

BHE Capstone Final Presentation

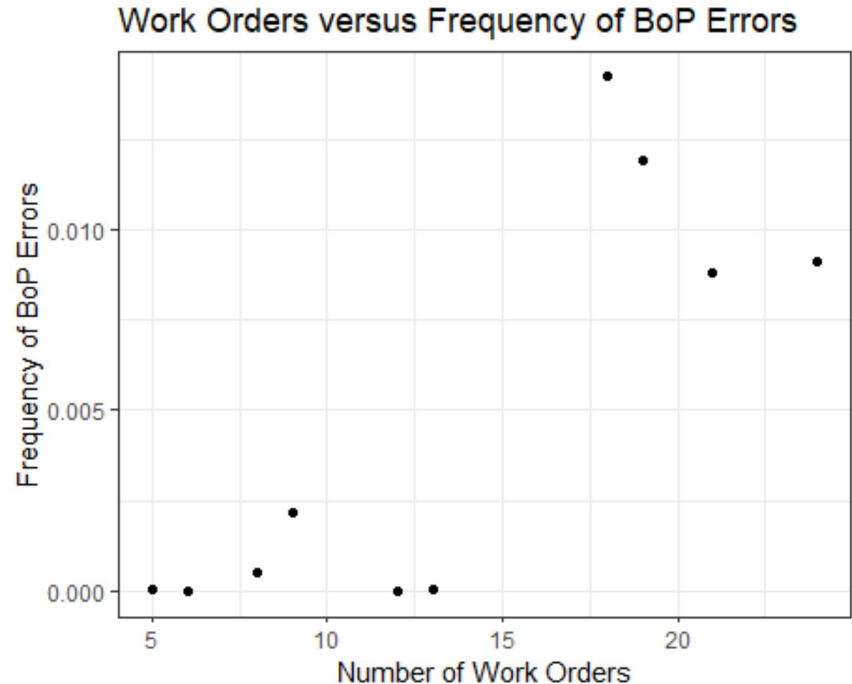
Maeve Emery, Anya Zarembski, Eric Wiederin

A dark blue diagonal gradient bar that starts from the bottom left corner and extends towards the top right corner, covering the lower half of the slide.

Research Question

What causes Balance of Plant errors and how can we predict them?

Balance of Plant errors are seeming correlated with a higher number of work orders on a given turbine. This gives us motivation to study these errors, as a higher number of work orders per Turbine is likely costly for BHE.



Data Preparation

- We first had to aggregate the data into one file and also include the weather data we received from the other section.
- We also had to impute the data with values that made sense for every column.
- We also created some variables to assist in model making, such as a binary Balance of Plant variable
- Finally, later in the project we created a dataset aggregating the data by day to see if we could make predictions for a BoP error occurring within a certain day.

K-Nearest Neighbors Creation

We felt it may be appropriate to use a Knn model to classify when Balance of Plant errors occurred based on the other numeric variables.

For the real-time data, we were able to create a model that predicted Balance of Plant errors with about 99% accuracy.

However, when this same method was applied to the data lagged by an hour as well as the daily dataset, the model only predicted no errors. This was accurate, as Balance of Plant errors are rare, but not helpful.

knn.359	0	1
0	54768	338
1	128	102

Real Time Model

knn.30	0	1
0	387	13
1	0	0

Daily Model

Random Forests

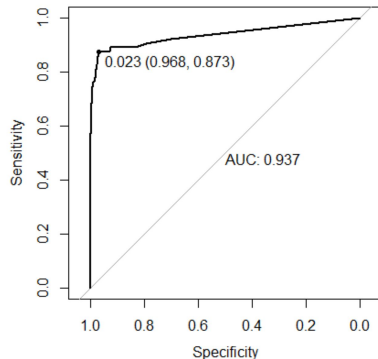
- Random forests were our next attempt to predict BoP errors.
- Compared real-time model to model predicting 24 hours in advance.
- Went from OOB error rate of 0.35% to 0.56%
 - Not a large increase - made us confident we could use this for future predictions
- Final random forest model predicting 24 hours in advance:
 - 0.55% OOB error rate
 - AUC: 0.937
 - Pi Star: 0.023

Confusion Matrix (Pi Star = 0.05)

	No Error	Error
No Error	28908	30
Error	130	144

Confusion Matrix (Pi Star: 0.023)

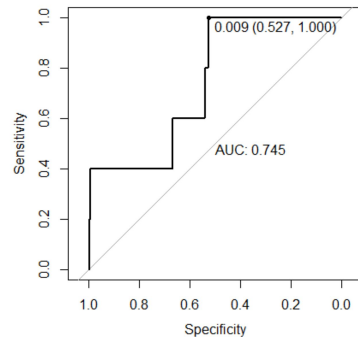
	No Error	Error
No Error	7004	8
Error	229	55



Daily Dataset

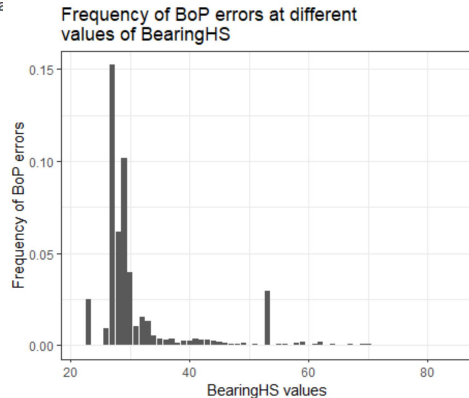
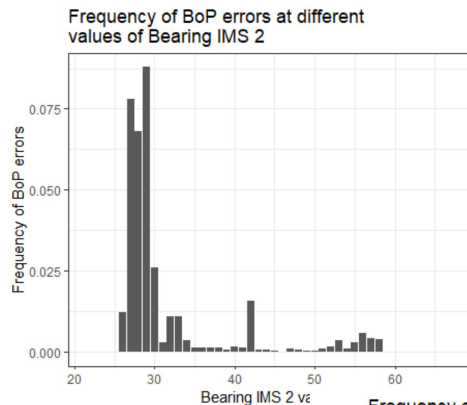
- Alternative to the dataset that would have an error for any error within the next given timeframe.
- Simplified version that has one entry for each turbine each day
- Summary statistics for each day
 - Extrema only
- Built random forest to predict if a balance of plant error will occur next day
- Results:
 - OOB Error: 3.01%

	No Error	Error
No Error	138	0
Error	124	5



Final Thoughts

- Predicting Balance of Plant errors 24 hours in advance with a random forest was most successful
- The most important variables in the occurrence of a balance of plant error are:
 - Hydraulic Pressure, Bearing IMS 2, Minimum Temperature, Bearing HS, & Generator RPM
- Further research ideas:
 - Extending the model to predict further in advance (48 hours, a week, etc)
 - Predicting specific types of balance of plant errors (Converter Tripped, Mainbreaker Cut Out)
 - Creating a forest entirely dependent on external factors (i.e. weather)
 - Would be able to use forecast data to predict further in advance



Thank You!

Questions?