Assignment 1 Summary

Introduction

This assignment implemented the k-Nearest Neighbors (KNN) classification algorithm to predict wine quality.

1. Correlation Analysis

The correlation matrix shows how different features relate to wine quality:

- Alcohol (0.3833) is the strongest positive predictor for high-quality wine.
- Volatile Acidity (-0.2254) and Density (-0.2687) was seen to negatively impact wine quality.
- Residual Sugar and Density (0.8389) are highly correlated, meaning wines with more sugar are usually denser.
- Citric Acid (-0.0007) has almost no effect on quality.

2. Model Performance

The KNN classifier was tested with different values of k, distance metrics, and weighting methods.

K	Metric	Weighting	Accuracy
1	Euclidean	Uniform	2152.5
5	Manhattan	Uniform	2304.6
9	Manhattan	Uniform	2391.3
11	Manhattan	Uniform	2407.5

Best Model:

- k = 11, Manhattan Distance, Uniform Weighting
- Accuracy: 74%
- Precision: 0.65
- Recall: 0.88
- F1 Score: 0.75

3. Confusion Matrix & Errors

- False Positives (1184722): Many low-quality wines were misclassified as high-quality.
- False Negatives (295568): The model correctly identified most high-quality wines.
- High Recall (0.88): The model got and recorded most high-quality wines.
- Lower Precision (0.65): It sometimes misclassifies wines as high-quality when they are not actually high-quality.