International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.6654
***Article Received**: 28 February 2023 **Revised**: 18 April 2023 **Accepted**: 28 April 2023*

Recommendation of Algorithm for Efficient Retrieval of Songs from Musical Dataset

Swathy Vodithala[1], Vaishnavi Gudimalla[2], Y. Bhavani[3], Preethi Madadi[4], Mohammed Sharfuddin Waseem[5]
[1]Associate Professor: Dept. of CSE(N),Kakatiya Institute of Technology and Science
Warangal,India
swathyvodithala@gmail.com
[2]Student: Dept. of CSE(IoT),Kakatiya Institute of Technology and Science
Warangal,India
vaishnavi.gudimalla5@gmail.com
[3]Associate Professor: Dept. of IT,Kakatiya Institute of Technology and Science
Warangal,India
yerram.bh@gmail.com
[4]Assistant Professor: Dept. of CSE,Kakatiya Institute of Technology and Science
Warangal,India
Preethi.muppidi@gmail.com
[5]Assistant Professor: Dept. of CSE,Kakatiya Institute of Technology and Science
Warangal,India
waseem7602@gmail.com

Abstract—Now-a-days, the research is more towards the entertainment like music, songs, movies, etc. There are many existing works that suggest good songs, movies to people depending on their mood, nature and time that has been savior for the society during the days of lockdown. The existing algorithms used in the literature for basic clustering are K-means, TSNE (T- distributed Stochastic Neighborhood Embedding), PCA (Principal Component Analysis).In this paper, the music dataset considered, consists of songs with attributes as song name, genres, artists, mode, tempo, valence, year, liveness, loudness, popularity, acousticness, danceability, duration, energy, explicit, instrumentalness, key. The important feature is extracted from the other features with the support of literature survey i.e., number of music listeners, types of the songs and type of the music. Later, the dataset is divided into clusters using traditional technique that is k-means based on genre, an important attribute which is selected from the above attributes. The different classifier models like Random Forest, Extra Trees, LightGBM, XGBoost, CatBoost classifier are applied on the clustered dataset and the results have been evaluated on each individual algorithm. Thus the paper recommends not only the group of relevant songs but also suggests the best accurate classification algorithm that can be used for any mentioned musical dataset. The paper also compares all the said ensemble algorithms by calculating the precision, recall, f1-score and support. The accuracy is also calculated for all said ensemble algorithms and based on the accuracy the best suitable algorithm is suggested.

Keywords-Musical dataset; Ensemble learning; Random Forest; Extra trees; LightGBM; XGBoost; CatBoost.

I. INTRODUCTION

All Musical datasets have become increasingly important in today's world, where music is a ubiquitous part of daily life. A musical dataset is a collection of metadata such as song titles, artists, genres, and other attributes that describe the music. Musical datasets provide a rich source of information, enabling them to better understand the characteristics of different songs and genres and make more accurate recommendations.

Musical datasets are also important for music genre classification, where they are used to train machine learning models to recognize the characteristics of different genres based on their audio features such as tempo, rhythm, and melody. Musical datasets are also useful in music emotion recognition, where machine learning models are used to predict the emotional response of listeners to different musical pieces.

Machine Learning is a subset of Artificial Intelligence which uses various algorithms for making predictions, allowing the machine learning model to automatically learn from training data, and predicts the outcome of new data based on historical data.

The machine learning techniques are classified into supervised, unsupervised, semi supervised and reinforcement learning which are used in regression, classification models and clustering methods.

Regression is used to predict continuous values like Salary, Age. The process of categorizing the data (which is a

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.6654
***Article Received**: 28 February 2023 **Revised**: 18 April 2023 **Accepted**: 28 April 2023*

supervised machine learning technique) makes it simple to store, organize, and retrieve the data. Classification is used to predict discrete values like Male or Female. Machine learning is mostly seen in Alexa, self-driving cars, recommendation system, spam detections.

II. EXISTING TECHNIQUES

This section discusses few existing techniques used for the classification of datasets like music, movies etc. Some of them are Random Forest classifier, Extra Trees classifier, LightGBM classifier, XGBoost classifier, CatBoost classifier.

A. *Random Forest Classifier*

Random forest is a popular machine learning algorithm used for classification and regression tasks. It is an ensemble learning method that combines multiple decision trees to make predictions for classification tasks. The algorithm works by randomly selecting subsets of the training data and features to build multiple decision trees. The final prediction is made by combining the predictions of all the trees using majority voting. Random forest classifier can handle a large number of input features, missing data, and is resistant to over fitting. It is useful for feature selection also. [9][10]

$$fi_j = \frac{\sum j{:}node\ j\ splits\ on\ feature\ i^{ni}j}{\sum_{k \in all\ nodes} ni_k} \qquad (1)$$

Lin et. al. published a paper "Music Genre Classification using Improved Random Forest Algorithm" that proposes a method for music genre classification using an improved random forest algorithm that incorporates feature selection and parameter optimization. The authors separate several audio components from music signals and evaluate how well their method performs in comparison to other widely used classification techniques. In terms of accuracy, precision, recall, and F1 score, they discover that the modified random forest approach performs better than these other algorithms. They also show that the best performance is achieved when spectral and rhythmic information are combined. Overall, the article offers insights into the particular audio qualities that are most instructive for categorizing music genres and may be helpful for creating intelligent music systems.[1]

Luo et. al. published a paper "Random Forest-Based Musical Emotion Classification" that proposes a method for classifying musical emotion using acoustic features extracted from music signals and a random forest classifier. The authors discover that the random forest classifier beats other widely used algorithms when they compare the performance of their classifier to those algorithms. Additionally, they look into how different feature subsets affect performance and discover that the best performance comes from combining all features. Overall, the paper shows that random forests can be a useful technique for classifying musical emotions and offers insights into the particular acoustic elements most useful for this endeavor.[2]

B. *Extra Trees Classifier*

Extra Trees Classifier is known as Extremely Randomized Trees, is an ensemble learning method that combines multiple decision trees to make predictions for classification tasks. The algorithm is similar to the random forest classifier, but it adds more unpredictability by choosing random thresholds for each feature and splitting each node only using a random subset of characteristics. Particularly for noisy or high-dimensional datasets, this method can lower overfitting and enhance generalization performance.In general, ExtraTreesClassifier is a quick and effective method that performs well in terms of generalization and accuracy across a range of classification problems.

$$\text{Entropy}\ (S) = \sum_{i=1}^{c} -p_i \log_2(p_i) \qquad (2)$$

M.S. Rahman et. al. published a paper "Music Genre Classification using Extra Trees Classifier" that proposes a method for music genre classification using the Extra Trees classifier, which is an extension of the Random Forest algorithm. Using the music dataset, they took out a number of features, including spectral features and Mel-frequency cepstral coefficients (MFCCs), and placed them into the Extra Trees model. The suggested approach was tested on three different music datasets, and the findings revealed that the Extra Trees classifier performed better than other well-known classification algorithms like SVM and k-NN, reaching high accuracy in classifying the musical genres.[3]

S.A. Shinde et. al. published a paper "Music Genre Classification using Ensemble of Extra Trees" that proposes an ensemble learning approach that involves training multiple Extra Trees classifiers with different subsets of features and combining their outputs using majority voting. An accessible music dataset was used to evaluate the method, and the findings revealed that the ensemble of Extra Trees classifiers outperformed a single Extra Trees classifier and other cutting-edge techniques in terms of accuracy, precision, recall, and F1-score. The work emphasizes the potency of ensemble learning with Extra Trees classifiers for categorization of musical genres.[4]

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.6654
***Article Received**: 28 February 2023 **Revised**: 18 April 2023 **Accepted**: 28 April 2023*

C. LightGBM Classifier

LightGBM is a gradient boosting framework that uses tree-based learning algorithms for classification and regression tasks. It can handle sparse, dense data and is made to be effective and scalable for huge datasets. To increase accuracy and cut down on computing costs, the technique makes use of gradient-based one-side sampling, leaf-wise tree development, and categorical feature management. In order to avoid overfitting and enhance generalization performance, LightGBM additionally offers early halting and regularization. Overall, LightGBM is a strong and effective algorithm that is capable of handling a range of classification and regression tasks on sizable datasets.

$$\widetilde{V}_j(d) = \frac{1}{n}\left(\frac{\left(\sum_{x_i \epsilon A_l} g_i + \frac{1-a}{b}\sum_{x_i \epsilon B_l} g_i\right)^2}{n_l^j(d)}\right) + \left(\frac{\left(\sum_{x_i \epsilon A_r} g_i + \frac{1-a}{b}\sum_{x_i \epsilon B_r} g_i\right)^2}{n_r^j(d)}\right) \qquad (3)$$

L. Chen et. al. published a paper"An Ensemble Learning Method for Music Genre Classification Based on LightGBM and Deep Neural Network" that proposes a method which uses a LightGBM classifier to select the most informative features from the audio signal and then feeds them into a DNN classifier for final prediction. Using many music genre classification datasets, the experimental results demonstrate that the proposed method performs better than the baseline methods and achieves state-of-the-art performance.[5]

D. XGBoost Classifier

XGBoost is an optimized gradient boosting algorithm that is widely used for classification and regression tasks. The algorithm can handle both sparse and dense data and is created to be effective and scalable for large-scale datasets. To avoid over fitting, XGBoost supports L1 and L2 regularization and builds the model using a combination of gradient descent and decision trees. The algorithm also includes a number of sophisticated features that make it a popular option for many machine learning applications including early halting, handling missing inputs, and parallel processing. In general, XGBoost is a strong and adaptable algorithm that can perform well in terms of accuracy and generalization on a variety of datasets.

$$\sum_{i=1}^{n} L(y_i, p_i) = \frac{1}{2}(y_i - p_i)^2 \qquad (4)$$

Y. Liu et. al. published a paper"Music Genre Classification using XGBoost Algorithm" that proposes a method which uses the GTZAN dataset for experimentation and compare the performance of XGBoost with other classifiers such as SVM and Random Forest.The results show that XGBoost outperforms the other classifiers with an accuracy of 96.6%.A feature selection step is also included in the study to enhance the XGBoost classifier's performance. The article shows that XGBoost is good in classifying music genres overall.[6]

E. CatBoost Classifier

CatBoost is a gradient boosting algorithm that is designed to handle categorical features and can be used for classification and regression tasks. Each decision tree in the ensemble is trained to minimize the loss function as part of the algorithm's gradient-boosting approach to decision tree construction. The algorithm has a number of sophisticated features, including strong outlier handling, automatic handling of categorical features, and handling missing values. CatBoost is a strong and effective method for managing huge datasets since it includes a built-in feature importance evaluation mechanism and allows parallel processing. Overall, CatBoost is a flexible approach that performs well on a variety of classification and regression tasks, particularly when working with categorical data, and it is very accurate and generalizable.

$$\hat{x}_k^i = \frac{\sum_{x_j \epsilon D_k} 1_{x_k^i = x_k^j} \cdot y_j + ap}{\sum_{x_j \epsilon D_k} 1_{x_k^i = x_k^j} + a};$$

$$if\ D_k = \{x_j : \sigma(j) < \sigma(i)\} \qquad (5)$$

A. Aziz et. al. published a paper "Music Genre Classification using CatBoost Algorithm" that presents a study on music genre classification using the CatBoost algorithm, which is a gradient boosting method designed for handling categorical features. The authors suggest a feature extraction method based on Statistical Spectrum Descriptors (SSD) and Mel Frequency Cepstral Coefficients (MFCC) to characterize audio signals, then use the CatBoost algorithm to identify the musical genres. The results demonstrate the effectiveness of the CatBoost algorithm for music genre categorization tasks, with the suggested method outperforming a number of other machine learning methods in terms of accuracy and F1-score. [7]

S. Kim et. al. published a paper "Music Genre Classification Using Convolutional Neural Network and CatBoost" that proposes a music genre classification method using a combination of convolutional neural network (CNN) and CatBoost algorithm.CatBoost is used as a classifier to categorise the music into distinct genres, while CNN is utilized to extract high-level features from the spectrograms of musical data. The experimental results show that the suggested

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.6654
***Article Received**: 28 February 2023 **Revised**: 18 April 2023 **Accepted**: 28 April 2023*

method outperforms state-of-the-art methods in terms of performance.[8]
Subheadings should be as the above heading "2.1 Subheadings". They should start at the left-hand margin on a separate line.

III. PROPOSED WORK

The working principle of the proposed work is determined with a flowchart as shown by Fig.1.

The proposed work contains the musical dataset with 19 features named as id, artists, year, valence, acounsticness, danceability, tempo, duration, energy, explicit, key, liveness, loudness, mode, name, popularity, release_date, speechiness, instrumentalness.

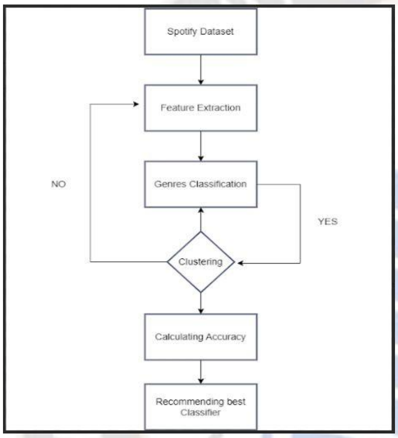
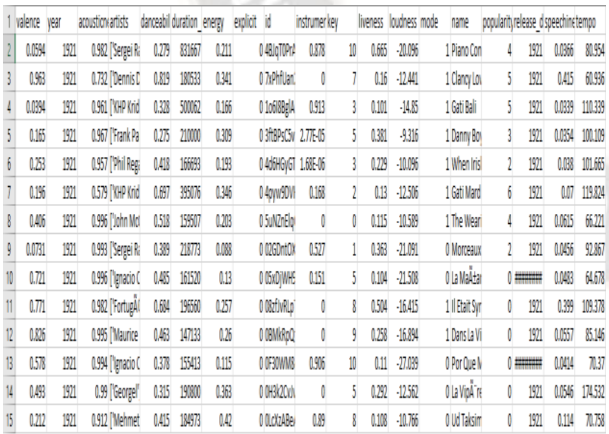


Fig 1. Flow Chart for proposed work

The sample dataset is shown below in Fig.2. that is taken from https://www.kaggle.com/datasets/vatsalmavani/spotify-dataset.



1	valence	year	acousticn	artists	danceabil	duration_	energy	explicit	id	instrumer	key	liveness	loudness	mode	name	popularity	release_d	speechine	tempo
2	0.0594	1921	0.982	['Sergei R	0.279	831667	0.211	0	4BJqT0Pr	0.878	10	0.665	-20.096	1	Piano Con	4	1921	0.0366	80.954
3	0.963	1921	0.732	['Dennis D	0.819	180533	0.341	0	7xPhfUan	0	7	0.16	-12.441	1	Clancy Lo	5	1921	0.415	60.936
4	0.0394	1921	0.961	['KHP Krid	0.328	500062	0.166	0	1o6I8BglA	0.913	3	0.101	-14.85	1	Gati Bali	5	1921	0.0339	110.339
5	0.165	1921	0.967	['Frank Pa	0.275	210000	0.309	0	3ftBPsC5v	2.77E-05	5	0.381	-9.316	1	Danny Bo	3	1921	0.0354	100.109
6	0.253	1921	0.957	['Phil Reg	0.418	166693	0.193	0	4d6HGyGT	1.68E-06	3	0.229	-10.096	1	When Iris	2	1921	0.038	101.665
7	0.196	1921	0.579	['KHP Krid	0.697	395076	0.346	0	4pyw9DV	0.168	2	0.13	-12.506	1	Gati Mard	6	1921	0.07	119.824
8	0.406	1921	0.996	['John Mc	0.518	159507	0.203	0	5uNZnElq	0	0	0.115	-10.589	1	The Wear	4	1921	0.0615	66.221
9	0.0731	1921	0.993	['Sergei R	0.389	218773	0.088	0	02GDntOX	0.527	1	0.363	-21.091	0	Morceaux	2	1921	0.0456	92.867
10	0.721	1921	0.996	['Ignacio C	0.485	161520	0.13	0	05xDjWHS	0.151	5	0.104	-21.508	0	La MaÃ±a	0	#######	0.0483	64.678
11	0.771	1921	0.982	['FortugÃ	0.684	196560	0.257	0	08zfJvRLp	0	8	0.504	-16.415	1	Il Etait Syr	0	1921	0.399	109.378
12	0.826	1921	0.995	['Maurice	0.463	147133	0.26	0	0BMkRpQ	0	9	0.258	-16.894	1	Dans La Vi	0	1921	0.0557	85.146
13	0.578	1921	0.994	['Ignacio C	0.378	155413	0.115	0	0F30WM8	0.906	10	0.11	-27.039	0	Por Que M	0	#######	0.0414	70.37
14	0.493	1921	0.99	['Georgel'	0.315	190800	0.363	0	0H3k2CvJ	0	5	0.292	-12.562	0	La VipÃ¨r	0	1921	0.0546	174.532
15	0.212	1921	0.912	['Mehmet	0.415	184973	0.42	0	0LcXzABe	0.89	8	0.108	-10.766	0	Ud Taksim	0	1921	0.114	70.758

Fig 2. Sample dataset

The proposed architecture shown in Fig.3., selects the important attribute that is to be used for clustering based on number of music listeners, type of music and songs. Later the clustering on the dataset is done by k-means algorithm. Later 5 algorithms are applied on clustered dataset to suggest the best algorithm for musical dataset.

The dataset consists of 19 attributes, but not all are taken into consideration. To determine which attributes are important, the following procedure is adapted. The music listeners, type and categories are initially analyzed and later the important attribute for clustering is selected

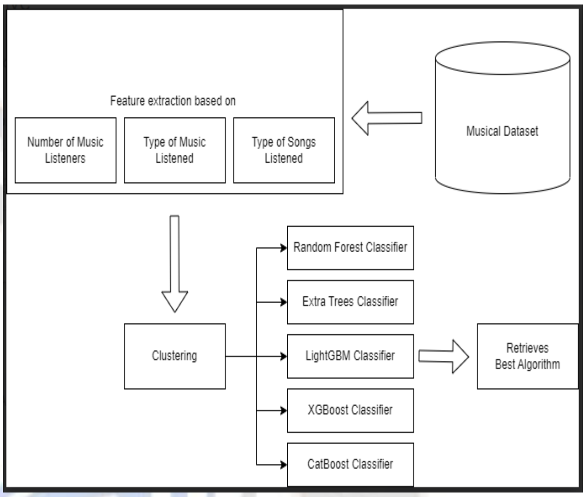


Fig 3. Proposed Architecture

A. Selecting the feature for clustering:

Number of music listeners:

The data from various sources such as music streaming services, radio stations, music industry reports, and surveys is obtained. From the data, computation of music listeners by comparing the number of listeners in different years is done. This confirms that the music listeners are increasing year by year.

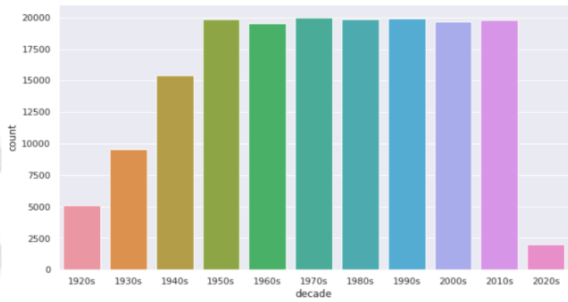


Fig 4. Number of Music Listeners

Type of Songs Listened:

The next task is to understand what type of songs the listeners want to hear.The top six attributes named as acousticness, danceability, energy, instrumentality, liveliness, and valenceare used to compute the type of songs the listeners hear.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.6654
***Article Received**: 28 February 2023 **Revised**: 18 April 2023 **Accepted**: 28 April 2023*

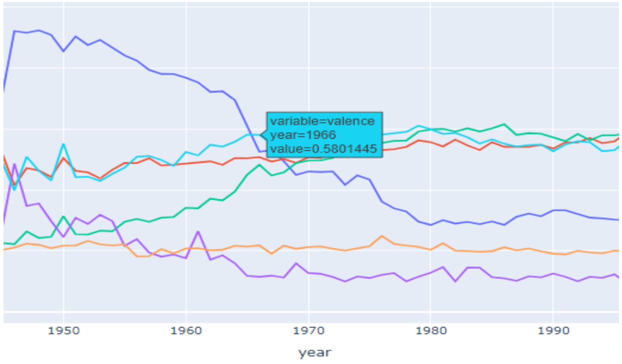


Fig 5. Type of Songs Listened

Type of Genres Listened:

The next level chooses about the genre that how many music listeners are listening to what type and category of music. The top 10 musical genres are chosen based on popularity, and the four most frequently utilized characteristics named as valence, energy, danceability, and acousticness are computed to determine the instruments that listeners prefer to hear in each genre.

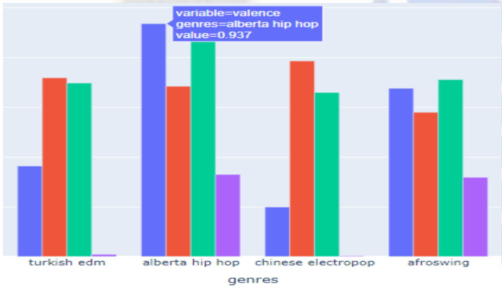


Fig 6. Type of Genres Listened

B. CLUSTERING THE MUSICAL DATASET:

Based on the Fig.4, Fig.5.5 and Fig.6., the genre is considered as an important attribute, and clustering of musical dataset is done based on genre. The attributes used for clustering are acousticness, danceability, energy, instrumentality, liveliness, and valence, which are all known to significantly impact a listener's perception and enjoyment of music.

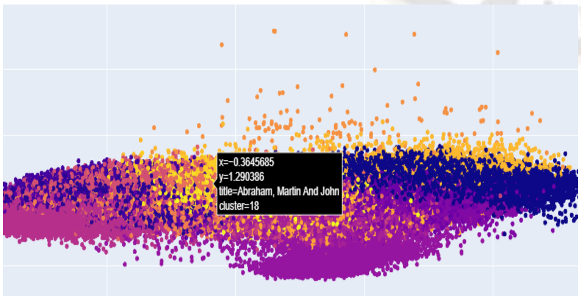


Fig 7. Clustering of dataset

The 5 algorithms choosen for proposed work are Random Forest Classifier, Extra Tree Classifier, LightGBM Classifier, XGBoost Classifier and CatBoost Classifier. In this paper, the proposed work recommends the best approach for clustering of any musical dataset. The clustered dataset is used and applied to find the results of individual algorithm. Out of the five algorithms, the optimal algorithm is recommended based on accuracy.

IV. EXPERIMENTATION RESULTS

The dataset used for the implementation has 1000 records and the class weights are 0 and 1.Each individual algorithm is reu through the dataset and the results have been recorded. The best algorithm recommended is the Random Forest algorithm for those who work in the above-mentioned genres and attributes. The Random Forest algorithm is often considered to be one of the best algorithms for analysing the dataset with a large number of variables and attributes, which is often the case with musical datasets. The algorithm uses an ensemble learning technique that involves the combination of multiple decision trees to make predictions about the type of music that listeners prefer.

Fig.8.represents the values of precision, recall, F1-score and support after applying the random forest classifier on the clustered dataset. The resultant graph for applying random forest classifier dataset is also shown in Fig.9.

```
==================================================
Modeling with : Random Forest Classifier

              precision    recall  f1-score   support

           0       0.75      0.49      0.60      6176
           1       0.85      0.95      0.90     18824

    accuracy                           0.83     25000
   macro avg       0.80      0.72      0.75     25000
weighted avg       0.83      0.83      0.82     25000


Accuracy : 0.83484
==================================================
```

Fig 8. Random Forest Classifier-results

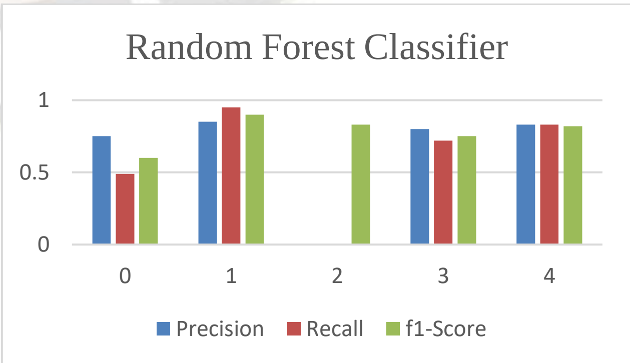


Fig 9. Random Forest Classifier

Fig.10.represents the values of precision, recall, F1-score and support after applying the Extra trees classifier on

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.6654
***Article Received**: 28 February 2023 **Revised**: 18 April 2023 **Accepted**: 28 April 2023*

the clustered dataset. The resultant graph for applying Extra trees classifier dataset is also shown in Fig.11.

```
==================================================
Modeling with : Extra Trees Classifier

              precision    recall  f1-score   support

           0       0.73      0.49      0.59      6176
           1       0.85      0.94      0.89     18824

    accuracy                           0.83     25000
   macro avg       0.79      0.72      0.74     25000
weighted avg       0.82      0.83      0.82     25000


Accuracy : 0.82944
==================================================
```

Fig 10. Extra Trees Classifier-Results

Fig 11. Extra Trees Classifier

```
==================================================
Modeling with : LightGBM Classifier

              precision    recall  f1-score   support

           0       0.75      0.52      0.61      6176
           1       0.86      0.94      0.90     18824

    accuracy                           0.84     25000
   macro avg       0.80      0.73      0.76     25000
weighted avg       0.83      0.84      0.83     25000


Accuracy : 0.83852
==================================================
```

Fig 12. LightGBM Classifier-results

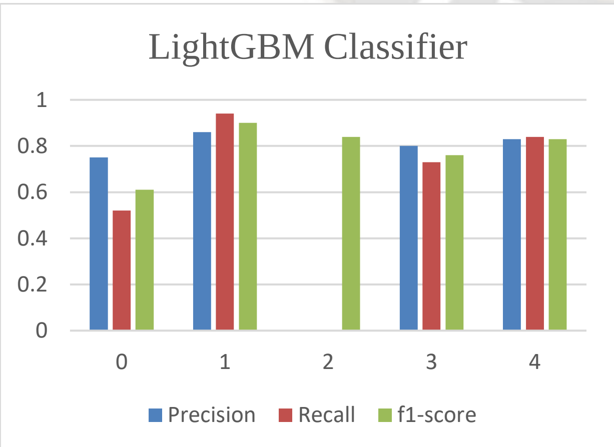


Fig 13. LightGBM Classifier

Fig.12.represents the values of precision, recall, F1-score and support after applying the LightGBM classifier on the clustered dataset. The resultant graph for applying LightGBM classifier dataset is also shown in Fig.13. Fig.14.represents the values of precision, recall, F1-score and support after applying the XGBoost classifier on the clustered dataset. The resultant graph for applying XGBoost classifier dataset is also shown in Fig.15.

```
==================================================
Modeling with : XGBoost Classifier

              precision    recall  f1-score   support

           0       0.72      0.57      0.63      6176
           1       0.87      0.93      0.90     18824

    accuracy                           0.84     25000
   macro avg       0.79      0.75      0.76     25000
weighted avg       0.83      0.84      0.83     25000


Accuracy : 0.8374
==================================================
```

Fig 14. XGBoost Classifier-Results

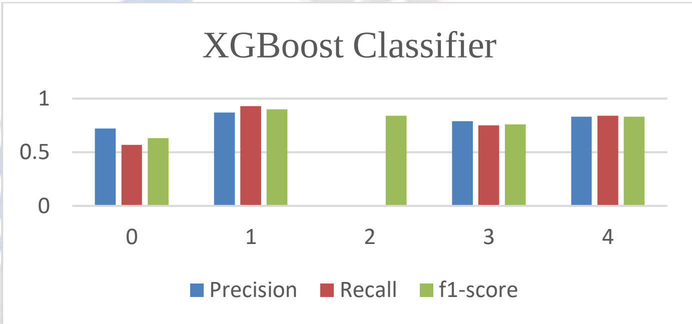


Fig 15. XGBoost Classifier

Fig.16.represents the values of precision, recall, F1-score and support after applying the CatBoost classifier on the clustered dataset. The resultant graph for applying CatBoost classifier dataset is also shown in Fig.17.

```
==================================================
Modeling with : CatBoost Classifier

              precision    recall  f1-score   support

           0       0.74      0.45      0.56      6176
           1       0.84      0.95      0.89     18824

    accuracy                           0.82     25000
   macro avg       0.79      0.70      0.72     25000
weighted avg       0.81      0.82      0.81     25000


Accuracy : 0.82416
==================================================
```

Fig 15. CatBoost Classifier-Results

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 5s
DOI: https://doi.org/10.17762/ijritcc.v11i5s.6654
***Article Received**: 28 February 2023 **Revised**: 18 April 2023 **Accepted**: 28 April 2023*

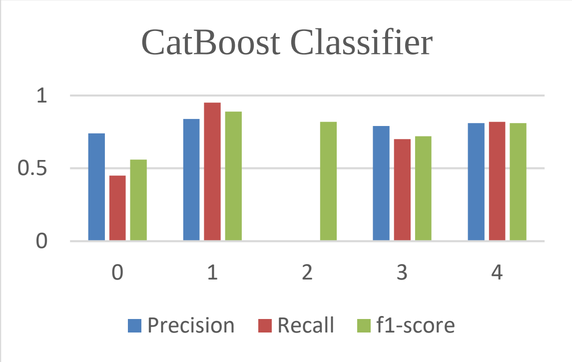


Fig 16. CatBoost Classifier

On applying all the five algorithms, the values of support are same whereas its differing in precision, recall and f1-score.based on the above analysis it is evident that the accuracy of random forest classifier is high and it best suits for the musical dataset. The pseudo code in Fig.17. demonstrates that the random forest classifier is having the high accuracy.

V. CONCLUSION AND FUTURE SCOPE

The musical dataset contains 19 features named as id, artists, year, valence, acounsticness, danceability, tempo, duration, energy, explicit, key, liveness, loudness, mode, name, popularity, release date, speechness, instrumentalness. The important feature used to cluster the datasets is extracted with the support of literature like number of music listeners, type of music and songs. The dataset is divided into clusters based on genre using k-means algorithm. Later, the 5 algorithms are applied on clustered dataset i.e., Random Forest Classifier, Extra Tree Classifier, LightGBM Classifier, XGBoost Classifier and CatBoost Classifier . The experimentation results clearly recommend the best approach of any musical dataset is random forest classifier. The clustered dataset is used and applied to find the results of individual algorithm. The future scope is to extend the work where it suggests the top rated relevant songs from the clusters.

```
rf = RandomForestClassifier(
        n_estimators = 1000,
        class_weight={0: 1.0050505050505052, 1: 0.9949494949494948}
)

rf.fit(X_train, y_train)
```

```
RandomForestClassifier
RandomForestClassifier(class_weight={0: 1.0050505050505052,
                                     1: 0.9949494949494948},
                       n_estimators=1000)
```

```
val_pred = rf.predict(X_val)
print("Accuracy :", accuracy_score(y_val, val_pred))

Accuracy : 0.8352
```

Fig 17. Results showing Accuracy for Random forest

REFERENCES

[1] Lin. L, “Music genre classification using improved random forest algorithm”, Journal of Ambient Intelligence and Humanized Computing, 12(10), 10147-10157, 2021.

[2] Luo. S, “Random Forest-Based Musical Emotion Classification. IEEE Access”, 10, 10759-10769, 2022.

[3] M.S. Rahman, M.S. Islam, M.K. Hasan, and T. Jahan, “Music Genre Classification using Extra Trees Classifier”, In Proceedings of the International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2-2021), pp. 57-62, 2021.

[4] S.A. Shinde, P.P. Shendge, P.M. Patil, “Music Genre Classification using Ensemble of Extra Trees”, In Proceedings of the 2021 International Conference on Electronics, Information, and Communication (ICEIC), pages 1-5, 2021.

[5] Chen. L, Xie. H, Zhang. Y, & Wang. Z, “ An Ensemble Learning Method for Music Genre Classification Based on LightGBM and Deep Neural Network”, IEEE Access, 9, 18411-18423, 2021.

[6] Liu. Y, Zhao. Y, & Liu. S, “Music genre classification using XGBoost algorithm”, Journal of Physics: Conference Series, 1862(1), 012026, 2021.

[7] Aziz. A, Shahzad. F, Abbas. G, Raza. A, & Alshomrani. S. M, “Music Genre Classification using CatBoost Algorithm”, Journal of Ambient Intelligence and Humanized Computing, 12(8), 8587-8602, 2021.

[8] Kim. S, Kim. S, & Ko. H, “Music Genre Classification Using Convolutional Neural Network and CatBoost”, IEEE Access, 8, 218413-218420, 2020.

[9] Somayeh arefi, mohammed ehsan basari, omid roozmand, “Feature selection for author identification of Persian online short texts”, Journal of information and communication technology”, Issue 47,vol 13,spring-summer 1400 PP:35-57

[10] Mahdiye baharloo, AlirezaYassi, “An improved method for detecting phishing websites using data mining on web pages”, Journal of information and communication technology”, Issue 43 ,vol 12,spring-summer 1399 PP:27-38