# MUSIC GENRE RECOGNITION

Submitted By:
Disha Umarwani (dhu200)
Vishwali Mhasawade (vvm248)

Submitted to:
Prof. Rumi Chunara(PhD)

## 1. Problem Motivation with appropriate Background source

A human being recognizes genre of a track within seconds and so to train a machine to do the same was the prime motivation for this project. Moreover, to analyze what audio features contribute to the genre of a track was the major focus of our study. To devise a methodology to accomplish this non- trivial task of Music Information Retrieval, a field concerned with browsing, searching and organizing large music collections with the help of Data Science is the major aim.

### 1.1 Background Source :

What role does genre play in recommending a song to the user?
What does genre of a song depend on?
Genre recognition is an active area of research and Spotify, iTunes, Sawan have been working on developing methods to recognize Genre and analyze the features that contribute to recognizing the genres of the tracks.

## 2. Data Set Description

The data is subset of FMA (Free Music Archive : A dataset for Music Analysis) dataset.

There are two files from which we extract relevant data.
1. echonest.csv - This dataset is provided by Spotify. It contains the features classified into five major categories 'audio_features', 'metadata', 'ranks', 'social_features', 'temporal_features' for the track. We have used the precomputed audio features for our analysis. The metadata, rank, social_features did not contribute much to the genre.

The audio features which are present in the dataset are:

- acousticness
- danceability
- energy
- instrumentalness
- liveness
- speechiness
- tempo
- Valence

2. tracks.csv - To explore a supervised learning approach we needed the genre for each song which was present in tracks.csv. It was combined with the audio features to form the training data.

### 2.1 Data sourced from:

The dataset mainly consist of 106,574 tracks [1,3] .
Link to the dataset : https://github.com/mdeff/fma

## 3. Clear Problem Statement

We aim to correlate the audio features of a track using classical statistical methods and visual analytical methods like Self Organizing Maps (Kohonen Maps). We also aim to train a model to recognize genre using these audio features.

## 4. Model Motivation

There are two models that we've used in project to predict genres ; Gradient Boost Classifier and XGBoost algorithms.
Both the algorithms follow the principle of gradient boosting. However there are underlying differences in the regularization of the cost function used to prevent overfitting the model over the training data.

### 4.1 Why these models ?

The dataset has imbalanced classes. Certain genres (Experimental, Blues, Instrumental) have less number of observations (data points) as compared to the others (Rock, Electronic, Hip-hop). Using statistical methods like multi-class linear classification for recognizing the genres creates a major issue while dealing with this imbalanced classes. These observations are considered as noise and not much is learnt from these observations.

To avoid this situation, we have used gradient boosting algorithms that adapt well to imbalanced datasets. Gradient boosting algorithms proceed by developing several weak classifiers that perform slightly better than an average classifier and combines these to form an ensemble which performs better with the imbalanced data.

# 5. Evaluation approach

## 5.1 Data Preparation

The data preparation stage involved extracting the features from echonest.csv and target from tracks.csv.

## 5.2 Feature Engineering

Feature engineering involved finding the correlation between the different audio features. Two different approaches were used for this : Pearson's correlation coefficient and Self Organizing Maps.

- Self organizing maps helped in visualizing the data in two dimensions.
- Clustering the data in 2-dimensional space also helped in understanding similar genres.

## 5.3 Data Preprocessing

The preprocessing involved :

- Removing Nan values
- Log transforming to remove the skewness in the data.
- Label encoding the genres.
- Standard Scaling the features.

## 5.3 Modelling

- In the modelling phase we trained a Gradient Boosting Classifier and then performed a Grid Search on the hyper parameters. We used the best estimator from grid search to predict the genre.
- We then used XGBoost Classifier as well. Both the models performed similarly.

## 5.4 Evaluation Metrics

For evaluating the quality of the Self Organizing maps two error metrics were used :

- Quantization error : Average distance between the observation (data point) and the Best Matching Unit (Winning Neuron) in the topological map.
- Topographic error : Proportion of data points for which the first and second Best Matching Unit are not adjacent.

For evaluating the performance of the models :

- Confusion matrix ( consisting of the predicted and the actual labels)

- Classification report for each individual genre stating the precision , recall , f1-score , support.

## 6. Assumptions/Limitations

Following are the limitations / assumptions made :

- There might be other audio features than the ones used that might be good predictors of the genre of the track.
- We have mainly considered the 12 main genres - Blues, Classical, Electronic, Experimental, Folk, Hip-Hop, Instrumental, International, Jazz , Old-Time / Historic, Pop, Rock. There are other genres as well like Techno, Disco and this system would not be able to recognize these.
- The major limitation or drawback is that the pre-computed features were used. Audio content analysis of the mp3 files of the tracks was not performed.

## 7. Problem in Scope of class

The methodology within the scope of class involved:

- Analyzing feature Importance using Decision Tree Classifier
- Grid Search for Hyperparameters to enhance the efficiency of Gradient Boost Algorithm

The major motivation for trying something beyond the scope like Self Organizing maps was to :

- Visualize the features in a 2-dimensional space and obtain information from that.
- Cluster the data points in this space to find the similarity between the genres and similar genre clusters would be close in the topological space.

## 8. What we changed from our original project

The original proposal was written keeping a very broad perspective in mind. We narrowed down a lot after our exploratory analysis stage.

- In the original proposal we aimed at using the features from features.csv which were the temporal features . The problem with doing so however is difficult to comprehend what those features actually mean since most of them are the results of signal processing operations on the audio files.
- For analyzing the audio features we planned to take two approaches , one being the classical statistical approach and the second one being the visual analysis technique of Self Organizing Maps. The idea behind this was to discover whether the results of the two approaches were similar.

- Unsupervised learning (K-means clustering) was also performed on the 2-dimensional data obtained from Self Organizing Maps which helped understand similar genres. This approach was later decided for analyzing the audio features.
- We limited on the number of questions we could answer like predicting the favorite songs and downloads of a song and did not dwell into that.
- Since we did not include audio files, their representation could not be learned with the help of Recurrent Neural Networks. However this representation was done with the help of Self Organizing Maps.

## Distribution of Grades

Disha Umarwani - 5
Vishwali Mhasawade - 5

## References :


1. https://github.com/mdeff/fma
2. https://arxiv.org/abs/1612.01840
3. https://github.com/sevamoo/SOMPY
4. http://ivape3.blogs.uv.es/2015/03/15/self-organizing-maps-the-kohonens-algorithm-explained/
5. Neural Networks : A Comprehensive Foundation - Simon Haykin