"Comparison of Various Prediction Methods for Renewable Energy (Solar and Wind)"

Major Project Report

Submitted in Partial Fulfilment of the Requirements for the Degree of

BACHELOR OF TECHNOLOGY

IN

ELECTRICAL ENGINEERING

By

Divyatej Mishra– 19BEE023

Prachi Kanakhara– 19BEE052

Under the Guidance of

Dr. Bhinal Mehta

Department of Electrical Engineering,
School of Energy Technology, Pandit Deendayal Energy University,
Gandhinagar 382 008

MAY 2023

Certificate of Originality of Work

We hereby declare that the B.Tech. Project entitled "Comparison of Various Prediction Methods for Renewable Energy (Solar and Wind)" submitted by us for the partial fulfilment of the degree of Bachelor of Technology to the Dept. of Electrical Engineering at the School of Energy Technology, Pandit Deendayal Energy University, Gandhinagar is the original record of the project work carried out by us under the supervision of Dr. Bhinal Mehta.

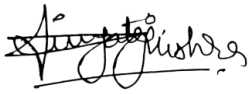
We also declare that this written submission adheres to university guidelines for its originality, and proper citations and references have been included wherever required.

We also declare that we have maintained high academic honesty and integrity and have not falsified any data in our submission.

We also understand that violation of any guidelines in this regard will attract disciplinary action by the institute.

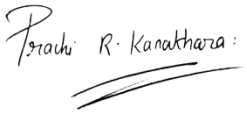
Divyatej Mishra

19BEE023



Prachi Kanakhara

19BEE052



Name of the Supervisor: Dr. Bhinal Mehta
Designation of the Supervisor: Assistant Professor
Signature of the Supervisor:

Place: Pandit Deendayal Energy University, Gandhinagar

Date:

Certificate from the Project Supervisor/Head

This is to certify that Major Project Report entitled "Comparison of Various Prediction Methods for Renewable Energy (Solar and Wind)" submitted by Mr. Divyatej Mishra, 19BEE023 and Ms. Prachi Kanakhara, 19BEE052 towards the partial fulfilment of the requirements for the award of degree in Bachelor of Technology in the field of Electrical Engineering from the School of Technology, Pandit Deendayal Energy University, Gandhinagar is the record of work carried out by him/her under my/our supervision and guidance. The work submitted by the student has in my/our opinion reached a level required for being accepted for examination. The results embodied in this major project work to the best of our knowledge have not been submitted to any other University or Institution for the award of any degree or diploma.

Name and Sign of the Supervisor

Name and Sign of the Industry Supervisor

Name and Sign of the HoD

Name and Sign of the Director

Place

Date

ACKNOWLEDGEMENT

We would like to express our deepest gratitude and appreciation to Dr. Bhinal Mehta for his invaluable guidance, support, and mentorship throughout the duration of our major project.

First and foremost, we are grateful for his willingness to oversee and supervise our project. His extensive knowledge, expertise, and dedication have been instrumental in shaping the direction and scope of our project. His insightful suggestions and constructive feedback have continuously challenged us to think critically and strive for excellence.

We are grateful for the time he has dedicated to meetings and discussions. His open-door policy and willingness to engage in meaningful conversations have not only strengthened our understanding of the subject matter but have also encouraged us to explore new ideas and perspectives.

Furthermore, we extend our appreciation to him for providing valuable resources, including academic references, research materials, and access to specialized equipment and facilities. His support has been pivotal in enabling us to conduct thorough investigations and experiments, contributing to the overall success of our project. We would also like to acknowledge the inspiring academic environment fostered by him. His passion for knowledge and commitment to academic excellence has been contagious, motivating us to push our boundaries and surpass our own expectations. We would like to express our heartfelt gratitude to him for his unwavering encouragement and belief in our abilities. His faith in us has been a constant source of motivation, instilling in us the confidence to overcome challenges and strive for innovation. We are profoundly grateful to him for his exceptional guidance, support, and mentorship throughout our major project. His expertise, dedication, and encouragement have played an integral role in our academic growth and success. We consider ourselves fortunate to have the opportunity to work under his guidance, and we are confident that the knowledge and skills we have gained will have a lasting impact on our future endeavors.

Divyatej Mishra Prachi Kanakhara

ABSTRACT

The rapid growth of renewable energy sources, such as solar and wind, has led to increasing demand for efficient and accurate prediction models to optimize their utilization. This project aims to compare various machine learning techniques for predicting renewable energy generation, specifically focusing on solar and wind energy.

The project starts by gathering historical weather and energy generation data from solar and wind farms. Feature engineering techniques are employed to extract relevant information, such as temperature, wind speed, irradiance, and previous energy production, which serve as inputs to the machine learning models.

Several machine learning algorithms, including random forests, artificial neural networks (ANN), and gradient boosting, are implemented and evaluated. The performance of these models is assessed based on key metrics such as accuracy, root mean square error (RMSE), and mean absolute error (MAE). Results demonstrate that machine learning techniques can effectively capture the complex relationships between weather conditions and renewable energy generation. The comparative analysis reveals the strengths and weaknesses of each model in accurately predicting solar and wind energy output.

Insights gained from this study can guide the selection and optimization of machine learning techniques for renewable energy forecasting. Overall, this project contributes to the field of renewable energy by providing a comprehensive comparison of machine-learning techniques for solar and wind energy prediction. The findings offer valuable insights for researchers, energy operators, and policymakers to make informed decisions and improve the integration and management of renewable energy sources in the power grid.

LIST OF ABBREVIATIONS

ABBREVIATIONS	FULL FORMS
ML	Machine Learning
XGBoost	Extreme Gradient Boosting
GW	Giga-Watt
PV	Photovoltaic
w.r.t	With Respect To
IEA	International Energy Agency
FDI	Foreign Direct Investment
GHGs	Green House Gases
AI	Artificial Intelligence
NWP	Numerical Weather Prediction
BP	Backpropagation
CSO	Cuckoo Search Optimization
DT	Decision Tree
RF	Random Forest
PVGIS	Photovoltaic Geographical Information System
MSE	Mean Square Error
R2	R Squared
ANN	Artificial Neural Network
LR	Linear Regression
GPR	Gaussian Process Regression
M5P	M5 Prime
RMSE	Root Mean Square Error
MAE	Mean Absolute Error
BRR	Bayesian Ridge Regression
CWT	Continuous Wavelet Transform
Catboost	Category Boosting
MdAE	Median Absolute Error
ELM	Extreme Learning Machines
SVR	Support Vector Regression
NREL	National Renewable Energy Laboratory
CFBP	Cascade Forward Back Propagation
SOM	Self-Organizing Map
RBF	Radial Basis Function
MLP	Multilayer Perceptron
IET	Institution of Engineering and Technology
MAPE	Mean Absolute Percentage Error
GWO	Grey Wolf Optimization
PSO	Particle Swarm Optimization
LM	Linear Model
ANF	Adaptive Neuro-Fuzzy
UNIX	Uniplexed Information Computing System
NE	Normalized Error
NSE	Nash–Sutcliffe Model Efficiency
NRMSE	Normalized Root Mean Square Error
NMBE	Normalized Mean Bias Error
NMAE	Normalized Mean Absolute Error
kW	Kilowatt
PCA	Principal Component Analysis
NASA	National Aeronautics and Space Administration
AnEn	Analog Ensemble

MRE	Mean Root Error
CORR	Correlation
LASSO	Least Absolute Shrinkage and Selection Operator
MPE	Mean Percent Error
MBE	Mean Bias Error
W	Watt
DL	Deep Learning
OLS	Ordinary Least Squares
ARIMA	Autoregressive Integrated Moving Average
ETS	Error-Trend-Seasonality
MLR	Multiple Linear Regression
SVM	Support Vector Machine
Wp	Watts Peak
PMIO	Port-Mapped I/O
VRE	Variable Renewable Energy
U. S	United States
U.K.	United Kingdom
DC	Direct Current
AC	Alternating Current
SCADA	Supervisory Control and Data Acquisition
sklearn	Scikit-Learn
°	Degree
kWh	Kilowatt Hour
m/s	Meter Per Second
KNN	K-Nearest Neighbors
GBDT	Gradient Boosting Decision Tree
LightGBM	Light Gradient-Boosting Machine
RSS	Residual Sum of Squares
LS Obj	Least Squares Objective
At	Actual Value
Ft	Forecast Value
Oi	Observed Values
NSM	National Solar Mission
Si	Simulated Values
INDC	Intended Nationally Determined Contributions
GDP	Gross Domestic Product
GBI	Generation Based Incentive
NIWE	National Institute of Wind Energy
MW	Mega Watts

LIST OF TABLES

LIST OF IMAGES

Table Of Contents

1. INTRODUCTION

1.1 Prologue

Renewable energy sources are becoming increasingly important in the global energy mix because of their potential to reduce greenhouse gas emissions and address the challenges posed by climate change. Solar and wind energy are two of the most prominent and rapidly growing sources of renewable energy. Solar energy is derived from the sun's radiation and can be harnessed through the use of photovoltaic (PV) cells, which convert sunlight directly into electricity. Wind energy, on the other hand, is generated by the movement of air caused by temperature and pressure differences in the Earth's atmosphere. We can capture this energy using wind turbines, which convert the wind's kinetic energy into electricity. Using solar and wind energy has increased significantly over the past decade, driven by declining costs, government incentives, and environmental concerns. According to the International Energy Agency (IEA), renewable energy sources, including solar and wind energy, accounted for over two-thirds of global net electricity capacity additions in 2019. In the same year, solar and wind energy contributed approximately 8% of the world's electricity generation. [1]

1.2 Problem Identification

Table 1.2.1: Global Solar Capacity

Global Electricity Power Generation Capacity Through Solar	849.5 GW (2021) **[2]**
Global Electricity Power Generation Capacity Annual Growth Rate Through Solar	26% (2012-2021) **[3]**
Share of Global Electricity Generation Through Solar	2% (2018) **[4]**

Table 1.2.2: Global Wind Capacity

Global Electricity Power Generation Capacity Through Wind	824.9 GW (2021) **[2]**
Global Electricity Power Generation Capacity Annual Growth Rate Through Wind	13% (2012-2021) **[2]**
Share Of Global Electricity Generation Through Wind	5% (2018) **[4]**

During the 1930s, the energy equivalent of approximately one barrel of oil was used to get 100 barrels of petroleum; by 2006, however, the ratio had been reduced to approximately 1 to 15, a trend that is expected to continue. As the ratio of an energy source nears one to one, the energy becomes inaccessible, despite remaining reserves. Even renewable energies such as solar and wind depend on fossil fuels to manufacture and transport related equipment. As with transportation, agriculture, and virtually every aspect of modern life, energy production itself depends on fossil fuels.

1.2.1 India's Dependence on fossil fuels

Coal production had been the government's focus to meet the energy demands of the world's second-largest population. In 2021, coal dominated primary energy consumption in the country.

According to the Central Electricity Authority, renewable sources of energy were expected to generate half of the country's power by 2030. The other half, however, was still expected to be generated through coal.

Natural gas and crude oil had been equally important fossil fuels along with coal.

The installed natural gas capacity in India was over 24 thousand megawatts as of February 2022. Unlike coal, the power sector was not the leading consumer of natural gas within the nation - it was the heavy industries instead. In 2020, the petroleum and natural gas sector had FDI inflows of around 59 billion U.S. dollars.

Production volume of petroleum products in India from the financial year 2012 to 2021, with an estimate of 2022.

1.2.2 Carbon Dioxide Emission

India is the third-largest emitter of GHGs in the world. India accounts for approximately 7% of today's global emissions. However, India has extremely low per capita emissions that are far below the global average. According to the World Bank data, in 2018, India had per capita emissions of 1.8 tonnes. This is projected to expand to 2.4 tonnes in 2030 per India's Paris Agreement obligations. In terms of sectoral GHG emissions, data from 2016 shows that electricity and heat account for the highest share of GHG emissions, followed by agriculture, manufacturing and construction, the transport sector, industry and land-use change and forestry. [5]

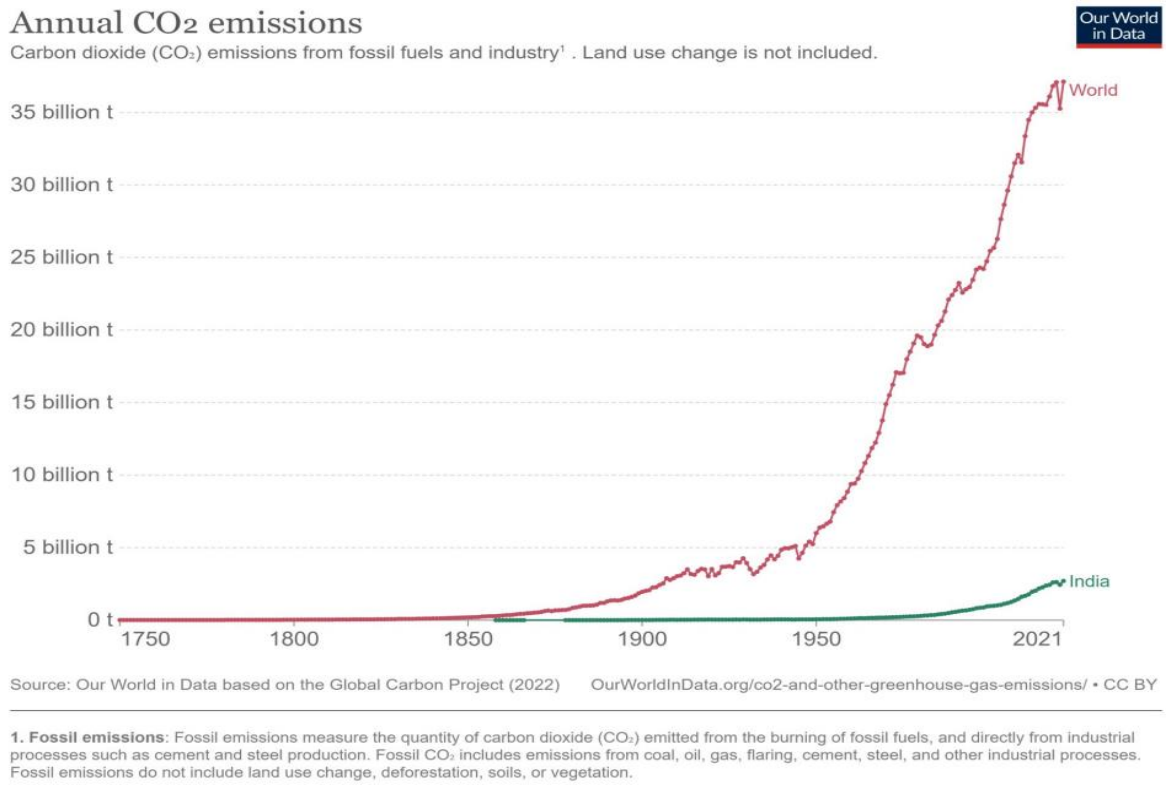


Fig 1.2.2.1: Annual CO2 Emission Of India

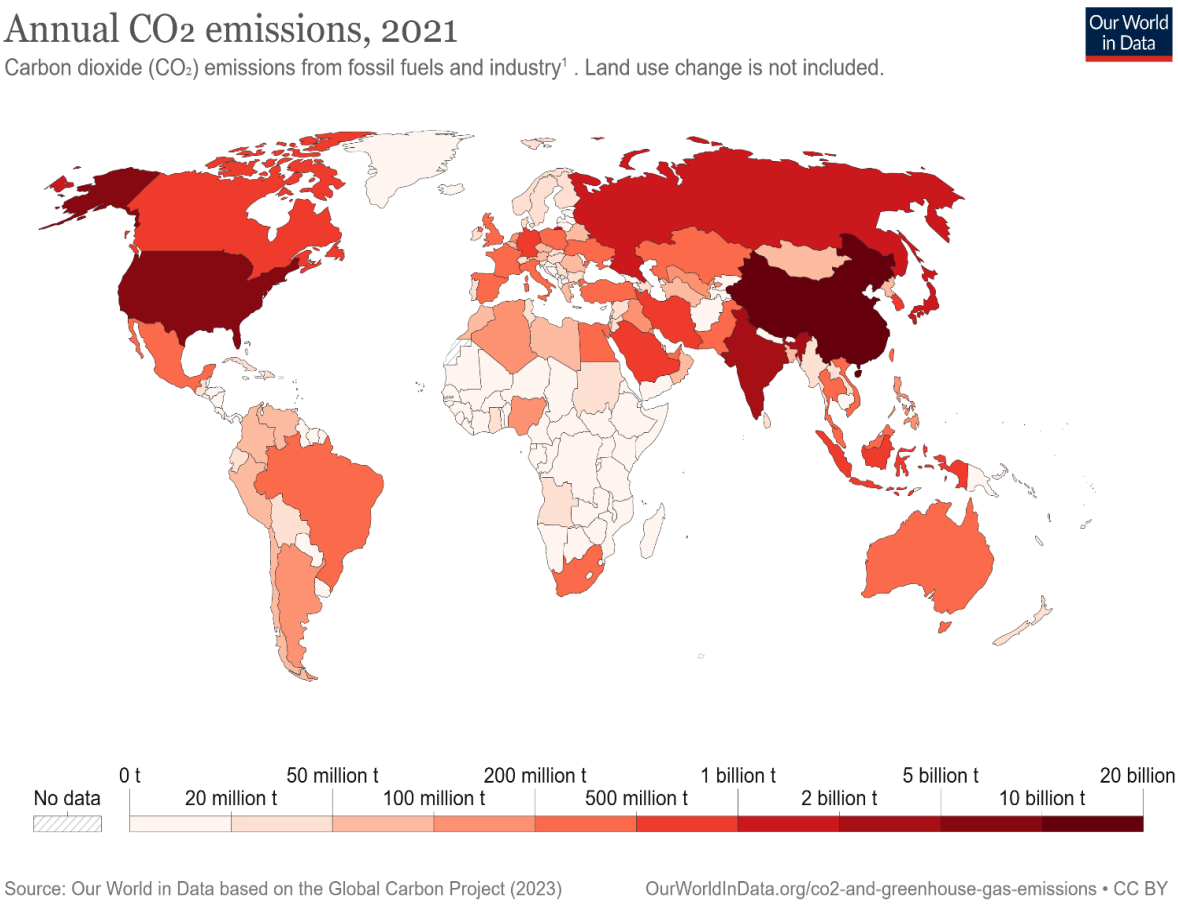


Fig 1.2.2.2: Annual CO2 Emission, 2021

1.3 Scope of Project

- Review previous studies on solar and wind energy output prediction methods, and compare them.
- Analyse the performance of various prediction methods for solar and wind energy output using a set of evaluation criteria.
- Identify the strengths and weaknesses of each prediction method and provide recommendations for the most effective methods for accurate energy output prediction.

1.4 Literature Review

1.4.1 Overview of previous studies on prediction methods for solar and wind energy

Solar energy has become more significant over the years and will continue to do so as a result of the strategic goals of carbon peaking and carbon neutrality [6][7].

Photovoltaic (PV) power generation is more unpredictable than conventional power due to the intermittent and volatile nature of sunlight, which causes several grid difficulties: frequency fluctuation [8][9] and voltage and current surges. [10][11] So, one of the key difficulties of PV

system engineering practice to address the mentioned problems is precisely forecasting the power generation of the PV system. [12][13]

These forecasting methods for PV power encounter several difficulties. At first, it is difficult for physical forecasting technology to identify output distinctive model parameters and collect reliable future weather forecast information. Second, to derive statistical laws, statistical forecasting technology requires vast amounts of historical data rather than specific geographic locations or other information about PV systems. About AI forecasting technology, fundamental flaws in the AI algorithm make it simple to get caught in the local optimum. The purpose of this study is to explain the aforementioned issues and provide some viewpoints on various PV power prediction techniques. [14]

The physical Prediction Method refers to a technology that deduces the variables affecting PV power generation from the principle before building a physical model. Using atmospheric physical data, such as wind speed, temperature, rainfall, humidity, and cloud image via a total sky imager, satellite and day length, physical method modelling is based on numerical weather prediction (NWP). [15][16][17]

To regress some unknown constants and further determine the functional relationship between the output power and the measurable unknown, the statistical method must collect a large amount of data on the PV power generation system's power output. The statistical method can be broken down into three categories based on the number of unknowns: the unary linear regression method, the multiple linear regression method, and the nonlinear regression method. Due to the AI algorithm's strong capacity for self-learning and self-adaptation, PV power forecasting is currently a popular area of research. The backpropagation (BP) neural network prediction model is constructed using the PV array generation sequence, weather type, irradiance intensity, and temperature, as described in the literature. [18] Yet, this technique requires countless authentic power information and huge computation. Additionally, due to the lack of historical data, it is not suitable for either newly constructed or under-construction power stations.

As for wind energy, utilizing wind energy resources effectively is difficult. The power system's dependability can be impacted by wind power's inherent fluctuation and output power's instability. The utilization of wind energy is somewhat constrained by the variability of its speed. As a result, numerous academics have been developing wind energy forecasting strategies to anticipate the state of wind energy. A popular area of research is forecasting wind speed and power, which is expected to help determine energy balance and the scheduling of power generation decisions. [19]

In recent years, the models for wind energy forecasting have undergone significant improvement and expansion. Preliminary divisions include uncertainty analysis and deterministic forecasting. The first can provide precise wind energy forecasting results over a predetermined period, while the second can provide probabilistic and confidence levels for wind energy uncertainty. The deterministic forecasting models fall into four additional categories: intelligent, physical, statistical, and hybrid. [20][21]

The physical approaches simulate the variation tendency of wind speed by taking into account the cause of wind speed (such as pressure, orography, altitude, and so on). Numerical weather prediction (NWP) is the most common physical method. The fluid mechanics and

thermodynamics equations are solved on the computer, and the state of the atmospheric motion in the future over a given period can be predicted. Based on a three-year dataset, a numerical weather prediction model with rapid high-resolution updates was built, and the predictions were in good agreement with the observations. [22] Additionally, cuckoo search optimization (CSO) and the results of the NWP simulation were combined to provide one-day forecasting. [23] The physical models do a great job of forecasting wind energy over the long term, but their use is limited because they require a lot of calculations and don't do a good job of forecasting wind energy over the short term. Therefore, physical methods are not widely employed in the field of wind energy forecasting. [24] Forecasting wind energy as a stochastic process is the consensus among statistical and intelligent methods. They are information-driven models utilizing the gathered verifiable breeze speed/power information or another connected exogenous information to gauge future qualities [25].

Table 1.4.1: Comparison of Prediction Methods

Paper	Prediction Method	Compared Method	Database	Inputs	Forecasting Horizon	Metric
[26]	XGBoost	DT, RFs XGBoost	PVGIS database for the city of Natitingou (Benin) for 12 years	Wind speed, sun position, temperature, direct irradiation, diffuse irradiation and reflected irradiation	3 Days	MSE, R2
[27]	ANN	LR. M5P DT, GPR	Own acquired data (Rooftop) in Qatar	Irradiance, relative humidity, ambient temperature, wind speed, PV surface temperature and accumulated dust	Day Ahead	RMSE, MSE, MAE, R2
[28]	Hybrid (BRR, CWT, Catboost)	-	The Australian weather	Temperature, relative humidity, horizontal irradiation, previous PV power, wind direction, and diffuse horizontal radiation.	24 Hours	RMSE, MSE, MAE, MdAE, R2

[29]	XGBoost	ELM, RF, SVR	The NREL hour1y weather and solar irradiance data for ten years	Dew point temp, Total Cloud Cover, Wind Speed, Sea-level pressure, solar irradiance	Day Ahead	MAE, RMSE
[30]	Ridge Regression	CFBP, SOM, RBF and MLP	Algeciras, Spain, obtained by European Commission for Energy and Transport (IET) PV Geographica l Information System	solar Irradiance incident on the PV panel, cell temperature, Linked turbidity, and wind speed.	Day Ahead	RMSE, MAE, MAPE, R2
[31]	Hybrid (ANN with GWO)	PSO, LM, ANF	Own acquired data – 5 kW grid-connected rooftop PV	UNIX time, date, time, radiation, temperature, pressure, humidity, wind speed, wind direction, sun rise time and sunset time	Day Ahead	NE, NSE, NRMSE , NMBE, NMAE, MSE
[32]	XGBoost + PCA	Random forest, ANN and XGBoost	Kaggle database, Hawaii, collected by NASA	Global horizontal irradiance, Per cent cloud cover and air temperature, solar azimuth and elevation	Day Ahead	RMSE, R2
[33]	ANN + AnEn	ANN and AnEn Combinatio n	3 solar power plant in Italy	clear sky index, image	72 Hours	RMSE, MRE, CORR, BIAS

[34]	LASSO, Rain Forest, Linear Regression , XGBoost, SVM, DL	Polynomial Regression	Own Data, Saudi Arabia, Abha City, King Khalid Univ.		5 Minutes	MSE
[35]	Fuzzy logic	Empirical models	Own Data, photovoltaic module of 210 W power output, Delhi India	Global solar radiation, Sunshine hours, Ambient temperature, Relative humidity, Wind speed, Dewpoint	Day Ahead	MPE, MBE
[36]	LASSO	OLS, ARIMA, ETS	Own Data. Hawaii Oahu Islaml	Horizontal irradiance, direct normal irradiance, diffuse horizontal irradiance, global tilt, air temperature, relative humidity, barometric pressure, wind speed, wind direction	5 Minutes	MAE, RMSE
[37]	SVM + GPR	SVM and GPR	PV modules in Port Harcourt	PV panel temperature, ambient temperature, solar flux, time of day and relative humidity	Day Ahead	RMSE, MAE, R2
[38]	ANN + MLR	ANN and MLR	Harare Institute of Technology, Harare, Zimbabwe,	PMIO, relative humidity,	Day Ahead	RMSE, R2

			100Wp PV	precipitation, wind speed, wind direction, ambient temperature, air pressure, maximum and minimum temperature, dew point, and clearness index		
[39]	SVR	Polynomial Regression and Lasso	Rooftop of Virginia Tech Research Center	Temperature, Dew Point, Relative Humidity, Visibility, Wind Speed, Wind Direction, Cloud Cover	Hourly	RMSE

1.4.2 Comparison of different prediction methods used in previous studies

Previous studies have used a variety of prediction methods to forecast solar and wind energy output. These methods can be broadly classified into three categories: statistical models, machine learning algorithms, and artificial neural networks. Statistical models, such as ARIMA and exponential smoothing, are effective in predicting short-term solar and wind energy output. However, these models may not be suitable for long-term prediction due to their inability to capture complex relationships between variables. Machine learning algorithms, such as SVMs and random forests, have been shown to outperform statistical models in some studies. These methods can identify non-linear relationships between variables and may be more suitable for long-term energy output prediction. However, they can be computationally expensive and require large amounts of training data.

Artificial neural networks have also been used for solar and wind energy output prediction and have shown promising results. ANNs are capable of learning complex relationships between variables and can handle large amounts of data. However, they can be computationally intensive and may require a significant amount of time to train. In addition to these methods, some studies have also used hybrid models that combine different prediction methods to improve accuracy. For example, some studies have used a combination of statistical models and ANNs to forecast energy output.

Overall, the choice of prediction method depends on several factors, including the length of the prediction horizon, the availability of data, and the computational resources available. Further research is needed to determine the most suitable prediction method for different energy output prediction scenarios.

1.4.3 Limitations of Previous Studies and Gaps In Research

Despite the significant progress made in developing and evaluating prediction methods for solar and wind energy output, several limitations and gaps in research exist. Some of these limitations include:

• Limited data sets: Many studies have used limited data sets, which may not be representative of the actual energy output of solar and wind energy systems. This can lead to inaccurate predictions and limit the applicability of the results to real-world scenarios.

• Lack of comparison: Some studies have focused on individual prediction methods without comparing their performance against other methods. This makes it difficult to determine the most suitable method for different prediction scenarios.

• Inadequate consideration of weather conditions: The accuracy of solar and wind energy output prediction is heavily dependent on weather conditions. However, some studies have not considered the impact of different weather conditions on prediction accuracy.

• Lack of long-term prediction studies: Most studies have focused on short-term prediction of solar and wind energy output. However, long-term prediction is critical for the effective integration of solar and wind energy into power grids.

• Limited research on uncertainty and risk analysis: Solar and wind energy output prediction is subject to significant uncertainty and risk. However, limited research has been conducted on how to incorporate uncertainty and risk analysis into prediction models.

These limitations and gaps in the research highlight the need for further research to improve the accuracy and reliability of solar and wind energy output prediction methods. Future studies should use larger and more representative data sets, compare the performance of different prediction methods, consider the impact of different weather conditions, and focus on long-term prediction and uncertainty and risk analysis

1.5 Prediction and Its Need

It involves the 4 following steps:

1. Predicting energy consumption

One of the most crucial parts of renewable energy forecasting is predicting how much energy users will consume. People can monitor usage trends to predict upcoming consumption changes, but this can be slow and inaccurate. Some machine learning algorithms can achieve this even with partial information, making them far more reliable than traditional approaches.

2. Predicting weather conditions

Different weather conditions produce different amounts of power. Machine learning applications can help predict these more accurately than traditional models. If renewables produce higher-than-average levels, energy companies can scale down fossil fuels and vice versa.

3. Predicting market movements

With enough high-quality data, consumer actions are surprisingly predictable. Machine learning algorithms can forecast long-term market movements, so renewable energy companies can understand their audience. Since it can take time to adjust production or marketing strategies, predicting these consumer trends early is crucial. As a result, sustainable energy will spread faster, helping the world move towards a greener future.

4. Predicting Potential Issues

If companies can predict when conditions may threaten the grid, they can prevent it, leading to considerable savings. One study found that AI-assisted predictive maintenance is up to 25.3% more efficient and 24.6% more precise. These savings can help make renewable installations more cost-effective, helping them grow further.

Globally, there is a growing need for renewable energy sources like solar and wind energy as greenhouse gas emissions must be reduced and there are worries about climate change. A fundamental obstacle to these energy sources' incorporation into the electrical grid is their fluctuation as a result of weather conditions. By more accurately and consistently forecasting the production of solar and wind energy systems, machine learning has the potential to solve this problem.

The reliability and stability of the electrical grid are one of the key reasons why it is necessary to estimate the production of solar and wind energy systems. Due to variations in meteorological factors like cloud cover and wind speed, solar and wind energy systems can provide very variable output. Due to this fluctuation, it may be challenging to maintain a balance between grid supply and demand, which may result in power outages and other disturbances.

To more accurately estimate future energy output, machine learning algorithms may be used to analyse previous data on weather patterns and energy output. These algorithms may be trained on massive datasets of historical weather and energy data to find patterns and correlations between meteorological conditions and energy output. The output of solar and wind energy systems may then be accurately predicted using this information.

The requirement to forecast solar and wind energy systems' production is crucial to improve their efficiency and cut costs. Energy suppliers may manage their resources more effectively and prepare for future demand if they can predict the output of these systems. This can assist guarantee that renewable energy sources are utilised effectively and lessen the need for pricey backup systems.

To increase their effectiveness and save costs, solar and wind energy plants must have their production predicted. If energy providers can forecast the output of these systems, they may be able to better manage their resources and get ready for future demand. This can help ensure the efficient use of renewable energy sources and reduce the need for expensive backup systems.

Scheduling, dispatch, real-time balancing, and reserve requirements are just a few of the system functions that are impacted by Variable Renewable Energy (VRE) predictions. Power system operators may successfully balance load and generation in intra-day and day-ahead scheduling by anticipating up and down ramps in VRE generation by incorporating VRE predictions into system operations. As a result, fuel costs are decreased, system dependability is increased, and the use of renewable resources is not restricted as much. In general, machine learning is essential for predicting the amount of solar and wind energy that will be produced, which is important for integrating renewable energy sources into the electrical grid. Machine learning may improve the performance of renewable energy systems, save costs, and maintain the stability and dependability of the grid by more accurately forecasting electricity output. Machine learning is projected to become more crucial to the development and integration of renewable energy sources as the demand for such sources rises.

2. MACHINE LEARNING AND ALGORITHMS

2.1 Machine Learning

2.1.1 Basics

Machine Learning is said as a subset of artificial intelligence that is mainly concerned with the development of algorithms which allow a computer to learn from the data and past experiences on its own.

Machine learning algorithms create a mathematical model with the use of historical sample data, or "training data," that aids in generating predictions or judgements without being explicitly programmed. Computer science and statistics are used with machine learning to create prediction models. Algorithms that learn from past data are created by machine learning or used in it. The performance will be higher the more information we supply.

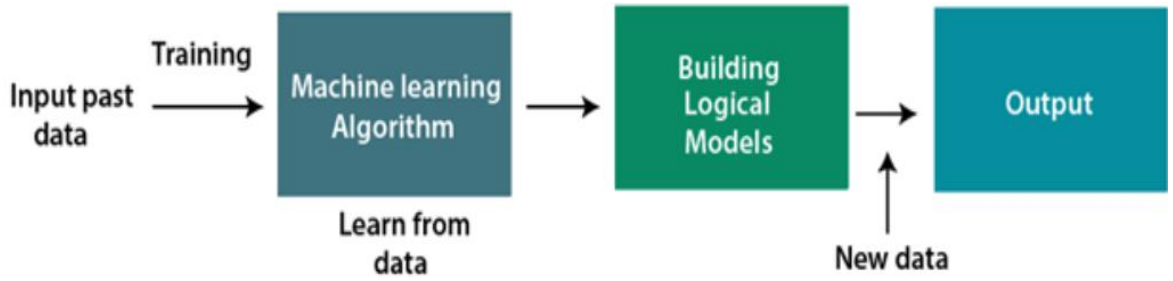


Fig 2.1.1.1: Machine Learning Basics

Machine learning can be classified into 3 types namely, Supervised, Unsupervised and Reinforcement learning.

In supervised learning, the machine learning system is trained with sample-labelled data, and then it makes predictions about the outcome based on those predictions. The system builds a model using labelled data to comprehend the datasets and learn about each one. After training and processing, the model is tested using sample data to see if it accurately predicts the desired outcome. In supervised learning, mapping input and output data is the main objective. Regression and classification are additional categories under which supervised learning falls.[40]

Unsupervised learning is when a computer learns alone. A set of unlabelled, unclassified, or uncategorized data is used to train the machine, and the algorithm is then required to operate independently on that data. Unsupervised learning's objective is to reorganise the input data into fresh features or a collection of objects with related patterns. There is no predefined outcome in unsupervised learning. The machine searches through a vast quantity of data to uncover insightful information. It can be further classified into clustering and association.

A machine in a reinforcement learning system receives a reward for each correct action and receives a penalty for each incorrect activity. With the help of this feedback, the agent automatically learns and performs better. The agent explores and engages with the environment during reinforcement learning. An agent performs better since its objective is to accrue the greatest reward points.[40]

Python is currently one of the most widely used programming languages for this activity, and it has supplanted several others in the business, in part due to its enormous library ecosystem. Some of the libraries that we ventured upon during the course of this project are numpy, pandas, seaborn, matplotlib, Keras, sklearn, etc.

2.1.2 Procedure

Identifying the problem that needs to be solved is the first step in creating a machine-learning model. This entails determining the model's objective, such as foreseeing a specific conclusion or making a choice based on information.

To create a machine learning model, the problem that needs to be solved must first be determined. This requires figuring out the model's goal, such as anticipating a particular outcome or making a decision based on available data.

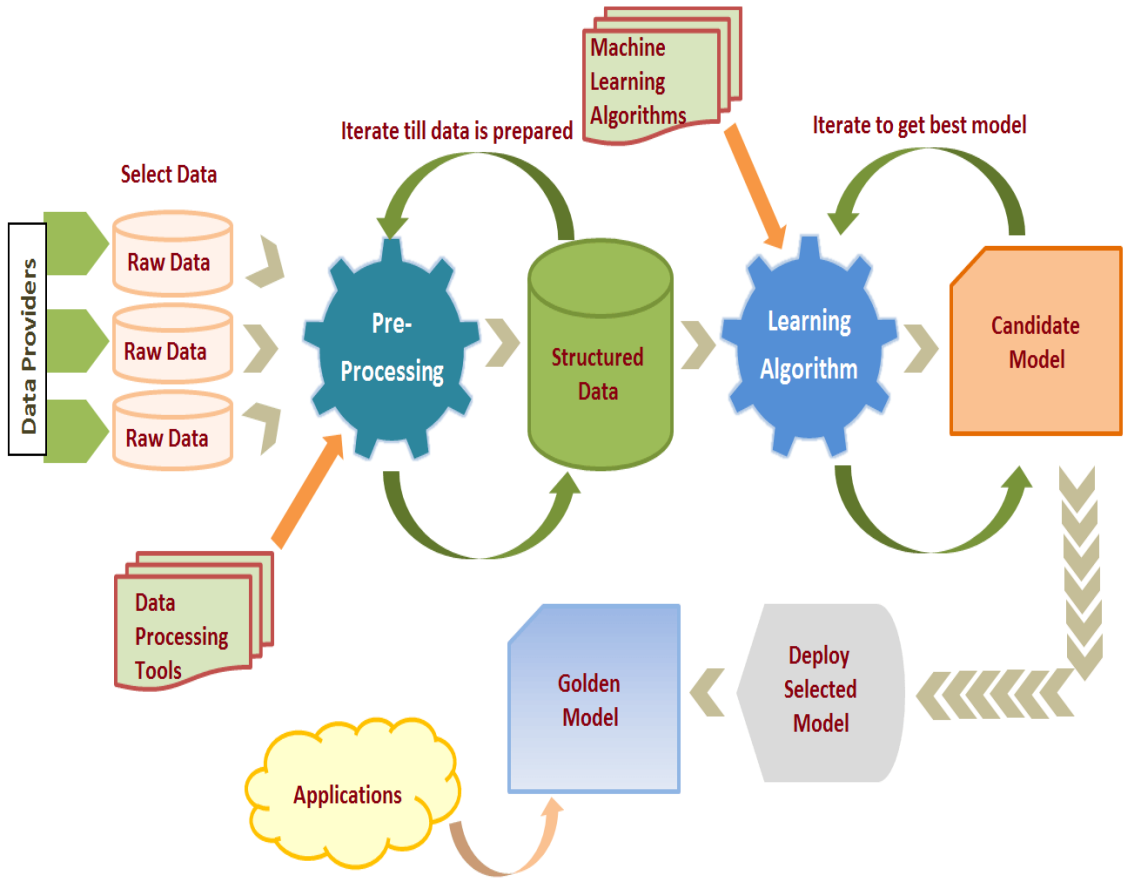


Fig. 2.1.1.2: Model Building Using ML

Once the model has been optimised, it may be put into use in a real-world setting to make judgements or predictions in response to fresh data.

Monitoring and maintaining the model is necessary to ensure it keeps performing properly over time once it has been implemented. If the underlying system changes or new data becomes available, this may require updating the model.

2.2 Methodology

Table 2.2.1: Research Methodology

Research Method	Quantitative Research Method
Sources	Secondary Sources
Research Design	Comparative

For solar, the dataset is from two solar power plants in India for over 34 days. It has two pairs of files - each pair has one power generation dataset and one sensor readings dataset. The power generation datasets are gathered at the inverter level. - each inverter has multiple lines of solar panels attached to it. The sensor data is gathered at a plant level–a single array of sensors is optimally placed at the plant. [41]

Table 2.2.2: Data Collected for Solar Energy

DATE_TIME	Date and time for each observation. Observations were recorded at 15-minute intervals.
PLANT_ID	This will be common for the entire file.
SOURCE_KEY	The source key in this file stands for the inverter id
DC_POWER	Amount of DC power generated by the inverter in this 15-minute interval. Units - kW
AC_POWER	Amount of AC power generated by the inverter in this 15-minute interval. Units - kW.
DAILY_YEILD	Daily yield is a cumulative sum of power generated on that day until that point in time.
TOTAL_YIELD	This is the total yield for the inverter until that point in time.
AMBIENT_TEMPERATURE	this is the ambient temperature at the plant.
MODULE_TEMPERATURE	There has a module (solar panel) attached to the sensor panel. This is the temperature reading for that module.
IRRADIATION	Amount of irradiation for the 15-minute interval.

For wind, data measured from Scada Systems like wind speed, wind direction, generated power etc. for 10 minutes intervals are used. This file was taken from a wind turbine's SCADA system that is working and generating power in Turkey. [42]

Table 2.2.3: Data Collected for Wind Energy

DATA_TIME	10 Minute time interval
WIND_SPEED (m/s)	Wind speed that turbine use for electricity generation
WIND_DIRECTION (°)	The wind direction at the hub height of the turbine (wind turbines turn in this direction automatically)
LV_ACTIVE_POWER (kW)	Power generated by the turbine for that moment
THEORETICAL_POWER_CURVE (kWh)	theoretical power values that the turbine generates with that wind speed, which is given by the turbine manufacture

2.3 Algorithms and Techniques

Table 2.3.1: Machine Learning Algorithms

Algorithm	Type of the task
K-nearest neighbour	Classification
Naive Bayes	Classification
Support vector machine	Classification
Linear regression	Classification/Regression
Random forest	Classification/Regression
K-means	Clustering
Principal component analysis	Feature extraction and dimensionality reduction
Canonical correlation analysis	Feature extraction
Neural networks	Classification/Regression

Each method has its strengths and weaknesses, and the choice of method depends on the specific requirements of the prediction task. The artificial neural network method provides the most accurate and reliable predictions, but it requires a large amount of data and can be computationally expensive. The statistical model is simple and can be used with relatively small amounts of data, but it assumes a linear relationship between the variables. The machine learning technique method can capture complex nonlinear relationships between the variables but can be computationally expensive and difficult to interpret. [43]

1. **Logistic Regression**

The supervised learning approach known as logistic regression is frequently used in situations involving binary categorization. Even though "regression" and "classification" are mutually exclusive terms, the word "logistic" refers to the logistic function, which is what this algorithm uses to perform the classification task. For many binary classification jobs, logistic regression is a popular choice since it is a straightforward yet highly successful classification technique. [43]

The logistic function, often known as the sigmoid function, is the foundation of logistic regression and converts any real-valued integer to a value between 0 and 1.

$$Sigmoid\ Function: \quad y = \frac{1}{1+e^{-x}} \qquad \text{… Eq. 2.3.1}$$

$$Linear\ Equation: \quad z = \beta_0 + \beta_1 x + \cdots + \beta_n x_n \qquad \text{… Eq. 2.3.2}$$

A logistic regression model employs the logistic function and log odds to perform a binary classification job on an input linear equation. The infamous logistic regression s-shaped graph will then appear.

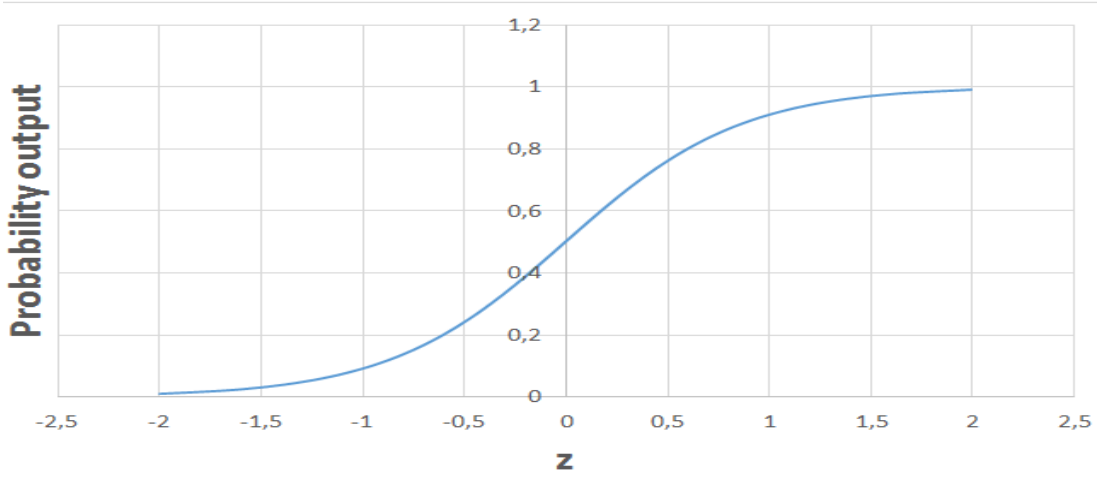


Fig. 2.3.1: Logistic Regression S-Shaped Graph

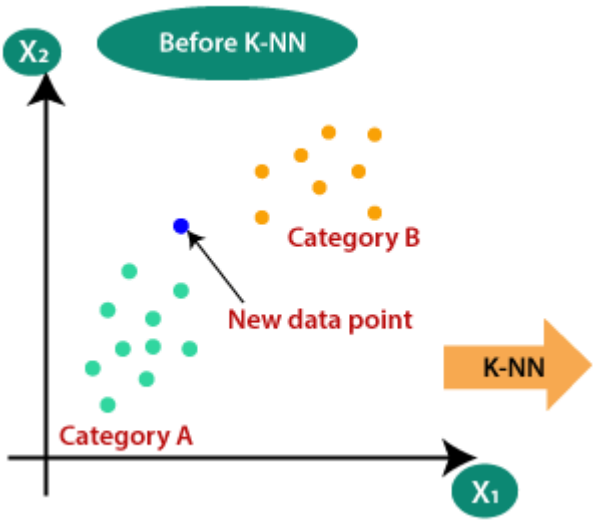
The probability that was calculated can be used "as is." The result may read, for instance, "The probability that this email is spam is 95%" or "The probability that a customer will click on this advertisement is 70%". However, probabilities are typically used to categorise data points. For instance, the forecast is a positive class (1) if the likelihood is higher than 50%. Any other case results in a negative class (0) prediction.

Choosing the positive class for all probability values greater than 50% is not always preferred. In the case of spam emails, we need to be nearly certain before we can label a message as such. We do not want the user to miss critical emails because emails marked as spam are automatically routed to the spam folder. Unless we are almost certain, emails are not labelled as spam. Conversely, when categorising a health-related issue, we must be much more sensitive. We do not want to overlook a cancerous cell, even if we have a slight suspicion that it may be. As a result, the value that marks the boundary between the positive and negative classes depends on the problem. [43]

2. KNN

One of the simplest machine learning algorithms, based on the supervised learning method, is K-Nearest Neighbour. The K-NN method makes the assumption that the new case and the existing cases are comparable, and it places the new instance in the category that is most like the existing categories. A new data point is classified using it based on similarity after all the existing data has been stored. This means that utilising this method, fresh data may be quickly and accurately sorted into a suitable category. Although this technique is most frequently employed for classification issues, it may also be utilised for regression. Since it is a non-parametric technique, it makes no assumptions about the underlying data. [43]

If there are two categories, Category A and Category B, and we have a new data point, x1, which category does this data point belong in? We require a K-NN method to address this kind of issue. K-NN makes it simple to determine the category or class of a given dataset. Consider the illustration below



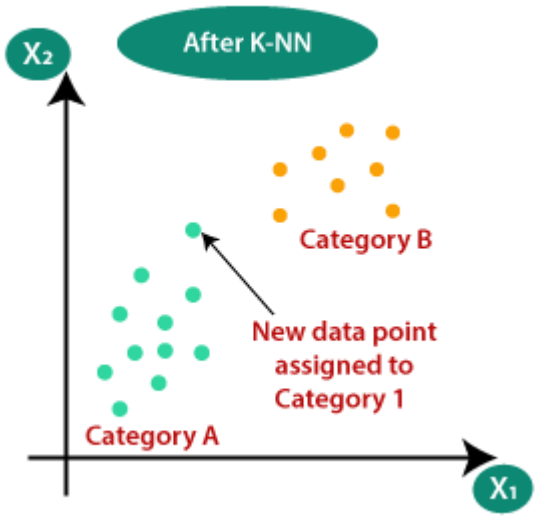


Fig. 2.3.2: Before KNN **Fig. 2.3.3: After KNN**

The choice of an ideal k value is crucial. The model is excessively particular and difficult to generalise if k is set too low. It also tends to be noise-sensitive. The model achieves a high level of accuracy on the train set but will perform poorly as a predictor for fresh, unforeseen data points. Consequently, we are likely to produce an overfit model. The model is overly generalised and is not a strong predictor on both the train and test sets, however, if k is too big. Underfitting is the term for this circumstance. [43]

kNN is a basic and intuitive model. It applies to non-linear jobs since it does not rely on any assumptions. Since the model must store every data point, kNN becomes incredibly slow as the number of data points rises. As a result, it is not memory efficient. The sensitivity of kNN to outliers is another drawback.

3. Decision Trees

Iteratively asking questions to divide up the data is the foundation of a decision tree. With a decision tree's visual representation, it is simpler to understand how the data will be divided up.

A decision tree to forecast client turnover is shown like this. The first split is based on the total monthly charges. The system then asks more questions to create distinct class labels. As the tree grows deeper, the inquiries get more focused. [43]

The decision tree technique seeks to maximise predictiveness at each division so that the model continuously learns more about the dataset. Usually, randomly dividing the features does not provide us with useful information about the dataset. Splits with higher node purity provide additional information. The distribution of various classes inside a node is negatively correlated with the purity of that node. The selection of the questions is done to promote purity or minimise impurity.

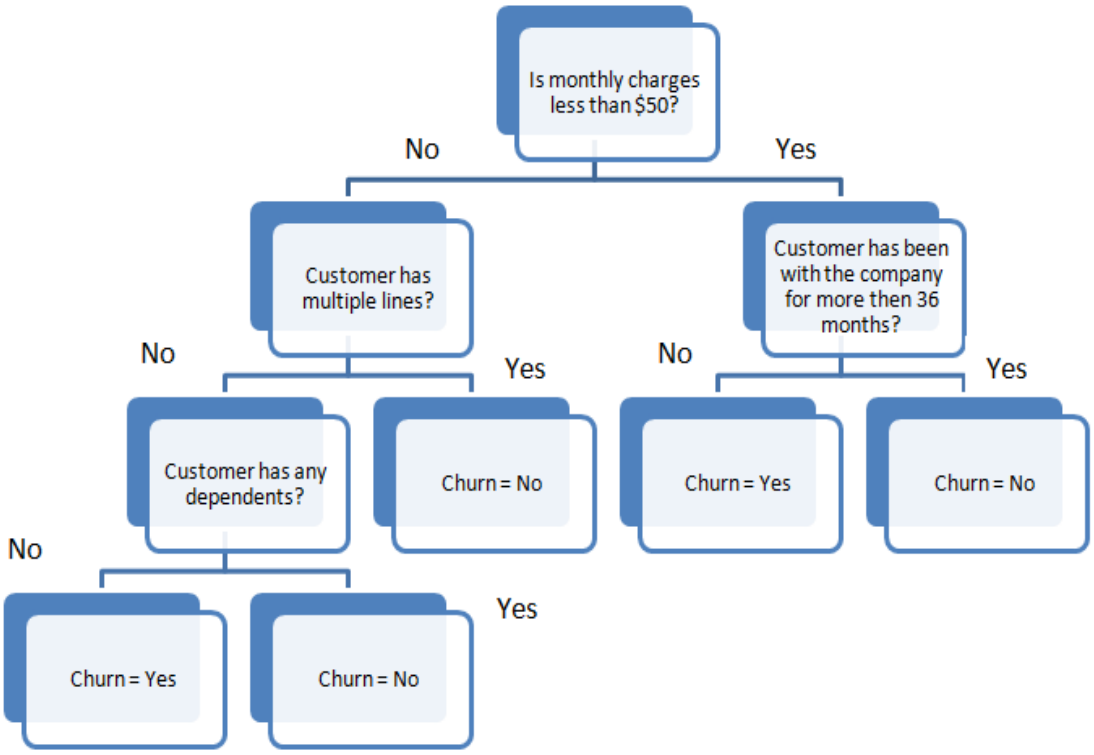


Fig. 2.3.4: Flow chart to Maximise Purity

How many inquiries are made? What time do we stop? When is our tree enough to address the classification issue we are facing? All of these inquiries point us in the direction of overfitting, one of the key ideas in machine learning. Up until all the nodes are pure, the model can continue to ask queries. This model, meanwhile, would be very specialised and would have trouble generalising. With the training set, it achieves high accuracy, but on new, previously unobserved data points, it performs poorly, which suggests overfitting. For the sci-kit-learn decision tree algorithm, the max_depth parameter controls the depth of a tree. [43]

How many questions are asked? When do we finish? When will our tree be sufficient to resolve the categorization problem we are experiencing? All of these questions lead us to overfit, one of the central concepts in machine learning. The model can keep posing questions until all nodes are pure. In contrast, this model would be quite specific and have difficulty generalising. It works well with the training set and reaches high accuracy, but struggles with fresh, previously unseen data points, which raises the possibility of overfitting. The max_depth option governs a tree's depth in the scikit-learn decision tree method.

Typically, the decision tree algorithm does not need features to be scaled or normalised. Working with a variety of feature data types (continuous, categorical, and binary) is also acceptable. On the downside, it is prone to overfitting and requires assembling to achieve good generalisation. [43]

4. Random Forest

A group of several decision trees is called a random forest. Decision trees are employed as parallel estimators in the bagging technique, which is used to construct random forests. When used to a classification issue, the outcome is determined by the majority vote of the findings from each decision tree. In a regression, the mean value of the target values in a leaf node

serves as the prediction. The mean value of the decision tree outcomes is taken into account by random forest regression.

Random forests are substantially more accurate than a single decision tree and lower the danger of overfitting. Additionally, decision trees in a random forest operate concurrently to prevent time from becoming a bottleneck.

Using uncorrelated decision trees is crucial to the effectiveness of a random forest. The total outcome won't change significantly from the outcome of a single decision tree if we utilise the same or very comparable trees. By using bootstrapping and feature randomization, random forests can produce decision trees that are not correlated.

Bootstrapping involves replacing training data with samples chosen at random. We refer to them as bootstrap samples. [43]

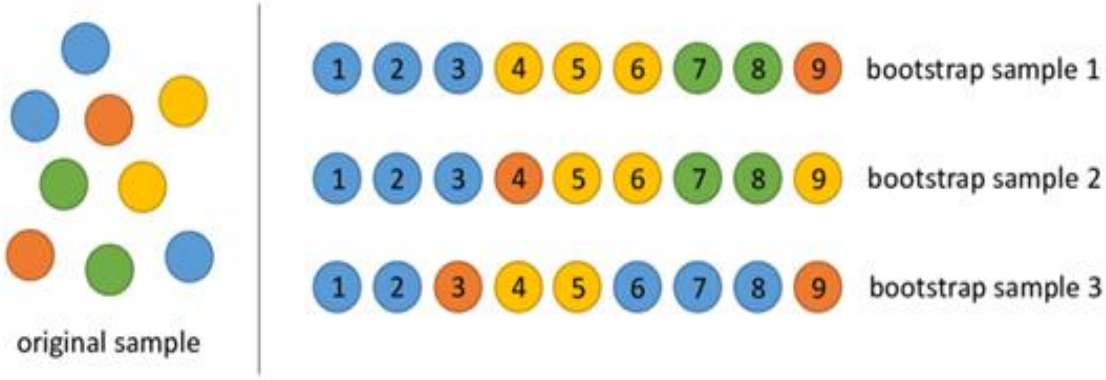


Fig. 2.3.5: Bootstrap Samples

By choosing features at random for each decision tree in a random forest, feature randomization is obtained. The max_features option allows you to regulate how many features are utilised for each tree in a random forest.

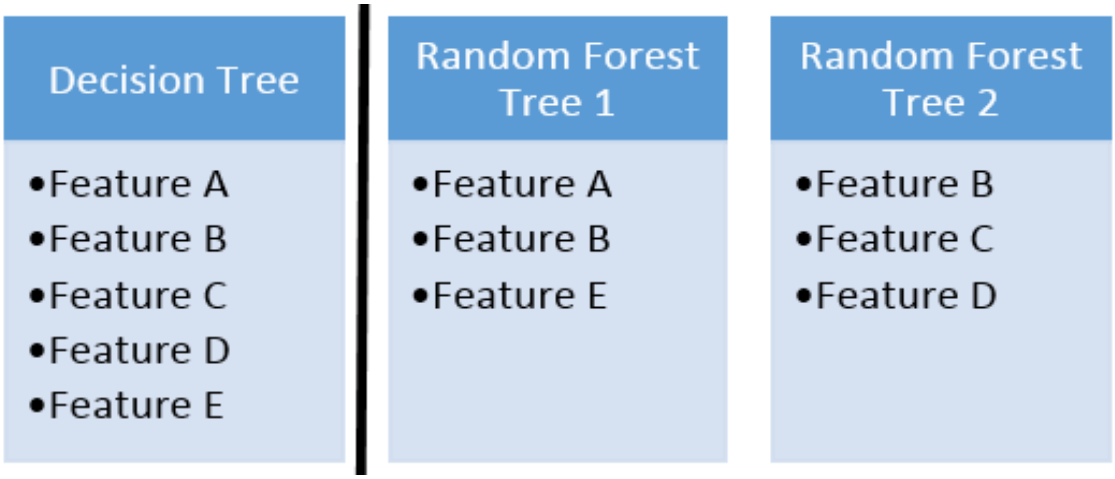


Fig. 2.3.6: Feature Randomization

The random forest does not require normalisation or scaling and is a very accurate model for many different issues. However, when compared to fast linear models (such as Naive Bayes), it is not a good option for high-dimensional data sets (such as text classification). [43]

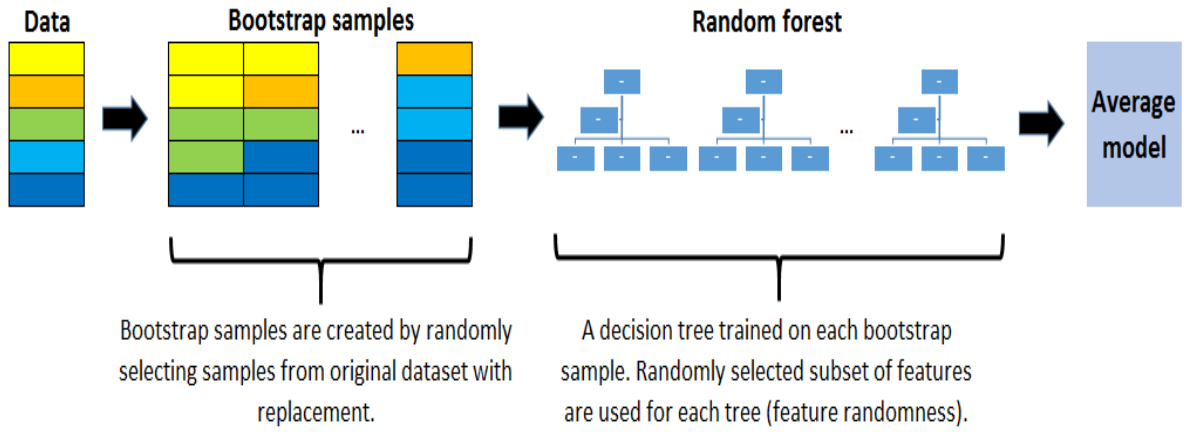


Fig. 2.3.7: Creating an Average Model

5. Gradient Boosting Decision Tree

It is an ensemble algorithm that combines many decision trees using the boosting approach.

Boosting is the process of successively combining learning algorithms to produce a strong learner from a large number of weak learners. Decision trees serve as weak learners in GBDT.

Every tree makes an effort to reduce the mistakes of the preceding tree. Although boosting's trees are poor learners, by adding numerous trees in succession and having each one concentrate on the mistakes made by the previous one, boosting becomes an extremely effective and precise model. Bootstrap sampling is not used in boosting, in contrast to bagging. Every time a new tree is introduced, the basic dataset is changed to suit the new tree. [43]

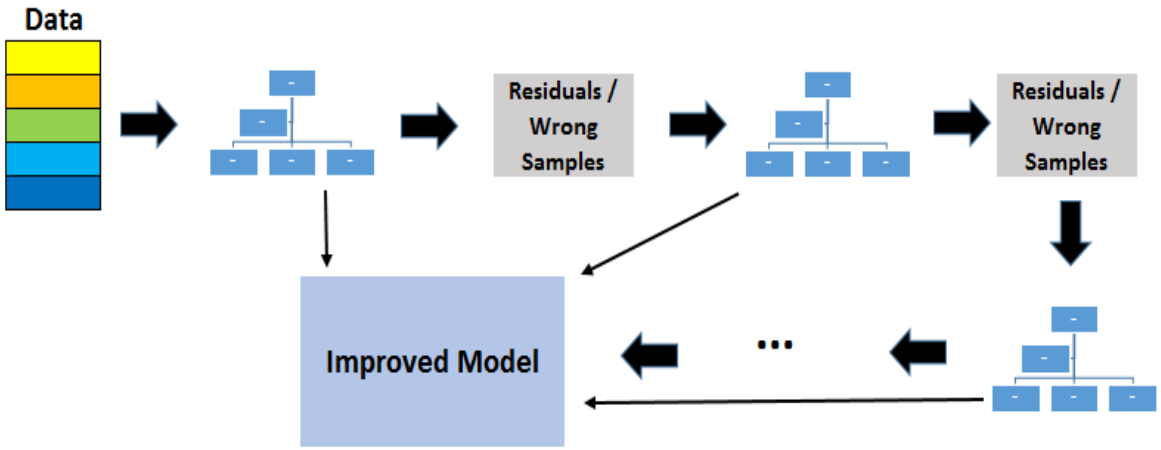


Fig. 2.3.8: Model Improvement

Due to the consecutive addition of trees, boosting algorithms acquire knowledge slowly. Slower learning models outperform faster learning ones in statistical learning.

Two essential hyperparameters for gradient-boosting decision trees are learning rate and n_estimators. The term "learning rate" simply refers to how quickly the model picks up new information. The overall model is altered by each additional tree. The rate of learning determines the extent of the alteration. The number of trees utilised in the model is indicated by the n_estimator. We need additional trees to train the model if the learning rate is poor.

However, we must be extremely selective when deciding how many trees to plant. Using an excessive number of trees increases the danger of overfitting.

When compared to random forests, GBDT is more effective at both classification and regression tasks and offers more precise predictions. It can handle mixed-type features and doesn't require any pre-processing. To avoid the model overfitting in GBDT, hyperparameter adjustment must be done carefully.

The GBDT method is so strong that several improved variants of it, including XGBOOST, LightGBM, and CatBoost, have been applied. [43]

Difference Between Random Forest and Gradient Boosting Decision Tree

The number of trees utilised in the model is one significant distinction between gradient-boosting decision trees and random forests. In random forests, adding more trees does not result in overfitting. After a certain point, adding more trees does not increase the model's accuracy, but neither does adding too many trees have a negative impact. Even though there is no overfitting risk associated with the number of trees in random forests, you still do not want to add an excessive number of trees for computational reasons.

However, in terms of overfitting, the number of trees in gradient-boosting decision trees is extremely important. Overfitting will result from adding too many trees, thus it's vital to stop at some time. [43]

6. XGBoost

The Gradient-Boosted Decision Tree (GBDT) machine learning toolkit XGBoost, which stands for "Extreme Gradient Boosting," is scalable and distributed.

Weights are significant in XGBoost. Each independent variable is given a weight before being put into the decision tree that forecasts outcomes. Variables that the tree incorrectly anticipated are given more weight before being placed into the second decision tree. These distinct classifiers/predictors are then combined to produce a robust and accurate model. It may be used to solve issues including regression, classification, ranking, and custom prediction. [43]

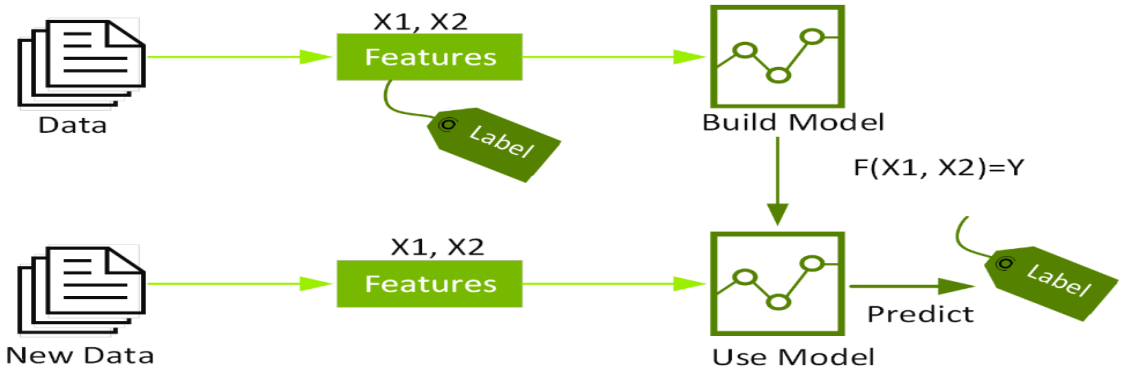


Fig. 2.3.9: Creating an Accurate Model Using XGBoost

7. Ridge and Lasso Regression

These regression techniques are used for creating models with limited resources in the presence of a large number of features. Large enough to cause computational challenges and to enhance the tendency of a model to overfit.

Ridge and Lasso appear to be working towards the same objective, yet their intrinsic characteristics and actual application cases diverge greatly.

They operate by penalising the size of feature coefficients and reducing the discrepancy between expected and actual observations. These methods are referred to as regularisation methods. The main distinction is in how they apply the coefficients' penalties:

- Ridge Regression:
    - Performs L2 regularization, i.e., adds penalty equivalent to the square of the magnitude of coefficients
    - Minimization objective = LS Obj + α * (sum of square of coefficients)
    - The parameter (alpha) balances the relative importance of minimising the RSS and the sum of squares of coefficients. has a range of possible values:
        - Alpha = 0
            - The objective becomes the same as simple linear regression.
            - We'll get the same coefficients as simple linear regression.
        - Alpha = infinity
            - The coefficients will be zero. Why? Because of infinite weightage on the square of coefficients, anything less than zero will make the objective infinite
        - 0 < alpha < infinity
            - The magnitude of α will decide the weightage given to different parts of the objective.
            - The coefficients will be somewhere between 0 and ones for simple linear regression. [43]
- Lasso Regression:
    - LASSO stands for Least Absolute Shrinkage and Selection Operator.
    - Performs L1 regularization, i.e., adds penalty equivalent to the absolute value of the magnitude of coefficients
    - Minimization objective = LS Obj + α * (sum of the absolute value of coefficients)
    - α (alpha) works similarly to that of the ridge and provides a trade-off between balancing RSS and the magnitude of coefficients. Like that of the ridge, α can take various values:
        - α = 0: Same coefficients as simple linear regression
        - α = ∞: All coefficients zero (same logic as before)
        - 0 < α < ∞: coefficients between 0 and that of simple linear regression

Here, LS Obj refers to the 'least squares objective,' i.e., the linear regression objective without regularization. [43]

8. Extra Trees Regression

It is a form of ensemble learning method that combines the findings of various de-correlated decision trees gathered in a "forest" to get its categorization outcome. It is conceptually very

similar to a Random Forest Classifier and only varies from it in how the decision trees in the forest are built.

The first training sample is used to build each decision tree in the Extra Trees Forest. The optimal feature to divide the data according to some mathematical criterion (usually the Gini Index) must then be chosen by each decision tree from a random selection of k features at each test node. There are several de-correlated decision trees produced as a result of this random sampling of characteristics.

The normalised total reduction in the mathematical criteria used in the decision of feature of split (Gini Index if the Gini Index is used in the construction of the forest) is computed for each feature during the construction of the forest to perform feature selection using the above forest structure. The Gini Importance of the feature is the name given to this value. The process of feature selection involves ranking each feature according to its Gini Importance, with the user selecting the top k features that appeal to them. [43]

Table 2.3.2: Strengths and Weakness of Machine Learning Algorithms

Technique	Strengths	Weakness
Linear Regression	Simplicity: Linear regression is easy to understand and interpret, making it a popular choice for simple modelling tasks. Linearity: It assumes that the relationship between the independent and dependent variables is linear, which makes it well-suited for modelling linear relationships. Efficiency: Linear regression can provide fast and efficient predictions for large datasets. Interpretability: The coefficients of a linear regression model can be interpreted as the slope of the line and the intercept, making it easy to understand the impact of the independent variables on the dependent variable.	Linearity assumption: Linear regression assumes that the relationship between the dependent and independent variables is linear. If the relationship is non-linear, the model may not fit the data well. Sensitivity to outliers: Linear regression is sensitive to outliers, which can significantly affect the coefficients of the model and the accuracy of the predictions. Multicollinearity: Linear regression assumes that the independent variables are not correlated with each other. If there is multicollinearity, where the independent variables are highly correlated, the model may not accurately estimate the impact of each variable on the dependent variable. Overfitting: Linear regression can easily overfit if the model is too complex or if there is not enough data to support the complexity of the model. Overfitting can lead to poor predictions of new data.
Decision Tree Regressor	Decision Tree Regressor is easy to understand and interpret. It produces a set of rules that can be easily visualized and understood by non-experts. It can handle both categorical and numerical data without the need for data normalization.	Decision Tree Regressor can easily overfit the data if the tree is too deep or if there are too many branches. This can be addressed by setting constraints on the tree's depth or the minimum number of samples required to split an internal node.

	Decision Tree Regressor is a non-parametric algorithm, which means it does not assume any particular distribution for the input variables. This makes it useful for modelling complex relationships between variables. It is computationally efficient and can handle large datasets with high dimensionality. Decision Tree Regressor can handle missing values and outliers.	It is sensitive to small variations in the data, which can lead to different trees being generated for different versions of the same dataset. Decision Tree Regressor is a greedy algorithm that makes locally optimal decisions at each step. This can lead to suboptimal solutions, particularly when dealing with complex relationships between variables. It is not suitable for problems where the output variable has a continuous range of values, as it can only produce a limited number of possible outputs (i.e., the discrete values of the output variable). Decision Tree Regressor can be biased towards features with a large number of levels or values. This can be addressed by using techniques such as random forests or gradient boosting.
Lasso regressor	Feature selection: Lasso performs feature selection by shrinking the coefficients of less important features to zero, making it useful for identifying the most important predictors in a dataset. Reduces overfitting: Lasso includes a regularization parameter that helps to prevent overfitting, making it a useful tool for modelling datasets with high dimensional or noisy data. Interpretable: Lasso produces models with easily interpretable coefficients, which can help to identify which variables are most strongly associated with the target variable. Handles correlated variables well: Lasso is capable of handling highly correlated variables by selecting one variable from each group of correlated variables.	Bias: Lasso introduces bias in the estimates of the coefficients due to the regularization penalty, which can lead to underestimation of the true effect sizes of the predictors. Hyperparameter tuning: Lasso requires careful tuning of the regularization parameter, which can be time-consuming and requires some expertise. May not work well with non-linear data: Lasso assumes a linear relationship between the predictors and the target variable, which may not be appropriate for highly non-linear data. May not work well with large datasets: Lasso can be computationally expensive for large datasets, especially when the number of predictors is high.
Extra Trees Regressor	Extra Trees Regressor is less prone to overfitting than other decision tree-based algorithms, such as the Decision Tree Regressor or the Random Forest Regressor, due to	Extra Trees Regressor can be computationally expensive when the number of trees and input features is large.

	the randomness introduced in the splitting criteria. It can handle both categorical and numerical data without the need for data normalization. Extra Trees Regressor is a non-parametric algorithm, which means it does not assume any particular distribution for the input variables. This makes it useful for modelling complex relationships between variables. It is computationally efficient and can handle large datasets with high dimensionality. Extra Trees Regressor can handle missing values and outliers effectively.	It requires more memory than other decision tree-based algorithms, as it needs to store a large number of trees. Extra Trees Regressor may not perform well when dealing with imbalanced data, meaning some classes or ranges of output variables are underrepresented in the training set. It may not perform well when the input features are highly correlated, as it may not capture the full complexity of the relationship between the input and output variables. Extra Trees Regressor is not as interpretable as other decision tree-based algorithms, such as the Decision Tree Regressor, since it relies on an ensemble of trees.
Gradient Boosting Regressor	Gradient Boosting Regressor can handle both categorical and numerical data without the need for data normalization. It is a powerful algorithm that can model complex relationships between variables and capture non-linear interactions. Gradient Boosting Regressor is an ensemble learning method that combines multiple weak learners (i.e., decision trees) to produce a strong learner. This leads to a more robust and accurate model than using a single decision tree. It can handle missing values and outliers effectively. Gradient Boosting Regressor is less prone to overfitting than other decision tree-based algorithms, such as the Decision Tree Regressor or the Random Forest Regressor, due to the regularization techniques used in the training process.	Gradient boosting cannot totally eliminate them since a dataset may contain several outliers. The gradient boosting classifier accepts the outlying values as well because it has a tendency to repair errors. As a result, it is a memory-intensive technique that is extremely sensitive to outliers in a dataset. The inclination of this approach to correct every error made by preceding nodes leads to overfitting of the model, which is another drawback. Gradient boosting models may require more processing resources and more time to train the entire model on CPUs.
KNeighbors Regressor	KNN Regressor is easy to understand and implement. It can handle both categorical and numerical data without the need for data normalization. KNN Regressor is a non-parametric algorithm, which	KNN Regressor can be computationally expensive, especially when dealing with large datasets or a high number of input features. It requires a large amount of memory to store the training data, as it needs to

	means it does not assume any particular distribution for the input variables. This makes it useful for modelling complex relationships between variables. It is robust to noisy data and can handle outliers effectively. KNN Regressor is versatile and can be used for both simple and complex regression problems.	compare each new data point with all the training examples. KNN Regressor can be sensitive to the choice of the number of neighbours (K) used for prediction, and the distance metric used to measure the similarity between data points. It can be biased towards features with a large number of levels or values. This can be addressed by using techniques such as feature scaling or feature selection. KNN Regressor may not perform well when the training data is imbalanced, meaning some classes or ranges of output variables are underrepresented in the training set.

2.4 Error Measurement

1. MAPE

The accuracy of a forecasting system is measured by the mean absolute percentage error (MAPE), also known as the mean absolute percentage deviation (MAPD). It may be computed as the average absolute per cent inaccuracy for each period with fewer real values divided by actual values. It expresses this accuracy as a percentage.

Since the variable's units are scaled to percentage units, which makes it easier to understand, the mean absolute percentage error (MAPE) is the most often used metric to anticipate error. If the facts are not excessive, it performs best (and no zeros). In regression analysis and model assessment, it is frequently employed as a loss function. [44]

The formula for MAPE is

$$M = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \qquad \text{… Eq. 2.4.1}$$

where:
n is the number of fitted points,
A_t is the actual value,
F_t is the forecast value.
Σ is summation notation (the absolute value is summed for every forecasted point in time).

Percentage errors are calculated in terms of absolute errors, without regard to sign. This avoids the problem of positive and negative errors cancelling each other out. [44]

2. R2

R-squared (R2) is a statistical measure that shows how much of a dependent variable's variance is explained by one or more independent variables in a regression model. R-squared measures how well the variation of one variable accounts for the variance of the second, as opposed to

correlation, which describes the strength of the relationship between independent and dependent variables. Therefore, if a model's R2 is 0.50, its inputs can account for around half of the observed variation.

The formula for R^2

$$R^2 = 1 - \frac{Unexplained\ Variation}{Total\ Variation} \quad \text{… Eq. 2.4.2}$$

3. MSE

The MSE evaluates the performance of either an estimator or a predictor (a function that maps arbitrary inputs to a sample of values of a random variable) (i.e., a mathematical function mapping a sample of data to an estimate of a parameter of the population from which the data is sampled). Depending on whether one is describing a predictor or an estimator, the definition of MSE changes. [44]

The Formula for MSE

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 \quad \text{… Eq. 2.4.3}$$

where,

MSE = Mean Squared Error

N = Number of Data Points

Y_i = Observed Values

$\hat{Y}_i$ = Predicted values

4. NRMSE

The normalized root mean squared error (NRMSE), also called a scatter index, is a statistical error indicator defined as

$$NRMSE = \frac{\sum(S_i - O_i)^2}{\sum Oi^2} \quad \text{… Eq. 2.4.4}$$

Where, O_i is observed values and Si is simulated values. It can also be calculated as RMSE/range or RMSE/mean. [44]

5. RMSE

One of the methods most frequently used to assess the accuracy of forecasts is root mean square error, also known as root mean square deviation. It illustrates the Euclidean distance between measured true values and forecasts.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \hat{x}_i)^2}{N}} \quad \text{… Eq. 2.4.5}$$

where ,

RMSD = Root Mean Squared Error

N = Number of Data points

x_i = actual observations

$\hat{x}_i$ = estimated observations

3. PARAMETER PREDICTION

Data Description

It contains measurements for 4 months (2016-09-01 to 2016-12-31).

The dataset includes observations of:

- Solar Irradiance (W/m2)
- Temperature (°F)
- Barometric Pressure (Hg)
- Humidity (%)
- Wind Direction (°)
- Wind Speed (mph)
- Sun Rise/Set Time

	UNIXTime	Data	Time	Radiation	Temperature	Pressure	Humidity	WindDirection(Degrees)	Speed	TimeSunRise	TimeSunSet
0	1475229326	9/29/2016 12:00:00 AM	23:55:26	1.21	48	30.46	59	177.39	5.62	06:13:00	18:13:00
1	1475229023	9/29/2016 12:00:00 AM	23:50:23	1.21	48	30.46	58	176.78	3.37	06:13:00	18:13:00
2	1475228726	9/29/2016 12:00:00 AM	23:45:26	1.23	48	30.46	57	158.75	3.37	06:13:00	18:13:00
3	1475228421	9/29/2016 12:00:00 AM	23:40:21	1.21	48	30.46	60	137.71	3.37	06:13:00	18:13:00
4	1475228124	9/29/2016 12:00:00 AM	23:35:24	1.17	48	30.46	62	104.95	5.62	06:13:00	18:13:00

Fig. 3.1: Parameters Data

Heatmap

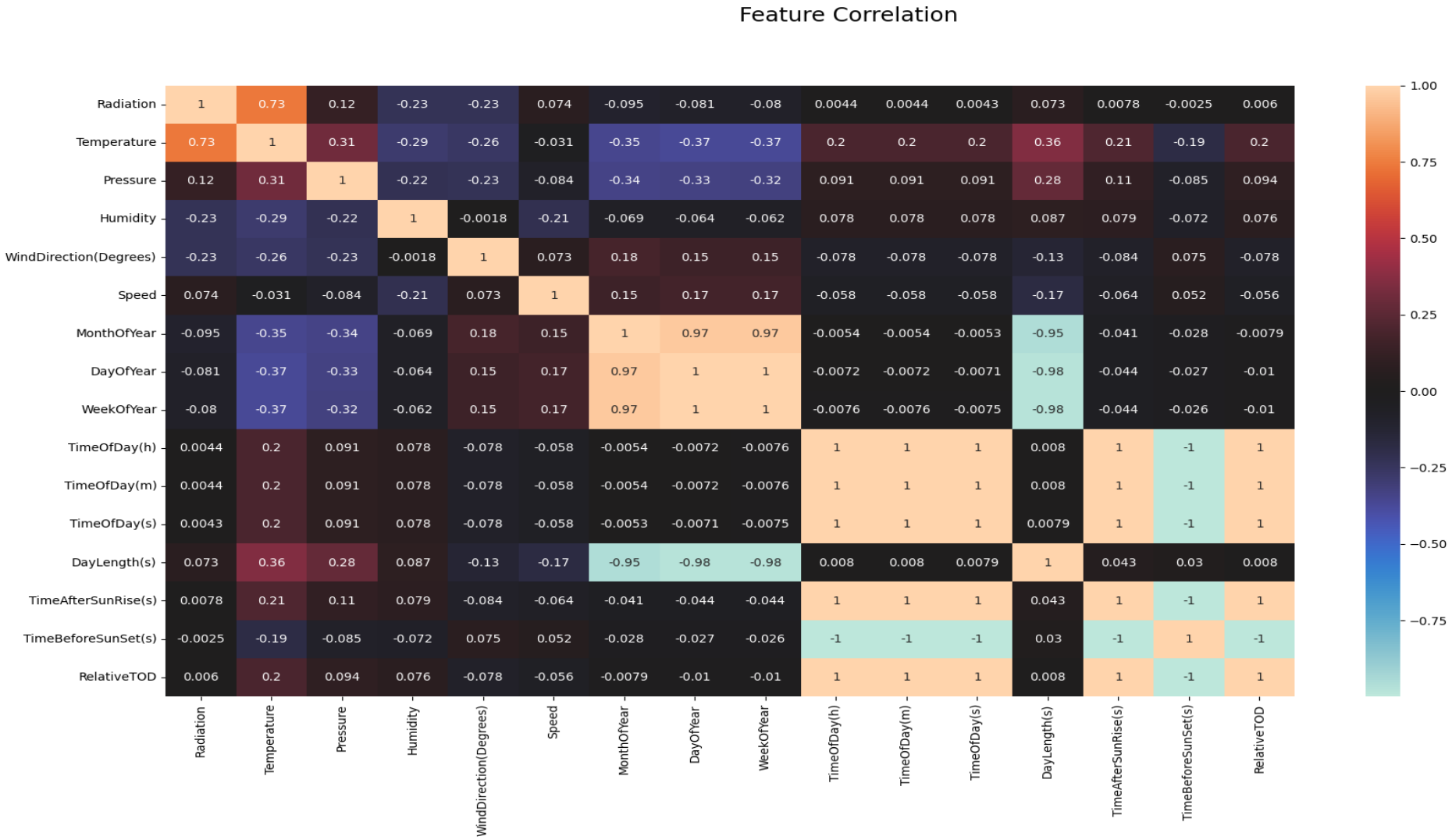


Fig. 3.2: Heatmap

Boxplots

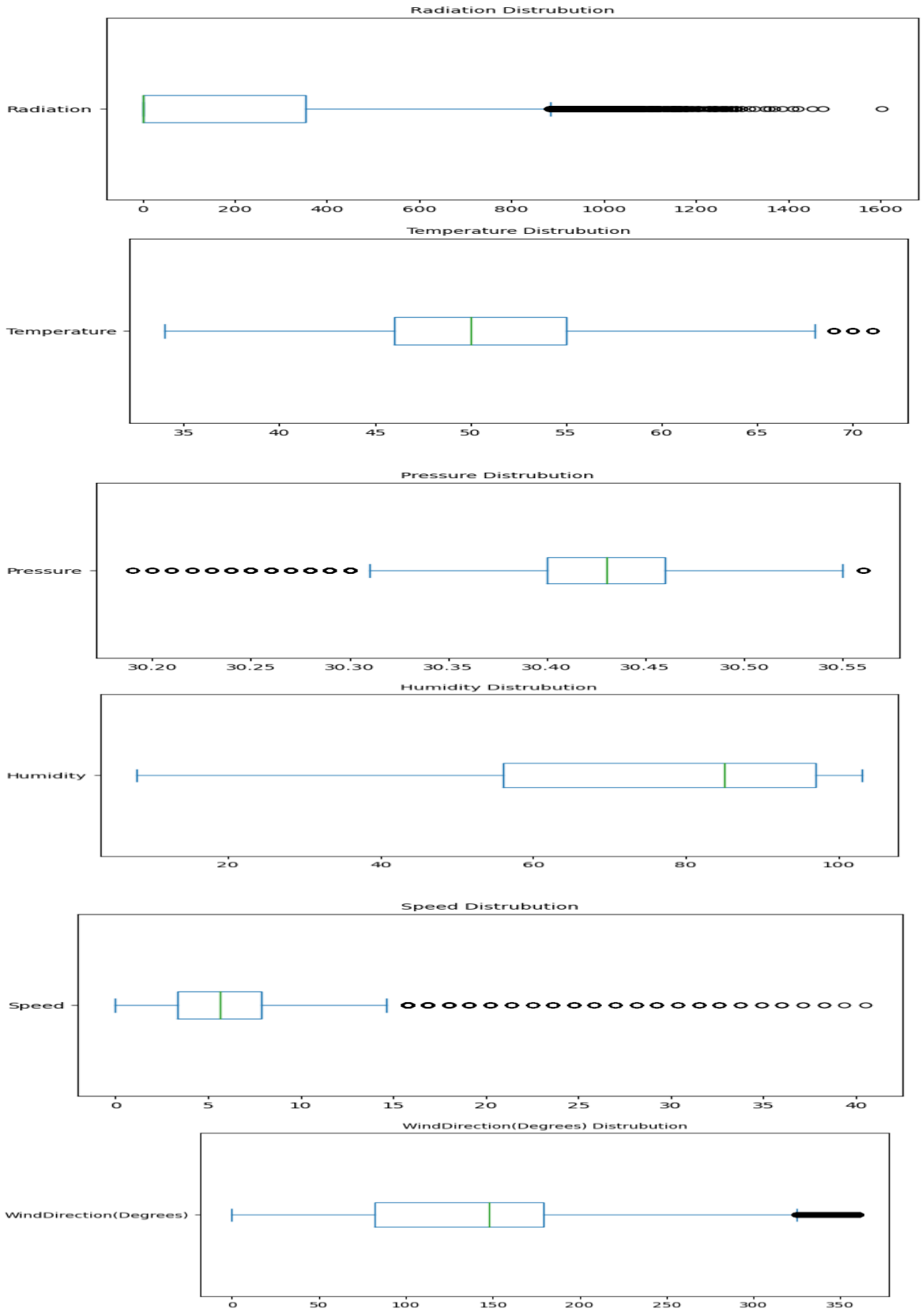


Fig. 3.3: Boxplots

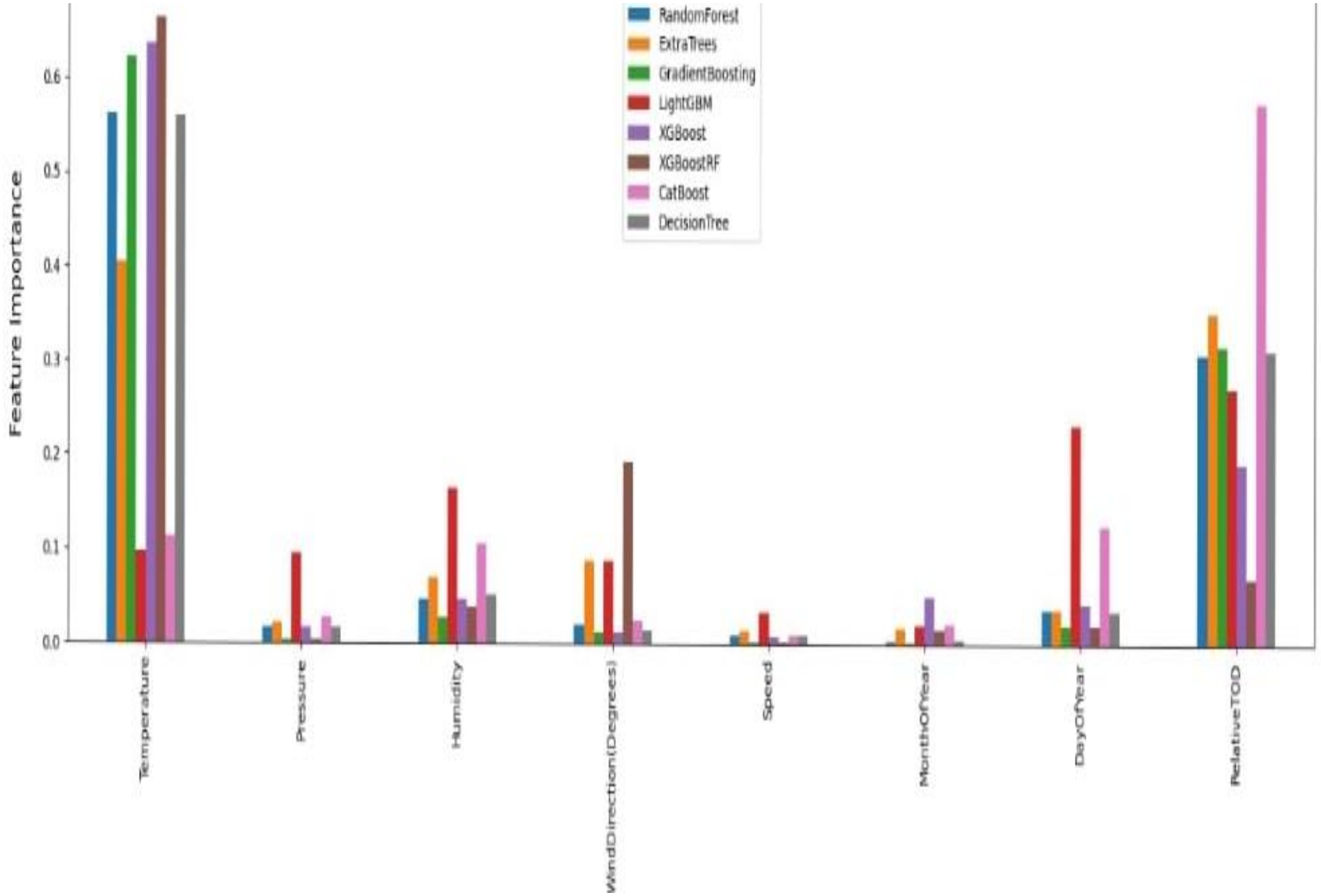


Fig. 3.4: Comparison of Parameters based on Techniques

4. PREDICTION FOR SOLAR ENERGY

4.1 Introduction to solar energy

4.1.1 Scenario

4.1.1.1 Global

Solar PV generation increased by a record 179 TWh (up 22%) in 2021 to exceed 1,000 TWh. Of all renewable technologies, it showed the second-largest absolute generation growth in 2021, right behind wind. In much of the world, solar PV is quickly becoming the most affordable alternative for new power generation, which is anticipated to spur investment in the years to come. However, to adhere to the Net Zero Emissions by 2050 Scenario, annual generation growth must average 25% from 2022 to 2030. This translates to a more than threefold increase in yearly capacity deployment through 2030, necessitating far more ambitious policy goals and more work from both public and commercial players, particularly in the areas of grid integration, policy mitigation, regulation and financing challenges. This is particularly the case in emerging and developing countries.

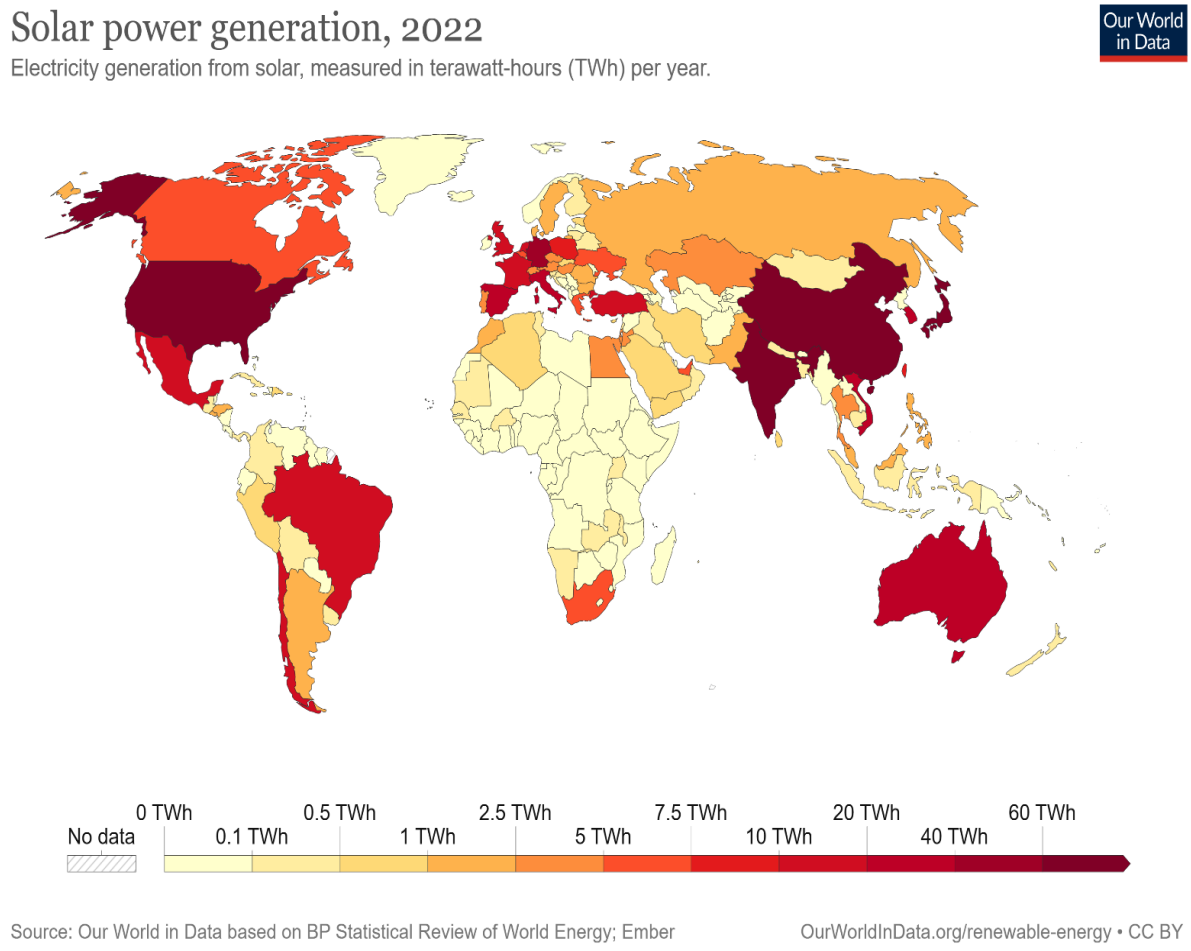


Fig. 4.1.1.1.1: Solar Power Generation, 2022

As a result of significant capacity increases in 2020 and 2021, China accounted for nearly 38% of the growth in solar PV power in 2021. The United States saw the second-largest generational growth (17% of the total), while the European Union saw the third-largest rise (10%). Despite the interruptions caused by COVID-19, supply chain delays, and increases in commodity prices that were encountered in 2021, solar PV managed to accomplish another record-breaking year of capacity growth (almost 190 GW). This should therefore result in a further acceleration in the rise of power generation in 2022.

However, increasing the annual average generation from the current 1000 TWh to a level of approximately
7400 TWh in 2030, which aligns with the Net Zero Scenario, requires annual average generation growth of about 25% during 2022–2030. Although this rate is comparable to the average annual growth seen over the previous five years, as the PV market expands, more work will be needed to keep this momentum going.
In 2021, utility-scale plants added 52% of the world's solar PV capacity, followed by residential (28%) and commercial and industrial (9%). The proportion of utility-scale plants decreased to its lowest level since 2012 as record-high distributed PV capacity additions in China, the US, and the EU in 2020–2021 was fueled by significant policy incentives. [45]

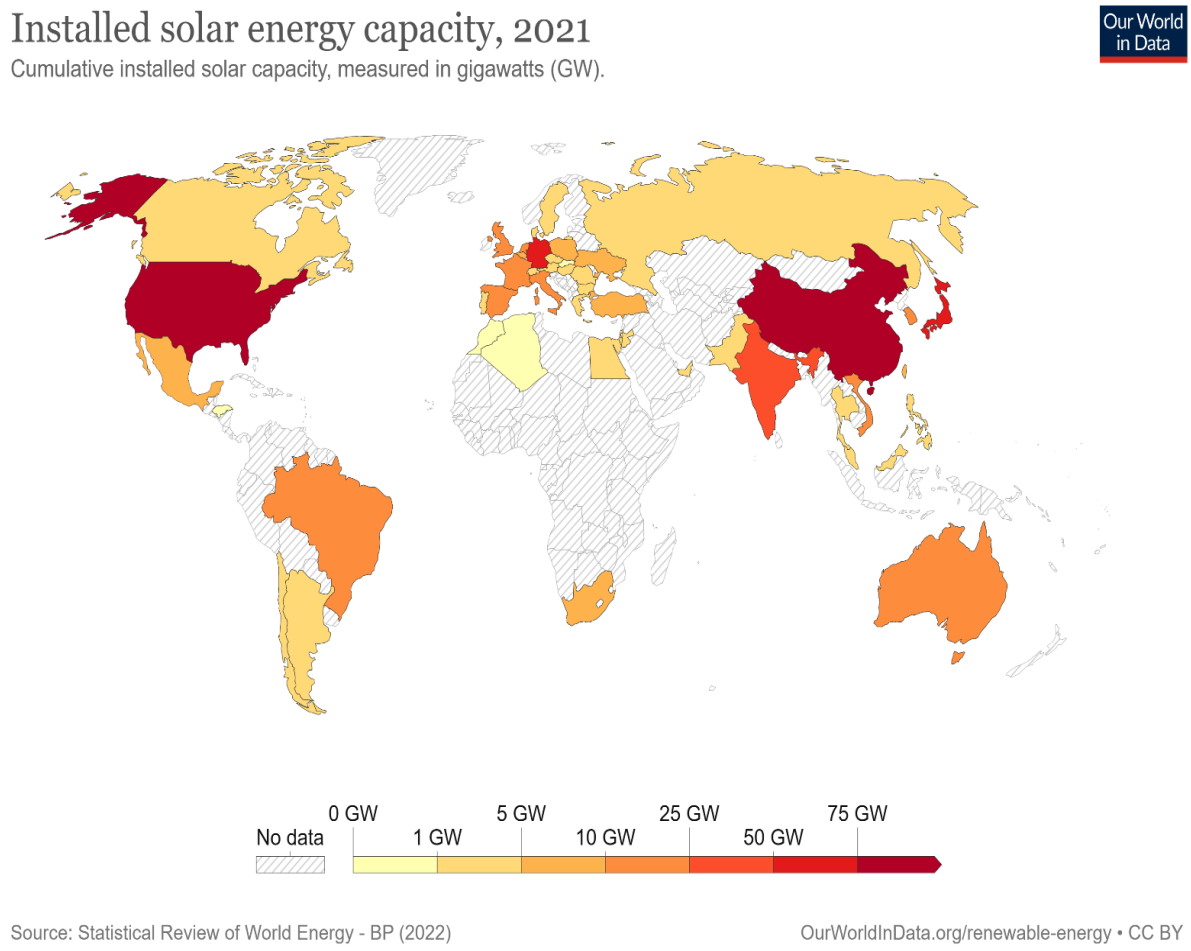


Fig. 4.1.1.1.2: Installed Solar Energy Capacity, 2021

4.1.1.2 India

India has a huge potential for solar energy. India's geographical surface receives around 5,000 trillion kWh of incident energy annually, with the majority of areas getting 4–7 kWh per square metre every day. If Solar energy can be successfully harvested in India, there could have an enormous scalability opportunity. Additionally, solar energy offers the option of distributed power generation and permits quick capacity expansion with minimal lead times. From the standpoint of a rural application, off-grid decentralised and low-temperature applications will be beneficial for addressing other energy demands for electricity, heating, and cooling in both rural and urban locations. Solar is the most secure source of energy from a security of supply standpoint since it is widely accessible.

The country's solar potential is estimated by the National Institute of Solar Energy to be 748 GW, assuming that solar PV modules will cover 3% of the wasteland area. The National Solar Mission is one of the main Missions of India's National Action Plan on Climate Change, which has given solar energy a prominent role. On January 11th, 2010, the National Solar Mission (NSM) was launched. The Government of India launched the National Solar Mission (NSM) as a significant project to promote ecological sustainability and solve the country's energy security issues. Additionally, India will make a significant contribution to the global effort to address the challenges of climate change.

This is consistent with India's Intended Nationally Determined Contributions (INDCs) goal to attain around 40% of cumulative installed capacity for electric power from non-fossil fuel sources by 2030.[46]

Table 4.1.1.2.1: Installed Capacity of Renewable Sources Of Energy In India (in GW)

Solar	Wind	Large Hydro	Biopower	Nuclear	Small Hydro
48.55	40.03	46.51	10.62	6.78	4.82

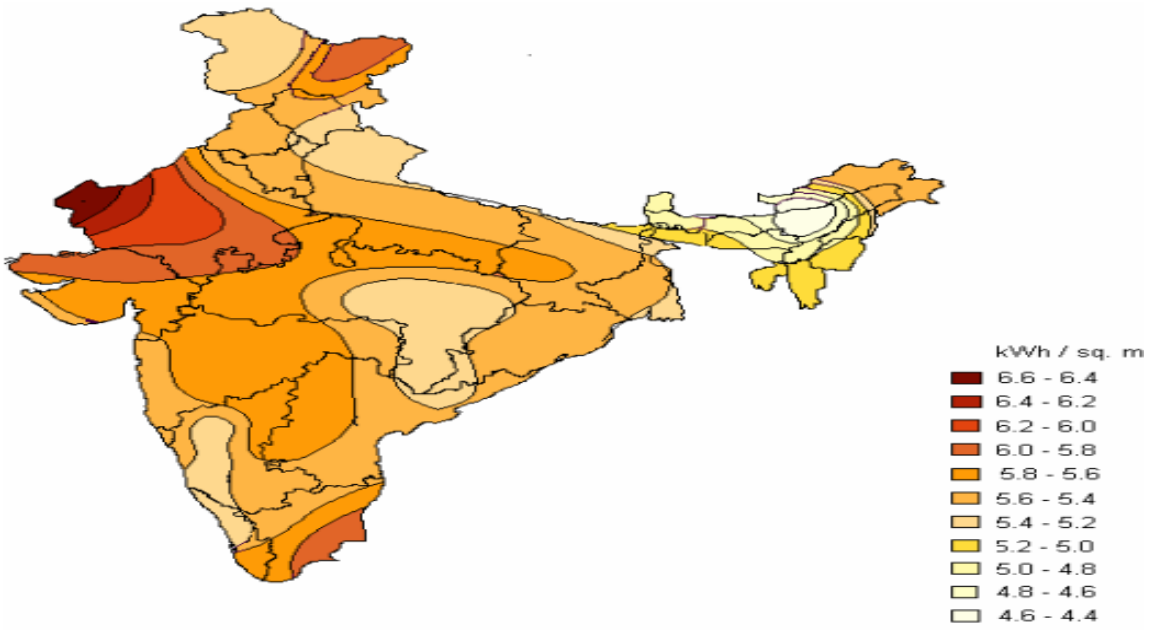


Fig. 4.1.1.2.1: Installed Capacity of Renewable Sources Of Energy In India (in GW)

4.2 Production process

Photovoltaic (PV) cells, sometimes referred to as solar cells, are devices used to produce electricity using solar radiation. A solar cell is a structure built of semiconductor materials, such as silicon, that collects photons from the sun and transforms them into electrons that may flow as a current.

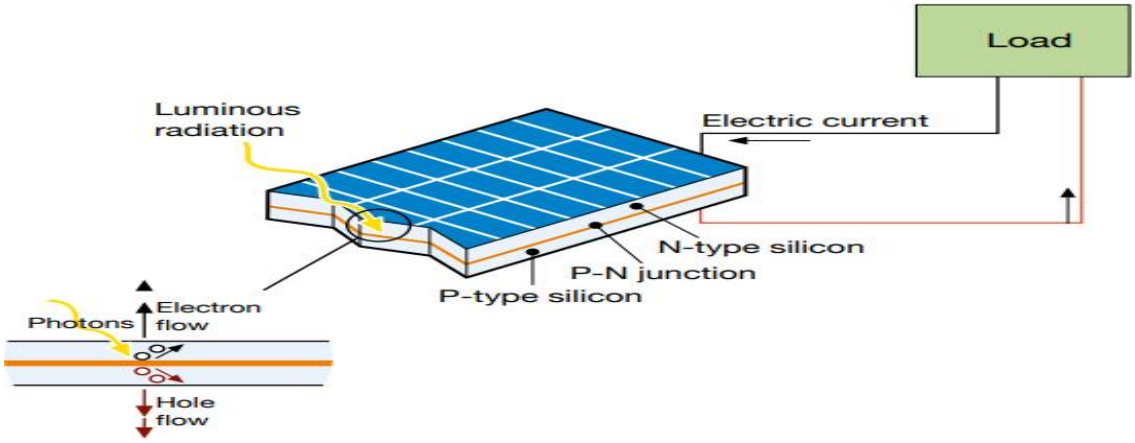


Fig. 4.2.1: PV Energy Production

When light strikes a solar cell, the excited electrons leave the semiconductor material and move to a conductive material, like a metal wire. Electric current, which may be utilised to power electrical appliances, is created by this movement of electrons.[47]

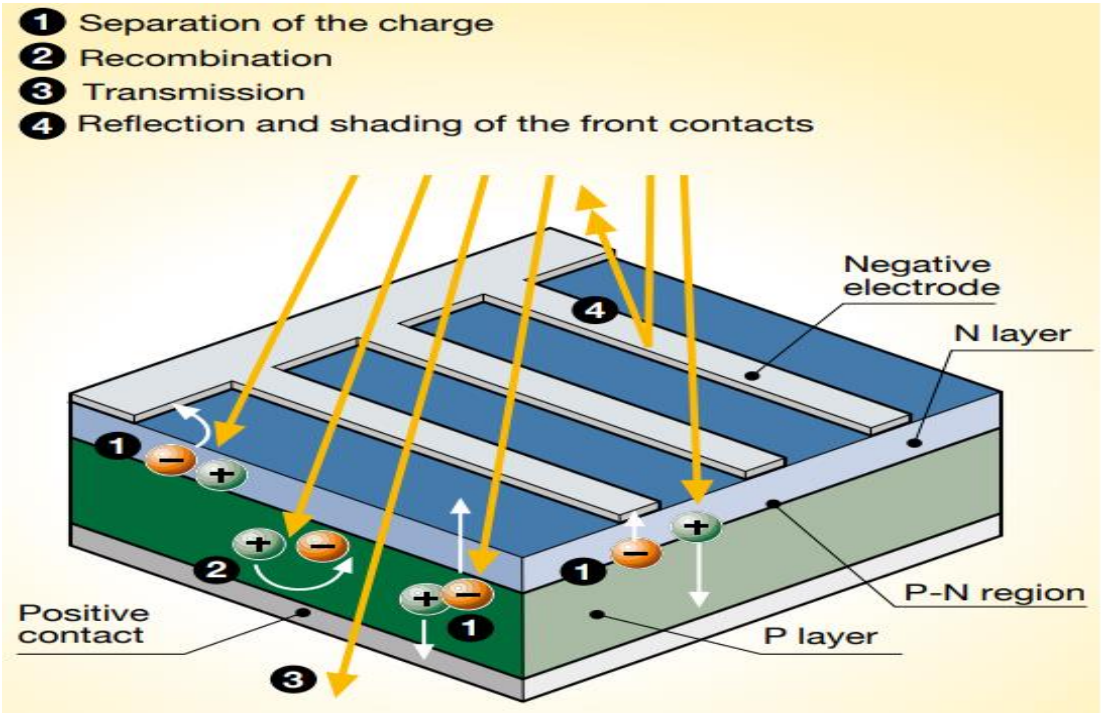


Fig. 4.2.2: Production Of Electric Current

Numerous solar cells are connected to form modules, which make up solar panels. To create a solar array, which can produce more power, many modules can be joined. To power buildings,

commercial establishments, and other electrical equipment, the solar panels' electricity can either be immediately supplied to the electrical grid or stored in batteries.[47]

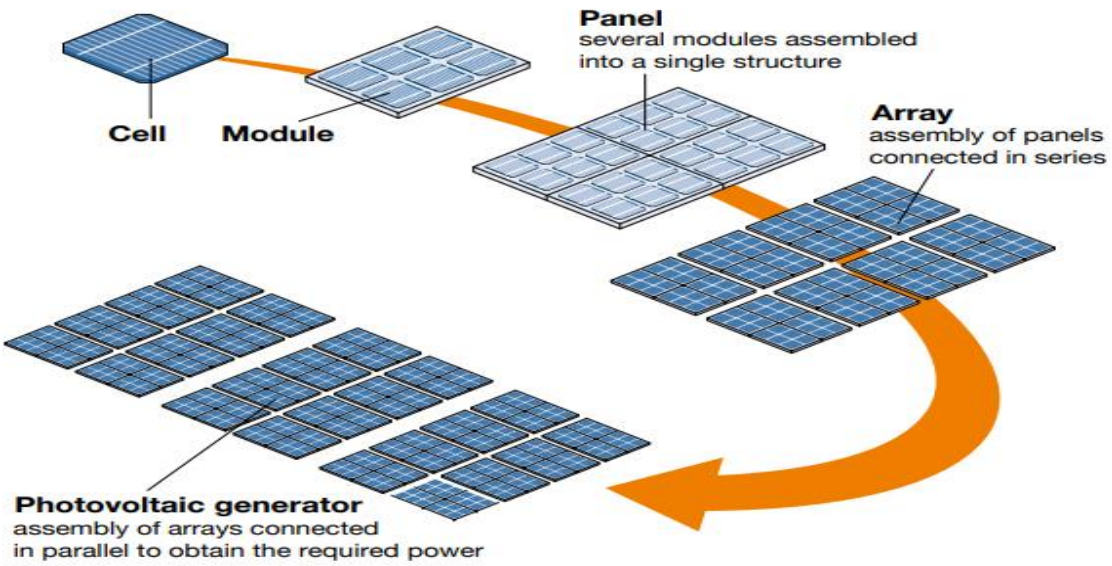


Fig. 4.2.3: PV Generators

In general, solar energy is a clean and sustainable energy source that can be utilised to create power without emitting harmful gases or depleting scarce resources.[47]

4.3 Prediction of Output Power

Description of Data

Generation Data

	DATE_TIME	PLANT_ID	SOURCE_KEY	DC_POWER	AC_POWER	DAILY_YIELD	TOTAL_YIELD
0	2020-05-15 00:00:00	4136001	4UPUqMRk7TRMgml	0.0	0.0	9425.000000	2.429011e+06
1	2020-05-15 00:00:00	4136001	81aHJ1q11NBPMrL	0.0	0.0	0.000000	1.215279e+09
2	2020-05-15 00:00:00	4136001	9kRcWv60rDACzjR	0.0	0.0	3075.333333	2.247720e+09
3	2020-05-15 00:00:00	4136001	Et9kgGMDl729KT4	0.0	0.0	269.933333	1.704250e+06
4	2020-05-15 00:00:00	4136001	IQ2d7wF4YD8zU1Q	0.0	0.0	3177.000000	1.994153e+07

Fig. 4.3.1: Data Generation

	PLANT_ID	DC_POWER	AC_POWER	DAILY_YIELD	TOTAL_YIELD
count	67698.0	67698.000000	67698.000000	67698.000000	6.769800e+04
mean	4136001.0	246.701961	241.277825	3294.890295	6.589448e+08
std	0.0	370.569597	362.112118	2919.448386	7.296678e+08
min	4136001.0	0.000000	0.000000	0.000000	0.000000e+00
25%	4136001.0	0.000000	0.000000	272.750000	1.996494e+07
50%	4136001.0	0.000000	0.000000	2911.000000	2.826276e+08
75%	4136001.0	446.591667	438.215000	5534.000000	1.348495e+09
max	4136001.0	1420.933333	1385.420000	9873.000000	2.247916e+09

Fig. 4.3.2: Data Generation Mean, Standards, Minimum And Maximum

Weather Data

	DATE_TIME	PLANT_ID	SOURCE_KEY	AMBIENT_TEMPERATURE	MODULE_TEMPERATURE	IRRADIATION
0	2020-05-15 00:00:00	4136001	iq8k7ZNt4Mwm3w0	27.004764	25.060789	0.0
1	2020-05-15 00:15:00	4136001	iq8k7ZNt4Mwm3w0	26.880811	24.421869	0.0
2	2020-05-15 00:30:00	4136001	iq8k7ZNt4Mwm3w0	26.682055	24.427290	0.0
3	2020-05-15 00:45:00	4136001	iq8k7ZNt4Mwm3w0	26.500589	24.420678	0.0
4	2020-05-15 01:00:00	4136001	iq8k7ZNt4Mwm3w0	26.596148	25.088210	0.0

Fig. 4.3.3: Weather Data

	PLANT_ID	AMBIENT_TEMPERATURE	MODULE_TEMPERATURE	IRRADIATION
count	3259.0	3259.000000	3259.000000	3259.000000
mean	4136001.0	28.069400	32.772408	0.232737
std	0.0	4.061556	11.344034	0.312693
min	4136001.0	20.942385	20.265123	0.000000
25%	4136001.0	24.602135	23.716881	0.000000
50%	4136001.0	26.981263	27.534606	0.019040
75%	4136001.0	31.056757	40.480653	0.438717
max	4136001.0	39.181638	66.635953	1.098766

Fig. 4.3.4: Weather Data - Mean, Standards, Minimum And Maximum

Pair Plots

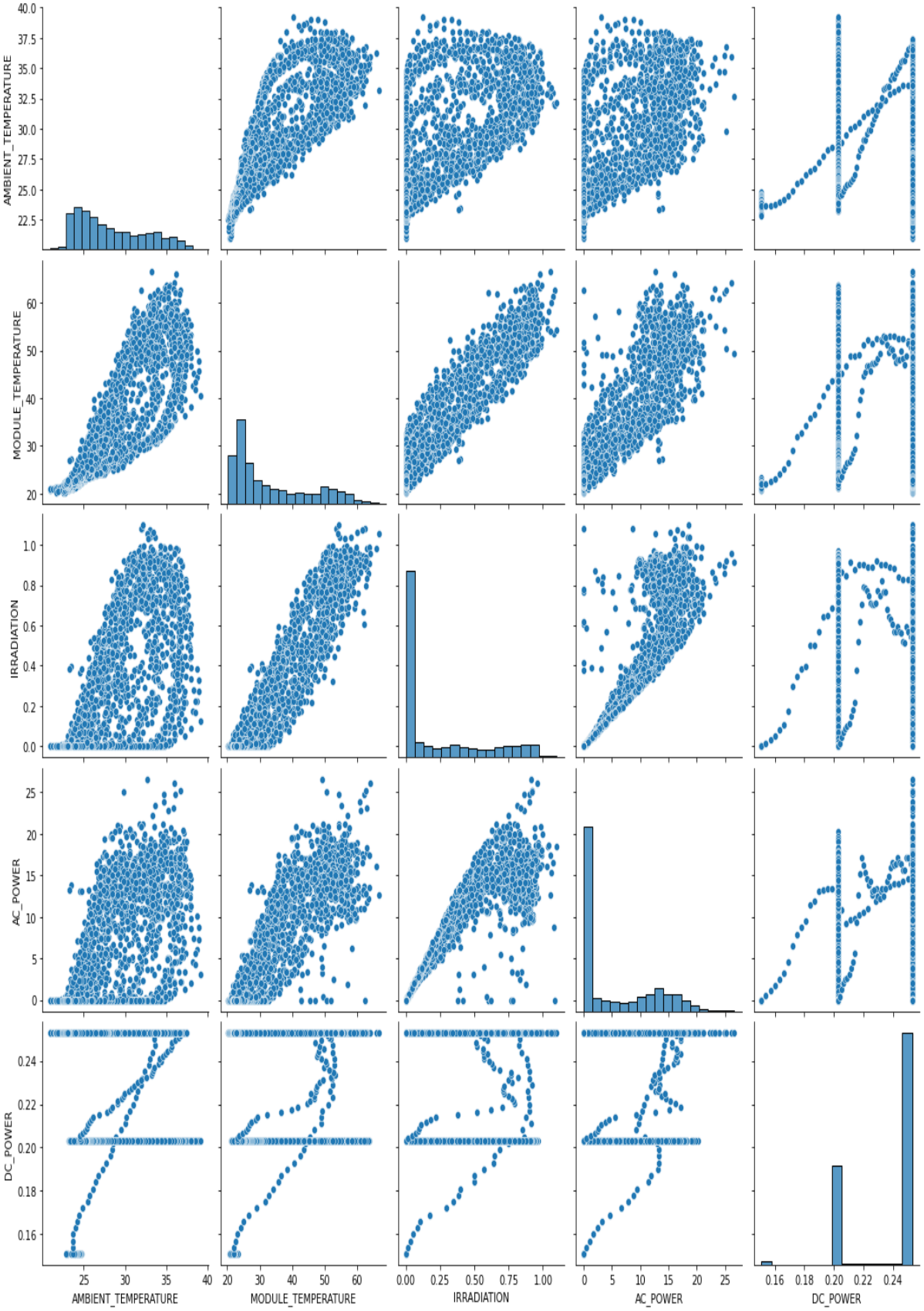


Fig. 4.3.5: Pair Plots

Boxplots

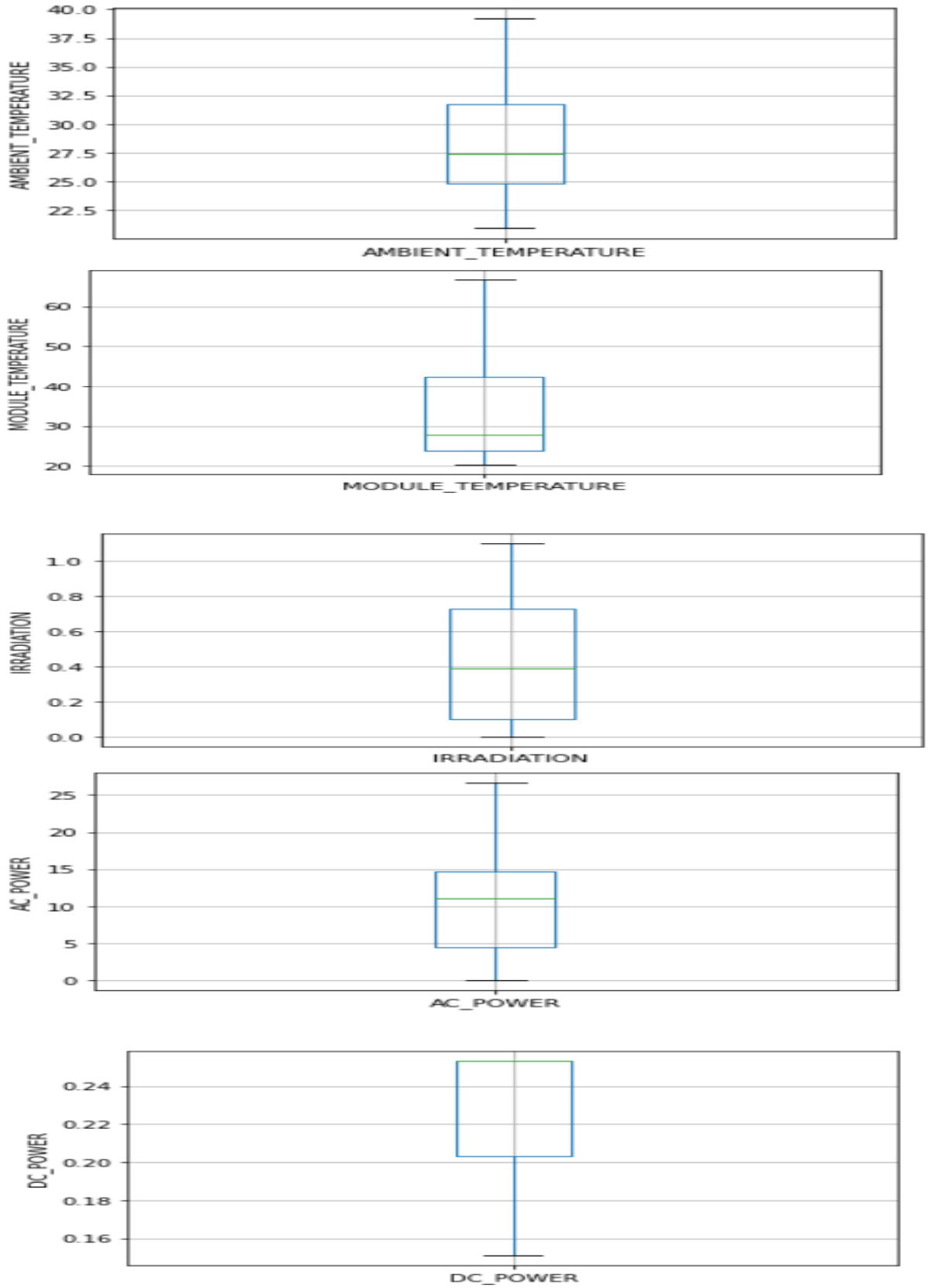


Fig. 4.3.6: Boxplots

Heatmap

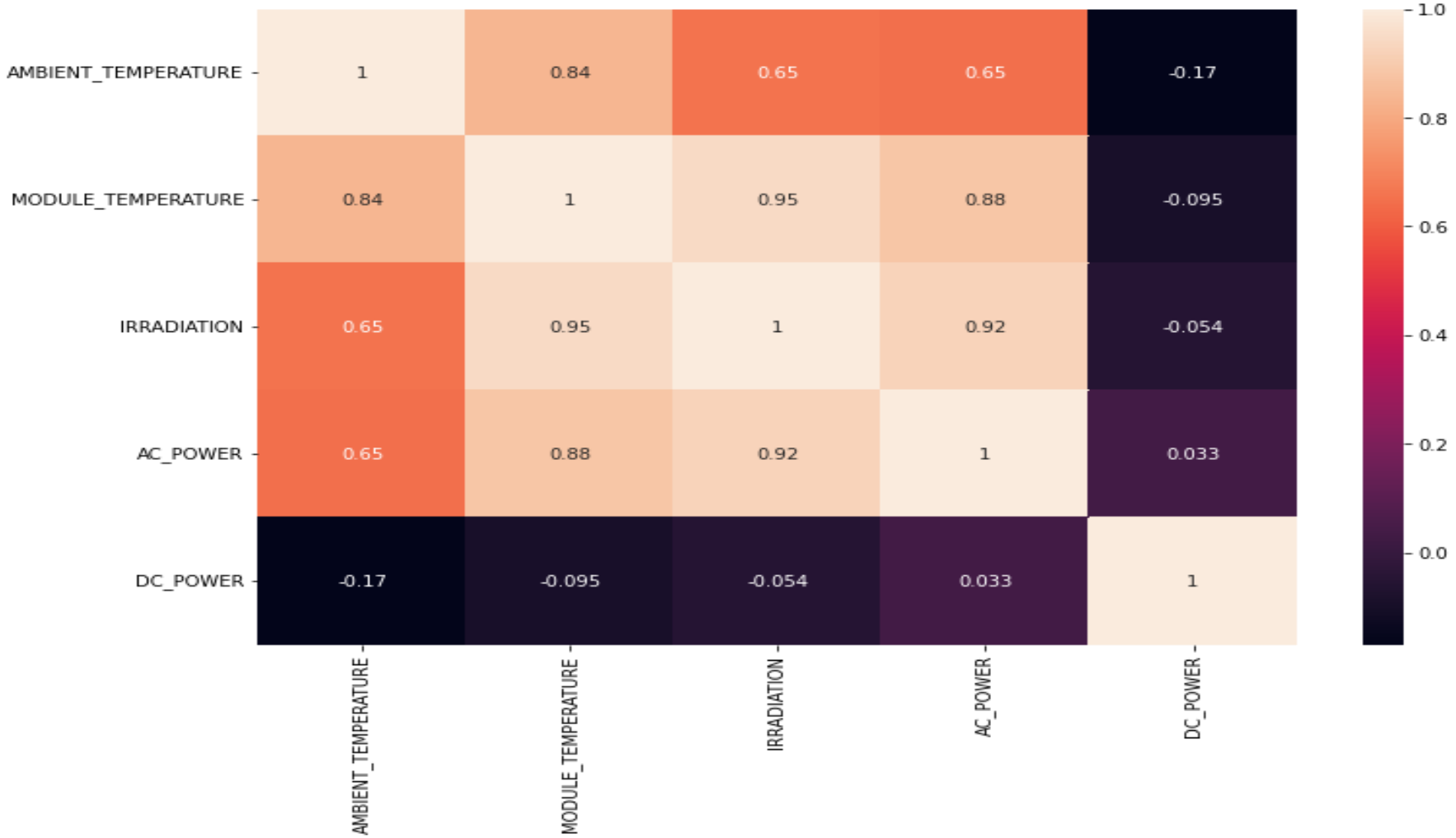


Fig. 4.3.7: Heatmap

5. PREDICTION FOR WIND

5.1 Introduction to wind energy

5.1.1 Scenario

5.1.1.1 Global

The production of energy from wind grew by a record 273 TWh in 2021 (up 17%). This growth rate was the highest among all renewable power technologies and was 55% higher than what was achieved in 2020. This exceptional growth in wind capacity additions, which reached 113 GW in 2020 compared to just 59 GW in 2019, allowed for such quick expansion. However, to meet the requirements of the Net Zero Emissions by 2050 Scenario, which calls for approximately 7900 TWh of wind electricity generation in 2030, average annual capacity additions must be increased to almost 250 GW, which is more than twice the record growth of 2020.

Nearly 70% of the increase in wind generation in 2021 was attributed to China, with the United States coming in second at 14% and Brazil at 7%. Despite near-record capacity increases in 2020 and 2021, the European Union experienced a 3% decline in wind power production in 2021 as a result of abnormally prolonged spells of low wind conditions. Due to regulatory deadlines in China and the US, capacity growth in 2020 increased by 90% and reached 113 GW, enabling record generation growth globally. [48]

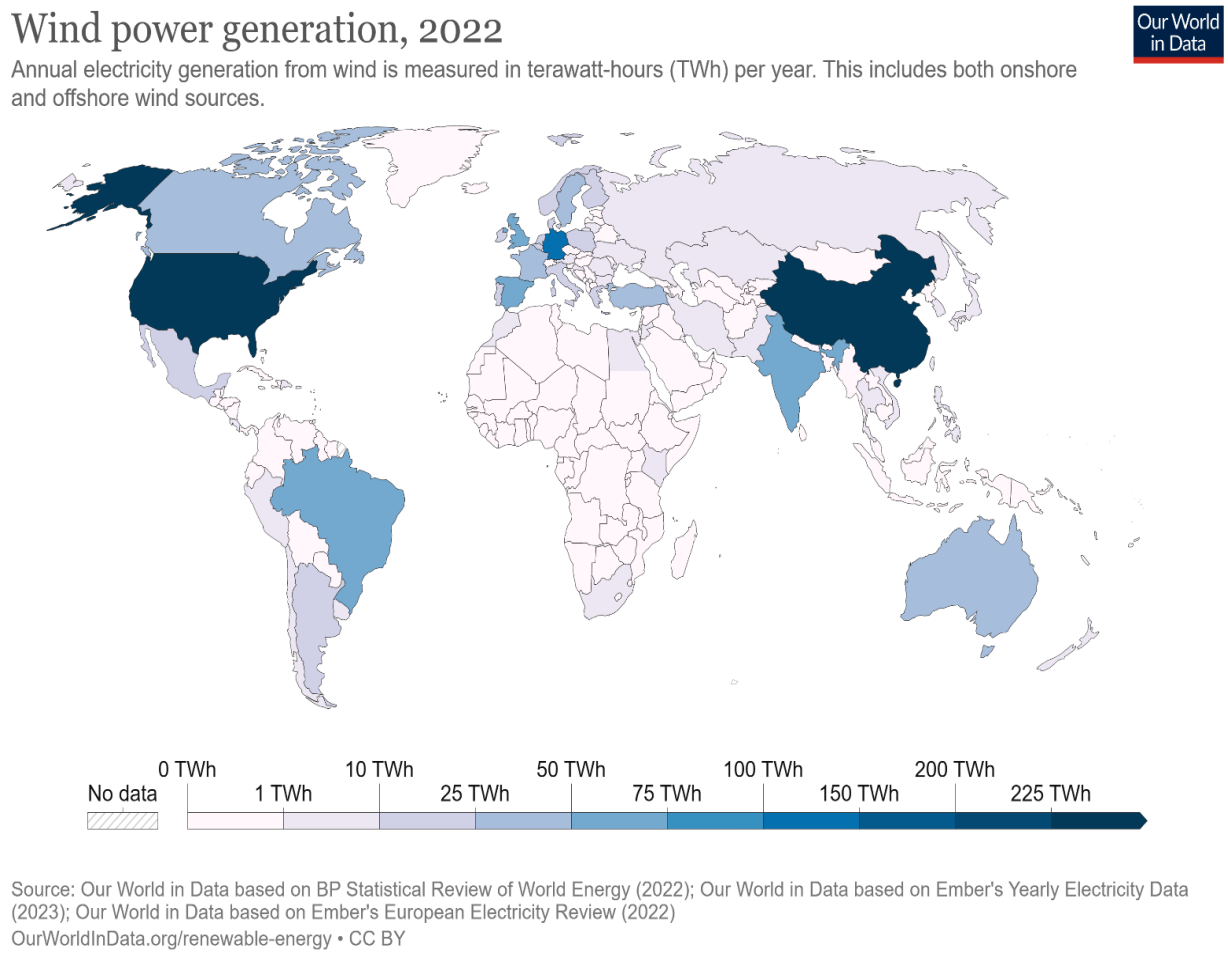


Fig. 5.1.1.1.1: Wind Power Generation

To achieve this level of sustained capacity growth, much more work is required. Making it easier to obtain onshore wind permitting and lowering the cost of offshore wind are the two most crucial areas for improvement. [48]

Wind power generation must increase on average by roughly 18% annually between 2022 and 2030 to match the Net Zero Scenario's 2030 wind power output level of over 7,900 TWh. The deployment is anticipated to stabilise after the unusually large capacity additions in 2020–2021, underscoring the necessity for significant efforts to move towards the Net Zero Scenario trajectory.

93% of the built wind power capacity in 2021—a total of 830 GW—was from onshore systems, while the remaining 7% came from offshore wind farms. While offshore wind is still in the early stages of expansion, with capacity only present in 19 countries, onshore wind is a mature technology that is used in 115 countries worldwide. [48]

However, as more nations prepare to build their first offshore wind farms, the offshore reach is anticipated to grow over the next few years.

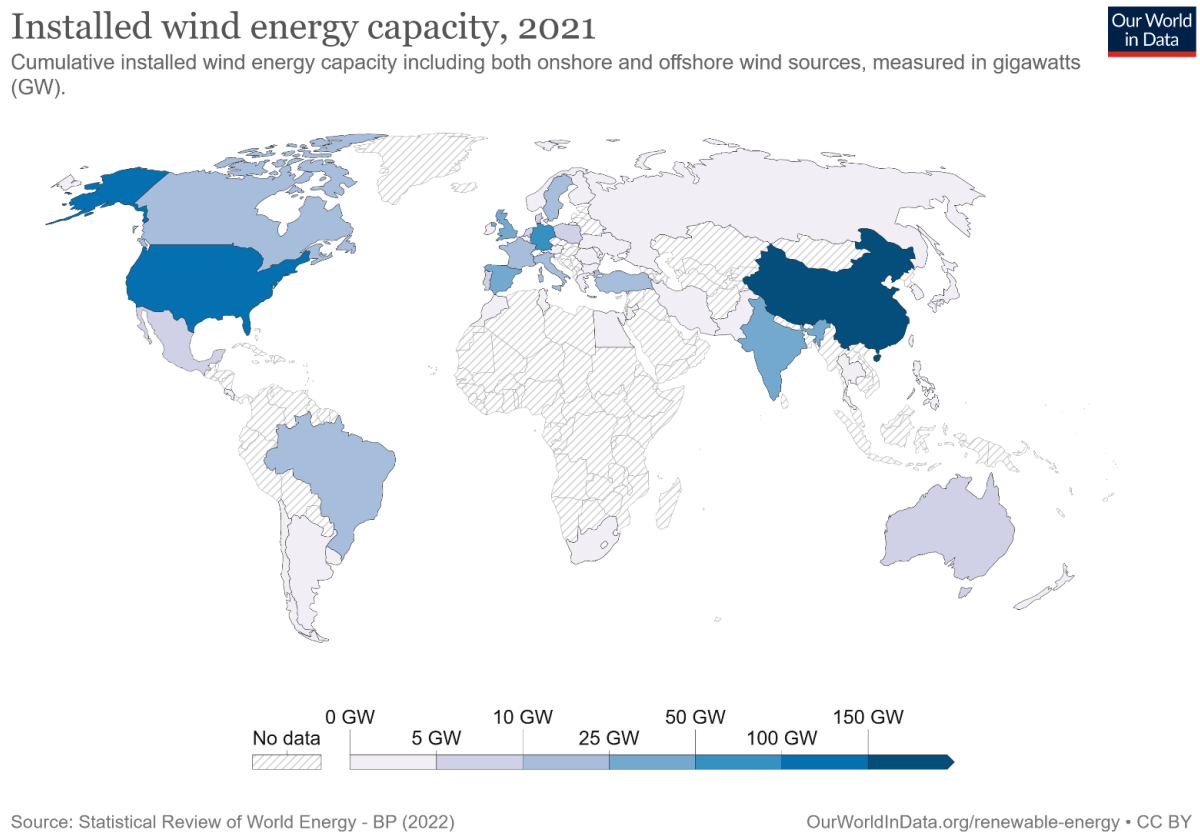


Fig. 5.1.1.1.2: Installed Wind Energy Capacity

In 2021, offshore technology contributed around 22% of the 94 GW increase in overall wind capacity, the greatest percentage ever and three times the average over the preceding five years. A slowdown in global onshore growth combined with record offshore capacity additions in China, which accounted for 80% of offshore growth, led to such a high percentage.

Onshore wind capacity expansion is anticipated to continue at a steady rate in the upcoming years, while offshore systems are anticipated to expand much more quickly in both their current

markets—the European Union and China—and new ones—the United States, Chinese Taipei, and Japan.

To generate approximately 8,000 TWh of yearly wind energy predicted by the Net Zero Scenario, both onshore and offshore farms will need more support. Focused efforts should be made to speed up permits, assist in finding suitable locations, cut costs, and shorten project development times.

By creating turbines with longer blades and taller towers, innovation in the onshore wind is aimed at boosting the technology's productivity, particularly in locations with little wind. However, the maximum height of onshore wind turbines is frequently limited in some areas for reasons related to the environment and public acceptance, which restricts the innovation potential.

Contrarily, there is no such constraint on turbine size in the offshore wind sector, therefore research is concentrated on creating bigger turbines, which enable decreases in the overall cost of power generation. Parallel to this, the advancement of cost-effective and secure offshore floating wind turbines is quickening.

A major instrument for the energy transition in nations like Japan, Korea, Portugal, France, and the West Coast of the United States, floating wind farms might unlock the enormous potential of ocean regions with sea depths too deep for fixed turbines.

The most crucial areas for improvement are streamlining onshore wind permits and lowering the cost of offshore wind to accomplish this level of continuous capacity expansion. [48]

5.1.1.2 India

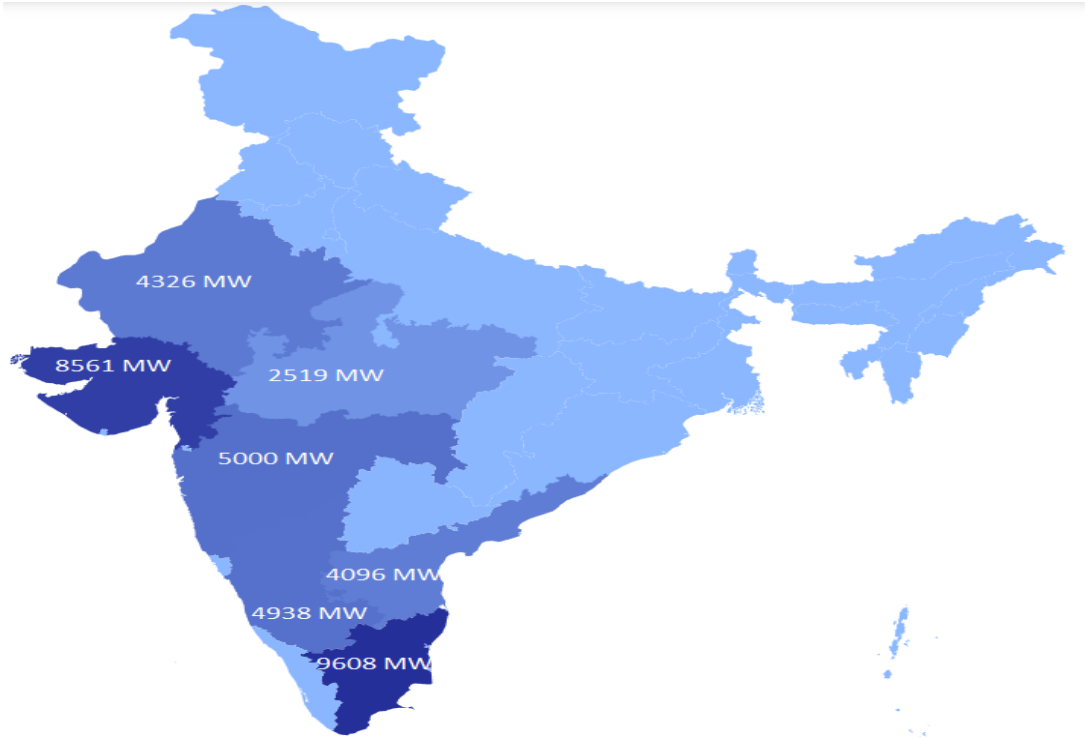


Fig. 5.1.1.2.1: India's Wind Energy Production Capacity

The domestic wind energy business in India is the one driving the sector's continuous growth. A robust ecology, project operating capabilities, and an annual generating base of roughly

10,000 MW (MW) have all been made possible by the growth of the wind sector. With a total installed capacity of 35.6 GW (as of 31 March 2019), the nation now has the fourth-largest wind capacity in the world and produced over 52.66 billion units in 2017–18.

By offering different monetary and financial incentives including accelerated depreciation advantages, concessional custom duty exemption on specific components of wind power generators, etc., the government is encouraging private sector investment in wind power projects all around the nation.
A thorough study of the wind resource is necessary before choosing possible sites because the wind is a site-specific, intermittent energy source. The government has set up more than 800 wind monitoring stations at heights of 50, 80, and 100 metres around the nation through the National Institute of Wind Energy (NIWE) and has distributed wind potential maps. The country has a 100 m above-ground capacity for 302 GW of gross wind energy, according to recent studies. [49]

5.2 Production Process

Wind turbines are used to convert wind energy into electricity. The kinetic energy of the wind is captured by a wind turbine and transformed into mechanical energy, which is utilised to produce electricity.

A wind turbine's fundamental parts are its blades, rotor, shaft, generator, and tower. The rotor revolves as a result of the blades being forced to rotate by the wind. The generator, which rotates to generate power, is linked to the rotor by a shaft.[50]

The size of the turbine and the wind speed are two elements that affect how much power a wind turbine can produce. To produce more power, wind turbines can be placed one at a time or in big clusters known as wind farms. The rotor blades of a wind turbine, which function similarly to an aeroplane wing or a helicopter rotor blade, convert wind energy into electricity using aerodynamic force. The air pressure on one side of the blade falls as the wind passes across it. Both lift and drag are produced by the different air pressure on the blade's two sides. The rotor spins because the force of the lift is greater than the force of the drag. If the generator is a direct drive turbine, the rotor is connected to it directly; otherwise, a gearbox that speeds up the rotation and permits a physically smaller generator is used. Electricity is produced as a result of the conversion of aerodynamic force into generator rotation.

Wind turbine electricity may be used to power buildings, commercial establishments, and other electrical equipment by being supplied directly to the electrical grid. It may also be used to power off-grid systems, isolated places, or battery storage.[50]

5.3 Prediction of Output Power

Description of Data

	Date/Time	LV ActivePower (kW)	Wind Speed (m/s)	Theoretical_Power_Curve (KWh)	Wind Direction (°)
0	01 01 2018 00:00	380.047791	5.311336	416.328908	259.994904
1	01 01 2018 00:10	453.769196	5.672167	519.917511	268.641113
2	01 01 2018 00:20	306.376587	5.216037	390.900016	272.564789
3	01 01 2018 00:30	419.645904	5.659674	516.127569	271.258087
4	01 01 2018 00:40	380.650696	5.577941	491.702972	265.674286
...	...	...	...	...	...
50525	31 12 2018 23:10	2963.980957	11.404030	3397.190793	80.502724
50526	31 12 2018 23:20	1684.353027	7.332648	1173.055771	84.062599
50527	31 12 2018 23:30	2201.106934	8.435358	1788.284755	84.742500
50528	31 12 2018 23:40	2515.694092	9.421366	2418.382503	84.297913
50529	31 12 2018 23:50	2820.466064	9.979332	2779.184096	82.274620

Fig. 5.3.1: Data Description

	LV ActivePower (kW)	Wind Speed (m/s)	Theoretical_Power_Curve (KWh)	Wind Direction (°)
count	50530.000000	50530.000000	50530.000000	50530.000000
mean	1307.684332	7.557952	1492.175463	123.687559
std	1312.459242	4.227166	1368.018238	93.443736
min	-2.471405	0.000000	0.000000	0.000000
25%	50.677890	4.201395	161.328167	49.315437
50%	825.838074	7.104594	1063.776282	73.712978
75%	2482.507568	10.300020	2964.972462	201.696720
max	3618.732910	25.206011	3600.000000	359.997589

Fig. 5.3.2: Data - Mean, Standards, And Minimum And Maximum

Pair Plot

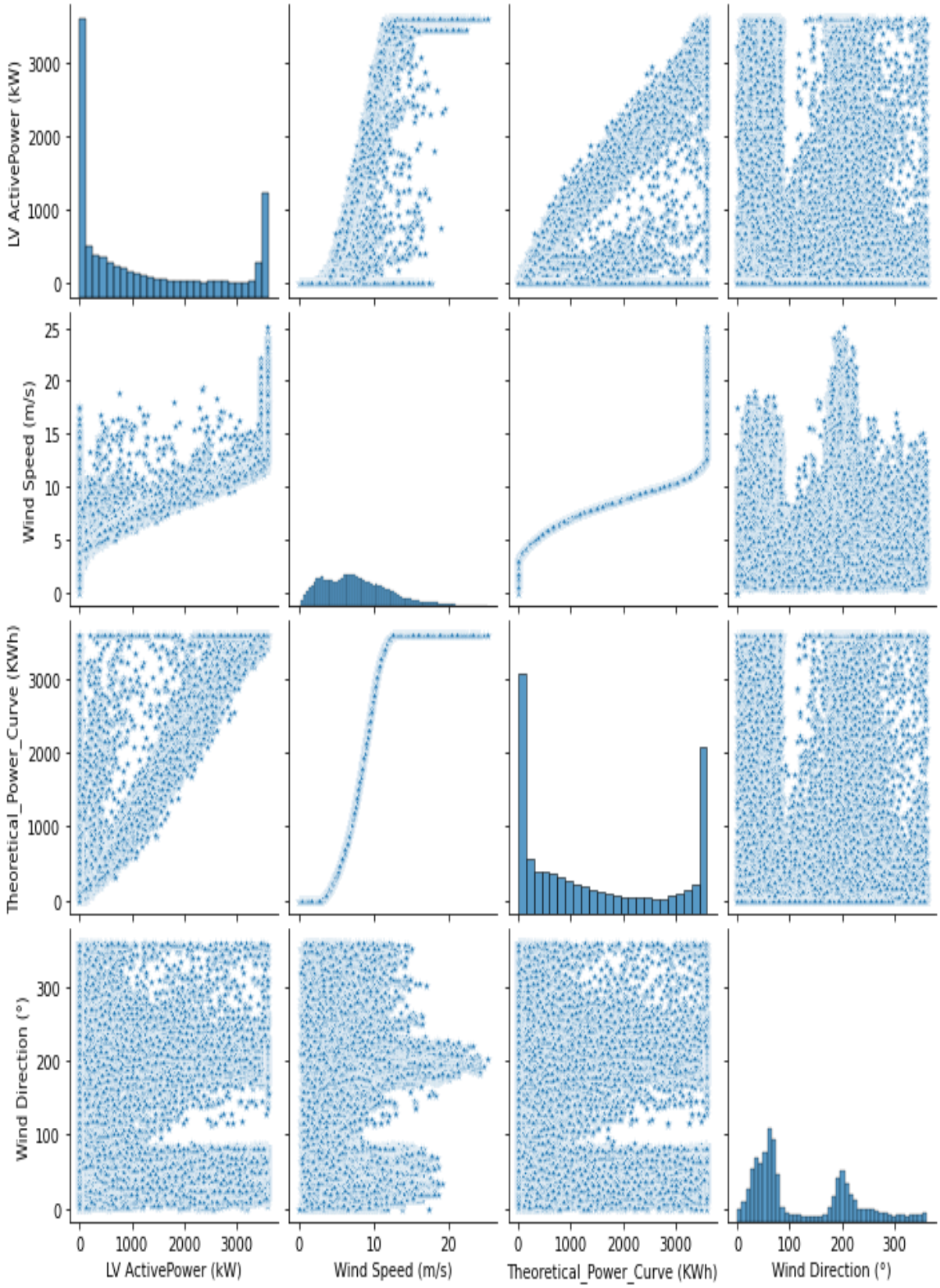


Fig. 5.3.3: Pair Plot

Boxplots

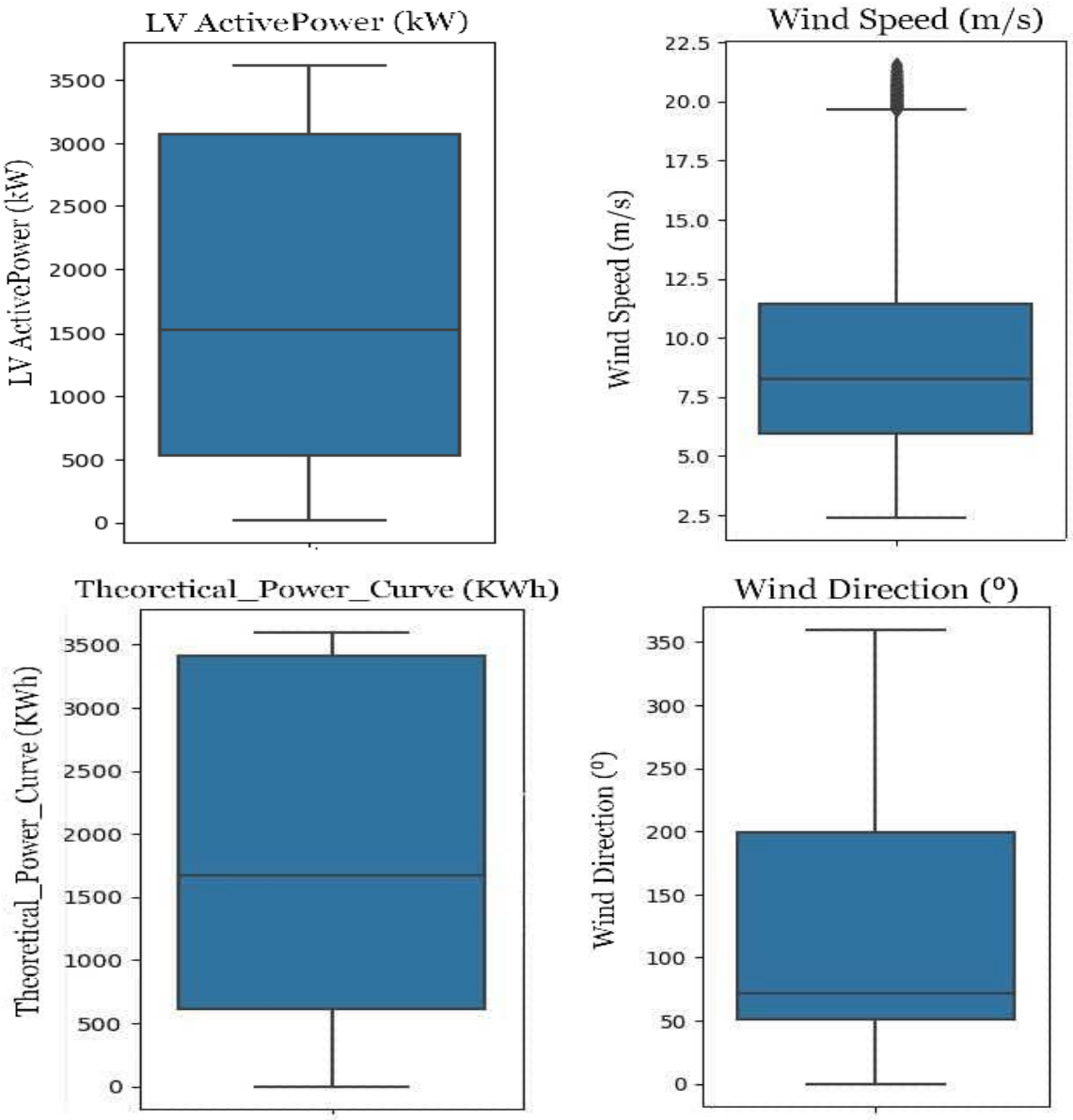


Fig. 5.3.4: Boxplot

6. COMPARISON AND ANALYSIS

6.1 Presentation of the Results Obtained from Different Prediction Methods

The results obtained from the three prediction methods are presented in tables and graphs to facilitate their comparison. The testing dataset is used to evaluate the performance of the methods.

The figures below show the Variation in actual and predicted values for Output Power for Solar (in MW) and Wind (in KW) respectively.

For Solar

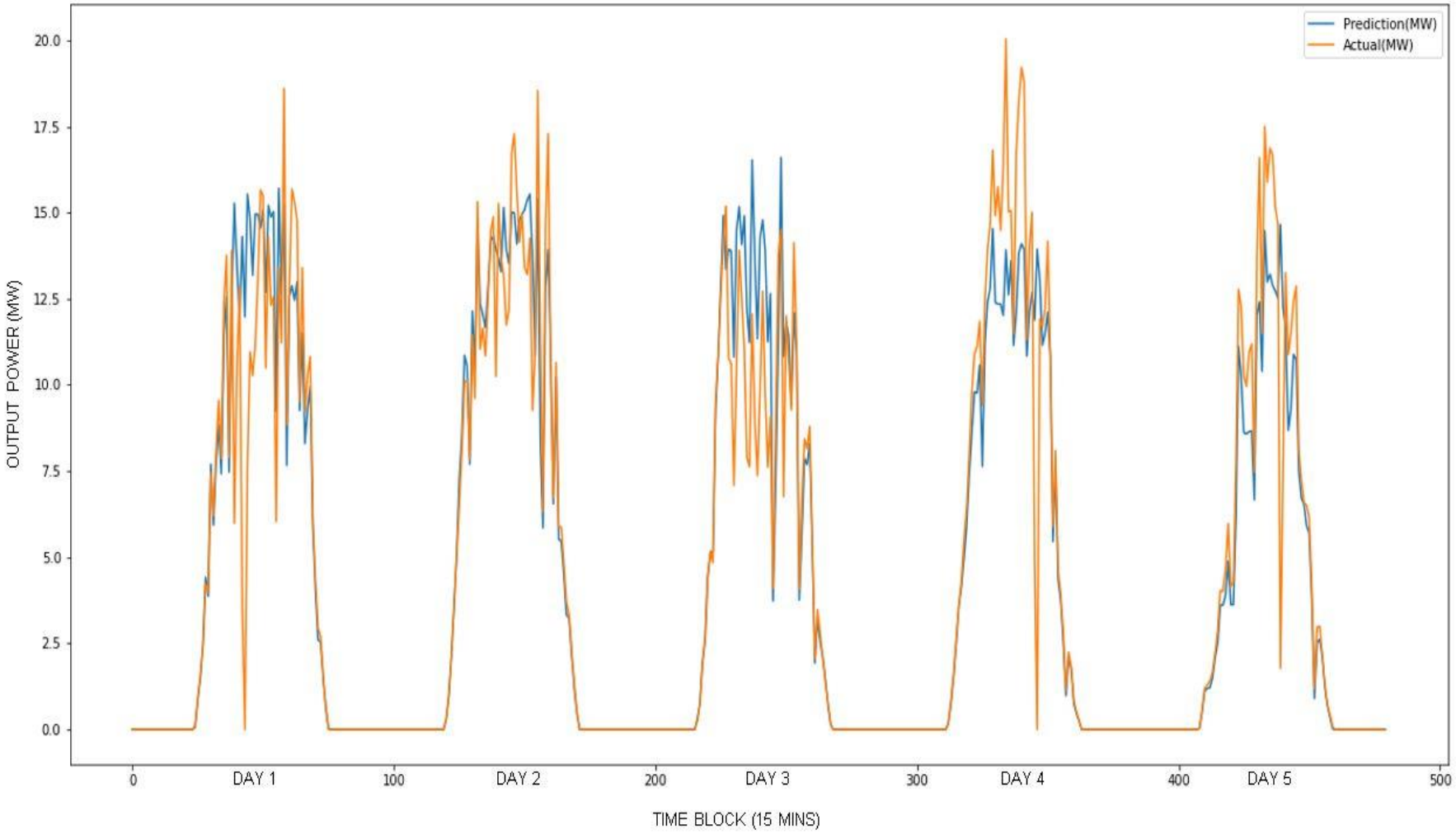


Fig. 6.1.1: Actual vs Predicted Values of Output Power

For Wind

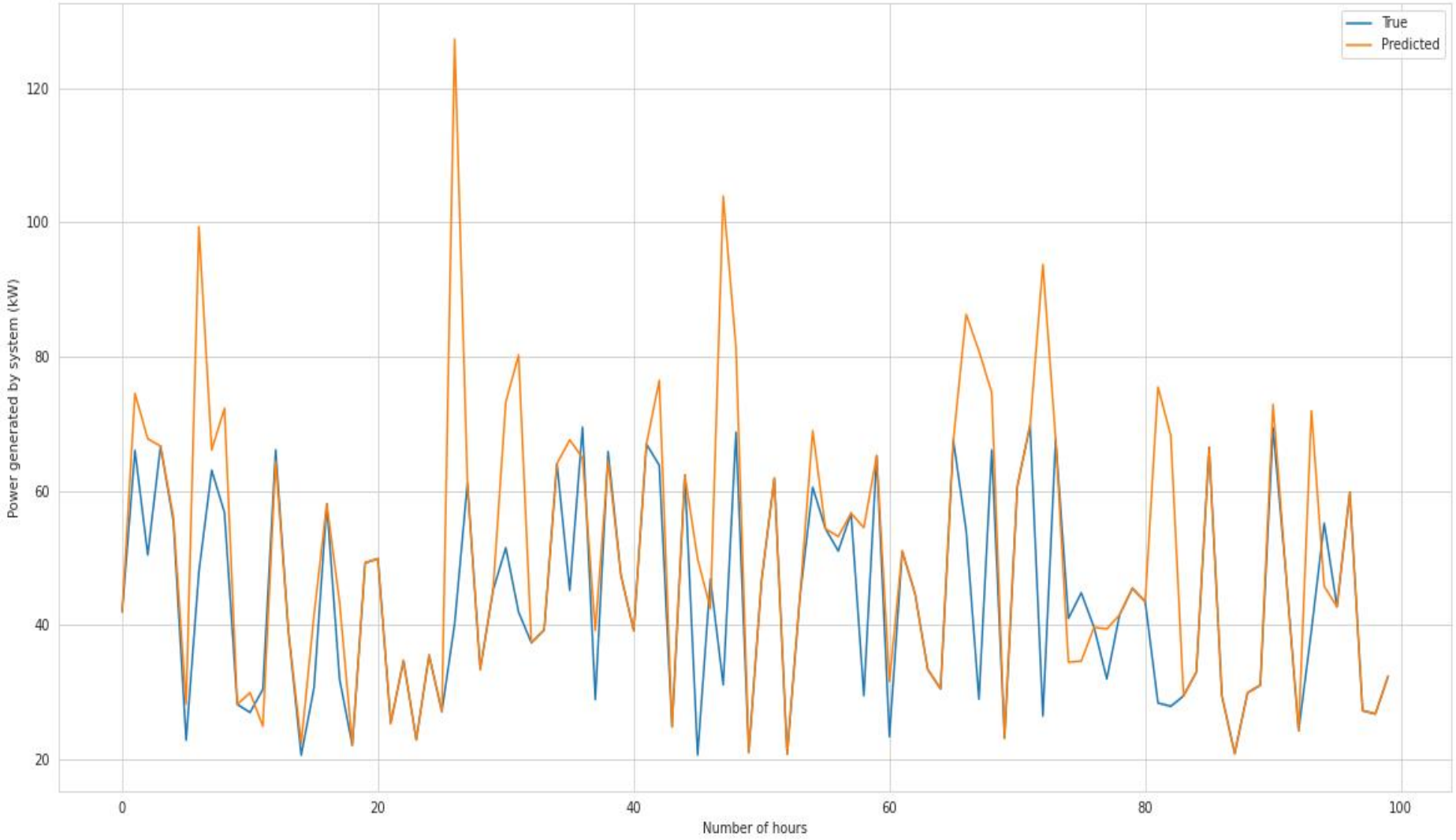


Fig. 6.1.2: Actual vs Predicted Values of Output Power

6.2 Comparison of Performance of the Prediction Methods Using Evaluation Criteria

For solar

Table 6.2.1: Comparison of Prediction Methods for Solar Energy

Technique	Mean Validation RMSE (%)
Linear Regression	2.4302998881730917
Decision Tree Regressor	2.3328394586081282
Random Forest Regressor	1.757966671219207
Ridge Regressor	2.43388927538925
Lasso Regressor	2.8508597011572148

For wind

Table 6.2.2: Comparison of Prediction Methods for Wind Energy

Technique	Mean Validation RMSE (%)
KNN	1.27622976983972
Decision Tree Regressor	1.4963184284627
Extra Trees Regressor	1.2205605621235
Random Forest Regressor	1.1624225595648
Gradient Boosting Regressor	1.1476226247425

7 CONCLUSION

Aiming to compare the performance of different techniques for solar and wind energy output forecasting, the study used historical data on solar and wind energy output to train and test the prediction models and evaluated the performance of each technique using various evaluation criteria.

It found that the accuracy of the predictions was influenced by several factors, including the amount and quality of the data used to train the models, the choice of input variables, and the evaluation criteria used to assess the performance of the models.

For power networks to effectively include variable renewable energy (VRE) sources like wind and solar, forecasting is a critical component.
Forecast time horizon, local weather conditions (which affect VRE resource variability), geographic scope, data availability (e.g., plant size, location, components), and data quality (e.g., consistency, precision, resolution) are all factors that have an impact on forecast performance. At shorter time intervals, forecast accuracy typically rises.
Frequent predictions, however, are only helpful, provided the time increments correspond to the periods in which system operators can take appropriate action. By adapting their methods to take into consideration regional circumstances and system operator requirements, practitioners can reduce forecast mistakes.

The project highlights the importance of accurate predictions for the effective integration of solar and wind energy into power grids, which can help policymakers and energy planners to make informed decisions and prioritize investments in new technologies that can improve the reliability and efficiency of renewable energy systems. The use of artificial neural networks and machine learning techniques can significantly improve the accuracy of predictions and can help to optimize the use of renewable energy sources in the power grid. Accurate and reliable predictions of energy output are critical for it.

With more accurate predictions, power grid operators can better manage the variability and uncertainty of solar and wind energy and can optimize the use of these renewable sources.

Artificial neural networks and machine learning techniques provide the most accurate predictions that can guide the development of better prediction models for solar and wind energy output.

Moreover, the use of these prediction models can enable better energy management, improve energy efficiency, and reduce greenhouse gas emissions, thus contributing to a more sustainable energy future. Overall, the study provides valuable insights into the best prediction methods for achieving this goal.

8 REFERENCES

[1] Shinn, L. (n.d.). Renewable Energy: The Clean Facts. NRDC. https://www.nrdc.org/stories/renewable-energy-clean-facts

[2] Renewable Capacity Statistics 2022. (2022, April 1). https://www.irena.org/publications/2022/Apr/Renewable-Capacity-Statistics-2022

[3] S. (n.d.). Global Solar Atlas. The World Bank Group. https://globalsolaratlas.info/map?c=61.438767,-57.187518,2andm=site

[4] Electricity - Fuels and Technologies - IEA. (n.d.). IEA. https://www.iea.org/fuels-and-technologies/electricity

[5] Annual CO_2 emissions by world region. (n.d.). Our World in Data. https://ourworldindata.org/grapher/annual-co-emissions-by-region

[6] Yang, B., Guo, Z., Yang, Y., Chen, Y., Zhang, R., Su, K., et al. (2021). Extreme Learning Machine Based Meta-Heuristic Algorithms for Parameter Extraction of Solid Oxide Fuel Cells. Appl. Energ. 303, 117630. https://doi.org/10.1016/j.apenergy.2021.117630

[7] Erdiwansyah, Mahidin, Husin, H. et al. A critical review of the integration of renewable energy sources with various technologies. Prot Control Mod Power Syst 6, 3 (2021). https://doi.org/10.1186/s41601-021-00181-3

[8] J. Liu et al., "Impact of Power Grid Strength and PLL Parameters on Stability of Grid-Connected DFIG Wind Farm," in IEEE Transactions on Sustainable Energy, vol. 11, no. 1, pp. 545-557, Jan. 2020, https://doi.org/10.1109/TSTE.2019.2897596

[9] Murty, V.V.S.N., Kumar, A. RETRACTED ARTICLE: Multi-objective energy management in microgrids with hybrid energy sources and battery energy storage systems. Prot Control Mod Power Syst 5, 2 (2020). https://doi.org/10.1186/s41601-019-0147-z

[10] Bozorg, M., Bracale, A., Caramia, P. et al. Bayesian bootstrap quantile regression for probabilistic photovoltaic power forecasting. Prot Control Mod Power Syst 5, 21 (2020). https://doi.org/10.1186/s41601-020-00167-7

[11] Yang, B., Zhong, L., Wang, J., Shu, H., Zhang, X., Yu, T., et al. (2021). State-of-the-art One-Stop Handbook on Wind Forecasting Technologies: An Overview of Classifications, Methodologies, and Analysis. J. Clean. Prod. 283, 124628. https://doi.org/10.1016/j.jclepro.2020.124628

[12] Huang, K., Li, Y., Zhang, X. et al. Research on the power control strategy of household-level electric power routers based on hybrid energy storage droop control. Prot Control Mod Power Syst 6, 13 (2021). https://doi.org/10.1186/s41601-021-00190-2

[13] Yang, B., Ye, H., Wang, J., Li, J., Wu, S., Li, Y., et al. (2021). PV Arrays Reconfiguration for Partial Shading Mitigation: Recent Advances, Challenges and Perspectives. Energy. Convers. Manag. 247, 114738. https://doi.org/10.1016/j.enconman.2021.114738

[14] Yang, B., Zhong, L., Wang, J., Shu, H., Zhang, X., Yu, T., et al. (2021). State-of-the-art One-Stop Handbook on Wind Forecasting Technologies: An Overview of Classifications, Methodologies, and Analysis. J. Clean. Prod. 283, 124628. https://doi.org/10.1016/j.jclepro.2020.124628

[15] Urquhart, B., Ghonima, M., Nguyen, D., Kurtz, B., Chow, C. W., and Kleissl, J. (2013). Sky-Imaging Systems for Short-Term Forecasting. Solar Energy. Forecast. Resource Assess., 195–232. https://doi.org/10.1016/B978-0-12-397177-7.00009-7

[16] Shen, Y., Yao, W., Wen, J., He, H., and Jiang, L. (2019). Resilient Wide-Area Damping Control Using Grhdp to Tolerate Communication Failures. IEEE Trans. Smart Grid 10 (3), 2547–2557. https://doi.org/10.1109/TSG.2018.2803822

[17] Tuohy, A., Zack, J., Haupt, S. E., Sharp, J., Ahlstrom, M., Dise, S., et al. (2015). Solar Forecasting: Methods, Challenges, and Performance. IEEE Power Energy. Mag. 13 (6), 50–59. https://doi.org/10.1109/MPE.2015.2461351

[18] Li, G., Liao, H., and Li, J. (2011). Discussion on the Method of Grid-Connected PV Power System Generation Forecasting. J. Yunnan Normal Univ. 31 (2), 33–38, 64. https://doi.org/10.3969/j.issn.1007-9793.2011.02.006

[19] Zhang Y, Wang J, Wang X. Review on probabilistic forecasting of wind power generation. Renewable Sustainable Energy Rev 2014;32:255–70. https://doi.org/10.1016/j.rser.2014.01.033

[20] Yan J, Liu Y, Han S, Wang Y, Feng S. Reviews on uncertainty analysis of wind power forecasting. Renewable Sustainable Energy Rev 2015;52:1322–30. https://doi.org/10.1016/j.rser.2015.07.197

[21] James EP, Benjamin SG, Marquis M. Offshore wind speed estimates from a high-resolution rapidly updating numerical weather prediction model forecast dataset. Wind Energy 2018;21(4):264–84. https://doi.org/10.1016/j.rser.2015.07.197

[22] Zhao J, Guo YL, Xiao X, Wang JZ, Chi DZ, Guo ZH. Multi-step wind speed and power forecasts based on a WRF simulation and an optimized association method. Apply Energy 2017;197:183–202. https://doi.org/10.1016/j.apenergy.2017.04.017

[23] Lei M, Shiyan L, Chuanwen J, Hongling L, Yan Z. A review on the forecasting of wind speed and generated power. Renewable Sustainable Energy Rev 2009;13(4):915–20. https://doi.org/10.1016/j.rser.2008.02.002

[24] E. Raafat Maamoun Shouman, 'Wind Power Forecasting Models', Wind Turbines - Advances and Challenges in Design, Manufacture and Operation. IntechOpen, Oct. 26, 2022. doi: 10.5772/intechopen.103034.

[25] Pinson P. Wind energy: forecasting challenges for its operational management. StatSci 2013;28(4):564–85. https://doi.org/10.1214/13-STS445

[26] Didavi, A. B., Agbokpanzo, R. G., and Agbomahena, M. (2021, December). Comparative study of Decision Tree, Random Forest and XGBoost performance in forecasting the power output of a photovoltaic system. In 2021 4th International Conference on Bio-Engineering for Smart Technologies (BioSMART) (pp. 1-5). IEEE.

[27] Khandakar, A., EH Chowdhury, M., Khoda Kazi, M., Benhmed, K., Touati, F., Al-Hitmi, M., and Jr SP Gonzales, A. (2019). Machine learning-based photovoltaics (P V) power prediction using different environmental parameters of Qatar. Energies, 12(14), 2782.

[28] Massaoudi, M., Refaat, S. S., Abu-Rub, H., Chihi, 1., and Wesleti, F. S. (2020, July). A hybrid Bayesian ridge regression-CWT-boost model for PV power forecasting. In 2020 IEEE Kansas Power and Energy Conference (KPEC) (pp. 1-5). IEEE.

[29] Li, X., Ma, L., Chen, P., Xu, H. Xing, Q., Yan, J., and Cheng, Y. (2022). Probabilistic solar irradiance forecasting based on XGBoost. Energy Reports, 8, 1087-1095

[30] Carneiro, T. C., Rocha, P. A., Carvalho, P. C., and Fernåndez-Ramirez, L. M. (2022). Ridge regression ensemble of machine learning models applied to solar and wind forecasting in Brazil and Spain. Applied Energy, 314, 118936.

[31] Kumar, A., Rizwan, M., and Nangia, U. (2020). A hybrid intelligent approach for solar photovoltaic power forecasting: impact of aerosol data. Arabian Journal for Science and Engineering, 45(3), 1715-1732.

[32] Munawar, U., and Wang, Z. (2020). A framework for using machine learning approaches for short-term solar power forecasting. Journal of Electrical Engineering and Technology, 15(2), 561-569.

[33] Cervone, G., Clemente-Harding, L., Alessandrini, S., and Delle Monache, L. (2017). Short-term photovoltaic power forecasting using Artificial Neural Networks and an Analog Ensemble. Renewable Energy, 108, 274-286.

[34] Mohana, M., Saidi, A. S., Alelyani, S., Alshayeb, M. J., Basha, S., and Anqi, A. E. (2021). Small-scale solar photovoltaic power prediction for residential load in Saudi Arabia using machine learning. Energies, 14(20), 6759.

[35] Perveen, G., Rizwan, M., and Goel, N. (2018). Intelligent model for solar energy forecasting and its implementation for solar photovoltaic applications. Journal of Renewable and Sustainable Energy, 10(6), 063702.

[36] Yang, D., Y e, Z., Lim, L. H. 1., and Dong, Z. (2015). Very short-term irradiance forecasting using the lasso. Solar Energy, 114, 314-326.

[37] Zazoum, B. (2022). Solar photovoltaic power prediction using different machine learning methods. Energy Reports, 8, 19-25.

[38] Chiteka, K., Arora, R., and Sridhara, S. N. (2020). A method to predict solar photovoltaic soiling using artificial neural networks and multiple linear regression models. Energy Systems 11(4), 981-1002.

[39] Alfadda, A., Adhikari, R., Kuzlu, M., and Rahman, S. (2017, April). Hour-ahead solar PV power forecasting using an SVR-based approach. In 2017 IEEE Power and Energy Society Innovative Smart Grid Technologies Conference (ISGT) (pp. 1-5). IEEE.

[40] Serafeim Loukas, P. D. (2022, June 10). What is machine learning: A short note on supervised, unsupervised, semi-supervised and... Medium. Retrieved June 10, 2022, from

https://towardsdatascience.com/what-is-machine-learning-a-short-note-on-supervised-unsupervised-semi-supervised-and-aed1573ae9bb

[41] Solar Power Generation Data. (2020, August 18). Kaggle. https://www.kaggle.com/datasets/anikannal/solar-power-generation-data

[42] S. (n.d.-b). Wind-Energy-Analysis-and-Forecast-using-Deep-Learning-LSTM/dataset at master · Sk70249/Wind-Energy-Analysis-and-Forecast-using-Deep-Learning-LSTM. GitHub. https://github.com/Sk70249/Wind-Energy-Analysis-and-Forecast-using-Deep-Learning-LSTM/tree/master/dataset

[43] Yıldırım, S. (2021, December 15). 11 Most Common Machine Learning Algorithms Explained in a Nutshell. Medium. https://towardsdatascience.com/11-most-common-machine-learning-algorithms-explained-in-a-nutshell-cc6e98df93be

[44] Zuccarelli, E. (2021, December 26). Performance Metrics in ML-Part 2: Regression | Towards Data Science. Medium. https://towardsdatascience.com/performance-metrics-in-machine-learning-part-2-regression-c60608f3ef6a

[45] Solar PV – Analysis - IEA. (n.d.). IEA. https://www.iea.org/reports/solar-pv

[46] Current Status | Ministry of New and Renewable Energy, Government of India. (n.d.). https://mnre.gov.in/solar/current-status/

[47] Solar energy | Definition, Uses, Advantages, & Facts. (2023, April 28). Encyclopaedia Britannica. https://www.britannica.com/science/solar-energy/Electricity-generation

[48] Wind Electricity – Analysis - IEA. (n.d.-b). IEA. https://www.iea.org/reports/wind-electricity

[49] Overview | Ministry of New and Renewable Energy, Government of India. (n.d.). https://mnre.gov.in/wind/current-status/

[50] Selin, N. E. (2023, April 24). Wind power | Capacity & Facts. Encyclopaedia Britannica. https://www.britannica.com/science/wind-power

[51] Lai, J., Chang, Y., Chen, C., & Pai, P. (2020). A Survey of Machine Learning Models in Renewable Energy Predictions. Applied Sciences, 10(17), 5975. https://doi.org/10.3390/app10175975