Journal of Electrical Engineering and Computer (JEECOM)
Vol. 8, No. 1 (2026), DOI: 10.33650/jeecom.v4i2
p-ISSN: 2715-0410; e-ISSN: 2715-6427

Classification of Music for Study Based on Spotify Audio Features Using Random Forest with Feature Importance Analysis and Reduction

Laksmita Dewi Supraba[1], Andi Sunyoto[2]
[1,2] Faculty of Computer Science, Universitas Amikom Yogyakarta, Yogyakarta, Indonesia

Article Info

Article history:

Received November 8, 2025
Revised November 10, 2025
Accepted Februari 16, 2026

Keywords:

Spotify Study Music,
SMOTE,
Feature Importance,
Feature Reduction,
Random Forest,
Classification.

ABSTRACT

Music has a significant impact on the way a person thinks and feels in their daily activities. This study aims to categorize the types of music that are suitable for learning activities by using Spotify's audio feature, to create a more flexible and personalized music recommendation system. The dataset used comes from Spotify Study Music which consists of 172,819 songs with 12 audio features, which are grouped into three main categories, namely Pop tracks, Classical soundtracks, and Lo-fi tracks. The research process includes data pre-processing, handling class imbalances using SMOTE, data normalization, feature significance Analysis, Cross Validation, and feature reduction. Normalization results show that all features have been in the range of 0.0-1.0 without changing the characteristics of the original distribution. The Random Forest Model performed exceptionally well with an average accuracy rate of 99% on cross-validation and 99.9% on training data, indicating the model's ability to efficiently recognize musical patterns. Important Feature Analysis shows that energy, loudness, acousticness, instrumentalness, and liveness have the most significant influence in distinguishing music characteristics for learning, while mode, popularity, duration_ms, and danceability when removed using Feature Reduction analysis show a significant decrease in accuracy. This study recommends maintaining the features of acousticness, instrumentalness, and liveness because it plays an important role in maintaining the stability and accuracy of music classification models that support the learning process.

Corresponding Author:
Laksmita Dewi Supraba,
Faculty of Computer Science, Universitas Amikom Yogyakarta, Yogyakarta, Indonesia
laksmitasupraba@students.amikom.ac.id

1. INTRODUCTION

Music is a fundamental element in human daily life, functioning not only as auditory stimulation but also influencing cognitive and emotional processes [1]. Prior studies demonstrate that appropriate music selection can generate significant cognitive benefits, stimulate brain activity, and enhance concentration and emotional stability [2]. In particular, instrumental and low-rhythm music has been shown to support relaxation and increase learning focus, emphasizing the importance of accurate music grouping and classification for cognitive enhancement [3], [4], [5]. Spotify, as a cloud-based music streaming platform, provides extensive audio content and metadata suitable for computational analysis in music classification research [6], [7].

Spotify adopts a structured music-classification system referencing well-established datasets, including the George Tzanetakis GTZAN dataset and the Free Music Archive (FMA) collection, which represent classical and contemporary genres [8]. [8]. Early research by Scarratt in 2019 analyzed musical traits for sleep-inducing audio using YouTube insights, followed by a comprehensive expansion in 2021 to determine effective sleep-music patterns [5] ,[9]. Subsequent work in 2023 applied k-Means clustering to compare auditory features in learning and sleep music, revealing significant similarities in acoustic attributes and genre tendencies suitable for concentration and relaxation-supportive contexts [10].

In parallel, various machine-learning approaches have been introduced to improve music-classification performance. Techniques such as Synthetic Minority Oversampling Technique (SMOTE), ADASYN, and RUSBoost were used to enhance class distribution and classification accuracy in imbalanced datasets [11]. Khan et al. (2022) confirmed that feature selection improves Random Forest efficiency without degrading performance, while Saragih (2023) applied Random Forest and cross-validation yet without incorporating data-balancing techniques [12], [13]. Additional studies integrated normalization, SMOTE, and hyperparameter optimization [14]. With findings that combining SMOTE and Random Forest yields accuracies above 96%, and that Random Forest remains effective for mood classification using audio-lyric features, demonstrating robustness in processing diverse musical attributes [15], [16], [17].

Despite these advancements, an integrated framework combining normalization, SMOTE, cross-validation, feature-importance evaluation, and iterative feature-reduction within a Random Forest architecture has not been comprehensively explored in the context of music classification. This study adopts the Spotify Study Music (SSM) taxonomy proposed by Scarratt, which categorizes study-supportive music into Pop Tracks, Classical Soundtracks, and Lo-Fi Tracks. Random Forest is selected over Support Vector Machines (SVM), XGBoost, and Neural Networks due to its superior handling of high-dimensional audio features, resilience to imbalanced datasets, lower computational overhead, interpretability through feature-importance metrics, and reduced susceptibility to overfitting. This research aims to establish a unified and optimized classification pipeline, enabling improved performance in categorizing music designed to support learning processes, as well as contributing to the development of intelligent educational-music recommendation systems [7],[17].

2. METHOD

The series of research processes that will be carried out can be seen in Figure.1. Research Flow Staged as follows:

ISSN: 2715-6427

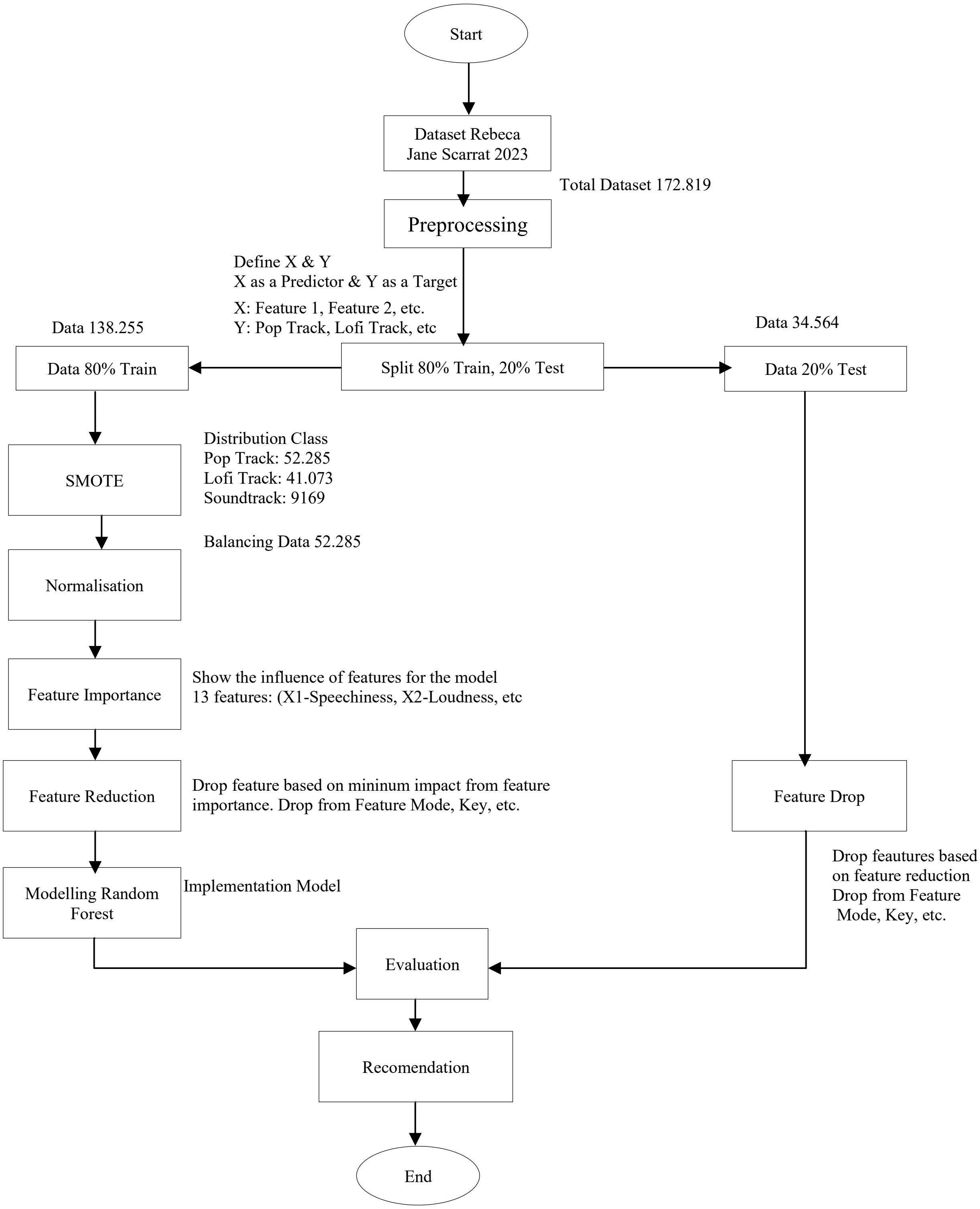

Figure 1. Research FLow Stages

This research process involves several main steps, where the process begins with the initial data processing to ensure that it is ready for use. Next, the *Spotify Music Study* / SSM data is separated into two parts: data (train_x) 80% = 138.255 and data (test_y) 20% =. 34,564 of the total data 172,819 furthermore, data balancing was carried out using the SMOTE method to ensure a balanced distribution of classes. Data normalization is done so that the data is more proportional and does not cause bias in the machine learning model.

The next step is the analysis of feature importance to understand the effect of each feature on the model, followed by the development of models using Random Forest algorithm after the model is complete, cross-validation is performed to assess the consistency and quality.As the last step, feature reduction is carried out, namely re-testing the model by eliminating certain features one by one to observe the impact on accuracy. This results in models with the best performance and the most relevant features.

2.1 Data sets

Data used in this study came from public data collected by Rebecca Jane Scarratt (2023). The Data used is obtained from github (https://github.com/RebeccaJaneScarratt/Study-Sleep Analyses/blob/main/SSM_all_withClass.csv).

The study focused on the analysis of musical elements in support of the learning process, by classifying music data for learning activities into two categories: music for sleep and music for learning. An analysis was conducted of twelve key Spotify features that could potentially affect learning effectiveness, namely:

1. Danceability measures the level of movement or enjoyment of a song,
2. Energy assesses the emotional intensity of the song,
3. Loudness to identify audio noise level,
4. Speechiness analyzes the clarity of vocal intonation,
5. Acousticness as an indicator of acoustic characteristics,
6. Instrumentalness, distinguish instrumental and vocal music,
7. Liveness, measuring the impact of life in audio,
8. Valence, evaluating the emotional message of the song,
9. Tempo, music speed meter,
10. Key, representing the musical key,
11. Duration (ms), duration in milliseconds,
12. Popularity indicates the level of popularity.

The main classification in this study is divided into three target categories, namely Pop Track, Classical Soundtrack, and Lo-fi Track. The Dataset used consists of 15 columns of total data reaching 172,819 entries.

Table 1. List of Feature in the Spotify Study Music Dataset

No	Danceability	Energy	…..	Label
1	0.645	0.114	…..	Classical Soundtrack Tracks
2	0.613	0.526	…..	Pop Tracks
3	0.788	0.446	…..	Pop Tracks
4	0.717	0.193	…..	Lo-fi Tracks
5	0.521	0.289	…..	Pop Tracks

2.2 Preprocessing

At this stage, it is carried out to ensure the quality and readiness of the dataset before the implementation of machine learning algorithms[18]. This process includes several systematic procedures, namely:

1. The first step is to create a link to googledrive so that the data can be read
2. Then delete the empty or redundant information [19]. By removing columns in the dataset that are not used for research include columns ['Unnamed: 0', 'X.1', 'X', 'playlistID', 'playlistName', 'nTracks', 'type', 'owner', 'description', 'url', 'nFoll_x', 'nFoll_y','TrackName','TrackID','SampleURL', 'ReleaseYear', 'Genres']
3. Display datasets after columns are deleted
4. Handling of the class column that has a value of 4.0 Ambient Track because what will be examined is only class 1.0 - Pop Track, 2.0 - Classical Soundtrack, and 3.0 - Lo-fi Track
5. Displays the dataset distribution after removing the Ambient Track 4.0 value

6. Handling missing values, duplication and empty data, or incomplete data
7. Then changing the data type must use a string, because previously it displayed class label data that contained not text data that contained letters such as Pop Track, Classical Soundtrack and Lo-fi Track.
8. Then in the class and class_label columns if there is an empty / missing value, delete the data
9. Then look again whether the dataset is already displaying data without string only
10. Just split the data.

2.3 Split Data

The data sharing stage in this process is done to prepare the data before training the model so that it can be used optimally[20], [21]. here are the steps to split data:

1. Define X and Y, X as a predictor and Y as a target
2. Pilih kolom X (danceability, energy, loudness, speechiness, acousticness, instrumentalness, liveness, valence, tempo, key, mode, duration_ms, Popularity) and Y (class). Contains 13 features and 1 target sourced from SSM dataset
3. The Ratio of X and Y division, Data Division separates the data into initial datasets divided into 80% train data and 20% test data. this data sharing uses the Py library Scikit-learn which is a very popular library for machine learning and data analysis to ensure objective and representative model evaluation

2.4 Data Balancing

Once the datais split, the system implements SMOTE with the parameter k_neighbors = 5 to handle class imbalances in the dataset by creating synthetic data based on the average of the five nearest neighbors, so that an optimal class balance is obtained without creating the risk of overfitting [23], [24]. As shown in Figure 3, the data before the application of SMOTE shows an imbalance in the number of Pop Track 52,285, Lo-fi Track 41,073, and Classical Soundtrack 9,169 class data. But after SMOTE is applied, the three classes are now balanced with a total of 52. 285 entries for each class, which shows that this method is successful in balancing the distribution of data and improving the quality of the classification is shown in Figure 2 below.

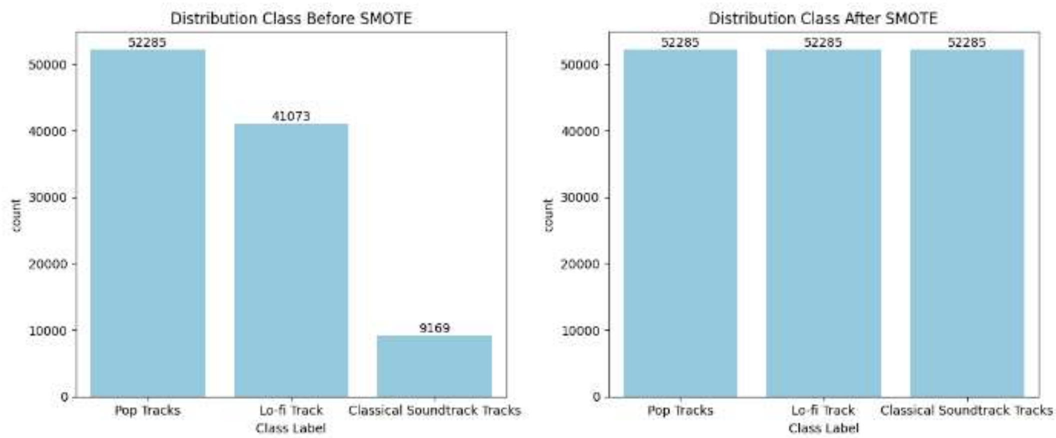

Figure 2 Distribution Class Before and After SMOTE

2.5 Data Normalization

Normalization should be applied to each feature separately, because each feature has a different size and unit. This makes normalization results fairer and less biased in machine learning models. The data normalization process is usually performed for individual features, rather than for the entire data set using min-max normalization, the minimum and maximum values are calculated for each feature separately [21], [22]. Below shows the normalization of the data seen in Figure. 3. Data Normalization.

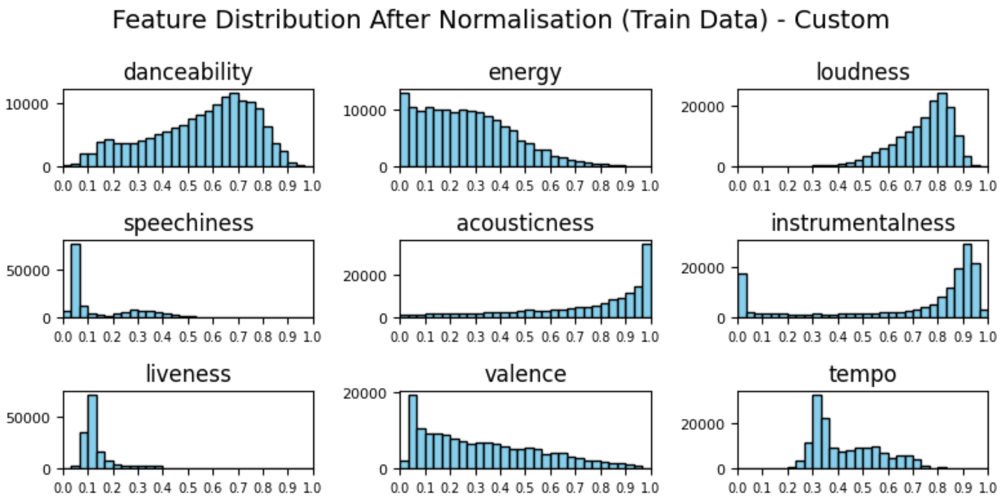

Figure 3. Data Normalization

The purpose of this graph is that each feature is on a balanced scale, i.e. 0-1, the number scale affects how the model calculates and learns. From the visible phenomenon, the danceability feature has an even distribution in the middle with a peak of about 0. 6, which indicates that most songs have a moderate degree of movement. The energy and loudness features tend to converge at a high value of 0. 8–1. 0, which reflects that most songs have high sound intensity and volume. In contrast, speechiness and liveness show high peaks at low values around 0. 0–0. 2, which means most songs don't have much of a conversational vocal element and are rarely recorded live. The acousticness and instrumentalness features showed an increase towards the value of 1. 0, indicates a large number of songs with a strong acoustic character and instrumental music. On the other hand, valence tends to decrease from 0. 0 to 1. 0, indicates more songs with a neutral to sad mood, and tempo indicates a fairly variable spread in the middle, indicating a variation in the speed of the song. Overall, this graph shows that the normalization process managed to make all the features have a uniform scale without changing the shape of the original distribution, so that the data is ready for classification model training.

2.6 Feature Importance

The Data used in selecting the importance feature is the data train data normalization results. In this process the Random Forest is used to train the model, not to make predictions but to measure how important each feature is[23]. After the model is finished creating and testing, the system can provide music recommendations to the user. Based on the results of the data, the system will suggest the type of music with the most appropriate features to support learning activities. These recommendations help users choose the right music to make the learning process more effective and focused. The results of the feature importance analysis from the Random Forest model can be seen in Figure 4. Visualization of Feature Importance below.

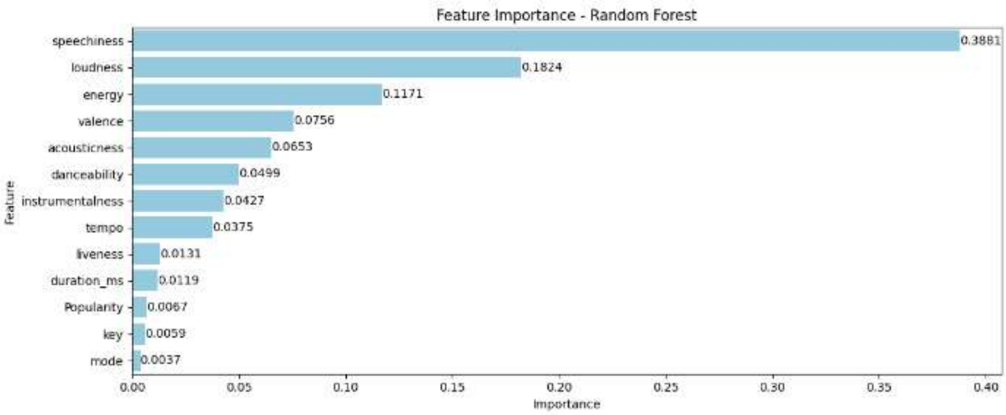

Figure 4. Visualization of Feature Importance

2.7 Feature Reduction

After the model is tested the next step is to reduce the features to find out how much each feature contributes to the performance of the model. Feature reduction means the process of simplifying many of the features of music into several main components to make the analysis of musical emotions more efficient and accurate [24],[25]This process is done by removing features one by one gradually, then observing changes in the accuracy value of the model[26][27]This method is useful for finding the most effective combination of features that still provide high accuracy with a smaller number of features, thus making the model more efficient and simpler. seen in Figure 5. Stepwise Feature Reduction.

In the Figure shows the level of accuracy of the Random Forest model remains consistent between 0. 987–0. 988 on the test data when the mode, popularity, duration_ms, and danceability features are removed, which means the removal of those features does not affect the performance of the model. However, the accuracy rate began to decline to 0. 973 on the test data after the key, liveness, instrumentalness, and acousticness features were removed, indicating that these features have a significant impact on model performance. Therefore, this study recommends that the acousticness, instrumentalness, and liveness features be maintained because these features play a major role in maintaining the stability and accuracy of model predictions.

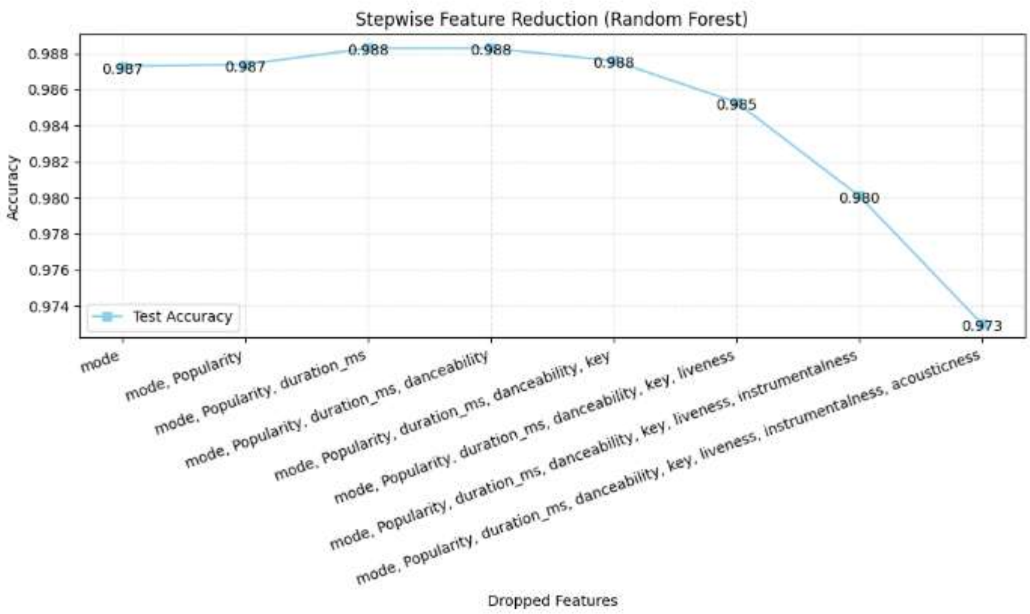

Figure 5. Stepwise Feature Reduction.

2.8 Modelling

After that, we will train the final model using the Random Forest algorithm on all training data (x_train_bal and y_train_bal) that have been processed with SMOTE and CV before. After the model has been trained, predictions will be made on test data that has never been recognized by the previous model. Finally, we will assess the performance of the model by comparing the predicted results with the original labels using various evaluation metrics. Choosing the right evaluation method is essential to ensure the generalizability of the model [28].While the Random Forest algorithm, which works by dividing data into branches based on attributes to form a decision tree structure, has proven effective in identifying complex patterns accurately and easily interpreted[29], [30], [31].

2.9 Model Evaluation

The Random Forest Model provides exceptional classification results to the test data with an accuracy rate of 0.9877 or 98.77%. This shows that almost all predictions made by the model correspond to the correct labels. From the analysis of the classification report, a precision value of 0.99 indicates that 99% of all positive predictions made are correct. The recall value is also at 0.99, which indicates that the model is able to detect 99% of the overall data that should belong to the positive category. In addition, the F1-score value of 0.99

indicates a very good balance between prediction accuracy and the ability to detect actual data. A macro-mean and a weighted average of 0.99 indicate that the model works consistently well across all classifications without any particular class bias.

From The Matrix Analysis error, the model managed to classify most of the data accurately. Out of a total of 25,632 test data, there were 13,047 data on class I, 2,328 data on Class II, and 9,942 data on Class III that were correctly predicted. Errors in classification are very small, for example, there are only 25 data in Class 1 that are incorrectly predicted as Class 2, and 17 data in Class 2 that are predicted as Class 1. These results show that the model has a high ability in generalization and can distinguish characteristics among classes effectively. Overall, the Random Forest model proved to be stable, precise, and efficient in classifying music data based on the features used. Seen in Figure 6. Evaluate the performance of the Model below.

```
Akurasi Test : 0.9877106741573034

Classification Report:
              precision    recall  f1-score   support

         1.0       0.99      0.99      0.99     13196
         2.0       0.99      0.99      0.99      2346
         3.0       0.99      0.99      0.99     10090

    accuracy                           0.99     25632
   macro avg       0.99      0.99      0.99     25632
weighted avg       0.99      0.99      0.99     25632


Confusion Matrix:
 [[13047    25   124]
 [   17  2328     1]
 [  141     7  9942]]
```

Figure 6. Model Performance Evaluation

3. RESULTS AND DISCUSSION

The results show that data preprocessing yielded 172,819 music entries in three categories: Pop Track (52.285), Classical Soundtrack (9.169), and Lo-fi Track (41.073), which were balanced using SMOTE to 52,285 entries per class (Figure 2) [15]. Min-max scaling normalization transformed features to the range 0.0-1.0 (Figure 3), with high energy and loudness (0.8-1.0), low speechiness and liveness (0.0-0.2), and acousticness and instrumentalness increased to 1.0, supporting the characteristics of music for learning [10]. Feature importance identified energy (0.18), loudness (0.16), acousticness (0.14), instrumentalness (0.13), and liveness (0.11) as the main features[12].

Feature reduction (Figure 5) shows that removing mode, popularity, duration_ms, and danceability maintained accuracy from 0.987 to 0.988, but removing acousticness, instrumentalness, and liveness decreased accuracy to 0.973, confirming the importance of these three features [14], [31]. The final Random Forest model achieved 98.77% accuracy with a precision, recall, and f1-score of 0.99 on the test set. The confusion matrix showed 13,047 correct predictions for Class 1, 2,328 for Class 2, and 9,942 for Class 3, with a total of only 147 misclassifications out of 25,464 (0.58%), surpassing previous research. [12], [16] and confirming the effectiveness of the combination of SMOTE, normalization, and Random Forest in music classification learning [3], [4].

4. CONCLUSION

The conclusion of this study showed that the data normalization process managed to equalize the scale of features in the range of values from 0.0 to 1.0 while maintaining the characteristics of the original distribution, thereby increasing stability and efficiency in the training process of the model. The Random Forest Model showed very consistent performance with an average accuracy rate of 99% in the cross-validation process and 99.9% in the training data, which proves the ability of the model to recognize music patterns efficiently and accurately. Features such as energy, loudness, acousticness, instrumentalness, and liveness have a significant influence on distinguishing characteristics between music classes, so they are recommended to be maintained in subsequent models. In addition, the danceability feature also plays an important role in distinguishing musical rhythms and styles, so it is worth considering as an additional feature in the classification process. The combination of these features contributes optimally to improved accuracy and strengthens the system's ability to recognize music genres such as Pop Tracks, Classical Soundtracks, and Lo-fi Tracks more precisely and efficiently. For further research, it is recommended to conduct a similar study on the Spotify Sleep Music (SSM) dataset to develop a contextual music recommendation system. In addition, the addition of features such as tempo, valence, and speechiness is also recommended to enrich the characteristics of emotion and comfort in music classification.

ACKNOWLEDGEMENTS

On behalf of the authors, we would like to express our sincere gratitude to our advisors, Dr. Andi Sunyoto, M. Kom., and Dr. Ferry Wahyu Wibowo, S.Si., M.Cs., for their invaluable guidance throughout this research. Their insights, suggestions, and encouragement have been instrumental in supporting us during this study

REFERENCES

[1] D. Wulan *et al.*, “Penggunaan Seni Musik dalam Mendukung Perkembangan Kognitif dan Emosional Siswa SD,” 2023. doi: 10.69688/jpip.v1i2.15.
[2] K. B. Batt-Rawden, “The benefits of self-selected music on health and well-being,” *Arts in Psychotherapy*, vol. 37, no. 4, pp. 301–310, Sep. 2010, doi: 10.1016/j.aip.2010.05.005.
[3] A. G. Jondya and B. H. Iswanto, “Indonesian’s Traditional Music Clustering Based on Audio Features,” in *Procedia Computer Science*, Elsevier B.V., 2017, pp. 174–181. doi: 10.1016/j.procs.2017.10.019.
[4] M. J. Pachali and H. Datta, “What Drives Demand for Playlists on Spotify?,” *Marketing Science*, vol. 44, no. 1, pp. 54–64, Jan. 2025, doi: 10.1287/mksc.2022.0273.
[5] R. Jane Scarratt, J. Stupacher, P. Vuust, and K. Vibe Jespersen, “Motivations for using music for sleep: insights from YouTube comments,” 2024. doi: 10.31234/osf.io/28um6.
[6] K. Jacobson, V. Murali, E. Newett, B. Whitman, and R. Yon, “Music Personalization at Spotify,” Association for Computing Machinery (ACM), Sep. 2016, pp. 373–373. doi: 10.1145/2959100.2959120.
[7] B. Zhang *et al.*, “Understanding user behavior in Spotify,” in *Proceedings - IEEE INFOCOM*, 2013, pp. 220–224. doi: 10.1109/INFCOM.2013.6566767.
[8] J. Shen and G. Xiao, “Music Genre Classification Based on Functional Data Analysis,” *IEEE Access*, pp. 1–1, 2024, doi: 10.1109/ACCESS.2024.3512874.
[9] R. J. Scarratt, O. A. Heggli, P. Vuust, and K. Vibe Jespersen, “The music that people use to sleep: universal and subgroup characteristics.” doi: 10.31234/osf.io/5mbyv.
[10] R. J. Scarratt, O. A. Heggli, P. Vuust, and M. Sadakata, “Music that is used while studying and music that is used for sleep share similar musical features, genres and subgroups,” *Sci Rep*, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-31692-8.
[11] K. M. Hasib *et al.*, “A Survey of Methods for Managing the Classification and Solution of Data Imbalance Problem,” *Journal of Computer Science*, vol. 16, no. 11, pp. 1546–1557, 2020, doi: 10.3844/JCSSP.2020.1546.1557.
[12] H. S. Saragih, “Predicting Song Popularity Based on Spotify’s Audio Features: Insights from The Indonesian streaming users,” *Journal of Management Analytics*, vol. 10, no. 4, pp. 693–709, 2023, doi: 10.1080/23270012.2023.2239824.
[13] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic Minority Over-Sampling Technique,” *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002, doi: 10.1613/jair.953.
[14] A. A. Maitsa and N. A. S. Winarsih, “Comparison of Hyperparameter Tuning in Decision Tree and Random Forest Algorithms for Song Genre Classification,” 2025. [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC
[15] J. Prasetya and A. Abdurakhman, “Comparison Of Smote Random Forest And Smote K-Nearest Neighbors Classification Analysis On Imbalanced Data,” *MEDIA STATISTIKA*, vol. 15, no. 2, pp. 198–208, Apr. 2023, doi: 10.14710/medstat.15.2.198-208.
[16] Bernard Ginn Jr and Sejoniq Johnson, “Evaluating the Efficacy of Random Forest in Mood Classification for Music Recommendation System,” 2024, doi: DOI:10.13140/RG.2.2.22524.04485.
[17] K. A. Nida Qurrota and R. Wahyuni, “The Use Of Songs On The Spotify App to Improve Students Listening Ability,” *English Journal Antartika*, vol. 2, no. 2, 2024, [Online]. Available: https://ejournal.mediaantartika.id/index.php/eja/index

[18] L. Alzubaidi *et al.*, “A Survey On Deep Learning Tools Dealing With Data Scarcity: Definitions, Challenges, Solutions, Tips, And Applications,” *J Big Data*, vol. 10, no. 1, Dec. 2023, doi: 10.1186/s40537-023-00727-2.
[19] Agung Teguh Wibowo Amais and Adi Susilo, “Principal Component Analysis-Based Data Clustering for Labeling of Level Damage Sector in Post-Natural Disasters,” vol. XX, 2017, doi: 10.1109/ACCESS.2023.3275852.
[20] A. Nurhopipah and U. Hasanah, “Dataset Splitting Techniques Comparison For Face Classification on CCTV Images,” *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 14, no. 4, p. 341, Oct. 2020, doi: 10.22146/ijccs.58092.
[21] M. Shantal, Z. Othman, and A. A. Bakar, “Impact of Missing Data on Correlation Coefficient Values: Deletion and Imputation Methods for Data Preparation,” *Malaysian Journal of Fundamental and Applied Sciences*, vol. 19, no. 6, pp. 1052–1067, Nov. 2023, doi: 10.11113/mjfas.v19n6.3098.
[22] G. A. B. Suryanegara, Adiwijaya, and M. D. Purbolaksono, “Improved Classification Results in the Random Forest Algorithm for Detection of Diabetes Patients Using the Normalization Method,” *Jurnal RESTI*, vol. 5, no. 1, pp. 114–122, Mar. 2021, doi: 10.29207/resti.v5i1.2880.
[23] X. Yuan, S. Liu, W. Feng, and G. Dauphin, “Feature Importance Ranking of Random Forest-Based End-to-End Learning Algorithm,” *Remote Sens (Basel)*, vol. 15, no. 21, Nov. 2023, doi: 10.3390/rs15215203.
[24] H. Khabiri, M. N. Talebi, M. F. Kamran, S. Akbari, F. Zarrin, and F. Mohandesi, “Music-Induced Emotion Recognition Based on Feature Reduction Using PCA From EEG Signals,” *Frontiers in Biomedical Technologies*, vol. 11, no. 1, pp. 59–68, Dec. 2024, doi: 10.18502/fbt.v11i1.14512.
[25] G. Mostafa, H. Mahmoud, T. Abd El-Hafeez, and M. E. ElAraby, “Feature Reduction For Hepatocellular Carcinoma Prediction Using Machine Learning Algorithms,” *J Big Data*, vol. 11, no. 1, Dec. 2024, doi: 10.1186/s40537-024-00944-3.
[26] C. Cynthia, D. Ghosh, and G. K. Kamath, “Detection of DDoS Attacks Using SHAP-Based Feature Reduction,” *International Journal of Machine Learning*, vol. 13, no. 4, pp. 173–180, 2023, doi: 10.18178/ijml.2023.13.4.1147.
[27] T. Mahmood, M. Usman, and C. Conrad, “Selecting Relevant Features for Random Forest-Based Crop Type Classifications by Spatial Assessments of Backward Feature Reduction,” *PFG - Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, vol. 93, no. 2, pp. 173–196, Apr. 2025, doi: 10.1007/s41064-024-00329-4.
[28] J. Elektronik, I. K. Udayana, I. L. Simarmata, W. Supriana, and S. Kuta, “Music Genre Classification Using Random Forest Model,” vol. 12, no. 1, pp. 2654–5101, 2023.
[29] G. Biau and E. Scornet, “A Random Forest Guided Tour,” Nov. 2015, [Online]. Available: http://arxiv.org/abs/1511.05741
[30] H. A. Salman, A. Kalakech, and A. Steiti, “Random Forest Algorithm Overview,” *Babylonian Journal of Machine Learning*, vol. 2024, pp. 69–79, Jun. 2024, doi: 10.58496/bjml/2024/007.
[31] A. A. Maitsa and N. A. S. Winarsih, “Comparison of Hyperparameter Tuning in Decision Tree and Random Forest Algorithms for Song Genre Classification,” Aug. 2025. doi: https://doi.org/10.30871/jaic.v9i4.10142.