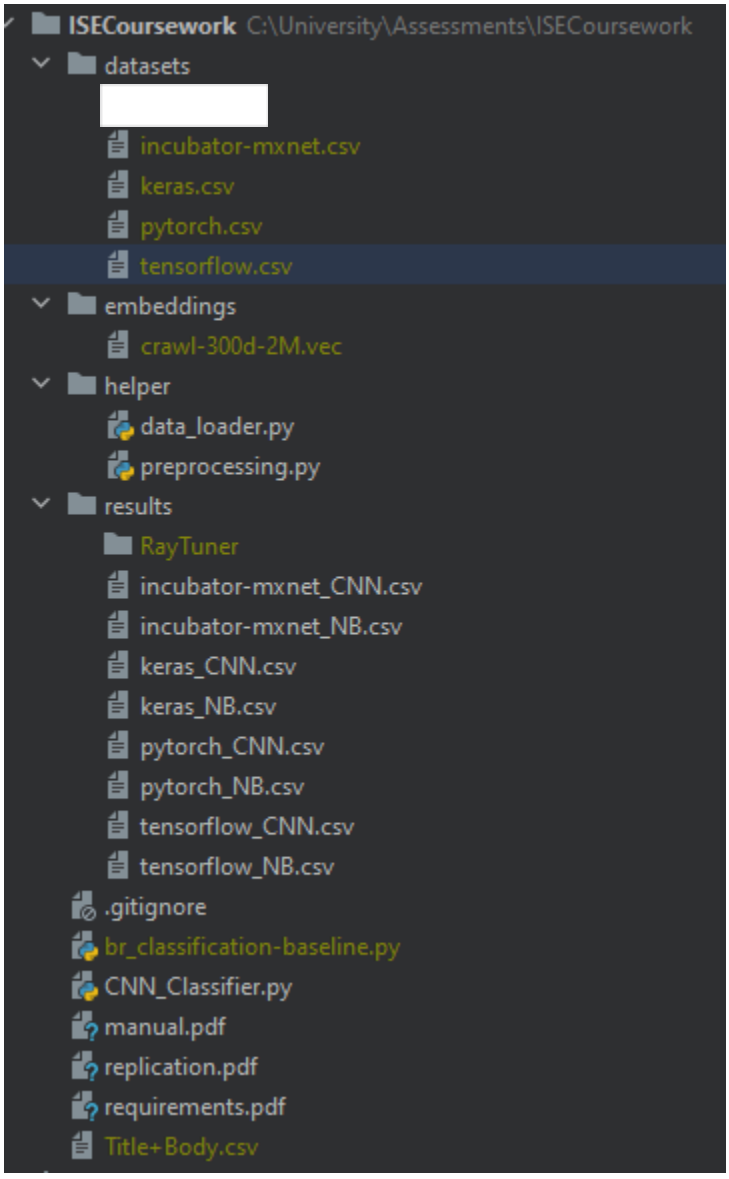
Requirements

Download the crawl-300d-2M.vec and put it in a new directory in the root directory named “embeddings”, the zip file can be found here https://fasttext.cc/docs/en/english-vectors.html

Download the datasets from lab 1 https://github.com/ideas-labo/ISE/tree/main/lab1/datasets and put them into a folder in the root directory named “datasets”

This is how the folder structure should look:



Using a python venv, preferably on pycharm, install the following packages using either conda or pip - I used python 3.12.6

```
import os
import nltk
```

```
import torch
import numpy 1.x
import pandas
import matplotlib <3.10.0
import time
import ray
import sklearn
import optuna
import copy
import re
import tqdm
```

Import the cuda version of pytorch if you have the available hardware

Pycharm

Setup your venv with your pycharm project and run “current file” while clicked on CNN_Classifier.py

Terminal

Go to the root directory and enter `python ./CNN_Revised` (note you will need to have your venv setup on the terminal or run the venv’s `python.exe` file instead of just `python`)

Here are the exact requirements and versions using pipreqs:

```
matplotlib==3.9.0 note 3.10.0 doesnt work for pycharm
nltk==3.9.1
numpy==2.2.3
optuna==4.2.1
pandas==2.2.3
ray==2.42.1
scikit_learn==1.6.1
torch==2.4.1
tqdm==4.66.5

Note: you may have to downgrade to numpy 1.26.4 as some of the module
(likely nltk) may not be compatible with numpy 2.x
```