Predicting Music Genres with Audio Features

Team Members: Kimberly Simmonds, Rohan Giri, Maxine Gardner, Carson Mullen, Akash Barathan

Presentation Outline

01
Introduction

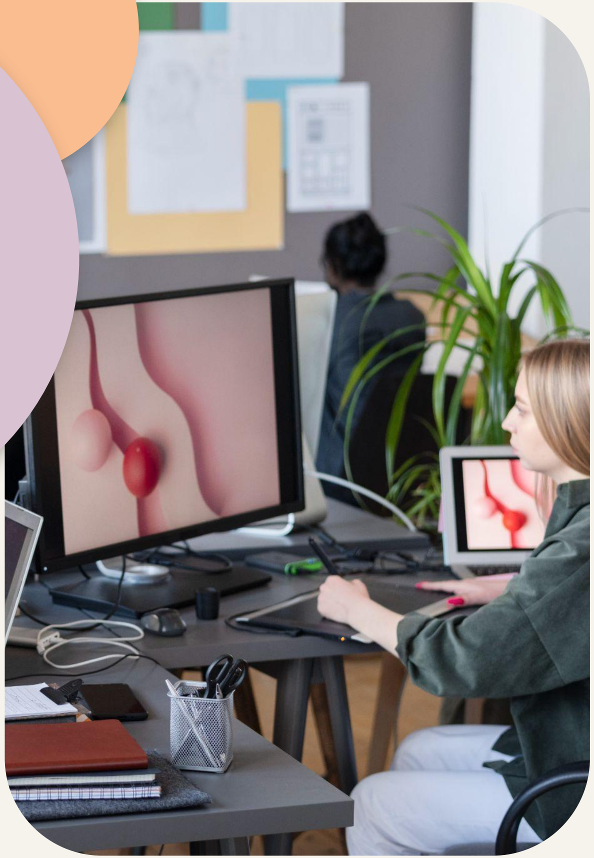
02
Data Overview

03
Model Evaluation

04
Application to Modern Dataset

05
Results

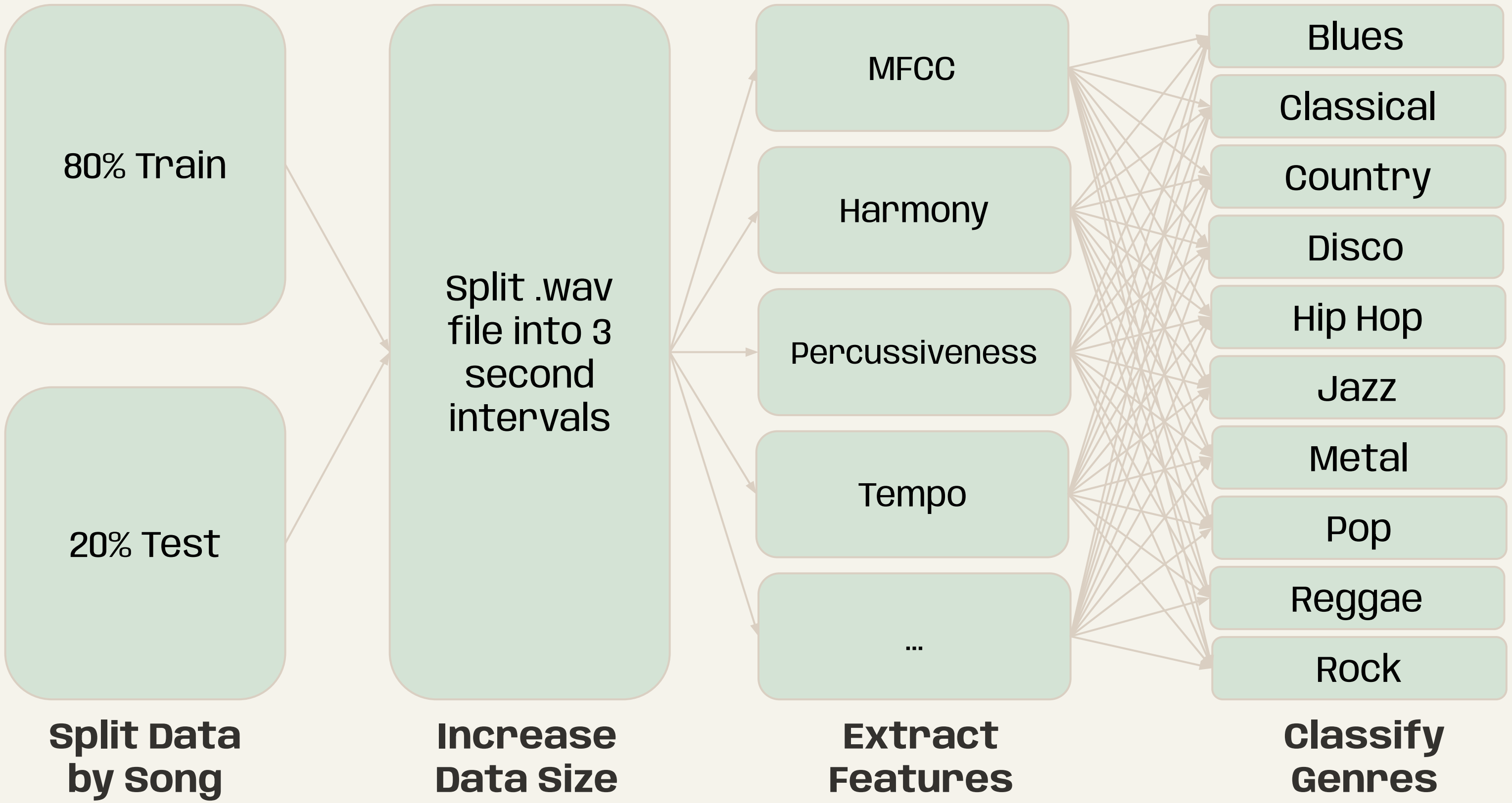
06
Future Scope & Conclusion



Introduction

Our project explores the intersection of music and artificial intelligence, classifying various music tracks into ten genres and using features such as tempo, harmony, and rhythm extracted using Librosa. Through combining machine learning algorithms with deep learning techniques, we aim to accurately categorize music tracks into their respective genres.

Workflow Diagram



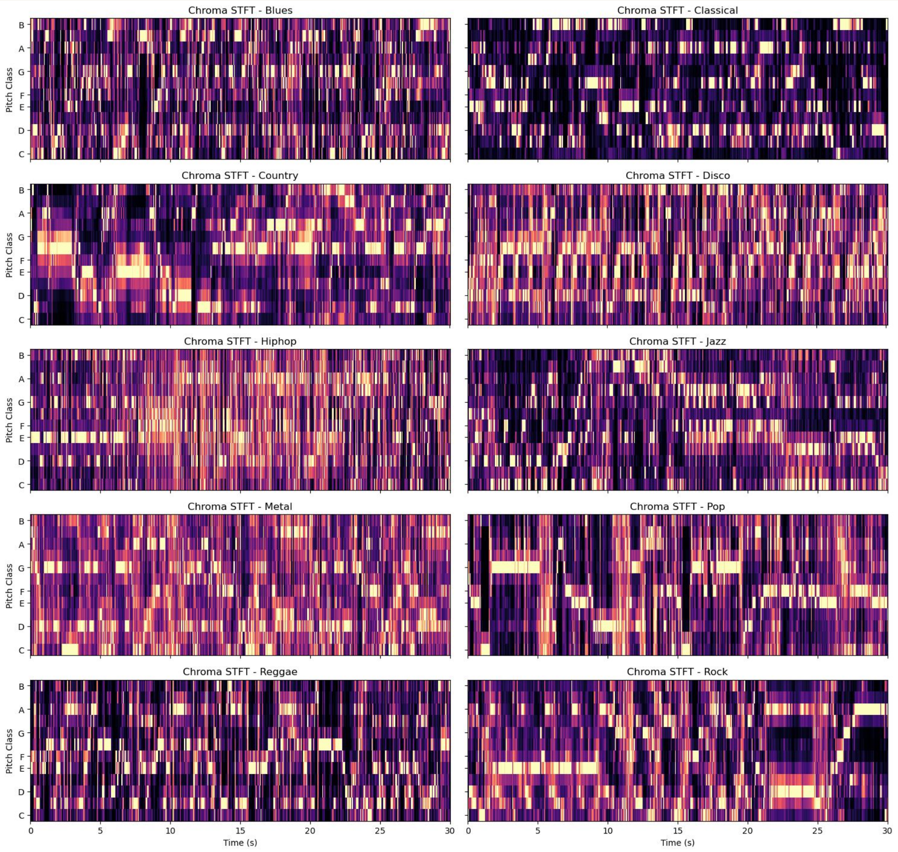
Data Overview

Data Pre-processing

Dataset	GTZAN Dataset - Kaggle 100 30-second audio clips from 10 different music genres up to 2001
Feature Engineering	187 musical features extracted from audio segments using Librosa
Youtube Data	Scraped 20 additional songs for each genre from Youtube starting 2002+

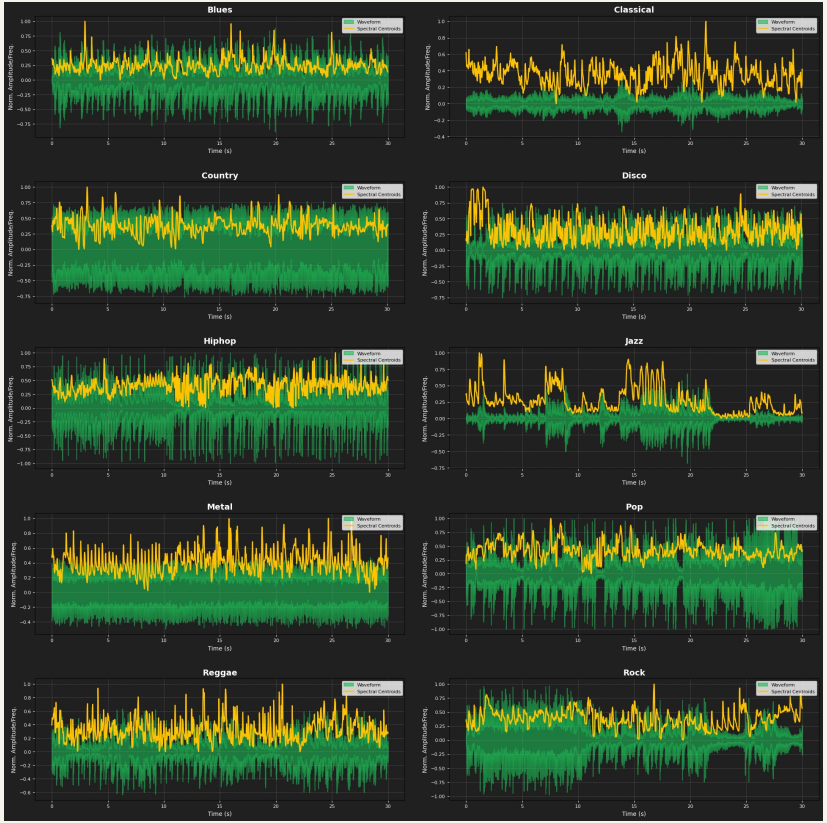
Feature Extraction: Chroma STFT

The Chroma STFT is a feature representation that uses a Short-Time Fourier Transform (STFT) to extract energy distribution over the 12 possible pitches.



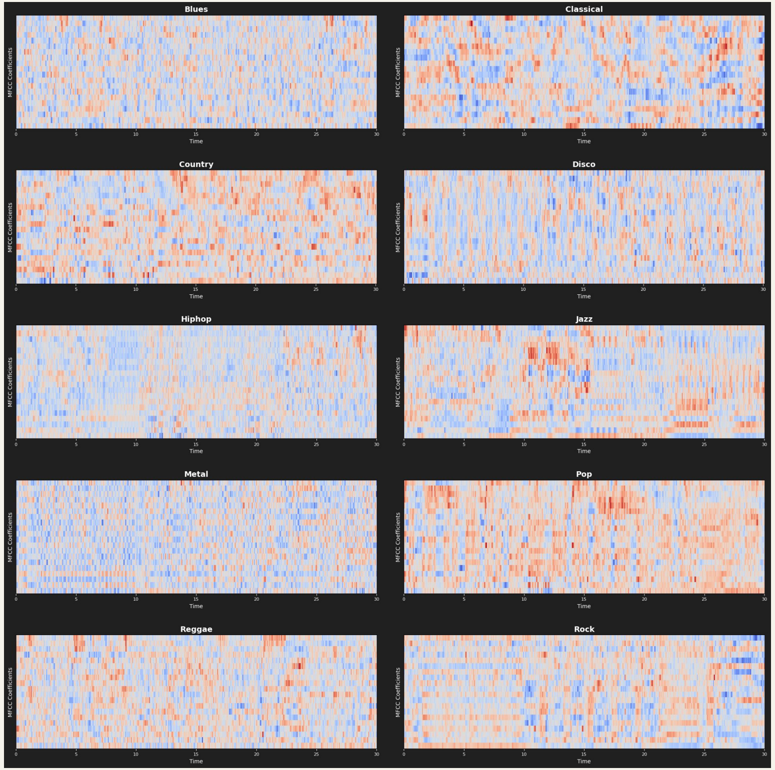
Feature Extraction: Spectral Centroid

The Spectral Centroid measures the “tonal character” or brightness in the sound over time. A higher spectral frequency represents a brighter sound, while a lower spectral frequency represents a darker sound.



Feature Extraction: MFCC

Mel Frequency Cepstral Coefficients (MFCCs) represents the Short-term power spectrum of the sound which can be used to describe the timbre of the sound. The Mel scale aligns with how humans perceive pitch and frequency. This makes MFCCs excellent at capturing perceptual characteristics of audio signals, which plays a critical role in distinguishing musical genres.



Feature Extraction: More

- Chroma_cqt
- Chroma_cens
- Tonnetz
- Rms
- Spectral_bandwidth
- Spectral_rolloff
- Zero_crossing_rate
- Harmony
- Percussive
- Tempo

Model Evaluation

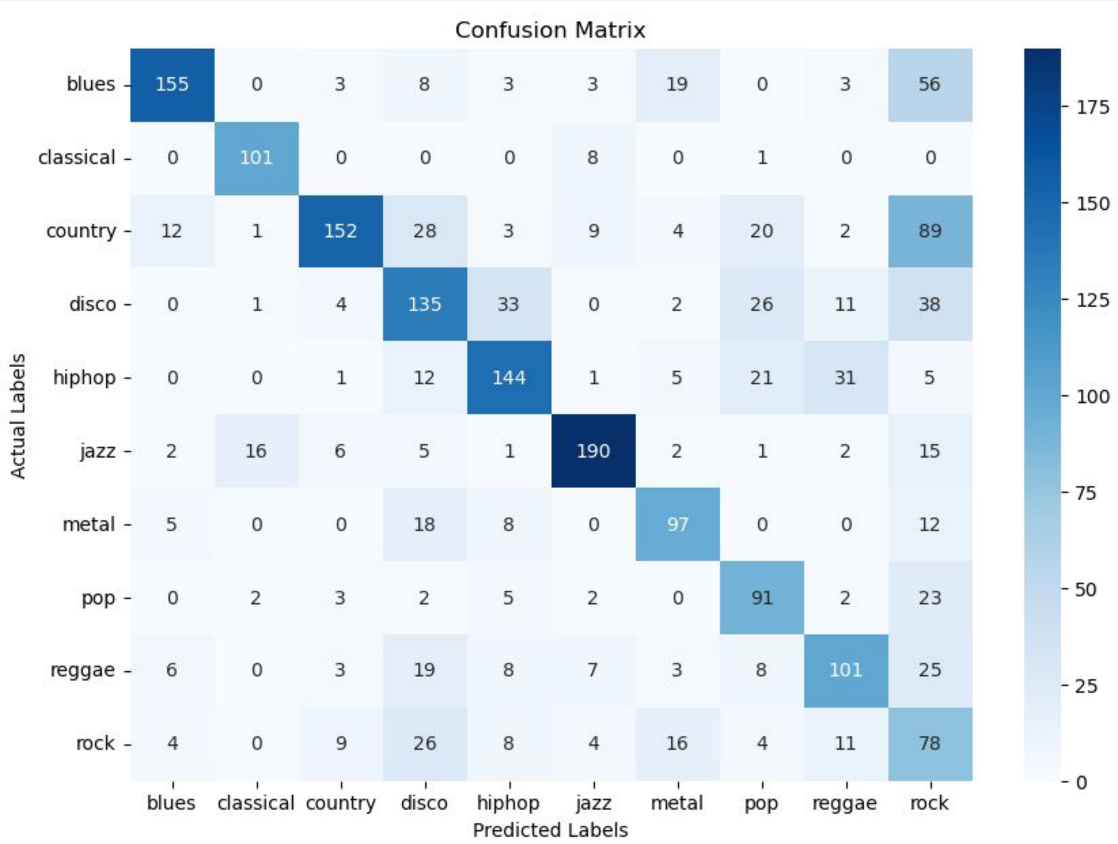
Baseline Models

Model	KNN (n=5)	Random Forest	Decision Tree
Accuracy	0.672	0.716	0.518
Precision	0.675	0.705	0.526
Recall	0.685	0.723	0.534
ROC-AUC	0.902	0.949	0.740

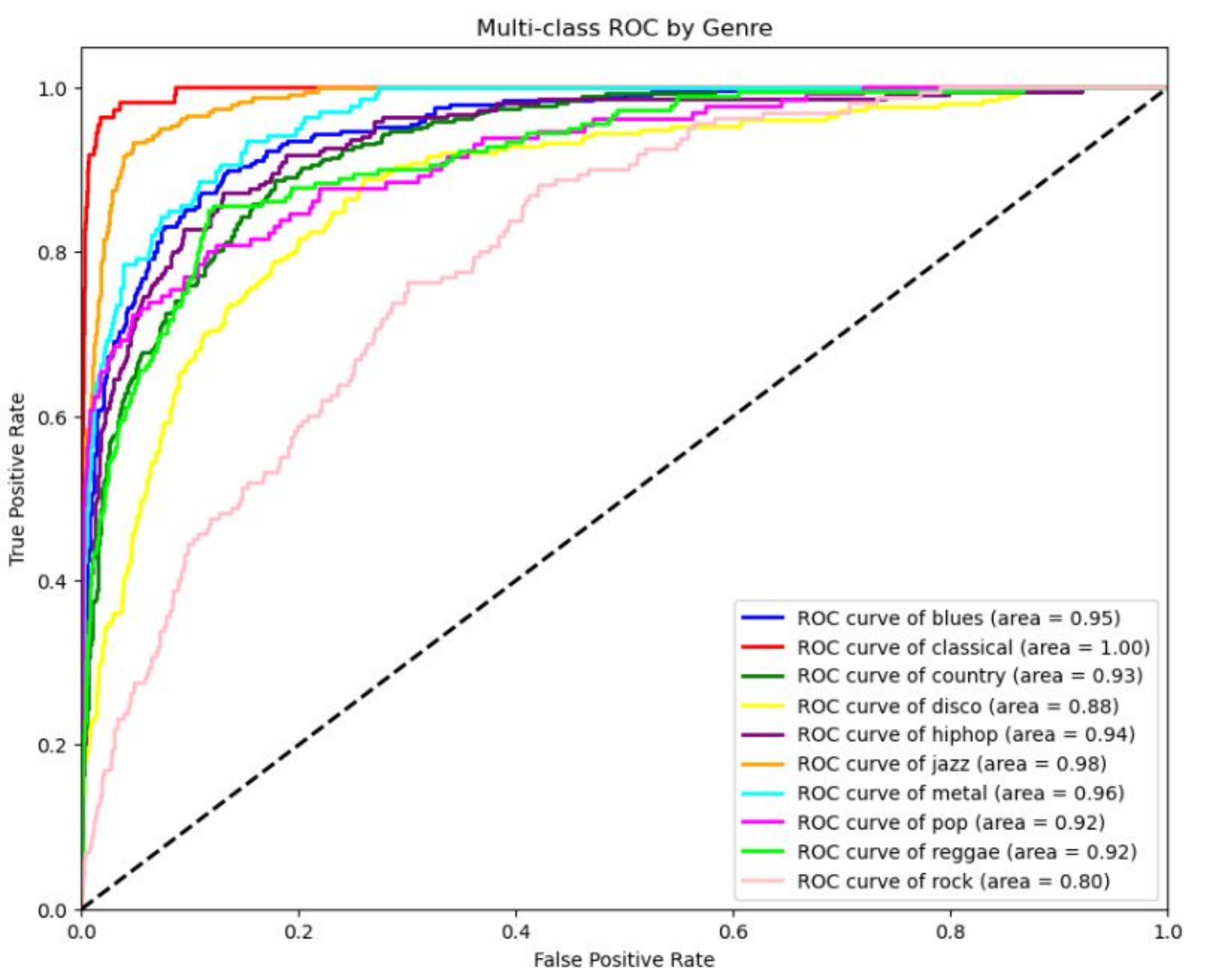
Convolutional Neural Network

Overall

Accuracy: 0.622
Precision: 0.660
Recall: 0.644
ROC-AUC: 0.93



Genre	ROC AUC
Blues	0.95
Classical	1.00
Country	0.93
Disco	0.88
Hiphop	0.94
Jazz	0.98
Metal	0.96
Pop	0.92
Reggae	0.92
Rock	0.80



Best Performing Model

XGBoost

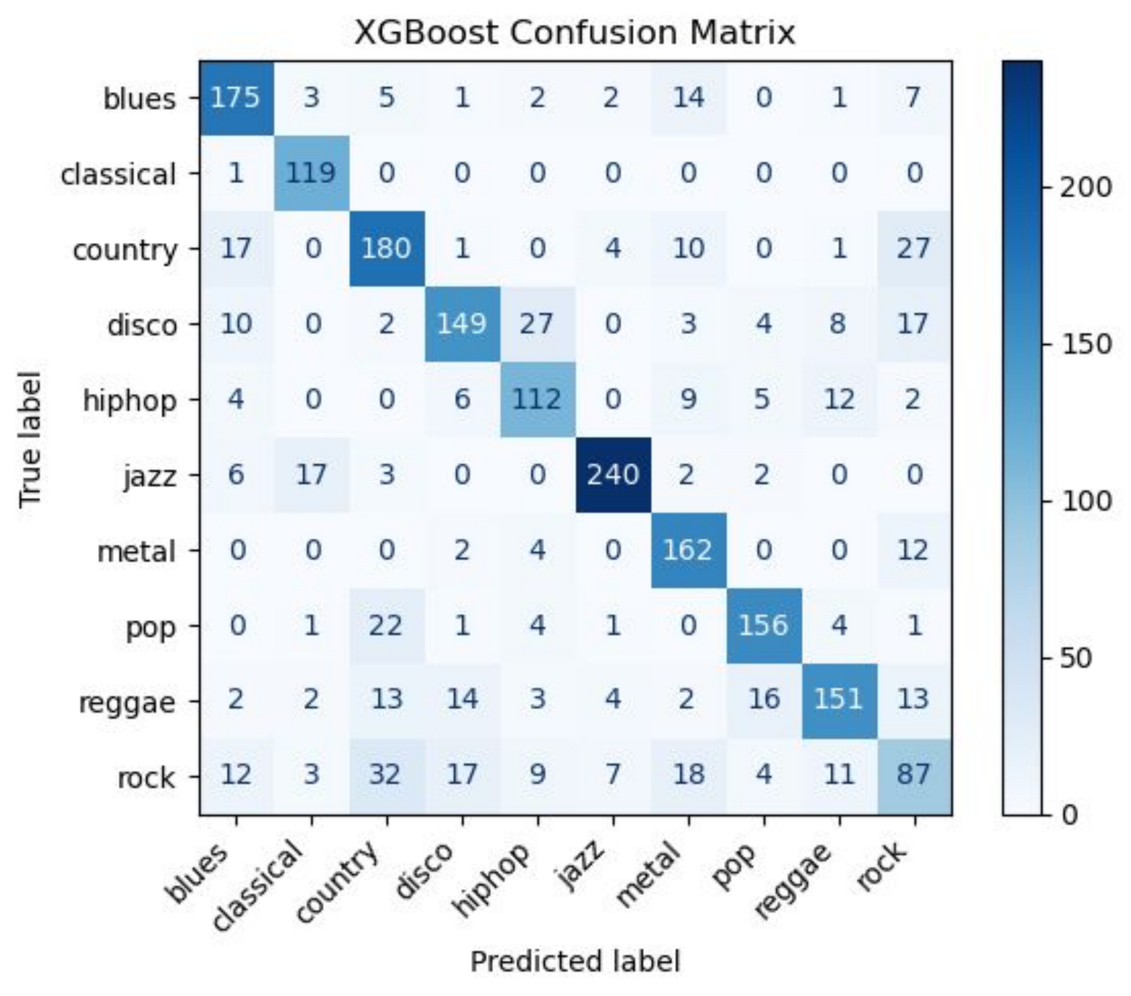
HyperOpt Tuned Model Parameters:

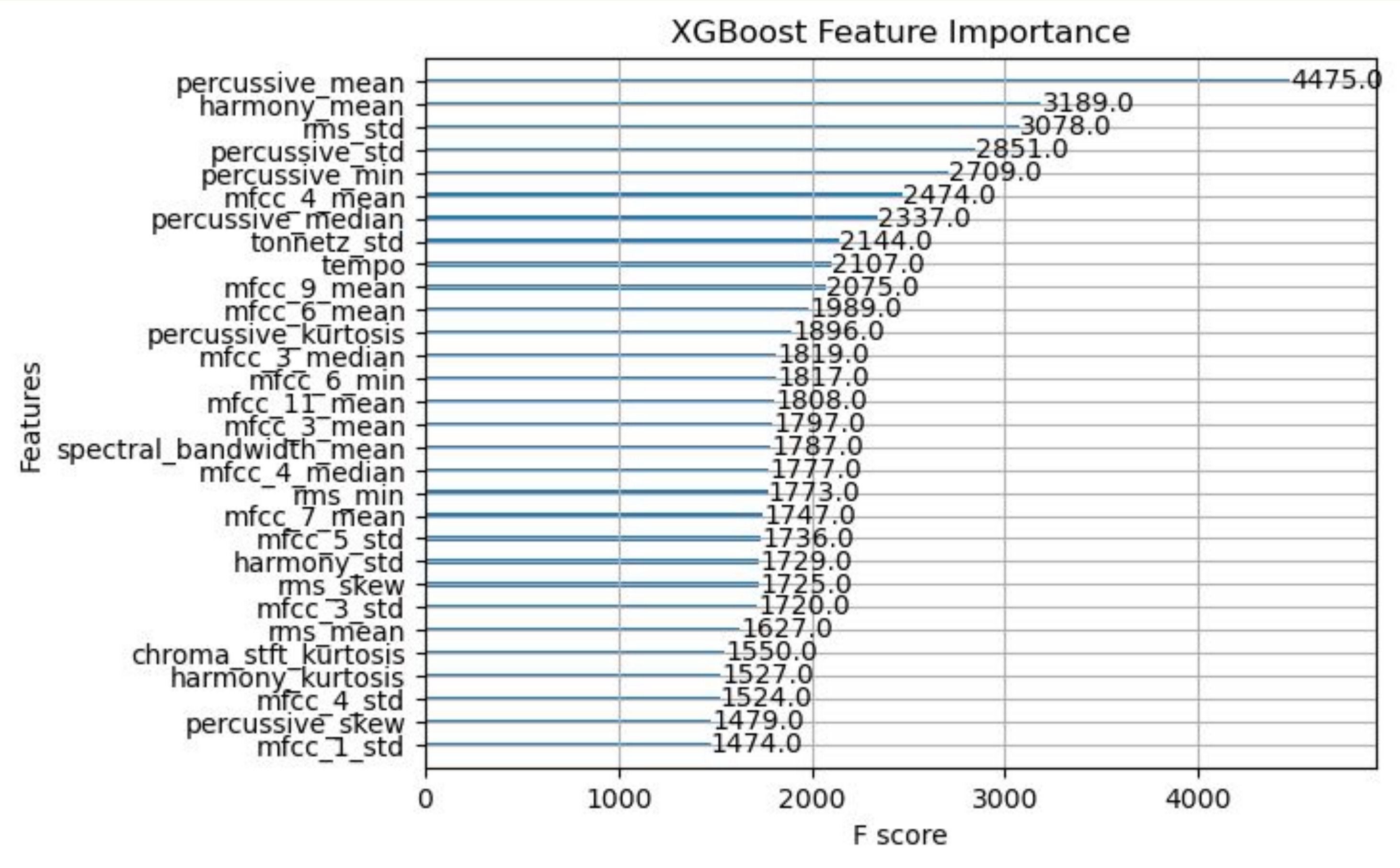
n_estimators: 1187
min_child_weight: 5
max_depth: 15
learning_rate: 0.048

Overall
Accuracy: 0.765
Precision: 0.760
Recall: 0.773
ROC-AUC: 0.965

Genre	Accuracy	Precision
Blues	0.77	0.83
Classical	0.82	0.99
Country	0.70	0.75
Disco	0.78	0.68
Hiphop	0.70	0.75
Jazz	0.93	0.89
Metal	0.74	0.90
Pop	0.83	0.82
Reggae	0.80	0.69
Rock	0.52	0.43

* Train accuracy/precision was 100%

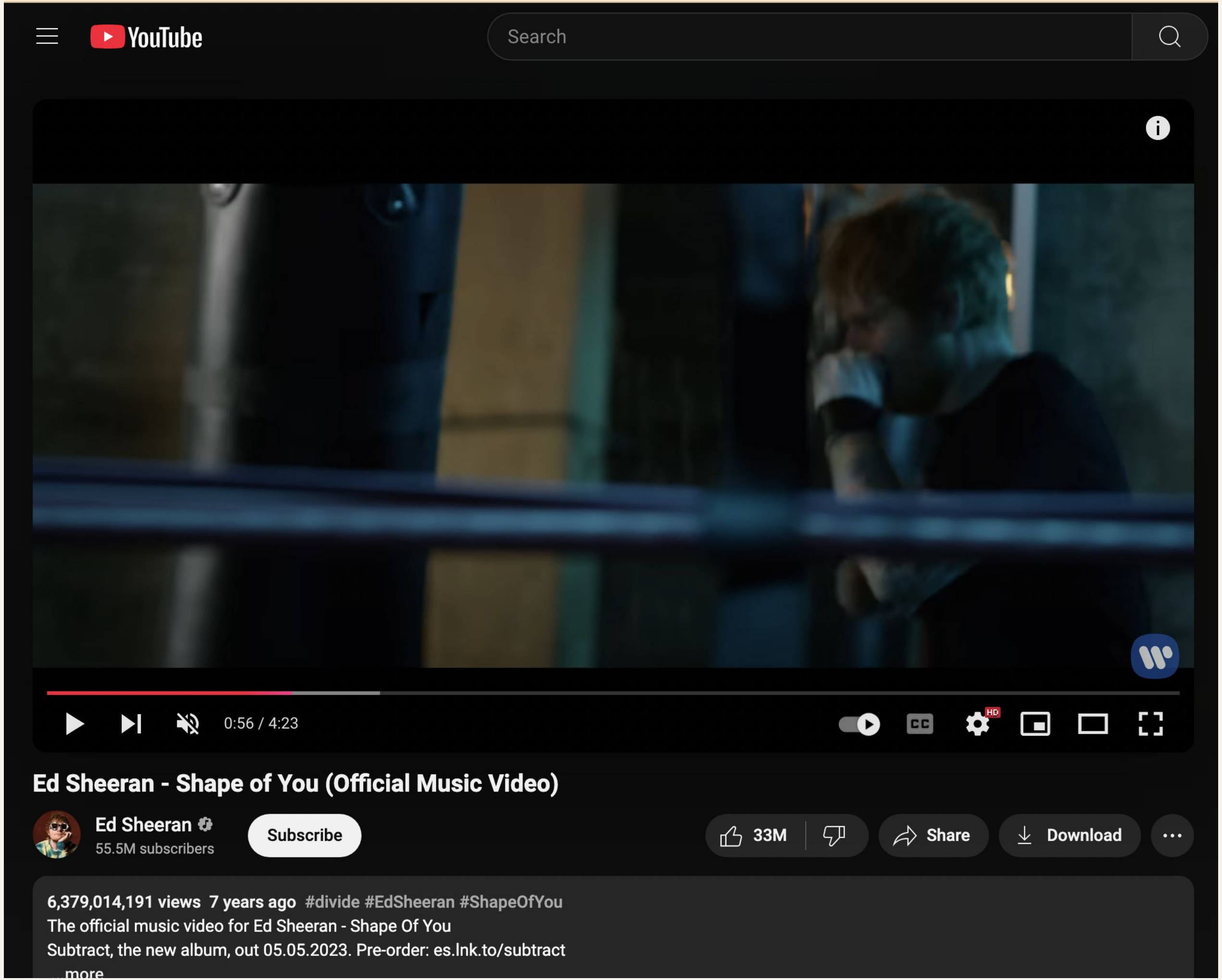

XGBoost Confusion Matrix
True label
Predicted label
blues
classical
country
disco
hiphop
jazz
metal
pop
reggae
rock
blues 175 3 5 1 2 2 14 0 1 7
classical 1 119 0 0 0 0 0 0 0 0
country 17 0 180 1 0 4 10 0 1 27
disco 10 0 2 149 27 0 3 4 8 17
hiphop 4 0 0 6 112 0 9 5 12 2
jazz 6 17 3 0 0 240 2 2 0 0
metal 0 0 0 2 4 0 162 0 0 12
pop 0 1 22 1 4 1 0 156 4 1
reggae 2 2 13 14 3 4 2 16 151 13
rock 12 3 32 17 9 7 18 4 11 87
200
150
100
50
0


XGBoost Feature Importance
Features
F score
percussive_mean
harmony_mean
rms_std
percussive_std
percussive_min
mfcc_4_mean
percussive_median
tonnetz_std
tempo
mfcc_9_mean
mfcc_6_mean
percussive_kurtosis
mfcc_3_median
mfcc_6_min
mfcc_11_mean
mfcc_3_mean
spectral_bandwidth_mean
mfcc_4_median
rms_min
mfcc_7_mean
mfcc_5_std
harmony_std
rms_skew
mfcc_3_std
rms_mean
chroma_stft_kurtosis
harmony_kurtosis
mfcc_4_std
percussive_skew
mfcc_1_std
4475.0
3189.0
3078.0
2851.0
2709.0
2474.0
2337.0
2144.0
2107.0
2075.0
1989.0
1896.0
1819.0
1817.0
1808.0
1797.0
1787.0
1777.0
1773.0
1747.0
1736.0
1729.0
1725.0
1720.0
1627.0
1550.0
1527.0
1524.0
1479.0
1474.0
0
1000
2000
3000
4000

Results

Models	Accuracy	Precision	Recall	ROC-AUC
KNN	0.672	0.675	0.685	0.902
Decision Tree	0.518	0.526	0.534	0.740
Random Forest	0.716	0.706	0.723	0.949
XGBOOST	0.765	0.760	0.773	0.965
CNN	0.622	0.660	0.644	0.93

What about newer music?


YouTube
Search
0:56 / 4:23
HD
Ed Sheeran - Shape of You (Official Music Video)
Ed Sheeran
55.5M subscribers
Subscribe
33M
Share
Download
6,379,014,191 views 7 years ago #divide #EdSheeran #ShapeOfYou
The official music video for Ed Sheeran - Shape Of You
Subtract, the new album, out 05.05.2023. Pre-order: es.lnk.to/subtract
...more

Apply Best Model to Modern Data

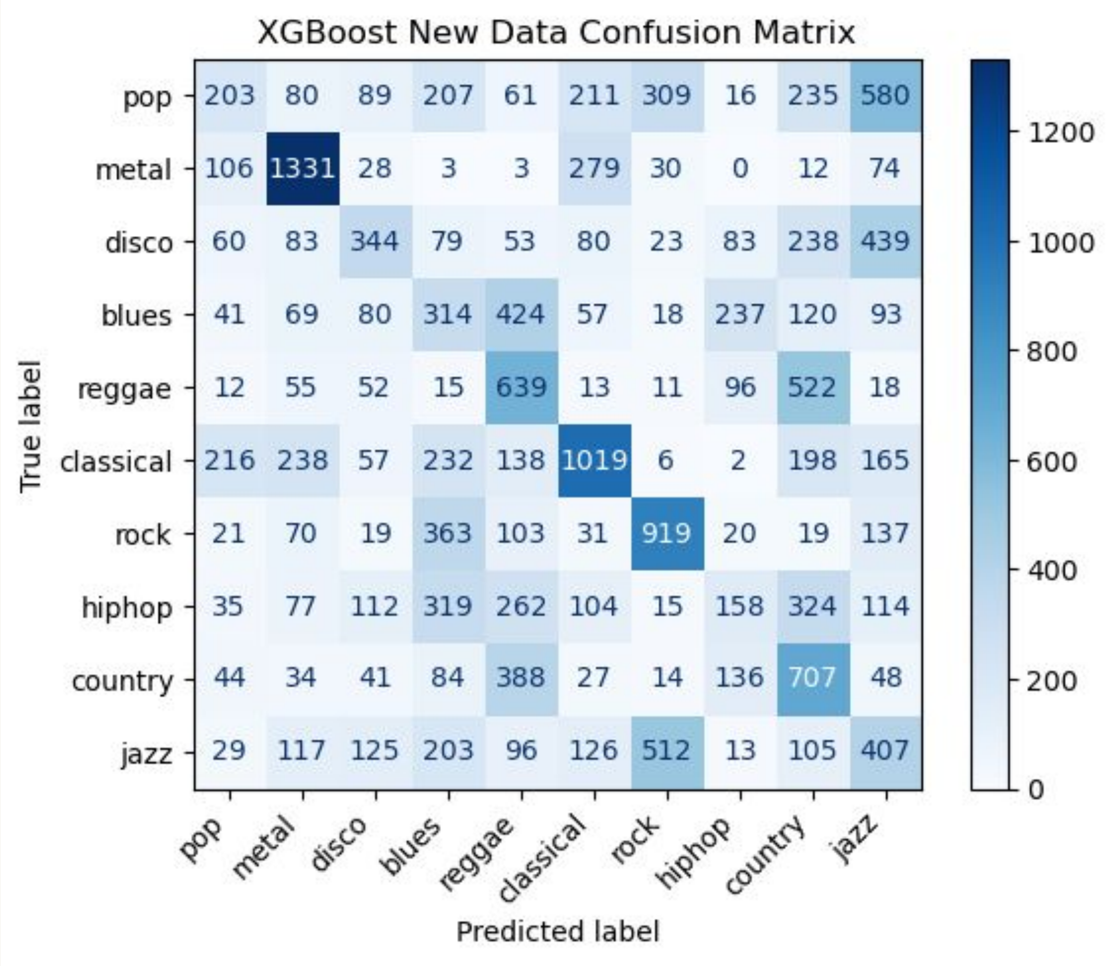
Genre	Accuracy	Precision
Blues	0.26	0.10
Classical	0.62	0.71
Country	0.36	0.23
Disco	0.17	0.22
Hiphop	0.29	0.45
Jazz	0.52	0.45
Metal	0.49	0.54
Pop	0.21	0.10
Reggae	0.29	0.46
Rock	0.20	0.23

Accuracy	0.356
Precision	0.342
Recall	0.350
F1 Score	0.334

The overall accuracy was 0.356 Precision 0.342 Recall 0.350 F1 Score 0.334

Genre	Blues	Classical	Country	Disco	Hiphop	Jazz	Metal	Pop	Reggae	Rock
Accuracy	0.26	0.62	0.36	0.17	0.29	0.52	0.49	0.21	0.29	0.20
Precision	0.10	0.71	0.23	0.22	0.45	0.45	0.54	0.10	0.46	0.23

Data Drift


XGBoost New Data Confusion Matrix
True label
Predicted label
pop metal disco blues reggae classical rock hiphop country jazz
pop 203 80 89 207 61 211 309 16 235 580
metal 106 1331 28 3 3 279 30 0 12 74
disco 60 83 344 79 53 80 23 83 238 439
blues 41 69 80 314 424 57 18 237 120 93
reggae 12 55 52 15 639 13 11 96 522 18
classical 216 238 57 232 138 1019 6 2 198 165
rock 21 70 19 363 103 31 919 20 19 137
hiphop 35 77 112 319 262 104 15 158 324 114
country 44 34 41 84 388 27 14 136 707 48
jazz 29 117 125 203 96 126 512 13 105 407
0
200
400
600
800
1000
1200

Future Scope

Applications

Music Recommendation Systems:

Enhance music streaming services with improved genre-based recommendations.

Digital Music Libraries Organization:

Automate and refine the categorization of large music collections.

Music Production Assistance:

Help producers align their tracks with genre-specific characteristics.

Enhancements

Real-Time Classification:

Optimize the model for immediate music genre classification in live settings.

Classify into Multiple Genres:

Many songs don't fit into exactly one bucket. Creating a binary classifier for each genre could be beneficial.

Global Scale Adaptation:

Expand the model to recognize and classify a diverse range of global music genres.

Conclusion

Importance of Model Selection:

- Evaluated tradeoffs between different models and observed that ensemble models can excel at classifying items with many features

Insights Regarding Audio Data:

- Music genre classification requires models that are able to capture non-linear patterns and adapt to overlapping features

Model Optimization:

- Learned how to balance the challenges of computation time, hyperparameter tuning, and the complexities of our dataset to achieve optimal model performance

Thanks!

Appendix/Resources

- https://www.kaggle.com/datasets/andradaolteanu/gtzan-dataset-music-genre-classification
- https://towardsdatascience.com/music-genre-classification-with-python-c714d032f0d8
- https://towardsdatascience.com/extract-features-of-music-75a3f9bc265d
- https://medium.com/bisa-ai/music-genre-classification-using-convolutional-neural-network-7109508ced47
- https://towardsdatascience.com/music-genre-classification-using-a-divide-conquer-crnn-2ff1cf49859f
- https://arxiv.org/pdf/1712.08370v1