**1. INTRODUCTION**

Our Project: We are tasked by Webee to analyze sensor data from Webee's rotating machinery to provide predictive maintenance for Webee customers.

What is Webee? Webee is a company that provides industrial companies with solutions for predictive maintenance.

The Goal: We need to build our own dataset for anomaly detection.

**2. BACKGROUND**

**IOT Sensors**

We use IoT sensors to collect real-time data, and most importantly the acceleration, temperature, and rotation of any given motor-driven machine. (picture)

**Dataset Limitations**

The datasets we use have many missing values and no specific time when a given machine is up and running. As such, many of the values shown are "estimated" (explained in section below).

**Operating Conditions**

Our team is based mainly on a group of machines data all within the same vicinity. As such, a given machine is not impacted by the noise from other machines.

**3. RESEARCH METHODOLOGY**

**Our Pipeline Workflow:**

**A - Raw Data Set**

- From raw CSV data:
  1. (E.g.) columns
  2. (E.g.) data based on time stamp
  3. (E.g.) calculate RMS values of velocity and acceleration

**B - Pre-processing & Variables**

- Reduce Columns
- Process data and time stamps
- Pair data based on time interval
- Calculate RMS values of velocity and acceleration

**C - Data Imputation**

- Fill missing velocity data on linear regression of acceleration data
- Fill missing temperature data based on interpolation

**D - GMM On/Off Subset**

- Velocity data follows a bi-modal distribution corresponding to when the machine is "on" and "off"
- Use Gaussian Mixture Model to determine a threshold value that splits the data into 2 subsets
- Use SVM for further analysis

**E - Feature Creation**

- Extract meaningful information from velocity and temperature data in training new features for the model
- Moving averages of velocity and temperature data
- Both X, Y, and Z
- YZ plane and RMS values
- Correlation
- Time of day
- E: Time between last data point

**F & H - Anomaly Detection**

**Anomaly Detection**

- Runs 2 anomaly detection algorithms and identifies anomaly based on thresholds. Data is trained for a given day based on training data of past 60 days
- Mahalanobis Distance Detection - Detects changes in the distribution of data relative to the distribution of data from previous days.
- Mahalanobis Distance Detection - Compares MD between data points to detect anomalies. If a data point exceeds the threshold, it is detected.
- Anomaly Score Chart - Weighted approach that considers anomalies detected in both algorithms.

**4. CONCLUSIONS**

**Total Pipeline Progress**

**5. FUTURE GOALS**

- Integrating Model to Webee's Software
  - To allow Webee's software to be used by a client by leveraging Webee's algorithm.
  - On-line Testing and Validation of Model
  - Compare the performance of the model with the actual data in real-time.

**6. REFERENCES & ACKNOWLEDGEMENTS**

- David Glass, Kail Lacy, and Shennhau Chang for guiding us
- Our CRP TA Luke Folket
- Dakot Morgan, Lucas Funes, and Daro Farfan from the Webee team

*Links:*