



GENRE PREDICTION FOR MUSIC RECOMMENDATION USING MACHINE LEARNING

Arpit Seth

Jain (Deemed-to-be) University, Bangalore, Karnataka, India

Article DOI: <https://doi.org/10.36713/epra4283>

ABSTRACT

Music applications are one of the most used applications in the world. Consumers can hear the song they like but difficult for them to find songs from the vast number of songs list. The flow of this paper is to increase the efficiency of music recommendation in terms of the genre based on the decision-tree which helps the users to get the music according to their preferences.

This model uses age and gender as an input set and genre as output. The model will predict the genre according to age and gender and the decision tree helps to reduce the complexity of the model.

1. INTRODUCTION

For every person music has a very special place in his heart and today's internet world made music easily accessible to people, nowadays people do not need to go a specific place to listen to music, they can listen anywhere easily in their mobile and they can even download in minimal charges.

Advanced technology has become the integral part of our life. To satisfy the need of the society, almost in each work, we use the technology. In current era computer science is major subject. It has many real life applications such as cloud computing [1], artificial intelligence [2], remote monitoring [3], internet of things [4, 5, 6, 7, 8, 9, 10, 11], SPP [12, 13, 14, 15, 16, 17, 18, 19], TP [20, 21, 22], internet Security [23], uncertainty [24, 25, 26] and so on. Technology is the mode by which user can store, fetch, communicate and utilize the information. So, all the organizations, industries and also every individual are using computer systems to preserve and share the information. The internet security plays a major role in all computer related applications. The internet security appears in many real-life applications, e.g., home security, banking system, education sector, defence system, Railway, and so on. In this manuscript we discuss about the protection of authentication which is a part of internet security.

Today internet world made listening music very cheap in old day's people has to buy Audio Cassette and music CD in addition with specific player and sometimes Audio Cassette and music CD of a company is not compatible with different company but now internet world made music easily accessible many company like gaana.com, Spotify, wink music, YouTube music etc.

They charge on monthly and yearly based with very minimal price and has many plans like family plan and student plan and some even give this feature for free.

Music is one of the most used mediums for entertainment. Music genre is used for optimizing the time of the users. Genre helps to classify the vast number of tracks and allocate them in the different categories. In many different ways, music can be divided into different genres, such as in mainstream music and art music, or religious music and secular music.

Genre helps the user to easily shortlist the track or music of their interest. Music is one of the vast spread industry shortlisting music tracks according to user's preference is difficult. The task is to classify various properties of provided music recordings, based on extracted sound characteristics. Music genres forecast: In this challenge, we invite you to create an



algorithm to identify the art genre of certain music track fragments

In this paper the dataset is divided into age and gender as an input set and genre is predicted as output set. In gender column “male” is denoted as ‘1’ and “female” is denoted as ‘0’. The data is used for prediction of genre based on the dataset. This helps the customer to reduce the time and can easily search for the music from the massive amount of tracks.

In a music streaming app or web application age and gender of user is input at the time of registration. With the help of algorithm, age and gender as an input genre will be predicted and music tracks are recommended according to the genre.

Decision Tree algorithm is part of the supervised learning algorithms family. Unlike other supervised learning algorithms, algorithms for decision tree can also be used to solve regression and classification problems. Decision tree helps in prediction on the given input set by certain patterns while model training.

Forms of Decision Trees Decision trees Styles are based on what kind of goal variable we have. It may be of two types:

- Categorical Variable Decision Tree: Decision Tree that has a categorical goal variable is labelled as Categorical Variable Decision Tree.
- Continuous Variable Decision Tree: Decision Tree has a constant target variable and is then called Continuous Variable Decision

The Pseudocode of Decision Tree Algorithm

- Place the best attribute of the dataset at the **root** of the tree.
- Split the training set into **subsets**. Subsets should be made in such a way that each subset contains data with the same value for an attribute.
- Repeat step 1 and step 2 on each subset until you find **leaf nodes** in all the branches of the tree.

This helps in reducing the complexity of the music streaming app. All the data is stored and the music is recommended to make the application user friendly. In order to predict the music our work recorded the data from users for better recommendation by applying feedback from users and learning their music listening behaviours.

1.1. MOTIVATION

In some papers we have found some problems searching for tracks. In our proposed work we overcome the drawback in other application. The main drawback is that recommendation of music is done according to common genre and languages. This technique is not sufficient of optimization. Even though several music recommendation systems have overcome much of the consumer preference issues, issues have been missed is recommendation of music with respect to age factor of user.

2. LITERATURE REVIEW

Generally, the recommendation system serves as information filtering and customization tool. The recommendation in our work sorts the music playlists according to user’s age. With the help of decision tree algorithm. Some scholars proposed recommendation based on content. In these type of proposed work recommendation is generated by categorizing user’s content. This works by extracting the previous preferred content. M. Daniel and J. C. Sally [27] uses lyrics features in music as content to identify the pattern of user for recommendation but that causes conflict and social issues. Steffen P. [28] proposed to create content-based analysis based on user’s playlist according to customer playlist and feedback. Advantage to content based filtering is that it is capable of recommending similar music tracks. No requirement of others profiles and feedback. The main disadvantage to content based filtering is that the content feature cannot be presented in application.

Collaborative filtering is another type of technique used by scholars for effective recommendation. This filtering approach uses correlation of users according to their profile. This filtering technique is widely used in music recommendation systems. The uses with common interest are used to recommend each other. M. Balabanovic and Y. Shoham [29] uses automated web browser of experimenting this approach. J. L. Herlocker and J. A. Konstan [30] uses task force recommendation. Task force recommendation assigns specific task. This uses both interest rating and interest association. The advantages of these type of filtering is it avoid inaccurate prediction of music items by trusting on other users browsing experience. Recommendation can be done for the users those content is not easy to analyse. In spite of the fact that collaborative filtering is widely used there are many problems which remain unsolved such as it is difficult to start recommendation which result as not an optimum performance.

Due to this advantages and disadvantages of both filtering technique. Some researchers proposed a hybrid filtering technique which combines both content based analysis and collaborative based analysis technique. Claudiu, W. N. Firan and P. Raluca [31] proposed a system in which users are collaboratively grouped together based on users rating and profile. And they recommendation is done based on content this reduce the time. H. Wang, W. Wang, J. Yang and P. S. Yu [32] uses clustering technique to form similar patterns among the users. The main advantage is that combination of these filtering technique effectively reduces the time for calculation.

2.1. DATASET DESCRIPTION

Dataset used in this project is classified in three column that is age, gender and genre. Data is collected by different survey and through feedback taken form

people and other online sources. This dataset is divided in two parts input set and output set. Age and gender are taken as input by the customer at the time of registration or login and genre is predicted according to the customer. Then music is recommended according to the prediction made by the proposed work.

Gender is denoted as 1 for **male** 0 for **female**. Genre which are used for working are Hip Hop, Jazz, classical.

This table is named as “music.csv”

Age	Gender	genre
20	1	Hip hop
23	1	Hip hop
21	0	Hip hop
24	0	Hip hop
32	1	classical
34	0	classical

Table 1. Data Set

central architecture, in which music recommendation is done based on age. During the lack of initial recommendation there is poor results in recommending music to new users in others work to overcome this problem we used age as a factor for recommendation. Further users’ feedback can be acquired to determine the music taste of new user. The core of the recommendation system is decision tree classification. With the help of this classification data set is divided into two tables. One is input set with consist of age and gender and other is output set with consist on genre. This work is divided into 2 phases.

In first phase gathering user’s personal information at the time of registration. Information such as age, gender and preferred music type. Then with these preferred types of music recommendation is done at initial phase for new uses. This problem of initial recommendation of music to new users is solved in second phase.

3. PROPOSED METHODOLOGY

The study discussed in this paper extends the understanding of decision-tree classification as its

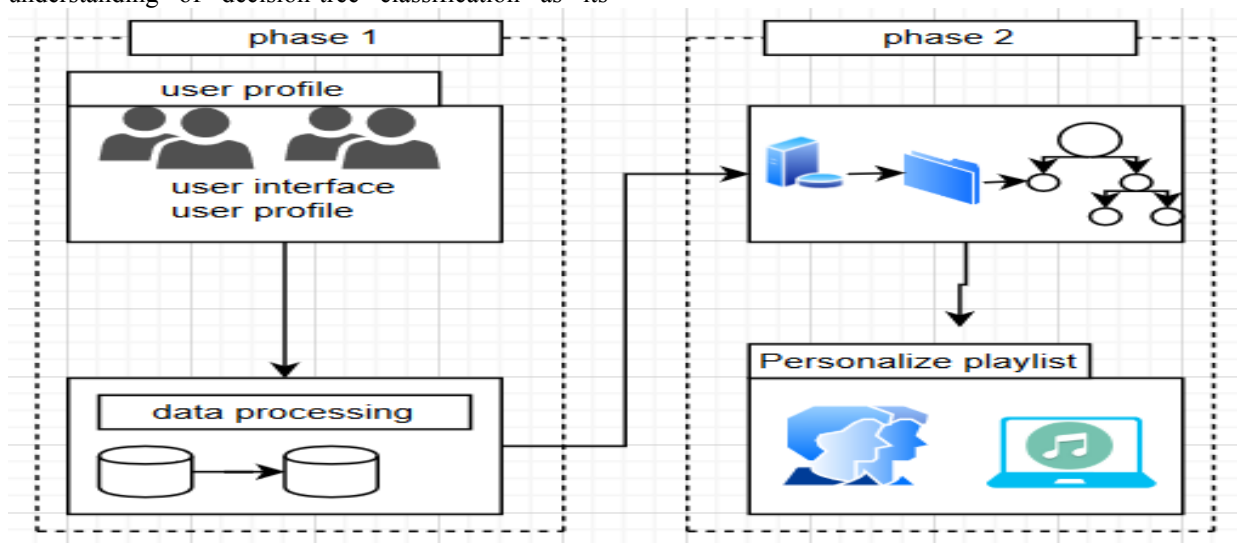


Figure 1 Two phases

In order to solve this problem, the proposed recommendation system uses information (age and gender) of the users which was gathered at the time of registration and predict the genre of the user for personalised user and then music is recommended according to the prediction made. This data used from the feedback of customers helps in making prediction with more accuracy which in results give good performance and optimize time as compared to others work.

3.1. DECISION TREE GENERATION

The allocation of data set to root node the data set is spliced into two different table. This action is

known as splitting action. This action is performed at the time of results generated by root node. For solving the attribute selection problem, researcher works and derive a solution which is getting information gain, Gini index. This function will calculate the value of each attribute and sort them and place them with respect to their values i.e., the attribute with higher value is the root node.

In decision tree classifier the data set is divided into two parts one is training set and other is testing set. This training set is then used of building the model and testing set is then used to test the accuracy of the model. Information gain or IG is a statistical property which measures how well the training examples are

differentiated by a given attribute according to their target classification.

Information gain $(Y, X) = E(Y) - E(Y|X)$
 Mathematical representation of information gain

$$Entropy = -\sum_j p_j \log_2 p_j$$

The Gini index can be interpreted as a cost feature used to determine splits in the dataset. It is determined by subtracting the sum of each class's squared percentages from one.

$$Gini = 1 - \sum_j p_j^2$$

3.2. BUILDING MODEL

Building the model for prediction is the practical approach for testing the accuracy of prediction. In this firstly the csv file of data set is fetched from the database. Then the attribute with high information gain is dropped from the original table. The model is build and then load for testing without keeping the expectation for high accuracy. The goal is to keep the system easy and simple for implementation. Getting of root and leaf nodes will help in understanding the model rules.

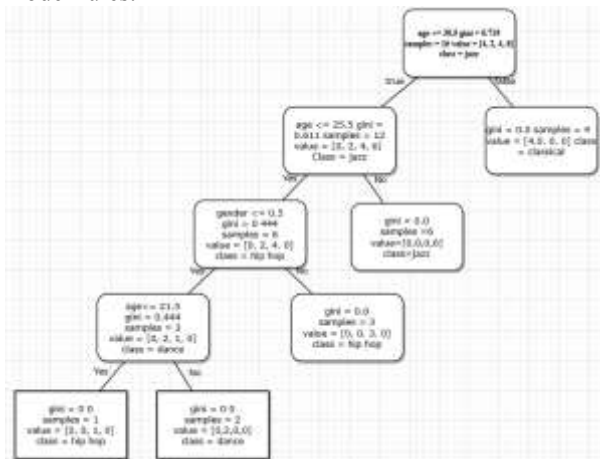


Figure 2 Decision tree algorithm

4. RESULT

In this proposed work the data is divides in two sets and test data set is used for checking the accuracy. The accuracy of model is checked by input the value form test data set and then results are compared to the predicted results.

Accuracy = $(TP + TN) / (\text{Total number of observation})$

```
Out[1]: array(['hip hop', 'dance'], dtype=object)
```

Figure 3 prediction by model

In this work the prediction is made on the basis of data set give with respect to age and gender and predict the genre.

Plot graph representation of accuracy:

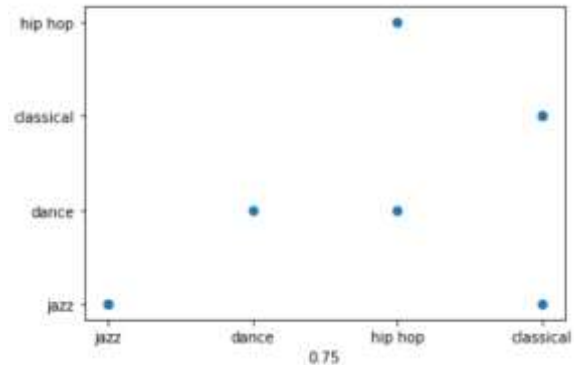


Figure 4 Accuracy Graph

5. CONCLUSION

In this paper, prediction of genre for music recommendation model is simple. Proposed system can be used to reduce problem faced during initial recommendation of music to new user. According to the results the proposed model overcome some drawbacks. The future scope of improvement includes:

1. There is no user interface provided in this project.
2. There can be more factors which will help in recommendation of music more effectively.

REFERENCES

- [1] X. Xu, "From cloud computing to cloud manufacturing," *Robotics and Computer-Integrated Manufacturing*, vol. 28, pp. 75-86, 2012.
- [2] M. Haenlein and A. Kaplan, "A brief history of artificial intelligence: on the past, present, and future of artificial intelligence," *California Management Review*, vol. 61, pp. 5-14, 2019.
- [3] H. Mohapatra, S. Panda, A. Rath, S. Edalatpanah and R. Kumar, "A tutorial on powershell pipeline and its loopholes," *International Journal of Emerging Trends in Engineering Research*, vol. 8, no. 4, 2020.
- [4] H. Mohapatra, "HCR using neural network," 2009.
- [5] H. Mohapatra and A. Rath, "Detection and avoidance of water loss through municipality taps in india by using smart tap and ict," *IET Wireless Sensor Systems*, vol. 9, no. 6, pp. 447-457, 2019.
- [6] H. Mohapatra and A. Rath, "Fault tolerance in WSN through PE-LEACH protocol," *IET Wireless Sensor Systems*, vol. 9, no. 6, pp. 358-365, 2019.
- [7] H. Mohapatra, S. Debnath and A. Rath, "Energy management in wireless sensor network through EB-LEACH," *International Journal of Research and Analytical Reviews (IJRAR)*, pp. 56-61, 2019.
- [8] V. Nirgude, H. Mahapatra and S. Shivarkar, "Face recognition system using principal component analysis & linear discriminant analysis method simultaneously with 3d morphable model and neural network BPNN method," *Global Journal of Advanced Engineering Technologies and Sciences*, vol. 4, p. 1, 2017.
- [9] M. Panda, P. Pradhan, H. Mohapatra and N. Barpanda,



- "Fault tolerant routing in heterogeneous environment," *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH*, vol. 8, pp. 1009-1013, 2019.
- [10] H. Mohapatra and A. Rath, "Fault-tolerant mechanism for wireless sensor network," *IET Wireless Sensor Systems*, vol. 10, no. 1, pp. 23-30, 2020.
- [11] D. Swain, G. Ramkrishna, H. Mahapatra, P. Patra and P. Dhandrao, "A novel sorting technique to sort elements in ascending order," *International Journal of Engineering and Advanced Technology*, vol. 3, pp. 212-126, 2013.
- [12] S. Broumi, A. Dey, M. Talea, A. Bakali, F. Smarandache, D. Nagarajan, M. Lathamaheswari and R. Kumar, "Shortest path problem using Bellman algorithm under neutrosophic environment," *Complex & Intelligent Systems*, vol. 5, pp. 409--416, 2019.
- [13] R. Kumar, S. Edalatpanah, S. Jha, S. Broumi, R. Singh and A. Dey, "A multi objective programming approach to solve integer valued neutrosophic shortest path problems," *Neutrosophic Sets and Systems*, vol. 24, pp. 134-149, 2019.
- [14] R. Kumar, A. Dey, F. Smarandache and S. Broumi, "A study of neutrosophic shortest path problem," in *Neutrosophic Graph Theory and Algorithms*, F. Smarandache and S. Broumi, Eds., IGI-Global, 2019, pp. 144-175.
- [15] R. Kumar, S. Edalatpanah, S. Jha and R. Singh, "A novel approach to solve gaussian valued neutrosophic shortest path problems," *International Journal of Engineering and Advanced Technology*, vol. 8, pp. 347-353, 2019.
- [16] R. Kumar, S. Edalatpanah, S. Jha, S. Gayen and R. Singh, "Shortest path problems using fuzzy weighted arc length," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, pp. 724-731, 2019.
- [17] R. Kumar, S. Edalatpanah, S. Jha, S. Broumi and A. Dey, "Neutrosophic shortest path problem," *Neutrosophic Sets and Systems*, vol. 23, pp. 5-15, 2018.
- [18] R. Kumar, S. Jha and R. Singh, "A different approach for solving the shortest path problem under mixed fuzzy environment," *International Journal of fuzzy system Applications*, vol. 9, no. 2, pp. 132-161, 2020.
- [19] R. Kumar, S. Jha and R. Singh, "Shortest path problem in network with type-2 triangular fuzzy arc length," *Journal of Applied Research on Industrial Engineering*, vol. 4, pp. 1-7, 2017.
- [20] R. Kumar, S. Edalatpanah, S. Jha and R. Singh, "A Pythagorean fuzzy approach to the transportation problem," *Complex and Intelligent System*, vol. 5, pp. 255-263, 2019.
- [21] J. Pratihari, R. Kumar, A. Dey and S. Broumi, "Transportation problem in neutrosophic environment," in *Neutrosophic Graph Theory and Algorithms*, F. Smarandache and S. Broumi, Eds., IGI-Global, 2019, pp. 176-208.
- [22] J. Pratihari, S. E. R. Kumar and A. Dey, "Modified Vogel's Approximation Method algorithm for transportation problem under uncertain environment," *Complex & Intelligent Systems (Communicated)*.
- [23] J. Sakhnini, H. Karimipour, A. Dehghantanha, R. Parizi and G. Srivastava, "Security aspects of Internet of Things aided smart grids: A bibliometric survey," *Internet of Things*, pp. 100-111, 2019.
- [24] S. Gayen, F. Smarandache, S. Jha and R. Kumar, "Interval-valued neutrosophic subgroup based on interval-valued triple t-norm," in *Neutrosophic Sets in Decision Analysis and Operations Research*, M. Abdel-Basset and F. Smarandache, Eds., IGI-Global, 2019, p. 300.
- [25] S. Gayen, F. Smarandache, S. Jha, M. Singh, S. Broumi and R. Kumar, "Introduction to plithogenic subgroup," in *Neutrosophic Graph Theory and Algorithm*, F. Smarandache and S. Broumi, Eds., IGI-Global, 2020, pp. 209-233.
- [26] S. Gayen, S. Jha, M. Singh and R. Kumar, "On a generalized notion of anti-fuzzy subgroup and some characterizations," *International Journal of Engineering and Advanced Technology*, vol. 8, pp. 385-390, 2019.
- [27] McEnnis, Daniel, Cunningham and S. Jo, "Sociology and Music Recommendation Systems," *ISMIR*, pp. 185--186, 2007.
- [28] Pauws, Steffen, v. d. Wijdeven and Sander, "User Evaluation of a New Interactive Playlist Generation Concept," *ISMIR*, pp. 638-643, 2005.
- [29] Balabanovic, Marko, Shoham and Yoav, "Learning information retrieval agents: Experiments with automated web browsing," *On-line Working Notes of the AAAI Spring Symposium Series on Information Gathering from Distributed, Heterogeneous Environments*, pp. 13-18, 1995.
- [30] Herlocker, L. Jonathan, Konstan and A. Joseph, "Content-independent task-focused recommendation," *IEEE Internet Computing*, vol. 5, pp. 40-47, 2001.
- [31] Firan, C. S, Nejd, Wolfgang, Paiu and Raluca, "The benefit of using tag-based profiles," *Latin American Web Conference (LA-WEB 2007)*, pp. 32--41, 2007.
- [32] Wang, Haixun, W. Wang, Yang, Jiong, Yu and S. Philip, "Clustering by pattern similarity in large data sets," *Proceedings of the 2002 ACM SIGMOD international conference on Management of data*, pp. 394-405, 2002.