

# Final Project Proposal

**Group Participants:** Umme Fatema Piu, Rahul Menon, Vivan Chopra

**Topic:** Classification of Music Genre.

## Data Source:

For the classification of music of different genres and in order to ensure testing can be carried out efficiently, the data set selected was ("GTZAN Dataset - Music Genre Classification") genre collection dataset as it had a large selection of genres, covering hip-hop, jazz, rock and many others which would create a good mix for classification.

## Intended Machine Learning algorithms:

As for the libraries that we intend to implement, scipy will be used since the audio files that we will classify are in WAV format and also allow us to perform the mathematical tasks and data plotting.

We will calculate distance by finding the K number of neighbors. We shall develop several functions to accomplish this exclusively for each functionality. The first thing we'll do is create a function that takes training data, existing instances, and the necessary number of neighbors. It will measure the separation between each point in the training set and every other point, locate the nearest K neighbors, and then return all neighbors. In order to make a project process clear and straightforward, we will build a function to determine the distance between two points. We are also planning to use Support Vector Machines to support our analysis in terms of classifications.

## Model Evaluation:

In order to evaluate the performance of the models we will use accuracy, F-score and area under the curve (AUC). We also need to test the accuracy and precision of the model and we will add an accuracy calculator function that will test the correct predictions by the total number of predictions.

## Role of each team member:

### Vivan Chopra:

- Will research about the machine learning algorithms and methods that we will use in the project such as K-Nearest Neighbors and find the maximum nearest neighbors count
- Will help in the model evaluation and work on the accuracy calculator for the program
- Will contribute towards the project report and documentation and ensure I am available for the group meetings project work

### Umme Fatema Piu:

- Will contribute on features statistics, data processing (i.e;handling missing values, variable transformations), cross validation for hyper parameter tuning, test accuracy check and project documentation.

### Rahul Menon

- Will Contribute in Group Meetings and documentation for the project.
- Will help in writing test cases for our functions to ensure our code has proper functionality.
- Will help in writing the main functionality in Python for the project.

**References:**

Kaggle. Retrieved October 21, 2022, from

<https://www.kaggle.com/datasets/andradaolteanu/gtzan-dataset-music-genre-classification?select=Data>  
Bahuleyan, H. (2018). Music Genre Classification using Machine Learning Techniques. *ArXiv*,  
*abs/1804.01149*.

Music Genre Classification Project Using Machine Learning Techniques. Retrieved October 21, 2022,  
from

<https://www.analyticsvidhya.com/blog/2022/03/music-genre-classification-project-using-machine-learning-techniques/>

DataFlair. Retrieved October 21, 2022, from

<https://data-flair.training/blogs/python-project-music-genre-classification/>