Replication Instructions

Follow the steps below to replicate the experiments and results presented in the report:

1. Setup Environment:

- Install Python 3.8 or later.
- Install dependencies using pip:

Bash: pip install pandas numpy matplotlib seaborn nltk scikit-learn xgboost transformers datasets torch sentence-transformers

2. Prepare Dataset:

- Download tensorflow.csv or pytorch.csv.
- Place the dataset in the data/ folder.

3. Run the Code:

- Open either notebook:
  - Run first Bug_Report_Classification_Bert.ipynb to run DistilBERT.
  - NB+SVM_+XG+DistilBERT.ipynb to run all models
  - Place it in notebooks/ folder.
- Run all cells top to bottom.
- Metrics will be printed and saved to results/*.csv. If you don't have results/ folder please create one.
- To change the dataset name-> search for project keyword in the .ipynb files and replace project = 'tensorflow' with project = 'pytorch'

4. Repeatability:

- Each model runs across 10 random seeds.
- Output metrics are averaged for stability.
- Use logged CSVs to verify and compare model performance.

5. Reproduce Figures:

- Confusion matrices and evaluation summaries are plotted in final cells.
- You can export plots as PNGs from notebook outputs.