Sentiment analysis using NLTK Twitter dataset (twitter\_samples) with a logistic regression model and an accuracy of %95 on the test set.

**Preprocessing data for sentiment analysis:**

1. Tokenizing the string
2. Lowercasing
3. Removing **stop words** and punctuations (stop words/punctuation lists can be modified)
4. Deleting Handles and URLs
5. **Stemming**: transforming any world to its stem: tuning to tun: tune, tuned, tuning

**Building word frequencies**:

* Building a dictionary where we can look up how many times a word appears in the list of positive or negative tweets

**Feature extraction:**

* Creating a (m by 3) matrix
* The first feature is the bias
* The second feature is the number of positive words in a tweet.
* The third feature is the number of negative words in a tweet.

**Training the model, Prediction, and Accuracy:**

* Training a logistic regression model from scratch

**Testing with your own tweet**

Tweet: This weeks tribute is up on insta! ❤️👑

Prediction: Positive sentiment