1. **Subject : -**

Sentiment analysis using TF-IDF.

* **Dataset Description : -**

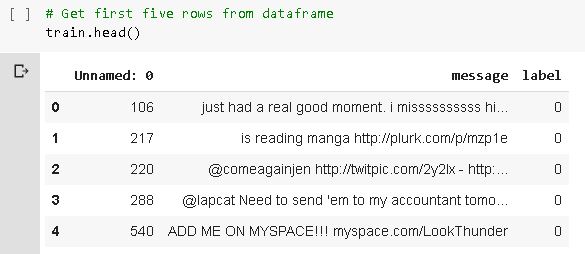
1. Dataset link : -

* https://drive.google.com/file/d/1q2lyvOvvRWJWAlc4h8dA5Q4FX3z9snCj/view?usp=sharing

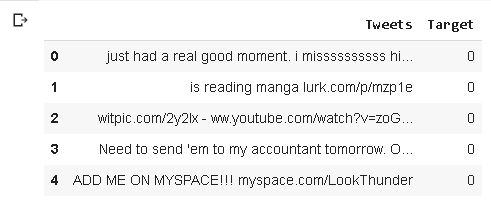
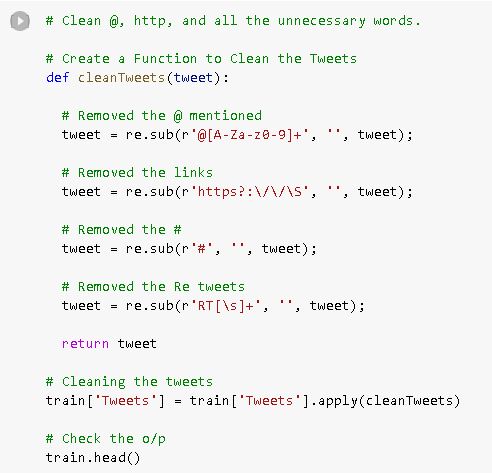
1. Shape of the dataset : -

* Shape of the DEPRESSION dataset is :- (10314, 3).

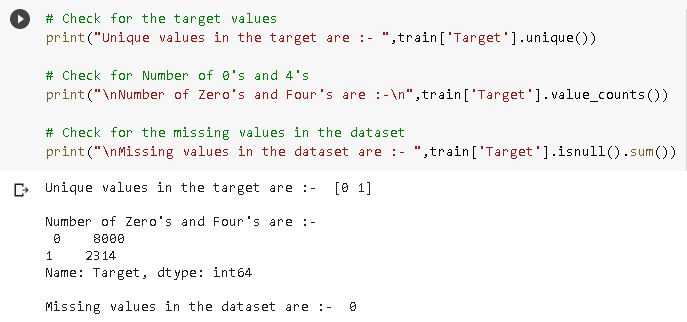
1. Dataset (first five rows) : -



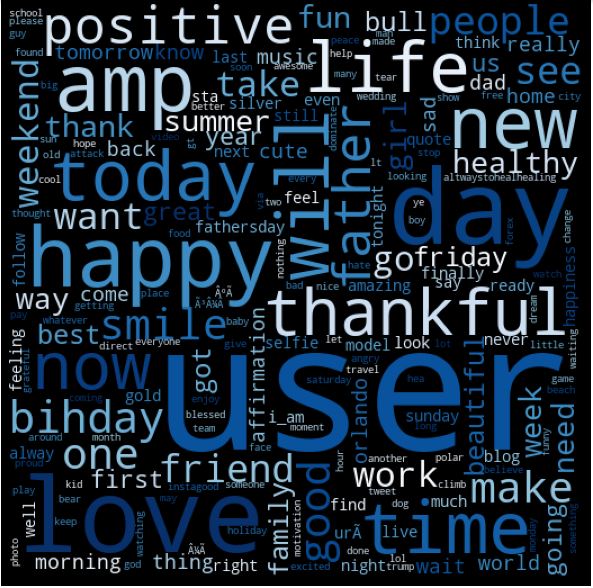
1. Data-Preprocessing : -



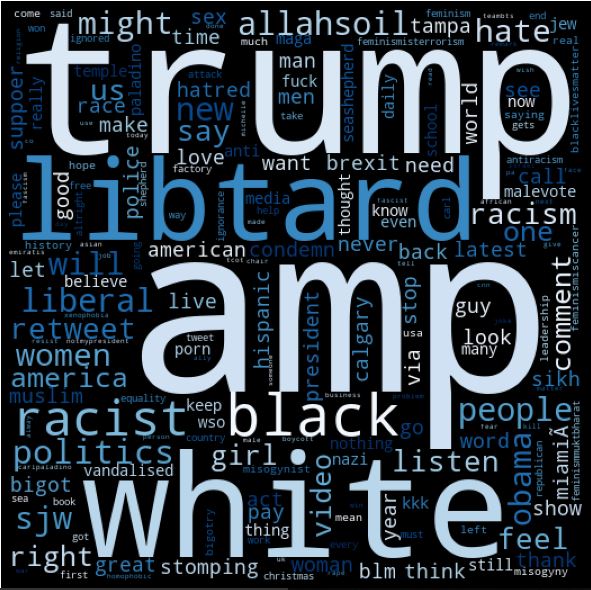
1. Dataset Description : -



1. WordCloud Analysis of Positive words : -



1. WordCloud Analysis of Negative Words : -



* **Description : -**

1. For making a model, we need to count each word, for that we decided to use a bag of words, after some research on this we get to know that Bag of Words will give less accuracy as compared to TF-IDF model.
2. Hence we shift to the TF-IDF model.
3. Now, while creating the IF-IDF matrix of words, first we divided the words into two parts. One is “Positive words” another is “Negative Words”.
4. After that we just use the formula of TF-IDF.

# Term frequency =

No. of repetition of words in sentence

----------------------------------------

No. of words in sentence

# Inverse Term Frequency =

No of sentences

---------------------------------------

No. of sentences containing Words

1. Calculate the overall probability of depressive tweets and positive tweets.
2. After that take a user query and now we need to decide that the given query is Depressed or not.
3. For this we created a “Classify” function. After that calculate the probability of positive and negative words in that query and then check which one is higher. And accordingly produced the output.

* **Accuracy of Algorithm : -**

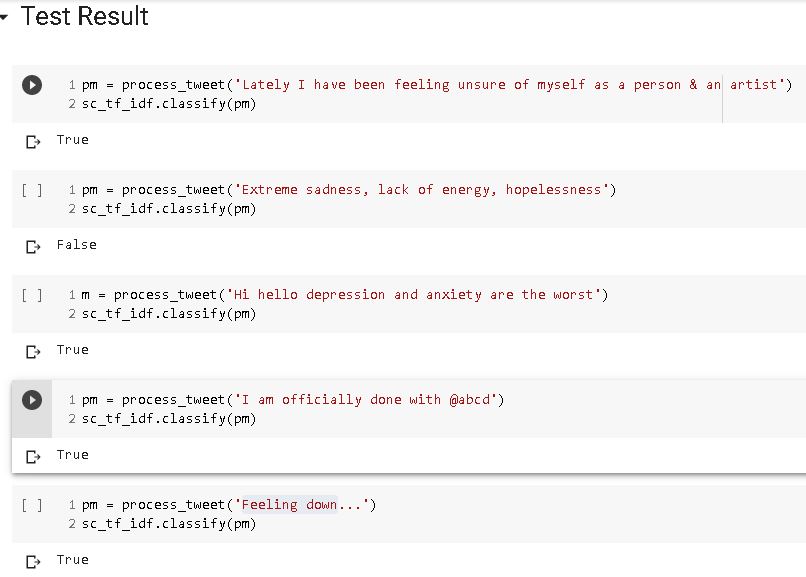
Precision: - 75.6%

Recall: - 66.7%

F-score: - 70.8%

Accuracy: - 88.67%

* **Output of Test Result : -**



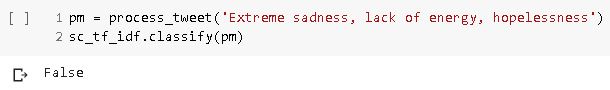
* **Advantages of the algorithm : -**

1. Accuracy of the algorithm is quite impressive.
2. Algorithm uses TF-IDF, hence predicts good output.

* **Disadvantages of the algorithm : -**

1. Predicts False for True (Depressive) statement.
2. Predicts True for False (Non-Depressive) statement.
3. Dataset is having very low rows and columns hence this problem occurs.

Sample Results to clarify the above analysis : -



* **Solution : -**

1. Use the Python 3 library named “TextBlob”.
2. Large dataset available.
3. Predicts “Polarity” and “Subjectivity”.

* **Code Link : -**

1. We used “Google colaboratory” to execute the code.
2. Link of the code : - https://colab.research.google.com/drive/12su\_xTbt2DwqJedUaiJoLyqEgjUhQ6fd?usp=sharing**.**