

# **Bosch Production Line Performance**

## **Capstone 1 – Proposal**

### **Tom Widdows**

#### **Introduction**

The objective of this project is to identify internal production problems at a Bosch manufacturing facility. Identifying these problems would allow Bosch to reduce waste, lower costs and increase profitability.

#### **Problem**

Internal failures on the Bosch production line costs Bosch time, money and quality. Because of the complexity of the production process, the underlying reason for a part failure is difficult to determine allowing defective parts to continue through the production process only to be scrapped. Identifying the source also allows Bosch to take proactive steps to eliminate existing production problems.

#### **Data**

The data for this project represents anonymized measurements of a part as it moves through production line(s). Each feature heading is coded to include the production line, the station on that line and a feature number. For example, feature column header 'L1\_S17\_F4548' would represent line 1, station 17 and feature 4548. The raw data has been separated into three comma delimited files representing numerical, date and categorical information.

Client / Source: Bosch / <https://www.kaggle.com/c/bosch-production-line-performance>

#### **Exploring Data Analysis (EDA)**

Because of the vast number of features in the Bosch dataset, my initial focus will be on feature reduction. I will also focus on feature engineering (consolidation) allowing removing other features ultimately resulting in feature reduction.

I will also examine a parts path through the production process and evaluate if there is any correlation with part failure.

#### **Solution**

The project seeks to reduce costs and improve quality by identifying product failures as close to the point of failure as possible. By utilizing the vast amount of measurement data at each stage along the production process, I should not only be able to identify the point of failure quickly, I should have a head start in correcting the underlying problem.

#### **Evaluation of the model**

The model will be evaluated using the Matthews correlation coefficient (MCC). The MCC is calculated directly from the confusion matrix using the formula:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

In this equation, TP is the number of true positives, TN the number of true negatives, FP the number of false positives and FN the number of false negatives.

[https://en.wikipedia.org/wiki/Matthews\\_correlation\\_coefficient](https://en.wikipedia.org/wiki/Matthews_correlation_coefficient)

### **Deliverables**

A model to predict internal failures on the Bosch production line presented in a Jupyter Notebook and supplemented with a slide presentation if required.