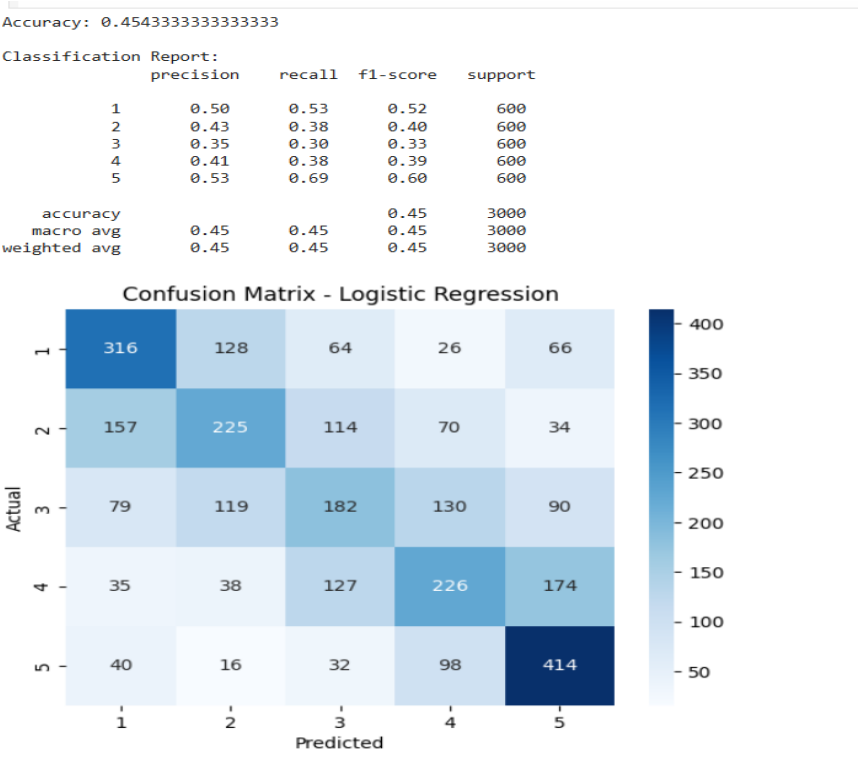


CROSS TEST REPORT

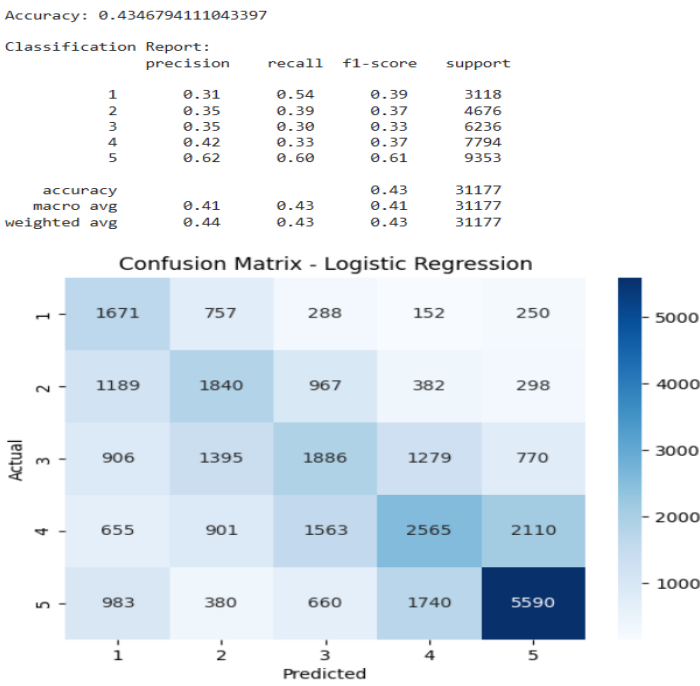
SUMMARY

In this project, two models were trained and evaluated. Model A achieved an accuracy of 45% whereas Model B achieved 43% accuracy. Although both models showed relatively low performance, Model A performed better than Model B.

Model A tested using imbalanced dataset:



Model B tested using the balanced dataset:



OBSERVATIONS.

- When trained on one dataset and tested on a different dataset, the performance of both models dropped sharply.
- The drop indicates poor generalization across domains.
- Model B, even though it scored 52% before cross test, still struggled heavily on unfamiliar data.
- The model accuracy dropped because each dataset uses different vocabulary, style, and context, so the model learned patterns specific to the training domain.
- This shows the need for combined or merged datasets for better cross domain performance.

RECOMMENDATION: choosing the model

Based on initial evaluation, Model A should be deployed because it has better accuracy (45%) compared to Model B (43%) on its own test set. Even though both models struggle in cross testing. Between the two, Model A is recommended for deployment because it performed better overall with a higher accuracy (45%) and slightly better consistency in predictions.

Although the performance is still not ideal, Model A is the best choice among the options tested and can serve as a baseline model. It can be deployed with the intention of further improvement through hyperparameter tuning and balanced training data in future work.