CIS465
Business Analytics Project

Aurubis

Capstone Project
Final Presentation

Yield Project

Team #18
MSBA 2020, Simon Business School

Agenda

1 Key Findings and Recommendations

1. Key Findings from Descriptive Analysis

Distribution of yields

- 80% - 100%

Yields by Operations

- 2– 6 operations per coil
- Popular operations: Rolling (CR), Annealing (AN)
- Low yield operations: HT, SW, WP

Yields by Statuses

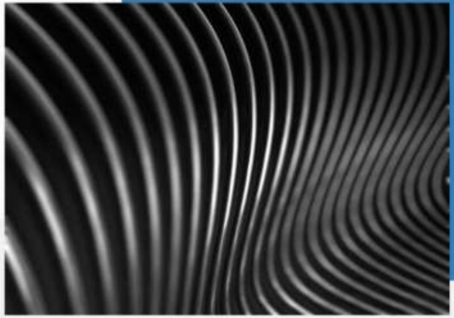
- Average yield: 88%
- Low yield statuses: PRIME, CHECK, BIRTH, E-WRK

Yields by Start Width and Gauge

- High yield combination: width 30 – 40, gauge 0.4 – 0.6
- Low yield combination: gauge 0 – 0.2

Aurubis

1. Key Findings from Predictive Analysis



Best machine learning model:

- Light GBM
- Mean Absolute Error (MAE) = 3.7769

Important features and impacts on yield:

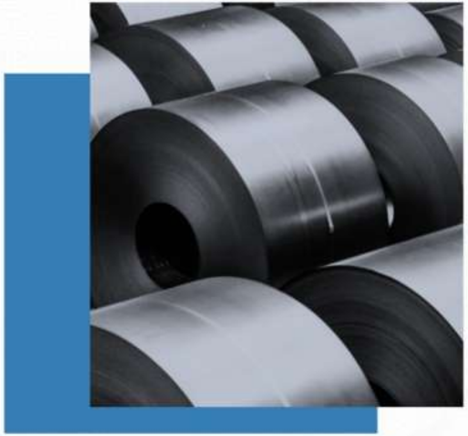
Category	Machines		Operations		Alloys		Others	
Impact on yield	+	-	+	-	+	-	+	-
Items	133	46, 47, 74, 134		AN	122	KLF5, 7036, 1102	INSWID	slits

Aurubis

1. Recommendations

To improve yield:

- Machines: more **133**, less 46, 47, 74, 134
- Operations: avoid Annealing (AN)
- Alloys: more **122**, less KLF5, 7036, 1102, if possible
- Others: use coils with greater **start width**, reduce total slits



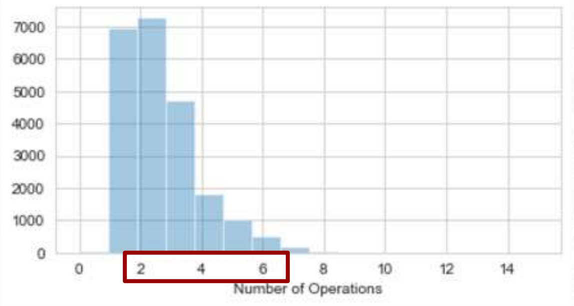
To accuracy of yield prediction:

- Optimize the operational system that records business operations and transactions
- Construct **data warehouse** to periodically get organized data for analysis

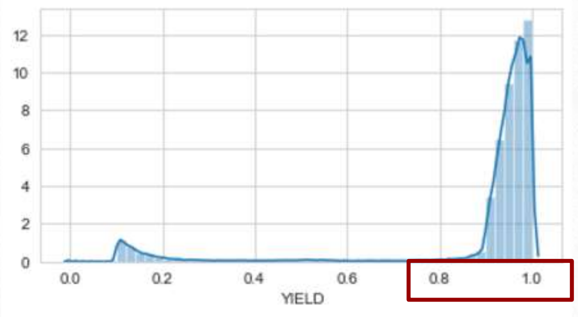
2 Exploratory Data Analysis

2. EDA - Operation & Yield

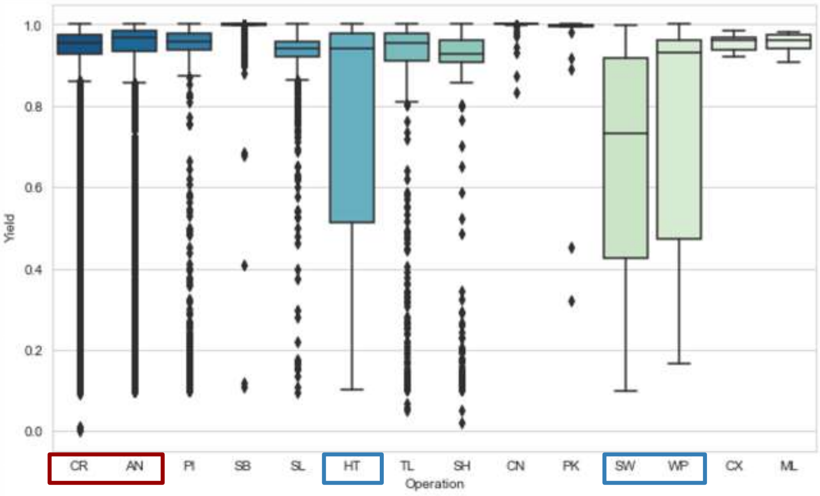
Most coils go through 2-6 operations



Yield by operation ranges mostly from 80% to 100%



Yield by Operations

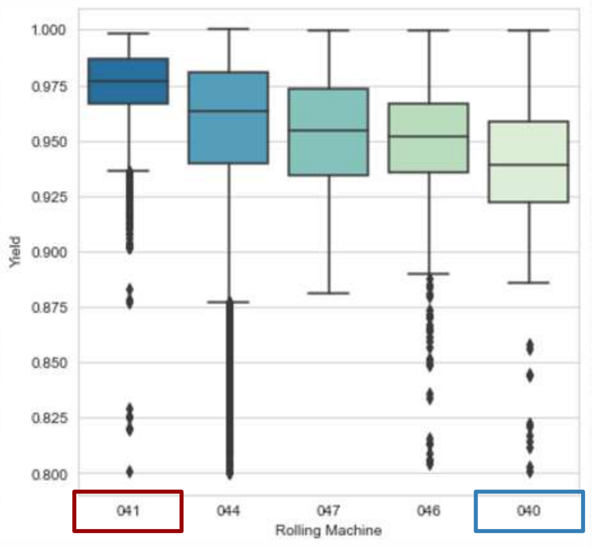


- Two most popular operations are **Rolling (CR)** and **Annealing (AN)**
- Most operations have **yield > 0.8**
- Yields of **HT**, **SW** and **WP** are lower than other operations

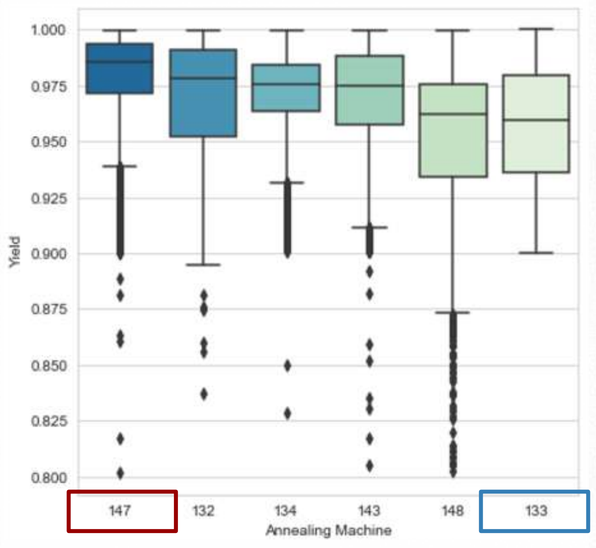
Aurubis

2. EDA - Rolling & Annealing

Performance of Rolling Machines



Performance of Annealing Machines

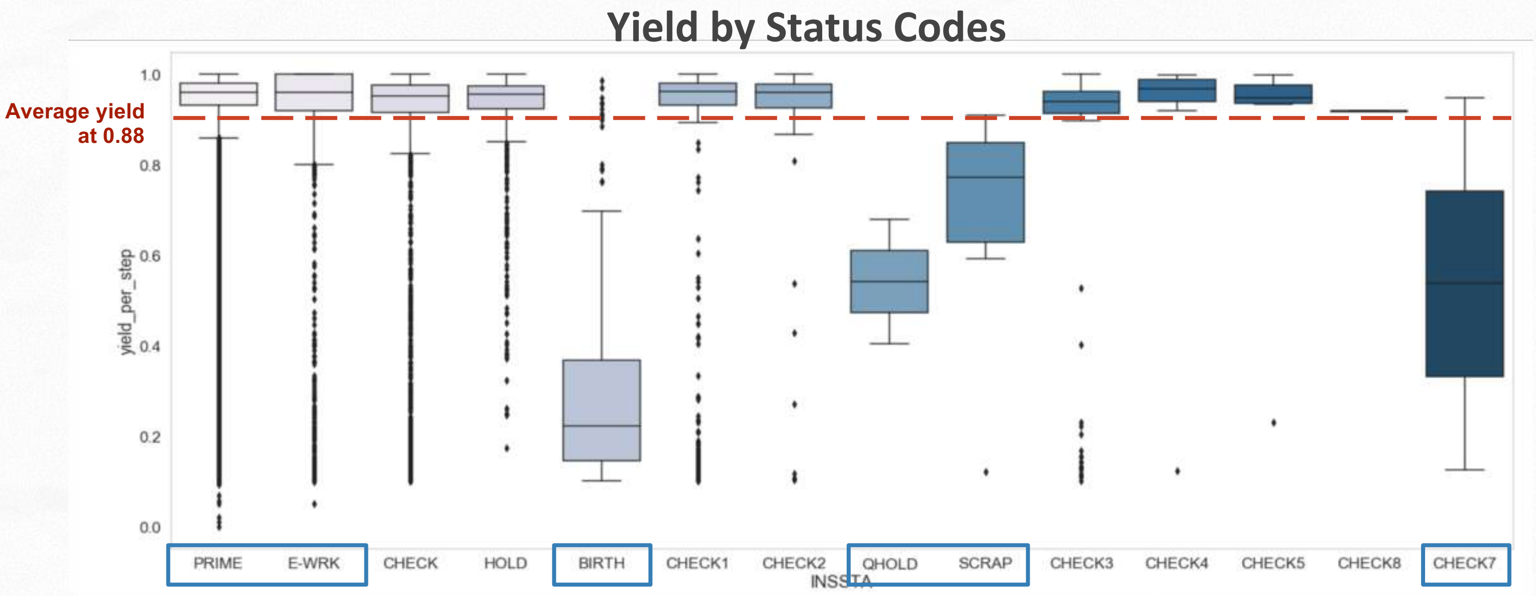


- Coils lose **2% - 7%** of weights when going through rolling machines
- Machine **41** is the **most** efficient
- Machine **40** is the **least** efficient

- Coils lose **1% - 6%** of weights when going through annealing operations.
- Machine **147** is the most efficient
- Machine **133** is the least efficient

Aurubis

2. EDA - Status Code and Yield

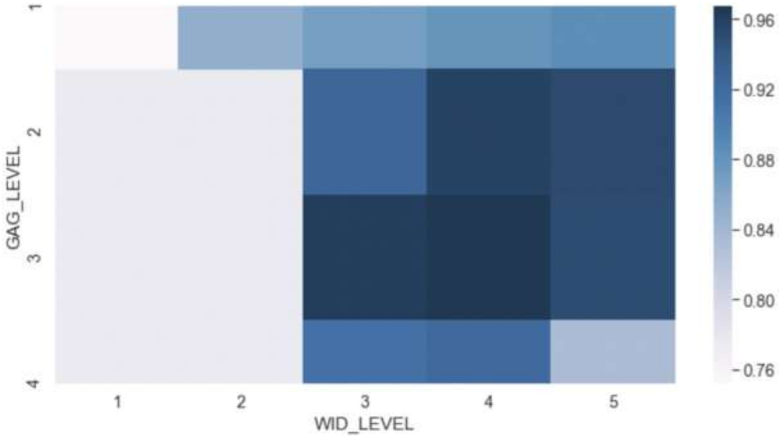


- Yield is typically lower than 0.88 when status code is **birth, hold, scrap and check7.**
- When only looking at rows where yield is lower than 0.88, **prime** and **E-WRK** also result in low yield of 0.4
    - Although prime and E-WRK are higher overall, there are many outliers in the dataset.

Aurubis

2. EDA - Width, Gauge and Yield

- Highest yield is generated at start width between **30 - 40,** start gauge between **0.4 - 0.6.**
- Lowest yield is generated at start gauge between **0 - 0.2**, indicating at the final stage of production, gauge becomes low and yield tends to be low.



Yield by Width and Gauge Level

Level	Width	Gauge
1	0-10	0-0.2
2	10-20	0.2-0.4
3	20-30	0.4-0.6
4	30-40	0.6-0.8
5	40-50	0.8-1

Width and Gauge Level Details

Aurubis

Aurubis

3

Feature Engineering

& Transform Data for Machine Learning

3. Feature Engineering

Step1: Calculate Yield at the coil level

- Create id_len, parent_id and child_id
- Calculate total_slits for each parent coil
- Use start weight and finish weight of each parent coil to calculate coil-level yield

	parent_id	start_weight	finish_weight	yield
0	420000	20674	18594	0.899391
1	420001	19938	16333	0.819189
2	420002	20235	17634	0.871460
3	420003	19664	18441	0.937805
4	420005	21788	20125	0.923674

Aurubis

3. Feature Engineering

Step2: Store expected feature at coil-level

1. Machine and Operations

- Get Machine and Operation dummies
- Get count of each dummy

  # of Machine Counts

  # of Operation Counts

2. Start Width, Length, Gauge and Standard Hour

	parent_id	SDOPCD_AN	SDOPCD_CN	SDOPCD_CR	SDOPCD_CX
0	420000	3	0	2	0
1	420001	2	0	1	0
2	420002	3	0	2	0
3	420003	2	0	2	0
4	420004	2	0	2	0

Aurubis

4 Predictive Analysis

4. Features and Models

Goal

Predict yield at coil level

Features (Coil-level)

- Standard hour
- Standard width
- Standard gauge
- Standard length
- Start width
- Start gauge
- Start length
- Alloy type
- # of slits
- # of machine counts
- # of operation counts

Models

- Lasso regression
- Ridge regression
- Light Gradient Boosting Machine

4. Result Comparison

<table>
<tr><th colspan="2"></th><th>Lasso Regression</th><th>Ridge Regression</th><th>Light GBM</th></tr>
<tr><td colspan="2">Mean Absolute Error</td><td>4.00691</td><td>4.01081</td><td>3.77688</td></tr>
<tr><td rowspan="4">Top Features that Decrease yield</td><td>Machines</td><td>46, 44, 74, 47</td><td>68 134, 132, 46, 74, 36, 47</td><td rowspan="8"><u>Machine:</u>
134, 143

<u>Operations:</u>
CR, AN, SL

<u>Alloys:</u>
1453, 260, 1102

<u>Other:</u>
INSWID, SDHOUR, SDLEN
SDGAG, total slits, SDWID</td></tr>
<tr><td>Operation</td><td>CR, AN, PI, SH</td><td>PI, AN, SB</td></tr>
<tr><td>Alloy</td><td>KLF5, 7036, 1102, 110</td><td>KLF5, 7036, 1102</td></tr>
<tr><td>Other</td><td>Total slits</td><td>Total slits</td></tr>
<tr><td rowspan="4">Top Features that Increase yield</td><td>Machines</td><td>133, 613</td><td>36, 27, 133, 612</td></tr>
<tr><td>Operation</td><td>-</td><td>CR</td></tr>
<tr><td>Alloy</td><td>122, 210, 260, 226</td><td>1453, 122</td></tr>
<tr><td>Other</td><td>INSWID, SDLEN, INSGAG</td><td>INSWID</td></tr>
</table>

Aurubis

4. Conclusion

- Use **machine 133** more, and use 46, 47, 74, and 134 less
- Optimize the **annealing** operation to generate higher yield
- Choose **alloy 122** rather than KLF5, 7036, and 1102 (But this will still depend on client's needs)
- Use coils with **higher width** and cut down number of slits

	Machines		Operations		Alloys		Other	
Impact on Yield	+	-	+	-	+	-	+	-
Features	133	46, 74, 47, 134	-	AN	122	KLF5, 7036, 1102	Start Width	Total Slits

Aurubis

5 Interactive Interface Demonstration

https://aurubis-buffalo-yield.herokuapp.com/

Q & A