INDIAN INSTITUTE OF TECHNOLOGY, KANPUR

DEPARTMENT OF ELECTRICAL ENGINEERING

Students-Undergraduate Research Graduate Excellence (SURGE) program 2022

Project Report

"Wind Power Forecasting Using Machine Learning Algorithms"

Under the guidance of **Prof.Gururaj Mirle Vishwanath**

Electrical Engineering Department, IIT Kanpur
gururajmv@iitk.ac.in
Tel: +91-512-259-7131
Office: ACES-103A, EED, IIT Kanpur

Submitted by:-
Satyansha Dev
Dept. of Electrical Engineering
Institute Roll no. **200897**
SURGE Roll no. **2230581**

CERTIFICATE

This is to certify that the project entitled “**Wind Power Forecasting Using Machine Learning Algorithms**” submitted by Satyansha Dev (2230581) as a part of Summer Undergraduate Research and Graduate Excellence 2022 offered by the Indian Institute of Technology, Kanpur, is a Bonafede record of the work done by her under my guidance and supervision at the Indian Institute of Technology, Kanpur from 1th June 2022 to 31st July 2022.

Gururaj MV

Prof. Gururaj Mirle Vishwanath
Assistant Professor
Department of Electrical Engineering
Indian Institute of Technology, Kanpur

Acknowledgments

Firstly, I will thank IIT Kanpur for giving me such an opportunity in which I learned a lot, and thanks again for allowing me to work under the supervision of such an inspiring mentor, Prof MV Gururaj. I am extremely grateful to sir for encouraging me at every moment. He also motivated me to learn Machine Learning, and explore data science. I am thankful to him for guiding and mentoring my project. His motivation gave me a chance to explore my calibre. Thanks to everyone in the department of Electrical Engineering, IIT Kanpur, for providing me with such an environment filled with motivation and knowledge that helped me develop my interest in research and further studies. This research-based internship was under constant support and inspiration from my friends and family, who helped to create an internship environment.
Thanks to everyone who supported and encouraged me.

Satyansha Dev

Wind Power Forecasting Using Machine Learning Algorithms

Abstract:- The scarcity of conventional resources has led to the exploration of renewable energy resources, non-abundant and non-exhaustible energy on a human scale, such as fossil fuels. In this regard, wind power is taking significant importance worldwide. Wind power prediction represents an active and important field in the renewable energy sector. Wind power production has increased rapidly. Having compatible wind turbines in the market forces the conventional powerplants to shut down. This affects the reserve markets, whose cost rises as the wind power capacity grows. The costs could be reduced by having wind turbines participate in the reserve markets. Therefore, forecasting the production of this type of energy is a crucial issue for a permanent balance between consumption and production. Since renewable energy sources are fixed into existing grids and combined with traditional sources, it is essential to know the amount of energy it will produce to minimize the wind farm's operational cost and the power grid's safe operation. This expected wind power is required to acquire consistent power generation from wind. The paper aims to utilize a combination of machine learning (ML) techniques for feature selection and regression (Decision Tree, Neural Network, Support Vector Machine, k Nearest Neighbor, and Random Forest, to investigate the advantages and the challenges associated with the algorithms within wind power forecasting). The proposed methodology would eventually be an ML model, which uses feature selection through irrelevancy and redundancy filters and then employ algorithms for auxiliary prediction. The data set used for training and testing the model has real-time daily values of wind speed, humidity, temperature, solar irradiance, and power output. The developed machine learning model would predict the amount of electricity produced from wind, a renewable resource. Hence, a step toward optimizing the entire ecosystem.

Keywords: Wind power; forecast; Machine Learning; Power grids; Regression; Data sets; Prediction algorithms

Contents

I. Introduction

Among the different types of green energy, wind energy is an essential part of electricity generation worldwide. Power production from conventional units is a large contributor to the emission of greenhouse gasses. Utilizing renewable energy sources instead of coal and oil-fired powerplants has become a global trend in recent decades. Renewable Energy is fuel-free, and the production cost has decreased to a competitive level. Many countries have substituted a share of the conventional units for RE. Biomass isn't included in the RE part. This lowered the emission of greenhouse gasses significantly and created change in players in the electricity market. RE sources have many benefits for society, but it also has a downside. Renewable energy is a fluctuating energy source and is therefore unreliable. Even though RE sources can have significant advantages by supplying low-cost clean energy, the downside is larger imbalances. Increasing power prediction accuracy is the best way to decrease imbalances and regulate cost. Forecasts are essential in a power system with large shares of RE. This paper will focus on wind power forecast and investigate methods used in the field. One of those methods is a Neural network, an effective way to solve the data fitting problem. Many works of literature have applied neural network methods to wind short-term power forecasting. Compared with traditional BP neural networks and RBF neural networks, an extreme learning machine neural network only needs to calculate the weight matrix from the hidden layer to the output layer, and the weight matrix of the hidden layer of the input layer is randomly generated. Therefore, it can be said that the training process of extreme learning machine neural networks is simple and easy to realize.

II. Importance of Wind Power Prediction

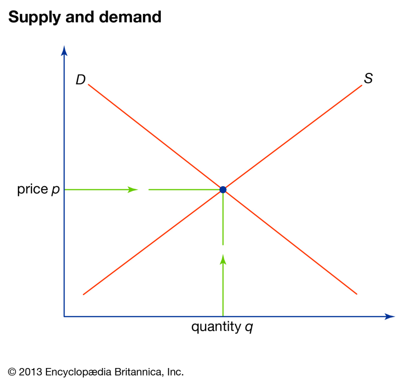
Wind power generation results are a directly related to wind speed, and, unlike conventional generation systems, it is not easily dispatchable. Therefore the fluctuations of wind generation require power substitution from various other sources that might not be available on short notice (12 hours for a nuclear one, and it takes 6 hours to fire up a coal plant). The situation worsens once wind power starts providing more than a small percentage of the electricity supplied to the grid. Better predictions allow utilities to deploy fewer spinning reserves, usually natural gas-based generators. Utilities on 2 different separate time scales typically request the forecasts:

1. short-term forecast (minutes to 6 hours), used to adjust the spinning reserves;
2. the utility uses long-term (day-to-week) forecasts to plan the energy buy or mix electricity from other energy providers.

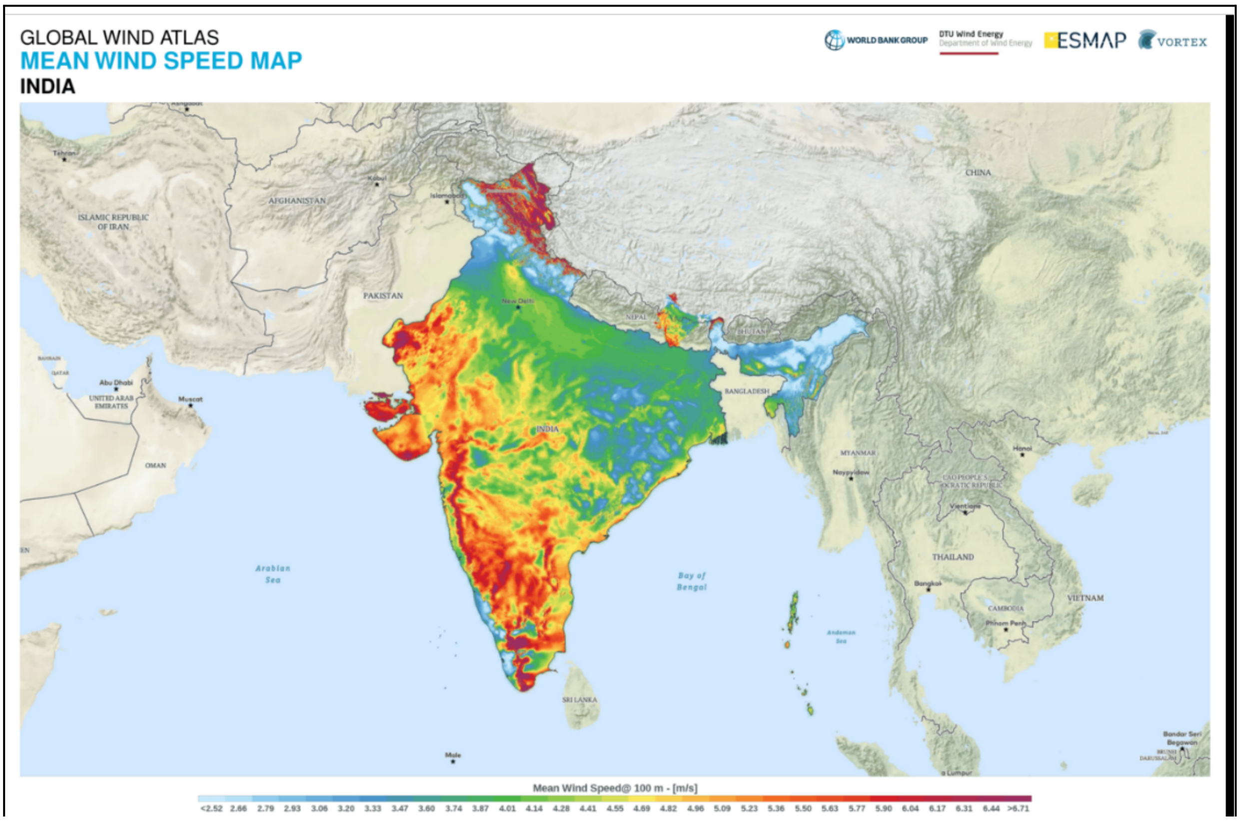
This information is required on a "day ahead" basis (e.g., by 6 am), but markets tend not to operate on the holidays and weekends, so sometimes, longer forecasts are used.

The challenges the utilities face when wind generation is injected into a power system depend on the share of that renewable energy. For Denmark, a country with very high shares of wind power in the electricity mix, the average wind energy penetration was 40-45% (that is 40-45% of the electricity consumption was met with wind energy), while rapid penetration (that is, the instantaneous wind energy power production compared to consumption to be completed at a given time) sometimes was above 100 percent (with occasional negative pricing for the electricity)

At any moment, the balance must be maintained in the electricity grid between electricity consumption and generation; otherwise, power quality or supply disturbances may occur.



The (TSO) transmission system operator is majorly responsible for the management of the balance of electricity on the grid: at any given time, electricity production needs to match the consumption. Therefore, production means should be scheduled to respond to the load profiles. The load corresponds to total electricity consumption over the area of interest, wing energy. Load forecasts generally give load profiles of high accuracy parallel to the market participation; wind power forecasts can be used for the optimal combined operation of wind and conventional generation, wind and hydro-power generation, or wind in combination with some energy storage devices. They also serve as the basis for quantifying the reserve needs for compensating for the eventual lack of wind energy production.



III. National Wind power potential in India

The National Institute of Wind Energy (NIWE) of the Indian government operates over 800 functional wind-monitoring stations over the whole nation. It has produced potential wind maps at heights of 50, 80, 100, and 120 metres. According to the most current study, the country has a 302 GW wind power potential at 100 metres and a 695.50 GW potential at 120 metres above the ground.

Installed wind power capacity and generation in India since 2007[12]

Financial year	06-07	07-08	08-09	09-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17	17-18	18-19[13]	19-20	20-21	21-22
Installed capacity (MW)	7,850	9,587	10,925	13,064	16,084	18,421	20,150	22,465	23,447	26,777	32,280	34,046	35,626	37,669	38,785	40,355
Generation (GWh)									28,214	28,604	46,011	52,666	62,036	64,485	59,824	68,640[14]

IIIA. History of Development of wind power in India

Famous power engineer Maneklal Sankalchand Thacker and the Indian Council of Scientific and Industrial Research (CSIR) started a project in December 1952 to investigate whether or not using wind power in the country was feasible. In India, wind energy development began at this point. The CSIR's Wind Power Sub-Committee was established to investigate the resources that may be employed in practice and the financial potential of wind energy. With the assistance of the Indian Meteorological Department, the Sub-Committee carefully reviewed the data that was available regarding surface winds in India and their velocity duration. Additionally, it began careful examinations of suitable sites to capture the greatest quantity of wind energy efficiently. It also began in-depth research into potential areas for harnessing the greatest quantity of wind energy, built and tested colossal windmills made of wood and bamboo, and successfully manufactured more.

Installed Wind Power Capacity

Fiscal year end	Cumulative capacity (in MW)
2005	6,270
2010	16,084
2011	18,421
2012	20,149
2013	21,264
2014	23,354
2015	26,769
2016	32,280
2017	34,046
2018	35,626
2019	37,669
2020	38,785
2021	40,355

In September 1954, a symposium on sun and wind energy in New Delhi by the CSIR and UNESCO. One of the participants was E. W. Golding, a British power engineer and authority in wind energy production. Convinced of the potential of wind energy in India, he made additional recommendations, including extensive wind velocity surveys in different regions of the nation, the full-time assignment of staff to experimental wind power studies, the establishment of a dedicated research lab, and the development of small- to medium-sized wind-powered electrical generators. Areas around Coimbatore and Saurashtra are promising locations for producing electricity from wind energy. The Wind Power Sub-Committee began constructing 20 wind velocity survey stations across India, putting its windmills to the test, and gathering data. Another option that the Indian government looked at was the construction of nearly 20,000 small to medium-sized wind-powered energy generators in rural areas. In 1960, the CSIR established a Wind Power Division as a component of Bangalore's newly constructed National Aeronautical Laboratory. From the 1960s through the 1980s, the NAL and other organizations repeated wind velocity measurements and produced more precise estimates of India's wind energy potential. Large-scale wind power development began in 1985 with the installation of a 40-kW Dutch machine connected to the grid as the first wind project at Veraval, Gujarat. Even though this machine's performance was generally subpar, it proved that operating wind turbines in India's grid-connected mode were technically feasible. The Indian government then arranged a number of wind turbine demonstration events. The government subsequently developed several demonstration wind farms in the nation's coastal areas and simultaneously began an extensive initiative to find locations suitable for wind projects.

IIIB. Electricity generation

Wind energy produced 62.03 TWh in the fiscal year 2018–19, or about 4% of all electricity generated, and makes up nearly 10% of India's installed power producing capacity. The five months from May to September, which coincide with the duration of the Southwest monsoon, account for 70% of the annual wind energy production. In India, solar energy is an excellent complement to wind energy because it is produced primarily during daylight, outside the rainy season.

IIIC. Wind power by state

There is a growing number of wind energy installations in states across India.

Installed wind capacity by state as of 31 March 2021[28]

State	Total Capacity (MW)
Tamil Nadu	9608.04
Gujarat	8561.82
Maharashtra	5000.33
Karnataka	4938.60
Rajasthan	4326.82
Andhra Pradesh	4096.65
Madhya Pradesh	2519.89
Telangana	128.10
Kerala	62.50
Others	4.30
Total	**39247.05**

IIID. Offshore wind power plants

Around 70 GW of offshore wind energy are present in India, mainly off Gujarat and Tamil Nadu coast. No offshore wind farm is being built or run as of May 2022. India began making plans in 2010 to enter the offshore wind energy market, and in 2014, design work on a 100 MW demonstration project off the Gujarat coast began.

The Nodal Ministry (MNRE) & Nodal Agency (NIWE) comes with the Expression of Interest requesting the bidders to create the first 1000MW commercial-scale offshore wind farm in India, along the coast of Gujarat, it appears that India is moving quickly toward the development of offshore wind energy. The planned area defined in the FOWIND & FOWPI study financed by European Commission is specified in the EoI issued on April 16, 2018. The offshore wind farm's proposed location in the Gulf of Khambhat might be 23–40 km offshore from the Pipavav port. The projected area is 400 sq km in size. Under the direction of NIWE, the wind measurements and other data collection are in progress.

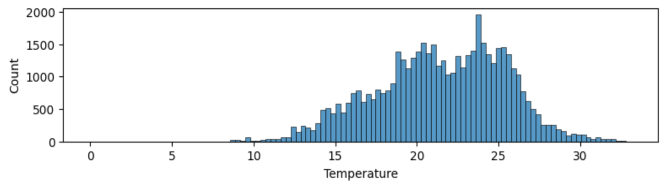
IV. Literature Review:

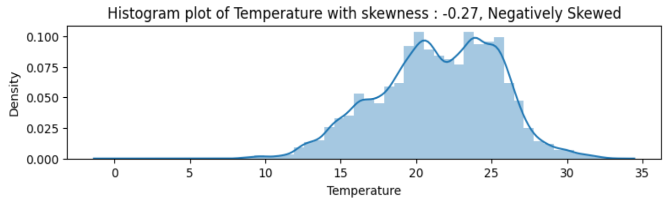
Numerous research has been conducted to anticipate wind energy's short- and medium-term production accurately. There have been several proposed statistics and machine learning algorithms. Inaccurate statistical models are used. A randomizable filter classifier method was found to be the best machine learning algorithm in a Fiji study on wind energy forecasting. For one day in advance forecasting, machine learning algorithms fared better in Italy than statistical approaches. A neural network-based covariance matrix adaptation-based algorithm was employed in Ireland to predict wind power production and the country's energy requirements. Using a brand-new hybrid model decomposition methodology is suggested, which significantly enhances multistep wind power generation's forecasting capabilities. The most effective preprocessing involves combining extreme machine learning techniques with hybrid preprocessing. Numerous studies have emphasized the significance of forecasting algorithms for wind energy production. They contend that machine learning algorithms are essential for predicting wind power output. If enough data is available, it can be expanded for long-term forecasting. They are employed for short-term prediction.

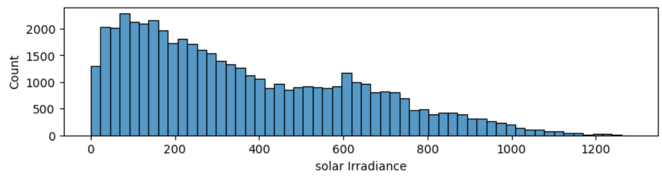
V. Case Study of Gujarat (India)

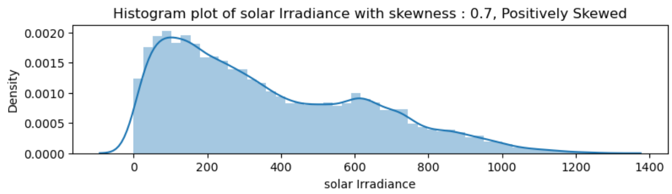
In 1986, Gujarat became the first state in India to put in a wind energy project. Gujarat had 3,093 MW of installed wind capacity as of February 2013. For India, there are a number of conflicting wind resource assessments. Due to the uncertainty surrounding these current wind energy evaluations, this analysis simulates the wind at current hub heights for one year using the Weather Research and Forecasting (WRF) model to produce more accurate estimates of wind resources in Gujarat. Accurate simulations of winds close to the surface and at elevations critical for wind energy are possible thanks to the WRF model. We concentrate on average wind speeds, which can be translated to wind power densities by the user using methods of their choice, as opposed to earlier resource assessments that published wind power density. We validated the model using conventional error metrics like root-mean-squared error (RMSE), mean absolute error (MAE), and bias, as well as against rank correlation, CRMSE, normalized RMSE, and the wind distribution error, using the data provided with attributes such as humidity, solar irradiance, wind speed, etc. (WDE). Similar to the data given, in the case of Gujarat, it is observed by the International Renewable Energy Agency that the highest wind speeds are in the Gulf of Kutch and the southern tip of the peninsula. It is also identified the Gulf of Khambhat has average wind speeds of 9 m/s. Conforming with the usual seasonal cycle expected in India and with validation results at five measurement sites, the wind speed is highest from May to August, 10 m/s, especially along the coastlines.

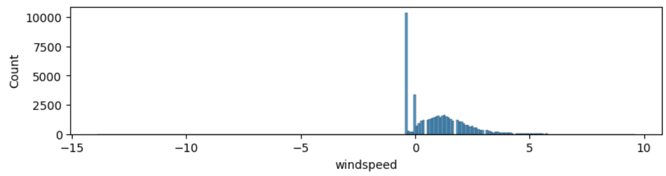
```
Mean Absolute Error: 17.17955615264725
Mean Squared Error: 1363.913218498463
Root Mean Squared Error: 36.9311957361045
```

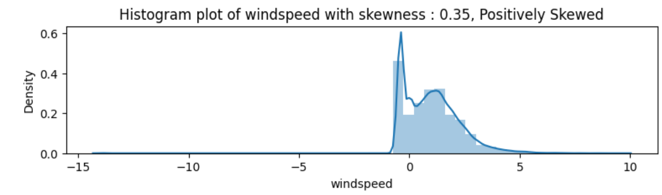


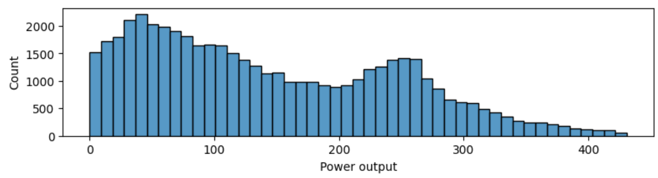


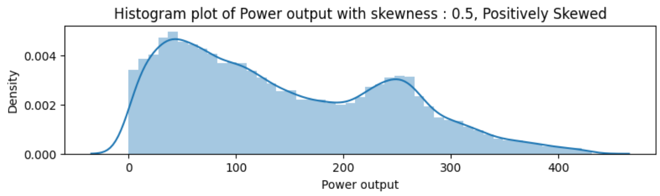


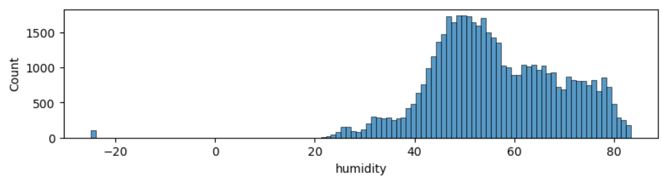


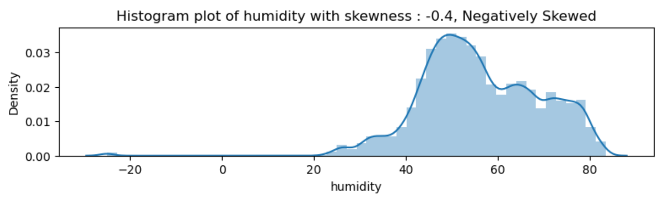


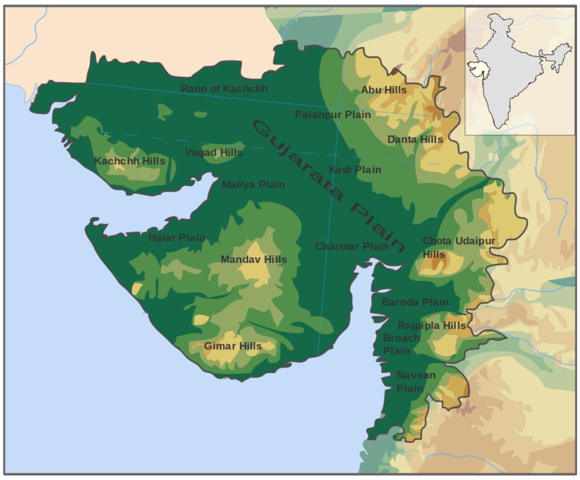












VI. Explanation and Prediction

Explaining and predicting are the same, or two separate concepts contested in science philosophy. The literature on philosophy of science is where the confusion between explanation and prediction first emerged. In his book on theory building, Dubin clarified the differences between the two ideas. Theories of social and human behavior address two distinct goals of science:

(1) prediction and

(2) understanding.

Why should explaining and predicting differ from one another? The explanation for this is that quantifiable data frequently misrepresent underlying ideas. The operationalization of theories and constructs into statistical models and quantifiable data results in differences in the capacity to conceptually describe phenomena and make predictions at a measurably accurate level. Consider a theory postulating that construct X causes construct Y, via the function F, such that Y = F(X). Measurable variables X and Y are operationalizations of X and Y, respectively. In the predictive context, we consider only X, Y, and f.

The disparity arises because explanatory modeling aims to match f and F as closely as possible for the statistical inference to apply to the theoretical hypotheses. The disparity manifests itself in different ways. Four major aspects are:

1. Causation–Association: In explanatory modeling, f represents an underlying causal function, and X is assumed to cause Y. In predictive modeling, f captures the association between X and Y.

2. Theory–Data: In explanatory modeling, f is carefully constructed based on F in a fashion that supports interpreting the estimated relationship between X and Y and testing the causal hypotheses. In predictive modeling, f is often constructed from the data. Direct interpretability in the relationship between X and Y is not required, although sometimes, transparency of f is desirable.

3. Retrospective–Prospective: Predictive modeling is forward-looking in that f is constructed for predicting new observations. In contrast, explanatory modeling is retrospective in that f is used to test an already existing set of hypotheses.

4. Bias–Variance: The expected prediction error for a new observation with value x, using a quadratic loss function,

$$\begin{aligned}\mathrm{EPE} &= E\{Y - \hat{f}(x)\}^2 \\ &= E\{Y - f(x)\}^2 + \{E(\hat{f}(x)) - f(x)\}^2 \\ &\quad + E\{\hat{f}(x) - E(\hat{f}(x))\}^2 \\ &= \mathrm{Var}(Y) + \mathrm{Bias}^2 + \mathrm{Var}(\hat{f}(x)).\end{aligned}$$

Bias is the result of misspecifying the statistical model f. When f is estimated using a sample, estimation variance is the result. The first term represents the error that occurs even when the model is correctly stated and accurately estimated. The source of the distinction between explanatory and predictive modeling is revealed by the decomposition as mentioned above: In explanatory modelling, the goal is to create an accurate representation of the underlying theory while reducing bias. In contrast, predictive modeling aims to reduce bias and estimation variance, often at the expense of theoretical precision in favor of increased empirical precision.

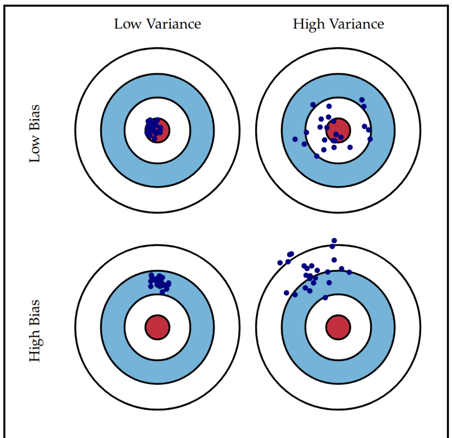
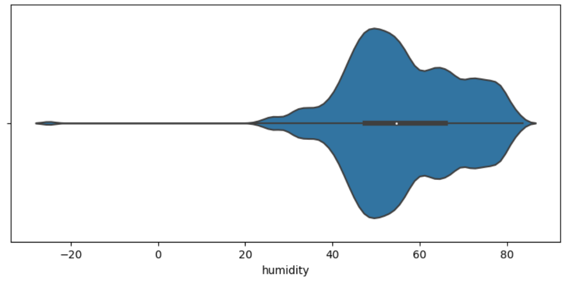
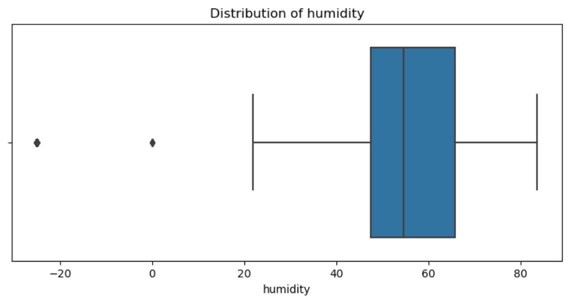
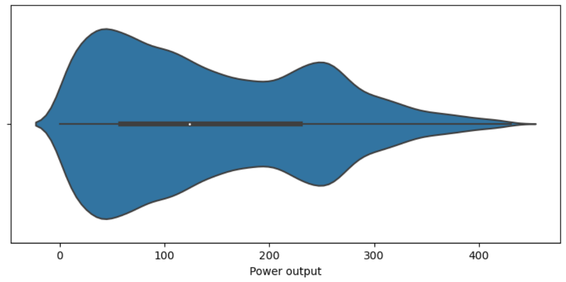
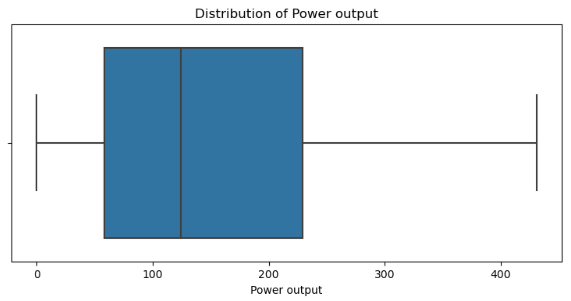


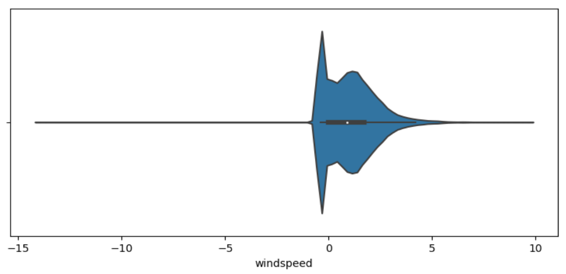
Fig. 1 Graphical illustration of bias and variance.

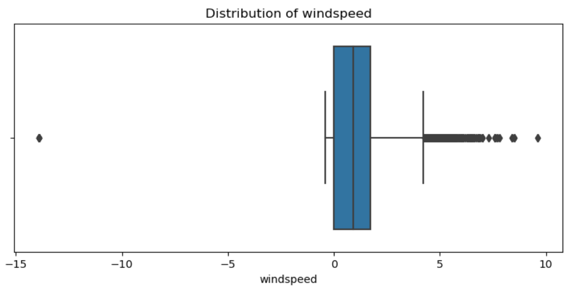

−20 0 20 40 60 80
humidity

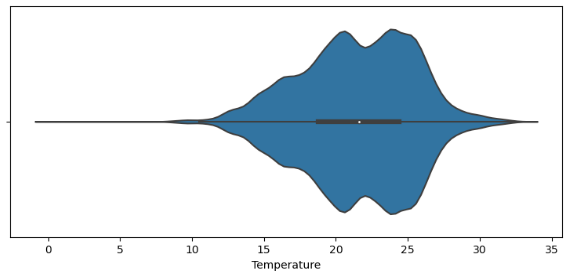

Distribution of humidity
−20 0 20 40 60 80
humidity

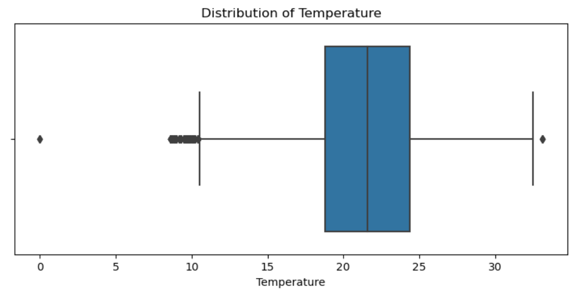

0 100 200 300 400
Power output

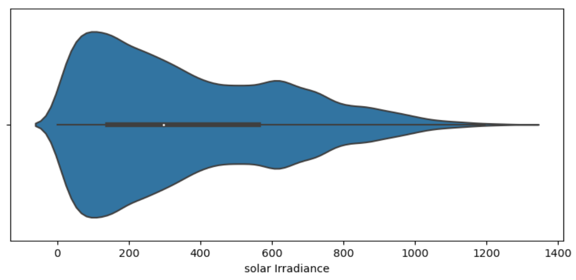

Distribution of Power output
0 100 200 300 400
Power output

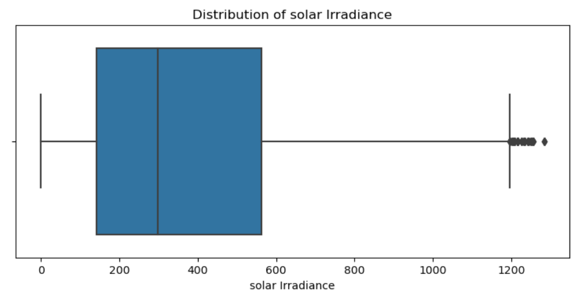

−15 −10 −5 0 5 10
windspeed


Distribution of windspeed
−15 −10 −5 0 5 10
windspeed


0 5 10 15 20 25 30 35
Temperature


Distribution of Temperature
0 5 10 15 20 25 30
Temperature


0 200 400 600 800 1000 1200 1400
solar Irradiance


Distribution of solar Irradiance
0 200 400 600 800 1000 1200
solar Irradiance

A. Explanatory Modeling:

The causal theory is almost always tested using statistical methods in many scientific disciplines, especially the social sciences. Statistical models are applied to data to assess causal hypotheses in the context of a causal theoretical model. In these models, an underlying effect, measured by variable Y, is considered to be the result of a series of underlying variables, measured by variable X. Define explaining as a causal explanation in light of this fact, and explanatory modeling as the use of statistical models to verify causal explanations.

B. Predictive Modeling:

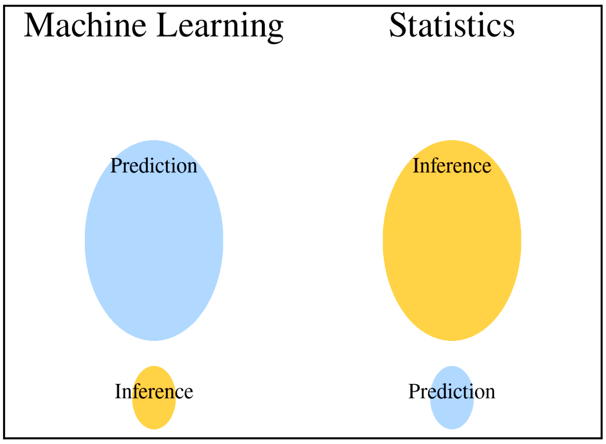
Applying a statistical model or data mining technique to the data allows for predicting upcoming or new observations. Nonstochastic prediction, in particular, seeks to forecast the output value (Y) of fresh observations given their input values (X). This concept also covers temporal forecasting, which involves predicting future values at time $t + k$, $k > 0$ using observations up until time t (the input) (the output). Point or interval forecasts, prediction regions, predictive distributions, or rankings of recent data are all examples of predictions.

C. Descriptive Modeling:

The objective of descriptive modeling is to characterize or depict the data structure concisely. In contrast to explanatory modeling, descriptive modeling does not rely on an underlying causal theory or does so less formally. Additionally, rather than the construct level, the measurable level is the main focus rather than the construct level. Descriptive modeling does not attempt to make predictions, unlike predictive modeling. Using a regression model can be descriptive when recording the relationship between the dependent and independent variables rather than for causal inference or prediction. We bring up this kind of modeling to distinguish it from causal explanatory and predictive modeling and to draw attention to the various methods used by statisticians and non-statisticians.

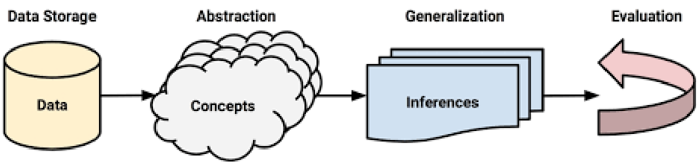
VII. Machine Learning

Machine learning is one of the newest global trends in the field of forecasting. The process of employing algorithms to evaluate data and create predictions is known as machine learning, which is the science of teaching computers to learn. Without having to be specifically coded, machine learning can learn from data.



The process of machine learning can be divided into three steps:

1. Data input,
2. Abstraction and
3. Generalization

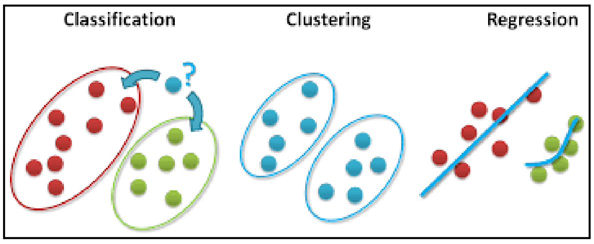


A. Overview of Machine Learning Methods

Machine learning consists of several algorithms, which can be classified into supervised or unsupervised learning. Supervised learning derives a function from labeled training data, while unsupervised learning uses unlabeled training data.

Further it can be divided into three sub-categories:

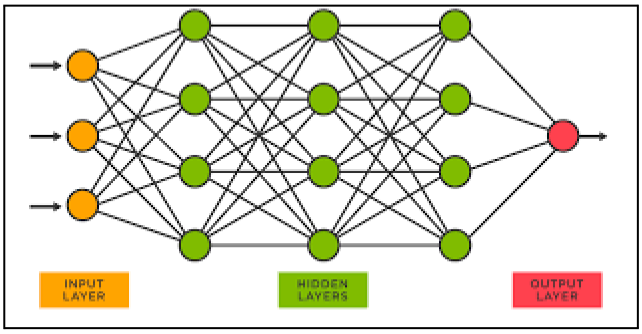
1. Classification
2. Regression and
3. Clustering



B. Brief Description of Machine Learning Methods

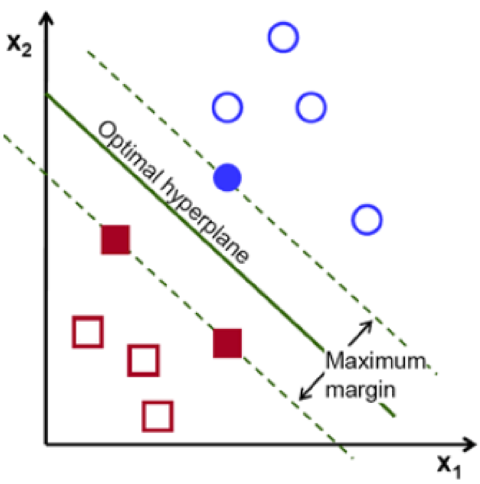
a. Neural Network:

It has nodes from the neighbouring layers, an input layer,, and an output layer. There are multiple nodes in each layer, and they are all linked to one or more hidden layers. The input of a neural network might be labeled data or unlabeled data.



b. Support Vector Machine:

A supervised learning tool is SVM. The goal is to locate hyperplanes in an N-dimensional space, where N is the total number of attributes that can be used to categorize data points. It measures the maximum distance between the data points and the hyperplane to distinguish between classes. Future data can be classified with greater assurance if the hyperplane and data points are as far apart as possible. Class 1 is represented by the squares, and class 2 by the circles. The ideal hyperplane is displayed.



c. k Nearest Neighbor:

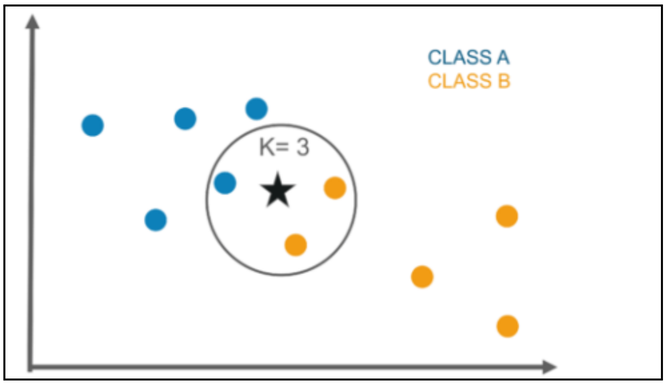
A straightforward supervised learning algorithm is kNN. The kNN method operates on the presumption that similar data points are nearby. A future datapoint can be evaluated by its closest neighbours by having a categorized dataset.

The number of "neighbours" to which the data point must be compared is denoted by the symbol k. The three data points closest to the new datapoint must be identified if k = 3.

The straight-line distance is called the Euclidean distance;

$$dist(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_n - q_n)^2}$$

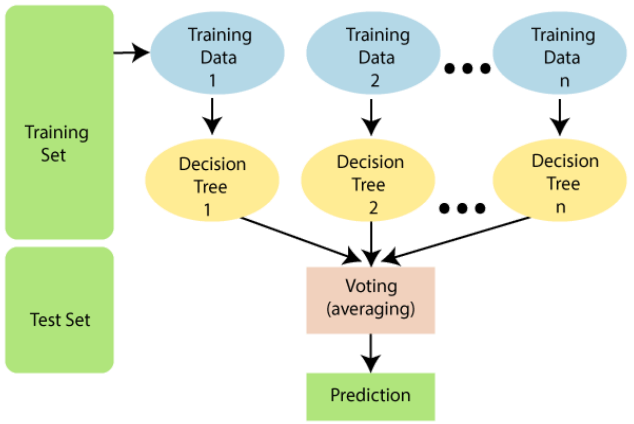
The star represents a new datapoint, and k is equal to three. The three nearest neighbors are found, and the new datapoint is classified in class B because most neighbors are in the respective class.



d. Random Forest:

The Random Forest method is made up of several Decision Trees. A decision tree is a model of decisions that resembles a tree, with each node holding a test for an attribute.

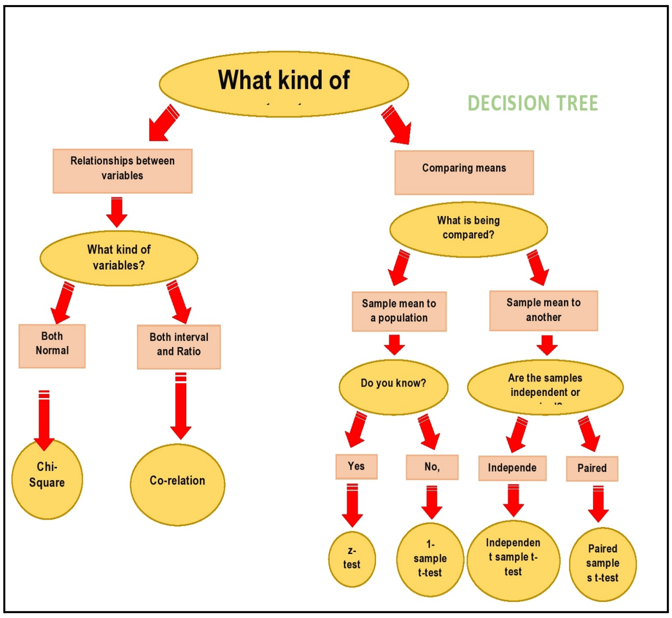
The class is represented by each leaf, while the branch symbolizes the test's results. In order to determine the overall class, Random Forest mixes many Decision Trees and relies on majority voting of all the prediction models.



e. Decision tree:

The classification and regression trees make up the decision tree. Until all nodes include only observations from the same data classes for all descendants, the root will be partitioned. A target variable Y is explained by the prediction using a group of explanatory factors X. The chi-square test is used to examine the modality of the explanatory factors X in order to determine whether variables have a high correlation with the target Y. We draw the conclusion

that the variable exhibits a significant positive correlation with the target variable Y for a chi-square test p-value less than 0.05. The significance of this criterion enables us to end the loop when all nodes have chi-square tests between variables X and Y greater than 0.05.



C. Performance Evaluation Criteria

The criteria used to evaluate our forecasting are given by (3), (4), (5), and (6), where N refers to the total number of online values that the data contain. These indices allow judgments on comparisons for future model improvement. However, it is very difficult to compare the models due to the different prediction horizons, the number of input parameters, and the weather conditions.

$$\frac{1}{N}\left|Y_i - \hat{Y}_t\right| = MAE \qquad (3)$$

$$\frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \hat{Y}_t\right)^2 = MSE \qquad (4)$$

$$\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \hat{Y}_t\right)^2} = RMSE = \sqrt{MSE} \qquad (5)$$

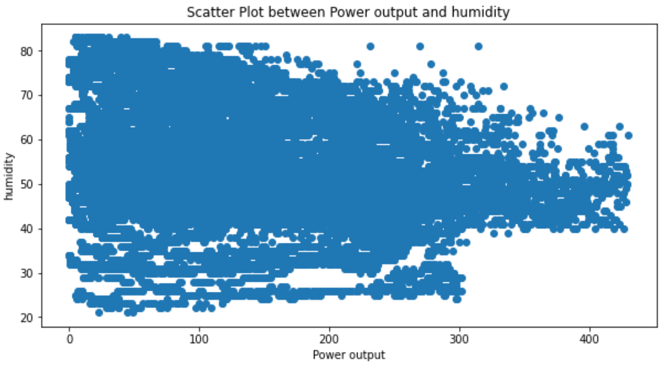
$$1 - \frac{\sum_{i=1}^{N}\left(Y_i - \hat{Y}_t\right)^2}{\sum_{i=1}^{N} Y_i^2} = R^2 \qquad (6)$$

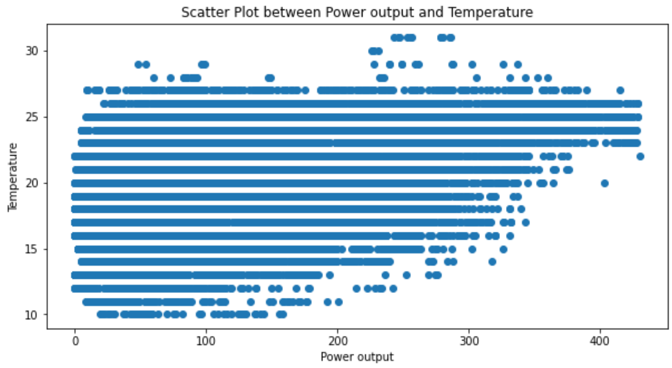
Nevertheless, the Mean Absolute Error (MAE) in (3) applies linear cost functions with a certain proportionality of the prediction errors.

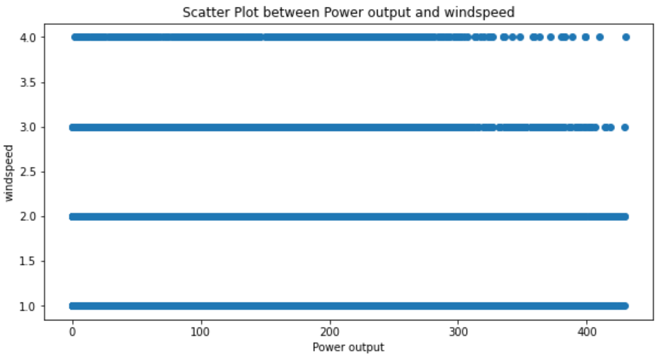
As for the Mean Square Error (MSE) in (4), it disfavors the highest deviations between prediction and observation but its square root (RMSE), according to equation (5), is very sensitive to these deviations.

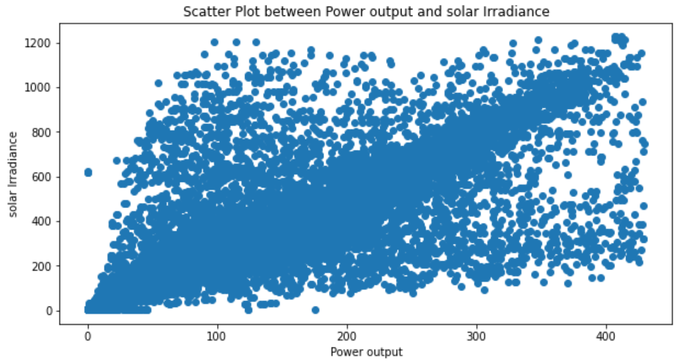
Thus it represents an excellent comparative parameter, especially in public use applications. The lower the (RMSE) or the MAE, the better the prediction for the production of our plant.

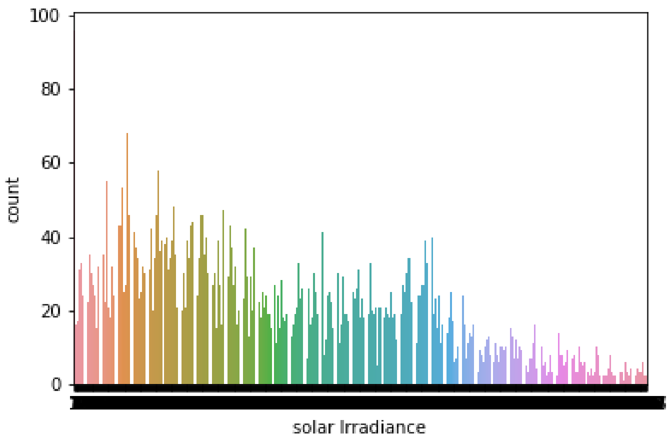
Mean squared error	$\text{MSE} = \frac{1}{n}\sum_{t=1}^{n} e_t^2$
Root mean squared error	$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2}$
Mean absolute error	$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n} \lvert e_t \rvert$

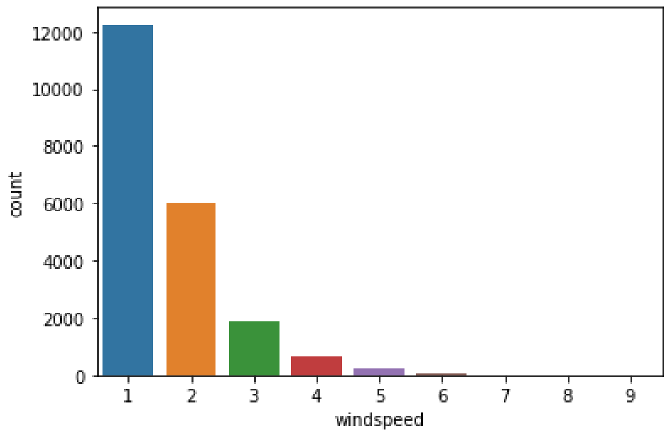

Scatter Plot between Power output and humidity
humidity
Power output

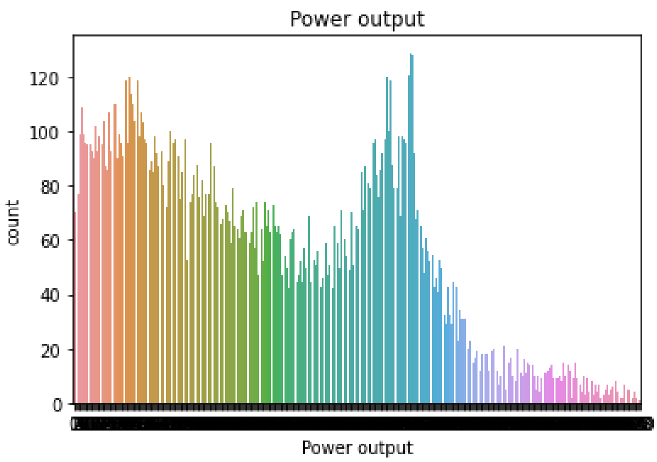

Scatter Plot between Power output and Temperature
Temperature
Power output

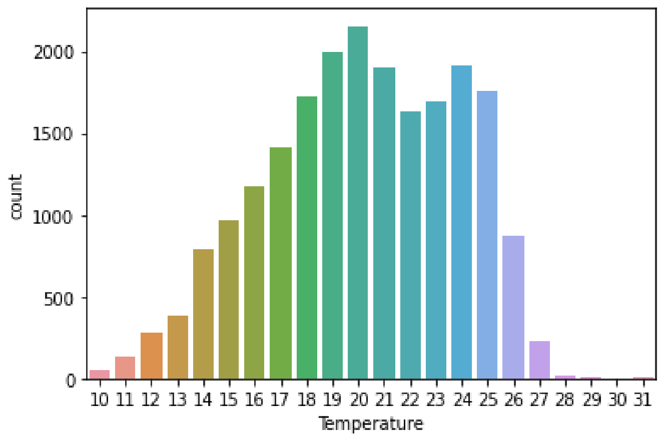

Scatter Plot between Power output and windspeed
windspeed
Power output

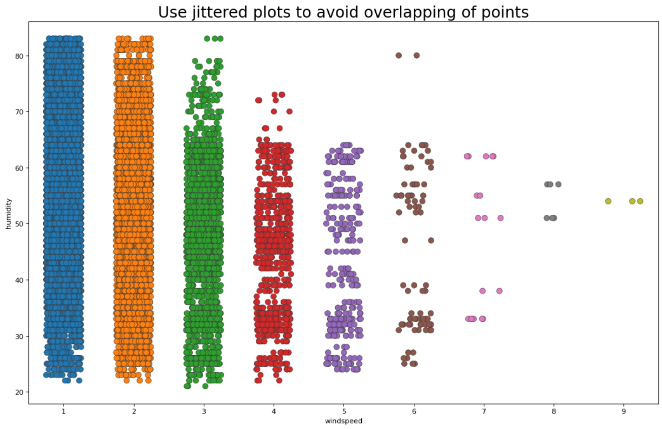

Scatter Plot between Power output and solar Irradiance
solar Irradiance
Power output

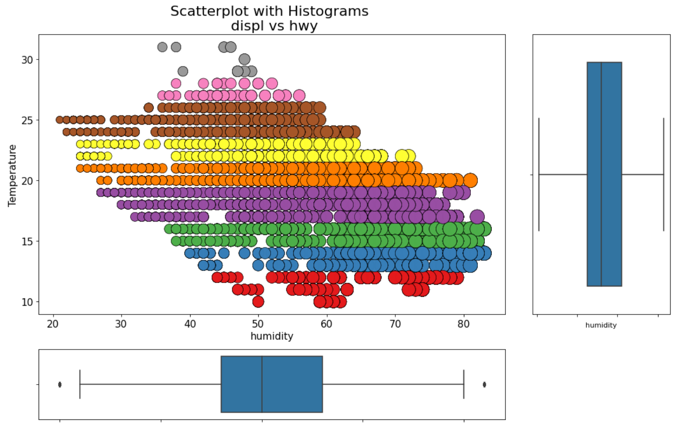

count
solar Irradiance

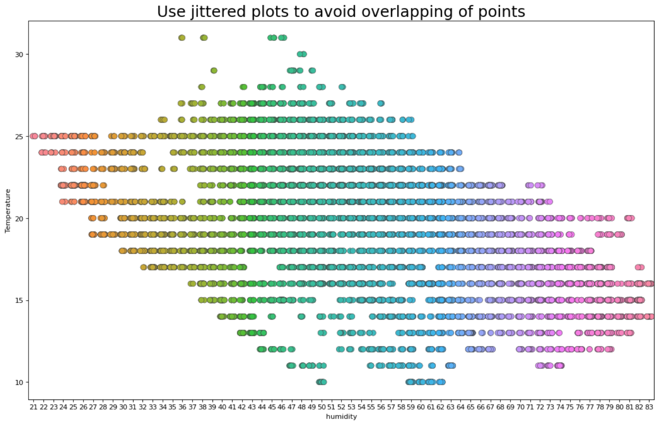

count
windspeed

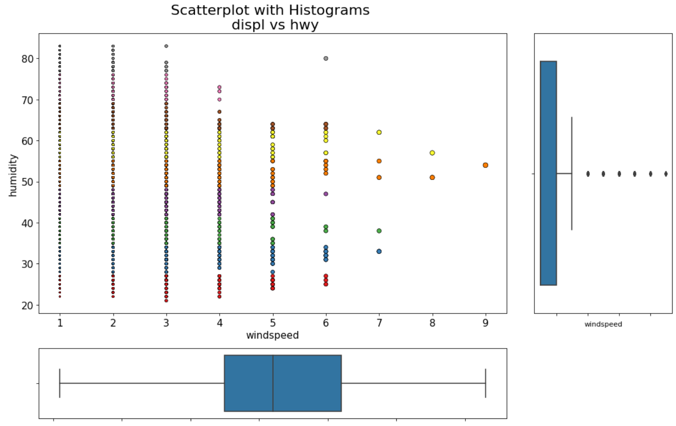

Power output
count
Power output

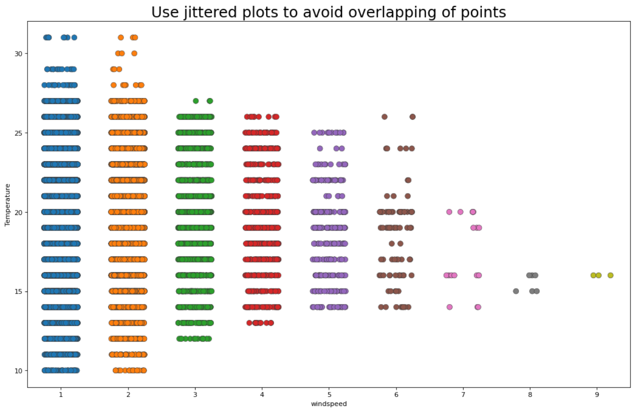

count
Temperature

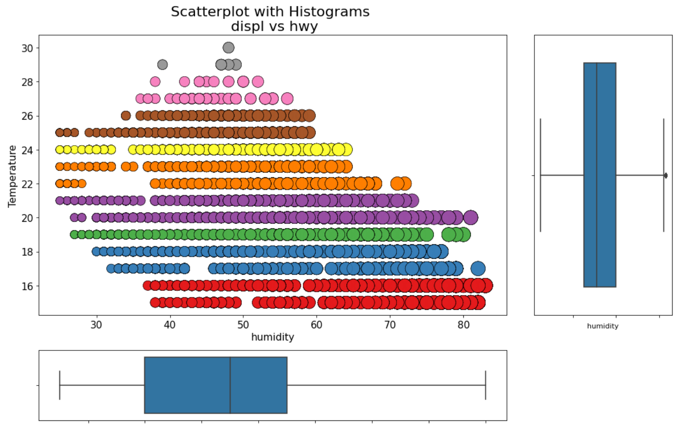

Use jittered plots to avoid overlapping of points
humidity
windspeed

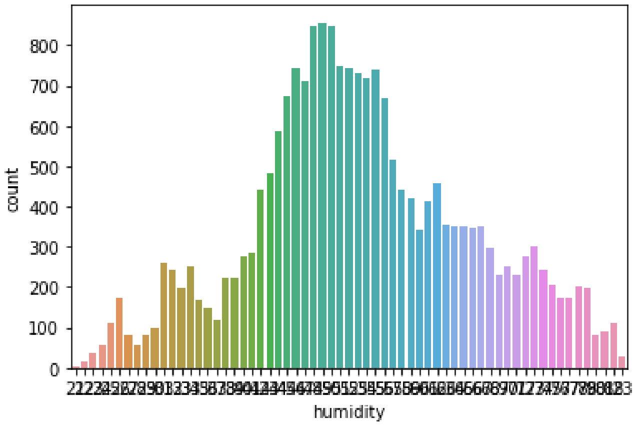

Scatterplot with Histograms
displ vs hwy
Temperature
humidity
humidity

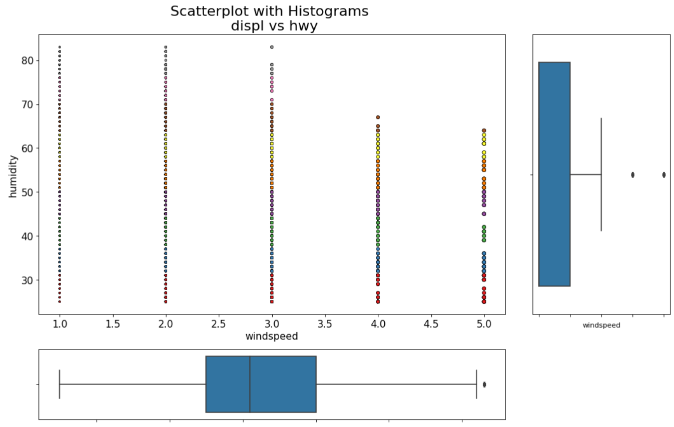

Use jittered plots to avoid overlapping of points
Temperature
humidity


Scatterplot with Histograms
displ vs hwy
humidity
windspeed
windspeed


Use jittered plots to avoid overlapping of points
Temperature
windspeed

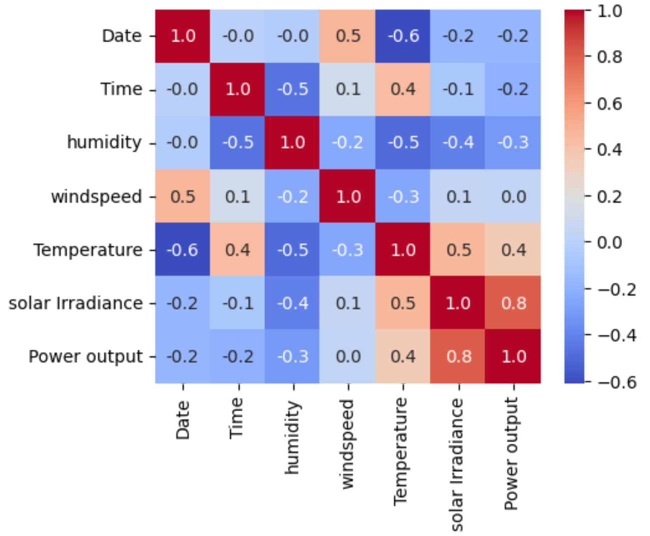

Scatterplot with Histograms
displ vs hwy
Temperature
humidity
humidity


count
humidity


Scatterplot with Histograms
displ vs hwy
humidity
windspeed
windspeed

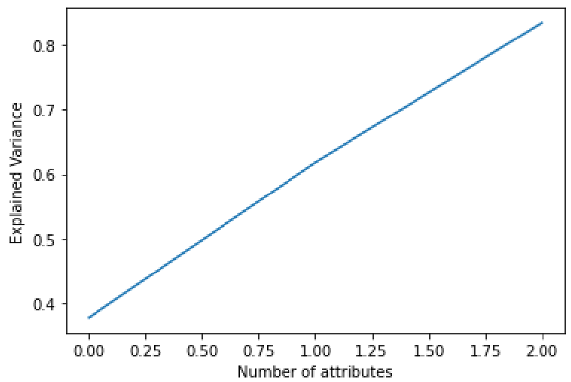
VIII. Prediction Algorithm

Random Forest is a technique that handled both low and high wind speeds quite successfully. The number of features significantly impacts the output of Random Forests. As the number of features rises, Random Forests error declines more quickly than kNN and SVM. Input that is irrelevant does not have a significant impact on the Random Forest. When new features are added, the Random Forests Mean Absolute Percentage Error (MAPE) reduces. The same input given to a neural network results in an increase in MAPE because the model is thrown off by the irrelevant data.



VIII.a. Selection of Neuron Number in Hidden Layer:

Deciding the number of neurons in the hidden layers is crucial in determining your overall neural network architecture. Though these layers do not directly interact with the external environment, they tremendously influence the final output.



The number of hidden layers and the number of neurons in each of these layers must be carefully considered. Using too few neurons in the hidden layers will result in underfitting. Underfitting occurs when too few neurons are in the hidden layers to adequately detect the signals in a complex data set. Using too many neurons in the hidden layers can result in overfitting. When the neural network's ability for processing information is so great that the training set's scant data is insufficient to train every neuron in the hidden layers, overfitting takes place. Even in cases where the training data is adequate, a second issue can arise. The amount of time required for training the neural network may rise to the point where it is impossible to do so. In the buried layers, a balance must be struck between having too many and not enough neurons.

- Between the size of the input layer and the output layer, there should be an appropriate number of hidden neurons
- 2/3 the size of the input layer plus the size of the output layer should be the number of hidden neurons
- Less than twice as many hidden neurons should exist as input layer neurons. It all boils down to trial and error in the end

IX. Challenges of Wind Power

- On a cost level, wind power must still compete with traditional generation sources. Wind projects must economically compete with the least expensive source of electricity, even if the cost of wind energy has fallen dramatically, and certain areas might not be windy enough to be cost-competitive.
- Good land-based wind sites are frequently found in distant areas, away from urban areas with high electricity demand. To get the electricity from the wind farm to the city, transmission lines need to be erected.
- The development of wind resources may not be the most financially successful use of the property. It's possible that land suitable for installation is worth more than producing electricity.
- Turbines could pollute with noise and sight. In comparison to traditional power plants, wind power plants have less of an influence on the environment, yet there are still concerns about the noise they make and how they affect the surrounding environment visually.

----> Fully grown tree

Training the simple model

```
[ ]  DTR_Regressor = DecisionTreeRegressor()
     DTR_Regressor.fit(X_train,y_train)

     DecisionTreeRegressor()

[ ]  X_train_predicted = DTR_Regressor.predict(X_train)
```

Visualizing decision tree

Wind farms may impact local fauna. Birds have been injured after colliding with rotating turbine blades. Through technological advancement or appropriately positioned wind turbines, the majority of these issues have been removed or significantly reduced. Research is still being done to create and enhance ways to lessen the effects of wind turbines on these species. Turbine blades have also killed bats in past.

X. Challenges in wind power forecast

In general, forecasting for wind energy can have a lot of problems. Typically, the algorithms are built on a secondary forecast, such as wind direction, speed, or temperature. These data are weather forecasts, adding even another level of uncertainty to the model.

Evaluation of classifier

1. Model evaluation on training data

```
[ ]  Evaluation_of_classifier(y_train, X_train_predicted)

     Accuracy : 100.0 %
     Error Rate : 0.0 %
```

2. Model evaluation on test set data

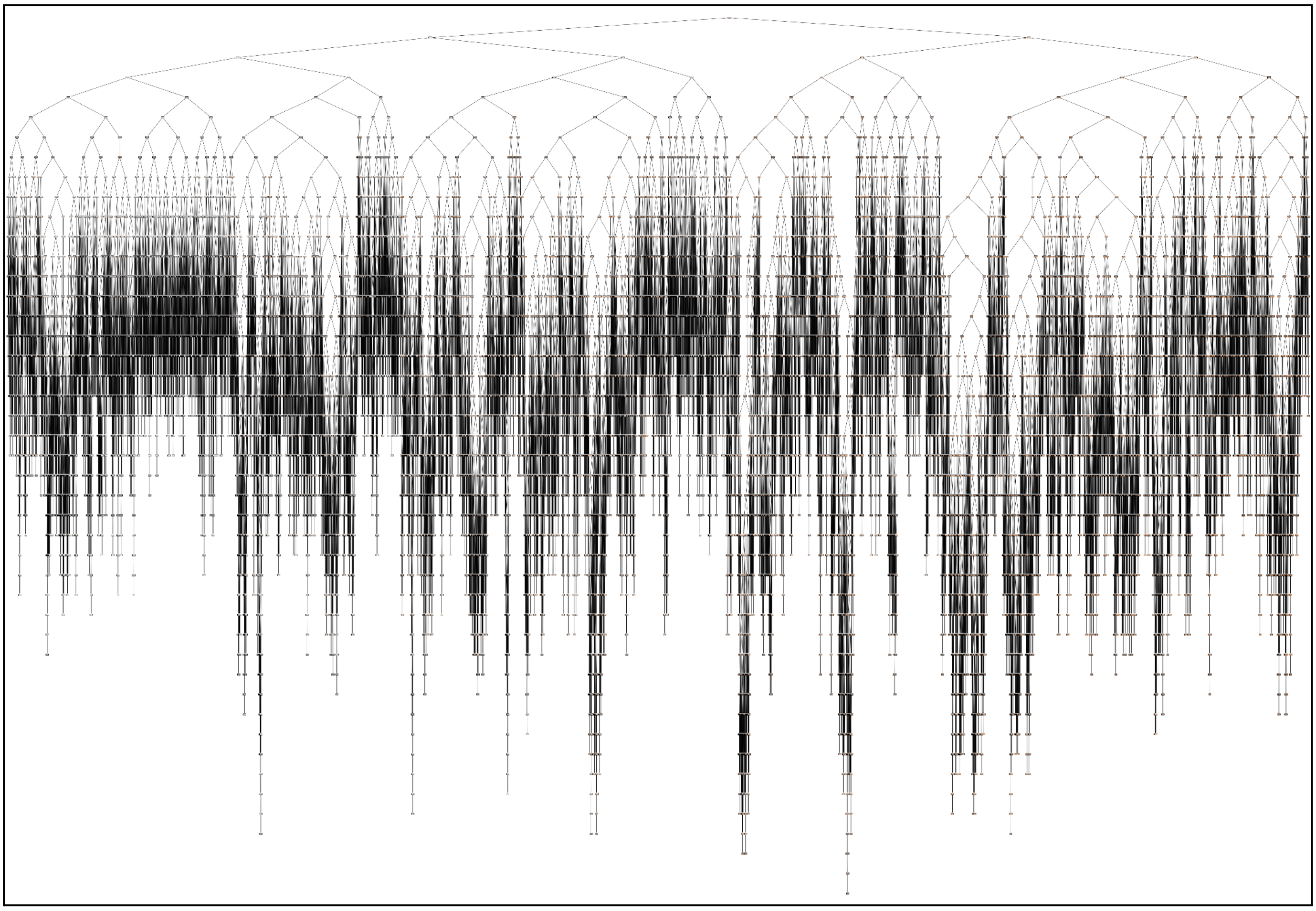
```
[ ]  y_predt = DTR_Regressor.predict(X_test)
     Evaluation_of_classifier(y_test, y_predt)

     Accuracy : 77.0 %
     Error Rate : 23.0 %
```

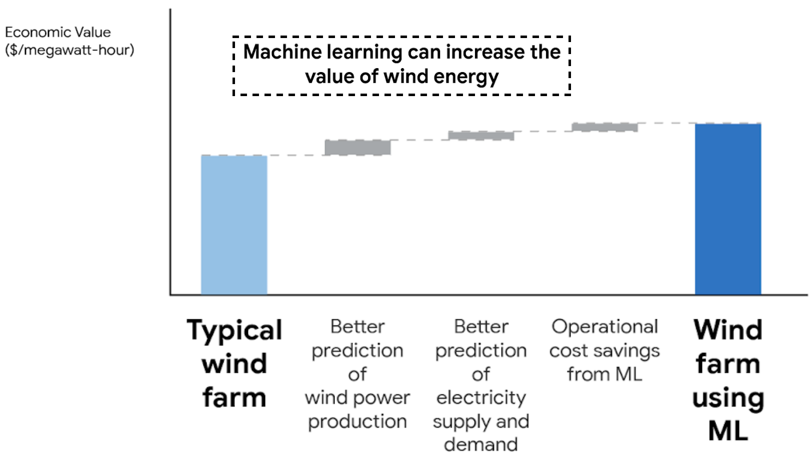
The output data will also contain errors if the input data contains errors. In other words, because the model's data are uncertain, there is a substantial risk that the wind power forecast would be inaccurate. One of the key factors keeping wind power out of the markets is its unavailability, which is crucial to the (transmission system operators) TSO. Analyzing the situation, it appears that the TSO, from the perspective of reserves, is unconcerned with the forecasts' accuracy. The TSO is cautious since there is no guarantee that their purchased services will be available. Allowing wind turbines to participate if the wind power forecast is higher than a certain threshold would be one way to address this problem. This permits mistakes to happen without affecting the level of service.

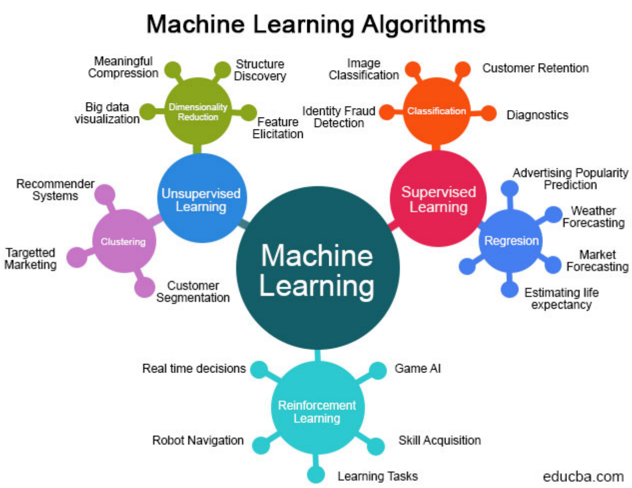
XI. Conclusion

The short-term forecasting of wind energy generation, is done in this article using machine learning, and the data is presented as a case study. The forecast for energy generation is crucial because it can have a big impact on cost-effectiveness and improved demand-side management. For this forecasting, a prediction algorithm is also used. The output of the algorithm is contrasted with that of Decision tree approach. The root mean accuracy and root mean square error (RMSE) of both techniques are determined. In the future, this method can be put up against a neural network-based forecasting algorithm. The cost increases are a result of the growth in wind energy. The flexibility of neural networks and decision trees has led to a rise in interest in these techniques. Additionally, the algorithms struggle with slow calculation times and irrelevant input data. Less research has been done on kNN and SVM, however the studies that have been done have produced some remarkable outcomes. With minimal features, kNN has a low computation time and error. The difficulty with kNN comes from its inability to anticipate outliers and make predictions across seasons. Additionally, it was discovered that Random Forrest could anticipate outliers and discard irrelevant input data. This essay can be used to determine which approach would work best in a certain situation. By providing a more reliable, autonomous, and human-error-free wind power prediction, a stronger forecasting model could be a step in the right way. The fundamental trend in the contemporary power system is the integration of large amounts of wind power into the grid. However, because wind power output is unpredictable and volatile, forecasting wind power output accurately is important to ensure the steady operation of the power grid. Using the random forest approach after figuring out the ideal hidden layer neuron number, we have introduced the fundamentals of machine learning neural networks in this study. To confirm the usefulness of the machine learning random forest, we compare the simulation results with the decision tree technique. The precision of the two algorithms can be seen from simulation results to be different.



Highly overfitted decision tree is obtained





XII. References

1. https://en.wikipedia.org/wiki/Wind_power_forecasting
2. https://en.wikipedia.org/wiki/Wind_power_in_India#:~:text=Wind%20power%20generation%20capacity%20in,Western%20and%20Northern%20Western%20regions
3. https://www.google.com/search?q=1080p+wind+turbine+hd&tbm=isch&hl=en-GB&sa=X&ved=2ahUKEwib05Psj5b5AhXqgWMGHfQdDMkQrNwCKAB6BQgBEPIB&biw=1519
4. https://www.google.com/search?q=random+forest+algo+hd+flowchart&tbm=isch&ved=2ahUKEwiG65D2iJf5AhVS0qACHSqiDQQQ2-cCegQIABAA&oq=random+forest+algo+hd+flowchart&gs
5. https://www.scirp.org/journal/paperinformation.aspx?paperid=106810
6. https://www.google.com/search?q=rmse+mae+mse&tbm=isch&ved=2ahUKEwiBmoKvlJf5AhWq_igGHaiDBvcQ2-cCegQIABAA&oq=rmse+mae+mse&gs_lcp=CgNpbWcQAzIGCAAQHhAIOgQIIxAnOggIABCxAxCDAToICAAQgAQQsQM6BAgAEEM6CwgAEIAEELEDEIMBOgcIABCxAxBDOgoIABCxAxCDARBDOgUIABCABDoGCAAQHhAFOgQIABAYOgQIABAeUK4IWKxUYIdWaABwAHgAgAGuAYgB9w-SAQQwLjEzmAEAoAEBqgELZ3dzLXdpei1pbWfAAQE&sclient=img
7. https://www.google.com/search?q=wind+power+prediction+with+machine+learning&oq=wind+power+predition&aqs=chrome.1.69i57j0i13l4j0i15i22i30j0i22i30j0i15i22i30l2j0i22i30.5719j0j9&sourceid=chrome&ie=UTF-8
8. https://www.researchgate.net/publication/268966331_Wind_Power_Prediction_with_Machine_Learning_chapter
9. https://www.researchgate.net/publication/341219336_Forecasting_of_Wind_Turbine_Output_Power_Using_Machine_learning
10. https://www.google.com/search?q=Prediction+of+wind+power+generation+based+on+time+series+wavelet+transform+for+large+Wind+Farm&oq=Prediction+of+wind+power+generation+based+on+time+series+wavelet+transform+for+large+Wind+Farm&aqs=chrome..69i57j69i61l2.432j0j4&sourceid=chrome&ie=UTF-8
11. https://www.google.com/search?q=A+double-stage+hierarchical+ANFIS+model+for+short-term+wind+power+prediction&oq=A+double-stage+hierarchical+ANFIS+model+for+short-term+wind+power+prediction&aqs=chrome..69i57j69i61l2.1041j0j4&sourceid=chrome&ie=UTF-8
12. https://www.google.com/search?q=Hour-ahead+wind+power+prediction+for+power+system+using+quadratic+fitting+function+with+variable+coefficients&oq=Hour-ahead+wind+power+prediction+for+power+system+using+quadratic+fitting+function+with+variable+coefficients&aqs=chrome..69i57j69i61l2.361j0j4&sourceid=chrome&ie=UTF-8
13. https://www.sciencedirect.com/science/article/pii/S0045790614001876
14. https://www.google.com/search?q=Wind+Power+Day-ahead+Prediction+Based+on+LSSVM+With+Fruit+Fly+Optimization+Algorithm&oq=Wind+Power+Day-ahead+Prediction+Based+on+LSSVM+With+Fruit+Fly+Optimization+Algorithm&aqs=chrome..69i57j69i61l2.492j0j7&sourceid=chrome&ie=UTF-8
15. https://pcmp.springeropen.com/articles/10.1186/s41601-017-0041-5
16. https://www.google.com/search?q=Wind+power+prediction+model+considering+smoothing+effects&oq=Wind+power+prediction+model+considering+smoothing+effects&aqs=chrome..69i57j69i61l2.323j0j4&sourceid=chrome&ie=UTF-8
17. https://www.google.com/search?q=Research+on+Distributed+wind+Power+Reactive+Voltage+Coordinated+Control+Strategy+Connected+to+Distribution+Network&oq=Research+on+Distributed+wind+Power+Reactive+Voltage+Coordinated+Control+Strategy+Connected+to+Distribution+Network&aqs=chrome..69i57j69i61l2.381j0j4&sourceid=chrome&ie=UTF-8
18. https://scholar.google.co.in/scholar?q=Ultra-short-term+Wind+Power+Forecasting+Based+on+Improved+LSTM&hl=en&as_sdt=0&as_vis=1&oi=scholart
19. https://www.google.com/search?q=Wind+power+prediction+using+wavelet+transform+and+chaotic+characteristics&oq=Wind+power+prediction+using+wavelet+transform+and+chaotic+characteristics&aqs=chrome..69i57j69i61l2.413j0j4&sourceid=chrome&ie=UTF-8
20. https://www.researchgate.net/publication/227421368_GIS-based_approach_for_the_evaluation_of_wind_energy_potential_A_case_study_for_the_Kujawsko-Pomorskie_Voivodeship
21. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8602198

22. https://www.semanticscholar.org/paper/Wind-power-prediction-model-considering-smoothing-Biyun-Suifeng/5d85e10dc8ca57c5270974271079b9bf60588575
23. https://scholar.google.com/scholar?as_q=Ultra-short-term+Multi-step+Wind+Power+Prediction+Based+on+Improved+EMD+and+Reconstruction+Method+Using+Run-length+Analysis%5BJ%5D&as_occt=title&
24. https://scholar.google.com/scholar?as_q=A+review+on+the+forecasting+of+wind+speed+and+generated+power%5BJ%5D&as_occt=title&hl=en&as_sdt=0%2C31
25. https://scholar.google.com/scholar?as_q=Current+methods+and+advances+in+forecasting+of+wind+power+generation%5BJ%5D&as_occt=title&hl=en&as_sdt=0%2C31
26. https://scholar.google.com/scholar?as_q=Recurrent+neural+networks+for+short-term+load+forecasting%5BJ%5D&as_occt=title&hl=en&as_sdt=0%2C31
27. https://www.google.com/search?q=random+forest+modeling&source=lmns&bih=656&biw=1519&hl=en-GB&sa=X&ved=2ahUKEwi2-vjo8Zr5AhWv9zgGHcBZBXYQ_AUoAHoECAEQAA
28. https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/#:~:text=Random%20forest%20is%20a%20Supervised,average%20in%20case%20of%20regression.
29. https://editor.analyticsvidhya.com/uploads/325745-Bagging-ensemble-method.png
30. https://www.analyticsvidhya.com/blog/2022/07/data-science-interview-series-part-2-random-forest-and-svm/
31. https://courses.analyticsvidhya.com/courses/Machine-Learning-Certification-Course-for-Beginners?utm_source=blog_navbar&utm_medium=start_here_button
32. https://towardsdatascience.com/table-of-contents-689c8af0c731
33. https://www.google.com/search?q=Random+Forest%3A+A+Classification+and+Regression+Tool+for+Compound+Classification+and+QSAR+Modeling&oq=Random+Forest%3A%E2%80%89+A+Classification+and+Regression+Tool+for+Compound+Classification+and+QSAR+Modeling&aqs=chrome..69i57j69i60.762j0j4&sourceid=chrome&ie=UTF-8
34. https://www.google.com/search?q=Application+of+Breiman%E2%80%99s+Random+Forest+to+Modeling+Structure-Activity+Relationships+of+Pharmaceutical+Molecules&oq=Application+of+Breiman%E2%80%99s+Random+Forest+to+Modeling+Structure-Activity+Relationships+of+Pharmaceutical+Molecules&aqs=chrome..69i57j69i61l2.953j0j4&sourceid=chrome&ie=UTF-8
35. https://towardsdatascience.com/10-must-read-machine-learning-articles-march-2020-80da9c380981
36. https://www.frontiersin.org/articles/10.3389/fdata.2018.00006/full
37. https://www.google.com/search?q=Very+short-term+probabilistic+wind+power+prediction+using+sparse+machine+learning&oq=Very+short-term+probabilistic+wind+power+prediction+using+sparse+machine+learning&aqs=chrome..69i57j69i61.353j0j4&sourceid=chrome&ie=UTF-8
38. https://www.google.com/search?q=Forecasting+Wind+Power+Generation+Using+Artificial+Neural+Network%3A+%E2%80%9CPawan+Danawi%E2%80%9D%E2%80%94A+Case+Study&oq=Forecasting+Wind+Power+Generation+Using+Artificial+Neural+Network%3A+%E2%80%9CPawan+Danawi%E2%80%9D%E2%80%94A+Case+Study&aqs=chrome..69i57j69i61.403j0j4&sourceid=chrome&ie=UTF-8
39. https://www.researchgate.net/publication/348795571_Wind_Energy_Prediction_Using_Machine_Learning
40. https://www.researchgate.net/figure/Random-Forest-logic_fig2_348795571
41. https://www.researchgate.net/publication/359704498_Exploring_Wind_Speed_for_Energy_Considerations_in_Eastern_Jerusalem-Palestine_Using_Machine-Learning_Algorithms
42. https://www.researchgate.net/publication/354075433_A_Machine_Learning-Based_Gradient_Boosting_Regression_Approach_for_Wind_Power_Production_Forecasting_A_Step_towards_Smart_Grid_Environments
43. https://www.researchgate.net/publication/360462221_A_Novel_Framework_Based_on_the_Stacking_Ensemble_Machine_Learning_SEML_Method_Application_in_Wind_Speed_Modeling
44. https://www.researchgate.net/publication/361570249_Machine_Learning-Based_Analysis_of_a_Wind_Turbine_Manufacturing_Operation_A_Case_Study
45. https://www.google.com/search?q=machine+learning+hidden+layers+selection+methods+explained+through&oq=machine+learning+hidden+layers+selection+methods+explained+through+&aqs=chrome..69i57.18059j0j4&sourceid=chrome&ie=UTF-8
46. https://towardsdatascience.com/beginners-ask-how-many-hidden-layers-neurons-to-use-in-artificial-neural-networks-51466afa0d3e