Espotifai
Automatic Playlist Recommender

Lucas Emanuel Resck Domingues
Lucas Machado Moschen
School of Applied Mathematics
Getulio Vargas Foundation

July 2020

1 Project statement

Question: Which song should we recommend based on a playlist and user/context information?

We can define playlist as a sequence of tracks (audio recordings). In this project we aim to study the playlist generation problem, that is, given a pool of tracks, a background knowledge about the user, and some metadata of songs and playlists, the goal is to create a sequence of tracks that satisfies some target as best as possible.

Our work can be followed on our GitHub repository.

2 Preliminary exploratory data analysis

1. **Inspect the available datasets.**

   The main dataset for this especific work should be The Million Playlist Dataset from RecSys Challenge 2018, but it's unavailable. So we searched for another datasets and API's in order to get data. If we use API's, we will clean and treat the data.

   The options are:

   - Spotify API with SpotiPy library to create a dataset;
   - Last.fm API to create dataset;
   - Million Song Dataset (subset with 10 thousand songs);
   - NowPlaying: Dataset obtained with information from Twitter, when a tweet is tagged with #nowplaying, #listento or #listeningto.
   - Million Playlist Song Dataset (Subset with 10 thousand songs);
   - FMA: A Dataset For Music Analysis Data Set;

- Others datasets.

2. **Analyse the available features of the datasets and how it can help us to solve our question.**

   We imagine some interesting features that can help us in this work:

   - Music genre;
   - Location;
   - Time of the day;
   - etc.

3. **Calculate some statistics about data.**

   - Percentage of listened genres;
   - Percentage of genders;
   - Percentege of countrys;
   - Count of missing data;
   - Distribution of artists auditions.
   - etc.

4. **Explore visually the influence of some categorical and quantitative variables on the features of the audios.**

   For example:

   - How age influences the musical genres heard?
   - How the country influences it?
   - How gender affects it?
   - etc.