

Analysis of Music Recommendation System using Machine Learning Algorithms

Prachi Singh¹, Mr. Pankaj K. Singh², Mr. Amit Ganguli³, Mr. Ajit Shrivastava⁴

¹M.Tech Scholar, Department of CSE, Sagar Institute of Science, Technology & Research Bhopal, (M.P.), India
^{2,3}Assistant Professor, Department of CSE, Sagar Institute of Science, Technology & Research Bhopal, (M.P.), India
⁴Professor & Head, Department of CSE, Sagar Institute of Science, Technology & Research Bhopal, (M.P.), India

ABSTRACT:- Today's World is known as Data World. Generation of Data is abundant in volume. Demand of useful data is every-where every business domain works upon data this day. Data mining played very important roles from last decade now this data mining is migrated to Machine Learning. Machine Learning comes by Artificial Intelligence and Mathematical Statistics. Machine learning is categorized into supervised, unsupervised and reinforcement. Supervised machine is using the learning algorithm to detect the discovery that is clearly due to the examples supplied to produce general interpretation, which then predicts future scenarios or events. In this Dissertation explains with the advancement of technology in music players, especially in intelligent cell phones, users have access to large archives. The quick and easy selection of favourite music in these large archives becomes one of the biggest problems for users. For example, choosing music in a silent forest is different from a crowded street or feelings of listening to music in the morning of a working day are different from a holiday afternoon. In this paper, a system is designed that it collects users' reference information such as weather, temperature, geographic location, etc., and according to their weighted composition, it recommends a suitable music that is currently available to the user. Here we analyse the music portal KKBOX in this work we analyse the different dependent attributes their behaviour and others. Finally, we apply two algorithms Random Forest & XGBClassifier.

Keywords: Classification Algorithm, Machine Learning, Deep Learning, Decision Tree & Random Forest, ensemble Techniques, CNN.

1. INTRODUCTION

In some previous years we all are talking about machine intelligent programs, and what this factor for our health issues, richness. Now days , machine related intelligent learnt program possibly pledge to protect growth of

population, address worldwide challenges such as environment change, and add much of income to the world's economy through expanding potency; so that it also essentially make changes in the nature of work, and Size, or define, do the choices in everyday life. Between these extremes, there is a potentially variable technology, which brings both its opportunities and challenges, and whose risks and benefits are needed to navigate because it is used to become more central to everyday activities.

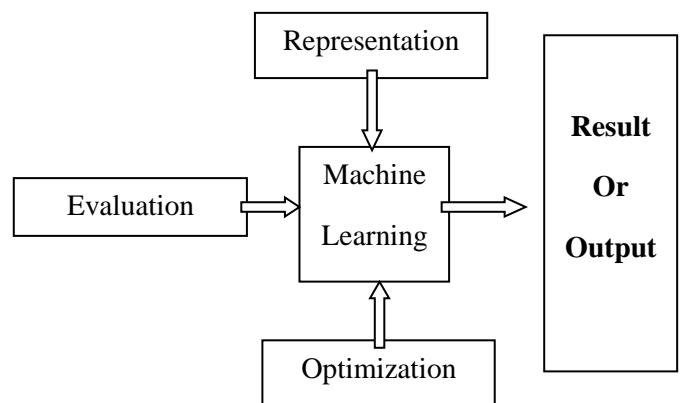


Figure 1: Major Component in Machine Learning

1.1 Data Mining Process Step

Data cleaning: It is a process of removing noise and inconsistent data.

Data integration: In this step data from multiple sources are combined.

Data selection: In this step data relevant for mining task is selected.

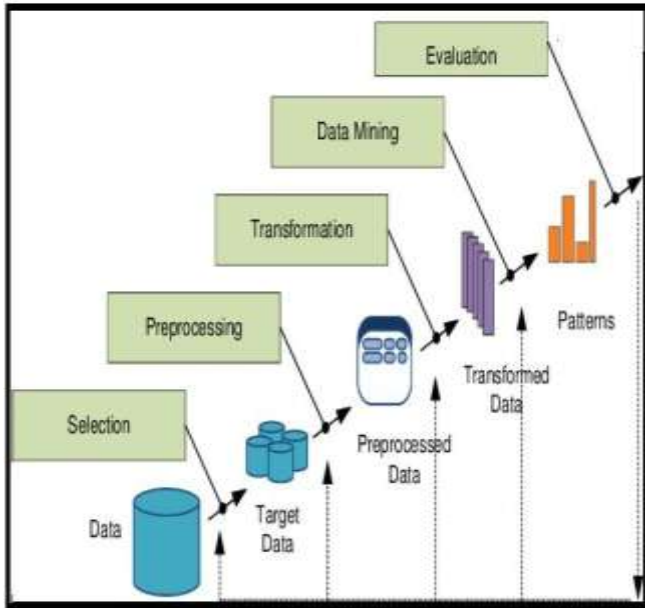


Figure 2: Process of Data Mining

Data transformation: In this step data will be transformed into form that is appropriate for mining.

Data mining: In this step some intelligent methods are applied for extracting data patterns.

Pattern evaluation: In this step we concentrate upon important patterns representing knowledge based on some measure are identified.

Knowledge presentation: In this step visualization and knowledge representation techniques are used to present the mined knowledge to the user.

1.2 Data Mining Techniques

Data mining Algorithms is categorized into different which is given below:

Classification

Classification is the frequently (most commonly) applied data mining mechanism, which explains a set of pre-classified examples to develop a (procedure) model that can (identifies or categories) classify the population (Dataset) of records at large.

Clustering

Clustering can be said as to find out of similar classes of objects. By using clustering mechanism, we can further find out dense and sparse (n Dimensional Space) regions in object space and can discover overall distribution trends (pattern) and relation among many coordinate points (correlations) among data attributes.

Association

Clustering Association play very important roles when we have abundant data with similar properties means with the above mechanism, we can divide the data which has similar properties associated with another object.

2. LITERATURE SURVEY

According to the paper, personalized music recommendation system (PMRS) based on the convolutional neural networks (CNN) approach. The CNN approach classifies music based on the audio signal beats of the music into different genres. In PMRS, we propose a collaborative filtering (CF) recommendation algorithm to combine the output of the CNN with the log files to recommend music to the user. The log file contains the history of all users who use the PMRS. The PMRS extracts the user's history from the log file and recommends music under each genre. We use the million-song dataset (MSD) to evaluate the PMRS. To show the working of the PMRS, we developed a mobile application (an Android version). We used the confidence score metrics for different music genre to check the performance of the PMRS. [1].

The personalized music recommender supports the user-favorite songs stored in a huge music database. In order to predict only user-favorite songs, managing user preferences information and genre classification are necessary. In our study, a very short feature vector, obtained from low dimensional projection and already developed audio features, is used for music genre classification problem. We applied a distance metric learning algorithm in order to reduce the dimensionality of feature vector with a little performance degradation. We propose the system about the automatic management of the user preferences and genre classification in the personalized music system. [2].

3. PROBLEM IDENTIFICATION

When we studied many Product Recommendation System related papers, we found that many techniques are available to find the prediction. We read number of research papers and find that every algorithm has some specification in terms of their performance. Author decided that he will give an algorithm which has better Accuracy. Since we know that deep learning or neural network gives better result in terms of performance so we will apply Random Forest & SGDCClassifier for better performance.

4. Block Diagram & Methodology

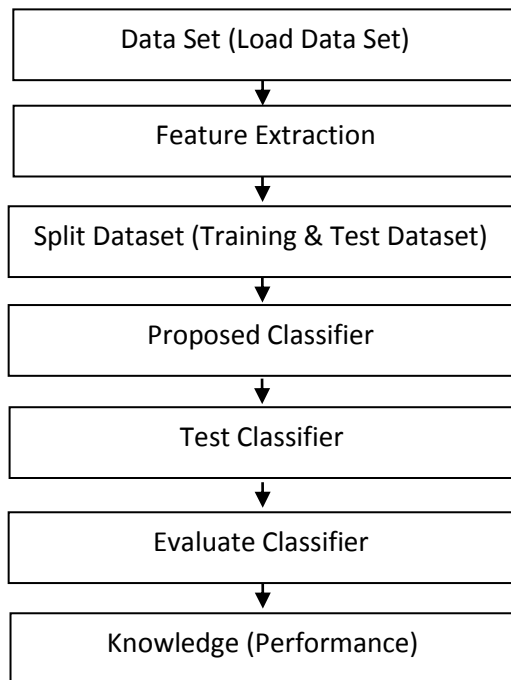


Figure 6: Block Diagram of Flow of Operation

5. Experimental Result

We used python programming Language to implement our logic we used number of libraries like NumPy, pandas, tweepy, matplotlib, seaborn and many more. This project is divided into two parts in first part we fetch data in another part we will process our data our overall process is given below:

Table01: Comparison of Different Algorithms

Algorithms	Train Size	Test Size	Accuracy
Random Forest	70	30	0.74
XGBClassifier	70	30	0.74
Random Forest	80	20	0.72
XGBClassifier	80	20	0.68
Random Forest	75	25	0.75
XGBClassifier	75	25	0.72

Explanation: In the above Table we compare to Different Algorithms with two different Dataset Splits.

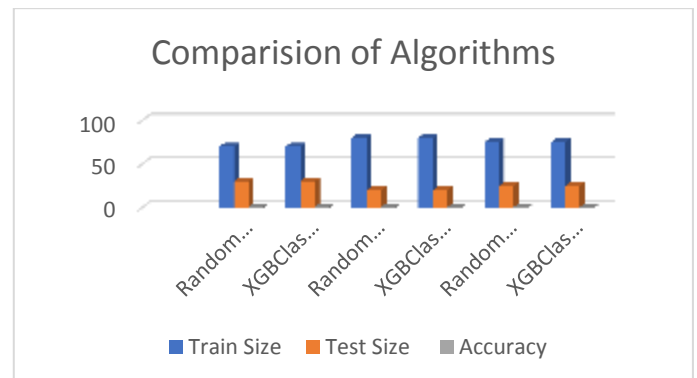


Figure 3: Graph Comparison

Explanation: By the analysis of above graph we can say that splitting of training and testing data may change the accuracy which is given above in terms of graphical values. We split our data into three different slot 70-30,80-20 & 75-25 . In every split we found the different values.

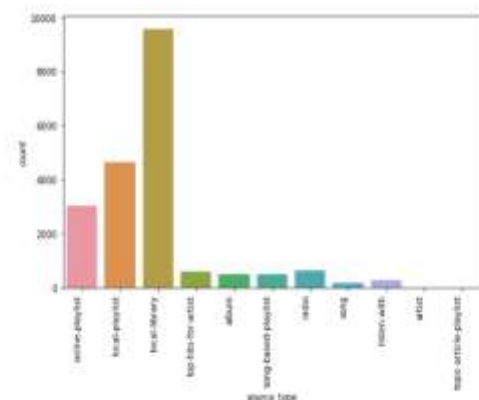


Figure 4: Music Data Comparison

Explanation: In the above figure where we analyze from train dataset where we concentrate over Source-type we found different attribute values like online-playlist, local-playlist , local-library ,top-hits-for-artist ,album ,song-based-playlist,radio,song,listen-with,artist. By data visualization we can say that in source-type attributes maxium values is local-library type.

6. CONCLUSION

We conclude the observation about the previous technique. Our observation different terms and condition. This represents our work in a new approach. In our work we have tried to improve Accuracy we applied ensemble techniques where we applied two different algorithms i.e. Random Forest algorithm and XGBClassifier algorithm in every algorithm we got different Accuracy %. Again, we adjust training and testing splitting (% values) by this mechanism we get better result.

REFERENCES

- [1] Z. Hyung, J. S. Park, and K. Lee, "Utilizing context-relevant keywords extracted from a large collection of user-generated documents for music discovery," *Info. Processing and Management*, vol. 53, no. 5, pp. 1185-1200, 2017.
- [2] J. B. Schafer, D. Frankowski, J. Herlocker, and S. Sen, "Collaborative filtering recommender systems," *The Adaptive Web: Methods and Strategies of Web Personalization*, pp. 291-324, 2007.
- [3] M. J. Pazzani and D. Billsus, "Content based recommendation systems," *The Adaptive Web: Methods and Strategies of Web Personalization*, pp. 325-341, 2007.
- [4] E. J. Humphrey, J. P. Bello, and Y. LeCun, "Moving beyond feature design: deep architectures and automatic feature learning in music informatics," in *Proc. 13th Int'l Conf. Music Info. Retrieval*, pp. 403-408, October 2012.
- [5] C. W. Hsu, and C. J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Tran. Neural Networks*, vol. 13, no. 2, pp. 415-425, 2002.
- [6] R. E. Fan, K. W. Chang, C. J. Hsieh, X. R. Wang, and C. J. Lin, "LIBLINEAR: a library for large linear classification," *J. Machine Learning Research*, vol. 9, pp. 1871-1874, 2008.
- [7] A. V. Oord, S. Dieleman, and B. Schrauwen, "Deep content-based music recommendation," in *Proc. 26th Int'l Conf. Neural Info. Process. Systems*, pp. 2643-2651, December 2013.
- [8] J. Salamon, and J. P. Bello, "Deep convolutional neural networks and data augmentation for environmental sound classification," *IEEE Sig. Processing Letters*, vol. 24, no. 3, pp. 279-283, March 2017.
- [9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Info. Processing Sys.*, pp. 1097-1105, 2012.
- [10] R. Jozefowicz, W. Zaremba, and I. Sutskever, "An empirical exploration of recurrent network architectures," in *Int'l Conf. Machine Learning*, pp. 2342-2350, June 2015.
- [11] J. S. Downie, "Music information retrieval," *Annual Review of Information Science and Technology*, 37:295-340, 2003.
- [12] R. Typke, F. Wiering, and R. Veltkamp, "A survey of music information retrieval systems," *Proc. ISMIR*, pp. 153-160, 2005
- [13] D. Jang, C.-J. Song, S. Shin, S.-J. Park, S.-J. Jang, S.-P. Lee, "Implementation of a matching engine for a practical query-by-singing/humming system," *Proc. ISSPIT*, pp. 258-263, 2011
- [14] S. W. Hainsworth and M. D. Macleod, "Particle filtering applied to musical tempo tracking," *EURASIP J. Applied Signal Processing*, vol. 15, pp. 2385-2395, 2004
- [15] S. Jo and C. D. Yoo, "Melody extraction from polyphonic audio based on particle filter," *Proc. ISMIR*, pp. 357-362, 2010