Genre Prediction Capstone Project Report

This project aims to predict the genre of songs given their audio features. Audio features, as well as other track-wise information, were collected from Spotify. Since Spotify's API doesn't offer genre information, I turned to The Sounds of Spotify, an independent music genre project that is described well here, to generate the dataset of songs and their genre labels that were used to train various predictive models.

In exploring and comparing the differences in the distributions of the audio features of songs from each genre, it became apparent that each major genre, as described by the data, is unique in their own ways, and the combinations of their subjective features are significantly different from one another. Pop songs are characterized as moderately danceable and energetic on average, and they are almost twice as likely to be in a major key rather than a minor one. R&B songs are also moderately danceable and energetic on average, but they are the only genre whose songs are more likely to be in a minor key than a major one. Hip Hop songs are characterized as highly danceable, energetic, and most importantly, very speechy on average. Latin songs are not only characterized as highly danceable, energetic, but they are also very positive on average comparatively to most other genres. EDM songs are very energetic on average, and they also tend to have much higher tempos than other genres. Reggae songs have the highest valence on average and are generally have higher speechiness than most other. Indie songs are very energetic on average and are over twice as likely to be in a major key than a minor one. Rock songs are moderately danceable, energetic, and valent on average. They're also three times as likely to be in a major key than a minor one. Metal songs are the most energetic on average comparatively to other genres and are not acoustic whatsoever. Rock songs are moderately danceable, energetic, and valent on average. Their features' averages resemble those of the rock genre, but there is an even smaller proportion of songs in a minor key. Jazz songs are very acoustic, and typically, they're not energetic. Classical songs are, in fact, the most acoustic and the least energetic of all genres. Lastly, both Jazz and Classical songs tend to be significantly longer than the other major genres.

To determine whether the audio feature representations of each genre are significantly different from one another, firstly, a series of independent t-tests were conducted on each pair of audio features between the two genres. Then, a logistic regression was conducted using the audio features of the genres as the independent variables and the genres themselves as the dependent variables. For each pair of similar genres (Rock vs Country, Rock vs Metal, Hip Hop vs R&B, Pop vs Rock, Pop vs Latin, Pop vs Indie, and Pop vs Country), all of the t-tests were highly significant ($p < 0.01$) and the Logistic Regressions all had generally high accuracies (above 65%).

To predict the genre of new songs, I sought to train various classification models on the audio features of the songs from each genre and compare their performance on unseen data. The Decision Tree Classifier was 66% percent accurate on average. The genres that the model learned best were Classical, Jazz, and Metal, whereas Pop was the most difficult genre to classify. I was sure that ensemble methods would outperform this model, but I thought that it was a good measure of just how much other algorithms improve performance overall. The Random Forest Classifier, with 498 estimators, as suggested by a randomized search of hyperparameters with cross validation, performed far better, averaging 75% accuracy. This model still picked up Classical, Jazz, and Metal the best, but its performance on the remaining genres significantly improved across the board, as well. On the other hand, the Gradient Boosting Classifier, with 455 estimators, a 0.1 learning rate, and a maximum tree

depth of 4, as suggested by a randomized search of hyperparameters with cross validation, was only 74% accurate on average. Lastly, the Extreme Gradient Boosting (XGB) Classifier, with 401 estimators, a 0.25 lambda rate, a 0.1 gamma rate, a 0.1 learning rate, and a maximum tree depth of 7, as suggested by a randomized search of hyperparameters with cross validation, was 75% accurate on average, and was the best performer comparatively to the other models.