

AI3011: Lab Assignment Report

Lab Assignment 06

Pratik Rana
U20220108



Declaration

I, Pratik, certify that this project is my own work, based on my personal study and research and that I have acknowledged all material and sources used in its preparation, whether they be books, articles, reports, lecture notes, and any other kind of document, electronic or personal communication. I also certify that this project has not previously been submitted for assessment in any academic capacity and that I have not copied in part or whole or otherwise plagiarised the work of other persons. I confirm that I have identified and declared all possible conflicts that I may have.

Dated: 01 - 03 - 2023

Pratik Rana

Answers

Q1: What is the importance of the 'k' value in KNN?

Ans. It determines the number of neighbours used for classification, impacting model complexity and decision boundary. Large K produces smoother boundaries but if K is too large, the algorithm will always predict the majority class.

In KNN for regression, When k is too small, the regression model might Overfit but if K is too large, the regression model might underfit.

Q2: How do you choose the optimal 'k' value in KNN?

Ans. Optimal k in KNN is often chosen using the hit-and-trial method by calculating the accuracy/other performance metrics for each value of K. It is usually chosen to be odd to avoid ties.

Q3: How does the choice of 'k' affect the bias-variance tradeoff in KNN?

Ans. Smaller k increases model complexity, reducing bias but increasing variance. Larger 'k' reduces variance but may increase bias.

Q4: Can KNN be sensitive to outliers? Explain.

Ans. Yes, It can be sensitive to outliers because it relies on distance metrics, where outliers can significantly affect neighbour selection.

Q5: How does the scale of features affect the performance of KNN?

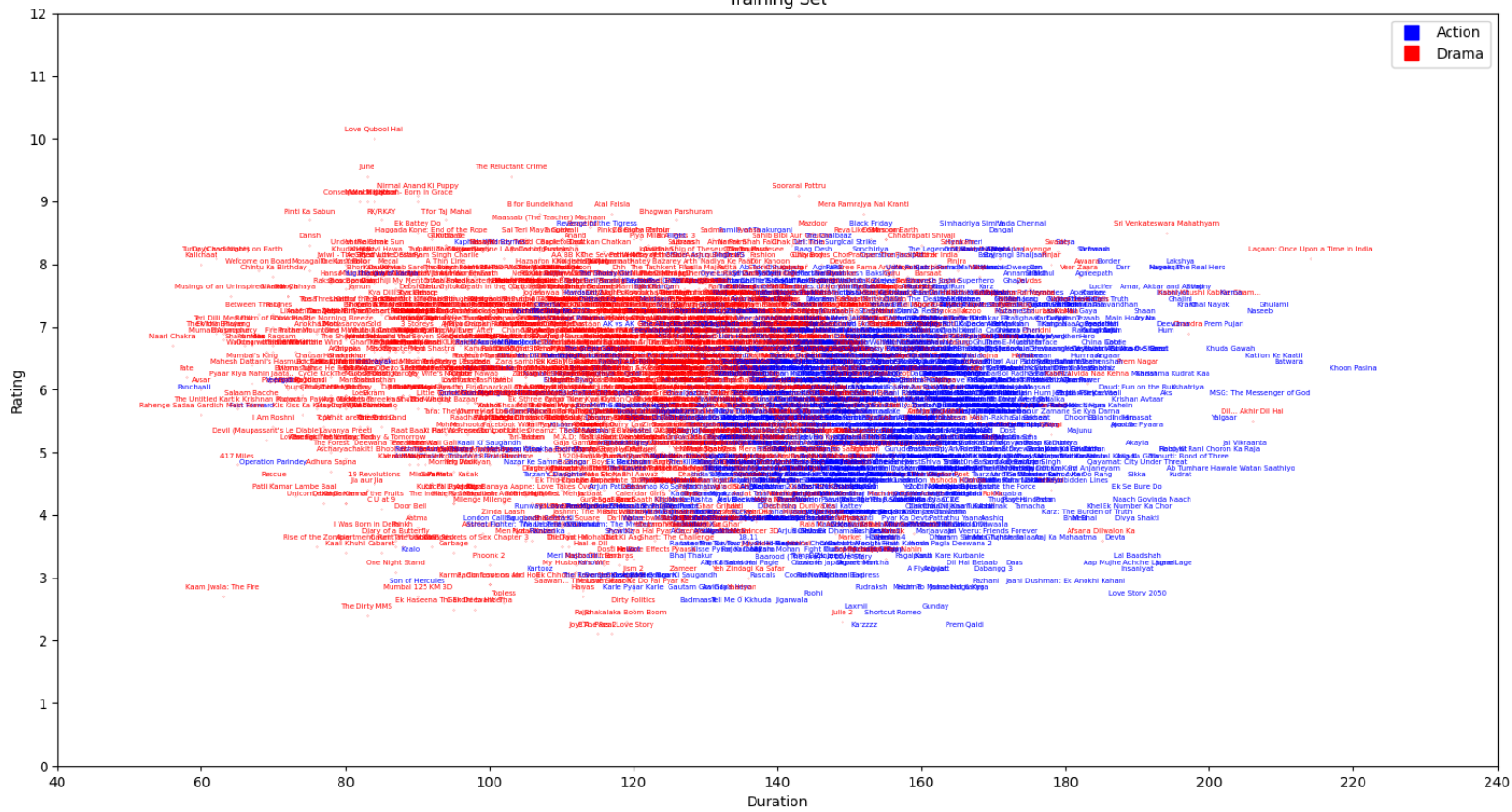
Ans. It affects KNN as it uses distance metrics. Large-scale features can dominate distance calculations, requiring normalisation for balanced performance.

References:

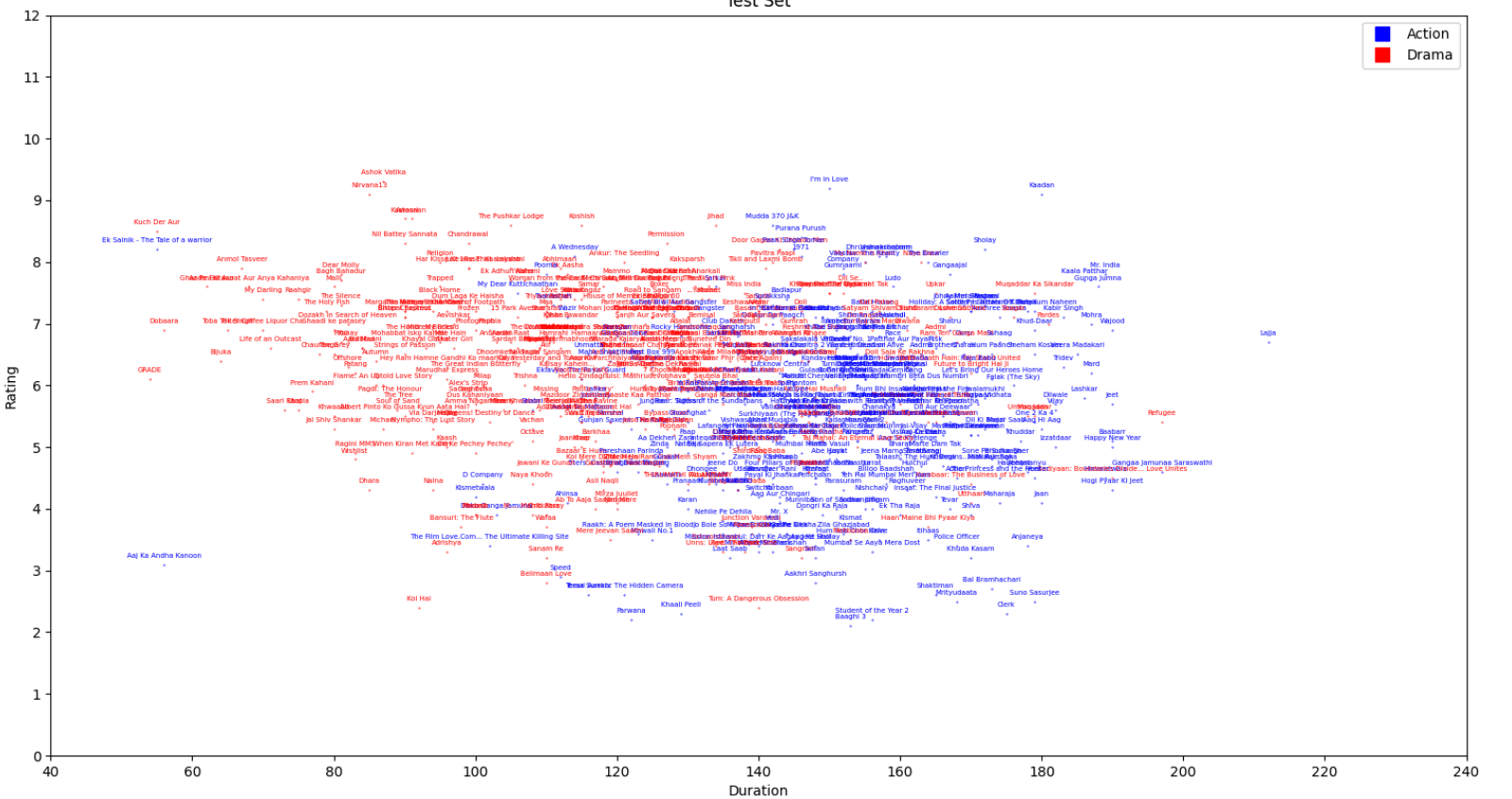
1. Lecture Slides

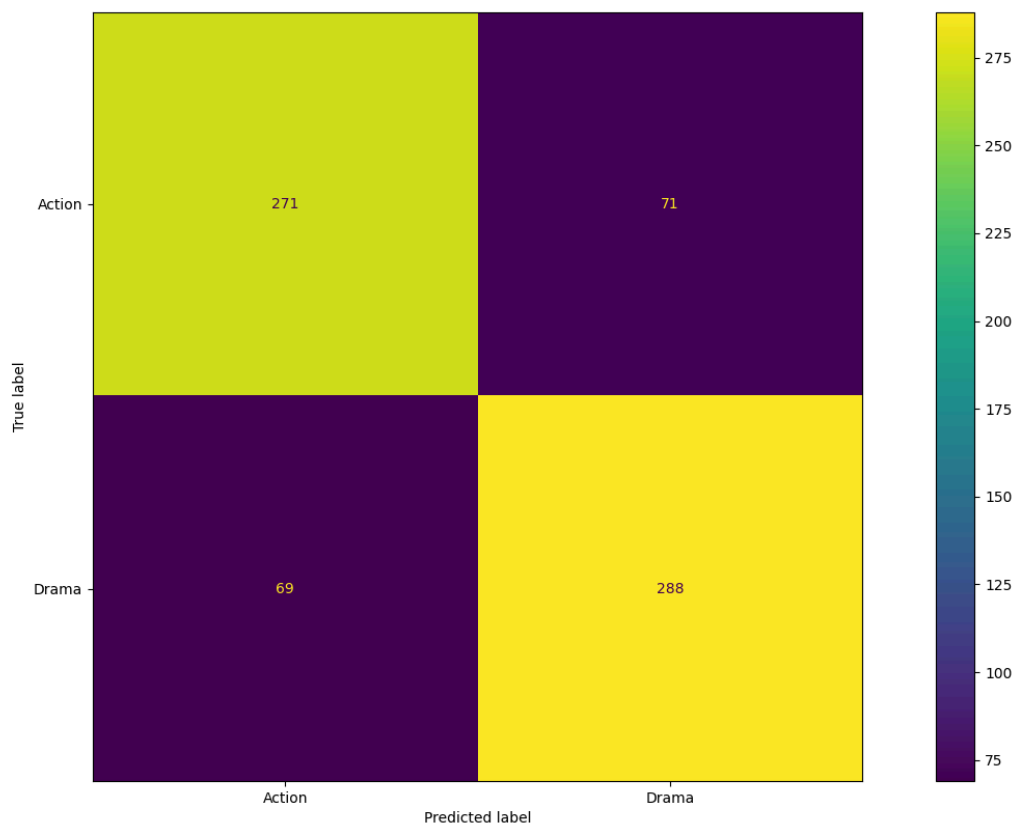
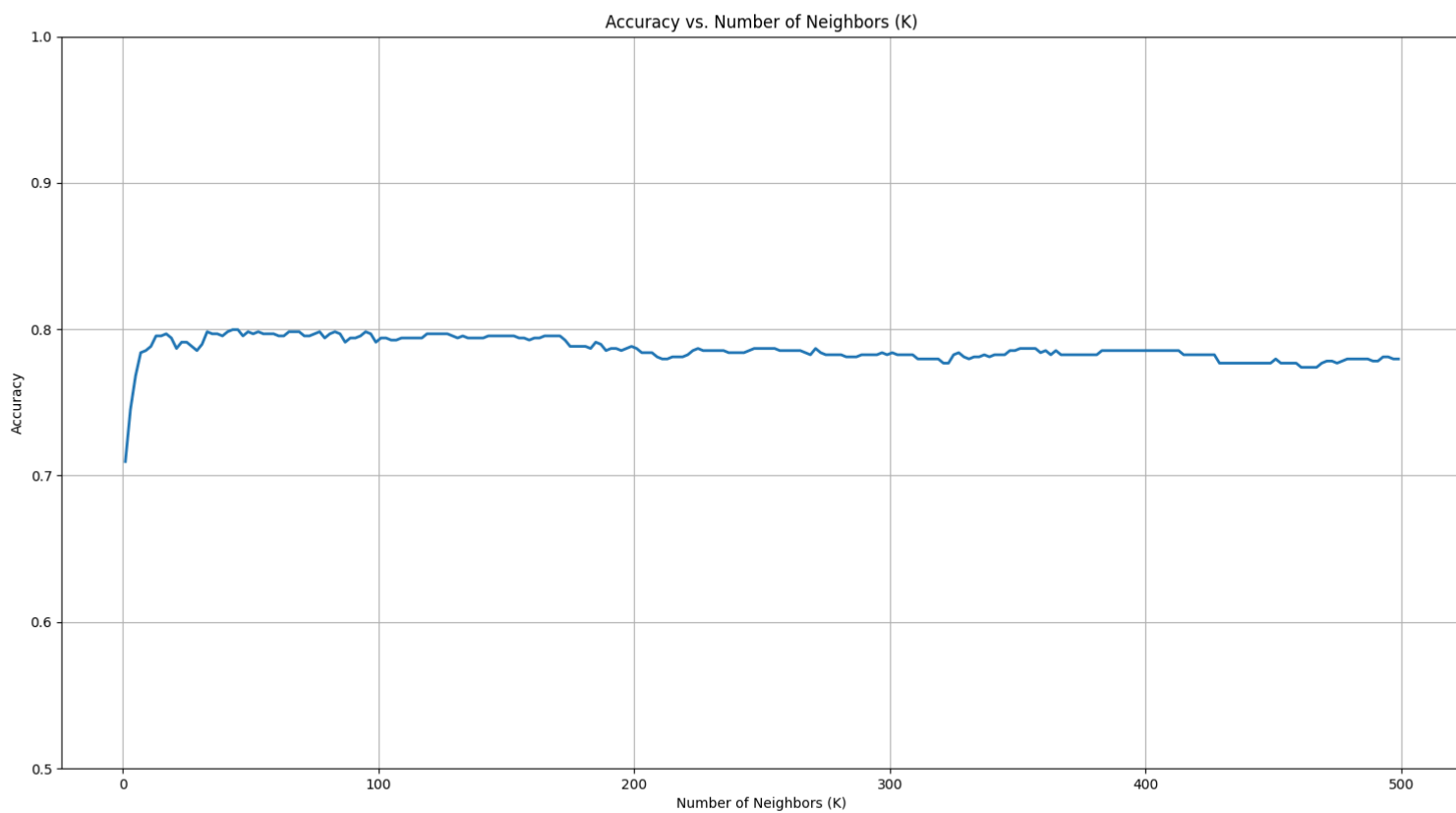
Pictures Below

Training Set



Test Set





Best K: 43 and Accuracy: 0.7997138769670958

Confusion Matrix:

Scores for each Genre:

Precision: [0.79705882 0.80222841]

Recall: [0.79239766 0.80672269]

F1 Score: [0.79472141 0.80446927]

Overall Score for both Genres:

Precision: 0.7996990855830419

Recall: 0.7997138769670958

F1 Score: 0.799699931522005