An Automated Platform for evaluating the factors related to Music Recommendation System

1st A Sheik Abdullah
VIT Chennai
Tamil Nadu, India
aa.sheikabdullah@gmail.com

2nd MK Thamaraimanian
VIT Chennai
Tamil Nadu, India
mktham568@gmail.com

3rd R Priyadarshini
VIT Chennai
Tamil Nadu, India
rspdarshini@gmail.com

4th Altrin Lloyd Hudson D
VIT Chennai
Tamil Nadu, India
altrinlloyd2002@gmail.com

5th Naga Pranava Shashank V
VIT Chennai
Tamil Nadu, India:
pranavanaga@gmail.com

***Abstract*—Listening to music has become one of the most frequently resorted to pastimes of people ranging from theyouth to the elder. While there are umpteen songs of different genres and artists from yesteryears in the podcast, it becomes essential that there is a recommendation System that analyzes the liking of a specific user with the help of the datasets genre and artists of the songs that he/she listened to in the past three days. The goal of this project is to create a system for recommending music that will analyze user interactions with the app or music platform in order to establish their musical preferences. Our system learns from users' previous listening history and recommends music they want to listen to in the future. Currently music service providers have generic, mood-based playlists, that are the same for all users. Here, we suggest improvements to these playlists by offering custom playlists for each user based on user input. Rich web application technologies have proliferated as a result of the rise in Internet usage as a source of information. Users can use these devices to listen to music without having to download it to their PC. Some people additionally employ their preferred methods to enhance the user experience.**

Keywords—1. MRS-Music Recommendation System 2. YT-YouTube 3. CF: -Collaborative Filtering 4. CBM: -Content-Based Model 5. EDM: -Electronic Dance Music 6. CART: -Classification Tree

I. INTRODUCTION

The main purpose of this study is to code a music recommendation system or shortly we call it as MSR, which helps recommend songs to the user based on his previous listening or interactions with the app or the music website. The recommendation of songs is mostly based on the analysis obtained from the users account from his past listening and liking to a particular type of music. Everyone has their own favorite album or band or a singer but users love to explore and always expect something new and which is suitable to their likings.

Music recommendation system is one of the zones where mere reading of the research papers gave us tremendous insights on latest trends, uses, disadvantages of both corroborative filtering as well as content-based filtering in the long run. We also understood the mechanism adopted by Spotify, YouTube Music, Gaana and other music platforms to provide quality music to their dedicated listener base and the algorithm adopted by them using Artificial Intelligence and their corresponding Machine vision. This worked as a case study towards the designing and structuring of our own Music recommendation System. The algorithms adopted by us and the statistical analysis pertaining to those have been furnished in the latter part of our report).

So, while using these filtering methods we will be able to produce music recommendation System that is robust innature as well as having the ability to satisfy the listener and provide him with varied music according to his aspirations. This is done using the K cluster algorithm that we have displayed in the forthcoming pages. We have taken many literary research papers written about Music recommendation systems in order to see and read the new perspectives offered by various authors in order to understand the efficiency of the MRS generated. We have also published our thoughts and reviewed the research papers.

II. PROPOSED METHODOLOGY

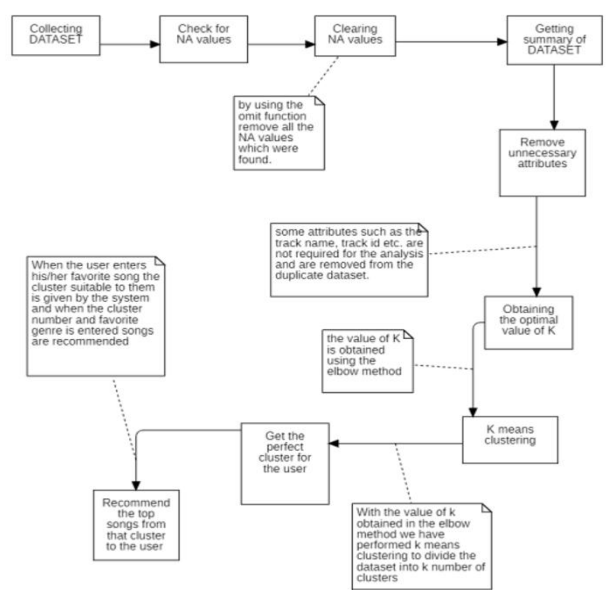

Fig. 1. Proposed methodology diagram

III. EXPERIMENTAL RESULTS AND DISCUSSION

A. K. Visvam Devadoss et al. (eds.), *Proceedings of the 6th International Conference on Intelligent Computing (ICIC-6 2023)*, Advances in Computer Science Research 107,
https://doi.org/10.2991/978-94-6463-250-7_2

A. *Elbow Method*

To recommend song to a user we have divided the whole data into clusters, But the problem is to determine how many clusters should we divide the data into, to solve this issue wehave used to elbow method to find the optimal number of clusters to be used, we have obtained a graph to check where the elbow is occurring the graph is given below.

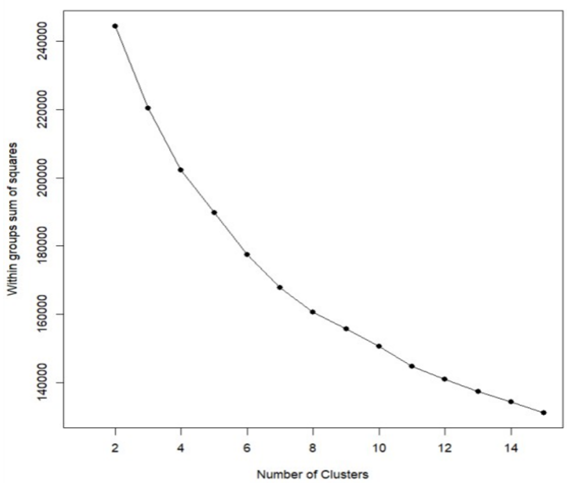

Fig. 2. WCSS vs Number of clusters

We have taken the max cluster size to be 15 but we couldn't find any elbow and after few trails and errors we have come to a conclusion that there no proper elbow, but the graph slightly tilts or becomes like an elbow at the value 7.

B. *K-means Clustering*

- So, we have taken the value of k=7 to continue with our k-means clustering which will further help us with the song recommendation. With k means clustering algorithm we have divided the dataset into 7 cluster, each cluster has some unique properties and is slightly different from the others.
- If a user's favorite song and artist is in a particular cluster than it is taken into assumption that he will like the songs which are there in that cluster in any particular genre.
- After we have divided the dataset into 7 clusters, we just have to enter the user's favorite song and artist so that the system will generate the suitable cluster for the user once the cluster is generated the user can give the cluster and genre to get the top 5 songs recommended by the system.
- For example, we can give the song memories by and artist maroon 5 the system has given us the cluster number 5, when we entered the cluster number and a genre the top 5 songs belonging to that cluster in the requested genre are suggested by the user.

IV. STATISTICAL INTERPRETATION AND ANALYSIS

The dataset which we are using now is the Spotify songs dataset released in the year 2021 and was taken from Kaggle,this dataset has the attributes such as:

1. Track_id
2. Track_name
3. Track_artist
4. Track_popularity
5. Track_album_id
6. Track_album_name
7. Track_album_release_date
8. Playlist_name
9. Playlist_id
10. Playlist_subgene
11. Playlist_genre
12. Danceability
13. Energy
14. Key
15. Loudness
16. Mode
17. Speechiness
18. Acousticness
19. Instrumentalness
20. Liveness
21. Valence
22. Tempo
23. Duration_ms

A total of 23 attributes are there in the dataset. We have used the glimpse() function to find out the number and what attributes are present in the dataset. We used the colSums(is.na()) function to find out the number of NA values the dataset there are 5 NAs in track_album_name, track_name and track_artist. We have removed the NA rows from the dataset to avoid unnecessary errors during further coding. The duration of songs which was given in the dataset is in milli seconds we have change it to minutes because most of the songs now when compared in terms of duration are taken into account as minutes and converting milliseconds to minutes is more sensible and easily understandable. We have separated the data into four sets based on the popularity, we have taken the track_popularity column in the dataset and separated as four groups.

Group 1: It has all the song with popularity between 0 and 20.

Group 2: It has all the song with popularity between 20 and 40.

Group 3: It has all the song with popularity between 40 and 60.

Group 4: It has all the song with popularity more than 60.We have found that there are a total of 4182 songs in group 1, 6162 songs in group 2, 8975 songs in group 3 and a total of 9033 songs in group 4.

Here is the summary of the major attributes required to suggest a song to a user.

TABLE I. SUMMARY STATISTICS

Attributes	Summary statistics					
	Minimum	*1st Quartile*	*Median*	*Mean*	*3rd Quartile*	*Maximum*
Popularity	0.0	21.0	42.0	39.34	58.0	100.0
Danceability	0.0	0.5610	0.670	0.6534	0.76	0.9830
Energy	0.000175	0.5790	0.722	0.698372	0.8430	1.0
Loudness	-46.448	-8.310	-6.261	-6.818	-4.709	1.275
Instrumentalness	0.0	0.0	0.0000207	0.0911294	0.0065725	0.994

V. VISUALIZATION

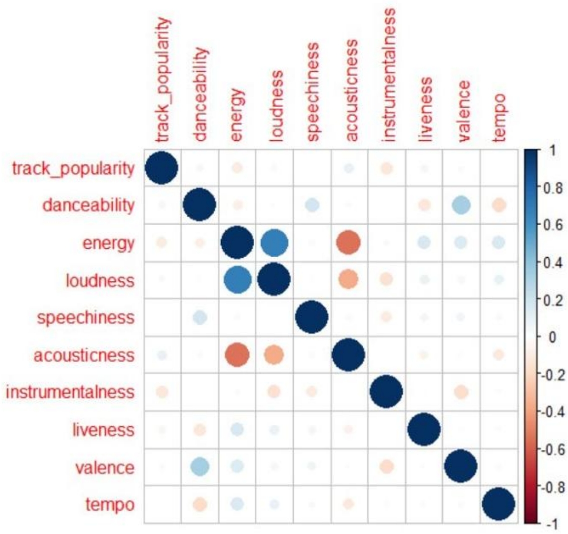

Fig. 3. Correlation plot

The plot in Fig. 3 shows that recognition and track functions are not properly correlated. The fact that several variables exhibit significant correlations with one another, however, suggests that this dataset is multicollinear.

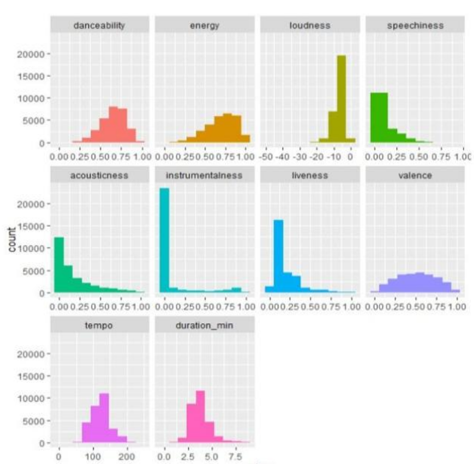

Fig. 4. Histogram

From the histograms in Fig. 4, we infer that:

- The users are mostly listening to songs of length not more than 3-4 minutes in an average, with the increase in the duration of the song the listener count is decreasing.
- Valence is typically allotted.
- Danceability and power are almost typically allotted.
- Most of the songs have a loudness level of -5dB.

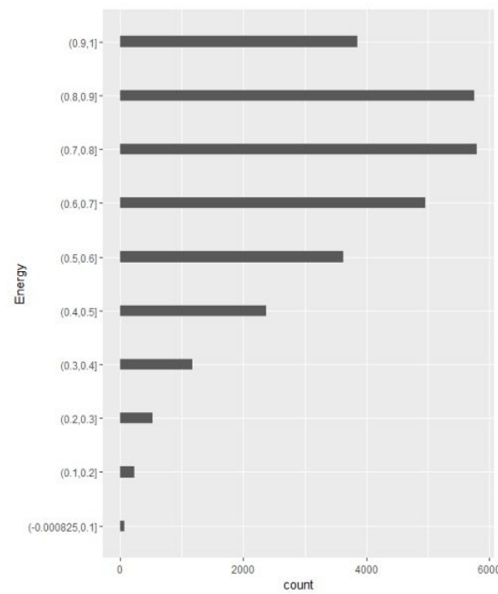

Fig. 5. Energy Distributions

From the energy histogram in Fig. 5, we proved that better energy songs are favored more by means of Spotify listeners.

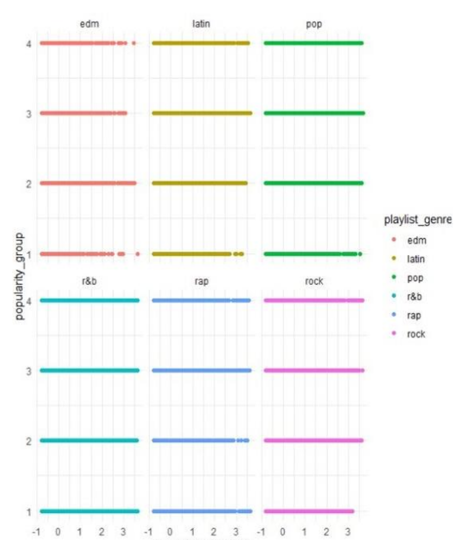

Fig. 6. Popularity by acousticness

From Fig. 6, we infer that acousticness does not affect track recognition as the extent of acousticness has been same across all reputation levels.

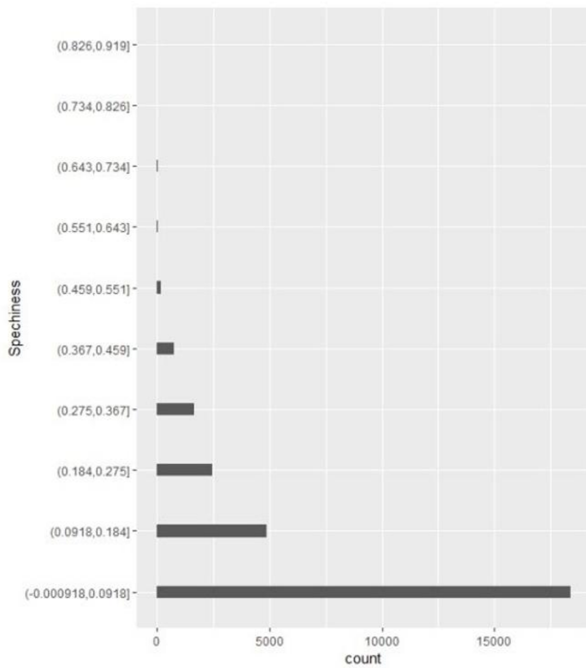

Fig. 7. Speechness distribution

The graph in Fig. 7 gives the distribution of Speechness. We know that we don't like songs that are speechier, this confirms our belief that most Spotify listeners love songs that are less speechy. That's the reason why Spotify doesn't keep many interesting songs in its database.

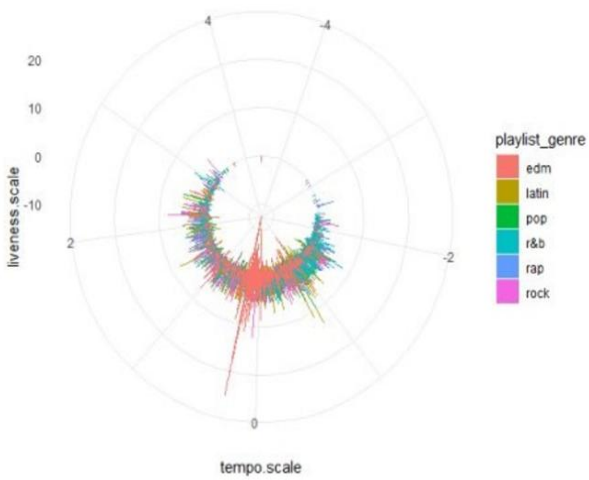

Fig. 8. Tempo and liveliness distribution across genre

Fig. 8 shows that EDM has far better tempo than the other genres, although liveness is virtually evenly distributed among all genres.

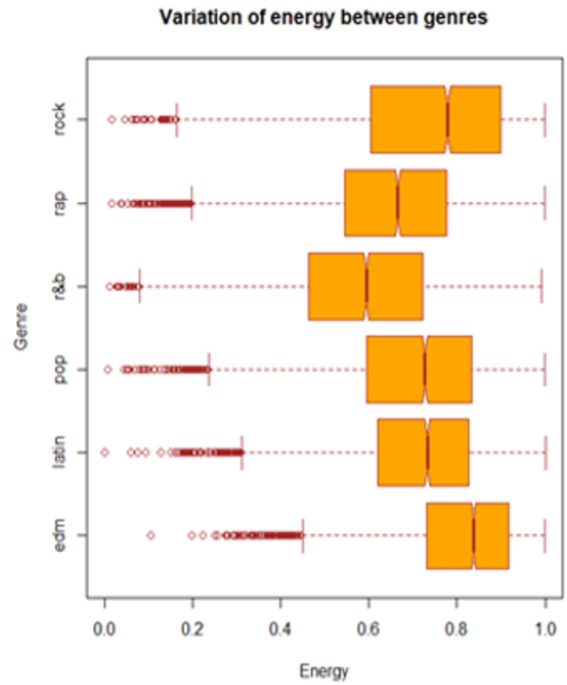

Fig. 9. Boxplot for Genre by Energy

The Fig. 9 represents the boxplot for Genre by Energy and from the graph we can infer that EDM has the highest energy.

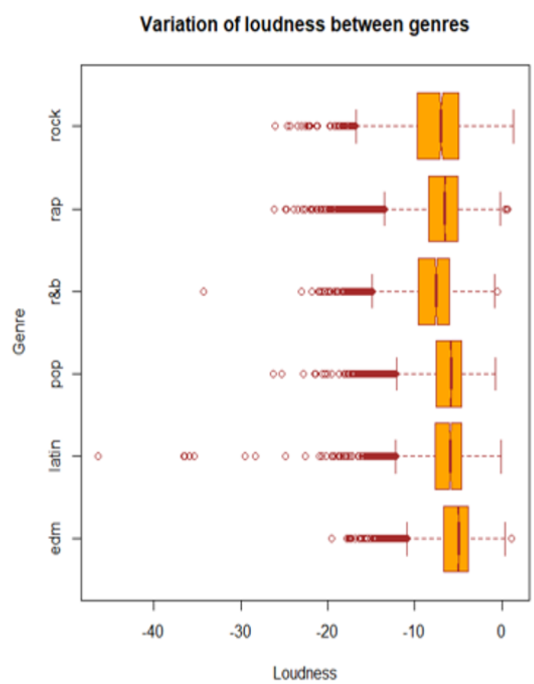

Fig. 10. Boxplot for Genre by Loudness

The Fig. 10 represents the boxplot for Genre by Loudness and from the graph we can infer that EDM genre are louder in nature.

VI. CONCLUSION

We have tried our best to explore various disciplines that are connected to the music recommendation system and have to simplify and explain the same. Our algorithm is robust and diverse in order to be considered as an efficient MRS. We have given our perspective by reviewing many research papers and have furnished our views about the same. We have provided the proposed methodology and explained the basis of our design. We have analyzed the results that were obtained from various experiments and published them. The statistical analysis involving sample test cases and our interpretation of them were presented in a streamlined manner. This enabled us to visualize various results that were obtained through graphs and charts. The visualization analysis was also presented in distinct and colorful way. In this way we have designed the intricate aspects of the MRS. From our analysis we have come to the conclusion that random forest and classification tree cannot be used for the dataset because of the existence of multicollinearity in it.

In this project, we are designing and implementing a music recommendation system. We use Kaggle's dataset to find relationships between users and songs, and learn from users' past listening histories and interactions with music platforms to provide feedback to users. Our system will recommend songs based on popularity, based on user - user similarity i.e., choose songs that similar users listen to, based on past history i.e., recommend songs that has been played many times by the user in a selected time span and also based on the input data about some favorite artists of a user. Our system helps users to find out new artists, albums or songs according to their likings which helps them make their own custom playlists more exciting.

REFERENCES

[1] Dushani Perera , Maneesha Rajaratne and Kasun Karunanayaka,"A Critical Analysis of Music Recommendation Systems and New Perspectives", Human Interaction, Emerging Technologies and Future Applications II , University of Colombo School of Computing , April 2020.

[2] Yading Song ,Simon Dixon.., "A Survey of Music Recommendation Systems and Future Perspectives ",The 09th International Symposium on Computer Music Modeling and Retrieval (CMMR) London, UK , June 2012.

[3] Varsha Verma ,Ninad Marathe ,Parth Sanghavi and Dr. Prashant Nitnaware , "Music Recommendation System Using Machine Learning" ,International Journal of Scientific Research in Computer Science Engineering and Information Technology ,November 2021

[4] Hamed Zamani, Ching-Wei Chen, Yashar Deldjoo and Mehdi Elahi , "Current challenges and visions in music recommender systems research", April 2018

[5] Aldiyar Niyazov , Elena Mikhailova and Olga Egorova , "Content-based Music Recommendation System " , 2021 29th Conference of Open Innovations Association (FRUCT) , May 2021.

[6] Yading Song , Marcus Pearce and Simon Dixon , " A Survey of Music Recommendation Systems and Future Perspectives ", 9th International Symposium on Computer Music Modeling and Retrieval (CMMR) , London, UK , June 201