Intelligent Software Engineering - Replication Manual

1. Introduction

This document provides guidelines for replicating the results obtained using 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 replicate the results, ensure your system meets the following requirements:

- **Python Version:** 3.8 or higher
- **Required Libraries:** (See requirements.pdf for installation)
    - numpy
    - pandas
    - scikit-learn
    - nltk
    - matplotlib
    - scipy
- **Operating Systems:** Compatible with Windows, Linux, macOS
- **Hardware Requirements:**
    - At least 8GB RAM
    - Sufficient disk space for datasets and FastText pre-trained embeddings

3. Installation

Follow these steps to set up the environment:

1. **Clone the repository:**
   git clone https://github.com/mariodamas/ise_coursework.git
2. **Navigate to the project directory:**
   cd ise_coursework
3. **Install dependencies (please see requirements.pdf):**
   pip install …
4. **Download FastText pre-trained embeddings:**
    - Link: FastText Embeddings
    - Extract and place it in the project directory.

4. Running the Experiments

To replicate the results, execute the following scripts in order:

4.1 Baseline Model

python baseline.py

- Generates baseline classification results.

4.2 SVM Model with TF-IDF

python svm_tf_idf.py

- Runs an SVM classifier using TF-IDF features.

4.3 SVM Model with Word Embeddings

python svm_word_embeddings.py

- Runs an SVM classifier using FastText word embeddings.

4.4 Aggregating Results

python results_mean.py

- Computes the mean results across multiple runs.

4.5 Statistical Testing

python results_mean.py

- Evaluates statistical significance of model improvements over the baseline.

5. Expected Outputs

After running the experiments, results will be available as CSV files:

- **Baseline results:** Files ending in _NB
- **SVM + TF-IDF results:** Files ending in _SVM_TF
- **SVM + FastText results:** Files ending in _SVM_WE
- **Statistical test results:** Displayed in the console log

6. Troubleshooting

Issue 1: Python script fails to execute

- Ensure dependencies are correctly installed (requirements.pdf)
- Check for missing files (datasets, embeddings, etc.)

Issue 2: Results do not match the original ones

- Ensure you are using the same dataset versions.
- Confirm preprocessing steps are followed correctly.
- Variability in random initialization may cause slight differences.

7. Contact & Support

For further assistance, visit the project repository or contact: **mxd489@student.bham.ac.uk**