Intelligent Software Engineering - User Manual

1. Introduction

This document provides guidelines to users for using our project based on the baseline from the Lab 1 and trying to improve it by SVM model. It includes steps for setting up the environment, running experiments, and interpreting the results.

2. System Requirements

To use this system, ensure you have the following installed:

- Python 3.8 or higher
- Required Python libraries (see requirements.pdf for details)
- Compatible operating systems: Windows, Linux, macOS

3. Installation

Follow these steps to install the system:

1. Clone the repository: “git clone https://github.com/mariodamas/ise_coursework.git”
2. Navigate to the project directory: “cd ise_coursework”
3. Install FastText pre-trained embeddings file in this link: https://dl.fbaipublicfiles.com/fasttext/vectors-crawl/cc.en.300.vec.gz
4. Make sure every dependency in requirements.pdf is installed: “pip install …”

4. Usage Instructions

4.1 Running the System

To start the system, run:

1. Run the baseline model (for every project):
   python baseline.py

2. Run the SVM model with TF-IDF (for every project):
   python svm_tf_idf.py

3. Execute SVM classification with word embeddings (for every project):
   python svm_word_embeddings.py

4. Run results_mean to obtain the results table that appears in the document
   python results_mean.py

5. Finally, check whether the baseline is beaten with the statistical test.
   python results_mean.py

This will launch the tool, allowing users to get data in csv for analysis.

4.3 Viewing Results

After processing, the system presents:

- Files ending by "_NB" will represent every baseline result.
- Files ending by "_SVM_TF" will represent every SVM + TF-IDF result.
- Files ending by "_SVM_WE" will represent every SVM + FastText result.
- Results for the statistical test will be displayed in the console log.

5. Troubleshooting & FAQs

Q1: The system does not start.

- Ensure Python and dependencies are correctly installed.
- Check error logs for missing libraries.

Q2: Results seem incorrect.

- Verify dataset consistency.
- Ensure the correct preprocessing steps have been followed.

6. Contact & Support

For further assistance, please refer to the project repository or contact: mxd489@student.bham.ac.uk