Intelligent Bug Report Classification: A Hybrid Approach Combining Tf-IDF, GloVe, and Meta Features

---

Requirements

I used PyCharm 2024.3.3 (Professional Edition - Build #PY-243.24978.54, built on February 12, 2025) to complete this assignment.

Python Version:

- Python: **3.13.2** (tags/v3.13.2:4f8bb39, Feb 4 2025, 15:23:48) [MSC v.1942 64 bit (AMD64)]
- Older versions (e.g., 3.8–3.9) may work, but are not officially tested.

Main Required Packages:

pandas==2.2.3

numpy==2.2.4

scikit-learn==1.6.1

nltk==3.9.1

scipy==1.15.2

tkinter – for GUI (preinstalled on Windows/macOS; install separately on Linux)

All installed packages and versions:

click==8.1.8
colorama==0.4.6
joblib==1.4.2
nltk==3.9.1
numpy==2.2.4
pandas==2.2.3
python-dateutil==2.9.0.post0
pytz==2025.2
regex==2024.11.6
scikit-learn==1.6.1
scipy==1.15.2
six==1.17.0
threadpoolctl==3.6.0
tqdm==4.67.1
tzdata==2025.2

Installation

Can use pip to download all the packages directly. I have a file called “packages.txt” with all the packages and versions. To download this way, do:

Pip install – r packages.txt

Other Required Files

glove.6B.100d.txt – GloVe word embeddings (100-dimensional).

Download this file and place it in the same directory as *improved_br_classification.py*

Can be downloaded from: https://nlp.stanford.edu/projects/glove/

Click download “latest code”

Scroll down till you see “glove.6B.zip” and click to download

Once downloaded, unzip the file.

You should see multiple options, but the one we want is **glove.6B.100d.txt**

Copy that file and place it in the same directory as *improved_br_classification.py*

If not already, the datasets all must be in the same directory as *improved_br_classification.py* and *br_classification.py* separetly (as *br_classification.py* is inside the lab1 folder).

My System

Windows 11

RAM: 16GB

Processor: Intel i9