Requirements

This document outlines the requirements and dependencies needed to run the Enhanced Bug Report Classifier.

Python Version

The code has been tested with Python 3.8 and 3.9. It may work with other versions, but these are recommended.

```
python>=3.8,<3.10
```

Required Packages

The following Python packages are required to run the classifier:

```
numpy>=1.20.0
pandas>=1.3.0
scikit-learn>=1.0.0
matplotlib>=3.5.0
seaborn>=0.11.0
scipy>=1.7.0
tqdm>=4.62.0
```

Package Installation

You can install all required packages using pip:

```
pip install numpy pandas scikit-learn matplotlib seaborn scipy tqdm
```

Or create a requirements.txt file with the following content:

```
numpy>=1.20.0
pandas>=1.3.0
scikit-learn>=1.0.0
matplotlib>=3.5.0
seaborn>=0.11.0
scipy>=1.7.0
tqdm>=4.62.0
```

And install using:

```
pip install -r requirements.txt
```

System Requirements

Recommended Hardware

- **CPU**: Multi-core processor (4+ cores recommended for faster training)
- **RAM**: 8GB minimum, 16GB recommended for larger datasets

- **Disk Space**: 1GB free space for datasets and results

Operating System

The code has been tested on: - macOS 12.0+ - Ubuntu 20.04 LTS - Windows 10/11

Optional Dependencies

The following packages are not strictly required but can enhance performance:

```
joblib>=1.1.0  # For parallel processing
numba>=0.54.0  # For improved computational performance
```

Compatibility Notes

Pandas Version Issues

If you encounter warnings related to numexpr or bottleneck versions with pandas, you can safely ignore them or install the recommended versions:

```
pip install numexpr>=2.8.4 bottleneck>=1.3.6
```

MacOS Notes

On macOS, you might need to install the XCode Command Line Tools for some dependencies:

```
xcode-select --install
```

Linux Notes

On Linux systems, ensure you have the required development libraries:

```
# Ubuntu/Debian
sudo apt-get install python3-dev build-essential

# CentOS/RHEL
sudo yum install python3-devel
```

Dataset Requirements

The code expects datasets in CSV format with specific columns: - `Title`: String column with bug report titles - `Body`: String column with bug report descriptions - `Comments`: String column with additional comments - `class`: Integer column (0 or 1) indicating if the bug is performance-related

Each dataset should correspond to a specific deep learning framework and be named accordingly.