

## Efficient Energy Management Using Artificial Intelligence (AI) and Machine Learning (ML) in Chemical Industry

Virendra Yadav<sup>1</sup>, Simmi Raj Chaudhary<sup>1</sup>, Ravindra Kumar<sup>2,\*</sup>

### Abstract

The globe is moving toward higher usage of renewable energy sources, particularly solar and wind energy, as a result of depleting fossil fuel supplies and growing environmental concerns. There are several forecasting methods available for effective wind energy utilization. This research article uses algorithms for predicting solar and wind energy as well as artificial intelligence (AI) techniques. A wind-coal coupling energy system planning scheme was designed to lower the high energy consumption and pollution generated by the coal chemical industries, as well as to increase the energy industry's rate of utilization. This plan called for the integration of a thorough assessment framework through cleaning, pre-processing, analysis, and algorithm operation check. This research, though centred on the chemical industry, yields findings applicable to artificial intelligence in a range of sectors and to industrial ecology at large. The machine learning algorithm implemented here, evaluated against multiple real-world network datasets, significantly streamlined pre-classification processing and yielded improved prediction accuracy relative to other algorithms. Industries need assistance in handling the complexity, uncertainty, and fuzziness inherent in this domain. New approaches are needed for every facet of the chemical industries. Climate change will cause variations in the generation and intermittency of solar and wind energy resources. Indian solar and wind park developers have noted variations in the seasonal and annual climates, as well as variations in sun irradiation and wind profiles.

**Keywords:** Solar power prediction, regression models, climate change, wind energy, solar PV, data analysis, machine learning

### INTRODUCTION

Over the past few decades, there has been a growing concern for energy efficiency in industrial activities, including mining, oil and gas extraction, chemicals, pulp and paper, and others. The chemical industry has employed AI and ML to lower environmental impacts, improve energy efficiency, and ensure operational safety [1]. A design strategy for a wind-coal coupling energy system was established in order to increase the energy industry's rate of utilization and decrease the high energy consumption and pollution generated by the coal chemical industries [2, 3]. Algorithms are used to forecast future usage and, based on those forecasts, decrease costs [3, 4]. A few significant algorithm are used which assist in making wise decisions. Decision tree and linear regression algorithms were used. The chemical industry is becoming more and more interested in leveraging AI to solve problems related to fault detection and diagnosis, control, optimization, and process modelling. Numerous

#### \*Author for Correspondence

Ravindra Kumar  
E-mail: rk@bietjhs.ac.in

<sup>1</sup>Student, Department of Chemical Engineering, Bundelkhand Institute Engineering and Technology, Jhansi, Uttar Pradesh, India

<sup>2</sup>Assistant Professor, Department of Chemical Engineering, Bundelkhand Institute Engineering and Technology, Jhansi, Uttar Pradesh, India

Received Date: March 04, 2025  
Accepted Date: March 06, 2025  
Published Date: March 27, 2025

**Citation:** Virendra Yadav, Simmi Raj Chaudhary, Ravindra Kumar. Efficient Energy Management Using Artificial Intelligence (AI) and Machine Learning (ML) in Chemical Industry. Journal of Modern Chemistry & Chemical Technology. 2025; 16(2): 33–50p.

factors are associated with the chemical industry's sustainability in terms of the environment, economy, and society. Accenture's survey reveals that 94% of chemical and advanced materials executives expect industry-wide digitalization, with AI at its core. While technical AI applications in the chemical sector are documented, the critical area of AI's sustainability implications remains unexplored. Though AI's role in general sustainable development is recognized, chemical industry-specific studies are needed [1–5].

## METHODOLOGY

### Data Preparation

1. *Data Collection*: Employ sensors, IoT devices, and SCADA systems to capture data as it occurs.
2. *Data Cleaning*: Address gaps, anomalies, and inconsistencies in the raw data through cleaning and preprocessing.

### Data Integration

Consolidate data from diverse sources into a centralized repository to facilitate comprehensive analysis.

### Model Development

1. *Predictive Maintenance Models*: Create machine learning models for predictive maintenance, concentrating on vital systems and equipment.
2. *Process Optimization Models*: Develop models to optimize parameters in real time while taking production quality and energy efficiency into account.

### Human-Machine Collaboration Interfaces

1. *User Interface Design*: Provide intuitive user interfaces so that operators can work with AI systems.
2. *Feedback Mechanism*: Create a feedback loop that incorporates operator insights for ongoing improvement.
3. *Energy Consumption Forecasting*: To anticipate patterns in energy usage, use time-series forecasting models [6, 7].

## ALGORITHMS OF AI AND ML IN DIFFERENT TYPES OF ENERGY

The application of machine learning (ML) in energy systems has experienced explosive growth over the last 20 years (Figure 1). From modelling and design to predictive analysis, ML techniques have become integral tools. This study aims to provide a comprehensive overview of the current landscape of ML models employed within the energy sector [6].

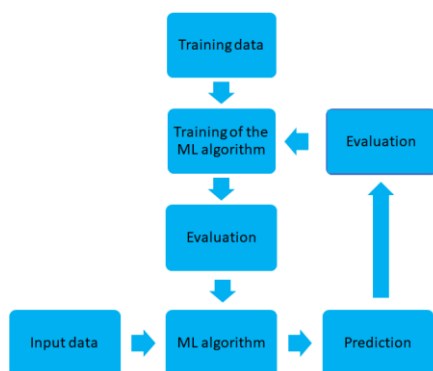


Figure 1. Machine learning process.

By employing an innovative methodology, this work pinpoints and organizes ML models according to their modelling type, energy source, and practical implementation. In addition, a comprehensive literature survey yields a detailed assessment of these models' efficacy and applications, culminating in an examination of key challenges and promising directions for subsequent investigation [8–10].

### ANN

Artificial neural networks (ANNs) provide a powerful architecture for handling intricate datasets through diverse machine learning techniques. Their versatility allows for applications spanning curve fitting, regression analysis, and predictive forecasting. The fundamental building block of an ANN is the neuron, which employs a transfer function to generate its output. Notably, ANNs excel in multi-variable contexts by simplifying complex relationships. In this study, an ANN was employed to validate the proposed load model. To account for inherent uncertainties within renewable energy systems, a chance-constrained modelling approach was adopted. Subsequently, genetic algorithms were leveraged to optimize the solution.

### Decision Trees

The learned function is represented by a decision tree, and the decision tree approach is used to approximate discrete-valued target functions. These techniques are among the most effective algorithms for inductive inference and are effectively applied in a wide range of energy systems. A decision tree approach is suggested to forecast a specific smart city's energy usage. Energy resource utilization is optimized through the use of decision trees. Consideration is given to an agricultural system made up of a synchronous generator, photovoltaic panels, and an electrical network for powering the electric motors. Forecasts of consumption are supplied based on various prediction algorithms in order to bolster the decision-making guidelines acquired during the process of learning. These choices confirm whether the forecasting algorithm chosen was the best option each period. The approach: The first step in the process is to obtain the energy predicted consumptions based on several algorithms. Next, based on modified decision tree depth updates, decision rules are used to support the forecasting algorithms' picks as the optimal options. This is known as decision rule-based learning. The number of splits the decision tree makes in order to produce the forecasts depends on its depth (Figure 2) [12].

### Deep Learning

The ability of deep learning to model complex, hierarchical patterns within data through stacked processing layers has fuelled its growing prevalence, particularly as computational resources and data volumes expand. This surge in deep learning's application coincides with the escalating challenges faced by modern electric power systems. Factors such as the rapid integration of renewable energy, increasing electricity consumption, and the imperative for enhanced grid resilience necessitate advanced technological solutions. Deep learning, as a specialized branch of machine learning, emerges as a promising approach to address these complexities.

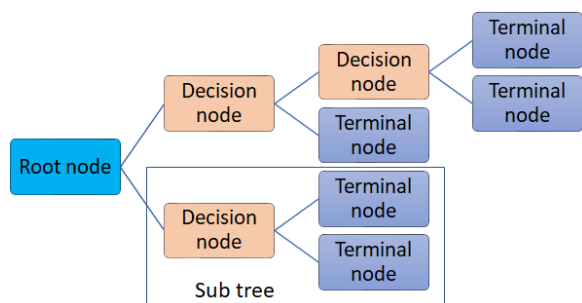


Figure 2. Decision tree.

This study explores the diverse applications of deep learning in electric power systems, including comprehensive data analysis, accurate load forecasting, rapid fault detection and diagnosis, robust power system security and stability evaluation, efficient integration and management of renewable energy, and proactive asset management and maintenance of the electrical grid. While deep learning offers significant potential, its implementation is not without obstacles. Key challenges include ensuring data accessibility and quality, managing computational demands, optimizing resource allocation, improving model interpretability, and seamlessly integrating these techniques with existing power system infrastructure and tools.

#### **K-nearest Neighbours (KNN)**

In machine learning, the KNN algorithm is widely used for its simplicity and effectiveness in classification and regression. The algorithm's core idea is that proximity implies similarity. When presented with a new data point, KNN searches the entire training set to find the 'K' most similar points, determined by a distance metric. For classification, the prediction is the most frequent class among these neighbours; for regression, it is the average or weighted average of their target values.

#### **Linear Regression**

In energy optimization challenges, linear regression is used to estimate energy usage and provide guidance for building design decisions. The objective is to identify the most energy-efficient building configuration that satisfies particular constraints and needs. Building configurations' predicted yearly heating energy demand is estimated using multivariate linear regression models, which makes it possible to compare various configurations more quickly and without the need for precise energetic simulations. An energy-efficient linear-regression-based distributed data gathering optimization strategy is proposed for wireless sensor networks, where nodes transfer constraints on model parameters instead of transmitting measurements, thereby extending the lifetime of the network and reducing communication costs. Additionally, big data analytics for energy demand management uses linear regression models to analyse the relationship between design parameters and building energy performance, helping architects with the early stages of building design and offering recommendations for energy optimization [6, 7, 9–11].

#### **PROCESS APPLIED IN WIND ENERGY**

Wind energy stands out as a leading contender among renewable resources, garnering global attention. Its potential to mitigate environmental degradation is substantial. However, the inherent variability of wind resources and the resulting operational instability present significant challenges to maximizing its utilization. Consequently, accurate wind power and wind speed forecasting are crucial for the effective deployment of this renewable energy source [13, 14].

#### **Data Frame**

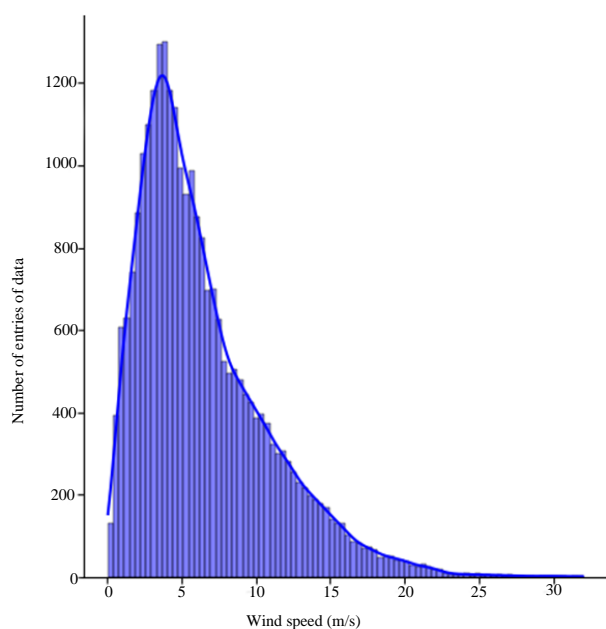
In this Firstly, we have collected different data from 2007 to 2009 known as dataset. Here we have some dataset of wind energy in which five columns and 26279 row are there, there are many default data, missing values, and some ununiformed data present. We will handle this type of data using cleaning, preprocessing and data visualization. As the data is collected, it is then further sent for data cleaning where unnecessary data is removed. Table 1 shows the dataset we have collected, which consist of different data of pressure, wind direction, wind energy power generation and wind speed.

#### **Analysis of Data Frame**

After the data is cleaned, it is sent for preprocessing, where it is transformed into a comprehensible format and the quality of the data is checked. We have gathered various data on pressure, wind direction, wind energy power generation, and wind speed, and we have plotted the data on this basis to create a graph that shows us the amount of power generated and the amount of wind speed that has been observed over time. The dataset is evaluated, and various graphs are created to help visualize the results. The wind speed and electricity generation are among the studied statistics, and some details about them are shown below. Now let us examine the many graphs that were plotted during the analysis.

**Table 1.** Data head and data tail of the collected data.

Date	Time	Wind direction (deg)	Wind speed (m/s)	Power generated by system (kW)
2007-01-01	00:00:00	104	3.429	0
2007-01-01	01:00:00	106	3.579	0
2007-01-01	02:00:00	102	4.307	1986.68
2007-01-01	03:00:00	99	4.562	2597.61
2007-01-01	04:00:00	98	4.553	2555.46
2007-01-01	05:00:00	98	4.695	2922.34
2007-01-01	06:00:00	98	4.713	2972.42
2007-01-01	07:00:00	98	4.491	2396.37
2007-01-01	08:00:00	96	4.549	2498.91



**Figure 3.** Wind speed vs. number of entries of data.

### Wind Speed

A histogram, which shows the frequency of various wind speed ranges and is frequently expressed in bins or intervals, may be included in the wind speed display.

Viewers can learn a great deal about the current wind conditions by studying the wind speed **displot**. This includes the range of wind speeds encountered, the frequency of occurrence for different speeds, and any patterns or outliers within the data (Figure 3).

### Wind Power Generation by System

The amount of electricity produced by wind turbines by how many entries, is shown visually in a wind power generating bar plot. The vertical axis in this type of plot typically represents the corresponding levels of power generation, often measured in megawatts (MW) or kW (Figure 4).

With the aid of machine learning, we can manage and regulate the maximum and minimum power generation on a large scale. This information provides some basic knowledge about the power generated by the system by analysing the entire column of power generated.

1. Maximum Power Generated by system (kW): 56452.4
2. Minimum Power Generated by system (kW): 0.0
3. Mean of System Power Generated: 15585.14
4. Median of System Power Generated: 5130.15
5. Mode of System Power Generated: 0.00

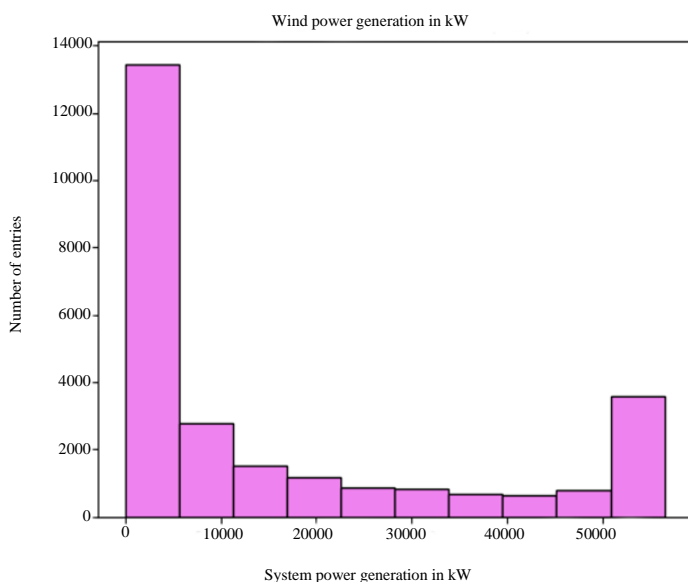
**Correlation Between Each Column**

Since the data has already been pre-processed, analysis of the data is currently being done through inspection, cleansing, transformation, and modelling in order to find relevant information that will help conclusions and decision-making. Here, we use the SNS to draw a graph inside each column.plt.show() pairplot(df). This is also used as a display of training data.

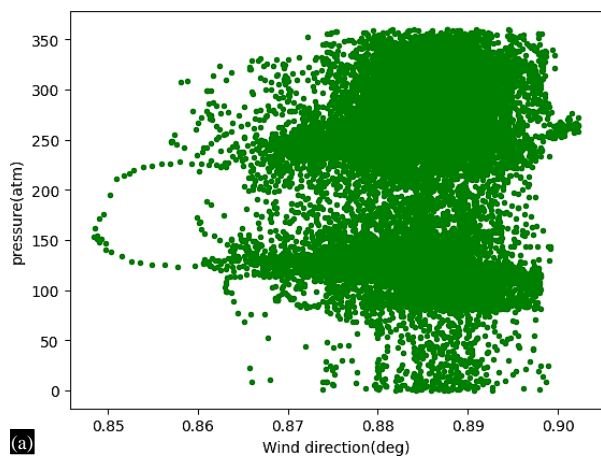
A graphical technique for visualizing the relationships between several variables in a dataset is called a pair plot, or pairs plot. A pair plot offers a thorough understanding of how factors related to power generation by a system, wind direction, wind speed, and column pressure interact with one another. A grid of scatterplots for numerical variables is produced when each variable in the dataset is displayed against every other variable in a pair plot (Figure 5(a-d)).

**Energy Prediction by Linear Regression**

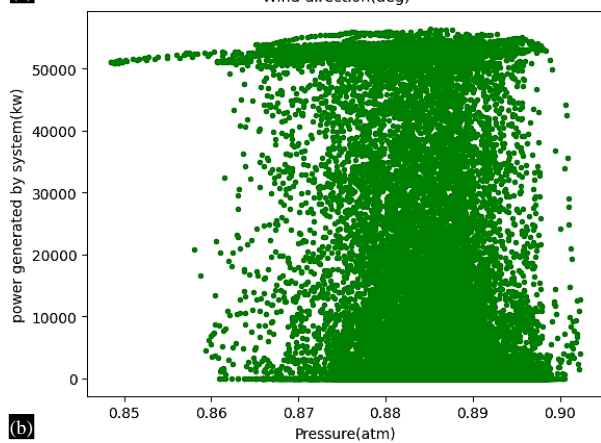
This graph uses linear regression analysis to show how wind speed and electricity generation in a wind energy system relate to one another. Power generation is the dependent variable (y-axis), and wind speed is the independent variable (x-axis).



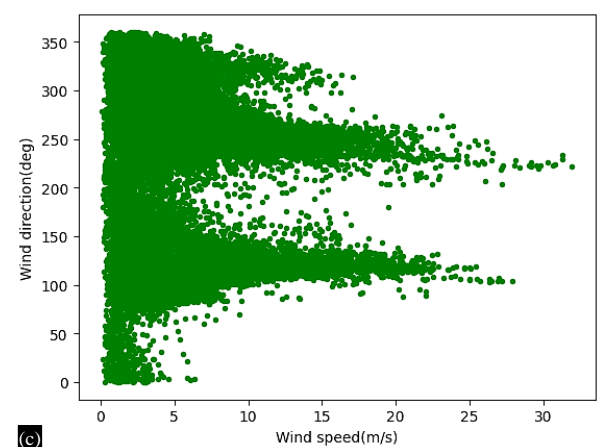
**Figure 4.** Wind power generation by system Information about power generated by the system.



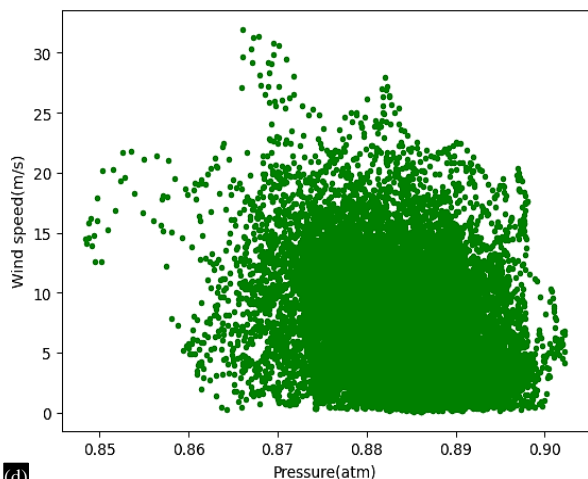
(a)



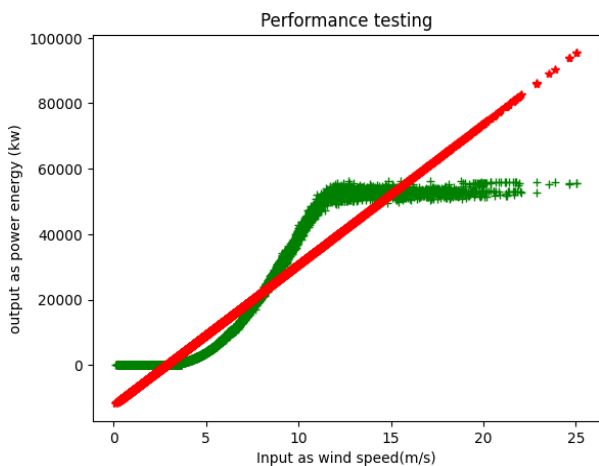
(b)



(c)



**(d)** **Figure 5.** (a) Pressure vs. Wind direction, (b) power generated by system vs. pressure, (c) wind direction vs. wind speed, (d) wind speed vs. pressure.



**Figure 6.** Wind energy prediction by linear regression.

The data points show the observed values that were gathered over a given time period from the wind turbines. Each point on the graph represents a combination of the wind speed and the equivalent power produced by the system. The graph shows a line graph of the recorded data points. In addition, the linear regression model that was fitted to the data is shown as a straight line on the graph (Figure 6).

**Accuracy of Predicted Value by Linear Regression**

1. An average absolute deviation of approximately 80.71 units was observed, indicating the model's predictive accuracy. This Mean Absolute Error (MAE) value, representing the typical difference between predicted and observed wind energy, suggests a strong agreement between the model's output and actual measurements.

2. The model's predictive accuracy was evaluated using Mean Squared Error (MSE), yielding a value of approximately 49,793.78. This metric, which represents the average of the squared deviations between predicted and observed wind energy values, suggests a high degree of precision, as a lower MSE generally signifies superior model performance.
3. A median absolute error of 0.0 units was observed, signifying that the model achieved perfect accuracy for half of the predicted wind energy values. This suggests a very strong alignment between the model's output and the actual measurements.
4. Accuracy Result: The accuracy result of the decision tree model is reported to be approximately 99.99%. This exceptionally high accuracy indicates that the model makes accurate predictions for the majority of cases, with very few errors.
5.  $R^2$  Score: The coefficient of determination,  $R^2$ , was found to be about 0.886, suggesting that wind speed accounts for a substantial portion of the variance in wind energy. In general, a higher  $R^2$  demonstrates improved model predictability.
  - Mean absolute error value is: 5116.1384743228
  - Mean squared error value is: 40626798.9171675
  - Median absolute error value is: 4614.788012560377
  - Accuracy result of linear regression is: 88.6313213416291
  - $R^2$  score value is: 0.886313213416291

#### WIND POWER PREDICTION BY DECISION TREE

The link between column pressure, wind direction, wind speed, and power generation by a system is investigated in this decision tree study. The decision tree approach predicts the output variable, in this case electricity generation, by recursively partitioning the feature space according to the chosen input variables. The decision tree's internal nodes represent decision points depending on the values of input features, whereas the root node represents the complete dataset. The dataset is divided into subsets by each decision node according to a particular attribute and the associated threshold value. Based on the majority class or the average value of the target variable within that subset, a power generation prediction is created at each leaf node of the decision tree. The training set of data that went into building the decision tree is the foundation for these forecasts. Using a hierarchical framework of rules or conditions to decide power generation levels based on input variables, the decision tree illustrates the decision-making process. Stakeholders can comprehend the circumstances under which power generation is anticipated to be high or low by observing the tree's branches (Figure 7).

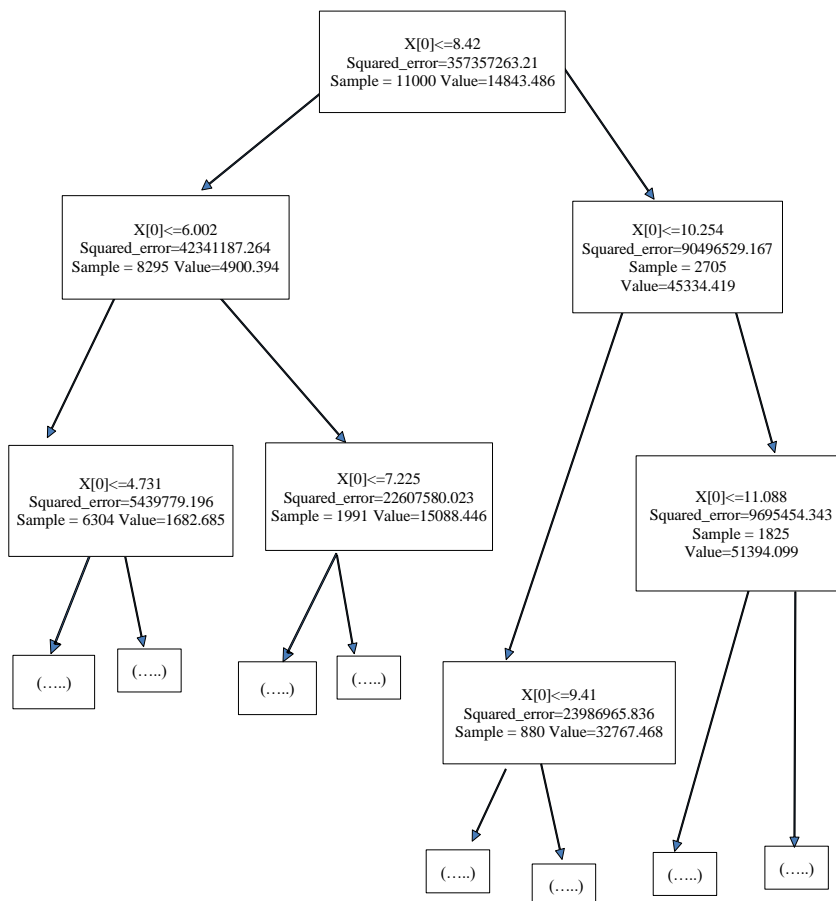
#### Accuracy of Predicted Value by Decision Tree for Wind Energy

1. *Mean Absolute Error (MAE)*: An average absolute deviation of approximately 80.71 units was observed. This metric, known as the Mean Absolute Error (MAE), quantifies the typical difference between predicted and actual wind energy values. The relatively low MAE suggests a strong alignment between the model's estimations and the recorded data.
2. *Mean Squared Error (MSE)*: A mean squared error (MSE) of 49793.78 units was obtained. This value represents the average of the squared discrepancies between predicted and observed wind energy output. Lower MSE values indicate superior model performance, reflecting a closer alignment between predictions and actual measurements.
3. The median absolute error is 0.0 units, meaning that exactly half of the absolute errors resulting from the difference between the actual and forecast wind energy levels are zero. This implies that the model's forecasts are extremely accurate and closely match the observed data.
4. *Accuracy Outcome*: The decision tree model's accuracy output is said to be roughly 99.99%. This very high accuracy shows that the model has very few errors and can predict the majority of situations accurately.
5.  *$R^2$  Score*: The  $R^2$  value is incredibly near to 1, at about 0.9999. This shows that the model has an excellent fit to the data, explaining nearly all of the variability in the dependent variable (wind energy) using the independent variable (wind speed) [13–16].
  - Mean absolute error value is: 80.70680956363636

- Mean squared error value is: 49793.780645305626
- Median absolute error value is: 0.0
- Accuracy result of decision tree is: 99.98606610645116
- R<sup>2</sup> score value is: 0.9998606610645117

**PROCESS APPLIED IN SOLAR ENERGY**

In order to make solar energy plants more competitive in the energy market and less dependent on fossil fuels for social and economic advancement, solar energy forecasting is essential. Predictions of solar power are useless based on statistics about sun radiation at night. In terms of solar energy, the chemical industry's integration of ML and AI technologies in solar energy applications facilitates more sustainable and effective energy production processes, improves operational resilience, and aids in the shift to cleaner and more renewable energy sources in the future. Artificial Intelligence (AI) and Machine Learning (ML) are being used more and more in the chemical industry to improve efficiency, streamline procedures, and cut costs in various areas of solar energy generation and management [12, 17, 18].



**Figure 7.** Wind energy predication by decision tree.

### Data Frame

First, we have gathered various data (referred to as the dataset) from 2016 to 2017. This dataset, which contains 7532 rows and 8 significant columns, contains a large number of default values, missing values, and inconsistent data. We will use data visualization, pre-processing, and cleaning to deal with this kind of data. After the data is gathered, it is sent for data cleaning, where extraneous data is eliminated. The dataset we have gathered is displayed in Table 2 and includes various statistics on "cloud coverage", "visibility", "temperature", "dew point", "relative humidity", "wind speed", "station pressure", and "solar energy". In this, firstly, we have collected different data 2016 known as dataset. Here we have some dataset of solar energy

### Analysis of Data Frame

Now, as the data is cleaned, it is then sent for pre-processing where the data is converted into an understandable format, where quality of data is checked; we have collected different data of pressure, wind direction, wind energy power generation and wind speed, on the basis of which we have plotted the graph which tells us that how much power has been generated and how much wind speed has been observed throughout. Analysis of dataset is done and different graphs have been drawn to understand the analysed data. The analysed data consist of wind speed, power generation and some information regarding them is presented below. Let us now understand the different graphs plotted during analysis.

**Table 2.** Data head and data tail of solar energy.

Date	Hour	Cloud coverage	Visibility	Temperature	Dew point	Wind speed	Station pressure	Solar energy
01-02-2016	6	0	9.92	0.37	-0.01	4.72	29.19	0
01-02-2016	7	0	10	0.47	-0.04	6	29.2	84.29
01-02-2016	8	0.14	9.92	0.89	0.3	2.76	29.22	531
01-02-2016	9	0.92	10	1.97	0.41	5.56	29.24	923.75
01-02-2016	10	0.46	10	3.15	1.07	2.12	29.27	1947.75
01-02-2016	11	0.42	10	4.5	1.05	0	29.27	3562
01-02-2016	12	0	10	6.16	1.26	0	29.26	4236.5
01-02-2016	13	0	10	7.56	1.41	0	29.25	3913
01-02-2016	14	0	10	8.47	0.63	3.72	29.24	3189.25
01-02-2016	15	0	10	8.78	-0.2	2.04	29.23	1687.75
01-02-2016	16	0	10	8.49	-0.21	3.2	29.24	453.93
01-02-2016	17	0	10	5.86	0.44	3.16	29.26	9.13
02-02-2016	6	0.92	5.04	2.07	0.54	14.08	29.19	0
02-02-2016	7	0.92	5.76	2.51	0.67	18	29.13	31.35
02-02-2016	8	0.92	5	2.63	1.06	17.92	29.08	116.1
02-02-2016	9	1	6.35	3.34	1.65	18.3	29.01	172.5
02-02-2016	10	0.99	5.22	3.95	2.22	20.18	28.95	205
02-02-2016	11	1	7.35	5.04	3.16	21.5	28.9	274
02-02-2016	12	1	5.93	6.34	4.93	20.46	28.85	325
02-02-2016	13	1	1	7.96	7.07	17.56	28.81	256.5
02-02-2016	14	1	0.92	9.99	9.27	15.76	28.76	133.75
02-02-2016	15	1	3.15	11.87	11.02	20.92	28.72	182.5
02-02-2016	16	1	4.68	12.79	11.72	17.96	28.71	115.5
02-02-2016	17	0.83	7.26	13.51	12.56	17.15	28.69	2.93
03-02-2016	6	0.16	10	2.47	0.22	15.36	28.83	0.68
03-02-2016	7	0.56	10	1.86	-0.17	11.72	28.85	303.38
03-02-2016	8	0.98	10	2.47	0.23	15.23	28.88	327.75

Commented [AG1]: Pl mention the units

### Weather Temperature

Given that it directly affects the output and efficiency of solar energy systems, weather temperature plays a critical role in solar energy analysis. The distribution of solar energy generation data over various temperature ranges can be visually represented by a `displot` graph, which can efficiently demonstrate this relationship (Figure 8).

### Energy Generated by Solar System

A bar plot graph analysis of a solar system's energy output gives an idea of how much energy was produced over a given time period. Here is how the energy produced by a solar system can be represented using a bar plot graph: A bar plot graph analysis of a solar system's energy production offers important insights into the system's reliability, efficiency, and performance over time. With the use of this visualization tool, stakeholders may maximize energy output and optimize the use of renewable energy resources by making well-informed decisions [17].

A bar plot of wind power generation data makes it simple for viewers to spot trends, patterns, and variances in the amount of electricity produced over time or at various locations. The ability to monitor, analyse, and optimize wind power producing operations makes this kind of visualization especially useful for stakeholders in the renewable energy sector, such as energy firms, legislators, and environmental organizations (Figure 9).

### Information About Energy Generated by Solar System

With the aid of machine learning, we can manage and regulate the maximum and minimum power generation on a large scale. This information provides some basic knowledge about the power generated by the system by analysing the entire column of power generated.

- Maximum Power Generated by system (kW): 4035.45
- Minimum Power Generated by system (kW): 0.0
- Mean of System Power Generated: 1793.61
- Median of System Power Generated: 1467.38
- Mode of System Power Generated: 0.0

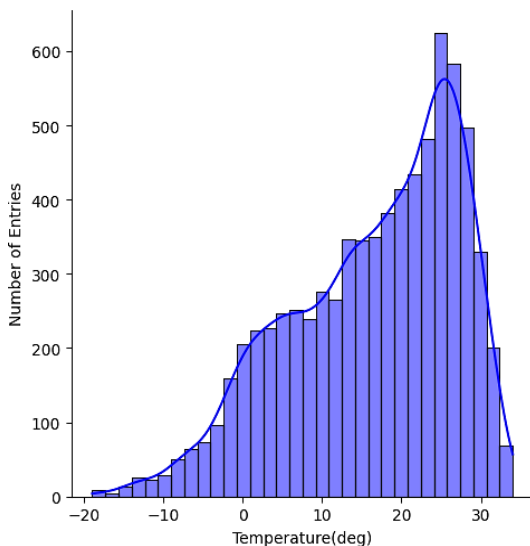


Figure 8. Weather temperature.

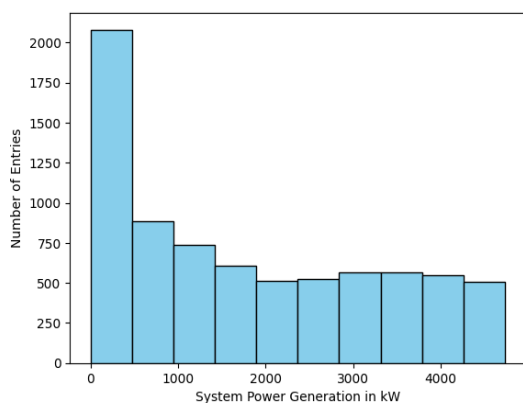


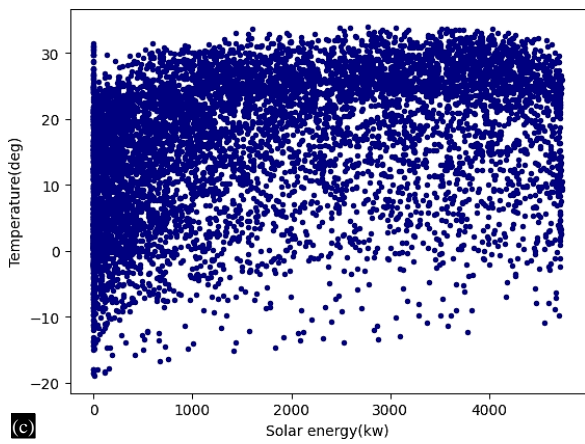
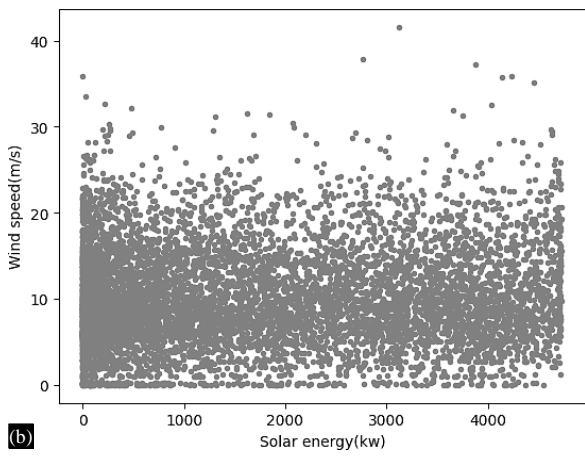
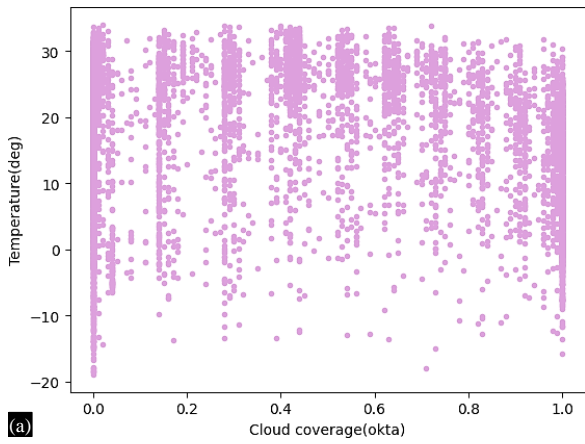
Figure 9. Solar energy generation.

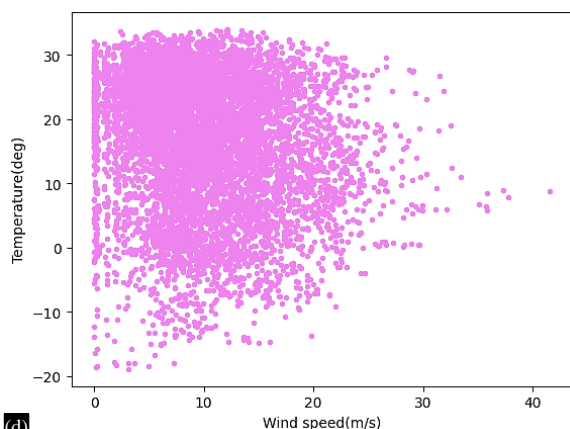
#### Correlation between Each Column

Since the data has already been pre-processed, analysis of the data is currently being done through inspection, cleansing, transformation, and modelling in order to find relevant information that will help conclusions and decision-making. Here, we use the SNS to draw a graph inside each column. `plt.show()` `pairplot(df)`. This is also used as a display of training data. With an emphasis on their function in the production of solar energy for chemical energy applications, this pair plot graph investigates the correlations between "Cloud Coverage", "Temperature", and "Solar Energy". A matrix of scatter plots and histograms is produced by plotting each variable in the dataset against every other variable (Figure 10(a-d)).

#### Accuracy of Predicted Value by Linear Regression for Solar Energy

1. Mean Absolute Error (MAE): The average size of errors between the expected and actual solar energy production values is represented by the MAE value, which is roughly 1154.51. This indicates that there is, on average, a 1154.51 unit difference between the model's predictions and the actual results. Better accuracy is shown by lower MAE values, which is the preferred value. On the other hand, the moderate mistakes of the model's predictions are indicated in this instance by the comparatively large MAE.
2. Mean Squared Error (MSE): This figure, which is roughly 1871706.70, is the average of the squared differences between the values of solar energy production that were predicted and those that were observed. Unlike MAE, MSE gives greater weight to errors that are larger. The relatively high MSE indicates that there are significant differences between the values that were observed and those that were predicted, indicating possible inaccuracies in the model's predictions.
3. Median Absolute Error: When all absolute mistakes between projected and actual solar energy output values are ordered in ascending order, the median absolute error value of roughly 1073.78 reflects the midpoint value in the dataset. According to this figure, which offers a reliable indicator of central tendency, around half of the absolute errors are below 1073.78 units.
4. Accuracy Result: The linear regression model's accuracy result, which is roughly 15.25%, indicates the percentage of accurate predictions it made. A low accuracy result means that there is a large percentage of errors in the model's predictions, making them less dependable.
5. The R-squared Score ( $R^2$ ), which is roughly 0.1525, represents the percentage of the variation in the solar energy output values that can be accounted for by the linear regression model. A score around 1 denotes a strong match between the model and the data, whereas a value near 0 shows little to no explanation of the data's variability by the model. The very low  $R^2$  score in this instance raises the possibility that the linear regression model is unable to adequately account for the variability in solar energy production.





**(d)** Figure 10. (a) Temperature vs. cloud coverage, (b) Wind speed vs. solar energy, (c) Temperature vs. Solar energy, (d) Temperature vs. Wind speed.

#### Solar Energy Prediction by Linear Regression

In chemical industries, solar energy generation is predicted using linear regression based on temperature data. The linear regression model is trained using historical temperature and accompanying energy generation data. Once trained, the model uses temperature inputs to estimate future times when energy will be generated. The expected energy generation values in kilowatts (kW) or another appropriate unit are plotted on the y-axis of a line plot graph, which is used to illustrate the temperature data on the x-axis.

The line plot graph displays a line connecting the predicted energy generation values corresponding to different temperature levels. This visualization enables stakeholders to observe the relationship between temperature and energy generation and understand how changes in temperature impact energy production in chemical industries.

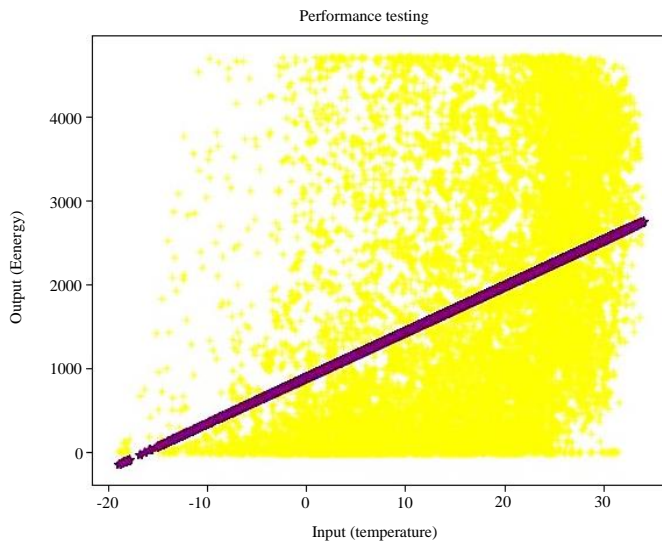
Once trained, the model can make predictions of energy generation for future time periods based on temperature inputs. These predictions are then visualized using a line plot graph, where the x-axis represents temperature values, and the y-axis represents predicted energy generation values in kilowatts (kW) or another relevant unit (Figure 11).

#### Solar Energy Prediction by Decision Tree

Historical data on cloud cover, visibility, temperature, wind speed, and matching solar energy generation in chemical companies is subjected to a decision tree algorithm. Based on the input attributes, the decision tree algorithm predicts the generation of solar energy by identifying patterns and relationships in the data. After it has been trained, the decision tree model can be used to forecast solar energy generation for upcoming times depending on the existing or anticipated values of wind speed, temperature, visibility, cloud cover, and visibility. A graph, which can be either a line plot or a scatter plot, is then used to display these predictions.

#### Accuracy of Predicted Value by Decision Tree for Solar Energy

1. The discrepancy between predicted and observed solar energy generation, quantified by the Mean Absolute Error (MAE), was approximately 728.04. This indicates a moderate level of prediction error for the decision tree model, as lower MAE values generally signify greater accuracy. While not negligible, the MAE suggests a reasonable alignment between the model's outputs and actual generation data.



**Figure 11.** Solar energy predication by linear regression.

2. *Mean Squared Error (MSE)*: The average of the squared discrepancies between the estimated and real solar energy generation numbers is represented by the MSE value, which is roughly 1003214.50. Even with larger deviations, a lower MSE still suggests that the model's predictions are generally more accurate than the MAE. Because the MSE is susceptible to outliers and greater errors, it is important to take into
3. *Median Absolute Error*: When all absolute errors between the values of projected and actual solar energy generation are sorted in ascending order, the median absolute error value of 561.13 denotes the midway value in the dataset. About half of the absolute errors are thought to fall below 561.13 units, according to this value, which offers a reliable indicator of central tendency.
4. *Accuracy Result*: The decision tree model's accuracy result, or roughly 54.57% of correct predictions, is indicative of its performance. Although 54.57% accuracy may not seem like much, it is important to take the problem's context into account and compare it to other models or a baseline. It suggests that although there is potential for improvement, the model's predictions are reasonably accurate.
5. *R-squared Score ( $R^2$ )*: The  $R^2$  score, of around 0.5457, shows how much of the variation in the values of solar energy generation can be accounted for by the decision tree model. A score nearer 1 signifies a strong match between the model and the data, whereas a value nearer 0 implies that the model's capacity to explain the variability in the data is limited. The decision tree model appears to be able to explain a moderate degree of variability in solar energy generation in this instance, based on the  $R^2$  score [12, 15, 17, 18].

Commented [AG2]: Pl complete

## CONCLUSION

The development of ML models is a prerequisite for the improvement of renewable energy systems. Recall that worries about sustainability and climate preservation have recently led to a rise in the popularity of renewable energy systems, making it more acceptable to harness the power of the sun, wind, and other renewable energy sources.

The primary characteristics of all renewable energy sources are their reliance on the environment and significant obstacles to management and planning. Forecasting power generation and demand is also

required because of the growth of the electrical infrastructure. However, if wind and solar energy oscillations and energy demand are not effectively handled, the system and its customers suffer.

ML models are becoming crucial in these kinds of energy systems as a result. The examination of these specific applications has shown how resilient the ML models appear to be. The findings indicate that the trend of research is shifting toward customized machine learning models created for specific applications. We have also talked about the future role of artificial intelligence and machine learning. As a result of these technologies becoming more widely available and user-friendly, many data steps, including the use of machine learning-based screening models, are becoming automated.

In manufacturing environments, advanced monitoring systems that alert users to abnormal or suboptimal behaviour, list correlated factors, and let engineers interactively visualize process data will soon become the norm. The ability to collect high-quality, well-organized production data will provide historians a competitive edge, particularly in plants where data ownership is deeply ingrained. When ML is combined with process engineering and control expertise, it can be applied to steady-state or batch-to-batch applications.

#### REFERENCES

1. Deshmukh M, Deshmukh S. Modeling of hybrid renewable energy systems. *Renew Sustain Energy Rev.* 2008; 12:235–49.
2. Elgerd OI. *Electric Energy Systems Theory: An Introduction.* New York, NY, USA: McGraw-Hill; 1982.
3. Gomez-Exposito A, Conejo AJ, Canizares C. *Electric Energy Systems: Analysis and Operation.* Boca Raton, FL, USA: CRC Press; 2018.
4. Quaschnig V. *Understanding Renewable Energy Systems.* Abingdon, UK: Routledge; 2016.
5. Isabella O, Smets A, Jäger K, Zeman M, van Swaaij R. *Solar Energy: The Physics and Engineering of Photovoltaic Conversion, Technologies and Systems.* Cambridge, UK: UIT Cambridge Limited; 2016.
6. Obulesu, O., Venkateswarulu, N., Sri Vidya, M., Manasa, S., Pranavi, K., Brahmani, C. Early Prediction of Healthcare Diseases Using Machine Learning and Deep Learning Techniques. In: Seetha, M., Peddoju, S.K., Pendyala, V., Chakravarthy, V.V.S.S.S. (eds) *Intelligent Computing and Communication. ICICC 2022. Advances in Intelligent Systems and Computing.* 2023;1447. Springer, Singapore. [https://doi.org/10.1007/978-981-99-1588-0\\_29](https://doi.org/10.1007/978-981-99-1588-0_29).
7. Abdel-Basset M, Hawash H, Chakraborty RK, Ryan M. Energy-net: a deep learning approach for smart energy management in IoT-based smart cities. *IEEE Internet of Things Journal.* 2021; 8(15):12422-35. <https://doi.org/10.1109/JIOT.2021.3063677>
8. Beck DAC, Carothers JM, Subramanian VR, Pfaendner J. Data science: Accelerating innovation and discovery in chemical engineering. *AIChE J.* 2016; 62:1402–16.
9. Industry 4.0: How to navigate digitization of the manufacturing sector. April 2015 [Internet]. Available from: [<https://www.mckinsey.com/capabilities/operations/our-insights/industry-four-point-o-how-to-navigae-the-digitization-of-the-manufacturing-sector>].
10. Chen MR, Zeng GQ, Lu KD, Weng J. A two-layer nonlinear combination method for short-term wind speed prediction based on ELM, ENN, and LSTM. *IEEE Internet Things J.* 2019; 6:6997–7010.
11. Adjerid H, Maouche AR. Multi-agent system-based decentralized state estimation method for active distribution networks. *Computers & Electrical Engineering.* 2020;86:106652. <https://doi.org/10.1016/j.compeleceng.2020.106652>.
12. Sakr HA, Fouda MM, Ashour AF, Abdelhafeez A, El-Afifi MI, Abdellah MR. Machine learning-based detection of DDoS attacks on IoT devices in multi-energy systems. *Egyptian Informatics Journal.* 2024; 28:100540. <https://doi.org/10.1016/j.eij.2024.100540>.
13. Ma J, Yang M, Han X, Li Z. Ultra-short-term wind generation forecast based on multivariate empirical dynamic modeling. *IEEE Trans Ind Appl.* 2017; 54:1029–38.

- 
14. Haidi T, Cheddadi B, Mariami FE, Idrissi ZE, Tarrak A. Wind energy development in Morocco: Evolution and impacts. *Int J Electr Comput Eng (IJECE)*. 2021;11(4):2811–9. doi:10.11591/ijece.v11i4.pp2811-2819.
  15. Kiesecker J, Baruch-Mordo S, Heiner M, Negandhi D, Oakleaf JR, Kennedy CM, Chauhan P. Renewable Energy and Land Use in India: A Vision to Facilitate Sustainable Development. *Sustainability*. 2019; 12:281.
  16. Haidi T, Cheddadi B. State of wind energy in the world: Evolution, impacts and perspectives. *International Journal on Technical and Physical Problems of Engineering*, 2022;14(2):6.
  17. Buwei W, Jianfeng C, Bo W, Shuanglei F. A solar power prediction using support vector machines based on multi-source data fusion. In: 2018 International Conference on Power System Technology (POWERCON); 2018 Nov; p. 4573–7.
  18. Didavi AB, Agbokpanzo RG, Agbomahena M. 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); 2021 Dec; p. 1–5.