



Analisis Hubungan Preferensi Genre Musik dan Kesehatan Mental pada Dataset MXMH

Jafar Jaya Priambudhi*, Imam Suharjo

Teknologi Informasi, Universitas Mercu Buana Yogyakarta, Yogyakarta

Jl Jembatan Merah 84 C, Gejayan, 55281, Yogyakarta, Indonesia

Email: ¹*221110097@student.mercubuana-yogya.ac.id, ²imam@mercubuana-yogya.ac.id

Email Penulis Korespondensi: 221110097@student.mercubuana-yogya.ac.id

Submitted: 09/06/2026; Accepted: 04/07/2026; Published: 05/07/2026

Abstrak—Kesehatan mental merupakan salah satu isu penting yang banyak dipengaruhi oleh berbagai faktor, termasuk kebiasaan mendengarkan musik. Dalam kajian psikologi musik, preferensi genre musik diketahui berkaitan dengan regulasi emosi dan kondisi psikologis individu. Penelitian ini bertujuan untuk menganalisis hubungan antara preferensi genre musik dan kondisi kesehatan mental menggunakan pendekatan berbasis data. Dataset yang digunakan adalah Music & Mental Health Survey (MXMH) yang terdiri dari 737 responden dengan variabel preferensi genre musik, durasi mendengarkan musik, serta indikator kesehatan mental yang meliputi anxiety, depression, insomnia, dan obsessive compulsive disorder (OCD). Tahapan penelitian meliputi data preprocessing, exploratory data analysis (EDA), penentuan jumlah cluster menggunakan Elbow Method dan Silhouette Score, clustering menggunakan algoritma K-Means, analisis hubungan genre musik dan kondisi mental, serta klasifikasi hasil clustering menggunakan Random Forest. Hasil penelitian menunjukkan bahwa responden dapat dikelompokkan ke dalam tiga cluster dengan karakteristik kondisi mental yang berbeda. Nilai Silhouette Score sebesar 0,2246 menunjukkan bahwa kualitas pemisahan cluster masih relatif rendah sehingga hasil segmentasi lebih bersifat eksploratif. Analisis korelasi menunjukkan adanya hubungan positif antara variabel anxiety dan depression, serta perbedaan pola preferensi genre musik pada kelompok dengan kondisi mental yang berbeda. Hasil feature importance menunjukkan bahwa variabel preferensi genre musik berkontribusi dalam membedakan karakteristik masing-masing cluster. Kontribusi penelitian ini adalah memberikan gambaran empiris mengenai hubungan antara preferensi genre musik dan kondisi kesehatan mental berdasarkan dataset MXMH melalui pendekatan segmentasi dan klasifikasi berbasis machine learning. Temuan penelitian ini menunjukkan bahwa preferensi musik berpotensi digunakan sebagai salah satu indikator untuk memahami pola kondisi psikologis individu, meskipun diperlukan validasi dan pengembangan metode lebih lanjut untuk memperoleh segmentasi yang lebih kuat.

Kata Kunci: Kesehatan Mental; Musik; Random Forest; Clustering; Genre Musik

Abstract—Mental health is a critical issue that is influenced by various factors, including music-listening habits. In the field of music psychology, music genre preferences are known to be associated with emotional regulation and an individual's psychological state. This study aims to analyze the relationship between music genre preferences and mental health using a data-driven approach. The dataset used is the Music & Mental Health Survey (MXMH), which consists of 737 respondents with variables including music genre preferences, duration of music listening, and mental health indicators such as anxiety, depression, insomnia, and obsessive-compulsive disorder (OCD). The research stages included data preprocessing, exploratory data analysis (EDA), determining the number of clusters using the Elbow Method and Silhouette Score, clustering using the K-Means algorithm, analyzing the relationship between music genre and mental health, and classifying the clustering results using Random Forest. The results showed that respondents could be grouped into three clusters with distinct mental health characteristics. A Silhouette Score of 0.2246 indicates that the quality of cluster separation is still relatively low, making the segmentation results more exploratory in nature. Correlation analysis revealed a positive relationship between the anxiety and depression variables, as well as differences in music genre preference patterns among groups with different mental health conditions. The feature importance results show that the music genre preference variable contributes to distinguishing the characteristics of each cluster. The contribution of this study is to provide an empirical overview of the relationship between music genre preferences and mental health conditions based on the MXMH dataset through a machine learning-based segmentation and classification approach. The findings of this study suggest that music preferences have the potential to be used as an indicator for understanding patterns of an individual's psychological state, although further validation and methodological development are needed to achieve a more robust segmentation.

Keywords: Mental Health; Music; Random Forest; Clustering; Music Genre.

1. PENDAHULUAN

Kesehatan mental menjadi salah satu isu penting di era modern, khususnya pada kelompok pelajar dan usia produktif yang rentan terhadap tekanan akademik, sosial, dan perubahan gaya hidup [1]. Berbagai kondisi seperti kecemasan (anxiety), depresi (depression), insomnia, dan gangguan obsesif kompulsif (obsessive compulsive disorder / OCD) dapat memengaruhi kualitas hidup, produktivitas, serta kemampuan individu dalam menjalankan aktivitas sehari-hari [2], [3]. Oleh karena itu, diperlukan pendekatan yang dapat membantu memahami faktor-faktor yang berkaitan dengan kondisi psikologis individu secara lebih komprehensif.

Salah satu faktor yang diketahui memiliki keterkaitan dengan kondisi psikologis adalah musik [4]. Dalam kajian psikologi musik (music psychology) dan musikologi, musik dipandang sebagai media regulasi emosi yang mampu memengaruhi suasana hati, tingkat stres, serta respons afektif seseorang. Karakteristik musik yang tersusun atas unsur melodi, ritme, harmoni, tempo, dan timbre membentuk berbagai genre dengan karakter emosional yang berbeda [5]. Genre musik dengan tempo lambat dan nuansa tenang cenderung digunakan untuk relaksasi dan pengurangan stres, sedangkan genre dengan energi tinggi dapat memberikan stimulasi emosional yang berbeda

[6]. Perbedaan respons tersebut menunjukkan bahwa preferensi seseorang terhadap genre musik tertentu berpotensi mencerminkan kondisi emosional dan psikologis yang dimilikinya.

Perkembangan layanan musik digital seperti Spotify menghasilkan data yang dapat digunakan untuk menganalisis perilaku mendengarkan musik secara lebih objektif [7]. Selain menyediakan fitur audio seperti valence, energy, dan danceability, platform digital juga merekam preferensi genre serta kebiasaan mendengarkan musik pengguna. Informasi tersebut membuka peluang untuk mengkaji hubungan antara pola konsumsi musik dan kondisi kesehatan mental menggunakan pendekatan berbasis data (data-driven approach).

Penelitian sebelumnya telah memanfaatkan metode machine learning untuk menganalisis data musik maupun kesehatan mental. Penelitian oleh Ipmawati [8] menggunakan algoritma K-Means dan Random Forest untuk menganalisis hubungan antara kebiasaan mendengarkan musik dan kondisi kesehatan mental, menghasilkan empat kluster utama dengan nilai Silhouette Score sebesar 0,24 dan akurasi klasifikasi sebesar 86%. [9] menunjukkan bahwa Random Forest memiliki performa yang lebih baik dibandingkan Support Vector Machine (SVM) dalam analisis pengaruh musik terhadap tingkat stres mahasiswa dengan akurasi masing-masing sebesar 85% dan 72%. Pada domain kesehatan mental, Faisti et al.[10] melaporkan bahwa Random Forest mampu menghasilkan akurasi sebesar 89,3% dalam proses klasifikasi kondisi mental.

Pada bidang analisis musik, metode K-Means telah digunakan untuk mengelompokkan lagu berdasarkan karakteristik audio dan suasana hati. Nuriska et al. [11] berhasil membentuk tiga kelompok suasana hati lagu Spotify dengan nilai Davies-Bouldin Index sebesar 0,299. Penelitian lain menunjukkan bahwa algoritma machine learning mampu mengklasifikasikan genre musik dengan akurasi mencapai 83% [12]. Selain itu, K-Means juga telah diterapkan untuk segmentasi data musik [13], [14] dan pengembangan sistem rekomendasi musik [15], sedangkan Random Forest digunakan untuk prediksi tren musik, analisis perilaku pengguna Spotify, serta klasifikasi emosi musik berdasarkan fitur audio [14], [16], [17], [18].

Meskipun berbagai penelitian telah menunjukkan keberhasilan penerapan machine learning pada domain musik dan kesehatan mental, sebagian besar penelitian masih berfokus pada analisis karakteristik lagu, fitur audio, sistem rekomendasi musik, atau klasifikasi kesehatan mental secara terpisah. Penelitian yang menjadikan preferensi genre musik sebagai representasi perilaku mendengarkan (listening behavior) dan mengaitkannya secara langsung dengan kondisi kesehatan mental individu masih relatif terbatas. Selain itu, sebagian besar penelitian hanya menggunakan pendekatan supervised learning atau unsupervised learning secara terpisah sehingga belum memberikan gambaran mengenai karakteristik kelompok pendengar sekaligus faktor-faktor yang membedakan kelompok tersebut.

Berdasarkan kondisi tersebut, penelitian ini menggunakan algoritma K-Means untuk mengelompokkan responden berdasarkan pola preferensi genre musik dan Random Forest untuk mengidentifikasi karakteristik serta variabel yang berkontribusi terhadap pembentukan kluster. Kontribusi utama penelitian ini adalah memberikan analisis empiris mengenai hubungan antara preferensi genre musik dan kondisi kesehatan mental menggunakan dataset Music & Mental Health Survey (MXMH), serta menunjukkan bagaimana pendekatan segmentasi dan klasifikasi dapat digunakan secara komplementer untuk memahami karakteristik kelompok pendengar musik. Hasil penelitian diharapkan dapat memperkaya kajian pada bidang analisis perilaku pengguna berbasis data musik dan memberikan referensi bagi penelitian lanjutan yang mengkaji hubungan antara pola konsumsi musik dan kesehatan mental.

2. METODOLOGI PENELITIAN

2.1 Jenis dan Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif berbasis data (data-driven) untuk menganalisis hubungan antara preferensi genre musik dan kondisi kesehatan mental. Analisis dilakukan melalui dua tahapan utama, yaitu segmentasi data menggunakan algoritma K-Means dan klasifikasi hasil segmentasi menggunakan algoritma Random Forest. Pendekatan ini dipilih karena memungkinkan identifikasi pola kelompok pendengar musik berdasarkan karakteristik preferensi genre serta evaluasi kemampuan model dalam mengenali karakteristik setiap kelompok yang terbentuk [16].

2.2 Dataset dan Variabel Penelitian

Dataset yang digunakan dalam penelitian ini adalah Music & Mental Health Survey (MXMH) yang diperoleh dari sumber dataset publik. Dataset ini berjumlah 737 responden yang berisi informasi terkait kebiasaan mendengarkan musik dan kondisi kesehatan mental. Sampel data pada dataset dapat dilihat di Tabel 1

Tabel 1. Sampel Data

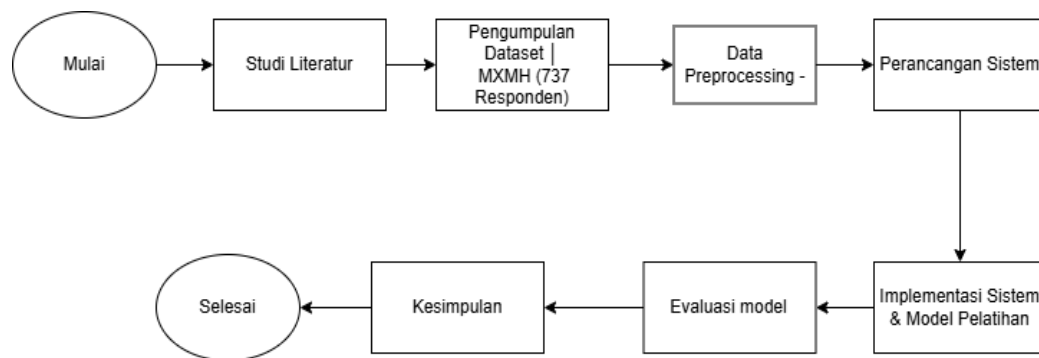
Age	Fav Genre	Hours per day	Anxiety	Depression	Insomnia	Ocd
18	Latin	3.0	3	0	2	0
63	Rock	1.5	7	2	2	1

Age	Fav Genre	Hours per day	Anxiety	Depression	Insomnia	Ocd
21	Kpop	1.0	5	3	5	3
61	Jazz	2.5	9	7	3	3
19	Rock	1.0	2	2	5	1

Variabel yang digunakan dalam penelitian ini meliputi preferensi genre musik, frekuensi mendengarkan musik per genre, durasi mendengarkan musik per hari (hours per day), serta kondisi kesehatan mental yang terdiri dari depression, anxiety, insomnia, dan obsessive compulsive disorder (OCD). Variabel kondisi mental berupa data numerik yang merepresentasikan tingkat kondisi masing-masing responden.

2.3 Tahapan Penelitian

Penelitian dilakukan melalui beberapa tahapan, yaitu pengumpulan dataset, data preprocessing, exploratory data analysis (EDA), penentuan jumlah cluster optimal menggunakan Elbow Method dan Silhouette Score, clustering menggunakan K-Means, analisis hasil clustering, analisis hubungan antara genre musik dan kondisi mental, klasifikasi menggunakan Random Forest, evaluasi model, analisis feature importance, serta visualisasi dan penyusunan laporan penelitian. Alur penelitian ditunjukkan pada gambar 1



Gambar 1. Tahapan Penelitian

2.4 Data Preprocessing

Tahap preprocessing dilakukan untuk memastikan kualitas data sebelum dianalisis. Proses ini meliputi penghapusan atribut yang tidak relevan seperti timestamp dan penanganan data yang memiliki nilai kosong (missing values) [19]. Selanjutnya, variabel kategorikal seperti Primary Streaming Service, While Working, Instrumentalist, Composer, Fav Genre, Exploratory, Foreign Languages, dan Music Effects ditransformasikan menggunakan metode One-Hot Encoding. Metode ini dipilih karena mampu merepresentasikan data kategorikal nominal tanpa menghasilkan hubungan ordinal semu yang dapat memengaruhi hasil analisis. Selain itu, data frekuensi mendengarkan musik dikonversi ke bentuk numerik dengan skala Never = 0, Rarely = 1, Sometimes = 2, dan Very Frequently = 3.

2.5 Exploratory Data Analysis (EDA)

EDA dilakukan untuk memahami pola awal dalam dataset[16]. Analisis yang dilakukan meliputi distribusi preferensi genre musik berdasarkan rata-rata frekuensi mendengarkan, serta analisis korelasi antar variabel kondisi mental (depression, anxiety, insomnia, dan OCD). Korelasi divisualisasikan menggunakan heatmap untuk melihat hubungan antar variabel.

2.6 Penentuan Jumlah Cluster

Sebelum proses clustering dilakukan, jumlah cluster optimal ditentukan menggunakan Elbow Method dan Silhouette Score. Elbow Method digunakan untuk mengamati perubahan nilai inertia terhadap jumlah cluster, sedangkan Silhouette Score digunakan untuk mengukur kualitas pemisahan antar cluster. Jumlah cluster yang menghasilkan nilai Silhouette Score tertinggi dipilih sebagai jumlah cluster optimal [19].

2.7 Clustering dengan K-Means

Sebelum proses clustering, data dinormalisasi menggunakan StandardScaler agar setiap fitur memiliki skala yang sama [14], [20]. Proses clustering dilakukan menggunakan algoritma K-Means dengan jumlah cluster yang diperoleh dari hasil evaluasi Elbow Method dan Silhouette Score

Tabel 2. Variabel Data

No	Variabel		
	Nama	Tipe	Keterangan
1	OCD	Numerik	Tingkat Kondisi OCD

No	Variabel		
	Nama	Tipe	Keterangan
2	Depression	Numerik	Tingkat Kondisi Depresi
3	Anxiety	Numerik	Tingkat Kondisi Kecemasan
4	Insomnia	Numerik	Tingkat Kondisi Insomnia
5	Frequency [Genre]	Kategorikal	Frekuensi Mendengar Genre

Pada Tabel 2 Fitur yang digunakan dalam proses clustering adalah frekuensi genre musik, yaitu Pop, Rock, Hip hop, Lofi, R&B, dan EDM. Hasil clustering kemudian dianalisis berdasarkan rata-rata kondisi mental pada setiap cluster untuk mengetahui karakteristik masing-masing kelompok responden.

2.8 Analisis Hubungan Genre dan Kondisi Mental

Analisis dilakukan untuk mengetahui hubungan antara frekuensi genre musik dan kondisi mental menggunakan korelasi Pearson yang divisualisasikan dalam bentuk heatmap [15]. Nilai korelasi positif menunjukkan bahwa genre tertentu lebih sering didengarkan pada kondisi mental tinggi, sedangkan nilai korelasi negatif menunjukkan hubungan yang berlawanan. Selain itu, dilakukan perbandingan antara kelompok responden dengan kondisi mental tinggi (nilai > 5) dan rendah (nilai ≤ 5) untuk mengidentifikasi perbedaan pola preferensi genre musik.

2.9 Klasifikasi Menggunakan Random Forest

Random Forest digunakan untuk mengklasifikasikan hasil clustering ke dalam beberapa kelas cluster. Dataset dibagi menjadi data pelatihan (training set) sebesar 80% dan data pengujian (testing set) sebesar 20% menggunakan metode train-test split. Random Forest merupakan metode ensemble learning yang terdiri dari banyak decision tree. Setiap pohon dibangun menggunakan subset data dan fitur yang dipilih secara acak sehingga mampu meningkatkan akurasi prediksi dan mengurangi risiko overfitting [16].

$$\hat{y} = \text{mode}(h_1(x), h_2(x), \dots, h_n(x)) \quad (1)$$

i mana $\hat{y}$ adalah hasil prediksi akhir, $h_i(x)$ adalah prediksi dari pohon ke-i, dan n adalah jumlah pohon ($n_{\text{estimators}}$).

2.10 Evaluasi Model

Evaluasi model dilakukan menggunakan beberapa metrik, yaitu accuracy, precision, recall, dan F1-score [14]. Accuracy digunakan untuk mengukur tingkat ketepatan prediksi secara keseluruhan:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

Precision dan recall masing-masing digunakan untuk mengukur ketepatan dan kelengkapan prediksi positif:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (4)$$

Sedangkan F1-score merupakan rata-rata harmonis antara precision dan recall:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

Selain itu, digunakan confusion matrix untuk melihat distribusi hasil prediksi terhadap data aktual.

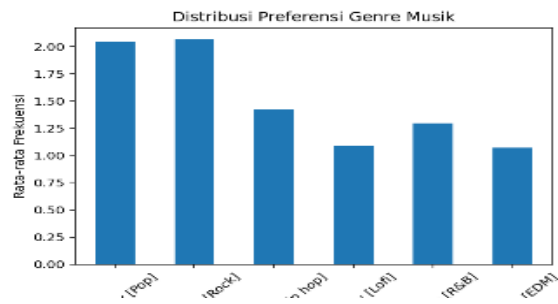
2.10 Feature Importance dan Visualisasi

Random Forest juga digunakan untuk menghitung feature importance guna mengetahui kontribusi masing-masing variabel dalam model. Hasil analisis ditampilkan dalam bentuk visualisasi seperti grafik distribusi genre, heatmap korelasi, grafik cluster terhadap kondisi mental, perbandingan preferensi genre, confusion matrix, serta grafik feature importance. Seluruh hasil analisis disusun dalam bentuk laporan berbasis HTML untuk memudahkan interpretasi dan penyajian hasil penelitian.

3. HASIL DAN PEMBAHASAN

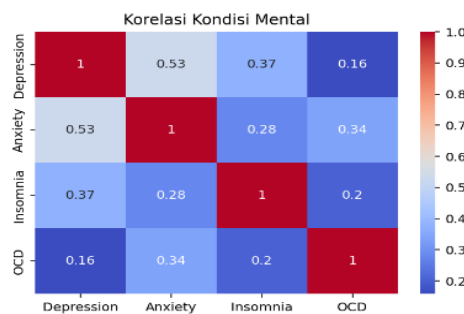
3.1 Hasil Analisis Data

Hasil Exploratory Data Analysis (EDA) menunjukkan bahwa distribusi preferensi genre musik memiliki variasi yang cukup signifikan. Berdasarkan visualisasi Gambar 2 rata-rata frekuensi mendengarkan, beberapa genre seperti Pop, Rock, Hip hop, Lofi, R&B, dan EDM menunjukkan tingkat konsumsi yang berbeda antar responden. Hal ini mengindikasikan adanya kecenderungan preferensi musik tertentu yang dapat menjadi indikator awal dalam analisis kondisi kesehatan mental.



Gambar 2. Distribusi Genre Musik

Selanjutnya dilakukan analisis korelasi terhadap variabel kesehatan mental yang terdiri dari Anxiety, Depression, Insomnia, dan OCD. Hasil visualisasi pada Gambar 3 menunjukkan adanya hubungan positif antara variabel Anxiety dan Depression. Temuan ini menunjukkan bahwa responden dengan tingkat kecemasan yang lebih tinggi cenderung memiliki tingkat depresi yang lebih tinggi pula. Variabel Insomnia juga menunjukkan korelasi positif terhadap Anxiety dan Depression, meskipun dengan kekuatan hubungan yang lebih rendah. Sementara itu, variabel OCD menunjukkan hubungan yang relatif lebih lemah dibandingkan variabel lainnya.

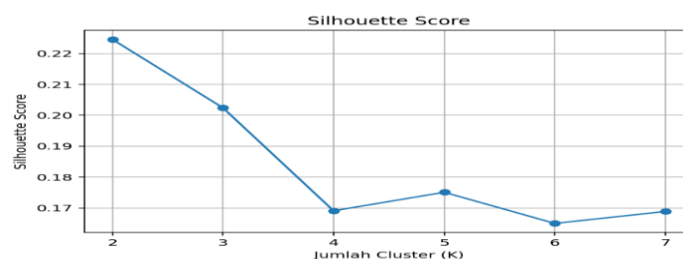


Gambar 3. Korelasi atau Hubungan Kondisi Mental

Hasil EDA menunjukkan bahwa data memiliki variasi yang cukup untuk dilakukan proses segmentasi menggunakan algoritma clustering serta analisis lanjutan menggunakan metode klasifikasi.

3.2 Hasil Clustering Menggunakan K-Means

Sebelum proses clustering dilakukan, seluruh fitur frekuensi genre dinormalisasi menggunakan StandardScaler untuk menghilangkan pengaruh perbedaan skala antar variabel [20]. Penentuan jumlah cluster dilakukan menggunakan Silhouette Score sebagai metode evaluasi kualitas clustering. Hasil evaluasi menunjukkan bahwa nilai Silhouette Score tertinggi diperoleh pada K=2 sebesar 0,2246. Namun, dalam penelitian ini digunakan K=3 karena pertimbangan interpretabilitas hasil dan kebutuhan analisis yang lebih rinci terhadap karakteristik responden. Penggunaan tiga cluster memungkinkan identifikasi variasi pola preferensi musik dan kondisi kesehatan mental yang lebih beragam dibandingkan segmentasi dua cluster yang cenderung menghasilkan kelompok yang terlalu umum. Selain itu, segmentasi menjadi tiga kelompok memberikan kemudahan dalam membedakan karakteristik responden berdasarkan tingkat kondisi mental yang relatif berbeda sehingga mendukung proses analisis hubungan antara preferensi genre musik dan kesehatan mental. Selain itu, penggunaan tiga cluster memberikan segmentasi yang lebih representatif terhadap keragaman pola preferensi musik responden. Pendekatan ini memungkinkan identifikasi kelompok dengan karakteristik yang lebih spesifik dibandingkan penggunaan dua cluster. Dengan demikian, analisis hubungan antara preferensi genre musik dan kondisi kesehatan mental dapat dilakukan secara lebih rinci. Hasil segmentasi tersebut selanjutnya digunakan sebagai dasar untuk mengeksplorasi karakteristik masing-masing kelompok responden.



Gambar 3. Elbow Method

Gambar 3 Menunjukkan Hasil clustering dengan K=3 menghasilkan nilai Silhouette Score sebesar 0,2022. Nilai tersebut menunjukkan bahwa kualitas pemisahan cluster masih relatif rendah, sehingga masih terdapat kemiripan karakteristik antar cluster. Oleh karena itu, hasil clustering pada penelitian ini digunakan sebagai pendekatan eksploratif untuk mengidentifikasi pola dalam data, bukan sebagai segmentasi yang bersifat definitif. Meskipun demikian, hasil pengelompokan tetap mampu menunjukkan adanya variasi karakteristik kondisi mental pada setiap cluster sebagaimana ditunjukkan pada Tabel .

Tabel 3. Karakteristik Kluster

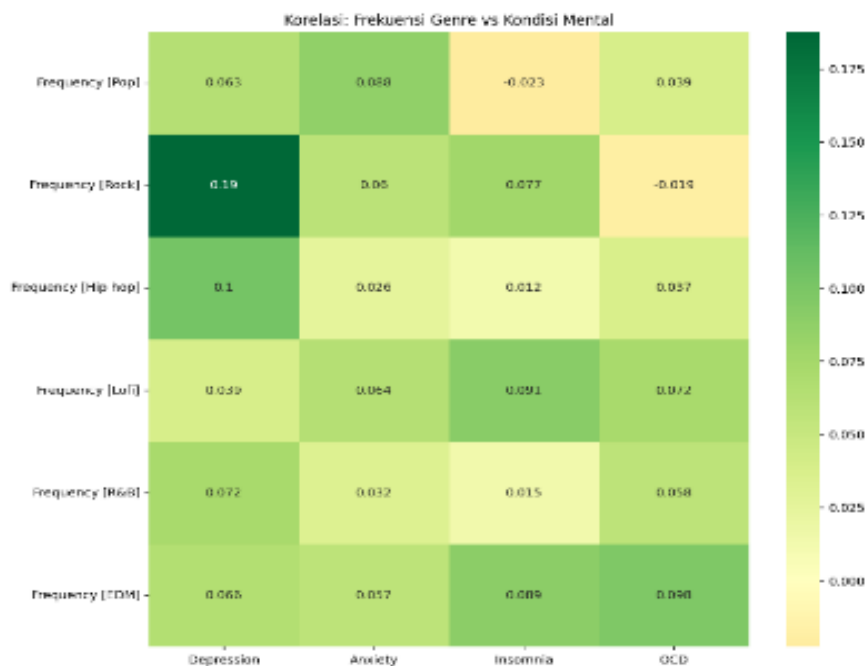
Cluster	Anxiety	Depression	Insomnia	OCD
0	6.00	5.30	3.95	2.80
1	5.75	4.90	3.85	2.40
2	5.95	4.00	3.35	2.80

Berdasarkan Tabel 3, terlihat adanya variasi karakteristik kesehatan mental pada masing-masing cluster. Cluster 0 cenderung memiliki tingkat kecemasan, depresi, insomnia, dan OCD yang lebih tinggi dibandingkan cluster lainnya. Sebaliknya, Cluster 1 menunjukkan nilai yang relatif lebih rendah dan lebih merata pada seluruh indikator kesehatan mental. Sementara itu, Cluster 2 memiliki tingkat kecemasan yang relatif tinggi, namun disertai tingkat depresi dan insomnia yang lebih rendah dibandingkan Cluster 0. Temuan ini menunjukkan bahwa segmentasi berdasarkan preferensi genre musik mampu membedakan kelompok responden dengan karakteristik psikologis yang berbeda.

Meskipun demikian, perbedaan antar cluster pada penelitian ini hanya diinterpretasikan secara deskriptif. Nilai Silhouette Score sebesar 0,2022 menunjukkan bahwa pemisahan antar cluster masih relatif lemah sehingga masih terdapat kemiripan karakteristik antar kelompok. Oleh karena itu, hasil clustering lebih tepat digunakan sebagai temuan eksploratif untuk mengidentifikasi pola awal hubungan antara preferensi musik dan kondisi kesehatan mental

3.3 Analisis Hubungan Genre Musik dan Kondisi Mental

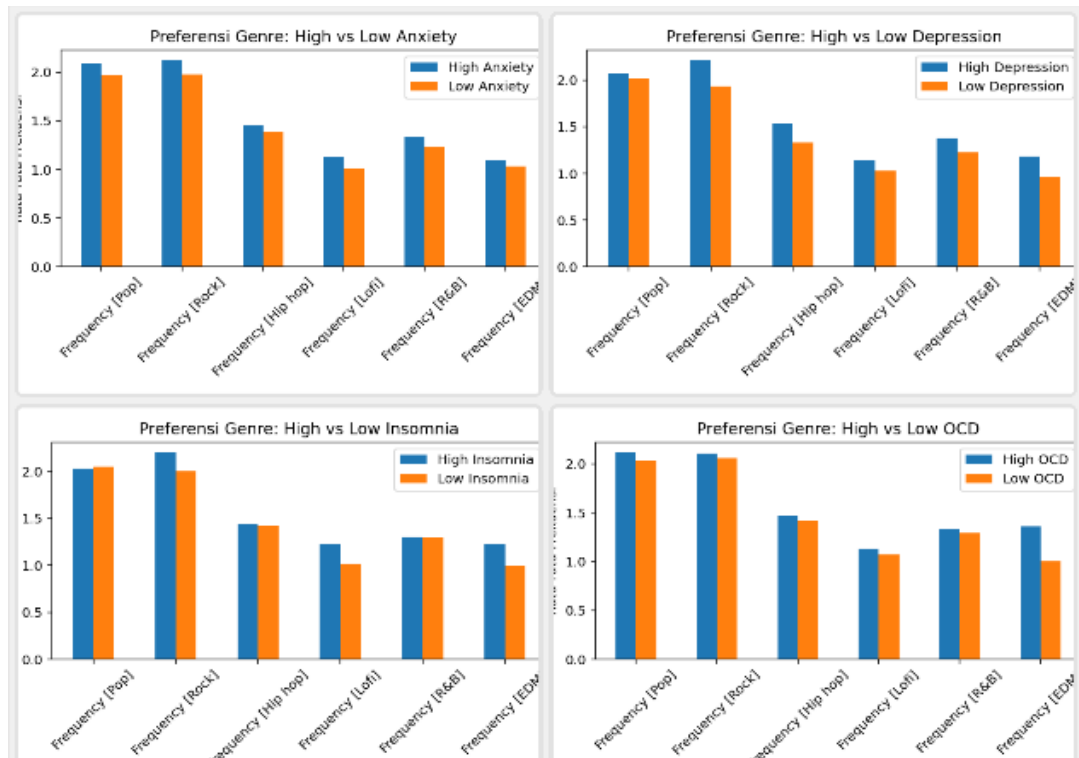
Analisis korelasi dilakukan untuk mengetahui hubungan antara frekuensi mendengarkan genre musik dengan kondisi kesehatan mental responden. Hasil korelasi divisualisasikan menggunakan heatmap sebagaimana ditunjukkan pada Gambar 4.



Gambar 4. Hubungan Genre dan Kondisi Mental

Hasil analisis menunjukkan bahwa beberapa genre memiliki hubungan positif maupun negatif terhadap variabel kesehatan mental. Nilai korelasi yang diperoleh secara umum berada pada kategori rendah hingga sedang, sehingga menunjukkan bahwa hubungan antara preferensi genre musik dan kondisi mental bersifat kompleks dan dipengaruhi oleh berbagai faktor lain di luar variabel penelitian.

Selain analisis korelasi, dilakukan pula perbandingan preferensi genre antara kelompok responden dengan skor kondisi mental tinggi (>5) dan rendah (≤5).



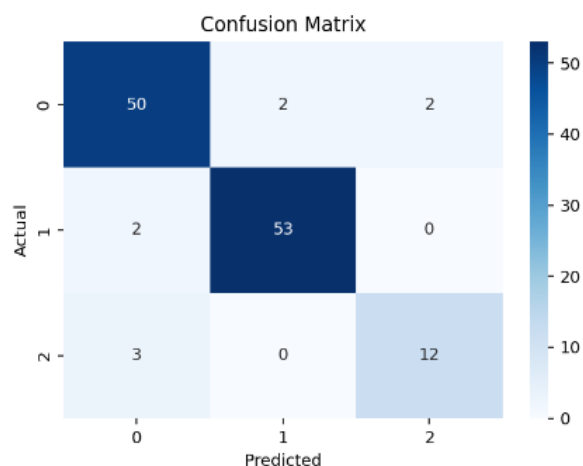
Gambar 5. Preferensi Genre Tiap Kondisi Mental

Hasil visualisasi Gambar 5 menunjukkan adanya perbedaan pola konsumsi musik pada masing-masing kelompok. Beberapa genre seperti Rock, Lofi, dan R&B cenderung memiliki frekuensi mendengarkan yang lebih tinggi pada kelompok dengan skor Anxiety dan Depression yang lebih tinggi. Sementara itu, genre Pop menunjukkan distribusi yang relatif merata pada berbagai kelompok kondisi mental.

Temuan ini menunjukkan bahwa preferensi musik dapat digunakan sebagai salah satu indikator perilaku yang berkaitan dengan kondisi psikologis individu, meskipun tidak dapat digunakan sebagai indikator tunggal untuk menentukan kondisi kesehatan mental seseorang.

3.4 Hasil Klasifikasi Menggunakan Random Forest

Setelah proses clustering selesai dilakukan, label cluster digunakan sebagai target klasifikasi pada algoritma Random Forest. Dataset dibagi menjadi data pelatihan (80%) dan data pengujian (20%) menggunakan metode train-test split untuk mengukur kemampuan generalisasi model. Hasil evaluasi menunjukkan bahwa model Random Forest menghasilkan nilai akurasi sebesar 93% pada data pengujian.



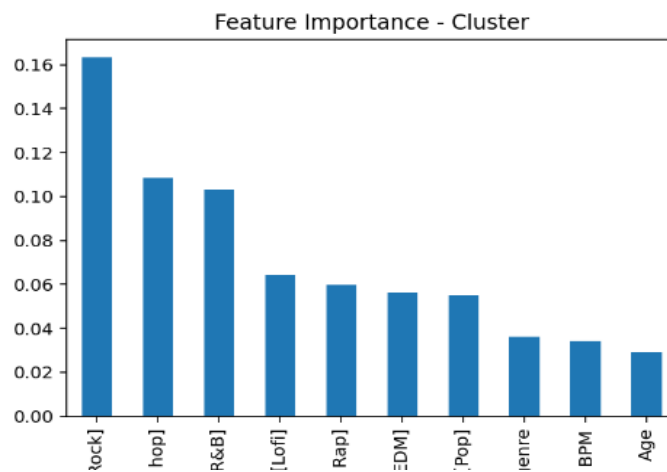
Gambar 6. Confusion Matrix

Berdasarkan Gambar 6, sebagian besar data berhasil diklasifikasikan ke dalam cluster yang sesuai. Kesalahan klasifikasi masih ditemukan pada beberapa data yang berada di area perbatasan antar cluster, yang mengindikasikan adanya kemiripan karakteristik pada sebagian responden. Tabel 4 menunjukkan hasil evaluasi model menggunakan metrik Precision, Recall, dan F1-Score.

Tabel 4. Random Forest

Kelas	Precision	Recall	F1-Score	Support
0	0.91	0.93	0.92	54
1	0.96	0.96	0.96	55
2	0.86	0.80	0.83	15
Accuracy			0.93	124
Macro Avg	0.91	0.90	0.90	124
Weighted Avg	0.93	0.93	0.93	124

Hasil tersebut menunjukkan bahwa model memiliki kemampuan yang baik dalam mengenali pola yang membedakan masing-masing cluster berdasarkan fitur yang digunakan. Selain proses klasifikasi, Random Forest juga digunakan untuk menghitung feature importance guna mengetahui variabel yang paling berkontribusi terhadap pembentukan model.


Gambar 7. Feature Importance

Hasil feature importance pada Gambar 7 menunjukkan bahwa beberapa variabel frekuensi genre musik memiliki kontribusi yang lebih besar dibandingkan variabel lainnya dalam membedakan karakteristik cluster. Temuan ini mengindikasikan bahwa pola preferensi musik memiliki peran penting dalam proses segmentasi responden yang dilakukan pada penelitian ini.

3.5 Pembahasan

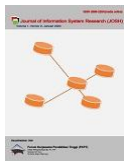
Hasil penelitian menunjukkan bahwa preferensi genre musik memiliki keterkaitan dengan kondisi kesehatan mental responden. Proses clustering menggunakan K-Means menghasilkan tiga klaster dengan karakteristik kesehatan mental yang berbeda. Meskipun nilai Silhouette Score sebesar 0,2022 menunjukkan kualitas pemisahan klaster yang relatif rendah, hasil segmentasi tetap mampu memberikan gambaran awal mengenai variasi pola preferensi musik dan kondisi psikologis responden.

Hasil penelitian ini konsisten dengan penelitian Ipmawati [8] yang menggunakan kombinasi K-Means dan Random Forest dalam analisis hubungan antara kebiasaan mendengarkan musik dan kesehatan mental. Selain itu, hasil korelasi menunjukkan adanya hubungan positif antara Anxiety dan Depression, sedangkan beberapa genre seperti Rock, Lofi, dan R&B cenderung lebih dominan pada kelompok dengan tingkat kecemasan dan depresi yang lebih tinggi. Temuan tersebut mendukung kajian psikologi musik yang menyatakan bahwa preferensi musik berkaitan dengan regulasi emosi dan kondisi psikologis individu [4], [5]

Hasil klasifikasi menunjukkan bahwa model mampu mengenali pola pada data yang digunakan. Namun, interpretasi performa model perlu dilakukan secara hati-hati karena label target berasal dari hasil clustering yang memiliki kualitas pemisahan relatif rendah. Kontribusi penelitian ini adalah memberikan analisis eksploratif mengenai hubungan antara preferensi genre musik dan kondisi kesehatan mental menggunakan kombinasi metode K-Means dan Random Forest pada dataset MXMH. Selain itu, hasil penelitian menunjukkan bahwa preferensi genre musik berpotensi digunakan sebagai indikator pendukung dalam memahami karakteristik psikologis individu. Penelitian selanjutnya disarankan menerapkan validasi model yang lebih komprehensif, pengujian signifikansi statistik antar klaster, serta penambahan fitur audio musik untuk meningkatkan kualitas analisis

4. KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma K-Means mampu mengelompokkan responden ke dalam tiga klaster berdasarkan preferensi genre musik dengan karakteristik kondisi kesehatan mental yang berbeda, sementara analisis menggunakan Random Forest dan feature importance menunjukkan bahwa preferensi musik dan variabel



kesehatan mental memiliki kontribusi dalam membedakan karakteristik setiap kluster. Kontribusi penelitian ini adalah memberikan gambaran empiris mengenai hubungan antara preferensi genre musik dan kondisi kesehatan mental menggunakan dataset Music & Mental Health Survey (MXMH) melalui pendekatan segmentasi dan klasifikasi berbasis machine learning, meskipun nilai Silhouette Score sebesar 0,2022 mengindikasikan bahwa kualitas pemisahan kluster masih relatif rendah sehingga hasil penelitian perlu diinterpretasikan secara hati-hati; oleh karena itu, penelitian selanjutnya disarankan menerapkan proses data cleaning yang lebih ketat, validasi model yang lebih komprehensif seperti k-fold cross validation, pengujian signifikansi statistik antar kluster, serta penambahan fitur audio musik seperti valence, energy, danceability, dan tempo untuk meningkatkan kualitas analisis dan interpretasi hubungan antara preferensi musik dan kesehatan mental.

UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih kepada Universitas Mercu Buana Yogyakarta, khususnya Fakultas Teknologi Informasi, atas dukungan akademik dan fasilitas yang diberikan selama pelaksanaan penelitian ini. Penulis juga menyampaikan apresiasi kepada penyedia dataset Music & Mental Health Survey (MXMH) yang digunakan sebagai sumber data dalam penelitian ini. Selain itu, ucapan terima kasih disampaikan kepada seluruh pihak yang telah memberikan bantuan, masukan, dan dukungan sehingga penelitian dan penyusunan artikel ini dapat diselesaikan dengan baik.

REFERENCES

- [1] E. Jurnal Penelitian Dan Pengabdian, R. Surya, F. Reni, R. Bakhtiar, and A. Dzaky Zaidan, “Musik Sebagai Media Komunikasi Bagi Kesehatan Mental,” *Ekasakti Jurnal Penelitian Dan Pengabdian*, vol. 5, no. 2, pp. 2025–2025, 2025, doi: <https://doi.org/10.31933/ejpp.v5i2.1317>.
- [2] S. Astuti et al., “Lagu sebagai Cermin Psikologis Generasi Muda: Studi Kesehatan Mental dalam Karya Musik Hindia,” *Jurnal Pengabdian Masyarakat dan Riset Pendidikan*, vol. 3, no. 4, pp. 2891–2895, May 2025, doi: [10.31004/jerkin.v3i4.800](https://doi.org/10.31004/jerkin.v3i4.800).
- [3] T. M. Z. Zahrani, Ulin Nihayah, and Ayu Faiza Algifahmy, “Upaya Menumbuhkan Kesehatan Mental Melalui Music Therapy Dalam Mengatasi Kecemasan,” *G-Couns: Jurnal Bimbingan dan Konseling*, vol. 8, no. 3, pp. 1884–1892, Jun. 2024, doi: [10.31316/gcouns.v8i3.5556](https://doi.org/10.31316/gcouns.v8i3.5556).
- [4] H. Jurnal et al., “Analisis Python Penggunaan Musik Sebagai Pengobatan Gangguan Mental,” *JUPTI*, vol. 2, no. 2, pp. 38–47, May 2023, doi: [10.55606/jupti.v2i2.1721](https://doi.org/10.55606/jupti.v2i2.1721).
- [5] Aisyah Khamini and Tri Setyo Andri Maulana, “Peran Musik dalam Regulasi Emosi dan Perubahan Mood : Kajian Literatur,” *JOURNAL SAINS STUDENT RESEARCH*, vol. 4, no. 1, pp. 968–972, Feb. 2026, doi: [10.61722/jssr.v4i1.8911](https://doi.org/10.61722/jssr.v4i1.8911).
- [6] N. Zulfatus Izza and V. Rega Natalia, “SWARA: Jurnal Antologi Pendidikan Musik Peran Musik Untuk Mengatasi Stres dan Krisis Mental Gen Z di Era Digital,” *SWARA: Jurnal Antologi Pendidikan Musik*, vol. 5, no. 1, pp. 113–124, Apr. 2025, doi: <https://doi.org/10.17509/swara.v5i1.84569>.
- [7] Muhammad Aulia Akbar and Anne Maryani, “Pengaruh Terpaan Lagu Hindia Terhadap Kesehatan Mental Gen Z,” *Bundang Conference Series: Communication Management*, vol. 5, no. 2, Jul. 2025, doi: [10.29313/bcsdm.v5i2.18694](https://doi.org/10.29313/bcsdm.v5i2.18694).
- [8] J. Ipawati, “Segmentasi Pendengar Musik Berdasarkan Kebiasaan dan Dampaknya Terhadap Kesehatan Mental Menggunakan K-Means dan Random Forest,” *JUTIK : Jurnal Teknologi Informasi dan Komputer*, vol. 11, no. 1, pp. 34–44, Apr. 2025, doi: [10.36002/jutik.v11i1.3748](https://doi.org/10.36002/jutik.v11i1.3748).
- [9] D. Muriyatmoko, A. Musthafa, M. Fa-Idzaa, U. Darussalam Gontor, and J. Timur Indonesia, “Seminar Nasional Amikom Surakarta (Semnasa) 2024 Perbandingan Metode Support Vector Machine dan Random Forest Dalam Menganalisis Pengaruh Musik Terhadap Penurunan Tingkat Stress Mahasiswa Semester 7 Saat Skripsi (Studi Kasus : Universitas Darussalam Gontor),” *Seminar Nasional Amikom Surakarta (Semnasa)*, vol. 2, pp. 128–135, Nov. 2024, Accessed: Jun. 17, 2026. [Online]. Available: <https://ojs.amikomsolo.ac.id/index.php/semnasa/article/view/533>
- [10] M. J. Faisti, R. Hadapiningradja Kusumodestoni, G. Wahyu, and N. Wibowo, “Mental Health Classification Using Naïve Bayes and Random Forest Algorithms,” Aug. 2025. doi: <https://doi.org/10.30871/jaic.v9i4.10144>.
- [11] D. Nuriska, B. Irawan, A. Bahtiar, and A. Rinaldi Dikananda, “Klasterisasi Data Lagu Terpopuler Spotify 2023 Berdasarkan Suasana Hati Menggunakan Algoritma K-Means,” *Jurnal Mahasiswa Teknik Informatika*, vol. 7, no. 6, pp. 3843–3850, Dec. 2023, doi: <https://doi.org/10.36040/jati.v7i6.8232>.
- [12] G. Zidane Dhamara and A. Nugroho, “Klasifikasi Genre Musik Menggunakan Machine Learning,” *Bulletin of Information Technology (BIT)*, vol. 6, no. 3, pp. 206–217, Sep. 2025, doi: [10.47065/bit.v5i2.2021](https://doi.org/10.47065/bit.v5i2.2021).
- [13] B. Hakim, F. Joanda Kaunang, C. Susanto, J. Salim, and R. Indradjaja, “Implementasi Machine Learning Dalam Pengelompokan Musik Menggunakan Algoritma K-Means Clustering,” 2025. [Online]. Available: <http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/indexBhustomyHakim|http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index|>
- [14] N. Aliya, A. Aini, R. Cinta Endynda, and H. Al Rosyid, “Jurnal Sains dan Teknologi Penerapan Algoritma K-Means untuk Klasterisasi Lagu Terpopuler 2025 Versi Spotify,” *Jurnal Sains Dan Teknologi*, vol. 02, no. 3, p. 147, 2026, doi: [10.62379/jsit.v2i3.939](https://doi.org/10.62379/jsit.v2i3.939).
- [15] M. A. Munajad, A. Ridwan, and T. G. Pratama, “Pengembangan Sistem Rekomendasi Musik dengan K-Means dan KNN Berbasis Cosine Similarity,” *Sainteks*, vol. 22, no. 2, pp. 153–165, Oct. 2025, doi: [10.30595/sainteks.v22i2.27815](https://doi.org/10.30595/sainteks.v22i2.27815).



- [16] I. Komang Dika Setiawan and I. Wayan Sudiarsa, “Analisis Klasifikasi Perilaku Pengguna Terhadap Customer Churn Pada Layanan Musik Spotify Menggunakan Metode Random Forest,” JMA), vol. 4, no. 1, pp. 3031–5220, Jan. 2026, doi: 10.62281.
- [17] N. Sofie Soraya, “Komparasi linear regression, random forest regression, dan multilayer perceptron regression untuk prediksi tren musik TikTok,” AITI: Jurnal Teknologi Informasi, vol. 20, no. Agustus, pp. 191–205, Apr. 2023, doi: <https://doi.org/10.24246/aiti.v20i2.191-205>.
- [18] P. Studi Informatika and F. Matematika dan Ilmu Pengetahuan Alam, “Analisis dan Klasifikasi Genre Musik Menggunakan Algoritma STFT dan Random Forest Merry Royanti Manalu a1 , Made Agung Raharja a2,” JNATIA, vol. 3 No. 1, no. 1, pp. 205–214, Nov. 2024, doi: <https://doi.org/10.24843/JNATIA.2024.v03.i01.p24>.
- [19] S. Naeem, A. Ali, S. Anam, and M. M. Ahmed, “An Unsupervised Machine Learning Algorithms: Comprehensive Review,” International Journal of Computing and Digital Systems, vol. 13, no. 1, pp. 911–921, Apr. 2023, doi: 10.12785/ijcds/130172.
- [20] N. Defany Marsya et al., “Jurnal Sains Informatika Terapan (JSIT) Penerapan Algoritma Random Forest Dalam Prediksi Emosi Musik Berdasarkan Karakteristik Fitur Audio Spotify,” JSIT, vol. 4, pp. 435–440, Aug. 2025, doi: <https://doi.org/10.62357/jsit.v4i2.648>.