WEBEE ENERGY MONITORING

by: Sumant Anantha, Tobias Bautista, Fenil Gala, Noah Kirkorsky, Shriya Gupta, Jimmy Dinh, Louis Caceres, Artharv Wayse, David Calalang, Sahithi Gokavarapu

1 INTRODUCTION

Webee is an asset intelligence company that specializes in innovation & emerging technologies. This year, the Webee Energy Monitoring team works with Energy Data from one of Webee's clients: Colgate. The goal of this project is to leverage the data in the Colgate Greenwood Plant in order to use energy more efficiently, cut down costs, and help shape the evolving world.

2 DATA

The raw data was collected from sensors placed in a factory.
We have 1 month of data, recorded every 8-10 seconds.
We averaged the data by minute, hour, and day long intervals reducing the size of the dataset.

- The dependent variables are Power Factor, THD Voltage, and Currents. The independent variable is Time.
- Data was analyzed by using Box Plots, Histograms, and Bar Graphs.

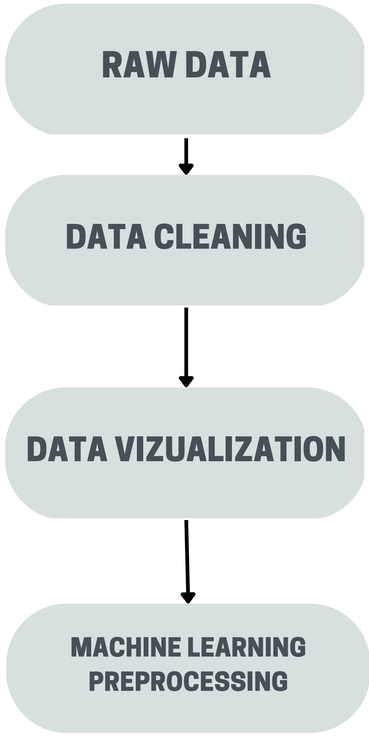
3 PREPROCESSING

- Given as a .csv (commas-separated value)
- Contained data for specific months

- Dropping rows with n/a values
- Analyzing outliers in the data
- Creating subsets with specific columns to be used

- The different subsets were analyzed through plots/graphs
- Some examples were box-whisker plots and scatter plots
- Data was also split/analyzed by splitting into time-series values

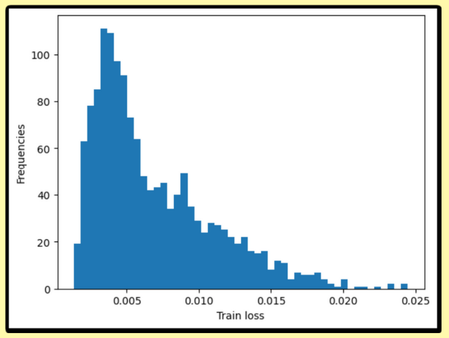
- Normalization of data (to values between 0 - 1)
- Splitting data into designated training, validation and testing sets



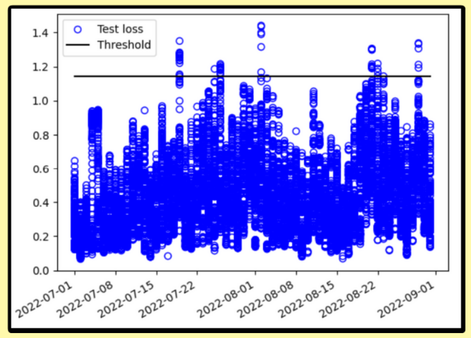
4 MODELLING

NEURAL NETWORKS

- Data is split between training, validation and test data.
- An autoencoder was used to deconstruct and reconstruct the data from each substation.
- This approach was applied to various parts of our data, such as the power factor and active energy analysis



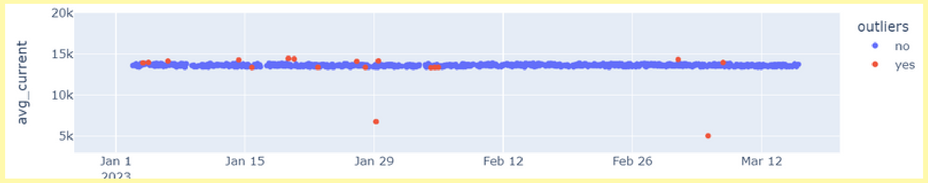
Power Factor in Substation 1



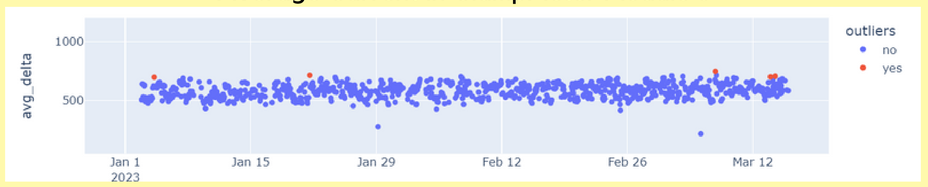
Active Energy in Substation 1

ISOLATION FOREST

- Current Data (Current1, 2, and 3) is used to analyze trends and detect potential anomalies
- The Isolation Forest Model split the data into branches, flagging the shortest ones as anomalies since it does not follow the trend
- This approach was applied to the average current values and the average difference between current values (difference refers to asynchronicity and inefficiency)



Average Current in ClampMeter Module 3



Average Difference in Currents in ClampMeter Module 3

5 EVALUATION

Neural Network

PROS:
- Able to predict losses based on input data trends
- Clear comparisons between individual substations

CONS:
- *Bias from interpretations of the input data*
- *Overly dependent on specific substations*

Isolation Forest:

PROS:
- *More systematic/ mathematical approach*
- *Less dependence on labelled input data*

CONS:
- *Hard to see* why anomalies are chosen
- Heavily dependent on input parameters

6 CONCLUSION / FUTURE GOALS

- Established a personalized threshold for the neural network models
- Gained new understanding of various predictive model approaches
- Benchmark the performance of the models against real anomalies in a device.
- Optimize the model by leveraging information from all the variables

7 REFERENCES

- https://learn.microsoft.com/en-us/azure/machine-learning/how-to-train-tensorflow
- https://pub.towardsai.net/autoencoder-for-anomaly-detection-using-tensorflow-keras-7fdfa9f3ad99

8 ACKNOWLEDGEMENT

Thank you to our corporate partner mentors from Webee:
- Lucas Petralli ,Yamil Abraham , Davor Margetic James Stone
- The Data Mine Staff