

Requirements

Repository Access

The code can be cloned from the public GitHub repository using either HTTPS or SSH:

HTTPS

```
git clone https://github.com/cg212/CG-ISE.git
```

SSH

```
git clone git@github.com:cg212/CG-ISE.git
```

Environment Setup

To run the code directly in your IDE:

Direct Installation

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

Core Dependencies

The project requires the following key libraries:

```
# Core dependencies
pandas>=1.3.0
numpy>=1.20.0
scikit-learn>=1.0.0
scipy>=1.7.0
joblib>=1.1.0

# Text processing
nltk>=3.6.0
gensim>=4.0.0
spacy>=3.2.0
regex>=2021.8.3

# Class imbalance handling
imbalanced-learn>=0.8.0

# NLP Models
en-core-web-sm @ https://github.com/explosion/spacy-
models/releases/download/en_core_web_sm-3.4.0/en_core_web_sm-3.4.0-py3-none-any.whl

# Visualisation
matplotlib>=3.4.0
seaborn>=0.11.0

# Utilities
tqdm>=4.62.3

# Statistical testing
statsmodels>=0.13.0
```

NLTK Resources

Before running the models, you need to download required NLTK resources:

```
python download_nltk_resources.py
```

This will download essential NLTK resources such as: - punkt (Tokenizer) - stopwords (Stopwords corpus) - wordnet (WordNet lexical database) - omw-1.4 (Open Multilingual WordNet)