# Music Genre Prediction

Reeya Gupta
Machine Learning Project

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

## Abstract

The project is inspired by the fact that music plays a significant role in our daily lives, and that different people have different tastes in various musical genres. It can be difficult for a music streaming service to suggest songs to users that fit their musical tastes. Building a music genre prediction system to aid music streaming platforms in making user-based song recommendations is the driving force behind this project .The goal is to create a tool that will help musicians, music fans, and industry professionals organize, classify, and suggest music to listeners based on their preferred genre. The project also seeks to clarify the connection between musical elements and genre categorization.

## Introduction

"Music Genre Prediction" aims to classify and predict the genre of a given song and categorizes music into different genres using machine learning algorithms. Python and various machine-learning libraries are used to implement the project.

## Methodology

Collecting and preprocessing the data required for the project: The data collection process involves querying the Spotify API to collect the features for all tracks in a given playlist, while the preprocessing process involves dropping unwanted columns, splitting the data into features and labels, encoding the labels, scaling the features, and splitting the data into training and testing sets.

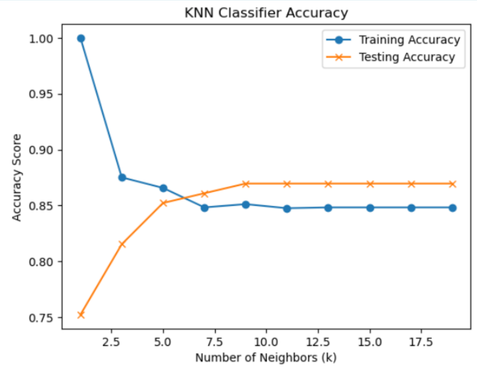
Exploring and analysing the data to understand its characteristics: The dataset used in the project contains audio features of over 4000 tracks from 16 different genres. The data preprocessing section of the code includes standardizing the data, encoding the target variable, and splitting the data into training and testing sets.

Extracting relevant features from the data for model building.

Building and evaluating machine learning models for music genre classification.

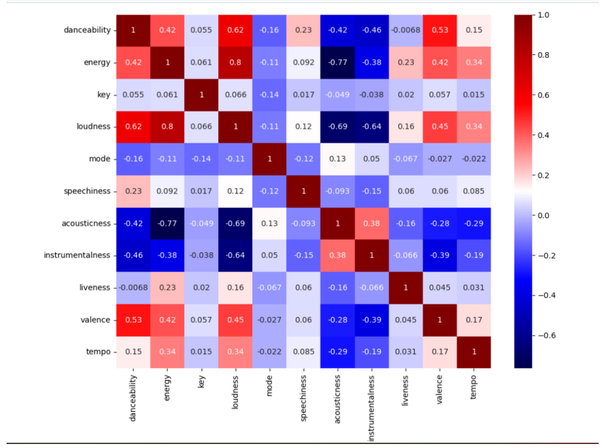
Comparing the performance of different models Discussing the results and insights obtained from the project and providing recommendations for future research.

## Results




K-Nearest Neighbors classification - plotting the accuracy of the model for different values of k.

Heatmap for Visualizing the correlation between various audio features of the music tracks in the dataset:




Evaluating the performance of a trained neural network model on a dataset:

```
In [420…
model_loss, model_accuracy = model.evaluate(
    X_test_scaled, y_test_categorical, verbose=2)
print(
    f"Normal Neural Network - Loss: {model_loss}, Accuracy: {model_accuracy}")
```

```
20/20 - 0s - loss: 1.6301 - accuracy: 0.4319 - 124ms/epoch - 6ms/step
Normal Neural Network - Loss: 1.630128026008606, Accuracy: 0.4318554997444153
```

## Conclusion

The models were trained on a dataset containing various audio features extracted from over several thousand songs across 16 different genres.

The project concluded that machine learning models such as K-Nearest Neighbors (KNN), Random Forest, and Support Vector Machine (SVM) can predict music genres with an accuracy of over 60%. Additionally, the project found that certain features such as tempo, loudness, and energy were more important in predicting genres than others.

The algorithms used in this project are:

Decision Tree: 68.25%

Random Forest: 73.67%

K-Nearest Neighbors (K-NN): 55.58%

Logistic Regression: 72.89%

Support Vector Machine (SVM): 68.03%

Based on these results, Random Forest is the best-performing algorithm with an accuracy score of 73.67%. It is followed closely by Logistic Regression with an accuracy score of 72.89%.