

Replication

To replicate the experiment and results:

1. Upload Datasets to Google Drive
Place the five datasets (tensorflow.csv, pytorch.csv, etc.) into your Drive in an accessible folder.
2. Open the Google Colab Notebook
Ensure all cells are run in sequence, beginning with mounting Google Drive.
3. Select Dataset
Modify the following line to switch datasets: `project = 'tensorflow'`
4. Run the Full Pipeline
Execute all code blocks of code: preprocessing, TF-IDF, SMOTE oversampling, train test split, evaluation of the model, visualising the results, statistical analysis.
5. Export Results to the CSV file; automatically saves based on which project dataset you are on.