Classifying a Song's Genre Using Lyrics

Analysis for music streaming companies I.e. Spotify/Pandora

Noah-John Hizon
7/29/22

Outline

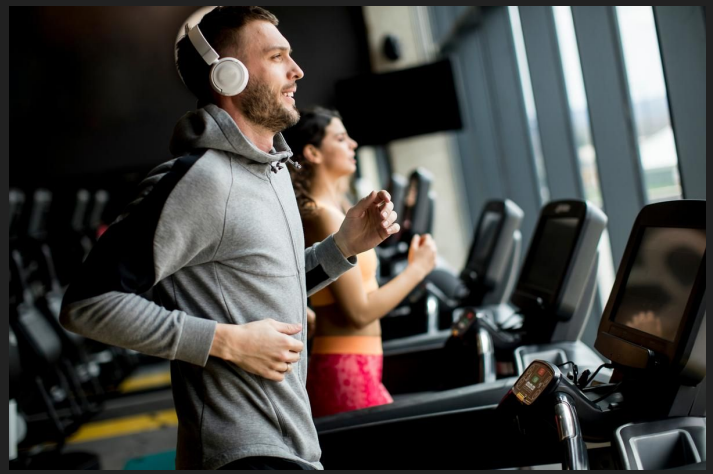
- Business Problem
- Data Understanding
- Classifier Model
- Classifier Results
- Recommendations
- Next Steps

Introduction

How do you find music?

What affects your choice of music at any given time?

- Setting
  - Dinner Party
  - Club
  - Studying
- Mood
  - Heartbreak
  - Exploring
  - Gym



Business Problem

Finding music via apps (Spotify, Apple Music, Pandora) is painstakingly difficult nowadays, especially when it comes to finding the right song/artist/album that fits a listener's mood

- Too much music on streaming services; market is oversaturated
- Curated, user-made playlists don't fit personal preference
- Radio stations composed of a song are too broad

Data Understanding

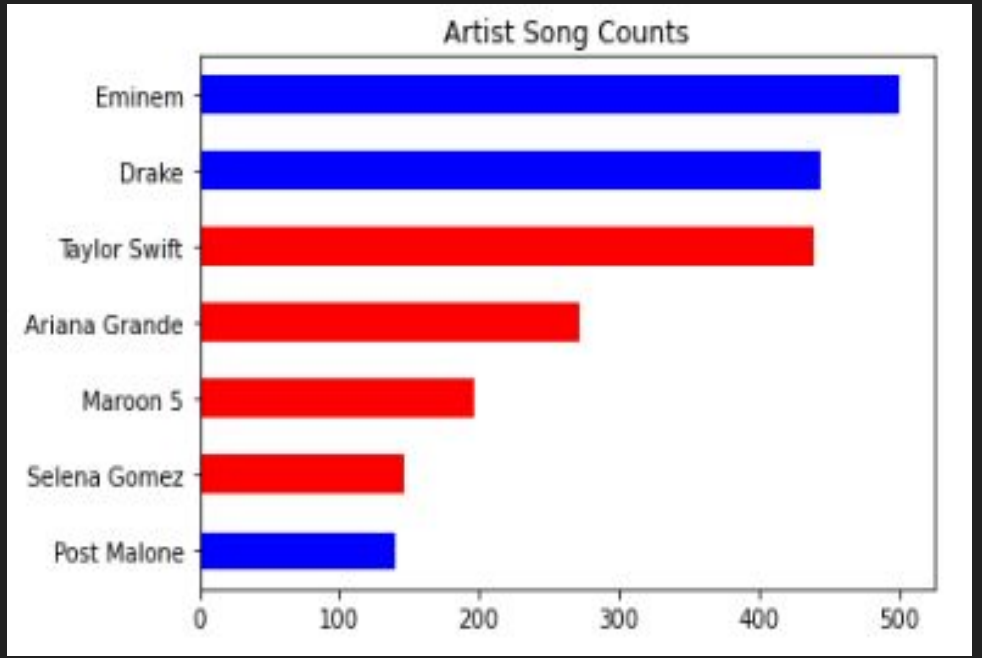
SOURCE: Ariana Grande, Drake, Eminem, Taylor Swift, Post Malone, Maroon 5, and Selena Gomez's lyric set for full discography through 2021 releases (Kaggle)

- Each row in the table represents a unique song, with lyrics stored in a singular text string
- **Columns of interest: Lyric, Genre designation**

Making the distribution of song count

<u>Model Record Count: 2,137 total</u>

In the chart below, Blue indicates Rap songs (1,084 total songs) and Red indicates Pop artists (1,053 total songs)



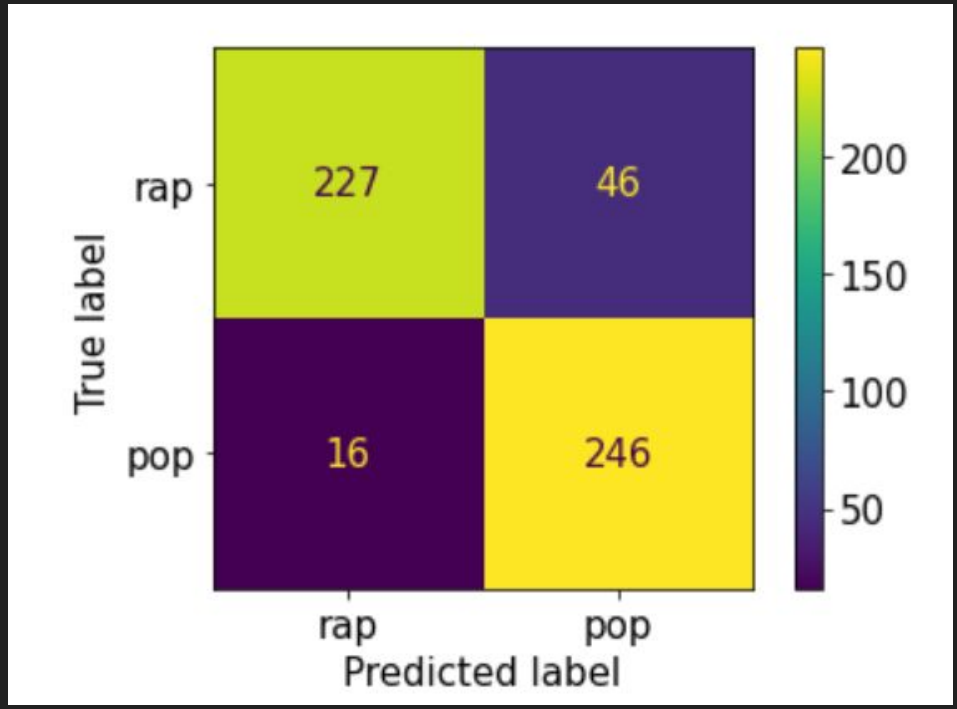
Modeling

An iterative process of classifier model comparison was used to collectively improve classification to a song's respective genre using lyrics.

Final Model: Multinomial Naive Bayes, TF-IDF Vectorizer

Interpretation of Confusion Matrix:

- 227 songs were correctly predicted as Rap
- 246 songs were correctly predicted as Rap
- 46 songs were incorrectly interpreted as Pop, when actual genre was Rap
- 16 songs were incorrectly interpreted as Rap, when actual genre was Rap

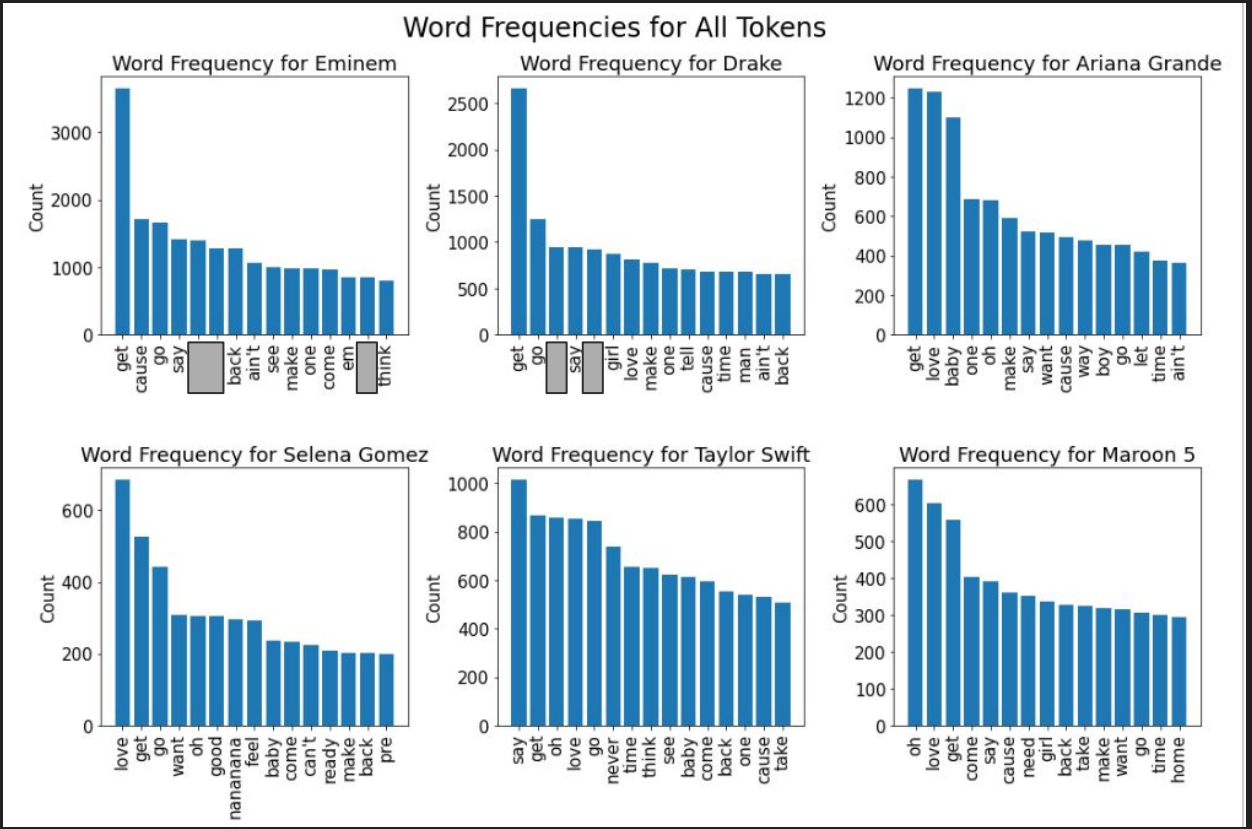


Confusion Matrix for Final Model

Classifier Results

Model was able to classify a song's genre with an accuracy of 92%

Model is also relatively accurate in the face of test data (88%)



Conclusion

From the model results, it can be seen that music genre can be successfully classified using a given song's lyrics, with high accuracy in the face of new (test) data

Lyrics contain invaluable information regarding the context/content of a song and should be implemented when it comes to the current state of recommendation systems

- Notes regarding results: Overlap between genres, Profanity, gender

Next Steps (Post-Proof of Concept)

- Broaden dataset
  - Expand dataset to other genres
  - Look back at rap/pop music from a different age (i.e. 90s)
  - Classifying by mood/other keywords instead of just genre
- Recommendation tool based on choice words depicting mood
  - Would require outside sentiment gathering for analysis
- Consider bigrams for context within different genres
  - Gender

Thank You!

Email: noahiz917@gmail.com

Github: noahiz917

LinkedIn: linkedin.com/in/noah-john-hizon/