User Manual

This manual provides step by step instructions on how to use the tool for analysing software project issues using BERT based embeddings and Logistic Regression. The tool processes textual data, extracts embeddings, trains a classifier and evaluates its performance.

Prerequisites

Before running the tool, ensure that the following dependencies are installed. Refer to requirements.pdf for details:

- Python 3.8+
- Transformers
- Pandas
- NumPy
- Scikit-learn
- NLTK
- PyTorch

Ensure that the datasets are stored in (.../datasets/).

Running the Tool

Firstly, open a terminal and clone the repository using:

- git clone <repository_link>

Install dependencies using pip:

- pip install pandas numpy torch transformers scikit-learn nltk

Run the script:

- python model.py

This will process the data, train the model and display classification results.

Configuring the Tool

Modify the project variable in the script to specify the dataset:

- where project = ‘caffe’ change to (pytorch, tensorflow, etc.)

Adjust the hyperparameters in the logistic regression section:

- param_grid = { ‘C’: [0.01,0.1,1,10,100]} #Regularisation strength

Output

The tool will print the results in a console and store them in a CSV file saved as ‘.../project_BERT_LogReg_results.csv’.