

Music Genre Classification using the GTZAN dataset

Applied Physics 157 Final Project

GTZAN dataset

Contains:

- 1000 audio files classified into 10 music genres (100 files each of 30s)
- Visualizations of .wav files via spectrogram
- 2 CSV files with features from audio files

Genres included:

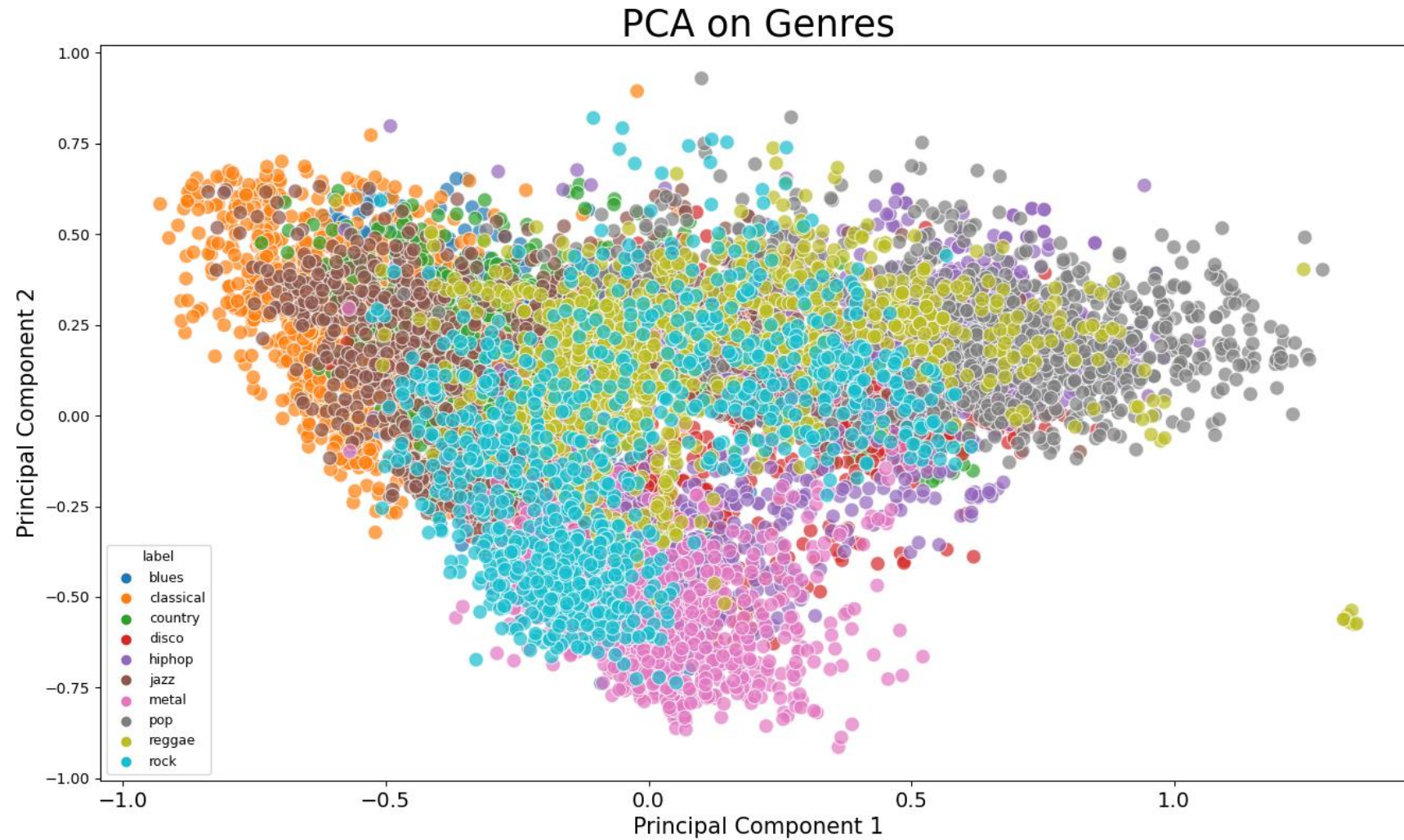
blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae, rock

GTZAN Features

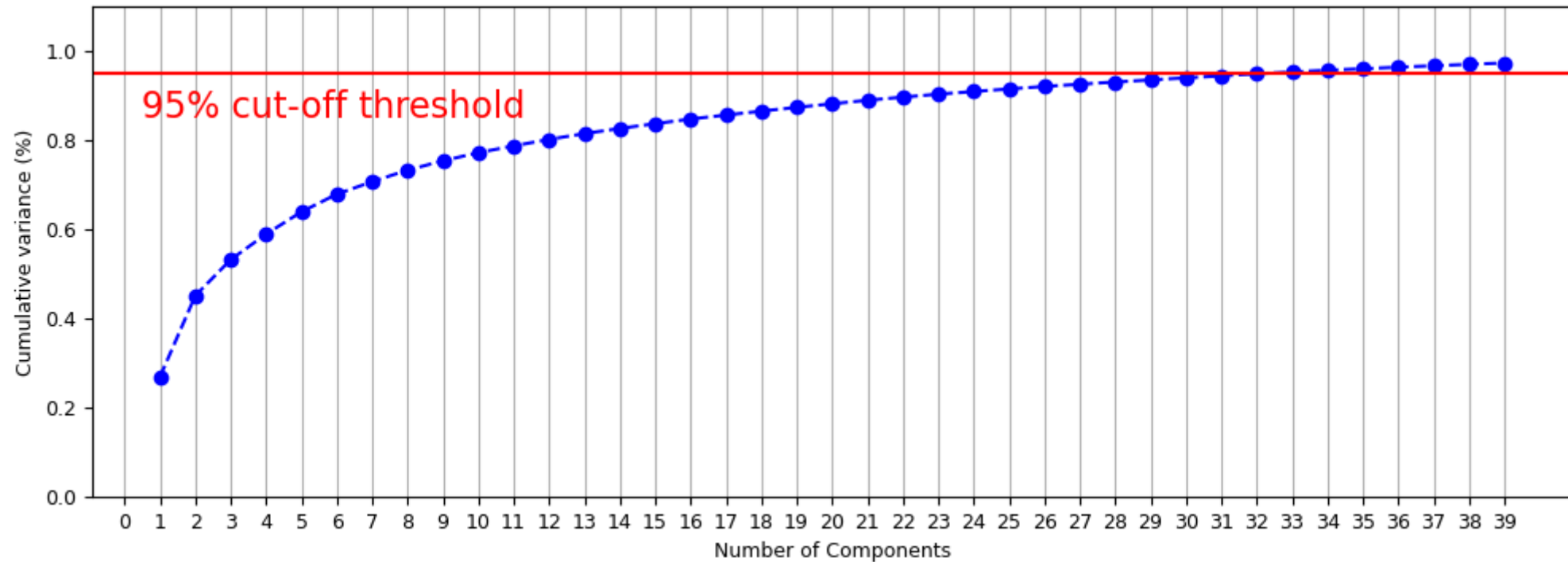
The GTZAN dataset contains 60 features including filename. Some pertinent ones are,

- **Tempo** – BPM
- **Chroma STFT** – Measure of different pitch intensities in the track
- **Spectral bandwidth** – Characterizes the frequency range of the track
- **MFCC** – Used for characterizing audio timbre and speech clarity
- **Perceptual sharpness** – Measure of sharpness and brightness of sound

Feature Correlations

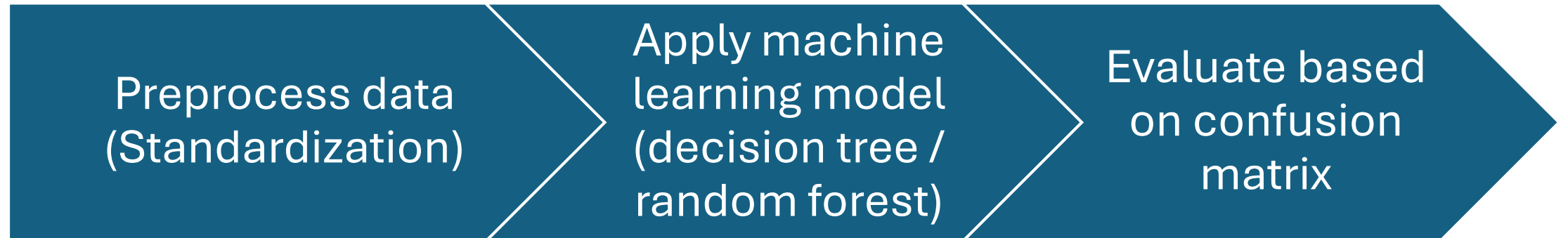


Feature Correlations

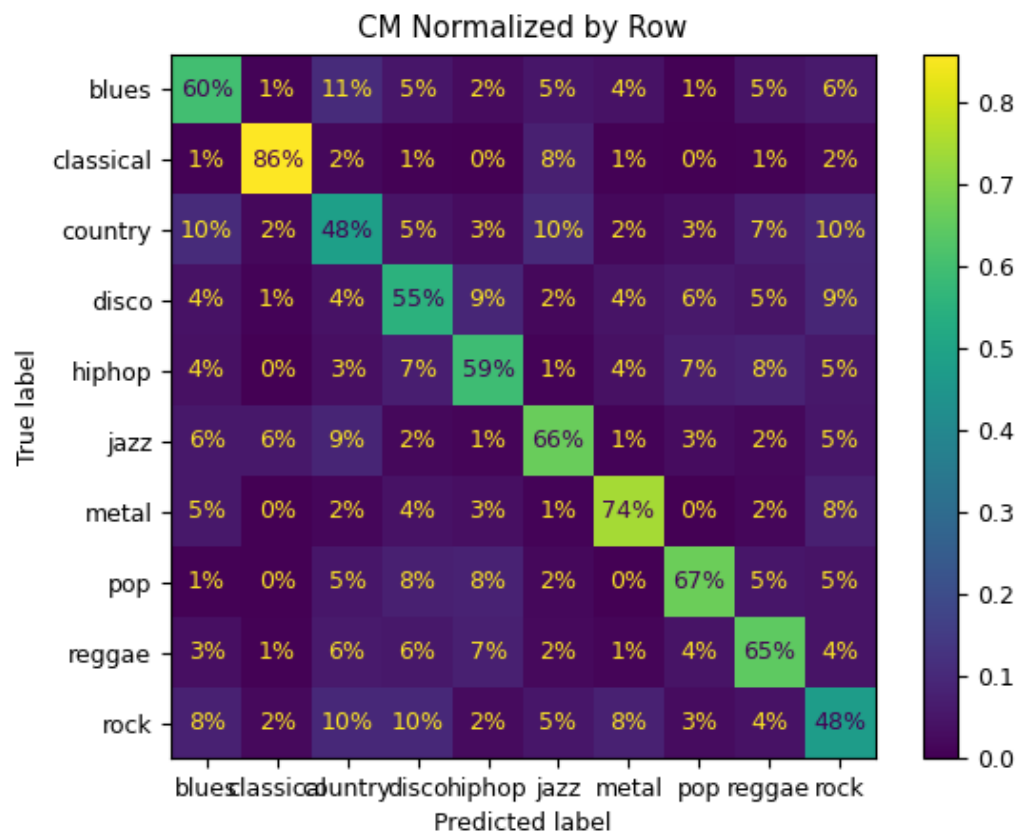


Number of components needed to explain
variance

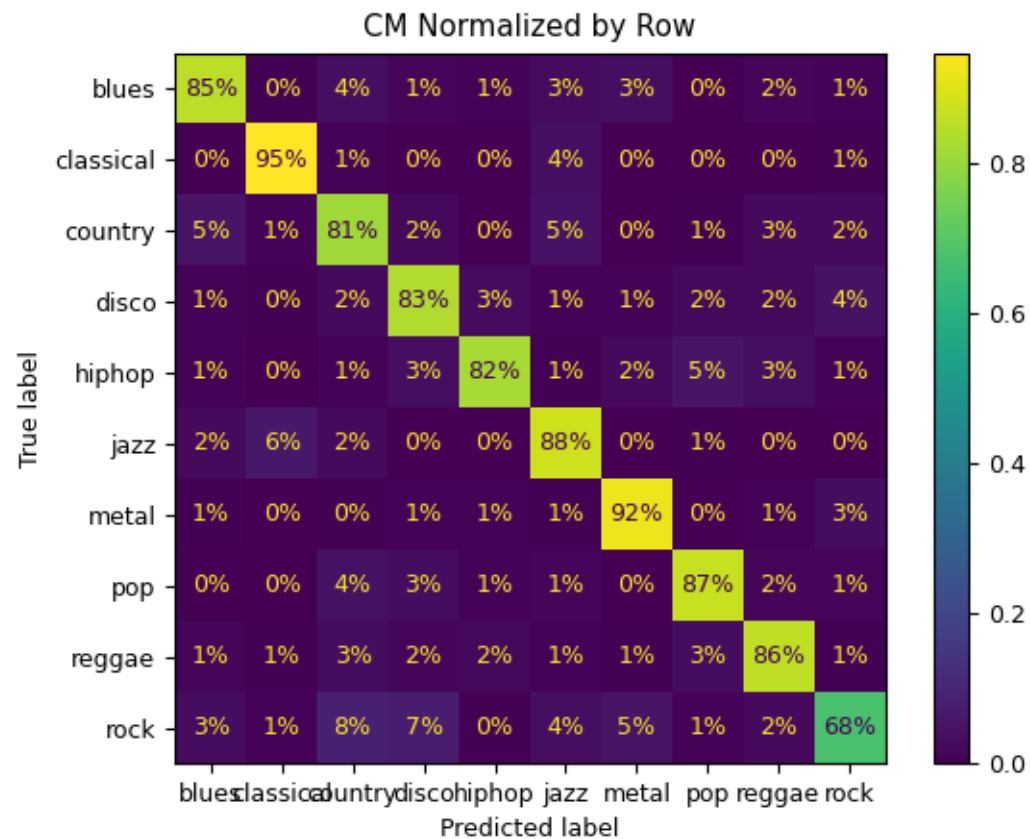
Methodology



Results (3sec audio files)

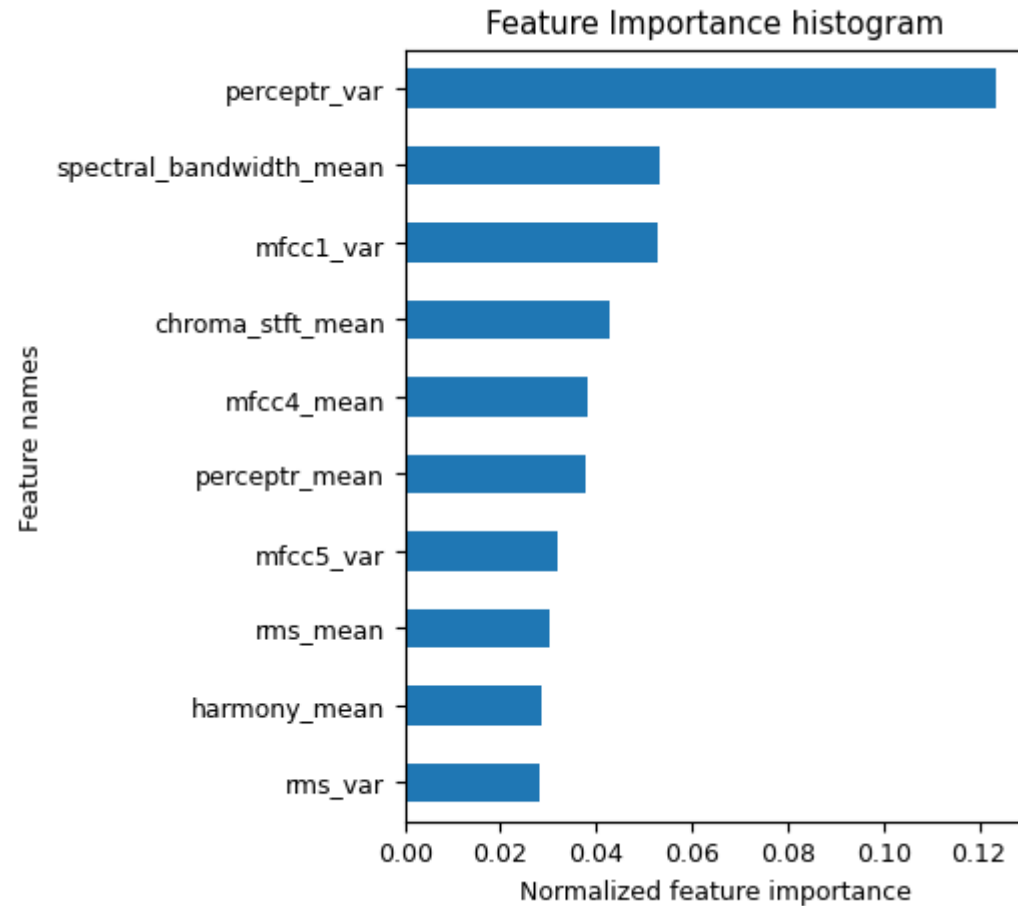


Confusion matrix for the decision tree classifier

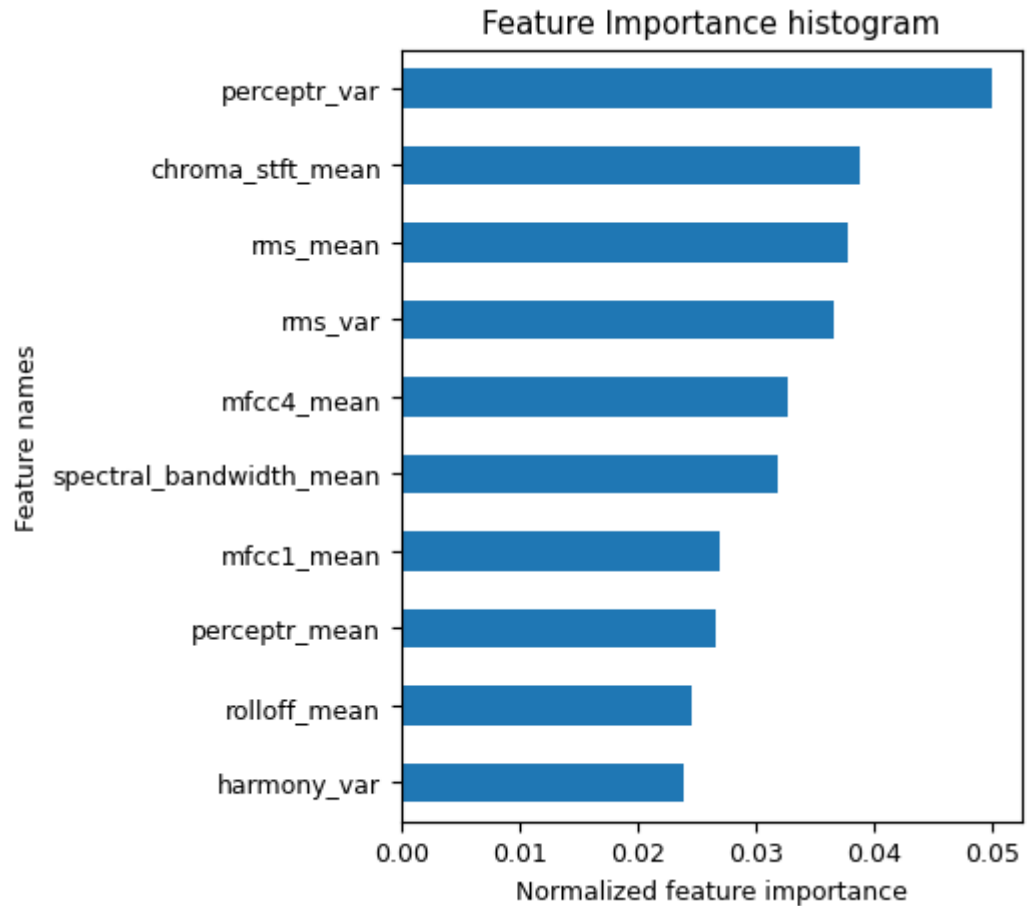


Confusion matrix for the random forest classifier

Results (3s audio files)

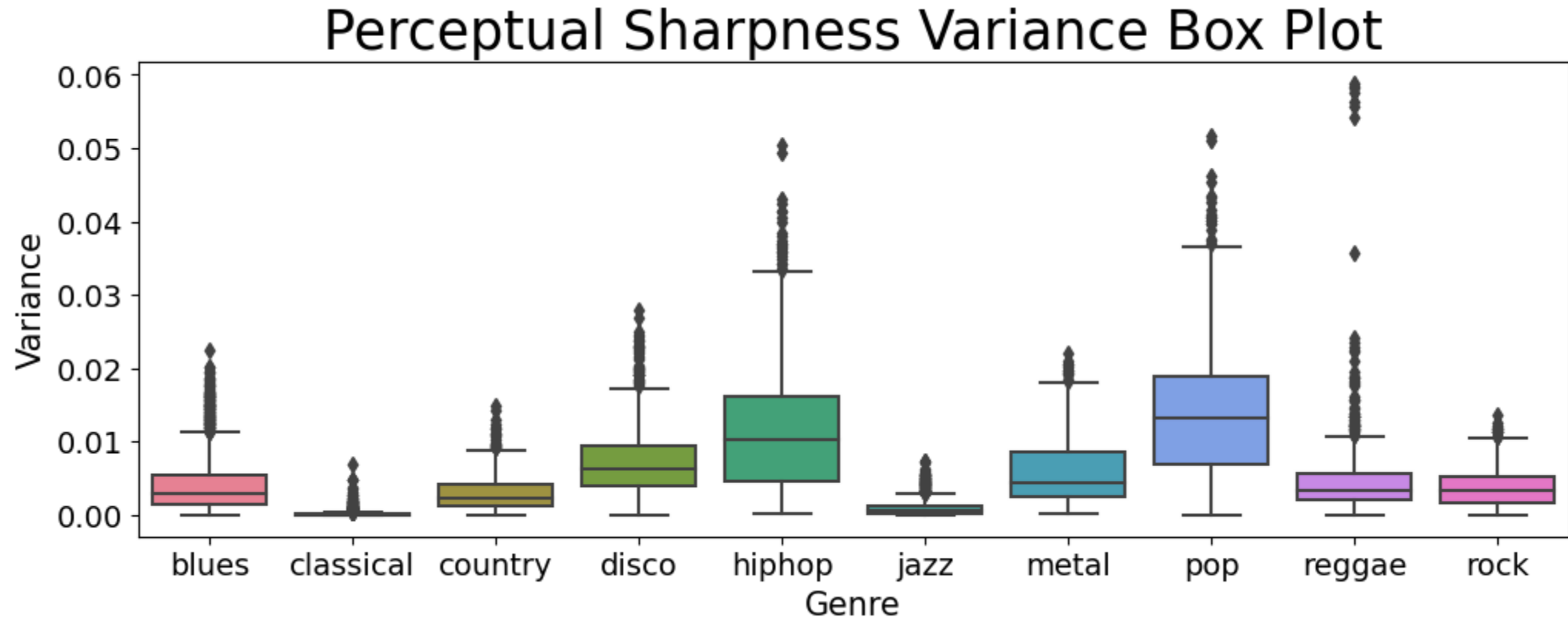


Feature importance plot for the decision tree classifier



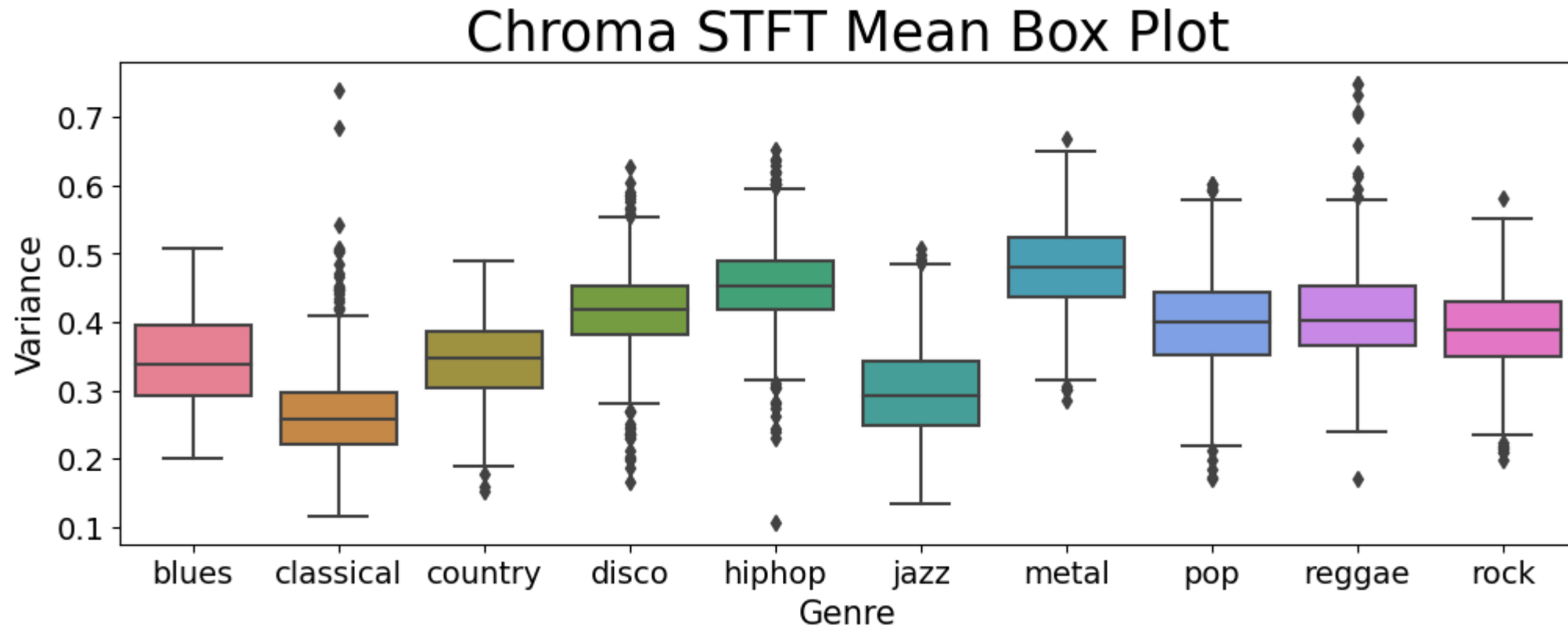
Feature importance plot for the random forest classifier

Results (3s audio files)



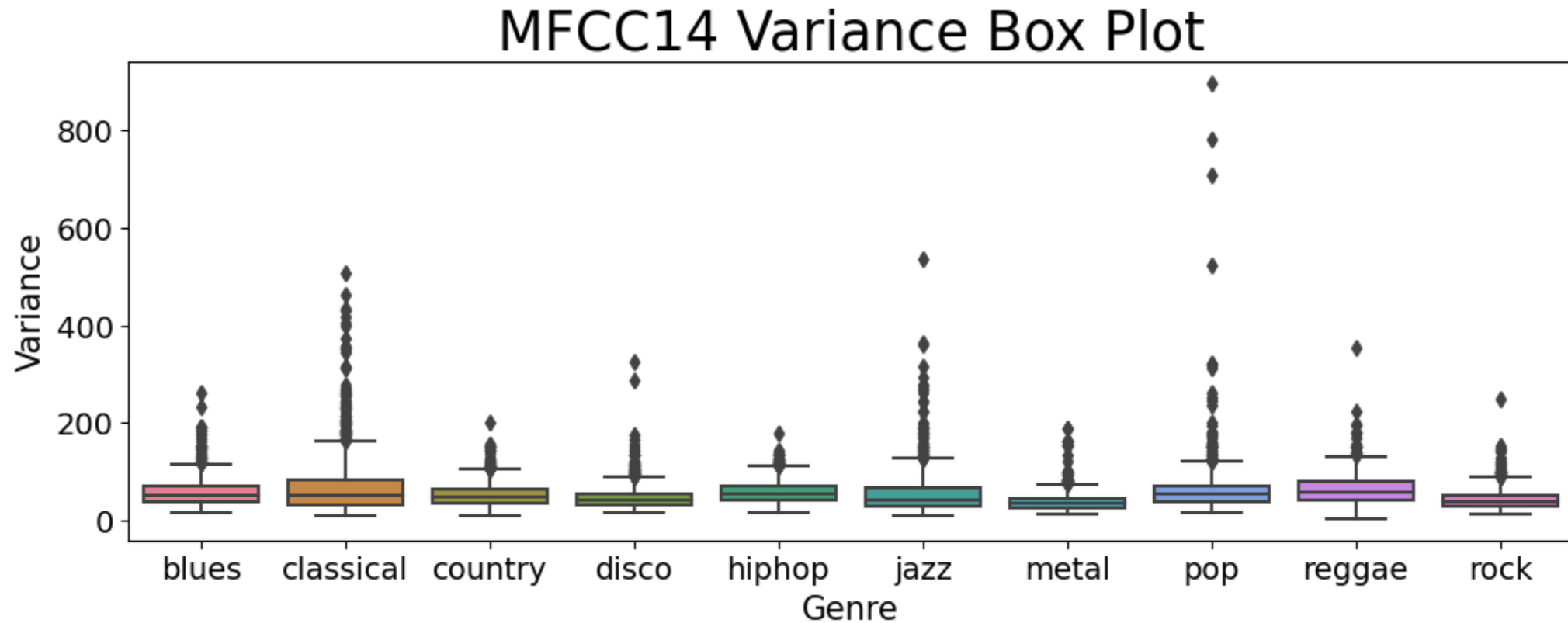
Highest feature importance for both decision tree and random forest classifier.

Results (3s audio files)



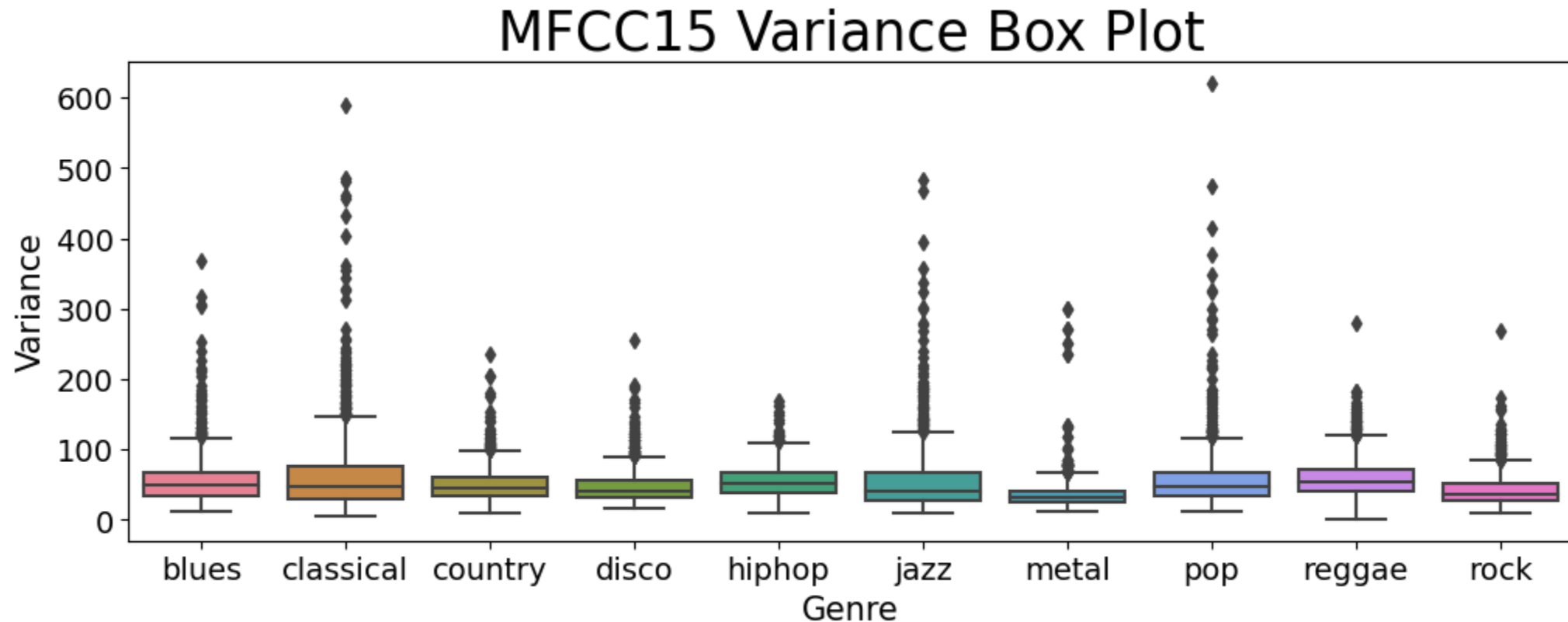
2nd highest feature importance for the random forest classifier.

Results (3s audio files)



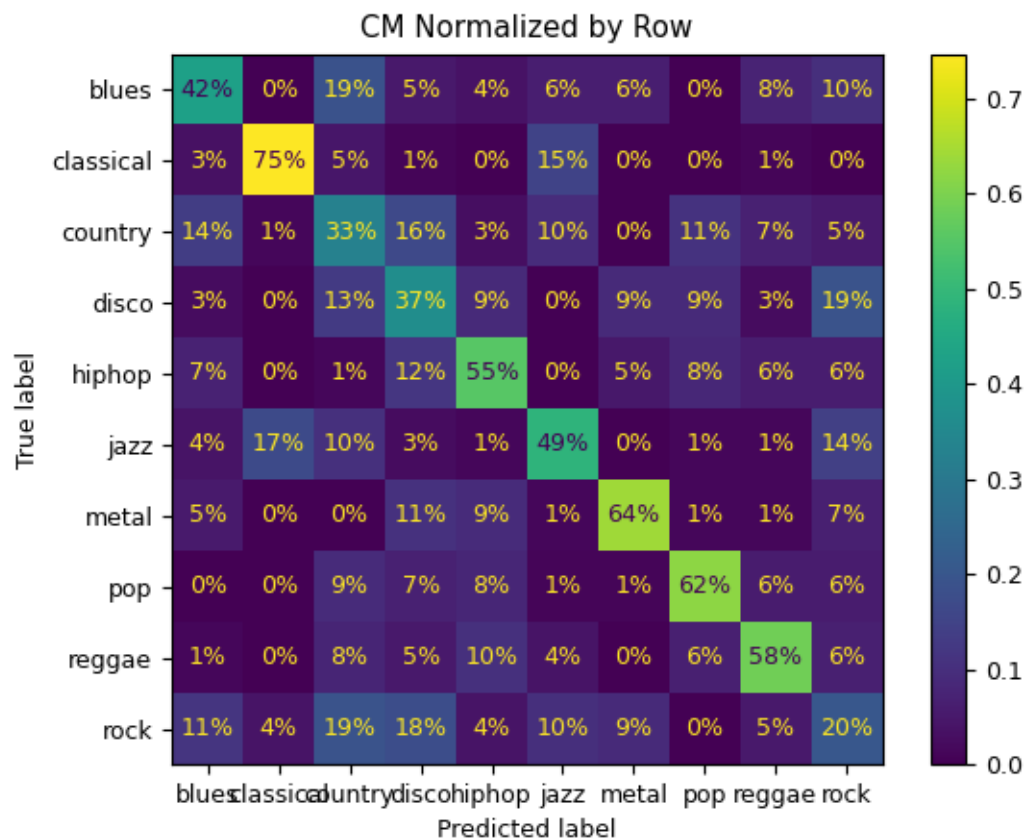
Lowest feature importance for the decision tree classifier.

Results (3s audio files)

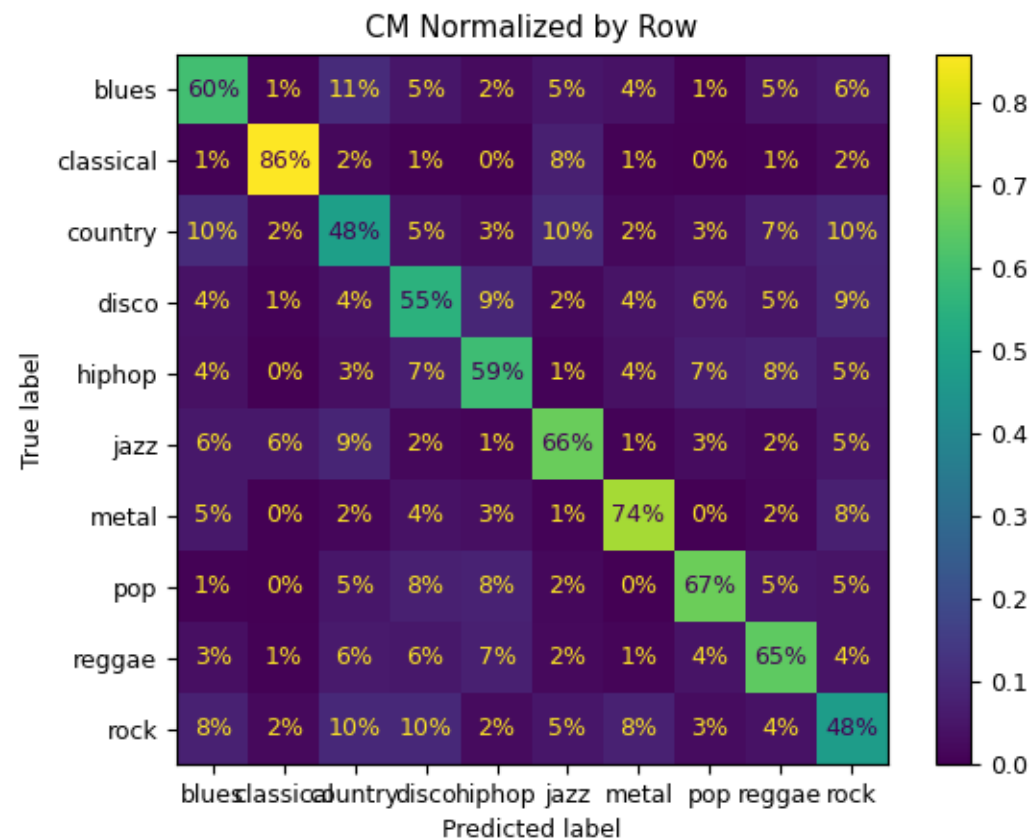


Lowest feature importance for the random forest classifier.

Versus 30s audio files

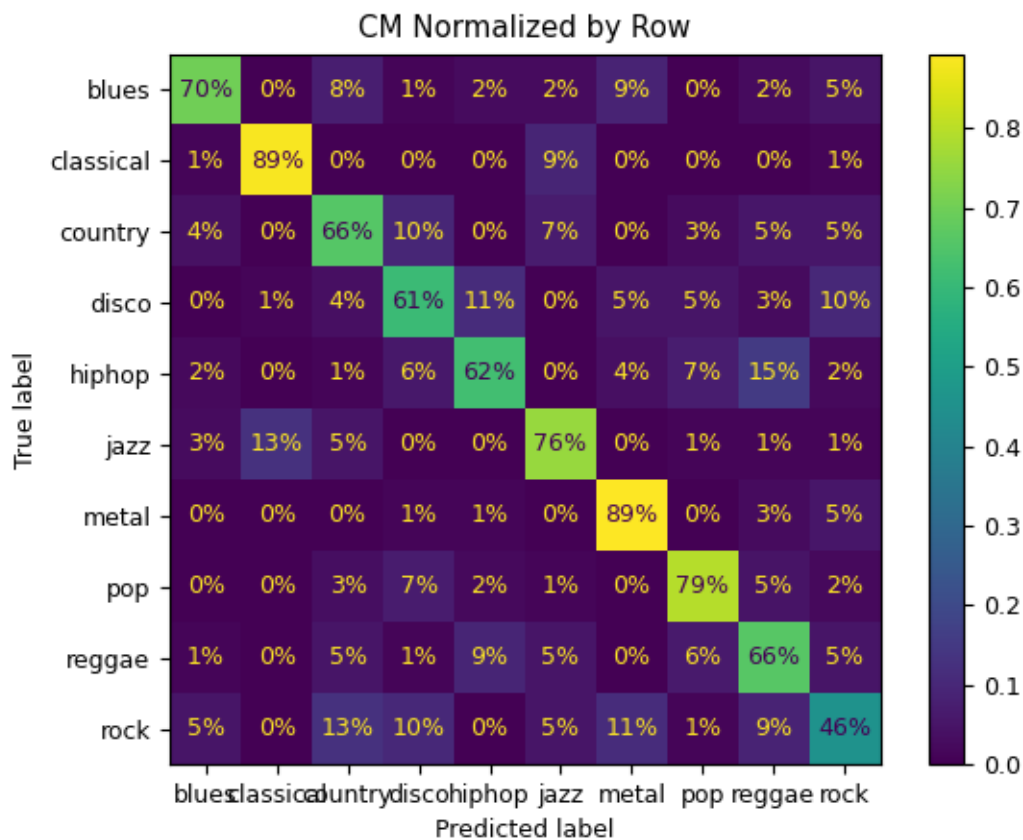


Confusion matrix for the decision tree classifier
(30s, 100 instances per genre)

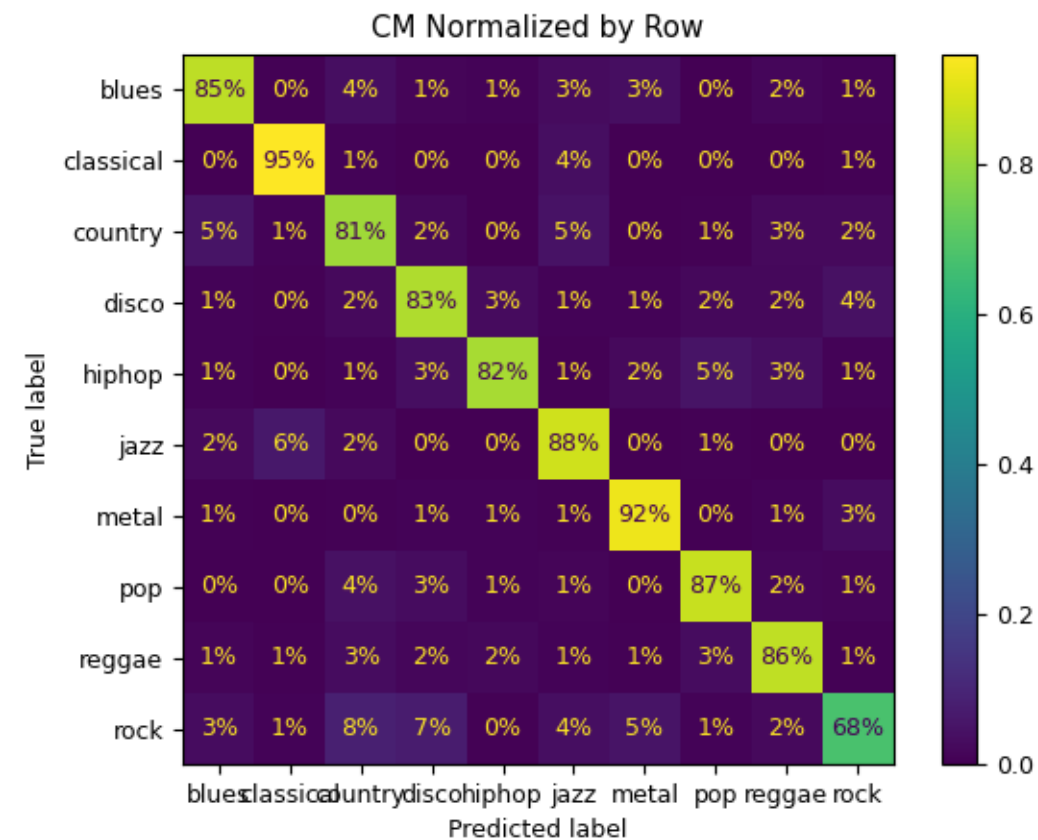


Confusion matrix for the decision tree classifier
(3s, 1000 instances per genre)

Versus 30s audio files



Confusion matrix for the random forest classifier
(30s, 100 instances per genre)



Confusion matrix for the random forest classifier
(3s, 1000 instances per genre)

Conclusions and Recommendations

- Principal component analysis can be used to visualize correlation between multi-dimensional instances.
- The random forest model classifies genres with at least 68% at the lowest and 95% at the highest.
- Hyperparameter testing can be done during modeling.
- Less important features may be removed or combined to something more relevant.

References

- <https://www.kaggle.com/datasets/andradaolteanu/gtzan-dataset-music-genre-classification/data>
- <https://www.kaggle.com/code/nkcong206/music-genre-classification-gtzan>

Thank you