Machine Learning for Wind Turbine O&M

Conor McKinnon, James Carroll, Alasdair McDonald, and Sofia Koukoura

Wind & Marine Energy Systems CDT, Rm 3.36, Royal College Building
University of Strathclyde, 204 George Street, Glasgow, G1 1XW

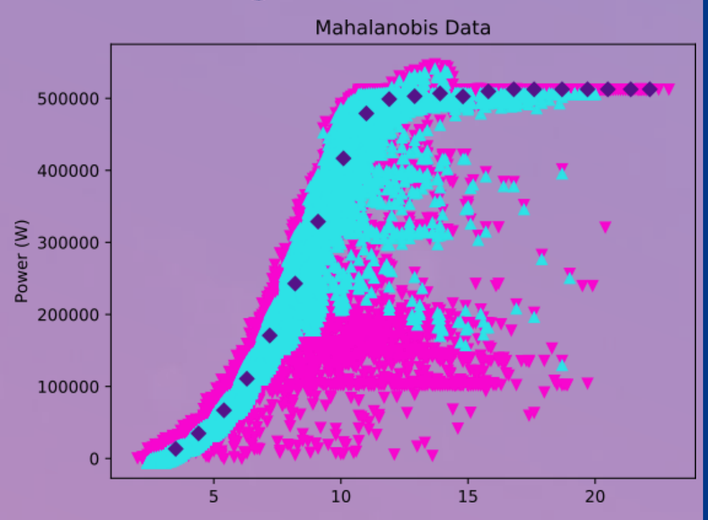
conor.mckinnon@strath.ac.uk

Data Pre-Processing

Data Cleaning: This involves removing "normal anomalies" from the data, such as negative power and curtailment. This can be done with clustering and a distance classifier.

Features: Can either be data driven or domain based. Data driven approaches look at the statistical relationships between features, whereas domain knowledge is typically from expert opinion.

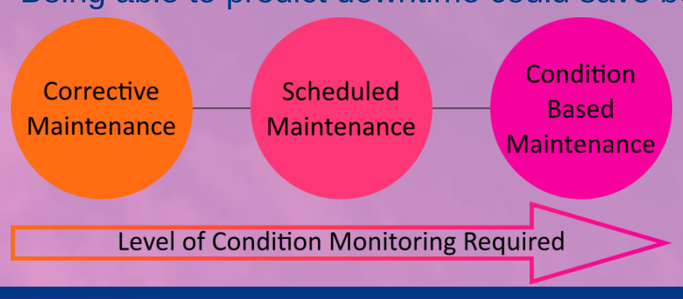
Re-evaluation: The results of the model's testing can then influence the previous two steps mentioned, and can also affect the aims and objectives.



Blue: Clean, Pink: Raw

Aims and Motivations

- More turbines being built offshore
- Vessel day rates can range from thousands to hundreds of thousands
- Access to turbines limited to weather windows
- Being able to predict downtime could save both time and money



Methodology

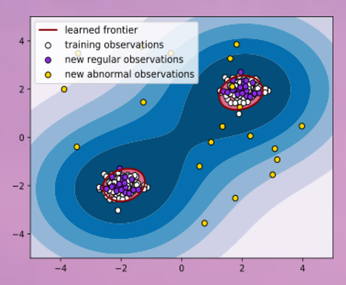
Training: This step is where data is fed into the model (typically both input and target variables), and then the model "learns" the patterns and relationships of the data. It uses a cost equation to determine how well constructed the model is.

Testing: This is where the model can truly be evaluated. The model is now fed data without the target variable, and has to "predict" what this variable is (either a number, or a label). The model is then given an accuracy measure to compare between model performance. Runtime is also a qualifier.

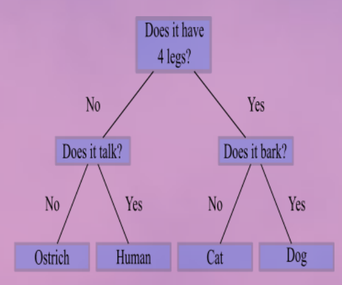
Model Examples

Classifiers: These models will predict a qualitative variable (a label, or an anomaly flag). Typically produces a boundary around/through data to separate them into classes. Can be for multiple classes or just one.

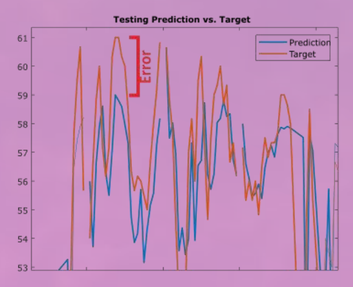
Regressors: These models predict quantitative variables (an input will be related to an output value). This can learn both temporal and spatial relationships between data. Can be used for forecasting/prediction, or even anomaly detection (as shown below).



(a) OCSVM



(b) Decision Tree

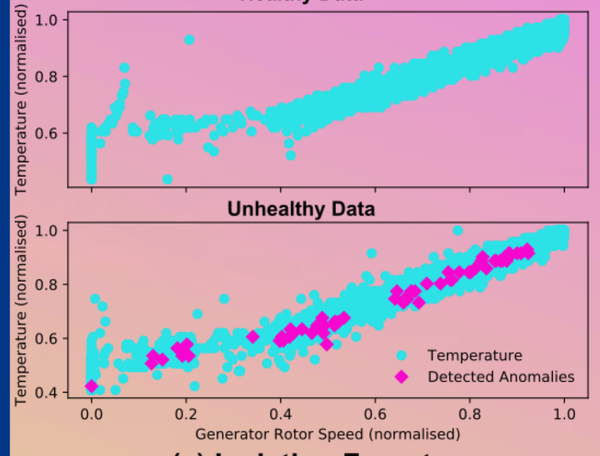


(c) NARX

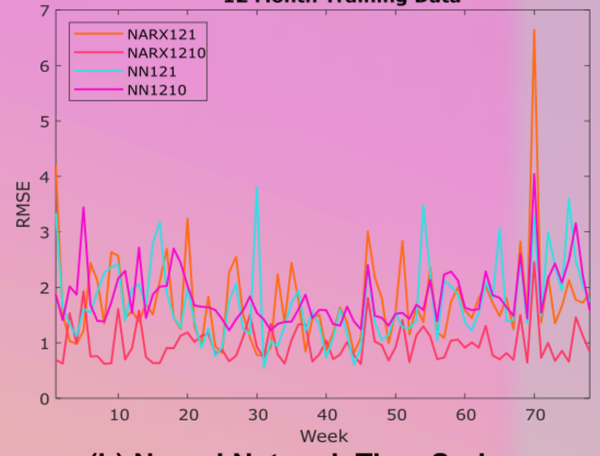
Results and Discussion

Machine learning can be used for wind energy for a number of applications, from fault diagnosis to fault prediction. One such area, that we have explored for the past 18 months, is anomaly detection. This looks at data (typically SCADA) and during training will find a boundary containing all of the "normal" data. This normal boundary is then applied to new unseen data to indicate anomalies, which can then be compared between multiple test cases.

Different methods of analysis can be used to assess the performance of the models. For example, the number of anomalies detected, or the peaks in the errors/anomalies over time. Context is also required, for example in figure (b) the NARX models appeared to perform worse, but during training were shown to better learn the data, so that when an anomaly was detected, you could be more sure of it.



(a) Isolation Forest



(b) Neural Network Time Series

Left: a) This graph shows the increase in anomalies from a healthy month to a test month a year later, for an isolation forest (decision tree) model.
b) The Neural Network time series looks at the weekly moving RMSE values between the values predicted by the Neural Networks and the actual target variables.