

Requirements

This guide gives instructions on how to clone the repository and pre-requisites for running the project.

The version of Python that was used in the development of this code was version 3.12.8.

To clone the repository on your local machine, run the following commands:

```
git lfs install
```

```
git clone https://github.com/Syperhanic/ISEProject.git
```

Note: 'git lfs install' installs Git LFS (Large File Storage), which is required before cloning as the GloVe embeddings text file (glove.6B.100d.txt) in the repository exceeds Git's maximum file size of 100 megabytes. If this causes issues, pull the repository without the file, and then download it manually using the following link: <https://nlp.stanford.edu/projects/glove/>. On the site, locate the 'Download pre-trained word vectors' section on the main page and find the install link for 'glove.6B.zip'. This is shown in the below image.

Download pre-trained word vectors

- Pre-trained word vectors. This data is made available under the [Public Domain Dedication and License](http://www.opendatacommons.org/licenses/pddl/1.0/) v1.0 whose full text can be found at: <http://www.opendatacommons.org/licenses/pddl/1.0/>.
 - [Wikipedia 2014](#) + [Gigaword 5](#) (6B tokens, 400K vocab, uncased, 50d, 100d, 200d, & 300d vectors, 822 MB download): [glove.6B.zip](#)

Once installed, place the glove.6B.100d.txt file in the root directory of the project.

The following pip install commands must be ran to install the required dependencies:

```
pip install pandas
```

```
pip install torch torchvision torchaudio torchtext --index-url  
https://download.pytorch.org/whl/cpu
```

```
pip install nltk
```

```
pip install scikit-learn
```

```
pip install optuna
```

For reference, after the above commands are ran, the following dependencies will be installed in the project:

```
certifi==2022.12.7 charset-normalizer==2.1.1
```

click==8.1.8 colorama==0.4.6
filelock==3.13.1
fsspec==2024.6.1 idna==3.4
Jinja2==3.1.4 joblib==1.4.2
MarkupSafe==2.1.5
mpmath==1.3.0
networkx==3.3 nltk==3.9.1
numpy==2.2.4
pandas==2.2.3
pillow==11.0.0 python-
dateutil==2.9.0.post0
pytz==2025.1
regex==2024.11.6
requests==2.28.1 scikit-
learn==1.6.1 scipy==1.15.2
setuptools==70.2.0
six==1.17.0 sympy==1.13.1
threadpoolctl==3.6.0
torch==2.6.0+cpu
torchaudio==2.6.0+cpu
torchtext==0.18.0+cpu
torchvision==0.21.0+cpu
tqdm==4.66.5
typing_extensions==4.12.2
tzdata==2025.2
urllib3==1.26.13