Name: Francesca Beller

Phone: 703-980-5238

Email: fbeller@regis.edu

Project Title: Spotify Recommendation Algorithm using Python

Type of Task: Machine Learning Predictive Algorithm

**Data**

The primary dataset that will be used will consist of two song sets, a “like” playlist and a “dislike” playlist. These will be pulled into the code using Spotify’s developer API and the *spotipy* Python library, which includes the following song information:

## audio features object

| KEY | VALUE TYPE | VALUE DESCRIPTION |
|---|---|---|
| acousticness | float | A confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic. |
| analysis_url | string | An HTTP URL to access the full audio analysis of this track. An access token is required to access this data. |
| danceability | float | Danceability describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. A value of 0.0 is least danceable and 1.0 is most danceable. |
| duration_ms | int | The duration of the track in milliseconds. |
| energy | float | Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy. For example, death metal has high energy, while a Bach prelude scores low on the scale. Perceptual features contributing to this attribute include dynamic range, perceived loudness, timbre, onset rate, and general entropy. |
| id | string | The Spotify ID for the track. |
| instrumentalness | float | Predicts whether a track contains no vocals. "Ooh" and "aah" sounds are treated as instrumental in this context. Rap or spoken word tracks are clearly "vocal". The closer the instrumentalness value is to 1.0, the greater likelihood the track contains no vocal content. Values above 0.5 are intended to represent instrumental tracks, but confidence is higher as the value approaches 1.0. |
| key | int | The key the track is in. Integers map to pitches using standard Pitch Class notation. E.g. 0 = C, 1 = C♯/D♭, 2 = D, and so on. |

| | | |
|---|---|---|
| liveness | float | Detects the presence of an audience in the recording. Higher liveness values represent an increased probability that the track was performed live. A value above 0.8 provides strong likelihood that the track is live. |
| loudness | float | The overall loudness of a track in decibels (dB). Loudness values are averaged across the entire track and are useful for comparing relative loudness of tracks. Loudness is the quality of a sound that is the primary psychological correlate of physical strength (amplitude). Values typical range between -60 and 0 db. |
| mode | int | Mode indicates the modality (major or minor) of a track, the type of scale from which its melodic content is derived. Major is represented by 1 and minor is 0. |
| speechiness | float | Speechiness detects the presence of spoken words in a track. The more exclusively speech-like the recording (e.g. talk show, audio book, poetry), the closer to 1.0 the attribute value. Values above 0.66 describe tracks that are probably made entirely of spoken words. Values between 0.33 and 0.66 describe tracks that may contain both music and speech, either in sections or layered, including such cases as rap music. Values below 0.33 most likely represent music and other non-speech-like tracks. |
| tempo | float | The overall estimated tempo of a track in beats per minute (BPM). In musical terminology, tempo is the speed or pace of a given piece and derives directly from the average beat duration. |
| time_signature | int | An estimated overall time signature of a track. The time signature (meter) is a notational convention to specify how many beats are in each bar (or measure). |
| track_href | string | A link to the Web API endpoint providing full details of the track. |
| type | string | The object type: "audio_features" |
| uri | string | The Spotify URI for the track. |
| valence | float | A measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry). |

This data will then be used to generate predictions on whether I will like songs from a test playlist. This test playlist will contain songs that I like and dislike, with a binary classifier assigned to each song (1 being a “liked” song, 0 being a “disliked” song). The generated prediction tags for like/dislike will be evaluated against the pre-created tags to determine model accuracy.

**Analysis**

Initial exploratory data analysis will be performed on both the “like” and “dislike” playlists to try to ascertain trends in the song traits (genre, liveness, tempo, etc.) that differ between the two sets. After exploratory data analysis is performed, multiple classifiers will be created using different machine learning algorithms and they will be evaluated against each other. Some of these algorithms include decision trees/random forest, K-nearest neighbors, and Gradient Boost. The model with the highest “success rate” will be the final algorithm.

**Anticipated Difficulties**

The primary anticipated difficulty is compiling the “like” and “dislike” playlists. While this is not difficult in terms of critical thinking, it is a high level of manual work regarding the number of songs required to sufficiently train an algorithm. Each training playlist will need 1000+ songs in order to be useful for training the models. However, I do not anticipate any major difficulties (as of now) in the actual coding for the project.

**Timeline**

Week 1: Project proposal

Week 2: Compiling “like” and “dislike” playlists for training and analysis

Week 3: Exploratory data analysis

Week 4: Initial model coding/training

Week 5: Evaluation of initial models

Week 6: Additional training/tuning/testing to improve model(s)

Week 7: Final model evaluations and beginning to put together project write-up and presentation

Week 8: Finish project write-up and presentation