

## Data Collection and Preprocessing Phase

Date	15 March 2024
Team ID	SWTID1720171463
Project Title	Predicting The Energy Output Of Wind Turbine Based On Weather Condition
Maximum Marks	6 Marks

## Data Exploration and Preprocessing Template

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	<div><div></div><table><thead><tr><th></th><th>ActivePower(kW)</th><th>WindSpeed(m/s)</th><th>Theoretical_Power_Curve (KWh)</th><th>Wind_Direction</th></tr></thead><tbody><tr><td>count</td><td>50530.000000</td><td>50530.000000</td><td>50530.000000</td><td>50530.000000</td></tr><tr><td>mean</td><td>1307.684332</td><td>7.557952</td><td>1492.175463</td><td>123.687559</td></tr><tr><td>std</td><td>1312.459242</td><td>4.227166</td><td>1368.018238</td><td>93.443736</td></tr><tr><td>min</td><td>-2.471405</td><td>0.000000</td><td>0.000000</td><td>0.000000</td></tr><tr><td>25%</td><td>50.677890</td><td>4.201395</td><td>161.328167</td><td>49.315437</td></tr><tr><td>50%</td><td>825.838074</td><td>7.104594</td><td>1063.776283</td><td>73.712978</td></tr><tr><td>75%</td><td>2482.507568</td><td>10.300020</td><td>2964.972462</td><td>201.696720</td></tr><tr><td>max</td><td>3618.732910</td><td>25.206011</td><td>3600.000000</td><td>359.997589</td></tr></tbody></table></div>		ActivePower(kW)	WindSpeed(m/s)	Theoretical_Power_Curve (KWh)	Wind_Direction	count	50530.000000	50530.000000	50530.000000	50530.000000	mean	1307.684332	7.557952	1492.175463	123.687559	std	1312.459242	4.227166	1368.018238	93.443736	min	-2.471405	0.000000	0.000000	0.000000	25%	50.677890	4.201395	161.328167	49.315437	50%	825.838074	7.104594	1063.776283	73.712978	75%	2482.507568	10.300020	2964.972462	201.696720	max	3618.732910	25.206011	3600.000000	359.997589
	ActivePower(kW)	WindSpeed(m/s)	Theoretical_Power_Curve (KWh)	Wind_Direction																																										
count	50530.000000	50530.000000	50530.000000	50530.000000																																										
mean	1307.684332	7.557952	1492.175463	123.687559																																										
std	1312.459242	4.227166	1368.018238	93.443736																																										
min	-2.471405	0.000000	0.000000	0.000000																																										
25%	50.677890	4.201395	161.328167	49.315437																																										
50%	825.838074	7.104594	1063.776283	73.712978																																										
75%	2482.507568	10.300020	2964.972462	201.696720																																										
max	3618.732910	25.206011	3600.000000	359.997589																																										
Overall Analysis	<div><div></div></div>																																													

## Data Preprocessing Code Screenshots

### Loading Data

```
path = r"T1.csv"

df = pd.read_csv(path)

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

### Data Transformation

```
y = df['ActivePower(kW)'] # 'Theoretical_Power_Curve (KWh)'
X = df[['Theoretical_Power_Curve (KWh)', 'WindSpeed(m/s)']] # 'ActivePower(kW)'

from sklearn.model_selection import train_test_split
train_X, val_X, train_y, val_y = train_test_split(X, y, random_state = 0)

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

forest_model = RandomForestRegressor(max_leaf_nodes = 500, random_state = 1)
forest_model.fit(train_X, train_y)
```

RandomForestRegressor

RandomForestRegressor(max\_leaf\_nodes=500, random\_state=1)

### Feature Engineering

Attached the codes in final submission.

### Save Processed Data

```
power_preds = forest_model.predict(val_X)
print(mean_absolute_error(val_y, power_preds))
print(r2_score(val_y, power_preds))
joblib.dump(forest_model, "power_prediction.sav")
```

```
162.90876721041636
0.9015207981707474

['power_prediction.sav']
```