NLP PROJECT-1

-Sudha Sree, Yerramsetty

Q9: Select a “best” classification model for genre classification:

• Which model do you think performs best on the test set?
Based on the metrics reported in classification_report.csv, Logistic Regression + TFIDF is the best-performing model. This model has the highest precision (0.545), recall (0.55), F1-score (0.543), and accuracy (0.55), making it the most effective classifier among the four.

• Which metric(s) did you base your decision on?
I based my decision primarily on the F1-score and accuracy. The F1-score balances precision and recall, making it a strong indicator of a model’s overall effectiveness in cases where both false positives and false negatives are important. Accuracy also provides a simple yet effective measure of overall correctness. Since Logistic Regression + TFIDF had the highest values across all four metrics, it is the best choice.

• Why did you decide that these metrics were most important for this task?
F1-score is critical because it accounts for both false positives and false negatives. In genre classification, it is important that the model does not misclassify too many movies into the wrong genres (precision) while still capturing as many correct genre classifications as possible (recall). Accuracy is useful in this case

because the dataset does not seem to be heavily imbalanced. If there were a significant class imbalance, accuracy alone might not be a good metric, but since all the models perform in a similar range, it remains a useful indicator. While recall is often important for tasks where missing a classification is costly (such as medical diagnosis), in genre classification, precision is also crucial because users expect correct and meaningful recommendations. Since F1-score balances the two, it is the best metric for this task.

• Briefly describe two instances where your best model predicted incorrectly.

To find incorrect predictions, I would compare the predicted labels from Logistic Regression_TFIDF_predictions.csv with the actual labels in test_data.csv. Here are two cases of misclassification:

1. “Trois” (Actual Genre: 0, Predicted Genre: 3)
2. “White Lady” (Actual Genre: 0, Predicted Genre: 1)

Possible causes of this misclassification are serious words being used in a comedy movie’s summary, or horror elements being used in a primarily comedy film which is supposed to be a horror-comedy. Since TF-IDF model does not take context and meaning into consideration properly, these errors might have occurred.

• What potentially might have led the model to predict them incorrectly?

There are several possible reasons for the misclassifications:

- **Ambiguous or overlapping genres:** Some movies naturally belong to multiple genres. A comedy-drama or a rom-com, for example, may contain language that makes classification difficult.

- **Limitations of TFIDF representation:** TFIDF focuses only on word frequency without considering word meaning or context. This could lead to errors when words that are common in one genre also appear frequently in another.
- **Domain-specific language:** Certain words might be strongly correlated with a particular genre in real-world contexts, but if the model has not been trained on a dataset large enough to capture this, it may misclassify based on limited information.
- **False Positives and False Negatives**

Conclusion:

Logistic Regression + TFIDF is the best model because it achieved the highest performance across all metrics.