

# Music Recommendation System

1<sup>st</sup> Disha Jain  
PESIUG20CS132

2<sup>nd</sup> Isha Adiraju  
PESIUG20CS169

3<sup>rd</sup> Juhi Bhattacharya  
PESIUG20CS182

**Abstract**—To analyze music currently being played and suggest similar music. The program suggests music based on user's listening habits and song's streams.

## I. INTRODUCTION

A music recommendation system is essentially a solution that allows music streaming platforms to offer their users relevant music recommendations in real-time. The goal of this project was to recommend songs to users based solely on their listening histories, with no information about the music. We applied various Filtering methods to achieve this: Popularity filtering, Content-Based filtering and Collaborative filtering.

## II. SCOPE

### A. What is within the scope?

- Recommendation system seek to predict or filter preferences according to user's choices. Useful model for various domains like e-commerce movies, videos, news, songs etc.
- We have employed ML algorithms and Data Analytics to implement the same for music recommendation.

### B. What is out of scope ?

The model will not be able to sense or empathise with the user's real emotions [using sentiment analysis] and sometimes this can be a limitation since the model can only recommend music based on the content filtering and popularity .

## III. DATASET AND FEATURES

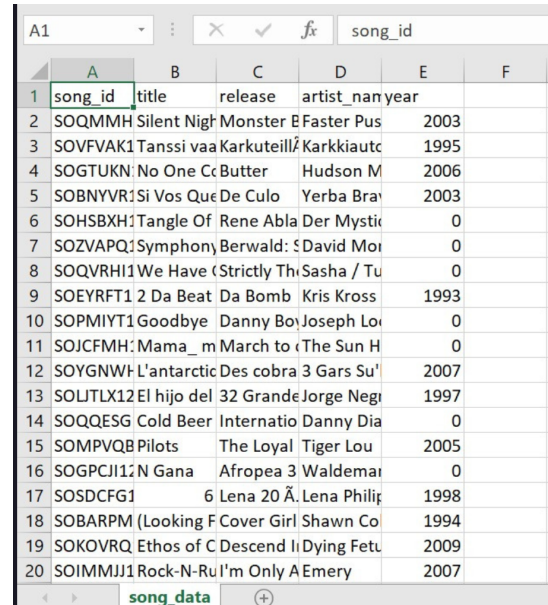
Dataset being used is Kaggle's Million Songs Dataset. Million Songs Dataset contains of two files: triplet\_file and metadata\_file. The triplet\_file contains user\_id, song\_id and listen time. The metadata\_file contains song\_id, title, release, year and artist\_name. Million Songs Dataset is a mixture of song from various website with the rating that users gave after listening to the song. The data is in the form of (user, song, play count) triplets. For example:

- (Isabelle, Hey Jude, 6)
- (Isabelle, Shake it Off, 12)
- (William, Whole Lotta Love, 15)
- (Emilien, Shake it Off, 9)

The training set contains 48 million such triplets, corresponding to 1.2 million users and 380K songs. The test set consists of 100K users and 157K songs.

### A. Pre-processing Of Data

- Two features are combined into a new feature called song
- Title and artist column can be eliminated for cleaner results
- Shortening the data set for quicker processing List of the most listened song in ascending order
- Percentage determines how much the song was listened by users in the data



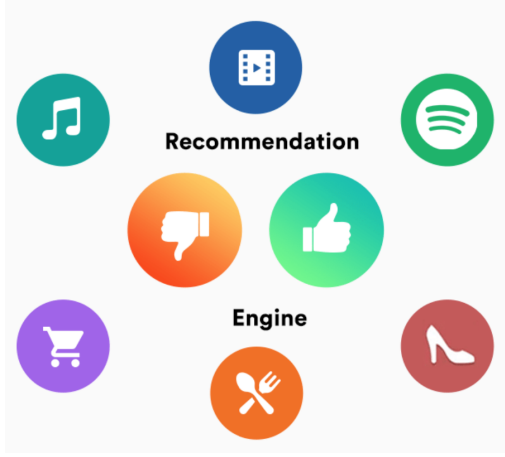
	A	B	C	D	E	F
1	song_id	title	release	artist_name	year	
2	SOQMMH	Silent Night	Monster	E Faster	Pus	2003
3	SOVFAK1	Tanssi vaa Karkuteilla	Karkkiaut			1995
4	SOGTUKN	No One Can Butter	Hudson M			2006
5	SOBNYVR1	Si Vos Quer De Culo	Yerba Br			2003
6	SOHSBXH1	Tangle Of	Rene Ab	Der Mysti		0
7	SOZVAPQ1	Symphony	Berwald: S	David Mo		0
8	SOQVRH1	We Have (Strictly Thi	Sasha / Tu			0
9	SOEYRFT1	2 Da Beat	Da Bomb	Kris Kross		1993
10	SOPMIYT1	Goodbye	Danny Boy	Joseph Lo		0
11	SOJCFMH	Mama_m	March to (The Sun H			0
12	SOYGNW1	L'antarctic	Des cobra	3 Gars Su'		2007
13	SOLJTLX1	2 El hijo del	32 Grande	Jorge Neg		1997
14	SOQQESG	Cold Beer	Internatio	Danny Dia		0
15	SOMPVQB	Pilots	The Loyal	Tiger Lou		2005
16	SOGPCJ1	12 N Gana	Afropea 3	Waldemar		0
17	SOSDCFG1		6 Lena 20 A	Lena Philip		1998
18	SOBARPM	(Looking F	Cover Girl	Shawn Co		1994
19	SOKOVRQ	Ethos of C	Descend I	Dying Fetu		2009
20	SOIMMJ1	Rock-N-Ru	I'm Only A	Emery		2007

## IV. ALGORITHMS USED

We used three different types of Filtering methods: Popularity filtering, Content-Based filtering and Collaborative filtering.

### A. Popularity Filtering

For popularity filtering, a simple approach involving arranging the songs in ascending order according to their listen count was used. This just suggests the top songs based purely on number of listen counts.



### B. Content Based Filtering

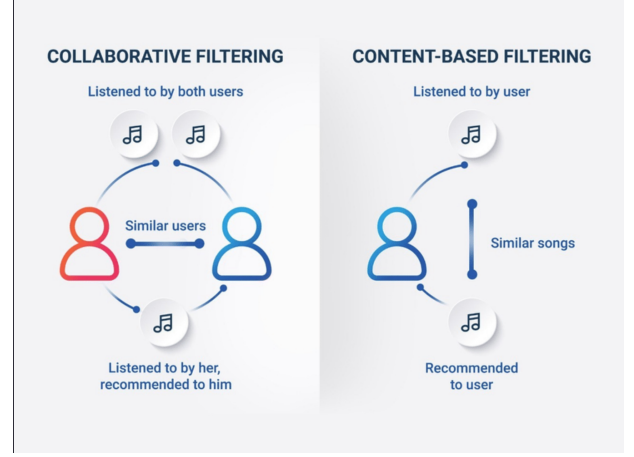
For Content Based Filtering, we have chosen to use the Singular Value Decomposition (SVD) Algorithm. SVD is basically a matrix factorization technique, which decomposes any matrix into 3 generic and familiar matrices. Specifically, the singular value decomposition of an  $m \times n$  complex matrix  $\mathbf{M}$  is a factorization of the form,

$$\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*, \quad (1)$$

where  $\mathbf{U}$  is an  $m \times m$  complex unitary matrix,  $\mathbf{\Sigma}$  is an  $m \times n$  rectangular diagonal matrix with non-negative real numbers on the diagonal,  $\mathbf{V}$  is an  $n \times n$  complex unitary matrix, and  $\mathbf{V}^*$  is the conjugate transpose of  $\mathbf{V}$ .

### C. Collaborative Filtering

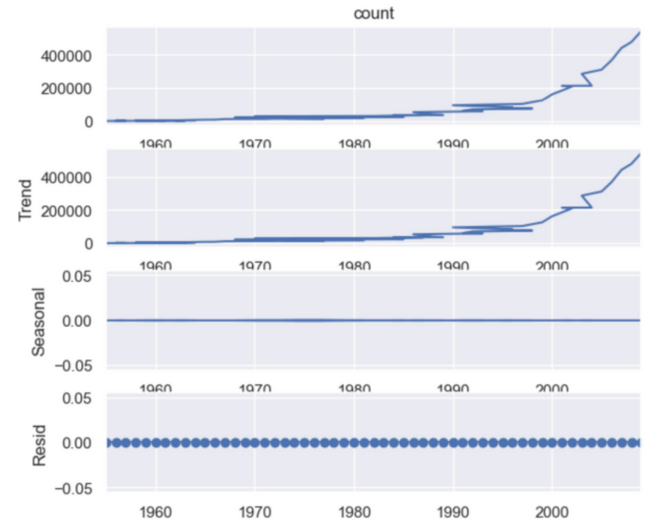
For collaborative filtering KNN was used. k-nearest neighbors algorithm is a non-parametric supervised learning method. It uses proximity to make classifications or predictions about the grouping of an individual data point. In this algorithm the input parameter was the user for which you want recommendations. KNN eliminates all the unique songs listened to by the user against other users since the only parameters that contribute are the common songs listened to. The algorithm checks common songs listened to against other users through an iterative approach. In the end it returns the same set of common songs in order.



## V. NOVELTY

For this project, we have chosen the ARIMA model as our novelty function. An ARIMA model (Autoregressive Integrated Moving Average) is a statistical analysis model. It uses the concept of time series data in order to better understand the data set or to predict future trends. ARIMA models use differencing to convert a non-stationary time series into a stationary one, and then predict future values from historical data. These models use “auto” correlations and moving averages over residual errors in the data to forecast future values.

In this project, we require such a forecasting model in order to predict future trends in music, as well as future personal recommendations to the users based on their history. The ARIMA model was first trained using the existing dataset, ie. until 2009 and then made to predict future music trends from the years 2010-2024.



In following example , it shows the graphs of previous trends and then the ARIMA model predicting trends for the coming years from 2010 to 2019.

```
forecat_10.head(10)
```

	year	listen_count
0	2010	6.051503e+05
1	2011	6.517776e+05
2	2012	7.230190e+05
3	2013	7.814937e+05
4	2014	8.383506e+05
5	2015	9.099236e+05
6	2016	9.691804e+05
7	2017	1.034339e+06
8	2018	1.105608e+06
9	2019	1.168523e+06

## VI. RESULTS AND ANALYSIS

- Output for Popularity filtering

	artist_name	score	Rank
649	Coldplay	29422	1.0
2850	The Black Keys	19862	2.0
1651	Kings Of Leon	18747	3.0
1107	Florence + The Machine	18112	4.0
1370	Jack Johnson	17801	5.0
2946	The Killers	16063	6.0
2374	Radiohead	14890	7.0
736	Daft Punk	14715	8.0
2073	Muse	14005	9.0
1554	Justin Bieber	13959	10.0
3174	Train	12535	11.0
978	Eminem	12138	12.0
1768	Linkin Park	11579	13.0
2205	OneRepublic	11476	14.0

- Output for Content-Based filtering

```
-----
Recommendation for user id 5
The number1 recommended song is 32603      Dungeon Master
Name: title, dtype: object BY 32603      EPMD / Nocturnal
Name: artist_name, dtype: object
The number2 recommended song is 19982      Breakfast At Tiffany's
Name: title, dtype: object BY 19982      Deep Blue Something
Name: artist_name, dtype: object
The number3 recommended song is 55866      Cat And Mouse
Name: title, dtype: object BY 55866      The Red Jumpsuit Apparatus
Name: artist_name, dtype: object
The number4 recommended song is 4931      Propane Nightmares
Name: title, dtype: object BY 4931      Pendulum
Name: artist_name, dtype: object
The number5 recommended song is 1967      Scatman (Ski-Ba-Bop-Ba-Dop-Bop)
Name: title, dtype: object BY 1967      Scatman John
-----
```

- Output for Collaborative filtering

```
recommendations *****
-----
              title      release artist_name
834983  Quiet Houses  Fleet Foxes  Fleet Foxes
-----
              title      release artist_name
939863  Meadowlarks  Fleet Foxes  Fleet Foxes
-----
              title      release artist_name
955922  Heard Them Stirring  Fleet Foxes  Fleet Foxes
-----
              title      release artist_name
102372  Tiger Mountain Peasant Song  Fleet Foxes  Fleet Foxes
-----
              title      release artist_name
333174  Sun It Rises  Fleet Foxes  Fleet Foxes
-----
              title      release artist_name
777445  Your Protector  Fleet Foxes  Fleet Foxes
-----
              title      release artist_name
650414  Oliver James  Fleet Foxes  Fleet Foxes
-----
              title      release artist_name
114578  Great Indoors  Room For Squares  John Mayer
-----
              title      release artist_name
605234  White Winter Hymnal  White Winter Hymnal  Fleet Foxes
-----
              title      release artist_name
```

## REFERENCES

- [1] Year of Publication: 2019 ,Authors: Namitha S J
- [2] Year of Publication: CME, MSE, Stanford University (2017) Authors: Emilien Dupont, Isabelle Rao, William Zhang
- [3] Year of Publication: Indian Institute of Technology, Kanpur (2020) Authors: Shefali Garg Fangyan SUN
- [4] Year of Publication: Dept. of Information Technology, PCE, Navi, 2021 Authors: Varsha Verma, Ninad Marathe, ParthSanghavi, Dr. Prashant Nitnaware