Journal of Artificial Intelligence and Capsule Networks (ISSN: 2582-2012)
www.irojournals.com/aicn/

Classification of Music Genres using Feature Selection and Hyperparameter Tuning

Rahul Singhal[1], Shruti Srivatsan[2], Priyabrata Panda[3]

[1]Department of Computer Science, Jaypee Institute of Information Technology, Noida, India
[2]Department of Computer Science, Sri Venkateswara College of Engineering, Sriperumbudur, India
[3]Department of Electrical Engineering, St. Thomas College of Engineering and Technology, Makaut, Kolkata, India

E-mail: [1]rahulsinghal1904@gmail.com, [2]shrutisrisvce@gmail.com, [3]priyabratapanda20000@gmail.com

Abstract

The ability of music to spread joy and excitement across lives, makes it widely acknowledged as the human race's universal language. The phrase "music genre" is frequently used to group several musical styles together as following a shared custom or set of guidelines. According to their unique preferences, people now make playlists based on particular musical genres. Due to the determination and extraction of appropriate audio elements, music genre identification is regarded as a challenging task. Music information retrieval, which extracts meaningful information from music, is one of several real - world applications of machine learning. The objective of this paper is to efficiently categorise songs into various genres based on their attributes using various machine learning approaches. To enhance the outcomes, appropriate feature engineering and data pre-processing techniques have been performed. Finally, using suitable performance assessment measures, the output from each model has been compared. Compared to other machine learning algorithms, Random Forest along with efficient feature selection and hyperparameter tuning has produced better results in classifying music genres.

Keywords: Feature selection, hyperparameter tuning, music genre classification, Music Information Retrieval (MIR)

1. Introduction

Music is an artform that is well appreciated worldwide. A genre is a prevalent method to form separate collections of music. The classification of varied instrumental material into

DOI: https://doi.org/10.36548/jaicn.2022.3.003
Received: 03.07.2022, received in revised form: 01.08.2022, accepted: 15.08.2022, published: 25.08.2022

distinct categories which aids in segregation of a particular piece of music is referred to as a music genre. Being artistic in nature, classification of music into a specific genre is sometimes ambiguous and some might even coincide with one other. Analysis is done on music based on factors like acoustics, danceability, pitch, and tempo [1].

Music genres are created by humans with characteristics related to the structure of their rhythm, instrumentation and the harmonic content of the music. Music Information Retrieval (MIR) refers to a form of interdisciplinary science used for extracting insights from music [2]. It can be addressed as a field concerning browsing, classifying, and searching large music collections.

1.1 Music Genres

Some of the music genres prevalent worldwide are:

1) Electronic: It is a type of music which makes use of electronic musical instruments and circuitry-based music technology. The rhythm of this genre is said to develop anticipation and produce dopamine.

2) Jazz: This music is uniquely identified by its complex harmony and rhythms apart from the heavy emphasis on improvisation. It is well known for reducing negative emotions and giving positive feelings.

3) Country: It is a genre filled with elements of folk and rural dance music, and is quite popular for thought provoking lyrics.

4) Rock: It perceives to denote loud and pessimistic vibrations. Rock music comprises of heavy beats and rhythmic electric guitar chords.

5) Classical: It is known for its Mozart effect. This genre comprises of symphony and opera which eliminate any kind of nervousness and alleviate the mood [3].

1.2 Music Features

Salient features of every genre of music are:

1) Sound or pitch: They are categorised as high and low pitches based on the duration of the piece, the volume at which they are played, and the timbre.

2) Melody: It refers to the central theme of a song which is repeated throughout the song.

3) Harmony: Harmony enhances the beauty of the melody.

4) Rhythm: It is the driving beat or pulse of a song, defining the movement of the piece of music [4].

This work comprises of several sections namely Introduction, Literature review, Dataset Description, Proposed Methodology, Experimental Results and Conclusion.

2. Related Work

The authors of [5] have presented a study on the relaxation effects of music listening using machine learning methods. The participants distributed by age, education level, sex, and presence of musical training were made to listen to music and their relaxation levels were recorded using VAS. A decision tree with an overall accuracy of 79% was produced, which suggests that ML techniques can be important to supporting music therapy practice. Random Forest (RF), Logistic Regression (LR), and Extreme Gradient Boosting (XGB) have been used in [6], RF being a much more effective toxicity classification classifier. An overall accuracy of 93% has been achieved in this study. According to the authors, the listener's mode could be improved if toxicity is filtered out by the recommendation system. In this study, several Machine Learning (ML) algorithms were used to classify the lyrics in terms of toxicity and nontoxicity from different music genres and artists.

In [7], along with briefly explaining the theoretical and scientific basis of the problem, it is shown how the application is used for practice, analysing distinct songs and comparing with online categorizations to discuss performance and challenges of music genre classification. The paper presented a web application whose job is retrieving songs from YouTube and classifying into music genres. Probabilistic Graphical Models and SVMs are some of the classifiers that were trained in a multi-label classification scenario. In [8], a detailed survey of music emotion recognition is given. Music Emotion Recognition (MER) can be used in various fields like automatic music composing, psychotherapy and recommendation systems. Commonly used datasets, emotion models and emotion recognition algorithms were involved in the study. The paper contributed to a detailed analysis of research papers on MER and the challenges faced by MER. The authors compared the performance of two models – a DL approach where a CNN model was trained and another, where four ML models were trained. The task of music genre classifications was studied using the data, basically audio clips from YouTube videos. The CNN and XGBoost models were beneficial in [9].

In [10], a music genre classification system and music recommendation engine, MusicRecNet has been proposed. It focused on the extraction of representative features obtained by a deep learning model. To evaluate this model, the GTZAN dataset containing different genres such as blues, classical, jazz, metal, etc. has been utilized. The proposed model has shown improved performance in music genre classification with a classification accuracy of around 90%.

Authors of [11] proposed an approach where multimodal data for music genre classification was learned and combined. The study showed the improved accuracy of the classifications based on the experiments on single and multi-label genre classifications. Results showed that a combination of learned data representations from different modalities give better results than any modalities in isolation. MSD-I and MuMu datasets were used to carry out the experiments. The paper [12] proposed an algorithm T-RECSYS, that used a mixture of content-based and collaborative filtering to produce an accurate real time recommendation system. This can be applied to various platforms such as Youtube (videos), Netflix (movies), Amazon (shopping), etc. It attained a precision score of around 88% in the Spotify Recsys Challenge. In [13], acoustic and visual features were extracted from audio files and were fused by the automated music genre classification. A final ensemble where all these features are fused is shown to produce better classifications than Latin Music Database and GTZAN genre collection. Bird species recognition and whale detection datasets were also assessed as part of the approach.

The problems encountered in the existing solutions are mostly around the number of features in classifying the genre of music or not taking into account sufficient records for each genre of music. This proposed work resolves these problems by selecting the three most important features for modelling, thereby significantly reducing the time and space complexity. There is sufficient data for each genre segregated based on varied characteristics of music.

3. Dataset Description

This research makes use of a Kaggle dataset that is just released [14]. The dataset contains details on 50005 songs that may be grouped into ten categories. Figure 1 depicts the frequency distribution indicating the number of songs in each genre. The dataset contains basic details for each song, including instance ID, artist name, track name, song key, acquired data, and mode (major or minor). Additionally, there is information on eleven musical traits (included in Table 1) for each song, which aids in identifying the genre of a certain song.

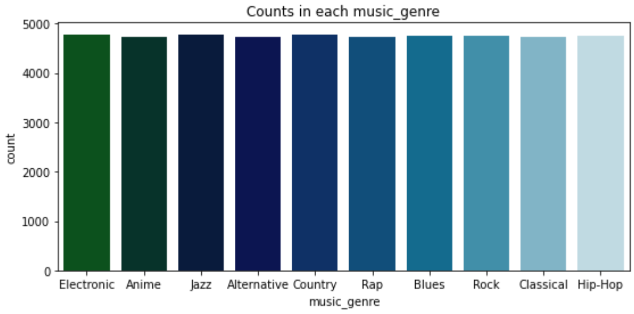

Figure 1. Number of songs present in each music genre

The following table elaborates the different characteristics of music considered for this study.

Table 1. Description of each music characteristic given in the dataset

S. No.	Music Characteristic	Description
1	Popularity	Based on the number of times a track is played and how recent they are, an algorithm calculates the popularity of a song.
2	Acousticness	The acousticness of a track is measured by a scale ranging from 0.0 to 1.0 measure, with the number 1.0 denoting a high level of certainty that the track is acoustic.
3	Danceability	Based on a variety of musical qualities such as pace, rhythm stability, beat strength, and general regularity, a track’s danceability rating indicates how ideal it is for dancing.
4	Music duration	Duration of a track indicates the span of the song which is measured in milliseconds.
5	Energy	Energy is an emotive measure of intensity and activity that ranges from 0.0 to 1.0. Usually energetic tunes have a quick, loud, and boisterous vibe.
6	Instrumentalness	Determines whether or not a music has vocals. In this context, noises like "ooh" and "aah" are viewed as instrumental. "Vocal" recordings, like raps or spoken words, are surely "vocal." The closer the instrumentalness score gets to 1.0, the more likely the track is without vocals. Instrumental recordings are represented by values over 0.5, but as the value approaches 1.0, confidence increases.
7	Liveness	The existence of an audience in the tape is detected. More the liveness scores, greater the likelihood of the track being performed live. If the number is more than 0.8, the music is almost certainly live.
8	Loudness	It decides a song's decibel level. Since decibels are measured in relation to a reference value, songs with lesser loudness values are considered to be quieter than those with higher loudness values.

9	Speechiness	The presence of spoken words in a track is detected by speechiness. When the attribute value is closer to 1.0, the more solely speech-like the recording is (e.g. talk show, audio book, poetry). Tracks with a value greater than 0.66 are almost entirely made up of spoken words. Tracks with values between 0.33 and 0.66 include rap music and may feature both music and spoken, either in portions or layered. Music and other non-speech-like recordings are most often represented by values below 0.33.
10	Tempo	Measured in beats per minute, the tempo is a track's total estimated pace. Tempo is the speed or pace of a composition in musical terms, and it is determined by the average beat duration.
11	Valence	A scale ranging from 0.0 to 1.0 describes how positive a track is musical. Tracks with a high valence, sound more positive whereas, tracks with a low valence, sound more negative.

4. Proposed Methodology

This section discusses the methodology proposed in this study. The flowchart given in Figure 2 portrays the steps in the proposed methodology.

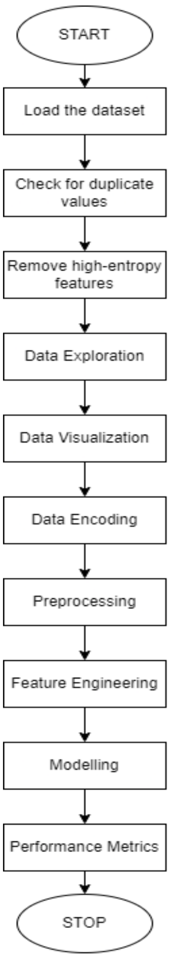

Figure 2. Flow Diagram of the Proposed Methodology

The work has been simulated using Python 3 in Jupyter Notebook. The steps in the flowchart have been explained elaborately below:

4.1 Data Exploration

Data exploration has been carried out to get a better understanding regarding the dataset. This process is begun by finding the number of tracks given by each artist to find the frequency distribution of songs among artists, which is given in Figure 3. Figure 3 shows the top 20 artists having the most number of songs. The number of songs belonging to each key is given in Figure 4 where it is inferred that most songs belong to the G key. The plot showing the number of songs belonging to each music genre is given in Figure 5 where it is seen that the songs per genre are balanced and represented in an equal manner. It is observed that all the features except music genre, key, and mode, have numerical data stored in them. The outliers present in each field belonging to music characteristics are also removed.

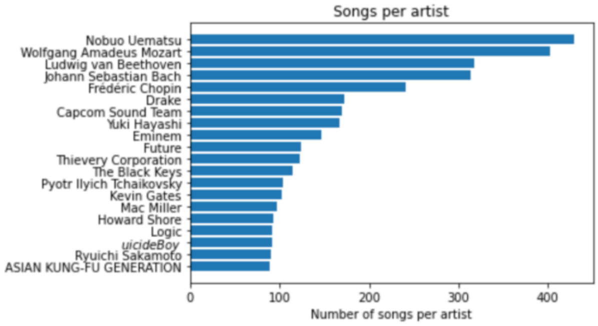

Figure 3. Top 20 artists having the most number of songs

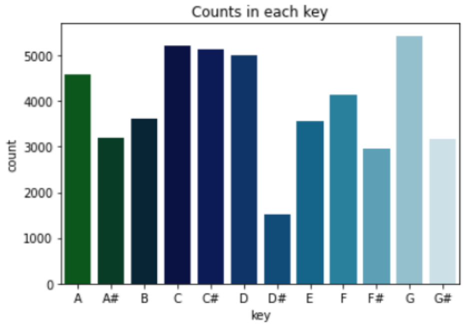

Figure 4. Number of songs present in each key

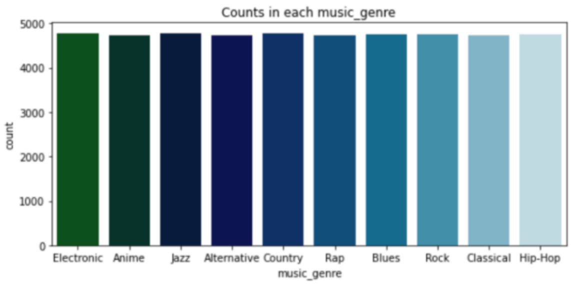

Figure 5. Number of songs belonging to each music genre

4.2 Data Pre-processing

Data pre-processing involves checking and removing the duplicate songs from the dataset. All samples with "empty_field" in artist_name are removed to make this visualisation more informative. It is observed that the tempo column has incorrect values for some songs. The tempo of a piece of music refers to how fast it should be played. Some samples contain a question mark (?) in this feature, and so all the songs with incorrect tempo values have been eliminated. Then, the labels are separated from features. Feature scaling is done next for normalising the data from independent features using StandardScaler() to make all values centred around 0 with standard deviation of 1.

4.3 Feature Engineering

Feature Engineering is carried out to extract features from raw data and keep the necessary features for efficient implementation of machine learning models. The field containing artist names is removed as it is not important for genre prediction. Then, features having low entropy are removed as these have less predicting properties and would not be useful. This includes the instance_id feature which stores ID for a song and is one of a kind for each sample; so, couldn't be utilised for modelling. Also, the obtained_data feature is not important since year has not been provided in this field and only the day is given; therefore, it is removed from the dataset. Similarly, index and track name fields are also removed, as each song has unique entries for them. LabelEncoder() is used to do one-hot encoding on the categorical features, assigning a number (integer) to each class. The labels (i.e., the music genre) are not encoded, allowing easier and more comprehensible interpretation. The values in the key and mode column are encoded. Then, the data are split into training, testing and validation sets. Stratified cross validation is used for ensuring that each fold has the same proportion of observations with a specified categorical value, such as the class result value.

4.4 Model implementation

Several machine models such as KNN, SVM, XGBoost, Random Forest have been implemented in this work for classifying songs into their respective music genres.

5. Results and Discussion

The dataset is split into training, validation and testing sets. Different machine learning models are evaluated and the results are tabulated below. From the table, it can be infered that without hyperparameter tuning, Extreme Gradient Boosting (XGBoost) has performed significantly better in the validation split, whereas Random Forest has done slightly better in the training set in terms of the performance metrics.

Table 2. Results of different Machine learning models on all features

Model	Accuracy (Test data)	F1- Score (Test data)	ROC Auc Score (Test data)	Accuracy (Validation data)	F1 Score (Validation data)	Accuracy (Training data)	F1 Score (Training data)
Logistic Regression	51.19	49.86	88.66	50.51	49.7	50.87	50.04
KNN	56.75	56.54	92.73	55.56	55.46	60.39	60.36
SVM	54.06	53.65	90.81	54.11	53.48	54.41	53.66
XGBoost	**99.60**	**99.60**	**99.99**	**99.74**	**99.74**	**84.66**	**84.68**
Random Forest	**99.60**	**99.60**	**99.71**	**99.72**	**99.72**	**97.2**	**97.2**

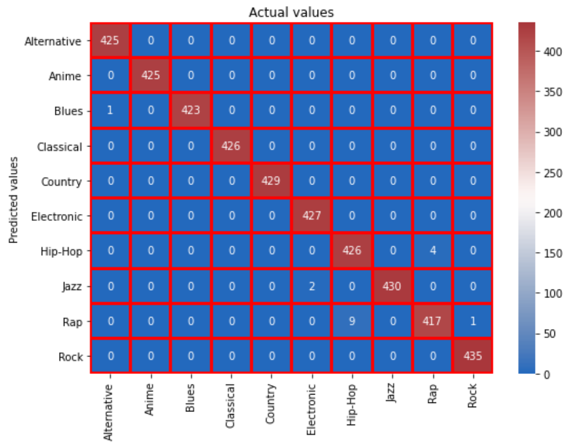

Figure 6. Confusion Matrix generated by Random Forest

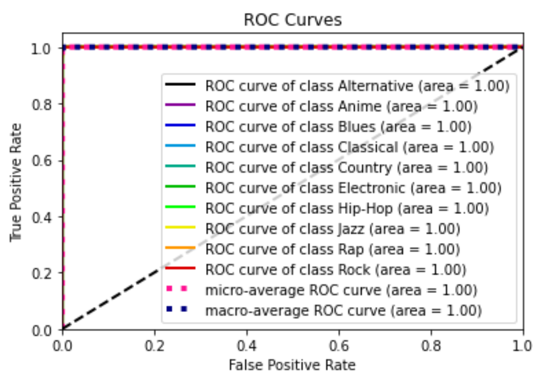

Figure 7. ROC Curves of each feature by Random Forest

From the above figure, it can be concluded that the model gives wrong predictions for genres of music which exhibit similar acoustic features.

5.1 Enhancing the performance of the best model

Hyperparameter tuning is performed on Random Forest classifier and the following results are obtained.

Table 3. Performance of Random Forest on improving hyper parameters

Model	Accuracy (Test data)	F1- Score (Test data)	ROC Auc Score (Test data)	Accuracy (Validation data)	F1 Score (Validation data)	Accuracy (Training data)	F1 Score (Training data)
Random Forest (Using Grid Search)	85.63	85.58	98.98	86.02	86.01	78.12	78.16
Random Forest (Using Randomized Search)	**98.8**	**98.8**	**99.98**	**99.74**	**99.74**	**87.4**	**87.49**

Tuning the parameters improves the performance in the validation set whereas, the performance achieved is similar in other splits.

5.2 Selection of important features

3 important features are selected based on feature importance score computed using Random forest, namely Popularity, Acousticness and Danceability. The performance is as follows:

Table 4. Performance of Random Forest on 3 important features

Model	Accuracy (Test data)	F1- Score (Test data)	ROC Auc Score (Test data)	Accuracy (Validation data)	F1 Score (Validation data)	Accuracy (Training data)	F1 Score (Training data)
Random Forest	99.48	99.48	99.99	99.71	99.71	96.77	96.77

In Table 4, there is a slight improvement in performance metrics when compared to the results with all features. By fitting only 3 features, the time and space complexity can be significantly optimized.

5.3 Memory and Time consumption

The time taken to execute the split has been calculated for the models and tabulated below. The table clearly shows that by reducing the number of features, the time taken for modelling is optimised to a great extent.

Table 5. Comparison of memory and time taken for different splits

Model	Memory (mB)	Training Time (ms)	Validation Time (ms)	Testing Time (ms)
Logistic Regression	4.9	3.68	0.16	0.14
KNN	4.9	1.83	0.21	0.19
SVM	4.9	64.64	0.99	0.95
XGBoost	4.9	35.12	4.53	4.39
Random Forest with all features	4.9	14.95	1.62	1.59
Random Forest with 3 features	**1.3**	**0.16**	**0.014**	**0.013**

6. Conclusion

Music genre classification has always been a challenging task and a lot of work has been published related to the same in the past few years. This research work has focussed on the comparison of results derived on implementing different machine learning models on a recently published dataset from Kaggle, classifying music genres. The study proposes the use of Random Forest Classifier with the most important three features for modelling along with appropriate hyperparameter tuning. It has produced comparable results using XGBoost on the testing set and validation set, and gave the best results on the training set. The results also shows that the overall time taken as well as memory consumed have been reduced considerably in comparison to other implemented machine learning models. Future directions in this area include trying out deep learning models on larger datasets. This work can also be extended to implement a music recommendation system which would depend on the mood of the user.

References

[1] Susino, Marco, and Emery Schubert. "Cultural stereotyping of emotional responses to music genre." Psychology of Music 47, no. 3 (2019): 342-357.

[2] Kasák, P., R. Jarina, and M. Chmulík. "Music information retrieval for educational purposes-an overview." In 2020 18th International Conference on Emerging eLearning Technologies and Applications (ICETA), pp. 296-304. IEEE, 2020.

[3] McCraty, Rollin, Bob Barrios-Choplin, Michael Atkinson, and Dana Tomasino. "The effects of different types of music on mood, tension, and mental clarity." Alternative therapies in health and medicine 4, no. 1 (1998): 75-84.

[4] Grzywczak, Daniel, and Grzegorz Gwardys. "Audio features in music information retrieval." In International conference on active media technology, pp. 187-199. Springer, Cham, 2014.

[5] Raglio, Alfredo, Marcello Imbriani, Chiara Imbriani, Paola Baiardi, Sara Manzoni, Marta Gianotti, Mauro Castelli, Leonardo Vanneschi, Francisco Vico, and Luca Manzoni. "Machine learning techniques to predict the effectiveness of music therapy: A randomized controlled trial." Computer methods and programs in biomedicine 185 (2020): 105160.

[6] Siddique, Md Abdus Salam, Md Imran Sarker, Robin Ghosh, and Kamal Gosh. "Toxicity Classification on Music Lyrics Using Machine Learning Algorithms." In 2021 24th International Conference on Computer and Information Technology (ICCIT), pp. 1-5. IEEE, 2021.

[7] Castillo, Jaime Ramirez, and M. Julia Flores. "Web-based music genre classification for timeline song visualization and analysis." IEEE Access 9 (2021): 18801-18816.

[8] Han, Donghong, Yanru Kong, Jiayi Han, and Guoren Wang. "A survey of music emotion recognition." Frontiers of Computer Science 16, no. 6 (2022): 1-11.

[9] Bahuleyan, Hareesh. "Music genre classification using machine learning techniques." arXiv preprint arXiv:1804.01149 (2018).

[10] Elbir, Ahmet, and Nizamettin Aydin. "Music genre classification and music recommendation by using deep learning." Electronics Letters 56, no. 12 (2020): 627-629.

[11] Oramas, Sergio, Francesco Barbieri, Oriol Nieto Caballero, and Xavier Serra. "Multimodal deep learning for music genre classification." Transactions of the International Society for Music Information Retrieval. 2018; 1 (1): 4-21. (2018).

[12] Fessahaye, Ferdos, Luis Perez, Tiffany Zhan, Raymond Zhang, Calais Fossier, Robyn Markarian, Carter Chiu, Justin Zhan, Laxmi Gewali, and Paul Oh. "T-recsys: A novel music recommendation system using deep learning." In 2019 IEEE international conference on consumer electronics (ICCE), pp. 1-6. IEEE, 2019.

[13] Nanni, Loris, Yandre MG Costa, Rafael L. Aguiar, Carlos N. Silla Jr, and Sheryl Brahnam. "Ensemble of deep learning, visual and acoustic features for music genre classification." Journal of New Music Research 47, no. 4 (2018): 383-397.

[14] Kaggle.com. 2022. Prediction of music genre. [online] Available at: <https://www.kaggle.com/vicsuperman/prediction-of-music-genre>