Natural Language Processing

Objective 1:

Given a tweets file that contains tweets along with their author, get the most frequent entities from the tweets

Approach:

Language: Python
Library: csv, Pandas
Development Environment: Jupyter

- Install the necessary libraries and import them
- Load the csv file (dataset) using the read method
- A variable is initialised to iterate through the loop followed by creation of a list to store all the entities
- Create a loop to append all the values to the list
- Using the count and set method count the occurrence of each entity from the list. The set method prevents the repetition of the entity.
- Created a Data Frame to display the output in the desired arrangement.

Objective 2:

To find out the sentiment/polarity of each author towards each of the entities.

Approach:

A pre-trained model/pipeline zero-shot classification was used for the sentiment analysis.

Zero-shot classification:

Zero-shot classification allows you to define your own labels and the classifier assigns a probability to each label. The scores will be independent and they fall between 0 and 1.

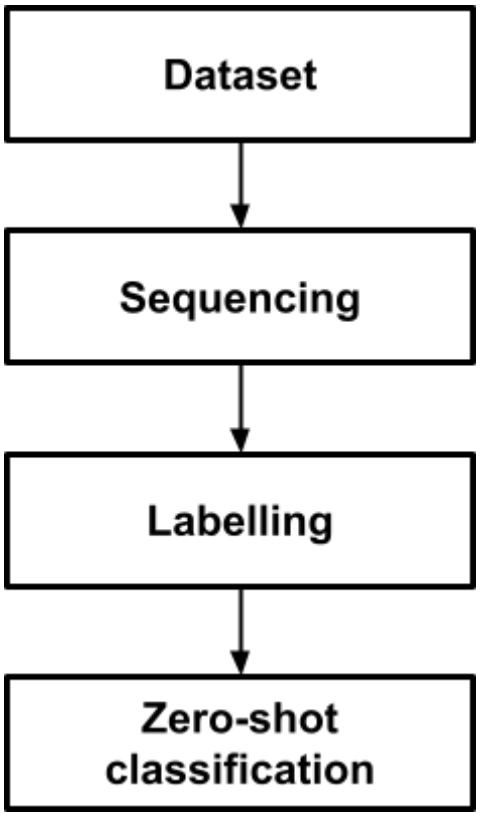

Fig 1: Proposed System

Language: Python
Libraries: Pandas, Transformers
Development Environment: Jupyter

Implementation:

- All the necessary libraries were imported
- Initialised a variable to define the zero-shot classification pipeline.
- Loaded the csv file (the data - tweets.csv)
- A string was created to get a sequence of the data
- The labels positive and negative were defined
- The classifier was called by passing the sequence and the sentiment labels as parameters
- A loop was created to iterate through the sequence and convert the result into a Data Frame followed by mapping the maximum score.
- Exported the output to csv file.

Future Enhancement:

Clean and pre-process the data and create a model using different methods like Random Forest, Naive Bayes, etc., followed by training the model and classification on the data based on positive and negative.

Reference:

Youtube - https://youtu.be/ljI7GoY5JgY