Replication

2.1 Prerequisites

Software:

Python 3.8+, Git, Jupyter Notebook.

Libraries: scikit-learn, nltk, gensim, imbalanced-learn.

2.2 Dataset Setup

Source:

The dataset comprises 2,500 GitHub issues from 5 repositories (PyTorch, Keras, etc.).

Download from GitHub REST API or use the preprocessed data in data/raw/.

Preprocessing:

```
# Example preprocessing steps (from your code)
df["text"] = df["title"] + " " + df["body"]
df["text"] = df["text"].str.lower().apply(remove_stopwords)
```

2.3 Code Execution

Train the Model:

```
python train.py --data data/processed/train.csv --model outputs/model.pkl
```

Evaluate Performance:

```
python evaluate.py --data data/processed/test.csv --model outputs/model.pkl
```

2.4 Reproducing Experiments

Hyperparameters:

Random Forest: n_estimators=200, max_depth=20, class_weight="balanced".

Word2Vec: vector_size=100, window=5.

Statistical Tests:

Run statistical_tests.ipynb to regenerate Wilcoxon test results and effect sizes.

2.5 Expected Results

Metrics:

F1-score: ~0.13 (PyTorch) to ~0.11 (Caffe).

Precision/Recall tables matching those in the report.

2.6 Notes on Variability

Stratified Splits: Use random_state=42 in train_test_split for reproducibility.

SMOTE: Ensure imbalanced-learn==0.9.0 to avoid sampling inconsistencies.