

Phase 2: Sentiment Analysis

In this phase, you are supposed to perform sentiment analysis on a dataset of labeled tweets. You are given a dataset of labeled tweets, on which you should:

- Clean the data (removing duplicate/useless data, handling upper/lower case, etc.)
- Use techniques taught in the course (Stemming, Lemmatization, etc.)
- Perform Feature extraction (BoW, TF-IDF)
- Train/test a supervised method (e.g. logistic regression) on each of the above features
- Use a pre-trained library for sentiment analysis
- Compare the three methods and discuss the results

Some useful libraries you could use: TextBlob, NLTK