Replication Instructions

Replication Guide for Experimental Results:

1. Open and run the provided Google Colab notebook.

2. Load the correct dataset (e.g., tensorflow.csv) and verify it contains columns 'Title', 'Body', and 'class'.

3. Ensure the text preprocessing, TF-IDF vectorization, and SMOTE steps are correctly included as per the script.

4. Random Forest hyperparameters are tuned using GridSearchCV with 3-fold cross-validation.

5. Each model is evaluated using Accuracy, Precision, Recall, F1-score, and ROC-AUC.

6. Repeat the training and evaluation 10 times using different random_state seeds (0–9).

7. Store each run's metrics and compute the average across all runs.

8. Generate the following outputs:

- Bar chart comparing model metrics

- ROC curve showing AUC

- CSV result file with all metrics (e.g., tensorflow_RF_results.csv)

9. Run the paired t-test between F1-scores of baseline and Random Forest to confirm statistical significance.

Ensure the model and dataset match the ones used in the report (TensorFlow is primary focus).

All results, plots, and CSVs used in the report can be regenerated with the above procedure.