

**Team Name: Faang**

**Swetha Mamidiga ([Smami3@unh.newhaven.edu](mailto:Smami3@unh.newhaven.edu))**

**Ajay Kumar ([Avatt1@unh.newhaven.edu](mailto:Avatt1@unh.newhaven.edu))**

**Sruthi Sundararajan ([ssund3@unh.newhaven.edu](mailto:ssund3@unh.newhaven.edu))**

## 2. Research question:

### MUSIC MAKES AN IMPACT

Using data mining techniques and Spotify to predict a person's mood.

Would it be possible to predict a person's mood to play the same type of songs if a person chooses one type of song? The data behind their music listening history would reveal any insight into their emotional states.?

### Selected dataset:

The dataset which we have selected is the actual data based on the people's moods, it includes Danceability, Acousticness, Energy, Instrumentalness, Liveness, Valence, Loudness, Speechiness and Tempo because they have more influence to classify the tracks.

## 3. List of modeling techniques:

In our project we have used Regression because data in the dataset is numerical. Regression is a data science task of predicting the value of target (numerical variable) by building a model based on one or more predictors. In Regression we have used Frequency table and Covariance Matrix. In Frequency table we use classifier as Decision Tree and in Covariance Matrix we use classifier as Logistic Regression.

For Frequency Matrix –

Decision Tree Classifier

For Covariance Matrix –

Logistic Regression Classifier

For similar function:

Random forest

## 4. Parameters and Hyper parameters:

Logistic regression:

Hyper parameters:

$C = 1e20$

Decision Tree:

Hyper parameters:

We used all hyper parameters here

```
ccp_alpha=0.0, class_weight=None, criterion='gini',  
max_depth=None, max_features=None, max_leaf_nodes=None,  
min_impurity_decrease=0.0, min_impurity_split=None,  
min_samples_leaf=1, min_samples_split=2,  
min_weight_fraction_leaf=0.0, presort='deprecated',  
random_state=None, splitter='best'
```

Random Forest :

Hyper parameters:

We used all the in built parameters

## 5.About hardware:

Google Research's Colaboratory, or "Colab" for short, is a product. Colab is a web-based Python editor that allows anyone to write and run arbitrary Python code. It's notably useful for machine learning, data analysis, and education. Colab is a hosted Jupyter notebook service that doesn't require any setup and offers free access to computational resources, including GPUs.

It is easy to sync the ipynb file to our google drive and also to the Github directly

## 6. Outcomes of data mining techniques from different perspectives using varied performance metrics:

For logistic regression:

**Outcomes:**

**Train set accuracy : 67.79**

**Test set accuracy : 65.74**

Decision Tree Classifier :

Outcomes:

**Train set accuracy : 99.86**

**Test set accuracy : 70.29**

Random forest :

Outcomes:

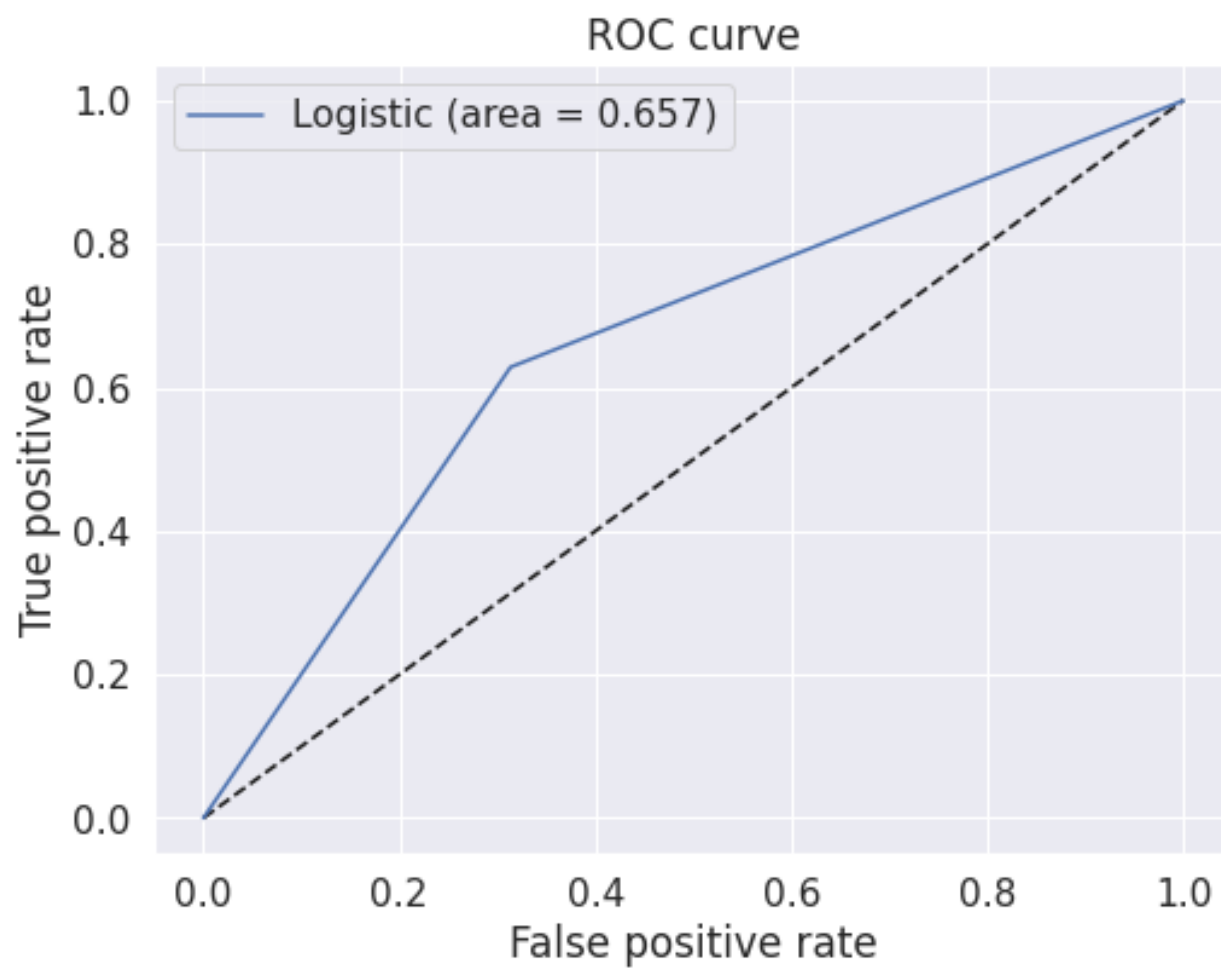
**Train set accuracy : 99.86**

**Test set accuracy : 77.82**

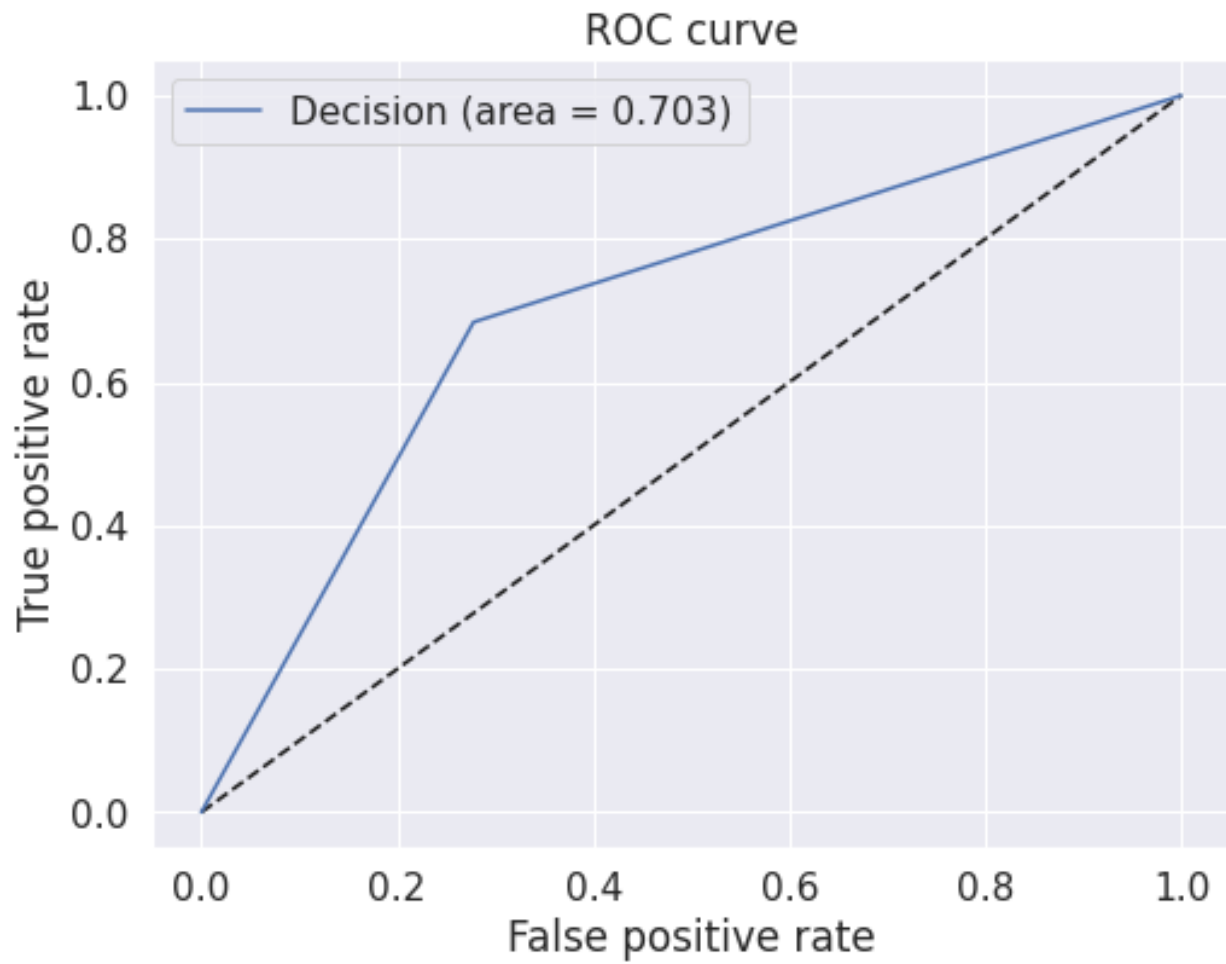
## 7. Visualization techniques used:

We have used confusion matrix which uses heat map to visualize the predicted data for the above 3 classifiers. Below are the three plots for the classifiers:

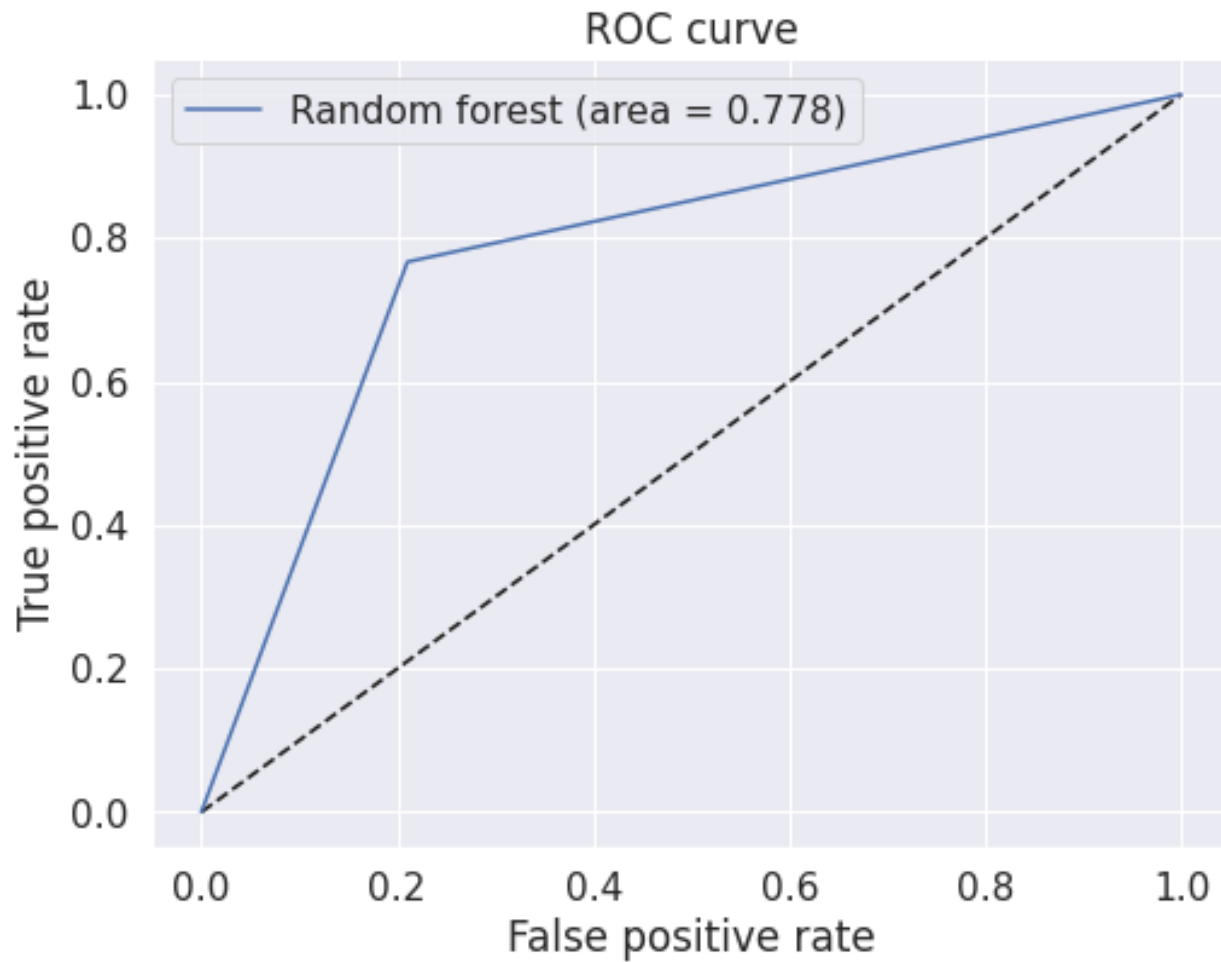
For logistic regression:



For Decision Tree Classifier :



For Random forest :



## 8.Conclusion:

The accuracy of the Random forest is greater when compared to decision tree and logistic regression. Therefore, we can conclude that Random forest is the best classifier for our project.

GitHub repository link:

[https://github.com/Ajay-kumarv/data\\_modeling.git](https://github.com/Ajay-kumarv/data_modeling.git)