Spotify Song Lyrics Genre Analysis
Evelyn Gaskin
7/10

Description

I want to see if I can train a model to predict trends in song's genres based on their lyrics and a couple statistics like tempo and song key or length. I'm interested in if different genres have different lyrical trends (besides classical or instrumental music, of course).
If possible, I'd like to use a user's whole sequential listening history to learn and predict genre trends similar to NetTalk, but if that is not compatible with the models we use in this class I'll just predict genres for singular songs, as opposed to a user's genre preference or next listening genre.

Data Features

I want to use a song's ID to get statistics like tempo, key, and loudness. I'll also use the song lyrics (and one-hot encoding using either the whole song, or chorus's or specific stanzas). The target output will be the song's genre.

Tempo	Key	Love	True	..	Word n	Target: Genre
110	9	2	5	..	0	Alternative

Data Gathering

Spotify allows users to get their listening data and history in the form of a JSON. Spotify also has an API that allows users to get playlist and listening data for users (amongst a host of other endpoints), and use a specific song's ID to get statistics for it called audio features and audio analysis. These include song genre, tempo, and key. I can use these features to generate data corresponding to a user's listening history.
For song lyrics, there's another API Some Random API that I can use to get song lyrics for a song's title. There are other lyric API's out there but this is one of the more reliable, straightforward ones I could find.

Spotipy is a python library I can use for the Spotify API, and the requests library for the other API requests.