Predicting the energy output of wind turbine based on weather condition

By Amratya Singh

Table of Contents

1.Introduction

1.1 Overview

Renewable energy such as wind and solar energy plays an increasing role in the supply of energy world-wide. This trend will continue because the global energy demand is increasing and the use of nuclear power and traditional sources of energy such as coal and oil is either considered as non-safe or leads to a large amount of CO_2 emission.Wind energy is a key-player in the field of renewable energy. The capacity of wind energy production was increased drastically during the last years. However, the production of wind energy is hard to predict as it relies on the rather unstable weather conditions present at the wind farm.In particular, the wind speed is crucial for energy production based on wind and the wind speed may vary drastically during different periods of time. Energy suppliers are interested in accurate predictions, as they can avoid overproductions by coordinating the collaborative production of traditional power plants and weather dependent energy sources.

1.2 Purpose

In recent years, the ML methods have become popular as they allow researchers to improve the prediction accuracy and are used for various engineering applications.
The ML methods have been used to increase the prediction accuracy of Energy Output, and the data derived from the literature sources were used Regression models tend to be used for the prediction of the Energy Output of wind turbine based on the various weather conditions. These models also demonstrate how they are related.. In this study, the ML regression methods were compared to predict the Energy Output. The study aimed to determine the most successful regression method by comparing the random forest and Linear Regression.

2.Literature Survey

2.1 Existing Problem

There exist a number of technological, environmental and political challenges linked to supplementing existing electricity generation capacities with wind energy. Here, mathematicians and statisticians could make a substantial contribution at the interface of meteorology and decision-making, in connection with the generation of forecasts tailored to the various operational decision problems involved. Indeed, while wind energy may be seen as an environmentally friendly source of energy, full benefits from its usage can only be obtained if one is able to accommodate its variability and limited predictability.

2.2 Proposed Solution

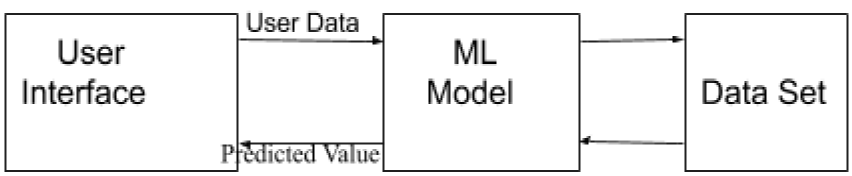
The solution was made possible as we had a lot of data , having data can make a huge difference . Using the dataset we can find out the correlation among all the values in the dataset and thus use those approximations and find out the resultant prediction which determines the energy output of the wind turbine . Thus, using this people can obtain all benefits from the wind energy which will help them to complete their tasks more efficiently.

3.Theoretical Analysis

3.1 Block Diagram

Three main components of our project are :-

1. The User Interface - A flask web application
2. ML model
3. Data Set



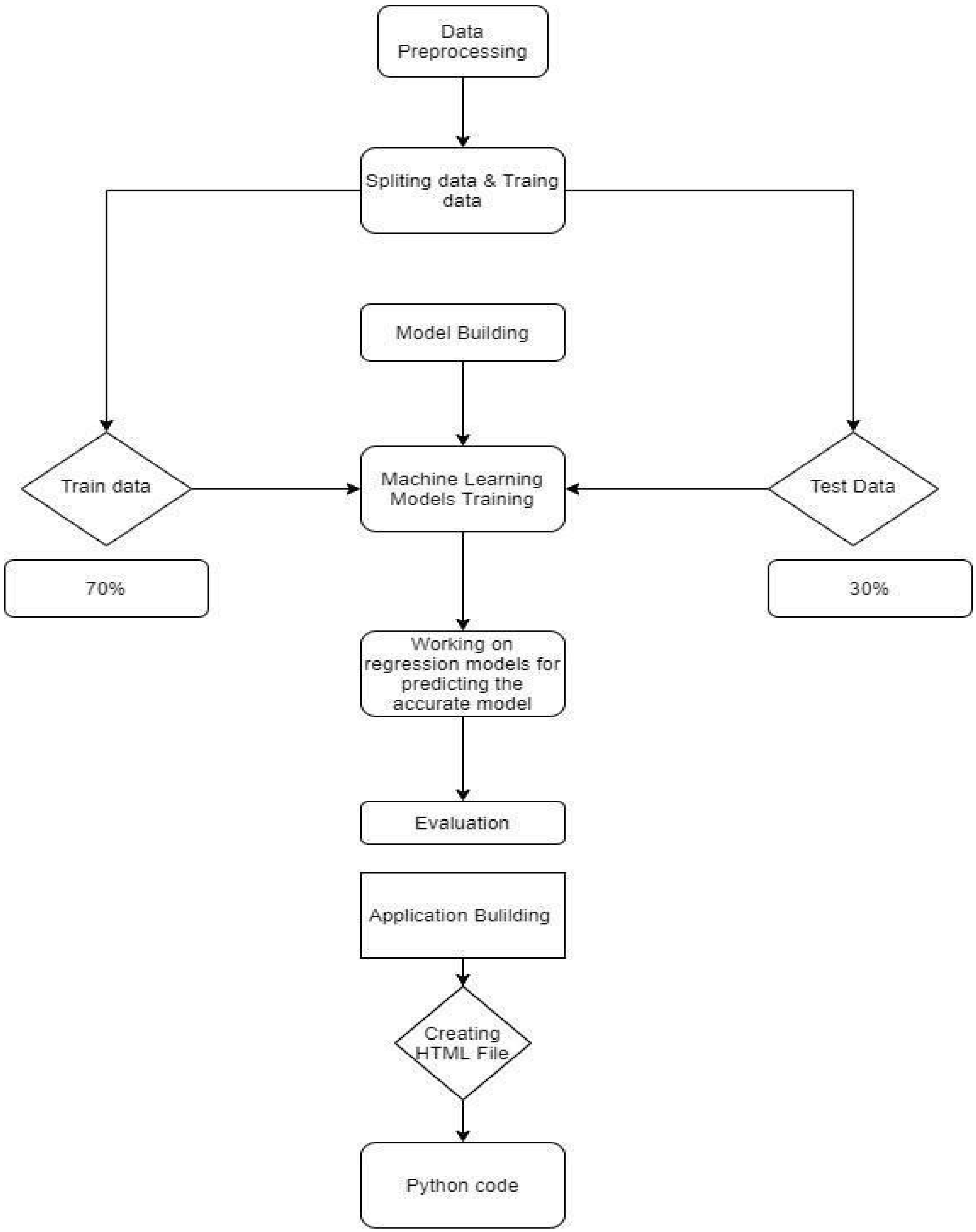
3.1 Hardware / Software designing

Python, Python Web Frameworks, Python for Data Analysis, Python For Data Visualization, Data Pre-processing Techniques, Machine Learning, Regression Algorithms

4.Experimental Analysis

The theoretical Power output for the present work was obtained from the experiments. For generating a reliable data bank on Power Output, we had considered three parameters active power, wind speed and wind direction.

5.Flowchart



Data Preprocessing
Spliting data & Traing data
Model Building
Train data
Machine Learning Models Training
Test Data
70%
30%
Working on regression models for predicting the accurate model
Evaluation
Application Bulilding
Creating HTML File
Python code

6.Result

We have analysed the various weather conditions and used Machine Learning to Predict the Energy Output. We have used Linear Regression and its variations, Lasso, Ridge and Random Forests to make predictions and compared their performance. Random Forest Regressor has the highest accuracy and is a good choice for this problem. Random Forest Regressor trains randomly initialized trees with random subsets of data sampled from the training data, this will make our model more robust. Our application is 90% efficient in predicting the energy output.

7. Advantages and Disadvantages

Advantages:

Using Machine learning to predict the Energy Output will be time efficient and will give more accuracy in predicting the approximately close value can be done easily. It's more trustworthy and cost effective .It also helps various organizations and people to utilize the wind power more efficiently and according schedule their works.

Disadvantages :

There is a 10% chance that the outcome will not predict the approximate value in that situation it can be troublesome.

8.Applications:

- Can predict the Energy output using the inputs provided.
- Implementable on the website.

9.Conclusion

- ➢ Compared to all other Machine Learning Models Random Forest was best suitable for this data.
- ➢ Random Forest Regressor gave the maximum accuracy when tested using r2_score confusion matrix.
- ➢ Maximum accuracy received is 90%.

10. Future Scope

This model can predict the outcome with many different inputs within seconds.Currently we are predicting energy output for the inputs given by the user based on the wind speed and wind direction , In future we can integrate it with the live data i.e current weather condition and can even more weather conditions like humidity, rain etc. to predict even more accurate.

11.Bibilograpghy

1. Flask https://pypi.org/project/Flask/
2. Virtual environments in python Pipenv & Virtual Environments
3. https://www.kaggle.com/berkerisen/wind-turbine-scada-dataset
4. https://pandas.pydata.org/
5. https://numpy.org/
6. https://seaborn.pydata.org/
7. https://towardsdatascience.com/understanding-random-forest-58381e0602d2

12. Appendix

A. Source Code

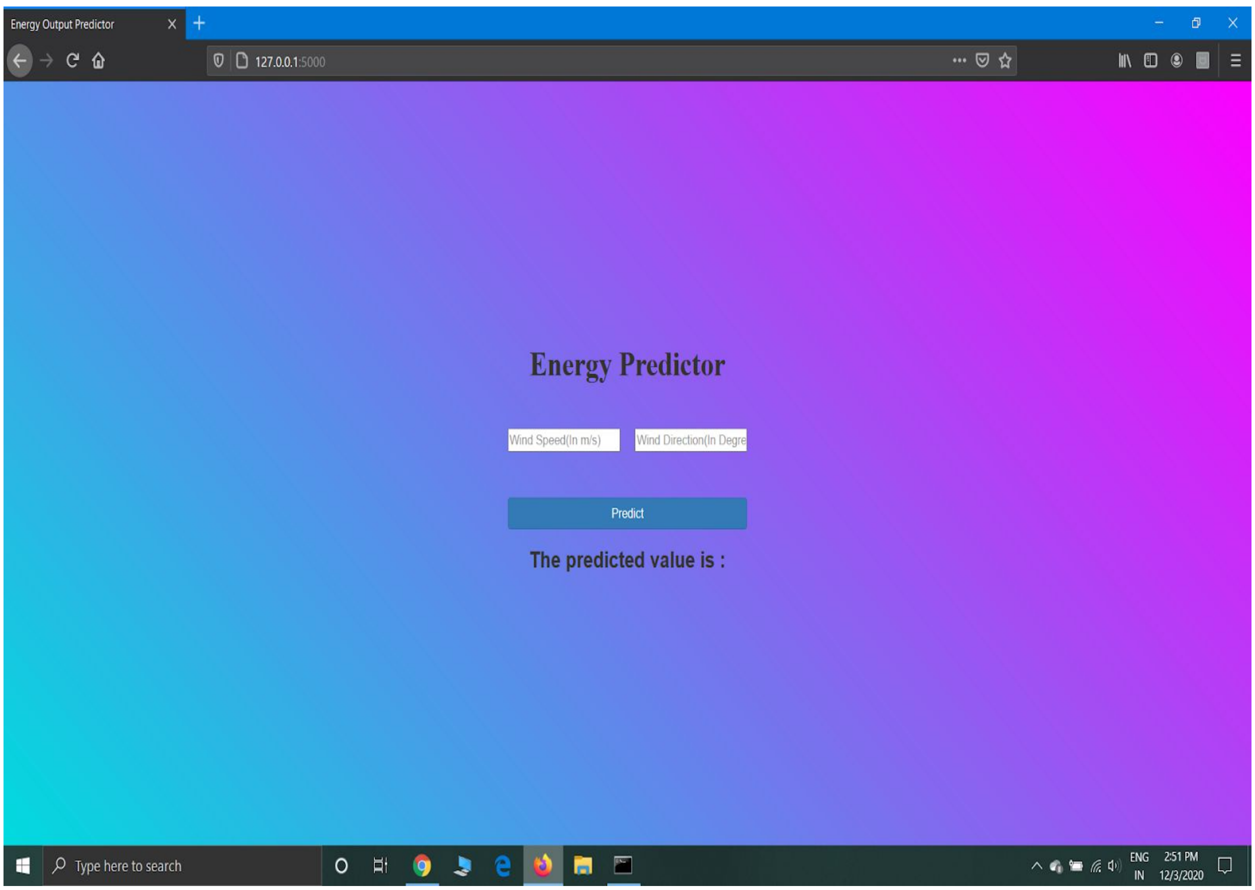
```
import pandas as pd
import numpy as np
import seaborn as sns
import datetime
import time
import matplotlib.pyplot as plt
%matplotlib inline

data = pd.read_csv('T1.csv')
data

data.shape
data.describe()

data.rename(columns={'Theoretical_Power_Curve
(KWh)':'Power','LV ActivePower (kW)':'ActivePower',"Wind
Speed (m/s)":"WindSpeed","Wind Direction
(°)":"WindDirection"},inplace=True)
data.head()

data['Time']=data['Date/Time'].apply(lambda x:
time.strptime(x,"%d %m %Y %H:%M")[4])
data.head()

temp=data['Time'][1:].values - data['Time'][0:-1].values
temp=np.array([0]+list(temp))
data['TimeDiff']=temp
data.tail()

data['Gust']=np.array([0]+list(data['WindSpeed'][1:].values
-data['WindSpeed'][:-1].values))
data.head()
data.dtypes

data_=data[data['TimeDiff'].isin([10,-50])]
data_.head()
```

```
len(data_)
data_=data_[data_['ActivePower']>=0]
data_.describe()

x=data_.sort_values(by='WindSpeed')['WindSpeed']
y=data_.sort_values(by='WindSpeed')['Power']
z=data_.sort_values(by='WindSpeed')['Gust']

plt.plot(x,y)
plt.plot(z,y)

sns.heatmap(data.corr(),annot=True)

x=data_[['WindSpeed','WindDirection','Gust']].values
y=data_['ActivePower'].values

x
y

plt.scatter(x[:,0],y)

from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test =
train_test_split(x,y,test_size=0.3,random_state=0)

x_train.shape, x_test.shape
data.head()

from sklearn.ensemble import RandomForestRegressor
LR =
RandomForestRegressor(n_estimators=10,criterion='mse',rando
m_state=0)
LR.fit(x_train,y_train)

y_pred = LR.predict(x_test)
y_pred
y_test

from sklearn.metrics import r2_score
r2_score(y_test,y_pred)
```

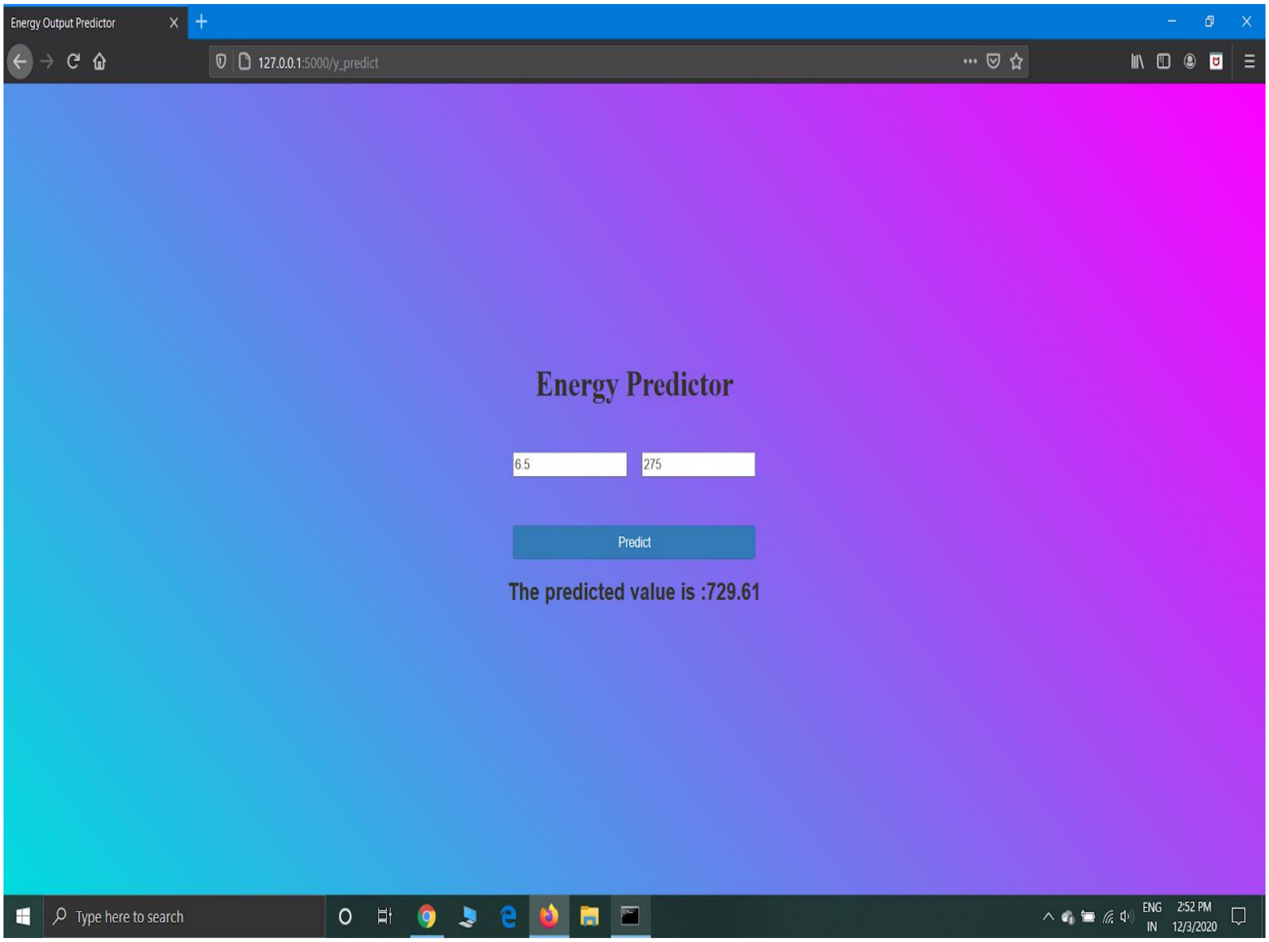
```
LR.predict([[6.8,275,0.5]])

import pickle
pickle.dump(LR,open("final.pkl",'wb'))

from joblib import dump
dump(LR,"final.save")
```

B. UI Output Screenshots

1.



2.


Energy Output Predictor
127.0.0.1:5000/y_predict
Energy Predictor
6.5
275
Predict
The predicted value is :729.61
Type here to search
ENG
IN
2:52 PM
12/3/2020