

207 Final Project

Spotify Recommender

Matthew Dodd

Rithvik Sunku

Angikaar Singh Chana

A dark blue diagonal gradient bar that starts from the bottom left corner and extends towards the top right corner, covering the lower half of the slide.

Introduction & Overview

- Our project was to create a song recommender system using Spotify
 - As a general overview, our project takes in songs from a user-provided playlist and returns similar songs that the model “recommends” for the user
- To operationalize our project, we utilized various tools and techniques:
 - Spotify API (Spotipy), Genius API
 - Machine Learning (scikit-learn, KMeans, GMMs)
 - TextBlob (Sentiment Analysis / NLP)

Problem Statements & Motivating Questions

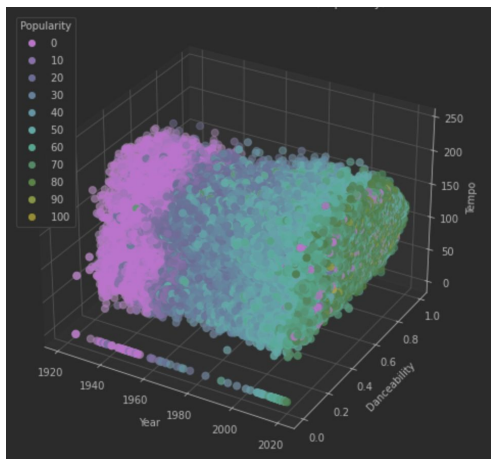
Our project can be summarized into a series of problem statements and motivating questions:

- Spotify users enjoy specific music styles but might want to diversity their playlists.
 - Can a Spotify user's liked songs' characteristics be utilized to recommend new songs based on similar features?
- Various song characteristics (e.g. acousticness, danceability) might have differing impacts on a listener's music preference.
 - Can a model identify and leverage characteristic trends to recommend songs that align with those preferences?
- Lyrics can influence the mood or sentiment of a song, affecting a listener's preference.
 - Can a model incorporate sentiment analysis to enhance song recommendations based on the mood or sentiment of a user's liked songs?

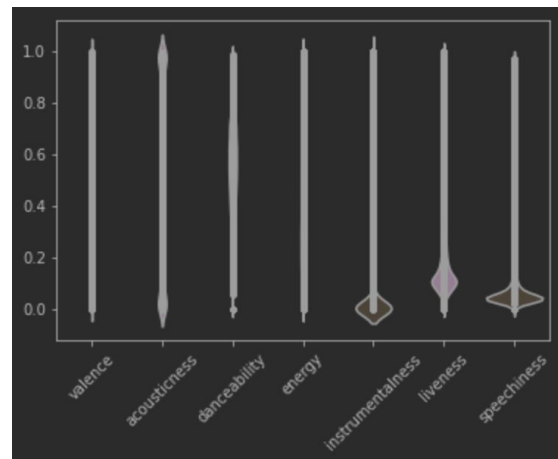
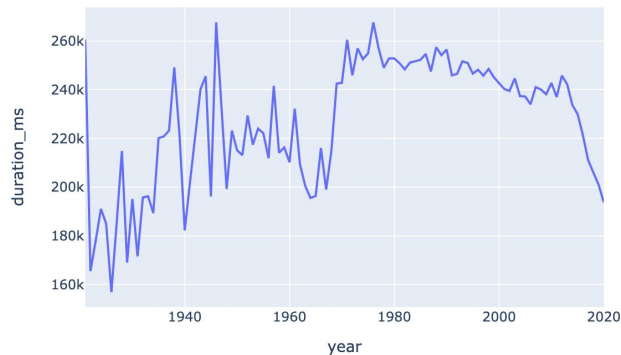
Data and APIs

- Our model uses one major dataset and two APIs:
 - `data.csv`, sourced from Kaggle, $n=170653$, contains various song features (e.g. acousticness, danceability) with which to train our model
 - `spotipy`, sourced from the Spotify API, allows a user's playlists to be input as the playlist to recommend from
 - `lyricsgenius`, sourced from the Genius API, allows a song's lyrics to be pulled for use in sentiment analysis (if applicable), yielding sentiment polarity and subjectivity

Exploratory Data Analysis



Trend of song duration over decades



Data Completeness

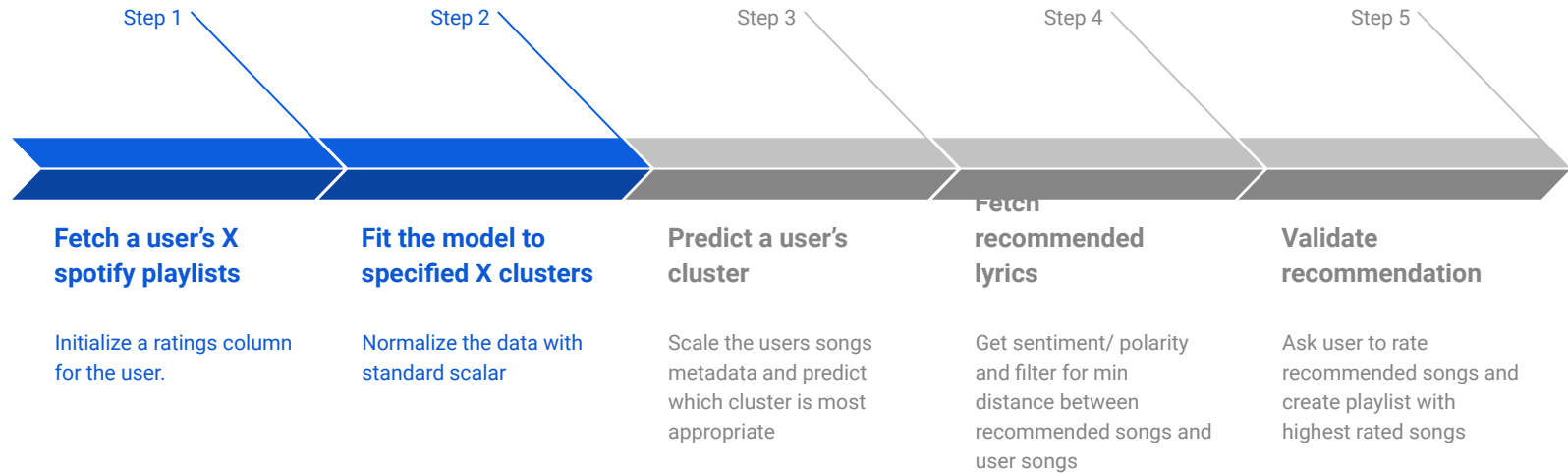
- Data Preparation
 - No NA's to deal with
 - No duplicate
 - Normalized the data with standard units
- Feature engineering
 - Feature selection
 - Adding lyrics and sentiment
 - Selecting number of clusters

```
""" Null values check """
```

```
data.isnull().sum()
```

```
valence      0
year         0
acousticness 0
artists      0
danceability 0
duration_ms  0
energy       0
explicit     0
id           0
instrumentalness 0
key          0
liveness     0
loudness     0
mode         0
name         0
popularity   0
release_date 0
speechiness  0
tempo        0
user_rating  0
dtype: int64
```

Methodology



Modelling Technique

- Our model uses two specific modelling techniques:
 - `Kmeans`: Unsupervised model that iteratively reduces newtonian distance between similar points and their centroid.
 - `Gaussian Mixture Models`, Unsupervised model that iteratively finds gaussian distributions to fit the data using likelihood estimation.
 - `Combined Cluster`: K-means algorithm is only used as a fallback option in cases where the GMM, algorithm may have performed poorly or generated incorrect cluster assignments.
 - `Post-predictive filtering`: sourced from the Genius API, allows a song's lyrics to be pulled for use in sentiment analysis (if applicable), yielding sentiment polarity and subjectivity used to filter best recommendations from cluster predictions
 - `Model Refinement`: looped process gathers `user_ratings` on recommended songs and informs future clustering

Model “Demo”

Song from playlist:

<https://open.spotify.com/track/3ZuT0Evo8chdVM6rPXXqgd?si=b89e240b09fa432f>

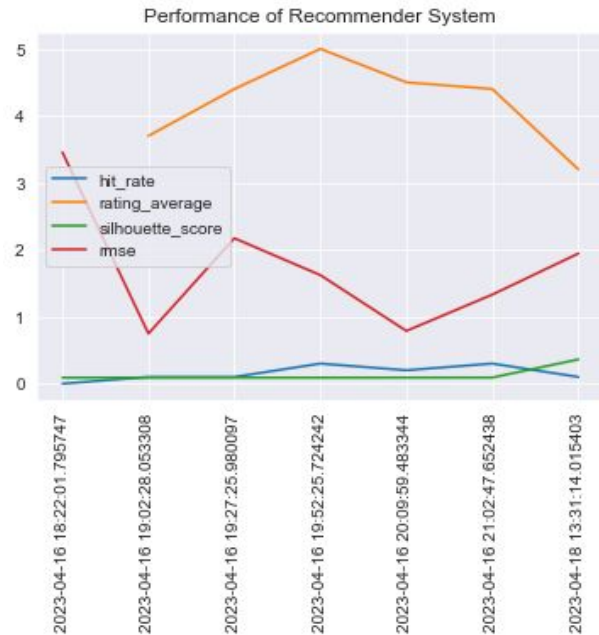
Song from recommender:

<https://open.spotify.com/track/5u6gmrJaLVWj1wjhdT2REM>

```
Your playlists:
1. Hot House
2. Darren After's USB
3. Alt vibey
4. squadveshirea
5. Eat clen n do tren
6. Mafia Baus
7. Deep
8. 1940s Hits (Billboard #1 Singles)
9. I Love My West Coast Classics
10. Gold School
Enter the number of the playlist you want to use for recommendations: 3
Retraining and Fitting Dataset...
Combined Silhouette score: 0.09028749565333252
Recommended songs:
1. Still a Nigga - Eazy-E
2. Gutter Girl - Hot Flash Heat Wave
3. What's Happenin! - Ying Yang Twins
4. Hooked - Why Don't We
5. I Love This Bar - Toby Keith
6. Obsessed - Mariah Carey
7. Who Made Who - AC/DC
8. Prospect (ft. Lil Baby) - iann dior
9. Rolling Stoned - Upchurch
10. Triple Trouble - Beastie Boys
Song Link Here: https://open.spotify.com/track/3q3dgg56jNdMv7mWntV2kV
On a scale of 1-10, how much do you like the song Still a Nigga by Eazy-E? Type esc to get new recommendations.4
Song Link Here: https://open.spotify.com/track/1sEzu2NasuG8s100HwYfN2
On a scale of 1-10, how much do you like the song Gutter Girl by Hot Flash Heat Wave? Type esc to get new recommendations.4
Song Link Here: https://open.spotify.com/track/5Sw4NiI7SniktUwPweTvs
On a scale of 1-10, how much do you like the song What's Happenin! by Ying Yang Twins? Type esc to get new recommendations.5
Song Link Here: https://open.spotify.com/track/2N2s2s2w0jmSIPNZcm8Jnr
On a scale of 1-10, how much do you like the song Hooked by Why Don't We? Type esc to get new recommendations.7
Song Link Here: https://open.spotify.com/track/2EEWdLpL97JcfN1BoMl
On a scale of 1-10, how much do you like the song I Love This Bar by Toby Keith? Type esc to get new recommendations.2
Song Link Here: https://open.spotify.com/track/3IcII2MMS7UARJJJPtEHXG8
On a scale of 1-10, how much do you like the song Obsessed by Mariah Carey? Type esc to get new recommendations.5
Song Link Here: https://open.spotify.com/track/6rvldt6EoZwz0gApECUCwd
On a scale of 1-10, how much do you like the song Who Made Who by AC/DC? Type esc to get new recommendations.3
Song Link Here: https://open.spotify.com/track/42jdzxx0dsavs7Ehr8fGE
On a scale of 1-10, how much do you like the song Prospect (ft. Lil Baby) by iann dior? Type esc to get new recommendations.6
Song Link Here: https://open.spotify.com/track/28B02YJaBmp2bygoM9H92n
On a scale of 1-10, how much do you like the song Rolling Stoned by Upchurch? Type esc to get new recommendations.3
Song Link Here: https://open.spotify.com/track/67wjzKe19et4fwb7ID6zAu
On a scale of 1-10, how much do you like the song Triple Trouble by Beastie Boys? Type esc to get new recommendations.5
1 highly rated songs added to your new playlist!
Are you satisfied with the recommendations? Say yes to exit (y/n) 
```

Metrics & Evaluation

- Tracked on each iteration of the recommender
 - Hit Rate: currently around 0.1
 - Rating Average: currently around 4.5
 - Silhouette Score: around 0.57
 - RMSE on user rated songs: fluctuating around 2
- Various ways to evaluate classifier (e.g. confusion matrix)



Takeaways

- Spotify's algorithm is really good and combines various data sources
- Having multiple users use the model has its tradeoffs
- Compute, caching, and efficient models are important for iterative feedback on recommender system
- Some genres are harder to cluster than others (e.g. indie)
- Users are very biased when looking for recommendations (usually want same artist or time period)



Future Improvements

- Explore additional features (e.g. genre, release year) to yield better predictions
 - Spotify users may prefer music from a certain genre or certain decade/era
- Experiment with more robust data and/or diverse playlists
 - Our model could be improved provided a longer timeline and access to more robust data than that which is publicly accessible
 - Could motivate collaborative filtering and more complex approaches
- Incorporate further NLP analysis such as social listening
 - Sentiment analysis could go beyond simply lyrics, could also involve other users with similar musical preferences