Business Analytics Project

Aurubis Yield Project Executive Summary

1. Project Vision and Objectives

The objective of this analytical project is to improve the yield (defined as the percentage of a product's cast weight that is converted into a finished product) and the accuracy of predicted yield, in order to achieve higher scheduling efficiency, on-time delivery performance, as well as lower production cost and less inventory in the plant. This summary presents the statistics and visualization results from both descriptive and predictive analysis on yield patterns and trends. Taking both coil history data and order data into the scale, we were devoted to performing analysis on the bigger picture to answer strategic questions for the business.

2. Summary of Findings and Recommendations

In the descriptive analysis, we first explored the distribution of yield per step and number of operations for each coil, then looked into the yield by operations, by status code, and by combination of start gauge and width.

Yields per step range mostly from 80% to 100%. To avoid bias, those around 10% to 20% are considered as outliers. Identifying coils by *INVID#*, we found that most coils go through two to six operations. The average yields of most operations are higher than 80%, as mentioned. Compared to other operations, HT, SW, and WP have relatively lower yields. Among all these operations, the most popular two are Rolling (**CR**) and Annealing (**AN**). In detail, coils lose 2% to 7% of weight when going through Rolling machines, among which the most efficient one is machine 41 and the least efficient one is 40. Going through Annealing machines, coils lose 1% to 6% of weight, with the most efficient machine being 147 and the least being 133.

By visualizing the yield by status code, we identified four statuses that typically show yields lower than average 88%, which are BIRTH, QHOD, SCRAP, and CHECK7. If we only focus on the observations with yields lower than 88%, the yields are lower at PRIME, CHECK, BIRTH, and E-WRK status. Considering the combination of start gauge and width level, the yields are the highest when start width is between 30 and 40 with gauge between 0.4 to 0.6, while yields are the lowest when start gauge is between 0 and 0.2, indicating that gauge becomes low and yield tends to be low at the final stages of production.

In the predictive analysis, we built machine learning models including Lasso Regression, Ridge Regression, and Light GBM based on the data with feature engineering. Employing mean absolute error (MAE) as the criteria of predictive model performance, we found that **Light GBM** reports the **lowest MAE (3.7769)**, then Lasso Regression (MAE = 4.0069) and Ridge Regression (MAE

= 4.1018). We identified the following features that have positive or negative impact on the coil-level yield by combining the coefficients and feature importance of three models.

Category	Machines		Operations		Alloys		Other	
Impact on yield	+	-	+	-	+	-	+	-
Items	133	46, 47, 74, 134		AN	122	KLF5, 7036, 1102	INSWID	slits

Based on the above analysis, we would recommend that Aurubis could use **machine 133** heavier rather than 46, 47, 74, and 134 to reduce the weight loss of coils and increase the yield. For operations, **avoiding performing Annealing** or optimizing the Annealing operation may contribute to generating higher yield. Choosing alloy **122** rather than KLF5, 7036, or 1102 will also help, hence we recommend using 122 if possible, but it still depends on clients' requirement of order. Apart from these three variable categories, **using coils with greater width** and cutting down the total slits also significantly lead to higher yield. In addition, based on our data pre-processing and feature engineering in this project, we recommend that Aurubis could optimize the operational system which records business operation and transaction, to get cleaner data for any further use.

3. High-level Overview of Data used, Methodology, and Techniques

After data preprocessing, exploratory data analysis, and feature engineering. We developed the final data used for the model. The goal for our prediction is to predict yield at the coil level. We calculated important features at the coil-level, such as width, gauge, length, alloy type, number of slits, number of machine counts and number of operation counts. To improve the yield and the accuracy of predicted yield, we did descriptive analysis and predictive analysis. In the descriptive analysis part, we conducted exploratory data analysis to extract insights from the date at the operation level. After feature engineering, we did predictive analysis at the coil level, generating three models, including Lasso Regression, Ridge Regression, and Light GBM. and compared the mean absolute error(MAE) for each model. We generate findings and recommendations from both the descriptive and predictive analyses.

(Please see the interactive interface App via: Aurubis Buffalo Yield Prediction App
Please see the script creating the App via: Script for Aurubis Buffalo Yield Prediction App)