Project Overview

This project aims to conduct sentiment analysis on multiple open - source projects (such as PyTorch, TensorFlow, Keras, etc.). Baseline models and improved models are used for experiments, and the experimental results are recorded.

Project Structure

```
ISE - coursework/
├── code/
│  ├── baseline/                    # Baseline model code (original_NB code)
│  ├── improved/                    # Improved model code (improved code for each project)
│  └── utils/                       # General utility scripts (e.g., data merging script)
├── results/                        # Result files
│  ├── caffe_NB.csv                 # Baseline results
│  ├── caffe_improved.csv           # Improved results
│  ├── incubator - mxnet_NB.csv
│  ├── incubator - mxnet_improved.csv
│  └── ... (similarly for other projects)
├── requirements.pdf                # Dependency description (Python libraries and versions)
├── manual.pdf                      # User manual
├── replication.pdf                 # Reproduction guide
└── README.md                       # Overall repository description
```

Usage Steps

Environment Preparation：

·Check the requirements.pdf file and install the required Python libraries and their corresponding versions.

·Ensure that git is installed in your environment for cloning the code repository.

Clone the Code Repository：

Run the following command in the terminal to clone the code repository:

```
git clone <repository address>
cd ISE - coursework
```

Run the Code

Baseline Model：

·Navigate to the code/baseline directory.

·Select the appropriate project code file (e.g., pytorch_nb_.py) as needed and run the following command:

```
python pytorch_nb_.py
```

Improved Model

·Navigate to the code/improved directory.

·Select the appropriate project code file (e.g., pytorch_improved_.py) as needed and run the following command:

```
python pytorch_improved_.py
```

View Results

After running the code, the results will be saved in the corresponding CSV files in the results directory. You can use Excel or other data analysis tools to open these files and view the experimental results.

Code Explanation

Data Processing

·remove_html function: Removes HTML tags from the text using regular expressions.

```
def remove_html(text):
    html = re.compile(r'<.*?>')
    return html.sub(r'', text)
```

·remove_emoji function: Removes emojis from the text using a regular expression pattern.

```
def remove_emoji(text):
  emoji_pattern = re.compile("["
                             u"\U0001F600-\U0001F64F" # emoticons
                             u"\U0001F300-\U0001F5FF" # symbols & pictographs
                             u"\U0001F680-\U0001F6FF" # transport & map symbols
                             u"\U0001F1E0-\U0001F1FF" # flags
                             u"\U00002702-\U000027B0"
                             u"\U000024C2-\U0001F251" # enclosed characters
                             "]+", flags=re.UNICODE)
return emoji_pattern.sub(r'', text)
```

·remove_stopwords function: Removes stopwords and technical terms from the text.remove_stopwords function: Removes stopwords and technical terms from the text.

```
def remove_stopwords(text):
    return " ".join([word for word in str(text).split() if word not in final_stop_words_list])
```

·clean_str function: Cleans the text by removing non - alphanumeric characters, handling special cases, and converting it to lowercas

```
def clean_str(string):
    string = re.sub(r"[^A-Za-z0-9(),.!?\'\`]", " ", string)
    string = re.sub(r"\r\n|\n|\r", " ", string) # Handle newlines
    string = re.sub(r"\s{2,}", " ", string) # Collapse multiple spaces
    string = re.sub(r"\\", "", string) # Remove backslashes
    string = re.sub(r"\'", "", string) # Remove apostrophes
    string = re.sub(r"\"", "", string) # Remove quotes
    return string.strip().lower()
```

Model Training

·Uses GaussianNB from the sklearn library as the classifier.

·Uses GridSearchCV for hyperparameter tuning.

Evaluation Metrics

Records evaluation metrics such as accuracy, precision, recall, and F1 - score.

Notes

·Before running the code, ensure that the data file paths are correct.

·The experimental results may be affected by the random seed and data distribution.