GARMENT PRODUCTIVITY

Group Members

Dalia Alotaibi 445202288

Jumana Alhumud 445201376

Fanar Barboud 445200050

Fatimah Alhebshi 445200431

Hessa Alarifi 445204385

CONTENT OF THE PRESENTATION

- THE PROBLEM
- DATA OVERVIEW
- DATA MINING TECHNIQUES
- RESULTS AND FINDINGS

THE PROBLEM

PROJECT AIM:

Analyze and predict garment factory employees' productivity using data mining.

OBJECTIVES:

- Classify employees as likely / unlikely to reach their productivity target
- Cluster employees with similar productivity patterns
- Use these insights to better understand and improve productivity performance

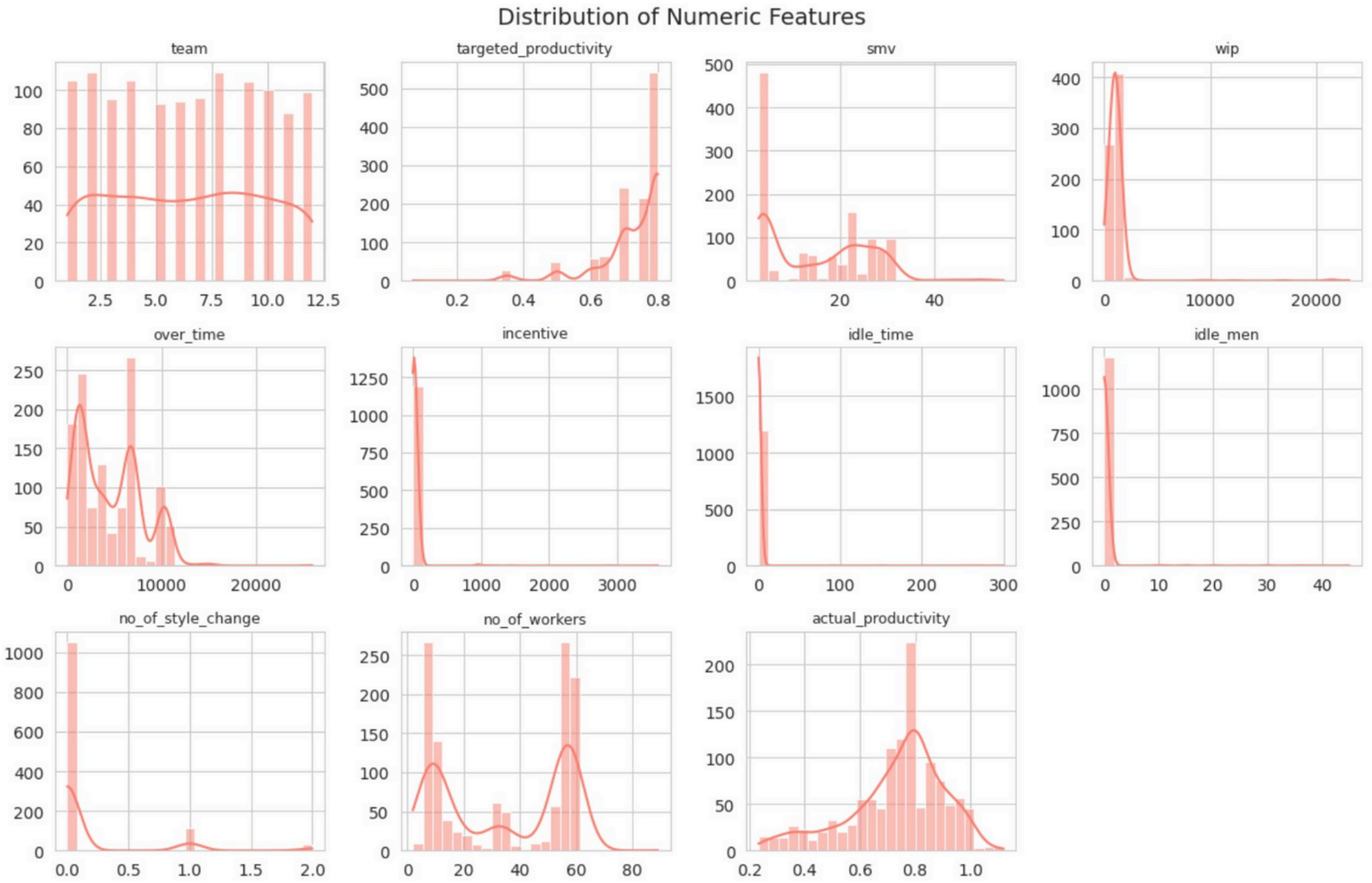
DATA OVERVIEW

- Real-world data from a garment manufacturing factory
- Size: 1,197 records (rows) and 15 features (columns)
- A discrete class label categorized into: Low, Medium, High

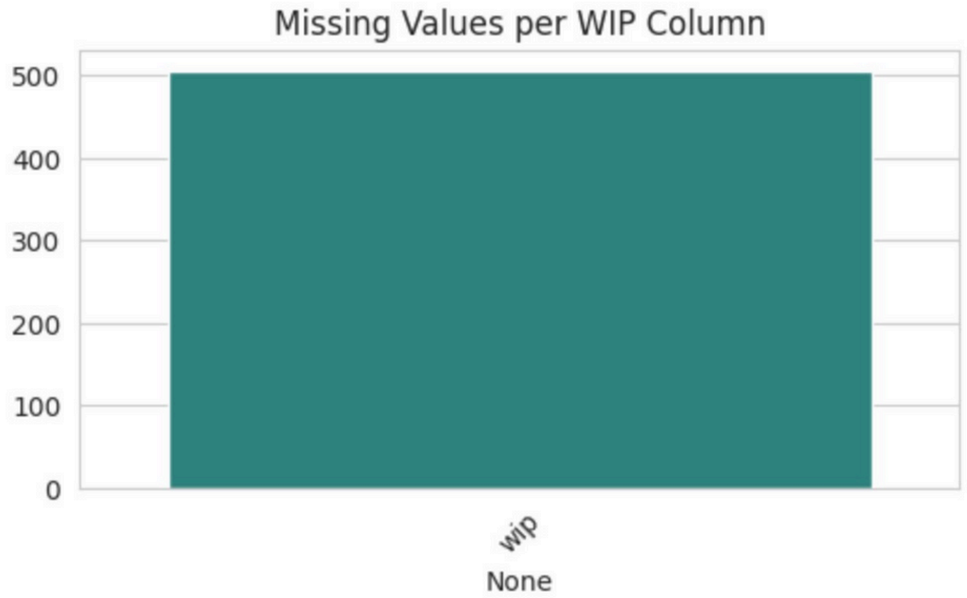
	date	quarter	department	day	team	targeted_productivity	smv	wip	over_time	incentive	idle_time	idle_men	no_of_style_change	no_of_workers	actual_productivity	Label
0	1/1/2015	Quarter1	sweing	Thursday	8	0.80	26.16	1108.0	7080	98	0.0	0	0	59.0	0.940725	High
1	1/1/2015	Quarter1	finishing	Thursday	1	0.75	3.94	NaN	960	0	0.0	0	0	8.0	0.886500	High
2	1/1/2015	Quarter1	sweing	Thursday	11	0.80	11.41	968.0	3660	50	0.0	0	0	30.5	0.800570	Medium
3	1/1/2015	Quarter1	sweing	Thursday	12	0.80	11.41	968.0	3660	50	0.0	0	0	30.5	0.800570	Medium
4	1/1/2015	Quarter1	sweing	Thursday	6	0.80	25.90	1170.0	1920	50	0.0	0	0	56.0	0.800382	Medium

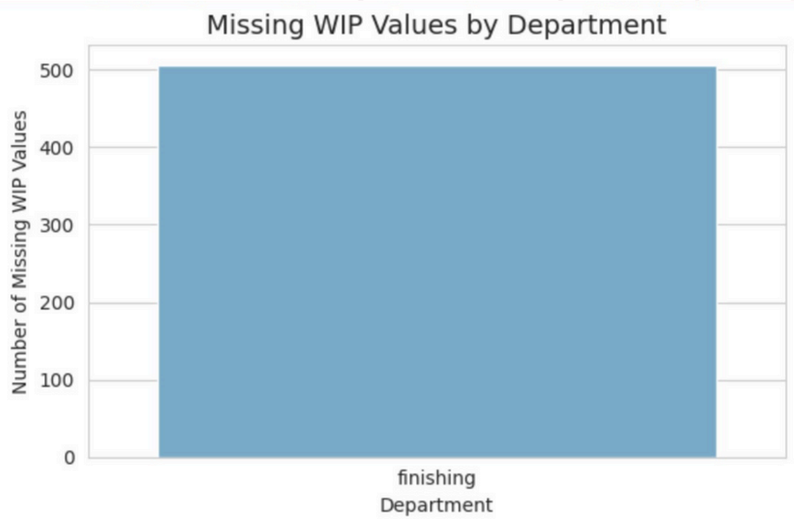
TABLE OF CONTENT

1. DATASET SUMMARY



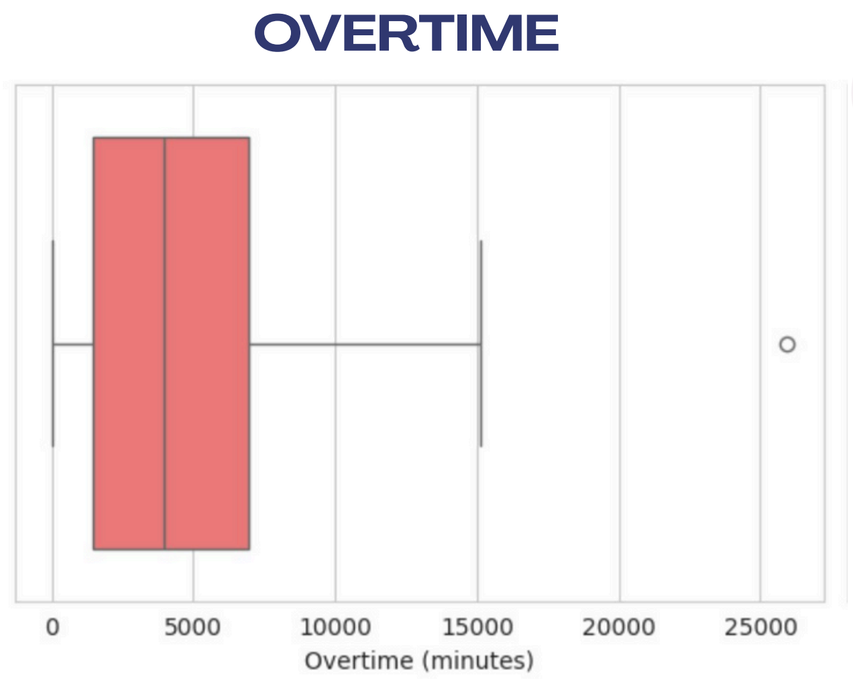
1. DATASET SUMMARY





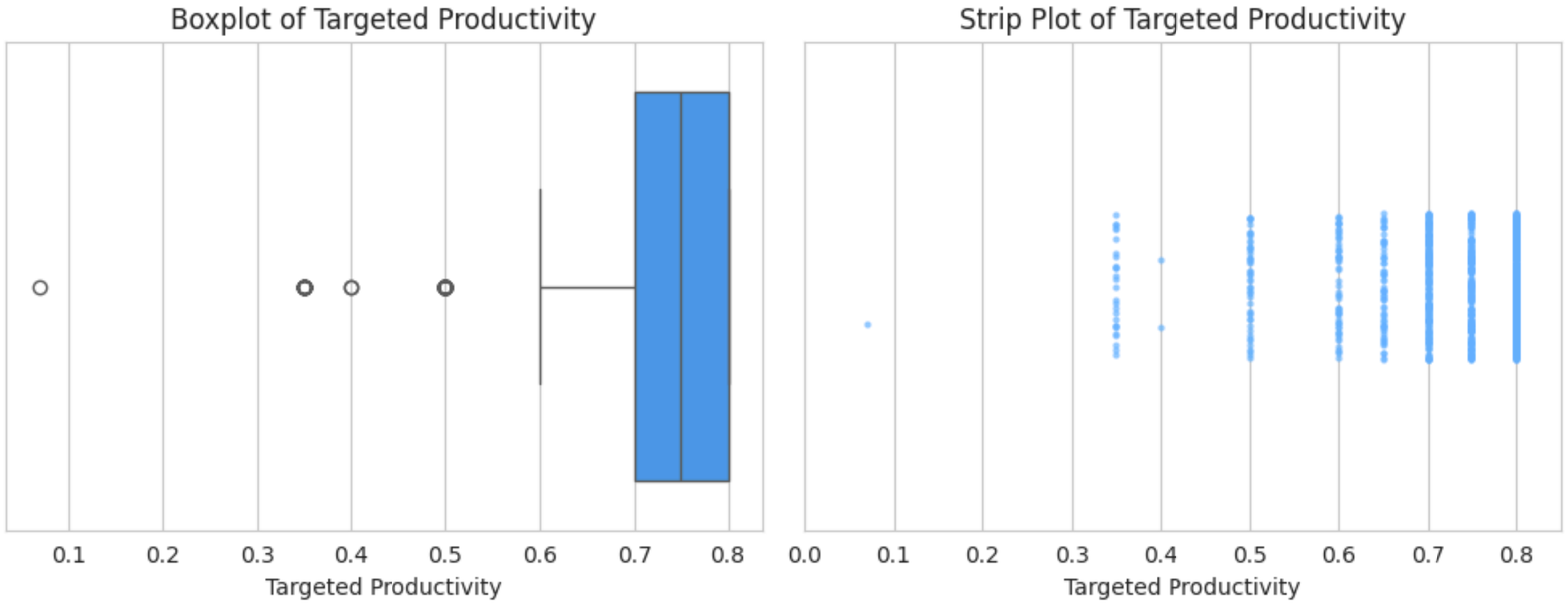
- This indicates that missing WIP data is not random!

1. DATASET SUMMARY



- extreme outlier indicates suggests an entry error rather than actual overtime.

1. DATASET SUMMARY



- extreme outlier also indicates entry error

3. CLASSIFICATION

- **Goal:** Predict the productivity level of each production day.
- **Classes:** High, Medium, Low.
- **Model Used:** Decision Tree Classifier.
- **Preprocessing:** All features were encoded, cleaned, and normalized to prepare the data for modeling.

3. CLASSIFICATION

Features Used for Classification

- targeted_productivity
- smv
- wip
- over_time
- incentive
- idle_time
- idle_men
- no_of_style_change
- no_of_workers
- actual_productivity

3. CLASSIFICATION

Decision Tree Setup

Algorithm: Decision Tree Classifier.

Splitting Criteria:

-Entropy (Information Gain)

-Gini Index

Train/Test Splits:

- 90% / 10%

- 80% / 20%

- 70% / 30%

3. CLASSIFICATION

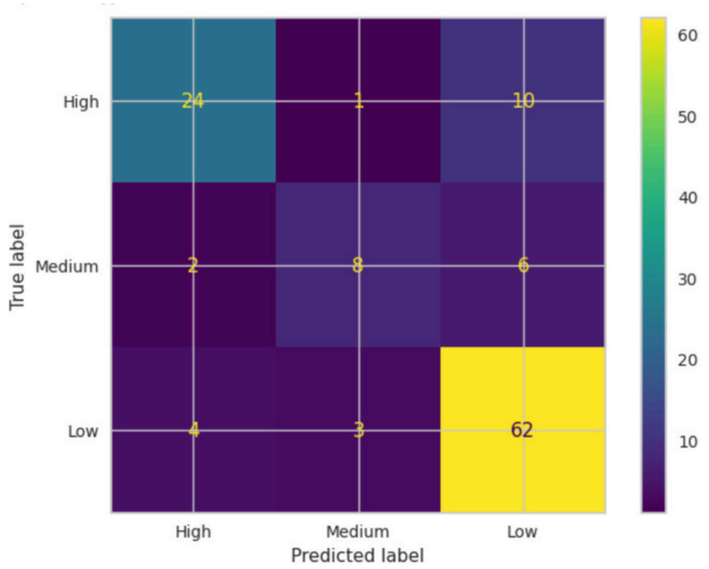
Evaluate the models of Gini Index:

Split	Accuracy	Error Rate	Sensitivity	Specificity	Precision
90% Train, 10% Test	0.7833	0.2167	0.5000	0.9615	0.6667
80% Train, 20% Test	0.7750	0.2250	0.5938	0.9357	0.5556
70% Train, 30% Test	0.7750	0.2250	0.5208	0.9357	0.5556

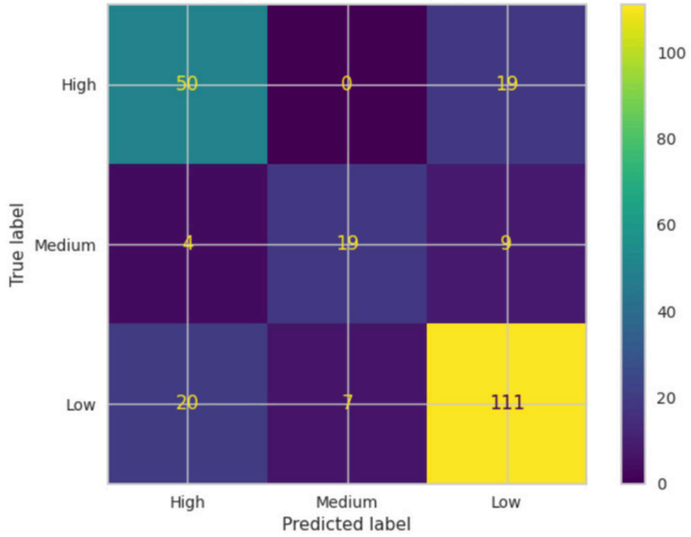
- Best accuracy and precision in 90/10.
- Best sensitivity in 80/20.
- Specificity is high in all splits.
- Gini performance is stable across 80/20 and 70/30.

CONFUSION MATRICES

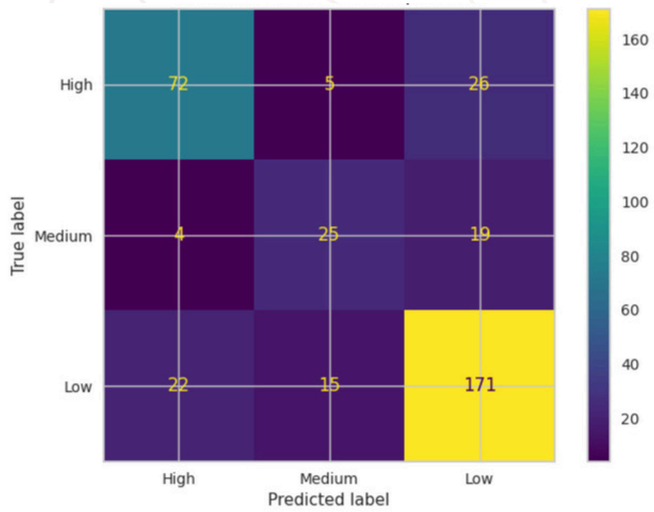
90/10 Split



80/20 Split



70/30 Split



3. CLASSIFICATION

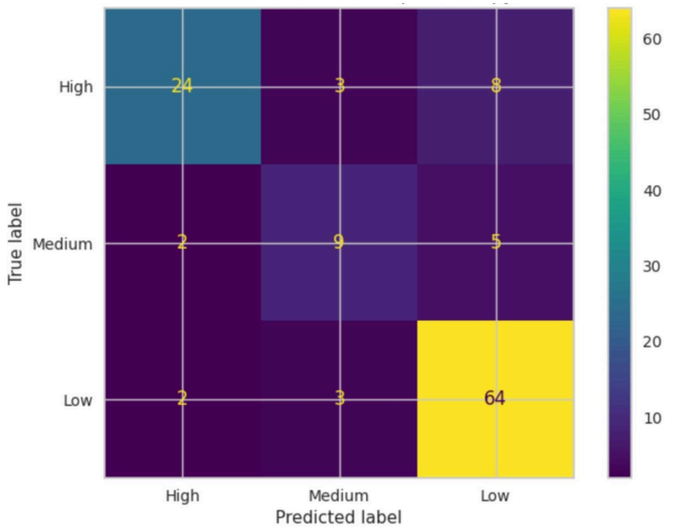
Evaluate the models of Entropy:

Split	Accuracy	Error Rate	Sensitivity	Specificity	Precision
90% Train, 10% Test	0.78.33	0.216	0.5625	0.9423	0.6
80% Train, 20% Test	0.78.33	0.216	0.5625	0.9806	0.81
70% Train, 30% Test	0.78.33	0.216	0.625	0.9296	0.577

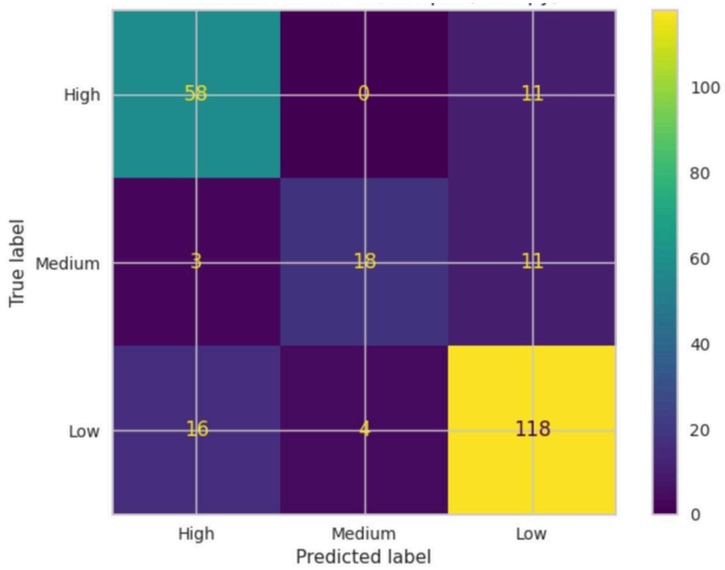
- Accuracy and error rate are identical across all splits.
- Best precision in 80/20.
- Best sensitivity in 70/30.
- Specificity highest in 80/20

CONFUSION MATRICES

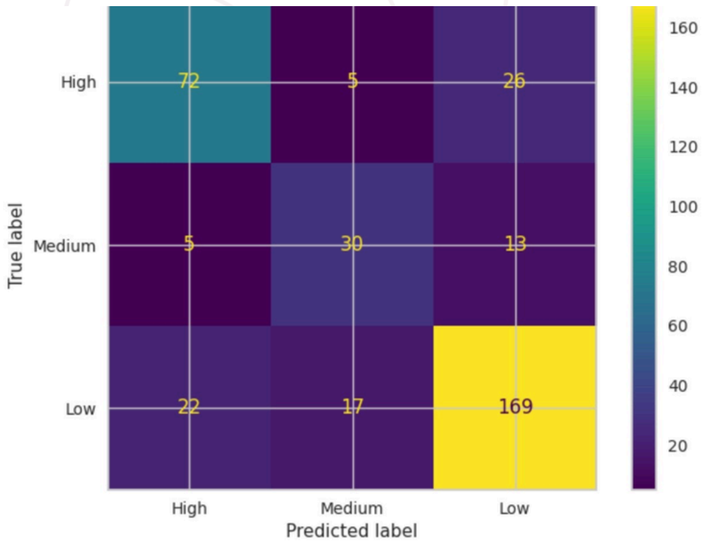
90/10 Split



80/20 Split



70/30 Split



3. CLASSIFICATION

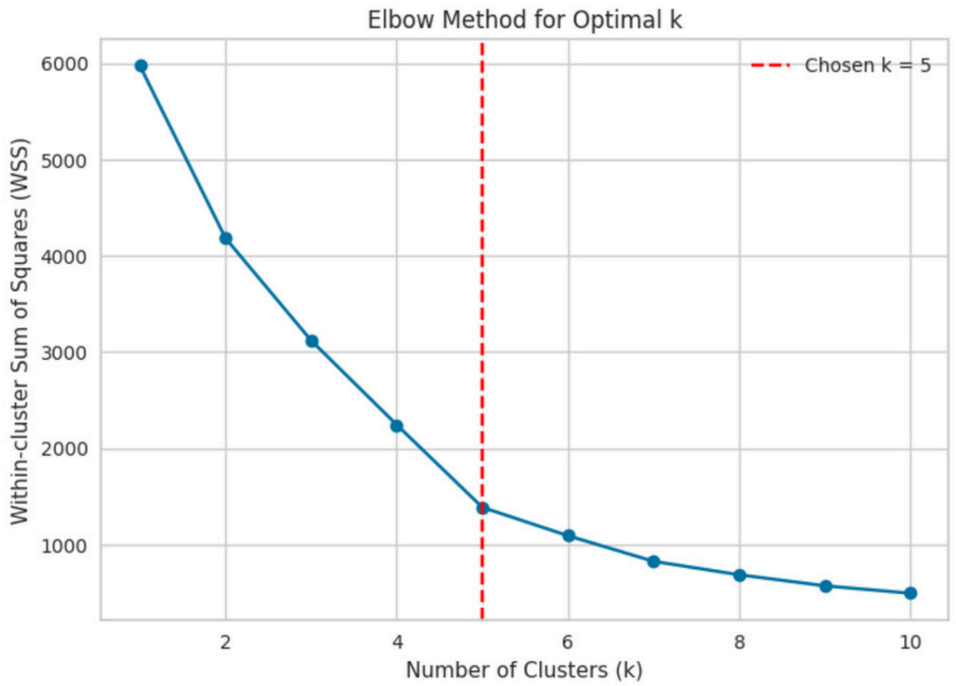
Metric	Entropy (80/20)	Gini (90/10)
Accuracy	**0.78.33**	**0.7833**
Error Rate	0.2167	0.2167
Sensitivity	0.5625	0.50
Specificity	**0.9806**	0.9615
Precision	**0.81**	0.6667

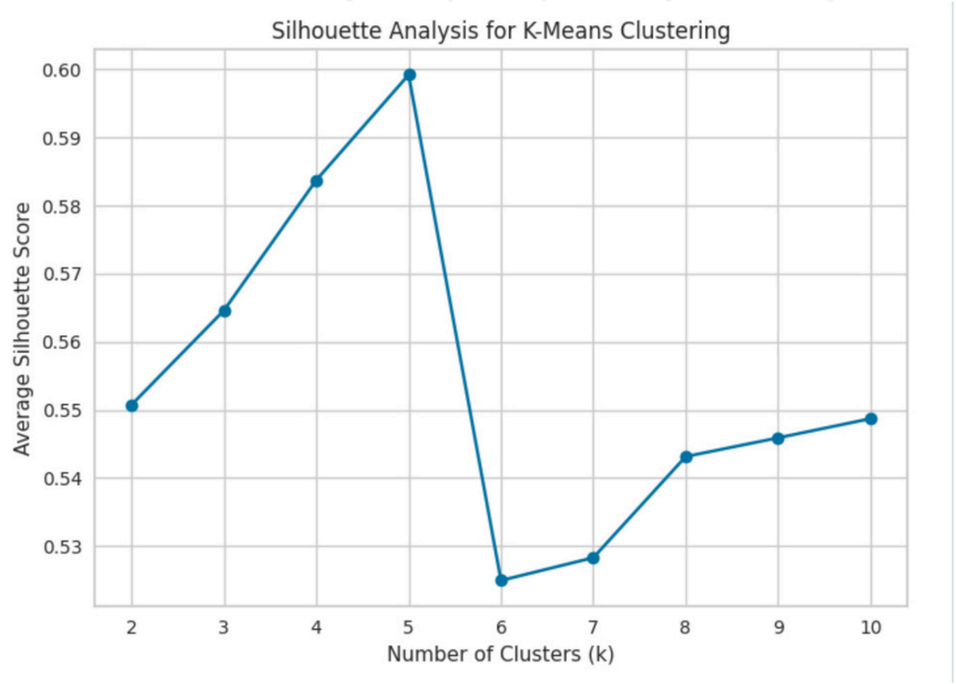
- Entropy (80/20) outperforms Gini in sensitivity, specificity, and precision.
- both criteria have the same accuracy and the same error rate

4. CLUSTERING

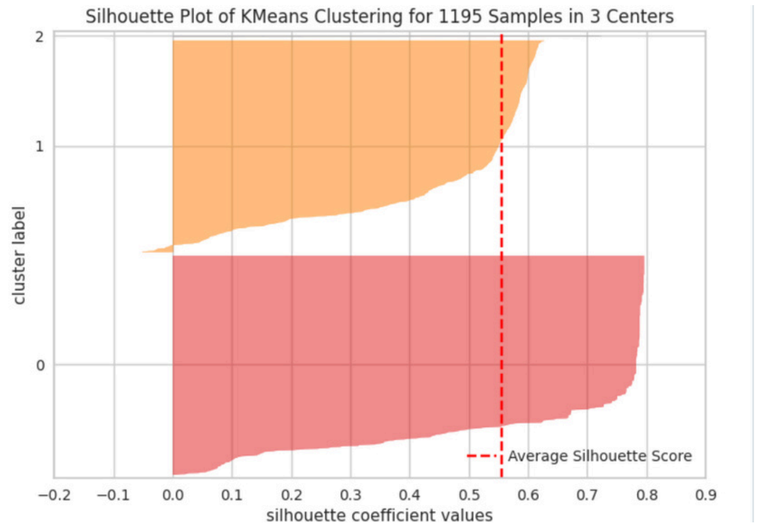
- **Goal**: Group production days into natural clusters based on productivity-related features.
- **Method Used**: Unsupervised Learning (Clustering).
- **Purpose**: Discover hidden patterns in productivity behavior without using labels.
- **Preprocessing**: All numeric features were normalized to ensure equal weight across clustering.

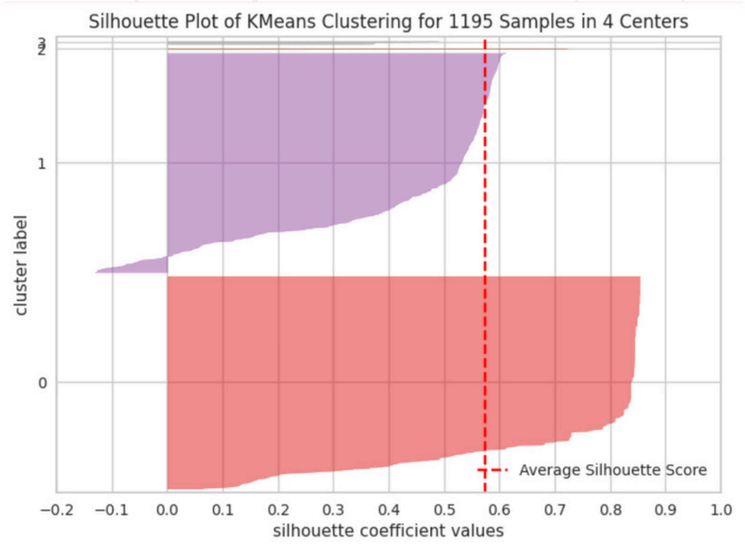
4. CLUSTERING

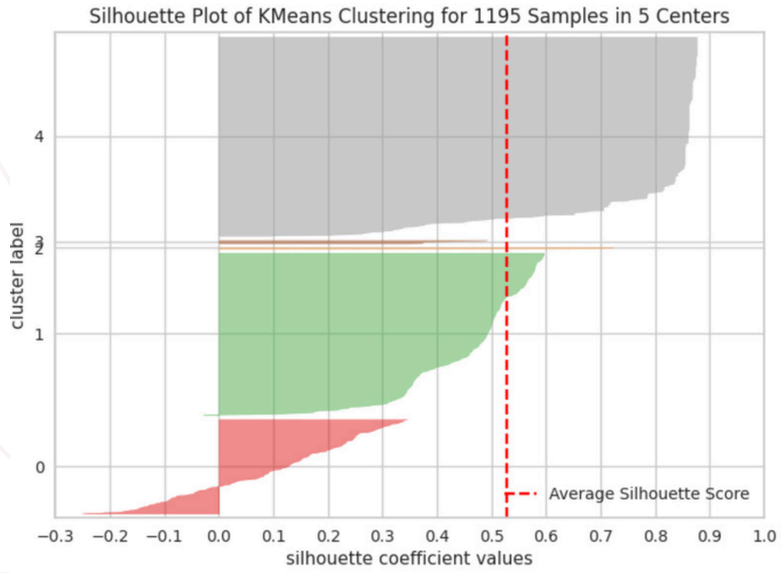


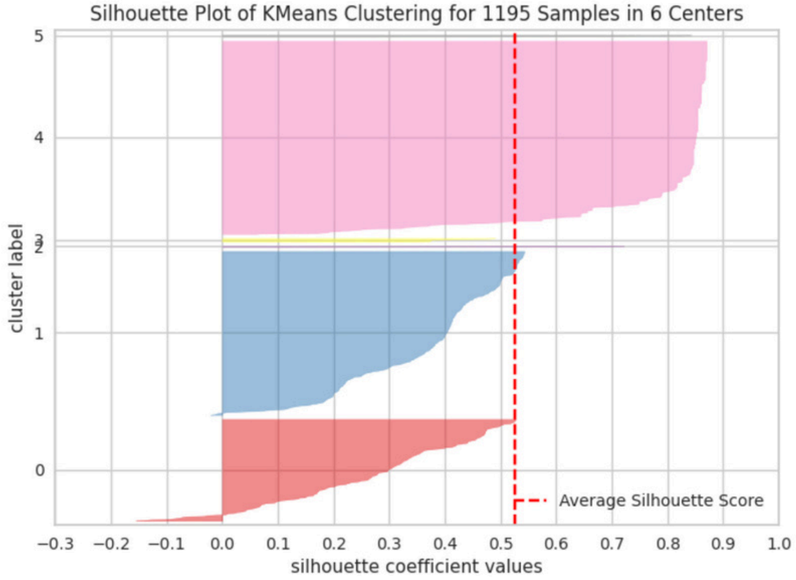


4. CLUSTERING

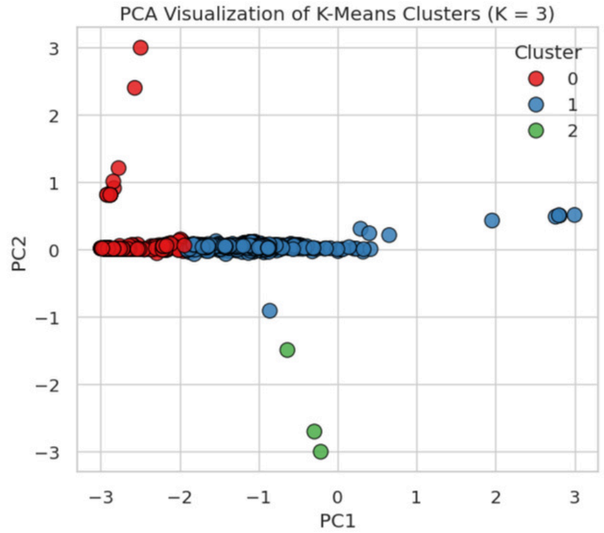


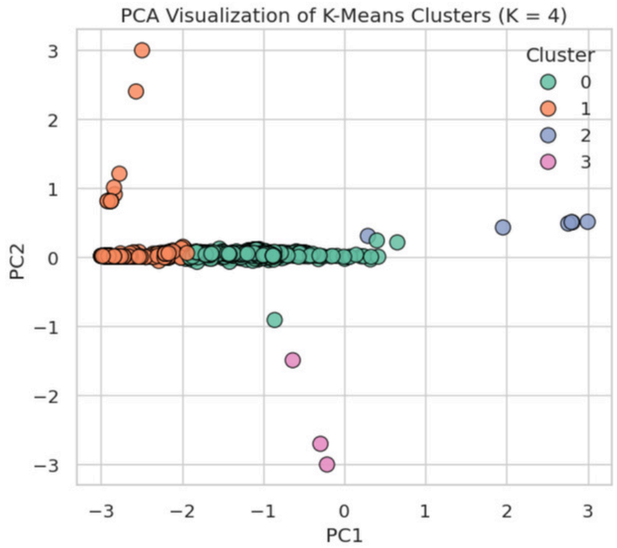


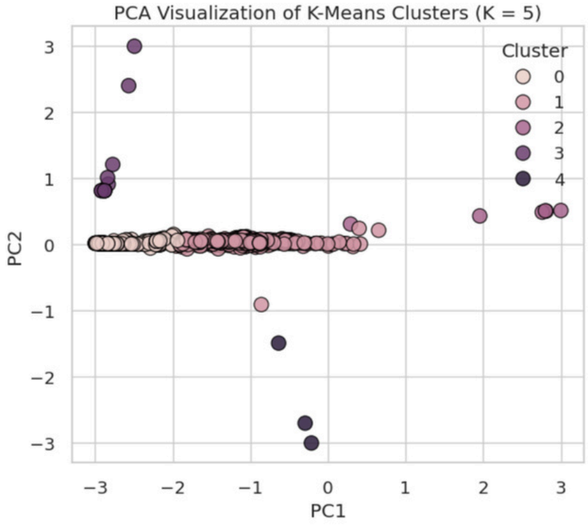


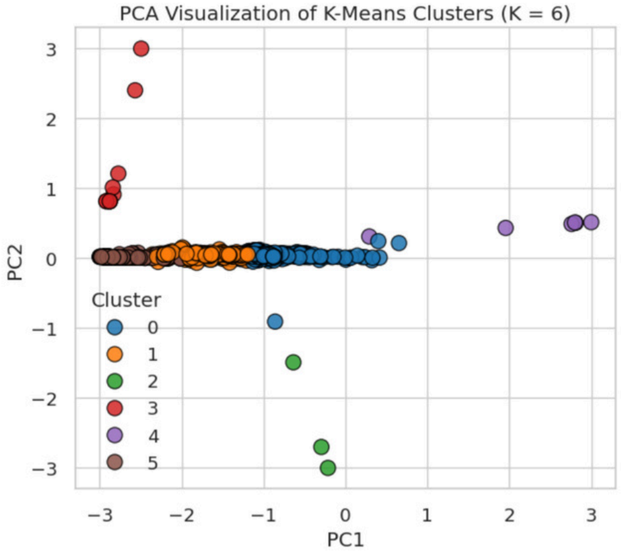


4. CLUSTERING

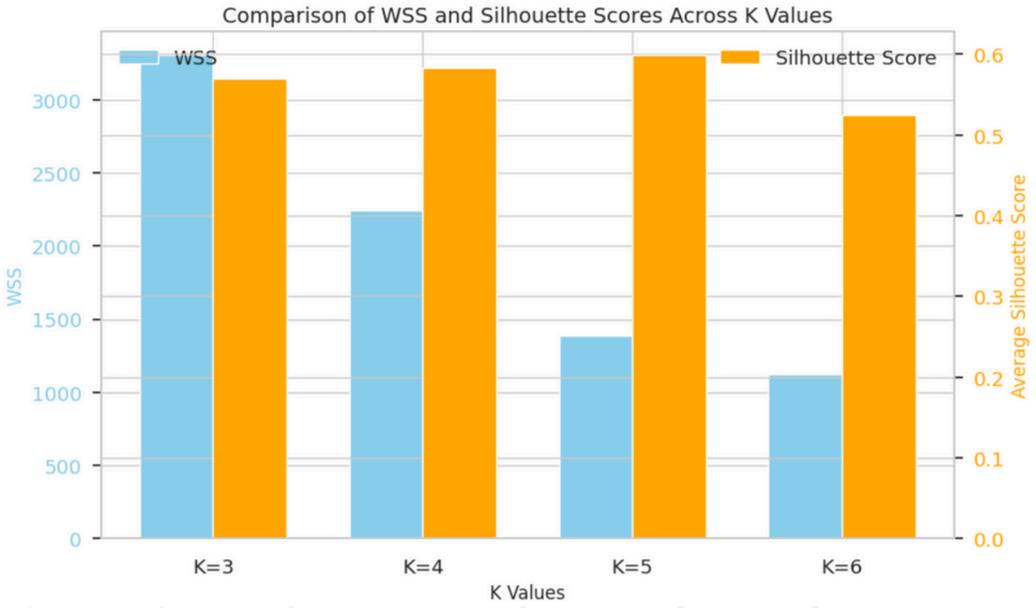








4. CLUSTERING



K	WSS	Average Silhouette Score
3	3306.566384	0.569758
4	2244.206017	0.584043
5	**1386.776440**	**0.599190**
6	1120.438397	0.524944

RESULTS & FINDINGS

- Production teams generally meet their target.
- High variability in SMV and WIP indicates inconsistent workload.
- Outliers mainly appear in style changes and targeted productivity.
- Low class achieves the highest prediction accuracy.
- Medium class remains the most challenging due to fewer samples.
- Entropy performs better in sensitivity and precision.
- Gini shows slightly higher overall accuracy.