

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:

- Features like alcohol content tend to have a strong positive correlation with quality.
- Volatile acidity and density were observed to negatively impact wine quality.
- Residual sugar and density showed a high correlation, indicating denser wines often contain more sugar.
- Citric acid had a negligible effect on wine quality.

2. Model Performance

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

K	Metric	Weighting	Accuracy	F1 Score
5	Manhattan	Distance	0.8357	0.8792
11	Euclidean	Distance	0.8347	0.8807
11	Manhattan	Distance	0.8337	0.8788
9	Euclidean	Distance	0.8316	0.8775
9	Manhattan	Distance	0.8265	0.8735

Best Model:

- Selected hyperparameters: k = 5, Distance Metric = Manhattan, Weighting = Distance
- Final Model Performance:
 - Accuracy: 83.57%
 - Precision: 86.94%
 - Recall: 88.92%
 - F1 Score: 87.92%

3. Confusion Matrix & Errors

- False Positives (88): Some lower-quality wines were incorrectly classified as high-quality.
- False Negatives (73): A small number of high-quality wines were misclassified.
- High Recall (0.8892): The model successfully identified most high-quality wines.
- High Precision (0.8694): The classifier effectively distinguished high-quality wines from lower-quality ones.