# Music Recommendations System

Kumar Saurabh (L00157097), MSc in Big Data and AI, Under the supervision of Dr. Shagufta Henna,
Letterkenny Institute of Technology,Letterkenny , Co. Donegal

***Abstract*— Music Recommendation has been very successful after the introduction of online music streaming websites and applications. These websites make all the songs available at your fingertips. You are just a click away from listening to your favorite songs. Here, the music recommender is very helpful in finding new and interesting songs for the user based on his liking in a huge collection of songs in the online music library. The online music library is a huge database and finding a song in that is not an easy task. Recommended systems are developed to help users find their choice of contents and make their work easier. However, the recommendation systems are still facing various challenges especially the music recommendation system. The purpose of this report is to identify the challenges, start building our agent from a small collection of data, and then implement it at a larger scale. A recommender is effective only when it has an ample amount of trained data. Only then we can say that it can be used on a large scale. This research will give the reader an idea about the challenges faced by the music recommendation system and it will guide him to identify interesting ideas that would be beneficial in building an effective recommended system.**

***Index Terms*— Music recommendation, streaming websites, applications, songs, music library, data.**

## I. Introduction

IN the modern age, music is one of the common platforms for entertainment. In the form of sounds that consist of melody, harmony, and rhythm, music is seen as the work of human imagination to communicate ideas and feelings. Music can be categorized among various genres, such as jazz, rock, pop, instrumental, classical, etc. Users across various online music service platforms have the same taste based on various aspects. In this recommended system, we will try to predict songs based on the ratings given by the user. Research in the music recommended system has increased a lot either it is in academics or in the industries. This has happened because of the online music streaming service. Most of the online music service providers such as Amazon Music, Gaana.com, iTunes, and Spotify offer an online library that includes huge amount of tracks based on different styles to meet the taste of users coming from various cultures and backgrounds. When we go on to filter the contents, based on our choices, we can see that the list becomes short and we are left with songs that are similar to our taste in music. I came to a scenario where I went to a party and music was playing on the beatbox. In the beginning, I thought I might be able to listen to my choice of music but I spend almost an hour and no songs were played of my choice.

This can be so irritating. At that point in time, it is understood that songs at a party can be played according to the choice of the host but it is intolerable when it comes to your own phone or laptop. You are listening to a song and all the other songs which are about to be played next are not according to your taste. Imagine how irritating it could be at that point in time. You will end up either searching for songs or close the app or website. Recommended systems are a solution to this. However, these recommendations are not very successful in predicting songs which are fit according to the preference of the user. Most of the time they give unacceptable recommendations. Managing songs which are as per someone's taste is easy when you are working on a small library i.e. personal computers but it can be a very difficult task when it comes to the online music library. The music library of all the online music apps present these days are huge and due to the race of being the best, new songs are being added to it everyday which makes it even more difficult for a user to search their type of songs. Building an effective recommended system is a difficult task as the recommended system should be able to learn from the user and predict results based on that. Else there wouldn't be any use of such recommended systems.

[1]GitHub Link

## II. LITERATURE REVIEW

This music recommendation software supports the user's favorite songs saved in a large music archive and recommends songs to anticipate the user's favorite songs and even discover more songs based on user preference and genre classification. Initially, we will try to separate music from recommending other fields such as movies, books, or objects. Music pieces are the most used sequence by sequence and not like movies. Nowadays, one of the most important tasks is to provide customized services to every person based on their needs [1].

To fulfill this purpose, the recommendation systems may be used as a mechanism to help consumers give numerous products and alternatives in the decision-making process. They are used to forecast and suggest achieving users' specific things. The key goal of the recommendation systems is to recommend items that can be valued by individuals of great potential. Different techniques are combined in new recommendation

[1]https://github.com/kadamdvishal/MusicRecommendations

systems with the goal of drawing correlations in available datasets.

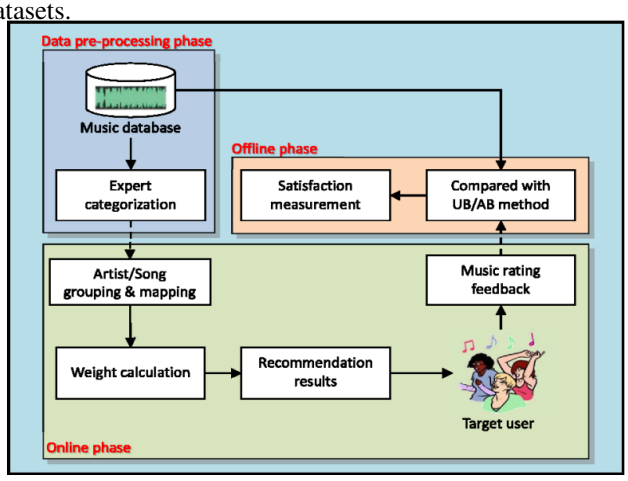


Fig 1: Format of MRS

The mixture of different algorithms makes the forecast more complex, as the predictions can consider different parameters. [2][3] A type of information filtering that aims to display items (movies, books, music) or social features (users, groups) that are likely to be of interest to a consumer is recommendation systems or recommendation systems. In other words, recommendation systems aim to filter the data that would be displayed to the user based on his recent actions, such as reading a book, viewing a movie, or listening to music.

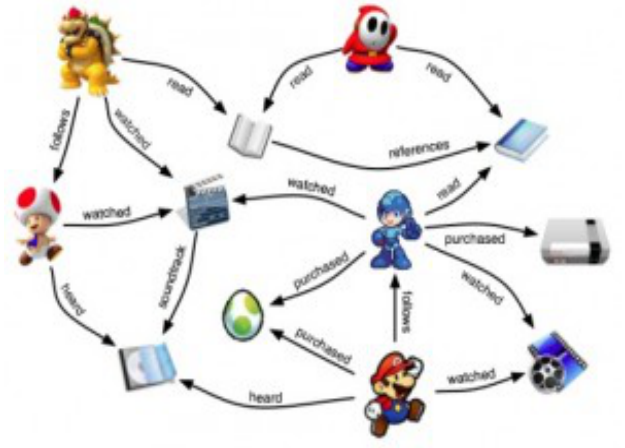


Fig 2: Filtering Using Recommendation System

An efficient algorithm is necessary to achieve this sort of knowledge filtering. Currently, many websites use recommendation services, such as Amazon, IMDB, Apple, etc. A person registered to a movies website, for example, declares that he likes war movies, either by regularly searching war movies or stating that he has seen lots of them. Then this pattern should be recognized by the suggestion method and he can suggest other war movies that he has not seen yet. This can provide an organization with considerable benefits and can improve its revenue since many people could enjoy stuff that they did not know existed. That's why, by recommending them, recommendation systems aim to identify similar items that a person enjoys to get them closer. In Hive, recommendation systems can be used to suggest other related users to users who have several preferences they didn't know before; Hive allows them the chance to meet. It can also be used for preferences, such that using Hive's tips, a person can discover several new items they might be interested in.

A common challenge in most of the recommended system is that when a new song is introduced in the music library, it is unable to give the correct prediction. Similarly, when a user is new, the recommended system gives the wrong prediction based on the taste of the user because of the insufficient amount of data. Another problem arises when there are not many ratings given for a particular song. It happens when there are a lot of users or songs in the music library so, it becomes very hard for the recommended system to give proper recommendation. We have tried to limit our collection of songs so that we can train our dataset properly in order to get correct prediction and further we can increase our data step by step and make an effective music recommended system.

## III. DATASET

We have created two datasets; one of them consists of songs collection which has various fields such as "SongID", Song Title, and Genres while the other dataset consists of fields such as "UserID", "SongID", "Song Ratings" and "Timestamp". We will be merging both of our datasets using the "SongID" column which is similar in both the datasets. After merging the data, we are going to implement various methodologies on the datasets and will be using the rating column to get our solution which is predicting a song for the user by analyzing the ratings he has given for various songs in the list of songs. Here, "UserID" is the unique ID for each user which will be helpful while we are training our agent for prediction. "SongID" is again a unique ID attached to each song. Song Rating is the ratings given to a song by the user and we will be using this column to create our recommender. Timestamp can be avoided in this report, but we have added it to make our dataset different which shows the time in micro-seconds

| | SongID | SongTitle | Genres |
|---|---|---|---|
| 0 | 1 | Smells Like Teen Spirit - Nirvana | Classical |
| 1 | 2 | Imagine - John Lennon | Jazz |
| 2 | 3 | One - U2 | Instrumental |
| 3 | 4 | Billie Jean - Michael Jackson | Rock |
| 4 | 5 | Bohemian Rhapsody - Queen | Classical |

Fig 3: Song Title dataset

| | UserID | SongID | SongRatings | Timestamp |
|---|---|---|---|---|
| 0 | 1 | 1 | 4 | 964982703 |
| 1 | 1 | 3 | 4 | 964981247 |
| 2 | 1 | 6 | 4 | 964982224 |
| 3 | 1 | 4 | 2 | 964982703 |
| 4 | 1 | 5 | 3 | 964981247 |

Fig 4: Songs Rating dataset

## IV. MAIN APPROACH

The recommendation framework has a user profile database containing the musical interest of the user. A step to be considered for this project is getting raw data into the machine learning environment. The machine learning approach which is used on the data is linear regression. Our main approach is to predict songs for a user based on song ratings using the regression techniques. Recommendation systems modify human relations with the network. It increases the degree of interactivity and boosts the experience of users. The primary purpose of using recommendation systems is to predict the accurate solution to the task it has been provided. In this report, we will be using various machine learning models to predict accurate results for our task. Predicting an appropriate solution is a major challenge here and we have tried to keep our dataset short in order to avoid any issue for our agent to predict the solution it has been asked for. recommendation framework has a user profile database containing the musical interest of the user. A step to be considered for this project is getting raw data into the machine learning environment. The machine learning approach which is used on the data is linear regression. Some different approaches to recommendation systems are available and each has its merits and demerits. There are various approaches for building recommendations system which are explained below:

### A. COLLABORATIVE FILTERING

Collaborative filtering or user-based filtering is an alternative to recommendation mechanisms and many websites use it extensively.

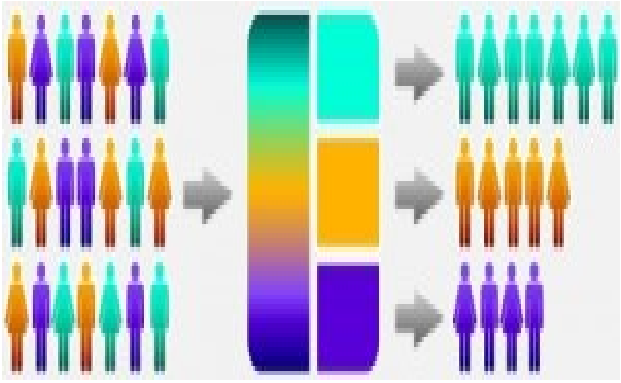

Fig 5: Collaborative Filtering

Its aim is to identify related users and then propose to one what the other wants. How close a user is to another user is determined by the distance between them. By using n-dimensional Pythagoras or certain other methods, distance may be calculated.

**Advantages :**

- More precise
- Very effective
- Very good recommendation results

**Disadvantages :**

- Not efficient to map every user to similar users within sites with many users
- Not effective for new user
- More complex algorithms

e.g. Last.fm is a website that uses a collaborative based approach.

### B. ITEM BASED

This kind of approach finds similarities between items instead of the users. This strategy proposes items that are similar to those that a consumer enjoyed in the past. In fact, different individual items are compared with items previously rated by the customer and it is proposed that the best matched items are compared. For eg, if a customer tends to buy sports shoes, then more sports shoes that the system has identified as close to those he has already purchased will be recommended by the system.

**Advantages :**

- Good for new user
- Better efficiency for items greater than users

**Disadvantages :**

- Scope is limited

e.g. Pandora is a website that uses item based approach.

### C. HYBRID APPROACH

A variation of the above approaches can lead to a mixed approach that can often produce improved outcomes for recommendations. This approach is also used by many web pages so they can avoid the drawbacks of one strategy while using the strengths of the other. Any of the most common hybrid recommendation systems are available on Netflix, Amazon,

and Google. In this recommended system, we have used a Linear Regression model to test and train our dataset.

Linear Regression is a commonly used technique for analysis. Its main idea is to predict a dependent variable. The association between one dependent variable and one or more independent variables is explained by these regression predictions. The formula y = c + b*x describes the simplest version of the regression equation for one dependent and one independent variable, where y = expected dependent variable score, c = constant, b = coefficient of regression, and x = score on the independent variable.

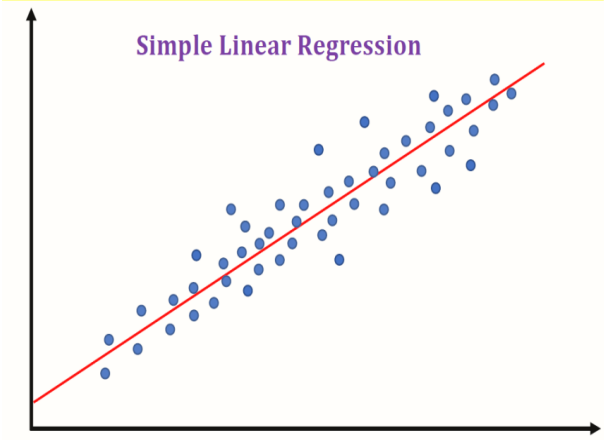


Fig 6: Linear Regression

**Types of linear regression:**
**Simple linear regression**
1 DV, 1 IDV
**Multiple linear regression**
1DV, 2+ IDV
**Logistic regression**
1DV, 2+ IDV
**Ordinal regression**
1DV, 1+ IDV
**Multinomial regression**
1DV, 1+ IDV
**Discriminant analysis**
1DV, 1+ IDV
DV= Dependent Variable (nominal),
IDV = Independent Variable(s) (interval or ratio)

When choosing the model for the study, model fitting is an important factor. Adding independent variables to a model of linear regression will often raise the model's explained variance (usually expressed as R2). However, by adding too many variables to the model, overfitting will occur, which decreases model generalizability. The razor of Occam explains the topic incredibly well-a straightforward model is normally superior to a more complicated model. Statistically, if a model contains a large number of variables, out of chance alone, any of the variables would be statistically important.

Our main approach is to predict songs for a user based on song ratings using the regression techniques. Recommendation systems modify human relations with the network. It increases the degree of interactivity and boosts the experience of users. The primary purpose of using recommendation systems is to predict the accurate solution to the task it has been provided. In this report, we will be using various machine learning models to predict accurate results for our task. Predicting an appropriate solution is a major challenge here and we have tried to keep our dataset short in order to avoid any issue for our agent to predict the solution it has been asked for. Regression is a commonly used tool for analyzing data. We are using the Linear Regression model for our recommended system. Linear Regression We will be using Google Colab as an environment to train and test our dataset. We will use the SongID column in the dataset to perform this task for our recommended system to get an accurate result. We have performed more than 50 tests on our agent to check if it is performing properly and we are happy that it has been able to recommend songs based on the ratings given by the user. We would further be using our recommended system to perform more difficult tasks by improving it for recommending other contents based on user preferences.

## V. EVALUATION METRIC

At present, most music personalized recommendation systems use content-based recommendation list to predict results for a user. However, regardless of the type of methods used, the essence is to use the prediction accuracy as an index for evaluating a music personalized recommendation system. We also access the performance of the recommender system in terms of freshness and popularity by recommending the best songs for each user. In our recommendation system, we are using the content-based filtering, so all our prediction is based on a single variable. We must be very precise while choosing our variable for the recommended system as it can decide whether our recommended system is effective or not [4].

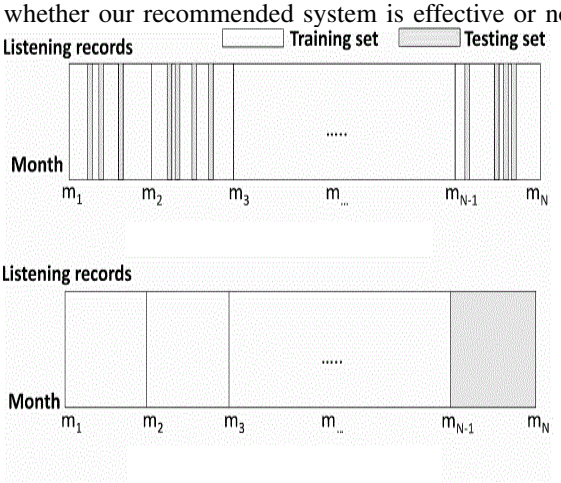


Fig 7: Real Life Strategy For MRS

Content-based filtering attempts to remove valuable data from the things of the client's assortment that could be helpful to speak to the client's requirements. This methodology comprehends the restriction of shared separating as it can suggest new things (even before the framework knows nothing from that thing), by looking at the genuine arrangement of client's things and figuring a separation with some kind of likeness measure. CBF calculations don't require evaluations of clients other than the clients that are objective. In this way, however

long a few snippets of data about the client's own inclinations are accessible, such strategies can be utilized in virus start situations. Besides, in the most extreme case, when another thing is added to the inventory, content-based techniques empower suggestions, since they can remove highlights from the new thing and use them to make suggestions [5].

## VI. RESULT AND ANALYSIS

We will be using regression models in our final code for the music recommendation system to predict songs for users. No cleaning for data is required as the data is not very large. We are getting random results using the recommended system based on the maximum ratings given for a song. We have tested. our agent multiple times and it is able to pick songs based on the change in the dataset so, we can say that our agent is performing as per the task assigned to it. We have merged both the dataset using the "SongID" column and further will be training our recommended system based on that. We will be using the filtering of data based on collaborative filtering which will make our data more accurate to be trained. In the end, we are getting the top 5 songs from the list. Then we are asked to select a song. When we select a song, the recommended system recommends songs based on the rating of that song.

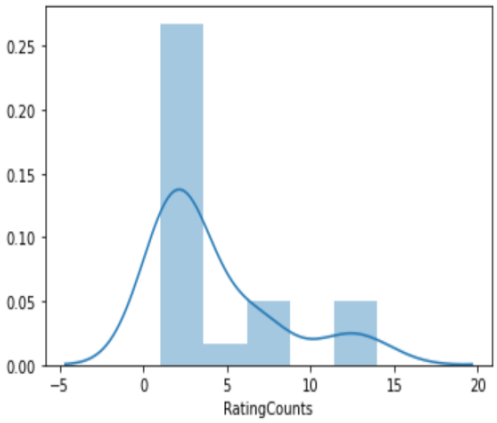


Fig 8: Rating Counts

## VII. FUTURE WORKS

In the current report, we are going to predict the songs for a user based on the ratings he has given for various songs. For the future, we propose to predict his choice of songs based on genres, mood and time he spends listening to a particular song. The mood of the user can be predicted by the type of song he listens to most of the time and similarly for the time, we calculate the time he is giving to a song for listening. In that way, we can skip a song for which he has given less time. The learning systems that are already present are not personalized. There are many users present who differ in taste, the amount of time they spend on the music streaming websites and the amount of data they share on the recommended system. Because of that, it is nearly impossible to predict songs for every user with a limited set of information. Most of the users do not provide ratings as well in the songs application so it is very difficult to make suggestions for them. The traditional recommender builds an initial profile of the user during the sign-in process and then they ask the user to share the preference of genres and based on that they recommend the latest songs present in the library. That is not at all accurate or based on that as the user has just made a profile on the application [7]. So, in the future, we will try to make the recommended system more accurate by collecting a lot of data when the user registers to the application so that we can predict accurate songs for the user and he listens to the songs which he likes. We are further trying to extend this dataset and keep it clean so that we can predict similar songs as per the user's taste.

## VIII. REFERENCES

[1] Szu-Yu Chou, Y. Yang and Yu-Ching Lin, "Evaluating music recommen- dation in a real-world setting: On data splitting and evaluation metrics," 2015 IEEE International Conference on Multimedia and Expo (ICME), Turin, 2015, pp. 1-6, doi: 10.1109/ICME.2015.7177456.

[2] Music Recommender System Based on Genre using Convolutional Recurrent Neural Networks Adiyansjah, Alexander A S Gunawan, Derwin Suhartonoa Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480.

[3] Novello A, McKinney MF, Kohlrausch A (2006) Perceptual Evaluation of Music Similarity. In: Proceedings of the 7th international conference on music information retrieval (ISMIR), Victoria, BC, Canada.

[4] Current challenges and visions in music recommender systems research · Markus Schedl1 · Hamed Zamani2 · Ching-Wei Chen3 ·Yashar Deldjoo4 · Mehdi Elahi5

[5] Elahi M (2011) Adaptive active learning in recommender systems. User Model Adapt Pers 414–417.

[6] Carenini G, Smith J, Poole D (2003) Towards more conversational and collaborative recommender systems. In: Proceedings of the 8th international conference on intelligent user interfaces, IUI'03. ACM, New York, NY, USA, pp. 12–18.

[7] Cremonesi P, Elahi M, Garzotto F (2017) User interface patterns in recommendation-empowered content intensive multimedia applications. Multimed Tools Appl 76(4):5275–5309.