Comparison of SVM, KNN, and NB Classifier for Genre Music Classification based on Metadata

DOI : 10.1109/isemantic50169.2020.9234199

De Rosal Ignatius Moses Setiadi
Department of Informatics Engineering
Dian Nuswantoro University
Semarang, Indonesia
moses@dsn.dinus.ac.id

Dewangga Satriya Rahardwika
Department of Informatics Engineering
Dian Nuswantoro University
Semarang, Indonesia
rahardwika@gmail.com

Eko Hari Rachmawanto
Department of Informatics Engineering
Dian Nuswantoro University
Semarang, Indonesia
eko.hari@dsn.dinus.ac.id

Christy Atika Sari
Department of Informatics Engineering
Dian Nuswantoro University
Semarang, Indonesia
atika.sari@dsn.dinus.ac.id

Candra Irawan
Department of Information System
Dian Nuswantoro University
Semarang, Indonesia
candra.irwawan@dsn.dinus.ac.id

Desi Purwanti Kusumaningrum
Department of Informatics Engineering
Dian Nuswantoro University
Semarang, Indonesia
desi.purwanti@dsn.dinus.ac.id

Nuri
Department of Electrical Engineering
Pati Technical College
Pati, Indonesia
nurisaja@sttp.ac.id

Swapaka Listya Trusthi
Department of Informatics Engineering
Pati Technical College
Pati, Indonesia
swapaka@sttp.ac.id

Abstract— **Music recommendations are one of the important things, such as music streaming platforms. Classification of music genres is one of the important initial stages in the process of music recommendation based on genre. Many music classifications are proposed by extracting audio features that require a not light computing process. This research aims to analyze and test the performance of music genre classification based on metadata using three different classifiers, namely Support Vector Machine (SVM) with radial kernel base function (RBF), K Nearest Neighbors (K-NN), and Naïve Bayes (NB). The Spotify music dataset was chosen because it has complete metadata on each of its music. Based on the results of tests conducted by the SVM classifier has the best classification performance with 80% accuracy, then followed by KNN with 77.18% and NB with 76.08%. The accuracy results are relatively the same as music classification based on audio feature extraction, so the classification with the extraction of metadata features can continue to be developed if the metadata in the dataset is well managed.**

Keywords—Music Classification, Music Metadata, Genre Music, Spotify, Feature Extraction

I. INTRODUCTION

Music is one type of entertainment that is widely used by humans to refresh the mind. Nowadays music can be played anywhere and anytime with various media. Nowadays music is also played streaming a lot like Spotify. The advantage of this method is that it can minimize the storage of music, play it on a variety of hardware, explore the collection of friends, besides that users will also find out about new music and music recommendations with similar genres being played by users. Music recommender is very useful for music streaming application users to be able to find songs that are by the user's preferences.

The recommendation process begins with the classification process of the music genre [1]–[3], where to do the classification required a feature extraction process that will be used for the training process. Some feature extractions used are content-based such as Mel-Frequency Cescal Coefficients (MFCC), Gaussian mixture model (GMM), Fast Fourier Transform (FFT), etc.[4]–[7]. The process of feature extraction and training requires computation that is quite heavy if the dataset is used very much, especially if the extracted features come from content or audio signal processing. But audio-based feature extraction processing can produce higher accuracy and effectiveness compared to using metadata. Some research has been done by [5]–[7] using audio-based feature extraction, produces an accuracy of around 71%, 76%, up to 82%. In the research [8] it is said that the classification of music genres uses inaccurate metadata due to the loss of some or all of the music metadata, especially music that is illegally downloaded on the internet.

Music streaming service certainly uses original music with better audio quality. Besides that, original music has complete metadata, so that genre classification based on metadata can be possible. In several previous studies, a classification based on texts such as lyrics was also tested[9]–[12], some research like [13], [14] also uses multimodal features where metadata is one of the features used to improve the accuracy of genre classification results. Logically processing metadata features does not require complex computations such as audio-based feature extraction. In this research, the classification of music genres based on metadata alone will be measured for accuracy in the Spotify music streaming dataset. Some classifiers such as SVM, KNN, and NB are used and tested for their performance to classify music genres based on metadata. All three are popular classifiers and are widely used in various classification studies such as research [6], [15]–[21]. The purpose of this research is to find out how effective the classification of music genres is based on the metadata in the Spotify music streaming dataset.

II. RELATED RESEARCH

In research conducted by Kumar et. al [6], several classifiers such as KNN and SVM have been tested to classify music genres using two kinds of feature extraction, namely FFT and MFCC. Based on the results of the SVM classifier experiment has the best performance when compared to the MFCC feature where the highest accuracy is 82.55%, but with the FFT feature, it produces an accuracy of 41.5%. Whereas, the KNN classifier with a value of n = 5 produces an accuracy that is nearly balanced where 65% is for the FFT feature and 67.5% for the MFCC feature.

In research conducted by Fulzele et. al [18], the SVM classifier is also tested to classify music genres in the GTZAN data set. In this research, seven features are used, namely SSD, MFCC, spectral roll-off, zero-crossing rate, chroma frequency, rhythm histogram, and spectral centroid. The accuracy of music genre classification resulting from the SVM classifier is 84%.

Kobayashi et al [19] proposed a classification of music genres using low-level audio features with SVM as the classifier. The data set used is GTZAN, where the audio signal normalization is pre-processed, then the undecimated transform, local feature extraction and integration feature are performed. The classification accuracy of the trial results reaches 81.5%.

Ali and Siddiqui [20] classify music genres with SVM and KNN classifiers. The dataset used in his research is GTZAN, and the extraction feature used is MFCC. Tests were carried out on Matlab on 10 types of genres in the GTZAN dataset. The resulting accuracy is 77% for the SVM classifier and 64.9 for the KNN classifier.

Ardiyansyah et al [21] propose a classification of music genres using the NB classifier. In the data set used there are four music genres, namely jazz, dangdut, pop, and rock. Extraction features used are short term energy (STE), zero-crossing rate (ZCR), spectral centroids, and spectral fluxes. In the classification process, MIRTOOLBOX is used in the Matlab application. Based on the trial results, the accuracy of each genre is 70.5% for the pop genre, 62.5% for dangdut genre, 30% for rock genre, and 10% for jazz genre.

Some of the literature studies above have given an overview of the effectiveness and accuracy of each proposed method, where all of the above literature uses features generated from audio extraction. This research attempts to examine how effective the use of the metadata feature is in the classification of music genres. This research will be tested using three classifiers namely SMV, KNN, and NB.

III. Method

In this section, the methods used in this research are explained. Fig. 1 shows the stages of the proposed method.

Fig. 1. Step by step of the method used

Based on Fig.1 in more detail the steps of the method used are as follows:

1. Data acquisition and selection: Take the Spotify music dataset using www.crowdai.org website. Spotify music dataset has a total of 228,159 music, consisting of 26 genres and each has 18 features. Because the number of features and the number of genres will greatly affect the complexity and length of time required for the calculation, then selected 5, 6, 7, and 8 genres in which each genre is taken 6000 sample music.
2. Preprocessing: In the first stage, training and testing data were distributed, 80% of training data, and 20% of testing data.
3. Feature extraction and selection: Next 18 metadata features are extracted except genre features, then 17 features are converted to numeric values, and normalized (specifically for training on SVM methods), then feature selection is performed using the chi-square method.
4. Classification: at this stage, the training process is done first based on data and the selection of features that have been determined, then the testing process is done using each method
5. Evaluation and measurement: the results of the testing process for each method are then measured for their accuracy.

IV. Implementation and Testing

This study uses the Spotify music dataset which has 228,159 music with 26 genres and 18 features [22]. The Python programming language is used and the Python Data Analysis Library(PANDAS) [23] is used to process data structures and perform data analysis. Besides that, Scikit Learn which is a package containing important modules of machine learning projects [24] is used in this research.

To reduce the computational time of 26 genres, 5,6,7, an 8 favorite genres were chosen Each genre is chosen 6000 music, then the data is separated into training data and testing data, where training data is 80% music for each genre, and testing data is 20% music for each genre. Table 1 addresses the list of genres used in this research.

TABLE I. Selected Genres and Used in This Research

Number of genres	Genres	Number of music
5	Pop, Electric, Rap, Opera, and Folk	30000
6	Pop, Electric, Rap, Opera, Dance, and Folk	36000
7	Pop, Electric, Rap, Opera, Dance, Ska, and Folk	42000
8	Pop, Electric, Rap, Opera, Dance, Ska, Rock, and Folk	48000

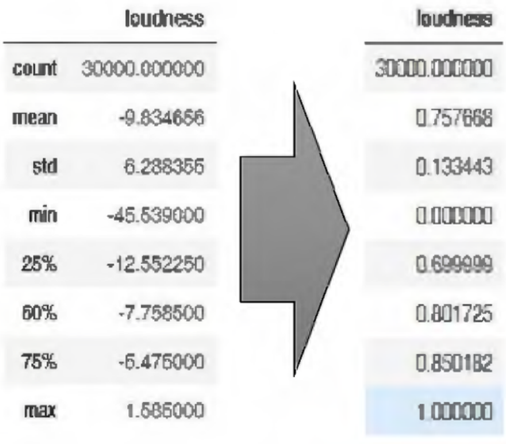
Furthermore, feature extraction is performed on each music except genre features. This genre feature will be used later as the target classification feature. Seventeen other features are artist_name, track_name, track_id, key, mode, time_signature with the object data type, popularity and duration_ms with integer data types, as well as acousticness, danceability, energy, instrumentalness, liveness, loudness, speechiness, tempo, and valence with type data float data. Because the type of data of all features is not the same, all of them are converted to float data types using PANDAS. The float data type was chosen because it is a numeric value that can be calculated and is more precise.

The number of features will make the training process will be slow, so the selection feature is done by the chi-square method. The results of the selection feature scores are presented in Table 2.

TABLE II. FEATURES SCORE

No	Features	Score
1	Acousticness	7333.825205
2	Instrumentalness	3664.597613
3	Popularity	2780.085255
4	Energy	2314.629554
5	Danceability	1032.640592
6	Mode	1048.192344
7	Speechiness	1032.152780
8	Valence	922.493163
9	Loudness	451.282083
10	Duration_ms	137.867351
11	Tempo	117.405865
12	Liveness	88.828267
13	Artist_name	86.447100
14	Time_signature	36.059803
15	Track_name	11.493590
16	Key	10.568912
17	Track_id	5.321589

Based on Table 1, then selected thirteen features with the highest score to be used for the learning process. Reduction of the number of features made to reduce the complexity of calculations. Furthermore, the value of each feature is normalized and converted in the range 0.00-1.00, this is done especially for the learning process that uses the SVM classifier because SVM can work optimally in that range. For example, the results of normalization can see Fig. 2



	loudness
count	30000.000000
mean	-9.834656
std	6.288355
min	-45.539000
25%	-12.552250
50%	-7.758500
75%	-5.475000
max	1.585000

loudness
30000.000000
0.757668
0.133443
0.000000
0.699999
0.801725
0.850182
1.000000

Fig. 2. An example of the normalization results in the loudness feature

In this research, the SVM kernel used is a radial basis function (RBF) wherein this kernel two variables are needed, namely gamma (γ) and C. Gamma value are used to determine the extent of the influence of each data with a borderline, while C is a hyperparameter which can determine the margin on the data used. In this study, the range of gamma variables was set between 0,0001 and 1, while the range of variable C was set between 0.01 and 100. The SVM-RBF classification was carried out by searching the grid to obtain the most optimal results from using hyperparameter. To do a grid search, you must first determine the value of cross-validation because the calculations made by SVM are based on cross-validation. K-fold cross-validation with comparison of 80% training data and 20% test data with free random conditions. Table 3 shows the results of calculating the accuracy of each genre number.

TABLE III. CLASSIFICATION RESULTS OF SVM-RBF CLASSIFIER

Number of genres	C	γ	Accuracy
5	100	0.1	80%
6	10	1.0	71%
7	10	1.0	72%
8	10	1.0	67%

In addition to using SVM-RBF as a classifier, this research also tested two other classifiers as a performance comparison. The two classifiers are KNN and NB. This classifier was chosen because it is considered quite reliable and popular to use but requires a relatively shorter running time than SVM-RBF. To make the comparisons commensurate, all stages of the proposed method are also carried out in the classification process using the KNN and NB classifier, except for the data normalization stage. The data normalization process is omitted because KNN and NB do not assume that the distribution of data is at a value near 0.00 with a variant unit of 1.00. In the KNN classifier, no single k value is used, but k values between 1 and 100 are used, and k values are sought with the most optimal accuracy. Table 4 shows the results of the KNN classifier classification. In the NB classifier, the same treatment is also carried out on the data used. Data separation, genre selection, encoding, and others have been carried out so that data can be executed directly. In contrast to SVM and KNN, NB has no specific variables, so the accuracy results obtained are truly objective of the data used. Table 5 shows the results of the NB classifier classification.

TABLE IV. CLASSIFICATION RESULTS OF KNN CLASSIFIER

Number of genres	k	Accuracy
5	32	75.61%
6	40	66.39%
7	45	68.40%
8	33	61.93%

TABLE V. CLASSIFICATION RESULTS OF NB CLASSIFIER

Number of genres	Accuracy
5	75.05%
6	64.97%
7	66.20%
8	59.32%

After several experiments with three types of classifiers, namely SVM-RBF, KNN, and NB, several things need to be highlighted regarding the comparison of the performance of each classifier used. The graph presented in Fig. 3 is a resume of the comparison of three classifiers in comparing accuracy results. It can be seen in Fig. 2 that the SVM classifier has the highest accuracy compared to another classifier, i.e. KNN and NB. The accuracy value produced by the method in this research is also almost the same as the performance of the method proposed in several previous studies using the audio feature extraction approach. Even so, the difference in the dataset both in terms of the amount of data, the number of features, the type of feature, the number of genres, etc. should be considered because differences in the dataset will affect the results and method performance. Table 6 presents a comparison of the results of the methods in previous studies and the proposed method.

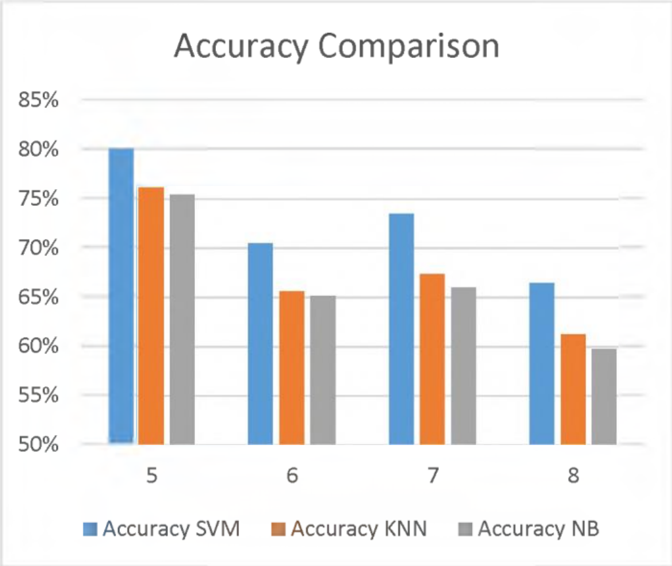


Fig. 3. Accuracy comparison of three classifier

TABLE VI. COMPARISON WITH THE PREVIOUS METHOD

Method	Accuracy (%)		
	SVM	***KNN***	***NB***
Method in [25]	-	-	72.69
Method in [18]	84	-	-
Method in [26]	83	-	-
Method in [27]	84	-	-
Method in [6]	82.55	67.50	-
Method in this research	80	77.18	76.08

Based on the results presented in Table 6 it appears that the classification of music genres produced based on metadata feature extraction can produce relatively the same performance as the Audio feature extraction approach, even the KNN and NB classifier appears to be superior. This shows that the classification based on metadata is considered ineffective as many say in various studies are not entirely correct, with a note if the metadata in the dataset used has been managed properly.

V. CONCLUSION

Music is one of the interactions that are needed by humans. Today's music streaming platform has grown tremendously and is increasingly being used. So that new music can be offered to a suitable listener, of course, a system is needed to provide recommendations where the classification process has an important role in it. This research proposes a classification method based on metadata features that are considered no better than audio feature extraction is applied to music classification. But because this research is applied to the Spotify music dataset, which is one of the world's leading streaming platforms, and it can be confirmed that the metadata in the dataset has been well managed. This research also performed a comparison of the performance of three classifiers to classify music genres based on metadata features, where SVM-RBF proved to be superior in terms of accuracy compared to KNN and NB. Another thing that can be concluded is the opportunity for music genre classification based on metadata turns out to have the accuracy that is not much different from the extraction of audio features, with metadata records must be complete and correct, this is quite possible to do on a professional music streaming platform. Besides the logic of the classification based on the metadata feature is relatively faster than the extraction of audio features, because it does not require a long-winded conversion process. Although this has not been tested in this research, it is therefore suggested that in the next research it is proposed to compare the performance of music genre classification based on metadata features and audio features head to head, both in terms of accuracy and computational time.

REFERENCES

[1] Adiyansjah, A. A. S. Gunawan, and D. Suhartono, “Music recommender system based on genre using convolutional recurrent neural networks,” in *Procedia Computer Science*, 2019, vol. 157, pp. 99–109.

[2] I. Andjelkovic, D. Parra, and J. O'Donovan, “Moodplay: Interactive music recommendation based on Artists' mood similarity,” *Int. J. Hum. Comput. Stud.*, vol. 121, pp. 142–159, Jan. 2019.

[3] A. Elbir, H. Bilal Çam, M. Emre Iyican, B. Öztürk, and N. Aydin, “Music Genre Classification and Recommendation by Using Machine Learning Techniques,” in *Proceedings - 2018 Innovations in Intelligent Systems and Applications Conference, ASYU 2018*, 2018.

[4] C. Kaur and R. Kumar, “Study and analysis of feature based automatic music genre classification using Gaussian mixture model,” in *Proceedings of the International Conference on Inventive Computing and Informatics, ICICI 2017*, 2018, pp. 465–468.

[5] S. Vishnupriya and K. Meenakshi, “Automatic Music Genre Classification using Convolution Neural Network,” in *2018 International Conference on Computer Communication and Informatics, ICCCI 2018*, 2018.

[6] D. P. Kumar, B. J. Sowmya, Chetan, and K. G. Srinivasa, “A comparative study of classifiers for music genre classification based on feature extractors,” in *2016 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics, DISCOVER 2016 - Proceedings*, 2016, pp. 190–194.

[7] S. Sugianto and S. Suyanto, “Voting-Based Music Genre Classification Using Melspectogram and Convolutional Neural Network,” 2020, pp. 330–333.

[8] A. J. H. Goulart, R. C. Guido, and C. D. MacIel, “Exploring different approaches for music genre classification,” *Egypt. Informatics J.*, vol. 13, no. 2, pp. 59–63, Jul. 2012.

[9] R. Mayer, R. Neumayer, and A. Rauber, “Rhyme and style features for musical genre classification by song lyrics,” in *ISMIR Conference*, 2008.

[10] K. Choi, J. H. Lee, and J. S. Downie, “What is this song about anyway?: Automatic classification of subject using user interpretations and lyrics,” in *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*, 2014, pp. 453–454.

[11] O. Coban and I. Karabey, “Music genre classification with word and document vectors,” in *Signal Processing and Communications Applications Conference*, 2017.

[12] H. Oğul and B. Kırmacı, “Lyrics mining for music meta-data estimation,” in *IFIP Advances in Information and Communication Technology*, 2016, vol. 475, pp. 528–539.

[13] S. Oramas, F. Barbieri, O. Nieto, and X. Serra, “Multimodal Deep Learning for Music Genre Classification,” *Trans. Int. Soc. Music Inf. Retr.*, vol. 1, no. 1, pp. 4–21, 2018.

[14] S. Oramas, O. Nieto, F. Barbieri, and X. Serra, “Multi-Label Music Genre Classification from Audio, Text, and Images Using Deep Features,” 2017.

[15] K. Leartpantulak and Y. Kitjaidure, “Music genre classification of audio signals using particle swarm optimization and stacking ensemble,” in *iEECON 2019 - 7th International Electrical Engineering Congress, Proceedings*, 2019.

[16] O. R. Indriani, E. J. Kusuma, C. A. Sari, E. H. Rachmawanto, and D. R. I. M. Setiadi, “Tomatoes classification using K-NN based on GLCM and HSV color space,” in *Proceedings - 2017 International*

Conference on Innovative and Creative Information Technology: Computational Intelligence and IoT, ICITech 2017, 2018, vol. 2018-Janua, pp. 1–6.

[17] N. Rahmayuna, D. S. Rahardwika, C. A. Sari, D. R. I. M. Setiadi, and E. H. Rachmawanto, "Pathogenic Bacteria Genus Classification using Support Vector Machine," in *2018 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, 2018, pp. 23–27.

[18] P. Fulzele, R. Singh, N. Kaushik, and K. Pandey, "A Hybrid Model for Music Genre Classification Using LSTM and SVM," in *2018 11th International Conference on Contemporary Computing, IC3 2018*, 2018.

[19] T. Kobayashi, Y. Suzuki, and A. Kubota, "Audio feature extraction based on sub-band signal correlations for music genre classification," in *Proceedings - 2018 IEEE International Symposium on Multimedia, ISM 2018*, 2019, pp. 180–181.

[20] M. A. Ali and Z. A. Siddiqui, "Automatic Music Genres Classification using Machine Learning," 2017.

[21] Ardiansyah, B. Yuliadi, and R. Sahara, "Music Genre Classification using Naïve Bayes Algorithm," *Int. J. Comput. Trends Technol.*, vol. 62, no. 1, pp. 50–57, Aug. 2018.

[22] B. Brost, R. Mehrotra, and T. Jehan, "The Music Streaming Sessions Dataset," 2018.

[23] W. McKinney, *Python for data analysis : data wrangling with pandas, NumPy, and IPython*, 2nd ed. O'Reilly Media, Inc, 2018.

[24] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[25] M. M. Panchwagh and V. D. Katkar, "Music genre classification using data mining algorithm," in *Conference on Advances in Signal Processing, CASP 2016*, 2016, pp. 49–53.

[26] G. Kour, M. Tech, S. Rbiebt, N. Kharar, and A. Mehan, "Music Genre Classification using MFCC, SVM and BPNN," 2015.

[27] L. Nanni, Y. M. G. Costa, A. Lumini, M. Y. Kim, and S. R. Baek, "Combining visual and acoustic features for music genre classification," *Expert Syst. Appl.*, vol. 45, pp. 108–117, Mar. 2016.