**Assignment 4: Report**

# **Topic:**

### A machine learning approach to classify music by mood based on song lyrics

# **Authors:**

1. Aayush Jain
2. Keshav Vyas
3. Krutika Kothawale
4. Dawna Grace

# **Objective:**

The objective of this project is to built a website on Flask framework which fetches top-k list of songs using Musixmatch API and displays each song as happy or sad based on TensorFlow model prediction to figure out overall mood of a country. The application is then hosted on Heroku platform

# 

# 

# **Dataset:**

For the purpose of this exercise, we will build models using these datasets:

Training Data:

* <https://raw.githubusercontent.com/rasbt/musicmood/master/dataset/training/train_lyrics_1000.csv>

Test Data:

* <https://github.com/rasbt/musicmood/blob/master/dataset/validation/valid_lyrics_200.csv>

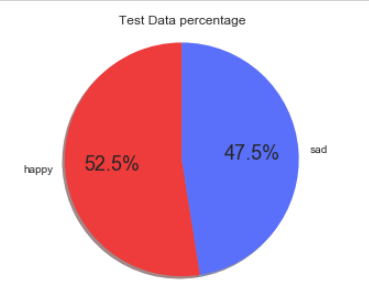
# **Data Preparation:**

## **EDA:**

1. Analysis for percentage of happy/sad songs in training and test datasets

fig1

The fig1. above shows that there are more sad songs than happy songs in the training data

fig2

The fig2. above shows that there are more sad songs than happy songs in the test data

1. Analysis for percentage of happy/sad songs in training data according to genre

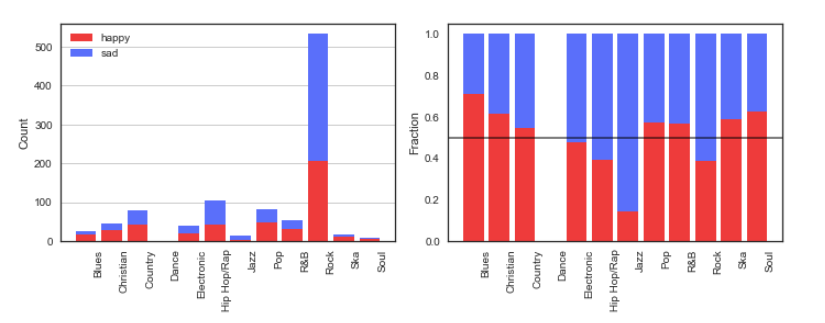
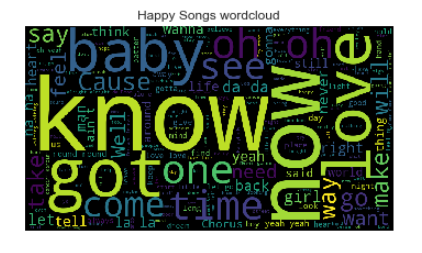


Fig3.

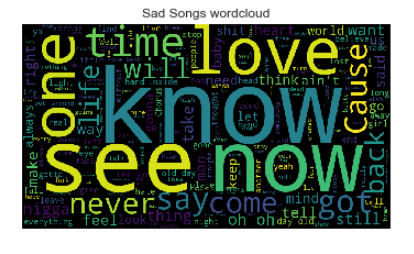
We can see observe that Jazz has highest sad songs and Blues have more happy songs

1. Analysis of happy and sad word frequency
   1. Happy wordcloud



Happy words with max number of occurance are : know, now, love, baby, got, one

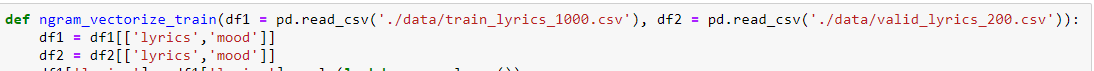
* 1. Sad wordcloud



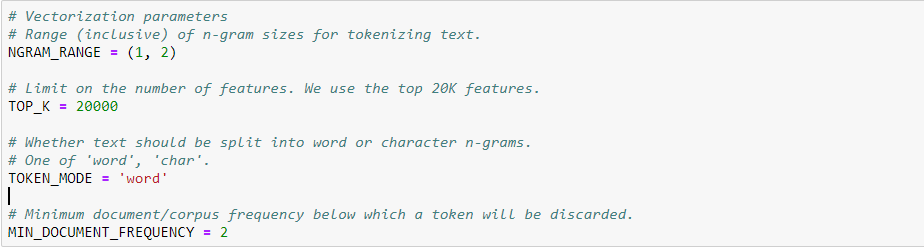
Happy words with max number of occurance are : see, now, love, got,never,cause

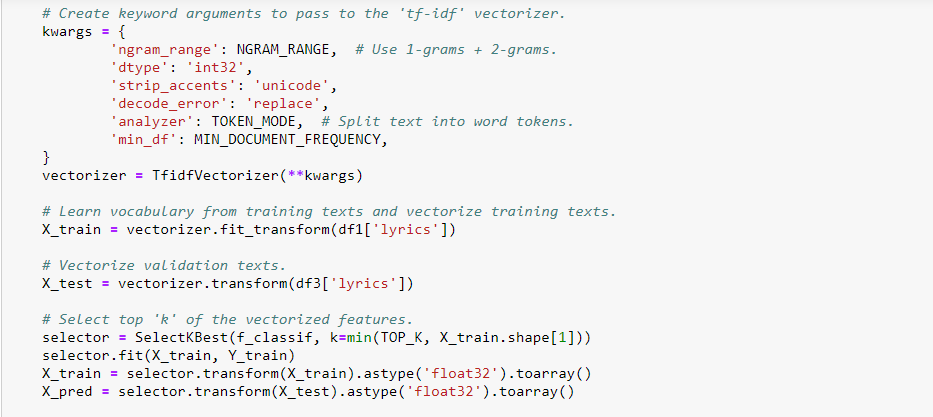
We can see **love and know** has **high occurrence in both happy and sad songs**

# **Load and vectorize data**



* Loaded the train and test datasets for from csv using read\_csv function of pandas. The train dataset has 1000 records and test has 200 records
* Extracted the column lyrics for both and cleaned the data to remove any special characters, stopwords and Lemmatized the dataset



****

* We used Tfifd vectorizor from sklearn to vectorize each work of each song lyrics in the dataset.
* We will use these vectors as input to our model

## 

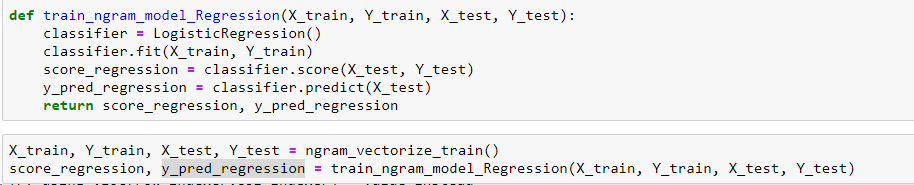
## **Predictive Modeling**

We have built 2 classification models to classify the song to happy or sad (1 or 0)

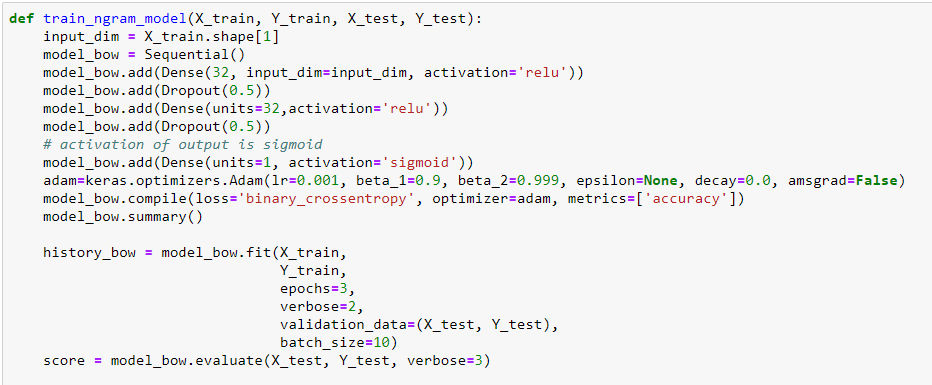
### **Logistic Regression Model**

### **Keras Model**

1. Logistic regression model



1. Keras Model



We trained the model using 3 layers and sigmoid activation function for the output layer as it has a value between 0 and 1

We used Adam optimizer and binary\_crossentropy with metrics of accuracy

After fitting the model at epochs = 3 we observed the highest accuracy with low loss and the curve flattened i.e. the accuracy continued to be the same

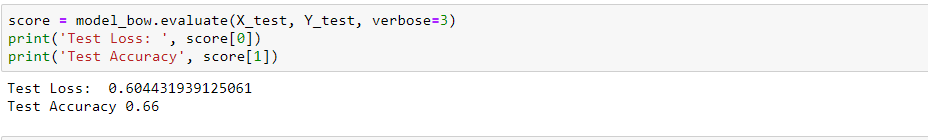
# **Metric calculation for Model:**

We calculated the following for merticis for both models and compared them:

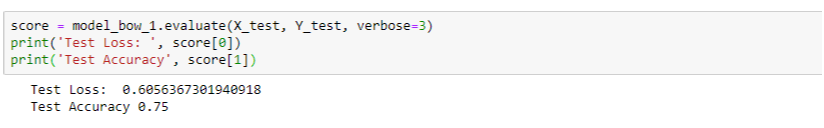
1. Accuracy and Loss
2. Confusion Matrix
3. F1 Score
4. ROC curve

**Accuracy**

Before hyperparameter tuning



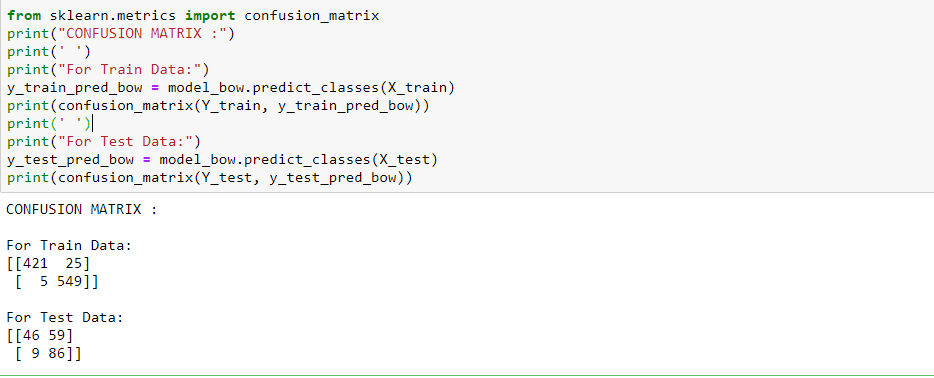
After hyperparameter tuning

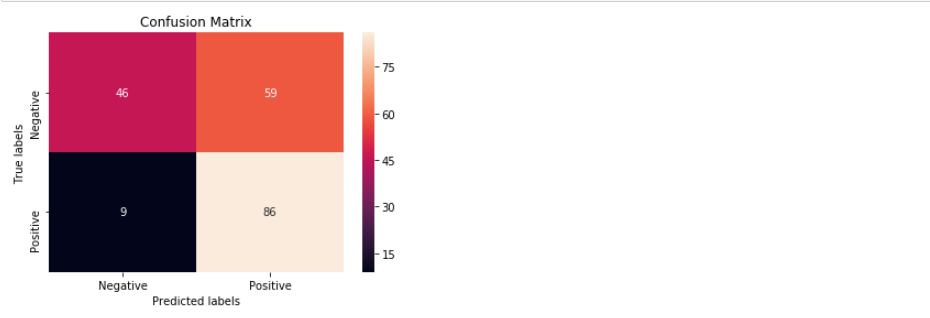


We have accuracy of 0.75 that is 75% after hyperparameter tuning.

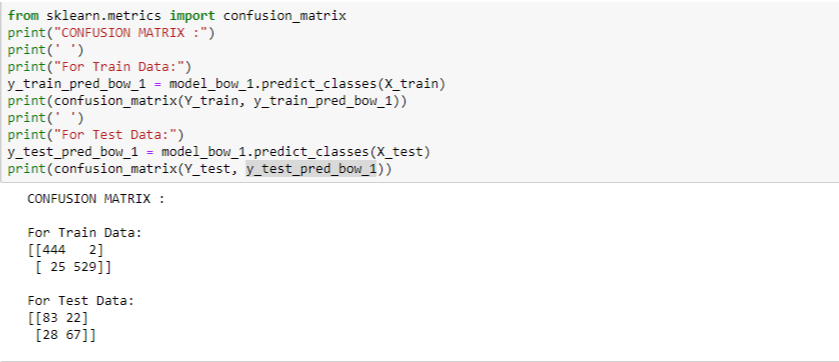
**Confusion Matrix**

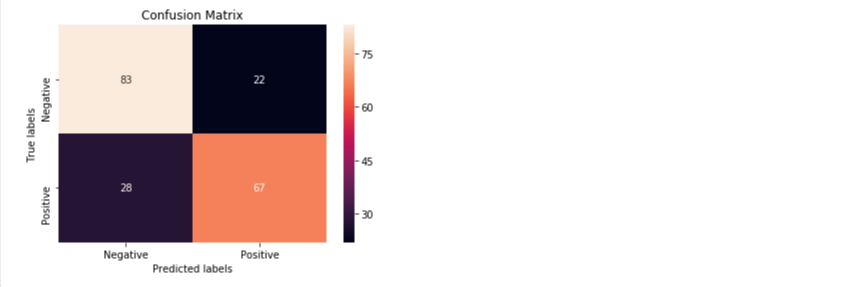
Before hyperparameter tuning





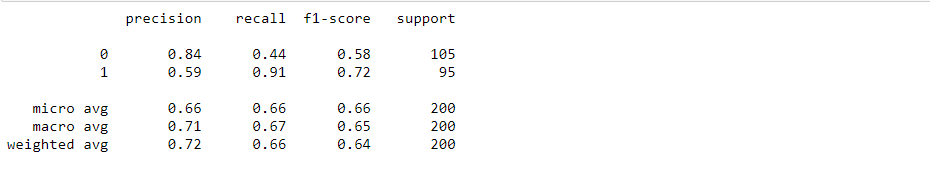
After hyperparameter tuning



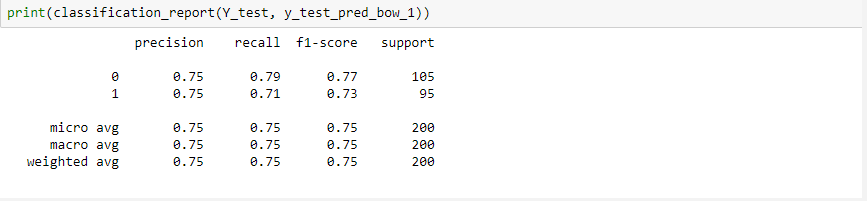


**F1 Score**

Before hyperparameter tuning

****

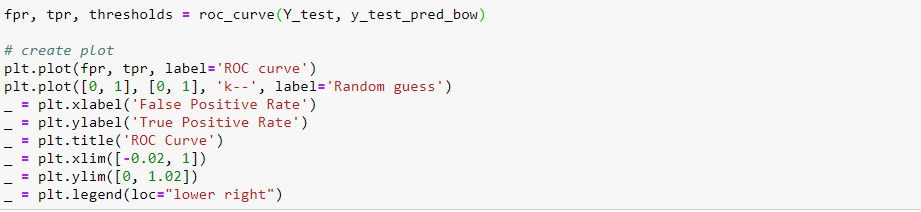
After hyperparameter tuning

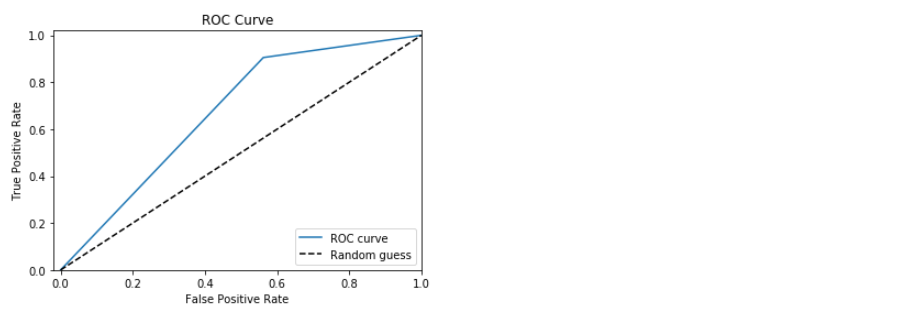


We can observe that there is a significant change in F1 score from 66% to 75% by changing the hyperparameters

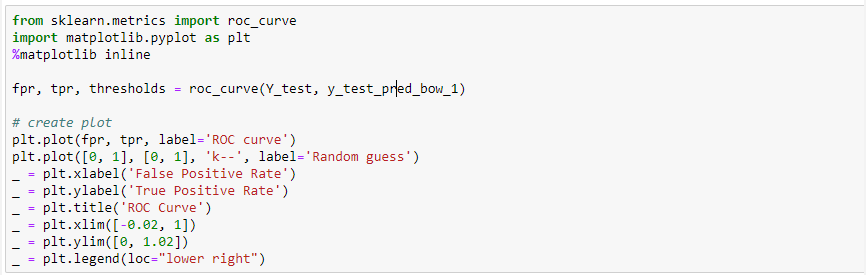
**ROC Curve**

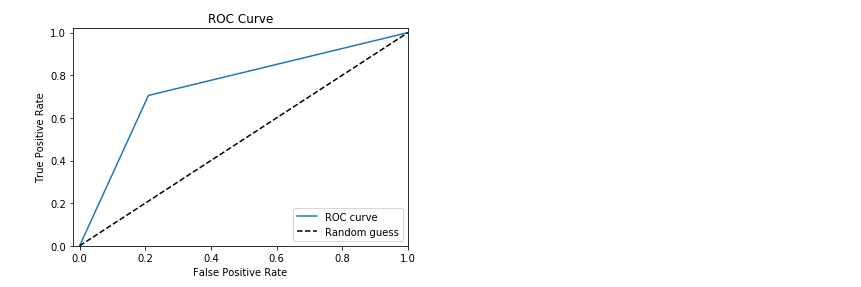
Before hyperparameter tuning

****

****

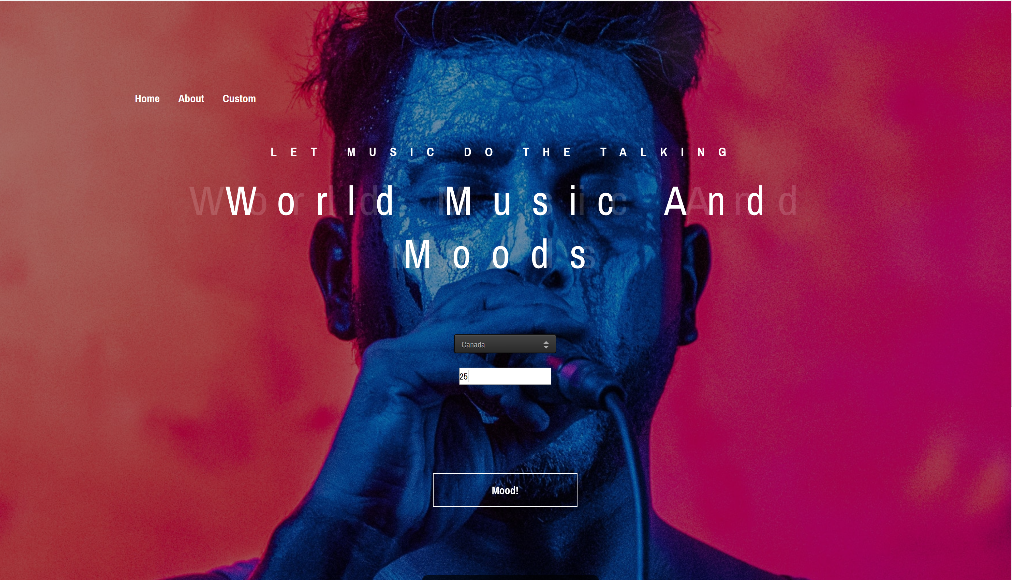
After hyperparameter tuning

****

****

## **Web application:**

* Designed python-based web application using micro web framework Flask
* Deployed the web application using cloud-based platform Heroku on Gunicorn server
* Application displays mood of the country and its songs based on top songs provided by MusixMatch API



Select the country UK, Canada, Australia and also put number of songs



Here we had selected Canada and top 25 songs. We can observe that the mood of Canada is SAD :(

