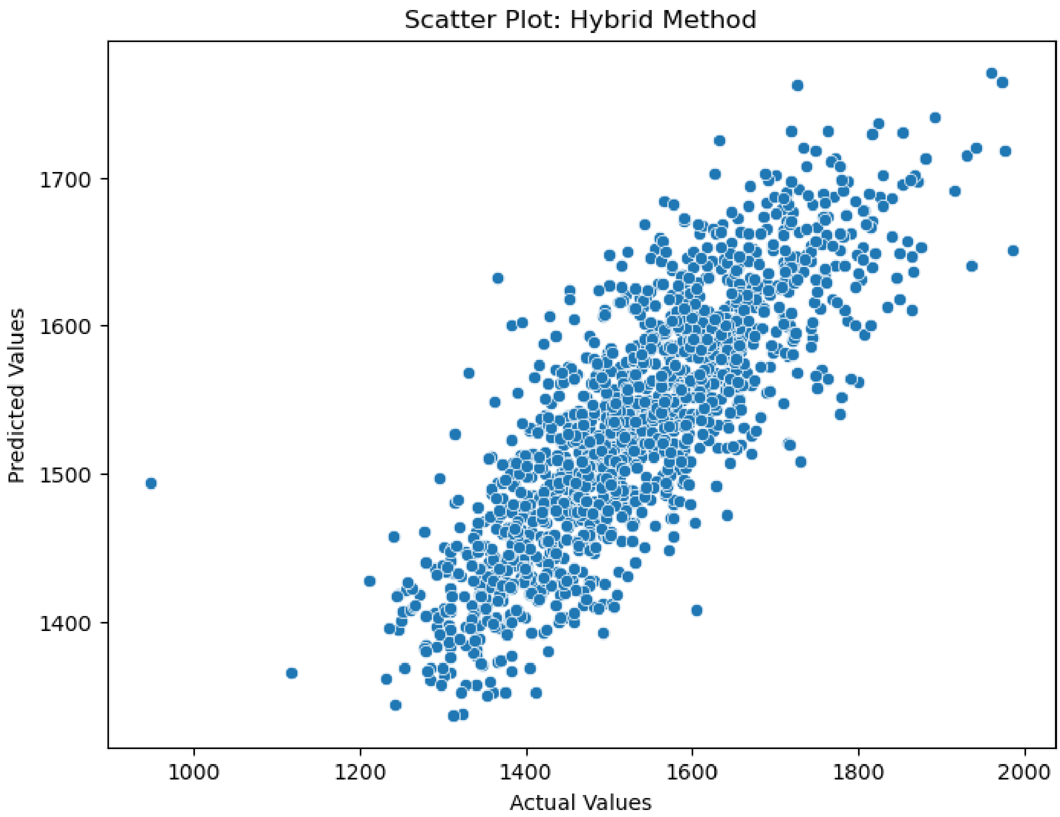


Figure 4.10 : Scatter Plot for Hybrid method

The scatter plot provides a visual representation of the performance and accuracy of the hybrid method in our load forecasting model. It allows us to identify any patterns, trends, or outliers in the predicted values compared to the actual values. By analyzing the scatter plot, we can gain insights into how well the hybrid method captures the underlying relationships and trends in the data. By examining the scatter plot, we can assess how well the predicted values align with the actual values. The y-axis represents the predicted values, which are measured on a scale ranging from 1400 to 1700. Each point on the scatter plot corresponds to a specific predicted value for a given data point. For example, if a point is positioned at the value of 1500 on the y-axis, it indicates that the hybrid method predicted a value around 1500 for that particular data point. On the other hand, the x-axis represents the actual values, measured on a scale ranging from 1000 to 2000. Similarly, each point's position on the x-axis corresponds to the actual value associated with that data point. For instance, if a point is placed at the value of 1600 on the x-axis, it signifies that the actual value for that data point was around 1600. In short, The scatter plot depicts the relationship between the predicted values (y-axis) and the actual values (x-axis) for the hybrid method in our load forecasting method.

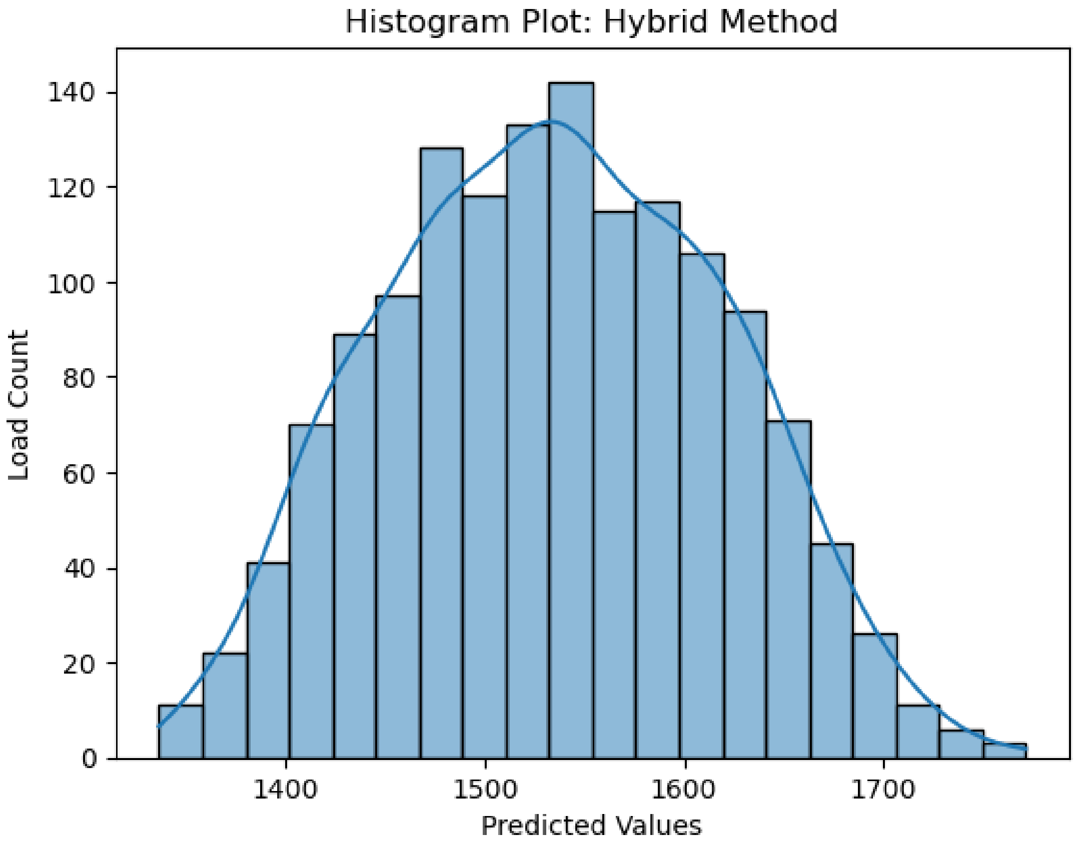


Figure 4.11 : Histogram Plot for Hybrid method

The histogram plot we provided represents the distribution of the predicted values from the Hybrid Method. The x-axis represents the range of predicted values, which in this case is from 1400 to 1700. The y-axis represents the load count, which indicates the frequency or count of occurrences for each predicted value range. Each bar on the histogram represents a range of predicted values. For example, the first bar represent the range of predicted values from 1400 to 1420, the second bar from 1420 to 1440, and so on. The height of each bar represents the load count, which indicates how many instances or occurrences of predicted values fall within that particular range. For instance, if the bar at 1400 has a height of 40, it means there are 40 instances where the predicted value was around 1400. The higher the bar, the higher the load count, indicating that more predicted values fall within that range. Conversely, lower bars represent ranges with fewer predicted values. By analyzing the histogram, we can observe the distribution pattern of the predicted values. It helps in understanding the concentration or spread of predictions within different value ranges. In our case, we provided load counts at specific intervals (0, 20, 40, 60, 80, 100, 120, 140), which indicates the frequency of predicted values falling within those ranges.

In summary, the histogram provides a visual representation of the predicted value distribution, allowing you to identify common ranges or concentrations of predicted values and assess the spread or variability of the predictions.

Chapter 5

Conclusions and Future Work

In this chapter we summarize the research works presented in this dissertation and make final concluding remarks with few directions for future works.

5.1 Summary of the Dissertation

In this chapter, a comprehensive summary of the dissertation on stlf is presented. The dissertation's objectives were to investigate and evaluate various models as well as to suggest a hybrid approach for precise stlf. The investigation was carried out with the intention of enhancing the effectiveness and dependability of load forecasting methods, which are very important in managing the electricity grid. The study started off with an overview of the importance of short-term load forecasting and its implications in the energy sector. The goals of the study were outlined, and they included the assessment of four different models and the creation of a hybrid approach that combined their advantages. Linear Regression, Random Forest, Neural Network & Support Vector Machine were the ones that were looked into. For the purpose of forecasting short-term load, each model was looked at in detail and evaluated. Their underlying assumptions, information needs, and constraints were carefully examined. The subsequent stage of the research involved the development of a hybrid method that integrated the strengths of the four individual models.

5.2 Future Research Directions

Although this dissertation has significantly advanced the field of short-term load forecasting, there are a number of promising lines of inquiry that could build on the findings and advance our understanding in this field. These new areas of investigation will address current issues, investigate cutting-edge methods, and advance load forecasting. Load patterns can be influenced by various dynamic factors, such as changes in consumer behavior, economic conditions, and government policies. Future research can focus on developing dynamic and adaptive load forecasting models that can quickly adapt to such changes and provide accurate predictions in real-time or near-real-time. These models can incorporate online learning algorithms, anomaly detection techniques, and adaptive forecasting methodologies to ensure robust and up-to-date load predictions.

Many power systems span across multiple regions or territories, each with unique load characteristics and dynamics. Future research can explore load forecasting models that can handle multi-region systems, accounting for interdependencies, regional variations, and cross-border power exchanges. Evaluating and comparing the performance of load forecasting models in such multi-region systems will provide insights into their scalability and effectiveness in diverse operational environments. To validate the effectiveness and practicality of load forecasting models, future research can focus on real-world

Bibliography

[1] B. Farsi, M. Amayri, N. Bouguila, and U. Eicker, "On short-term load forecasting using machine learning techniques and a novel parallel deep LSTM-CNN approach," *IEEE Access*, vol. 9, pp. 31191–31212, 2021, doi: 10.1109/ACCESS.2021.3060290.

[2] Y. Wang, "Short-term Power Load Forecasting Based on Machine Learning."

[3] S. Mishra, "SHORT TERM LOAD FORECASTING USING COMPUTATIONAL INTELLIGENCE METHODS Electronics & Communication Engineering (Specialization in Telematics and Signal Processing)," 2008.

[4] P. C. de Lima e Silva, C. A. Severiano, M. A. Alves, R. Silva, M. Weiss Cohen, and F. G. Guimarães, "Forecasting in non-stationary environments with fuzzy time series," *Applied Soft Computing Journal*, vol. 97, Dec. 2020, doi: 10.1016/j.asoc.2020.106825.

[5] N. Ahmad, Y. Ghadi, M. Adnan, and M. Ali, "Load Forecasting Techniques for Power System: Research Challenges and Survey," *IEEE Access*, vol. 10, pp. 71054–71090, 2022, doi: 10.1109/ACCESS.2022.3187839.

[6] "10984209".

[7] S. Singh, S. Hussain, and M. A. Bazaz, "Short term load forecasting using artificial neural network," in *2017 4th International Conference on Image Information Processing, ICIIP 2017*, Institute of Electrical and Electronics Engineers Inc., Mar. 2018, pp. 159–163. doi: 10.1109/ICIIP.2017.8313703.

[8] A. K. Loder, "Short Term Load Forecasting Using Recurrent and Feedforward Neural Networks." [Online]. Available: www.eia.gov

[9] A. Baliyan, K. Gaurav, and S. Kumar Mishra, "A review of short term load forecasting using artificial neural network models," in *Procedia Computer Science*, Elsevier B.V., 2015, pp. 121–125. doi: 10.1016/j.procs.2015.04.160.

[10] S. K. Acharya, Y. M. Wi, and J. Lee, "Short-term load forecasting for a single household based on convolution neural networks using data augmentation," *Energies (Basel)*, vol. 12, no. 18, Sep. 2019, doi: 10.3390/en12183560.

[11] E. Jaafar, A. Nainoon, and I. O. Habiballah, "Load Forecasting using Machine Learning Methods: Review." [Online]. Available: www.ijert.org

[12] I. Moghram and S. Rahman, "Analysis and Evaluation of Five Short-Term Load Forecasting Techniques," *IEEE Power Engineering Review*, vol. 9, no. 11. pp. 42–43, 1989. doi: 10.1109/MPER.1989.4310383.

[13] E. A. Feilat, D. T. Al-Sha'abi, M. A. Momani, and E. A. Feilat, "Long-term load forecasting using neural network approach for Jordan's power system," vol. 1, no. 2, pp. 43–50, 2017, doi: 10.28964/EngPress-1-108.

[14] D. Zhang, W. Guan, J. Yang, H. Yu, W. C. Xiao, and T. Yu, "Medium—And Long-Term Load Forecasting Method for Group Objects Based on the Image Representation Learning," *Front Energy Res*, vol. 9, Aug. 2021, doi: 10.3389/fenrg.2021.739993.

[15] M. Sarhani and A. El Afia, "Electric load forecasting using hybrid machine learning approach incorporating feature selection."

[16] R. Hariharan and M. K. Kumar, "A research on electric load forecasting factors effecting and methods involved," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 12, pp. 1462–1466, Oct. 2019, doi: 10.35940/ijitee.L3951.1081219.

[17] G. Singh, D. S. Chauhan, A. Chandel, D. Parashar, and G. Sharma, "Factor Affecting Elements and Short term Load forecasting Based on Multiple Linear Regression Method." [Online]. Available: www.ijert.org

[18] S. M. Anowarul Haque, M. Asiful Islam, and M. Simulink ---, "Short-Term (Seven Day Basis) Load Forecasting Of a Grid System in Bangladesh Using Artificial Neural Network," vol. 15, pp. 15–25, doi: 10.9790/1676-1504011525.

[19] S. Salehimehr, B. Taheri, and M. Sedighizadeh, "Short-term load forecasting in smart grids using artificial intelligence methods: A survey," *The Journal of Engineering*, vol. 2022, no. 12, pp. 1133–1142, Dec. 2022, doi: 10.1049/tje2.12183.

[20] C. A. Christodoulou, V. Mladenov, L. Ekonomou, and V. Mladenov, "A short-term load forecasting method using artificial neural networks and wavelet analysis Dealing with the emerging European electric energy challenges using advanced simulation capabilities View project Special Issue '10th Anniversary of Electronics-Hot Topic in Artificial Intelligence Circuits and Systems' View project Lambros Ekonomou UBITECH ENERGY Sprl A short-term load forecasting method using artificial neural networks and wavelet analysis," 2016. [Online]. Available: https://www.researchgate.net/publication/306095419

[21] E. L. Taylor, "Short-term Electrical Load Forecasting for an Institutional/ Short-term Electrical Load Forecasting for an Institutional/ Industrial Power System Using an Artificial Neural Network Industrial Power System Using an Artificial Neural Network." [Online]. Available: https://trace.tennessee.edu/utk_gradthes

[22] X. Ji *et al.*, "A Hybrid Residential Short-Term Load Forecasting Method Using Attention Mechanism and Deep Learning," *Buildings*, vol. 13, no. 1, Jan. 2023, doi: 10.3390/buildings13010072.

[23] L. Wu, C. Kong, X. Hao, and W. Chen, "A Short-Term Load Forecasting Method Based on GRU-CNN Hybrid Neural Network Model," *Math Probl Eng*, vol. 2020, 2020, doi: 10.1155/2020/1428104.

[24] M. Jacob, C. Neves, and D. Vukadinović Greetham, "Short Term Load Forecasting," 2020, pp. 15–37. doi: 10.1007/978-3-030-28669-9_2.

[25] E. A. Madrid and N. Antonio, “Short-term electricity load forecasting with machine learning,” *Information (Switzerland)*, vol. 12, no. 2, pp. 1–21, Feb. 2021, doi: 10.3390/info12020050.

[26] F. Elahe and M. Jin, “Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000. An Adaptive and Parallel Forecasting Strategy for Short-Term Power Load Based on Second Learning of Error Trend”, doi: 10.1109/ACCESS.2017.DOI.

[27] “Short-Term Load Forecasting.”

[28] M. Madhukumar, A. Sebastian, X. Liang, M. Jamil, and M. N. S. K. Shabbir, “Regression Model-Based Short-Term Load Forecasting for University Campus Load,” *IEEE Access*, vol. 10, pp. 8891–8905, 2022, doi: 10.1109/ACCESS.2022.3144206.

[29] https://www.aemo.com.au/

[30] https://www.iso-ne.com/isoexpress/

Turnitin Originality Report

Processed on: 02-Sep-2023 15:24 +06
ID: 2156257864
Word Count: 11452
Submitted: 1

221-44-236 By Md. Sakib Rahman

Similarity Index	Similarity by Source
19%	Internet Sources: 15% Publications: 10% Student Papers: 11%

2% match (Internet from 26-Oct-2021)
https://github.com/scikit-learn/administrative/blob/master/jmlr_2011_paper/pedregosa11a.tex

1% match (student papers from 05-Mar-2022)
Submitted to Green University Of Bangladesh on 2022-03-05

1% match (student papers from 03-Apr-2022)
Submitted to Green University Of Bangladesh on 2022-04-03

1% match (Md Fazla Elahe, Min Jin, Pan Zeng. "An Adaptive and Parallel Forecasting Strategy for Short-Term Power Load Based on Second Learning of Error Trend", IEEE Access, 2020)
Md Fazla Elahe, Min Jin, Pan Zeng. "An Adaptive and Parallel Forecasting Strategy for Short-Term Power Load Based on Second Learning of Error Trend", IEEE Access, 2020

1% match (student papers from 11-Feb-2018)
Submitted to Otago Polytechnic on 2018-02-11

1% match ()
Taylor, Eric Lynn. "Short-term Electrical Load Forecasting for an Institutional/Industrial Power System Using an Artificial Neural Network", TRACE: Tennessee Research and Creative Exchange, 2013

1% match (Internet from 22-Sep-2022)
https://turcomat.org/index.php/turkbilmat/article/download/11594/8513/20586

< 1% match (Internet from 08-Feb-2023)
https://www.researchgate.net/publication/342590596_Software_Defect_Prediction_Using_Metaheuristic-based_Feature_Selection_and_Classification_Algorithms

< 1% match (Internet from 30-Jan-2023)
https://www.researchgate.net/publication/221226477_Feature_Selection_with_Imbalanced_Data_for_Software_Defect_Prediction

< 1% match (Internet from 17-Mar-2023)
https://www.researchgate.net/publication/351949863_Foreign_Currency_Exchange_Rate_Prediction_Using_Bidirectional_Long_Short_Term_Me

< 1% match (Internet from 05-Feb-2023)
https://www.researchgate.net/publication/343091063_A_Comprehensive_Review_of_the_Load_Forecasting_Techniques_Using_Single_and_Hyb

< 1% match (Internet from 30-Jan-2023)
https://www.researchgate.net/publication/4354849_A_Combination_Method_for_Short_Term_Load_Forecasting_Used_in_Iran_Electricity_Marke

< 1% match (Internet from 12-Feb-2023)
https://www.researchgate.net/publication/3267507_Short-Term_Load_Forecasting_Based_on_an_Adaptive_Hybrid_Method

< 1% match (Md Fazla Elahe, Min Jin, Pan Zeng. "An Adaptive and Parallel Forecasting Strategy for Short-Term Power Load Based on Second Learning of Error Trend", IEEE Access, 2020)
Md Fazla Elahe, Min Jin, Pan Zeng. "An Adaptive and Parallel Forecasting Strategy for Short-Term Power Load Based on Second Learning of Error Trend", IEEE Access, 2020

< 1% match (Internet from 29-Jun-2023)
http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/9324/21651.pdf?isAllowed=y&sequence=1

< 1% match (Internet from 26-Oct-2022)
http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/8567/181-35-300.pdf?isAllowed=y&sequence=1

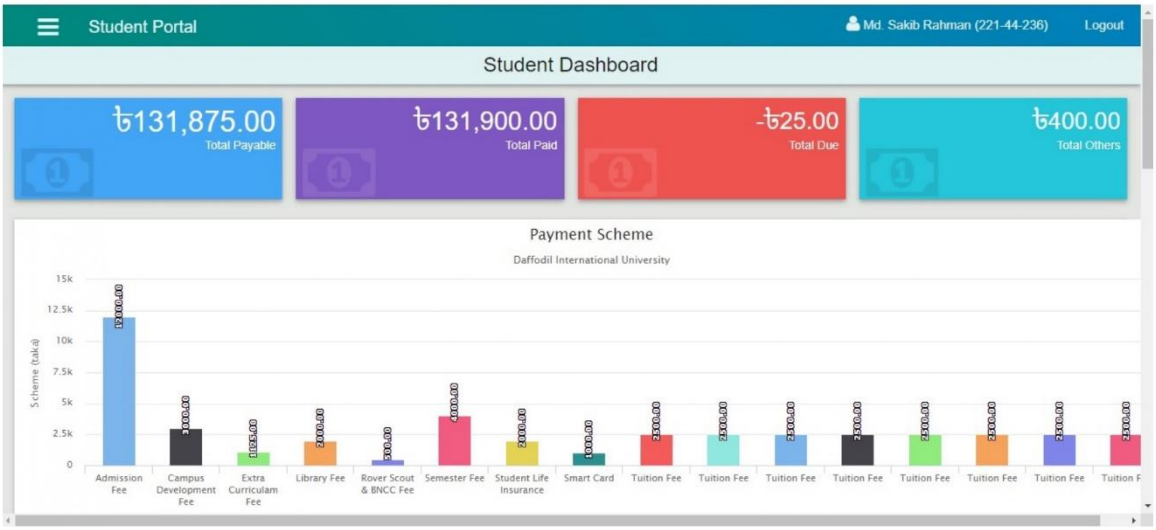
< 1% match (student papers from 10-Jul-2023)
Submitted to Liverpool John Moores University on 2023-07-10

< 1% match (student papers from 06-Mar-2023)
Submitted to Liverpool John Moores University on 2023-03-06

< 1% match (student papers from 27-Feb-2023)
Submitted to Liverpool John Moores University on 2023-02-27

< 1% match (student papers from 20-Aug-2023)
Submitted to Liverpool John Moores University on 2023-08-20

< 1% match (Internet from 30-Jun-2023)
https://as-proceeding.com/index.php/ijanser/article/download/787/745/1416


Student Portal
Md. Sakib Rahman (221-44-236)
Logout
Student Dashboard
৳131,875.00
Total Payable
৳131,900.00
Total Paid
-৳25.00
Total Due
৳400.00
Total Others
Payment Scheme
Daffodil International University
Scheme (taka)
15k
12.5k
10k
7.5k
5k
2.5k
0
Admission Fee
Campus Development Fee
Extra Curriculam Fee
Library Fee
Rover Scout & BNCC Fee
Semester Fee
Student Life Insurance
Smart Card
Tuition Fee
Tuition Fee
Tuition Fee
Tuition Fee
Tuition Fee
Tuition Fee
Tuition Fee