



Klasifikasi Genre Musik Menggunakan Algoritma *Support Vector Machine*

(*Music Genre Classification Using Support Vector Machine Algorithm*)

Jersy Pattikawa¹, Samuel Peilouw², Margin Koritelu³, Citra Fathia Palembang^{4*},
Emanuella M. C. Wattimena⁵

^{1,2,3,4,5}Program Studi Ilmu Komputer, Fakultas Sains dan Teknologi, Universitas Pattimura

Jl. Ir. M. Putuhena, Ambon, 97233, Indonesia

*Corresponding Author: *fpchiet@gmail.com

Manuscript submitted:
15th April 2026

Manuscript revision:
20th April 2026

Accepted for publication:
5th May 2026

Abstract

The development of digital music platforms such as Spotify and YouTube Music has led to a significant increase in the volume of music data available online. This situation presents challenges in organizing and retrieving songs that match user preferences. One viable solution is automatic music genre classification using machine learning techniques. This study aims to develop a music genre classification model using the Support Vector Machine (SVM) algorithm based on text features and artist metadata. The dataset used consists of the attributes artist_name, tags, and playlist genre. The research stages include data cleaning, text feature transformation using the Term Frequency-Inverse Document Frequency (TF-IDF) method, artist name encoding using Label Encoder, train-test data splitting, SVM model training, and evaluation using accuracy, precision, recall, and F1-score metrics. To improve model performance, parameter optimization was carried out using GridSearchCV. The results show that the SVM model is capable of classifying music genres with an accuracy of 67%. The Pop genre achieved the best classification performance, while the Rock and R&B genres exhibited relatively high misclassification rates due to data distribution imbalance and the limitations of the features used. The findings indicate that the SVM algorithm is sufficiently effective for text-based music genre classification; however, the addition of audio features and data balancing techniques is still required to further improve model performance.

Keywords: Machine Learning; Music Genre Classification, Support Vector Machine (SVM); Text-Based Classification; TF-IDF.



This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).

How to cite this article:

J. Pattikawa, S. Peilouw, M. Koritelu, C. F. Palembang, and E. M. C. Wattimena, "Klasifikasi Genre Musik Menggunakan Algoritma Support Vector Machine", *algorithm*, vol. 2, no. 1, pp. 11-18, May 2026.

Copyright © 2026 Author(s)

Journal homepage: <https://ojs3.unpatti.ac.id/index.php/algorithm>

Research Article [Open Access](#)

1. PENDAHULUAN

Musik merupakan bentuk hiburan yang memiliki peran penting dalam kehidupan masyarakat yang terus berkembang seiring dengan kemajuan teknologi dan budaya. Di Indonesia, musik telah menjadi bagian tak terpisahkan dari keseharian, dengan pasar musik streaming yang besar dan beragam genre favorit seperti Pop, Rock, Hiphop, dan Klasik [1]. Kemunculan platform digital dan layanan streaming musik seperti Spotify dan YouTube Music telah memicu ledakan jumlah data musik digital, yang pada gilirannya menimbulkan tantangan dalam pencarian, pengelompokan, dan rekomendasi lagu yang relevan bagi pengguna [2][3].

Kebutuhan untuk mengelompokkan dan merekomendasikan musik berdasarkan genre atau preferensi pengguna menjadi semakin penting. Sistem klasifikasi genre musik secara otomatis tidak hanya dapat membantu pengguna menemukan lagu yang sesuai secara cepat dan akurat, tetapi juga meningkatkan pengalaman pengguna secara keseluruhan [1][3]. Genre musik sendiri berfungsi sebagai label deskriptif yang menggambarkan karakteristik umum lagu, seperti struktur ritmis, harmoni, dan instrumentasi, yang dapat dimanfaatkan dalam proses klasifikasi [3].

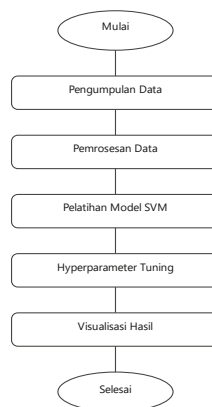
Selain klasifikasi genre, pengembangan sistem rekomendasi musik berbasis kecerdasan buatan juga menjadi fokus utama yang bertujuan untuk memprediksi lagu yang disukai atau kemungkinan besar akan didengarkan oleh pengguna, menggunakan pendekatan yang lebih kompleks dibanding sistem rekomendasi tradisional yang hanya memberikan hasil benar atau salah [4],[5]. Penerapan machine learning dalam sistem ini memungkinkan peningkatan akurasi dan efektivitas dalam memberikan rekomendasi yang personal.

Untuk menunjang kedua pendekatan tersebut klasifikasi dan rekomendasi pemahaman mendalam terhadap karakteristik data musik menjadi hal yang krusial. Proses analisis dan pengelompokan yang tepat akan membantu menyusun basis sistem yang efisien dan relevan dalam konteks aplikasi nyata pada platform digital.

Support Vector Machine (SVM) adalah algoritma pembelajaran mesin yang digunakan untuk tugas klasifikasi dan regresi. SVM bekerja dengan mencari hyperplane optimal yang memisahkan data dari berbagai kelas dengan margin terbesar sehingga meningkatkan akurasi klasifikasi [6],[7],[8]. Algoritma ini efektif dalam menangani data berdimensi tinggi dan mampu mengatasi masalah non-linear melalui penggunaan kernel.

2. METODE PENELITIAN

Penelitian ini merupakan studi kuantitatif terapan yang mengimplementasikan algoritma SVM dalam mengklasifikasikan genre musik dengan menggunakan data teks dan metadata, serta menilai kinerja model melalui pengukuran kuantitatif yang menggunakan metrik evaluasi klasifikasi. Penelitian dilakukan melalui beberapa tahapan sebagaimana ditunjukkan pada Gambar 1.



Gambar 1. Diagram alur penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder yang berbentuk dataset lagu dengan format CSV yang diambil dari sumber online, dengan atribut utama yaitu *artist_name*, *tags* (keterangan lagu), dan *playlist_genre* (*label genre*). Proses pengolahan data dilakukan melalui beberapa langkah, dimulai dengan mengimpor dataset menggunakan pustaka *pandas*, kemudian melanjutkan dengan membersihkan data untuk menghilangkan baris yang kosong atau duplikat agar kualitas data tetap terjaga. Setelah itu, kolom *artist_name* diubah menjadi nilai numerik dengan metode *LabelEncoder*, sedangkan kolom *tags* diproses menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk menciptakan representasi fitur numerik dari informasi teks. Setelah kedua fitur tersebut dipersiapkan, data kemudian dibagi menjadi dua bagian, yaitu 80% untuk data latih, dan 20% untuk data uji dengan menggunakan fungsi *train_test_split* agar model dapat dilatih dan dinilai dengan baik, kemudian dievaluasi performanya menggunakan data pengujian. Untuk meningkatkan performa model, dilakukan proses *tuning hyperparameter* menggunakan *GridSearchCV*, dan terakhir adalah visualisasi hasil prediksi model.

3. HASIL DAN PEMBAHASAN

3.1 Pemrosesan Data

Pemrosesan data dimulai dengan membaca dataset lagu dari file CSV menggunakan pustaka *pandas*. Data kemudian disaring untuk hanya menyertakan kolom *artist_name*, *tags*, dan *playlist_genre*, yang merupakan fitur dan label utama dalam klasifikasi. Baris yang mengandung nilai kosong dihilangkan untuk menjaga kualitas data. Setelah itu, nama artis dikonversi menjadi bentuk numerik dengan *LabelEncoder*, sehingga dapat digunakan oleh model *machine learning*. Kolom *tags*, yang berisi deskripsi lagu dalam bentuk teks, diolah menggunakan teknik TF-IDF untuk mengubah teks menjadi vektor numerik sebelum dilakukan klasifikasi [9]. Fitur *tags* dan *artist_name* kemudian digabung menjadi satu matriks fitur akhir yang siap digunakan untuk pelatihan model. Label target, yaitu genre lagu (*playlist_genre*), juga di-*encode* ke dalam bentuk numerik agar dapat dikenali oleh algoritma klasifikasi.

3.2 Pelatihan Model

Setelah data selesai diproses, tahap berikutnya adalah pelatihan model. Dataset dibagi menjadi dua bagian, yaitu 80% untuk data pelatihan dan 20% untuk data pengujian, menggunakan fungsi *train_test_split* dari pustaka *sklearn*. Model yang digunakan dalam proyek ini adalah SVM. Model dilatih pada data pelatihan (*X_train*, *y_train*) dengan menggunakan parameter default. Pada tahap ini, model mempelajari pola hubungan antara fitur-fitur seperti *tags* dan *artist_name* terhadap genre lagu, sehingga mampu mengenali pola yang mungkin tidak terlihat secara kasat mata.

3.3 Evaluasi Model

Model yang telah dilatih kemudian dievaluasi performanya menggunakan data pengujian. Prediksi dilakukan terhadap data uji (*x_test*), dan hasil prediksi dibandingkan dengan label asli (*y_test*) untuk mengukur akurasi. Evaluasi dilakukan menggunakan *classification_report*, yang menampilkan berbagai metrik seperti *precision*, *recall*, *f1-score*, dan *accuracy*. Metrik-metrik ini membantu memahami seberapa baik model dalam mengklasifikasikan lagu ke dalam genre yang benar. Jika nilai akurasi atau *f1-score* masih rendah, maka perlu dilakukan perbaikan, seperti *tuning parameter model* atau pengolahan data yang lebih baik.

```
Akurasi Model: 0.67
Classification Report:
      precision    recall  f1-score   support

   Edm         0.75     0.60     0.67         5
   Latin        0.60     0.68     0.64        22
   Pop         0.64     0.91     0.75        86
   R&B         0.00     0.00     0.00         12
   Rap         0.86     0.57     0.69        42
   Rock        1.00     0.08     0.14        13

 accuracy          0.64
 macro avg         0.64     0.47     0.48       180
 weighted avg     0.67     0.67     0.63       180
```

Gambar 2. *Classification_report*

Gambar 2. Menampilkan hasil evaluasi sebuah model untuk mengklasifikasikan genre musik. Model tersebut diuji menggunakan 180 data dan berhasil mencapai akurasi sebesar 67%. Pada *classification_report*, terlihat bahwa model bekerja dengan tingkat keberhasilan yang berbeda-beda untuk setiap genre. Genre Pop memiliki data terbanyak yaitu 86 sampel, diikuti Rap dengan 42 sampel, Latin 22 sampel, Rock 13 sampel, R&b 12 sampel, dan Edm hanya 5 sampel. Dari sisi performa, model paling kesulitan mengenali genre R&b karena seluruh nilainya nol, yang berarti model sama sekali tidak mampu mengidentifikasi genre tersebut. Rock juga bermasalah meski dari sisi precision nilainya sempurna 1.00, namun recall-nya hanya 0.08 yang berarti hampir semua lagu Rock terlewat. Sementara Pop menjadi genre yang paling mudah dikenali dengan recall 0.91. Secara keseluruhan, rata-rata tertimbang model ini berada di angka 0.67 untuk precision dan recall, yang sejalan dengan nilai akurasi.

3.4 Hyperparameter Tuning

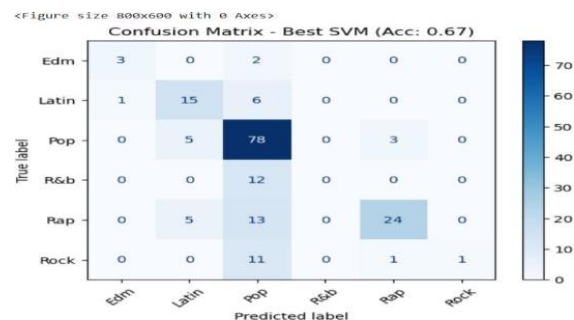
Untuk meningkatkan performa model, dilakukan proses *tuning hyperparameter* menggunakan *GridSearchCV* [10]. Parameter yang diuji meliputi nilai C, jenis kernel, dan gamma, yang semuanya mempengaruhi cara kerja algoritma SVM. *Grid Search* akan mencoba berbagai kombinasi parameter dan melakukan *cross-validation* untuk menentukan kombinasi parameter terbaik berdasarkan akurasi. Setelah proses tuning selesai, model terbaik digunakan kembali untuk memprediksi data uji, dan akurasi dihitung kembali untuk dibandingkan dengan model sebelumnya. Biasanya, proses tuning ini dapat secara signifikan meningkatkan performa model, terutama jika dataset memiliki kompleksitas yang tinggi.

Tabel 1. *Hyperparameter Tuning* pada model SVM menggunakan *Grid Search* dengan *cross-validation*

<i>Param_grid</i>	
<i>C</i>	0.1, 1, 10
<i>kernel</i>	<i>linear, rbf</i>
<i>gamma</i>	<i>scale, auto</i> (hanya untuk kernel non-linear)

3.5 Visualisasi

Langkah terakhir adalah visualisasi hasil prediksi model. *Confusion_matrix* digunakan untuk menunjukkan jumlah prediksi benar dan salah dari masing-masing kelas (genre). Matriks ini divisualisasikan menggunakan *ConfusionMatrixDisplay*, yang memberi gambaran visual tentang performa klasifikasi model. Setiap baris mewakili label sebenarnya, sedangkan setiap kolom mewakili label prediksi. Warna atau nilai diagonal utama (kiri atas ke kanan bawah) menunjukkan jumlah prediksi yang tepat. Visualisasi ini sangat membantu untuk mengidentifikasi genre mana yang sering salah diklasifikasikan dan memahami kelemahan model secara lebih intuitif.



Gambar 3. *Confussion Matrix*

Gambar 3 *confusion matrix* menunjukkan performa model SVM terbaik dengan akurasi 67% dalam mengklasifikasikan genre lagu. Genre Pop paling sering terklasifikasi dengan benar, namun juga menjadi genre yang paling sering diprediksi secara keliru oleh model untuk genre lain seperti Rock, Rap, dan R&B. Genre seperti R&B dan Rock memiliki akurasi sangat rendah karena sebagian besar lagu dari genre tersebut diklasifikasikan sebagai Pop, menandakan adanya ketimpangan data dan kesamaan representasi fitur. Kesalahan klasifikasi yang cukup tinggi ini menunjukkan bahwa fitur teks dan nama artis belum cukup membedakan genre secara optimal, sehingga diperlukan penambahan fitur audio atau teknik pengolahan lanjutan agar performa model meningkat.

Hal ini menunjukkan bahwa meskipun model SVM umumnya menghasilkan klasifikasi yang baik, terdapat kekurangan yang cukup signifikan, terutama dalam identifikasi genre-genre tertentu. Genre seperti R&B dan Rock sering kali salah terklasifikasikan karena jumlah datanya lebih sedikit dibandingkan dengan genre seperti Pop, sehingga model cenderung bias terhadap genre yang lebih banyak. Selain itu, fitur yang digunakan, yaitu deskripsi teks dan nama artis, belum cukup silam untuk membedakan karakteristik antar genre secara mendalam. Dengan demikian, guna meningkatkan akurasi dan keadilan dalam klasifikasi, model perlu dikembangkan lebih lanjut, misalnya dengan menyeimbangkan jumlah data di setiap genre dan menambahkan fitur-fitur baru, seperti fitur audio (contohnya tempo, ritme, dan nada), agar model dapat mengenali setiap genre dengan lebih akurat.

Selain *confusion matrix*, pada penelitian ini juga dibangun *dashboard* visualisasi interaktif menggunakan *Streamlit* yang bertujuan untuk mengeksplorasi data lagu populer secara lebih mendalam dan mudah dipahami.

3.6 Visualisasi Dashboard

Visualisasi Dashboard ini menampilkan tiga visualisasi utama. Pertama, tabel Top 10 Lagu Terpopuler dalam genre Pop yang memperlihatkan nama lagu, artis, dan skor popularitasnya.



Dashboard Lagu Populer

Top 10 Lagu Terpopuler dalam Genre Pop

	track_name	track_artist	track_popularity
96	blinding lights	Theweeknd	1
95	memories	Maroon5	1
77	tusa	Karolg	1
388	falling	Trevordaniel	0.9545
79	everything i wanted	Billieeilish	0.9545
154	don't start now	Dualipa	0.9545
21	ritmo (bad boys for life)	Theblackeyedpeas	0.9091
52	bad guy	Billieeilish	0.8636
82	yummy	Justinbieber	0.8636
97	someone you loved	Lewisalpaldi	0.8182

Gambar 4. *Dashboard* visualisasi data Top 10 Lagu Terpopuler

Gambar 4. Dashboard ini dibuat menggunakan *Streamlit* atau *Dash* dengan desain antarmuka yang sederhana dan informatif. Kedua, grafik batang horizontal yang memvisualisasikan perbandingan popularitas kesepuluh lagu tersebut secara sekilas, dengan masing-masing artis dibedakan melalui warna yang berbeda.



Gambar 5. Grafik Popularitas Lagu

Selanjutnya, berdasarkan gambar 5, sumbu Y (vertikal) menampilkan daftar judul lagu dari atas ke bawah, dengan yang paling populer berada di bagian atas. Sementara itu, sumbu X (horizontal) menunjukkan nilai popularitas lagu dalam skala 0 hingga 1, dengan interpretasi data sebagai berikut:

1. Lagu dengan judul *blinding lights* oleh The Weeknd, *memories* oleh Maroon 5, dan *tusa* oleh Karol G memiliki nilai popularitas maksimal dalam dataset yaitu 1.0
2. Lagu dengan skor popularitas tinggi tapi di bawah puncak yaitu dengan judul *falling* oleh Trevor Daniel, *everything i wanted* oleh Billie Eilish dan *don't start now* oleh Dua Lipa memiliki skor 0.9545 tetap tergolong sangat populer.
3. Lagu dengan popularitas terendah dengan judul *someone you loved* oleh Lewis Capaldi berada pada angka 0.8182 namun tetap masuk dalam 10 besar.

Ketiga, grafik *scatter plot* yang menunjukkan hubungan antara tanggal rilis album dengan tingkat popularitas lagu, sehingga dapat dilihat bagaimana tren popularitas bergerak dari akhir 2019 hingga awal 2020.



Gambar 6. Grafik Popularitas Lagu

Kemudian, Gambar 6 digunakan untuk menganalisis hubungan antara tanggal rilis lagu dengan tingkat popularitas lagu tersebut. Sumbu X (horizontal) merepresentasikan waktu rilis album atau lagu (*track_album_release_date*) mulai dari November 2018 hingga awal 2020, sumbu Y (vertikal) menampilkan nilai popularitas lagu (*track_popularity*) dalam skala 0-1, di mana nilai 1 menunjukkan tingkat popularitas tertinggi. Kemudian untuk setiap titik mewakili satu lagu, dan warna menunjukkan artis yang menyangikan lagu tersebut. Sedangkan untuk garis tren (*trendline*) menunjukkan arah umum hubungan antara waktu rilis dan popularitas lagu.

Berdasarkan grafik tersebut, dapat diinterpretasikan sebagai berikut:

1. Lagu dengan popularitas tinggi:

- a. Lagu dari artis seperti The Weeknd, Dua Lipa, dan Maroon 5 yang dirilis menjelang akhir 2019 memiliki tingkat popularitas mendekati 1, menunjukkan penerimaan yang sangat baik dari pendengar tak lama setelah rilis.
 - b. Beberapa lagu dari Billie Eilish tetap sangat populer meskipun dirilis lebih awal pada pertengahan 2019.
2. Garis tren yang menanjak :
Garis tren naik secara perlahan, mengindikasikan bahwa semakin baru tanggal rilis sebuah lagu, cenderung semakin tinggi popularitasnya. Hal ini dapat disebabkan oleh efek pemasaran, peningkatan jumlah pengguna platform digital, atau relevansi lagu terhadap tren saat itu.
3. *Outlier*
Lagu dari Trevor Daniel (warna ungu) dirilis cukup awal (akhir 2018 atau awal 2019) tetapi tetap menunjukkan popularitas yang tinggi, menandakan potensi viral atau efek delay popularitas.

4. KESIMPULAN

Penelitian ini berhasil mengklasifikasikan genre musik secara otomatis menggunakan algoritma Support Vector Machine (SVM) dengan memanfaatkan fitur *tags* dan metadata artis_name yang ditunjukkan dengan tingkat akurasi sebesar 67% kinerja terbaik pada genre yang umum seperti Pop. Namun, model ini kurang efektif pada genre yang jarang seperti Rock dan R&B disebabkan oleh ketidakseimbangan data serta keterbatasan fitur yang digunakan. GridSearchCV mampu meningkatkan performa, tetapi tidak sepenuhnya mengatasi bias dalam klasifikasi. Secara keseluruhan algoritma SVM terbukti efektif untuk klasifikasi genre yang berbasis teks, namun akurasi dan keadilan model masih bisa diperbaiki dengan menambah fitur audio dan mengatur agar distribusi genre dalam dataset lebih seimbang.

REFERENSI

- [1] L. Novea and B. H. Prasetio, "Sistem Prediksi Genre Musik dan Penyediaan Tautan Rekomendasi Daftar Putar Menggunakan Teknik STFT dan Decision Tree Machine Learning," *Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 1, pp. 1–15, 2017, [Online]. Available: <https://repository.ub.ac.id/id/eprint/227821/>.
- [2] B. Hakim, F. J. Kaunang, C. Susanto, J. Salim, and R. Indradjaja, "Implementasi Machine Learning Dalam Pengelompokan Musik Menggunakan Algoritma K-Means Clustering," *IDEALIS Indones. J. Inf. Syst.*, vol. 8, no. 1, pp. 74–83, 2025, doi: 10.36080/idealism.v8i1.3357.
- [3] G. A. V. M. Giri and M. L. Radhitya, "Klasifikasi Genre Musik Menggunakan Teknik Pembelajaran Mesin," *J. Teknol. Inf. dan Komput.*, vol. 9, no. 1, pp. 1–9, 2023, doi: 10.36002/jutik.v9i1.2176.
- [4] M. A. Madani, G. Helmi, and Y. Hendra, "Sistem Rekomendasi Musik Menggunakan Machine Learning," *J. Technol. Comput.*, vol. 1, no. 1, pp. 40–49, 2024, [Online]. Available: <https://http/ojs.adzkie.ac.id/index.php/jtech>Journalhomepage:<https://ojs.adzkie.ac.id/index.php/jtech>.
- [5] Z. K. A. Baizal and R. Andiety, "Content-Based Music Recommender System Using Deep Neural Network," *Build. Informatics, Technol. Sci.*, vol. 6, no. 2, pp. 1111–1119, 2024, doi: 10.47065/bits.v6i2.5762.
- [6] M. A. Ramadhani *et al.*, "Implementasi Algoritma Support Vector Machine (SVM) Untuk Diagnosis Kesehatan Manusia Berbasis Web Application," *J. Ners*, vol. 9, no. 1, pp. 896–902, 2025, [Online]. Available: <https://journal.universitaspahlawan.ac.id/index.php/ners/article/view/31481>.
- [7] Y. Anne Mudya and T. M. Ridho, "Metode Support Vector Machine Untuk Analisis Sentimen

- Aplikasi Threads di Google Play Store,” *Indones. J. Comput. Sci.*, vol. 13, no. 5, pp. 76–83, 2024, doi: 10.33022/ijcs.v13i5.4446.
- [8] F. Putrawansyah and T. Susanti, “Penerapan Metode Support Vector Machine Terhadap Klasifikasi Jenis Jambu Biji,” *JIKO (Jurnal Inform. dan Komputer)*, vol. 8, no. 1, p. 193, 2024, doi: 10.26798/jiko.v8i1.988.
- [9] C. F. Palembang, E. M. C. Wattimena, and V. I. Fataruba, “Analisis Sentimen Dosen terhadap Kebijakan Tunjangan Kinerja di Perguruan Tinggi Negeri Menggunakan Algoritma Naive Bayes,” *ALGORITHM J. Comput. Sci. Comput. Intell.*, vol. 1, no. 1, pp. 21–28, 2025, doi: 10.30598/algorithm.v1i1.21-28.
- [10] A. Hamied Nababan and M. Y. Hutagalung, “Hyperparameter Tuning Pada Model Stance Detection Menggunakan GridSearchCV,” *J. Sains dan Teknol.*, vol. 5, no. 1, pp. 205–209, 2023, [Online]. Available: <https://doi.org/10.55338/saintek.v5i1.1505>.