Data-Driven Optimization of Aluminum Rolling Parameters for Quality Improvement

Final report for the BDM capstone Project

Submitted by

Name: Aditya Rajput

Roll Number: 23F1002621

IITM Online BS Degree Program,

Indian Institute of Technology, Madras,

Chennai Tamil Nadu, India, 600036

Table of Contents

1. Executive Summary

This project, conducted for Hindalco Industries, aims to minimize defects in aluminum sheet rolling by identifying optimal Aluminum Rolling process parameters and understanding their impact on defect occurrence. Minimizing scoring defects is crucial for **production efficiency** as it reduces material wastage, rework, and downtime, ensuring smoother operations. It also enhances **cost-effectiveness and product quality**, leading to higher customer satisfaction and improved profitability for Hindalco's aluminum rolling process.

The dataset comprises **12,459 entries** representing the **primary data collected over a span of 2 years**. Given the significant class imbalance (the dataset contained only **198 defect cases out of 12,459** samples) a **statistical data analysis approach** was adopted over traditional machine learning techniques.

The project was structured into **Two key phases (each focused on solving one of the problem statements identified in the proposal stage)**:

1. **Optimization of Process Parameters (Problem Statement 1)** – Conducted **ANOVA, T-tests, Chi-square tests, and correlation analyses** for each width category to determine optimal parameter ranges for minimizing defects. **Response Surface Methodology (RSM)** was used to derive optimal settings.
2. **Significance Analysis of Process Parameters (Problem Statement 2)** – Used **Random Forest feature importance, and PCA** to identify the most influential parameters and interactions affecting defect formation.

The study resulted in **optimized process parameter settings per width**, and a **ranked list of key influencing factors** for defect minimization. These insights provide a **data-driven framework** for enhancing quality control and reducing defects in aluminum rolling operations.

2. Detailed Explanation of Analysis Process

Analysis Link: (Google Colab)

https://colab.research.google.com/drive/1mlO52j32FMbEAYyCBxxiFAUB9NEnHiFS?usp=sharing

2.1 - Introduction

The dataset provided for analysis contained 131 features. Initially, 131 features were available, but after discussions with **domain experts at Hindalco (Video Included in Midsem Report)**, it was suggested that many parameters were indirectly related to a few key process variables. By focusing on these critical parameters, we could derive meaningful insights that would inherently account for the influence of the remaining variables. Based on this expert input, 16 essential features were decided as the primary focus for the analysis.

These included: **(See Midsem Report for Details on the Metadata)**

- 1 Width feature (Categorical, 7 Unique Values) – Represents customer-defined sheet widths (800, 865, 775, 760, 810, 780, 770).
- 13 Numerical Process Parameters – Including MaxMillSpeed, EntryTensionSetpoint, MeanRollBend, BUR_Bot_Diam etc.
- 1 Categorical Process Parameter BUR_Bot_ID (10 unique values).
- Scoring (Target Variable, Binary: 'Yes' for defects, 'No' for defect-free sheets).

Given the severe **class imbalance** (98.4% 'No' and only 1.6% 'Yes'), traditional classification-based **machine learning models struggled** to correctly classify defect cases. Even after applying **resampling techniques** such as **SMOTE (Synthetic Minority Over-sampling Technique)** and **random over-sampling**, the models failed to **effectively learn patterns** that differentiate defect ('Yes') cases from non-defect ('No') cases. Most models exhibited **high accuracy** simply due to the dominance of the majority class, but their **true performance metrics** revealed critical shortcomings.

Key issues included:

- **Low ROC-AUC Scores**: Indicating that the models were unable to effectively separate the two classes.
- **Near-Zero Recall for the 'Yes' Class**: Meaning the models **almost never** identified defective cases correctly.
- **Overfitting to the Majority Class**: Even with class weighting adjustments, the models prioritized predicting the non-defect ('No') class, ignoring the minority 'Yes' cases.

Due to these challenges, a **pure machine learning approach** was deemed **ineffective** for defect classification. Instead, a **statistical data analysis approach** was adopted, focusing on **hypothesis testing, feature interactions, and parameter optimization** to derive meaningful insights for defect minimization. Before proceeding with these statistical tests, the dataset was thoroughly examined for **missing values**, ensuring clean and reliable data for hypothesis testing.

2.2 - Optimization of Process Parameters (Problem Statement 1)

Since the **Width** feature had **seven unique values** (800, 865, 775, 760, 810, 780, 770), the **original dataset was divided into seven sub-dataframes**, each corresponding to a different width category. This segmentation allowed for a **more granular analysis**, ensuring that process parameter optimizations were tailored **specifically to each width** rather than treating all widths as a single group.

This segmentation ensured that **hypothesis testing, parameter optimization, and defect minimization strategies** were conducted **individually for each width**, leading to **more precise and actionable insights** for Hindalco's rolling process.

To determine the optimal process settings for minimizing defects, the following hypothesis tests were performed for all widths:

- Chi-Square Test: Assessed whether the categorical variable BUR_Bot_ID significantly influenced defect occurrence.
- ANOVA & T-tests: Compared the means of numerical process parameters between defect and non-defect cases to determine statistically significant differences.

- Correlation Analysis (Pearson/Spearman): Evaluated the relationships between process parameters and defect occurrence.
- Two-Way ANOVA: Identified interaction effects between key process parameters affecting defect formation.
- Response Surface Methodology (RSM): Used to determine the optimal operating ranges for process parameters to minimize defects.

From this analysis, we identified parameter thresholds and optimal settings for each width category, helping operators fine-tune machine settings to reduce scoring defects.

2.3 - Significance of Process Parameters (Problem Statement 2)

A more holistic approach was taken to assess process parameter significance **across the entire dataset**. To determine the most influential process parameters and their interactions affecting defect formation, we conducted:

- Random Forest Feature Importance (on balanced subsamples): Used as an exploratory method to assess parameter significance.
- Principal Component Analysis (PCA): Reduced dimensionality and identified the most dominant process parameters.

This approach led to a ranked list of process parameters based on their influence on scoring defects, providing a data-driven strategy for process improvement.

3. Results and Findings

3.1. Chi-Square Test

The **Chi-Square test** was conducted to examine whether the **only** categorical process parameter **BUR_Bot_ID** has a statistically significant relationship with defect occurrence (**Scoring**). This test helps determine whether certain **BUR_Bot_ID** values contribute more to defects than others, which is crucial for identifying process conditions that increase defect risks.

The **Chi-Square test results** for each width category are as follows:

Width	Chi-Square Statistic	p-value	Significant? (α = 0.05)
760	149.5760	0.0000	Significant
770	6.7970	0.5587	Not Significant
775	35.3526	0.0001	Significant
780	5.6213	0.6896	Not Significant
800	116.2782	0.0000	Significant
810	16.9082	0.0180	Significant
865	471.6581	0.0000	Significant

- **For Widths 760, 775, 800, 810, and 865** certain **BUR_Bot_ID** values contribute more to defects than others.
- **For Widths 770 and 780** defects occur independently of **BUR_Bot_ID** selection.

3.2 - ANOVA Test

ANOVA (Analysis of Variance) was conducted to determine whether the mean values of numerical process parameters significantly differ between the 'Defect' and 'Non-Defect' groups. A lower p-value (< 0.05) indicates a statistically significant difference, implying that the process parameter may influence defect formation.

These results suggest that certain process parameters, such as MaxMillSpeed, MeanRollForce, BUR_Bot_Diam, and BUR_Top_ID1, are statistically significant in influencing defect rates for specific width categories. These parameters should be considered for further optimization to minimize defect occurrence. Also, the parameter ‘MeanCoolantFlow’ was not found to be significant for any of the widths.

ANOVA Test Results: Significant Process Parameters by Width

Width	Significant Process Parameters (p-value)
760	MaxMillSpeed (0.0183), EntryTensionSetpoint (0.0254), MeanRollForce (0.0449), BUR_Top_ID1 (0.0097)
770	No significant parameters observed
775	MaxRollBend (0.0320), BUR_Top_ID1 (0.0059), MaxRollForce (0.0344), MeanTilt (0.0274)
780	No significant parameters observed
800	MaxMillSpeed (0.0127), MeanRollForce (<0.0001), BUR_Top_ID1 (0.0034), WKR_Bot_Diam (<0.0001), BUR_Bot_Diam (<0.0001), MaxRollForce (0.0053)
810	MaxMillSpeed (0.0023), ExitTensionSetpoint (0.0011)
865	MaxMillSpeed (0.0345), MaxRollBend (<0.0001), MeanRollBend (0.0001), BUR_Top_ID1 (0.0002), BUR_Bot_Diam (<0.0001)

Process Parameter	760	770	775	780	800	810	865
MaxMillSpeed	Yes	No	No	No	Yes	Yes	Yes
EntryTensionSetpoint	Yes	No	No	No	No	No	No
ExitTensionSetpoint	No	No	No	No	No	Yes	No
MaxRollBend	No	No	Yes	No	No	No	Yes
BUR_Top_ID1	Yes	No	Yes	No	Yes	No	Yes
MeanRollForce	Yes	No	No	No	Yes	No	No
MaxRollForce	No	No	Yes	No	Yes	No	No
MeanRollBend	No	No	No	No	No	No	Yes
MeanTilt	No	No	Yes	No	No	No	No
WKR_Bot_Diam	No	No	No	No	Yes	No	No
BUR_Bot_Diam	No	No	No	No	Yes	No	Yes

3.3 - T-Test

The T-test was conducted to validate the findings from ANOVA by comparing the mean values of each process parameter between defect (Scoring = 1) and non-defect (Scoring = 0) cases for each width category. This statistical test helps identify which parameters show significant differences and may contribute to defect occurrence. The test assumed unequal variances. A p-value threshold of **0.05** was used to determine statistical significance. Parameters with p-values below this threshold indicate a significant difference in mean values between defect and non-defect cases.

Width	Significant Parameters (p-value)	Key Observations
760	MaxMillSpeed (0.0203), EntryTensionSetpoint (0.0167), MaxRollBend (0.0137), MeanRollForce (0.0347), BUR_Top_ID1 (<0.0001)	Multiple parameters show strong differentiation between defect and non-defect cases.
770	BUR_Top_ID1 (0.0002)	Only BUR_Top_ID1 was significant, other parameters lacked strong differentiation.
775	BUR_Top_ID1 (0.0017), MaxRollForce (0.0360), MeanTilt (0.0180)	BUR_Top_ID1 appears as a key influencing factor across multiple widths.
780	BUR_Top_ID1 (0.0002)	Most other parameters did not show significant differences.
800	MaxMillSpeed (0.0211), MeanRollForce (<0.0001), BUR_Top_ID1 (<0.0001), WKR_Bot_Diam (<0.0001), BUR_Bot_Diam (<0.0001), MaxRollForce (0.0062)	This width category shows strong differentiation across multiple process parameters.
810	EntryTensionSetpoint (<0.0001), ExitTensionSetpoint (<0.0001), MaxRollBend (0.0337), BUR_Top_ID1 (0.0020), WKR_Bot_Diam (<0.0001), BUR_Bot_Diam (<0.0001), MaxRollForce (<0.0001), MeanTilt (0.0284)	Entry and Exit Tension Setpoints are notably significant.
865	MaxMillSpeed (0.0464), MaxRollBend (<0.0001), MeanRollBend (<0.0001), BUR_Top_ID1 (<0.0001), BUR_Bot_Diam (<0.0001)	This width category exhibits strong differentiation in roll parameters.

1. **BUR_Top_ID1 consistently shows significant differences** across all width categories, suggesting a strong correlation with defect occurrence.

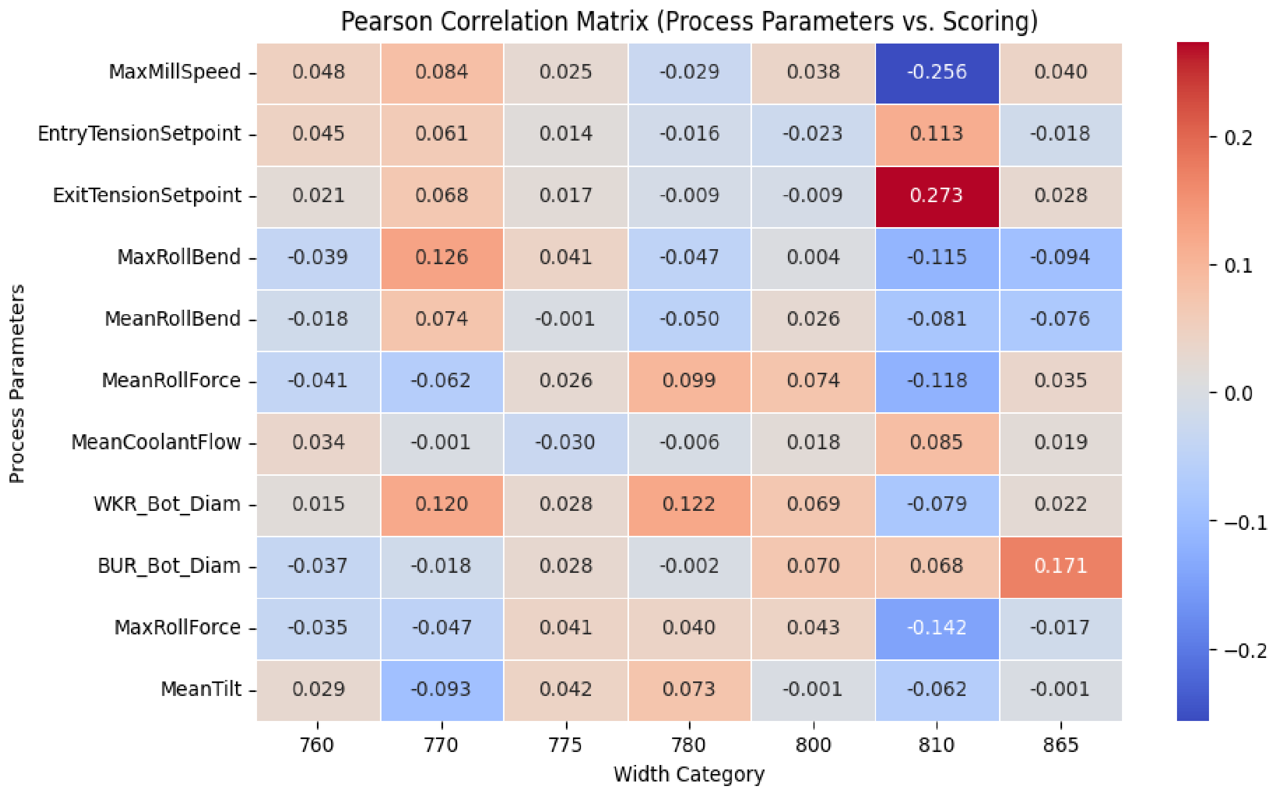
2. **MaxMillSpeed, MaxRollBend, and MeanRollForce appear in multiple width categories**, indicating their potential impact on quality.
3. **Entry and Exit Tension Setpoints are significant for certain widths (810, 760, 800),** implying that tension variations may influence defect rates.
4. **Larger widths (800, 810, 865) exhibit more significant parameters,** suggesting that defects in these categories may be more sensitive to process variations.

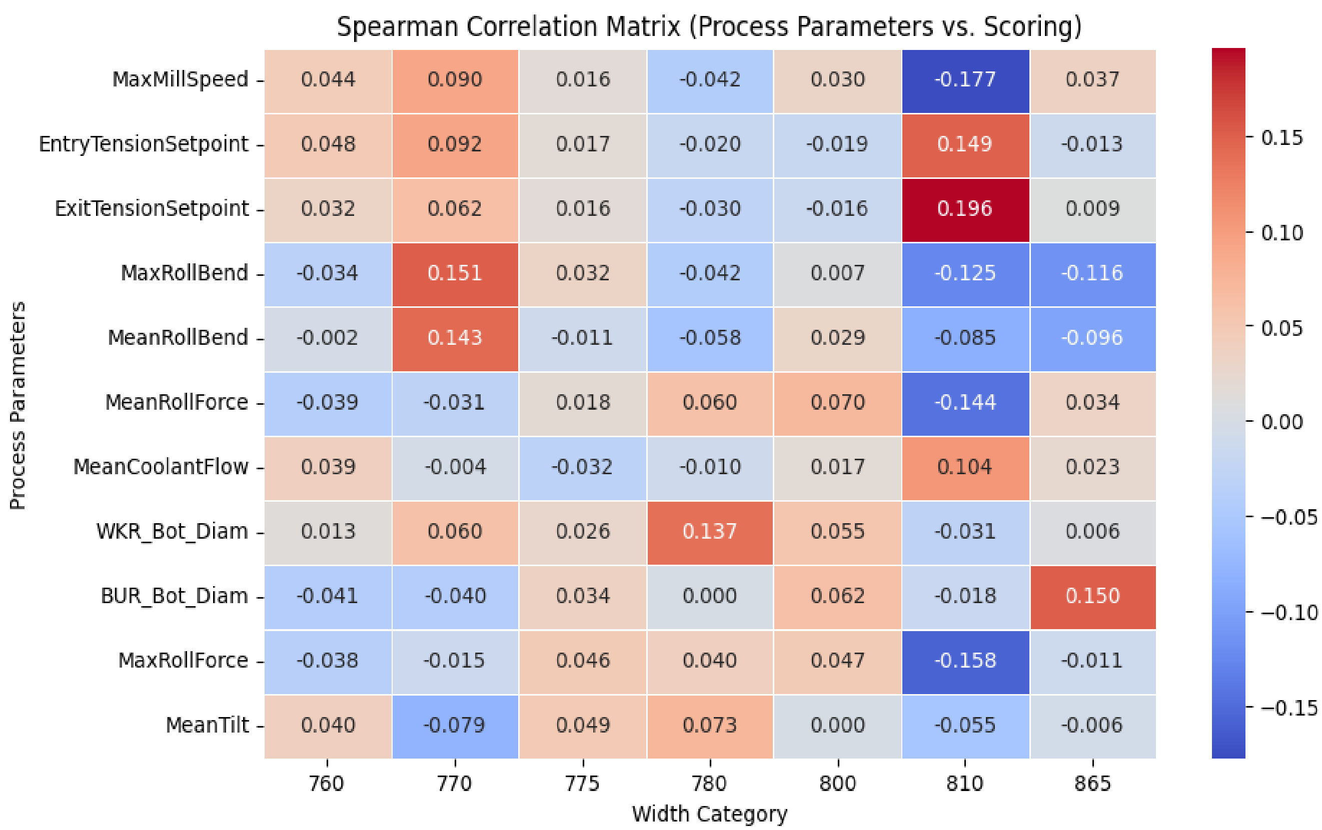
3.4 - Pearson/Spearman Correlation

Correlation analysis was performed to examine relationships between numerical process parameters and defect occurrence (Scoring) across width categories using **Pearson's** and **Spearman's** correlation coefficients.

- **Overall Weak Correlations:** Most parameters showed weak correlations with defects, indicating no single factor strongly influences defect occurrence.
- **Notable Trends by Width:**
    - **Width 770:** Max Roll Bend (~0.126 Pearson, ~0.151 Spearman) and WKR_Bot_Diam (~0.12 Pearson) exhibited moderate correlations.
    - **Width 780:** Mean Roll Force (~0.099 Pearson) and WKR_Bot_Diam (~0.122 Pearson) showed higher correlations.
    - **Width 810:** MaxMillSpeed (-0.256 Pearson) and Exit Tension Setpoint (0.273 Pearson) had the strongest correlations, suggesting mill speed reductions may contribute to defects while higher exit tension could be protective.
- **Pearson vs. Spearman:** Spearman correlations were slightly higher, indicating potential non-linear relationships.

No single parameter consistently drives defect formation, but **MaxMillSpeed, Exit Tension Setpoint, WKR_Bot_Diam, and Mean Roll Force** warrant further analysis, particularly regarding interactions. Correlation heatmaps were used for visualization.


Pearson Correlation Matrix (Process Parameters vs. Scoring)
Process Parameters
MaxMillSpeed 0.048 0.084 0.025 -0.029 0.038 -0.256 0.040
EntryTensionSetpoint 0.045 0.061 0.014 -0.016 -0.023 0.113 -0.018
ExitTensionSetpoint 0.021 0.068 0.017 -0.009 -0.009 0.273 0.028
MaxRollBend -0.039 0.126 0.041 -0.047 0.004 -0.115 -0.094
MeanRollBend -0.018 0.074 -0.001 -0.050 0.026 -0.081 -0.076
MeanRollForce -0.041 -0.062 0.026 0.099 0.074 -0.118 0.035
MeanCoolantFlow 0.034 -0.001 -0.030 -0.006 0.018 0.085 0.019
WKR_Bot_Diam 0.015 0.120 0.028 0.122 0.069 -0.079 0.022
BUR_Bot_Diam -0.037 -0.018 0.028 -0.002 0.070 0.068 0.171
MaxRollForce -0.035 -0.047 0.041 0.040 0.043 -0.142 -0.017
MeanTilt 0.029 -0.093 0.042 0.073 -0.001 -0.062 -0.001
760 770 775 780 800 810 865
Width Category
0.2
0.1
0.0
−0.1
−0.2


Spearman Correlation Matrix (Process Parameters vs. Scoring)
Process Parameters
MaxMillSpeed 0.044 0.090 0.016 -0.042 0.030 -0.177 0.037
EntryTensionSetpoint 0.048 0.092 0.017 -0.020 -0.019 0.149 -0.013
ExitTensionSetpoint 0.032 0.062 0.016 -0.030 -0.016 0.196 0.009
MaxRollBend -0.034 0.151 0.032 -0.042 0.007 -0.125 -0.116
MeanRollBend -0.002 0.143 -0.011 -0.058 0.029 -0.085 -0.096
MeanRollForce -0.039 -0.031 0.018 0.060 0.070 -0.144 0.034
MeanCoolantFlow 0.039 -0.004 -0.032 -0.010 0.017 0.104 0.023
WKR_Bot_Diam 0.013 0.060 0.026 0.137 0.055 -0.031 0.006
BUR_Bot_Diam -0.041 -0.040 0.034 0.000 0.062 -0.018 0.150
MaxRollForce -0.038 -0.015 0.046 0.040 0.047 -0.158 -0.011
MeanTilt 0.040 -0.079 0.049 0.073 0.000 -0.055 -0.006
760 770 775 780 800 810 865
Width Category
0.15
0.10
0.05
0.00
−0.05
−0.10
−0.15

3.5 - Two-Way ANOVA

To understand the combined influence of process parameters on defect rates, a Two-Way ANOVA analysis was conducted for each width category. This analysis identified statistically significant interactions (p-value < 0.05) between process parameters, highlighting key factors affecting defect formation.

1. **Speed-Related Interactions: MaxMillSpeed** showed significant interactions with **BUR_Bot_Diam, MaxRollForce, MeanRollBend, and MeanCoolantFlow** across multiple widths (760, 775, 780, 810, and 865). This suggests that rolling speed's impact on defects is influenced by roll diameters and force-related parameters.
2. **Tension Interactions: EntryTensionSetpoint and ExitTensionSetpoint** interacted with **MaxRollForce, MeanRollForce, MeanCoolantFlow, and MeanTilt** in various width categories. This indicates that optimal tension settings must be adjusted in conjunction with force and cooling parameters to reduce defects.
3. **Roll Bending and Force Interactions: MaxRollBend and MeanRollBend** exhibited strong interactions with **MeanCoolantFlow, BUR_Bot_Diam, and MeanTilt**, particularly in widths 775, 800, and 865. This suggests that bending forces, in combination with cooling and roll diameters, significantly influence defect formation.
4. **Cooling and Diameter Effects: MeanCoolantFlow** was involved in interactions with **EntryTensionSetpoint, MeanRollBend, and BUR_Bot_Diam**, confirming its role in defect minimization. Cooling efficiency must be carefully controlled alongside roll diameter adjustments.
5. **Roll Diameter Dependencies: WKR_Bot_Diam, BUR_Bot_Diam, and MaxRollForce** exhibited strong interactions in widths 775, 800, and 865. This indicates that roll configurations must be optimized based on force settings to minimize defects.

These insights enhance the understanding of defect formation and guide the final optimization recommendations for aluminum rolling operations.

3.6 - Response Surface Methodology (RSM)

Response Surface Methodology (RSM) is a statistical technique used for modeling and analyzing the relationships between multiple input variables and a response variable. It is particularly useful for optimizing processes by finding the ideal parameter settings that minimize or maximize an outcome. In this study, RSM was applied to determine the optimal process parameters that minimize defects in the aluminum rolling process. 'BUR_Top_ID1' was excluded from this study as optimal values of an identifier are meaningless.

Steps Followed:

1. **Polynomial Feature Expansion:**
    - A second-order polynomial model was used to account for interactions and quadratic effects between process parameters.
2. **Generalized Linear Modeling (GLM):**
    - A logistic regression model was fitted using a Binomial family function to predict defect probability based on expanded polynomial features.
3. **Optimization via L-BFGS-B Algorithm:**
    - The model was used as an objective function in a constrained optimization framework to find process parameter values that minimize defect probability.

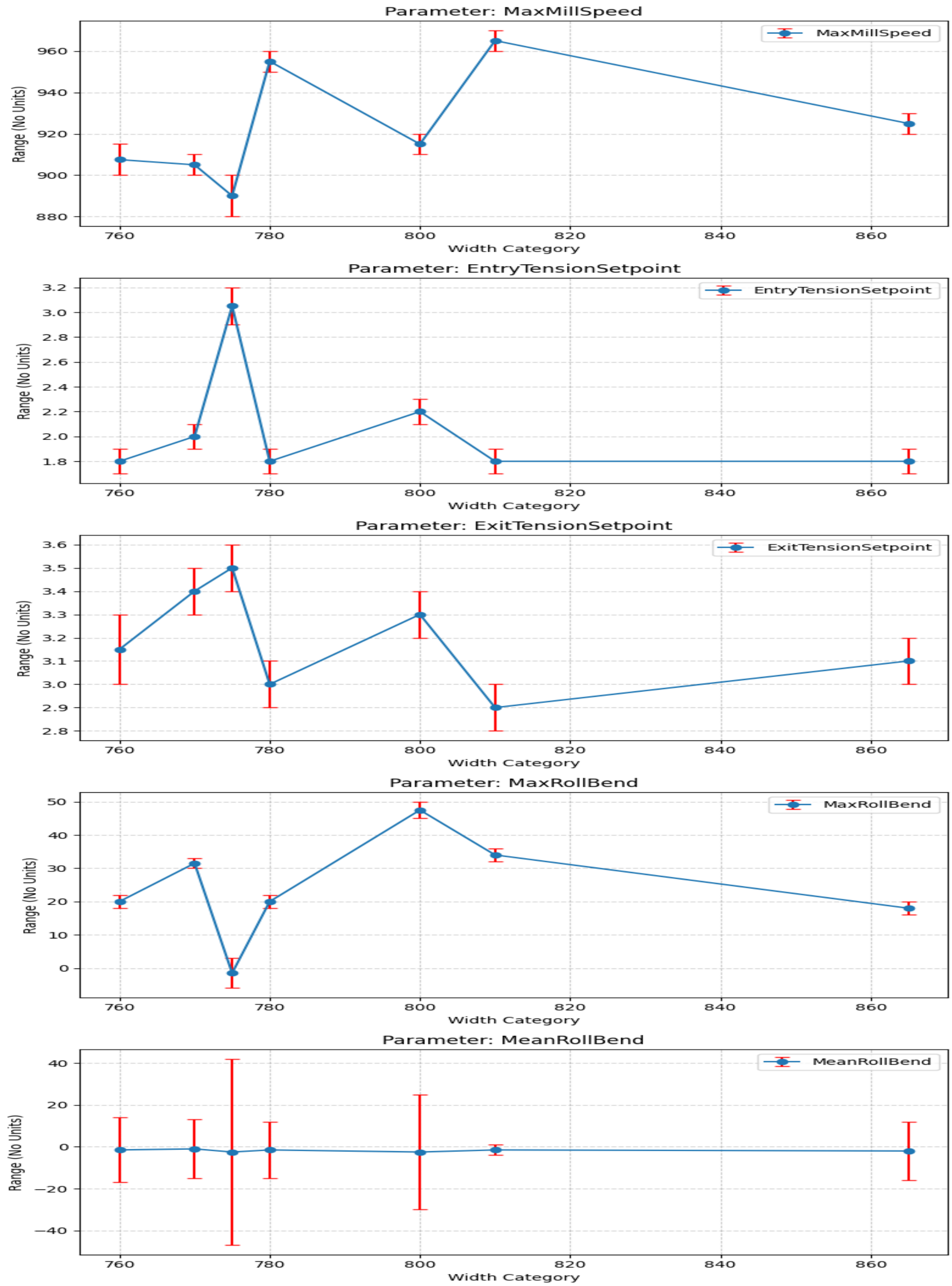
- The optimized parameter values for each width category were identified, showing distinct trends in tension settings, roll bending, and force distribution.
- In general, **Mean Roll Force**, **Entry Tension Setpoint**, and **MaxMillSpeed** exhibited strong influence over defect probability.
- The identified **ideal operating ranges** for minimizing defects vary across width categories but follow consistent trends in process stability and force distribution.

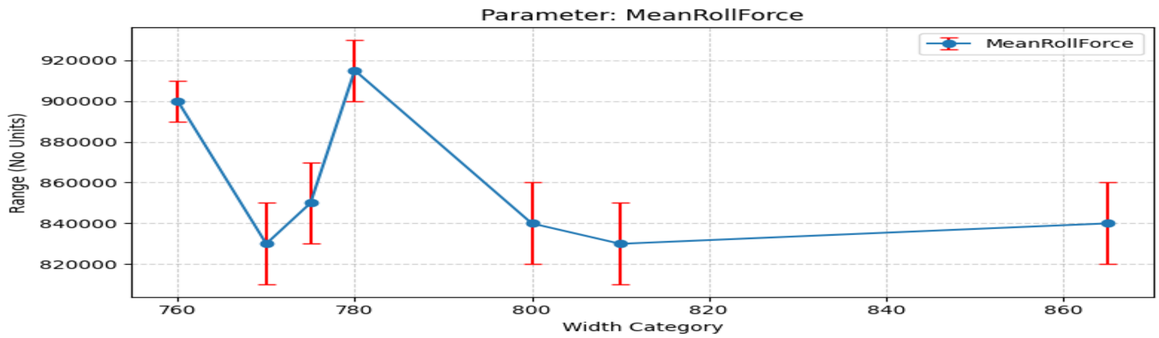
Important: In the absence of explicit unit specifications from the Hindalco Team, I have chosen to present only the parameter ranges in the final report. This decision avoids potential misinterpretations or inconsistencies that could arise from assuming incorrect units. It is assumed that the standard industry practices apply, and the stakeholders can refer to their internal guidelines if needed.

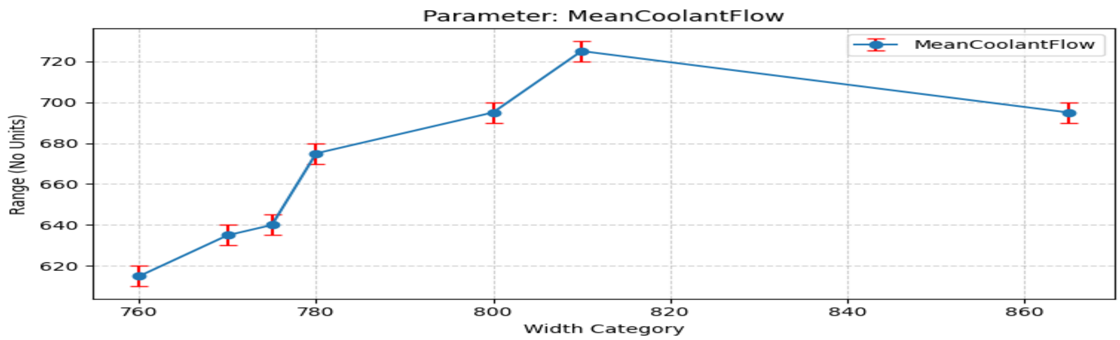
Table: Recommended Operating Ranges to Minimize Defects

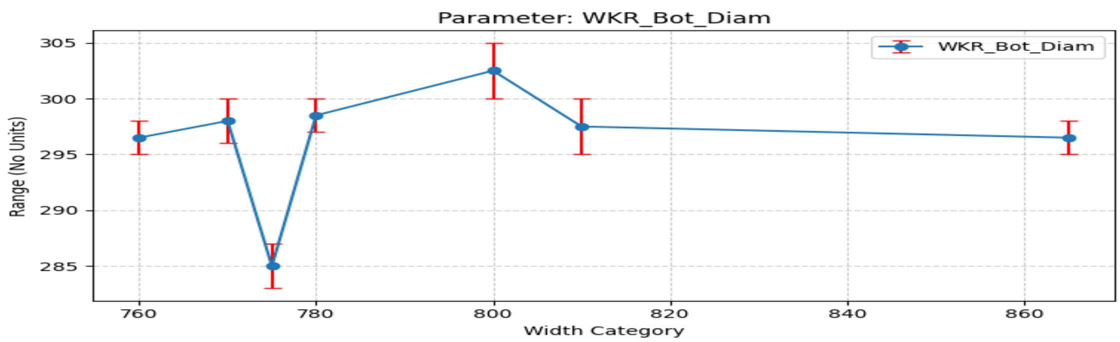
Process Parameter	Width 760	Width 770	Width 775	Width 780	Width 800	Width 810	Width 865
Max Mill Speed	**900 - 915**	900 - 910	880 - 900	950 - 960	**910 - 920**	**960 - 970**	**920 - 930**
Entry Tension Setpoint	**1.7 - 1.9**	1.9 - 2.1	2.9 - 3.2	1.7 - 1.9	2.1 - 2.3	1.7 - 1.9	1.7 - 1.9
Exit Tension Setpoint	3.0 - 3.3	3.3 - 3.5	3.4 - 3.6	2.9 - 3.1	3.2 - 3.4	**2.8 - 3.0**	3.0 - 3.2
Max Roll Bend	**18 - 22**	30 - 33	**-6 to -3**	18 - 22	45 - 50	32 - 36	**16 - 20**
Mean Roll Bend	-17 to -14	-15 to -13	-47 to -42	-15 to -12	-30 to -25	-4 to -1	**-16 to -12**
Mean Roll Force	**890k - 910k**	810k - 850k	830k - 870k	900k - 930k	**820k - 860k**	810k - 850k	820k - 860k
Mean Coolant Flow	610 - 620	630 - 640	635 - 645	670 - 680	690 - 700	720 - 730	690 - 700
WKR_Bot_ Diam	295 - 298	296 - 300	283 - 287	297 - 300	**300 - 305**	295 - 300	295 - 298
BUR_Bot_Diam	790 - 800	790 - 800	788 - 792	794 - 796	790 - 795	795 - 800	**792 - 795**
Max RollForce	1.75M - 1.80M	1.72M - 1.77M	**1.65M - 1.70M**	1.72M - 1.76M	**1.65M - 1.70M**	1.58M - 1.61M	1.60M - 1.65M
MeanTilt	0.02 - 0.03	0.04 - 0.05	**-0.27 to - 0.25**	0.02 - 0.03	-0.26 to - 0.23	0.00 - 0.01	0.02 - 0.03

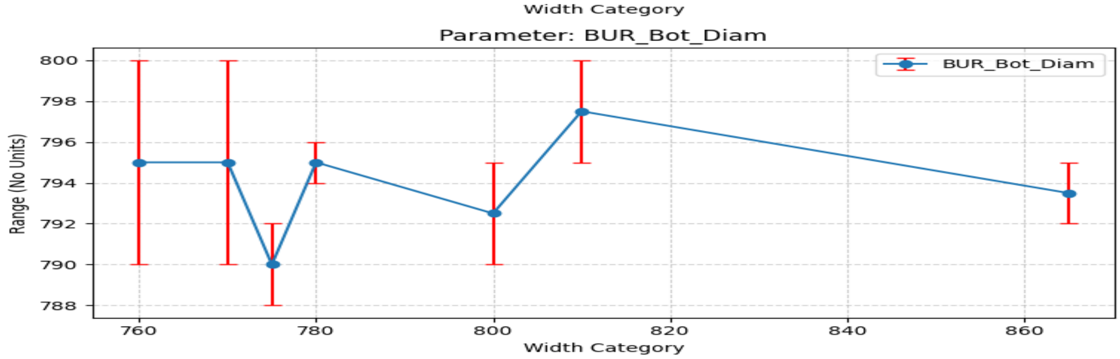
- The ranges provide flexibility for real-world operations, ensuring that mill operators can adjust within reasonable limits. Operators should use these ranges as guidelines while continuously monitoring quality outcomes. The **Bold** features are the ones identified to be significant for those widths by the ANOVA and t-tests.
- The tabular data outlining optimal process parameter ranges for each width category was visualized as **range bar charts** to provide a clear representation of the recommended operating limits, aiding in process control and defect reduction.

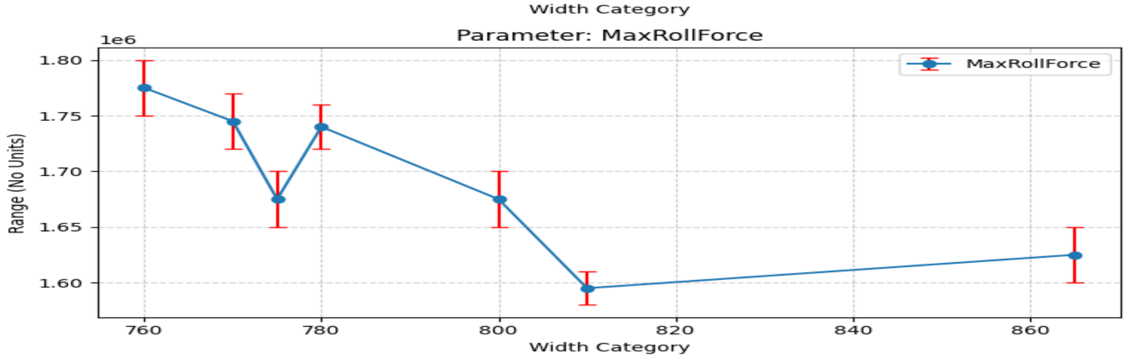

Parameter: MaxMillSpeed
MaxMillSpeed
Range (No Units)
Width Category
Parameter: EntryTensionSetpoint
EntryTensionSetpoint
Range (No Units)
Width Category
Parameter: ExitTensionSetpoint
ExitTensionSetpoint
Range (No Units)
Width Category
Parameter: MaxRollBend
MaxRollBend
Range (No Units)
Width Category
Parameter: MeanRollBend
MeanRollBend
Range (No Units)
Width Category


Parameter: MeanRollForce
MeanRollForce
Range (No Units)
920000
900000
880000
860000
840000
820000
760
780
800
820
840
860
Width Category


Parameter: MeanCoolantFlow
MeanCoolantFlow
Range (No Units)
720
700
680
660
640
620
760
780
800
820
840
860
Width Category


Parameter: WKR_Bot_Diam
WKR_Bot_Diam
Range (No Units)
305
300
295
290
285
760
780
800
820
840
860
Width Category


Parameter: BUR_Bot_Diam
BUR_Bot_Diam
Range (No Units)
800
798
796
794
792
790
788
760
780
800
820
840
860
Width Category


Parameter: MaxRollForce
MaxRollForce
1e6
Range (No Units)
1.80
1.75
1.70
1.65
1.60
760
780
800
820
840
860
Width Category

Note: To gain a comprehensive understanding of process parameter significance, for further analysis the entire dataset is used instead of segmented width categories. This broader approach helps identify the most influential parameters impacting defect probability across all rolling conditions. By doing so, we further progress towards solving Problem Statement 2, which focuses on determining the significance of process parameters and their interactions in influencing defects.

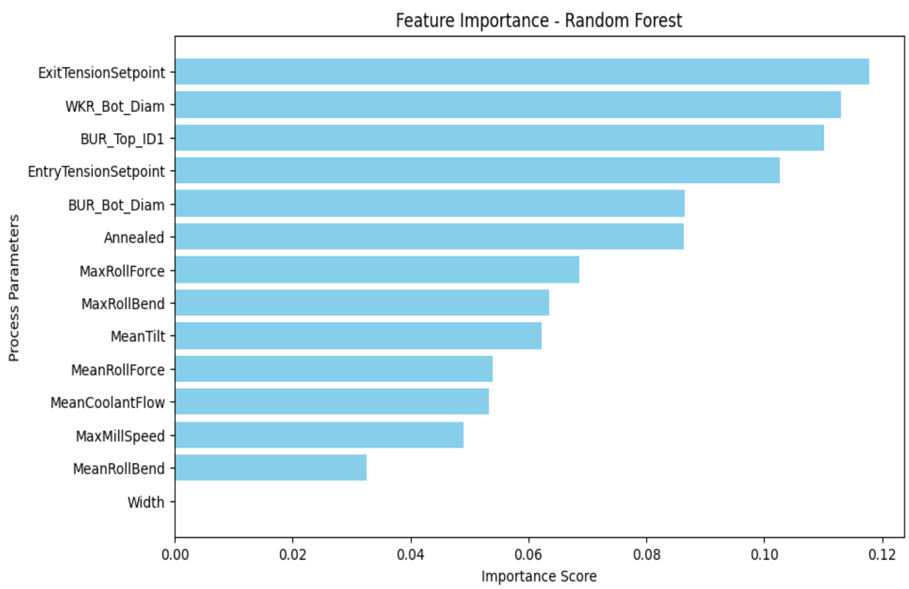
3.7 - Feature Importance Analysis

To identify the most influential process parameters in determining defect occurrence, we employed a **Random Forest feature importance analysis**. While machine learning was not used for defect prediction, this exploratory analysis provides insights into the relative significance of various parameters.

1. **Feature Selection**:
    - The categorical variable **BUR_Bot_ID** was excluded for this analysis.
    - The target variable was **Scoring (0 = No Defect, 1 = Defect)**.
2. **Handling Class Imbalance**:
    - The dataset was highly imbalanced, so **SMOTE (Synthetic Minority Over-sampling Technique)** was used to generate synthetic samples and balance the classes.
3. **Random Forest Model**:
    - A **Random Forest Classifier (100 estimators, random state = 42)** was trained on the balanced dataset.
    - Feature importance scores were extracted based on the model's learned decision

The **Exit Tension Setpoint** emerged as the most significant factor, followed closely by **WKR Bottom Diameter** and **BUR Top ID1**. Notably, **tension parameters, roll diameters, and force-related variables** played a crucial role in defect probability.

Rank	Feature	Importance Score
1	Exit Tension Setpoint	**0.118**
2	WKR Bottom Diameter	**0.113**
3	BUR Top ID1	**0.110**
4	Entry Tension Setpoint	**0.103**
5	BUR Bottom Diameter	**0.086**
6	Annealed	**0.086**
7	Max Roll Force	**0.069**
8	Max Roll Bend	**0.064**
9	Mean Tilt	**0.062**

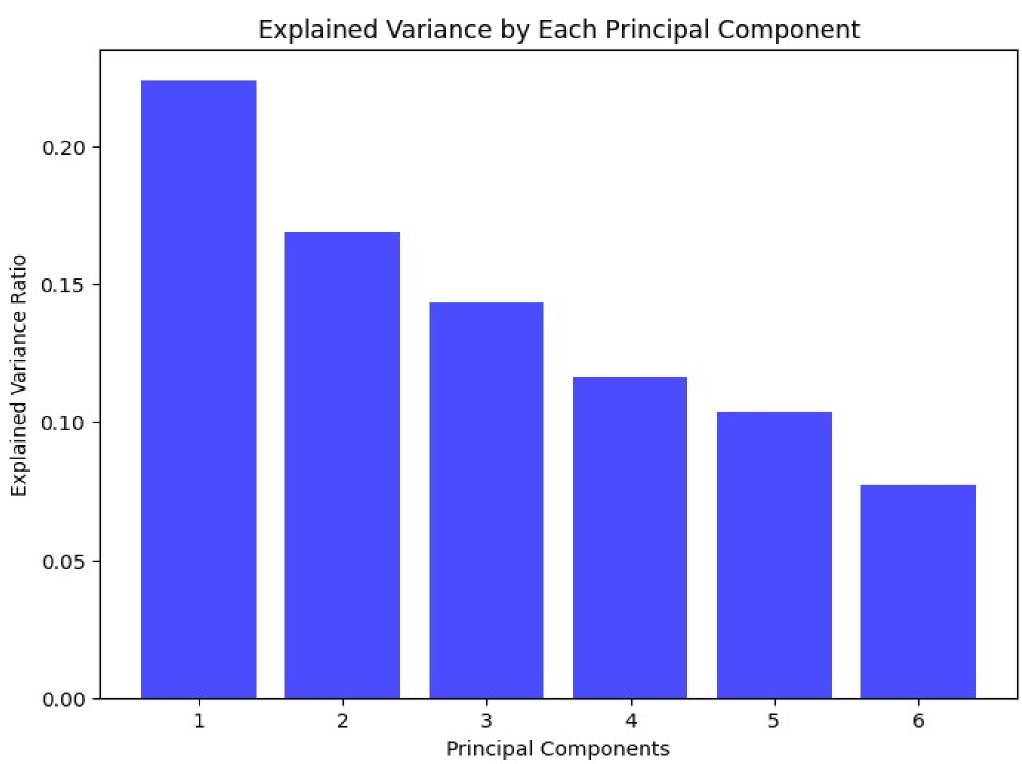


- **Tension settings** (Exit & Entry) have a strong influence on defect rates, indicating the need for precise control in this area.
- **Work and Backup Roll Diameters** are significant, suggesting that roll wear and selection impact product quality.
- **Max Roll Force and Roll Bending** also contribute, implying that force control mechanisms may need further optimization.

3.8 - Principal Component Analysis (PCA)

To identify dominant process parameters and reduce dimensionality while retaining maximum variance, **Principal Component Analysis (PCA)** was performed on key numerical features. PCA helps uncover the most influential parameter combinations affecting defect occurrence.

- The **cumulative explained variance** tells us how much of the total variance in the data is captured by the first few principal components.
- Result: [0.224,0.393,0.537,0.654,0.757,0.835]
  This means:
  - **PC1 alone** explains **22.4%** of the variance.
  - **PC1 + PC2** together explain **39.3%** of the variance.
  - **PC1 + PC2 + PC3** together explain **53.7%** of the variance.
  - **PC1 to PC6** together explain **83.5%** of the total variance.
- If we aim to retain around **85-90%** of variance while reducing the number of features, selecting the first **5-6 components** seems reasonable.



4. Interpretation of Results and Recommendations

4.1 – Solution to Problem Statement 1 – Optimal Parameter Settings

Based on the statistical analysis (ANOVA, T-tests, and Two-Way ANOVA), we identified the most critical process parameters for each width category. The recommended operational adjustments for optimizing rolling quality are as follows:

Width 760 Operational Recommendations:

- Maintain **Max Mill Speed** within **900-915** to prevent excessive force variations.
- Set **Entry Tension Setpoint** in the **1.7 - 1.9** range to ensure smooth material entry and minimize defects.
- Ensure **Mean Roll Force** stays within **890k – 910k** to maintain uniform thickness and avoid excessive work hardening.

Width 770 Operational Recommendations: No significant features identified. Maintain current process settings, but monitor for potential variations in secondary parameters.

Width 775 Operational Recommendations:

- Keep **Max Roll Bend** within -6 to -3 to minimize edge defects and ensure uniform strip flatness.
- Set **Max Roll Force** between **1.65M - 1.70M** to maintain stable deformation control.
- Adjust **Mean Tilt** within **-0.27 to -0.25 degrees** to reduce skewing and edge waviness.

Width 780 Operational Recommendations: No significant features identified. Continue operating within the existing process conditions, but monitor interaction effects from ANOVA.

Width 800 Operational Recommendations:

- Set **Max Mill Speed** between **910 - 920** to maintain steady strip elongation.
- Maintain **Mean Roll Force** at **820k - 860k** to prevent excessive strain and ensure proper rolling conditions.
- Ensure **WKR_Bot_Diam** stays between **297 - 300** and **BUR_Bot_Diam** remains in **790 - 795** range to ensure proper roll alignment and contact.

- Adjust **Max Roll Force** within **1.65M - 1.70M** to optimize rolling force distribution.

Width 810 Operational Recommendations:

- Keep **Max Mill Speed** between **960 - 970** for efficient rolling with minimal surface defects.
- Adjust **Exit Tension Setpoint** to **2.8 - 3.0** to maintain uniform tension and prevent strip flutter.

Width 865 Operational Recommendations:

- Maintain **Max Mill Speed** in **920 - 930** range for stable deformation control.
- Adjust **Max Roll Bend** to **16 - 20** to minimize waviness and edge cracking.
- Set **Mean Roll Bend** between **-16 to -12** to balance roll pressure distribution.
- Ensure **BUR_Bot_Diam** stays within **792 - 795** to achieve proper roll gap control.

4.2 – Solution to Problem Statement 2 – Feature Importance Analysis

Random Forest Feature Importance: From these rankings, **tension parameters** and **roll geometry** emerge as the most critical factors.. These findings suggest that careful monitoring and control of tension and roll configuration are essential to minimize scoring defects.

Principal Component Analysis (PCA): In terms of loadings, **roll force parameters**, **tension settings**, and **roll diameters** contribute heavily to the top principal components. Consequently, if the goal is to reduce the dataset size while preserving most of the relevant information, **retaining the first five or six principal components** is advisable.

Final Recommendations regarding feature importance:

- **Tension control** (both entry and exit) and **roll geometry** (WKR/BUR diameters) are primary drivers of defect formation.
- PCA confirms that these variables explain a large portion of the variance, reinforcing their importance in process optimization.

By focusing on these parameters—particularly tension setpoints, roll diameters process engineers can more effectively **minimize scoring defects** and improve overall product quality.