Assumptions :

It is assumed that you have following packages installed in your system : nltk, wordsegment, sklearn, preprocessor, Spacy, re Download the Glove twitter model from https://nlp.stanford.edu/projects/glove/
It is zip file you will need to extract it manually code to extract the zip file is not written in the script.

Key decisions :
This is a classification problem in NLP. To start with the problem first we need to convert the text into numbers for this purpose I used glove embeddings. Before creating the embeddings there are alot of mis-spelled words, hashtags, punctuations, tags, links, emoticons, stopwords are present in the tweets. Thus, I cleaned the data, for the purpose of cleaning the data I used regex (re) and nltk module. with regex expressions I eliminated the URls, punctuations and converted word written as 'luv' into love.

To gain the information from the hashtags I used a wordsegment module which segments the words which are not otherwise segmented. Eliminated stopwords using nltk stop_words, and lammetized them. After the cleaning, convert the words into their corresponding vectors using CountVectorizer following the Tfidf transformer. This model has been made on a very large amount of data for most of the words present in tweets there is a vector. There are 3 dimensionality vector models namely : 50, 100 , 200. Here I have used a 200d vector model in which for every word there is a 200 length vector, I added all the vectors corresponding to each word present in a tweet.Thus created a vector of length 200 for each tweet. After the conversion of text into numbers the task is to classify these vectors into positive and negative tweets. I plotted the vectors into 2D space and they were linearly separable so I used SVM to classify these vectors. Experiments: I tried converting the text into numbers using tf-idf vectorizer as well as glove and classification using different classification models such as SVM,logisitc regression,multinomial NB,randomforest classifier