

# Music Genre Classification

Hyeonjung Ko, Yash Shetty

November 8th, 2019

## Abstract

Classifying raw music files into distinct genres is a challenging task with numerous practical applications. In this project, we compare the performance of three different classifiers—Logistic Regression, Random Forest, and CNN/RNN—in the task for music genre classification for a given audio file.

## 1 Introduction

Through this project, we aim to classify unseen music data into genre classes. We aim to maximize accuracy while minimizing the length of the given music data. Music has always been a part of people’s lives. There can be numerous potential applications for music genre classification.

With the rising demand for multimedia data on the internet as well as multimedia databases, there is a massive need for the automatic classification, indexing and retrieval of information. In particular, this is true for music files with the ever-increasing number of online music stores and music search applications that need an efficient way to sort and store their big data. Two approaches that could be used to achieve this end are (a) keyword based indexing, which would involve the highly tedious and inefficient procedure of human annotation or (b) content-based indexing that could be achieved using automatic classification. This where our models can kick in and save time, effort and money.

## 2 Proposed Project

We plan on implementing 3 different classification methods: Logistic Regression, Random Forest, and CNN/RNN. We will be using the GTZAN Genre Collection dataset. The audio file data will be converted into spectrograms, which then will be used to extract features. Additional datasets from the Spotify API and Audio Set could be utilized to train models with the same types of classifiers for a comparative study of performance and prediction accuracy.

GTZAN Genre Collection dataset contains 1000 audio tracks each with length of 30 seconds. It contains 10 genres, each represented by 100 tracks. The tracks are all 22050Hz Mono 16-bit audio files in .wav format. For analysis, audio file to spectrogram conversion, and additional feature extraction, we will use the Librosa Python library.

With logistic regression, we want to compute  $k$  different decision boundaries, each classifying a class versus all else. With random forest classifier, we want to use the results of multiple decision trees to classify data. The larger the number of decision trees in the classifier model, the more we can counter-balance the errors from the individual ones. With the CNN/RNN, we will train the model end-to-end. It will predict the genre of the given audio file using just the audio spectrogram.