Replication Guide

This document provides step-by-step instructions to replicate the experiments presented in the **Bug Report Classification** project.

Setup

```
# Install dependencies
pip install pandas numpy matplotlib seaborn scikit-learn tensorflow nltk tqdm

# Create required directories
mkdir -p datasets model_results
```

Run Analysis

From the project root, run the following:

```
python bugreport.py
```

This will process all projects listed in the `PROJECTS` variable inside `bugreport.py`.

Verify Results

- Check output files in `model_results/{project}/`
- Key metrics in `{project}_final_comparison.csv`

Each project folder contains:

- `*_final_comparison.csv` — performance metrics (accuracy, precision, recall, F1, AUC)
- `*_confusion_matrix.png` — visual confusion matrix
- `*_history_*.png` — training curves for deep learning models
- (Optional) `*.h5` — saved model weights

You can open the `.csv` and `.png` files to verify the model performance.

Notes

- All datasets are included in the `datasets/` folder.
- If you'd like to analyze a different subset of projects, modify the `PROJECTS` list in `bugreport.py`.
- All experimental result files, including performance summaries, training history plots, and confusion matrices, have been uploaded under the directory `model_results/` in the GitHub repository for review.