

# Recommending songs in spotify

---

A Predictive Analysis  
Data analysis course project  
Yara Mohammadi Bahram  
24/1/2021



# Dataset

Scraping over 14000 songs from 163 playlists (600 from my playlist) in JSON format from Spotify's API

To remove:

- Songs never repeating across playlists
- Songs with missing information
- Playlists with less than 10 remaining songs



2261 songs remained in total from 87 playlists (70 from our playlist)

# Extracted stuff

Song information:

Song ID	Genres	Explicit	Popularity	Dancability	Loudness	Tempo	Acousticness	Energy	Valence	Liveness	Speechiness
---------	--------	----------	------------	-------------	----------	-------	--------------	--------	---------	----------	-------------

Preprocess all by Standard Scaler

Playlist information:

Playlist ID	Song ID	Added at
-------------	---------	----------

# Genres, genres everywhere...

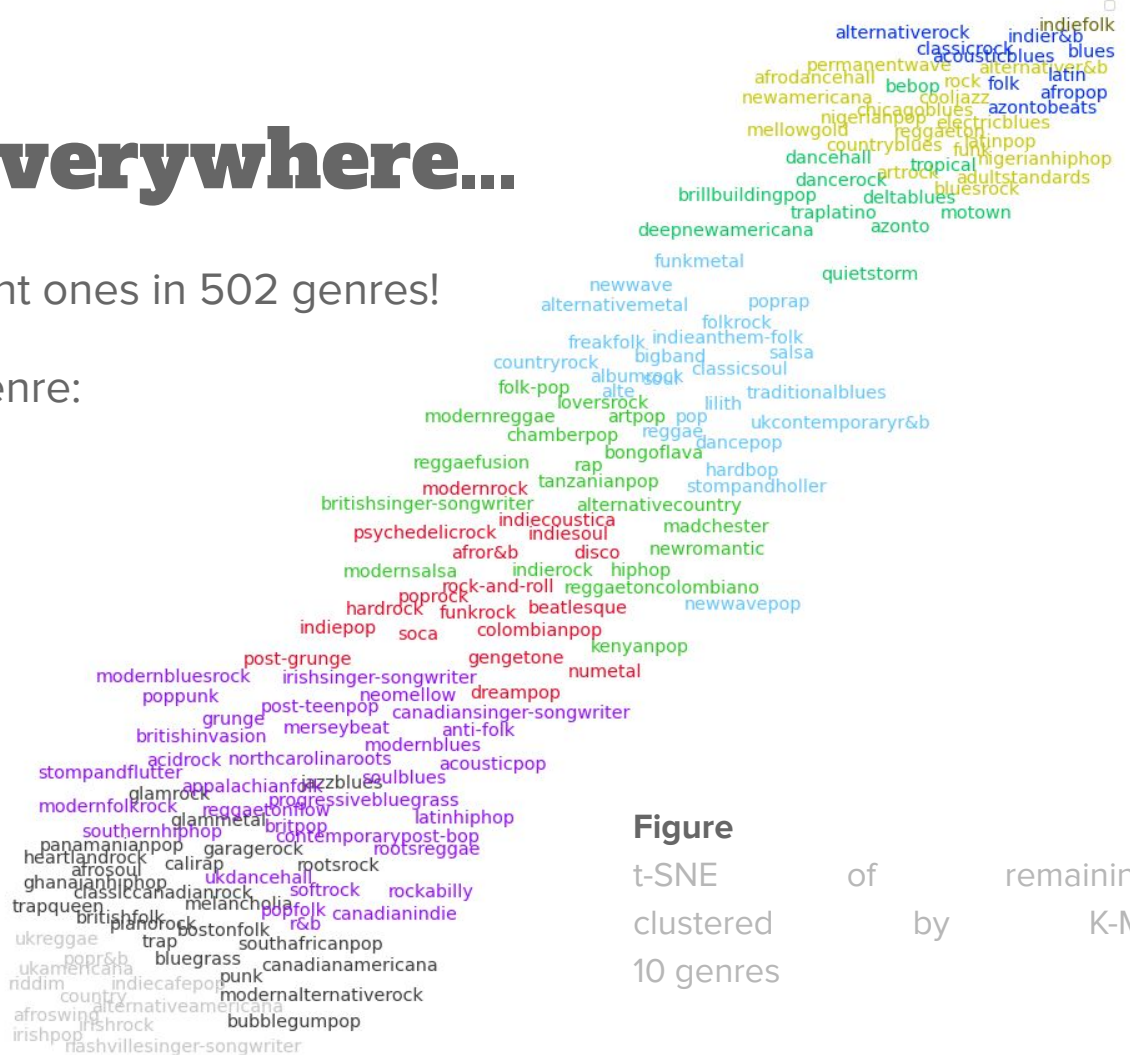
163 genres after removing infrequent ones in 502 genres!

A song may have more than one genre:

[‘album rock’, ‘classic rock’, ‘hard rock’, ‘rock’]

Embedding:

- Word2Vec (Dim: 20)
- TF-IDF weighted average



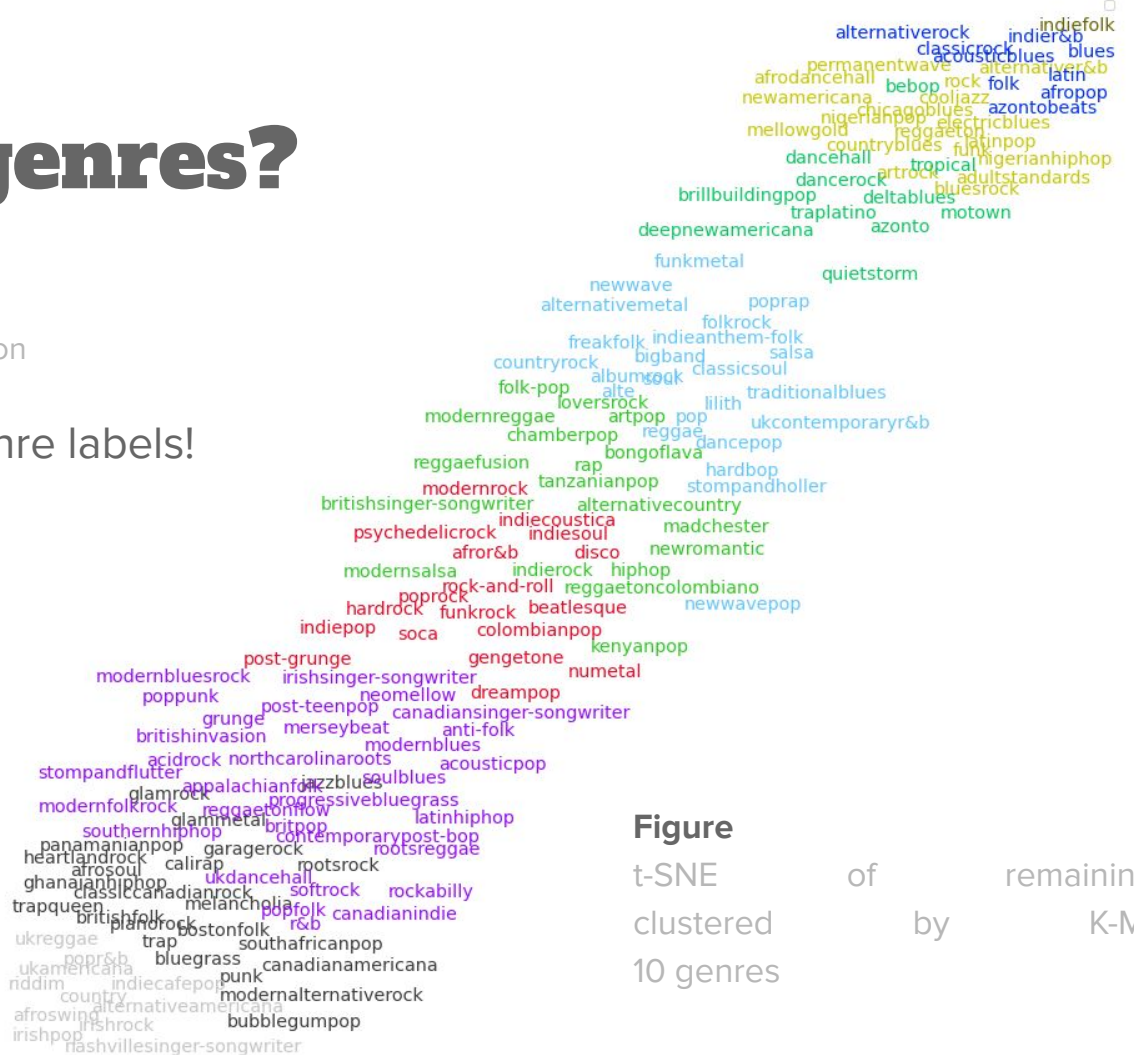
# Can we predict genres?

- Train random forest
  - Grid search and 5 fold cross validation

98% accuracy in predicting new genre labels!

max\_depth: 10,

n\_estimators: 100



# **Content Based Recommendation**

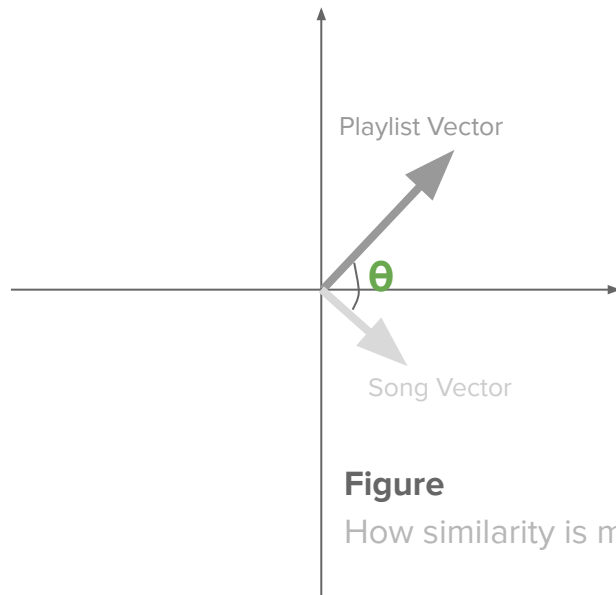


# KNN based on cosine similarity

Create Playlist vector by:

- Averaging my song vectors
- Weighted averaging
  - “Added At” column... Remember?

Song-feature matrix (1090 rows, 33 columns)



# Collaborative Filtering Recommendation





# Methods used

- KNN (same approach, only slightly different)
  - Finding K Nearest Playlists
  - Playlist-song interaction matrix (87 rows, 1090 columns)!
- Matrix Factorization
  - Train SVD with SGD optimization
  - $(87, 20) \times (20, 1090)$     a little better!
- Auto Encoder
  - Encoder + Decoder layer with 20 latent neurons and a custom loss function

5 Fold cross validation used in all

# Evaluation

No labels on recommended songs...

- RMSE & MAE
  - Not perfect for comparing, but good for making sure model is trained.
- Experiment on our playlist
  - Get 10 songs from each recommender
  - Do we like those songs?

# Results

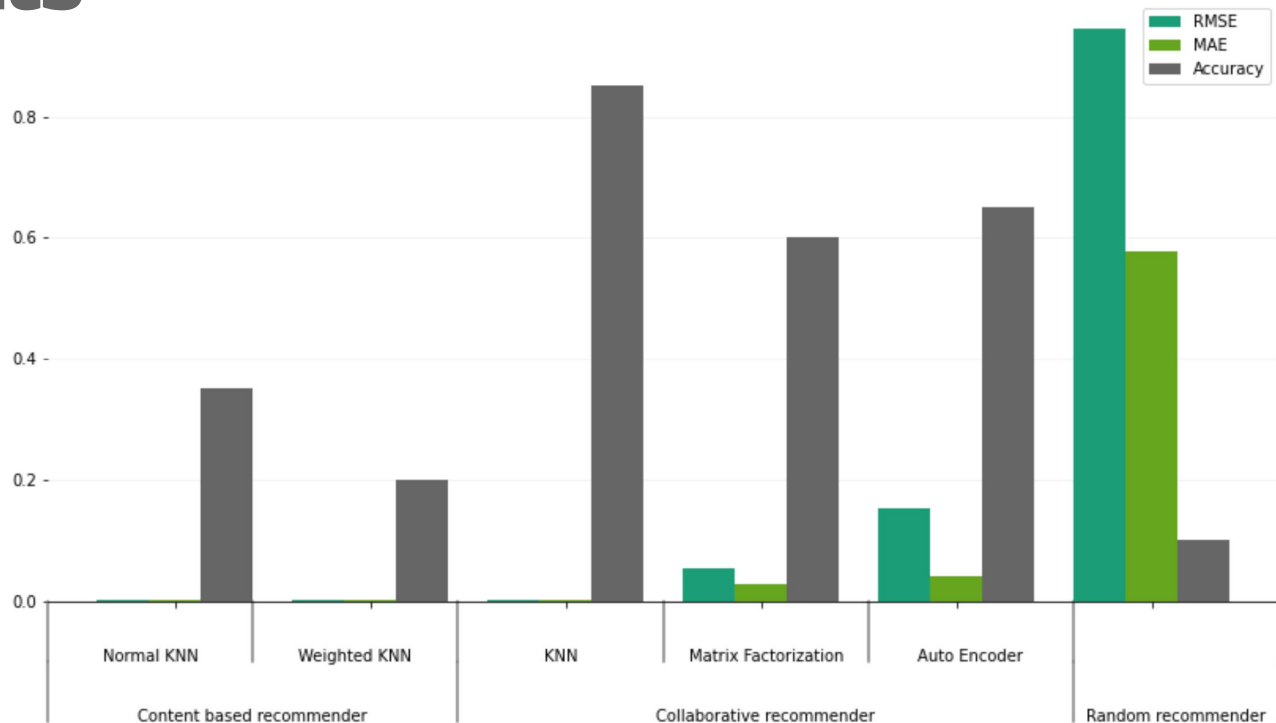


Figure 3: Which recommender Works best?

# Thank you



Follow me on spotify!

**El-Bahramino**