Project Proposal

Song Recommendation System

Danish James

Domain

Music is one of the most widely available forms of art and impacts many people on a very wide scale. Many people listen to music on a day to day basis and it is a subject of discussion for many reasons. However, listening to the same music all the time will get boring for anyone, and thus people are always on the lookout for something new, something they haven't experienced before. For my project, I aim to create a song recommendation system that will introduce the users to new music that they have not heard of before, but would probably like. The full product would consist of a website where people can connect with their spotify accounts. The model would then collect their song history, see what was liked and disliked, and generate a playlist based off of their history as a suggestion.

An **MVP** would consist of taking in a single song and generating similar songs for a playlist suggestion. For this, the MVP would take the song, analyze its low-level features such as tempo, energy, loudness, etc as well as features like the artist, the genre, and lyrical analysis and utilize that to find songs with similar features and lyrical content. The full model will also include a user database that would be the initial step of trimming down song choices. The primary objective is to find users with similar tastes and use their musical history as a filter for songs to look at.

Data

The data will consist of several steps. The first step is to create a user database, along with their liked songs. With an account on last.fm, one can see the usernames of listeners for specific artists. Utilizing the last.fm API, I will scrape together a list of usernames across a variety of different artists and genres, and then pool them together along with their liked songs into a database. From there, I will get a database for the songs that will contain their low-level features, additional features such as the artist and the genre, and their spotify ID for tracking purposes. These two databases combined would be the source of my training data for my model.