Music Genre Identification

Problem Statement & Background:

Music has the power to uplift us, reduce stress, inspire creativity, promote physical activity, and connect people together. No matter the purpose, there is a genre for everything. By exploring genres that one particularly enjoys, they can find new artists and songs that resonate with themselves and enrich their musical experiences to their own needs.

Given that radio shuffle features from music applications often fall short of recommending related songs, there is a need for optimal music genre identification. Challenges of recommending music based on genre include the nonexclusive and broad nature of music. For instance, music is a culmination of genres–classifying audio as a single genre is rare. Moreover, a genre can be defined by features beyond the audio, such as culture or location.

Considering these complexities, the goal of this project was to build a machine learning model that can accurately classify a song into one of nine different musical genres based on its acoustic features. While the nine genres included in this project may not be sufficient for a radio recommender, a fine-tuned model to hundreds or thousands of subgenres can capture specifics and recommend similar songs. Additionally, this technology could be used for automatic genre tagging in large music libraries to assist in organization and structure.

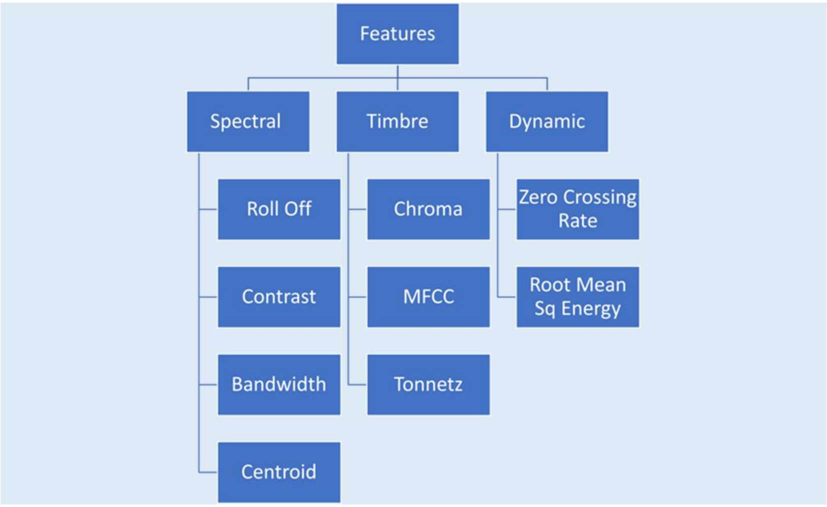
Data:

Given licensing issues with sharing commercial music, biases towards specific genres or cultures, and limited data sets, finding a musical data set aligning with specific criteria can be difficult. To overcome this challenge, I used the Free Music Archive (FMA) dataset, which is a collection of about 100,000 high-quality, legal audio tracks spanning 161 genres and subgenres. The FMA dataset is a prime option because it is open-source.

The acoustic features were extracted from the FMA audio–over 340 days of audio–in a paper called "FMA: A Dataset for Music Analysis," Defferrard et. al. (2017). While the FMA dataset has grown since 2017, I am working with a snapshot of the library from 2017. The paper collected metadata such as artist information, publication year, and song titles along with the acoustic features. However, my project focused solely on classifying music based on acoustic features, and therefore, I opted not to utilize the metadata.

The features data included 106,000 tracks, each with 518 acoustic features describing it. There are nine acoustic features that be organized into three main bins: spectral features, timbre features, and dynamic features (Figure 1). Spectral features provide information about the sound's brightness, while timbre features describe the tonality, which is determined by the harmonic content of the sound. Dynamic features, on the other hand, capture the loudness or energy of the audio. Each one of these features had the following seven statistical measures calculated on it: mean, standard deviation, kurtosis, minimum, maximum, median, and skewness.

Figure 1. Nine Acoustic Musical Features

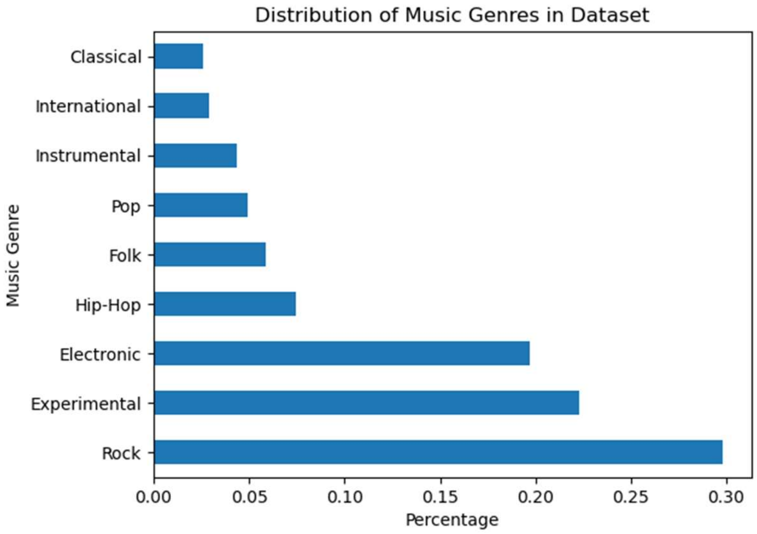


Data Cleaning & Exploratory Data Analysis:

The original data came in three different tables: features, tracks, and genres. All three tables were merged and the metadata was removed to achieve my desired dataset, which included only the features and the genre. After removing two redundant chroma features, the final dataset contained 350 features. An inclusion criterion was songs with one parent genre, which decreased the dataset to around 50,000 samples.

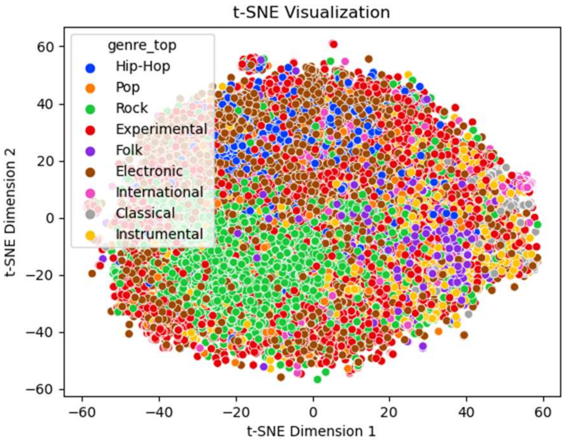
In the exploratory analysis, histograms of each statistical measure per genre were produced to analyze variance between genres. All of the features for the models were used because each feature could potentially contribute to the classification task and, thus, removing some could lead to loss of information. The final dataset had a skewed distribution with a large class imbalance (Figure 2).

Figure 2. Music Genre Distribution



The t-SNE dimensionality reduction technique was employed to create a 2-D visualization of the genres (Figure 3). While some genres were clearly clustered in specific areas, others such as international and experimental appeared to be dispersed throughout the graph, making the classification task more challenging due to their ambiguous nature.

Figure 3. 2-D Visualization of the Genres



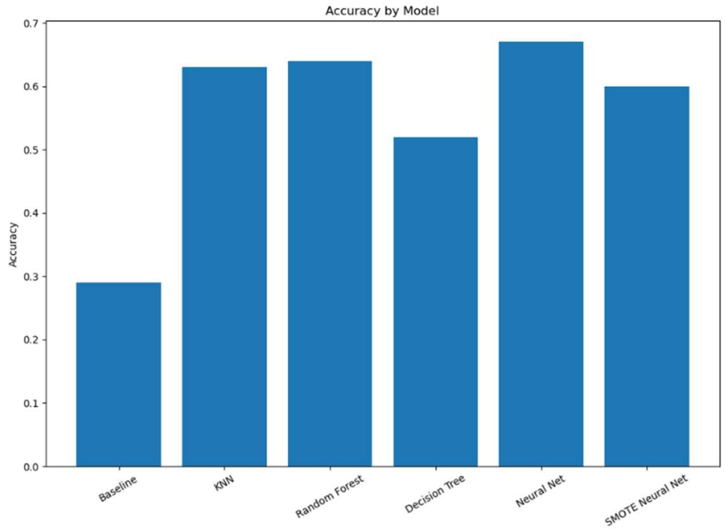
Modeling:

The modeling technique was to cast a wide net of models and determine which ones had the best performance across two main metrics: accuracy and weighted f1-score. Using weighted f1 is particularly useful to determine success in an imbalanced dataset because it considers the trade-off between precision and recall for each class. It provides a single metric that combines the performance of all classes, weighted by their support (the number of samples in each class). To address the data imbalance, a model was trained with synthetic minority over-sampling technique (SMOTE), which creates artificial minority class sample to balance the training data. In total, 5 different classification models were trained: decision tree, random forest, KNN, neural net, and a SMOTE trained neural network. For each of these models, both a baseline with the default model parameters and an optimized model with tuned the hyperparameters were trained. This approach led to insightful information regarding which models to further pursue for future improvements or updates to this project.

Findings & Conclusion:

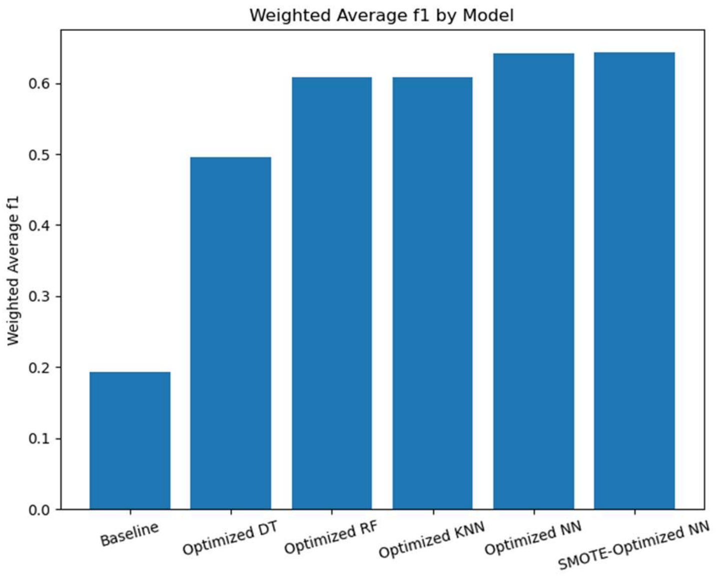
The optimized accuracies of each model were plotted in Figure 4. Based on accuracy, the neural network with 3 hidden layers and nearly 50,000 trainable parameters was the top performer, achieving an accuracy rate of almost 70%. It's worth noting that all the models performed well above the baseline accuracy of 30%.

Figure 4. Optimized Accuracies of each Model



The weighted F1 score for each model is depicted in Figure 5, revealing that the two top performers were both neural networks. While the SMOTE-trained NN had a slightly higher F1 score, its accuracy was slightly lower. On the other hand, the decision tree model performed the worst across both metrics, which was somewhat expected given its simple nature.

Figure 5. Weighted F1 Score of each Model



To conclude, the neural networks outperformed the other models in this genre identification task. Although I did not test any other models using SMOTE, I believe their performance would correspond to their architectures and capabilities. The results suggest that neural networks hold great promise as an effective approach for this task.

Next Steps:

I have several ideas that I believe can improve the performance of the models. I was particularly impressed with the random forest model and would like to evaluate its performance after incorporating SMOTE. Moreover, I intend to rerun the models after removing the international and experimental categories, which could potentially allow the models to discern clearer boundaries between genres. This would be a challenging task but could lead to exciting new insights. Additionally, I am intrigued by the idea of training a model using spectrograms of song clips and comparing its performance to running the models with only spectral data. This could be an entirely new project but would be a fascinating avenue to explore.