SITEKNIK
Sistem Informasi, Teknik dan Teknologi Terapan
E-ISSN: 3032-3991
Vol. 2. No. 2 April 2025. Pages. 67-75

Performance Analysis of SVM and Random Forest Algorithms in the Case of the Influence of Music on Mental Health

Karisma Septa Kresna[1], Kusnawi[2]
[12] Informatics, Faculty of Computer Science, Amikom University Yogyakarta, Indonesia

Keywords: Data *Mining*, Mental Health, *Random Forest*, SVM, Data Music Therapy
Submitted: 16/02/2025
Revised: 21/02/2025
Accepted: 09/03/2025

Abstrak

Mental health disorders are conditions that impress a person's behavior, mindset, and emotions. According to WHO data, the rate of mental disorders in Asia has increased significantly in the past two decades, with about one-fifth of the world's adolescent population experiencing stress each year. Music has long been known to have a positive influence on mental health, and music therapy is used as one approach to assist individuals in improving social, mental, and physical conditions. In this study, the authors used data mining techniques to identify relevant patterns regarding the influence of music on mental health. Two classification algorithms, namely the Support Vector Machine (SVM) and Random Forest, is used to analyze and characterize the data. SVM is known to excel at managing high-dimensional data, while Random Forest is effective at handling data with missing outliers and features. This study purpose to oppose the performance of the two algorithms in classifying the influence of music on mental health to identify the superior algorithm in this context. The Random Forest algorithm gets 93% accuracy and SVM gets 95% accuracy, the hyperparameter tuning on the SVM algorithm has a better performance than Random Forest with an accuracy score of 97% for SVM, while for Random Forest it gets an accuracy score of 94%. The results of the study are expected to provide insight into the use of music as a mental health therapy tool.

Corresponding Author:
Karisma Septa Kresna
Informatics, Faculty of Computer Science, Amikom University Yogyakarta
Jl. Ring Road Utara, Ngringin, Condongcatur, Depok District, Sleman Regency, Special Region of Yogyakarta 55281Email: kresna@students.amikom.ac.id

INTRODUCTION

Mental health disorders are conditions that impact the way a person behaves, thinks, and feels emotions. From data sourced from WHO (2012), the rate of mental disorders cases in Asia has increased in the last two decades. In a global context, the World Health Organization reports that about one-fifth of the world's adolescent population experiences

This is an open access article under the CC BY-SA 4.0 license

stress each year (Fadilah et al., 2024). In more serious situations, this circumstance can increase the risk of developing a tendency to end one's own life (Sub'haan et al., 2023). Triggering factors for mental health disorders come from internal factors such as character, physical health, mindset, and how to cope with stress, as well as internal factors that include social, economic, and environmental conditions around the individual (Aanda et al., 2022).

Music has a highly expressive quality, unique motivation, and a means to build relationships (Allyssa et al., 2023). Music therapy is a way that can be used to help individuals of different ages to be better socially, mentally, and physically (Fadilah et al., 2024). This study refers to the thepry of daya mining. A semi-automated method that utilizes various techniques, including statostics, mathematical science, machine learning and also machine learning, to draw out discover valuable information and also knowledge from big database (Zai, 2022). Data mining can also collect information and process information sourced from big data obtained through databases, social media, or other sources. The purpose of data mining is to identify patterns, trends in data, and uniqueness that are useful as tools for business decision-making. Data mining has stages, namely data preprocessing, data exploration, model making, and evaluation (Rahman Wahid et al., 2023).

Classification is the process of grouping data that has similar characteristics into a predetermined class. In general, the data classification process is represented by classes whose characteristics have been determined (Septhya et al., 2023). The classification process an utilize methods such as Support Vector Machine (SVM) and Random Forest. This techniques offer advantages, palticularly in handling upper size. Support Vector Machine (SVM) is a method approach lest relies on ribbed functions in upper size attribute room and is trained using optimization algorithms. Its accuracy is affected by the kernel function and the parameters used. SVM is known to be fast and effective for text classification (Isnain et al., 2021). Various studies have proven that SVM is capable of generating predictions with an excellent level of accuracy (Dessy Kusumaningrum & Imah, 2020).

Among the various classification methods, Random Forest is a very suitable choice to apply in the field of fault diagnosis (Sebayang et al., 2023). Random Forest is an algorithm lest combines much decision trees to improve classification accuracy. This algorithm has low errors, is efficient in handling big data, and is able to overcome missing data with bootstrap methods (Intan Permata & Esther Sorta Mauli Nababan, 2023). Random Forest has several advantages that can improve high-accuracy results on lost data and ressiting outliers (Rahayu et al., 2023). In the research (Rahayu et al., 2023), (Rijal et al., 2024) using the Random Forest algorithm for classification shows the best performance. Therefore, the Random Forest algorithm was implemented in this study.

One way to maximize algorithm performance is by hyperparameter tuning, which is a variable whose value has an impact on the learning process and model parameters. GridSearchVC is a method that works to search for optimal combinations of hyperparameters. This method works by evaluating various combinations of hyperparameters to find the best configuration (Misriati Point, 2024). This process is often combined with k-fold cross-validation to identify optimal combinations of hyperparameters, and this method is known as Grid Search Cross-Validation or GridSearchCV (Khusna et al., 2024). Some commonly used techniques in hyperparameter tuning are grid search, Bayesian optimization, and also random searching (Purnama, 2024). Based on the problems that have been presented, this study compares the Support Vector Machine (SVM) and Random Forest algorithms with hyperparameter tuning in classifying the influence of music on mental health. This study purpose to identify the most superior algorithms and provide academic insights into the application of classifications upon the field of mental health.

RESEARCH METHODS

The study used *the Support Vector Machine* (SVM) and *Random Forest algorithms* to classify the effects of music on mental health. The analysis was conducted based on the type of music, frequency of listening, as well as the individual's mental health condition. The model was evaluated using accuracy, precision, recall, and F1-score metrics, while literature studies supported understanding of music therapy, mental health, and the classification methods used. In this research method, there are a number of stages that need to be implemented. The flow of the research method can be seen in the figure below:

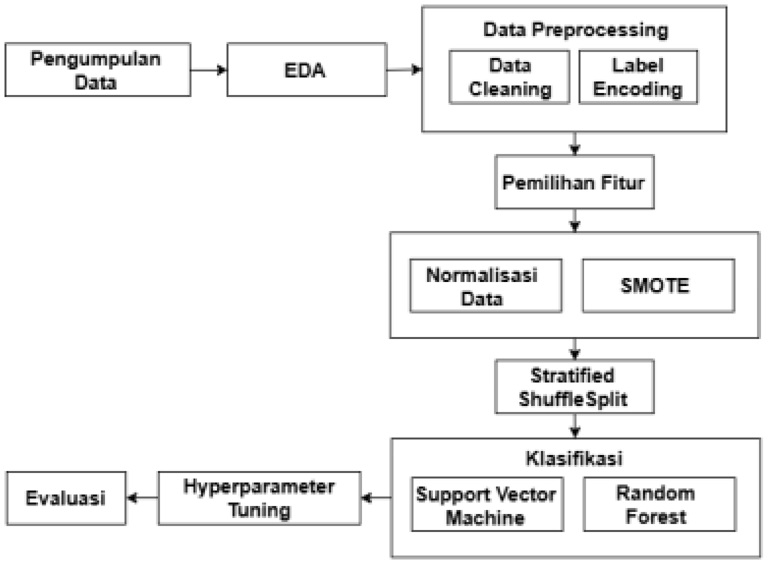

Figure 1. Research Flow

The research process consist of several key stages. First, data collection is conducted through literature review and the use of the mxmh_survey_result dataset from Kagg;e, which contains information abput the impact of music on mental health. Next, Exploratory Data Analysis (EDA) is performed to eamine the data charateristics before preprocessing. During the preprocessing stage, data is cleaned by removing null values and converted using label encoding techniques. After that, feature selection is carried out using the correlation matrix method to determine the relevant variables.

The next sep involves data normalization using MinMaxScaler and the application of SMOTE techniques to address class imbalance. One the data is prepared, the dataset are split for training and also examine using StartifiedShuffleSplit method. The design I sthen built using the Support Vector Machine (SVM) algorithm by different kernel types and Random Forest on specific parameters. To increase model performance, hyperparameter tuning is conduct apply the RandomsizedSearchCV method. Finally, the model is evaluated using a flurry matrix and performance metrics such as accuracy, fidelity, withdrawal, and F1-score to assess its effectiveness.

RESULTS AND DISCUSSION

Data Preprocessing

The preprocessing boards is bring out to spotless and improve the quality of the data that will be used for modeling. The data used comes from a *mxmh_survey_results* dataset in Kaggle, which has gone through *a data cleaning* process by removing *null values* and *encoding labels* to turn categorical data into numerical.

Feature Selection

This correlation analysis aims to find factors that show a strong relationship with the 'Music effects' feature. Figure 2 shows that the correlation of each feature with the target of 'Music effects' does not have a strong correlation. From the results of the correlation analysis, because the 34 features have a low correlation, it was decided to use all features, except 'Timestamp' and 'Permissions'.

```
While working                  0.146556
Exploratory                    0.144517
Anxiety                        0.120924
Frequency [R&B]                0.116698
Fav genre                      0.115240
Frequency [Gospel]             0.095802
Instrumentalist                0.090223
Frequency [Pop]                0.082634
Frequency [K pop]              0.082255
Composer                       0.078434
Frequency [Country]            0.077789
Frequency [Lofi]               0.076136
Frequency [EDM]                0.059854
BPM                            0.059445
Age                            0.059195
OCD                            0.052105
Frequency [Rap]                0.049186
Frequency [Jazz]               0.048376
Primary streaming service      0.042288
Hours per day                  0.041468
Frequency [Hip hop]            0.039574
Frequency [Classical]          0.036409
Frequency [Metal]              0.026720
Frequency [Latin]              0.025410
Foreign languages              0.016577
Frequency [Video game music]   0.016214
Frequency [Rock]               0.010445
Depression                     0.009313
Insomnia                       0.008440
Frequency [Folk]               0.005147
Permissions                         NaN
Name: Music effects, dtype: float64
```

Figure 2. Correlation Analysis

Data Balancing

After seeing the unbalanced distribution of data, the data normalization process was carried out before the SMOTE process. Figure 4 represents the result of a balanced data distribution after the SMOTE process.

```
Music effects
0    465
1    136
2     15
Name: count, dtype: int64
```

Figure 3. Data Sharing Before SMOTE

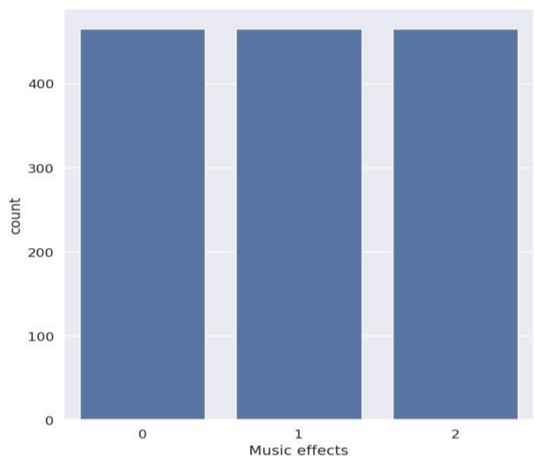

Figure 4. Data Sharing After SMOTE

SVM Model Creation

In this study, the Support *Vector Machine* (SVM) model is created using a library from sklearn.svm with parameters c=10, kernels rbf, linear, poly, sigmoid and class_weight='balanced', then fit X_train and y_train by adding .values.ravel() to adjust the target according to machine learning needs.

Table 1. SVM Evaluation Results

SVM	Accuracy	Precission	Recall	F1-Score
RBF	95%	95%	95%	95%
Linear	76%	75%	75%	75%
Polynomial	89%	90%	89%	89%
Sigmoid	43%	49%	43%	40%

Performance measurements of the SVM model were performed before hyperparameter tuning using the accuracy and *confusion matrix* from the sklearn.metrics library. From table 1, the highest accuracy of the SVM kernel rbf model is 95%, precission 95%, recall 95%, and f1-score 95%. From these results, the SVM kernel rbf model will be hyperparameter tuned and its performance will be compared with the *Random Forest model*.

Creating a Random Forest Model

In figure 5, the creation of the Random Forest specimen utilizes the sklearn.ensemble library with parameters n_estimators=100 and random_state=42, and then fits it with X_train and y_train.

```
RandomForestClassifier
RandomForestClassifier(random_state=42)
```

Figure 5. Development of the Random Forest Model

Hyperparameter Tuning SVM

This study applies hyperparameter tuning to the *Support Vector Machine (SVM) specimen of the rbf kernel using the GridSearchCV technique* from the sklearn.model_selection library.

```
Parameter hyperparameter terbaik: {'C': 10, 'gamma': 1, 'kernel': 'rbf'}
Estimator hyperparameter terbaik: SVC(C=10, gamma=1)
```

Figure 6. SVM's Best Parameters and Estimators

Figure 4.10 shows the best parameter and estimator configuration results during the hyperparameter process on the SVM model. The best parameter was obtained: {'C' : 10, 'gamma':1,'kernel':'rbf'} and the best estimator: SVC(C=10,gamma=1).

Hyperparameter Random Forest

In this process, parameters n_estimators, max, depth, min_samples_split, bootstrap, criterion are used with a predetermined range of values. The parameter n_iter=100 is set to specify many combinations to be evaluated randomly. The cv=5 parameter serves to set many 5-folds on the cross-validation process.

Table 2. Best Hyperparameter Random Forest

Parameters	Best Hyperparameters
n_estimators	200
max_depth	40
min_samples_split	2
min_samples_leaf	1
Bootstrap	False
Criterion	Gini

Table 4.16 is the result of the execution of the *Random Forest hyperparameter process* by obtaining the best combination of hyperparameters: {'n_estimators': 200, 'min_samples_split': 2, 'min_samples_leaf': 1, 'max_depth': 40, 'criterion': 'gini', 'bootstrap': False}.

Evaluation

After performing hyperparameter process on the *Support Vector Machine* and *Random Forest models*, the next stage is to visualize the evaluation by applying a confusion matrix. In Figure 4.11, the confusion matrix visualization on the SVM model is obtained, obtaining the result of 91 positive correct data in class 0 and obtaining the result of predicting 1 incorrect data which should be class 1. Furthermore, the model obtained the results of 86 positive true data of class 1, and 7 false data which should be class 0. Meanwhile, in class 2, the model got positive correct prediction results of 93 data.

```
Evaluasi Hyperparameter Tuning model SVM :
Accuracy: 0.967741935483871
Classification Report:
              precision    recall  f1-score   support

           0       0.93      0.98      0.95        93
           1       0.99      0.92      0.96        93
           2       0.99      1.00      0.99        93

    accuracy                           0.97       279
   macro avg       0.97      0.97      0.97       279
weighted avg       0.97      0.97      0.97       279
```

Figure 7. Classification Report Hyperparameter SVM

```
Evaluasi Model Random Forest dengan Hyperparameter Tuning
Accuracy: 0.9354838709677419
Classification Report:
              precision    recall  f1-score   support

           0       0.86      0.98      0.91        93
           1       0.97      0.83      0.90        93
           2       0.99      1.00      0.99        93

    accuracy                           0.94       279
   macro avg       0.94      0.94      0.93       279
weighted avg       0.94      0.94      0.93       279
```

Figure 8. Classification Report Hyperparameter Random Forest

From figure 4.12, it can be seen that the results of the visualization of the hyperparameter tuning evaluation in Random *Forest* are 91 positive results, predicting class 1 as class 0 by 15 false positive data, predicting class 0 as class 1 by 2. In class 1, the model gets the correct class prediction result of 77 data, estimated as class 0 of 15 data. Meanwhile, in grade 2, the correct prediction results were obtained of 93 data and 1 data was estimated as class 1.

Table 3. Comparison of Methods Before Hyperparameter

Method	Accuracy	Precission	Recall	F1-Score
SVM kernel rbf	95%	95%	95%	95%
SVM linear kernel	76%	75%	75%	75%
SVM polynomial kernel	89%	90%	89%	89%
SVM kernel sigmoid	43%	49%	43%	40%
Random Forest	93%	93%	93%	93%

Table 4. Comparison of Methods After Hyperparameter

Method	Accuracy	Precission	Recall	F1-Score
SVM	97%	97%	97%	97%
Random Forest	94%	94%	94%	93%

In table 3, a comparison of methods before the hyperparameter tuning process is presented, it can be seen that the SVM kernel rbf method is better at processing data, it can be seen from the accuracy, precission, recall, and f1-score that are better than *Random Forest*. Based on the output listed in table 4, that shall be concluded that the *Support Vector Machine* kernel rbf model gets better performance results than the *Random Forest Classifier* model. This is evidenced by the higher accuracy, precission, recall, and f1score values of SVM kernel rbf in the *Random Forest Classifier model*.

CONCLUSIONS AND SUGGESTIONS

Conclusion

After conducting studies and research that discussed the performance analysis of SVM algorithms and *Random Forest* In the case of the influence of music on mental health, this study obtained several results described in the following conclusions:

Based on the results of the research, after the selection of features using 30 features, the two algorithms get an increase in performance. *The Random Forest* algorithm got *93% accuracy*, *93% precision*, *93% recall*, 93% *f1-score* and SVM kernel rbf got *95%* accuracy, *95% precision*, 95% *recall*, 95% *f1-score* from these results, the SVM kernel rbf algorithm was superior in classifying the influence of music on mental health.

The hyperparameter tuning in the SVM kernel rbf algorithm has a better performance than *Random Forest* in classifying the influence of music on mental health with an *accuracy* score of 97% for SVM, while for Random Forest it gets an *accuracy value* of 94%.

This study concludes that the application of hyperparameter tuning to the SVM kerbel rbf and *Random Forest* algorithms gets an improvement in performance in the built model. The SVM kernel rbf algorithm consistently provides better performance results than *Random Forest*, both in implementations with default parameters and after hyperparameter optimization.

Suggestion

Considering that there are still shortcomings in the research carried out, the author provides several suggestions that can serve as a reference in conducting similar studies in the future, as follows:

Evaluation in the study will be better if other evaluation methods, such as AUC-ROC.

Exploring more about different preprocessing techniques, for example the use of dimension subtraction techniques such as PCA (Principal Component Analysis).

To see the results of the study clearly such as informative and interesting visualizations, it is considered to use graphs that represent the comparison of the performance of the models used in addition to using heatmaps for *matrix confusion*.

To get more accurate and representative analysis results, the amount of data needs to be increased.

REFERENCE

Aanda, S., Annisafitri, A., Angelia, M., Augilera, S. C., & Nurdiantami, Y. (2022). A Literature Study of the Influence of Music on Student Mental Health. *Journal of Education and Counseling*, *4*(5), 2580–2588. https://journal.universitaspahlawan.ac.id/index.php/jpdk/article/view/7002

Allyssa, J. P., Malik, R., & Drew, C. (2023). The Influence of Music on Mental Health in Tarumanagara University Students Class of 2020. ... *Health...*, *4*(September), 1889–1896. http://journal.universitaspahlawan.ac.id/index.php/jkt/article/view/16245%0Ahttp://journal.universitaspahlawan.ac.id/index.php/jkt/article/download/16245/13306

Dessy Kusumaningrum, & Imah, E. M. (2020). A Comparative Study of Mental Workload Classification Algorithm Based on EEG Signals. *Journal of Intelligent Systems*, *3*(2), 133–143. https://doi.org/10.37396/jsc.v3i2.69

Fadilah, W. F., Sayekti, S., Sunaryanti, H., Tri, R., School, H., & Mamba's Health Sciences, T. (2024). *The Effect of Music Therapy on Student Mental Health*. *5*(2), 445–452. http://jurnal.globalhealthsciencegroup.com/index.php/JLH

Indra Budi Trisno, M. A. R. (2023). ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING WEBINAR. *Journal of Independent Service*, *2*(11), 2307–2315. https://doi.org/10.31862/9785426311961

Intan Permata, & Esther Sorta Mauli Nababan. (2023). Application Of Game Theory In Determining Optimum Marketing Strategy In Marketplace. *Journal of Research in Mathematics and Natural Sciences*, *2*(2), 65–71. https://doi.org/10.55606/jurrimipa.v2i2.1336

Isnain, A. R., Sakti, A. I., Alita, D., & Marga, N. S. (2021). Public Analysis Sentiment on the Jakarta Government's Lockdown Policy Using the SVM Algorithm. *Journal of Data Mining and Information Systems*, *2*(1), 31. https://doi.org/10.33365/jdmsi.v2i1.1021

Khusna, N. F., Rahmah, A., & Nur, R. K. (2024). *Implementation of Random Forest in Stunting Case Classification in Toddlers with Hyperparameter Tuning Grid Search*. *2024*(Same), 791–801.

Rahayu, K., Fitria, V., Septhya, D., Rahmaddeni, R., & Efrizoni, L. (2023). Text Classification to Detect Depression and Anxiety in Twitter Users Based on Machine Learning. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *3*(2), 108–114. https://doi.org/10.57152/malcom.v3i2.780

Rahman Wahid, M. A., Nugroho, A., & Halim Anshor, A. (2023). Prediction of lung cancer with linear regression algorithm. *Bulletin of Information Technology (BIT)*, *4*(1), 63–74. https://doi.org/10.47065/bit.v4i1.501

Rijal, M., Aziz, F., & Abasa, S. (2024). Depression Prediction: Recent Innovations in Mental Health Journal Pharmacy and Application. *Journal Pharmacy and Application of Computer Sciences*, *2*(1), 9–14. https://doi.org/10.59823/jopacs.v2i1.47

Sebayang, E. R. B., Chrisnanto, Y. H., & Melina. (2023). Mental Health Data Classification in the Technology Industry Using the Random Forest Algorithm. *IJESPG Journal*, *1*(3), 237–253.

Septhya, D., Rahayu, K., Rabbani, S., Fitria, V., Rahmaddeni, R., Irawan, Y., & Hayami, R. (2023). Implementation of Decision Tree Algorithm and Support Vector Machine for Lung Cancer Disease Classification. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *3*(1), 15–19. https://doi.org/10.57152/malcom.v3i1.591

Sub'haan, F., Sinaga, S., & Winangsit, E. (2023). Music Therapy to Improve Mental Health: A Literature Review in Psychodynamic Perspectives. *Assertive: Islamic Counseling Journal*, *02*(1), 1–12. https://ejournal.uinsaizu.ac.id/index.php/assertive/article/view/8017%0Ahttps://ejournal.uinsaizu.ac.id/index.php/asertif/

Titik Misriati, R. A. (2024). *Optimization of Random Forest and Support Vector Machine with Hyperparameter GridSearchCV for PrimaKu Review Sentiment Analysis*. *5*(4), 1342–1351. https://doi.org/10.47065/josh.v5i4.5347

Zai, C. (2022). Implementation of Data Mining as Data Processing. *Journal of Data Portal*, *2*(3), 1–12. http://portaldata.org/index.php/portaldata/article/view/107