ENHANCING MICROGRID ENERGY MANAGEMENT USING MACHINE LEARNING

ME420 Mechanical Engineering Individual Research Project

November 2023

by

PIYUMI BHAGYA S. (E/17/259)

This report is submitted for the completion of the requirements of the course ME 420 which is offered for the Degree of Bachelor of the Science of Engineering (B.Sc.Eng) of the University of Peradeniya

Department of mechanical Engineering

Faculty of Engineering

University of Peradeniya

Sri Lanka

ENHANCING MICROGRID ENERGY MANAGEMENT USING MACHINE LEARNING

By

PIYUMI BHAGYA S. - E/17/259

Supervised by

Dr. Ubaya Higgoda
Dr. Damayanthi Herath
Dr. Lahiru Jayasuriya

Declaration

I confirm that I have written this report without any external help and not using sources other than those I have listed in the report. I confirm also that this report or similar version of it has not been submitted to any other examination board and has not been previously accepted as part of an exam for a qualification. Each direct quotation or paraphrase of an author is clearly referenced.

Place, 27th Dec 2023

Piyumi Bhagya S.

Name of the project supervisor/s:

Dr. Damayanthi Herath
Dr. Ubaya Higgoda
Dr. Lahiru Jayasuriya

Abstract

This research project focuses on the vital task of forecasting wind and photovoltaic (PV) energy for microgrid operations. Microgrids, integral to localized energy systems, depend on accurate predictions of renewable energy for optimal functionality. Using machine learning techniques, the study emphasizes developing practical forecasting models tailored to the dynamic characteristics of wind and PV energy sources. The objective is to improve prediction accuracy, addressing challenges related to intermittency and variability, and thereby enhancing the overall efficiency and sustainability of microgrid operations.

Moreover, the project extends its scope to a preliminary investigation of optimization methods for utilizing forecasted data in microgrid operations. This initial exploration seeks to identify, assess, and optimize strategies for integrating forecasted renewable energy seamlessly. Anticipated outcomes aim to offer practical insights for refining both renewable energy forecasting methodologies and optimization technologies, contributing to the advancement of resilient and environmentally conscious microgrid systems.

Acknowledgement

I extend my deepest appreciation to my outstanding supervisor, Dr. Damayanthi Herath, for her invaluable guidance and unwavering support throughout this research project. Gratitude also goes to my other supervisors, Dr. Ubaya Higgoda and Dr. Lahiru Jayasooriya, whose expertise shaped the trajectory of my work.

Special thanks to Dr. Hansini for her instrumental expertise in the success of this project. I am also thankful to all the lecturers who contributed to shaping my research methodology and providing constructive critiques on my results.

I want to express my heartfelt gratitude to the Faculty of Engineering at the University of Peradeniya for the opportunity and continuous support in navigating the complexities of this research endeavor.

My sincere thanks to colleagues and friends for their immeasurable assistance and countless opportunities during this research journey. Additionally, I am indebted to my family and friends for their unwavering encouragement and support throughout my academic pursuits.

This project has been a collective effort, and I am truly grateful for the support and encouragement received from all quarters.

Table of contents

Declaration i

Abstract ii

Acknowledgement iii

Table of contents iv

List of Symbols viii

1 Introduction 1

1.1 Aim of the project 2

1.2 Objectives 3

1.3 Problem formulation 3

1.3.1 Research question 3

1.3.2 Scope of the Problem 3

1.4 Proposed Methodology 3

2 Literature Review 5

2.1 Introduction 5

2.1.1 Background 5

2.2 Machine learning For Renewable Energy Forecasting 6

2.3 PV and Wind Energy Forecasting 7

2.4 Data Preprocessing and Feature Engineering 8

2.5 Evaluation Metrics and Model Selection 8

2.6 Optimization of Microgrid Operation 9

2.7 Conclusion of Literature Review 10

3 Methodology 11

3.1 Data Collection 11

3.1.1 Data Sources and Data Information 11

3.1.2 PV generation Data 15

3.1.3 Weather data 16

3.1.4 Hourly data 16

3.2 Pandas Profiling 17

3.3 EDA on PV Generation Data 17

3.4 EDA on Weather Data 18

3.4.1 Feature selection 18

3.5 Data Preprocessing 21

4 Weather Data Preprocessing 22

4.1 Visualize data 22

4.2 Missing Data Handling 23

4.3 Outlier detection and handling 23

4.3.1 Outlier Detection Methods for Weather Data 23

4.3.2 Time series decomposition 24

4.3.3 Manual Examination with Actual Data Plots 25

4.4 Feature Engineering 26

4.4.1 Wind 26

4.4.2 Time 27

4.5 Split Data 28

4.6 Normalizing data 29

5 PV Generation Data Preprocessing 30

5.1 Handling Negative Values 30

5.2 Visualize Data 30

5.3 Outlier Detection 32

5.3.1 Box Plot 32

5.3.2 Time Series Decomposition 33

5.3.3 Manual Examination with Actual Data Plots 34

5.4 Normalizing data 34

5.5 Visualize normalized data 34

5.6 Generalized Input Data for Forecasting Model 35

5.7 Missing Data Handling 35

5.8 Outlier Handling 35

5.9 Resampling data 35

5.10 Data Integration 35

6 Data Windowing and Forecasting Models 37

6.1 Data Windowing 37

6.2 Single shot models 37

6.2.1 Baseline model 37

6.2.2 CNN model 38

6.2.3 LSTM model 39

6.3 Training Procedure 41

6.3.1 Model Compilation 41

6.3.2 Custom Metrics 41

6.3.3 Training Execution 41

6.4 Relevant Equations for Neural Network Training 42

6.5 Hyperparameter Tunning 42

6.5.1 Hyperparameter Tuning for CNN Model 43

6.5.2 Hyperparameter Tuning for LSTM Model 43

6.5.3 Hyperparameter tunning of training 43

7 Optimization 45

7.1 Microgrid Optimization 45

7.2 Microgrid operation scenarios 45

7.3 Optimization methods and algorithms 46

8 Results and Discussion 49

8.1 Performances of Wind forecasting models 49

8.1.1 Loss Comparison 49

8.1.2 MAE Performance 50

8.1.3 RMSE Performance 50

8.1.4 Rsquare Performance 51

8.1.5 Discussion 51

8.1.6 Conclusion 52

8.2 Performances of PV Energy forecasting models 52

8.2.1 Loss Performance 52

8.2.2 MAE Performance 53

8.2.3 RMSE Performance 53
8.2.4 Rsquare Performance 54
8.2.5 Discussions and conclusions 54
8.3 Check for Overfitting Underfitting of NN models 55
8.3.1 Overfitting 55
8.3.2 Underfitting 55
8.3.3 Wind Forecasting CNN model 56
8.3.4 PV Energy Forecasting CNN model 56
8.3.5 Discussion 57
8.4 Visualize Forecasted Data 57
8.4.1 Actual vs Predicted Data 57
9 Conclusions **61**
References **63**
Appendix **65**

List of Symbols

STL - Seasonal-Trend decomposition using LOESS

EDA - Exploratory Data Analysis

ML - Machine Learning

DL - Deep Learning

MG - MicroGrids

PV - Photo Voltic

EV - Electro Voltic

MAE - Mean Absolute Error

RMSE- Root Mean Square Error

MAPE-Mean Absolute Percentage Error

ISO - International Standard Organization

IQR - Inter Quartile Range

NaN - Not a Number

CNN - Convolutional Neural Network

LSTM - Long Short Term Memory

List of Figures

Figure 3.1 Heat map correlation plot of features from Pandas Profiling Report 19
Figure 3.2 Overview of Variables in weather data- Pandas Profiling 21
Figure 4.1 Line plot of Weather Data over time 22
Figure 4.2 Sample of last 10 days of Weather Data over time 22
Figure 4.3 Seasonal decomposition Plots of Weather data 25
Figure 4.4 Heat Map of wind speed with wind direction 26
Figure 4.5 Distribution of wind vectors Heat Map 26
Figure 4.6 Day singnal 27
Figure 4.7 Year Signal 27
Figure 4.8 FFT of dwpt 28
Figure 4.9 Violin plot of the standardized data 29
Figure 5.1 Line plots of PV Generation data over time 30
Figure 5.2 Sample data of Last 20 days 31
Figure 5.3 Box plot for PV data 32
Figure 5.4 Seasonal Decomposition plots for *'CUP_PV'* 33
Figure 5.5 Zoomed Seasonal Decomposition plots for *'CUP_PV'* 34
Figure 5.6 Box Plot of Normalized PV data 34
Figure 6.1 Window representaion for the baseline model 38
Figure 6.2 Window representaion for the CNN model 38
Figure 6.3Window representaion for the LSTM model 40
Figure 8.1 Losss Comparision of Wind Forecasting models on the data sets 49
Figure 8.2MAE Performance of Wind Forecasting models on the data sets 50
Figure 8.3RMSE Performance of Wind Forecasting models on the data sets 50
Figure 8.4Accuracy Performance of Wind Forecasting models on the data sets 51
Figure 8.5 Losss Comparision of PV Energy Forecasting models on the data sets 52
Figure 8.6 MAE Performance of PV Energy Forecasting models on the data sets 53
Figure 8.7 RMSE Performance of PV Energy Forecasting models on the data sets 53
Figure 8.8 Accuracy Performance of PV Energy Forecasting models on the data sets 54
Figure 8.9 accuracy vs loss performances on train and validation data sets over epoches 56
Figure 8.10accuracy vs loss performances on train and validation data sets over epoches 56
Figure 8.11 3 split window samples of the Test datasets on PV generation forecasting using CNN model 58

Figure 8.12 3 split window samples of the Test datasets on wind speed forecasting using CNN model 58
Figure 8.13 Actual vs Predicted plot sample on test dataset of wind speed forecasting CNN model 59
Figure 8.14 Actual vs Predicted plot sample on test dataset of PV energy forecasting CNN model 59

List of Tables

Table 3.1 On-Campus Solar PV Generators data 13

Table 3.2 Meteorological Parameters 14

Table 3.3 Meteorological Data Units 14

Table 3.4 Weather Condition Codes 15

Table 3.5 Selected On-Campus Solar PV Generators data 16

Table 3.6 Data Description of PV combined data 17

Table 3.7 Data Description of Weather data 18

Table 3.8 Sample of the Weather data set from Pandas Profiling Report 19

Table 6.1 Summary of evaluation metric equations 42

Table 7.1Microgrid operation scenarios 46

Table 7.2 microgrid operation scenarios summarized 46

Table 8.1 Test data Performance Evaluation of Wind Forecasting Models 51

Table 8.2 Test data Performance Evaluation of PV Energy Forecasting Models 54

1 Introduction

Renewable energy sources, exemplified by solar and wind power, are assuming an increasingly pivotal role in addressing the burgeoning global energy demand through sustainable means. Despite their environmental merits, the intermittent and variable nature of these energy sources poses significant challenges for their seamless integration into conventional power systems. A critical component in addressing this challenge lies in the accurate forecasting of renewable energy generation, which is indispensable for the efficient planning and operation of power systems, including microgrids.

The inherent intermittency and variability of solar and wind power introduce complexities in their integration into power systems. The challenges extend beyond the simple fluctuation in energy output and delve into the intricacies of matching supply with demand in real-time. Traditional forecasting methods, relying on time series analysis and numerical weather prediction models, have historically grappled with the nuanced and dynamic nature of renewable energy generation.

In response to the limitations of traditional forecasting, recent years have witnessed a paradigm shift with the advent of machine learning (ML) and deep learning techniques. These methodologies offer a promising avenue for significantly enhancing the precision and reliability of renewable energy forecasting. By leveraging advanced algorithms, such as artificial neural networks, support vector machines, and decision trees, ML has demonstrated its potential to revolutionize the accuracy of predictions, thereby facilitating the seamless integration of renewable energy sources into power systems.

A burgeoning body of research has embarked on exploring the application of ML algorithms specifically for forecasting renewable energy generation from solar and wind sources. This exploration is pivotal, considering that traditional forecasting models often falter in capturing the intricate patterns inherent in renewable energy systems. Studies in this domain have not only showcased the efficacy of ML techniques in improving accuracy but have also illuminated novel avenues for enhancing the overall management of microgrids.

This report presents the research project on forecasting renewable energy for microgrids, which aims to develop and evaluate a convolutional neural network (CNN) model that can predict the energy output of solar and wind power sources based on weather data. The motivation for this project stems from the growing importance of renewable energy sources in meeting the global energy demand in a sustainable way, as well as the challenges posed by their intermittent and variable nature for the integration into conventional power systems. Accurate forecasting of renewable energy generation is essential for the efficient planning and operation of power systems, especially microgrids, which are small-scale power networks that can operate independently or in coordination with the main grid. Microgrids offer several benefits, such as increased reliability, resilience, flexibility, and reduced greenhouse gas emissions, but they also require careful management and coordination of their energy resources.

This project aimed to forecast renewable energy for microgrids using machine learning techniques. The methodology of this project involved the collection and analysis of historical data on renewable energy sources, the selection and optimization of suitable machine learning models, the evaluation and comparison of the model performance, and investigation of the application of the model to forecast future energy demand and supply. The outcomes of this project demonstrate the feasibility and benefits of using machine learning techniques for forecasting renewable energy for microgrids. The model developed in this project can be used as a tool for planning and managing microgrid operations and for enhancing the reliability and sustainability of microgrid systems. The future scope of this project includes extending the model to incorporate more renewable energy sources and scenarios, exploring different machine learning algorithms and techniques, and applying the model to real-world microgrid systems. This report will provide a comprehensive overview of the methodology, results, and conclusions of this project, as well as the future directions of this research.

The results of this project demonstrate the feasibility and effectiveness of using a CNN model for forecasting renewable energy generation for microgrids. The model developed in this project achieved high accuracy and robustness in predicting the solar and wind power output for different time horizons. The model also showed its applicability and usefulness in supporting the optimal operation and management of microgrids, by providing reliable information on the expected energy supply and demand balance. The future scope of this project includes improving the model by incorporating more input features, such as geographical and topographical factors, as well as exploring other forecasting methods, such as deep learning and ensemble learning. This report will provide a detailed description of the methodology, results, and conclusions of this project, as well as the future directions of this research.

1.1 Aim of the project

The project's objective is to apply machine learning methods to predict renewable energy production and explore optimization approaches and techniques to optimize the operation of Microgrids. Aim of this project is to use machine learning models and optimization strategies to increase the reliability of renewable energy forecasts and improve the overall performance and efficiency of Microgrids.

1.2 Objectives

The main objectives of this project are:

1. To forecast PV energy and wind energy for a microgrid system using machine learning techniques.
2. To evaluate the performance of the machine learning models on the microgrid data set.
3. To Investigate optimization methods that can use the forecasted data to optimize the microgrid operation and reduce the cost and carbon emissions.

1.3 Problem formulation

Based on our introductory discussion, the problem can be summarized with the following

1.3.1 Research question

How well time series and machine learning techniques perform in short-term (24 hours) forecasting of solar PV energy and wind speed and how to optimise microgrid operation using forecasted data?

1.3.2 Scope of the Problem

The project focuses on forecasting wind speed and PV generation, crucial parameters for renewable energy systems. The temporal scope covers daily forecast from hourly data. The geographical scope encompasses San Diego California to ensure the models are tailored to local weather patterns.

1.4 Proposed Methodology

The proposed methodology for this research consists of the following steps,

Literature survey:

Conduct a comprehensive review of the existing literature on microgrid forecasting to identify the current state-of-the-art methods and the research gaps.

Data collection and Data Preprocessing:

Collect Historical data on microgrid load, generation, weather, and other relevant variables from various sources and pre-process to ensure data quality, consistency, and compatibility. Use data preprocessing techniques like EDA on data set, sorting, cleaning, Feature engineering, normalizing, standardizing.

Selection of Forecast Models and training:

Based on the characteristics of the data and the forecasting objectives forecast models such as artificial neural networks, support vector machines, random forests, etc. will be selected. From the model select a most suitable model doing further survey and train using appropriate algorithms and parameters. Implement a baseline model to compare performance as well.

Model optimization and performance evaluation:

The forecast models will be optimized using various techniques such as cross-validation to improve their accuracy and robustness. The performance of the models will be evaluated using various metrics such as mean absolute error, root mean square error, mean absolute percentage error, R square etc.

Investigation on Use the forecasted data to optimise the microgrid operation:

Investigate how forecasted data on load and generation will be used as inputs to an optimization model that aims to minimize the operational cost and environmental impact of the microgrid while satisfying the technical and operational constraints. Consider numerous factors such as battery storage, demand response, renewable energy sources, etc. Give conclusions and recommendations.

2 Literature Review

2.1 Introduction

Renewable energy sources, such as solar and wind power, are becoming increasingly important for meeting the world's growing energy demand in a sustainable manner [4, 5]. However, the intermittent and variable nature of these energy sources presents challenges for their integration into power systems [1, 2]. Accurate forecasting of renewable energy generation is essential for effective planning and operation of power systems, including microgrids [3, 8].

Traditional forecasting methods, such as time series analysis and numerical weather prediction models, have limitations in their ability to accurately forecast renewable energy generation [6, 7]. In recent years, machine learning and deep learning techniques have emerged as promising approaches for improving the accuracy of renewable energy forecasting [1, 10].

A growing body of research has investigated the application of machine learning algorithms, such as artificial neural networks, support vector machines, and decision trees, for forecasting renewable energy generation from solar and wind sources [11]. These studies have demonstrated the potential of machine learning techniques to improve the accuracy of renewable energy forecasting and support the effective manaagement of microgrids .

This literature review aims to provide a comprehensive overview of the current state of research on the use of machine learning for forecasting renewable energy generation. The review will cover key studies and findings in this field [1- 10], and discuss the implications of these findings for the optimization of microgrid operation.

2.1.1 Background

Renewable energy sources, such as solar and wind power, have the potential to provide clean and sustainable energy to meet the world's growing demand. The integration of these energy sources into power systems, including microgrids, is essential for reducing greenhouse gas emissions and mitigating the impacts of climate change [4, 5].

However, the integration of renewable energy sources into power systems presents several challenges. One of the main challenges is the intermittent and variable nature of renewable energy generation, which can cause fluctuations in power supply and affect the stability of the grid [1, 2]. To address this challenge, accurate forecasting of renewable energy generation is essential for effective planning and operation of power systems [3, 8].

Traditional forecasting methods, such as time series analysis and numerical weather prediction models, have been widely used for renewable energy forecasting. However, these methods have limitations in their ability to accurately capture the complex relationships between weather variables and renewable energy generation [6, 7]. As a result, there is a need for more advanced forecasting techniques that can improve the accuracy of renewable energy forecasting [1, 10].

This background section provides an overview of the importance of renewable energy generation and the challenges associated with its integration into power systems. The next section will discuss the potential of machine learning techniques to address these challenges and improve the accuracy of renewable energy forecasting.

2.2 Machine learning For Renewable Energy Forecasting

Machine learning is a subfield of artificial intelligence that involves the development of algorithms that can learn from data and make predictions or decisions without being explicitly programmed. Machine learning techniques have been widely applied in various fields, including renewable energy forecasting [1, 10].

In recent years, there has been growing interest in the application of machine learning techniques for improving the accuracy of renewable energy forecasting. Machine learning algorithms, such as artificial neural networks, support vector machines, and decision trees, have been used to model the complex relationships between weather variables and renewable energy generation [11-17]. These algorithms can learn from historical data to make accurate predictions of future renewable energy generation [1, 2].

Several studies have demonstrated the potential of machine learning techniques to improve the accuracy of renewable energy forecasting. For example, Khayaty et al. (2021) developed an intelligent microgrid energy management system based on a deep learning approach, which was able to accurately forecast renewable energy generation and support effective microgrid operation [2]. Similarly, Raja et al. (2021) conducted a review of machine learning techniques for renewable energy forecasting and found that these techniques can significantly improve forecasting accuracy compared to traditional methods [4].

In addition to improving forecasting accuracy, machine learning techniques can also provide valuable insights into the factors that affect renewable energy generation. For example, Maduabuchi et al. (2023) used machine learning algorithms to estimate renewable energy potential based on climatic and weather data and found that these algorithms were able to identify important relationships between weather variables and renewable energy generation [5].

Overall, the existing literature suggests that machine learning techniques have significant potential to improve the accuracy of renewable energy forecasting and support the effective integration of renewable energy sources into power systems [1-27]. Further research is needed to develop and evaluate advanced machine learning algorithms for renewable energy forecasting, and to investigate their potential applications in microgrid management and optimization.

2.3 PV and Wind Energy Forecasting

Photovoltaic (PV) and wind energy are two of the most widely used renewable energy sources. These energy sources have the potential to provide clean and sustainable energy to meet the world's growing demand. However, the accurate forecasting of PV and wind energy generation is essential for their effective integration into power systems [4, 5].

PV and wind energy generation are affected by a range of weather variables, such as solar irradiance, wind speed, and air temperature. Traditional forecasting methods, such as time series analysis and numerical weather prediction models, have limitations in their ability to accurately capture the complex relationships between these variables and renewable energy generation [6, 7]. As a result, there is a need for more advanced forecasting techniques that can improve the accuracy of PV and wind energy forecasting [1, 10].

Machine learning techniques have emerged as promising approaches for improving the accuracy of PV and wind energy forecasting. Several studies have investigated the application of machine learning algorithms, such as artificial neural networks, support vector machines, and decision trees, for forecasting PV and wind energy generation [11-17]. These studies have demonstrated the potential of machine learning techniques to accurately model the complex relationships between weather variables and renewable energy generation [1, 2].

For example, Rodríguez et al. (2018) used artificial neural networks to predict solar energy generation using weather forecasts for microgrid control. Their results showed that the machine learning-based approach was able to accurately forecast solar energy generation and support effective microgrid operation [16]. Similarly, Husein and Chung (2019) used a long short-term memory recurrent neural network to forecast day-ahead solar irradiance for microgrids. Their results showed that the deep learning-based approach was able to significantly improve the accuracy of solar irradiance forecasting compared to traditional methods [18].

Overall, the existing literature suggests that machine learning techniques have significant potential to improve the accuracy of PV and wind energy forecasting [1-27]. Further research is needed to develop and evaluate advanced machine learning algorithms for PV and wind energy forecasting, and to investigate their potential applications in microgrid management and optimization.

2.4 Data Preprocessing and Feature Engineering

Data preprocessing and feature engineering are important steps in the development of accurate forecasting models [1, 4]. These steps involve preparing the data for analysis by cleaning, transforming, and selecting relevant features.

Data preprocessing may involve filling missing data, removing outliers, and normalizing or scaling the data [1, 4]. Feature engineering involves selecting relevant features and transforming them into a format that can be used by the ML or DL algorithm [1, 4]. This may involve encoding categorical variables, creating interaction terms, or applying mathematical transformations to the data.

In the context of renewable energy forecasting, data preprocessing and feature engineering may involve selecting relevant weather and time-related features, such as temperature, wind speed, and time of day [1, 4]. These features may be transformed or combined to create new features that are more predictive of renewable energy generation [1, 4].

In conclusion, data preprocessing and feature engineering are essential steps in the development of accurate forecasting models. Careful attention to these steps can improve the performance of ML and DL algorithms in forecasting renewable energy generation [1, 4].

2.5 Evaluation Metrics and Model Selection

Evaluation metrics are used to assess the performance of renewable energy forecasting models [1, 4]. Common evaluation metrics used in renewable energy forecasting include mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE) [1, 4]. These metrics provide a measure of the accuracy of the forecasting model and can be used to compare the performance of different models. Model selection involves choosing the best ML or DL algorithm for a given forecasting task [1, 4]. This may involve comparing the performance of different algorithms using evaluation metrics, as well as considering other factors such as computational complexity and interpretability [1, 4]. Cross-validation is a commonly used technique for model selection, which involves splitting the data into training and validation sets and assessing the performance of the model on the validation set [1, 4]. In conclusion, evaluation metrics and model selection are important steps in the development of accurate forecasting models. Careful attention to these steps can improve the performance of ML and DL algorithms in forecasting renewable energy generation [1, 4].

2.6 Optimization of Microgrid Operation

Machine Learning has been shown to be effective for forecasting energy demand. According to research it shows that ML was able to produce more accurate forecasts than traditional methods. As Electrical load forecasting algorithms are needed for prediction of the energy demand for the day ahead, a few weeks up to a year or a period of over a year [9]. So, the methods that are being used have to be more accurate and efficient. The choice of ML model will depend on the specific needs of the energy planner or decision-maker. For example, if the planner is interested in forecasting short-term demand, then an SVR algorithm may be the best option. However, if the planner is interested in forecasting long-term demand, then an ANN may be more appropriate.

In literature it has used many ML methods for forecasting energy demand. Enea Mele (2019) have given a review about short-term forecasting such as Support Vector Machines (SVM), k-Nearest Neighbors (kNN), Random Forest and Artificial Neural Networks (ANN) and compare their performance efficiency, capabilities and limitations.[9].

In [1] it shows that there have been proposed Multiple deep learning techniques in past. These include artificial neural networks (ANN), convolutional neural networks (CNN), recurrent neural networks (RNN), Long short-term memory networks (LSTM) and bidirectional long short-term memory networks (BLSTM).

Maciej Slowik and Wieslaw Urban have developed a universal forecasting tool for energy consumption by end-use consumers. This model allows the end-users to be equipped with an energy demand prediction, enabling them to participate more effectively in the smart grid energy market. A single, long short-term memory (LSTM)-layer-based artificial neural network model for short-term energy demand prediction was developed.[4].

Among these methods deep learning Deep Learning models are a good alternative to learn patterns from customer data and then forecast demand for different forecasting horizons. deep Learning uses multiple layers of neurons composed of complex structures to model high-level data abstractions [12]. most commonly used deep learning-based methods for energy management and power forecasting, namely, artificial neural networks (ANN), deep neural networks (DNN), convolutional neural networks (CNN), and recurrent neural networks (RNN) [13].

Another paper [14] considers a load forecasting problem in residential areas as well as in commercial buildings. A deep RNN is employed for medium to long term energy consumption forecasting. Simulation results show the effectiveness of the proposed deep RNN based model over MLP for load demand prediction of commercial buildings.

In research work [1] also adapts DL based methods for load forecasting. In there a hybrid forecasting method is developed by combining the best features of CNN and K-means clustering. The results show that their hybrid CNN–K-means forecasting algorithm has higher accuracy.

So, there is more research in forecasting energy demand of microgrid. And this existing literature suggests that machine learning techniques have significant potential to improve the accuracy of energy demand forecasting [10]. Further research is needed to develop and evaluate advanced machine learning algorithms for energy demand forecasting, and to investigate their potential applications in microgrid management and optimization.

2.7 Conclusion of Literature Review

In conclusion, this literature review has provided a comprehensive overview of the current state of research on machine learning for forecasting renewable energy generation. The review has covered key studies and findings in this field and discussed the implications of these findings for the optimization of microgrid operations.

The existing literature suggests that machine learning techniques have significant potential to improve the accuracy of renewable energy forecasting and support the effective integration of renewable energy sources into power systems. Machine learning algorithms, such as artificial neural networks, support vector machines, and decision trees, have been used to model the complex relationships between weather variables and renewable energy generation. These algorithms have demonstrated their ability to accurately forecast renewable energy generation and support effective microgrid operation.

Further research is needed to develop and evaluate advanced machine learning algorithms for renewable energy forecasting, and to investigate their potential applications in microgrid management and optimization. There is a need for more research on the use of machine learning techniques for forecasting PV and wind energy generation, and for optimizing microgrid operation using forecasted data.

Overall, this literature review has highlighted the significant potential of machine learning techniques to improve the accuracy of renewable energy forecasting and support the sustainable development of power systems. The findings from this review provide a foundation for future research in this field and can help to guide the development of advanced machine learning algorithms for renewable energy forecasting and microgrid management.

3 Methodology

The methodology employed in this project is instrumental in achieving the objectives outlined in the problem formulation. As the world pivots towards sustainable energy solutions, the accurate forecasting of critical parameters, such as wind speed and photovoltaic (PV) generation, becomes paramount for optimizing renewable energy production. This section provides an overview of the structured and systematic approach adopted to address the forecasting challenges and derive meaningful insights from the collected data.

3.1 Data Collection

Accurate microgrid forecasting hinges on a foundation of meticulous data collection and preparation.

In the data collection process, the importance of information cannot be overstated. Comprehensive information provides context, ensuring accurate interpretation and analysis of collected data. It contributes to data accuracy, facilitates research reproducibility, addresses ethical considerations, supports data validation, and promotes methodological rigor. Clear and detailed information enhances effective communication of research findings and is fundamental to maintaining the integrity and credibility of the research process.

3.1.1 Data Sources and Data Information

The primary data sources consist of two indispensable opensource datasets.

1. (PV) generation data

The first dataset used in this research comes from historical photovoltaic (PV) generation data found in an open-source multi-year dataset by Silwal et al. (2021) [3]. For more detailed information, the associated paper provides a comprehensive explanation of the data. This dataset publicly presents a comprehensive power dataset covering parts of the University of California, San Diego microgrid. This sophisticated microgrid integrates various distributed energy resources (DERs) such as solar power plants, electric vehicles, buildings, a combined heat and power gas-fired plant, and electric and thermal storage. Most datasets in this collection provide 15-minute averages of real and reactive power spanning from January 1, 2015, to February 29, 2020. Serving as the primary input, this dataset offers valuable insights into the variability of PV generation within a microgrid context.

Here is a summary of the data included in this paper,

- **Data scarcity and availability**: The authors address the lack of open-source, high-resolution power data for microgrid research and share a comprehensive dataset from parts of the University of California, San Diego microgrid.

- **Microgrid features and components**: The microgrid consists of various distributed energy resources, such as solar, electric vehicles, buildings, gas-fired plants, and storage, that can operate independently or in coordination with the utility grid.

- **Data characteristics and quality**: The dataset covers five years of 15-min power data for different generators, loads, and storage devices.

- **Power system data:** The paper provides power generation, consumption, and storage data from different components of the UC San Diego microgrid, such as buildings, EV charging stations, solar PV generators, battery energy storage systems, and thermal energy storage systems. The data covers a period of about five years from 2015 to 2020 and has a resolution of 15 minutes.
- **Data analysis:** The paper presents some basic analysis of the data, such as the campus thermal load and storage, the demand charge calculation, and the thermodynamic states and processes of the chilled water loop. The paper also provides some examples of potential applications of the data for microgrid research, such as optimization, forecasting, and smart charging.

For this Project only the PV Generation data is considered for Forecasting PV energy.
Table 01 shows the On Campus Solar PV Generators data infomation given in the paper.

Facility/function	**PV (DC, AC kW) Rating**	**Data file name**	**Start date**	**End date**	**Missing days**
Biomedical Sciences Library	Unk, 390	Bsb_librarypv.csv	21 Jan, 2015	29 Feb, 2020	0.5
Biomedical Sciences Building	284, Unk	bsb_buildingpv.csv	21 Jan, 2015	30 Aug, 2018	0.2
Bio-Engineering Hall	Unk, 74	Bioengineeringpv.csv	21 Jan, 2015	29 Feb, 2020	26.7
Campus Service Complex	Unk, 54	Csc_buildingpv.csv	15 Jan, 2016	29 Feb, 2020	49.4
Central Utility Plant	Unk, 65	Cup_pv.csv	01 Jan, 2015	29 Feb, 2020	0.4
Engineering Building Unit II	43, 35	Ebu2_a_pv.csv	27 April, 2015	29 Feb, 2020	0.3
Engineering Building Unit II	37, 31	Ebu2_b_pv.csv	27 April, 2015	29 Feb, 2020	0.3
Electric Shop	Unk, Unk	Electricshoppv.csv	24 Oct, 2015	29 Feb, 2020	0.2
Fleet Services	29, 24	Garagefleetspv.csv	18 Mar, 2016	29 Feb, 2020	0.4
Gilman Parking	195, 200	Gilmanparkingpv.csv	09 May, 2015	29 Feb, 2020	812.1
Hopkins Parking	338, 350	Hopkinsparkingpv.csv	29 Aug, 2015	29 Feb, 2020	1

Keeling Apartments	41, Unk	Keelinga_pv.csv	15 May, 2017	29 Feb, 2020	0.1
Keeling Apartments	Unk, Unk	Keelingb_pv.csv	15 May, 2017	29 Feb, 2020	0.1
Otterson Hall	18, Unk	Kyoceraskylinepv.csv	14 Feb, 2016	29 Feb, 2020	200.7
Leichtag Biomedical Research	Unk, 50	Leichtagpv.csv	01 Jan, 2015	29 Feb, 2020	0.3
MESOM Laboratory	61, Unk	Mesom_pv.csv	31 Mar, 2016	29 Feb, 2020	14.3
Mayer Hall	Unk, 120	Mayerhallpv.csv	01 Jan, 2015	29 Feb, 2020	0.4
Oslar Parking	268, Unk	Oslerparkingpv.csv	17 Dec, 2018	29 Feb, 2020	185.4
Price Center	63, 75	Pricecentera_pv.csv	27 April, 2015	29 Feb, 2020	0.3
Price Center	66, 75	Pricecenterb_pv.csv	23 May, 2015	29 Feb, 2020	0.2
SD Supercomputing Center	Unk, 65	Sdsc_pv.csv	01 Jan, 2015	29 Feb, 2020	0.3
Structural and Material Engineering	Unk, 120	Sme_solarpv.csv	14 Oct, 2016	29 Feb, 2020	0.1
Powell Structural Research Lab	6.5, Unk	Powellpv.csv	01 Jan, 2015	03 Mar, 2016	0.1
Birch aquarium	49, Unk	Stephenbirchpv.csv	12 April, 2016	29 Feb, 2020	0.1

Table 3.1 On-Campus Solar PV Generators data

2. Weather Data

The second data set is from Historical weather and climate data from the San Diego International Airport, CA, US[21]. In leveraging Meteostat for our forecasting project, we incorporated San Diego International Airport weather data. Meteostat, operating as a global data aggregator, follows international WMO standards and relies on reputable national weather services. The integration seamlessly accessed raw, on-site measurements, offering a mix of real observations and model data by default, with the flexibility for users to opt-out of model data. Despite potential small inconsistencies and variations in aggregation methods, Meteostat's data, sourced from reliable channels, is generally accurate, providing a robust foundation for our forecasting analysis.

Data information summary is as folllows,

- **Time format:** Meteostat follows the ISO 8601 standard and uses Coordinated Universal Time (UTC).
- **Time ranges:** JSON API endpoints require the start and end parameters to be dates in the format YYYY-MM-DD.
- **Meteorological parameters:** The abbreviations and meanings of various meteorological parameters, such as temperature, precipitation, wind, humidity, pressure, snow, sunshine, and weather condition.
- **Meteorological data units:** The text shows the units used for each meteorological parameter, such as °C, mm, hPa, km/h, etc.
- **Weather condition codes:** The text provides a table of integer values and corresponding weather conditions, ranging from clear to storm.

The Meteostat API uses abbreviations to describe meteorological parameters as given in Table 3.2.

Code	**Meaning**
TEMP	Air Temperature
TAVG	Average Temperature
TMIN	Minimum Temperature
TMAX	Maximum Temperature
DWPT	Dew Point
PRCP	Total Precipitation
WDIR	Wind (From) Direction
WSPD	Average Wind Speed
WPGT	Wind Peak Gust
RHUM	Relative Humidity
PRES	Sea-Level Air Pressure
SNOW	Snow Depth
TSUN	Total Sunshine Duration
COCO	Weather Condition Code

Table 3.2 Meteorological Parameters

Units of the Parameters are shown in Table 3.3 below.

Parameter(s)	**Unit**
Temperature	°C
Precipitation	mm
Sunshine Duration	Minutes
Air Pressure	hPa
Wind Speed, Peak Wind Gust	km/h
Wind Direction	Degrees
Visibility, Cloud Height	m
Relative Humidity	%

Table 3.3 Meteorological Data Units

Weather conditions are indicated by an integer value between 1 and 27 according to this list as shown in Table 3.4 below.

Code	**Weather Condition**
1	Clear
2	Fair
3	Cloudy
4	Overcast
5	Fog
6	Freezing Fog

7	Light Rain
8	Rain
9	Heavy Rain
10	Freezing Rain
11	Heavy Freezing Rain
12	Sleet
13	Heavy Sleet
14	Light Snowfall
15	Snowfall
16	Heavy Snowfall
17	Rain Shower
18	Heavy Rain Shower
19	Sleet Shower
20	Heavy Sleet Shower
21	Snow Shower
22	Heavy Snow Shower
23	Lightning
24	Hail
25	Thunderstorm
26	Heavy Thunderstorm
27	Storm

Table 3.4 Weather Condition Codes

3.1.2 PV generation Data

From the comprehensive dataset presented in Table 3.1, a meticulous selection process was undertaken to choose 11 PV generation datasets. Key factors, including the starting date, end date, and the number of missing days, were meticulously considered during this selection. This careful consideration ensures that by incorporating information on starting and ending dates, along with minimizing missing data, the selected PV generation datasets form a representative and reliable subset for further analysis and model development.

Facility/function	**PV (DC, AC kW) Rating**	**Data file name**	**Start date**	**End date**	**Missing days**
Biomedical Sciences Library	Unk, 390	Bsb_librarypv.csv	21 Jan, 2015	29 Feb, 2020	0.5
Central Utility Plant	Unk, 65	Cup_pv.csv	01 Jan, 2015	29 Feb, 2020	0.4
Engineering Building Unit II	43, 35	Ebu2_a_pv.csv	27 April, 2015	29 Feb, 2020	0.3
Engineering Building Unit II	37, 31	Ebu2_b_pv.csv	27 April, 2015	29 Feb, 2020	0.3
Electric Shop	Unk, Unk	Electricshoppv.csv	24 Oct, 2015	29 Feb, 2020	0.2
Fleet Services	29, 24	Garagefleetspv.csv	18 Mar, 2016	29 Feb, 2020	0.4
Hopkins Parking	338, 350	Hopkinsparkingpv.csv	29 Aug, 2015	29 Feb, 2020	1

Mayer Hall	Unk, 120	Mayerhallpv.csv	01 Jan, 2015	29 Feb, 2020	0.4
SD Supercomputing Center	Unk, 65	Sdsc_pv.csv	01 Jan, 2015	29 Feb, 2020	0.3
Birch aquarium	49, Unk	Stephenbirchpv.csv	12 April, 2016	29 Feb, 2020	0.1

Table 3.5 Selected On-Campus Solar PV Generators data

In Table 3.5, missing data is represented as missing days. Considering a 15-minute interval, one missing day corresponds to 24*4 data points. The selected data exhibits relatively low missing values. Notably, while the End Date remains consistent, variations in the Starting Date are observed among different PV generators, emphasizing the diverse temporal characteristics of the chosen datasets.

3.1.3 Weather data

Hourly weather data from January 21, 2015, to February 29, 2020, has been considered for analysis. This timeframe corresponds with the selected PV generation data range, allowing for a examination of the weather conditions and their impact on photovoltaic energy production.

3.1.4 Hourly data

In this study, the decision to consider hourly data stems from its inherent advantages in forecasting and analysis. Hourly data offers a more aggregated and smoother representation of patterns, simplifying the analysis and interpretation process. This choice aligns with common forecasting practices, striking a balance between capturing energy generation variations and avoiding excessive detail. Additionally, the use of hourly data contributes to computational efficiency, reducing the complexity of the forecasting task and mitigating the impact of short-term fluctuations or noise in the dataset. Overall, the selection of hourly data is a strategic decision aimed at optimizing the forecasting process

To align the photovoltaic (PV) data with hourly intervals, which is consistent with the selected hourly weather data, a conversion process is applied. The PV data, initially recorded in 15-minute intervals, undergoes a transformation to hourly granularity by calculating the mean of every four consecutive 15-minute data points. This conversion not only harmonizes the temporal resolution of both datasets but also ensures a synchronized analysis, facilitating a meaningful exploration of the relationship between hourly weather conditions and the corresponding photovoltaic energy generation.

The equation for converting 15-minute interval PV data to hourly data by calculating the mean of every four consecutive 15-minute data points can be expressed as follows:

Hourly PV Data =(PV Data_t + PV Data_(t+1) + PV Data_(t+2) + PV Data_(t+3)) / 4

Where *Hourly PV Generation* is the average PV generation for the hour starting at time *t. PV Data_t* is the PV generation data point at time *t* (15-minute interval). *t+1, t+2,* and *t+3* represent the subsequent 15-minute intervals within the same hour. The division by 4 calculates the mean of the four data points.

3.2 Pandas Profiling

In the realm of data analysis, leveraging the Pandas Profiling library emerges as a powerful tool to comprehend the intricacies of our dataset. This tool generates a comprehensive report that encapsulates statistical summaries, distribution analyses, and correlation matrices, shedding light on both the structure and patterns within the data. The report serves as a quick yet profound exploration, uncovering potential issues like missing values, duplicate entries, and outliers. With Pandas Profiling, the initial stages of data analysis become more efficient, providing a solid foundation for informed decision-making and subsequent modeling endeavors.

The detailed Pandas profiling reports are provided in the Appendix section at the end of the report for reference.

3.3 EDA on PV Generation Data

In the Exploratory Data Analysis (EDA) phase for PV Generation Data, the initial step involves sorting the data chronologically based on datetime.

Duplicate entries are then systematically eliminated, ensuring a refined dataset.

To create a comprehensive overview, data from the 10 PV generation datasets is amalgamated into a single data frame.

	EBU2_A_PV	EBU2_B_PV	ElectricShopPV	GarageFleetsPV	HopkinsParkingPV	StephenBirchPV	BSB_LibraryPV	CUP_PV	MayerHallPV	SDSC_PV
count	169796	169792	152537	138499	157844	136167	178033	178042	178036	178050
mean	7.693715	6.268381	15.346192	5.49539	57.871618	8.18325	24.883366	12.867729	27.246169	11.104686
std	11.226992	9.151862	23.408834	7.865599	86.308393	11.90596	36.346223	19.020354	40.393563	16.466745
min	0	0	-0.119	0.003	-0.759	0	-0.259	-0.068	-0.606	-0.001
max	34.977	29.837	86.224	24.533	328.436	45.762	143.878	72.969	164.853	65.313

Table 3.6 Data Description of PV combined data

However, the presence of varying starting dates poses a challenge, resulting in missing values at the initial timestamps. Therefore, handling missing values in the dataset requires attention to two distinct issues: actual data gaps and discrepancies arising from varying starting dates.

A closer look at Table 3.5 reveals some instances of negative values, which contradicts the inherent nature of power generation—power generation cannot be negative. To rectify this, it is crucial to replace these incorrect values with suitable alternatives. This corrective action not only ensures data consistency but also aligns with the fundamental principle that power generation values should be non-negative.

The variance in mean and standard deviation across the 10 PV generators is attributed to their distinct power ratings, signifying the maximum power each generator can produce. Since power ratings vary among the generators, it naturally results in variations in mean and standard deviation values. This diversity in power capacity highlights the need to consider individual characteristics when analyzing and interpreting the data. Taking these differences into account is essential for accurate assessments and insightful conclusions regarding the performance of each PV generator within the microgrid.

3.4 EDA on Weather Data

As the very first step data distribution was obtained as shown in table 3.7 to take an initial idea about the data first.

	temp	dwpt	rhum	prcp	snow	wdir	wspd	wpgt	pres	tsun	coco
count	44783	44780	44780	43250	0	40288	44782	0	44320	0	15543
mean	18.506866	11.488899	66.537204	0.025794	NaN	236.851594	8.926316	NaN	1015.731209	NaN	3.139484
std	4.108475	5.695723	17.210477	0.343637	NaN	94.976269	6.566615	NaN	3.568052	NaN	1.754997
min	3.9	-20.7	4	0	NaN	0	0	NaN	999.3	NaN	0
max	37.8	22.9	100	30	NaN	360	59.4	NaN	1032.5	NaN	25

Table 3.7 Data Description of Weather data

3.4.1 Feature selection

In refining the machine learning model for renewable energy forecasting in microgrid operation, careful consideration was given to feature selection.

Looking at the count values in Table 3.7 , *'snow,' 'wpgt,' 'tsun,'* and *'coco'* features were excluded due to their minimal contribution or the availability of data, making them unsuitable for meaningful analysis.

Still the missing values count shows considerably high missing values for wdir and prcp. Missing values count is shown infront of each remaining features as shown below.

- *temp* 1
- *dwpt* 4
- *rhum* 4
- *prcp* 1534
- *wdir* 4496
- *wspd* 2
- *pres* 464

From the pandas profiling Heat map Correlation plot was observed to see the correlation of *prcp*.

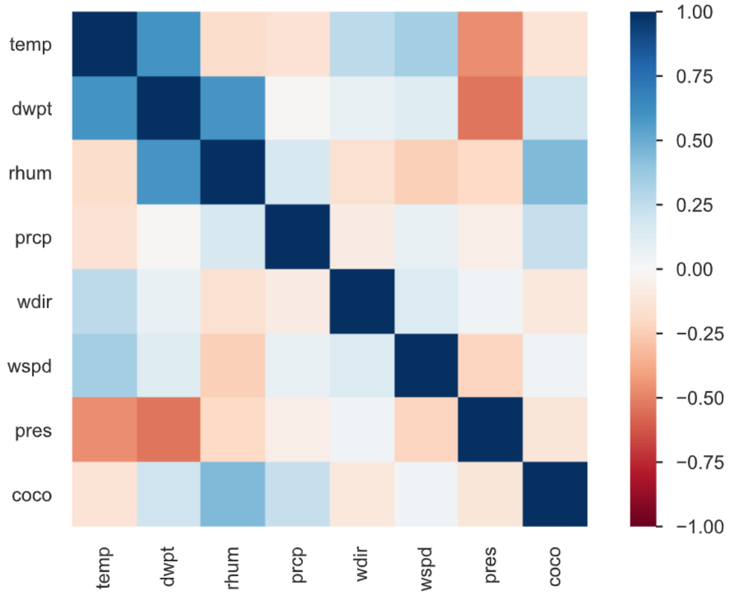


Figure 3.1 Heat map correlation plot of features from Pandas Profiling Report

The 'prcp' feature, characterized by a substantial number of missing values and lower correlation with other features which can observed from table 3.8 , was also excluded from the dataset. This decision aimed to enhance the model's reliability in accurate forecasting.

Sample

First rows | Last rows

	time	temp	dwpt	rhum	prcp	snow	wdir	wspd	wpgt	pres	tsun	coco
44774	12/31/2015 14:00	7.2	1.7	68.0	0.0	NaN	30.0	5.4	NaN	1019.5	NaN	NaN
44775	12/31/2015 15:00	7.2	1.7	68.0	0.0	NaN	NaN	0.0	NaN	1019.9	NaN	NaN
44776	12/31/2015 16:00	8.3	1.7	63.0	0.0	NaN	NaN	0.0	NaN	1020.4	NaN	NaN
44777	12/31/2015 17:00	10.6	2.2	56.0	0.0	NaN	NaN	0.0	NaN	1020.6	NaN	NaN
44778	12/31/2015 18:00	13.9	2.2	45.0	0.0	NaN	NaN	0.0	NaN	1020.9	NaN	NaN
44779	12/31/2015 19:00	15.6	0.6	36.0	0.0	NaN	NaN	0.0	NaN	1020.5	NaN	NaN
44780	12/31/2015 20:00	17.8	-3.1	24.0	0.0	NaN	300.0	0.0	NaN	1019.3	NaN	NaN
44781	12/31/2015 21:00	17.8	0.4	31.0	0.0	NaN	300.0	11.2	NaN	1018.4	NaN	NaN
44782	12/31/2015 22:00	17.2	1.2	34.0	0.0	NaN	300.0	13.0	NaN	1017.9	NaN	NaN
44783	12/31/2015 23:00	16.7	2.3	38.0	0.0	NaN	320.0	16.6	NaN	1017.6	NaN	NaN

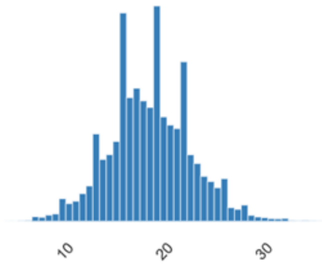
Table 3.8 Sample of the Weather data set from Pandas Profiling Report

From the Pandas Profiling Report the variable overviews was observed as follows.

temp

Real number (ℝ)

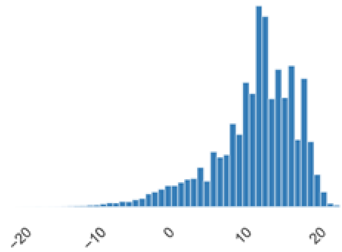
Distinct	118	Minimum	3.9
Distinct (%)	0.3%	Maximum	37.8
Missing	1	Zeros	0
Missing (%)	< 0.1%	Zeros (%)	0.0%
Infinite	0	Negative	0
Infinite (%)	0.0%	Negative (%)	0.0%
Mean	18.506866	Memory size	350.0 KiB



dwpt

Real number (ℝ)

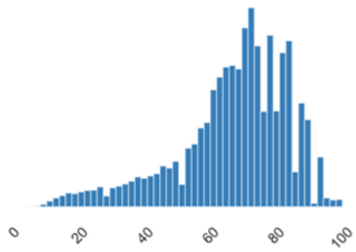
Distinct	355	Minimum	-20.7
Distinct (%)	0.8%	Maximum	22.9
Missing	4	Zeros	97
Missing (%)	< 0.1%	Zeros (%)	0.2%
Infinite	0	Negative	2235
Infinite (%)	0.0%	Negative (%)	5.0%
Mean	11.488899	Memory size	350.0 KiB



rhum

Real number (ℝ)

Distinct	96	Minimum	4
Distinct (%)	0.2%	Maximum	100
Missing	4	Zeros	0
Missing (%)	< 0.1%	Zeros (%)	0.0%
Infinite	0	Negative	0
Infinite (%)	0.0%	Negative (%)	0.0%
Mean	66.537204	Memory size	350.0 KiB

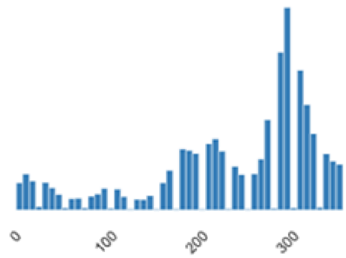


wdir

Real number (ℝ)

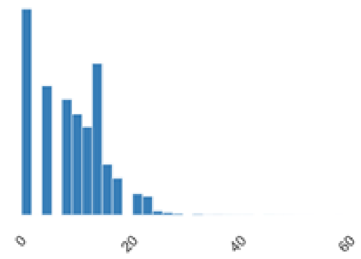
MISSING ZEROS

Distinct	357	Minimum	0
Distinct (%)	0.9%	Maximum	360
Missing	4496	Zeros	532
Missing (%)	10.0%	Zeros (%)	1.2%
Infinite	0	Negative	0
Infinite (%)	0.0%	Negative (%)	0.0%
Mean	236.84241	Memory size	350.0 KiB



wspd
Real number (ℝ)

Distinct	32	Minimum	0
Distinct (%)	0.1%	Maximum	59.4
Missing	2	Zeros	9934
Missing (%)	< 0.1%	Zeros (%)	22.2%
Infinite	0	Negative	0
Infinite (%)	0.0%	Negative (%)	0.0%
Mean	8.9263164	Memory size	350.0 KiB



pres
Real number (ℝ)

HIGH CORRELATION MISSING

Distinct	297	Minimum	999.3
Distinct (%)	0.7%	Maximum	1032.5
Missing	464	Zeros	0
Missing (%)	1.0%	Zeros (%)	0.0%
Infinite	0	Negative	0
Infinite (%)	0.0%	Negative (%)	0.0%
Mean	1015.7312	Memory size	350.0 KiB

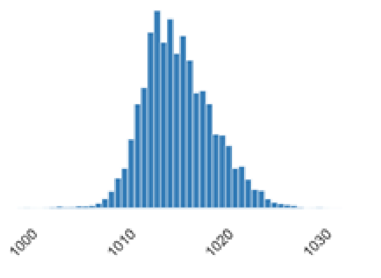


Figure 3.2 Overview of Variables in weather data- Pandas Profiling

The 'wdir' feature, representing wind direction, contain numerous missing values. As it can be observed from the *wdir* and *wspd* columns of the sample data in Table 3.8, this was primarily attributed to instances where wind speed registered as zero, leading to the absence of a directional indicator.

The dataset was arranged chronologically by sorting entries based on the datetime column. Next Identified and eliminated duplicate entries to ensure data integrity, retaining only unique timestamps.

3.5 Data Preprocessing

In the realm of renewable energy forecasting, the journey from raw data to accurate predictions is paved through meticulous data preparation and preprocessing. This transformative process involves shaping, refining, and augmenting datasets to extract meaningful insights, enabling robust models and informed decision-making. From integrating diverse sources like PV generators and weather data to handling outliers, missing values, and scaling features, each step in the data preparation is a crucial stride toward ensuring the quality and reliability of the information used for forecasting. This introduction sets the stage for a detailed exploration of the steps undertaken to harness the full potential of the data and lay the groundwork for precise time series forecasting.

Continued refinement of the PV generation data and Weather data will be detailed separately in the subsequent chapters.

4 Weather Data Preprocessing

After initial EDA on weather data, the data is further preproceesed as below.

4.1 Visualize data

An examination of the data plots was conducted to get an idea on the data and determine appropriate methods for filling in missing values and detect obvious outliers.

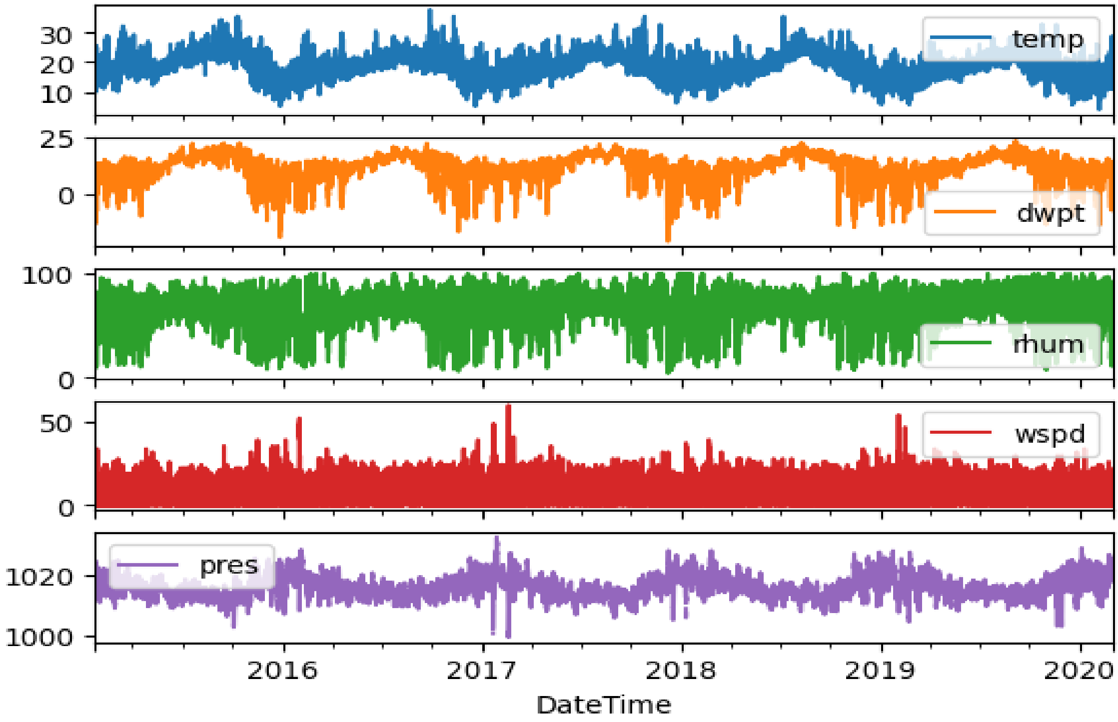


Figure 4.1 Line plot of Weather Data over time

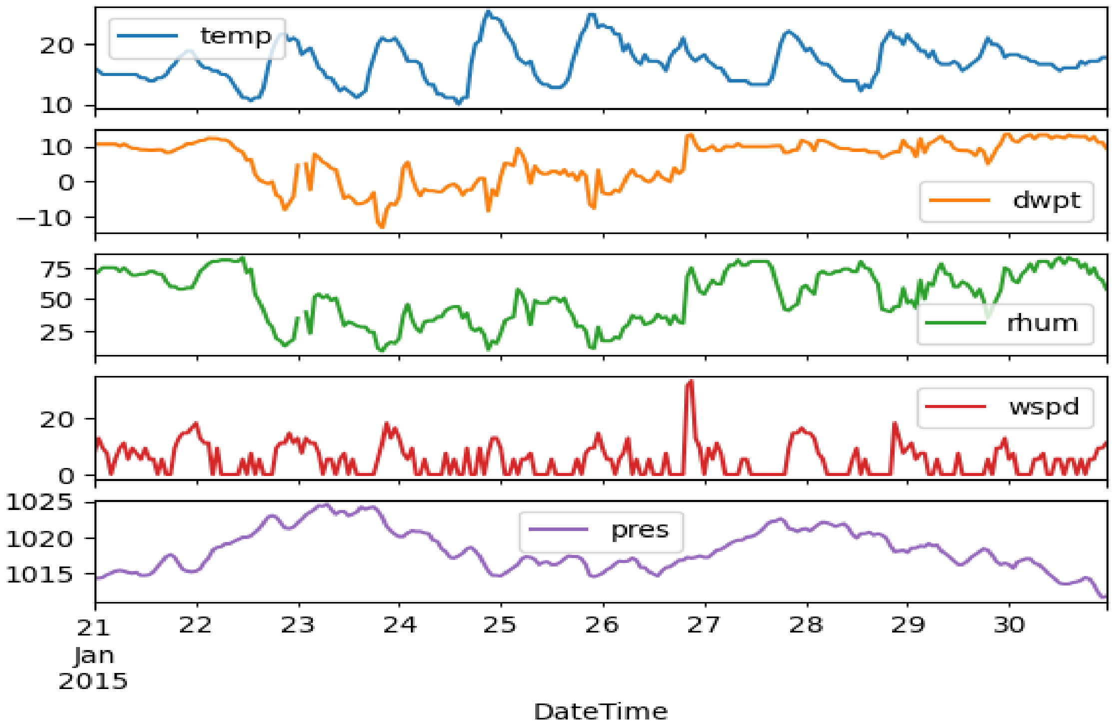


Figure 4.2 Sample of last 10 days of Weather Data over time

The data variations of temperature, relative humidity, wind speed, pressure, and dewpoint were observed over time. The two plots, displaying all data and sample data, clearly depict the highly random and fluctuating nature of the weather data.

Upon careful observation, it is apparent that the peak points within the data do not stand out as obvious outliers. This is primarily due to the simultaneous occurrence of peaks in other features at those specific points. The interrelated nature of these peaks suggests a more nuanced relationship, indicating that the presence of peaks may be a characteristic feature rather than an anomaly within the dataset.

4.2 Missing Data Handling

- *temp 1*
- *dwpt 4*
- *rhum 4*
- *wdir 4496*
- *wspd 2*
- *pres 464*

Upon examining the missing value count, it is evident that *wdir* (wind direction) has a significant amount of missing data. This observation aligns with previous findings in section, *3.4.1 Feature selection* where missing data was identified due to zero wind speed values.

To address the missing values in *wdir*, a potential solution is to incorporate the wind vector as a feature in subsequent analyses. Additionally, for cases where wind speed is zero, setting the corresponding wind direction missing values to zero can be considered as a suitable approach.

To further fill in the remaining missing weather data values, linear interpolation method is proposed. Given the hourly nature of the data, it is reasonable to assume that weather conditions during a 3-hour interval are likely to be similar to the preceding and succeeding hours. Therefore, employing interpolation for the missing values aligns with this temporal pattern.

4.3 Outlier detection and handling

4.3.1 Outlier Detection Methods for Weather Data

Several methods can be considered for detecting outliers in weather data, especially in the context of time series regression models.

- **Z-Score or Standard Score:** Identify points beyond a threshold assuming a normal distribution.

- **Modified Z-Score:** Robust alternative using median and median absolute deviation.

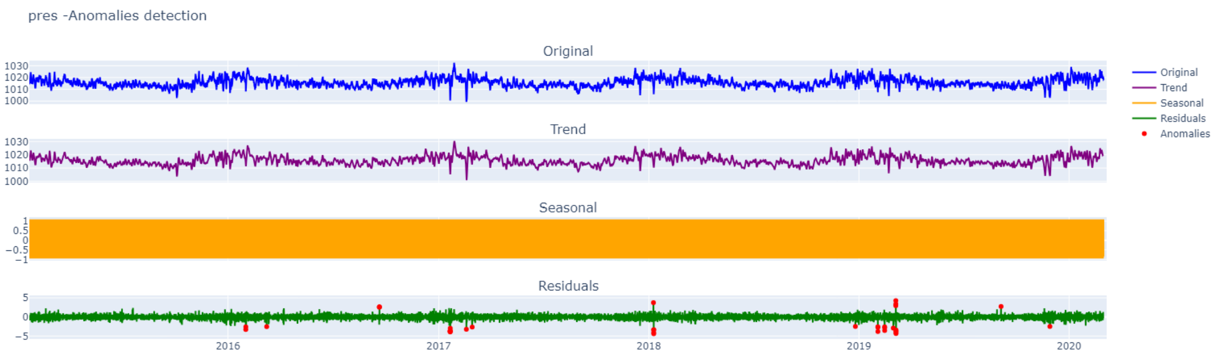
- **IQR (Interquartile Range) Method:** Define a range based on interquartile range, less sensitive to extreme values.
- **Moving Averages:** Smooth time series and spot significant deviations.
- **Seasonal Decomposition:** Detect outliers in residual component after decomposition.
- **Machine Learning Models:** Identify deviations from predicted values.
- **Visual Inspection:** Visually inspect data for unusual patterns.
- **Consistency Checks:** Ensure weather parameters fall within reasonable ranges.

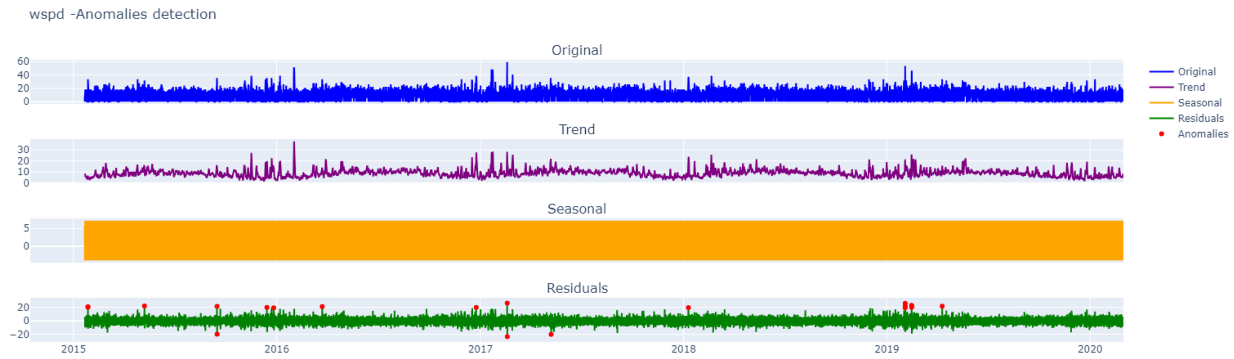
The choice depends on data characteristics and detection goals, with potential enhancement through combining methods or leveraging domain-specific knowledge.

4.3.2 Time series decomposition

Time series decomposition is a pivotal phase in unraveling the fundamental components of a dataset. The utilization of Seasonal-Trend decomposition using LOESS (STL) emerges as a powerful method, dissecting a time series into its core elements—trend, seasonality, and residuals. Anomalies, representing deviations from anticipated patterns, are often discerned in the residual component.

In the STL method, a straightforward threshold has been established at 5.0 times the standard deviation of the residuals and aligns with a daily seasonality assumption, allowing effectively identify anomalies. Obtained plots are shown in Figure 4.3 below.





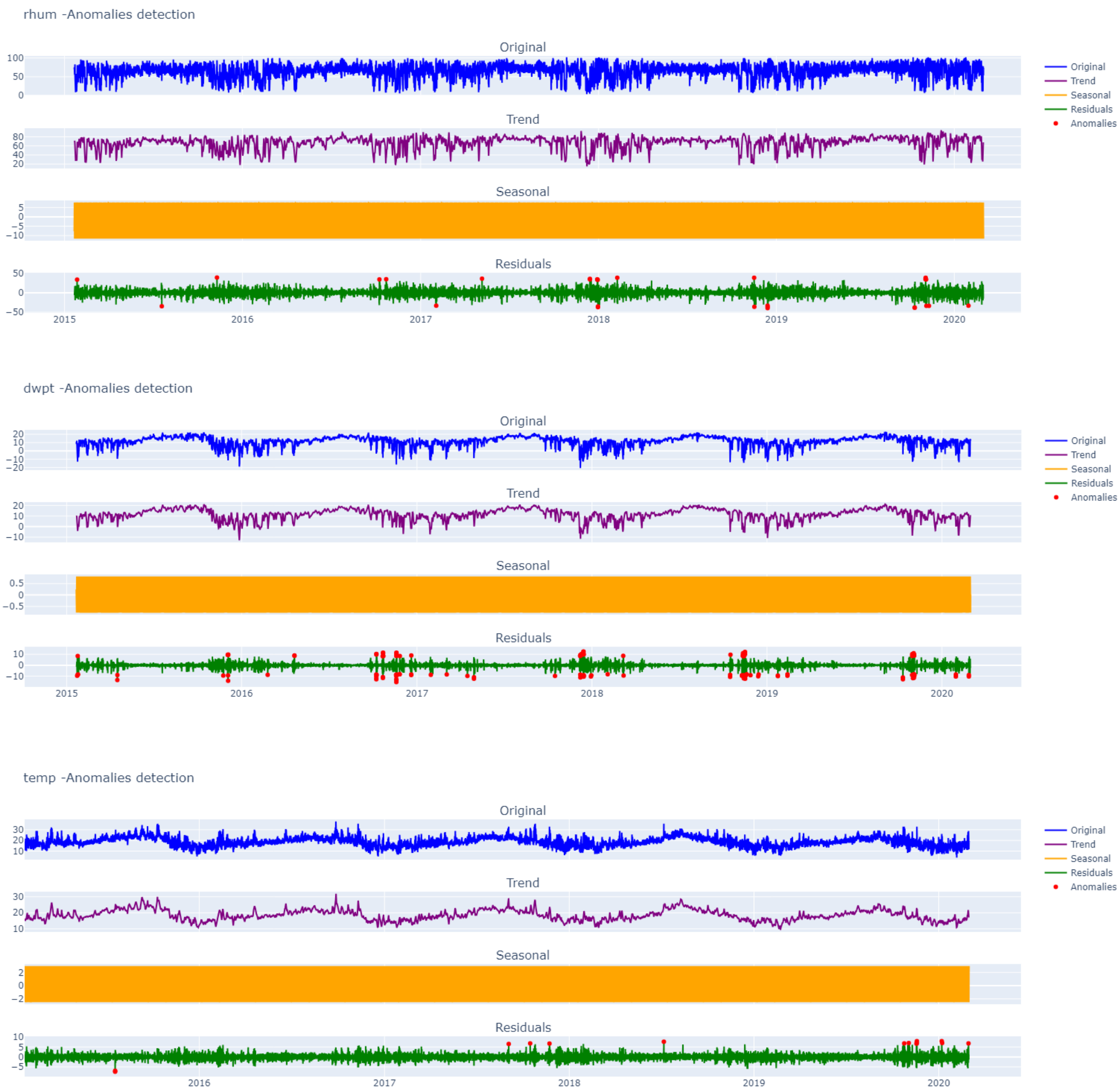


Figure 4.3 Seasonal decomposition Plots of Weather data

To facilitate a closer examination of the outlier-marked regions, the plots have been zoomed in. This allows for a more detailed visual inspection, aiding in the thorough analysis of these specific areas.

4.3.3 Manual Examination with Actual Data Plots

Upon reviewing residuals following STL decomposition, points identified as possible outliers (highlighted in red) at 5 times the standard deviation aligned with actual data upon visual inspection. Detected outliers were addressed by employing the interpolation method, effectively replacing these anomalous data points. This approach ensures a smoother transition and maintains the overall coherence of the weather data.

4.4 Feature Engineering

4.4.1 Wind

wdir (deg)—gives the wind direction in units of degrees. Angles do not make good model inputs: 360° and 0° should be close to each other and wrap around smoothly. And also the direction shouldn't matter if the wind is not blowing.Figure 4.4 shows the initial data distribution along with wind speed and wind direction.

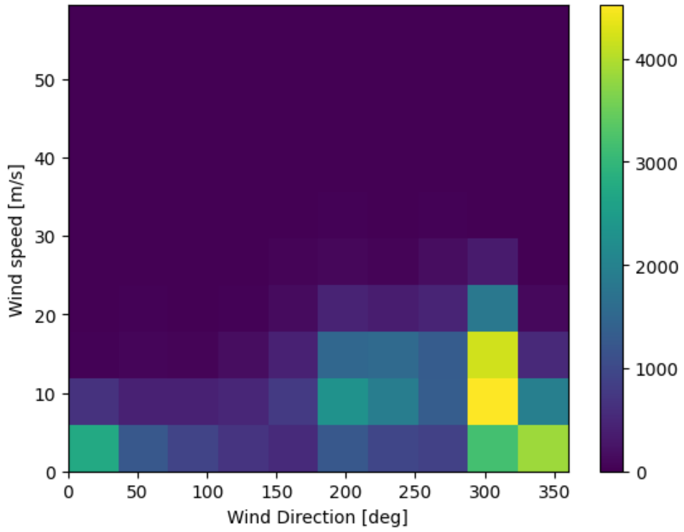


Figure 4.4 Heat Map of wind speed with wind direction

Recognizing the challenges posed by traditional angular measurements, a more interpretable approach involves converting wind direction and wind speed columns into a wind vector. This transformation aligns with the concept that 360° is equivalent to 0° in circular measurements. By representing the data in terms of wind vectors, the model gains a simpler and more accurate interpretation of the distribution. This not only enhances the model's understanding but also contributes to improved performance in capturing the nuances of wind-related features.

The distribution of wind vectors are shown in figure 4.5 below.

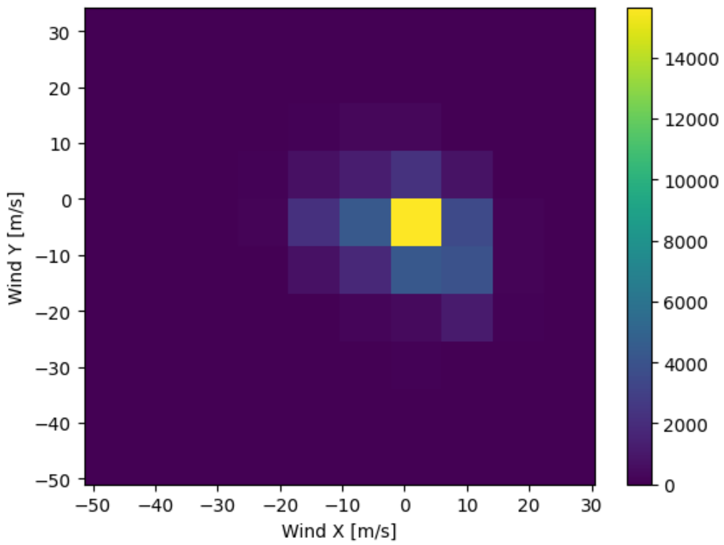


Figure 4.5 Distribution of wind vectors Heat Map

4.4.2 Time

The Date Time column, while valuable, is not optimally represented in its string form. To enhance its utility, the first step involves converting it into seconds. However, treating time in seconds directly may not offer meaningful insights for the model. Given the inherent daily and yearly periodicity in weather data, addressing this periodic nature is crucial.

To extract usable signals, a strategic approach involves employing sine and cosine transforms. This technique effectively clears the "Time of day" and "Time of year" signals, providing the model with more interpretable and relevant temporal features for improved accuracy in capturing time-related patterns.Figure 4.6 and Figure 4.7 shows the day and year signals obtained.

```
day = 24*60*60 seconds
year = (365.2425)*day seconds
```

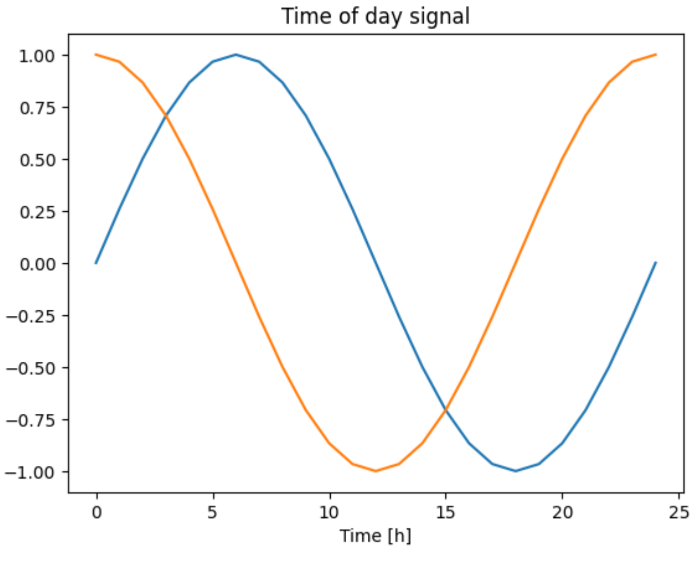


Figure 4.6 Day singnal

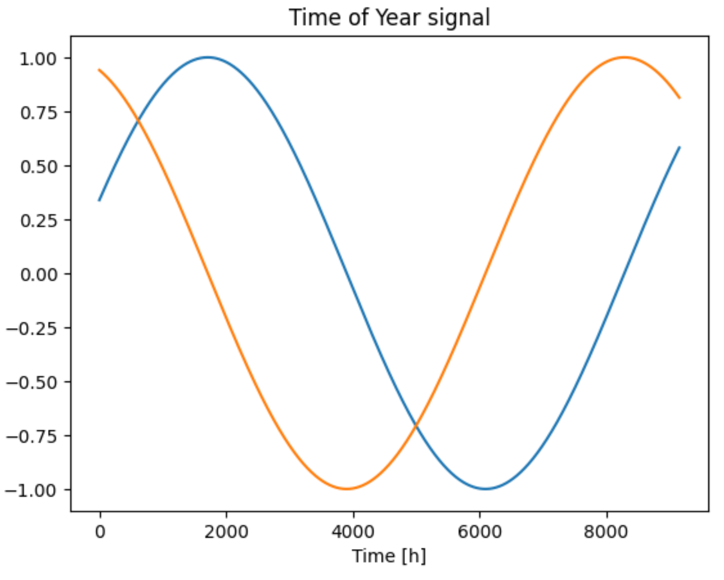


Figure 4.7 Year Signal

By transforming time-related data using sine and cosine transforms, the model gains access to crucial frequency features, offering valuable insights into periodic patterns. In cases where the relevant frequencies are known in advance, this approach proves particularly effective.

For scenarios where the frequencies are unknown, a robust method involves utilizing the Fast Fourier Transform (FFT) to extract important frequency features. This technique allows the model to identify and leverage key frequencies that contribute significantly to the underlying patterns in the data. A practical demonstration of this is evident in the tf.signal.rfft of temperature over time, revealing prominent peaks near frequencies corresponding to 1/year and 1/day.

Figure 4.8 shows theFFT for dewpoint. Similarly the plots can be obtained for other features and check for important frequencies for all the features.

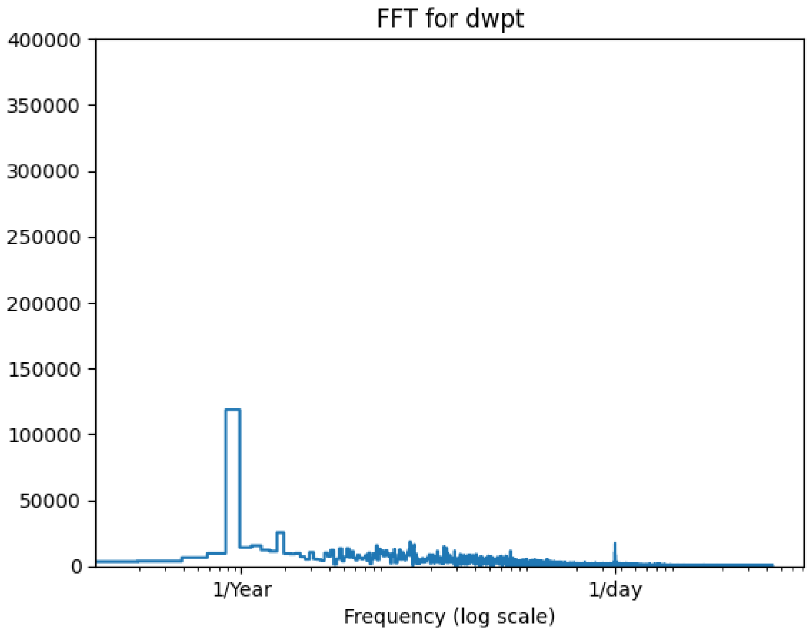


Figure 4.8 FFT of dwpt

It can be obsereved that yearly and daily frequency is signifinant.

Following the Feature Engineering the new features are as follows,

temp	dwpt	rhum	wspd	pres	Wx	Wy	Day sin	Day cos	Year sin	Year cos

4.5 Split Data

The dataset has been divided into three parts: training, validation, and test sets, with a distribution of (70%, 20%, 10%) respectively. This split allows the model to learn from the majority of the data during training, validate its performance on a separate portion, and finally, test its capabilities on an unseen data set.

Following the splitting of the data, the shape of the datasets is as follows,

```
no of features: 11
train data shape: (31348, 11)
val data shape: (8957, 11)
test data shape: (4479, 11)
target feature: (wspd)
```

4.6 Normalizing data

To ensure consistency and prevent data leakage, the dataset, excluding the target feature, has been normalized using the standardization method. It's important to note that the mean and standard deviation for this normalization are computed exclusively using the training data. This precautionary measure ensures that the models are unaware of the values in the validation and test sets, maintaining the integrity of the training process.

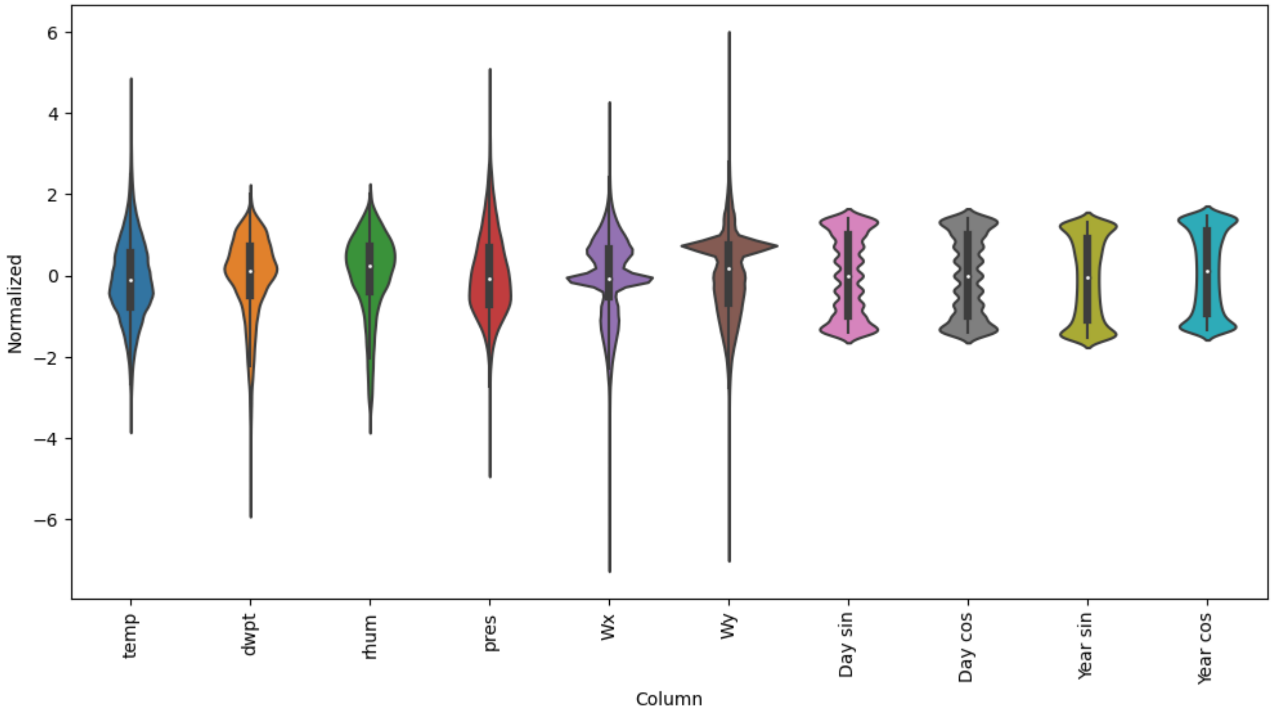


Figure 4.9 Violin plot of the standardized data

This violion plot shows the standardized data distibution.The tails normally represents outliers but they are not actual outliers in here as already discussed in outlier detection section.

5 PV Generation Data Preprocessing

From the EDA on the dataset it was observed negative values and missing values in the 10 PV Generators. Also it was noted that starting dates were different in the datasets

5.1 Handling Negative Values

Upon reviewing negative data values along with their corresponding datetime entries, it became evident that a potential criterion for replacement could involve substituting negative values with their respective magnitudes.

5.2 Visualize Data

An examination of the data plots was conducted to determine an appropriate method for filling in missing values and detect obvious outliers.

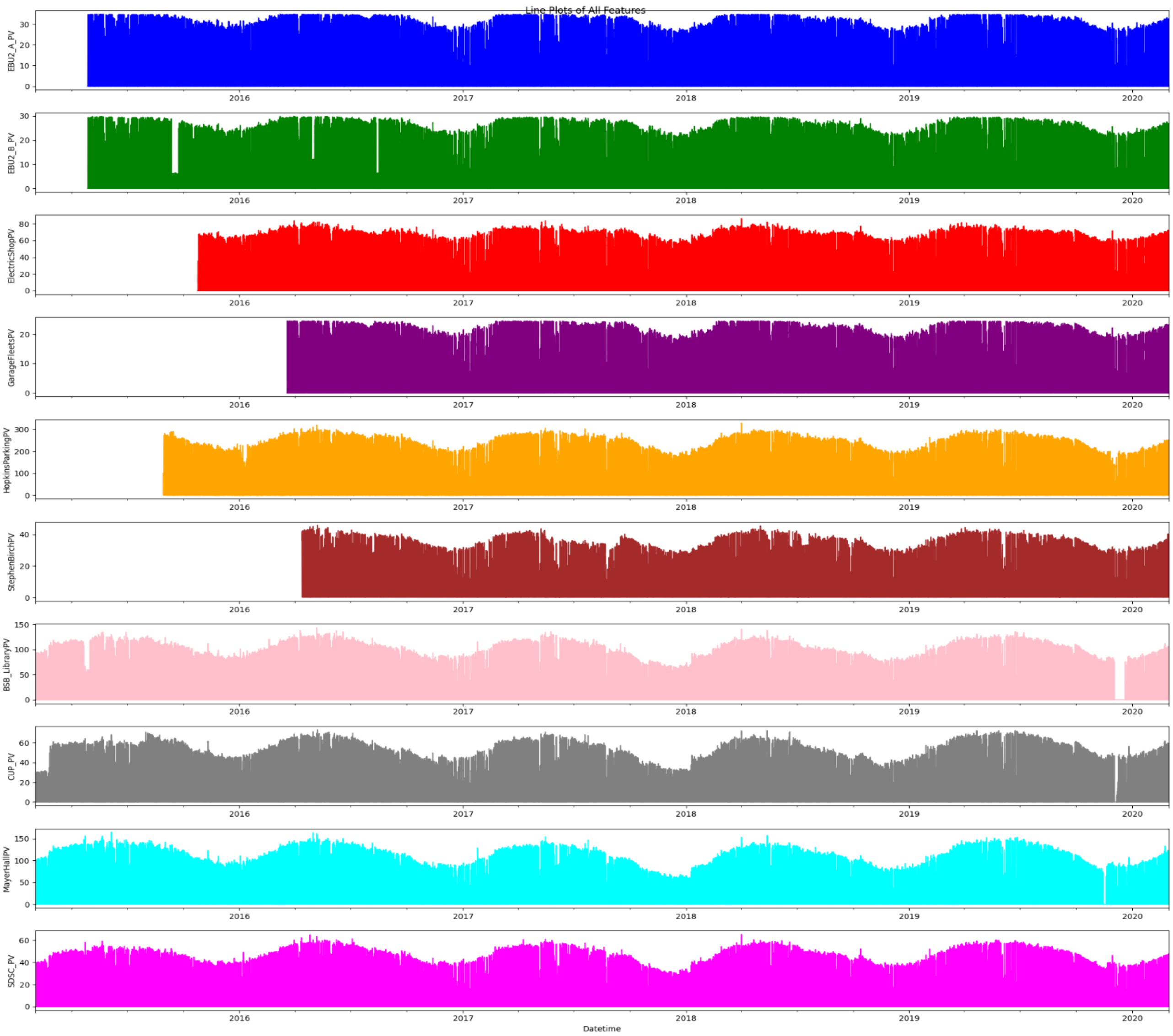


Figure 5.1 Line plots of PV Generation data over time

Observations from the line plot reveal consistent patterns across all plots over the years. Although the starting dates vary, most of them commence from mid-year 2016. Notably, there are differences in maximum generations, even though the overall patterns remain similar. This divergence can be attributed to varying maximum power generation capacities. Normalizing the data by considering the maximum power rating and utilizing the rated power is deemed a suitable approach. Additionally, reviewing sample data will contribute to a clearer visualization of the dataset.

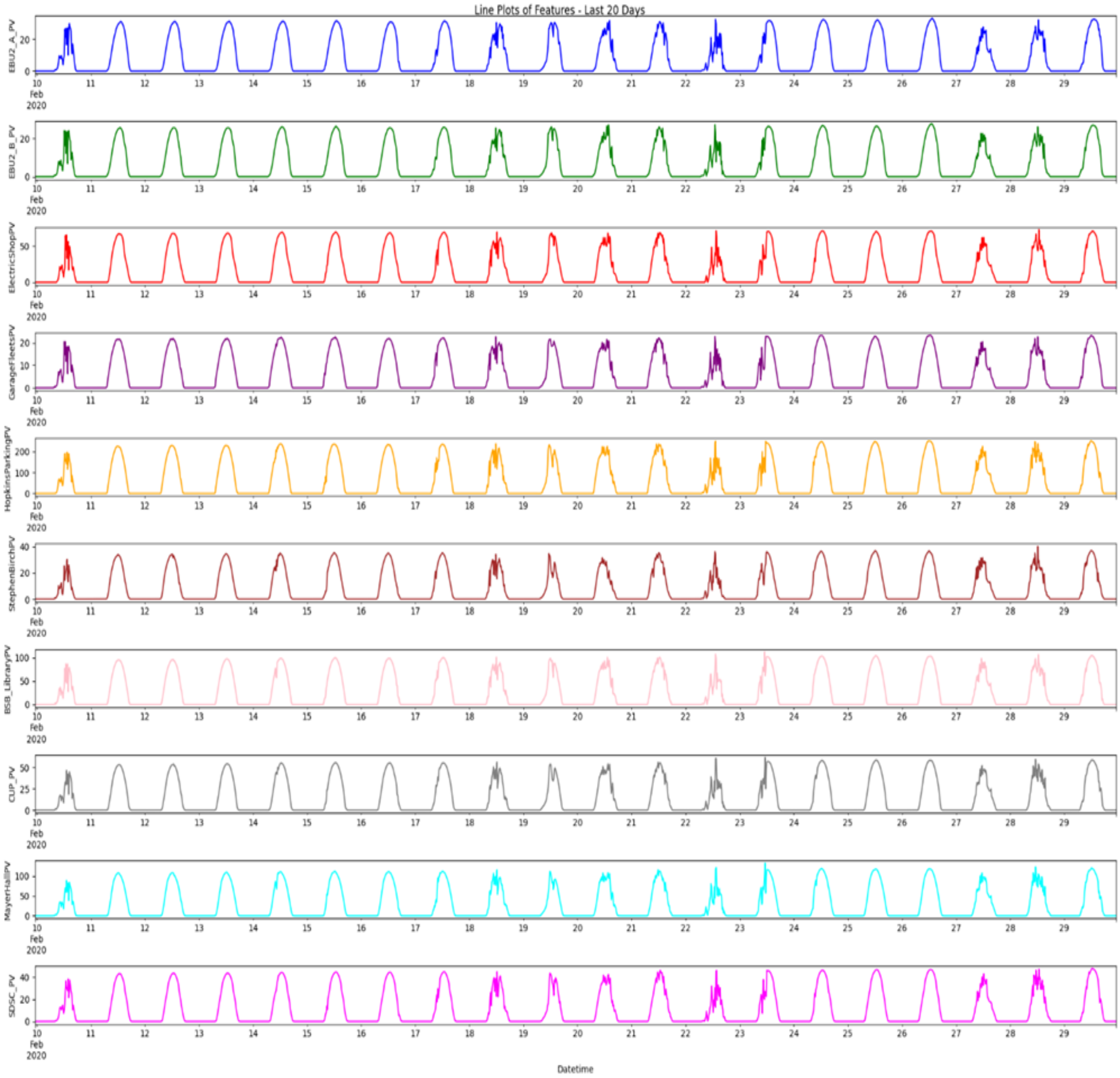


Figure 5.2 Sample data of Last 20 days

The last 20 days data of the pv generators looks almost same eventhough the maximum power generation is different.The plot itself shows that if the data is mapped to a same range the data will be nearly same.

This observation underscores the importance of normalizing the data. Normalization, by considering a consistent range or utilizing appropriate scaling techniques, becomes a prudent approach. It ensures that the inherent patterns and trends in the data are more effectively captured and comparable across different generators, contributing to a more unified and meaningful analysis.

5.3 Outlier Detection

5.3.1 Box Plot

A boxplot, or box-and-whisker plot, is a visual representation that provides a concise summary of the distribution of a dataset. The rectangular box in the plot spans the interquartile range (IQR), encapsulating the middle 50% of the data, with a line inside indicating the median. The whiskers extend from the box, denoting the range of the data within a certain multiple of the IQR. Any data points lying beyond the whiskers are considered potential outliers and are plotted individually. Boxplots offer a clear visualization of the data's variability, skewness, and the presence of extreme values, aiding researchers and analysts in making informed interpretations of the dataset's characteristics.

Box plot for PV data is shown in Figure 4.3 below.

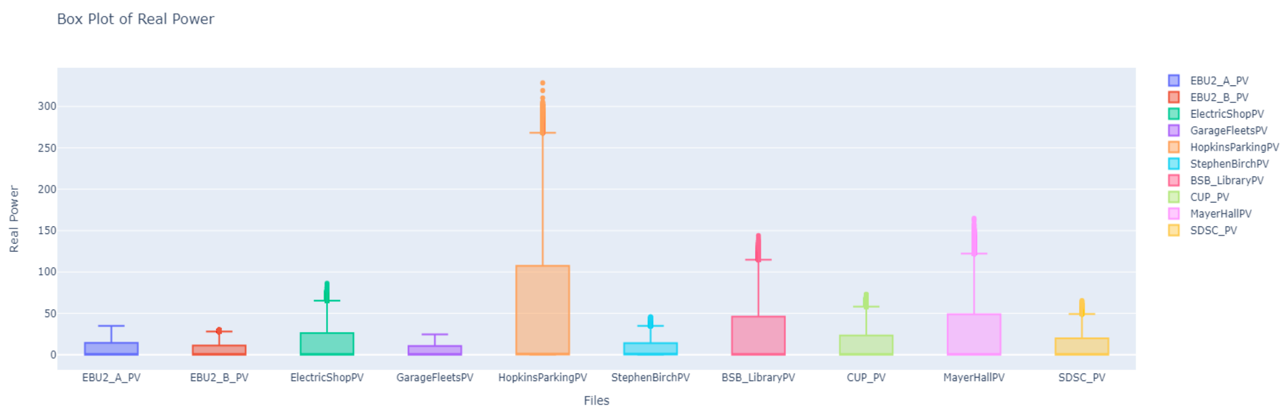


Figure 5.3 Box plot for PV data

In the analysis of the PV generation data, it was observed that the tails of the plot showcase the maximum peak PV generations over the years. It is important to note that these points, although appearing as outliers, are not actual anomalies but rather represent the extreme values of PV generation. To address this, a more refined visualization approach is necessary for effective outlier detection. This will ensure that the identification of outliers is accurate and meaningful, avoiding misinterpretation of the extreme values as anomalies. The upcoming sections will explore and implement improved visualization techniques to enhance the outlier detection process in the PV generation dataset.

5.3.2 Time Series Decomposition

In the STL method, a straightforward threshold has been established at 5.0 times the standard deviation of the residuals and aligns with a daily seasonality assumption, allowing effectively identify anomalies. Obtained plot for *'CUP_PV'* PV generator is shown in Figure 4.4 below.

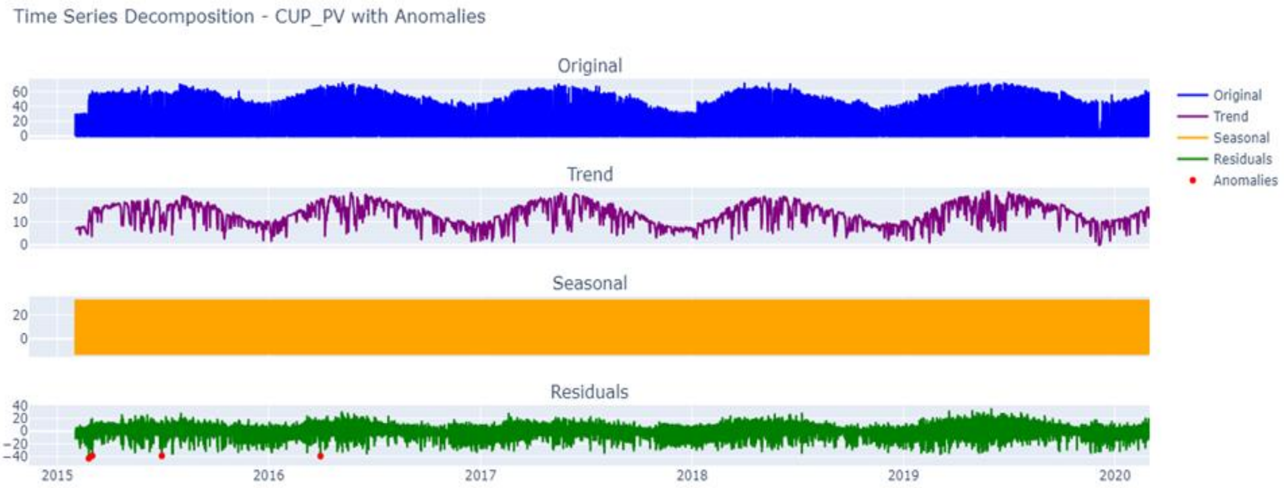
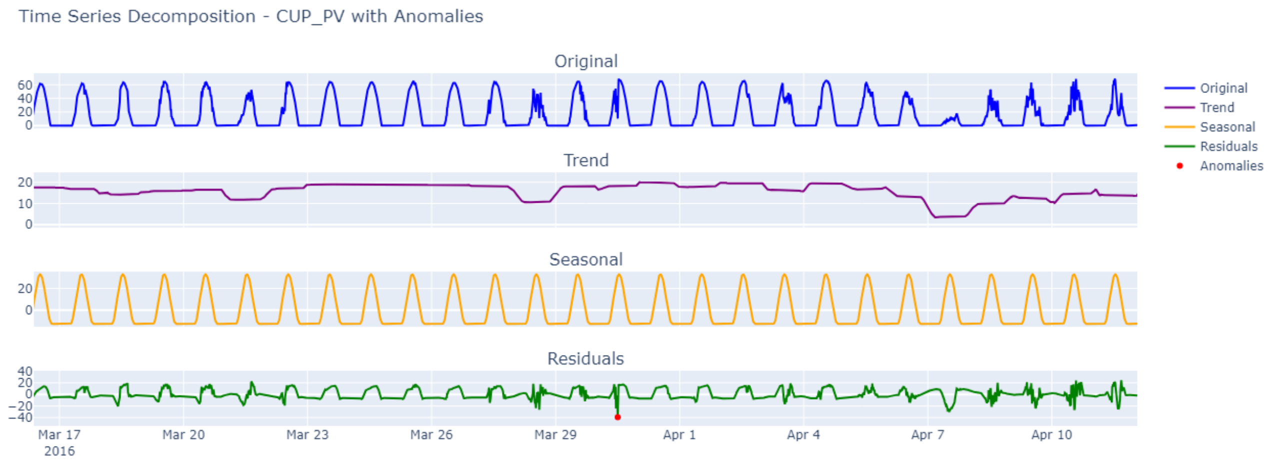


Figure 5.4 Seasonal Decomposition plots for *'CUP_PV'*

Zoomed-out perspectives are illustrated in the figures below, providing a broader view of the anomaly pointers in the datasets.



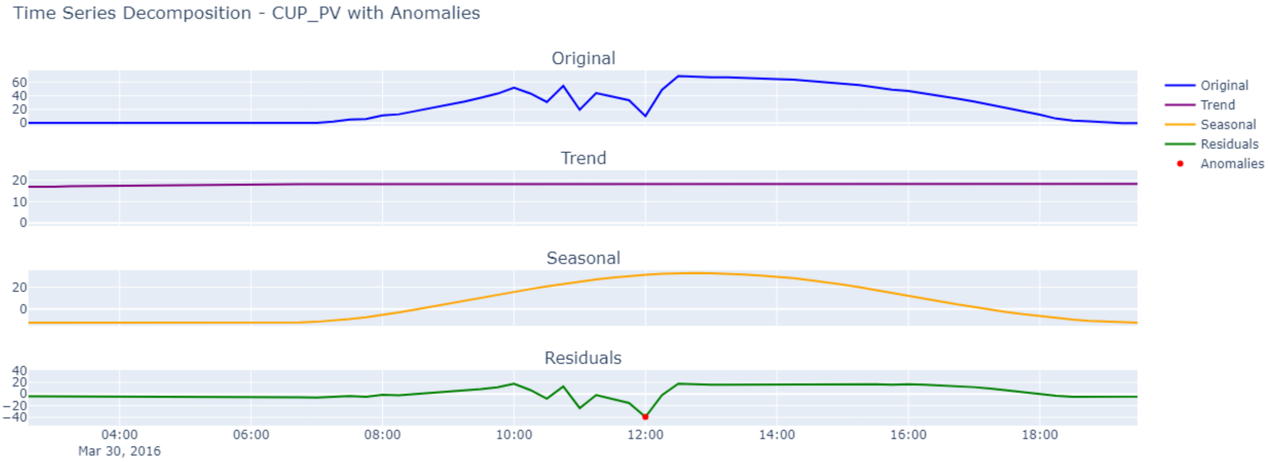


Figure 5.5 Zoomed Seasonal Decomposition plots for *'CUP_PV'*

Similarly, STL plots for other PV generators were obtained and reviewed for the detection of obvious outliers.

5.3.3 Manual Examination with Actual Data Plots

Upon reviewing residuals post STL decomposition, points identified as possible outliers (highlighted in red) at 5 times the standard deviation do not align with actual data upon visual inspection. These outliers exhibit similarities with neighboring values, challenging their classification as significant anomalies. Examining actual data values reveals that fluctuations in PV generation energy are normal, making it difficult to easily identify anomalies as outliers, as it might lead to replacing actual values with incorrect ones.

5.4 Normalizing data

The observed variance in mean and standard deviation among the 10 PV generators is linked to their unique power ratings, indicating the maximum power each generator can produce. As power ratings differ, variations in mean and standard deviation naturally occur. Recognizing this diversity in power capacity emphasizes the importance of considering individual characteristics during data analysis and interpretation.

Normalized the data to a 0 to 1 range using the Min-Max Scaler for consistent and comparable scaling, ensuring accurate assessments that consider the inherent differences in power capacity among the generators.

5.5 Visualize normalized data

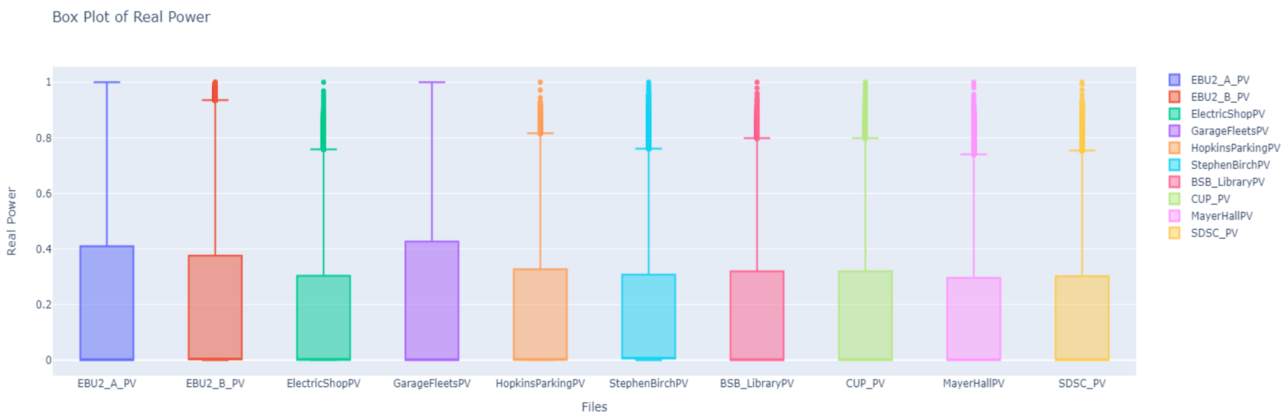


Figure 5.6 Box Plot of Normalized PV data

Following the data normalization, distinct variations in box plot distributions were observed as shown in Figure 4.6, particularly within the same geographical area. These differences can be attributed to variations in starting times for data collection. Additionally, environmental conditions, including the presence of shadows, exhibited variability across different stations. The distribution patterns in box plots were collectively influenced by these factors, emphasizing the importance of a thoughtful approach in selecting input data for forecasting models. The current alignment of data with the generalized weather information from San Diego Airport highlights the need to consider regional nuances for more accurate interpretations.

5.6 Generalized Input Data for Forecasting Model

Having taken the average of normalized PV generation values for each generator and datetime, the dataset has been enhanced. This refined data, coupled with the inclusion of general weather information for the city, aligns with the aim of creating a more comprehensive and broadly applicable forecasting model. This strategic method ensures better adaptability and reliability in the context of the city. The approach of averaging, where missing data is ignored, contributes to the improvement of the dataset.

5.7 Missing Data Handling

The approach of averaging data naturally addresses missing data concerns, as it involves ignoring them. This contributes to a more substantial improvement in the dataset's overall quality.

5.8 Outlier Handling

The process of averaging data inherently addresses outliers, providing a commendable approach for obtaining accurate and improved data. This strategic method significantly contributes to enhancing the overall quality of the dataset.

5.9 Resampling data

This resampling process operates under the assumption that significant changes within a one-hour interval are minimal. By averaging four consecutive data points over 15-minute intervals, it was aimed to capture a representative value for each hour while considering the relatively stable nature of the data during this timeframe. This simplification helps create a smoother and more manageable hourly dataset for analysis.

5.10 Data Integration

Merge the preprocessed weather dataset (without the standardization and splitting steps) with normalized PV generation data based on the timestamp. By integration of datasets, the model gains valuable insights into the dependencies and correlations between weather conditions and PV energy production, enhancing the accuracy and effectiveness of the forecasting model.

After merging a specific start time for the dataset has been designated considering PV generation data as,

```
start_datetime = '2016-01-01 00:00:00'
```

In the interest of refining the dataset for the specific modeling context, the 'wspd' column has been removed. This decision is based on the understanding that 'wspd' is not the target variable in this scenario, and the information related to wind speed is already incorporated within the 'wind vector' feature. Removing redundant or non-essential columns streamlines the dataset, focusing on the variables crucial for the forecasting model.

Data normalization part and the Data splitting part is similar to section 4.5 and 4.6 of weather data Preprocessing but the target variable here is the *MeanPower* data (Average normalized PV generation data).

6 Data Windowing and Forecasting Models

The criteria outlined below are applied to both PV generation forecasting and wind energy forecasting. For PV energy forecasting, the target variable is the Normalized Mean Power, while for wind energy forecasting, the target variable is Wind Speed. By examining the relationship between the target variable and energy, calculations for energy forecasting can be derived. This unified approach ensures consistency in the forecasting methodologies for both PV and wind energy.

6.1 Data Windowing

The objective is to predict 24 hours into the future based on the information from the preceding 24 hours. This multi-step prediction framework involves leveraging a sliding window of 24-hour intervals in the past data as input to forecast the subsequent 24 hours. By aligning the model's input and output windows, the training process enables the model to understand the temporal patterns and dependencies within the data, ultimately enhancing its ability to make accurate predictions for the specified time horizon. Time series forecasting is done using TensorFlow.

```
Input width= 24 time steps
Output width/labels = 24 time steps
```

Time series Data Frame is converted to a *tf.data.Dataset* of *(input_window, label_window)* pairs using the *tf.keras.utils.timeseries_dataset_from_array* function.

The models implemented in this study adopt a *single-output*, *multi-time-step* forecasting approach using *single-shot models*. In single output model the model will predict only one feature (target feature) whereas multiple output models will predict multiple features. In multi-step prediction, the model learns to predict a range of future values, differing from a single-step model where only a single future point is predicted. This distinction allows for the prediction of a sequence of future values.

6.2 Single shot models

One high-level approach to this problem is to use a "single-shot" model, where the model makes the entire sequence prediction in a single step. This can be implemented efficiently as a *tf.keras.layers.Dense.*

6.2.1 Baseline model

The baseline model serves as a fundamental reference point for comparison. Typically, it could involve simple statistical methods or straightforward algorithms.

A simple baseline for this task is to repeat the previous day, assuming tomorrow will be similar since this task is to predict 24 hours into the future using 24 hours of the past.

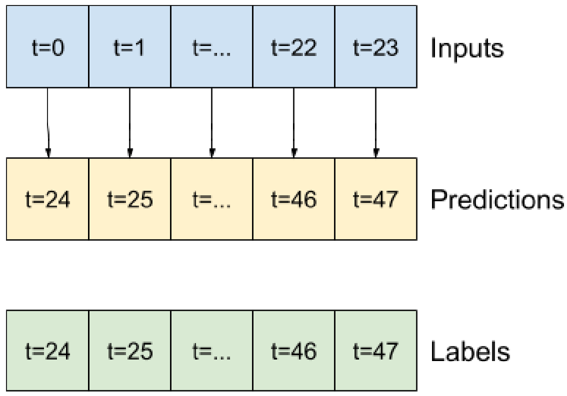


Figure 6.1 Window representaion for the baseline model

6.2.2 CNN model

CNN model is the selected Neural Network for the forecasting from the research gap. In the realm of time series forecasting, the Convolutional Neural Network (CNN) model stands out as a powerful architecture capable of extracting intricate temporal patterns from sequential data. Unlike traditional neural networks, CNNs are adept at capturing local dependencies through convolutional operations, making them particularly well-suited for tasks where the arrangement of input data holds significance.

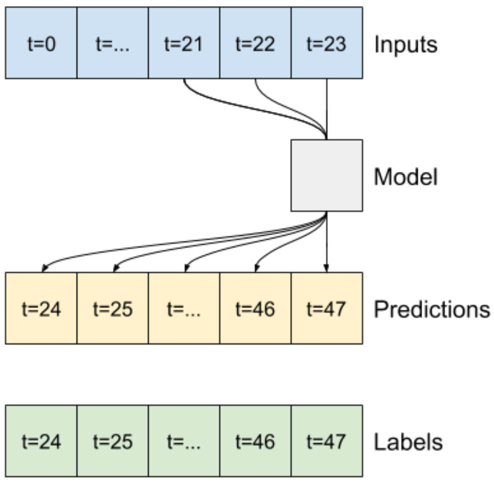


Figure 6.2 Window representaion for the CNN model

Figure 6.2 illustrates how the data windowing is used in CNN model in this task.

Model Architecture

The Convolutional Neural Network (CNN) architecture deployed for time series forecasting demonstrates a strategic design to effectively capture intricate temporal patterns. Comprising a sequence of essential layers, this model is adept at discerning and utilizing relevant information for accurate predictions.

Lambda Layer:

The initial lambda layer selects the specific time steps from the input data, focusing on recent and pertinent information for forecasting.

Convolutional Layer:

Utilizing a kernel size of the lamda layer input time steps and Rectified Linear Unit (ReLU) activation, the convolutional layer performs local operations on sequential data to extract complex temporal features.

Batch Normalization:

Following convolution, batch normalization stabilizes training by normalizing inputs, contributing to faster convergence and improved model performance.

Dropout Layer:

To prevent overfitting, a dropout layer with a suitable dropout rate is incorporated, selectively dropping connections during training for enhanced generalization.

Dense Layer:

The densely connected layer employs ReLU activation and He initialization to refine features extracted by previous layers, enhancing the model's understanding of temporal dependencies.

Reshape Layer:

Tailoring the output to match the desired prediction sequence length, the reshape layer ensures predictions align with the forecasting task requirements.

This CNN model is crafted to leverage convolutional operations and sequential data processing, demonstrating proficiency in making accurate predictions for energy generation forecasting based on historical data. The thoughtful design of each layer contributes to the model's ability to discern relevant temporal patterns, ultimately enhancing forecasting precision.

6.2.3 LSTM model

LSTM model is also implemented for performance comparison of the selected CNN model. The Long Short-Term Memory (LSTM) model is a specialized type of recurrent neural network (RNN) designed to address challenges related to learning long-term dependencies within sequential data. Unlike traditional RNNs, LSTMs are equipped with memory cells and gating mechanisms, enabling them to capture and store information over extended time intervals. This unique architecture makes LSTMs particularly effective in tasks where understanding temporal dependencies is crucial.

In the context of energy generation forecasting, the LSTM model proves valuable due to its ability to discern intricate patterns and trends in time-series data. By leveraging memory cells that selectively retain and propagate information, LSTMs excel at capturing the nuanced relationships inherent in energy generation sequences. The sequential nature of renewable energy data, influenced by factors such as weather patterns and time of day, aligns seamlessly with the strengths of LSTM models.

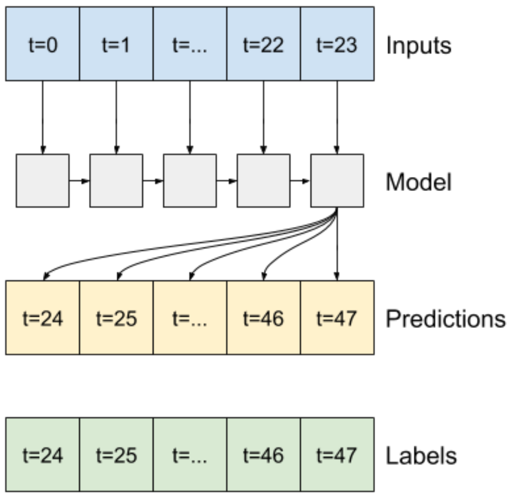


Figure 6.3Window representaion for the LSTM model

Figure 6.2 illustrates how the data windowing is used in LSTM model in this task.

Model Architecture

The architectural choices are tailored to optimize the model's ability to understand and forecast energy generation patterns. The key components of the LSTM model include:

LSTM Layer:

The LSTM layer with specific no of units is employed to learn and remember patterns across temporal sequences. The LSTM layer provides the final output at the last time step.

Dense Layer:

A dense layer is introduced with an initialization strategy that enhances learning. The layer outputs values aligned with the desired prediction sequence length, facilitating effective forecasting.

Reshape Layer:

Tailoring the output to meet the specific requirements of the forecasting task, the reshape layer ensures that predictions conform to the designated format.

The LSTM model is adept at capturing long-range dependencies, making it well-suited for forecasting tasks where temporal patterns play a crucial role. The intricate design of the architecture allows the model to discern complex patterns within sequential data, contributing to its effectiveness in the energy generation forecasting framework.

6.3 Training Procedure

The training of neural network models involves critical steps for effective model learning and performance evaluation. The following outline provides a high-level overview of the training procedure:

6.3.1 Model Compilation

- **Loss Function:** Mean Squared Error (MSE), a suitable choice for regression tasks, quantifies the average squared difference between predicted and actual values.
- **Optimizer:** The Adam optimizer is employed for efficient weight updates during training.
- **Metrics Monitoring:** The training process monitors metrics such as Mean Absolute Error, RMSE, R-squared, to evaluate model performance.

6.3.2 Custom Metrics

- **R-squared Metric:** In the context of time series regression models, R2 is used to assess how well the model fits the data.

6.3.3 Training Execution

- The model is trained using the **fit** method, utilizing the training data.
- The number of epochs is capped at a predefined maximum to control the duration of training.
- Validation data is employed to monitor the model's generalization to unseen data.
- Early stopping is implemented to halt training if no improvement in the validation loss is observed within a specified patience period.

This comprehensive training procedure is encapsulated in a function, facilitating consistent application across the neural network architectures. The training history, containing valuable insights into the model's learning progress, is obtained for further analysis and interpretation.

6.4 Relevant Equations for Neural Network Training

Table 6.1, illustrates a summary of the equations of the used metrics with their respective formula. In each formula, R represents the actual or real value while P represents the predicted or forecasted value. Also, n represents the number of observations.

Evaluation Metrics	Formulas
Mean Square Error (MSE)	MSE = $\left(\frac{1}{n}\right)\sum_{i=1}^{N}(R_i - P_i)^2$
Root Mean Square Error (RMSE)	RMSE = $\sqrt{MSE}$
Mean Absolute Error (MAE)	MAE = $\left(\frac{1}{n}\right)\sum_{i=1}^{n}\lvert R_i - P_i\rvert$
R2 value (Coefficient of Determination)	$\sqrt{}\,(1 - \frac{SS_res}{SS_tot} = 1 - \frac{\sum_{i=1}^{n}(P_i - R_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2})^2$

Table 6.1 Summary of evaluation metric equations

6.5 Hyperparameter Tunning

The Convolutional Neural Network (CNN) model underwent meticulous hyperparameter tuning to optimize its performance in energy generation forecasting. Key hyperparameters were fine-tuned to achieve the best results:

The Convolutional Neural Network (CNN) model used in this study undergoes careful hyperparameter tuning to optimize its performance in energy generation forecasting. Key hyperparameters include the convolution width, the number of convolutional filters, the kernel size, and dropout rate. The convolution width is critical for capturing relevant temporal patterns, and its adjustment influences the receptive field of the model. Additionally, the number of filters and kernel size in the convolutional layer determine the model's ability to extract features effectively. Fine-tuning these parameters enhances the CNN's capacity to discern intricate relationships within the data. Moreover, dropout regularization is applied to prevent overfitting, and its optimal rate is determined through experimentation, striking a balance between model complexity and generalization.

The Long Short-Term Memory (LSTM) model undergoes rigorous hyperparameter tuning to refine its ability to capture temporal dependencies for energy generation forecasting. The crucial parameters include the number of LSTM units. The number of LSTM units impacts the model's memory and its capability to retain information over time.

Optimizing the learning rate ensures efficient convergence during training, avoiding overshooting or slow convergence. Additionally, selecting an appropriate optimizer, such as Adam or RMSprop, contributes to the model's overall performance. By systematically adjusting these hyperparameters.

6.5.1 Hyperparameter Tuning for CNN Model

The Convolutional Neural Network (CNN) model underwent meticulous hyperparameter tuning to optimize its performance in energy generation forecasting. Key hyperparameters were fine-tuned to achieve the best results:

- **Convolution Width (CONV_WIDTH):** Set to 12 for capturing relevant temporal patterns.
- **Number of Convolutional Filters:** Adjusted to 256 to enhance feature extraction.
- **Kernel Size:** Set to match the convolution width (12) for effective pattern recognition.
- **Dropout Rate:** Tuned to 0.01 for regularization and preventing overfitting.

These hyperparameter values were carefully selected through iterative experimentation, ensuring the CNN model's optimal configuration for accurate and reliable energy generation predictions.

6.5.2 Hyperparameter Tuning for LSTM Model

The Long Short-Term Memory (LSTM) model underwent rigorous hyperparameter tuning to refine its ability to capture temporal dependencies for energy generation forecasting. The following hyperparameter values were determined through experimentation:

- **Number of LSTM Units:** Set to 128 for memory and effective information retention in Wind speed forecasting and 256 for PV energy forecasting

Values were systematically adjusted to enhance the LSTM model's ability to extract intricate patterns and achieve superior forecasting accuracy.

6.5.3 Hyperparameter tunning of training

Hyperparameter tuning is crucial for optimizing the neural network model's configuration. The selected hyperparameters and their values include:

- **Maximum Epochs:** Set to 50, determining the maximum number of training epochs before early stopping is triggered.
- **Early Stopping Patience:** A patience of 4, specifying the number of epochs with no improvement after which training will be stopped.
- **Learning Rate and Optimizer**: The Adam optimizer is used with its default learning rate.

These hyperparameters are chosen based on empirical observations and best practices. Adjusting these values may impact the model's convergence and generalization abilities, and careful consideration is required to achieve optimal performance. The provided values serve as a starting point, and further experimentation may be conducted for fine-tuning.

7 Optimization

The Optimization Methods Investigation chapter focused on exploring various optimization methods for microgrid operation, without actual implementation. The objective was to understand how these methods could potentially enhance the efficiency of the microgrid based on the predicted availability of renewable energy sources such as photovoltaic (PV) and wind power.

Various scenarios were identified, each with its own optimization objectives and constraints. Existing optimization methods and algorithms for microgrid energy management, such as linear programming, dynamic programming, heuristic methods, and stochastic methods, were reviewed. The most suitable optimization method for each scenario was selected based on the forecasted renewable energy generation data.

7.1 Microgrid Optimization

Microgrid optimization is a critical aspect in harnessing the full potential of self-contained energy systems, as highlighted in the provided information. Microgrids, designed for independent operation, are instrumental in reshaping the energy landscape by offering localized, sustainable, and resilient power solutions. The primary goals of microgrid optimization, as outlined, revolve around maximizing the integration of renewable energy sources, minimizing operational costs, and maintaining grid stability, even in the face of disruptions. The optimization strategies prioritize clean energy sources such as solar and wind, aiming to reduce reliance on fossil fuels and mitigate environmental impact.

However, the process of microgrid optimization is not without its challenges. The inherent complexities and uncertainties associated with accounting for unpredictable energy demand, variable weather conditions, and diverse technical constraints pose significant hurdles. To address these challenges, optimization methods need to be adaptable to different microgrid types, various energy scenarios, and evolving grid conditions. The continuous research and development mentioned in the source material become crucial in refining optimization techniques. Integration of advanced technologies, including artificial intelligence and machine learning, can further enhance the adaptability and efficiency of microgrid optimization strategies.

7.2 Microgrid operation scenarios

Four microgrid operation scenarios are represented below using different configurations of a microgrid, each with its own set of operational characteristics and challenges. The choice of scenario would depend on various factors such as the availability of renewable energy sources, the demand-supply balance, the need for grid stability, and the cost considerations.

Scenario A	The microgrid is not connected to the main grid but includes a Battery Energy Storage System (BESS) and a generator.
Scenario B	The microgrid is connected to the main grid but does not include a BESS or a generator.
Scenario C	The microgrid is connected to the main grid, includes a BESS, but does not have a generator.
Scenario D	The microgrid is connected to the main grid and includes both a BESS and a generator.

Table 7.1Microgrid operation scenarios

The summarized table is shown below in table7.1.

Scenario	**Connected to Main Grid**	**Includes BESS**	**Includes Generator**
A	No	Yes	Yes
B	Yes	No	No
C	Yes	Yes	No
D	Yes	Yes	Yes

Table 7.2 microgrid operation scenarios summarized

The investigated methods to optimize microgrid operation under each of these scenarios using forecasted renewable energy data is explained in next section.

7.3 Optimization methods and algorithms

Algorithms play a pivotal role in optimizing microgrid operations, particularly when aiming to reduce costs and emissions. Depending on the objectives, a microgrid can be optimized for either cost reduction, emission reduction, or a balance between both. For single-objective optimization, each objective is considered individually, while in multi-objective optimization, weights are assigned to different objectives based on their priority.

The choice of algorithm also depends on the accuracy of forecasting. In scenarios where energy forecasting is accurate, deterministic optimization algorithms are employed. These algorithms make decisions based on precise predictions of energy generation and consumption. On the other hand, when forecasting is uncertain due to the variability of renewable energy sources, stochastic optimization algorithms are more suitable. Stochastic algorithms consider the probability distribution of variables, accommodating the inherent unpredictability in renewable energy production.

Renewable energy sources exhibit time dependency and dynamic characteristics. As such, mathematical algorithms, particularly linear algorithms, can be employed to optimize microgrid operations over time. Linear programming, for instance, allows for the efficient allocation of resources to meet energy demands while considering various constraints.

Here’s a detailed explanation summary:

- **Objective**: The primary goal is to reduce cost and emission, or one of them at a time. This can be achieved through the use of different optimization algorithms.
- **Depending on Objective**:
  - **Single Objective**: If the focus is on one objective at a time, single objective optimization can be used.
  - **Multi Objective Optimization**: If multiple objectives need to be considered simultaneously, weights can be assigned accordingly in multi-objective optimization.
- **Depending on Accuracy**:
  - **Accurate Forecasting**: If the forecasting is accurate, deterministic optimization algorithms can be used. These algorithms provide a definite outcome based on the given input.
  - **Uncertain Forecasting**: If the forecasting is uncertain, stochastic optimization algorithms can be used. These algorithms take into account the randomness and uncertainty in the forecasting.
- **Renewable Energy Variation**:
  - **Time Dependency and Dynamic**: Renewable energy generation varies with time, so the algorithms need to account for this time dependency and dynamic nature.

- **Mathematical Algorithms**: Linear mathematical algorithms can be used to handle this time-dependent variation.

In summary, the choice of optimization algorithm in microgrid operations depends on the specific objectives, the accuracy of energy forecasting, and the dynamic nature of renewable energy sources. Whether employing deterministic or stochastic algorithms and considering single or multiple objectives, the goal is to achieve a balanced and efficient microgrid operation that minimizes costs, reduces emissions, and maximizes the utilization of renewable energy.

Although the optimization methods were not implemented, the investigation provided valuable insights into their potential effectiveness. It highlighted the significant role these methods could play in improving the performance of the microgrid, leading to more efficient use of renewable energy and potentially reducing reliance on non-renewable energy sources. However, it also underscored the challenges associated with optimization, such as the need for accurate forecasting of renewable energy availability and the complexity of the optimization algorithms. Despite these challenges, the research concluded that machine learning-based optimization methods hold great promise for the future of renewable energy management and sustainability. The findings from this research could serve as a valuable reference for future studies and applications in this field.

8 Results and Discussion

After Hyperparameter tunning of the forecasting models, the performances obtained were as follows.

8.1 Performances of Wind forecasting models

All the performances on all data sets (training, validation, test) on all 3 forecasting models are shown in the following sections.

8.1.1 Loss Comparison

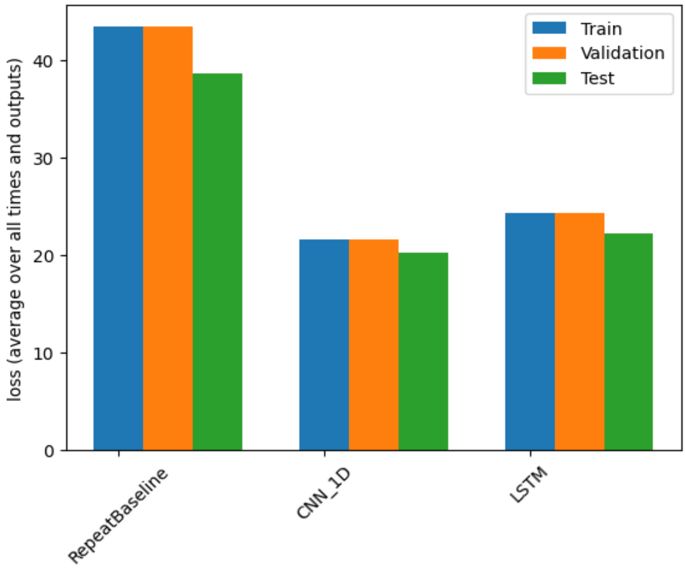


Figure 8.1 Losss Comparision of Wind Forecasting models on the data sets

There is a good loss performance of NN models on PV generation data compared to the baseline model after hyperparameter tunning. CNN model outperform other two models considering loss function. The unseen data perform better than train and validation data on NN models on wind data.

8.1.2 MAE Performance

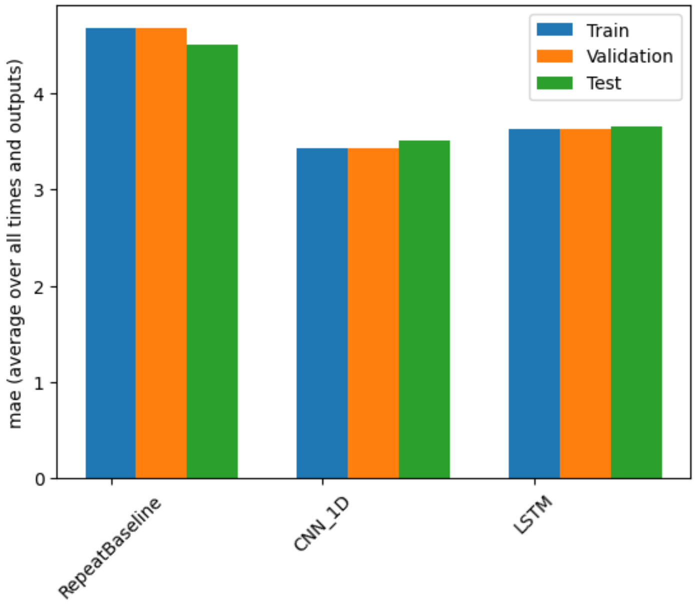


Figure 8.2MAE Performance of Wind Forecasting models on the data sets

Similarly, the MAE performance of NN models on Wind data is good compared to the baseline model after hyperparameter tunning. The unseen data perform better than train and validation data on NN models but not in baseline model. CNN model outperform other two models considering MAE.

8.1.3 RMSE Performance

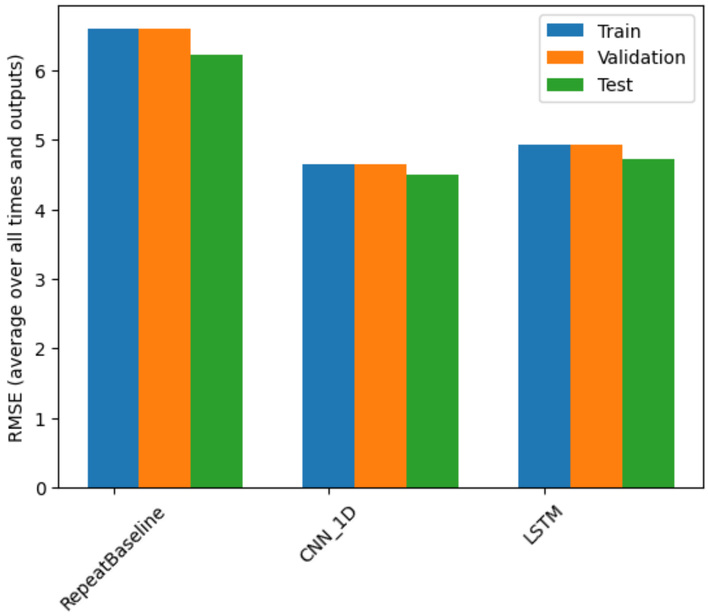


Figure 8.3RMSE Performance of Wind Forecasting models on the data sets

Similarly, the RMSE performance of NN models on Wind data is good compared to the baseline model after hyperparameter tunning. The unseen data perform better than train and validation data on NN models. CNN model outperform other two models considering RMSE.

8.1.4 Rsquare Performance

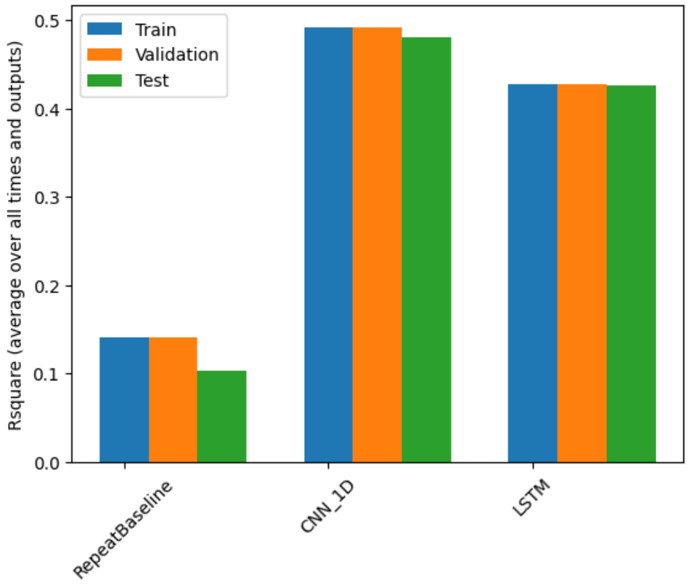


Figure 8.4Accuracy Performance of Wind Forecasting models on the data sets

R^2 performance of NN models on Wind data is very good compared to the baseline model after hyperparameter tunning. But the values lie around 0.4. CNN model outperform other two models considering R^2 value.

8.1.5 Discussion

The following Table 8.1 shows the performance values of all 3 wind forecasting models on the unseen test data set.

	MSE	MAE	RMSE	Rsquare
RepeatBaseline	38.63705	4.50533	6.21587	0.10405
CNN	20.21994	3.50545	4.496661	0.47996
LSTM	22.29595	3.65524	4.721859	0.42607

Table 8.1 Test data Performance Evaluation of Wind Forecasting Models

The CNN model can explain around 48% of the test data while LSTM model can explain around 42% on wind data indicates that these models are able to explain a portion of the variance in the test data.

The relatively low percentages suggest that there is a significant amount of unexplained variability in the data that these models are not capturing.However,It's important to note that R2 is just one metric, and it may not provide a complete picture of the model's performance. Looking at the other metrics the selected model performance is acceptable.

8.1.6 Conclusion

The complexity inherent in wind data, influenced by various factors such as weather patterns and geographical features, contributed to the challenges faced by the models. Difficulties arose from the observed noise and intricate patterns within the data, posing obstacles for traditional linear regression models. This necessitated the adoption of advanced neural network architectures capable of capturing non-linear relationships and spatial-temporal dynamics. The limitations of the models in explaining a significant proportion of the variance underscored the need for further refinement, potentially through feature engineering and the exploration of alternative model architectures. These findings prompt a deeper investigation into the factors influencing wind patterns and call for a holistic approach to model development, incorporating domain expertise and considering the nuanced dynamics of the data.

8.2 Performances of PV Energy forecasting models

8.2.1 Loss Performance

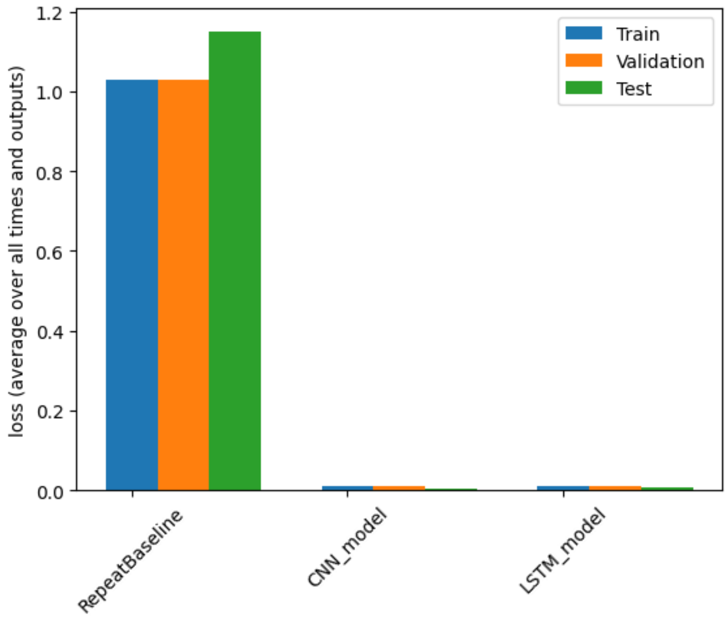


Figure 8.5 Losss Comparision of PV Energy Forecasting models on the data sets

There is an excellent loss performance of NN models on PV generation data compared to the baseline model after hyperparameter tunning. If it is observed carefully the unseen data perform better than train and validation data on NN models.

8.2.2 MAE Performance

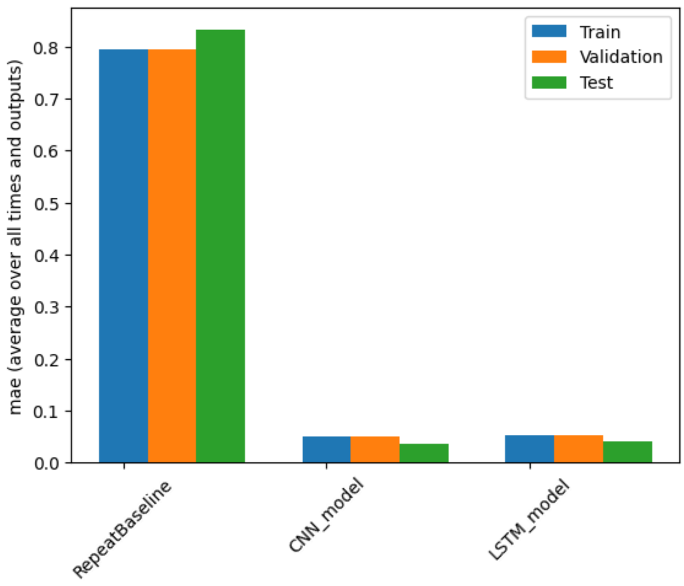


Figure 8.6 MAE Performance of PV Energy Forecasting models on the data sets

Similarly, the MAE performance of NN models on PV generation data is very good compared to the baseline model after hyperparameter tunning. On careful observation, it can be seen that the unseen data perform better than train and validation data on NN models.

8.2.3 RMSE Performance

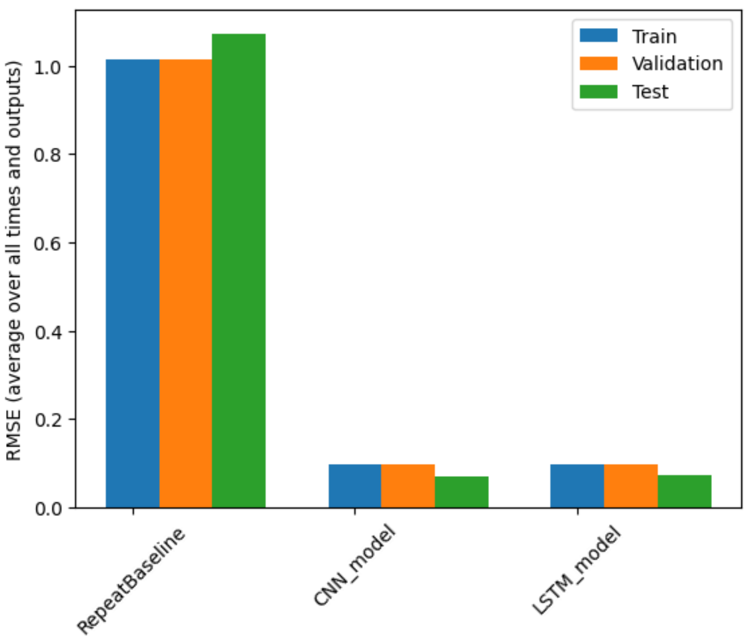


Figure 8.7 RMSE Performance of PV Energy Forecasting models on the data sets

Similarly, the RMSE performance of NN models on PV generation data is very good compared to the baseline model after hyperparameter tunning. The unseen data perform better than train and validation data on NN models.

8.2.4 Rsquare Performance

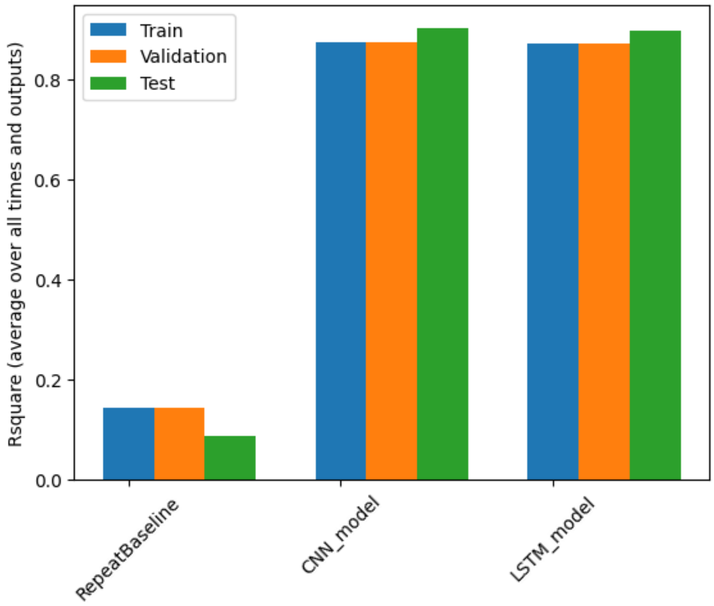


Figure 8.8 Accuracy Performance of PV Energy Forecasting models on the data sets

Similarly, the Rsquare performance of NN models on PV generation data is very good compared to the baseline model after hyperparameter tunning. The unseen data perform better than train and validation data on NN models.

8.2.5 Discussions and conclusions

The following Table 8.1 shows the performance values of all 3 PV Energy forecasting models on the unseen test data set.

	MSE	**MAE**	**RMSE**	**Rsquare**
RepeatBaseline	1.149171	0.83254	1.071994	0.08693
CNN	0.005156	0.03458	0.071806	0.9021
LSTM	0.005457	0.03938	0.073873	0.89728

Table 8.2 Test data Performance Evaluation of PV Energy Forecasting Models

R^2 value shows the accuracy of the time series regression models. The CNN model can explain around 90% of the test data while LSTM model can explain around 89% on PV generation data.

In the realm of time series regression modeling for renewable energy applications, the performance of different architectures, such as CNN and LSTM, was evaluated for predicting PV (photovoltaic) generation data. The R2 values revealed strong accuracy for both models, with the

CNN model exhibiting an impressive explanatory power of approximately 90%, outperforming the LSTM model by a slight margin. This outcome suggests that the CNN architecture effectively captures the intricate patterns and trends within the PV generation data. The findings underscore the versatility of neural network models in handling the complexities of time series data, with the CNN model standing out as particularly effective in this specific context. Such insights contribute to the ongoing exploration of optimal modeling techniques for accurate forecasting in renewable energy scenarios.

8.3 Check for Overfitting Underfitting of NN models

8.3.1 Overfitting

Overfitting is a common challenge in machine learning where a model becomes overly attuned to the nuances and details of the training data, including noise and outliers, at the expense of its ability to generalize to new, unseen data. One of the telltale signs of overfitting is a notable disparity in accuracy between the training and testing datasets. While the model may exhibit high accuracy on the training data, its performance tends to suffer when applied to new data. In the context of evaluating the R2 metric, if the R2 value is markedly higher on the training dataset compared to the testing dataset, it serves as a clear indicator of potential overfitting. This discrepancy underscores the model's struggle to extend its learned patterns beyond the training set, emphasizing the need for measures to enhance generalization.

8.3.2 Underfitting

Conversely, underfitting occurs when a model is too simplistic, failing to capture the intricate patterns within the data. This results in poor performance on both the training and testing datasets. Signs of underfitting manifest as low accuracy across both datasets, indicating the model's inadequacy in grasping the underlying relationships present in the data. Specifically, in the evaluation of R2 values, if these values are consistently low on both the training and testing datasets, it signals that the model lacks the necessary complexity to effectively represent the complexities within the data. Addressing underfitting may involve adjusting the model architecture, increasing complexity, or refining training strategies to ensure that the model can better capture the inherent patterns and nuances in the dataset.

Visualizing the accuracy and loss performances on both the training and validation datasets over epochs on wind forecasting as shown in Figure 8.9 and PV generation forecasting in Figure 8.10 of CNN models provides a valuable insights into the training progress and potential issues such as overfitting or underfitting. Below is a generalized interpretation of such plots:

8.3.3 Wind Forecasting CNN model

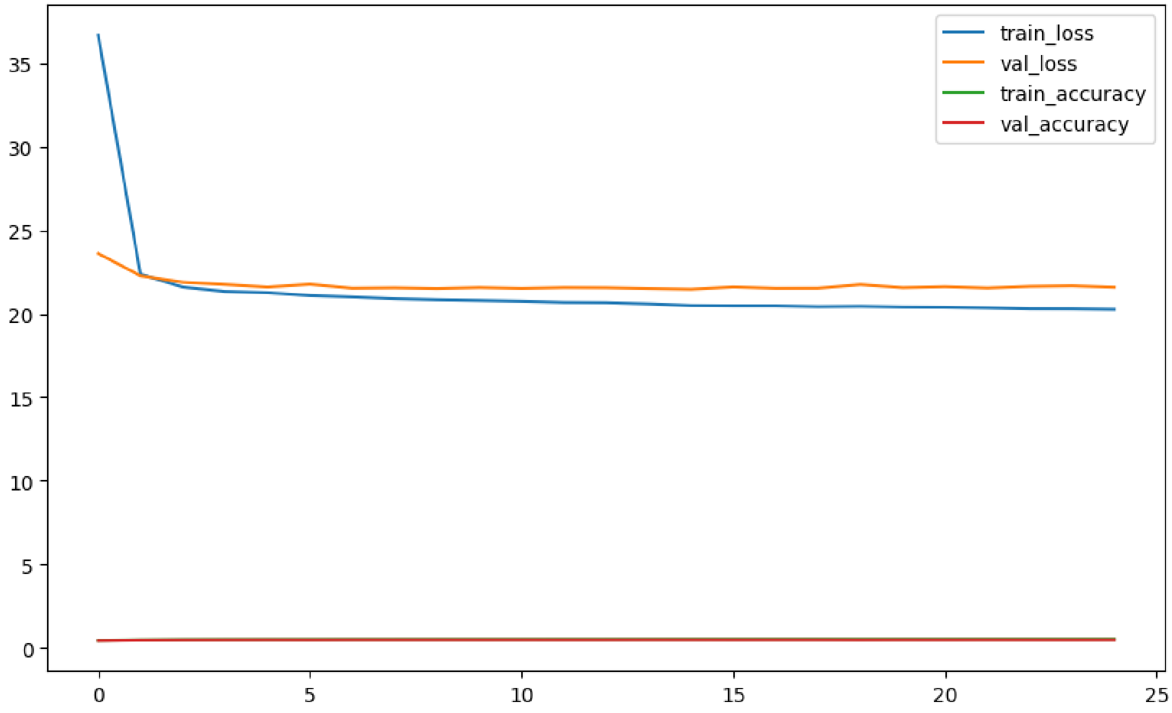


Figure 8.9 accuracy vs loss performances on train and validation data sets over epoches

8.3.4 PV Energy Forecasting CNN model

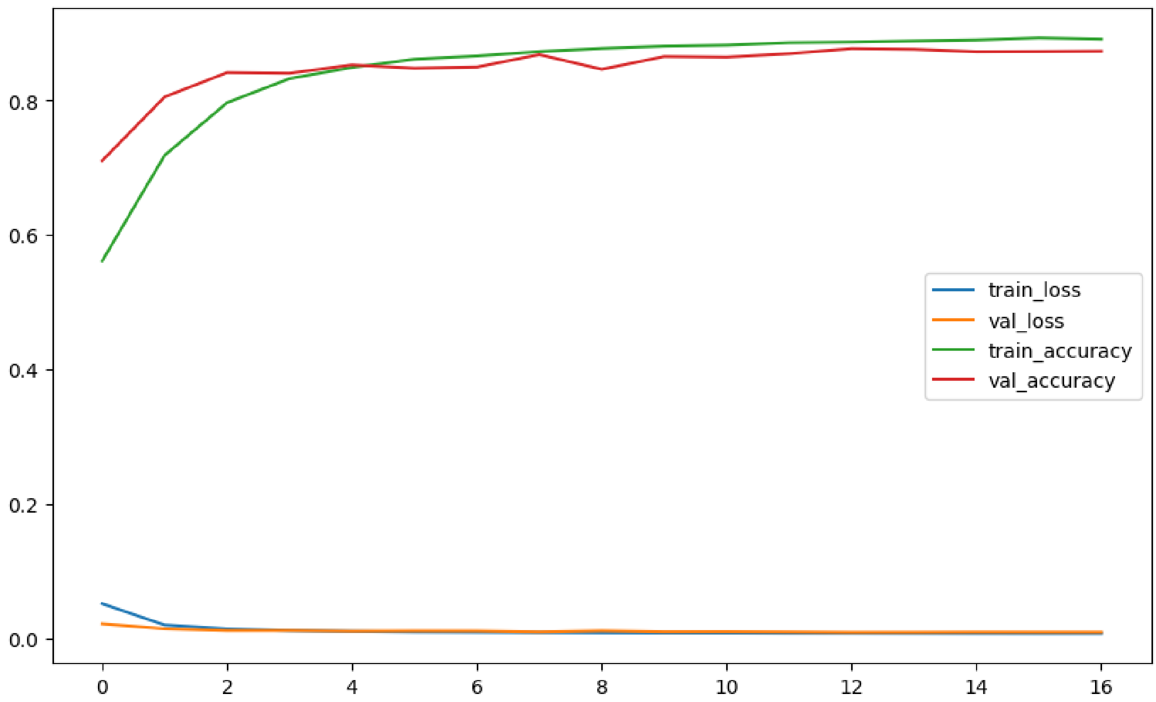


Figure 8.10accuracy vs loss performances on train and validation data sets over epoches

8.3.5 Discussion

In the plotted graphs, accuracy and loss curves for the training and validation datasets over epochs reveal trends in the model's learning process. The accuracy curves depict how well the model is performing, while the loss curves represent the extent of errors during training. Initially, both training and validation accuracy may increase, and losses may decrease as the model learns from the training data. However, it's crucial to monitor these trends as training progresses.

The observed trends in the accuracy and loss curves are indicative of a well-performing model that demonstrates effective learning and generalization. In the context of accuracy curves, witnessing a steady increase in both training and validation accuracy is a positive sign. It suggests that the model is successfully learning and adapting to the patterns present in the training data, and importantly, it is able to generalize well to unseen data, as reflected by the consistent rise in validation accuracy.

Simultaneously, the behavior of the loss curves adds further support to the model's proficiency. The consistent decrease in both training and validation loss over epochs implies that the model is becoming increasingly adept at minimizing errors during training. This reduction in loss signifies improved overall performance and a more accurate representation of the underlying patterns in the data. The alignment of decreasing loss in both training and validation sets indicates that the model is not overfitting to the training data but rather generalizing effectively to new, unseen data.

The observed patterns in both accuracy and loss curves in wind forecasting and PV generation Forecasting plots suggest a well-balanced learning process, showcasing the model's ability to learn from the training data while maintaining strong generalization to validation or testing datasets.

8.4 Visualize Forecasted Data

8.4.1 Actual vs Predicted Data

The following two figures show the simple visualization of the 3 split window samples of the test datasets in PV generation forecasting and Wind speed forecasting. The plots include inputs, labels and predictions clearly.

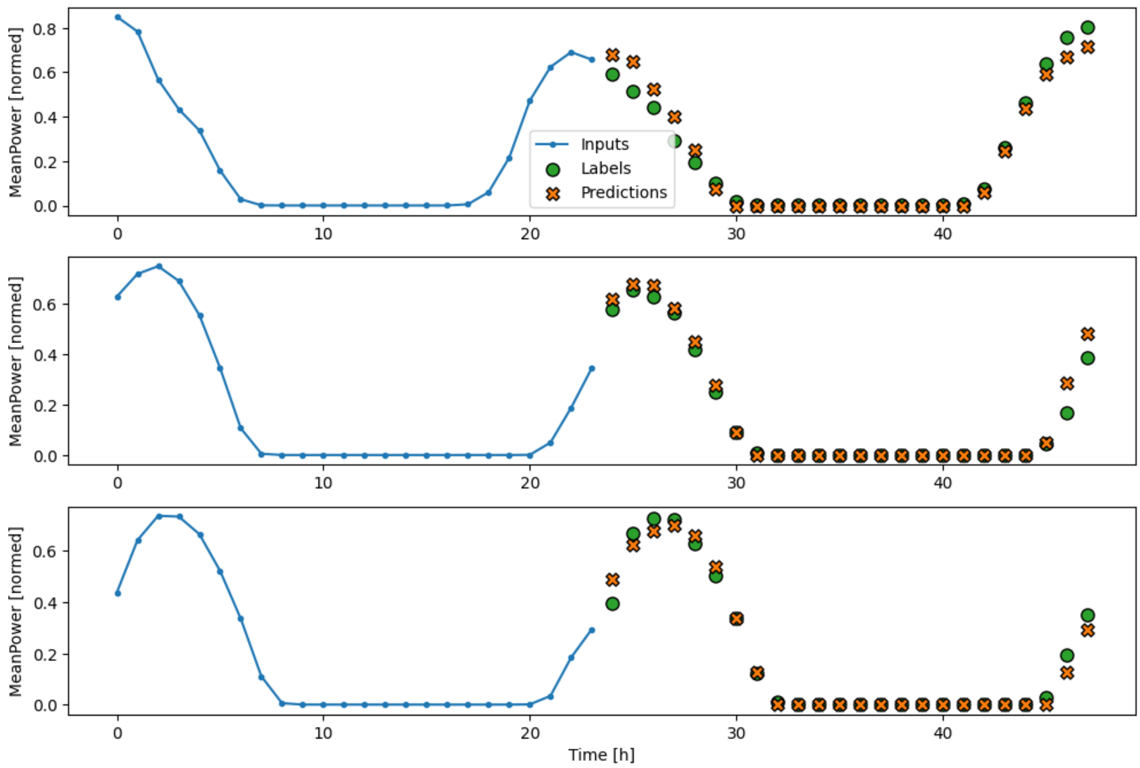


Figure 8.11 3 split window samples of the Test datasets on PV generation forecasting using CNN model

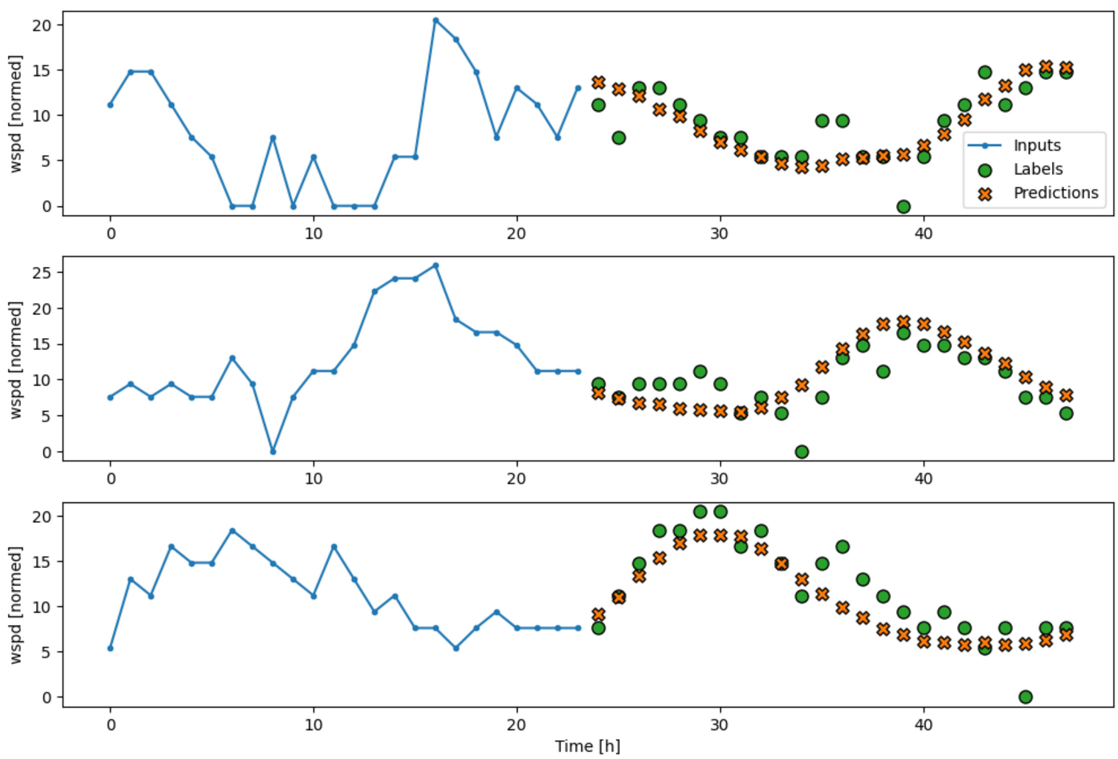


Figure 8.12 3 split window samples of the Test datasets on wind speed forecasting using CNN model

It can be observed that the CNN models have captured the complex patterns of the data in a good way.

Lets look at the actual vs predicted plot samples for wind speed forecasting and PV energy forecasting using CNN model.

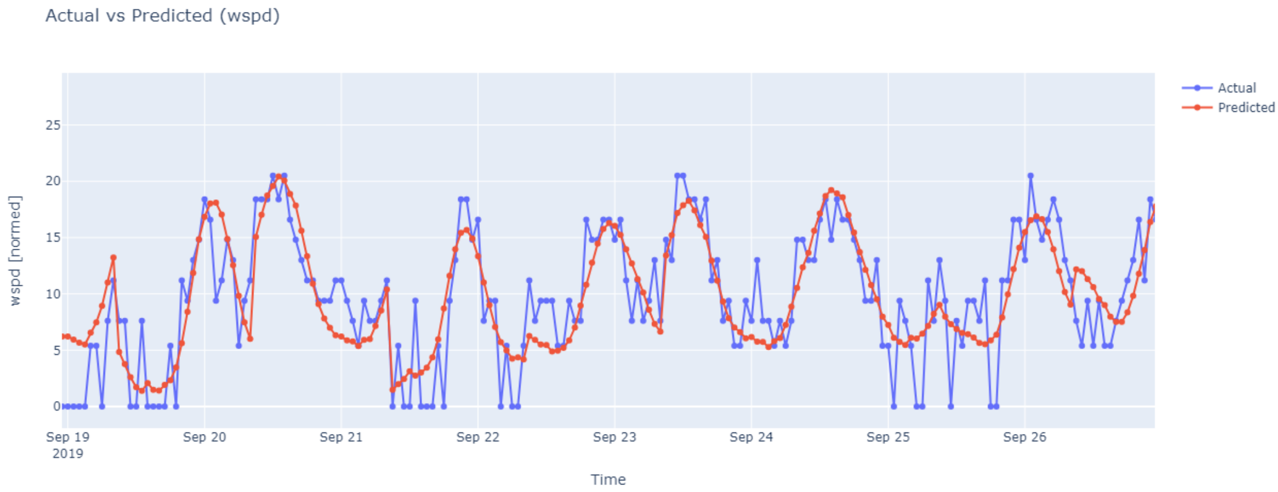


Figure 8.13 Actual vs Predicted plot sample on test dataset of wind speed forecasting CNN model

It can be observed that the fluctuation is very high for wind speed however the model has been able to capture the variation up to some extent.

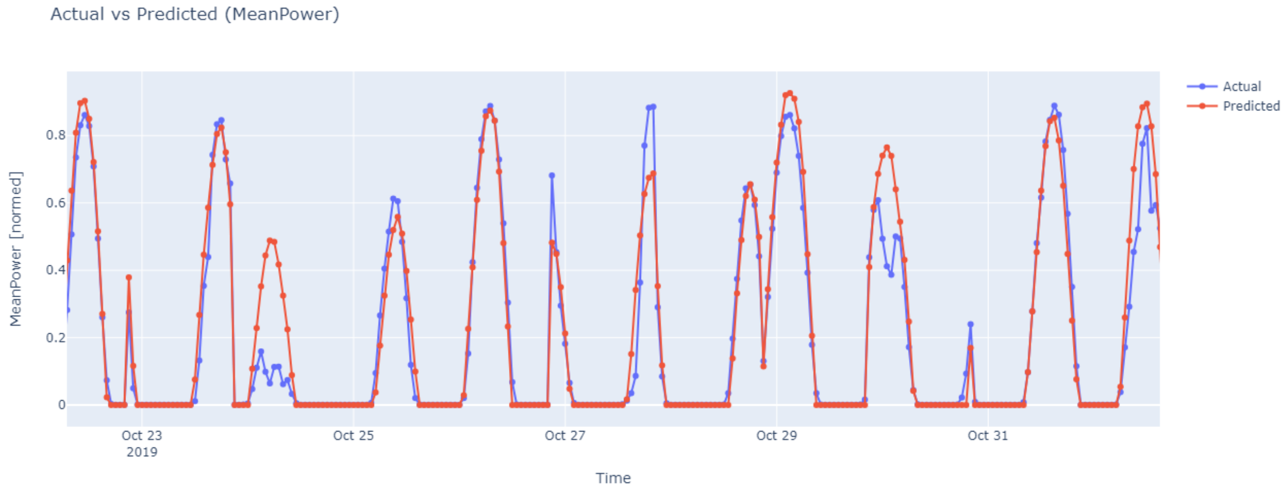


Figure 8.14 Actual vs Predicted plot sample on test dataset of PV energy forecasting CNN model

It can be observed that the fluctuation has a very uncertain special pattern for PV generation.The model has been able to capture the variation very much closely.

In Figure 8.13, which depicts the Actual vs Predicted plot sample on the test dataset for wind speed forecasting using the CNN model, a notable observation is the high level of fluctuation in wind speed. Despite the inherent unpredictability associated with wind speed variations, the CNN model exhibits commendable performance by capturing the variability to a considerable extent.

Turning attention to Figure 8.14, which represents the Actual vs Predicted plot sample on the test dataset for PV energy forecasting using the CNN model, a distinct pattern of uncertain fluctuations in PV generation is evident. Remarkably, the CNN model demonstrates a close alignment with the observed variations, showcasing its ability to accurately capture the nuanced patterns in PV energy generation.

These visualizations underscore the robust performance of the CNN model in both wind speed and PV energy forecasting, effectively navigating the challenges posed by the intricate and dynamic nature of environmental data. The models' capability to closely track fluctuations enhances their reliability in predicting future values, contributing to the efficacy of renewable energy forecasting in practical applications.

9 Conclusions

In essence, this study serves as a steppingstone towards harnessing the potential of machine learning in renewable energy forecasting, contributing to the ongoing global efforts in sustainable energy utilization and environmental stewardship. The research project embarked on a journey to explore the potential of machine learning techniques in forecasting renewable energy generation, with the ultimate goal of optimizing microgrid operation1. The project was rooted in the understanding that renewable energy sources, such as solar and wind power, are inherently variable and unpredictable. This variability poses significant challenges for the efficient operation of microgrids, which are small-scale power grids that can operate independently or in conjunction with the main electrical grid.

To address these challenges, the project leveraged machine learning techniques to predict the output of photovoltaic (PV) and wind power based on weather data. The project collected and processed data from a microgrid located in San Diego, California, and used this data to train a Convolutional Neural Network (CNN) model. The CNN model was chosen due to its ability to process spatial-temporal data and its proven effectiveness in various forecasting tasks.

The performance of the CNN model was evaluated using several metrics, including Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). The results showed that the CNN model achieved better accuracy and lower errors than the baseline model and a Long Short-Term Memory (LSTM) model. This demonstrated the potential of CNNs in accurately forecasting renewable energy generation.

As we move forward, the insights gained from this study provide valuable contributions to the field of renewable energy forecasting. The robustness of the CNN and LSTM models positions them as promising tools for real-world applications, aiding decision-makers in optimizing energy production and consumption. The significance of accurate forecasting cannot be overstated, particularly in the context of transitioning towards sustainable and resilient energy systems.

While our current research has provided valuable insights into the forecasting of wind energy using machine learning models based on hourly data, there remains a rich landscape for future exploration. One avenue for potential improvement is the investigation of the impact of utilizing more granular data intervals, such as 15-minute increments, on model performance. Although our study focused on hourly resolutions, the examination of shorter intervals could unveil intricate wind patterns that contribute to more precise forecasting. Future research endeavors could delve into the dynamics of wind behavior within these finer time resolutions, seeking correlations that may further enhance the accuracy of our predictive models. The exploration of increased temporal granularity stands as an exciting prospect for refining forecasting models and advancing our comprehension of renewable energy dynamics. While our current work has focused on hourly data, this proposition opens the door to continual refinement, showcasing the adaptability of machine learning methodologies in optimizing renewable energy solutions for microgrid applications.

In the realm of future work, further refinements to model architectures, hyperparameter tuning, and exploration of additional data features present avenues for improvement. Continuous efforts in model enhancement and adaptation to evolving data patterns will play a pivotal role in advancing the reliability and precision of renewable energy forecasting models.

In addition to forecasting, the project also investigated various methods to optimize microgrid operation using the forecasted data2. The aim was to meet the demand and other technical constraints, such as emissions and cost, while maximizing the use of renewable energy. The project identified four possible microgrid operation scenarios and suggested some optimization algorithms and techniques for each scenario.

Despite not implementing the optimization methods, the investigation provided valuable insights into their potential effectiveness. The research underscored the significant role these methods could play in improving the performance of the microgrid, leading to more efficient use of renewable energy and potentially reducing reliance on non-renewable energy sources.

However, the research also highlighted the challenges associated with optimization. These include the need for accurate forecasting of renewable energy availability, the complexity of the optimization algorithms, and the need to adapt the methods to different types of microgrids and energy scenarios.

In conclusion, the research project demonstrated the potential of machine learning techniques in predicting renewable energy generation and optimizing microgrid operation1. It highlighted the challenges and opportunities in this field and provided valuable insights that could guide future research and applications. The project underscored the role of machine learning in advancing renewable energy management and sustainability, and it concluded with a positive outlook on the future of this field. The findings from this research could serve as a valuable reference for future studies and applications in renewable energy management and sustainability. The research also identified areas for future exploration and improvement, emphasizing the ongoing evolution and potential of machine learning in renewable energy management and sustainability.

References

[1] Benti, N. E., Chaka, M. D, Semie, A. G. (2023). Forecasting renewable energy generation with machine learning and deep learning: Current advances and future prospects. Sustainability, 15(9), 7087.

[2] Khayaty, M. S., Movludiazar, A, Fotouhi, R. (2021). Intelligent microgrid energy management system based on deep learning approach. In 2021 11th Smart Grid Conference (SGC). ResearchGate.

[3] Silwal, S., Mullican, C., Chen, Y.-A., Ghosh, A., Dilliott, J, Kleissl, J. (2021). Open-source multi-year power generation, consumption, and storage data in a microgrid. Journal of Renewable and Sustainable Energy, 13(2), 025301.

[4] Raja, S. S. S., Ahmad, A., Alqaraghuli, M. A, Gao, W. (2021). A review on the application of machine learning techniques for renewable energy forecasting. Renewable and Sustainable Energy Reviews, 149, 111367.

[5] Maduabuchi, C., Nsude, C., Eneh, C., Eke, E., Okoli, K., Okpara, E., Idogho, C., Waya, B, Harsito, C. (2023). Renewable energy potential estimation using climatic-weather-forecasting machine learning algorithms. Energies, 16(4), 1603.

[6] Sweeney-Lejeune, L., Browell, J, McMillan, D. (2020). The future of forecasting for renewable energy. Wiley Interdisciplinary Reviews: Energy and Environment, 9(6), e375.

[7] Kim, J.-H., Kim, J.-H, Kim, J.-H. (2020). A deep learning-based forecasting model for renewable energy scenarios to guide sustainable energy policy: A case study of Korea. Renewable and Sustainable Energy Reviews, 121, 109688.

[8] Widodo, D. A., Iksan, N., Udayanti, E. D, Djuniadi. (2021). Renewable energy power generation forecasting using a deep learning method. IOP Conference Series: Earth and Environmental Science, 700(1), 012026.

[9] Meenal R., Binu D., Ramya K.C., Michael P.A., Kumar K.V., Rajasekaran E, Sangeetha B. (2022). Weather Forecasting for Renewable Energy System: A Review.Archives of Computational Methods in Engineering,29,2875-2891.

[10] Ahmed, F., El-Shafie, A, Jaafar, O. (2019). A review of deep learning for renewable energy forecasting. Energy Conversion and Management, 198, 111814.

[11] Aupke, P., Kassler, A., Theocharis, A., Nilsson, M, Uelschen, M. (2021). Quantifying Uncertainty for Predicting Renewable Energy Time Series Data Using Machine Learning. Engineering Proceedings, 5(1), 50. https://doi.org/10.3390/engproc2021005050

[12] Heydari, A., Astiaso Garcia, D., Keynia, F., Bisegna, F, De Santoli, L. (2019). Renewable Energies Generation and Carbon Dioxide Emission Forecasting in Microgrids and National Grids using GRNN-GWO Methodology. Energy Procedia, 159, 154-159.

[13] Dutta, S., Li, Y., Venkataraman, A., Costa, L. M., Jiang, T., Plana, R., Tordjman, P., Choo, F. H., Foo, C. F, Puttgen, H. B. (2017). Load and renewable energy forecasting for a microgrid using persistence technique. Energy Procedia, 143, 617-622.

[14] Aslam, S., Javaid, N., Khan, Z. A., Alhussein, M, Alamri, A. (2019). Deep Learning Based Techniques to Enhance the Performance of Microgrids: A Review. In 2019 International Conference on Frontiers of Information Technology (FIT) (pp. 1-6). IEEE.

[15] Saez, D., ´ Avila, F., Olivares, D., Ca ´ nizares, C, Mar ˜ ´ın, L. (2015). Fuzzy Prediction Interval Models for Forecasting Renewable Resources and Loads in Microgrids. IEEE Transactions on Smart Grid, 6(2), 548-556.

[16] Rodr´ıguez, F., Fleetwood, A., Galarza, A, Fontan, L. (2018). Predicting ´ solar energy generation through artificial neural networks using weather forecasts for microgrid control. Renewable Energy, 126©, 855-864.

[17] Moradzadeh, A., Moayyed, H., Zakeri, S., Mohammadi-Ivatloo, B, Aguiar, A. P. (2021). Deep Learning-Assisted Short-Term Load Forecasting for Sustainable Management of Energy in Microgrid. Inventions, 6(1), 15.

[18] Husein, M, Chung, I.-Y. (2019). Day-ahead solar irradiance forecasting for microgrids using a long short-term memory recurrent neural network: A deep learning approach. Energies, 12(10), 1856.

[19] Slowik, M, Urban, W. (2022). Machine Learning Short-Term Energy Consumption Forecasting for Microgrids in a Manufacturing Plant. Energies, 15(9), 3382.

[20] Suresh, V., Janik, P., Rezmer, J, Leonowicz, Z. (2020). Forecasting Solar PV Output Using Convolutional Neural Networks with a Sliding Window Algorithm. Energies, 13(3), 723.

[21] “San Diego International Airport - Weather History & Climate Data.” https://meteostat.net/en/station/72290?t=2018-12-04/2018-12-04

Appendix

- Pandas Profiling reports:
  https://drive.google.com/drive/folders/1NGlvD6c2yJCSgyz47Oh4ys4Xk-foobPZ?usp=drive_link
- Downloaded raw Data:
  https://drive.google.com/drive/folders/19UhGVNpQpkvYvaH5_iZB22IkUkFlQggB?usp=drive_link
- Github repository for Google colab Notebooks: https://github.com/Piyumi22/FYRP.git