



Live Presentation - Music Recommendation System

By Dabidin Keshika
17 December 2022

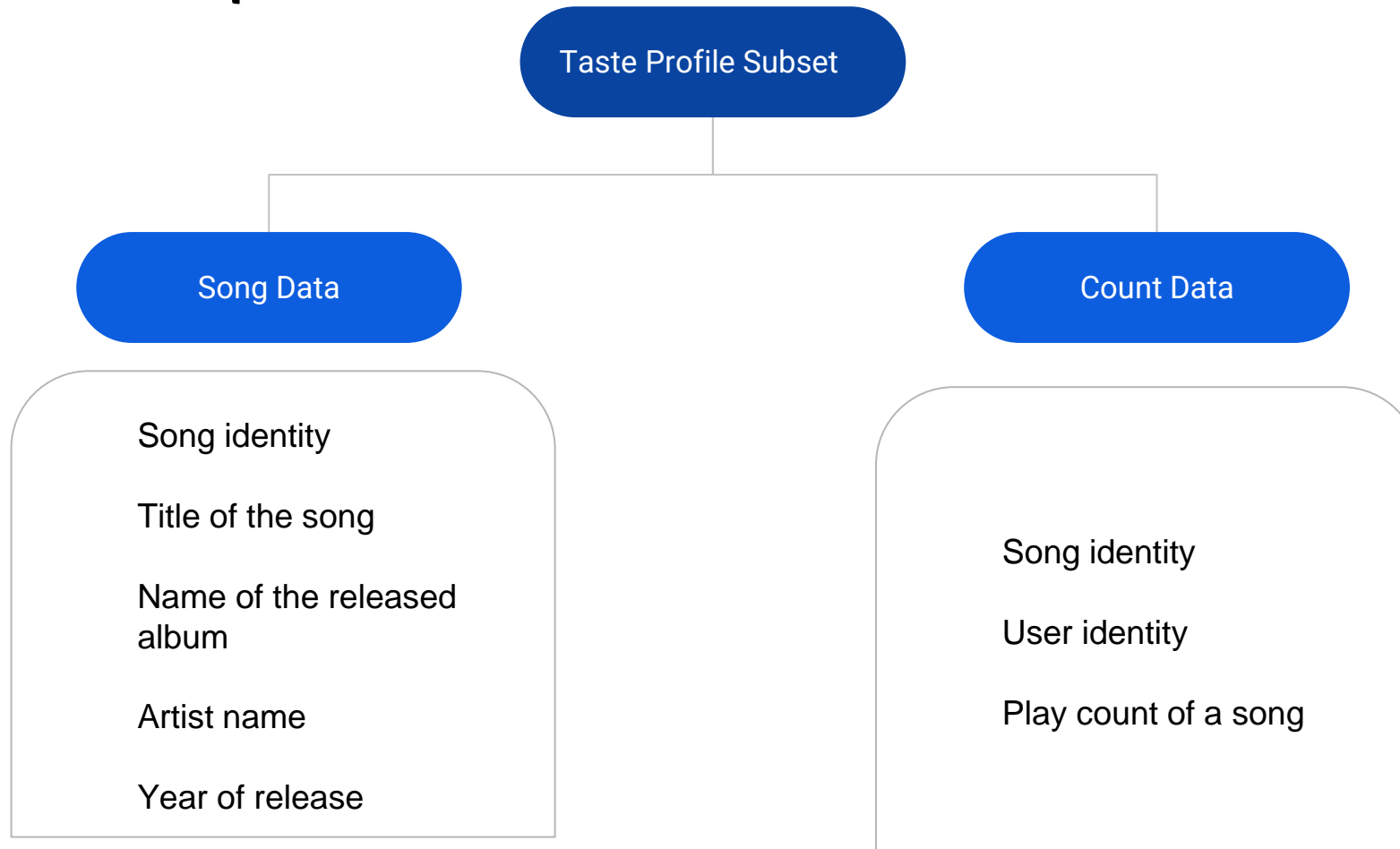
Executive Summary

- **Objective** : Build a music recommendation system for Spotify.
- **Dataset** : Million Song Data.
- **Four types** of algorithms tested : popularity, collaborative filtering, cluster based and content based.
- **Optimization** of the algorithms and performance evaluation.

Problem definition

- Increasing number of online companies are utilizing recommendation systems to increase user interaction and enrich business potential.
- The benefits of building a recommendation system is to namely :
 - ✓ Improve user retention and engagement
 - ✓ Understand changing trend of the customers' tastes
- Build the focus on the streaming music industry and develop an industry level music recommendation system under different scenarios and for different users.

Data Description



Data Source : Taste Profile subset released by the Echo Nest of the Million Song Dataset.

<http://millionsongdataset.com/>

Data Description

- 563 unique songs
- 232 unique artists
- 3155 unique users

= 1,776,265 play counts in the dataset.

- 117,876 play counts in the dataframe, i.e., not every user has listened to every song in the dataset.

→ possibility of building a recommendation system to recommend songs to the users which they have not interacted with.

Solution Design

- Different recommendation algorithms :

1. **Popularity Based -**

Finding the top n songs for a recommendation based on the average play count of songs.

1. **User-to-User and Item-to-Item Similarity Based Collaborative Filtering -**

Personalized recommendation system built with the user-item interaction.

1. **Matrix Factorization (Collaborative Filtering) -**

Personalized recommendation system based on the past behavior of the user.

1. **Cluster Based Filtering -**

Clustering similar users together and recommending similar songs to similar clusters.

1. **Content Based -**

Using the song features to extract information and build the recommendation system.

Solution Design

- Performance evaluation of the recommendation algorithms :
 1. **Precision@k** : fraction of recommended items that are relevant in top k predictions.
 1. **Recall@k** : fraction of relevant items that are recommended to the user in top k predictions.
 1. **F1-score@k** : harmonic mean of Precision@k and Recall@k.
 1. **RMSE** : Root Mean Squared Error.

Analysis and Key insights

- Popularity based recommendation system

1. Predict top n songs for example, top 10 songs of the day or the week.
2. Does not take into account customer's preferences.
3. Example :

Song Identity	Average Count	Playing Frequency
8582	1.948069	751
352	2.184492	748
2220	2.220196	713
1118	1.817221	662
4152	1.930982	652

Analysis and Key insights

- Collaborative filtering based recommendation system

Model	RMSE	F1_Score	Prediction for the song 1671 for the user 6958 of play count=2	Which one performed better between the optimized model and the baseline model?
User-user similarity-based baseline	1.0878	0.504	1.80	Baseline model
User-user similarity-based optimized	1.0596	0.482	1.53	
Item-item similarity-based baseline	1.0394	0.397	1.36	Optimized
Item-item similarity-based optimized	1.0328	0.489	1.96	
Matrix Factorization baseline	1.0252	0.498	1.27	Optimized
Matrix Factorization optimized	1.0141	0.502	1.34	



Analysis and Key insights

- Cluster based recommendation system**

Model	RMSE	F1_Score	Prediction for the song 1671 for the user 6958 of play count=2	Which one performed better between the optimized model and the baseline model?
Cluster Based baseline model	1.0487	0.472	1.29	Optimized model
Cluster Based optimized model	1.0654	0.465	1.91	



→ RMSE is lower in the item-item similarity optimized model. Hence, we will keep the latter.

Analysis and Key insights

- Content based recommendation system

Example :

The recommendations for the song with title 'Learn To Fly' are:

'Everlong', 'The Pretender', 'Nothing Better (Album)', 'From Left To Right', 'Lifespan Of A Fly', 'Under The Gun', 'I Need A Dollar', 'Feel The Love', 'All The Pretty Faces', and 'Bones'

→ **Features used : Song title, Artist Name and Release**

The songs recommended are of either the same group or of similar genre, for example, rock.

Conclusion

Business proposal

- Keep both the content-based recommendation system and the item-item similarity optimized recommendation system as both models would complement each other.
- Include a top 10 song category using the popularity-based recommendation.

Further Insights

- Make use of the original Million Song Dataset as well as other datasets to get other features and use it on the recommendations.

Some examples of features to extend the study would be the loudness of the song and the hating score of the song.

- Use more sophisticated implementation or try other levels of thresholds for optimization as well as trying new libraries.



Thank you for your
attention.

