Bag of words:

The bag-of-words model is a simplifying representation used in natural language processing and information retrieval (IR). In this model, a text (such as a sentence or a document) is represented as the bag (multiset) of its words, disregarding grammar and even word order but keeping multiplicity.

Step 1:

Go through all the data, and create a dictionary which contains all the words appeals in the twitters, and save the dictionary as W.

Step 2:

Go through the data again, for each tweet, count the words frequency in the order of dictionary W. Therefore, after go through all tweets, we would get a M\*N matrix which M is corresponding each tweets and the class label, while N records the word frequency in each tweets.

Step 3:

Use this matrix as our training data.