Short-term Wind Power Forecasting using K-Means Clustering with Deep Neural Network

By Devansh Ojha

I. INTRODUCTION

The development and utilization of renewable energy has been one of the hottest spots around the world. Wind power generation is rapidly expanding into a large-scale industry due to the cleanness and wide availability, and has been characterized as a fluctuating and intermittent power. An accurate and reliable wind power forecasting approach is essential for power quality, reliability management while reducing the cost of supplying spinning reserve.

II. PROBLEM DEFINITION

A wind power forecast corresponds to an estimate of the expected production of one or more wind turbines in the near future. In the electricity grid at any moment balance must be maintained between electricity consumption and generation - otherwise disturbances in power quality or supply may occur. Wind generation is a direct function of wind speed and, in contrast to conventional generation systems, is not easily dispatchable. Fluctuations of wind generation thus receive a great amount of attention.

III. OBJECTIVE

A data mining approach for wind power forecasting, which consists of the K-means clustering method and bagging neural network. The historical data are clustered according to the meteorological conditions and historical power. Pearson correlation coefficient is used to calculate the distance between the forecasting day and the clusters.

IV. SCOPE/IMPORTANCE OF PROJECT

Wind power forecasting (WPF) is significant to guide the dispatching of grid and the production planning of wind farm effectively. The intermittency and volatility of wind leading to the diversity of the training samples have a major impact on the forecasting accuracy. As the balance between consumption and generation of power must be kept therefore fluctuations of wind generation is a very important field of study.

V. METHODOLOGY

To deal with the training samples dynamics and improve the forecasting accuracy, a data mining approach consisting of MinMax normalization, K-means clustering and deep neural network is proposed for short-term WPF. Based on the similarity among historical days, K-means clustering is used to reduce the dataset by taking the centroids of the various clusters. Further, this reduced dataset is used to predict the power generated in the future, when given normalized values for the parameters.

VI. RESULTS

Description of the initial data:

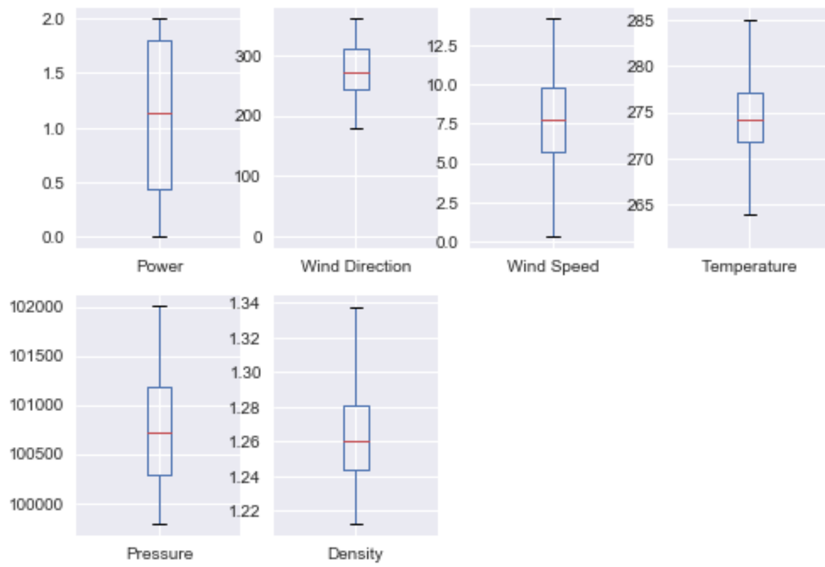
WIND FORECASTING DATASET

The number of tupls vs attributes of the dataset are:
(3000, 6)

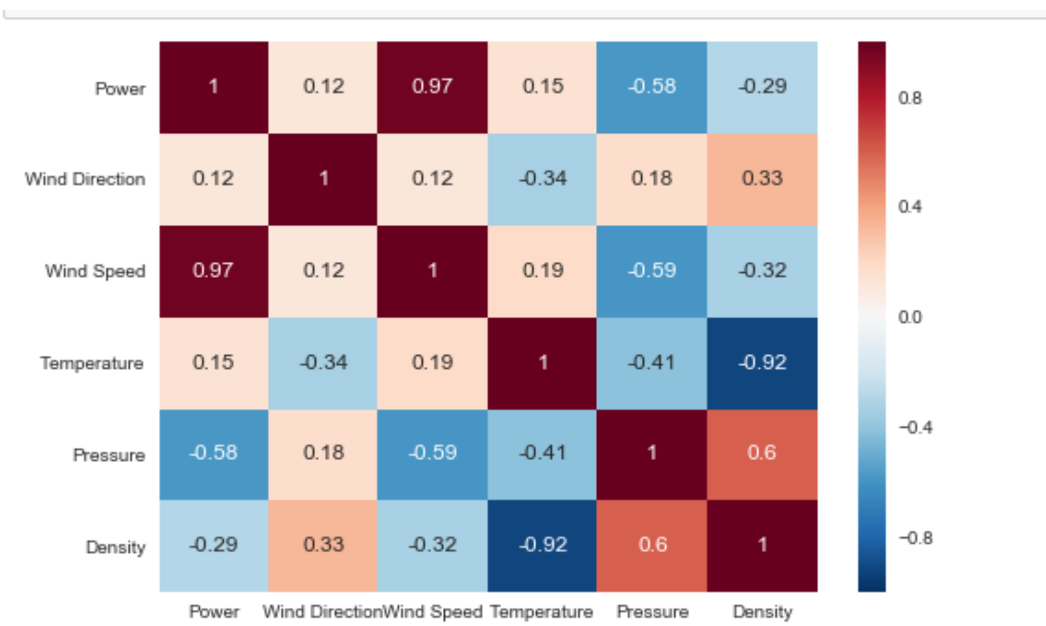
Initial few tuples are:

	Power	Wind Direction	Wind Speed	Temperature	Pressure	Density
0	0.668	319.914	6.519	277.955	100892.168	1.240
1	0.637	319.876	6.418	277.878	100894.912	1.240
2	0.607	320.195	6.320	277.799	100897.656	1.240
3	0.563	319.953	6.177	277.665	100906.816	1.240
4	0.521	320.701	6.039	277.494	100906.816	1.241

	Power	Wind Direction	Wind Speed	Temperature	Pressure	Density
count	3000.000000	3000.000000	3000.000000	3000.000000	3000.000000	3000.000000
mean	1.084307	268.376159	7.594241	274.029058	100795.292325	1.262555
std	0.700270	56.247664	2.800689	4.840671	562.721775	0.028089
min	0.000000	0.026000	0.314000	261.569000	99780.680000	1.212000
25%	0.438000	243.926250	5.660250	271.787000	100290.648000	1.244000
50%	1.136500	270.897000	7.783500	274.121500	100712.712000	1.260000
75%	1.796000	309.282750	9.778000	277.054000	101187.662000	1.281250
max	2.000000	359.982000	14.163000	285.093000	102001.824000	1.338000



A heatmap to check which to attributes are most correlated

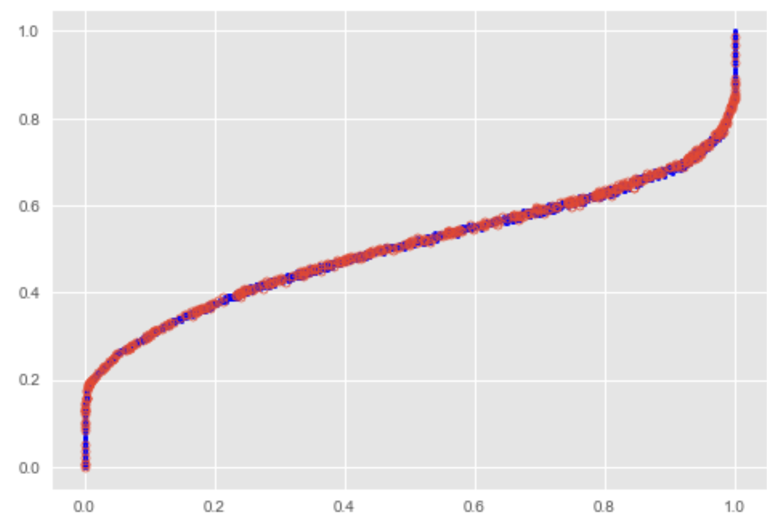


	Power	Wind Direction	Wind Speed	Temperature	Pressure	Density
Power	1	0.12	0.97	0.15	-0.58	-0.29
Wind Direction	0.12	1	0.12	-0.34	0.18	0.33
Wind Speed	0.97	0.12	1	0.19	-0.59	-0.32
Temperature	0.15	-0.34	0.19	1	-0.41	-0.92
Pressure	-0.58	0.18	-0.59	-0.41	1	0.6
Density	-0.29	0.33	-0.32	-0.92	0.6	1

Thus, we see that Wind Power is most correlated to Wind Speed

Representation of the dataset after KMeans:

(Blue points-> data
Red circles-> centroids)



On running the neural network for the given test values of

Wind Direction, Wind Speed, Temperature, Pressure, Density:

```
--
[ 0.83228  0.31758  0.70522  0.78958  0.39894]
P5: 0.11202756315469742
```

VI. CONCLUSION

Our approach for wind power forecasting consists of normalized data values in the range [0,1] to get similar values. Further we clustered the data to reduce the total dataset for an accurate and quicker prediction of the actual power generated by the turbines. Finally we applied a Deep Neural Network that works on back propagation to correct and predict the Wind Power generated by the turbines. Our approach gives a prediction that is consistent with the actual values to a good accuracy and can hence be applied to predict powers when the actual output is not known.