JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 9. NO. 1 AUGUST 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v9i1.4324

MUSIC RECOMMENDATION SYSTEM BASED ON COSINE SIMILARITY AND SUPERVISED GENRE CLASSIFICATION

Jamie Mayliana Alyza[1]; Fandy Setyo Utomo[2*]; Yuli Purwati[3]; Bagus Adhi Kusuma[4]; Mohd Sanusi Azmi[5]

Department of Informatics Engineering[1,2,3,4]
Universitas AMIKOM Purwokerto
https://amikompurwokerto.ac.id
jamiealyza@gmail.com[1], fandy_setyo_utomo@amikompurwokerto.ac.id[2*],
yulipurwati@amikompurwokerto.ac.id[3], bagus@amikompurwokerto.ac.id[4]

Faculty of Information and Communication Technology[5]
Universiti Teknikal Malaysia Melaka
https://www.utem.edu.my/
sanusi@utem.edu.my[5]

(*) Corresponding Author

Abstract— *Categorizing musical styles can be useful in solving various practical problems, such as establishing musical relationships between songs, similar songs, and finding communities that share an interest in a particular genre. Our goal in this research is to determine the most effective machine learning technique to accurately predict song genres using the K-Nearest Neighbors (K-NN) and Support Vector Machine (SVM) algorithms. In addition, this article offers a contrastive examination of the K-Nearest Neighbors (K-NN) and Support Vector Machine (SVM) when dimensioning is considered and without using Principal Component Analysis (PCA) for dimension reduction. MFCC is used to collect data from datasets. In addition, each track uses the MFCC feature. The results reveal that the K-Nearest Neighbors and Support Vector Machine offer more precise results without reducing dimensions than PCA results. The accuracy of using the PCA method is 58% and has the potential to decrease. In this music genre classification, K-Nearest Neighbors (K-NN) and Support Vector Machine (SVM) are proven to be more efficient classifiers. K-Nearest Neighbors accuracy is 64,9%, and Support Vector Machine (SVM) accuracy is 77%. Not only that, but we also created a recommender system using cosine similarity to provide recommendations for songs that have relatively the same genre. From one sample of the songs tested, five songs were obtained that had the same genre with an average accuracy of 80%.*

Keywords: *K-Nearest Neighbors, Support Vector Machine, Music, Genre, Classification.*

Intisari—*Pengategorian gaya musik dapat memecahkan berbagai masalah praktis, seperti membangun hubungan antara genre dan lagu, mengidentifikasi lagu yang mirip, dan menemukan komunitas dengan minat yang sama dalam genre tertentu. Penelitian ini bertujuan untuk menemukan metode pembelajaran mesin yang paling efektif untuk memprediksi genre lagu secara akurat. Kami menggunakan algoritma K-Nearest Neighbour (K-NN) dan Support Vector Machine (SVM). Selain itu, jurnal ini juga membandingkan kedua metode ini dengan menggunakan Analisis Komponen Utama (PCA) dan tanpa menggunakan Analisis Komponen Utama (PCA). Fitur MFCC digunakan untuk mengumpulkan data dari dataset. Selain itu, pada masing-masing trek juga menggunakan fitur MFCC. Hasil penelitian mengungkapkan bahwa K-Nearest Neighbour dan Support Vector Machine memberikan hasil yang lebih tinggi tanpa menggunakan metode PCA. Hasil akurasi menggunakan metode PCA adalah 58% dan berpotensi akan mengalami penurunan. Dalam klasifikasi genre music ini, K-Nearest Neighbour (K-NN) dan Support Vector Machine (SVM) terbukti menjadi classifier yang lebih efisien. Akurasi K-Nearest Neighbour adalah 64,9% dan akurasi Support Vector Machine adalah 77%. Tidak hanya itu, peneliti juga membuat recommender system menggunakan cosine similarity, untuk memberikan rekomendasi lagu yang memiliki genre relative sama. Dari satu contoh lagu yang diujikan, diperoleh lima lagu yang memiliki kesamaan genre dengan rata – rata akurasi 80%.*

Kata Kunci: *K-Nearest Neighbors, Support Vector Machine, Musik, Genre, Klasifikasi.*

INTRODUCTION

The speed of every individual or music lover can have a playlist of up to hundreds of songs, but professional music lovers can have a large collection of music. The majority of music files are organized based on the song's title or the artist's name [1], that can lead to trouble finding songs related to certain genres. The state-of-the-art music database continues to achieve a reputation for dedicated records or personal audio libraries. Because of the increase in web access and net speed, there has been an increase in the number of people involved with sound collections too. However, having an extensive music collection in a warehouse requires a painstakingly long process, especially as manually audio sorting genres. Genres and sub-genres have also been used to categorize music based on the music and the lyrics [2]. This will make the classification process more difficult. For example, when they were first discovered, songs in the jazz genre differed from jazz in recent years, which had experienced improvisation, sometimes combined with pop.

According to the Big Indonesian Dictionary (KBBI), music is the science or art of arranging notes or sounds in sequences to create a unified and continuous sound. The term "music" comes from the Greek word "mousikos," symbolizing the God of beauty and arts. According to Encyclopaedia Britannica, music is an art that combines vocal or instrumental sounds to express beauty and emotion, following cultural standards of rhythm, melody, and harmony.

The music combines tone, rhythm, and harmony to create a unified composition that impacts emotions and cognition. It is both a pleasure to learn and listen to in daily life, adding excitement and enhancing activities. Furthermore, music is classified into genres based on its inherent characteristics, allowing for categorizing and identifying different musical styles [3].

Music genres classify music based on similarities in geography, engineering, context, style, and themes. Clear definitions are important as songs can contain multiple genres but are often labelled under only one. Genre classification is common and effective, but it can be challenging for non-experts due to the presence of multiple genres in a single piece of music [4]. To address this, applications are being developed to aid in accurate genre descriptions. However, accurately defining genres remains a challenge within the realm of music applications. Genre identification in music is done manually by an expert [5] however, this raises several problems, including inefficiency in identifying musical genres manually because it requires time and experts who understand music, so there are no mistakes in making the identification. Automatic identification of musical genres can help reduce or replace the role of humans in presenting genres in music.

In this case, we use an automated technique to improve the accuracy of genre classification by considering audio composition details. We combine rhythm, tempo, energy distribution, tone, timbre, and other features to classify genres. Tone and rhythm are suggested as indicators of the genre, but the lack of audio descriptor poses limitations. This leads to quick classification execution, making it challenging to discern with limited information and resulting in different audio arrangements [4].

At this stage, testing is carried out on several types of music that are often heard by the general public. There are 6 types or genres of music used: blues, Country, hip-hop, pop, reggae and rock. Testing is used using PCA. PCA or Principal Component Analysis is a linear transformation to determine a new coordinate system from a dataset. PCA techniques can reduce the dimensions of the dataset without losing important information from the dataset [6].

According to Angelo, Mauricio and Raul, 2020 [1]the weakness of the PCA method is that it could be more optimal in the separation between classes, and the main component is a linear combination of all variables with a load assigned to each variable. The resulting load value is usually not 0, so the principal component results obtained are difficult to interpret [7]. The experiment showed that the accuracy of PCA is 58%, if PCA is still used for this classification, the accuracy will be lower [1].

That's why this paper uses k-nearest neighbors (k-NN) and Support Vector Machine (SVM) if the PCA method only uses 6 genres, in this method 10 genres are used, that is blues, classical, rock, jazz, reggae, metal, country, pop, disco, and hip-hop [2]. The k-nearest neighbors by default, operates in a non-linear manner and has the ability to perceive both linear and non-linear distribution of information. It tends to perform excellently when there is a large amount of data. The application of the Support Vector Machine is possible in both linear and non-linear approaches. The Support Vector Machine is perfect as soon as we obtain a limited collection of points across multiple dimensions because it easily discovers the linear separation that should exist. Support Vector Machine is fine with strangers as it just uses the very connected points to locate a linear splitting [1]. We utilize MFCC (Mel Frequency Cepstral Coefficient) to extract audio signal features, especially in audio signal processing. MFCC represents the spectral envelope's shape with a concise set of features (typically 10-20) [8].

VOL. 9. NO. 1 AUGUST 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v9i1.4324

Not only classifying genres, but our experiment also discusses recommender systems that use cosine similarity. Cosine similarity is a measurement utilized to assess the degree to which documents resemble each other, regardless of their magnitude. From a mathematical perspective, cosine similarity determines the cosine value of the angle formed by projecting two vectors onto a space with multiple dimensions. The cosine similarity is beneficial because it allows for two comparable documents to have a lesser angle between them, even if the Euclidean distance widely separates them due to their size (such as when one document contains the word song 50 times and another contains it only 10 times). The more accurate the angle, the greater the resemblance[9].

MATERIALS AND METHODS

We adopted the framework from Fandy, Nanna and Mohd, 2020 [10], for classifying music genres. The stages in this system are as follows in the Figure.1

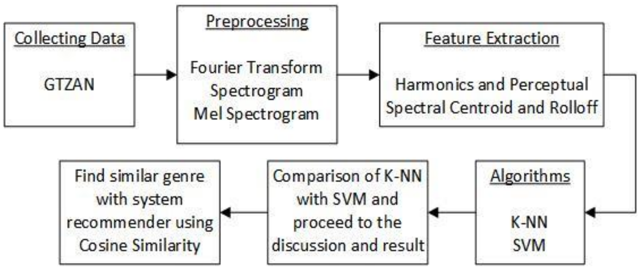

Figure 1. Flowchart

a. Data collection

G. Tzanetakis proposed GTZAN, an extensively utilized music compilation dataset for genre categorization. Around 2000-2001, various sources like CDs, DVDs, radios, and microphone recordings were used to compile the data files. This collection comprises 1000 music tracks (the overall summary of the GTZAN dataset is shown in Table.1), each with a duration of 30 seconds, sampled at a frequency of 22050 Hz and with a bit depth of 16. There are 9991 rows or records, each containing 59 columns or features/attributes. We split the dataset into 70% - 30% training and testing. So that there are approximately 7000 data trains and 3000 data testing. The GTZAN contains various music genres such as pop, reggae, metal, jazz, blues, disco, classical, hip hop, rock, and country. Each of these genres has a track record of 100 music tracks. The .wav format is utilized to examine every audio track [11].

Data is drawn from 10 genres to 100 audio libraries, all 30 seconds long, from the famous GTZAN dataset, an optical description of every audio file. Data classification can be accomplished by utilizing a neural network as a method. Since NNs (like K-NN, which will be used today) commonly utilize a visual depiction, the audio files converted into Mel Spectrograms can be used. The following 2 CSV files contain featured audio files. One file has an average variance of 30 seconds long for every song, which is calculated via several attributes which can be derived from the audio recording. from the audio file. Alternate files have identical frameworks, except the music is previously divided within 3-second soundtracks (thus, data magnitude fed towards the author categorization model is increased by ten times)[11].

Table 1. Overall Summary GTZAN Dataset

No	Music Class	Number of classes label
1	Pop	1000
2	Reggae	1000
3	Metal	1000
4	Jazz	1000
5	Blues	1000
6	Disco	1000
7	Classical	1000
8	Hip hop	1000
9	Rock	1000
10	Country	1000

b. Preprocessing

1) Fourier Transform

The researcher only captures the amplitudes produced by many waves when the researcher records sound, as shown in Figure 2. A mathematical concept called the Fourier transform separates a signal into its individual frequencies. In addition to identifying the spectral components found within a signal, the Fourier transform also establishes the amplitude of each frequency present in the signal [12].

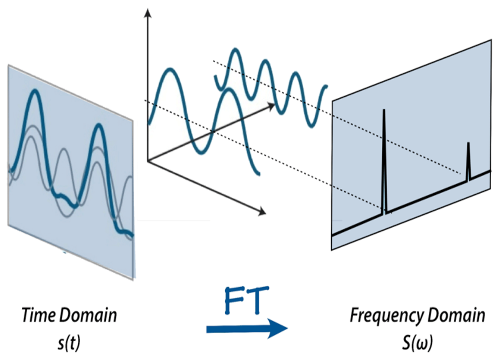

Figure 2. Fourier Transform

P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v9i1.4324

2) The Spectrogram

A spectrogram, also called a sonograph or sound gram, visually represents a signals frequency spectrum. It often uses logarithmic axes for the frequency. Spectrograms are created using the Fourier Transform, squaring the magnitude of the Short-Time Fourier Transform STFT to indicate sound power for specific frequency and time. [13]. An example of a spectrogram image can be seen in Figure 3.

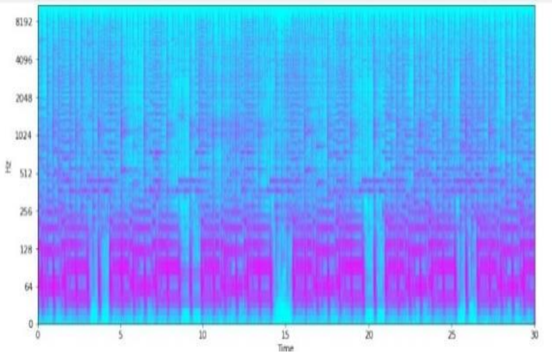

Figure 3. Spectrogram

3) Mel Spetrogram

The frequencies in a spectrogram are converted into the mel scale to create a mel spectrogram [13]. this widely used method in speech technology applies principles from the human hearing system. It involves linear and logarithmic filtering of sound signals to handle different frequency ranges [14]. An example of a Mel spectrogram image can be seen in Figure 4.

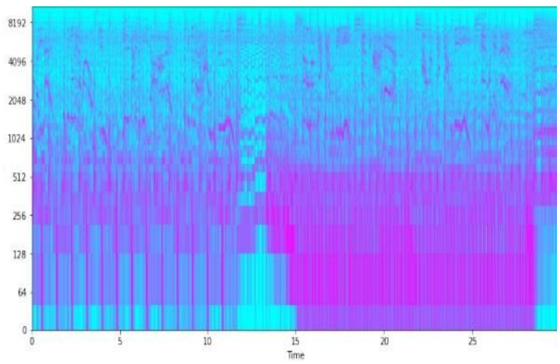

Figure 4. Mel Spectrogram

c. Feature extraction

1) Harmonics and Perceptual

The value by which the signal changes from positive to negative. Harmonics are characteristics that can't be distinguished by the human ear (representing the colour of the sound), as shown in Figure 5. Perceptual understanding of shock waves means rhythms of sound and emotion.

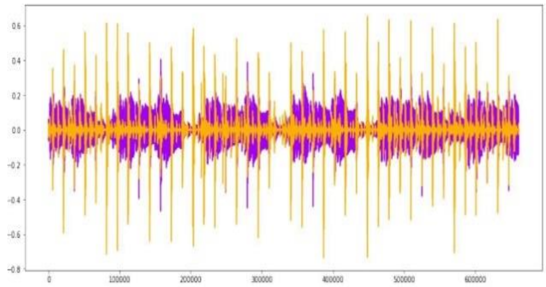

Figure 5. Harmonics and perceptual

2) Spectral Centroid and Roll off

It displays the location of the sounds center of mass, which is established by computing a weighted average while accounting for the sounds frequencies, that are located below given percentage of the total spectral energy, like 85%. An example of a Spectral Centroid and Rolloff image can be seen in Figure 6.

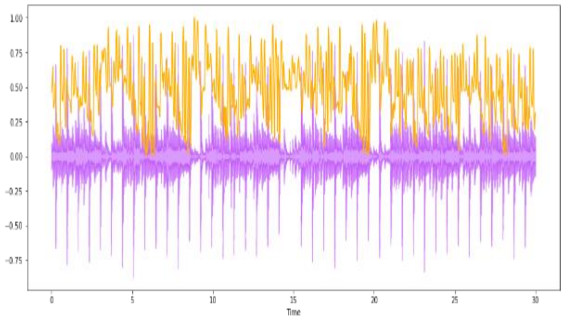

Figure 6. Spectral Centroid and Rolloff

Mel-Frequency Cepstral Coefficients (MFCC) of a signal is a small set of features (usually around 10-20) that briefly describe the overall shape of the spectral envelope. It models the characteristics of the human voice. An example of a Mel-Frequency Cepstral Coefficients (MFCC) image can be seen in Figure 7.

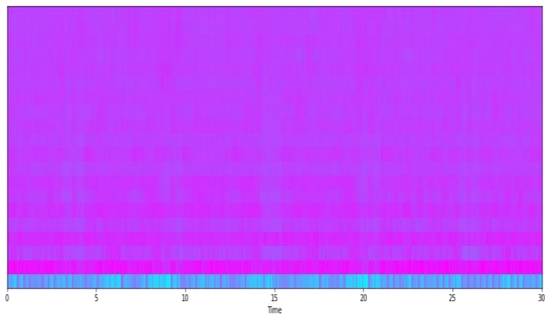

Figure 7. Mel- Frequency Cepstral Coefficients (MFCC)

3) In exploring audio data, we use libros, the parent source of the audio file. Explore audio data or EDA using the Reggae.0036.wav file, as shown in Figure 8. The vibrational ebb in various intencity (y),

VOL. 9. NO. 1 AUGUST 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v9i1.4324

and the sample rate (sr) is the audio data count samples carried per second, measured in Hz or kHz.

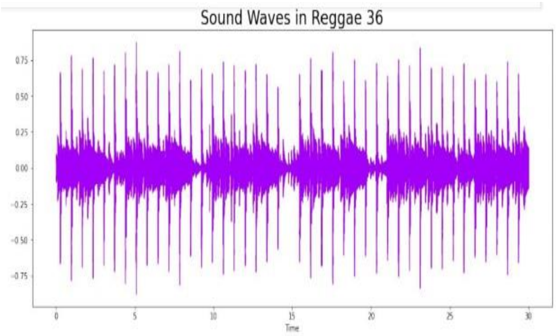

Figure 8. Sound Waves in Reggae 36

d. Algorithms

There are two algorithms that we use:

1) K-Nearest Neighbors (K-NN)

K-nearest neighbor K-NN classification is a basic and straightforward method, often chosen as the initial approach for classification studies with limited prior knowledge about data distribution. [15]. This kind of algorithm is used to solve classification and regression problems. It is a sufficient but lazy algorithm that performs no training until it receives the training sets. The K-nn algorithm is the basic logic of the implementation [16]. The researcher has a training set A, denoted as A = {a1, a2, a3,……….an}, with corresponding classes C = {c1, c2, c3,……….cs}, as well as a test set T, denoted as T = {t1, t2, t3,……….tn}. The objective is to find the initial k neighbors in the training set that is closest to the test set. The program iterates through sets A and C, calculating the KL divergence between elements ai and corresponding elements from set CJ. Distances and class lists are stored and sorted. The k nearest neighbors are returned. Each neighbor contributes a vote for their class; the class with the highest votes is the final output. The process repeats in a loop[17].

2) Support Vector Machine (SVM)

The SVM is a valuable statistic machine learning technique that has been successfully applied in the pattern detection zone. Support Vector Machine derive from the theory of minimizing structural risks through the theory proposing that knowledge is gained through computational processes [4]. In order to calculate the geometric distance between a data point and the hyperplane, it is necessary for the researcher to normalize it based on the magnitude of w. The resulting distance can be described as:

$$d\left((w,b),x\right) = \frac{yi\,(xi.w+b)}{||w||} \geq \frac{1}{||w||} \text{.......(1)}$$

The one that naturally selects is the hyperplane that provides the most significant geometric distance from the nearest data points. [4]

The researcher employed SVM for data classification, using a mercer kernel to project data into higher-dimensional space. Understanding the outcome in this expanded space was important. SVM selected support vectors and weights to improve the boundary between training orders and the classifier boundary. Using complete tracks as samples solved the track categorization issue. SVM is suitable for direct and indirect strategies, identifying divisions with incomplete points in different dimensions and effectively handling outliers with relevant data points[18].

3) Confusion Matrix

A confusion matrix is a useful tool in data mining and machine learning for evaluating the accuracy of label predictions. It is commonly used in classification models to determine the appropriate labels based on data attributes. The confusion matrix is presented as a table, showing the frequency of correct and incorrect predictions. Rows correspond to actual labels, and columns represent predicted labels by the model [19]. The classification process in the Confusion Matrix is represented by four terms, which include: [20]

a) TP (True Positive) is the amount of data in which the actual class and predictions are positive classes.
b) FN (False Negative) is the total data whose actual class is positive, while the predicted class is negative.
c) FP (False Positive) is the amount data whose actual class is negative while the predicted class is positive.
d) TN (True Negative) is the amount of data whose actual class is negative, while the precision class is negative.

4) Cosine Similarity

Cosine similarity is an approach that enables us to evaluate the similarity of content across different datasets. In this instance, cosine similarity serves as a measure to assess the magnitude that can be employed to interpret the closeness of the distance relying on the likeness between artists utilizing available datasets.

The usual way to determine the similarity between two items is by counting the number of events they have in common during another session. As a result, the calculation of item similarity can be subsequently determined [21].

P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v9i1.4324

$$\cos\theta_{QD} = \frac{\sum_{i=1}^{n} Q_i D_i}{\sqrt{\sum_{i=1}^{n}(Qi)^2 \cdot \sqrt{\sum_{i=1}^{n}(Di)^2}}} \text{..........(2)}$$

Equation 1 computes the distance using cosine similarity based on the similarity of Q to D. Q and D represents artist being tested, and their similarity will be evaluated in relation to the similarity between the two artist. In this equation, n represents a portion of the data related to two artist, specifically Q and D [21].

RESULTS AND DISCUSSION

The authors used the PCA method to classify music genres in previous experiments. With 6 genres of music, that is: blues, country, hip-hop, pop, reggae, and rock. But using the PCA method several weaknesses make the accuracy low, namely 58% [1]. Then, we decided to change the method of using K-NN and SVM and upgrade the genre to 10 genres, which include blues, classical, rock, jazz, reggae, metal, country, pop, disco, and hip-hop.

Classification accuracy varies based on different genres and machine learning. Each accuracy in each genre is shown in the following of Table 2.

Table 2. Accuracy of each genre and algorithms

Genre	K-NN	SVM
Blues	49%	83%
Classical	90%	94%
Country	62%	70%
Disco	55%	66%
Hip-hop	57%	74%
Jazz	73%	90%
Metal	76%	83%
Pop	71%	80%
Reggae	54%	71%
Rock	62%	59%

In general, our observation revealed that SVM proves to be a more efficient classifier with a 77% accuracy rate, while the accuracy of K-NN is 64,9%.

Here the confusion matrix is used to predict the accuracy of the algorithm. To begin with, the authors employed the K-NN (K-Nearest Neighbors) for categorizing [19] [22].

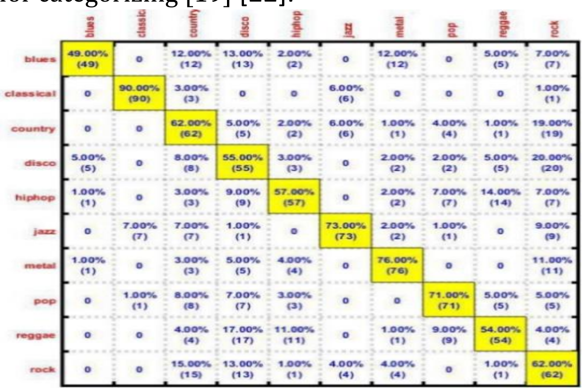

Figure 9. Matrix K-NN

As shown in Figure 9, the result of the training recognition rate is 57,6%, and validating recognition rate is 64,9% [22].

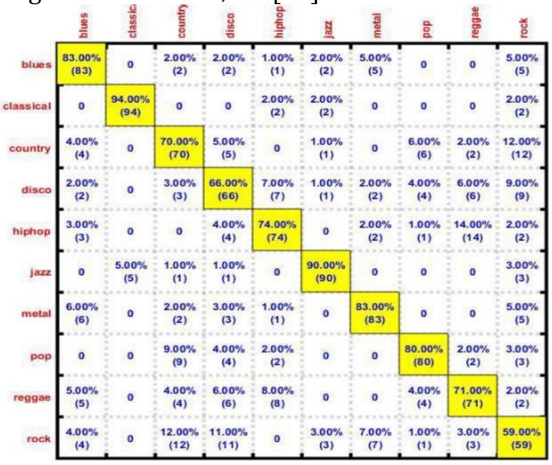

Figure 10. Matrix SVM

Figure 10 shows the result of the training recognition rate is 99,01% and validating recognition rate is 77.00%. See Table 2 for details on each genre's accuracy and algorithm.

The recommender system computes the cosine similarity between every pair of songs in the dataset. This generates a matrix of dimensions 100 x 100 (with duplicated information since the similarity of a song A to song B is equivalent to the similarity of song B to song A). The recommender system enables any given vector to locate the most suitable match in a descending order of a ranking, ranging from a better match to the very best match. The cosine similarity library will be used to accomplish this task for audio files.

The function of "Find similar songs()" is pre-defined and designed to receive a song's name as input and provide the top 5 songs that closely resemble it as a output.

Pop example:

```
# pop.00019 - Britney Spears "Hit me baby one more time"
find_similar_songs('pop.00019.wav')

ipd.Audio(f'{general_path}/genres_original/pop/pop.00019.wav')
```

```
*******
Similar songs to  pop.00019.wav
filename
pop.00023.wav     0.862836
pop.00034.wav     0.860499
pop.00078.wav     0.829135
pop.00088.wav     0.824456
pop.00091.wav     0.802269
Name: pop.00019.wav, dtype:
float64
```


Accredited Rank 2 (Sinta 2) based on the Decree of the Directorate General of Higher Education, Research and Technology. Ministry of Education, Culture, Research and Technology of the Republic of Indonesia No.225/E/KPT/2022, 07 December 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 9. NO. 1 AUGUST 2023
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v9i1.4324

From the code above, it can be concluded that there are 5 songs that have the same relative genre as the sample song used, namely (pop.00019), with accuracy details in the following of Table 3.

Table 3. Result of Cosine Similarity

No	Audio Name	Accuracy
1	Pop.00023.wav	0.862836
2	Pop.00034.wav	0.860499
3	Pop.00078.wav	0.829135
4	Pop.00088.wav	0.824456
5	Pop.00091.wav	0.802269

CONCLUSION

The study aimed to enhance music genre classification accuracy by considering audio composition details. The traditional PCA method yielded low accuracy (58%), leading to the adoption of K-Nearest Neighbors (K-NN) and Support Vector Machine (SVM) algorithms with an expanded set of 10 music genres. SVM outperformed K-NN, achieving 77% accuracy compared to K-NN's 64.9%. The confusion matrix confirmed the algorithms' effectiveness, with K-NN's training recognition rate at 57.6% and validating recognition rate at 64.9%, and SVM's training recognition rate at 99.01% and validating recognition rate at 77%. Moreover, a recommender system using cosine similarity successfully recommended similar songs based on audio features. We tested the genre of the song from Britney Spears with the title Hit me baby one more time, and the result of the song's genre is pop. Overall, the study demonstrated SVM's effectiveness in music genre classification and the value of cosine similarity for building a song recommender system. By carrying out this research, it is hoped that a wide range of potentials and benefits will be opened up for the community, especially for music lovers. This research seeks to bridge tensions and strengthen connections between songs, identify songs that have something in common, and limit community groups with the same interests in certain musical genres.

REFERENCE

[1] A. C. M. da Silva, M. A. N. Coelho, and R. F. Neto, "A Music Classification model based on metric learning applied to MP3 audio files," *Expert Syst Appl*, vol. 144, pp. 3–35, Apr. 2020.

[2] H. C. Ceylan, N. Hardalaç, A. C. Kara, and F. Hardalaç, "Automatic Music Genre Classification and Its Relation with Music Education," *World Journal of Education*, vol. 11, no. 2, p. 36, Apr. 2021.

[3] A. Viloria, O. B. P. Lezama, and D. Cabrera, "Segmentation process and spectral characteristics in the determination of musical genres," *Procedia Comput Sci*, vol. 175, pp. 96–101, Aug. 2020.

[4] B. R. Ismanto, T. M. Kusuma, and D. Anggraini, "Indonesian Music Classification on Folk and Dangdut Genre Based on Rolloff Spectral Feature Using Support Vector Machine (SVM) Algorithm," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 15, no. 1, p. 11, Jan. 2021.

[5] Sarofi Muhammad Abid As, Irhamah, and Mukarromah Adatul, "Identifikasi Genre Musik dengan Menggunakan Metode Random Forest," *Jurnal Sains dan Seni ITS*, vol. 9, no. 1, pp. 1–8, Jun. 2020.

[6] V. Fauziah Ramadhani, B. Hidayat, and Azizah, "STEGANALISIS UNTUK FILE AUDIO BERFORMAT MP3 DENGAN METODE LEAST SIGNIFICANT BIT (LSB) PADA KLASIFIKASI PRINCIPAL COMPONENT ANALYSIS (PCA)," *ITN Malang*, vol. 3, no. 1, pp. 1–6, 2017.

[7] R. Mayapada, G. M. Tinungki, and N. Sunusi, "Penerapan Sparse Principal Component Analysis dalam Menghasilkan Matriks Loading yang Sparse," *Jurnal Matematika Statistika & Komputasi*, vol. 15, no. 2, pp. 44–54, Jan. 2019.

[8] Doshi Sanket, "Music Feature Extraction in Python," *Towards Data Science*, Dec. 30, 2018. https://towardsdatascience.com/extract-features-of-music-75a3f9bc265d#:~:text=MFCC%20%E2%80%94%20Mel%2DFrequency%20Cepstral%20Coefficients&text=%3D'time')-,,a%20number%20of%20frames%20available. (accessed Jul. 19, 2023).

[9] S. Kumar, "EVALUATING THE EFFECTIVENESS OF MUSIC RECOMMENDATIONS USING COSINE SIMILARITIES OF VARIOUS COMBINED FEATURE SETS CONSISTING OF METADATA AND USER GENERATED TAGS," North Dakota State University of Agriculture and Applied Science, 2022.

[10] F. S. Utomo, N. Suryana, and M. S. Azmi, "Stemming impact analysis on Indonesian Quran translation and their exegesis classification for ontology instances," *IIUM Engineering Journal*, vol. 21, no. 1, pp. 33–50, Jan. 2020.

[11] J. Ho Kin Pou, H. Keshav Rao, G. Bhambhani, J. Joseph, and S. J. Prakash Asst, *Music Genre*

Classification using Machine Learning. Bengaluru: River Publishers, 2020.

[12] Kartik Chaudhary, "Understanding Audio data, Fourier Transform, FFT and Spectrogram features for a Speech Recognition System," *Towards Data Science*, Jan. 19, 2020. https://towardsdatascience.com/understanding-audio-data-fourier-transform-fft-spectrogram-and-speech-recognition-a4072d228520 (accessed Jul. 19, 2023).

[13] D. Lionel, R. Adipranata, and E. Setyati, "Klasifikasi Genre Musik Menggunakan Metode Deep Learning Convolutional Neural Network dan Mel-Spektrogram," *Universitas Kristen Petra*, vol. 7, no. 1, pp. 1–5, 2020.

[14] U. Ayvaz, H. Gürüler, F. Khan, N. Ahmed, T. Whangbo, and A. A. Bobomirzaevich, "Automatic Speaker Recognition Using Mel-Frequency Cepstral Coefficients Through Machine Learning," *Computers, Materials and Continua*, vol. 71, no. 2, pp. 5511–5521, Nov. 2022.

[15] S. Uddin, I. Haque, H. Lu, M. A. Moni, and E. Gide, "Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction," *Sci Rep*, vol. 12, no. 1, Dec. 2022.

[16] X. Mu, "Implementation of Music Genre Classifier Using KNN Algorithm," *Highlights in Science, Engineering and Technology CSIC*, vol. 34, pp. 149–153, Feb. 2023.

[17] Indrianti Ismayani, "Klasifikasi dengan KNN (K-Nearest Neighbors) menggunakan Python," Jun. 2019. https://medium.com/@16611086/klasifikasi-dengan-knn-k-nearest-neighbors-menggunakan-python-bf772e8d4ef5 (accessed Jul. 19, 2023).

[18] S. Y. Yehezkiel and Y. Suyanto, "Music Genre Identification Using SVM and MFCC Feature Extraction," *IJEIS (Indonesian Journal of Electronics and Instrumentation Systems)*, vol. 12, no. 2, p. 115, Oct. 2022.

[19] KANTINIT, "Confusion Matrix: Pengertian, Cara Kerja dan Contoh Soal," *Kantinit*, Jan. 09, 2023. https://kantinit.com/kecerdasan-buatan/confusion-matrix-pengertian-cara-kerja-dan-contoh-soal/ (accessed Jul. 19, 2023).

[20] Trivusi, "Metriks Evaluasi Sistem Menggunakan Confusion Matrix," *Trivusi*, Jul. 16, 2022. https://www.trivusi.web.id/2022/04/evaluasi-sistem-dengan-confusion-matrix.html (accessed Jul. 20, 2023).

[21] M. Ludewig, I. Kamehkhosh, N. Landia, and D. Jannach, *Effective nearest-neighbor music recommendations*. Vancouver: Association for Computing Machinery, 2018.

[22] H. Abdulbar, P. P. Adikara, and S. Adinugroho, "Klasifikasi Genre Lagu dengan Fitur Akustik Menggunakan Metode K-Nearest Neighbor," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 8, pp. 8259–8268, Aug. 2019.

Accredited Rank 2 (Sinta 2) based on the Decree of the Directorate General of Higher Education, Research and Technology. Ministry of Education, Culture, Research and Technology of the Republic of Indonesia No.225/E/KPT/2022, 07 December 2022. Published by LPPM Universitas Nusa Mandiri