

# Classifying Genres in Spotify

---

Caroline Shi

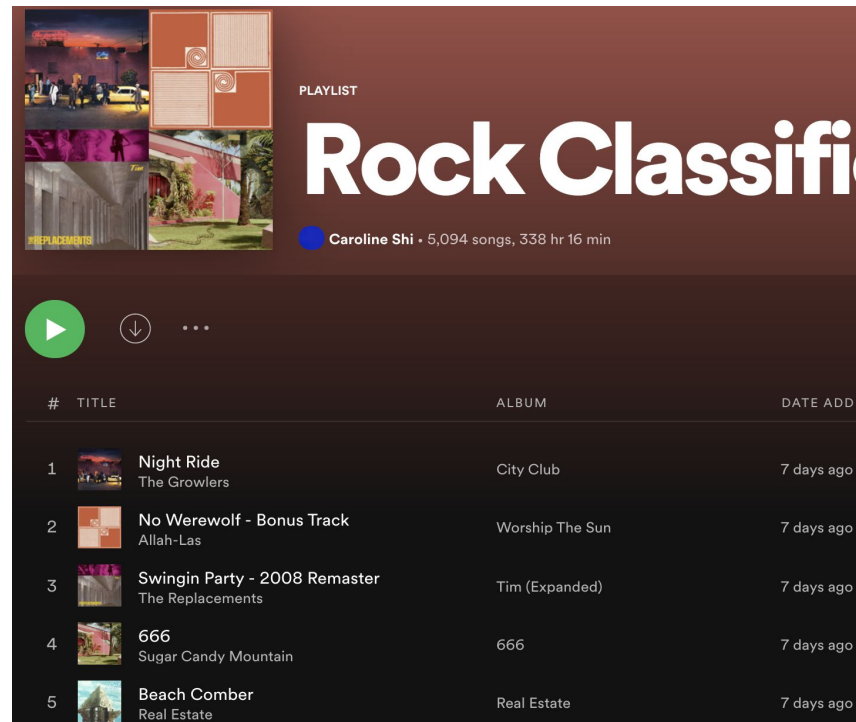
# Objective

- How accurately can we classify music genres for tracks based off of their features?
- Are some features that are more important for classifying one genre vs another?



# Grabbing Data from Spotify

- 10 → 27k tracks from Spotify API
  - Balanced dataset
  - Dropping nulls during API
- 18 total features
  - 14 numeric track characteristics
  - No feature engineering
  - Complexity

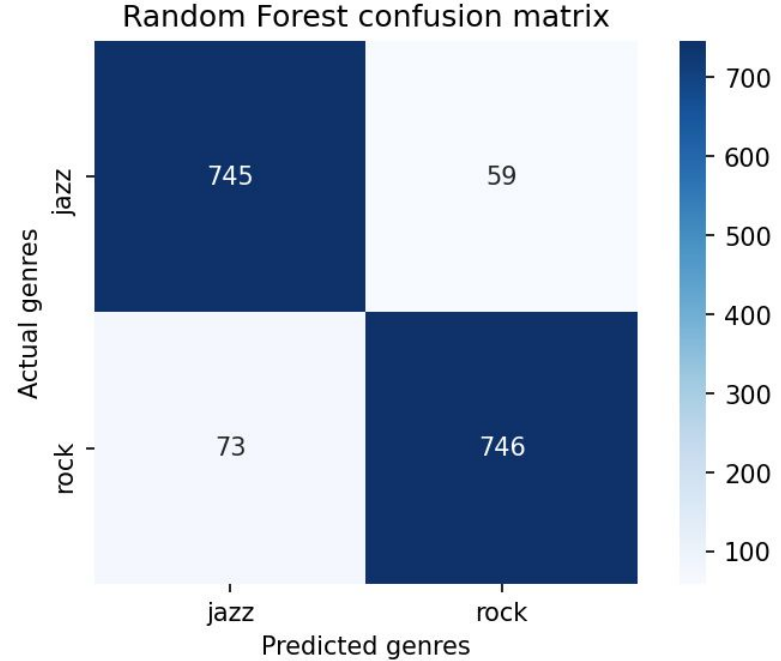


# Baseline Binary Model

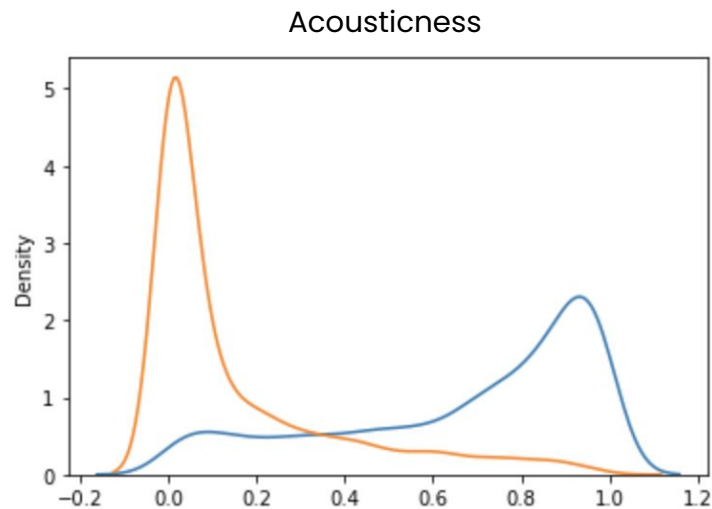
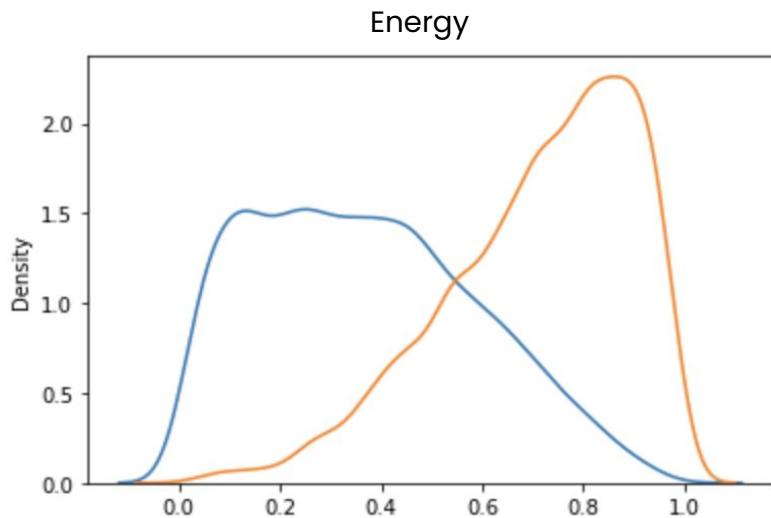
Classification Model	Accuracy
KNN	0.904



Random Forest	0.919
---------------	-------



# Looking at Feature Distribution Between Genres



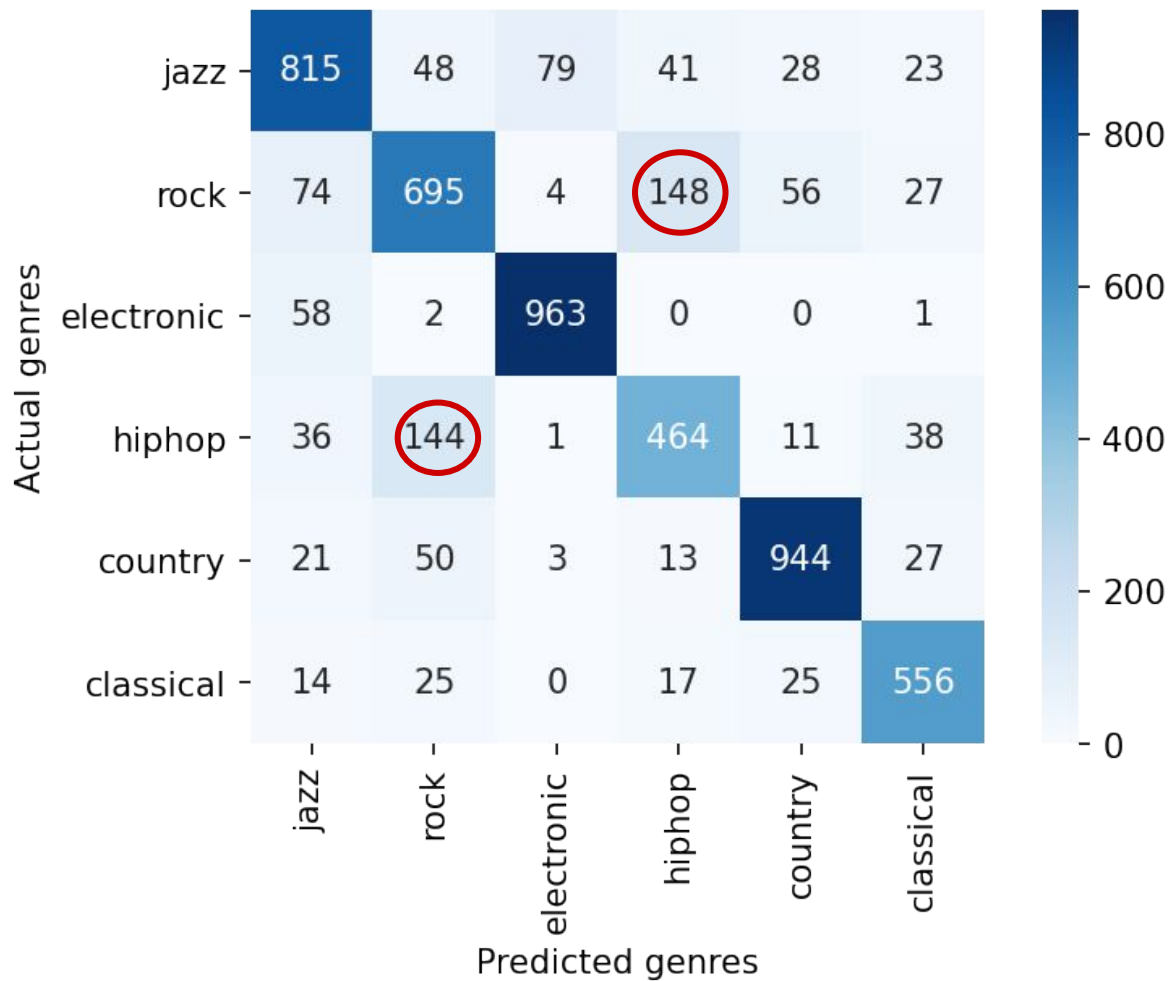
# Final Six Class Model

Classification Models	Accuracy
KNN	0.741



XGBoost	0.816
---------	-------

Test set, XGBoost confusion matrix



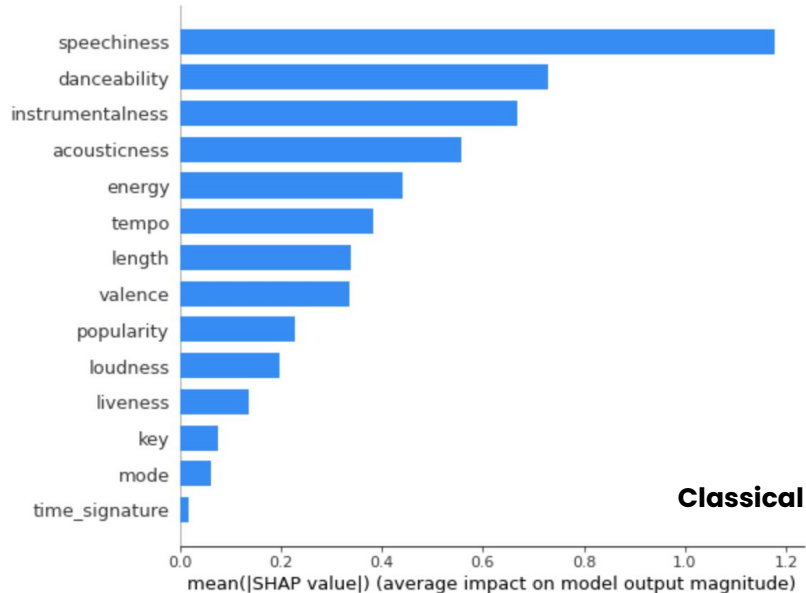
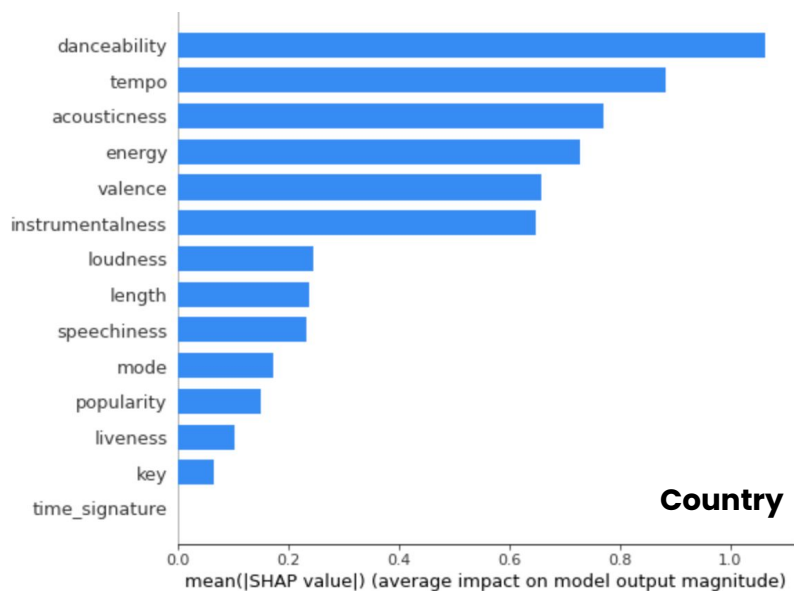
# Song Predictions



Track Title	Artist	Actual Genre	Predicted Genre
Kids	MGMT	Rock	Rock
Tell Me When To Go	E-40	Hip-Hop	Hip-Hop
Sweet Caroline	Neil Diamond	Rock	Hip-Hop
In a Silent Way	Miles Davis	Jazz	Electronic



# Comparing Feature Importance Between Genres



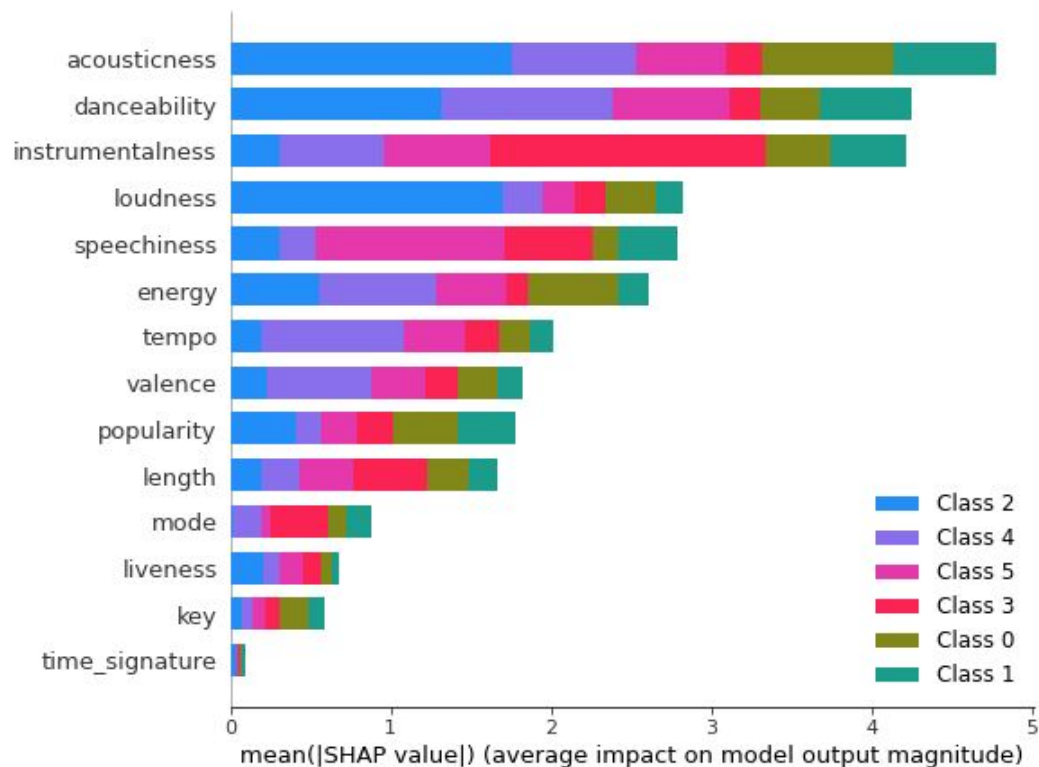
# Future Work

- Tuning XGBoost
- Looking further into feature importance per genre
- More features outside of track features?
- More genres



**Questions?**

# Appendix



# Appendix 2

