WEBEE MACHINE MONITORING

THE DATA MINE CORPORATE PARTNERS SYMPOSIUM 2022

Students:
Shi Quan Pang
Param Patel
James Stone
CRP TA: Luke Foltz

1 | INTRODUCTION

Our Project
Our team was tasked by Webee to analyze sensor data (from Webee products attached to industrial rotating machines) in order to provide Prescriptive Maintenance for Webee customers.

What is Webee?
A SaaS company that provides industrial remote monitoring solutions for other companies.

The Goal
Without any ground truth in our datasets, our team is tasked with an unsupervised time-series anomaly detection problem.

2 | BACKGROUND

IIOT Sensors
Webee utilizes IIOT sensors that collect data, most importantly the acceleration, temperature, velocity, and current of any given motor-driven machine. *(pictured)*

Dataset Limitations
The datasets we were given have many missing values and no specifics as to when a given machine is on/off. As such, many of the values are at a "baseline" level (explained in section below).

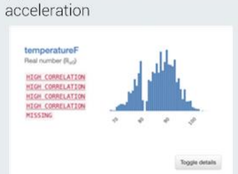
Operating Conditions
Our team focused mainly on a group of machine data all within the same vicinity. As such, a given machine A could have been impacted from the noise a machine B generated.

3 | RESEARCH METHODOLOGY

Our Pipeline Workflow:

A - Raw Data Set

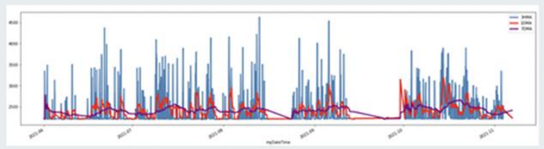
From raw CSV data:
(1) Drop columns
(2) Process date and time stamps
(3) pair data based on '5min' interval
(4) calculate RMS values of velocity and acceleration



B - Pre-processing & Variables

From raw CSV data:
(1) Drop columns
(2) Process date and time stamps
(3) pair data based on '5min' interval
(4) calculate RMS values of velocity and acceleration

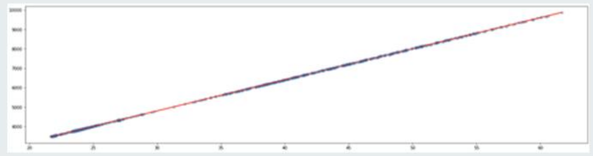
This image shows the different moving averages (by time -- denoted by line color) within our datasets.



D - Data Imputation

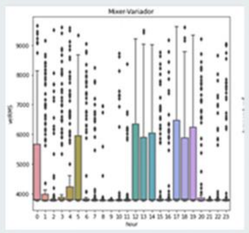
(1) Fill missing velocity data based on linear regression of acceleration data
(2) Fill missing temperature data based on interpolation

This image shows the correlation between acceleration and velocity in respect to our dataset.



E - GMM On/Off Subset

(1) Velocity data follows a bi-modal distribution corresponding to when the machine is 'on' and 'off'
(2) Use Gaussian Mixture Model to determine a threshold value that splits the data into the 2 subsets
(3) Use 'on' dataset for subsequent analysis



This image displays the different times a certain dataset will report to appear "off" (low power, undistinguishable data) or "on".

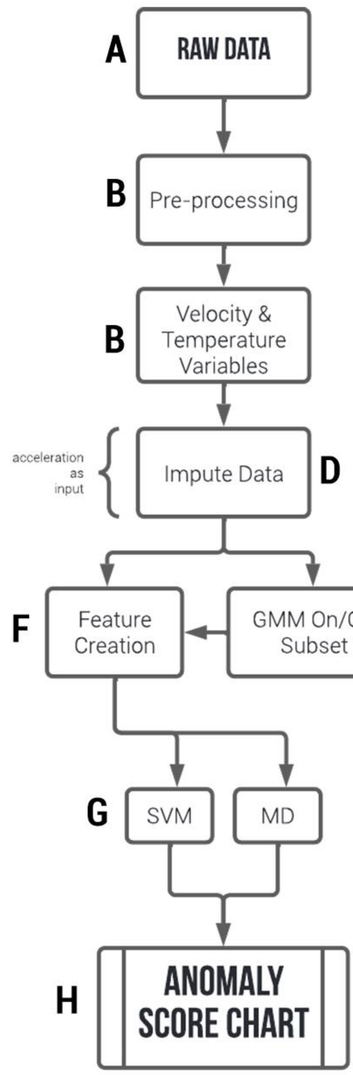
F - Feature Creation

Extract more meaningful information from velocity and temperature data by creating new features consisting of:
(1) Moving averages of velocity and temperature data
(2) Both 'on' and 'on and off' data
(3) XYZ-plane and RMS values
(4) CUSUM
(5) Time of day
(6) Time between last data point

G & H - Anomaly Detection

Anomaly Detection:
Run 2 anomaly detection algorithms and identify anomaly risk based on their results. Data is trained for a given day based on training data of past 60 days
(1) One-Class Support Vector Machine - Classifies new data as similar or different to the training set based on user-defined fraction of anomalies 'nu'
(2) Mahanabolis Distance Detection - Compares MD between a data point and the distribution. Anomaly is detected if p-value that exceeds threshold is detected
(3) Anomaly Score Chart: Weighted approach that considers anomalies detected in both algorithms.

This image displays the use of MD and its threshold (denoted by red line).



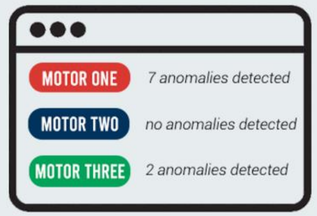
Abnormality Points

This graph displays the abnormality score by day. Red denotes high abnormalities. Yellow is moderate warning. Green falls within normal operation.

4 | CONCLUSIONS

Total Pipeline Progress

Shown below is a future vision of what a Webee client may see from the front-end when the pipeline is implemented. Within a given day, the anomaly risk is calculated from the past based upon the weighted score of all models.



5 | FUTURE GOALS

Integrating Model to Webee's Software
As shown above, we hope that our model/pipeline can be utililzed by a client by merging it with Webee's software (front-end).

On-line Testing and Validation of Model
Sometime in the near future, Webee aims to test the model with a company in a real-world setting.

Improving Model using Labelled Data
With future use-cases, labelled data will give a ground truth and more accurate understanding of the data for better results.

Enabling Prescriptive Maintenance
A goal for the future, we want to be able to predict and prescribe when maintenance is needed.

6 | REFERENCES & ACKNOWLEDGEMENTS

- David Glass, Kali Lacy, and Shuennhau Chang for guiding us
- Our CRP TA Luke Foltz
- Dávor Margetic, Lucas Funes, and Darío Farfán from the Webee team

Links:
https://towardsdatascience.com/support-vector-machine-svm-for-anomaly-detection-73a8d676c331
https://stackoverflow.com/questions/40874263/how-to-get-standard-deviation-from-each-components-in-sklearn-gmm-after-fit
https://towardsdatascience.com/machine-learning-for-anomaly-detection-and-condition-monitoring-d4614e7de770
https://stats.stackexchange.com/questions/360036/cusum-algorithm-and-first-derivative