

Personalized Spotify Recommendation algorithm - Aryamaan Parikh

This project aims to provide personalized playlists based on textual user input. The user would write (in natural language) about how they feel, or the vibe of the playlist they wish for, and the algorithm will create a **personalized** playlist, based on what the user wished for, for their enjoyment.

This is something I personally would very much benefit from, as I really do not enjoy searching for new music of a certain mood, that I would like. In addition to this, my interest lies in the domain of personalized content discovery, and this project would help me gain good experience in the same.

The data is available in fragments across the internet, but to start with, I would use the data published for a spotify competition (Spotify stream activity dataset). The subset of this data has around 20,000 rows, with Song ID. I would then have to fetch other song features using the Spotify API (BPM, loudness, danceability, etc..) The intention of this data gathering activity is to get as much information about the **song** as possible. I would also need the data of users (1000 users, approx) so I can implement user-user collaborative filtering. In addition to this, I would also need to retrieve my own personal account's playlists, in order to build some sort of **user vector**, to capture my taste.

The approach involves modules, but I am yet to determine the exact order of them. I have described the modules in brief below.

1. User clustering module : This would cluster similar users to myself, and provide me with a way to leverage their playlists, for potentially new content that I may like.
2. NLU module : This would extract emotions and relevant phrases from the user's input, and help reduce the search space, within which recommendations would be derived from. Could potentially involve leveraging **OOB semantic similarity**, to handle synonyms and paraphrases. (Google news)
3. Song vectorizer module : This module would create an item vector for each song in the dataset.
4. User vectorizer module : This module would create a user vector for each user in the dataset (To determine their taste)

Will add more as things become clearer. This will leverage the spotify API, and **hopefully**, I can embed it into a very bare bones mobile app, with a simple UI (Text box + button "Find me music!", would return a link to a spotify playlist created publicly for you.) As of now, in prelim steps, working on a local machine with data of the order of 20,000. Will scale as needed.