TextBlob.

Intuitive interface to NLTK.

Some Applications of NLP.
Machine Translation
Chatbots

TextBlob is easy to learn and showcases characteristics like sentiment analysis for performing NLP tasks. It employs NLTK to accomplish its tasks. It supports complex evaluation and strategies on textual data.

Token - word in a text document

Tokenization - Reducing paragraph into words with TextBlob.

Lemmatization - converting the token(word) to its root form contained in pre-defined dictionary.

POS tagging(Parts of Speech) - Labelling a token(word) to a particular POS on the account of its context.

Sentiment.

Intensity of each word(token) in the text.

Sentiment Analysis.

It aids us to interpret emotions of public and accumulate information. It is a procedure of examining data and categorizing it according to the requirement of the research. By examining these emotions, it is easy to understand what people like, want and their major concerns.

Polarity[-1 to 1] - Negative sentiment(-1) and Positive sentiment(1).

Subjectivity[0 to 1] - Extent of opinion and verifiable information present in the text.

Polarity and subjectivity 'zero' or neutral - Data irrelevant to interest of research as it does not contain distinguishing text, links and/or hashtags. These do not contribute to our analysis at all.

Tweepy - Allows to connect with the *Twitter API* and allows streaming of Twitter data in real time.

Lexicon-based Approach.

Recognizing moods & emotions in text where each word is assigned a score.

WordStat Sentiment Dictionary - Pre-defined dictionary which includes positive & negative word patterns.

Analysis.

I. Analyse sentiment polarity on Donald Trump's dataset.
II. Analyse sentiment polarity on Joe Biden's dataset.
III. Examine positive(1), negative(-1) & neutral(zero) replies on Donald Trump's Tweets.

IV. Examine positive(1), negative(-1) & neutral(zero) replies on Joe Biden's Tweets.
V. Discard rows in Donald Trump's dataset whose reply is neutral(zero).
VI. Discard rows in Joe Biden's dataset whose reply is neutral(zero).
VII. Perform data visualization of the balanced datasets.

spaCy - Is yet another library used to perform NLP tasks. But TextBlob is employed in our project due to its simple interface.

Speed comparison - spaCy > TextBlob > NLTK