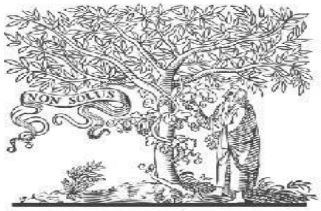
International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

COPY RIGHT



ELSEVIER SSRN

2023 IJIEMR. Personal use of this material is permitted. Permission from IJIEMR must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. No Reprint should be done to this paper, all copy right is authenticated to Paper Authors

IJIEMR Transactions, online available on 05th Apr 2023. Link :http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=Issue 04

10.48047/IJIEMR/V12/ISSUE 04/08

Title **MUSIC RECOMMENDATION SYSTEM USING MACHINE LEARNING**

Volume 12, ISSUE 04, Pages: 52-58

Paper Authors

Dr. Sanjay Gandhi Gundabatini, R. Bindu Sneha, M. Vamsi, N. Yagna Yaswanth, K. Krishna Vamsi

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

MUSIC RECOMMENDATION SYSTEM USING MACHINE LEARNING

Dr. Sanjay Gandhi Gundabatini[1], Professor, Department of CSE,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.

R. Bindu Sneha[2], **M. Vamsi[3]**, **N. Yagna Yaswanth[4]**, **K. Krishna Vamsi[5]**
[2,3,4,5] UG Students, Department of CSE,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.
[1] sanjaygandhi.g@gmail.com[1], bindusneha02@gmail.com[2],
vamsimedabalimi415@gmail.com[3], yaswanth.yagna@gmail.com[4],
vamsikomma044@gmail.com[5]

Abstract

The rise of network technology has led to notable advancements in music recommendation systems, making online music platforms the preferred choice for people to listen to their favorite songs. Nevertheless, these systems encounter a range of obstacles, including difficulties in data storage, suboptimal computational efficiency, issues with cold start, and sparse data due to the vast amount of information involved [1]. To tackle these challenges, the objective of this study is to create and establish a music recommendation system that effectively addresses these kinds of concerns. The approach is implemented using K Nearest Neighbor Classification and Random Forest Classifier. Since this paper proposes another approach for the recommendation, the other approach utilises the Spotify API calls. It aims at the implementation using the various machine learning algorithms which includes K Nearest Neighbor Classification, Decision Tree Classifier and the Random Forest Classifier. This model recommends the music to the user by taking certain factors into account like genre, year range and music features like acousticness, danceability, energy, tempo, valence. The Spotify API is another approach that will recommend the songs based on the previous listening history, preferences, likes of a particular user which comes under the content based filtering.

Keywords : Content based filtering, K-Nearest Neighbor, Machine Learning, Random Forest Classifier, Recommendations, Spotify API.

1. Introduction

A music recommendation system is an application that offers tailored music suggestions to users based on their preferences, listening habits, and previous activity. The advent of digital music streaming platforms and online music stores has resulted in an explosion of available music, making it increasingly challenging for users to discover new songs they would enjoy listening to. To address this challenge, music recommendation systems have been developed using various machine learning techniques. These systems analyze datasets of user behavior and music

metadata to generate personalized music recommendations. The most common techniques used in music recommendation systems include collaborative filtering, content-based filtering, and hybrid filtering. Collaborative filtering recommends music based on the behavior of similar users, while content-based filtering offers suggestions based on the attributes of the music, such as genre, tempo, and mood. Hybrid filtering combines both approaches to provide more precise recommendations. This particular implementation primarily focuses on personalized recommendations and uses Content-Based filtering.

In this study, we implemented the K-Nearest Neighbor model, Decision Tree and the ensemble learning technique i.e Random Forest Classifier. These models work on the large datasets and extracts the information for the smooth recommendation.

2. Literature Survey

According to the research survey, Since 1995, the number of internet users worldwide has grown by around 40% to reach 3.2 billion. This increase in the availability of information has opened up many opportunities, but it has also made decision-making more complex for users. With such an enormous amount of data available, making certain decisions can be challenging. While it is important to be well-informed, excessive information can impede the decision-making process. Recommendation systems were developed to alleviate this confusion and improve the browsing experience by providing users with tailored suggestions. Music recommendation system is one among those[2]. In one study, researchers discuss the fundamental concepts, design, advantages, and drawbacks of content-based recommender systems. They also provide an overview of classical and advanced techniques for representing items and user profiles, along with widely adopted techniques for learning user profiles. The paper also explores future research directions, including the role of User Generated Content in evolving vocabularies and the challenge of delivering successful recommendations to users. In another study, researchers highlight the K-Nearest Neighbor (K-NN) model as an item-based approach for recommendations, which differs from user-based algorithms that search for neighbors between individuals. The K-NN method is a non-parametric learning technique that uses a database to categorize data points based on their similarity in item attributes. It is a great starting point for building a recommendation system since it makes no assumptions about the distribution of the data underlying the recommendations. Also, certain researchers discovered that Random Forest Classifier being the ensemble learning technique is more widely used for the implementation of music recommendation systems. As observed, many studies were implemented using the machine learning algorithms such as Random Forest and KNN, utilising the

previous listening data of the particular user in their Spotify account.

3. Problem Identification

While music recommendation systems have become an important tool for music lovers to discover new and relevant music, there are several challenges that these systems face. Some of the key problems in music recommendation systems include data sparsity, cold-start problem, limited music diversity, scalability, overfitting, contextualization, user engagement, unintended biases, lack of transparency, multi-objective optimizations and privacy concerns. These problems highlight the need for the study we devoloped and innovation in music recommendation systems to improve their accuracy, diversity, and overall user experience.

4. Methodology

4.1 Dataset Preparation:

The dataset is obtained from the kaggle named SpotGenTrack which is widely used by all the developers for music recommendation and related projects. This file contains data sources and features extracted files which contain spotify_artists, spotify_tracks and spotify_albums.csv files. The required features are then extracted into another file which is stored in the features extracted folder. The dataset contains 101938 songs, 75503 albums, 40734 artists, 11 genres and 9 audio features.

4.3 Data Pre-processing:

During this stage, the extracted features are filtered based on the input given by the user in terms of genre, year range and for the features of music are stored into a new file. This stored data is used for the implementation of the system.

4.4 Attribute Subset Selection:

The attributes that are used for the consideration of recommendation are selected and cleaned in such a way that the features tend to give the perfect results upon execution. Out of 45 attributes, only 28 attributes are considered in the process of feature selection.

4.5 Models:

Here, three machine learning models are being used for the recommendation of music.

4.5.1 K-Nearest Neighbor

The k-Nearest-Neighbors (kNN) method is a simple yet effective classification strategy. Its limitations include low efficiency because it is a lazy learning method, which prevents it from being used for large repositories of dynamic web mining, as well as its dependence on choosing an appropriate value for k. In this paper, we suggest a novel kNN-based classification technique to address these problems. Our method builds a kNN model directly from the data, which forms the foundation for categorization and lessens reliance on k. Based on the size of the genre data, the ideal number of k is automatically chosen, improving classification accuracy and speed. [6].

Here, the nearest neighbors k value is defined based on the length of the genre_data being choosen. So, based on the k value, the nearest neighbors will be selected. The recommendation of first

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

page was implemented by using the k-Nearest Neigbor Classification model.

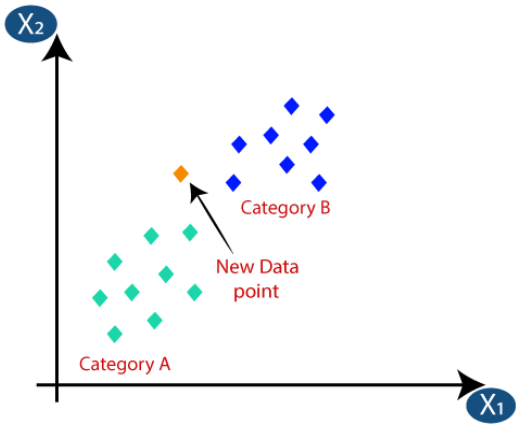


Figure 1: KNN model

4.5.2 Decision Tree Classifier

Decision Trees have emerged as a model-based approach for developing recommender systems. Using decision trees for building recommendation models offers several advantages, including efficiency, interpretability, and flexibility in handling various input data types. The decision tree constructs a predictive model that links the input to a forecasted value based on the input's attributes. Each interior node in the tree corresponds to an attribute, and each arc from a parent to a child node represents a possible value or set of values of that attribute. The tree construction starts with a root node and the input set. An attribute is assigned to the root, and arcs and sub-nodes are generated for each set of values. The input set is then divided by the values, ensuring that each child node only receives the part of the input set that matches the attribute value specified by the arc to the child node. The process is then repeated recursively for each child until further splitting is no longer feasible. Therefore, the decision tree classifier is a valuable tool for developing recommender systems.[7]. This model is implemented for the other Spotify account page recommendations.

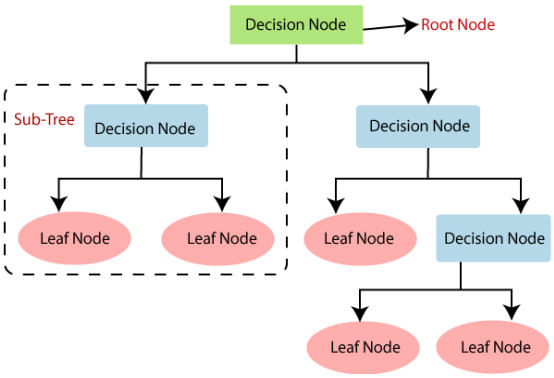


Figure 2: Decision Tree Classifier model

4.5.3 Random Forest Classifier

Random forest is a highly accurate method that operates efficiently on large datasets. It has the capability to predict missing data with high precision, even when a substantial portion of the data is missing without any preprocessing. Random forest combines bagging and random feature selection techniques to generate decision trees, which are then combined with individual learners. A random subset of training data is used to generate these trees. Once the forest has been trained, the test rows are passed through it, and each tree generates an output class. The output class for the random forest is determined by taking the mode of these classes. This is how a random forest classifier operates in a music recommendation system[8].

Since, random forest classifier is an ensemble learning technique in machine learning, it is widely used for the

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

recommendations for ots higher accuray and performance.

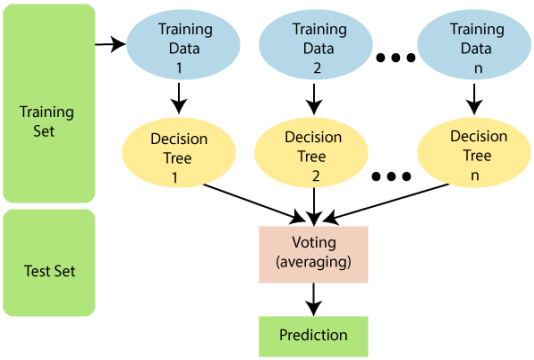


Figure 3: Random Forest Classifier model

5. Implementation

To implement the proposed solution, Python language is being utilized due to its various modules and packages. The implementation of one approach includes the usage of kNN and Random Forest Classifier, while the other approach uses Decision Tree Classifier in combination with the aforementioned machine learning models. A numerical dataset is applied to all the above models, which results in the creation of two user pages. One page is dedicated to inputting the genre, year range, and basic music features. In case the user is not interested, he/she can navigate to another page, which provides recommendations based on the user's previous listening history, preferences, and past data available in their Spotify account.

Pseudo Code:

Input: User Input (genre, year range, features)

Output: Outputs recommended songs

Begin

```
# Define a size for your train set
train_size = int(0.8 * len(df))
# Split dataset
train_set = shuffle_df[:train_size]
test_set = shuffle_df[train_size:]
y = train_set.favorite
# Train / Split Data
oversample = SMOTE()
X_train, y_train =
oversample.fit_resample(X, y)
#Decision Tree Classifier Model
dt = DecisionTree Classifier
(max_depth=30).fit(X_train, y_train)
#Random Forest Classifier
rf = RandomForestClassifier(n_estimators
= 10, max_depth = 20).fit(X_train, y_train)
```

End

6. Results & Conclusion

Upon implementation using the above models, the recommended songs will be generated based on the users taste. The user can simply input the genre, year range and features (if he wishes) to get the recommended songs. He can also be recommended with the songs by redirecting towards the Spotify page of the particular user based on his previous listening history, preferences and song data.

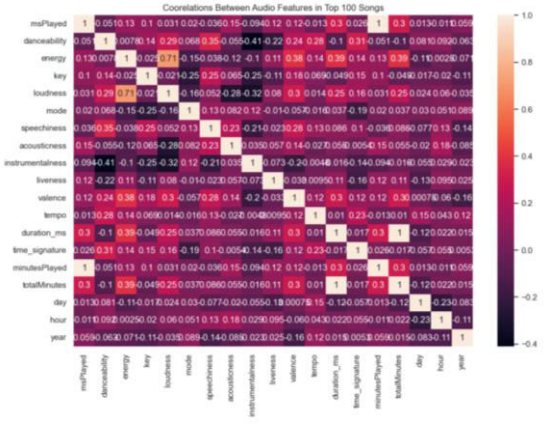


Figure 4: Confusion matrix for all the features of the dataset.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

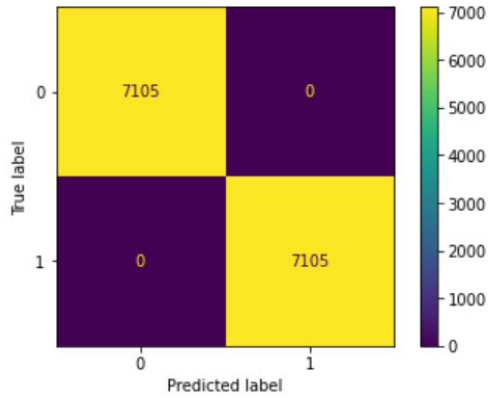


Figure 5: Confusion matrix for Decision Tree Classifier Model.

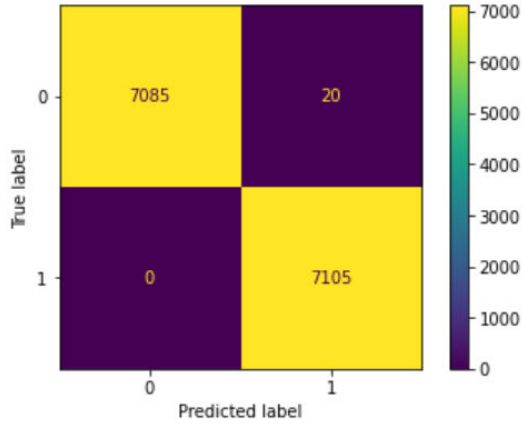


Figure 6: Confusion matrix for Random Forest Classifier Model.

Figure 7: The user application page

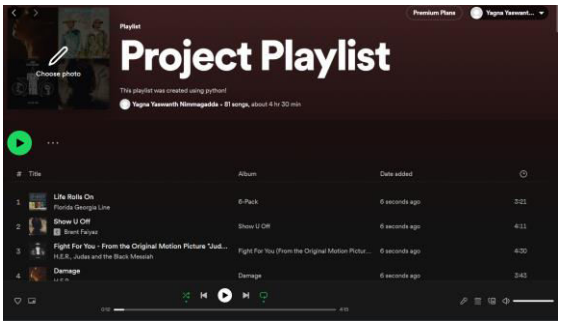


Figure 8: User Application Page on Spotify.

7. Limitations & Future Scope

This approach is implemented for a particular user with a spotify account as a personalised music recommendation.

Since it is for a particular user, the information has to be updated dynamically within the implementation. Later, there is a possibility for the development of this project as it can be improved by extending to all the users by taking the user's spotify information into account.

References

[1] Pengfei Sun (2022). Music Individualization Recommendation System Based on Big Data Analysis, Comput Intell Neurosci, Hindawi Limited.

[2] Achin J., Vanita J., Nidhi K.(2016) A LITERATURE SURVEY ON RECOMMENDATION SYSTEM BASED ON SENTIMENTAL ANALYSIS, Advanced Computational Intelligence: An International Journal (ACII), Vol.3, No.1

[3] Varsha V., Ninad M., Parth., Dr. Prashant N., (2021, November). Music Recommendation System Using Machine Learning, International Journal of Scientific Research in Computer Science

Engineering and Information Technology, ISSN : 2456-3307, doi : CSEIT217615.

[4] Ramasuri, A., Ajay, K., Bhoomireddy, P., Athukumsetti, J., Achanta, S., Mudunuru, V., (2021, July), Music Recommendation System With Advanced Classification, IJERT Volume 10, Issue 7.

[5] Gongde, G., Hui, W., David, B., Yaxin, B., Kieran, G. (2004, August). KNN Model-Based Approach in Classification. In School of Computer Science, Queen's University Belfast.

[6] Berrimi, M., Hamdi, S., Cherif, R. Y., Moussaoui, A., Oussalah, M., & Chabane, M. (2021, March). COVID-19 detection from Xray and CT scans using transfer learning. In *2021 International Conference of Women in Data Science at Taif University (WiDSTaif)* (pp. 1-6). IEEE.

[6]2. H. W., Nearest Neighbours without k: A Classification Formalism based on Probability, technical report, Faculty of Informatics, University of Ulster, N.Ireland.

[7] Amir, G., Amnon, M., Lior, R., Alon, S., Arnon, S. A Decision Tree Based Recommender System.

[8] Kasimalla, B., Dr. K, V., (2019, June). An Efficient Recommendation System using Random Forest in Machine Learning. In 2019 JETIR June 2019, Volume 6, Issue 6.

[9] Knees, Peter, and Markus, S., ”A survey of music similarity and recommendation from music context data.” ACM Transactions on Multimedia Computing, Communications, and Applications 10,2.

[10] Ferretti, N., Stefano, S, ”Clustering of Musical Pieces through Complex Networks: an Assessment over Guitar Solos.” IEEE MultiMedia (2018).

[11]Millecamp, Martijn, et al. ”Controlling Spotify recommendations: effects of personal characteristics on music recommender user Interfaces.” Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization. ACM, 2018.

[12]O Bryant, Jacob. ”A survey of music recommendation and possible improvements.” (2017).

[13]Schedl, Markus, Arthur Flexer, and Julin Urbano. ”The neglected user in music information retrieval research.” Journal of Intelligent Information Systems 41.3 (2013): 523-539.