Prediction of Wind Power: Do Machine Learning (ML) Algorithms Outperform Linear Regression Algorithms?

By: Priya Purohit

Summary of Purpose

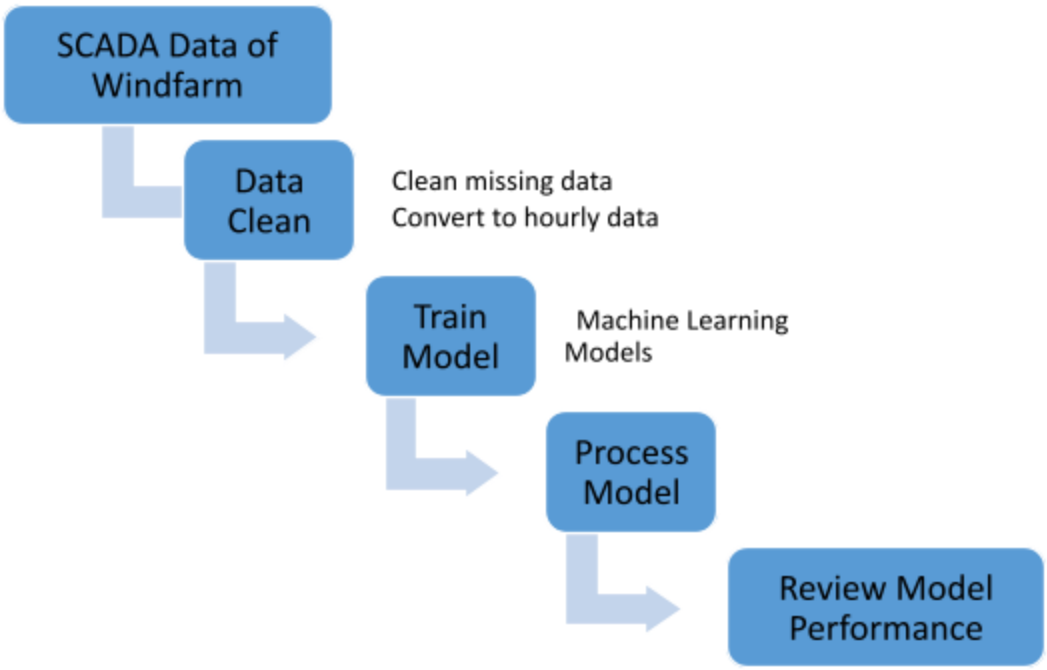
Renewable energy from wind power is increasingly becoming a major source of electricity in the utility fleet (Burton, 2012). However, compared to traditional sources of electricity, the randomness of the electricity generated by wind power due to the intermittent nature of wind and other physical factors poses a challenge to the electricity grid operators. It is critical for the safe operation of the electric grid system that utilities can accurately forecast the wind power that would be generated (Burton, 2012).

This research investigates whether wind power forecasting using the machine learning algorithm Decision Tree Regression (DTR) dominates the forecasting performance of the Linear Regression (LR) statistical method. The algorithms are implemented using Python libraries. The results are tabulated and presented.

Procedure

The wind power and the meteorological information was used from a freely available SCADA database for a wind farm located in Turkey (Kaggle, 2024). The dependent variable is the actual wind power generated by the farm in kilowatt. The independent variables considered are the wind speed (m/s), wind direction (◦), and rated power (kW). The data recorded is for one year between 1-Jan-2018 and 31-Dec-2018 with a ten-minute interval.

The input file was read and pre-processed to check for any null values (there were none) using pandas library. Using sklearn libraries (Scikit-Learn, 2024), the data was then randomly split with 80% of the data used as the training data to train the models. Using the trained fitted model, testing was performed on the 20% of the testing data to test the forecasting performance. The result of the forecasted wind power and the actual wind power are graphed using matplotlib libraries (Matplotlib, 2024). The comparison of the forecasting performance is using the Mean Square Error (MSE) measure. The MSE of both LR and DTR are tabulated.



Results

Measure	Linear Regression Model	Decision Tree Regression Model
Mean Squared Error (MSE)	165665.36	158582.15
Root Mean Square Error	407.02	398.22
Mean Absolute Deviation Error (MADE)	186.19	157.67

515376961

	Training data set			Testing data set		
Model	Mean Squared Error (MSE)	Root Mean Square Error	Mean Absolute Deviation Error	Mean Squared Error (MSE)	Root Mean Square Error	Mean Absolute Deviation Error

		(RMSE)	(MADE)		(RMSE)	(MADE)
Linear Regression	162490.04	403.10	184.78	165665.36	407.02	186.19
Decision Tree Regression Model	120201.87	346.70	141.23	158582.15	398.22	157.67

Analysis

DTR learning outperforms the LR training. This is seen by comparing the RMSE and MADE measurements of the training and testing data sets in the table above. The difference between the LR learning algorithm and testing algorithm is low compared to the improvement seen in the DTR learning algorithm and DTR testing algorithm measurements.

Conclusion

It can be concluded that DTR is better able to learn from the data set compared to the LR algorithm. The reviewing MADE results show that for the given parameters DTR outperformed the LR by 15%.

References

1. Tony **Burton**, Nick Jenkins, David Sharpe, Ervin Bossanyi. Wind Energy Handbook ISBN:9780470699751. (2012)
2. **Kaggle** database. https://www.kaggle.com/datasets/berkerisen/wind-turbine-scada-dataset. Accessed on 19-Feb-2024 (2024)
3. Scikit-Learn. https://scikit-learn.org/0.21/modules/tree.html. Accessed on or about 1-Feb-2024 (2024)
4. Matplotlib. https://matplotlib.org/. Accessed on or about 1-Feb-2024 (2024)