Optimizing Plate Production:

An Analysis of Scale Defects and Recommendations on Improvement

Seon Young Jhang

With POSCO

Business Problem

Client Issue:
A recent surge in **scale defect occurrences** reported.

Root Cause:
Investigation revealed a significant increase in scale defects during the **rolling process** at OO Factory.

Objective:
- Collect and analyze relevant data.
- Identify the root cause of the defects.
- Propose optimal conditions and solutions to mitigate defect occurrences.

Overview of Datasets

[Data Description]

Data size: 720 rows x 21 columns
Date: Aug 1st – 2nd , 2008 (2 days)

Target Variable

- Scale: Indicates the presence of scale (oxidized iron) defects.

Some Feature Variables

- PLATE_NO: Product number.
- ROLLING_DATE: Rolling process timestamp.
- SPEC: Product specification.
- STEEL_KIND: Steel type (e.g., carbon, titanium).
- PT_THK: Target plate thickness.
- PT_LTH: Target plate length.
- HSB: Hot Scale Breaker application status.
- FUR_NO: Furnace number.
- FUR_HZ_TEMP: Heating zone material temperature.
- FUR_SZ_TIME: Soaking zone time (minutes).

[Summary Statistics for 'SCALE']

	SCALE
Non-Defective	489 *(68%)*
Defective	231 *(32%)*

[Null Values]

```
df.isnull().sum()
```

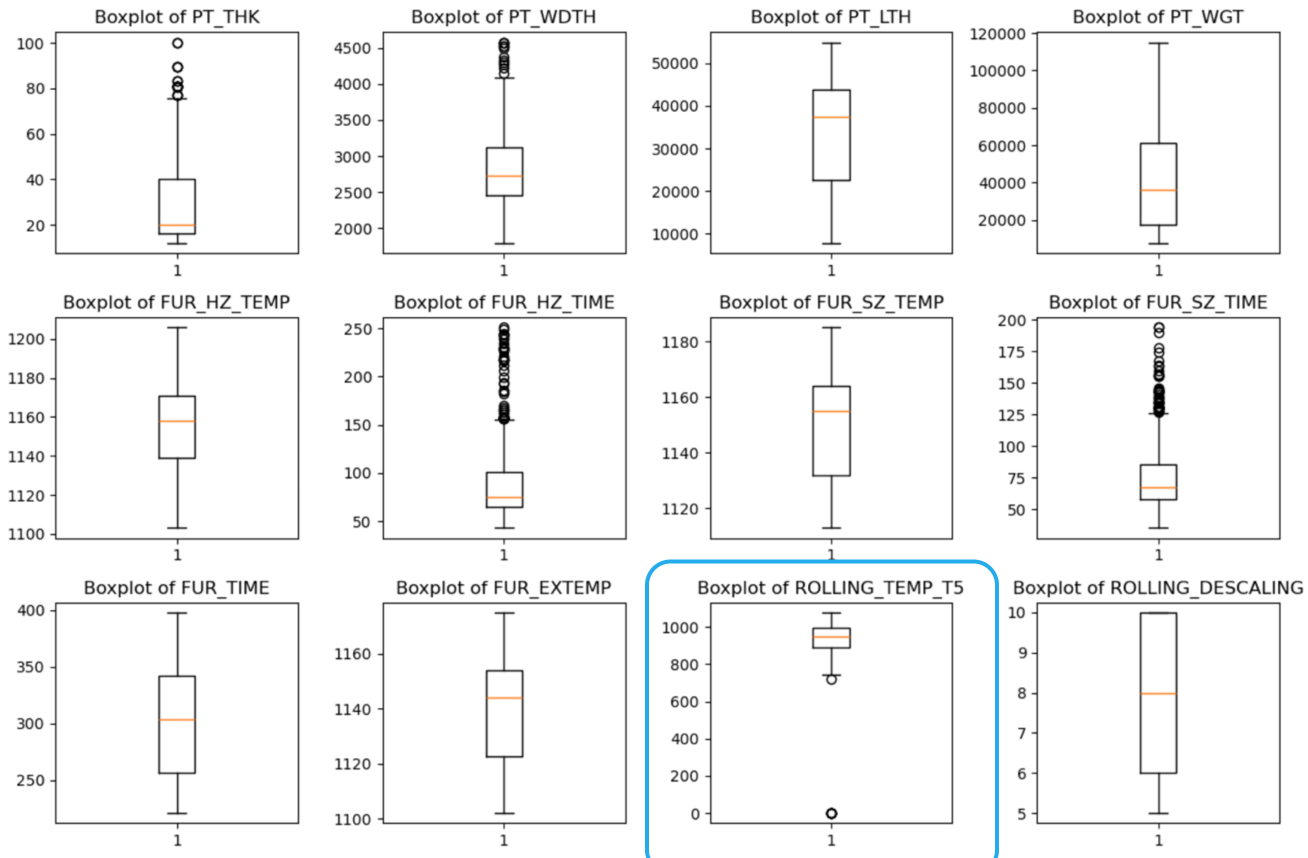
```
PLATE_NO            0
ROLLING_DATE        0
SCALE               0
SPEC                0
STEEL_KIND          0
PT_THK              0
PT_WDTH             0
PT_LTH              0
PT_WGT              0
FUR_NO              0
FUR_NO_ROW          0
FUR_HZ_TEMP         0
FUR_HZ_TIME         0
FUR_SZ_TEMP         0
FUR_SZ_TIME         0
FUR_TIME            0
FUR_EXTEMP          0
ROLLING_TEMP_T5     0
HSB                 0
ROLLING_DESCALING   0
WORK_GR             0
dtype: int64
```

There are no
Null values for this dataset

With POSCO

Overview of Datasets

[Outliers]



In the context of rolling process, a temperature of zero is physically impossible as the rolling temperature refers to the heat level at which the steel plate is processed.

```
df['ROLLING_TEMP_T5'].describe()
```

```
count       720.000000
mean        933.920833
std         107.863887
min           0.000000
25%         889.750000
50%         951.000000
75%         994.250000
max        1078.000000
Name: ROLLING_TEMP_T5, dtype: float64
```

```
df[df['ROLLING_TEMP_T5']==0]
```

	PLATE_NO	ROLLING_DATE	SCALE	SPEC	STEEL_KIND	PT_THK	PT_WDTH	PT_LTH	PT_WGT	FUR_NO	...	FUR_HZ_TEMP	FUR_HZ_TIME	FUR_SZ
41	PB562815	2008-08-01:02:23:09	0	GL-E36-TM	T8	55.51	3765	11398	37400	3	...	1132	95	
184	PB562958	2008-08-01:10:00:06	0	JS-SM490YB	C0	16.09	2221	43596	61150	3	...	1169	64	
185	PB562959	2008-08-01:09:94:29	0	JS-SM490YB	C0	16.09	2221	43596	61150	3	...	1163	57	
494	PB563268	2008-08-02:01:10:40	0	PILAC-BT33	T5	80.82	2085	12919	17090	2	...	1133	89	
495	PB563269	2008-08-02:01:10:36	0	PILAC-BT33	T5	80.82	2085	13176	17430	1	...	1130	92	
496	PB563270	2008-08-02:01:10:23	0	NV-D32-TM	T0	40.35	2497	23592	37320	3	...	1119	109	

6 rows × 21 columns

There are six rows that have 'ROLLING_TEMP_T5' as '0'

Replace 0 values with the mean value, grouped by 'SPEC'

Data Visualization

[Histogram]

Histogram plots show the conditions under which a plate is more likely to be defective. The analysis revealed that smaller plate widths, higher furnace temperatures, higher furnace soak temperatures, and higher furnace extract temperatures were associated with a higher likelihood of defects.

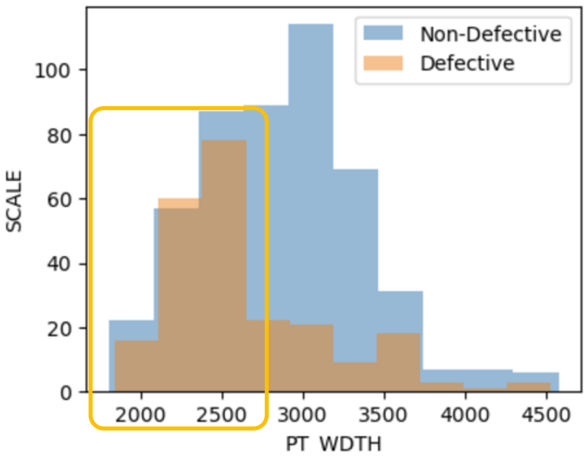
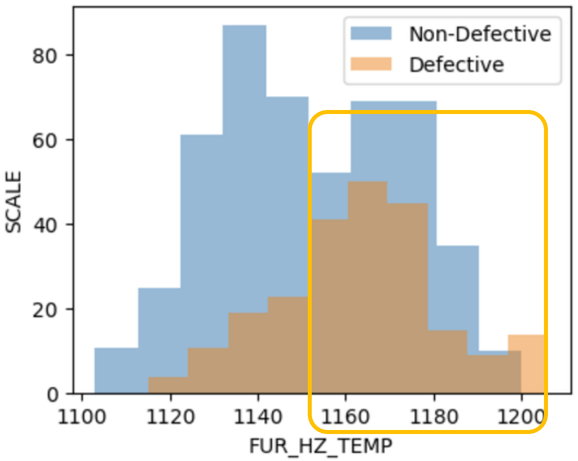
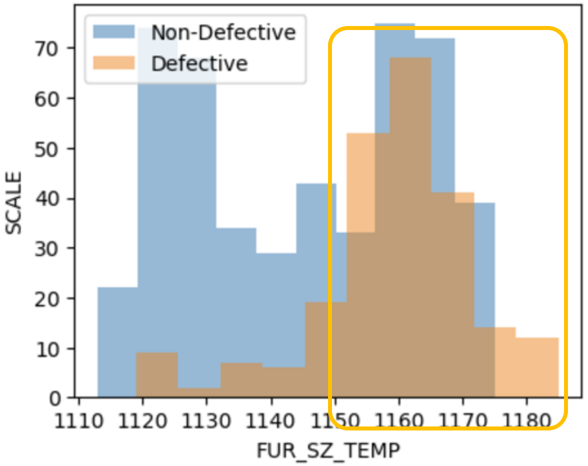


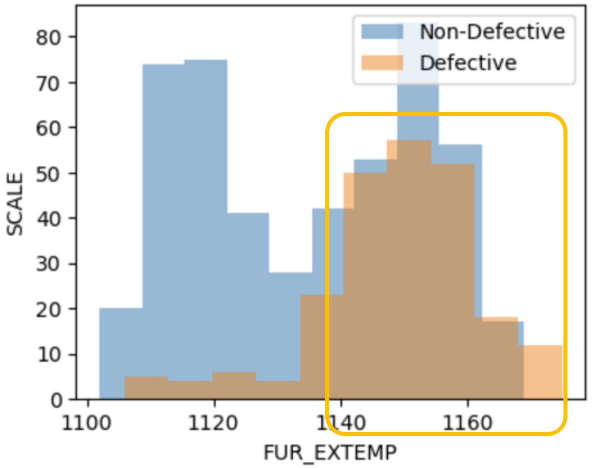
Plate Width



Material temperature in the heating zone of the furnace (°C)



Material temperature in the soaking zone of the furnace (°C)



Calculated furnace exit temperature

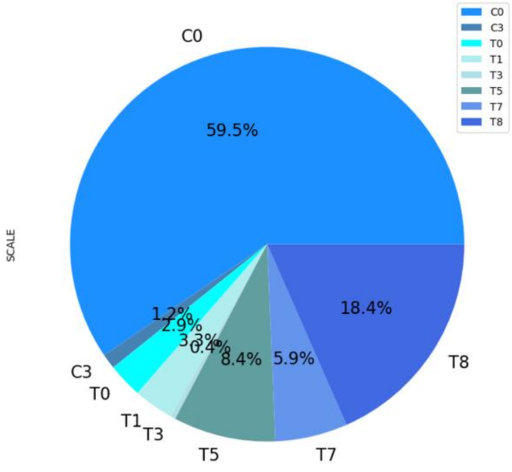
Data Visualization

[Pie Chart]

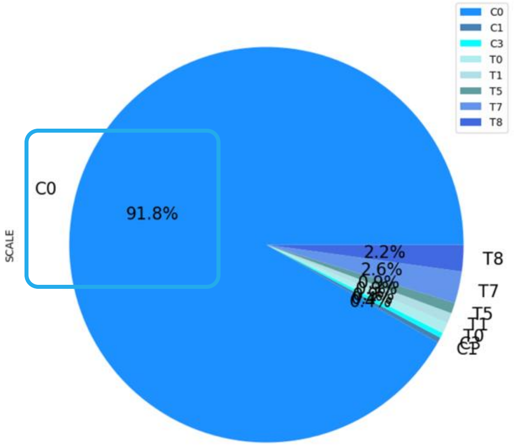
Steel Type

Although the absolute number of CO plates is the highest, the majority of defective plates among the steel types were of the CO type.

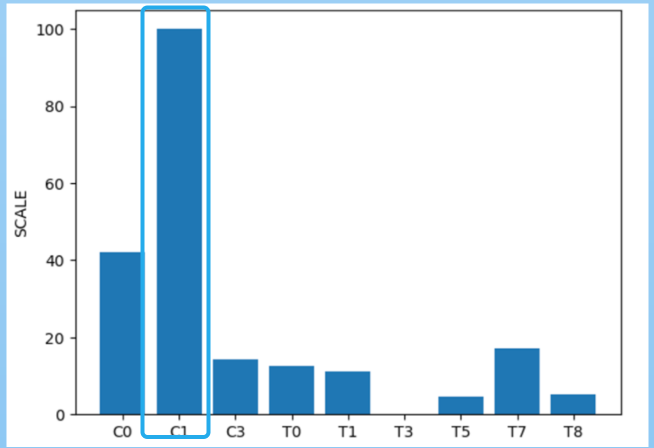
Non-Defective



Defective



	STEEL_KIND	SCALE_PROB
0	C0	42.147117
1	C1	100.000000
2	C3	14.285714
3	T0	12.500000
4	T1	11.111111
5	T3	0.000000
6	T5	4.651163
7	T7	17.142857
8	T8	5.263158

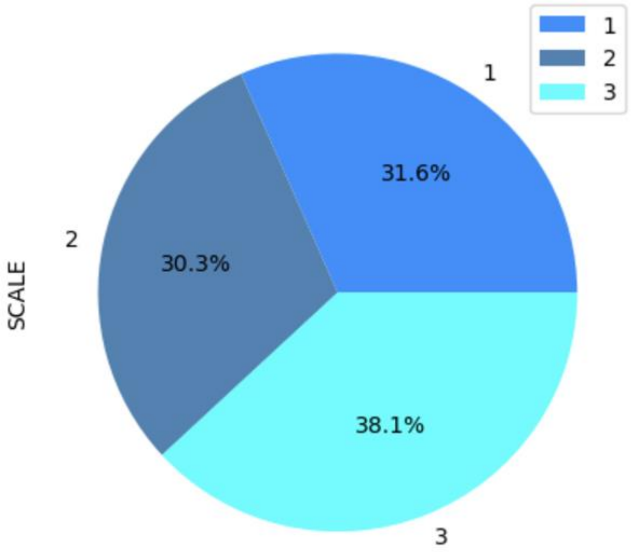


Evaluating the defect rate in relative terms rather than absolute numbers. Above table shows that CO has 42.1% of defect rate, and T7 also has a high percentage of 17.1%.
While C1 shows a 100% defect rate, there is no significant concern since it is based on just one case.

Data Visualization

[Pie Chart]

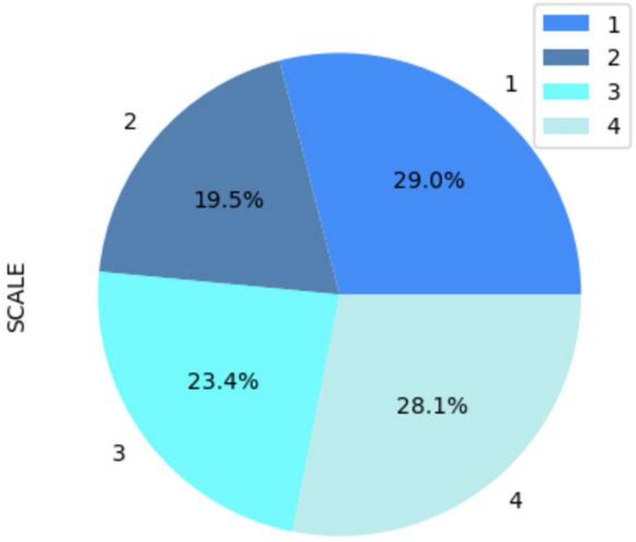
Furness Number



	FUR_NO	FUR_PROB
0	1	30.416667
1	2	29.535865
2	3	36.213992

There was no significant difference observed between different furnace numbers.

Working Group



	WORK_GR	WG_PROB
0	1	35.449735
1	2	27.272727
2	3	31.395349
3	4	33.505155

There was no significant difference observed between different working groups.

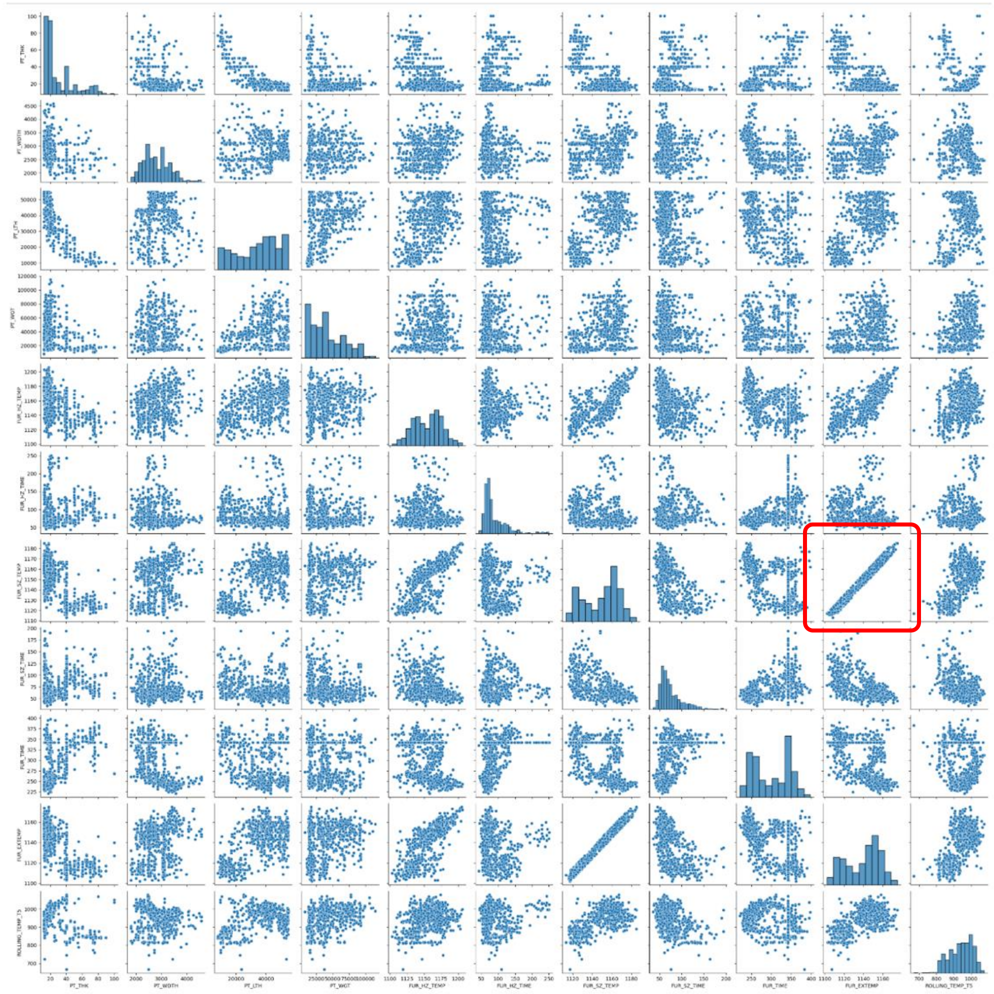
ANOVA Analysis

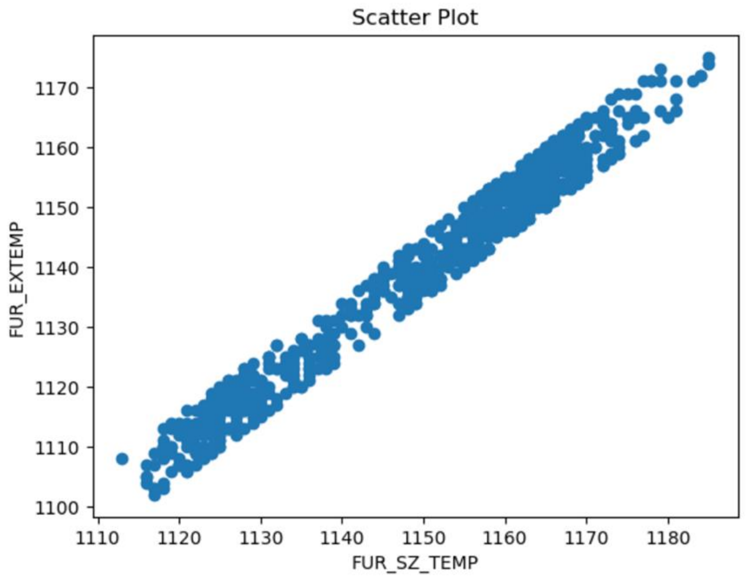
```
F-statistic: 0.9827921566698729
P-value: 0.40028114545320315
There is no statistically significant difference between WORK_GR groups.
```

With POSCO

Data Visualization

[Pair Plots]





In the heatmap, the correlation between FUR_SZ_TEMP and FUR_EXTEMP is observed to be 0.99 (through Heatmap analysis). Given the potential for multicollinearity, it would be wise to remove one of these variables to avoid issues with variance inflation (VIF).

Data Visualization

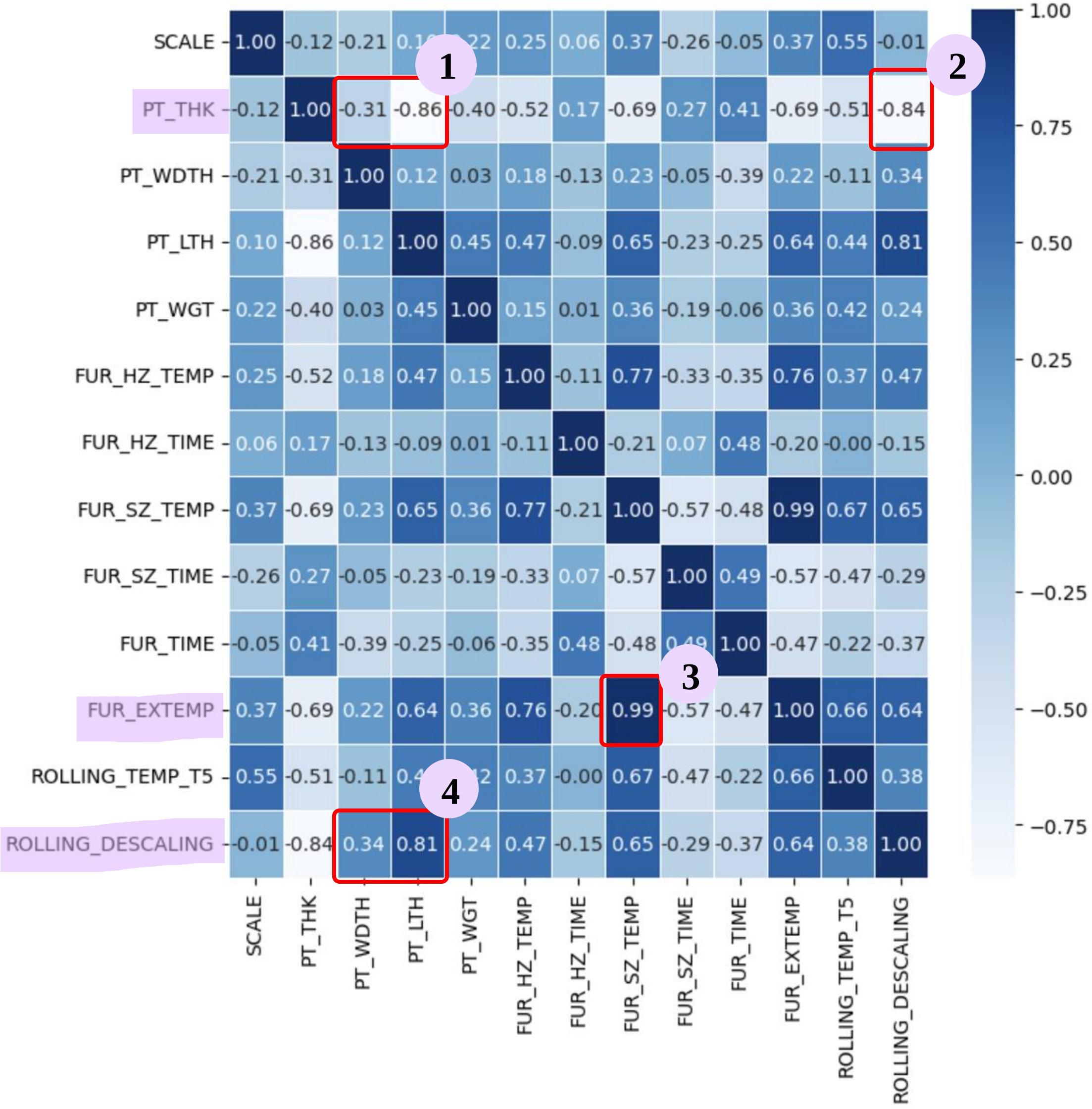
[Heatmap]

1. A negative correlation is observed between *PT_THK* (plate thickness) and both *PT_LTH* and *PT_WGT*. This implies that as the thickness of the steel increases, the length and width tend to decrease.

2. A negative correlation is observed between *PT_THK* and *ROLLING_DESCALING*. This indicates that thicker steel generally requires fewer descaling operations.

3. Based on these findings, we decided to remove one of the two variables: *FUR_SZ_TEMP* and *FUR_EXTEMP*. Since they show a perfect positive linear relationship, retaining both in the model could cause multicollinearity issues.

4. Conversely, a strong positive correlation is observed between *PT_WDTH* and *ROLLING_DESCALING*, as well as *PT_LTH* and *ROLLING_DESCALING*. This suggests that wider or longer products may require more descaling operations.

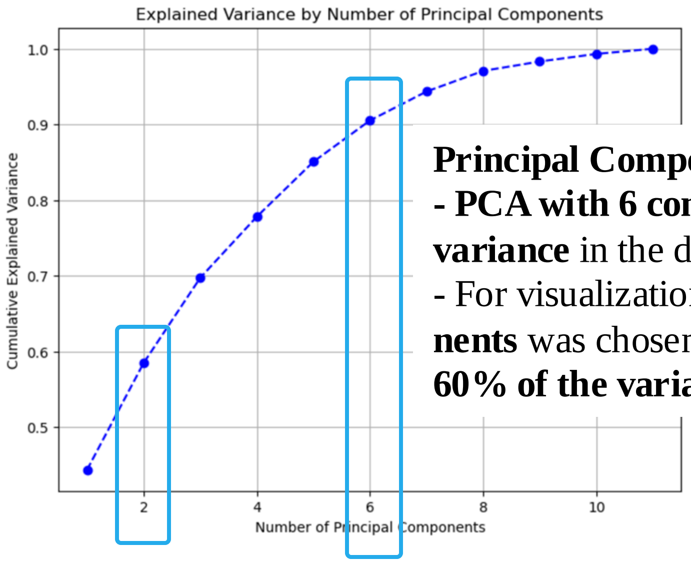


[Machine Learning] Clustering (K-Means)

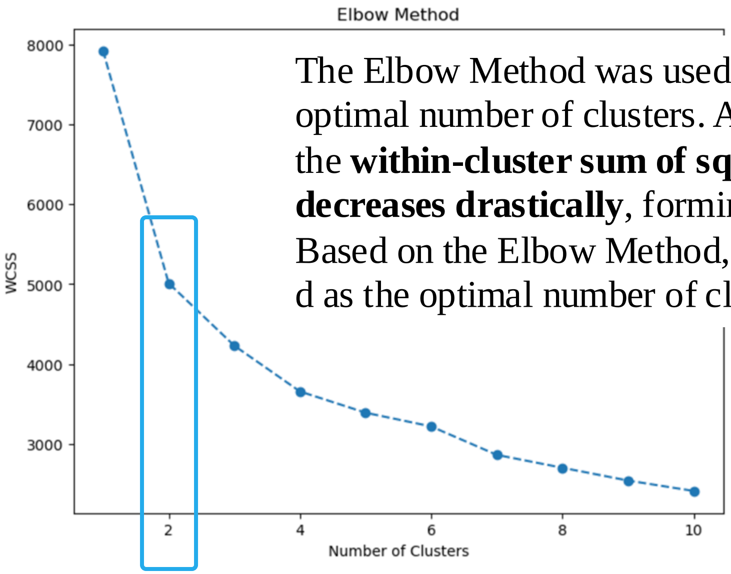
[Clustering] K-Means

Select Relevant Features
The following features were selected from the dataset:
PT_THK, PT_WDTH, PT_LTH, PT_WGT, FUR_HZ_TEMP, FUR_HZ_TIME, FUR_SZ_TEMP, FUR_SZ_TIME, FUR_TIME, ROLLING_TEMP_T5, ROLLING_DESCALING

Scaling
Scaled the selected features to a range of 0 - 1 to ensure equal contribution to analysis



Principal Component Analysis (PCA):
- PCA with 6 components explains 90% of the variance in the data, reducing dimensionality.
- For visualization purposes, **PCA with 2 compo nents** was chosen, which explains approximately **60% of the variance** in the data.



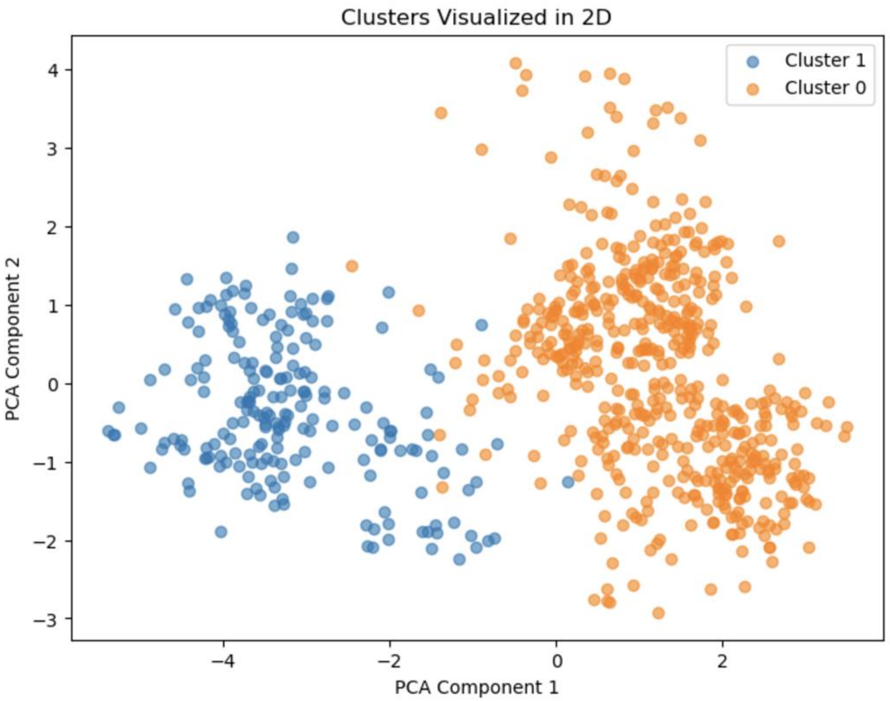
The Elbow Method was used to determine the optimal number of clusters. As shown in the graph, the **within-cluster sum of squares (WCSS) decreases drastically**, forming an 'elbow' shape. Based on the Elbow Method, **2 clusters** was selecte d as the optimal number of clusters.

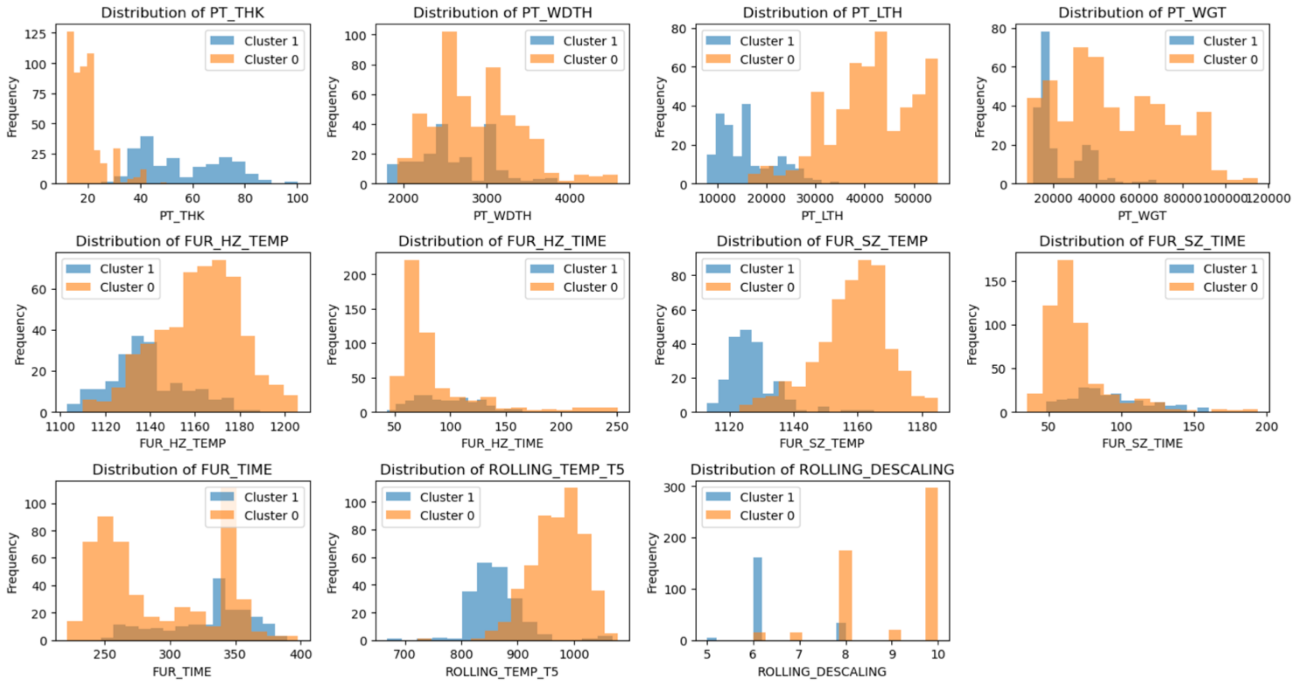
[Machine Learning] Clustering (K-Means)

[Clustering] K-Means

Cluster 0
Count: 520 (out of 720, *72%*)
Non-Defective : 59%
Defective : 41%

Cluster 1
Count: 200 (out of 720, *28%*)
Non-Defective: 92%
Defective: 8%





Implications of K-means Clustering

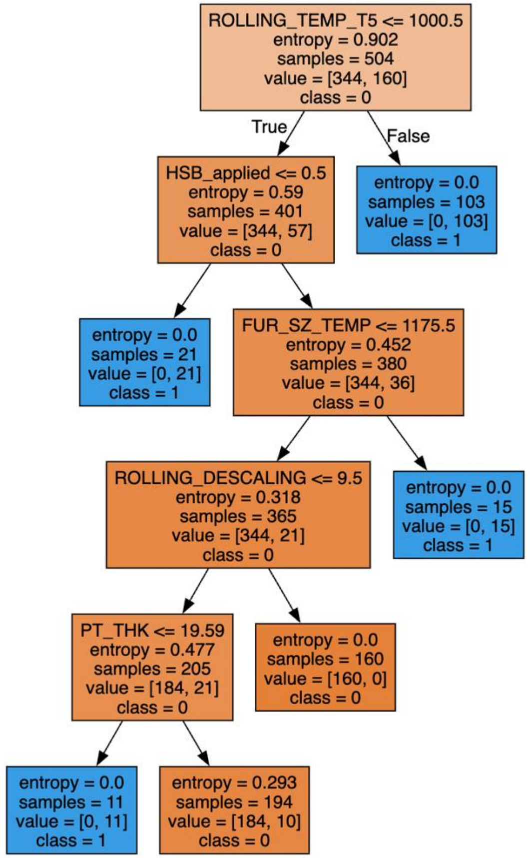
Clusters Represent Different Process Conditions:

- K-means clustering groups the data into clusters based on similarities in the features.
- The two clusters identified (Cluster 0: 58.85, Cluster 1: 91.50) likely represent two distinct sets of conditions in the rolling process that influence the occurrence of scale defects.

[Machine Learning] Decision Tree

[Decision Tree]

Decision Tree Graph



Hyperparameter Tuning

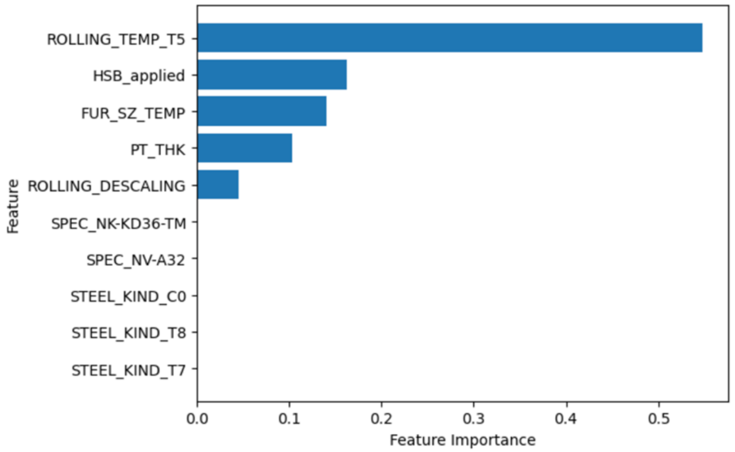
Used grid search to identify the optimal combination

Criterion: Entropy
Max Depth: 5
Min Samples per Leaf: 10

Performance

Train Accuracy	Test Accuracy
98%	99%

Feature Importance



Among the features, ROLLING_TEMP_T5 has the highest importance, with a value greater than 0.5. It is followed by HSB_applied, FUR_SZ_TEMP, and PT_THK.

With POSCO

[Machine Learning] SVM

[SVM]

```
svm_uncustomized = SVC(random_state = 1234)
svm_uncustomized.fit(df_train_x, df_train_y)

print("Accuracy on training set: {:.3f}".format(svm_uncustomized.score(df_train_x, df_train_y)))
print("Accuracy on test set: {:.3f}".format(svm_uncustomized.score(df_test_x, df_test_y)))
```

```
Accuracy on training set: 0.683
Accuracy on test set: 0.671
```

After Scaling

```
svm_scaled = SVC(random_state=1234)
svm_final = svm_scaled.fit(df_scaled_train_x, df_train_y)
print("Accuracy on training set: {:.3f}".format(svm_scaled.score(df_scaled_train_x, df_train_y)))
print("Accuracy on test set: {:.3f}".format(svm_scaled.score(df_scaled_test_x, df_test_y)))
```

```
Accuracy on training set: 0.885
Accuracy on test set: 0.806
```

Both the accuracy on the training and test sets improve for the SVM model after applying scaling.

Following scaling, the test set performance increases to 80.6%.

Performance for scaled SVM Model

Train Accuracy	Test Accuracy
88.5%	80.6%

Conclusion / Recommendations

Actionable Insights for Optimization with K-Means Model

1. By comparing the two clusters, you can **optimize the rolling process** to achieve results more similar to Cluster 0, which has fewer defects.
2. If Cluster 0 involves more ideal process settings, replicating those conditions in the factory could help reduce the defect rate and improve plate quality.

Improvement Strategy

1. With this understanding, you can propose adjustments to the **production environment or rolling process** for Cluster 1 (high defect occurrence) to bring it closer to the conditions in Cluster 0 (low defect occurrence).
2. This could involve adjustments in machine parameters, material handling, or environmental control to ensure a more consistent, defect-free outcome.

These insights can guide you in **targeting specific areas for process improvement** and **reducing defect occurrences**.

Actionable Insights for Decision Tree Model

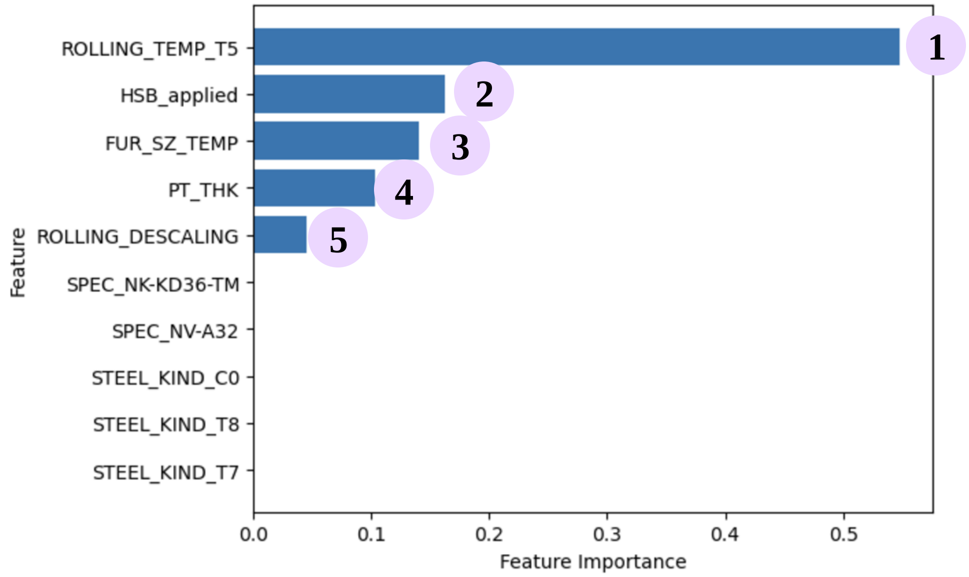
The Decision Tree model performs better than the SVM model. This may be because the plate process procedure is relatively constant, and simpler models like decision trees tend to perform better in such cases. The implications of this decision tree model can help predict future defective or non-defective cases.

```
Accuracy: 0.991
Confusion matrix:
[[145   0]
 [  2  69]]
              precision    recall  f1-score   support

           0      0.986     1.000     0.993       145
           1      1.000     0.972     0.986        71

    accuracy                          0.991       216
   macro avg      0.993     0.986     0.989       216
weighted avg      0.991     0.991     0.991       216
```

With this predictive model, it will be possible to prevent defective products in advance by predicting defect rates in the future.

Conclusion / Recommendations

Feature Importance for Decision Tree Model



Recommendations

To reduce the occurrence of scale defects and enhance process efficiency in the plate rolling process, the following measures are recommended, with a focus on balancing quality, cost, and energy efficiency:

1. **Lower the heating furnace temperature** cautiously, ensuring that the material reaches the optimal rolling temperature to maintain product quality.
2. **Optimize HSB application** based on specific plate characteristics to improve surface quality without negatively affecting material strength or equipment wear.
3. **Adjust the heating furnace crack zone temperature** by carefully reducing it to minimize scale formation, while maintaining the necessary temperature distribution for effective rolling.
4. **Increase the frequency of descaling operations**, but do so strategically to avoid excessive wear on equipment and unnecessary surface damage.
5. **Optimize plate thickness**, considering specific product requirements and strength, while also accounting for potential risks of scale formation.

Thank You

With POSCO