MIE368 - Analytics in Action:

Project Abstract

## Content-Based Spotify™ Playlist Recommendation System

November 27, 2024

*Team 20*

### Hussein Ismail (1008747910)

### Marco Mastrangelo (1009133518)

### Devraj Solanki (1009065707)

### Yahia Hussein (1008956444)

Providing good song recommendations to users is a challenge many music streaming platforms face [1]. On popular streaming platforms like Spotify, user recommendations typically rely on the preferences of similar users. However, this approach falls short when users seek highly personalized or content-specific suggestions. To address this issue, the team proposes acontent-based recommendation system [2] that suggests new songs to add to a playlist based on its existing content. For data collection, a “song database” was created by querying the top 10 songs of the top 2500 most-listened-to artists on Spotify using the Spotify Web API [3]. The API extracted general information about each track including track name, artist name(s), release date etc. as well as audial metrics such as danceability, valence, acoustic-ness etc. Additionally, playlists were also queried, extracting general information and audial metrics for each song. The song recommendation system consists of a two-part machine learning model. The first part involved training a denoising autoencoder to generate song embeddings of each track in the song database. Then, the ‘target’ playlist (which we want to create recommendations for) was aggregated into one “mean” song, and encoded to generate a playlist embedding. Finally, using cosine similarity, the top 20 songs most similar to the playlist embedding were identified as song recommendations. The second part of the team's model assessed the quality of the top 20 songs recommended by the first part of the model. For this part, a dataset was created with every song from the ‘target’ playlist and an equal number of random songs not in it. An XGBoost classifier was trained on this dataset to predict if a song should be in the ‘target’ playlist. The top 20 recommendations were evaluated using this model to determine if any of the 20 recommendations belong to the playlist/ are good quality recommendations for the playlist. The team's two-part machine learning model was tested on a variety of playlists, including genre-based and utility-focused playlists (e.g., workout, study). The XGboost model was fit to each of these playlists resulting in an average test accuracy of 0.84 and average precision of 0.86 across the playlists tested. The model performed best on genre-specific playlists like pop and old-school rap, which had test precisions of 0.93 and 0.94 respectively. From these results, it was concluded that playlists that shared a specific genre or music type will typically result in a more accurate classification model, and thus better quality recommendations.

**References**

[1] A. M. A. Ahmed, “5 Challenges in Building Recommender Systems,” *Medium*, Nov. 04, 2023. https://awadrahman.medium.com/5-challenges-in-building-recommender-systems-77acbc0948cb

[2] A. Kharwal, “Content Based Filtering and Collaborative Filtering: Difference,” *The Clever Programmer*, Apr. 20, 2023. https://thecleverprogrammer.com/2023/04/20/content-based-filtering-and-collaborative-filtering-difference/#:~:text=Content%2Dbased%20filtering%20is%20like,with%20similar%20preferences%20have%20liked.

[3] Spotify, “Web API | Spotify for Developers,” *developer.spotify.com*. https://developer.spotify.com/documentation/web-api.