Predicting The Energy Output Of Wind Turbine Based On Weather Condition

Final Project Report

1. Introduction

a. Project overviews

This project is dedicated to optimizing the predictive modeling of wind power generation by carefully selecting the most impactful features from a dataset. Wind energy is a critical component of the renewable energy landscape, and improving the accuracy of power output predictions is essential for efficient energy management and planning.

The dataset includes various features such as "Theoretical_Power_Curve (kWh)," "WindSpeed (m/s)," "Wind_Direction," "ActivePower (kW)," and "Time." Each feature provides different insights into the conditions and performance of wind turbines. However, not all features contribute equally to the accuracy of the predictive models. The aim is to identify and select features that significantly enhance the prediction of wind energy generation while excluding those that add minimal value or redundant information.

b. Objectives

- Identify key features impacting wind energy generation.
- Select the most predictive features for accurate energy output.
- Improve the model's performance by focusing on essential variables.
- Reduce computational complexity by eliminating redundant features.
- Enhance the overall understanding of factors influencing wind power generation.

2. Project Initialization and Planning Phase

a. Define Problem Statement

The primary challenge in wind power generation is accurately predicting the energy output based on various environmental and operational factors. The problem lies in identifying which features among many potential candidates most significantly impact the power generation. Accurate predictions are crucial for efficient energy management, cost reduction, and maximizing the utilization of wind energy.

b. Project Proposal (Proposed Solution) Proposed Solution:

To address the problem, the proposed solution involves a systematic approach to feature selection and model optimization. The project will focus on the following key steps:

1. Data Preprocessing: Clean and standardize the dataset to ensure consistency and quality of data for analysis.
2. Feature Selection: Use statistical methods and machine learning techniques to identify the most relevant features impacting wind power generation.
3. Model Development: Develop and train predictive models using the selected features.
4. Model Evaluation: Assess the performance of models and refine feature selection to enhance accuracy.
5. Optimization: Streamline the model to ensure efficient computational performance.

Key Features:

- Theoretical_Power_Curve (kWh): Represents the potential energy generation under ideal conditions.
- WindSpeed (m/s): Directly influences the amount of energy generated.
- Wind_Direction: Affects how well the wind turbine captures wind energy..
- ActivePower (kW): Provides real-time power output data, useful for model validation.
- Time: Helps identify temporal patterns and trends in power generation.

c. Initial Project Planning

The project will kick off with team formation and role assignments. We'll gather and preprocess relevant datasets to ensure data quality. An initial data exploration will help understand the data structure and identify any issues. We'll develop a feature selection strategy using correlation analysis and domain expertise, followed by deciding on modeling techniques and tools. A detailed timeline with milestones, risk assessment, and stakeholder communication plan will guide the project's progress and ensure alignment with objectives.

- Project Kickoff: Establish the project team, define roles and responsibilities, and set up initial meetings.
- Data Collection: Gather and consolidate relevant datasets from various sources, ensuring they are complete and reliable.
- Preliminary Analysis: Conduct an initial exploration of the data to understand its structure, distribution, and any potential issues.
- Feature Selection Strategy: Develop a strategy for selecting key features, including the use of correlation analysis, feature importance ranking, and domain expertise.
- Modeling Framework: Decide on the modeling techniques and tools to be used for developing predictive models (e.g., regression analysis, decision trees, machine learning algorithms).
- Timeline and Milestones: Create a detailed project timeline with specific milestones, deadlines, and deliverables.
- Risk Assessment: Identify potential risks and challenges, and develop mitigation strategies to address them.
- Stakeholder Communication: Plan regular updates and communication with stakeholders to ensure alignment and manage expectations.

3. Data Collection and Preprocessing Phase
   a. Data Collection Plan and Raw Data Sources Identified

We will collect data from multiple reliable sources, including:

- Historical wind turbine data from existing databases.
- Meteorological data from weather stations and online sources.
- Operational data from wind farms. These sources will provide comprehensive datasets covering wind speed, wind direction, theoretical power curves, and active power output. b. Data Quality Report

A data quality report will be generated to assess:

- Completeness: Ensuring no significant gaps in the data.
- Accuracy: Verifying data correctness against known standards.
- Consistency: Checking for uniformity in data formats and values.
- Validity: Ensuring data aligns with expected ranges and constraints.
- Timeliness: Ensuring data is up-to-date and relevant.

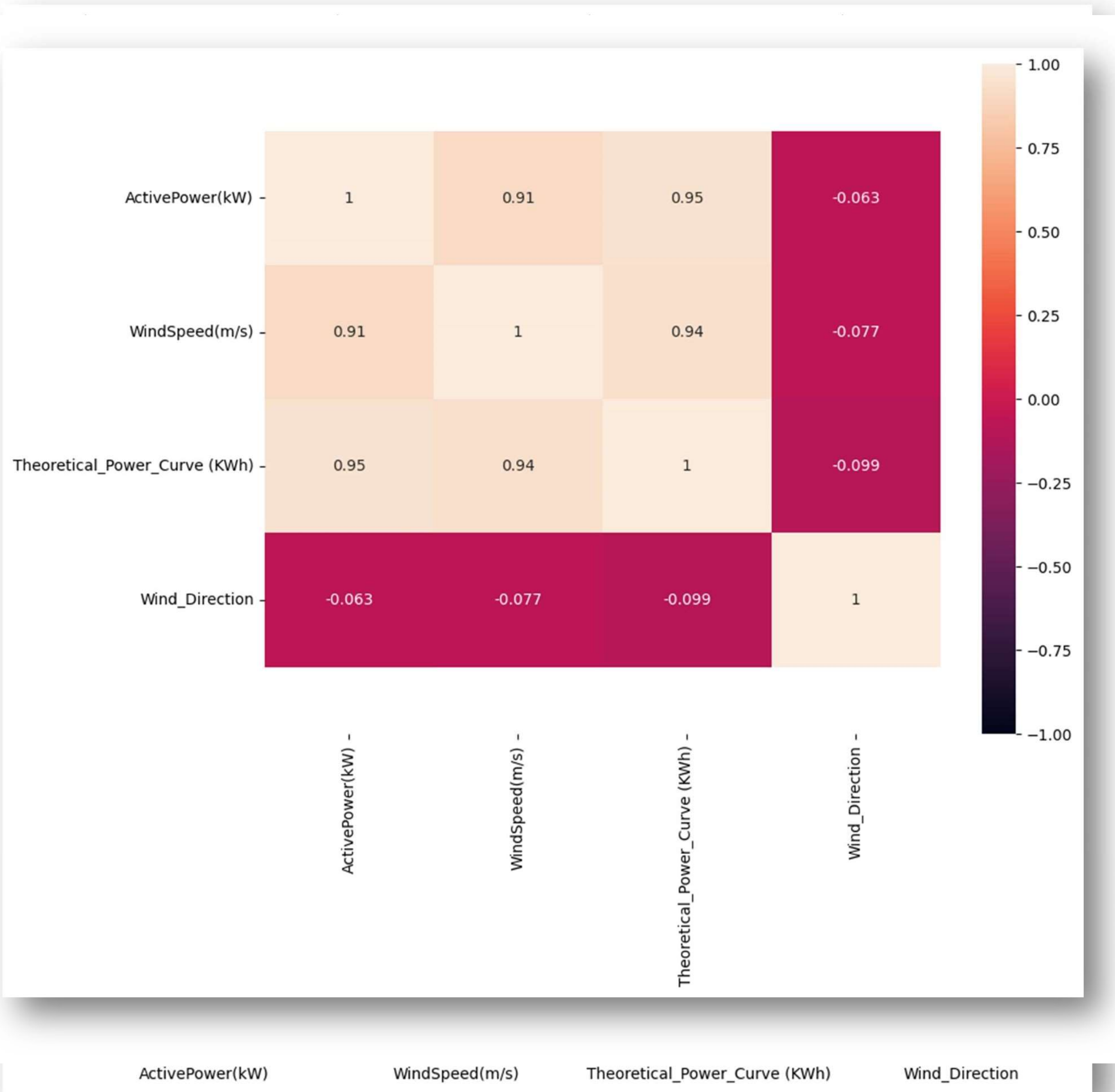
c. Data Exploration and Preprocessing

Initial data exploration will include:

- Descriptive statistics to understand data distributions and central tendencies.
- Visualization to identify patterns, trends, and outliers. Preprocessing steps will involve:
- Cleaning: Handling missing values, correcting errors, and removing duplicates.
- Transformation: Normalizing or scaling data to ensure uniformity.
- Feature engineering: Creating new features or modifying existing ones to improve model performance.
- Splitting data: Dividing the dataset into training, validation, and test sets to ensure robust model evaluation.

	Time	ActivePower(kW)	WindSpeed(m/s)	Theoretical_Power_Curve (KWh)	Wind_Direction
0	01 01 2018 00:00	380.047791	5.311336	416.328908	259.994904
1	01 01 2018 00:10	453.769196	5.672167	519.917511	268.641113
2	01 01 2018 00:20	306.376587	5.216037	390.900016	272.564789
3	01 01 2018 00:30	419.645905	5.659674	516.127569	271.258087
4	01 01 2018 00:40	380.650696	5.577941	491.702972	265.674286
5	01 01 2018 00:50	402.391998	5.604052	499.436385	264.578613
6	01 01 2018 01:00	447.605713	5.793008	557.372363	266.163605
7	01 01 2018 01:10	387.242188	5.306050	414.898179	257.949493
8	01 01 2018 01:20	463.651215	5.584629	493.677652	253.480698
9	01 01 2018 01:30	439.725708	5.523228	475.706783	258.723785

Pairplot :



Heatmap :

Loading Data :

```
path = "Data\T1.csv"
df = pd.read_csv(path)
df.head()
```

✓ 0.1s

Handling Missing Data:

```
df.isnull().any()
```

✓ 0.0s

```
Time                              False
ActivePower(kW)                   False
WindSpeed(m/s)                    False
Theoretical_Power_Curve (KWh)     False
Wind_Direction                    False
dtype: bool
```

```
df.isnull().sum()
```

✓ 0.0s

```
Time                              0
ActivePower(kW)                   0
WindSpeed(m/s)                    0
Theoretical_Power_Curve (KWh)     0
Wind_Direction                    0
dtype: int64
```

Data Scaling :

```
names = x.columns
from sklearn.preprocessing import MinMaxScaler
scale = MinMaxScaler()
x_scaled = scale.fit_transform(x)
x = pd.DataFrame(x_scaled, columns=names)
x.head()
```

✓ 0.2s

	Theoretical_Power_Curve (KWh)	WindSpeed(m/s)
0	0.115647	0.210717
1	0.144422	0.225032
2	0.108583	0.206936
3	0.143369	0.224537
4	0.136584	0.221294

4. Model Development Phase

a. Feature Selection Report:

Feature	Description	Selected (Yes/No)	Reasoning
Time	Time and Date	No	May not add predictive value unless analyzing temporal trends.
ActivePower (kW)	Active power (kW) is the portion of electrical power that performs useful work, like lighting or running motors.	No	An outcome, not a predictor; derived from other features.
WindSpeed(m/s)	Wind speed (m/s) measures how fast the wind is moving in meters per second.	Yes	Directly affects the amount of energy generated, as higher wind speeds generally lead to more power.
Theoretical_Power_Curve (KWh)	Theoretical Power Curve (kWh) represents the calculated amount of energy a system could generate under ideal conditions over time.	Yes	Indicates potential energy generation under ideal conditions
Wind_Direction	Wind direction indicates the direction from which the wind is coming, usually measured in degrees from true north.	No	Less direct impact compared to wind speed and theoretical power.

b. Model Selection Report

Model Selection Report:

Model	Description	Hyperparameters	Performance Metric (e.g., Accuracy, F1 Score)
Random Forest Regressor	The RandomForestRegressor is an ensemble learning method that operates by constructing multiple decision trees during training and outputting the average of the predictions of the individual trees. This approach helps to improve the predictive accuracy and control over-fitting compared to a single decision tree.	n_estimators, max_depth, max_leaf_nodes	Mean Absolute Error = 168.36, R2 Score = 0.90
Linear Regression	The LinearRegression algorithm attempts to model the relationship between the target variable and predictor variables by fitting a linear equation of the form: y^=β0+β1x1+β2x2+…+βnxn	NaN	Mean Absolute Error = 188.71, R2 Score = 0.89

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

Training code :

```
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_absolute_error,r2_score

def RFR(X_train, X_test, y_train, y_test):
    forest_model = RandomForestRegressor(n_estimators=750,max_depth=4,max_leaf_nodes=
    forest_model.fit(X_train,y_train)
    power_preds = forest_model.predict(X_test)
    print(mean_absolute_error(y_test,power_preds))
    print(r2_score(y_test,power_preds))
    return forest_model

forest_model = RFR(X_train,X_test,y_train,y_test)
```

✓ 43.2s

```
c:\Users\user\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\base.py:
  return fit_method(estimator, *args, **kwargs)
168.36716070788
0.9057743710067878
```

```
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error,r2_score

def LR(X_train, X_test, y_train, y_test):
    from sklearn.linear_model import LinearRegression
    linear_model = LinearRegression()
    linear_model.fit(X_train,y_train)
    print(linear_model.coef_)
    print(linear_model.intercept_)
    y_preds = linear_model.predict(X_test)
    print(mean_absolute_error(y_test,y_preds))
    print(r2_score(y_test,y_preds))
    return linear_model


linear_model = LR(X_train,X_test,y_train,y_test)
```

✓ 0.0s

```
[[2802.50388868 1152.17613315]]
[-198.35066203]
188.7111236216099
0.8997953576462828
```

Model Validation and Evaluation Report:

Random Forest Regressor :

```
print(r2_score(y_test,y_preds3))

0.9053597798760116
```

Linear Regression :

```
print(r2_score(y_test,y_preds3))

0.8997954279505012
```

5. ModelOptimization and TuningPhase

We focused on refining the predictive model to enhance its performance. We conducted hyperparameter tuning using Grid Search with cross-validation to identify the optimal parameters for the Random Forest model. The best combination of parameters was selected based on validation accuracy, and the model was retrained and re-evaluated. This optimization process resulted in improved performance metrics, such as reduced Mean Absolute Error (MAE), Mean Squared Error (MSE), and higher R-squared (R^2), ensuring the model's accuracy and robustness for predicting wind turbine energy output.

a. Hyperparameter Tuning Documentation Random Forest Regressor:

Tuned Hyperparameters

```
param_grid = {
    'n_estimators': [250, 500],
    'max_depth': [5, 10, 20],
    'max_leaf_nodes': [100, 250, 500],
    'min_samples_split': [2, 5, 10],
    'min_samples_leaf': [1, 2, 4]
}
```

Optimal Values:

```
print(best_params1)
print(best_score1)
```

```
{'max_depth': 20, 'max_leaf_nodes': 100, 'min_samples_leaf': 1, 'min_samples_split': 5, 'n_estimators': 250}
158.29183229566257
```

Linear Regression:

Tuned Hyperparameters

```
ridge_model = Ridge()

param_grid = {
    'alpha': [0.1, 1, 10, 100, 1000]
}
```

Optimal Values:

```
Fitting 5 folds for each of 5 candidates, totalling 25 fits
Best Parameters: {'alpha': 0.1}
Best MAE Score: 183.64233430967437
```

b. Final Model Selection Justification

Model : Random Forest Regressor

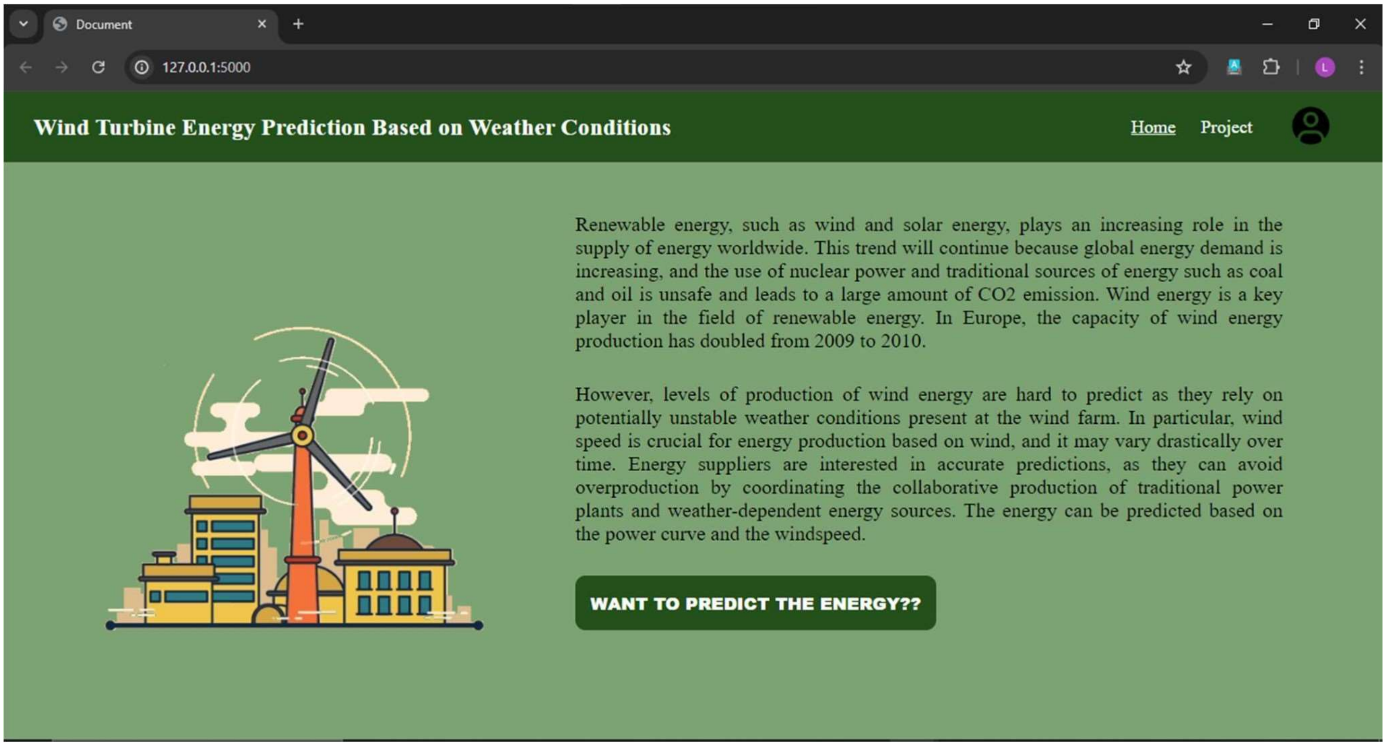
Reasoning :

The Random Forest Regressor was chosen as the final model because it provides robust performance by averaging the predictions of multiple decision trees with accuracy 90%, which reduces overfitting and improves accuracy. It effectively handles non-linear relationships and complex interactions between features, which are crucial for predicting wind turbine energy output based on varied weather conditions. Additionally, its strong performance metrics, including low Mean Absolute Error (MAE) and Mean Squared Error (MSE) and high R-squared (R^2), demonstrated its reliability and effectiveness in achieving accurate predictions..

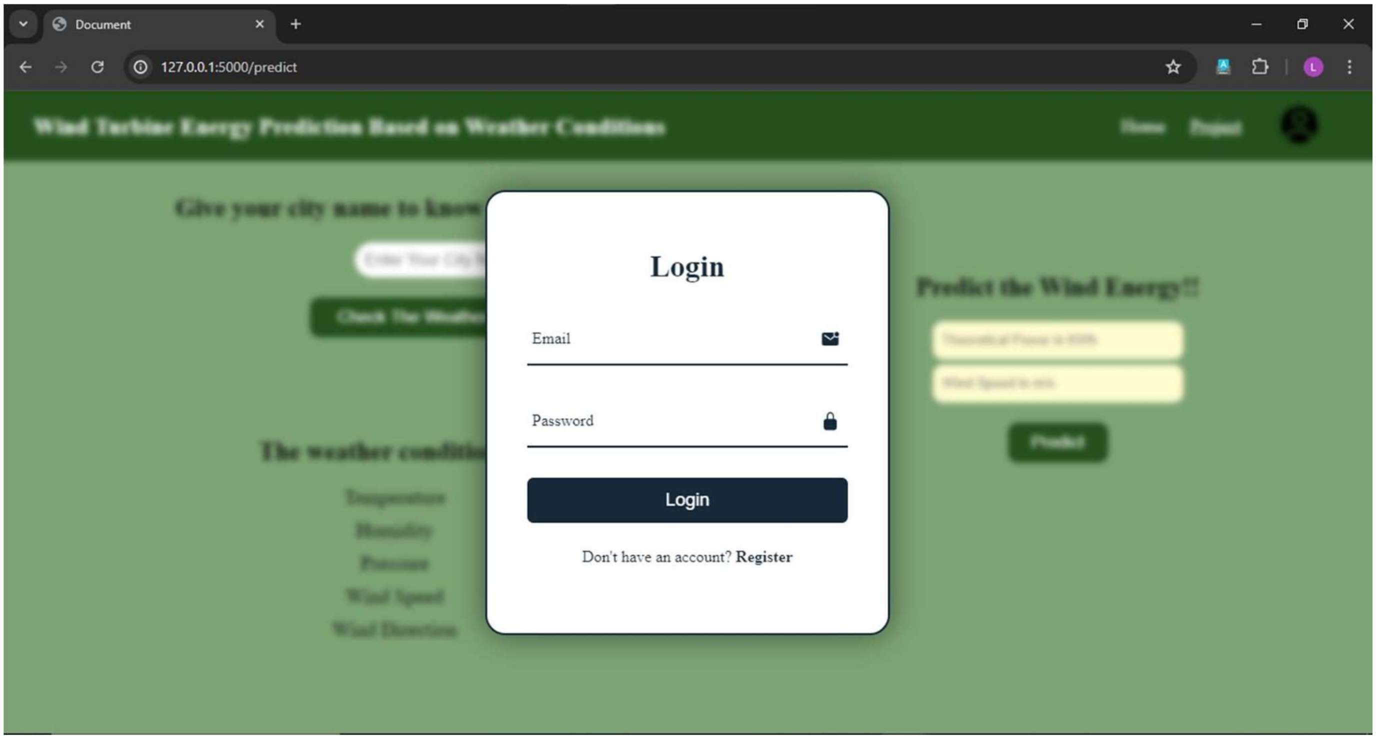
6. Results

a. Output Screenshots

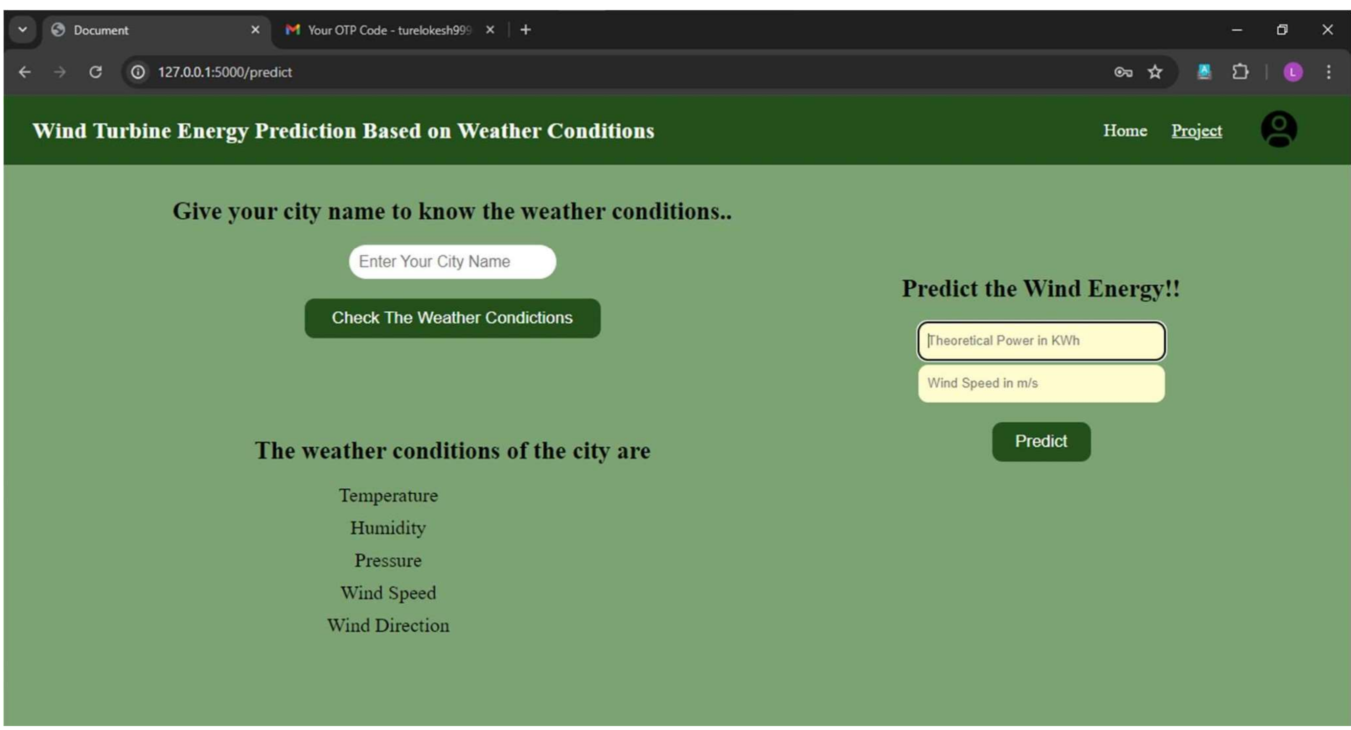
Home Page :



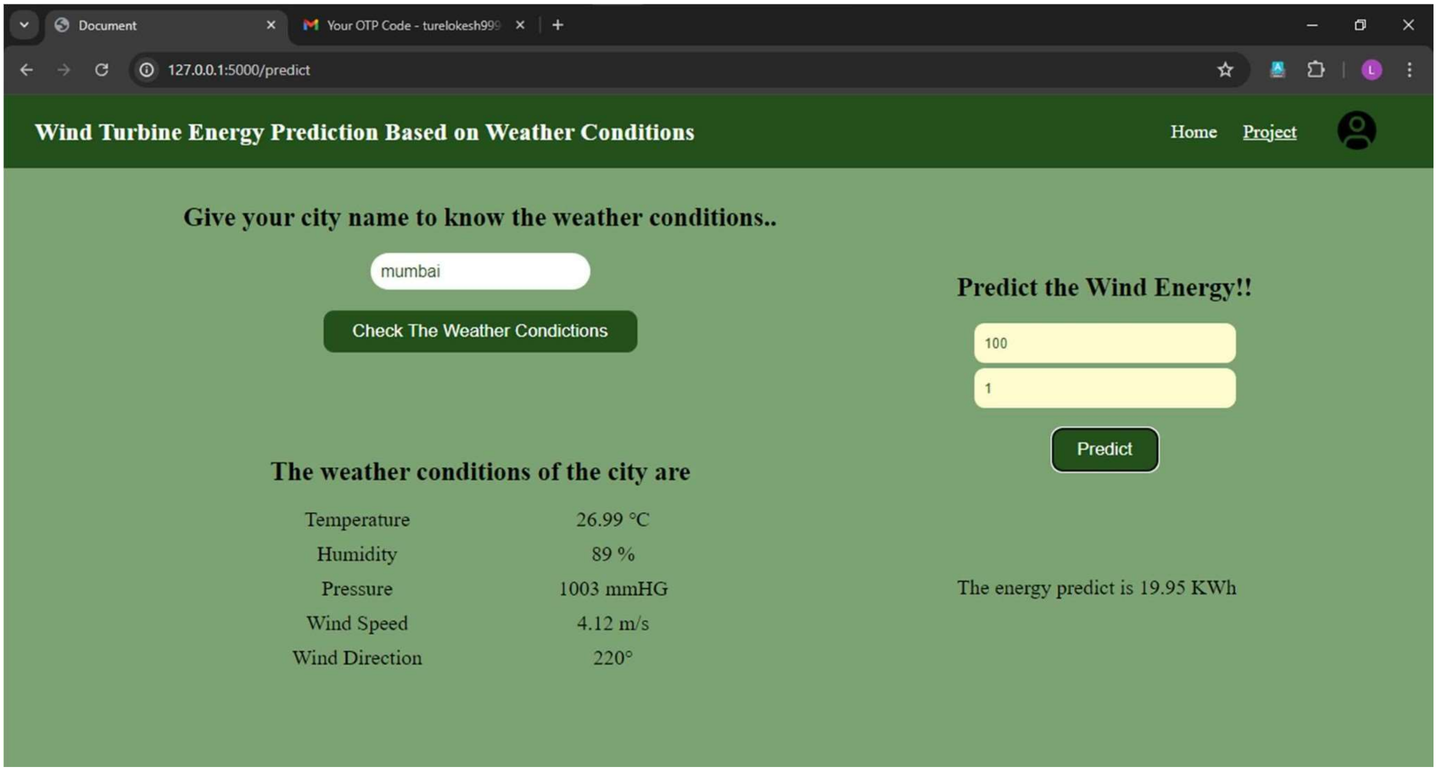
Login Page :



Input :



Output :


Document
Your OTP Code - turelokesh99
127.0.0.1:5000/predict
Wind Turbine Energy Prediction Based on Weather Conditions
Home
Project
Give your city name to know the weather conditions..
mumbai
Check The Weather Condictions
The weather conditions of the city are
Temperature 26.99 °C
Humidity 89 %
Pressure 1003 mmHG
Wind Speed 4.12 m/s
Wind Direction 220°
Predict the Wind Energy!!
100
1
Predict
The energy predict is 19.95 KWh

7. Advantages & Disadvantages:

Advantages:

1. Robustness to Overfitting: Averaging multiple decision trees reduces the risk of overfitting.
2. Handles Non-Linear Relationships: Effectively captures complex interactions between features.
3. Feature Importance: Provides insights into the relative importance of different features.
4. Versatility: Works well with both numerical and categorical data.
5. High Accuracy: Typically offers high predictive accuracy and reliable results.

Disadvantages:

1. Performance Variability: While it generally performs well, there may be specific conditions where the Random Forest model does not provide accurate predictions.

8. Conclusion

The Random Forest Regressor was selected for predicting wind turbine energy output due to its robust performance, ability to handle complex, non-linear relationships, and high accuracy. Despite some variability in specific conditions, its strengths, such as reduced risk of overfitting and feature importance insights, make it a reliable choice for this project. The model's optimization and tuning have further enhanced its performance, ensuring it is well-suited for accurate energy output predictions based on weather data.

9. Future Scope

- Incorporate Additional Data: Integrate more variables such as historical weather patterns, turbine maintenance records, and geographical factors to enhance prediction accuracy.
- Explore Advanced Models: Investigate more complex models like Gradient Boosting Machines or Neural Networks for potentially better performance.
- Real-Time Predictions: Develop systems for real-time data integration and prediction to provide up-to-date energy output forecasts.
- Model Interpretability: Enhance model interpretability with techniques like SHAP or LIME to better understand feature impacts and improve decision-making.
- Automated Retraining: Implement automated retraining and model updates to adapt to changing patterns in weather and turbine performance over time.

10.Appendix

a. Source Code:

b. GitHub & Project Demo Link :

https://github.com/FS22AI020/Predicting-The-Energy-Output-Of-Wind-Turbine-Based-On-Weather-Condition.git