To replicate the results reported in the coursework:

1. Make sure the following files are in the same root directory:

- “radomforest.ipynb”

- “base\_model.ipynb”

- A folder named `dataset/` containing all required bug report datasets

2. Install the required packages (see requirements.pdf)

3. Run the Baseline Model:

- Launch Jupyter and open “base\_model.ipynb”

- “# Choose the project (options: 'pytorch', 'tensorflow', 'keras', 'incubator-mxnet', 'caffe')”

Modify the line below this comment to select the dataset you want to run. For example, to run the result for 'pytorch', set projects = ['pytorch']

- Execute all cells in order

- The results will be saved in files with the suffix “\_BaseModel\_NB.csv”.

4. Run the Random Forest:

- Launch Jupyter and open “randomforest.ipynb”

- “# Choose the project (options: 'pytorch', 'tensorflow', 'keras', 'incubator-mxnet', 'caffe')”

Modify the line below this comment to select the dataset you want to run. For example, to run the result for 'pytorch', set projects = ['pytorch']

- Execute all cells in order

- The results will be saved in files with the suffix “\_randomforest\_NB.csv”.

5. Evaluation

- The notebook evaluates five individual projects and a merged dataset

- It computes and prints Accuracy, Precision, Recall, and F1-score for both Naive Bayes and Random Forest models

- Results for each project will also be saved as CSV files, automatically named with “\_NB.csv” suffix

Note:

- Ensure that the dataset files and expected filenames match the notebook configuration.