

Report-Hybrid Powered Industrial Data Center

by User System

Submission date: 26-Aug-2024 11:45PM (UTC+0800)

Submission ID: 2438471790

File name: Report-Hybrid_Powered_Industrial_Data_Center.docx (3.62M)

Word count: 8426

Character count: 48100

List of Tables

Chapter 3

Table 1. Hardware Tools Table

Table 2. Software Tools Table

Chapter 8

Table 3: Initial Model Results for Windmill Power Production

Table 4. Optimized Model Results for Windmill Power Production

Table 5. Initial Model Results for Solar Power Production

Table 6. Optimized Model Results for Solar Power Production

Table 7. Model Performance on Hybrid Power Generation System

Table 8: A table showing the risk assessment of the project

List of Figures

Chapter 3

Fig 1. Software block diagram

Chapter 6

Fig 2. Distribution of Numerical Features

Fig 3. Triangle Heatmap of Feature Correlations

Fig 4. Top 10 Most Important Features Based on Correlation

Fig 5. Panel Area vs Generated Power

Fig 6. Irradiance vs Produced power

Fig 7. Panel Efficiency vs Power Output
Fig 8. Model Training Performance
Fig 9. Univariate Analysis
Fig 10. Correlation Matrix
Fig 11. Top 10 Most Important Features Based on Correlation
Fig 12. Air Density vs Generated Power
Fig 13. Turbine Area vs Generated Power
Fig 14. Turbine Efficiency vs Generated Power
Fig 15. Wind Speed vs Generated Power
Fig 16. Model Performance
Fig 17. Cross-Validation Results
Fig 18. Average Production by Season
Fig 19. Annual Energy Production
Fig 20. Average Monthly Production for Each Source
Fig 21. Average Production by Day of the Week
Fig 22. Average Production by Hour of the Day
Fig 23. Average Hourly Energy Production (2020-2023)
Fig 24. Distribution of Produced Energy

Chapter 8

Fig 25. Random Forest CM
Fig 26. Linear Regression CM
Fig 27. Hyperparameter Tuning Results (Solar)
Fig 28. Hyperparameter Tuning Results (Wind)
Fig 29. Hyperparameter Tuning Results (Hybrid)

Chapter 1

Keep Blank

Chapter 2

2.1 Introduction

As worldwide data processing and storage are rapidly growing, industry data centers are becoming the main infrastructure hubs. Nonetheless, these facilities require an enormous amount of energy, which is causing valid concerns about the environment and consequently costly operations. [1] To mitigate these problems, the inclusion of the sun and the wind as two possible renewable energy sources has quickly become a realistic solution for data centers in their quest for sustainability and the utilization of power in a more efficient way. [2, 3] This particular study will delve into the modeling and simulation of an industrial data center that runs on both solar and wind energy to render it more advanced technology-wise and lessen its environmental impact.

When arguing about the use of solar-powered as opposed to wind-powered data centers, many factors are important, with each one contributing to the efficiency and sustainability of these energy sources. Solar energy, because it depends on photovoltaic panels, provides a predictable and relatively stable output of energy during the daylight hours. The conversion can be used to convert the unpredictable sunlight into electricity. [4] It can be an added advantage for data centers, which need a continuous flow of power so they run. Also, solar panels, by and large, present less of a problem with a space deficit in comparison with wind turbines, hence they are a possible solution for populated areas where space is limited. [5] However, solar panel efficiency is often directly related to geographic location and weather conditions, and therefore, it can produce insufficient or irregular amounts of energy in places with less sunlight, possibly necessitating the use of other power sources or large energy storage solutions.

It is the wind energy that captures the wind as the wind turbines generate electricity throughout the day and night, provided enough wind is available. [6] This quality can be very

useful in places with strong and regular wind storms. However, wind turbines occupy a much larger area and may be visually and acoustically disturbing at times, thus it might be a hard one to apply in densely populated areas. The variation in wind speed, where it has both high and low periods, can lead to changes in energy output, thus needing robust energy storage systems or auxiliary power arrangements to guarantee a continuous power supply. Having said this, horizontal-axis wind turbines are mostly placed in suitable areas where the wind is favorable and thus the capacity factor is high. [7]

Primarily, the decision to choose solar and wind power for data centers depends on the surrounding geographic and manmade factors. Solar energy works well in predictable and sunny sites and is more efficient in populated areas with restricted space. On the other hand, Wind is more suitable for areas with high wind speeds. It can pose problems in those regions with space constraints and noise control. Nevertheless, in the long term, both of them that by the way of greening data center they can sustain the earth. But choosing the right options and identifying and dealing with the drawbacks of each source are the elements that need to be considered in a comprehensive way. [8]

The primary goal of this work is to evaluate solar PV and wind turbine technologies and assess their interdependence in supplying power to data centers. We plan to make an energy demand analysis by making an in-depth study of the energy load behavior of the center so that the required capacity and the best combination of solar and wind power can be introduced or employed. We have designed the system by making a hybrid system that combines both solar and wind sources, which should come coupled with storage solutions that can not only absorb but also supply a stable and reliable power system. Finally, we will optimize the research through the use of advanced optimization tactics to heighten the productivity of the hybrid system. These tactics focus mainly on the choice of configuration and the right operational settings for both energy sources. [9]

We will be using different machine learning models to analyze the data from all three types of energy sources: solar, wind, and hybrid power stations. We will work out the best models for analyzing and break-down the best infrastructure to optimize the data centers' power generation method.

25

2.2 Keep Blank

2.3 Keep Blank

2.4 Keep Blank

Develop Project Plan and Timeline & Gantt Chart: Keep Blank

Chapter 3

3.1 Resource List

Table 1. Hardware Tools Table

Component	Unit Price (in USD)	Quantity	Total Cost (in USD)
INTEL 13TH GENERATION CORE I5 13500 RAPTOR LAKE PROCESSOR WITH 14 CORES AND 24 THREADS	239.13	1	239.13
MSI MAG B760M MORTAR DDR5 LGA1700 GAMING MOTHERBOARD	183.62	1	183.62
TEAM T-FORCE DELTA TUF GAMING RGB 16 GB 6000 MHZ DDR5 GAMING RAM	67.47	2	134.94
MAXGREEN 1200 VA OFFLINE UPS (PLASTIC BODY)	52.10	1	52.10
GIGABYTE C301 GLASS E-ATX GAMING CASE (BLACK)	57.22	1	57.22
MSI GEFORCE RTX 4060 TI VENTUS 3X OC GRAPHICS CARD WITH 16 GB GDDR6 MEMORY	555.13	1	555.13
SAMSUNG 980 PRO 500 GB PCIE GEN4 M.2 NVME SSD	73.45	1	73.45
COOLERMASTER MWE 750-WATT GOLD V2 POWER SUPPLY	999.23	1	999.23
DEEPCOOL LE520 240 MM ALL-IN-ONE ARGB LIQUID CPU COOLER	56.37	1	56.37
Total			2351.19

Table 1 outlines the key hardware components critical for stable performance and efficiency in a machine learning project. The CPU handles all computations required for data processing and model training. A gaming motherboard supports high-performance components, ensuring system stability. Gaming RAM is essential for managing large datasets and complex computations by enabling fast data acquisition and storage. The GPU

accelerates parallel computations, particularly in deep learning, dramatically speeding up model training. Solid-state drives (SSDs) enhance data management and system responsiveness by providing faster read and write speeds. A dependable power source, coupled with an uninterruptible power supply (UPS), guarantees stable power delivery and shields the system from electrical surges, thereby preventing data loss or hardware damage. The gaming case facilitates smooth air circulation and cooling, preventing overheating. [10]

Table 2. Software Tools Table

Tool	Functions	Why selected this tool
Dataset	Collection of attributes and records for training and testing	Project relevance and authenticity to the target data centers
Pandas	Understanding dataset characteristics and informing preprocessing steps	Provides essential insights and visualizations for data understanding
Matplotlib, Seaborn	Builds static, and interactive visualizations	Provides essential visualizations for data analysis
Random Forest	Combines multiple decision trees	Improves predictive accuracy and controls overfitting
Linear Regression	Models the relationship between one or more independent variables and a dependent variable	Fits a linear equation to the observed data
Support Vector Regression	Models the relationship between one or more independent variables and a dependent variable	Finds a function that maximizes the margin of error while fitting the data within a specified tolerance

Gradient Boosting	Builds models sequentially, each correcting its predecessors' errors	Improves predictive performance and reduces overfitting
K-Nearest Neighbors	Classifies data points by giving them the most frequent label among their K-nearest neighbors in the feature space	Predicts a target variable's value by averaging the values of the K-nearest neighbors for a given data point
CatBoost Regressor	Builds models sequentially, each correcting its predecessors' errors	Handles categorical features efficiently and robustly, and improves predictive accuracy
AdaBoost Classifier	Combines multiple weak classifiers to create a strong classifier	Emphasizes previous model errors and assigns more weight to misclassified instances
Logistic Regression	Describes the connection between one or more independent variables and a dependent variable	Predicts the probability of a binary outcome
GridSearchCV Optimizer	Optimization of the model's learning process	Efficient and adaptive learning rate
Mean Squared Error (MSE)	Measures the average squared difference between predicted and actual values	Quantifies model's prediction accuracy
Mean Absolute Error (MAE)	Assesses the mean absolute deviation between predicted and actual values 5	Provides a straightforward indication of prediction accuracy
R ² Score	Measures the fraction of variance in the dependent variable that can be predicted from the independent variables	Reflects the quality of fit of a regression model

3.2 Specifications of Software Block Diagram using ML

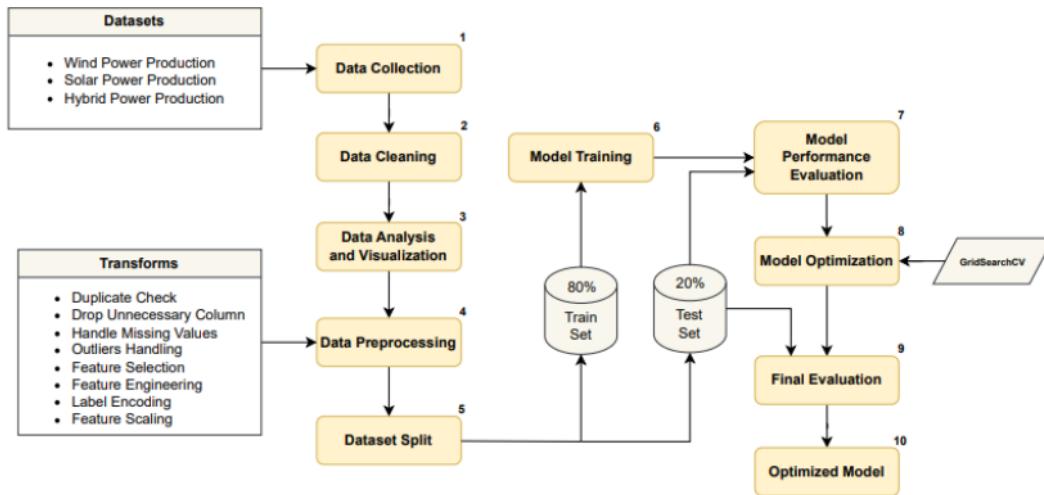


Fig 1. Software block diagram

Initially, we collected data sets from solar, wind, and hybrid power stations. Second, we cleaned the data by removing duplicate entries, changing data types, and filtering outliers. Thirdly, we examined and represented the data better through Python libraries. Fourthly, we pre-processed the data using eight methods. The fifth stage involved splitting the dataset into 80% for training and 20% for testing. In the sixth stage, machine learning models were trained on the training set. The seventh stage saw the evaluation of these models using the test set. The eighth stage utilized GridSearchCV for model optimization. The ninth stage re-evaluated the optimized models with the test set. Finally, in the tenth stage, we formalized the model with the best performance scores.

Chapter 5

5.1 Data Collection

For this project, we have used three different datasets that contain data about the performance and usage of both solar and wind-powered generators, two common renewable energy sources.

The first dataset [11] that we used, contained the data for the power generated by solar panels from a **Canadian power plant**, and has the records of 4213 and 21 attributes namely:

3

- temperature_2_m_above_gnd: Represents the temperature 2 meters above the ground.
- relative_humidity_2_m_above_gnd: Indicates the relative humidity 2 meters above the ground.
- mean_sea_level_pressure_MSL: Shows the mean sea level pressure.
- total_precipitation_sfc: Represents the total precipitation.
- snowfall_amount_sfc: Indicates the snowfall amount.
- total_cloud_cover_sfc: Represents the total cloud cover.
- high_cloud_cover_high_cld_lay: Shows the high cloud cover.
- medium_cloud_cover_mid_cld_lay: Indicates the medium cloud cover.
- low_cloud_cover_low_cld_lay: Represents the low cloud cover.
- shortwave_radiation_backwards_sfc: Indicates the shortwave radiation backward.
1
- wind_speed_10_m_above_gnd: Shows the wind speed 10 meters above the ground.
- wind_direction_10_m_above_gnd: Represents the wind direction 10 meters above the ground.
- wind_speed_80_m_above_gnd: Indicates the wind speed 80 meters above the ground.
- wind_direction_80_m_above_gnd: Shows the wind direction 80 meters above the ground.
- wind_speed_900_mb: Represents the wind speed at 900 mb.
- wind_direction_900_mb: Indicates the wind direction at 900 mb.
1
- wind_gust_10_m_above_gnd: Shows the wind gust 10 meters above the ground.
- angle_of_incidence: Represents the angle of incidence.
- Zenith: Indicates the zenith angle.

- Azimuth: Represents the azimuth angle.
1
- generated_power_kw: Shows the power generated in kilowatts.

The second dataset [12] contains records for wind power generated by windmills in the USA, with 28,200 records for training and 12,086 for testing. Attributes include:

- tracking_id: Represents the unique identification number of a windmill.
- datetime: Shows the date and time of a record.
- wind_speed(m/s): Indicates the wind speed in meters per second.
- atmospheric_temperature(°C): Represents the temperature in degrees Celsius at the windmill location.
- shaft_temperature(°C): Shows the temperature of the shaft in degrees Celsius.
- blades_angle(°): Represents the angle of the blades of a wind turbine in degrees.
- gearbox_temperature(°C): Indicates the temperature of the gearbox in degrees Celsius.
- engine_temperature(°C): Shows the engine temperature in degrees Celsius.
- motor_torque(N-m): Represents the motor torque in Newton meters.
- generator_temperature(°C): Indicates the generator temperature in degrees Celsius.
- atmospheric_pressure(Pascal): Represents the atmospheric pressure in Pascals.
- area_temperature(°C): Shows the temperature in degrees Celsius within a 100-meter radius of the windmill.
- windmill_body_temperature(°C): Represents the temperature of the windmill body in degrees Celsius.
- wind_direction(Â°): Represents the direction of the wind (in degrees).
- resistance(ohm): Represents the resistance against the wind.
- rotor_torque(N-m): Represents the torque of a rotator (in Newton meter).
- turbine_status: Represents the status of the turbine (masked).
- cloud_level: Represents the following levels of the cloud in the sky on a particular day:
Extremely Low, Low, and Medium.
- blade_length(m): Represents the length of the blades of a windmill (in meters).
- blade_breadth(m): Represents the breadth of the blades of a windmill (in meters).
- windmill_height(m): Represents the height of the blades of a windmill (in meters).

- windmill_generated_power(kW/h): Represents the power generated (in Kilo watts per hour).

The final dataset [13] contains the data of solar and wind energy production records on an hourly basis for the **French grid** since 2020. The dataset is well organized and it can be used in various fields of machine learning such as time series forecasting, anomaly detection, price signal analysis, etc. The dataset consists of 59806 records and 9 attributes namely:

- Date and Hour: Date and Hour range.
- Date: Day date granularity.
- StartHour: Recording Start Hour.
- EndHour: Recording End Hour.
- Source: Production Source (Wind or Solar).
- Production: Production (in MWh).
- dayOfYear: Day of the Year.
- dayName: Day Name.
- monthName: Month Name.

5.2 Data Preprocessing

We have used the following data preprocessing techniques to set up the dataset before training and testing the data with machine learning models. [14]

Duplicate Rows Check: Identifying and removing duplicate records is crucial to prevent errors in analysis or modeling. Duplications can arise from common or rare errors, so after detecting and understanding the root cause, duplicates should be eliminated before proceeding to data visualization and distribution.

Dropping Unnecessary Columns: Evaluate and filter out irrelevant columns using techniques like correlation analysis. This process ensures that only essential data remains, improving the dataset's focus and reducing complexity.

Handling Missing or Null Values: Properly addressing missing or null values is vital to prevent inaccurate analysis or models. Detect missing values and correct them using techniques like imputation (mean, median, mode), or remove data with excessive missing values. Algorithms that handle missing data can also be utilized.

Outliers Management: Detect and manage outliers to prevent them from skewing results. Depending on their impact, outliers can be transformed, removed, or capped to ensure accurate analysis.

Feature Selection: Select the most important features to prevent overfitting and improve model performance. Use statistical tests, correlation analysis, and model-based methods to retain relevant features and exclude unnecessary ones.

Feature Engineering: Create or modify features to enhance model performance. This includes deriving new features, performing transformations, and converting categorical data to numerical forms where necessary.

Label Encoding: Transform categorical data into numerical values through label encoding to make it compatible with machine learning algorithms.

Feature Scaling: Standardize or normalize feature values to maintain consistency across all features, enhancing the stability and performance of machine learning models.
12

Chapter 6: Design and Development

6.1 Solar Power Generation Model Development

While building models for solar power generation, we utilized various Python libraries to visualize the relationships between the independent variables and the target variable. We have also implemented the formula: **Psolar = irradiance × panel area × panel efficiency**, in our calculations. A collection of figures describing the different relationships are listed below:

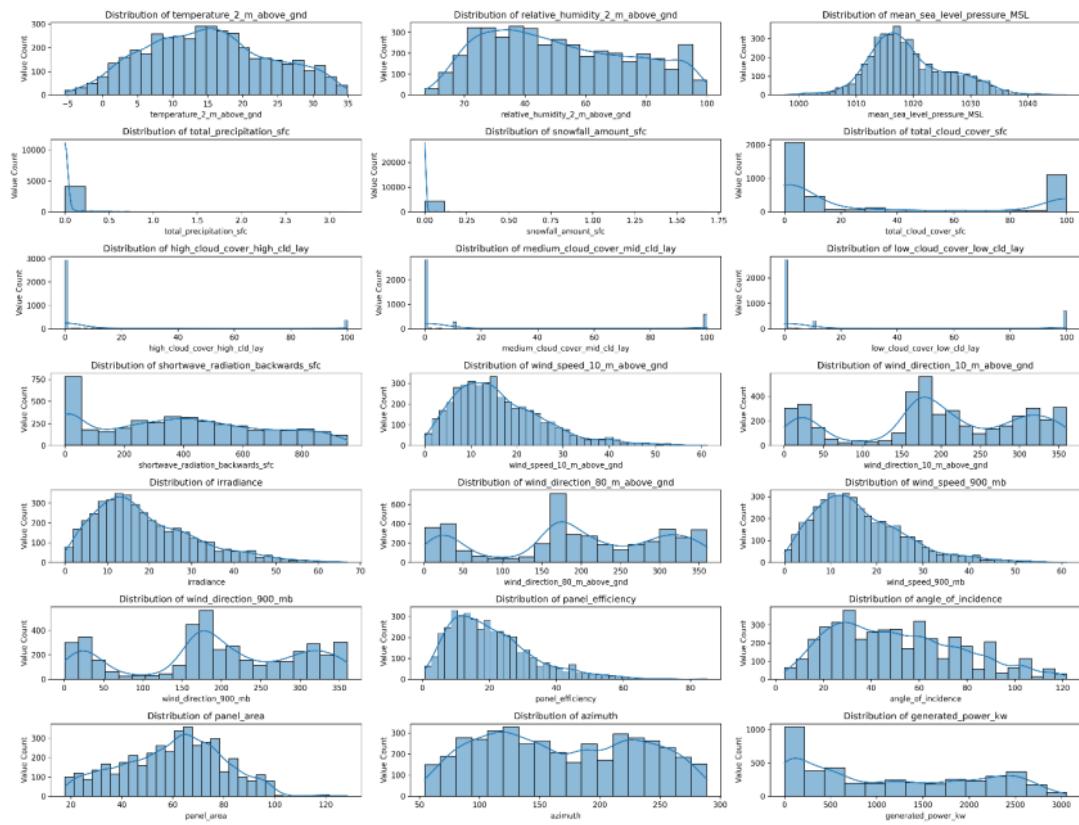


Fig 2. Distribution of Numerical Features

Figure 2 shows how the numerical features are distributed in their respective ranges and how the graphs are skewed, indicating the spreading or clustering of data.

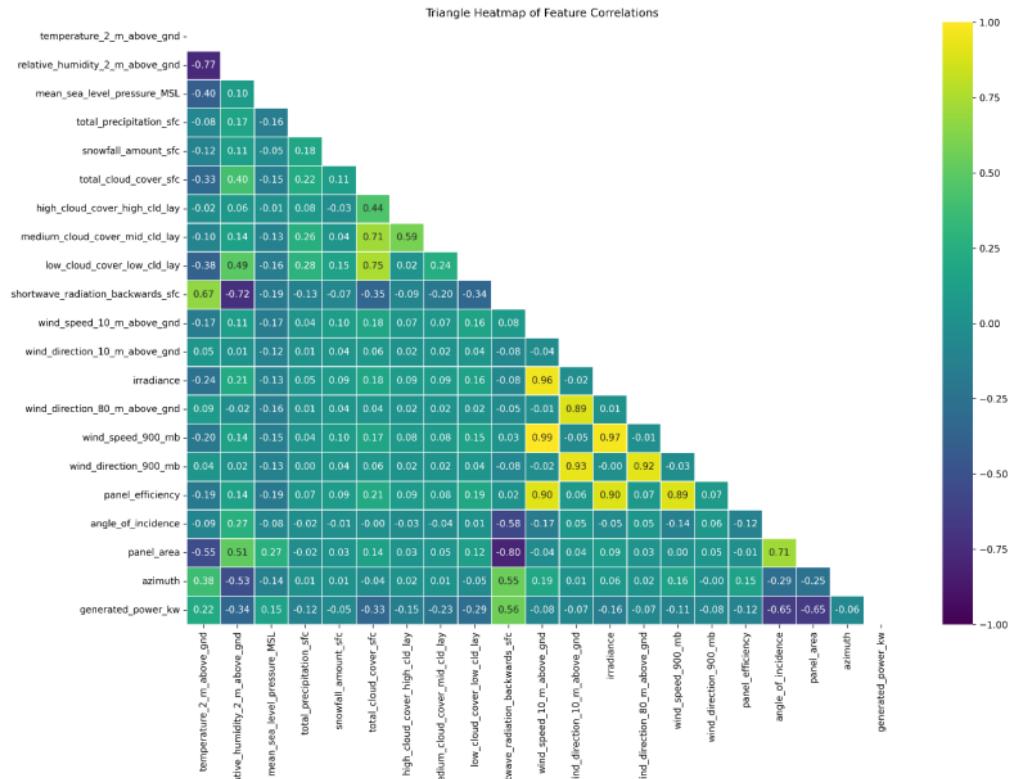


Fig 3. Triangle Heatmap of Feature Correlations

Figure 3 shows the heatmap of the feature correlations i.e. how strong the relation is between each feature. In this case, darker colors indicate stronger negative correlations between features, while lighter colors signify stronger positive correlations.

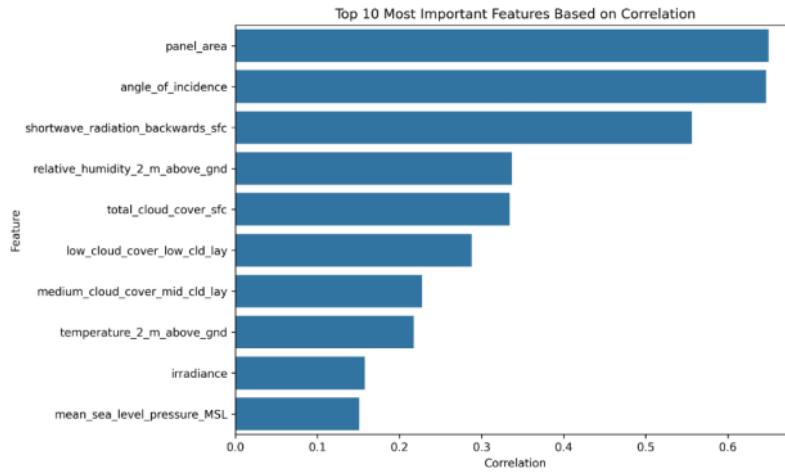


Fig 4. Top 10 Most Important Features Based on Correlation

Figure 4 shows the top 10 features most strongly correlated with the target variable. It was found that 'panel_area' had the highest correlation, followed by 'angle_of_incidence' and 'shortwave_radiation_backwards_sfc'.

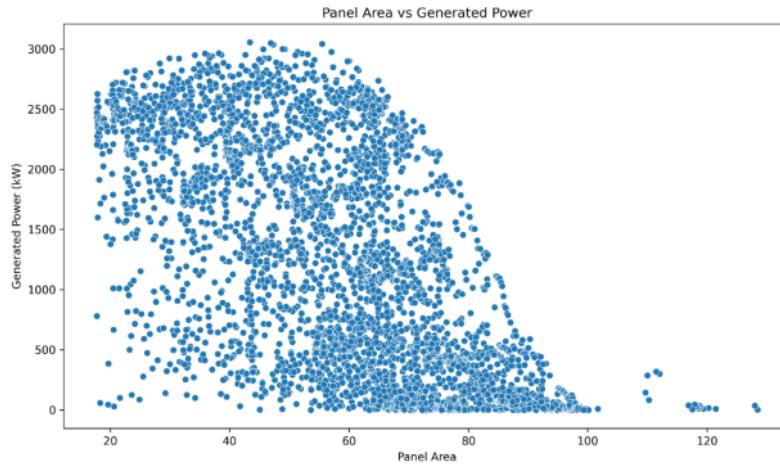


Fig 5. Panel Area vs Generated Power

Figure 5 shows the graph for panel area against generated power that indicates how much power is generated over a range of panel areas. It can be seen that panel area between 20

and 80 had generated the most amount of power and an approximation of 50 panel area had generated the highest amount of power that is, 3000 kw.

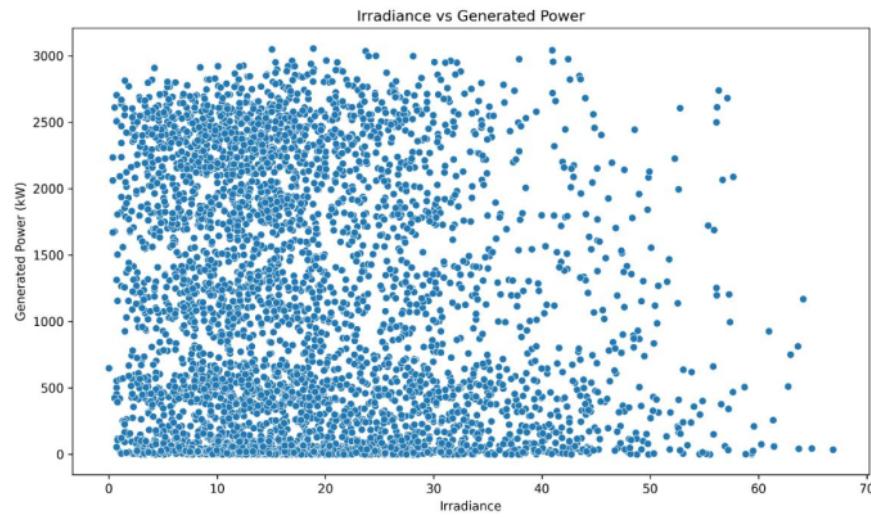


Fig 6. Irradiance vs Produced power

Figure 6 illustrates the graph for irradiance against generated power that represents the change in power generation with respect to change in irradiance. It can be seen that irradiance within the range of 0 to 40 had the densest amount of power generation and an approximate irradiance of 20 had the highest power generation of 3000 kW.

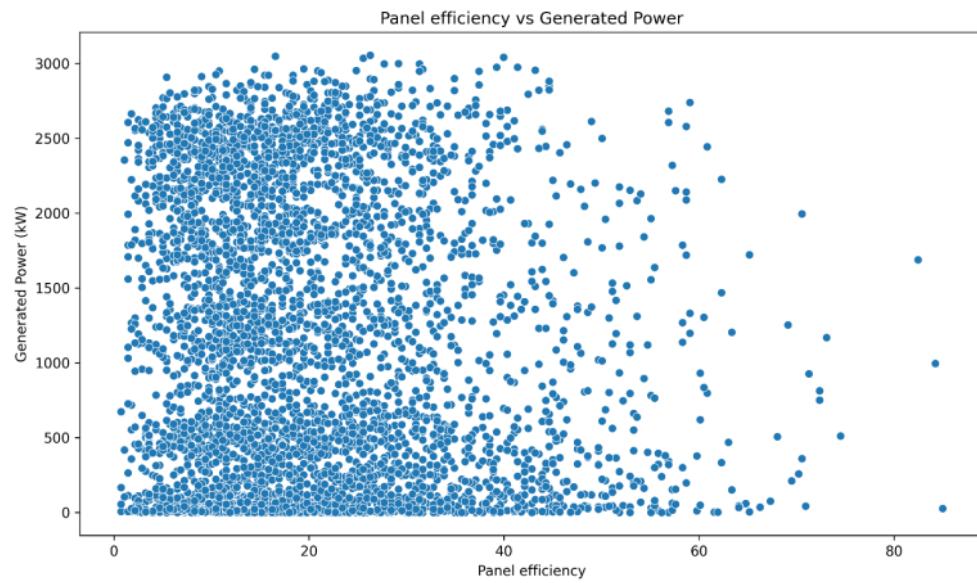


Fig 7. Panel Efficiency vs Power Output

Figure 7 represents the scatter diagram of panel efficiency against generated power depicts how much power is generated for each change in panel efficiency. It can be seen that the graph is skewed to the left, meaning more variation of power generation occurred within the range of 0 to 40-panel efficiency, where an approximation of 20-panel efficiency generated the highest amount of power of 3000 kW.

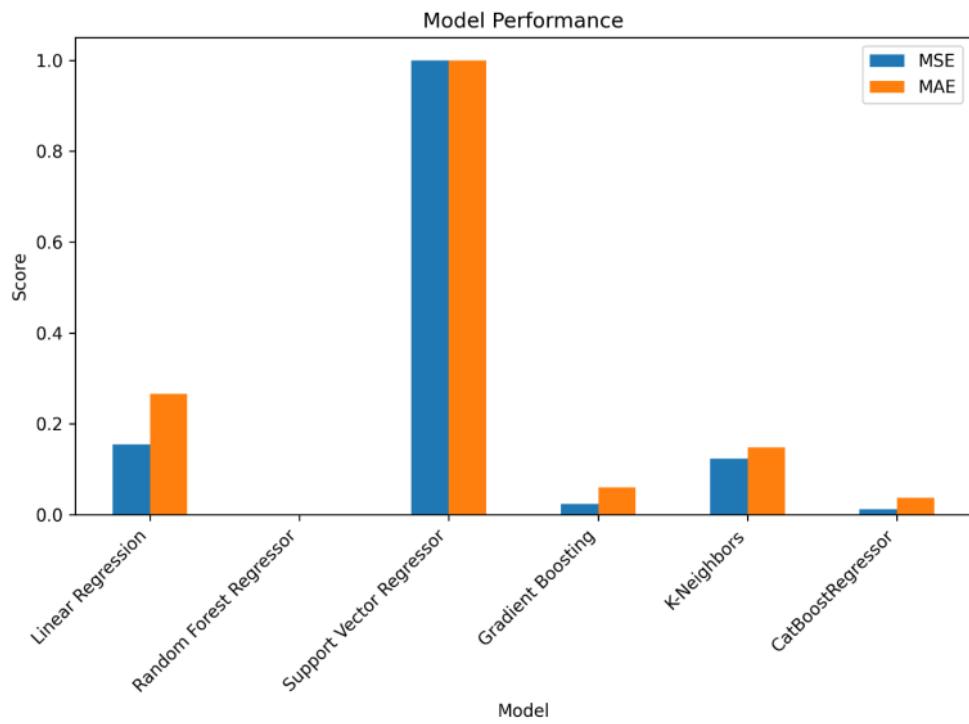


Fig 8. Model Training Performance

Figure 8 represents the training performances of every model we have used for the solar power generating plant. It can be seen that SVC had the highest MSE and MAE, meaning it had the most errors. In contrast, Random Forest Regressor had the lowest MSE and MAE, indicating that it is the best model to use.

6.2 Wind Power Generation Model Development

In developing the models for wind power generation, we utilized various Python libraries to visualize the connections between the independent variables and the target variable. We have also implemented the formula: **P_{wind} = 0.5 × air density × turbine area × (wind speed³) × turbine efficiency**, in our calculations. A collection of figures describing the different relationships are listed below:

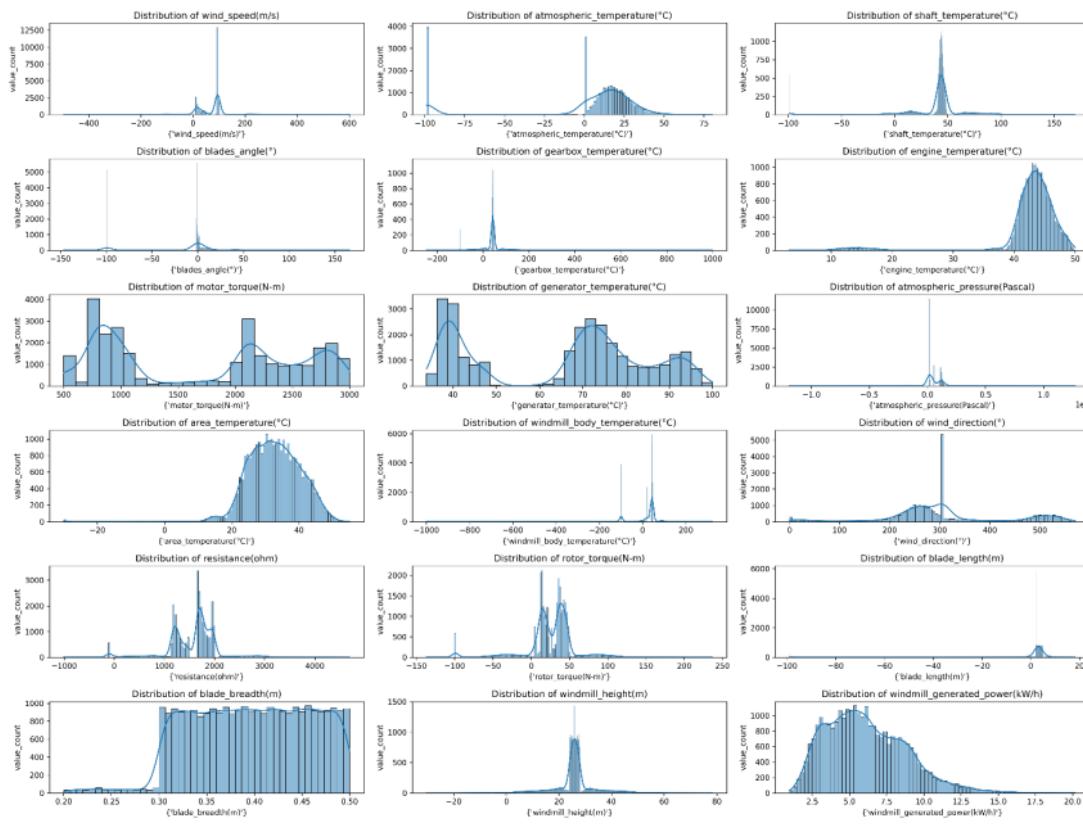


Fig 9. Univariate Analysis

Figure 9 represents the univariate analysis that indicates how each feature is distributed and how much the graph is skewed determining how the data is clustered.

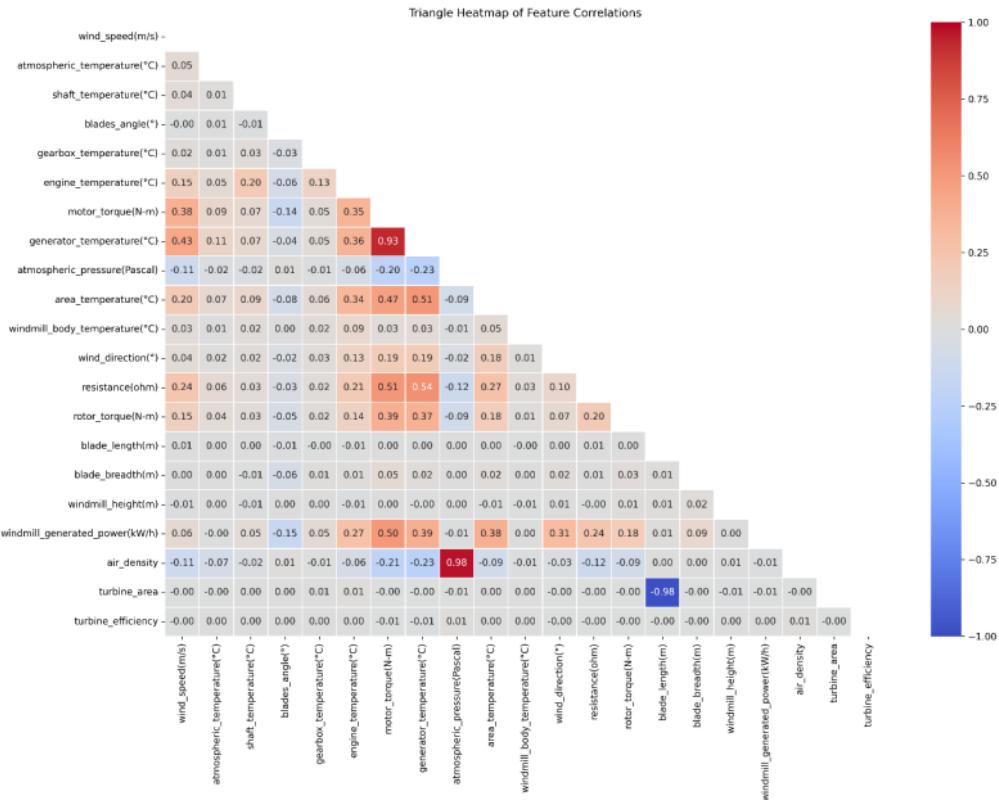


Fig 10. Correlation Matrix

Figure 10 represents the correlation matrix for the power generation by the wind. In this case, the bluer the color is, there is more negatively correlation and the redder the color is, there is more positively correlation between the features.

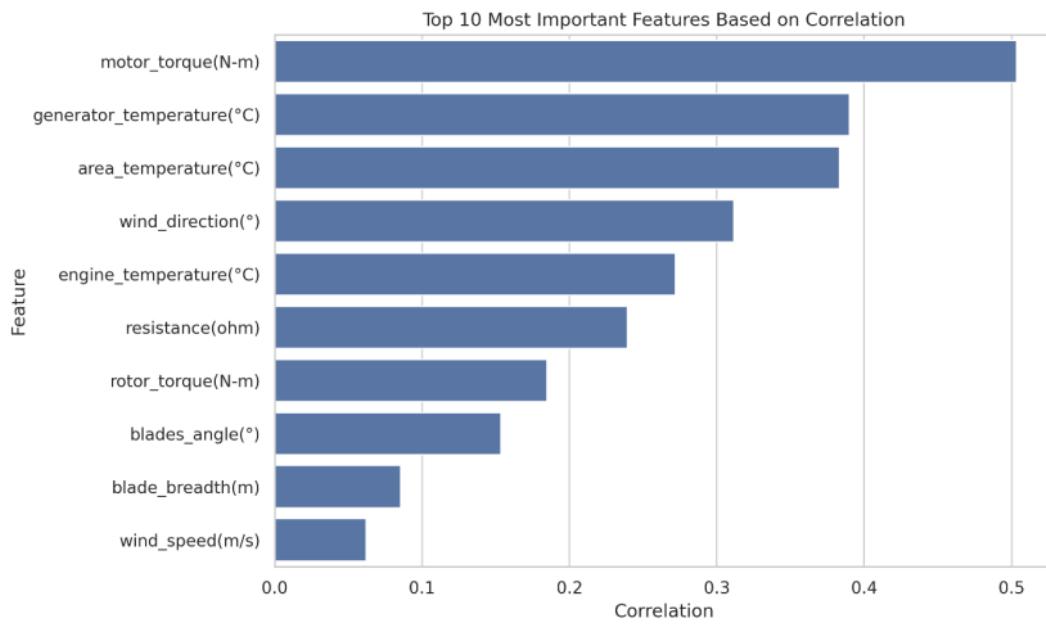


Fig 11. Top 10 Most Important Features Based on Correlation

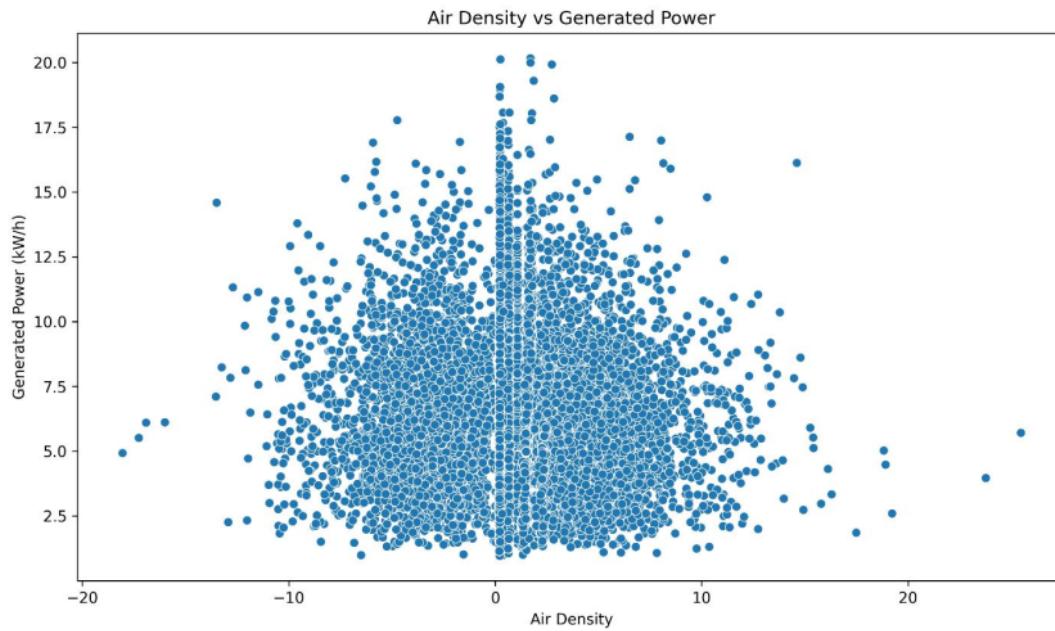


Fig 12. Air Density vs Generated Power

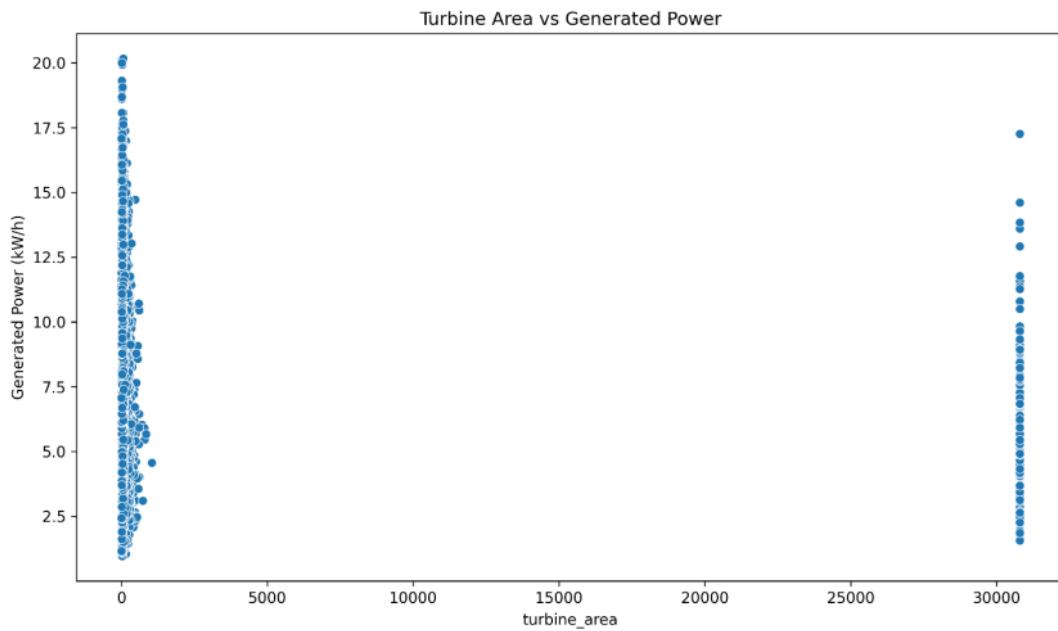


Fig 13. Turbine Area vs Generated Power

Figure 11 represents the top 10 most important features based on correlation, indicating that 'motor_torque(N·m)' had the highest correlation followed by 'generator_temperature(°C)' and 'area_temperature(°C)'.

Figure 12 represents the graph for air density against power generation, where it can be seen that the data is clustered in the center. Most of the power is generated when the air density was within the range of -10 to 10 and the maximum power of 20 kW/h was generated when the air density was 0.

Figure 13 represents the graph for turbine area against power generation which indicates that approximately 0 to 1000 area of turbine generated the most power and only a certain amount of power was generated for turbine area of 31000. The highest amount of power, 21.0 kW/h, was generated at turbine area of 0.

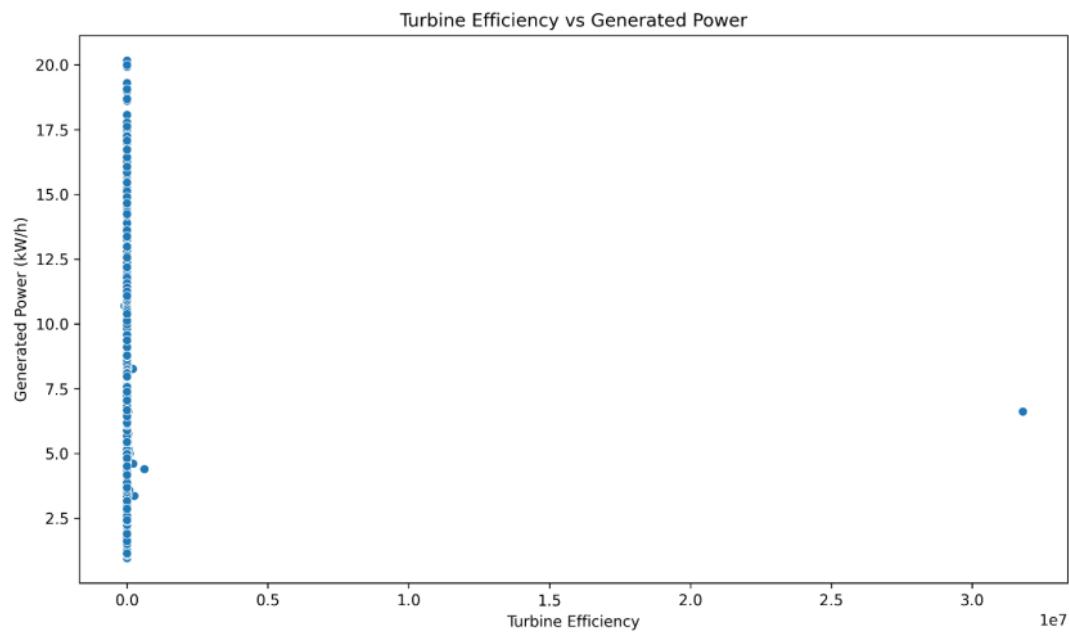


Fig 14. Turbine Efficiency vs Generated Power

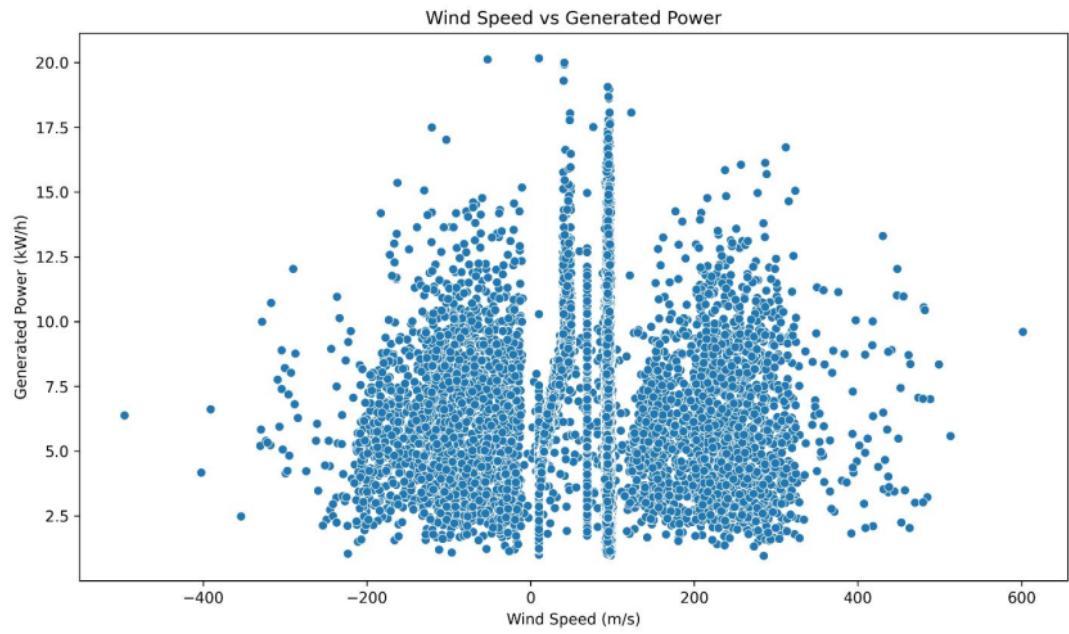


Fig 15. Wind Speed vs Generated Power

Figure 14 represents the graph for turbine efficiency against power generation indicating that almost all the power generations occurred at a turbine efficiency of 0.0 and there is only one point at 3.4. The highest power 21.0 kW/h was generated at a turbine efficiency of 0.0.

Figure 15 displays the graph of wind speed versus power generation, showing an almost symmetrical distribution, with most power generation data concentrated between -200 and 300 m/s. The highest power of 20.0 kW/h was generated at a wind speed of 0 m/s.
27

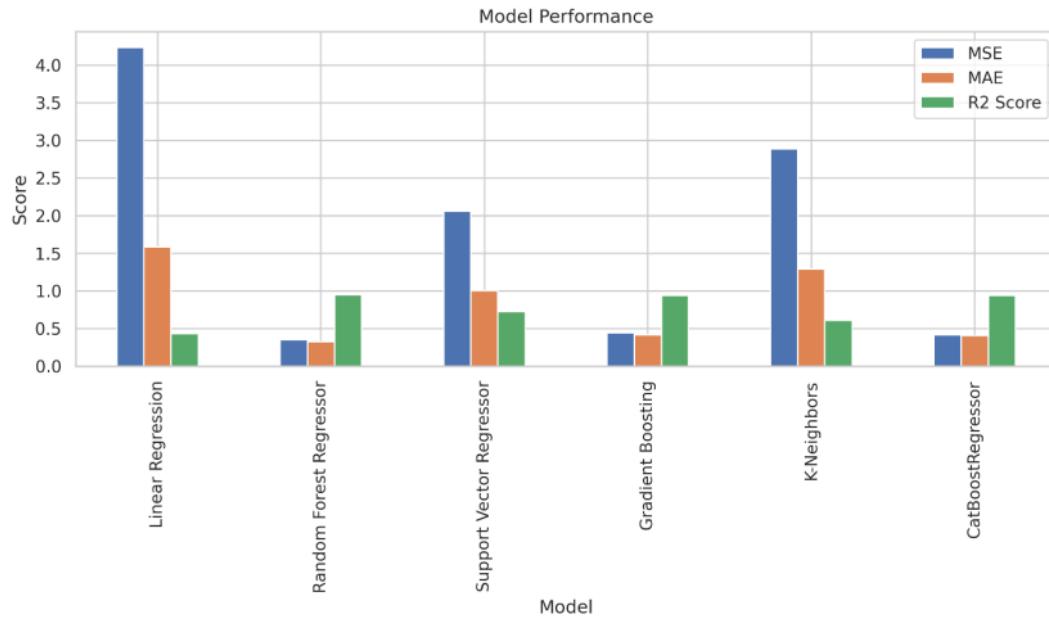


Fig 16. Model Performance

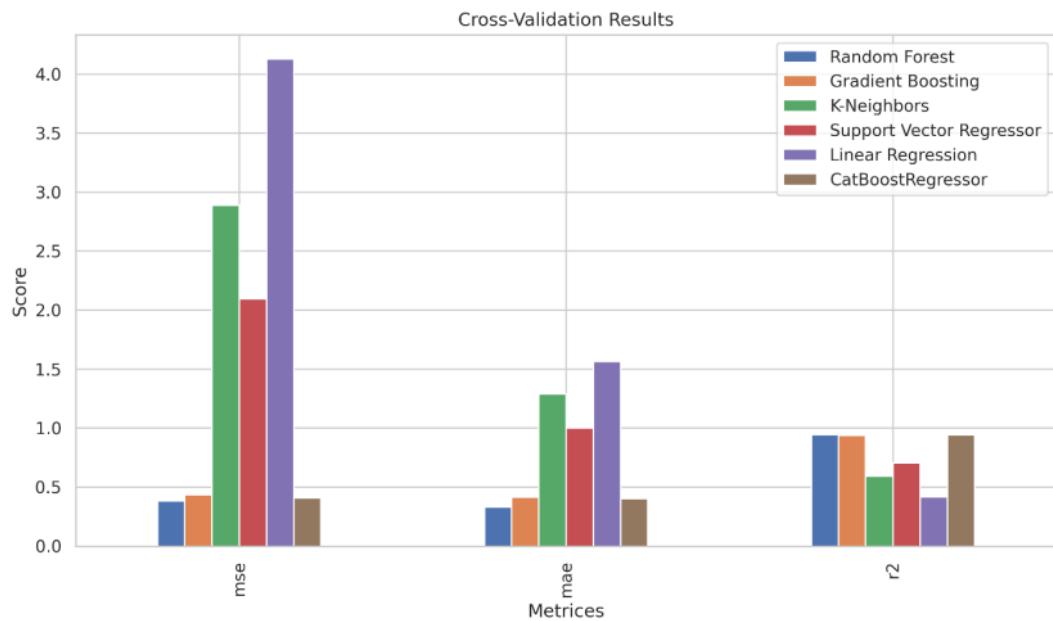


Fig 17. Cross-Validation Results

Figure 16 represents the models' performances on the wind-powered stations depicting that linear regression had the highest MSE and MAE and the lowest R2 score compared to others, meaning it will be the worst model to work on. On the contrary, random forest regressor had the lowest MSE, MAE and the highest R2 score which indicates that it is the best model to train the wind powered plant dataset.

Figure 17 represents the cross-validation results for all the models we have used the dataset to train on. It can be seen that linear regression still yielded the highest MSE and MAE and lowest R2 score indicating it is the least accurate model. On the contrary, random forest produced the lowest MSE and MAE and the highest R2 score, meaning it had the best cross-validation scores.

6.3 Hybrid Data Center Consumption Model Development

During the development of models for hybrid data centers, we employed various Python libraries to illustrate the relationships between the independent variables and the target variable. A collection of figures describing the different relationships are listed below:

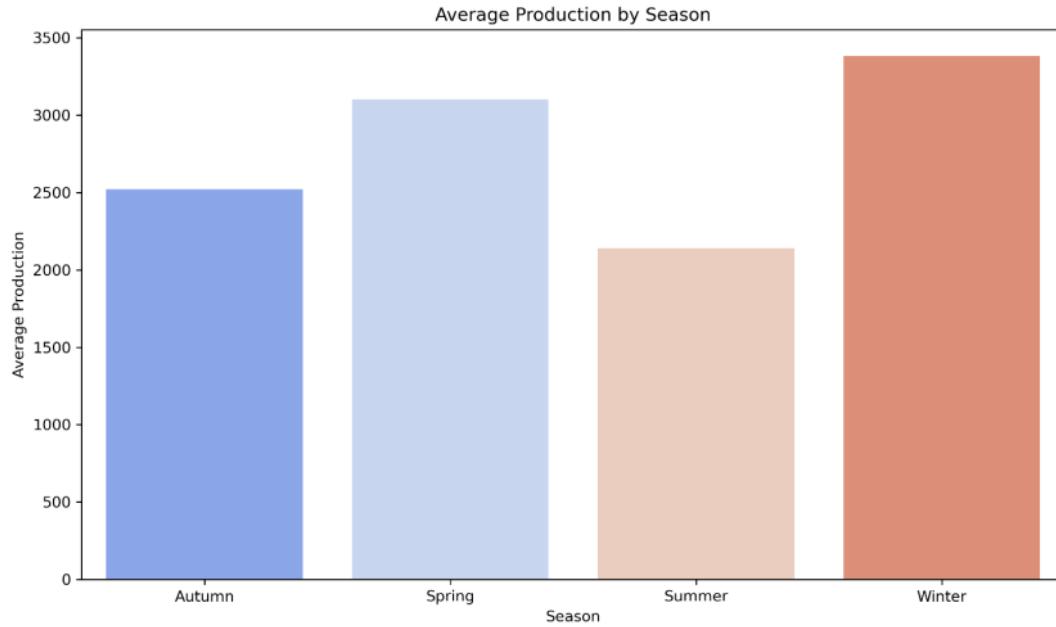


Fig 18. Average Production by Season

Figure 18 represents the average production of power in each season. It can be seen that in winter, maximum power is generated followed by spring, autumn and summer. During the winter, the sky tends to be clear and so more wind can flow, moreover, sunlight can easily fall on the solar panels. Hence, the most power generation takes place in winter altogether.

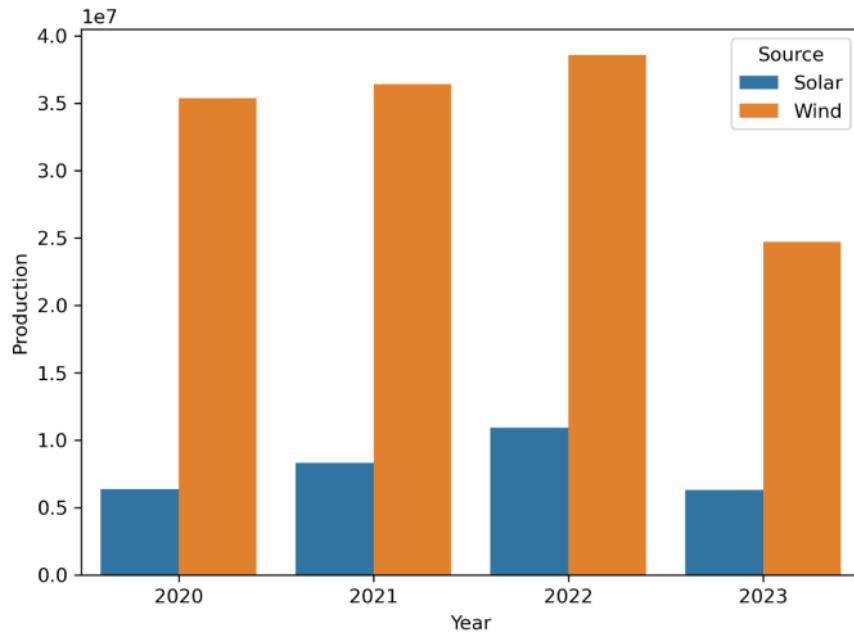


Fig 19. Annual Energy Production

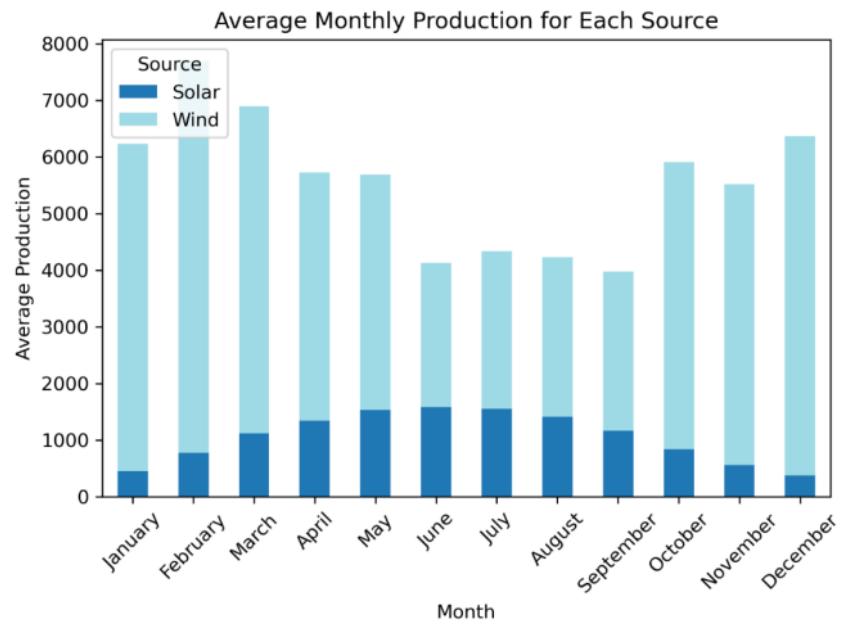


Fig 20. Average Monthly Production for Each Source

Figure 19 illustrates the annual energy production, showing that 2022 saw the highest output from both solar and wind sources. Over the years, the production increased for both until 2022, after that, it fell in 2023.

Figure 20 shows the average monthly power production of both solar and wind resources. It can be seen that wind produced more power than solar. The highest amount of power produced was in February by wind and by solar in June. Overall, wind was the most efficient and power-generating compared to solar power generation.

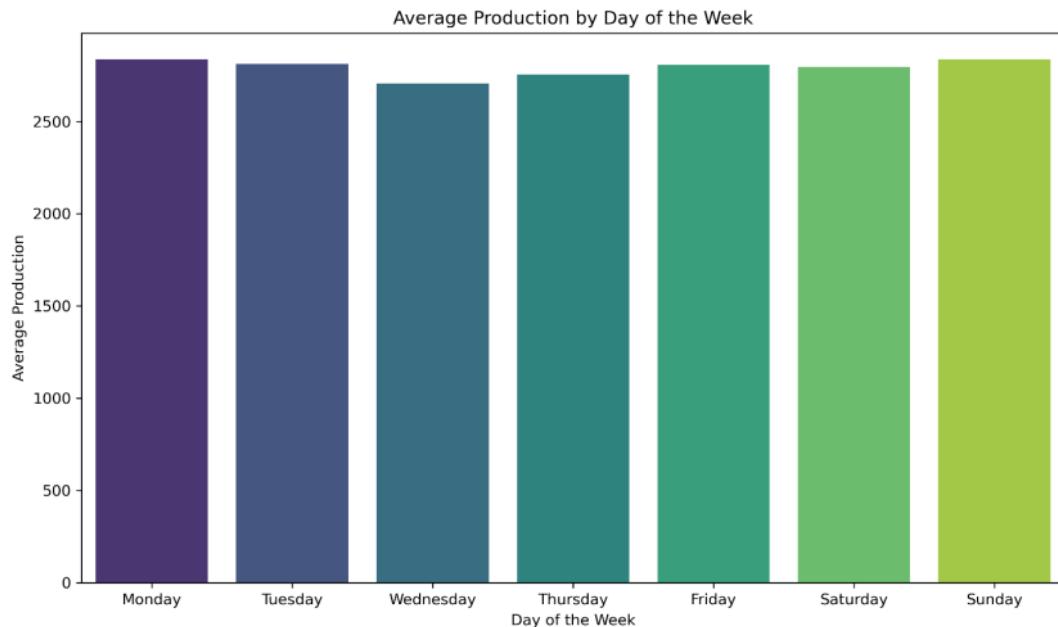


Fig 21. Average Production by Day of the Week

Figure 21 represents the average power production by each day of the week. As all the production level is pretty similar for all the days, it can be easily accepted that power production does not depend much on any specific day.

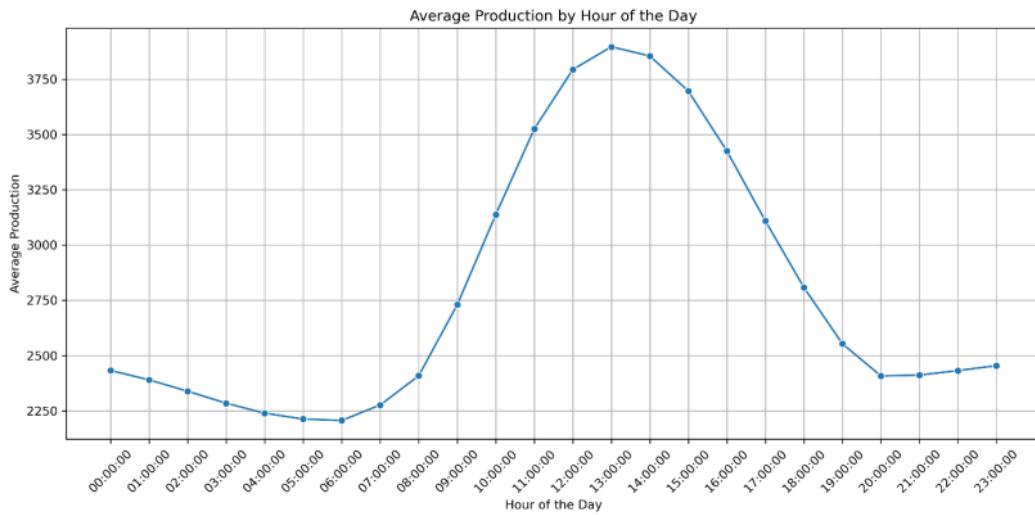


Fig 22. Average Production by Hour of the Day

Figure 22 represents the graph for average power production by hour of the day. It can be seen that most of the power is generated around the 9:00:00 to 18:00:00, for about 9 hours. This shows that it is more effective to generate power during the nights around these times.

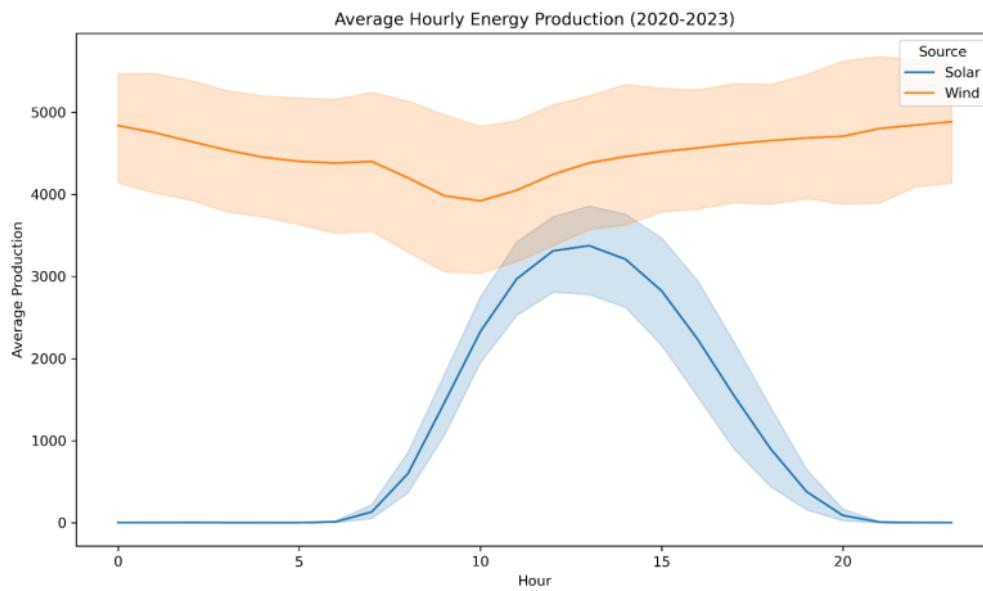


Fig 23. Average Hourly Energy Production (2020-2023)

Figure 23 shows that in the year of 2020-2023, most of the energy is generated by the wind compared to solar. But the production by wind fell during the hours of 7 to 18, which could mean that it was day and therefore, energy was efficiently generated by the solar panels.

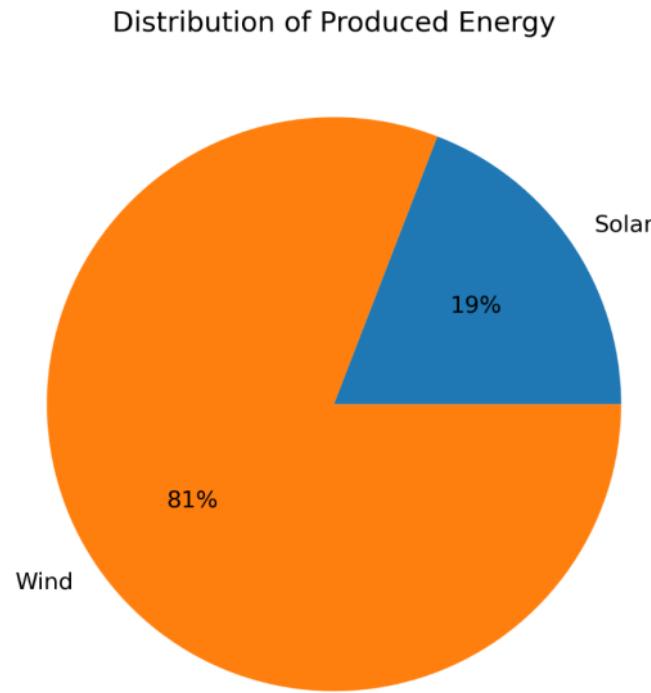


Fig 24. Distribution of Produced Energy

Figure 24 depicts that 81% of energy is produced by wind and 19% energy is produced by solar.

So, it can be effectively said that wind energy is more productive in producing energy compared to solar energy. But a factor should be considered as well i.e. a machine cannot be run 24/7, therefore, it is best to use the hybrid system in such a way that wind is used to generate energy when there is no sun and kept at rest during daylight as solar would be the best option to harness energy then.

Chapter 7: Testing

7.1 Integrate Models

Supervised Machine Learning Algorithm Regressor Models for Wind/Solar Power

Production: These models categorize power production into groups such as "low," "medium," or "high" based on input variables like temperature, wind speed, and sunshine. The process starts with data collection, preparation, and splitting into training and testing subsets. The model is trained to classify power production based on input features, with its effectiveness assessed through metrics such as accuracy, precision, recall, and F1 score. It is subsequently evaluated on new data to verify its generalization capabilities.

For wind power, the model factors in terrain and wind interactions, while for solar power, it considers radiation intensity and weather changes. Classification models offer benefits in swiftly and accurately classifying power production, though their effectiveness is contingent on the quality and relevance of the data.²⁶

Random Forest Classifier: Well-suited for classifying weather conditions impacting wind and solar power production, it employs an ensemble of decision trees to enhance accuracy and manage complex relationships between variables such as temperature and wind speed.

Gradient Boosting Classifier: Effective in identifying potential outages or failures by building a chain of models that correct previous mistakes. It improves forecasting accuracy by learning from historical data.

K-Neighbors Classifier: Useful for anomaly detection in power production data. It classifies data by proximity to other data points, making it a cost-effective method for identifying unusual conditions.

Support Vector Classifier: Best for binary classification tasks, such as determining if power production will be high or low. It selects an optimal hyperplane to separate data points, handling complex input-output relationships.

Linear Classifier: Suitable for simple sorting tasks with linear relationships. It quickly determines if power production will be above or below a threshold, offering straightforward computation.

CatBoostClassifier: Excels in handling datasets with categorical features like weather conditions. It uses a gradient-boosting algorithm to manage and encode categorical variables effectively, making accurate predictions based on complex interactions.

GridSearchCV: Used to optimize hyperparameters for models like Random Forest, K-Neighbors, Support Vector, Linear, and CatBoostClassifiers. It systematically tests combinations of parameters to improve model performance in tasks such as predicting energy production, identifying anomalies, and optimizing simple classifications.

Supervised Machine Learning Algorithm Regressor Models for Hybrid Model: In hybrid solar and wind power plants, supervised machine learning regression models are crucial for predicting and optimizing power output. These models learn from historical data to forecast future power production based on variables like solar irradiance, wind speed, and temperature. The process begins with data collection and preprocessing, ensuring quality and relevance. The data is then used to train a regression model, which maps environmental conditions to power output. This training involves tuning the model to minimize errors and accurately reflect how different factors influence power generation.

Following training, ^{the} model's performance is assessed on new data to confirm it generalizes effectively. Its effectiveness is evaluated using metrics such as accuracy and error rates. Various regression models can be applied, including simple linear regression for direct relationships, polynomial regression for non-linear patterns, and ensemble methods for combining multiple models to improve accuracy. Neural networks, particularly deep learning models, are powerful in handling large, complex datasets and provide highly accurate predictions.

Once validated, the model predicts live power output using real-time weather data, aiding in energy production decisions and grid integration. These regression models enhance the management and optimization of hybrid solar and wind power plants by providing accurate forecasts and improving efficiency.

Decision Tree Classifier: A straightforward algorithm that classifies data by making decisions based on specific features. It can forecast operational states or maintenance requirements in hybrid power plants using inputs such as weather conditions and energy output. Decision trees are clear and easy to understand, making them valuable for strategic system management.

Random Forest Classifier: An ensemble technique that generates multiple decision trees and combines their outcomes for greater stability and accuracy. It mitigates overfitting and enhances generalization, proving effective for complex interactions between features such as solar irradiance and wind speed.

Logistic Regression: Used for binary classification tasks, it predicts event probabilities based on input features. In hybrid power plants, it can forecast the likelihood of equipment failure or energy production surpassing thresholds, aiding operational planning.

Gradient Boosting Classifier: An ensemble approach that constructs a series of models in sequence, with each one addressing the errors of its predecessor. It identifies complex patterns in data, making it well-suited for tasks demanding high predictive accuracy in hybrid power plants.

AdaBoost Classifier: Another ensemble method that improves performance by focusing on misclassified instances. It is useful in fault detection or performance evaluation where classes are underrepresented, enhancing classification accuracy in hybrid energy systems.

K-Nearest Neighbors (KNN) Classifier: A non-parametric technique that classifies data points by identifying the majority class among their nearest neighbors. It is useful for categorizing operational modes or forecasting energy output in hybrid power plants, although it can be computationally demanding with large datasets.

GridSearchCV: A tool for hyperparameter optimization, it evaluates predefined hyperparameter values through cross-validation. In hybrid power plants, GridSearchCV fine-tunes parameters for models like Decision Trees, Random Forest, Gradient Boosting, and KNN, ensuring optimal performance and preventing overfitting.

These models, ranging from basic decision trees to advanced ensemble methods, provide various advantages for managing and optimizing hybrid solar and wind power plants. The selection of a model depends on specific task needs, such as accuracy, interpretability, and the capacity to manage complex data interactions.

Chapter 8: Analysis

8.1 Validation Result Analysis

Table 3. Initial Model Results for Windmill Power Production

	MSE	MAE	R2 Score
Linear Regression	4.237552	1.585079	0.432065
Random Forest Regressor	0.352166	0.325774	0.952801
Support Vector Regressor	2.062976	1.002725	0.723511
Gradient Boosting	0.442132	0.419548	0.940744
K-Neighbors	2.892978	1.295511	0.612271
CatBoostRegressor	0.417199	0.408115	0.944085

The table titled "Initial Model Results for Windmill Power Production" in the text shows three different regression model results which have been examined about their prediction of wind power production from windmills. The table consists of the following: Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared (R^2) score for each model. 19

8

Mean Squared Error (MSE) measures the average squared difference between the predicted and actual values, with lower values of the better model performance. It is the model "Random Forest Regressor" which has been the most accurate in predicting windmill power production more times. It has the smallest squared error with MSE = 0.352166, which is the lowest among other models, meaning there is less difference between the actual and predicted windmill power production. So, the Random Forest Regressors are valid and the most accurate among them. On the other hand, linear regression is the one that shows the highest MSE. As a result, it gives 4.237552 as the value of the model's prediction error, which is higher than the other models.

13

The Mean Absolute Error (MAE) is a measure showing the average size, in terms of the difference between the predicted and actual values, of the errors in the predictions. The Random Forest Regressor is the model that performs best according to the metric of MAE which is the lowest of 0.325774. This confirms its legitimacy. In contrast, the Linear Regression variety displays the most MAE of 1.585079, thereby indicating that, on average, its predictions are more divergent from the actual values compared to the predictions given by the other models.

7

The R-squared (R^2) score indicates the proportion of variance in the dependent variable that is predictable from the independent variables. The R^2 score is a statistical parameter that sheds light on the model's frequency of fitting a dataset. The answer is the Random Forest Regressor first with a high R^2 score of 0.952801 because it captures about 95% of the variance in the power output. After that, models changing and tight are set having also high R^2 ones: The R^2 cores for Gradient Boosting and CatBoostRegressor are 0.940744 and 0.944085, correspondingly tell that they are promising in the prediction of good performance. As the opposite Linear Regression has only the highest R^2 score indicator having 0.432065. This follows that Linear Regression can be put as a model using Varying coefficient regression for grouping with R being the variable. Hence, from the variance perspective, the data is well represented by the output from one model which is $R^2=0.432065$ and it is the inverse of the other values which are increasing for the other models.

For the most part, the Random Forest Regressor model is the most stable one in the context of windmill power production prediction. Its unmatched results in the MSE, MAE, and R² scoring systems are the manifestation of its competence to capture the data's patterns and give forecasts with high precision.

Table 4. Optimized Model Results for Windmill Power Production

	MSE	MAE	R2 Score
Random Forest	0.422544	0.377724	0.943369
Gradient Boosting	0.308694	0.323037	0.958628
K-Neighbors	2.737085	1.265830	0.633164
Support Vector Regressor	1.670192	0.913549	0.776154
Linear Regression	4.235872	1.585038	0.432290
CatBoostRegressor	0.334904	0.346686	0.955115

The table titled "Optimized Model Results for Windmill Power Production" enumerates the performance indicators of the various regression models after optimization, concentrating on the one that best predicts wind turbine power generation. Among the metrics included are mean-square error (MSE), mean absolute error (MAE), and R-squared (R²) score.

¹¹

Mean Squared Error (MSE) measures the average squared differences between predicted and actual values, with smaller values reflecting better models. In this case, the model with the least MSE of 0.308694 is Gradient Boosting, which leads the race by the slimmest of margins in completing the predictions with the smallest average squared error of the rest of the optimized ones. It is, therefore, an indication of the possibility of data leakage or multicollinearity for Gradient Boosting which may have lent it to overfitting. A disturbingly high MSE of 167.60 is evidence of this. Error is the difference between the truth and what the model believes in this case the lower the value the better the model is performing.

²

Mean Absolute Error (MAE) is a mean of the differences between the actual and predicted values as with the negative errors' signs removed, the average positive errors. Equally, norming the value gives a new normalized Value, in which every parameter that you passed in becomes a new parameter that is normed according to the amount it contributes to the

prediction. Like min-max, the gradient is either zero or infinity for the activation cells. In those cases, the regularization term should be active to limit the rate of learning and stop the model from diverging.

The R-squared score is a measure of the amount of variance in the dependent variable explained by the independent variables. A high R^2 score suggests a well-fit model for data. The model with the highest R^2 score of 0.958628 is the Gradient Boosting model, implying that it explains most of the windmill's energy production variance, and therefore, it seems the most effective model. The model called CatBoostRegressor, on the other hand, has a rating of 0.955115 which of the above is ideal for highway energy harvesting. Linear Regression on the other hand has a low R^2 score of 0.432290 which shows that it is far less effective and the most inadequate model.

In general, the commanded Gradient Boosting model demonstrates it standing out as the best model for predicting windmill power production, being the best in all three performance metrics compared to others: MSE, MAE, and R^2 . This means that the optimization process has, indeed, brought this change. CatBoostRegressor is also good at MAE and R^2 , but the Gradient Boosting algorithm outperforms it in the general case of prediction. Linear Regression, which still has a poor performance notwithstanding its optimization progress, is now showing more room for improvement compared to the other models.

Table 5. Initial Model Results for Solar Power Production

	MSE	MAE
Linear Regression	0.155032	0.266655
Random Forest Regressor	0.000000	0.000000
Support Vector Regressor	1.000000	1.000000
Gradient Boosting	0.024038	0.060821
K-Neighbors	0.123471	0.148265
CatBoostRegressor	0.012205	0.037256

Table "Initial Model Results for Solar Power Production" shows a summary view of how well the different regression models do in solar power production without any type of optimization. In these results, resulting metrics include the Mean Squared Error and the Mean Absolute Error. Mean Squared Error (MSE) is the average of the squared difference between predictions and the observed/target data points. The objective is to reduce the value of MSE while fitting a model to the dataset. It is clear from Table 3 that the value of MSE for the Random Forest Regressor attains the minimum value of 0.000000; it leads to the perfect prediction of solar power production without error and, hence, gives an excellent sign of aptitude. CatBoostRegressor is also having an efficient performance with a quite lower value of MSE: 0.012205, which is my average squared error while predicting. MSE for Gradient Boosting: 0.024038, showing fairly good accuracy, just marginally less than the most optimum CatBoostRegressor. The Support Vector Regressor holds an accuracy of 1.000000 with an MSE, making this model comparatively more erroneous in its predictions than others.

Mean Absolute Error represents the average magnitude of errors in the predictions. A lower MAE shows that the model makes its estimations much closer to the actual ones. The same applies here, where Random Forest Regressor recorded zero errors in its predictions, while its MAE is equal to 0.000000. It scores very well; its MAE equals 0.037256, which means a pretty low average error is being fulfilled. On the other hand, Gradient Boosting with an MAE of 0.060821, while pretty respectable, is much lower, considering that most algorithms show about the same performance in purely accurate predictions. In contrast, the Support Vector Regressor has the highest MAE equal to 1.000000, which means the model's predictions deviate, on average, from the actuals by a great extent. On the whole, the MSE and MAE for the prediction of solar power production with the Random Forest Regressor model are really outstanding, even in this baseline evaluation. CatBoostRegressor also did well, particularly in MAE; being high in value gives a high predictive accuracy. Though good enough, Gradient Boosting's accuracy is lesser than the best-preferred models, that is, CatBoost and XGBoost. K-Neighbors and Support Vector Regressors are less effective, as reflected by their higher MSE and MAE, with Support Vector Regressors being the worst among the listed.

Table 6. Optimized Model Results for Solar Power Production

6	MSE	MAE
Random Forest	0.048085	0.148474
Gradient Boosting	0.076384	0.000000
K-Neighbors	0.595918	0.526512
Support Vector Regressor	1.000000	0.944024
Linear Regression	0.950576	1.000000
CatBoostRegressor	0.000000	0.038685

The following is the table "Optimized Model Results for Solar Power Production", which displays the analysis of different models created through simulation and optimized to go W.R.T. the solar power generation. The two key performance metrics presented in this table are Mean Squared Error and Mean Absolute Error.

18
The mean squared error is one of the metrics used to estimate the mean squared difference between the forecasted and factual values. The lower the MSE is, then the model is better optimized with respect to prediction accuracy. Out of all, MSE values of 0.000000 really make CatBoostRegressor stand out very accurately in predicting the production of solar power, with no squared error. Thus, this is indeed a remarkable feat that signifies the CatBoostRegressor has been optimized to the data that it conforms to. The Random Forest model also exemplifies a good fit with its MSE being very low at 0.048085, which suggests quite accurate predictions, yet not as improved as the CatBoostRegressor. The Support Vector Regressor and Linear Regression models are on the other end of the scale, being the models that have the highest MSE values (1.000000 and 0.950576, respectively), thus their predictions are the least accurate compared to the other models.

2
Mean Absolute Error (MAE) calculates the errors' average absolute value that could be made as predictions besides their orientation without regard to their sign. It is not essential to mention that the lower MAE is, the closer to the foreseen figures the model's predictions are. Gradient Boosting results in a perfect MAE of 0.000000, which is extremely good and indicates no error. CatBoostRegressor also stands out with the MAE value of 0.038685 and implies very low average prediction error. In contrast, only the Linear regression model has

the highest MAE with a value of 1.000000, which points to a fault in the predicted values. The Support Vector Regressor also has a large MAE value of 0.944024, which suggests considerable prediction error. On the whole, solar power production forecasting is dominated by CatBoost Regressor which is the #1, being the best track record in both MSE and MAE. Its capability to obtain almost the same errors in both criteria shows its superlative correctness. In the meantime, the Gradient Boosting model also does an excellent job, especially in terms of MAE. In contrast, its MSE is a little higher than CatBoostRegressor's. The Random Forest model although it is effective, has not yet achieved near-perfect performance as CatBoostRegressor and Gradient Boosting. The Linear Regression and Support Vector Regressor are not as effective as in the case of the other models, demonstrated by their higher MSE and MAE scores implying that they display bigger prediction errors.

Table 7. Model Performance on Hybrid Power Generation System

Model	Accuracy
Decision Tree	89%
Random Forest	90%
Linear Regression	90%
Gradient Boosting	88%
K-Nearest Neighbors	88%
AdaBoost	82%

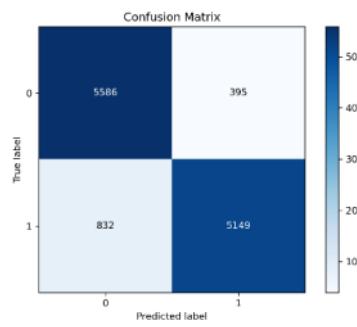


Fig 25. Random Forest CM

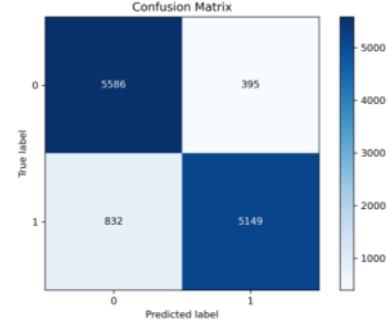


Fig 26. Linear Regression CM

Fig 25 shows the confusion matrix of random forest where it depicts the results of the model's performance. The results are as follows: True Negative = 5586, True Positive = 5149, False Negative = 395, False Positive = 832.

Fig 26 shows the confusion matrix of linear regression where it depicts the results of the model's performance. The results are as follows: True Negative = 5586, True Positive = 5149, False Negative = 395, False Positive = 832.

These two confusion matrixes produce the results for the best-performing models having the highest possible accuracy of 90%.

8.2 Analysis of Test Optimization

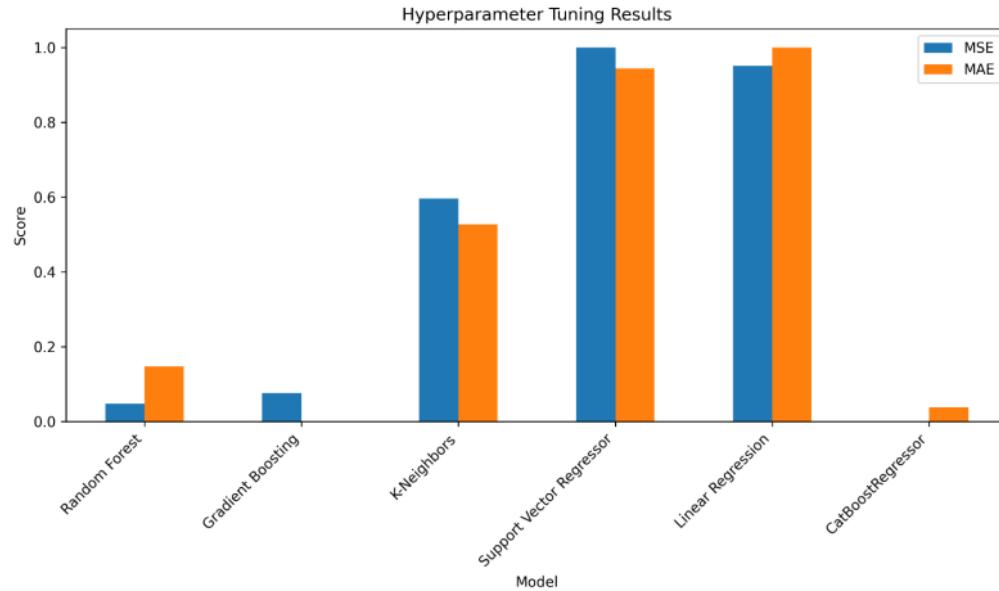


Fig 27. Hyperparameter Tuning Results (Solar)

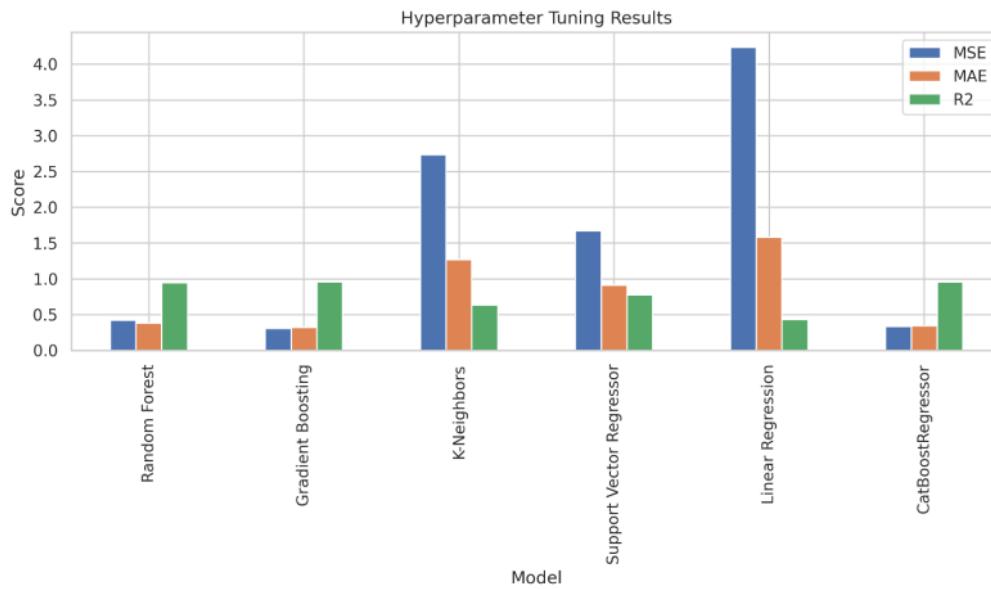


Fig 28. Hyperparameter Tuning Results (Wind)

Figure 27 represents how the performance results changed after tuning the hyperparameters of the solar-powered generator dataset. It can be seen that SVC and Linear Regression had the highest MSE and MAE respectively, indicating they caused the highest errors and hence should not be used. On the other hand, CatBoostRegressor and Gradient Boosting had the lowest MSE and MAE respectively, and should be considered more efficient and reliable to use.

Figure 28 shows the bar graph for the hyperparameter tuning results for all the models used in training the wind-powered generator dataset. It is found that linear regression still had the highest MSE and MAE and the lowest R2 score and hence should not be used. On the other hand, gradient boosting and catboostregressor yielded the lowest MSE and MAE and the highest R2 score, indicating they should be given priority as they would predict more accurately.

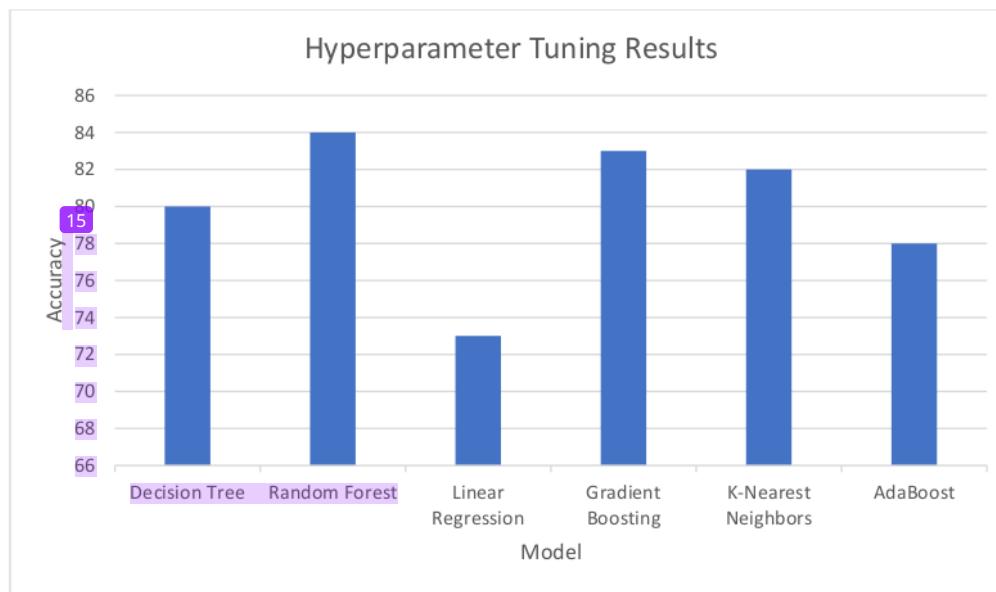


Fig 29. Hyperparameter Tuning Results (Hybrid)

Figure 29 represents the hyperparameter tuning results for the hybrid system. It can be seen that the tuned models produced unfavored results compared to the untuned models and hence it was unwise to use these models.

Chapter 9

9.1 Monitoring Procedures for the model

To facilitate successful operations and constant model enhancements, a broad monitoring framework shall be used. Performance evaluation involves measuring different standards of validation such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared (R^2) scores at training time as well as thereafter. Cross-checking these metrics with the benchmarks is important for detection of over-fitting or under-fitted cases. A separation dataset will be used for conducting model evaluations periodically paying particular attention to accuracy in confusion matrices and classification reports that help to detect performance problems. To maintain data integrity, monitoring of data quality is done by monitoring data drift alongside routine training checks. Furthermore, technical facilities are going to run continuous performance checks for example when it comes to bottlenecks handling or any limitations that might arise. Also, algorithm efficiency and complexity would be managed through consistent reviewing as well as optimizing the model's computations. Integration and deployment will involve close monitoring through gradual testing with the assistance of a real-time performance-tracking system that guarantees interoperability and smooth operation throughout the entire process. User feedback will also be solicited on a continuous basis for analysis purposes which will help in identifying usability issues thereby enhancing acceptance rate. Moreover, by constantly assessing and enhancing the computation proficiency of the model, both the performance and intricacy of algorithms will be supervised. The integration and deployment process will be subjected to continuous incremental tests and real-time performance tracking to confirm compatibility and seamless functioning. Any usability problems that may arise will be dealt with through the active collection and examination of user feedback to assist in fostering adoption. To avail stakeholders of continuous information, revise monitoring processes, and ascertain effectiveness of the model continuously, regular reports as well as periodic assessments will be done.

9.2 Risk Assessment:

Here, the risk assessment table 8 for the project on the Optimization and Simulation of a Hybrid Solar and Wind-Powered Industrial Data Center which looks into various models and their abilities to predict wind and solar power generation performance is given below:

Table 8: A table showing the risk assessment of the project

Risk No.	Description of the Risk	Probability of the Risk	Effect on the Project	Contingencies to Mitigate the Risk
GR1	Technical Risks	MEDIUM TO HIGH	Technical issues could delay the project, impacting model training and deployment, or result in reduced system accuracy. 2	Conduct regular technical reviews and validations during each phase of the project to identify and address potential issues early.
GR2	Data Quality Risks	MEDIUM	Poor data quality could lead to inaccurate model predictions and ineffective waste sorting.	Implement rigorous data preprocessing and augmentation techniques to enhance the quality and reliability of input data. 20
GR3	Resource Risks	MEDIUM	Insufficient computational resources or expertise may delay model training and system deployment.	Ensure access to adequate computing resources and allocate sufficient time for model training and validation.

GR4	Schedule Risks	MEDIUM	Delays in project phases could lead to missed deadlines and potential cost overruns.	Utilize project management tools to monitor progress and adjust resource allocation as needed to stay on schedule.
SR1	Model Performance Risk	HIGH	Inadequate model performance could result in poor classification accuracy, affecting the overall effectiveness of the waste sorting system.	Continuously monitor model performance using validation data and implement early stopping to prevent overfitting.
SR2	Algorithm Complexity Risk	MEDIUM TO HIGH	High algorithmic complexity might make the model difficult to interpret and maintain, increasing debugging and testing time.	Simplify model architecture where possible and modularize complex algorithms for easier troubleshooting and updates.
SR3	Data Drift Risk	MEDIUM	Changes in waste types or distributions over time could lead to a decrease in model accuracy.	16 Implement monitoring mechanisms to detect data drift and retrain models periodically with updated data.

SR4	Testing and Validation Risk	MEDIUM	Insufficient testing may result in unrecognized errors, leading to system failures during deployment.	Develop and execute a comprehensive testing plan that covers all possible scenarios, including edge cases.
SR5	Integration Risk	MEDIUM	Difficulty integrating the model with existing waste sorting systems may cause delays or reduced effectiveness. 9	Plan for integration early in the project, and conduct incremental testing to ensure compatibility with existing systems. 9
SR6	User Adoption Risk	LOW TO MEDIUM	Users may resist adopting the new system due to unfamiliarity or perceived complexity.	Provide comprehensive training and support for users, and involve them in the development process to ensure the system meets their needs.

Table 8 lists out possible hazards connected with the efficiency of models used to predict both solar energy production as well as wind generation and offer effective methods for minimizing these dangers.

9.3 Government Risk Assessment (Local): Keep Blank

9.4 Government Risk Assessment (Global): Keep Blank

Chapter 10

IDC Operators: Keep Blank

IDC Managers: Keep Blank

IDC on AWS: Keep Blank

References

- [1] D. Turek and P. Radgen, "Optimized data center site selection—mesoclimatic effects on data center energy consumption and costs," *Energy Efficiency*, vol. 14, no. 3, Mar. 2021, doi: <https://doi.org/10.1007/s12053-021-09947-y>.
- [2] Enida Sheme, Neki Frasher, S. Holmbacka, S. Lafond, and Drazen Luzanin, "Datacenters powered by renewable energy: A case study for 60 degrees latitude north," Sep. 2016, doi: <https://doi.org/10.1109/softcom.2016.7772131>.
- [3] I. Goiri, W. Katsak, K. Le, T. D. Nguyen, and R. Bianchini, "Designing and Managing Data centers Powered by Renewable Energy," *IEEE Micro*, vol. 34, no. 3, pp. 8–16, May 2014, doi: <https://doi.org/10.1109/mm.2014.6>.
- [4] Mohd. R. S. Shaikh, "A Review Paper on Electricity Generation from Solar Energy," *International Journal for Research in Applied Science and Engineering Technology*, vol. V, no. IX, pp. 1884–1889, Sep. 2017, doi: <https://doi.org/10.22214/ijraset.2017.9272>.
- [5] A. O. M. Maka and J. M. Alabid, "Solar energy technology and its roles in sustainable development," *Clean Energy*, vol. 6, no. 3, pp. 476–483, Jun. 2022, doi: <https://doi.org/10.1093/ce/zkac023>.
- [6] Z. Gan, "Research on Wind Power Generation Technology in New Energy Power Generation," *IOP Conference Series: Earth and Environmental Science*, vol. 651, no. 2, p. 022013, Feb. 2021, doi: <https://doi.org/10.1088/1755-1315/651/2/022013>.
- [7] C. Huang, C. Liu, M. Zhong, H. Sun, T. Gao, and Y. Zhang, "Research on Wind Turbine Location and Wind Energy Resource Evaluation Methodology in Port Scenarios," *Sustainability*, vol. 16, no. 3, pp. 1074–1074, Jan. 2024, doi: <https://doi.org/10.3390/su16031074>.
- [8] R. Yadav and Mr. M. Sharma, "Hybrid Power Generation System Using Solar -Wind Energy: A Review," *International Journal of Trend in Scientific Research and Development*, vol. Volume-2, no. Issue-3, pp. 941–946, Apr. 2018, doi: <https://doi.org/10.31142/ijtsrd11115>.

[9] A. Rahaman, K. N. Noor, T. A. Abir, S. Rana, and M. Ali, "Design and Analysis of Sustainable Green Data Center with Hybrid Energy Sources," *Journal of Power and Energy Engineering*, vol. 09, no. 07, pp. 76–88, 2021, doi:

<https://doi.org/10.4236/jpee.2021.97006>.

[10] niloshandj, "Functions of Computer Hardware Components," *Functions of Computer Hardware Components*, 2020.

<https://computerhardwarecomponentsandfunctions.wordpress.com/>

[11] "Solar energy power generation dataset," www.kaggle.com.

<https://www.kaggle.com/datasets/stucom/solar-energy-power-generation-dataset>

[12] N. Gupta, "Predict the Power(Kw/h) produced by Windmills," Kaggle.com, 2021.

<https://www.kaggle.com/datasets/emnikkhil/predict-the-powerkwh-produced-by-windmills> (accessed Aug. 25, 2024).

[13] H. Upton, "Wind & Solar Daily Power Production," Kaggle.com, 2020.

<https://www.kaggle.com/datasets/henriupton/wind-solar-electricity-production> (accessed Aug. 25, 2024).

[14] G. Lawton, "Data Preprocessing: Definition, Key Steps and

Concepts," *SearchDataManagement*, Jan. 2022.

<https://www.techtarget.com/searchdatamanagement/definition/data-preprocessing>

Report-Hybrid Powered Industrial Data Center

ORIGINALITY REPORT



PRIMARY SOURCES

- | Rank | Source | Percentage |
|------|--|------------|
| 1 | Arti Jain, Rajeev Kumar Gupta, Mohit Kumar.
"chapter 16 Solar Power Forecasting Using
Machine Learning Techniques", IGI Global,
2024
Publication | 1 % |
| 2 | fastercapital.com
Internet Source | 1 % |
| 3 | Submitted to University of Wolverhampton
Student Paper | 1 % |
| 4 | Richard J. Roiger. "Data Mining - A Tutorial-
Based Primer", CRC Press, 2017
Publication | <1 % |
| 5 | www.fastercapital.com
Internet Source | <1 % |
| 6 | Luka Jovanovic, Nebojsa Bacanin, Rejitha
Ravikumar, Milos Antonijevic, Goran Radic,
Miodrag Zivkovic. "Chapter 46 Generative
Adversarial Networks for Synthetic Training
Data Replacement in Phishing Email Detection
Using Natural Language Processing", | <1 % |

-
- 7 Submitted to Queen's College <1 %
Student Paper
- 8 Haowen Zheng, Yao Lu, Zekun Sun, John Panneerselvam, Xiang Sun, Lu Liu. "Energy optimisation in cloud datacentres with MCTIDE: Mixed Channel Time-series Dense Encoder for workload forecasting", Applied Energy, 2024 <1 %
Publication
- 9 Submitted to University of Wales Institute, Cardiff <1 %
Student Paper
- 10 Submitted to University of Glamorgan <1 %
Student Paper
- 11 www.ijraset.com <1 %
Internet Source
- 12 Oubah Isman Okieh, Serhat Seker, Seckin Gokce, Martin Dennenmoser. "An Enhanced Forecasting Method of Daily Solar Irradiance in Southwestern France: A Hybrid Nonlinear Autoregressive with Exogenous Inputs with Long Short-Term Memory Approach", Energies, 2024 <1 %
Publication

13	epdf.pub Internet Source	<1 %
14	medium.com Internet Source	<1 %
15	Submitted to CSU, San Jose State University Student Paper	<1 %
16	Submitted to Instituto de Empress S.L. Student Paper	<1 %
17	jsss.copernicus.org Internet Source	<1 %
18	www.nature.com Internet Source	<1 %
19	Submitted to University of Hull Student Paper	<1 %
20	"Portfolio risk management model based on machine learning", Financial Engineering and Risk Management, 2023 Publication	<1 %
21	om.opensooq.com Internet Source	<1 %
22	Luiza Araujo Costa Silva, Luis Gonzaga Baca Ruiz, David Criado-Ramón, Joao Gabriel Bessa et al. "Assessing the Impact of Soiling on Photovoltaic Efficiency using Supervised	<1 %

Learning Techniques", Expert Systems with Applications, 2023

Publication

-
- 23 hrcak.srce.hr <1 %
Internet Source
-
- 24 www.oneteam.it <1 %
Internet Source
-
- 25 www3.db-thueringen.de <1 %
Internet Source
-
- 26 Gurwinder Singh, Harun. "Chapter 6 Multiple Linear Regression Based Analysis of Weather Data for Precipitation and Visibility Prediction", Springer Science and Business Media LLC, 2023 <1 %
Publication
-
- 27 Kaijun Li, Liming Kang, Kexin Guo, Linye Song, Xinghui Zhang, Wei Xu. "Risk assessment of respiratory droplet infections caused by coughing in various indoor dynamic environments", Journal of Building Engineering, 2023 <1 %
Publication
-
- 28 Moin Ahmed, Zhiyu Mao, Yun Zheng, Tao Chen, Zhongwei Chen. "Electric Vehicle Range Estimation Using Regression Techniques", World Electric Vehicle Journal, 2022 <1 %
Publication
-

29

<link.springer.com>

Internet Source

<1 %

30

Malhar Khan, Muhammad Amir, Adil Hussain, Yaseen Badar, Muhammad Sharif, Muhammad Faisal. "Enhancing Solar Power Forecasting in Multi-Weather Conditions Using Deep Neural Networks", 2023 2nd International Conference on Emerging Trends in Electrical, Control, and Telecommunication Engineering (ETECTE), 2023

<1 %

Publication

Exclude quotes

On

Exclude matches

Off

Exclude bibliography

On