1

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

School of Computer Science and Engineering
J Component report

Programme : B.Tech
Course Title : Foundations of Data Analytics
Course Code : CSE3505
Slot : F2

Title: Music Recommendation System

Team Members: Sri HarshaVardhan Palla 20BCE1308
Rukksana A 20BCE1968
Siddharth M 20BPS1007

Faculty: Dr. Sheik Abdullah A

Sign:

Date: 18/11/2022

INDEX

Abbreviations Used:-

1.MRS:-Music Recommendation System

2.YT:-Youtube

3.CF:-Collabrative Filtering

4.CBM:-Content-Based Model

5.CBF:-Content-Based Filtering

6.NA:- not applicable

7.EDM:-Electronic Dance Music

8.CART:-Classification Tree

List of Figures:-

ABSTRACT

Listening to music has become one of the most frequently resorted to pastimes of people ranging from the youth 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. This project aims to envision a music recommendation system with such a function. Therefore there is a need for a good recommendation system. The objective of this project is to develop a music recommendation system which will determine the musical preferences of the users based on the analysis of their interaction during use . Our system learns from the users past listening history and recommends them songs which they would probably like to hear in 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 .

The growth of technology and resources have led to the development of rich webpages and web-applications. Among these applications are present the music service providers are spotify, amazon music, gaana, etc. . Theses music providers allow users to listen music online and provide good music recommendations with the help of their music recommender engines to enhance the user experience.

INTRODUCTION

The main objective of this project is to develop a music recommendation system. The system should able to estimate what artist or group would match user preferences to the user at a given time. It is a fact that we do not always want to hear the same artists or genres or the same tracks again and again, we do have favourite songs and artists, but sometimes we appreciates a new song.

This system helps users find new artists, albums or songs making the musical catalogue available for the user according to his preference. The user will receive song recommendations related to his/her interaction with the system.

Music recommendation systems are one of the most used systems by music listeners nowadays. In that sense there is a need for the listener to be aware of different kinds of music that suits his interests in order for him to pick and choose the kind of music that he wants to listen. He must be able to filter out the content based on his preferences. In order to do that, we have use filtering method. The music recommendation system that we have designed will incorporate majorly two types of filtering namely corroborative filtering and content based filtering. The former is based on suggestion that pertaining to individual preferences and likings and his past interactions and the latter is based on general hits and likes by the audience involving trending songs or songs that have the most no of clicks etc.

So while using these filtering methods we will be able to produce music recommendation System that is robust in nature 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.

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 what is the algorithm adopted by their system 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.

LITERATURE REVIEW

1.Music Recommendation System -Sj Namitha(2019)

This research paper talks in detail about how big data technology has made it possible that music listeners could get access to music as they want. The advancement of cloud technologieseases users to get access to an unlimited number of songs.

This paper generically describes that big data analysis is the propeller to the varied MRS that exists today. The advanced technologies have been incorporated in the latest streaming music companies like YT music,Spotify etc.

2.Learning in Music Recommendation systems.Marcus Schedl (2019)

From this research paper we come to know that there are various kinds of Research on music recommendation systems (MRS) is spiraling.

This paper observes that the research on MRS is spiraling with the help of deep learning. It calls for neural network to be increasingly adopted into MRS keeping in mind its potential.

3. Current Visions in Music Recommendation systems-Hamed Zamani 2018

The author of this research paper remarks that while today’s MRSs considerably help users to find interesting music in these huge catalogs, MRS research is still facing substantial challenges

4.Content based MRS Niyazov(2021)

This research paper discusses 2 different approaches involved in building content based MRS. One using acoustic analysis w.r.t the notes of the songs and the other involving deep learning to improve the results obtained

5.Survey of MRS Yading Song(2020)

This paper surveys a general framework and state of art approaches for recommendation of music. Context based model and emotional model,this paper says has been paid increasing attention in recent years for being able to connect to the users directly.

6.Music system preference -Varsha Verma(2021)

This paper analyses the reasons behind the increasing use of content based interactions in recent years compared to the traditional corroborative filtering as it exploits interactions between users and items like clicks or ratings

7.Music system using neural networks-Dushan perera(2020)

This model argues that beyond deep learning there are times when MRS uses cosine similarity of extracted data from one music vector and another. It is generally obtained by considering 2 different music sources from different genres and their feature vectors.

8.Critical analysis of MRS -T Bharadwaj(2021)

This author argues that the preferences of an individual varies depending on his mood,situation,context,pattern and other external factors. The paper states that there's some amount of fluidity and flexibility in cases pertaining to an individual's preferences in MRS and that has to be kept in mind while designing it.

9.Exploring MRS M Jitendra(2021)

The author exemplifies that recommendations are classified into two types based on the number of users the system suggests to. When the interest of a user is taken into account to provide the suggestion for that one particular user, then it is called as the personal recommender system. Since a personal taste changes for different users, this type of recommenders is used for a single user for his taste. The other type of recommender is the public recommender. When a system takes into account the interests of multiple users to make recommendations.

This paper argues that there's a variety in approach towards personal MRS and a MRS that designed as per public demand and recommendations. It is impossible for an individual to comply with the public demand at all times.When a system grabs the interests of users on a large scale, the popularity of item plays a major role in recommendations.

10.MSR - Dushani Perera(2020)

This paper talks in detail about hybrid filtering which combines the individual advantages of each content based and corroborative filtering. It says that the challenge of implementing this lies in the data collection as large amounts of data would be needed for the implementation.

PROPOSED METHODOLOGY

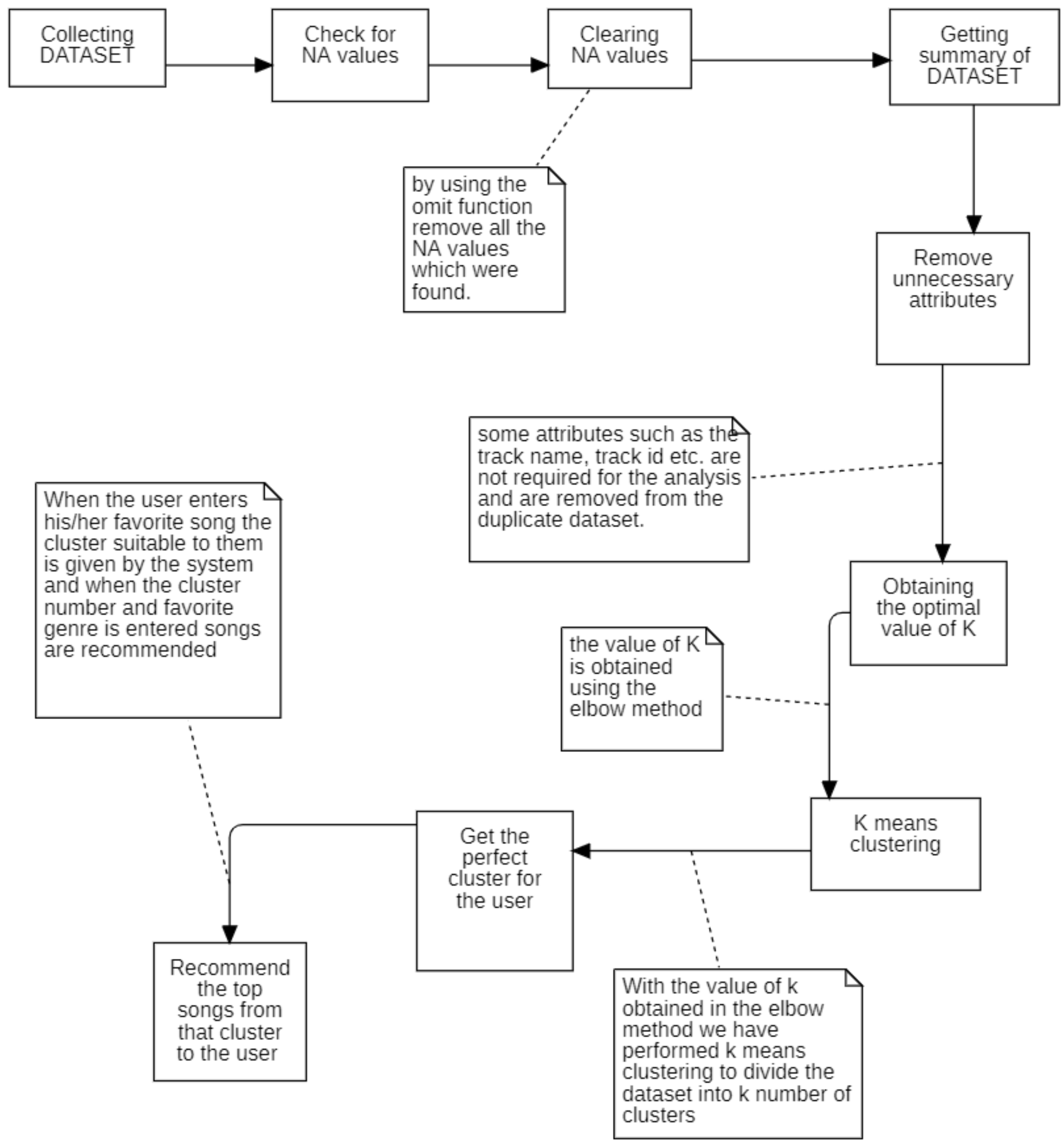

Fig 0.1

EXPERIMENTAL RESULTS AND DISSCUSSION

The dataset which we are using now is the spotify_songs dataset released in the year 2021 and was taken from kaggel.

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 we have 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(Fid 1.1).

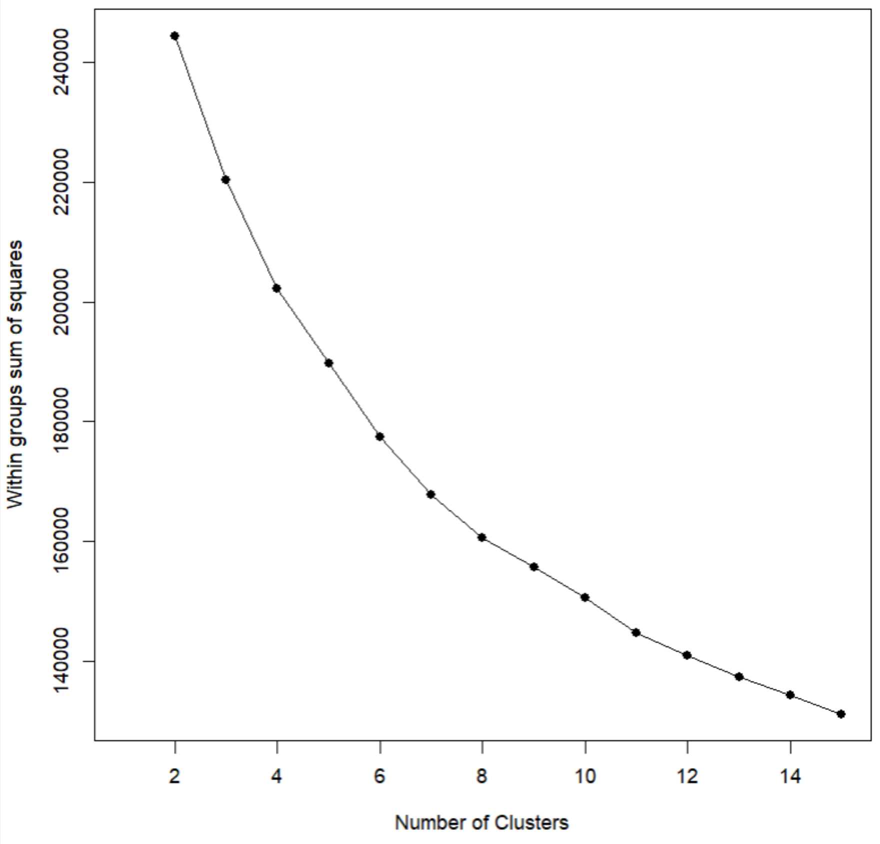

Fig 1.1

- We have taken the max cluster size to be 15 but we couldn’t fine any elbow and after few trail 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.

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 users favourite 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 users favourite 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 given the song memories by and artist maroon 5 the system has given us the cluster number 3, 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.

Sample Result:-

We have give the favourite song of the user as track_name == "Memories - Dillon Francis Remix", track_artist == "Maroon 5"

Popularity group	Cluster
4	3

Table 1.1

When we have given the cluster number as 3 and genre as r&b the top 5 songs of the genre in cluster 3 are displayed.

STATISTICAL ANALYSIS AND INTERPRETATION

The dataset which we are using now is the spotify_songs dataset released in the year 2021 and was taken from kaggel, 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_subgenre
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 hav used the glimpse() function to findout the number and what attributes are present in the dataset, we hav also found that there are more than 32000 observations.

We used the colSums(is.na()) function to findout the number of NA values the dataset there are 5 NAs in track_name, track_artist, and track_album_name. 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 ms(milli seconds) we have change it to mins cause most of the songs now when compared in terms of duration are taken into account as minutes and coverting ms to mins 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 hav 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:-

Popularity:-

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.00	21.00	42.00	39.34	58.00	100.00

Danceability:-

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0	0.5610	0.6700	0.6534	0.7600	0.9830

Energy:-

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.000175	0.579000	0.722000	0.698372	0.843000	1.000000

Loudness:-

Min. 1st Qu. Median Mean 3rd Qu. Max.

-46.448 -8.310 -6.261 -6.818 -4.709 1.275

Instrumentalness:-

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.00 0.0000 0.0000207 0.0911294 0.0065725 0.9940000

VISUALIZATION ANALYSIS

Correlation plot:-

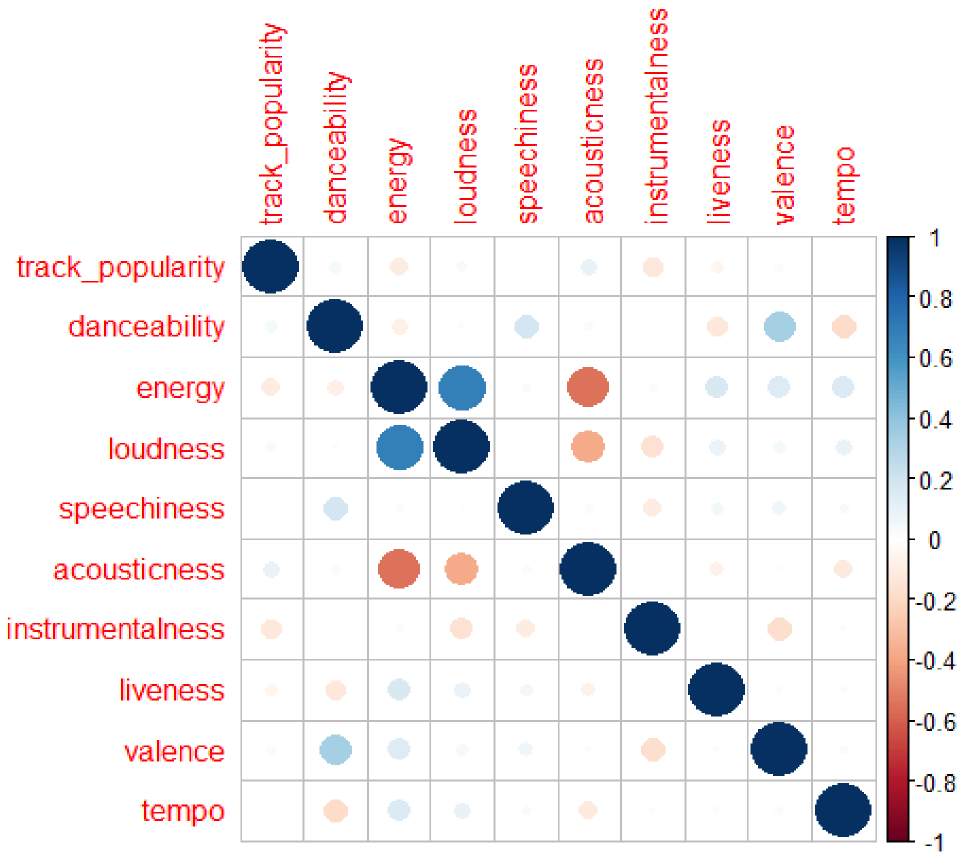

Fig 2.1

The plot indicates recognition does now not have sturdy correlation with track functions. However we determined some variables have strong correlation with each other, indicating that this dataset has multicollinearity and may not be suitable for various type algorithms.

Histogram:-

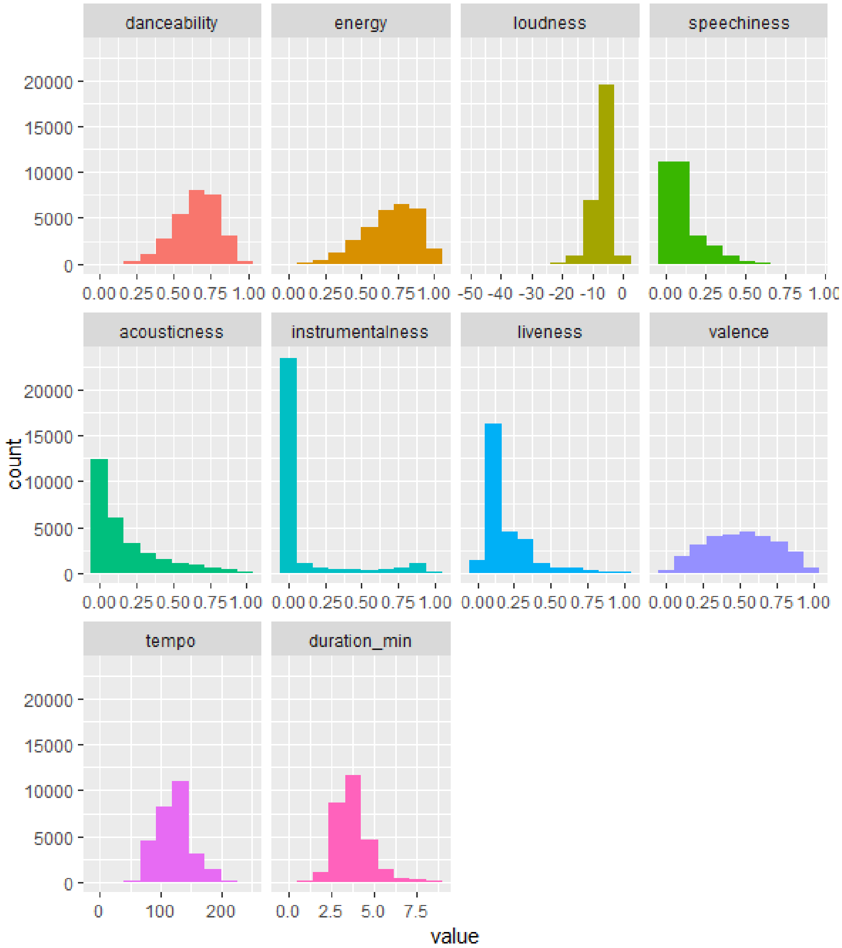

Fig 2.2

From the histograms, we are able to look at that:

- Majority observations have a fee no larger than 0.1 in instrumentalness, and that is the purpose why the distinction among suggest and median of instrumentalness is pretty big
- Majority of songs listened to have a length of approximately 3-4 mins with songs longer than that period having decrease frequency of listeners

- Valence is typically allotted
- Danceability and power are almost typically allotted
- Most of the tracks have a loudness of -5dB
- Majority tracks have speechiness index much less than 0.2 indicating that less speechy songs are greater favoured with the aid of listeners

Energy Distribution:-

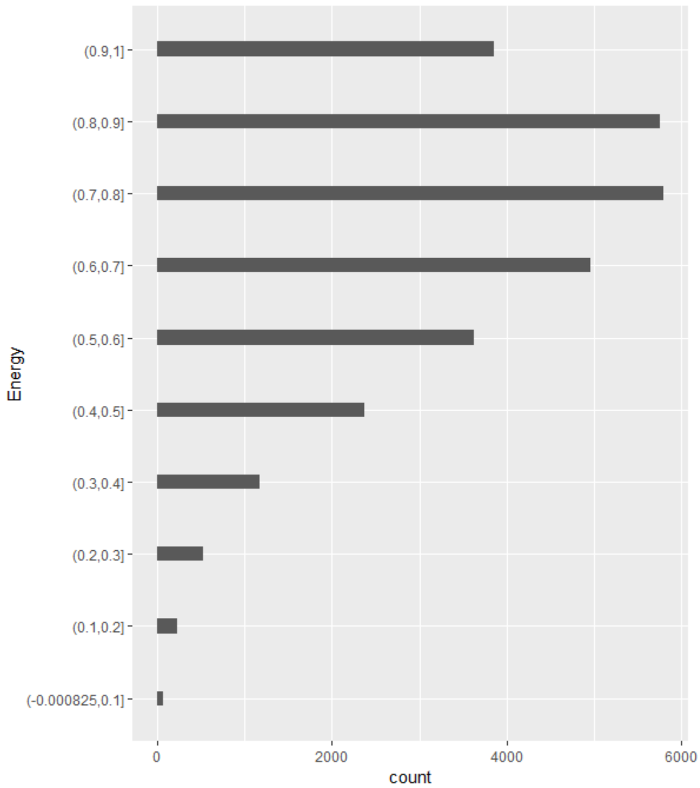

Fig 2.3

Supports the findings from the energy histogram. Hence proved that better energy songs are favoured more by means of Spotify listeners.

Popularity by accousticness:-

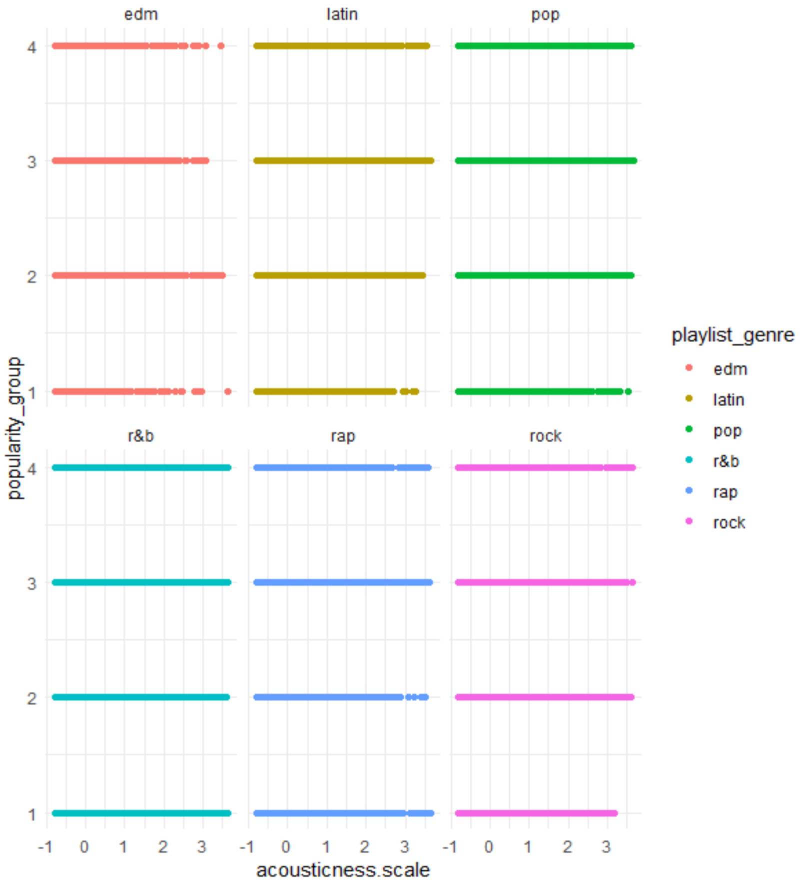

Fig 2.4

Accoustiness does not effect track recognition as the extent of accousticness has been uniform across all reputation levels.

Speechness Distribution:-

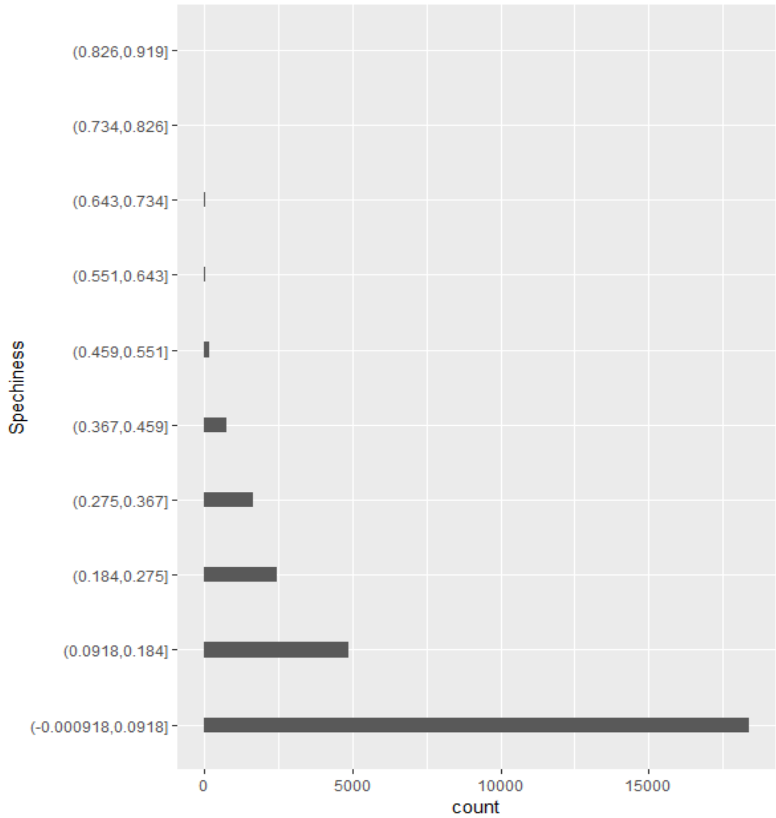

Fig 2.5

This graph also helps the findings of our histogram for speechiness. As all of us recognize how we do not like speechier tracks, this affirms our belief that less speechy songs are greater favoured by most Spotify listeners. So, Spotify does not maintain speechier songs in their database.

Tempo and liveliness distribution across genre:-

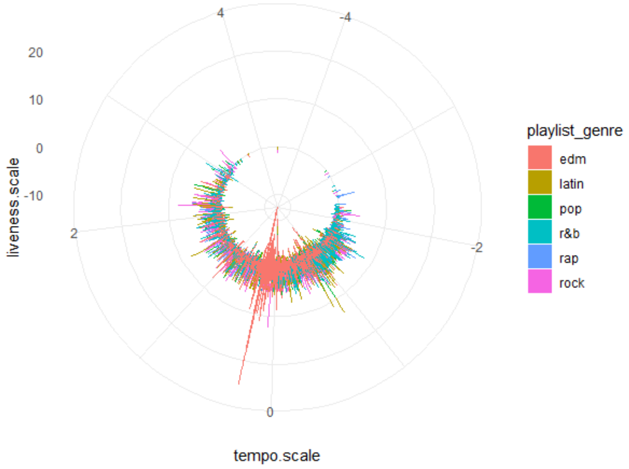

Fig 2.6

Tempo is way better for EDM genre in comparison to the others whilst Liveness is almost uniformly dispensed across all genres.

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 were also presented in distinct and colorful way. In this way we have designed the intricate aspects of the MRS.

In this project, we are designing and implementing a music recommendation system. We used Datasets provided by Kaggle to find correlations between users and songs and to learn from the previous listening history of users to provide recommendations for songs which users would prefer to listen most in future. 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 favourite artists of an user.

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 9th 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 2012 .

[7] Smt. Namitha S J , "Music Recommendation System" IJERT , Volume 08 , B N M Institute of Technology, Bangalore, July 2019 . ,

[8] Mukkamala S. N.V. Jitendra; Radhika, Y , "Singer Gender Classification using Feature-based and Spectrograms with Deep Convolutional Neural Network" , International Journal of Advanced Computer Science and Applications; West Yorkshire Vol. 12, Iss. 2, (2021).

[9] Markus Schedl, Peter Knees, Brian McFee and Dmitry Bogdanov , "Music Recommendation Systems: Techniques, Use Cases, and Challenges" , Springer , 22 November 2021 .

[10] Victor Ivamoto , "Movie Recommendation System in R " ,HarvardX's Data Science, December 2019 .