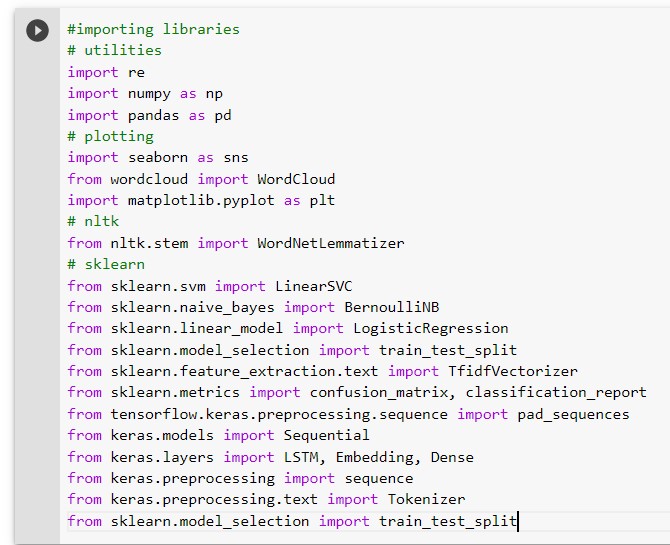
**Twitter Sentiment Analysis**

**Dataset Description:**

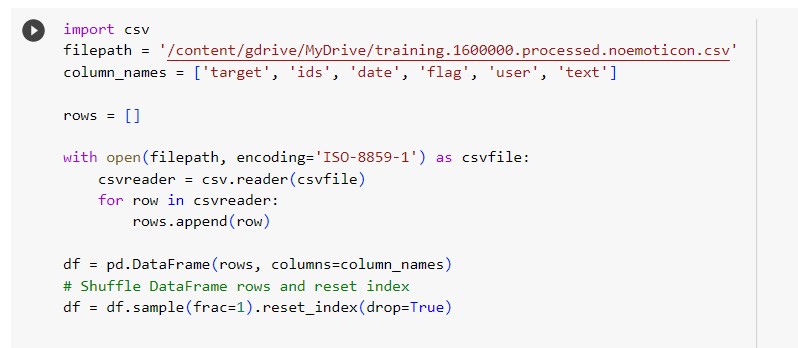
Sentiment140 is a popular dataset for sentiment analysis that consists of 1.6 million tweets. The tweets were collected in February 2009, during which time Twitter was still a relatively new platform. The dataset was created by Stanford University researchers as a benchmark for evaluating sentiment analysis algorithms. Each tweet in the dataset is labeled with a sentiment score of either positive, negative, or neutral. The sentiment scores were assigned based on the presence of emoticons in the tweet. If a tweet contained a positive emoticon (e.g., :) or <3), it was labeled as positive. If it contained a negative emoticon (e.g., :( or </3), it was labeled as negative. Tweets without any emoticons were labeled as neutral. Sentiment140 has become a widely used dataset in the field of natural language processing and sentiment analysis. It has been used for tasks such as sentiment classification, opinion mining, and emotion detection. However, it is worth noting that the dataset is now over a decade old and may not be representative of current sentiment on Twitter.

**Importing Necessary Libraries:**

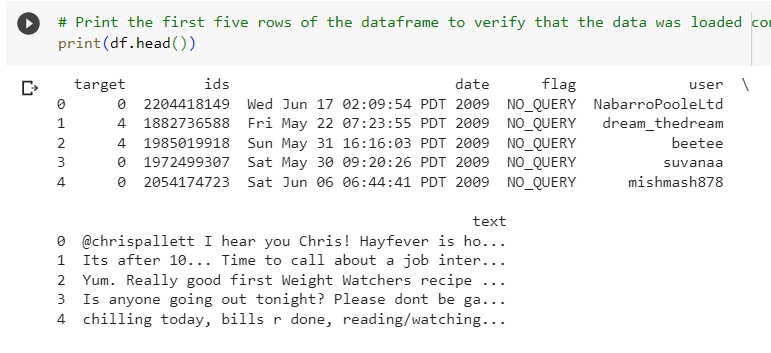


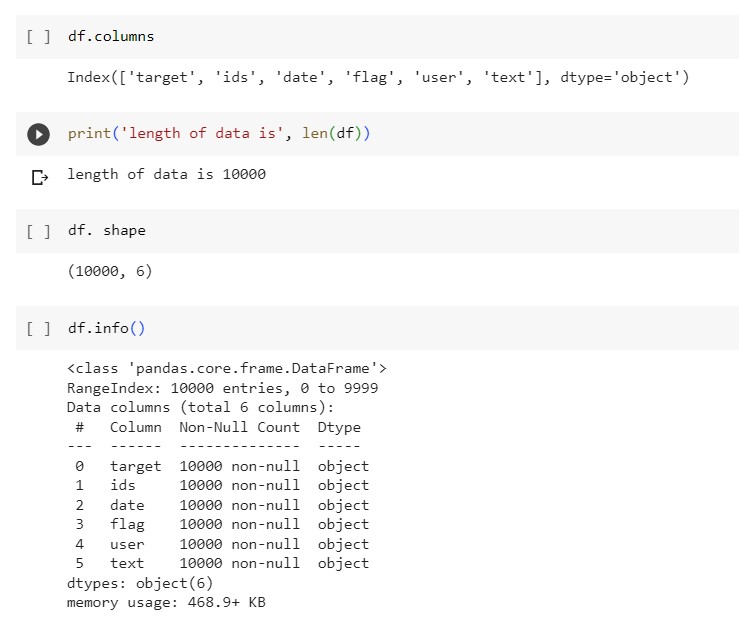
**Reading Data:**

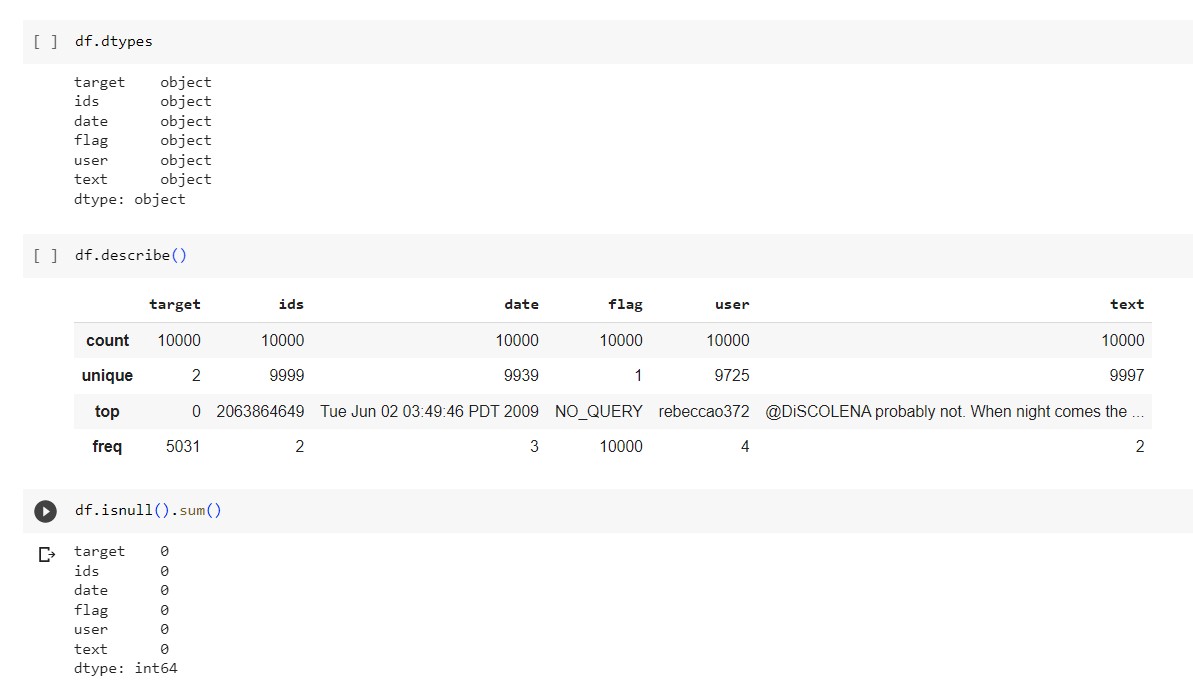
As the data contain first 0.8 million negative and next 0.8 million positive tweets thats why i have shuffled data.



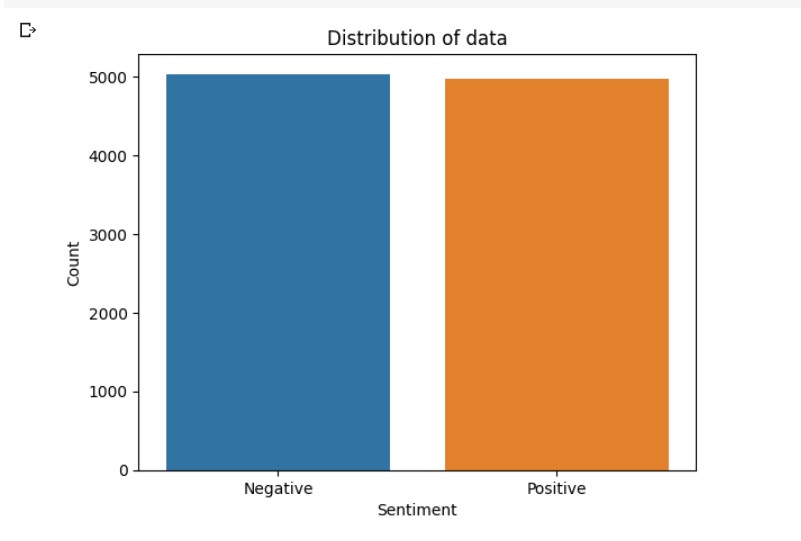
**Exploratory Data Analysis:**

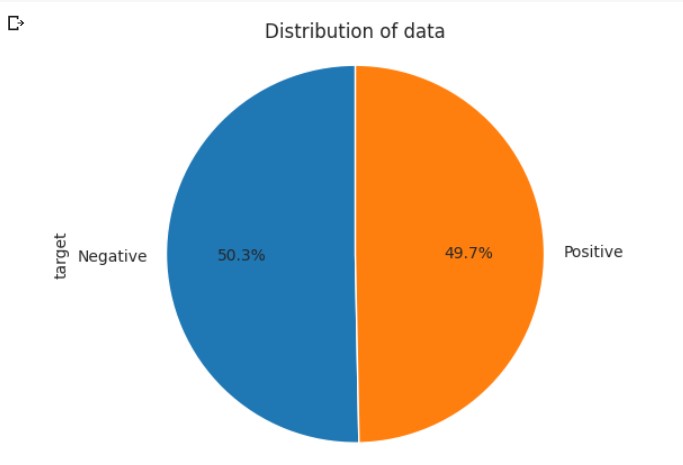






**Data Distribution:**





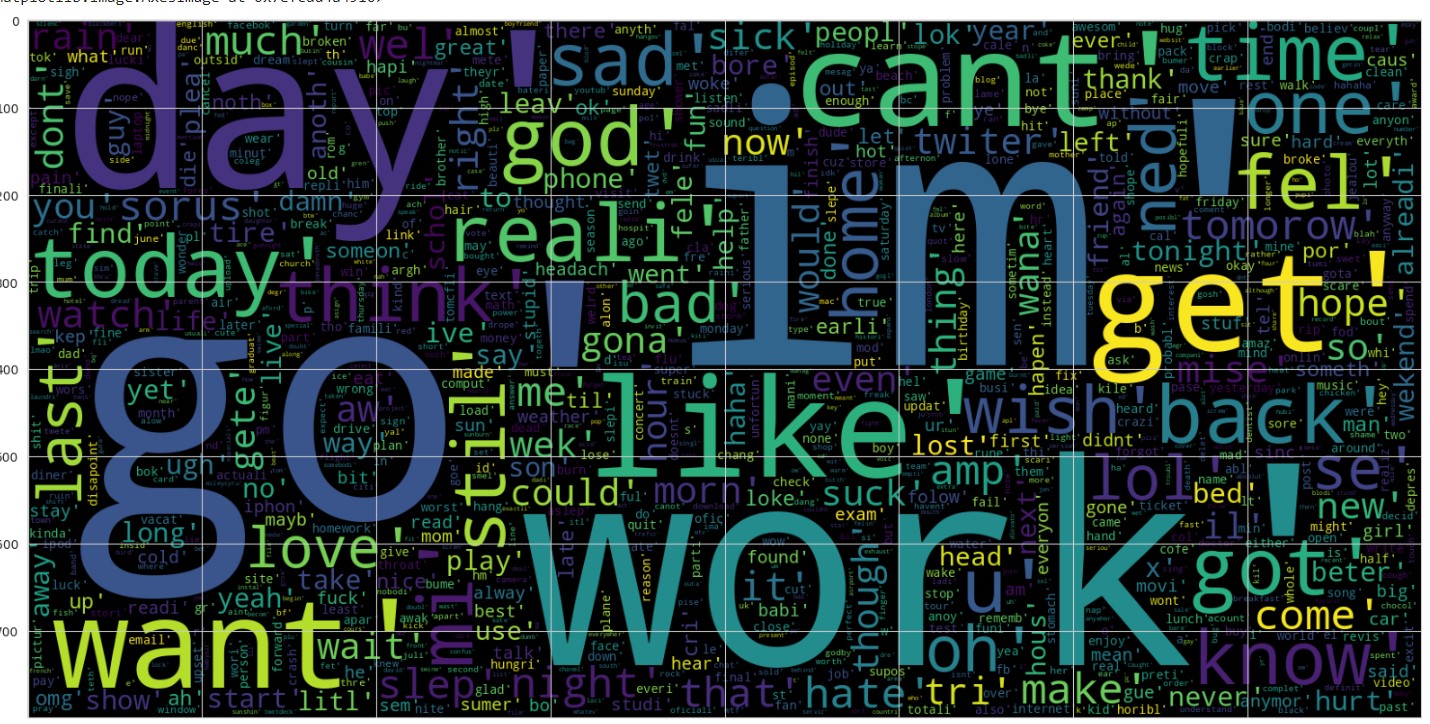
**Preprocessing:**

I have done many preprocessing steps:

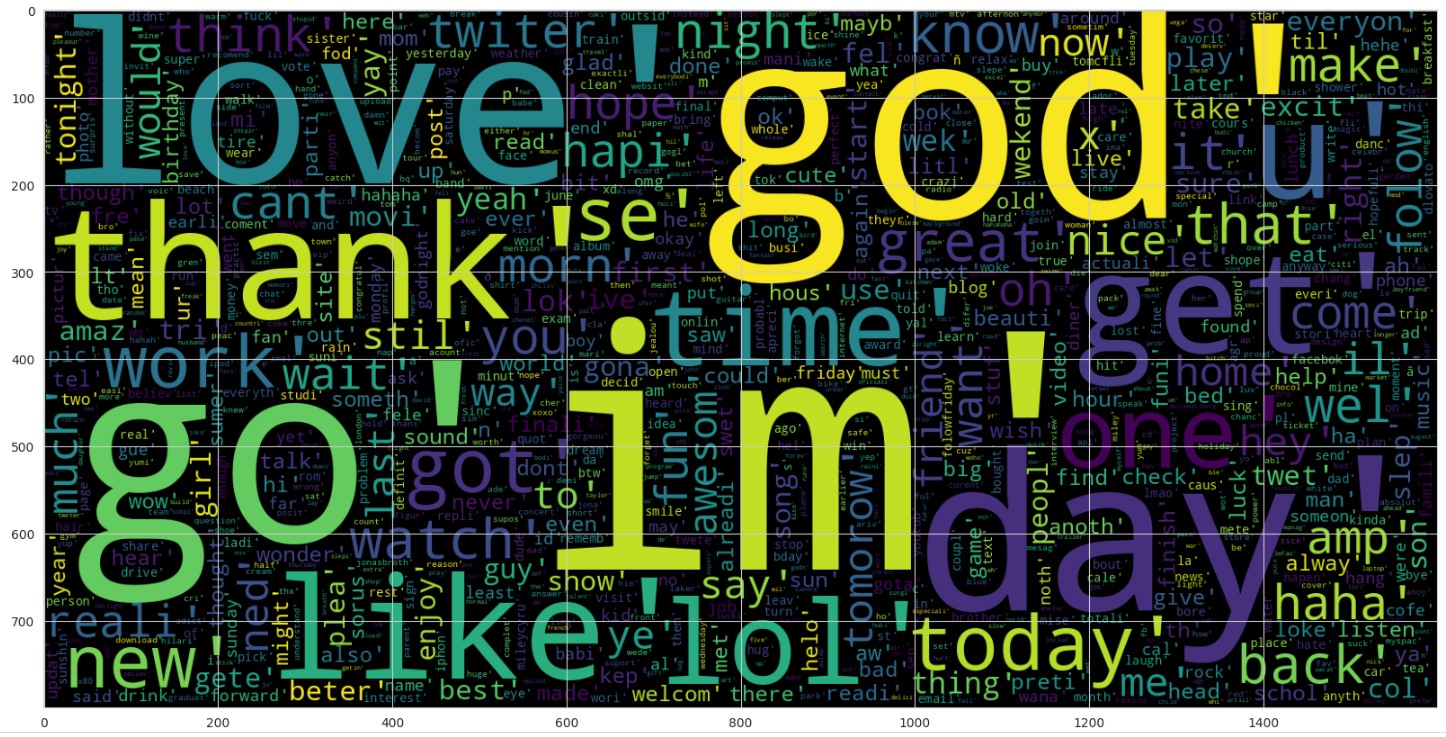
* Removing Stop words
* Removing URLs
* Removing Punctuation
* Removing Repeating Characters
* Removing Numeric Values
* Tokenizing
* Stemming
* Lemmatization



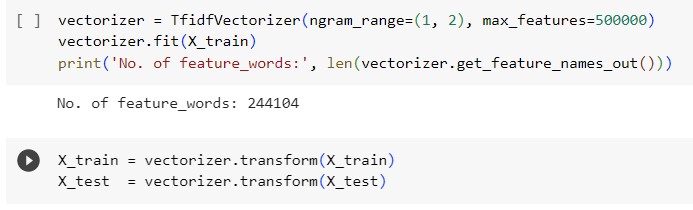
**Word Cloud for Negative tweets:**



**Word Cloud for positive tweets:**



**Tfid-Vectorizer:**

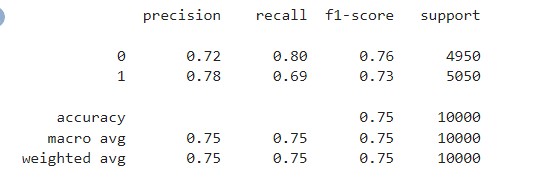


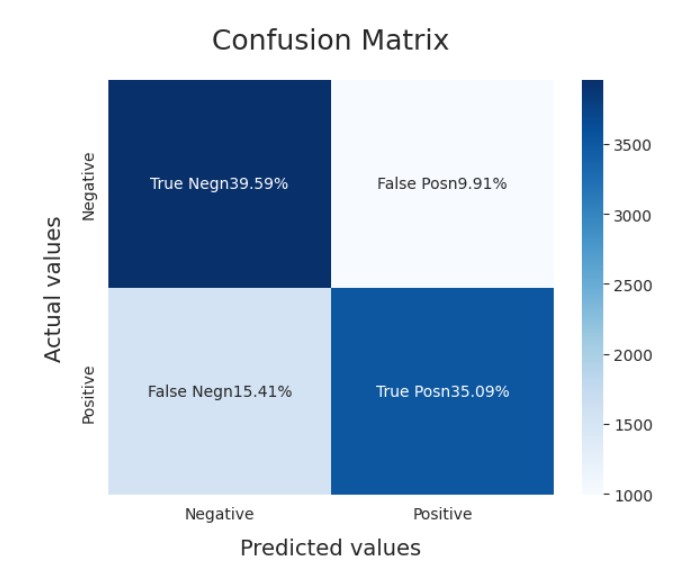
**Evaluation Function:**

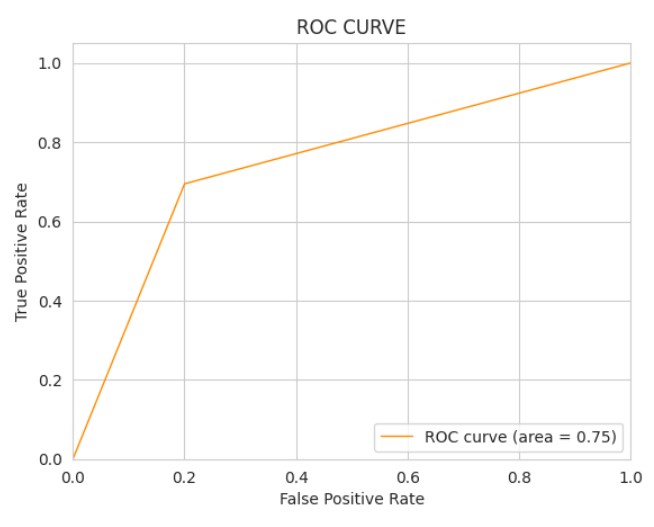


**Models:**

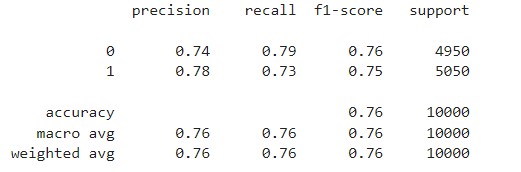
* **Naive bayesian :**

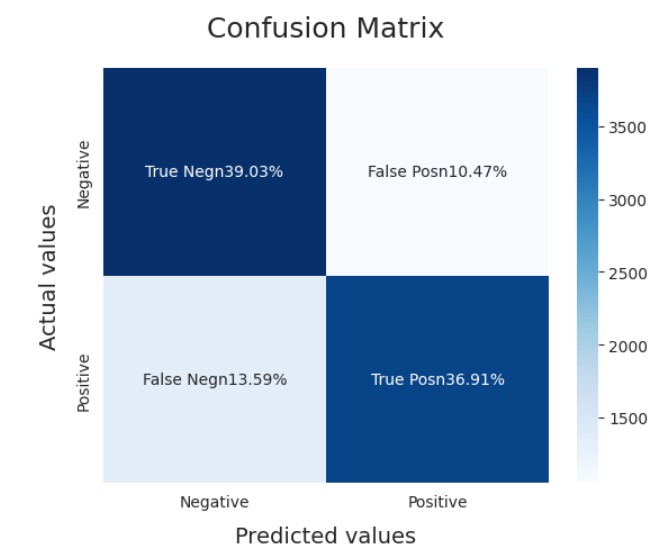


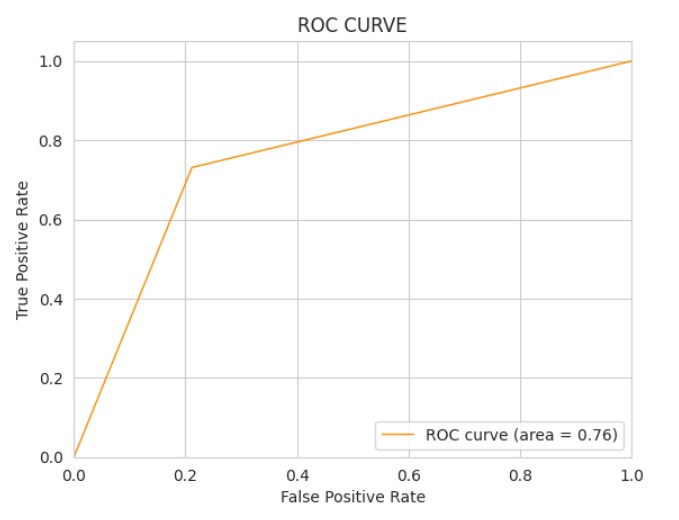




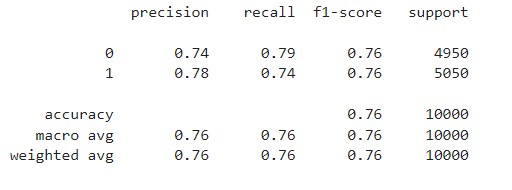
* **SVM:**

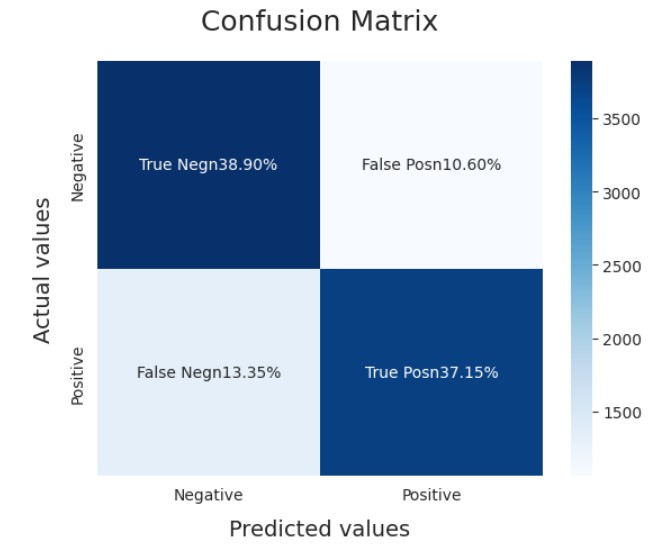


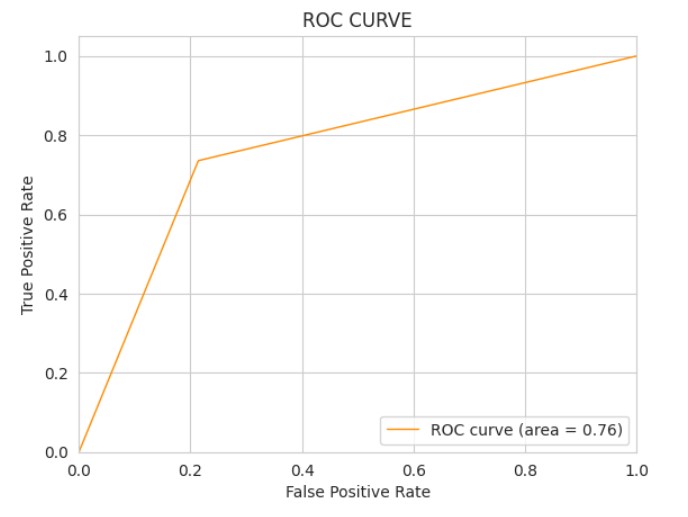




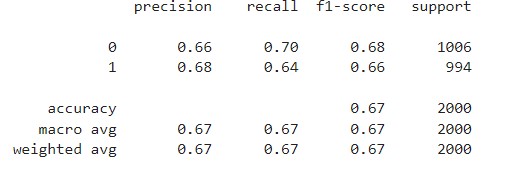
* **Logistic Regression:**

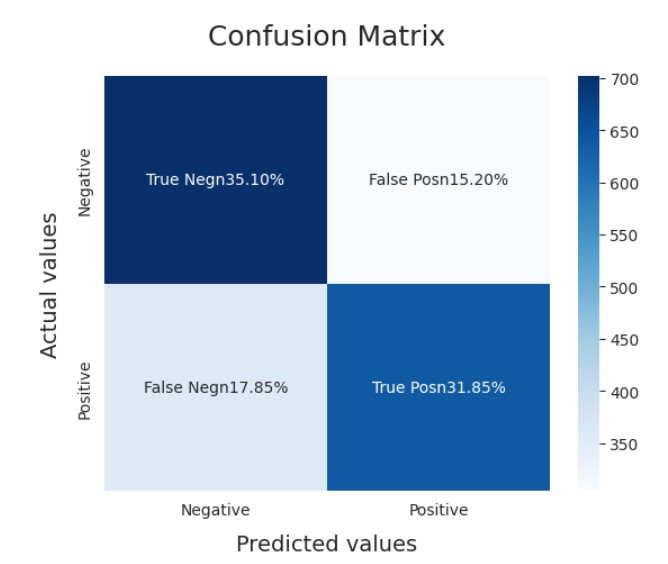


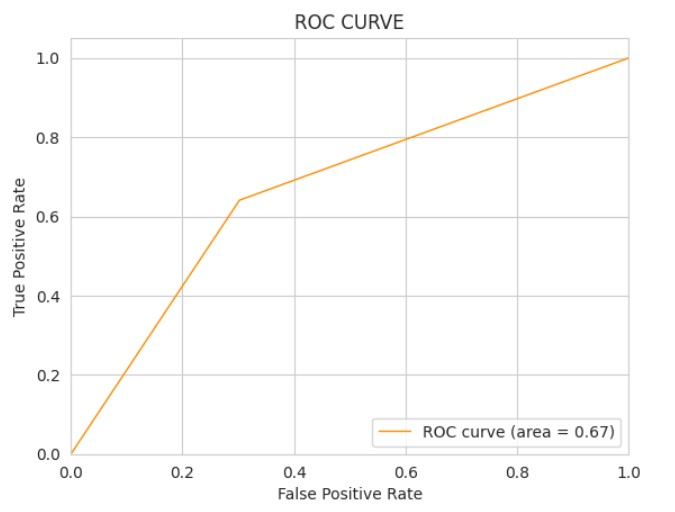




* **KNN:**







* **LSTM:**

