PREDICTIVE MAINTENANCE TOOLBOX

IE3100M Systems Design Project AY2019/2020
Department of Industrial Systems Engineering & Management

Group 10 Members: Jonathan Loh, Isaac Francis s/o Benert, Lie Yi Liang, Muhammad Zikry Bin Khazid
Supervising Professor: He Shuangchi
Micron Supervisor: Tan Peng

NUS National University of Singapore

Micron

PROJECT OVERVIEW

Problem Description

To develop an efficient and effective methodology to extract relevant sensors and predict sensor values to be used for Predictive Maintenance of tools.

Inaccurate Maintenance Dates

Ineffective sensors will lead to machine failures which causes:

1. Scrapping of Wafers
2. Manpower Wastages

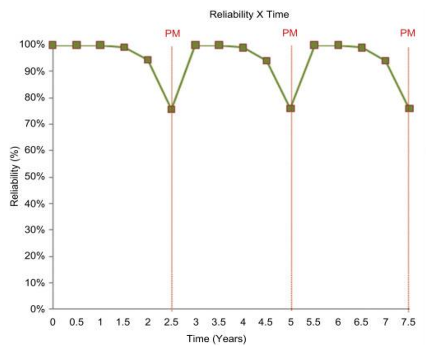
Problem Background

There are 12 tools in FAB10W each having nearly 200 sensors measuring different performance indicators:

Temperature | Gas Flow | Water Level

Saw-Tooth Trend

The data of an ideal sensor will follow a **saw-tooth trend,** indicating that the tool is reaching its maximum capacity before a maintenance takes place, bringing the tool back to optimal condition.



Current Solution Implemented By Micron

Currently, to extract useful sensors from the pool of sensors, Micron employees are tasked to **manually** sieve out these sensors by identifying a saw-tooth trend in those sensors.

Disadvantages

Disadvantages of Current Solution

1. It is a time consuming process to manually look through 200 sensors within each tool to identify sensors that follows a saw-tooth trend
2. It will be ineffective as there is no statistical evidence that one sensor is better than the other
3. There are no benchmark of rules that states what an ideal sensor should be which is what the Micron employees should look out for.

Objectives

1 Ranking System

- Come up with a suitable methodology to **rank** the sensors based on their relevance with predictive maintenance
- Using the **Coefficient of Determination** values to accurately filter out ineffective sensors using a level-system based on the criteria

2 Modelling Relevant Sensor

- Modelling relevant sensors based on historical data to set **accurate** predictive maintenance dates of tools
- Setting appropriate **parameters** to ensure that model created is accurate and precise

3 User-Interface

- Providing a user-interface terminal to allow Micron employees **easy access** to utilize this tool
- The user-interface created must be **user friendly** and **intuitive** for users without any background knowledge on computing

DATA PREPROCESSING

Step 1
Removal of any columns or rows that are not useful in the analysis such as the count and range of data

Step 2
Filling in of missing values that are not recorded by the sensor using the forward fill method

Step 3
Using the Moving Average 5 (MA5) method to smoothen the data set by removing the **white noise**

Step 4
Combining all 12 datasets into 1 to conduct a comprehensive analysis on the sensors' data

The purpose of data preprocessing is to process the raw data and remove unnecessary information that is irrelevant to our study such as white noise. Furthermore, data preprocessing allows us to fill in missing values using forward fill which will allow us to identify specific trends in the dataset effectively.

PROJECT METHODOLOGY

Sensor Ranking System (SRS)

Criteria of a Good Sensor

1. Sensor's data must follow a saw-tooth trend
1. Sensors' data are reactive to maintenance dates
1. Consistent trends between maintenance periods

Statistical Methods For SRS

Coefficient of Determination calculates the proportion of the variance in the dependent variable that is predictable from the independent variable(s).

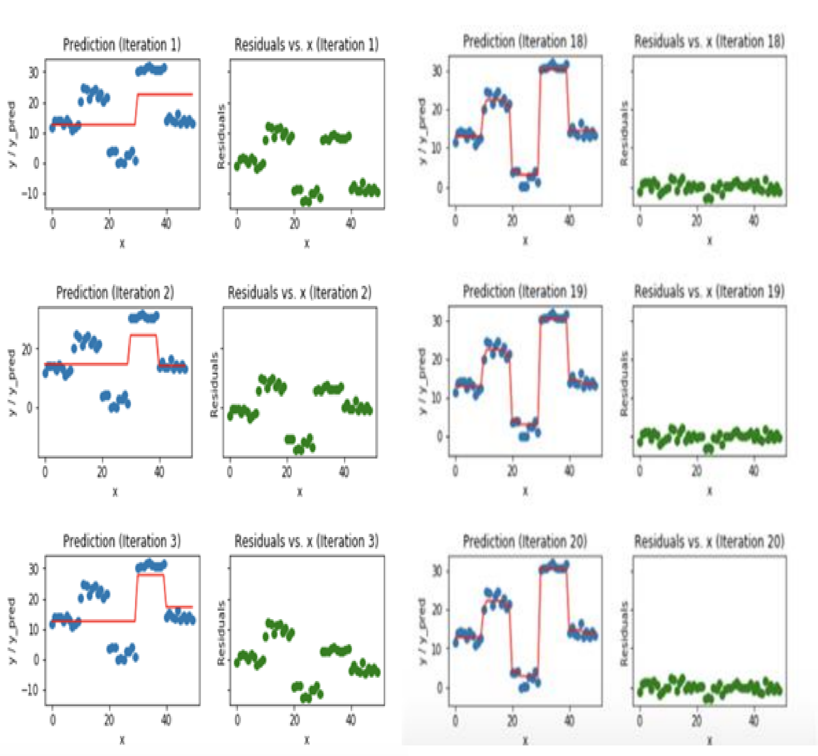
$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

- Does the sensor have 50% or more data present? — No → **Level 1**: Sensors that are not working — Yes ↓
- Does the sensor have a $-0.8 < r^2 < 0.8$ compared to a straight line graph? — No → **Level 2**: Sensors that exemplifies a straight line graph — Yes ↓
- Are the sensors in the top 95% of average range = (max - min) / number of maintenance periods? — No → **Level 3**: Sensors that have unresponsive signals — Yes ↓
- Are the sensors in the top 15 in terms of absolute average r^2? — No → **Level 4**: Sensors that do not follow a consistent trend — Yes ↓
- Does the sensor have $r^2 < 0.9$ between two max or min points within each maintenance period? — No → **Level 5**: Sensors that show a slight irregular trend — Yes ↓

Level 6
Most useful sensors ranked based on the **days** that the shift of y value that occurs prematurely before maintenance and **average r^2 values** of the best fit line within maintenance periods

Predictive Maintenance Methodology (PMM)

Gradient Boosting Regressor



How Gradient Boosting works

1. Forms an ensemble of individual regression models that together create a final model with the least prediction error, by first fitting an initial regression model to the sensor data in Iteration 1
1. In Iteration 2, a second regression model is built in that aims on accurately predicting situations where the first model predicts poorly
1. By Iteration 20, the residuals are distributed around 0, which show that the predictions are very close to true values, as compared to the first 3 iterations
1. This process is repeated with the final prediction model built correcting the prediction flaws of the combined boosted ensemble of all previous models

Strengths of Gradient Boosting

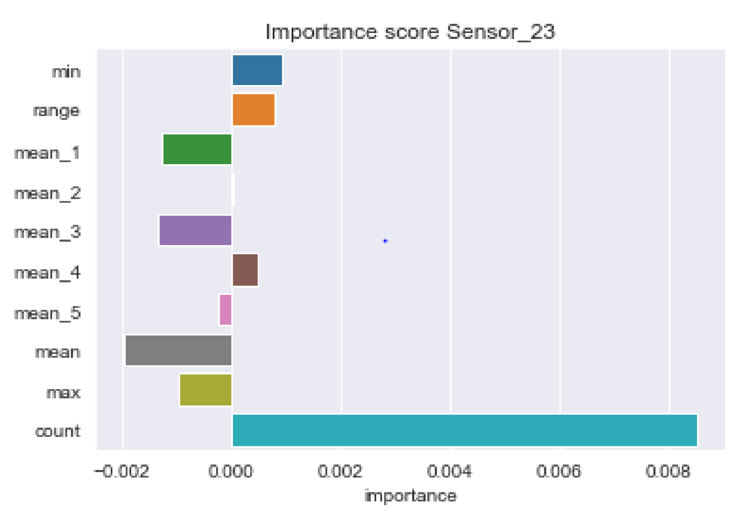
- Can optimize on different loss functions and provides several hyperparameter tuning options
- No data pre-processing is kept at minimal since it often works well with categorical and numerical values as it is

Performance Metrics of Predictive Models

- Mean Absolute Error is the amount of error between the actual and predicted values
- Mean Square Error measures the average of the squares of the error
- R-square value is a statistical measure of how close the data are to the fitted regression line

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad \text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|x_i - x| \quad R^2 = \frac{SS_{\text{reg}}}{SS_{\text{tot}}} = \frac{SS_{\text{reg}}/n}{SS_{\text{tot}}/n}$$

Steps in Feature Engineering



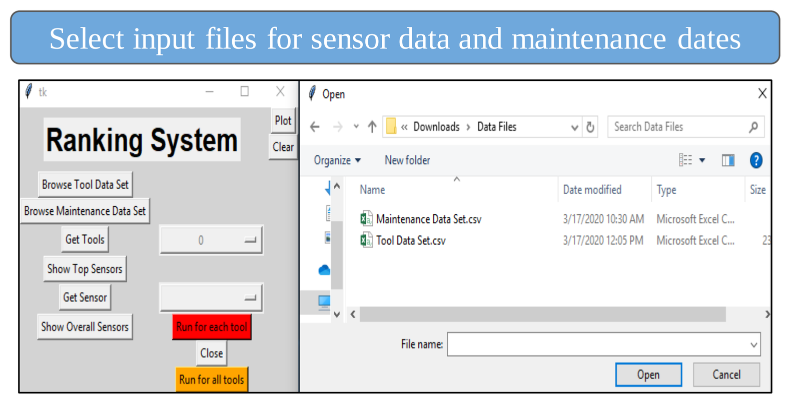
1. Fit Gradient Boosting Model using all variables and calculate its model.score() from the sklearn library. This will be the benchmark score
1. Drop columns one by one and calculate the score of the model. The difference between the score and the benchmark score will be calculated
1. Features that produces a negative difference value will be deemed as unimportant features and will be excluded from the analysis

Strengths of Feature Selection

- Selecting the most optimal features will lead to better accuracy
- Increase in prediction model flexibility

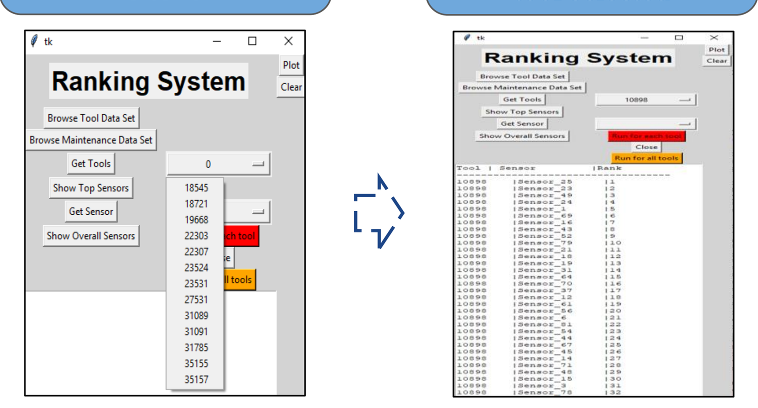
Graphic User-Interface (GUI)

Select input files for sensor data and maintenance dates

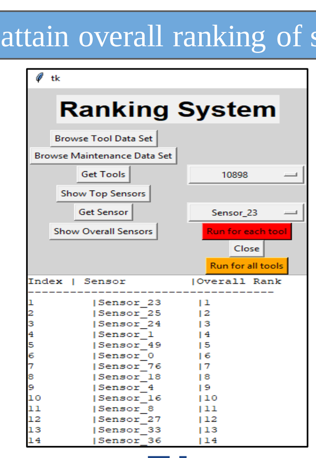


Select a tool whose sensors the user would like to rank

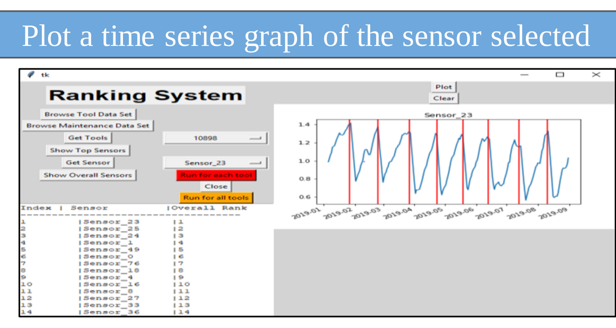
Run the program to rank the sensors of tool selected



Run model to attain overall ranking of sensors across



Plot a time series graph of the sensor selected



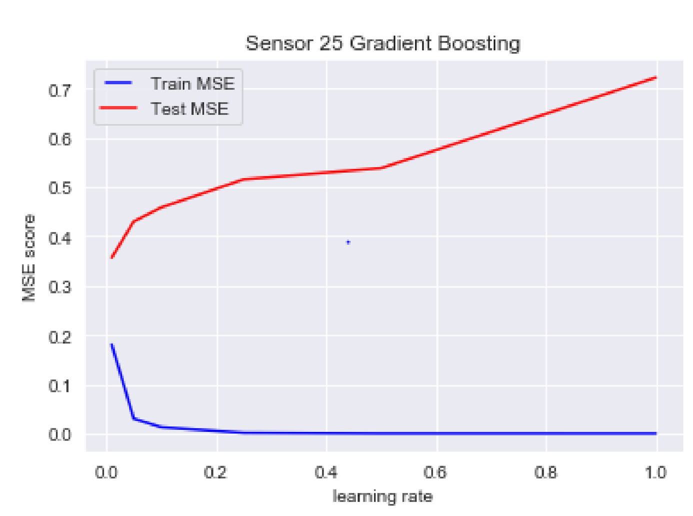
SRS Findings

Sensor 159 & 168
Both of these sensors were only used in tools 1 & 12. Hence, these sensors are omitted from our analysis as they are not used in all 12 tools

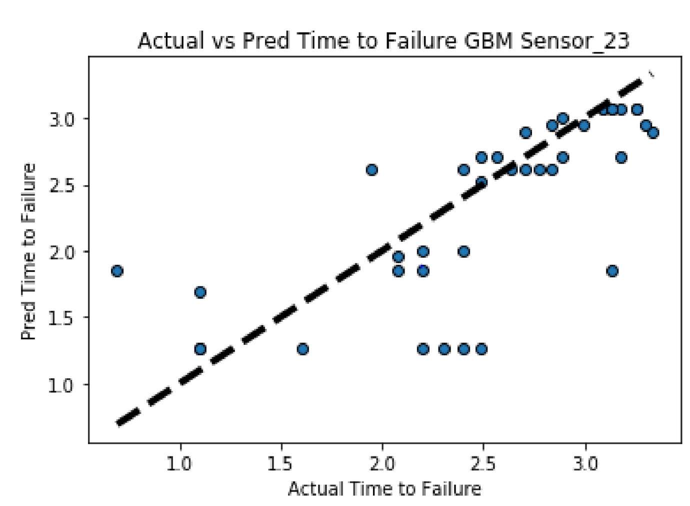
Tools 11 & 12
Tools 11 only started in February and tool 12 only started in March. Hence they have insufficient data to conduct thorough analysis

Sensor 8
Sensor 8 measured the time of the maintenance period. Hence, this sensor is omitted from our analysis as it is irrelevant even though its data fits the criteria of a good sensor

Validation Curve for Hyperparameter Optimization



Prediction Maintenance Model of Sensor 23



Limitations

Sensor Ranking System
SRS only conducts analysis on sensors with 50% or more data present, and on tools with at least two maintenance dates that are at least 10 days apart

Predictive Maintenance Model
The performance indicators that the sensors measure, such as pressure, temperature, gas flow and RF(radio-frequency) power, were not specified by Micron, due to the sensitivity of releasing such information. Hence, the model might have limited predicting capabilities

Graphic User Interface
Due to the time complexity of our algorithm, the SRS GUI requires a longer time to output the SRS results to the end users

RESULTS & ACHIEVEMENTS

Gradient Boosting Model Results

Sensor Name	R^2	MSE	MAE
1 (Micron's)	0.17	0.74	0.86
23	0.58	0.28	0.39
24	0.53	0.29	0.35
25	0.57	0.29	0.37

Recommendations

Sensor Ranking System
Conduct further research to find other ways of including data from all sensors and tools

Predictive Maintenance Model
Micron could provide the information of performance indicators that the sensors measure as pseudonyms

Graphic User Interface
A device with greater processing capabilities could be used

Achievements

Increased Accuracy in Maintenance Date Prediction

High Compatibility

Cost Savings

Skillset

Data Analytics:
Python programming, Big Data Pre-processing, Time Series Data Modelling & Analytics

Project Management:
Problem Solving, Project Management, Time Management, Framework Development, Scope Management

Systems Engineering:
Random Forest Feature Selection, Support Vector Machine, Gradient Boosting, Moving Average 5 (MA5), Grid Search, Cross- Validation

Future Work Exploration

- The proposed solution can be adapted into other industries that uses performance measurement sensors to set accurate predictive maintenance dates
- More research could be done on identifying other performance indicators of a tool that could be monitored by the sensors to provide a more holistic prediction model