ISSN No:-2456-2165

User Based Spotify Recommendation System using Machine Learning Algorithms

Dr. Mahaboob Basha. Sk.[1],
Professor,
NRI Institute of Technology, A.P, India-521212

L.Vyshnavi[2],
UG Scholar, Dept. of IT,
NRI Institute of Technology, A.P-521212

S. Sriharsha[3],
UG Scholar, Dept. of IT,
NRI Institute of Technology, A.P-521212

G.Dhathrik[4]
UG Scholar, Dept. of IT,
NRI Institute of Technology, A.P-521212

Abstract:- We have described a personalized music recommendation system using K-nearest neighbour that is KNN and machine learning methods in this paper. We present a collaborative filtering and content filtering recommendation algorithm to combine the output of the network with the log files to recommend music to the user in a personalized music recommendation system. The recommended system includes log files that store the past or viewed history of the user's music playlist. The propound music exhortation system pulls the consumer's the beyond records from the log file and provides track tips for each recommendation. Content-based approaches make suggestions based on the audio characteristics. Speedy development of cell phones and internet has made possible for us to access various music resources freely. While the music industry may favour certain types of music more than others, it is salient to understand that there isn't a single human culture on earth that has existed without music. In this paper, we have sketched, implemented and examined a song recommendation system. We have used Song text provided to find relationship between users and songs and to seek from the preceding listening history of users to deliver recommendations for songs which users may prefer to listen mostly. The dataset bottles up over 10,000 songs and listeners are advocated the first-class available songs based totally at the mood, style, artist and top charts of that yr. With a powerful interactive UI, we show the listener the cover songs that were played the maximum and top charts of the year. Listener also have an option to select his/her favourite artist and albums on which songs are recommended to them by utilizing the dataset. A recommendation system plays a important role in providing a well user experience in an application by providing the most suitable and personalized services for each and every user. Currently, Spotify has one fifty-five million premium subscribers and three forty five million active users. Spotify's recommendation system has also played a dominant role in the success of Spotify. In the modern years, music and movie flowing services have grown extremely. Currently, Netflix and Spotify have a bulk number of users, which has made these spurting services victorious. A recommendation system plays a vital role in providing a well user experience in an application by recommending the most acceptable and personalized services for each and every user.

Keywords: *K-NN, SVM, Multiple Linear Regression, Random Forest Regression, Popularity Model, Content-Based Model, Collaborative Filtering.*

I. INTRODUCTION

Everyone's taste in music is unique, which means that no matter what music you make, someone is always bound to enjoy listening to it. While the music industry may favor certain types of music compared to others, it is important to understand that there isn't a single human culture on earth that has existed without music. Music is such a great bliss to us, regardless of whether we are renowned recording artists, karaoke singers, or merely fans of music. The number of songs available increases the listening capacity of every single individual. According to disclose of MarsBands.com, there are at least ninety-seven million songs. These are the only songs officially released. If we included songs everyone knows or the incredibly old Celtic songs with no names, we would nearly reach 200 million songs, since the website mostly does not add Happy Birthday or an inexpressible song from 1400 A.C. This is when we only add artists whose names are officially on music charts.

Starting there, let's say that there are presently around one million songwriters active that we know about. Using the same percentage as above, we can estimate that there have been approximately 15.3 million songwriters throughout history. To get an innovative idea, there are four million songs on Spotify that have never been played. In total, there must be billions of songs there, and Spotify itself is by no means the limit of music. There are trillions and trillions of songs in the world, so many that an estimate is impossible, and the potential for an infinitely greater number that have not yet been made, creating a world of music for us to enjoy. Keeping this general idea in mind, one can see that the number of songs is too high for a person, even if listening to music is his or her best hobby. People sometimes find it difficult to pick from millions of songs. Moreover, music service providers need an efficient way to manage songs and help their customers discover music by giving quality recommendations. This means it not only gives the user freedom to select the songs he or she wants to listen to but also recommends songs according to their previous listening history. Thus, there is a solid need for a good recommendation system. In order to effectively access, discover, and present music content to the final user, techniques for searching, retrieving, and recommending need to be appropriate for music content. There has been some work done by both savants and the industry to provide music recommendation services.

Fig 1 Spotify Music

- *Technologies Used:*

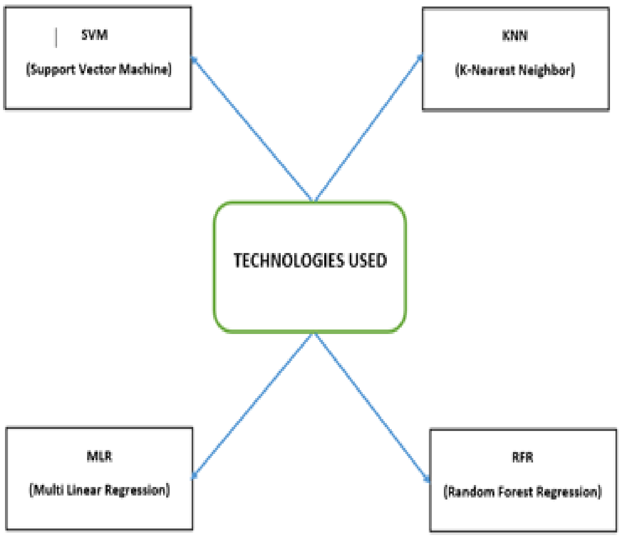

Fig 2 Technologies Used

- *SVM:*

A supervised machine learning method called SVM (Support Vector Machine) can be utilized to resolve classification and regression issues. However, it is usually used to address categorization issues. The value of each feature corresponds to a certain coordinate in the SVM algorithm, and each piece of data is represented as a point in n-dimensional space (where n is the number of features you have).

A linear model called the Support Vector Machine, or SVM, can be used to address classification and regression problems. It is helpful for a variety of applications and can solve both linear and nonlinear problems. SVM is a fundamental idea:

The approach uses a line or hyperplane to partition the data into classes. It separates the words into categories during the training phase, such as happy, sad, and so on, based on the training dataset, and then predicts the mood of the input song based on the words and the correspondingly score. The top similarity songs' moods are predicted using SVM and arranged in increasing order, after which songs with the same mood are suggested. (fig:2)

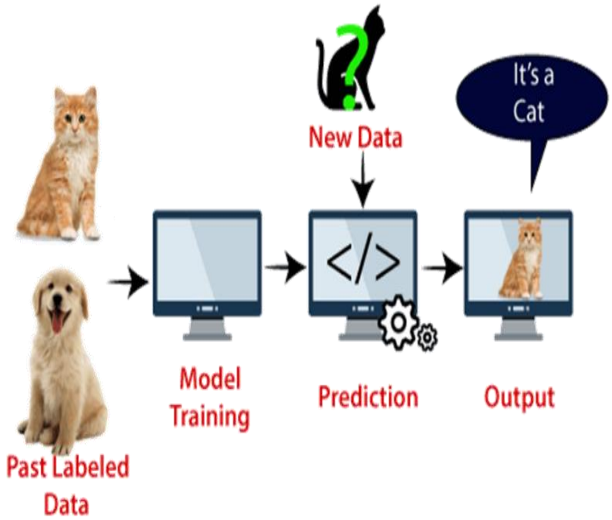

Fig 3 SVM

- *K-NN:*

The K-Nearest Neighbor (K-NN) model for recommendations is an item-based strategy that searches for neighbors between objects, in contrast to user-based algorithms that look for neighbors between users. The best model for implementing item-based collaborative filtering and a great place to start when developing a recommendation system is K-Nearest Neighbor. A non-parametric learning technique is the K-NN approach. This technique uses a database with categorized data points to draw conclusions for new samples. K-NN makes no assumptions about the distribution of the underlying data and only relies on the similarity of item attributes. K-NN ranks the "distance" between each item in the database and the target item when it arrives to a decision about an item.

The top K items are then suggested as the most comparable items.

The K-Nearest Neighbors method's algorithm is as follows: 2012 (Han et al.)

- ✓ Establish the parameter k (number of nearest neighbors).
- ✓ Determine the separation between all training data and the data that will be examined.
- ✓ Sort the distances created (in ascending order) and find the one that is the closest.
- ✓ Include the proper class (c).
- ✓ Determine how many classes are closest neighbors, and then identify the class as the data evaluation. [fig:4]

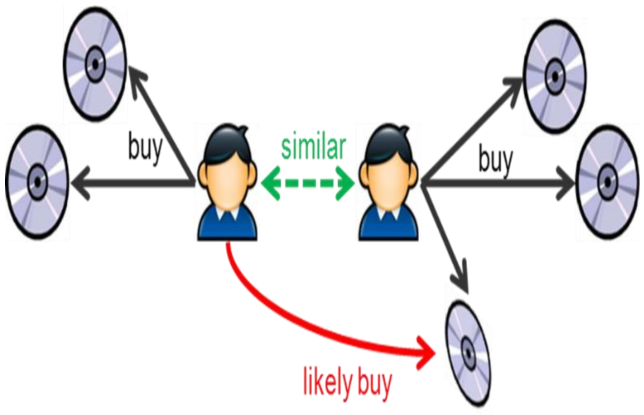

Fig 4 K-NN

- *Multiple Linear Regression:*

Similar to linear regression, multiple regression attempts to predict a value based on two or more factors, but with more than one independent value.

Multiple linear regression, or simply "multiple regression," is a statistical method for predicting the result of a response variable that makes use of a number of explanatory variables. One explanatory variable is used just once in multiple regression, an extension of linear (OLS) regression. [fig:4]

- *Random Forest Regression:*

Every decision tree has a significant variance, but when we mix them all in parallel, the resulting variance is low since each decision tree is perfectly trained on the sample data in question. As a result, the outcome is dependent on numerous decision trees rather than just one. The majority voting classifier is used to determine the final output in a classification challenge. The final output in a regression problem is the mean of all the outputs. Aggregation describes this section. Using several decision trees and a method called Bootstrap and Aggregation, Random Forest an ensemble methodology capable of handling both regression and classification problems.

The central idea is to use a combination of decision trees to determine the final output rather than relying on individual decision trees. Random Forest has multiple decision trees as its base learning models. We randomly perform row sampling and feature sampling from the dataset, forming sample datasets for every model. This part is called Bootstrap. Nifty500 index represents the top 500 companies based on full market capitalization from the companies listed in National Stock Exchange. Inflation is represented by the Consumer Price Index, with 2001 as the base year.

For deposits lasting more than a year, the term "deposit rate" is used as a proxy for the "interest rate." The gold price is expressed as the spot price in rupees per ounce. For each of these variables, monthly data were gathered from January 2000 to December 2018. These data were retrieved from the Center for Monitoring the Indian Economy's databases. For every one of these variables, there were 228 observations. The gathered data was trained and modeled using machine learning methods. Eighty percent of the data were used for the model's training, and the remaining twenty percent were used for testing.

In this work, linear regression, random forest regression, and gradient boosting regression were employed as machine learning techniques. Regression analysis is a statistical technique for determining the relationship between various variables. When one of the independent variables changes while the other variables remain constant, regression analysis is used to determine how the value of the dependent variable changes. Multiple linear models are linear regression models with more than one independent variable. Below is an illustration of a multiple linear regression where Y is the dependent variable and X1, X2 are the independent variables.

$$Y = a + b1*X1 + b2*X2 + ... + bp*Xp$$

As a result, machine learning has adapted linear regression, which was created in the field of statistics and is examined as a model for understanding the relationship between input and output numerical variables. Now, it functions as both a statistical and a machine learning method. Different machine learning applications can employ decision trees. A decision tree builds a tree that is used for regression and classification. However, trees that have been grown very deeply to learn extremely irregular patterns have a tendency to over fit the training sets. The tree might develop in a totally different way as a result of a small amount of data noise. Ensemble methods include building numerous models and combining them to achieve better outcomes. Ensemble approaches refer to both Random Forests and gradient boosted trees.

Multiple weak learners, such decision trees, are combined using ensemble methods to create a powerful learner, like random forest. Third International Conference on Trends in Electronic and Informatics Proceedings (ICOEI 2019) ISBN: 978-1-5386-9439-8; IEEE Xplore Part Number: CFP19J32-ART; 19/$31.00 ©2019 IEEE 1361 Both gradient-boosted trees and random forests can be applied to classification and regression applications. Different machine learning applications can employ decision trees. A decision tree builds a tree that is used for regression and classification. However, trees that have been grown very deeply to learn extremely irregular patterns have a tendency to overfit the training sets.

The tree might develop in a totally different way as a result of a small amount of data noise. With the aid of many decision trees and a method called Bootstrap Aggregation, also referred to as bagging, a Random Forest is an ensemble methodology capable of handling both regression and classification tasks. This method's fundamental principle is to integrate several decision trees to get the final result rather than depending solely on one decision tree. To lower the variance and preserve the low bias produced by a Decision Tree model, the Random Forest performs bootstrapping on Decision Trees. [fig:2]

➢ *Software Requirements Specification:*

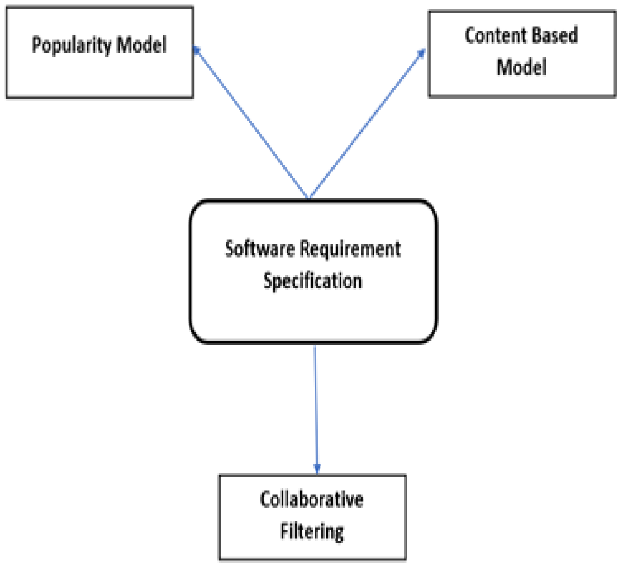

Fig 5 SRS

➢ *Popularity Model:*

It is a straightforward model that ranks the songs in the training set in order of decreasing popularity and suggests the most well-liked ones. This approach disregards the preferences of the user. [fig:5]

➢ *Content Based Model:*

A representation of the item and a profile of the user's preferences serve as the foundation for content-based filtering techniques. Recommendation is treated as a user-specific classification problem by content-based recommenders, which also learn a classifier for the user's preferences based on product attributes.

The system primarily employs two sorts of data to build user profiles: a model of the user's preferences and a history of the user's interactions with the recommender system. Based on song metadata, we constructed a K-Nearest Neighbor model to suggest songs. We first built a library of songs based on various information elements (artist, genre, etc.), after which we suggested songs that were comparable. From among the songs found in the user's profile, we choose the k closest neighbors. The nearest neighbor algorithm, which is based on a ball tree, is used to alleviate the computing shortcomings of the brute-force method. A data structure created using the ball tree algorithm, which divides data into a series of nesting hyperspheres, can be particularly effective on highly structured data, even in very high dimensions. [fig:5]

➢ *Collaborative Filtering:*

The propose behind collaborative filtering is that people who have formerly agreed will do so again and that they will continue to enjoy the same kinds of things.

Recommendations are generated by the algorithm solely based on data from rating profiles for many persons or things. It generates recommendations using this neighborhood by identifying peer users or items with rating histories similar to the current user or item. We have created a collaborative filtering model based on items. For training, the listen count parameter is used as implicit feedback.

When comparing two items, we examine the set of items the target user has rated, determine how similar they are to the target item I, and then choose the K items that are the most comparable. The cosine similarity function is used to calculate the similarity between two items by first obtaining the user ratings for both items.

The forecast is then calculated by taking a weighted average of the target user's ratings on these similar products once we have determined how similar the items are. The formula used to determine rating is very similar to user-based collaborative filtering, with the exception that weights are assigned to items rather than persons. [fig:5]

➢ *Existing System:*

Reinforcement learning, or RL, is a type of machine learning model that responds to its happening environment in an effort to maximize the ultimate, long-term reward, whatever that may be. In our case, that reward is our users' long-term satisfaction with Spotify. RL isn't about short-term solutions. It's always playing the long game. In other words, rather than handing users the "empty calories" of a content diet that will only satisfy them in the moment, RL aims to push them toward a more live able, diverse, and fulfilling content diet that will last one's time. This could mean playing an advanced dance track we think might fit a user's present mood, or it could mean suggesting a calming, ambient piece to help them study.

After that e-commerce gained popularity of selling products and the development of new generation mobile phones and innovative inventions like tablets, shopping become easier than before. Therefore, new kinds of advertising techniques came up. Recommendation systems are one of the most popular techniques nowadays. They are useful for both the company and the user because they increase product sales while reducing the time spent shopping. However, they have some problems because of the huge amount of data. The main problems are being unable to get really relevant results and being unable to get results in a reasonable amount of time.

- *Disadvantages:*

Up until now, a lot of methods have been developed to resolve the problems stated above, such as collaborative filtering and content-based filtering. However, they have some weaknesses. For example, collaborative filtering has the problem of "cold starts," which means that the recommendation system cannot produce any suggestions or recommendations. This difficulty occurs when items are provided in the system but there are few customers and few or no rankings. And in the other example of content-based filtering, if the content lacks enough information to distinguish the items precisely, the recommendation cannot be made. On the other hand, our system should find the most accurate recommendations. Our potential users, whom we desire to help with their problems, are companies that use internet utilities to sell their products. And by proposing our product to these companies, we will reach users of the websites that companies use. Therefore, we will reach both customers and companies with our system.

➢ *Proposed System:*

The creation of a music suggestion app is the major goal of this endeavor. Users can choose and listen to songs stored on the device using the program. A log is established each time a person listens to a certain song. We construct a recommendation engine using a variety of ways to propose songs to users.

The primary goal of this proposed system is to increase the functionality of the current recommendation system. Traditional music recommendation systems produce recommendations via collaborative filtering or content-based filtering. Collaborative filtering and content-based filtering are combined in hybrid techniques to take advantage of both of their advantages and disadvantages. Better user profiles are created through user modelling.

Users and objects are connected via context awareness in situations like working or dancing. Items are tagged with user reviews in tag-based recommendations. The goal of recommendation in the long tail is to reduce the popularity bias. It typically goes hand in hand with collaborative filtering and disregards item popularity for content-based filtering. Recommendation networks give the recommendation strategies some new properties. The creation of playlists can be seen as a variation on top-N suggestions that satisfies user needs. Group suggestion includes some pre- or post-processing, either by combining several user profiles with shared interests or by combining different lists of recommendations into a single list. The system we propose consists of three models.

- *Advantages of Proposed System:*

The main aim of any recommendation engine is to stimulate demand and actively engage users. Primarily a component of an ecommerce personalization strategy, recommendation engines dynamically populate various products onto websites, apps, or emails, thus enhancing the customer experience.

It provides personalization and thus boosts user engagement. The recommender system is helpful to both service providers and users. It saves time for the user in finding and selecting a perfect song and at the same time, it also helps service providers retain customers for a longer time on their platform

➢ *System Architecture:*

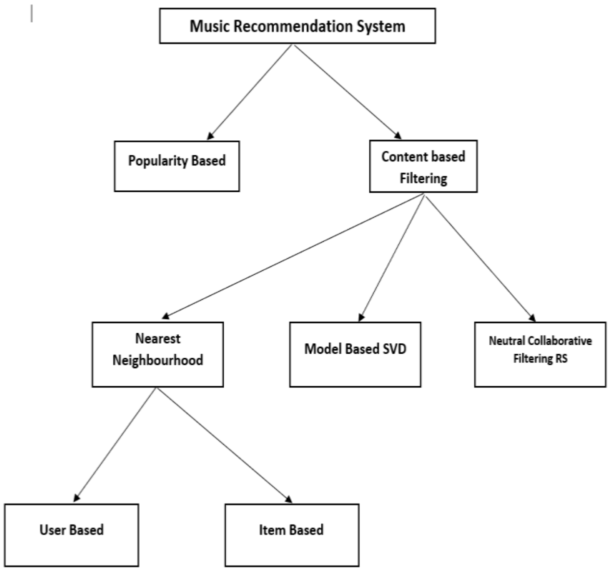

Fig 6 System Architecture

➢ *Future Scope:*

In coming days, the recommendation system plays a vital in maintaining and developing music systems likewise improving the user experience.

II. CONCLUSION

To improve the quality of music recommendations, music recommender systems should first take music genre information into account. Based on the attributes of the songs, the music recommender can make song recommendations. By calculating the similarity score for each recommended song, the music recommender can detect plagiarism in the dataset used. By comparing the lyrics of the supplied song with all the other songs in the dataset, the mood of the song is predicted. The anticipated mood and similarity scores are then used to recommend music depending on the mood.

Because different music recommender systems operate in various ways, the complicated nature of machine learning systems like the Music Recommendation System cannot have a uniform framework. Our findings allow us to identify additional music aspects for future research in order to increase the recommender system's accuracy, such as employing tempo gramme to record the local tempo at a certain time.

REFERENCES

[1]. Balabanović, Marko, and Yoav Shoham. "Fab: content-based, collaborative recommendation." *Communications of the ACM* 40.3 (1997): 66-72.
[2]. Bernhards son, Erik. “Recommendations at Spotify v4” [Presentation Slides], URL: http:// de.slideshare. net/MrChrisJohnson/algorithmic-music-recommendations-at-spotify, [Online; last accessed June 2016]

[3]. Bernhards son, Erik. “Nearest neighbour methods and vector models – part 1“. URL: https://erikbern.com/2015/09/24/nearest-neighbor-methods -vector-models-part-1/. [Online; last accessed June 2016]
[4]. Dieleman, Sander. “Recommending music on Spotify with deep learning”.URL: http://benanne.github.io/ 2014/08/05/ spotifycnns.html August 05, 2014. [Online; last accessed June 2016].
[5]. Hu, Yifan, Yehuda Koren, and Chris Volinsky. "Collaborative filtering for implicit feedback datasets." *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on*. Ieee, 2008.
[6]. Koren, Yehuda, Robert Bell, and Chris Volinsky. "Matrix factorization techniques for recommender systems." Computer 8 (2009): 30-37.
[7]. Pazzani, Michael J., and Daniel Billsus. "Content-based recommendation systems." The adaptive web. Springer Berlin Heidelberg, 2007. 325-341.
[8]. Spotify Developers. Audio Features [Table File]. URL: https://developer.spotify.com/web-api/console / get-audio-features-track/#complete. [Online; last accessed June 2016]
[9]. Spotify Webpage. N.p. URL: https://www.spotify. com/de/. [Online; last accessed June 2016]
[10]. Statista.N.p.URL: http://www.statista.com/statistics/ 244995/number-of-paying-spotify-
[11]. Steck, Harald, van Zwol, Roelof, Johnson, Chris. “Interactive Recommender Systems” [Presentation Slides], URL: http://de.slideshare.net/MrChrisJohns on/interactive-recommender- systems-with-Netflix–and-Spotify,[Online;lastaccessed June 2016]
[12]. Kaggle: Your Machine Learning and Data Science Community
[13]. Welcome To Colaboratory - Colaboratory (google.com)

BIOGRAPHIES

Dr. Sk. Mahaboob Basha is presently working as Professor in the department of Information Technology at NRI Institute of Technology, Vijayawada. He received his M.Tech degree from Jawaharlal Nehru Technological University, Kakinada (JNTUK) and Ph.D in Computer Science and Engineering from Acharya Nagarjuna University(ANU). He has published over 10 research papers in International Conferences and Journals. He has more than 20 years of experience in teaching.

L. Vyshnavi is currently studying B.Tech with specification of Information Technology in NRI Institute of Technology. She done a summer internship project on machine learning.

S. Sriharsha is currently studying B.Tech with specification of Information Technology in NRI Institute of Technology. He done a summer internship project on machine learning.

G. Dhathrik is currently studying B.Tech with specification of Information Technology in NRI Institute of Technology. He done a summer internship project on machine learning.