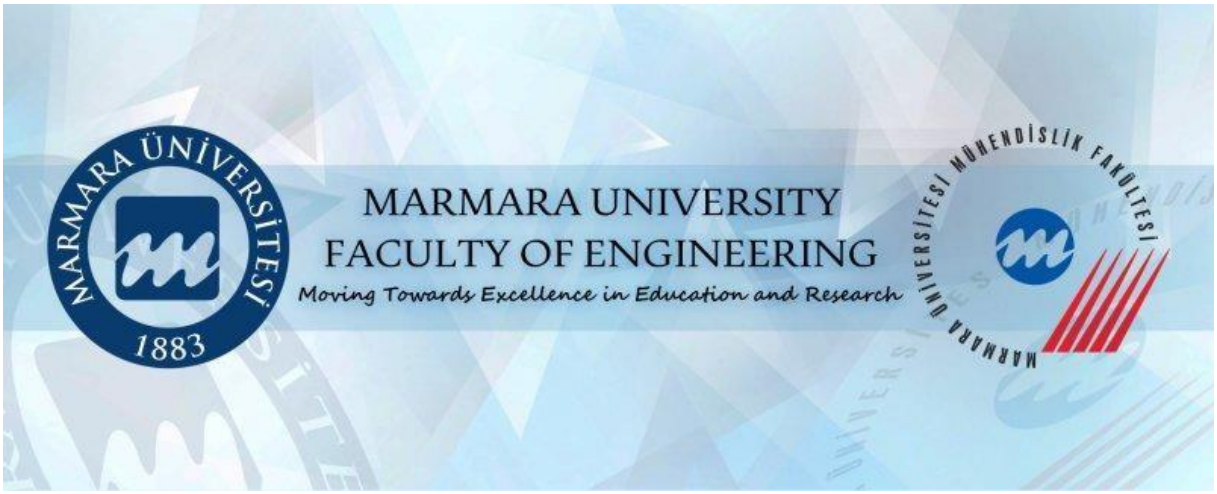

CSE4062 2020 SPRING PROJECT

ANALYSIS OF ELECTRIC PRODUCTION OF A WIND TURBINE

Group: 8

Name	ID Number	Mail Addresses
Furkan Can Ercan	150316044	furkanercan98@gmail.com
Muhammed Avcı	150414007	muhammedavci96@gmail.com
Yasin Gök	150115058	ygok.96@gmail.com

Project Scope

In this project we basically aim to generate an appropriate prediction model for our dataset by using regression algorithms. Our project based on a dataset which includes a year of measurements of a wind turbine plant and it contains active power generation of the turbine, wind speed and wind direction with respect to date and time, theoretical capacity of power generation of the turbine, the month of measurements, detailed time and date and day-night classification of the measurement time. It is a comprehensive dataset and it includes measurements per every 10 minutes during a whole year from 1st of the January 2018 00:00 to 31st of the December 2018 23:50

These columns of the dataset demonstrates a wind turbine, active power generation means the real electricity production of the turbine in that specific date and time in terms of kilowatt(kW) , wind speed explains the speed of the wind in terms of meter/second, wind direction explains the angle of the wind in terms of degrees, theoretical capacity of power generation explains the theoretical calculation of electricity production in terms of kilowatt/hour (kW/h),day-night classification explains the time between 00:00-08:00 as night and 08:00-00:00 as day and lastly month classification explains months as 1-12 as in a 1 year period.

By using this dataset we aim to build an efficient and a highly accurate model to predict active power generation of the plant by using regression algorithms. We are planning to arrange the dataset and check for the missing values and prepare the data as our first step and then to analyze the dataset, to select the important features for our model to predict our target variable, try and experiment different regression algorithms as simple linear regression, multiple linear regression, polynomial regression, Lasso Regression. We will try not to overfit or underfit our data by making our model as simple as it is possible but also complex enough to comprehend the training data. After these steps we will calculate the errors of different models with different measures such as R squared and standard error of estimate and chose our best model. In the end we will interpret the data via our most accurate model and try to make generalizations and to reach a conclusion.

Dataset Summary

	Date	Time	LV ActivePower (kW)	Wind Speed (m/s)	Theoretical_Power_Curve (KWh)	Wind Direction (°)	Month	Day/Night
0	01 01 2018	00:00:00	380.047791	5.311336	416.328908	259.994904	1	0
1	01 01 2018	00:10:00	453.769196	5.672167	519.917511	268.641113	1	0
2	01 01 2018	00:20:00	306.376587	5.216037	390.900016	272.564789	1	0
3	01 01 2018	00:30:00	419.645904	5.659674	516.127569	271.258087	1	0
4	01 01 2018	00:40:00	380.650696	5.577941	491.702972	265.674286	1	0
5	01 01 2018	00:50:00	402.391998	5.604052	499.436385	264.578613	1	0
6	01 01 2018	01:00:00	447.605713	5.793008	557.372363	266.163605	1	0
7	01 01 2018	01:10:00	387.242188	5.306050	414.898179	257.949493	1	0
8	01 01 2018	01:20:00	463.651215	5.584629	493.677652	253.480698	1	0
9	01 01 2018	01:30:00	439.725708	5.523228	475.706783	258.723785	1	0
10	01 01 2018	01:40:00	498.181702	5.724116	535.841397	251.850998	1	0
11	01 01 2018	01:50:00	526.816223	5.934199	603.014076	265.504700	1	0
12	01 01 2018	02:00:00	710.587280	6.547414	824.662514	274.232910	1	0
13	01 01 2018	02:10:00	655.194275	6.199746	693.472641	266.733185	1	0
14	01 01 2018	02:20:00	754.762512	6.505383	808.098139	266.760406	1	0
15	01 01 2018	02:30:00	790.173279	6.634116	859.459021	270.493195	1	0
16	01 01 2018	02:40:00	742.985291	6.378913	759.434537	266.593292	1	0
17	01 01 2018	02:50:00	748.229614	6.446653	785.281010	265.571808	1	0
18	01 01 2018	03:00:00	736.647827	6.415083	773.172863	261.158691	1	0
19	01 01 2018	03:10:00	787.246216	6.437531	781.771216	257.560211	1	0
20	01 01 2018	03:20:00	722.864075	6.220024	700.764700	255.926498	1	0
21	01 01 2018	03:30:00	935.033386	6.898026	970.736627	250.012894	1	0

Our dataset includes 8 columns and 50531 rows of data about a wind turbine. We have 8 columns which means 8 different attributes listed as:

Date: Demonstrates the specific date of the measurement in dd/mm/yyyy format. It is in text type.

Time: Demonstrates the specific time of the measurement in hh/mm/ss format. It is also in text type.

LV Active Power (kW) : It demonstrates the measured value of electricity production of the wind turbine in terms of kilowatts in corresponding time and date. It is a numeric attribute. It is also our target attribute. We will try to create a regression model to predict electricity production in the future.

Wind Speed: It demonstrates the speed of the wind in terms of meter/second in corresponding time and date. It is a numeric attribute.

Theoretical Power Curve: It demonstrates the theoretical calculation of the electricity production of the wind turbine in terms of kilowatt/hour. It is a numeric attribute.

Wind Direction: It demonstrates the angle of the wind in terms of degrees. It is a numeric attribute.

Month: It demonstrates the month of the corresponding measurement took place. It consists 12 different categories for each to represent 12 months. It is a categorical attribute.

Day/Night: It demonstrates the category of the corresponding measurement took place. It assumes from time 00:00 to 08:00 as night -0- and the rest of the day as day -1-. It is a categorical attribute.

```
Type of the  Date column  is : object
Type of the  Time column  is : object
Type of the  LV ActivePower (kW) column  is : float64
Type of the  Wind Speed (m/s) column  is : float64
Type of the  Theoretical_Power_Curve (KWh) column  is : float64
Type of the  Wind Direction (°) column  is : float64
Type of the  Month column  is : int64
Type of the  Day/Night column  is : int64
```