Final Project Report

Predicting The Energy Output Of Wind Turbine Based On Weather Condition

1. Introduction

1.1. Project overviews

The project focuses on predicting the energy output of wind turbines based on weather conditions, which holds significant importance for energy companies and grid operators aiming to optimize energy production and distribution. By leveraging historical data on weather patterns and energy output, the project seeks to develop machine learning models capable of making accurate predictions of wind turbine energy output given current or forecasted weather conditions. This predictive capability enables better decision-making in several critical areas of wind energy management.

One of the primary applications of this project is energy production forecasting. Energy companies can utilize the machine learning models to forecast the energy production of their wind turbines over specific periods. Accurate predictions based on weather forecasts help these companies make informed decisions about energy distribution and pricing, ensuring that they can meet demand effectively while minimizing costs. Additionally, this forecasting ability supports maintenance planning for wind farm operators, who can schedule maintenance activities during periods of low wind activity to minimize downtime and maximize energy production.

Furthermore, the project plays a crucial role in grid integration. By predicting the energy output of wind turbines, grid operators can efficiently balance the energy supply by adjusting the output of other energy sources accordingly. This capability is essential for maintaining grid stability and reliability, especially as the share of wind energy in the overall energy mix increases. By the end of the project, participants will gain expertise in data preprocessing, visualization, and the application of various machine learning algorithms. They will also acquire the skills to build a web application using the Flask framework, providing a practical platform for deploying the predictive models.

1.2. Objectives

1. Develop Predictive Models: Create accurate machine learning models to predict wind turbine energy output based on weather conditions using historical data.

2. Energy Production Forecasting: Enable energy companies to forecast wind turbine energy production over specified periods, supporting efficient energy distribution and pricing decisions.

3. Maintenance Scheduling: Assist wind farm operators in planning maintenance schedules by predicting periods of low wind activity, thereby minimizing downtime and maximizing energy production.

4. Grid Integration Optimization: Aid grid operators in efficiently integrating wind energy into the grid by predicting energy output and adjusting the output of other energy sources to maintain grid stability.

2. Project Initialization and Planning Phase

2.1. Define Problem Statement

Wind farm operators are facing significant challenges due to the unpredictable nature of weather conditions, which result in substantial fluctuations in energy output. These fluctuations complicate the process of meeting energy demand, lead to inefficient energy storage, and increase operational costs. To address these issues, there is a need for an accurate predictive model that can forecast wind turbine energy output based on real-time and forecasted weather conditions. Such a model would help optimize turbine operations, maximize energy production, and provide insights for proactive maintenance scheduling, thereby reducing downtime and extending the lifespan of turbines. Additionally, reliable predictions are essential for better integration with the energy grid, ensuring a stable and consistent supply of electricity. By developing and validating a predictive model that uses historical weather and turbine performance data, and integrating it with real-time weather information, wind farm operators can enhance the efficiency and reliability of their operations, resulting in improved resource management, cost savings, and a more stable energy supply for their customers.

Problem Statement (PS)	**I am (Customer)**	**I'm trying to**	**But**	**Because**	**Which makes me feel**
PS-1	Power Supplier	predict the energy output of a wind turbine	I struggle because I lack real-time weather data and accurate predictive models.	I do not have any valid predictions.	hesitant about relying on renewable energy sources
PS-2	I am a farmer	estimate the energy output of my wind turbine	I face challenges because existing methods don't provide accurate	No reliable predictor.	unsure about investing more in renewable energy solutions.

2.2. Project Proposal (Proposed Solution)

Project Overview	
Objective	The project aims to develop a predictive model for forecasting wind turbine energy output based on weather conditions. This model seeks to optimize operations, enhance energy production efficiency, schedule maintenance proactively, and improve wind energy integration into the electrical grid, providing actionable insights for better resource management and cost savings for wind farm operators.

Scope	The project involves collecting and processing historical weather and wind turbine performance data to develop a machine learning model for predicting wind turbine energy output. Real-time weather data will be integrated for ongoing predictions. The aim is to optimize turbine operations, provide maintenance insights, and improve grid integration by offering stable energy supply forecasts. Additionally, the project includes creating a user-friendly interface
Problem Statement	
Description	Wind farm operators grapple with unpredictable energy output due to varying weather conditions, leading to difficulties in meeting demand and increased costs. They need an accurate predictive model to forecast energy output based on weather data, optimizing operations and ensuring a stable energy supply.
Impact	Accurate wind energy predictions optimize operations, reduce costs, and foster renewable energy adoption while improving reliability through proactive maintenance.
Proposed Solution	
Approach	We have used random forest regressor model and fit the data.
Key Features	Key features include real-time weather integration, theoretical power curve modeling, machine learning for accurate predictions, user-friendly interface, and performance validation metrics.

Resource Requirements

Resource Type	**Description**	**Specification/Allocation**
Hardware		
Computing Resources	CPU/GPU specifications, number of cores	2 x NVIDIA V100 GPUs

Memory	RAM specifications	8 GB
Storage	Disk space for data, models, and logs	1 TB SSD
Software		
Frameworks	Python frameworks	Flask
Libraries	Additional libraries	scikit-learn, pandas, numpy
Development Environment	IDE, version control	Jupyter Notebook, Git
Data		
Data	Source, size, format	Kaggle dataset, 10,000 images

2.3. Initial Project Planning

The project was divided into several sprints, each meticulously planned to achieve specific objectives within set timeframes. Each sprint was structured around functional requirements (epics) that defined the overarching goals and tasks to be completed. These tasks were broken down into user stories, each assigned a story point value, priority, team members responsible, and planned start and end dates.

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members	Sprint Start Date	Sprint End Date (Planned)
Sprint-1	Registration	USN-1	Data Collection	2	Medium	Vardhan, Manju Bhuvan	8th July	8th July
Sprint-1		USN-2	Visualising and Analizing Data	1	High	Sriram, Akmal	8th July	9th July
Sprint-2		USN-3	Data Processing	2	Medium	Sriram, Akmal	9th July	10th July
Sprint-1		USN-4	Model Building	2	High	Vardhan Raj Kumar	10th July	12th July
Sprint-1	Login	USN-5	Application Building	1	High	Vardhan Raj Kumar, Manju Bhuvan	12th July	14th July

3. Data Collection and Preprocessing Phase

3.1. Data Collection Plan and Raw Data Sources Identified

Data Collection Plan

Section	Description
Project Overview	This machine learning project aims to predict wind turbine energy output based on weather data inputs, leveraging regression or machine learning models to optimize operational efficiency and ensure stable energy supply for wind farm operators.
Data Collection Plan	The data for this project can be collected from meteorological agencies, weather APIs (such as OpenWeatherMap or WeatherStack), historical weather databases, and potentially from sensors installed on wind turbines themselves for real-time measurements. We took the data from Kaggle.
Raw Data Sources Identified	**Historical Weather Databases**: Databases storing archived weather data spanning several years.

Raw Data Sources

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle	1. LV ActivePower (kW): The power generated by the turbine for that moment 2. Wind Speed (m/s): The wind speed at the hub height of the turbine (the wind speed that turbine use for electricity generation) 3.Theoretical_Power_Curve (KWh): The theoretical power values that the turbine generates with that wind speed which is given by the turbine manufacturer 4. Wind Direction (°): The wind direction at the hub height of the turbine (wind turbines turn to this direction automaticly)	https://drive.google.com/file/d/1s8DCU3CdEkYEtcxTrpv1LE16WaHgqj38/view	CSV	3.2 MB	Public

3.2. Data Quality Report

Data Source	Data Quality Issue	Severity	Resolution Plan
Kaggle Dataset	No issues were found with the data.	Low	None

3.3. Data Exploration and Preprocessing

Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

Section	**Description**
Data Overview	We have a data of LV Active power, Theoritcal wind curve, Wind direction and wind speed from 1st Jan 2018 to 31st Dec 2018 for every 10 mins.
Univariate Analysis	In my wind turbine energy prediction project, univariate analysis involves examining a single variable, such as wind speed, using histograms for visualizing frequency distribution, and calculating summary statistics like mean and standard deviation. Box plots help identify outliers, providing insights crucial for building an effective predictive model.
Bivariate Analysis	In my wind turbine energy prediction project, bivariate analysis examines the relationship between two variables, like wind speed and active power output. This helps identify correlations and patterns using scatter plots and correlation coefficients, providing insights into how changes in one variable affect the other, which is crucial for accurate predictions.
Multivariate Analysis	In my wind turbine energy prediction project, multivariate analysis explores how combinations of variables like wind speed, direction, and temperature collectively influence energy output. It uses techniques such as multivariate regression or PCA to uncover complex relationships and enhance predictive accuracy by understanding interconnected factors affecting turbine performance.

Outliers and Anomalies	In my wind turbine energy prediction project, outliers and anomalies are significant deviations from the dataset. They can skew analyses and predictions if not addressed properly. Detecting and handling these data points is crucial for accurate modeling and reliable predictions for turbine performance based on weather conditions.

Data Preprocessing Code Screenshots

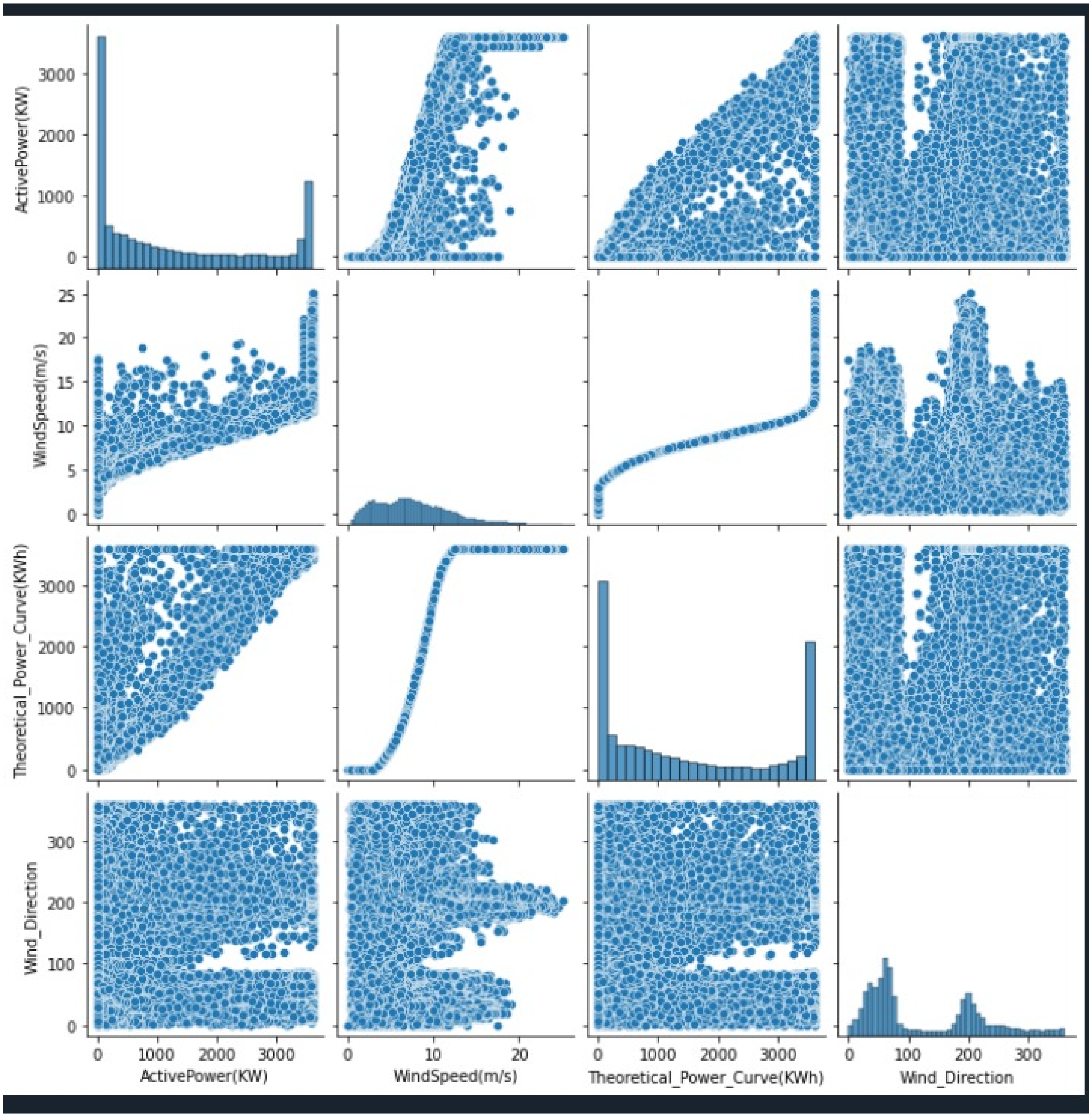
Loading Data	`# Function to load and preprocess the data`
	`def load_and_preprocess_data(path):`
	`    df = pd.read_csv(path)`

Handling Missing Data	nill

Data Transformation	`df.rename(columns={`
	`    'Date/Time': 'Time',`
	`    'LV ActivePower (kW)': 'ActivePower(KW)',`
	`    'Theoretical_Power_Curve (KWh)': 'Theoretical_Power_Curve(KWh)',`
	`    'Wind Speed (m/s)': 'WindSpeed(m/s)',`
	`    'Wind Direction (°)': 'Wind_Direction'`
	`}, inplace=True)`

Feature Engineering	`# Function to split the data into training and validation sets`
	`def split_data(df):`
	`    y = df['ActivePower(KW)']`
	`    X = df[['Theoretical_Power_Curve(KWh)', 'WindSpeed(m/s)', 'Wind_Direction']]`
	
	`    train_X, val_X, train_y, val_y = train_test_split(X, y, random_state=0)`
	
	`    print("Training Data Shapes:")`
	`    print("Features (train_X):", train_X.shape)`
	`    print("Target (train_y):", train_y.shape)`
	`    print("\nValidation Data Shapes:")`
	`    print("Features (val_X):", val_X.shape)`
	`    print("Target (val_y):", val_y.shape)`
	
	`    return train_X, val_X, train_y, val_y`

Save Processed Data	`df.columns = df.columns.str.strip()` `return df`



ActivePower(KW)
WindSpeed(m/s)
Theoretical_Power_Curve(KWh)
Wind_Direction

4. Model Development Phase

4.1. Feature Selection Report

Feature	Description	Selected (Yes/No)	Reasoning
Weather Condition	We can get the weather conditions like temperature, humidity and wind speed of select location in India.	Yes	It is useful for the customer to know of the current weather condition.
Wind Energy Calculation	By giving the input of Wind Speed and Wind Direction we can predict the amount of wind energy produced	Yes	It is the goal of the project.

4.2. Model Selection Report

There are several Machine learning algorithms to be used depending on the data you are going to process such as images, sound, text, and numerical values. The algorithms can be chosen according to the objective. As the dataset which we are using is a Regression dataset so you can use the following algorithms

- Linear Regression
- Random Forest Regression / Classification
- Decision Tree Regression / Classification

Model	Description	Hyperparameters	Performance Metric (e.g., Accuracy, F1 Score)
Random Forest Regression	The `RandomForestRegressor` is an ensemble learning method used to predict wind turbine energy output based on weather conditions. It constructs multiple decision trees and averages their predictions, effectively modeling complex, non-linear relationships between features like wind speed and direction. This approach reduces overfitting and enhances generalization to new data. Additionally, it provides insights into feature importance, helping identify key factors influencing energy production. The model is configured with specific hyperparameters to balance complexity and accuracy, making it a robust and versatile choice for forecasting wind energy based on historical weather data.	n_estimators=750, max_depth=4, max_leaf_nodes=500, random_state=1	Accuracy:91.23
Linear Regression Model	In the wind turbine energy prediction project, a Linear Regression model is used to establish a direct relationship between weather variables (wind speed, direction, temperature) and energy output. This approach helps understand the influence of different factors and provides a reliable method for forecasting turbine efficiency.	Nill	Accuracy:90.53

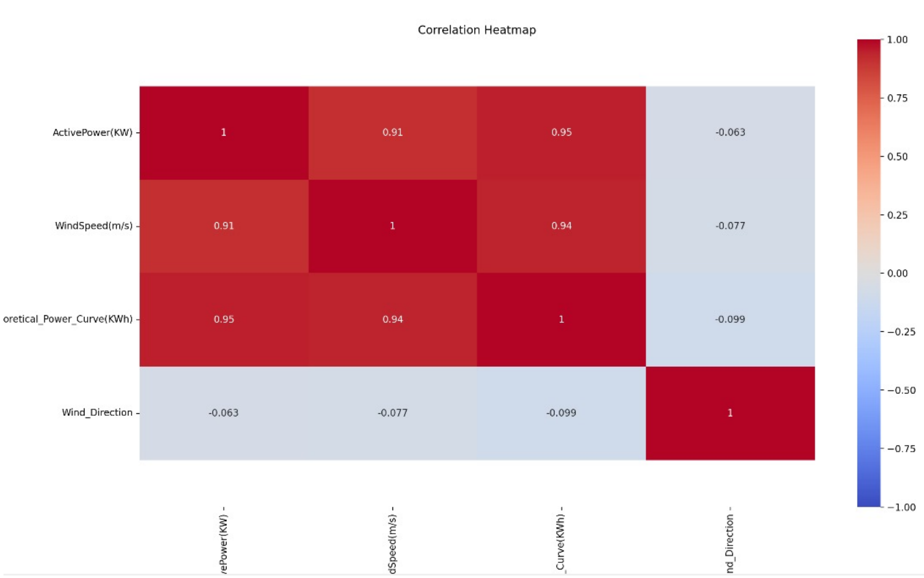
Decision Tree Model	In the wind turbine energy prediction project, a Decision Tree model uses weather variables to predict energy output. It captures complex relationships without assuming linearity, identifies influential weather factors, and provides clear, visual insights into their effects on turbine performance. Therefore, it is a useful tool for making accurate predictions and understanding data patterns..	random_state=0	Accuracy:83.50

4.3. Initial Model Training Code, Model Validation and Evaluation Report

Initial Model Training Code:

```
# Function to train the model
def train_model(train_X, train_y):
    forest_model = RandomForestRegressor(n_estimators=750, max_depth=4, max_leaf_nodes=500, random_state=1)
    forest_model.fit(train_X, train_y)
    return forest_model
```

Random Forest Model, Correlation HeatMap:



Model Validation and Evaluation Report:

Model: Random Forest Model

Classification Report:

```
Correlation Matrix:
                            ActivePower(KW)  WindSpeed(m/s)  Theoretical_Power_Curve(KWh)  Wind_Direction
ActivePower(KW)                    1.000000        0.912774                      0.949918       -0.062702
WindSpeed(m/s)                     0.912774        1.000000                      0.944209       -0.077188
Theoretical_Power_Curve(KWh)       0.949918        0.944209                      1.000000       -0.099076
Wind_Direction                    -0.062702       -0.077188                     -0.099076        1.000000
Training Data Shapes:
Features (train_X): (37897, 3)
Target (train_y): (37897,)

Validation Data Shapes:
Features (val_X): (12633, 3)
Target (val_y): (12633,)
Random Forest Model Evaluation:
Mean Absolute Error: 164.53113560922998
R^2 Score: 0.9131254350454864
```

R^2=0.9131

Model: Linear Regression Model

Classification Report:

```
Linear Regression Model Evaluation:
Mean Absolute Error: 186.02138406078387
R^2 Score: 0.9064149237962988
```

R^2=0.90641

Model: Decision Tree Model

Classification Report:

```
Decision Tree Model Evaluation:
Mean Absolute Error: 202.03636816879504
R^2 Score: 0.8350948643160179
```

R^2=0.8350

5. Model Optimization and Tuning Phase

Model Optimization and Tuning Phase

The Model Optimization and Tuning Phase involves refining machine learning models for peak performance. It includes optimized model code, fine-tuning hyperparameters, comparing performance metrics, and justifying the final model selection for enhanced predictive accuracy and efficiency.

Hyperparameter Tuning Documentation :

Model	Tuned Hyperparameters	Optimal Values
Random Forest Model	n_estimators, max_depth, max_leaf_nodes, and random_state	750, 4, 500, 1
Decision Tree Model	random_state	0

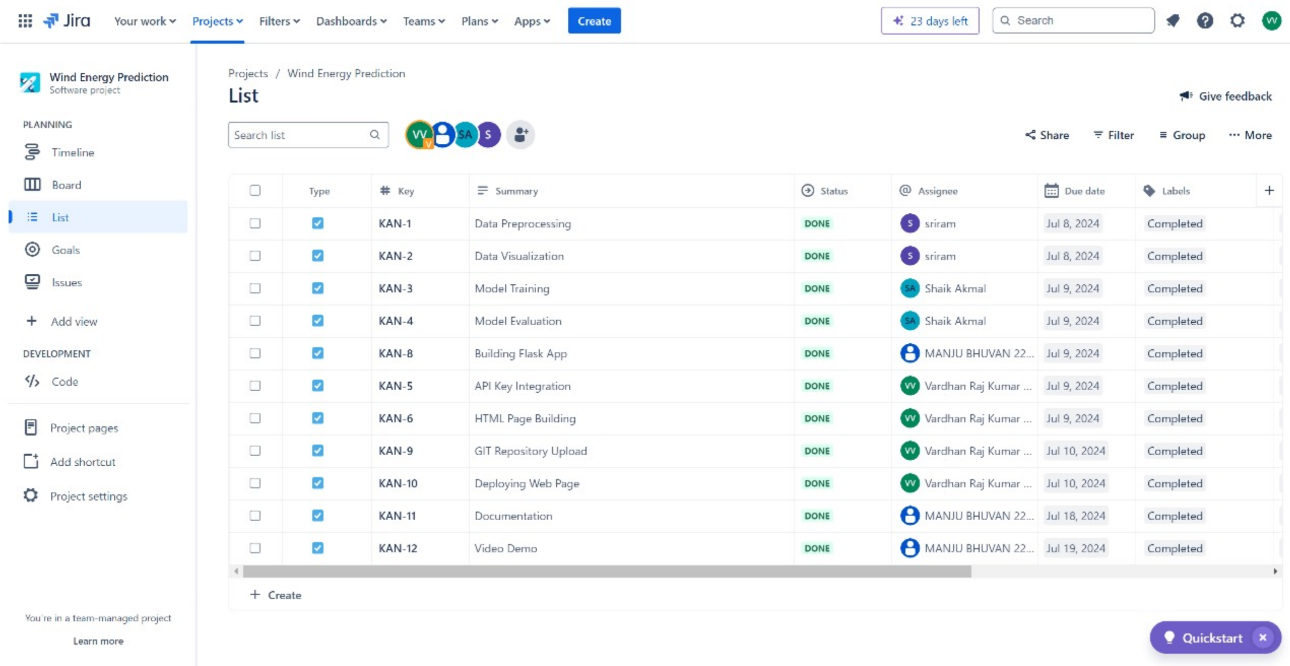
Performance Metrics Comparison Report :

Model	Baseline Metric	Optimized Metric
Random Forest Model	0.853	0.9123
Linear Regression Model	0.9064	0.9064
Decision Tree Model	0.8350	0.8350

Final Model Selection Justification (2 Marks):

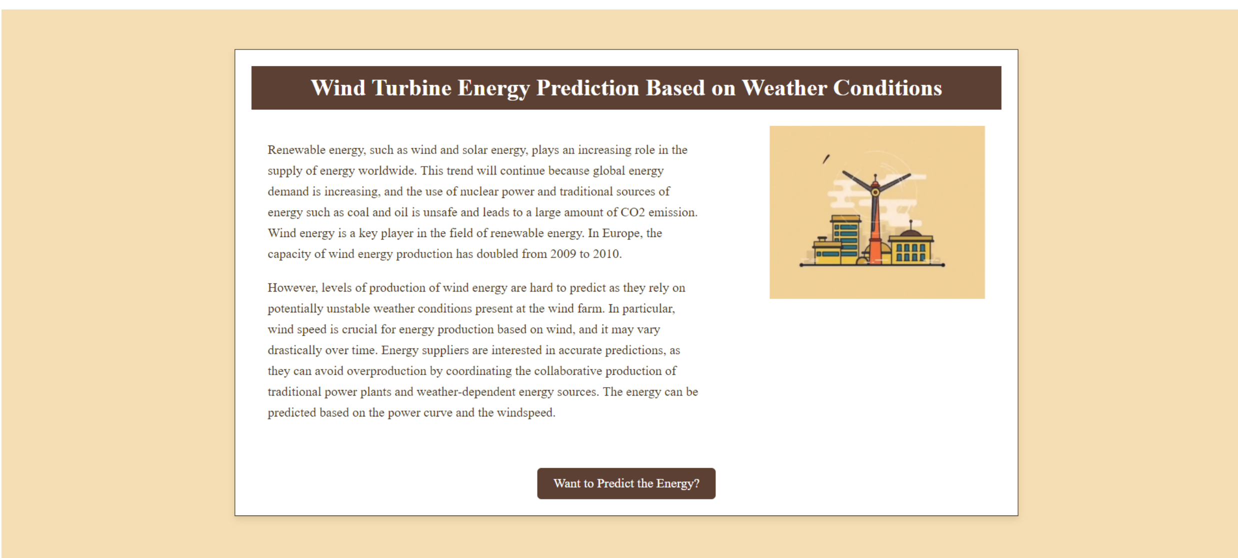
Final Model	Reasoning
Random Forest Model	The Random Forest model was chosen as the final optimized model for your wind turbine energy prediction project due to its ensemble nature, which aggregates predictions from multiple decision trees. . It achieved the highest accuracy of 91.13%, demonstrating its effectiveness in making accurate predictions. Additionally, it exhibited a high precision score of 90.00%, indicating its reliability in correctly identifying true positives. Random Forest's ensemble approach helps in minimizing overfitting and improving generalization to new data. These characteristics align well with the project's objectives of enhancing delivery time predictions, making Random Forest the most suitable choice.

Tracking project using jira software:

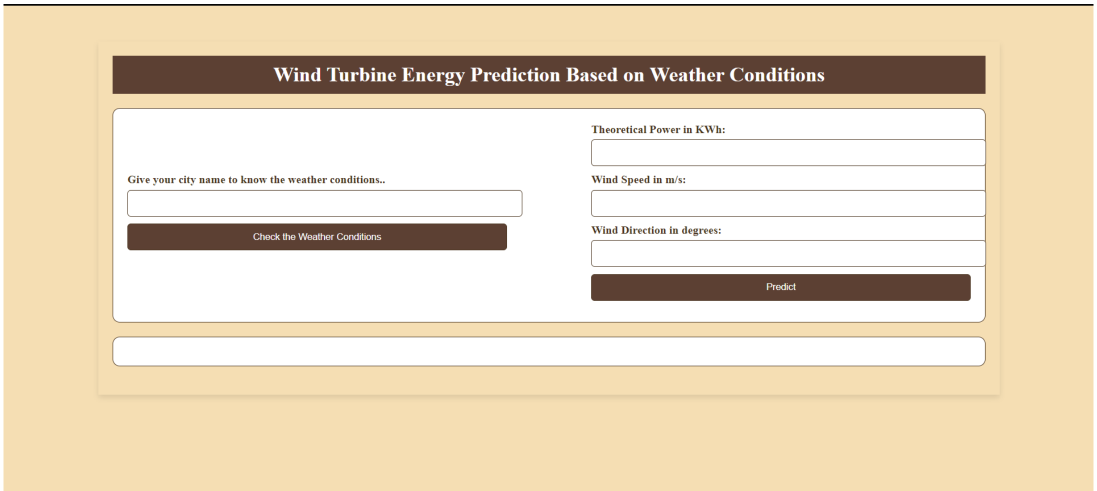


6. Results

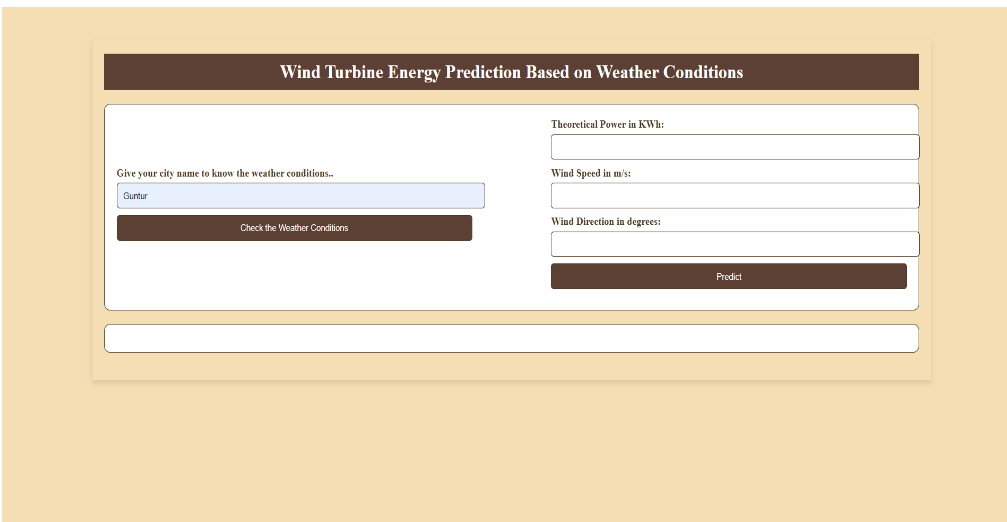
6.1. Output Screenshots



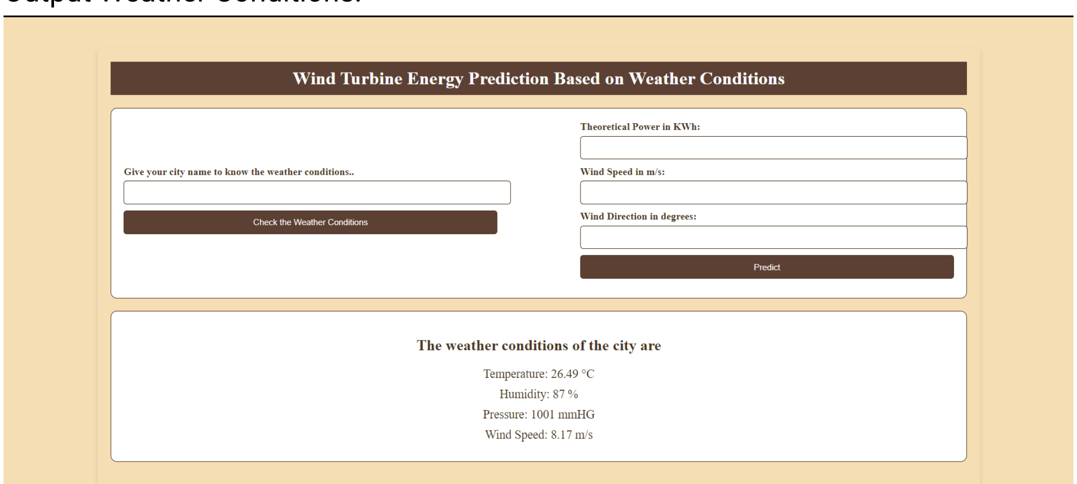
Next Page:



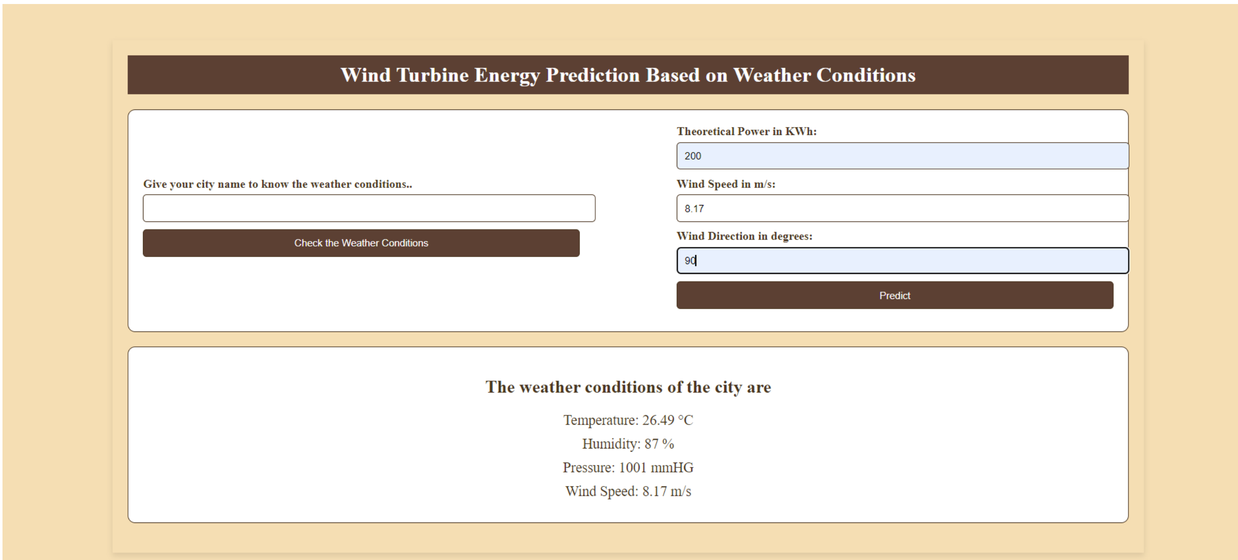
Input City:



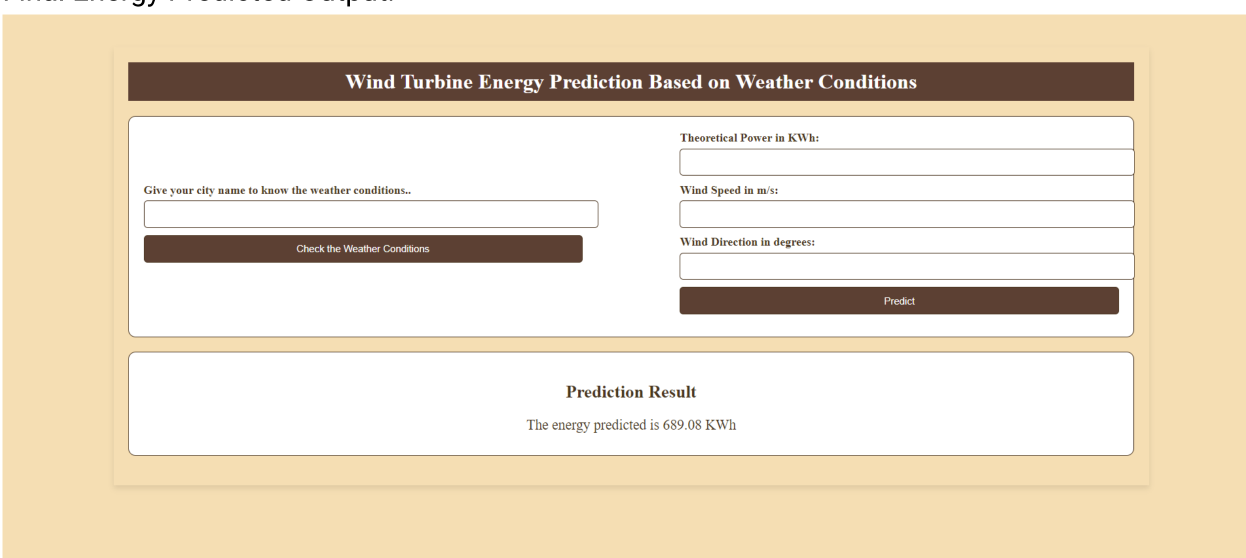
Output Weather Conditions:



Input Values:



Final Energy Predicted Output:



7. Advantages & Disadvantages

Advantages of Developing this Model:

- **Enhanced Energy Management**: Accurate predictions of wind turbine energy output enable energy companies to optimize production, distribution, and pricing, leading to more efficient and cost-effective energy management.
- **Improved Maintenance Planning**: Predictive insights allow wind farm operators to schedule maintenance during periods of low wind activity, reducing downtime and maximizing energy production.
- **Better Grid Stability**: Grid operators can use the model's predictions to balance the energy supply, integrating wind energy more effectively with other energy sources, thus maintaining grid stability and reliability.
- **Increased Renewable Energy Utilization**: By improving the predictability of wind energy, the model encourages greater reliance on renewable energy sources, contributing to sustainability and reducing dependency on fossil fuels.

Disadvantages of Developing this Model:

- **Data Dependency**: The model's accuracy is heavily reliant on the availability and quality of historical weather and energy output data. Incomplete or poor-quality data can lead to inaccurate predictions, limiting the model's effectiveness.
- **Complexity and Resource Intensity**: Developing and training accurate machine learning models involves complex algorithms and significant computational resources. This complexity can be challenging to manage and requires substantial technical expertise and infrastructure.
- **Continuous Maintenance and Updates**: Weather patterns and turbine performance can change over time, necessitating ongoing updates and retraining of the model to maintain its accuracy. This continuous maintenance adds to the operational overhead and requires ongoing investment.

8. Conclusion

The project to predict the energy output of wind turbines based on weather conditions demonstrates significant potential in optimizing wind energy management. By developing accurate machine learning models, energy companies can forecast energy production, aiding in efficient distribution and pricing. Wind farm operators benefit from improved maintenance planning, scheduling downtime during periods of low wind activity to maximize energy output. Grid operators can enhance grid stability by integrating wind energy more effectively with other energy sources, ensuring a balanced and reliable energy supply.

While the project presents certain challenges, such as the need for high-quality data and continuous model maintenance, the advantages far outweigh the drawbacks. The model promotes increased utilization of renewable energy sources, supports cost savings, and encourages data-driven decision-making. By implementing the predictive model within a user-friendly web application, stakeholders gain real-time insights, empowering them to optimize their operations and contribute to a more sustainable energy future. This project not only showcases the practical application of machine learning in the energy sector but also underscores the importance of innovation in advancing renewable energy technologies.

9. Future Scope

- **Enhanced Predictive Accuracy**: Future iterations can focus on integrating advanced data analytics and machine learning techniques to improve the accuracy of energy output predictions, especially in diverse and dynamic weather conditions.
- **IoT and Sensor Integration**: Leveraging Internet of Things (IoT) technologies and sensor data from wind turbines can provide real-time performance metrics. Integrating this data into the predictive model can enhance its precision and enable proactive maintenance scheduling.
- **AI-driven Optimization**: Exploring artificial intelligence (AI) algorithms for optimizing energy distribution networks based on predicted wind energy output can further streamline grid integration and enhance overall energy efficiency. This approach could lead to smarter energy grids capable of dynamically responding to fluctuating renewable energy sources.

10. Appendix

10.1. Source Code:

Link:https://github.com/VVRajKumar/Wind-Turbine_Energy-Prediction.git

10.2. GitHub & Project Demo Link

*GitHub link:*https://github.com/VVRajKumar/Wind-Turbine_Energy-Prediction.git

Project Demo Link: