GROUP 20

D3 PRESENTATION

CUSTOMER SATISFACTION IN FOOD DELIVERY

GROUP MEMBERS:

Abdallah Alshaqra (H00401376)
Syeda Zainab (H00386212)
Kanishka Agarwal (H00388192)
Sai Suhas Varanasi (H00403009)

R2 DATA PREPROCESSING AND EDA

Exploratory Data Analysis

Delivery Dataset:

- Strong correlation between delivery time and satisfaction (~0.81).
- Delivery time distribution peaks around ~26 minutes.
- Visualized dependencies on ratings, weather, and traffic.

Restaurant Dataset:

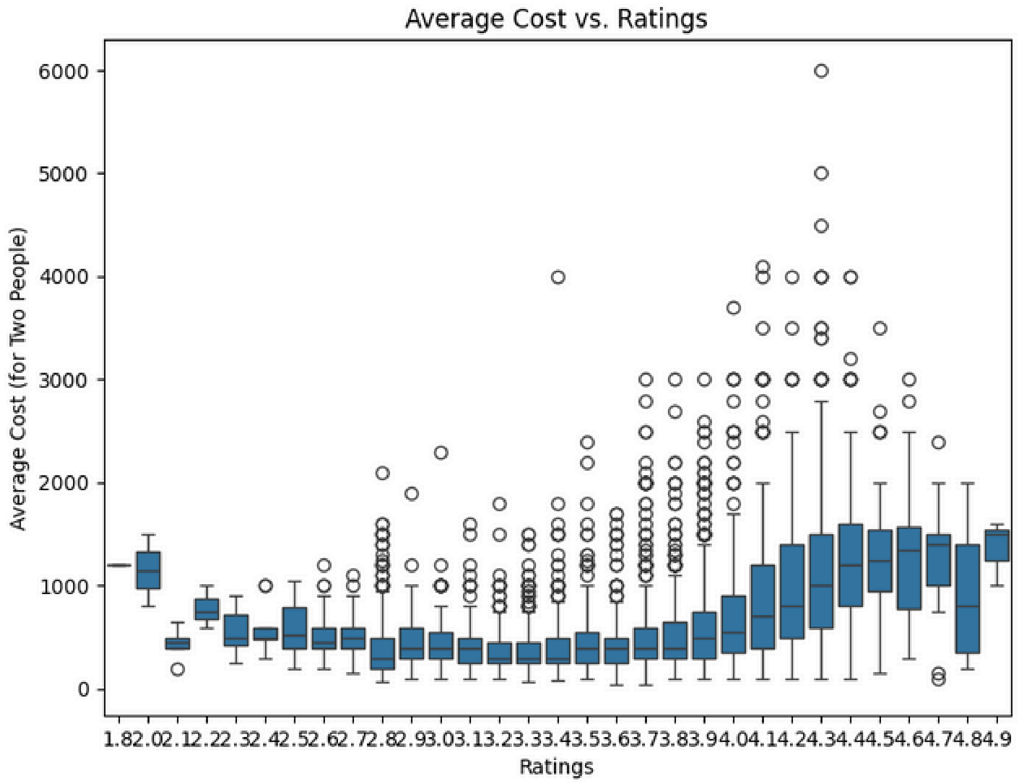
- High ratings weakly correlated with average cost.
- Most ratings cluster between 3.0 and 4.0, with outliers.
- Online orders are slightly more preferred than dine-in.

Data Preprocessing

- Resolved missing values: ~1900 in delivery ratings, traffic, and weather; gaps in restaurant ratings and cost.
- Standardized columns and corrected data types.
- Engineered satisfaction features and encoded categorical variables.
- Output: Two clean, consistent datasets for integration and modeling.

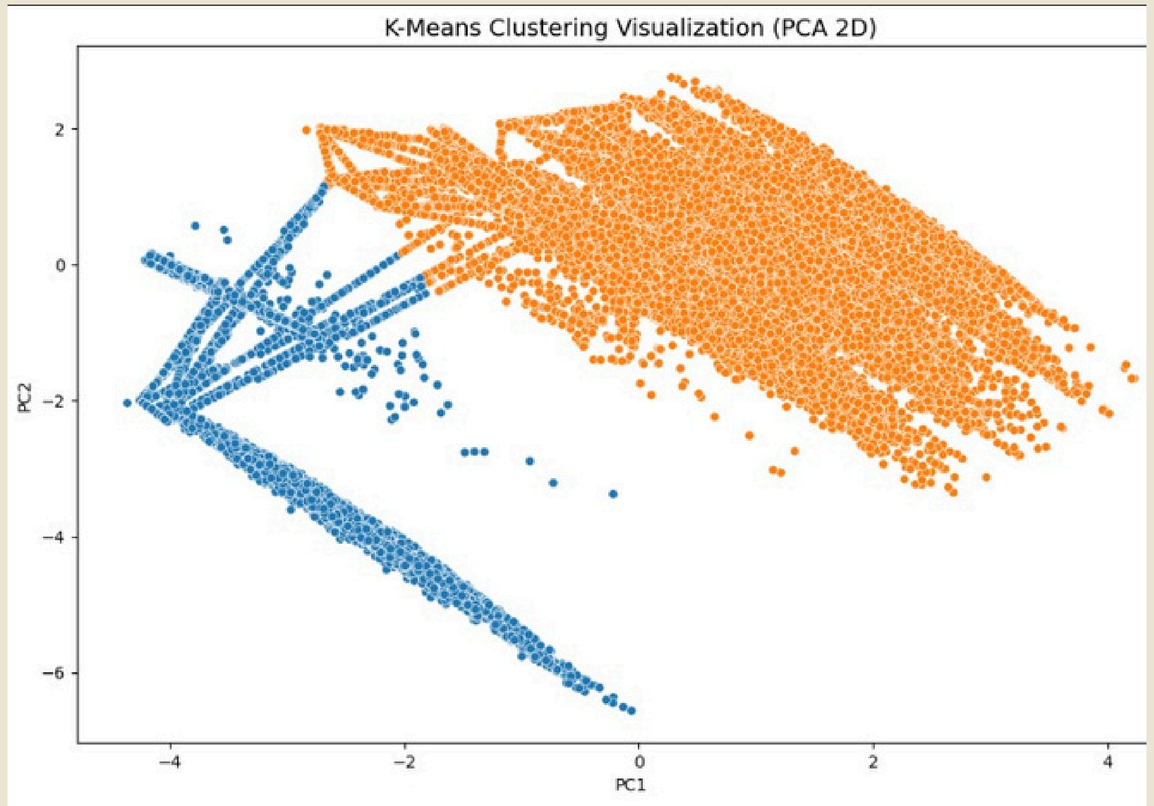
Introduction

This Section aims to analyze customer satisfaction in food delivery services using two datasets: delivery operations and restaurant data. Insights from EDA guide preprocessing to clean and prepare data for modeling.



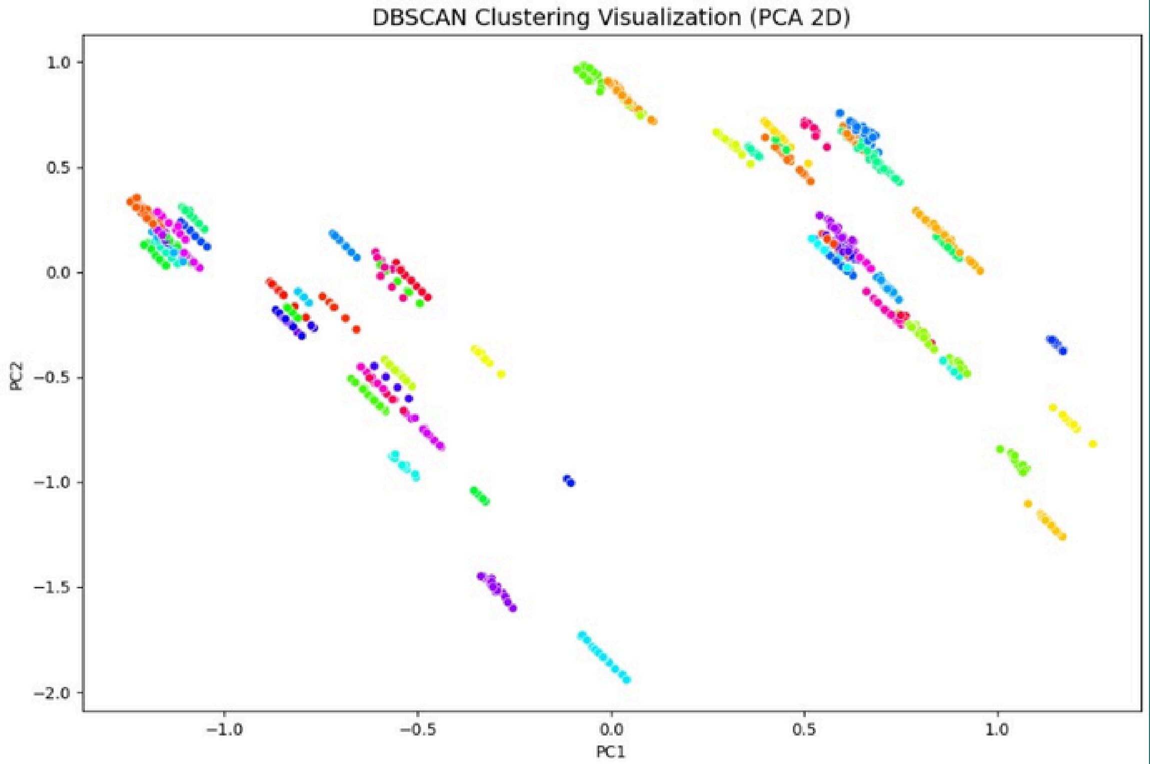
R3 CLUSTERING

K-Means Clustering for Delivery



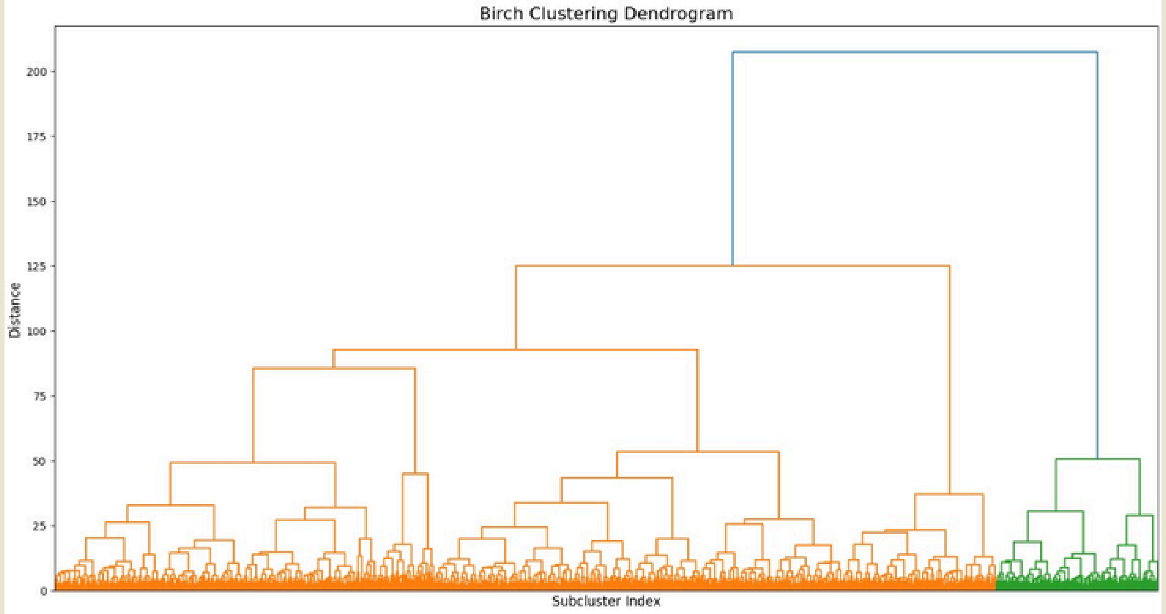
K-Means Clustering forms spherical clusters by minimizing intra-cluster variance. For delivery data, it identified two clusters validated by the elbow method. For restaurants, it also found two clusters, though they were less cohesive, showing its limitations with irregular data.

DBSCAN Clustering for Delivery



DBSCAN Clustering focuses on density to identify clusters of varying shapes and sizes. In the delivery dataset, it detected multiple small clusters with noise. For restaurants, it revealed granular insights into irregular clusters, excelling with non-linear patterns.

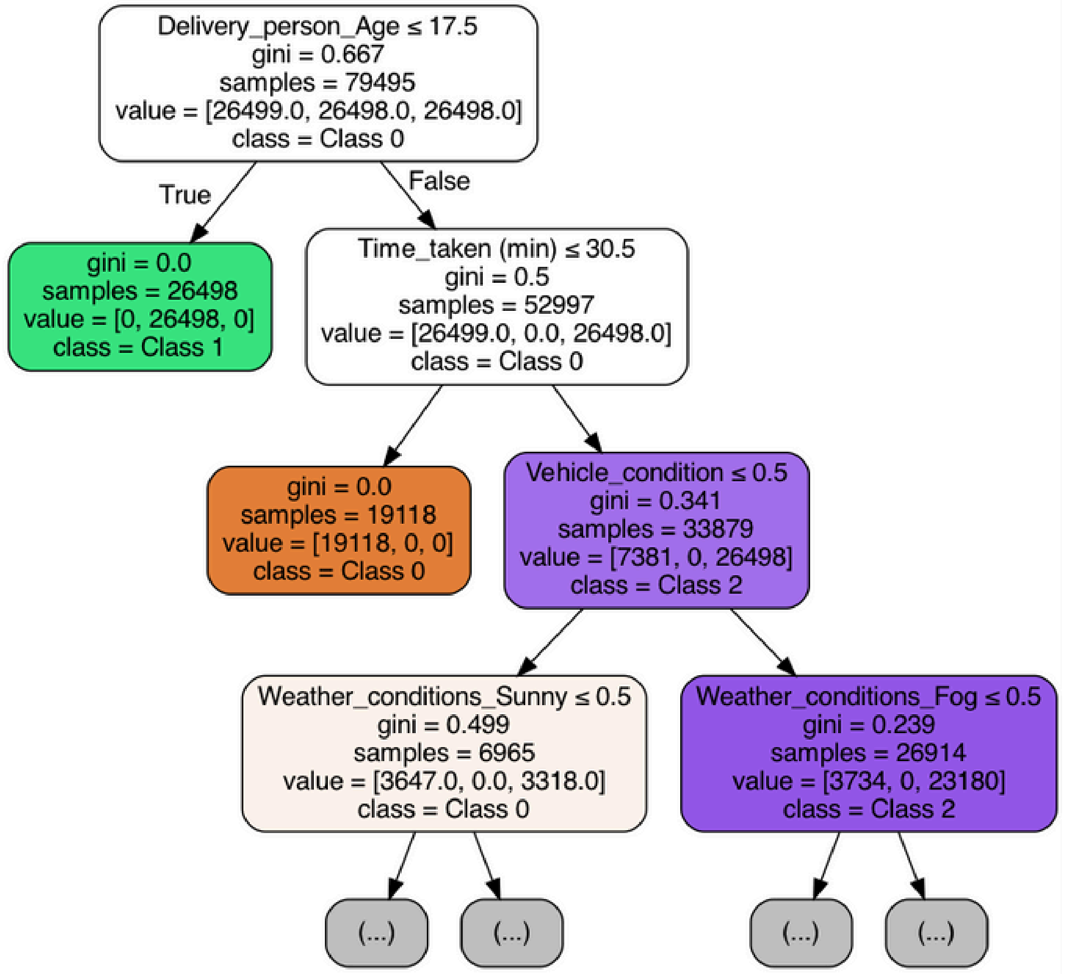
Birch Clustering for Restaurant



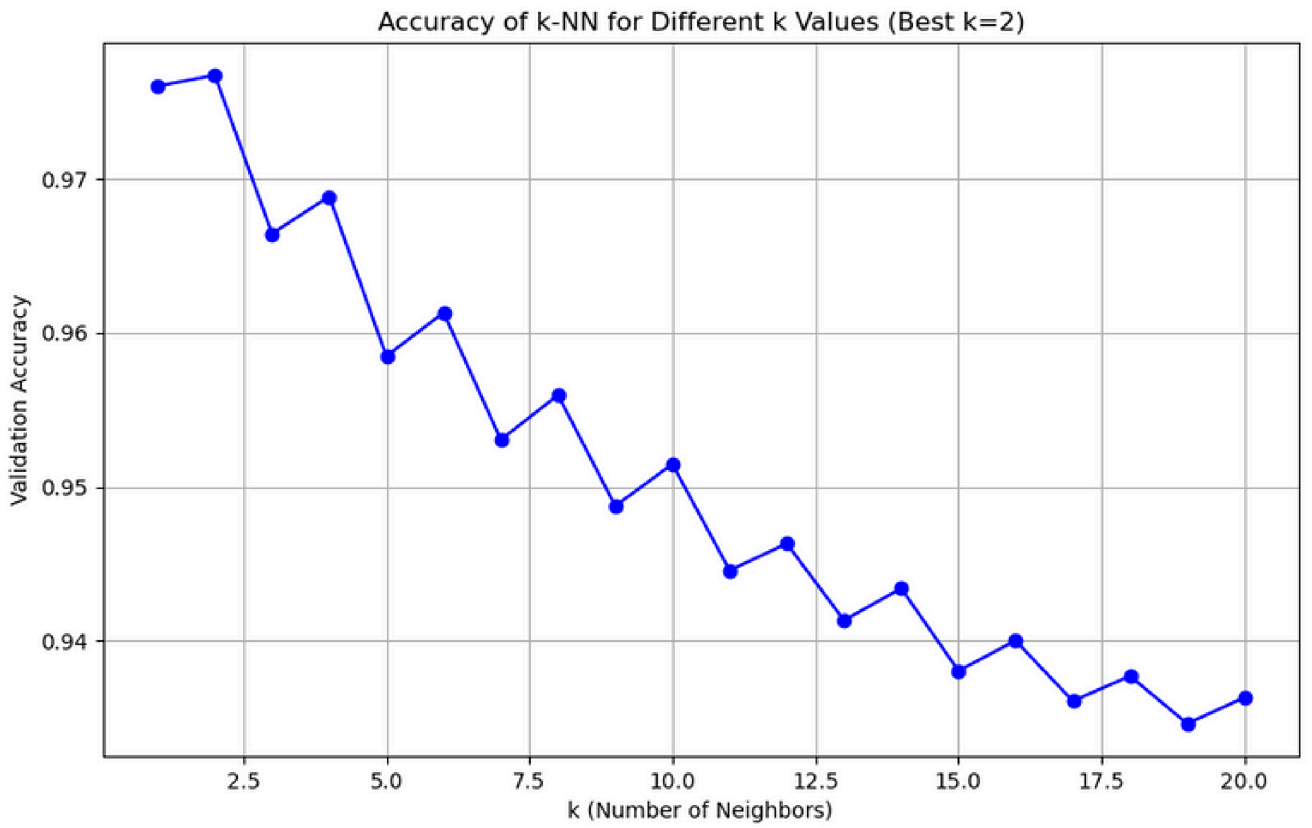
Birch Clustering combines hierarchical clustering with centroid-based methods. For delivery data, it produced two clear hierarchical clusters, while for restaurants, it highlighted consistent large and small clusters. It balances scalability and precision, making it effective for structured datasets.

R4 BASIC CLASSIFIERS AND DECISION TREES

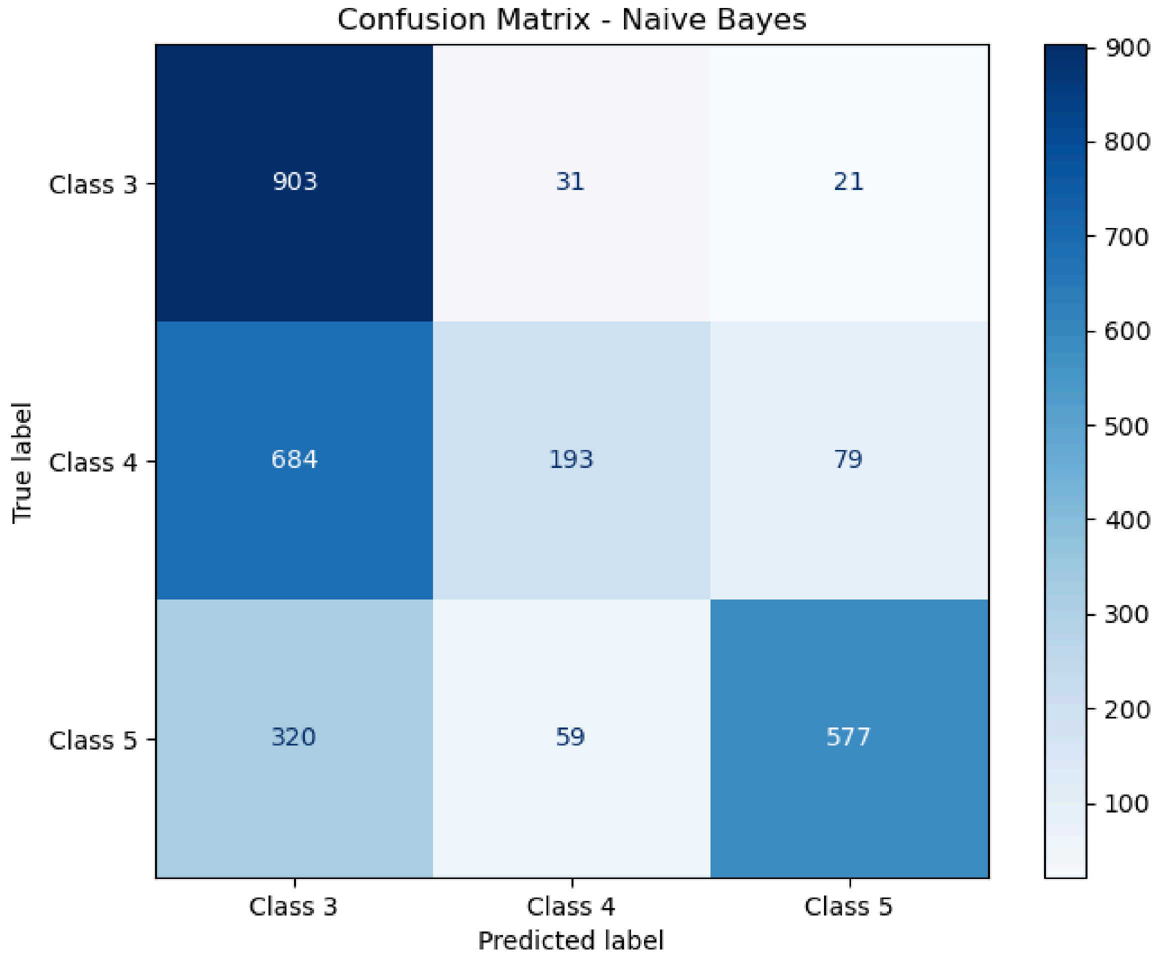
Decision Tree for Delivery



KNN accuracy for Delivery



Confusion Matrix for Restaurant



Decision Tree

The Decision Tree model splits data into decision nodes, highlighting significant features. Hyperparameter tuning, including tree depth, which improves accuracy and interpretability for classification tasks.

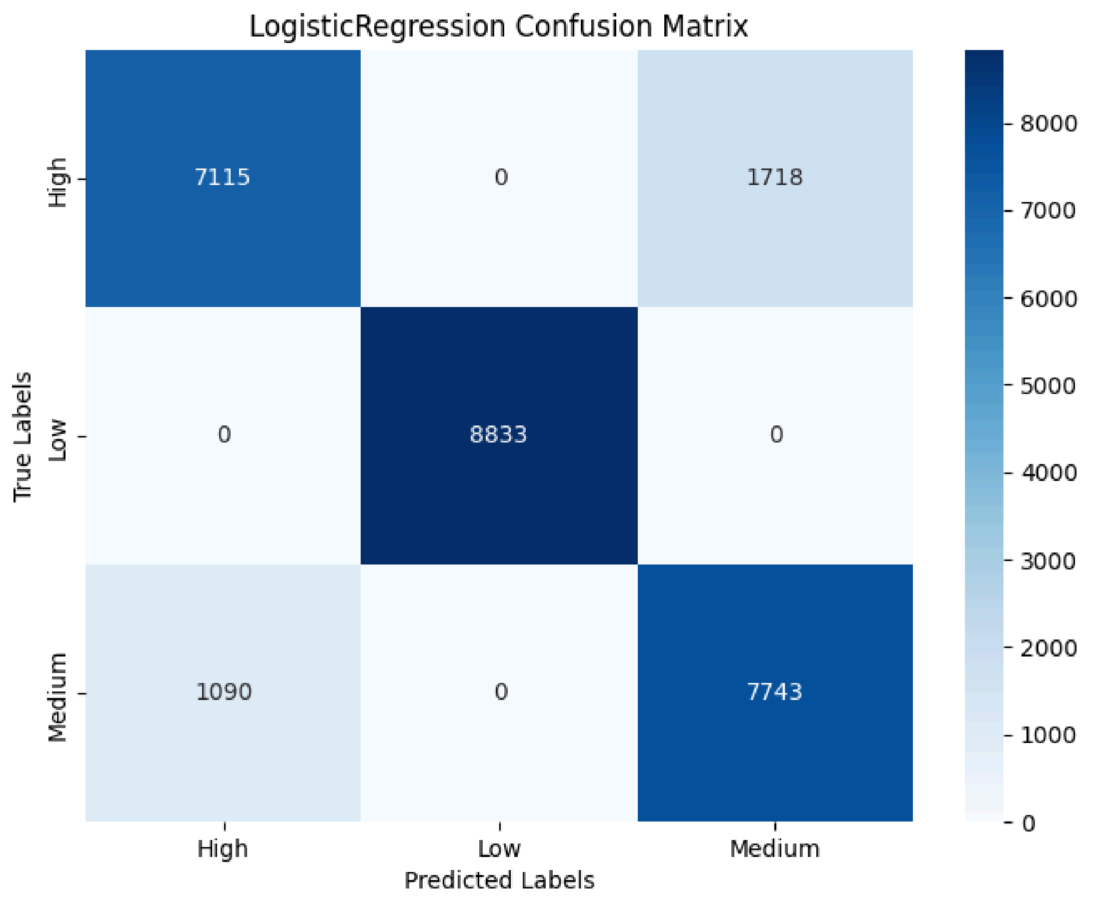
KNN (k-nearest neighbour)

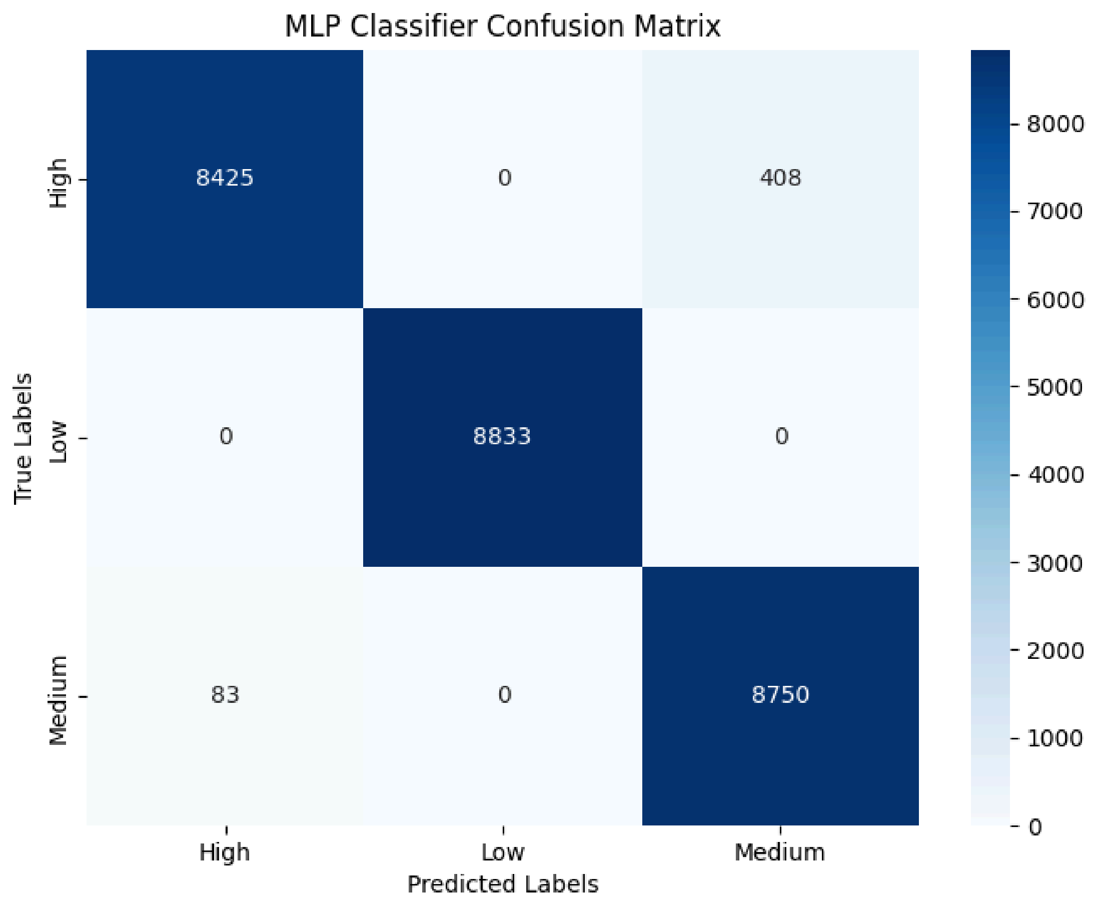
The k-NN model classified data based on similarity to neighbors. Different k-values were tested, achieving strong results for well-separated classes with scaled features.

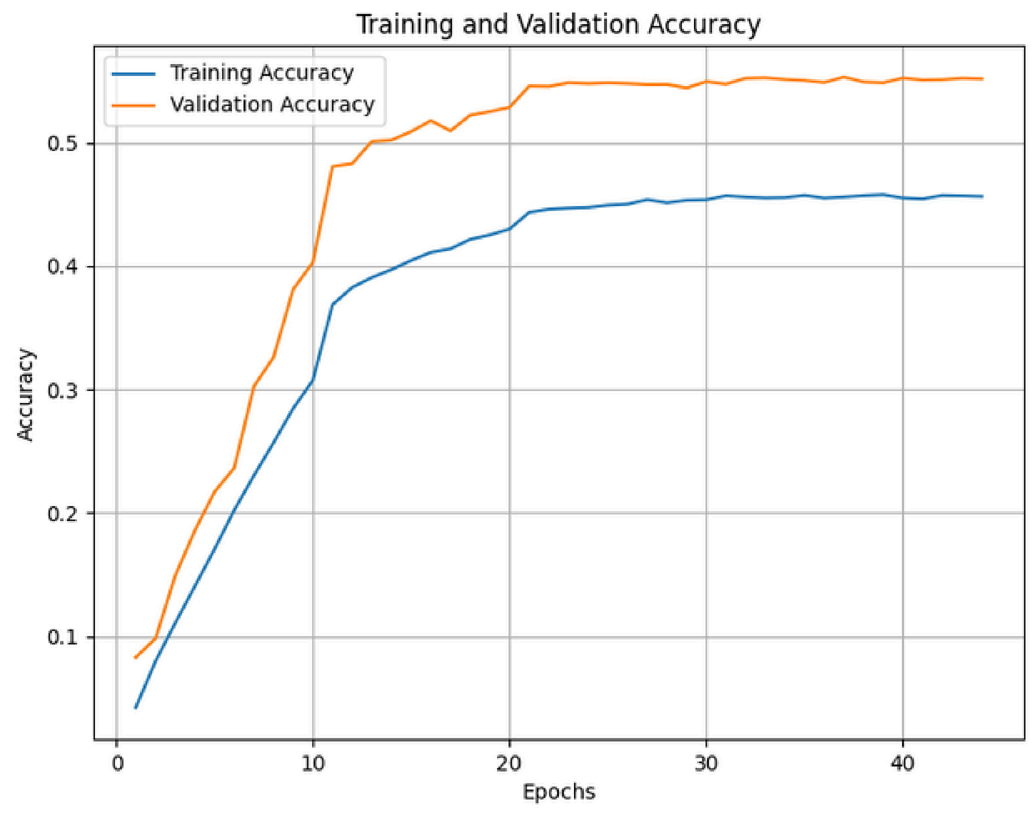
Naïve Bayes

The Naive Bayes model used probabilistic calculations with feature independence assumptions. It served as a baseline, performing well with categorical features but struggled with complex, non-linear relationships.

R5 NEURAL NETWORKS AND CNNS







Logistic regression

The Logistic Regression part evaluates a baseline model using, hyperparameter tuning via grid search, learning curves, and classification metrics (accuracy, confusion matrix, and ROC curve) to assess prediction performance.

MLP

We developed a tailored MLP classifier, which included steps for, training, and assessment, evaluation with features like learning curves, classification reports, and confusion matrices to analyze model performance."

CNNs

The CNN part involves complex feature extraction and classification tasks, showcasing its strength in capturing spatial patterns, which are particularly effective for image or sequential data.

THANK YOU