

TOOL MANUAL: BUG REPORT CLASSIFICATION

Author: Javier Rodríguez Fernández (jxr497@student.bham.ac.uk)

Professor: Tao Chen

School of Computer Science, University of Birmingham, B15 2TT, UK

1 TOOL OVERVIEW

This tool automates bug report classification using an SVM model with TF-IDF preprocessing. It includes:

- **Solution_Javier.py:** Trains the SVM model and outputs metrics.
- **StatisticalAnalysis_Javier.py:** Generates statistical comparisons and plots.

2 PREREQUISITES

Before using the tool:

- 1) Install dependencies listed in **requirements.pdf**.
- 2) Download datasets (tensorflow.csv, pytorch.csv, etc.) and place them in the **project root folder**.

3 USING THE SOLUTION CODE

STEP 1: Configure the script

- 1) Set the project variable to your dataset name (e.g. project = 'tensorflow')
- 2) Ensure REPEAT = 30 (do not modify this value)

STEP 2: Run the Code

- 1) Open a terminal (Command Prompt or PowerShell)
- 2) Navigate to the project directory: "cd C:\path\to\project_folder"
- 3) Execute the script: "python Solution_Javier.py"

STEP 3: Output

- A CSV file (e.g., tensorflow_SVM.csv) will be generated in the project folder.
- This file contains metrics like **F1-score**, **Accuracy**, and **AUC** averaged over 30 runs.

4 STATISTICAL ANALYSIS

STEP 1: Configure the analysis script

- 1) Open "StatisticalAnalysis_Javier.py"
- 2) Set the project variable to match your dataset (e.g. project = 'tensorflow')

STEP 2: Run the Analysis

- 1) Ensure both baseline and solution results are generated.
- 2) In the terminal, run: “python StatisticalAnalysis_Javier.py”

STEP 3: Output

These files will be generated:

- 1) **Pareto Front Plot** (pareto_front_F1_vs_AUC_*.png): Shows SVM’s dominance in F1 vs. Accuracy.
- 2) **Mean Metric Plot** (mean_points_plot_*.png): Compares SVM and Naive Bayes metrics.
- 3) **Wilcoxon Test Results** (p_values_plot_*.png): Displays statistical significance ($p < 0.05$).

5 TROUBLESHOOTING

ISSUE	SOLUTION
CSV files not found	Ensure datasets are in the root folder and filenames match (e.g., tensorflow.csv).
Dependency errors	Verify installations with pip list and cross-check requirements.pdf.
Plots not generating	Ensure matplotlib is installed and the project variable is set correctly.
Statistical test errors	Confirm both {project}_SVM.csv and {project}_NB.csv exist.

6 SUPPORT

For issues, contact: jxr497@student.bham.ac.uk