I collected the tweets using snscrape. snscrape is a scraper for social networking services (SNS). It scrapes things like user profiles, hashtags, or searches and returns the discovered items, e.g., the relevant posts. And I also proceeded to preprocess the tweets by removing the links, hashtags, and non-ascii characters.

I continued to convert the tweets as I progressed with their preprocessing to lower-case, remove square bracket, removes number and punctuation. Next for the ANN models I Created a word to index dictionary using tokenizer. Tokenization breaks the raw text into words, sentences called tokens. These tokens help in understanding the context or developing the model for the NLP. The tokenization helps in interpreting the meaning of the text by analyzing the sequence of the words.

We chose Glove to aid with the development of the embedding layer on our ANNs for natural language processing. GloVe is an acronym that stands for global vectors for word representation. It is a Stanford-developed unsupervised learning approach for producing word embeddings by aggregating a corpus's global word-word co-occurrence matrix. The obtained embeddings reveal fascinating vector space linear substructures of the word.

I built the LSTM model with sequential for simplicity. Modeled the embedding layer, a LSTM with two dense layers. I then compiled, trained, tested and evaluated the models by creating plots of their training and validation histories.

We compared our model’s prediction score and we helped each other write the report abstract, introduction, methods, results, Dataset, and the conclusion. I worked on the social, ethical, legal and professional considerations in text mining.