We set out to tackle an interesting challenge: can we predict a song's genre just from its audio characteristics? Working with 50,000 tracks from Spotify (5,000 songs in each of 10 genres), we built a machine learning system that analyzes features like danceability, energy, and loudness to guess the correct musical genre. To get the best results, we created a five-step process: cleaning the data, exploring patterns through dimensionality reduction, grouping similar songs with clustering, training classification models, and finally experimenting with enhanced features.

When we first looked at our data, we found a few issues that needed fixing. Some songs had missing information - the duration field used "-1" to mark missing data, and some tempo entries just had "?" instead of actual values. Rather than throwing out entire genres, we filled in these gaps with the median value for each genre, which kept our dataset balanced. We also removed five rows that had other missing information. We then transformed our categorical data into a format our models could use. Musical keys (like C major or A minor) were converted to one-hot encoding, as was the mode (major/minor). We turned genre names into simple numbers from 0-9. For our testing strategy, we carefully split the data so that each genre had exactly 4,500 songs for training and 500 for testing. Finally, we standardized the numerical features using only information from the training set to avoid any "peeking" at the test data.

To better understand our high-dimensional data, we tried visualizing it in 2D using several techniques. PCA gave us a rough separation between Classical and Hip-Hop music, but most other genres were jumbled together. With t-SNE, Classical music formed a tight group, but everything else blended into one big blob. UMAP did the best job at naturally separating Classical and Hip-Hop, while Rock and Electronic music spread out more widely. Not surprisingly, supervised LDA created the clearest separation between genres.

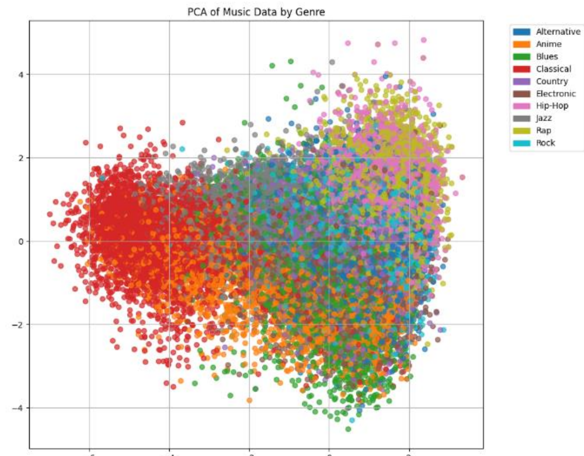
Encouraged by what we saw, we used K-Means clustering to see if songs would naturally group together by genre. The silhouette score peaked at k=2 clusters - basically just separating Classical music from everything else. When we looked at how genres distributed across 7 clusters, we found that Classical music was very pure (over 83% in a single cluster), and Hip-Hop and Rap also formed reasonably distinct groups at cluster 2 (shown on the P(cluster | true genre) when k = 7 graph). But other genres remained pretty mixed up. This told us something interesting: unsupervised clustering can capture some genre distinctions (especially for very distinctive genres like Classical), but we'd need supervised classification to really nail the boundaries between most genres.

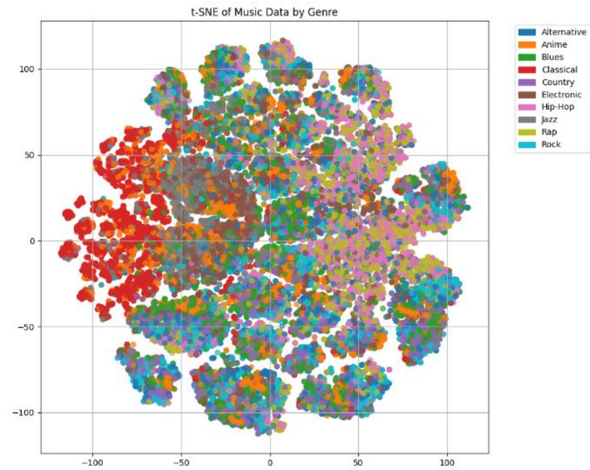
Now came the real test - could we train models to accurately predict genres? First, we tried seven different machine learning algorithms on the original training dataset: Logistic Regression, Decision Tree, Random Forest, linear SVM, k-Nearest Neighbors, XGBoost, and a neural network (MLP). Eventually, I chose to model on neural network with PCA dataset because it has the highest AUC of all 7 models. Linear SVM and XGBoost weren't far behind (AUC around **0.923-0.927**). Our neural network (MLP) working **with PCA-reduced features** gave us the best test results: a **macro-AUC of 0.929** with **57.2% accuracy**. We also looked at how performance improved with a learning curve that shows how model reacts to more training data. Performance rose from an AUC of 0.903 with just 5,000 samples to 0.926 with 25,000 samples, then basically flattened out. As a final experiment, we **added clustering information** as extra features on original dataset as comparison. Simpler models like Logistic Regression and Decision Trees got slightly better results with these new features. But our more powerful models barely changed or even performed a tiny bit worse. This told us something valuable: cluster information helps simple models that can't capture complex relationships on their own but doesn't add much for sophisticated models.

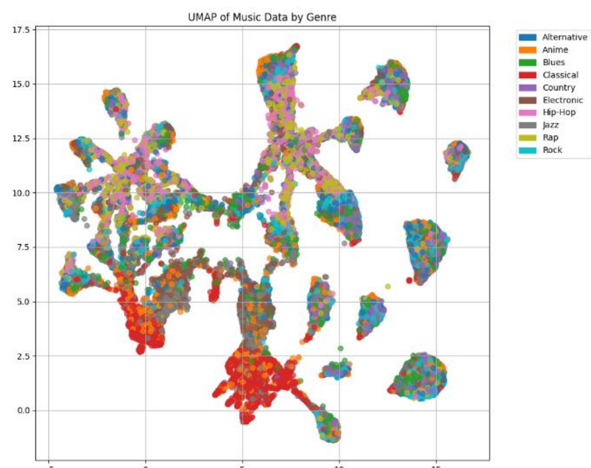
Our journey through the data revealed some fascinating insights about music genres. Classical and Hip-Hop seem to have truly distinctive audio signatures that machines can identify fairly easily. Other genres have much more overlap in their audio characteristics, making them harder to tell apart using just these Spotify features. To get better results in the future, we'd need either richer features (perhaps incorporating lyrics, artist information, or cultural context) or more sophisticated model architectures. This project showed us both the power and limitations of using audio features alone to classify music - some genre boundaries are more about cultural context and history than purely acoustic differences!

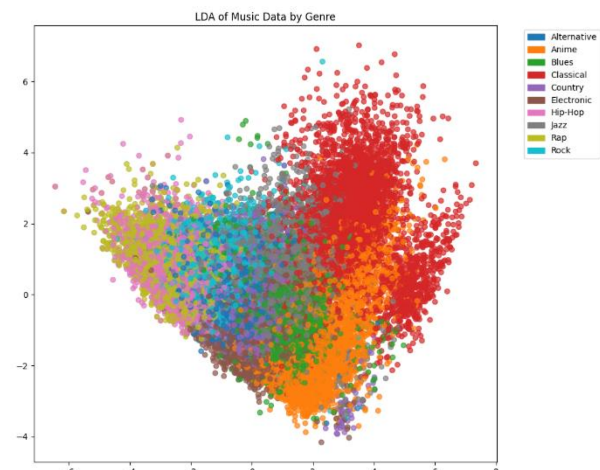
Final Macro-AUC: 0.928 Overall Accuracy: 0.572

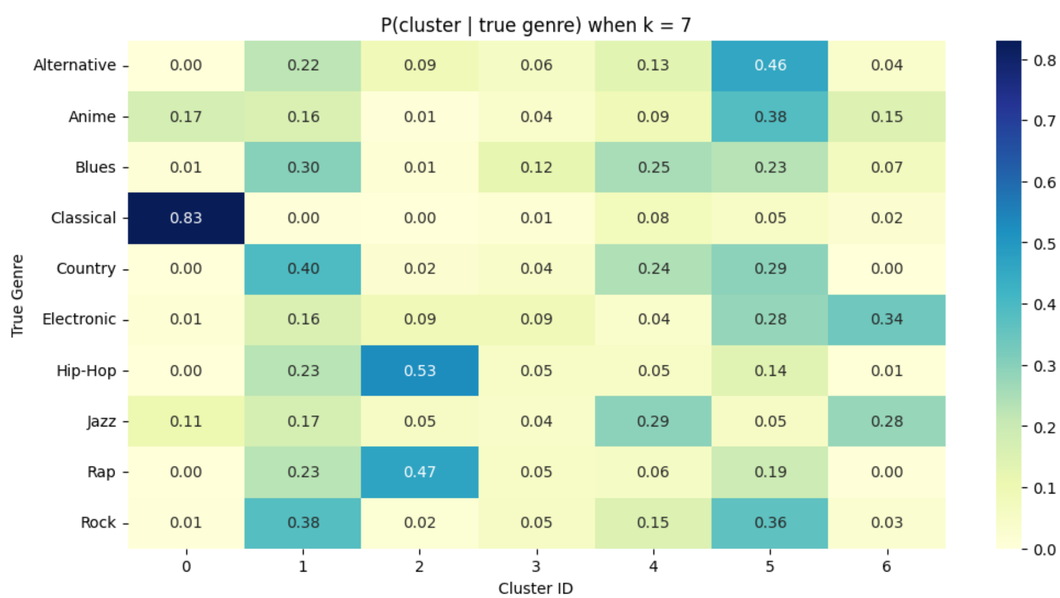
Plots:











Extra Credit:

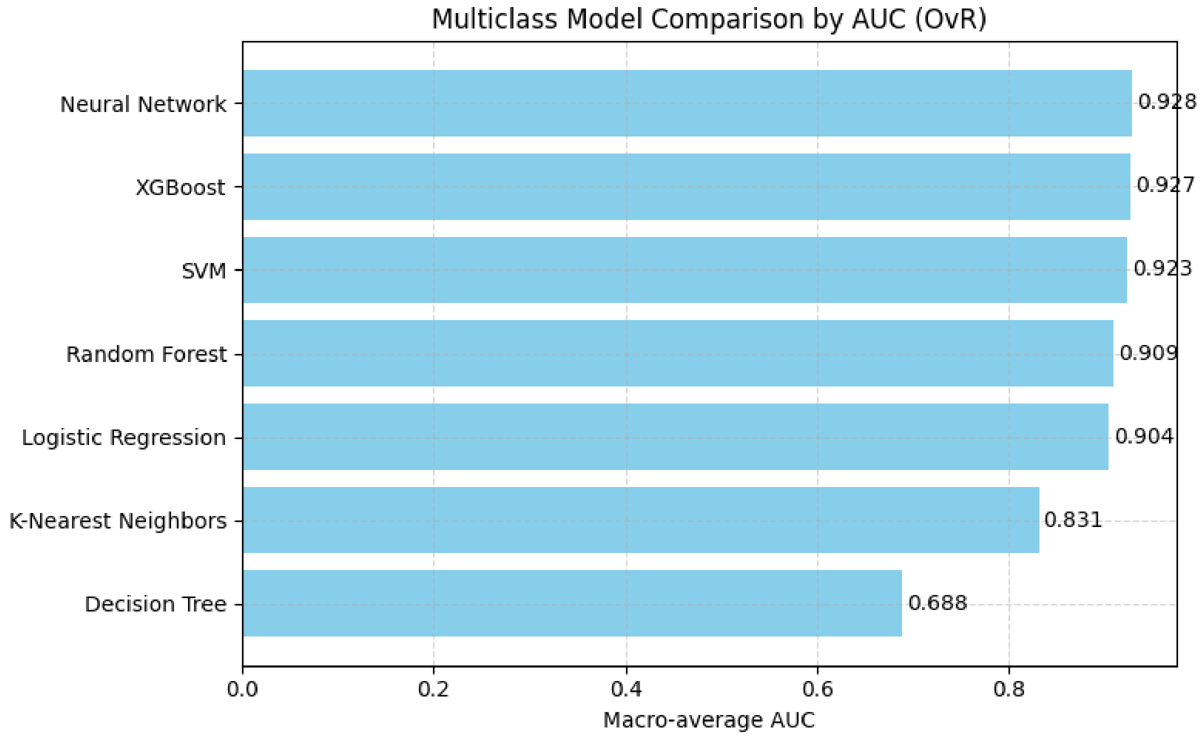


Figure 6: Model Comparison on original training set

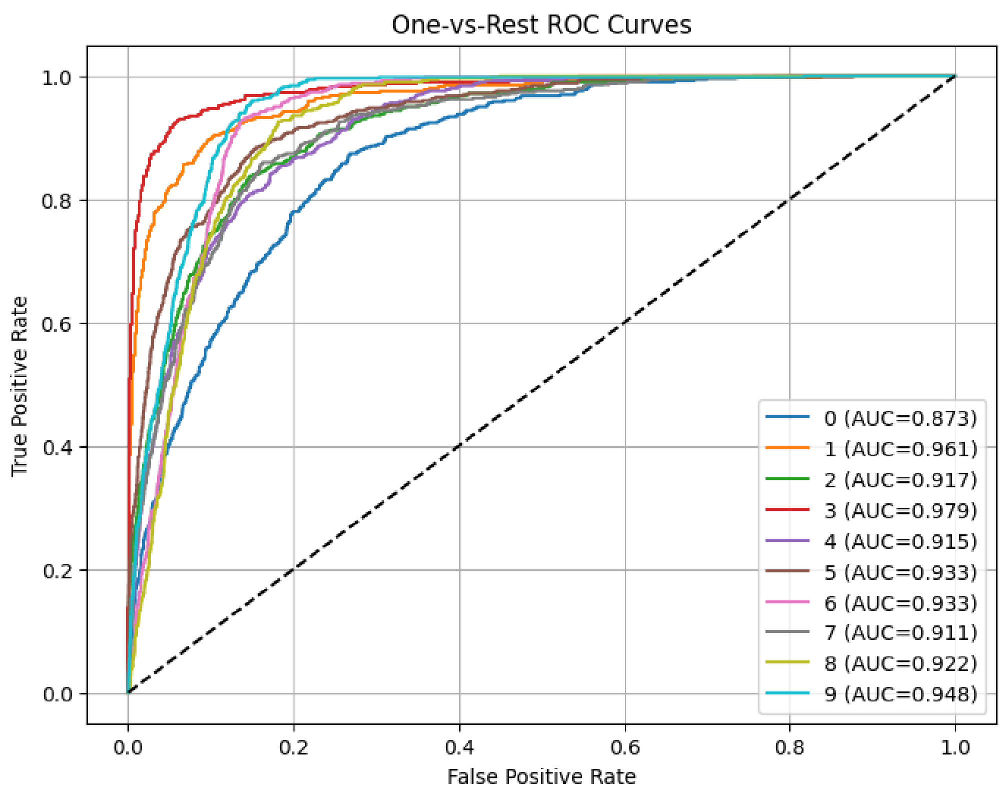


Figure 7: Neural Network AUC-ROC plot