Synesthete: A deep learning engine that sees sound

Vivas Kaul
Sprint 1 Deliverable
7/28/2023

Table of Contents

Project Overview

- **Objective:** Create a novel music recommendation engine that will utilize machine learning/deep learning to provide suggestions of similar songs to a user based on a single audio track as an input. The audio can come via a pre-saved audio file or as a recording in real time.
- Current methods
  - Pre-existing data about a user gathered over time
  - Make recommendations based on other users thought to be similar when no previous data exists (cold-start).
- This method relies entirely on song similarity rather than user similarity.

Proposed Solution

1. Create spectrographic images of audio data via encoder/transformer
2. Vectorize images into n-dimensional vectors
3. Collate vectors into singular data set
4. Train convolutional neural network model (CNN)
5. Accept user input
    a. Audio recording captured from phone
    b. Audio file provided to model
6. Convert into image via encoder
7. Perform similarity calculation and provide top five most similar matches
    a. Cosine similarity given that two vectors are being compared

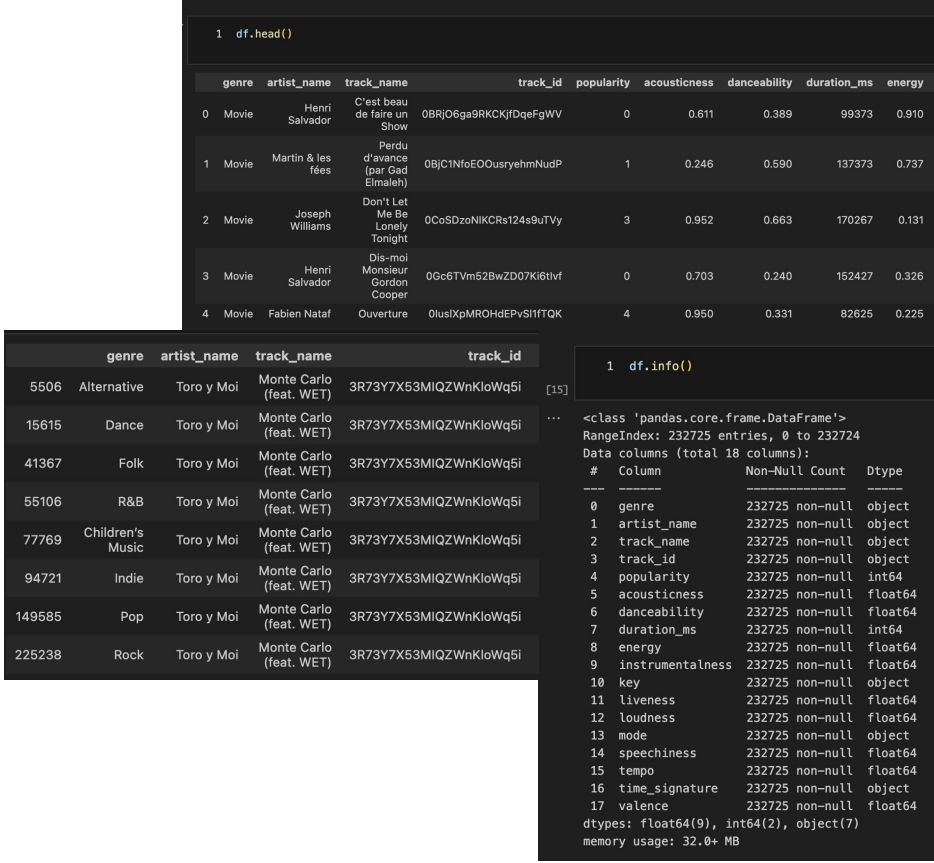
***Note: These steps are subject to change as new information or challenges emerge**

Potential Impact

- User
  - Improving ease of finding similar music
  - Discovery of new artists
- Business
  - Expanded functionality (value add)
  - Increased engagement
- Standalone app
  - Possible disruptor
  - Marketing/advertising tool

Data Overview

- The data set was obtained from Kaggle: https://www.kaggle.com/datasets/zaheenhamidani/ultimate-spotify-tracks-db?resource=download
- 232,725 rows each representing a song
- 26 genres.
  - Roughly 10,000 songs per genre
- The unique track_ids were extracted and used within the spotify_dl tool to pull the mp3 data needed for the model.



```
1  df.head()
```

	genre	artist_name	track_name	track_id	popularity	acousticness	danceability	duration_ms	energy
0	Movie	Henri Salvador	C'est beau de faire un Show	0BRjO6ga9RKCKjfDqeFgWV	0	0.611	0.389	99373	0.910
1	Movie	Martin & les fées	Perdu d'avance (par Gad Elmaleh)	0BjC1NfoEOOusryehmNudP	1	0.246	0.590	137373	0.737
2	Movie	Joseph Williams	Don't Let Me Be Lonely Tonight	0CoSDzoNIKCRs124s9uTVy	3	0.952	0.663	170267	0.131
3	Movie	Henri Salvador	Dis-moi Monsieur Gordon Cooper	0Gc6TVm52BwZD07Ki6tIvf	0	0.703	0.240	152427	0.326
4	Movie	Fabien Nataf	Ouverture	0IuslXpMROHdEPvSl1fTQK	4	0.950	0.331	82625	0.225

	genre	artist_name	track_name	track_id
5506	Alternative	Toro y Moi	Monte Carlo (feat. WET)	3R73Y7X53MIQZWnKloWq5i
15615	Dance	Toro y Moi	Monte Carlo (feat. WET)	3R73Y7X53MIQZWnKloWq5i
41367	Folk	Toro y Moi	Monte Carlo (feat. WET)	3R73Y7X53MIQZWnKloWq5i
55106	R&B	Toro y Moi	Monte Carlo (feat. WET)	3R73Y7X53MIQZWnKloWq5i
77769	Children's Music	Toro y Moi	Monte Carlo (feat. WET)	3R73Y7X53MIQZWnKloWq5i
94721	Indie	Toro y Moi	Monte Carlo (feat. WET)	3R73Y7X53MIQZWnKloWq5i
149585	Pop	Toro y Moi	Monte Carlo (feat. WET)	3R73Y7X53MIQZWnKloWq5i
225238	Rock	Toro y Moi	Monte Carlo (feat. WET)	3R73Y7X53MIQZWnKloWq5i

```
[15]  1  df.info()

...  <class 'pandas.core.frame.DataFrame'>
     RangeIndex: 232725 entries, 0 to 232724
     Data columns (total 18 columns):
      #   Column            Non-Null Count   Dtype
     ---  ------            --------------   -----
      0   genre             232725 non-null  object
      1   artist_name       232725 non-null  object
      2   track_name        232725 non-null  object
      3   track_id          232725 non-null  object
      4   popularity        232725 non-null  int64
      5   acousticness      232725 non-null  float64
      6   danceability      232725 non-null  float64
      7   duration_ms       232725 non-null  int64
      8   energy            232725 non-null  float64
      9   instrumentalness  232725 non-null  float64
      10  key               232725 non-null  object
      11  liveness          232725 non-null  float64
      12  loudness          232725 non-null  float64
      13  mode              232725 non-null  object
      14  speechiness       232725 non-null  float64
      15  tempo             232725 non-null  float64
      16  time_signature    232725 non-null  object
      17  valence           232725 non-null  float64
     dtypes: float64(9), int64(2), object(7)
     memory usage: 32.0+ MB
```

Next Steps for Sprint 2

- Two areas of focus
  - Gathering MP3s for the model
    - This involves downloading at least 10,000 songs from the list of track_ids that were sourced from the Kaggle dataset.
      - spotify_dl
  - Using some kind of audio encoder/transformer to convert the mp3s into vectorized data directly or some kind of spectrographic image which could then be converted into a vectorized form.
    - Initially going to try the Hugging Face transformer.
      - Has tutorials online
- Longer Term (Sprint 3) - Create CNN
  - Tensorflow
  - Pytorch

Questions?