Music Genre Classification Using Random Forest Model

Ivan Luis Simarmata[a1], I Wayan Supriana[a2]

[a]Informatics Department, Faculty of Mathematics and Natural Sciences, University of Udayana
South Kuta, Badung, Bali, Indonesia
[1]ivan_luis030602@protonmaill.com
[2]wayan.supirana@unud.ac.id (Corresponding author)

Abstract

Music genre is a grouping of music based on their style. To group music into certain genres is a long and boring task to do manually because one must listen to each song individually and determine which genre does this song belong to. This process can be made automatic using classification models like Random Forest. The Random Forest model is a mutated version of the decision tree model, where Random Forest uses multiple decision trees to get a single result. In this paper the model that will be tested is the Random Forest model and XGB Classification model for comparison. The XGB Classification model is used to compare because it is similar to the Random Forest model. XGB Classification is a mutated decision tree model which uses CART as its tree. The results show that with the Random Forest model, an accuracy of 72% is achieved when all audio features are included, and with the XGB Classification, an accuracy of 73% is achieved with some audio features dropped.

Keywords: Classification, Decision Tree, Random Forest, Accuracy, XGB Classification

1. Introduction

Music Genre is a way to categorize or classify music based on the style and features of the music [1]. Classifying music based on genre can be done manually by listening to each song individually. But doing so will consume a lot of time and effort making it an ineffective method. So, an automatic process is required to help classify music [2]. Music Genre Classification have been a problem that has been studied by the Music Information Retrieval community.

There are many classification models that can be used, including Support Vector Machine, K-Nearest Neighbor, Decision Tree, and many more. One of the famous models is the K-Nearest Neighbor [2]. But for this research, the model that will be used is the Random Forest model which is the mutated version of the Decision Tree model.

Previous research conducted by [3] using decision tree to classify Latin music genre. They used 2 types of decision tree model which is the Categoric Attributes and Regression Trees (CART) and the C4.5 algorithm. By using these models, they only achieved an accuracy between 55% - 62%.

Another research conducted by [2] uses the Modified K-Nearest Neighbor model to classify music. The dataset that was used is the GTZAN dataset. From this research it was concluded that the Modified K-Nearest Neighbor model was able to classify music with an accuracy of 55.3%.

Based on both studies, the authors goal is to use the Random Forest model and the XGB Classification model to classify music genre since both models are mutated versions of the Decision Tree model [4]–[6]. The dataset that will be uses is the GTZAN data set based on the research done by [2].

2. Reseach Methods

The process of the system will start by analyzing the feature of the audio from the dataset to see which features are related. After that the features will be preprocessed. For the first scenario all features will be included for training and testing, and for the second scenario some features will be dropped to see which model between Random Forest and XGB Classification is more accurate. The flow process of the system can be seen in figure 1.
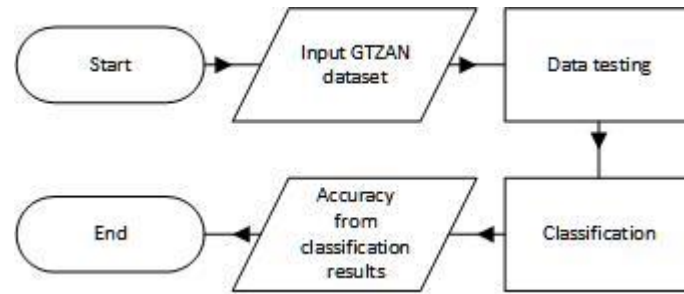
Figure 1. Flowchart of music classification system

2.1   Research Data

The dataset that will be used is the GTZAN dataset that can be accessed from Kaggle. This dataset was made from 2002 and have been used in many researches for music genre recognition (MGR) [7]. The dataset consists of 1000 audio that has been grouped into 10 genres with 100 audios each. The audios are 30 seconds long.

2.2   Audio Features

The audio features that are analyzed will be MFCC, Spectral Centroid, Spectral Bandwidth and Rolloff.

    a. Mel Frequency Cepstrum Coefficient (MFCC)
        Series of short-term power spectrum in an audio file is called MFCC [2]. It is often used for speech recognition and speaker identification [8].
    b. Spectral Centroid
        Spectral centroid is the average of frequencies weighted by amplitude. A higher centroid value means a higher brightness of the high frequnecy [2].
    c. Spectral Bandwidth
        Spectral bandwidth is the width of the frequency band that can be measure from a certain number of decibels below the spectral maximum [9].
    d. Rolloff
        Rolloff is the steepness of a transition from the stop band to the pass band. It is often referred as the measurement of the spectral shape of the audio [2].

2.3   Classification

In this research there are 2 models that will be used to classify music genre, and these 2 models are both mutated version of the Decision Tree model.

    a. Random Forest Model
        Random Forest is a model that uses multiple decision trees to obtain a single result [10]. This model uses bootstrapping, an ensemble learning method, to generate multiple decision trees based of the given dataset, and then averaging the result for a prediction/classification [11].
    b. XGB Classification Model
        The XGB Classification Model is similar to a decision tree model, but the tree that is used are CART trees, and each node contains real value scores of whether an instance belongs to a group [6].

3.   Result and Discussion\

In this research there are 2 scenarios for the experiment. The first scenario includes all features to be tested, while the second scenario, 3 features will be dropped.

3.1.   Results

To display the results from the classification, a confusion matrix is used. This will show the precision, recall and f-1 score of each genre, which shows how accurate the model is.

a. **Precision**
Precision shows the percentage of correct positive predictions relative to total positive predictions. For music classification, precision is an important metric.

$$Precision = \frac{TP}{TP+FP} \tag{1}$$

b. **Recall**
Recall is the percentage of correct positive predictions relative to total actual positives.

$$Recall = \frac{TP}{TP+FN} \tag{2}$$

c. **F-1Score**
The harmonic mean of precision and recall. The closer it is to 1 means the better the model.

$$F-1\,Score = 2 \times \frac{(Recall \times Precision)}{(Recall+Precision)} \tag{3}$$

d. **Accuracy**
Accuracy shows how many positive predictions are actually positive from the total positive predictions.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{4}$$

The results for the first scenario can be seen in table 1 and 2.

Table 1. Random Forest (including all features)

Genre	Precision	Recall	f-1 Score	Support
Blues	0.93	0.61	0.74	23
Classical	0.83	1.00	0.91	43
Country	0.74	0.55	0.63	31
Disco	0.57	0.44	0.50	18
Hiphop	0.62	0.68	0.65	19
Jazz	0.65	0.60	0.63	25
Metal	0.78	0.82	0.80	17
Pop	0.80	1.00	0.89	32
Reggae	0.53	0.56	0.55	16
Rock	0.41	0.44	0.42	16
Accuracy:				0.72

In table 1 it is shown that the Blues genre achieved the highest precision, meaning that out of all the Blues music that the model predicted, only 93% are actually Blues. Out of all the music that are actually

Blues, the model only predicted 61% correctly. Resulting with a 74% on the f-1 score, which means that this model is not that great for classifying Blues music, but great at classifying Classical music.

Table 2. XGB Classification (including all features)

Genre	Precision	Recall	f-1 Score	Support
Blues	0.65	0.57	0.60	23
Classical	0.91	1.00	0.96	43
Country	0.67	0.52	0.58	31
Disco	0.50	0.56	0.53	18
Hiphop	0.63	0.63	0.63	19
Jazz	0.58	0.56	0.57	25
Metal	0.82	0.82	0.82	17
Pop	0.80	1.00	0.89	32
Reggae	0.67	0.62	0.65	16
Rock	0.36	0.31	0.33	16
Accuracy:				0.70

In table 2 it is shown that the Classical genre achieved the highest precision and recall, resulting a high f-1 score. This shows that the XGB Classification model is accurate in classifying Classical music.

The results for the second scenario can be seen in table 3 and 4.

Table 3. Random Forest (3 Features Dropped)

Genre	Precision	Recall	f-1 Score	Support
Blues	0.86	0.52	0.65	23
Classical	0.86	1.00	0.92	43
Country	0.65	0.48	0.56	31
Disco	0.67	0.56	0.61	18
Hiphop	0.50	0.68	0.58	19
Jazz	0.65	0.68	0.67	25
Metal	0.80	0.94	0.86	17
Pop	0.78	1.00	0.88	32
Reggae	0.75	0.56	0.64	16
Rock	0.46	0.38	0.41	16
Accuracy:				0.72

Table 4. XGB Classification (3 Features Dropped)

Genre	Precision	Recall	f-1 Score	Support
Blues	0.76	0.57	0.65	23
Classical	0.88	1.00	0.93	43
Country	0.68	0.55	0.61	31
Disco	0.58	0.61	0.59	18
Hiphop	0.78	0.74	0.76	19
Jazz	0.65	0.60	0.63	25
Metal	0.71	0.88	0.79	17
Pop	0.82	1.00	0.90	32
Reggae	0.56	0.56	0.56	16
Rock	0.38	0.31	0.34	16
Accuracy:				0.73

In table 3 and 4 it shows that there are no significant differences showing that even with some dropped features, both models perform similarly.

3.2. Discussion

Based on the results that can be seen in tabel 1, 2, 3 and 4, both models have achieved a higher accuracy compared to the decision tree model based of the research conducted by [3]. But between the random forest model and the XGB Classification model, the random forest model achieved a higher accuracy of 72% if all audio features were included, as seen in table 1. But the XGB Classification model will achieve a higher accuracy of 73% with some audio features dropped from testing, as seen in table 4. It can also be seen that in both models, the Classical genre achieved the highest f-1 score, and Rock with the lowest. This shows that both models can classify Classical music with a high accuracy, but still struggles in classifying Rock music.

4. Conclusion

Based on the results of the research, it is shown that the random forest model and the XGB Classification model are more accurate compared to the decision tree model. The random forest model was able to achieve 72% of accuracy with all audio features in testing, while the XGB Classification model was able to achieve 73% of accuracy with some audio features being dropped. In the future, this research can be developed by using a newer updated dataset because the GTZAN dataset has been made and used from 2002.

References

[1]   E. Hughes, "What is a Music Genre? (And How Are They Defined?)," Feb. 25, 2022. https://www.musicalmum.com/what-is-a-music-genre/ (accessed Oct. 02, 2022).

[2]   I. N. Giri, L. A. Putri, Gst. A. Giri, I. G. Putra, I. M. Widiartha, and I. W. Supriana, "Music Genre Classification Using Modified K-Nearest Neighbor (MK-NN)," *Jurnal Elektrik Ilmu Komputer Udayana*, vol. 10, no. 3, pp. 261–270, 2022.

[3]   G. M. Bressan, B. C. F. de Azevedo, and E. A. S. Lizzi, "A decision tree approach for the musical genres classification," *Applied Mathematics and Information Sciences*, vol. 11, no. 6, pp. 1703–1713, Nov. 2017, doi: 10.18576/amis/110617.

[4]   S. Thongsuwan, S. Jaiyen, A. Padcharoen, and P. Agarwal, "ConvXGB: A new deep learning model for classification problems based on CNN and XGBoost," *Nuclear Engineering and Technology*, vol. 53, no. 2, pp. 522–531, Feb. 2021, doi: 10.1016/J.NET.2020.04.008.

[5]   T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System", doi: 10.1145/2939672.2939785.

[6]   B. T, "Beginner's Guide to XGBoost for Classification Problems," Apr. 07, 2021. https://towardsdatascience.com/beginners-guide-to-xgboost-for-classification-problems-50f75aac5390 (accessed Oct. 03, 2022).

[7]   B. L. Sturm, "The GTZAN dataset: Its contents, its faults, their effects on evaluation, and its future use," 2013. [Online]. Available: http://imi.aau.dk/

[8]   S. Gupta, J. Jaafar, W. F. wan Ahmad, and A. Bansal, "Feature Extraction Using Mfcc," *Signal Image Process*, vol. 4, no. 4, pp. 101–108, Aug. 2013, doi: 10.5121/sipij.2013.4408.

[9]   T. L. Szabo, "TRANSDUCERS," *Diagnostic Ultrasound Imaging*, pp. 97–135, 2004, doi: 10.1016/B978-012680145-3/50006-2.

[10]   C. Lindner, "Automated Image Interpretation Using Statistical Shape Models," *Statistical Shape and Deformation Analysis: Methods, Implementation and Applications*, pp. 3–32, Mar. 2017, doi: 10.1016/B978-0-12-810493-4.00002-X.

[11]   N. Beheshti, "Random Forest Classification," *Towards Data Science*, Jan. 29, 2022. https://towardsdatascience.com/random-forest-classification-678e551462f5 (accessed Oct. 28, 2022).