

# Implementasi Deep Learning Menggunakan Kerangka Kerja Cross-Industry Standard Process For Data Mining (CRISP- DM) Pada Peminat Genre Musik Di Aplikasi Spotify

Julius Yoshua Siahaan<sup>1</sup>, Ichsan Ramdhani<sup>2\*</sup>

<sup>1,2</sup>Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pamulang, Tangerang Selatan, Indonesia

Email : <sup>1</sup>[juliusyosuasiahaan@yahoo.co.id](mailto:juliusyosuasiahaan@yahoo.co.id), <sup>2\*</sup>[dosen02110@unpam.ac.id](mailto:dosen02110@unpam.ac.id)

(\* : coressponding author)

**Abstrak** - Penelitian ini bertujuan memetakan peminat genre musik pada tingkat kumpulan lagu Spotify melalui implementasi deep learning menggunakan kerangka kerja Cross-Industry Standard Process for Data Mining (CRISP-DM). Dataset berasal dari rekomendasi halaman utama Spotify yang dikompilasi ke dalam satu playlist penelitian pada tahun 2024, kemudian diekspor melalui Exportify. Setelah pembersihan data, diperoleh 7.858 track dengan 15 fitur audio dan metadata numerik. Label genre dibentuk secara bertingkat dari genre awal, distandardisasi menjadi genre utama, lalu dikelompokkan ke delapan meta\_genre, yaitu Rock/Metal/Punk, Pop/Indo/Malay, Other, Urban, Country/Folk, EDM/Dance, Latin, dan Jazz/Classical. Pemodelan dilakukan dengan k-Nearest Neighbor (k-NN), Gaussian Naive Bayes, dan model gabungan berbasis rata-rata probabilitas. Evaluasi menggunakan accuracy, precision, recall, dan weighted F1-score pada data validasi dan data uji, disertai pengujian kestabilan melalui beberapa pengulangan stratified split. Hasil menunjukkan bahwa k-NN memberikan kinerja terbaik pada data uji dengan accuracy 65,39% dan weighted F1-score 62,61%. Model gabungan mencapai weighted precision tertinggi pada validasi sebesar 67,45%, tetapi masih sensitif terhadap variasi pembagian data. Deployment pada seluruh dataset memperlihatkan dominasi Rock/Metal/Punk sebesar 62,4% dan Pop/Indo/Malay sebesar 11,7%. Temuan ini menunjukkan bahwa CRISP-DM mampu menata proses klasifikasi genre secara sistematis, sementara hasil deployment dapat digunakan untuk membaca peminat genre pada tingkat kumpulan lagu, bukan pada tingkat pengguna individual.

**Kata Kunci:** *Deep Learning, CRISP-DM, Spotify, K-Nearest Neighbor, Naive Bayes, Peminat Genre*

*Abstract - This study maps music genre interest at the level of Spotify song collections by implementing deep learning within the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework. The dataset was compiled from Spotify home-page recommendations in 2024 and exported through Exportify. After data cleaning, 7,858 tracks remained with 15 numerical audio-feature and metadata variables. Genre labels were organized hierarchically from genre awal to genre utama and then merged into eight meta\_genre classes: Rock/Metal/Punk, Pop/Indo/Malay, Other, Urban, Country/Folk, EDM/Dance, Latin, and Jazz/Classical. Modeling used k-Nearest Neighbor (k-NN), Gaussian Naive Bayes, and a probability-averaging ensemble. Performance was evaluated with accuracy, precision, recall, and weighted F1-score on validation and test sets, followed by stability assessment using repeated stratified splits. The results show that k-NN achieved the best overall test performance with 65.39% accuracy and 62.61% weighted F1-score. The ensemble obtained the highest weighted precision on validation at 67.45%, yet remained sensitive to data-split variation. Deployment over the full dataset indicates that Rock/Metal/Punk dominates the collection at 62.4%, followed by Pop/Indo/Malay at 11.7%. These findings indicate that CRISP-DM provides a traceable structure for genre classification, while the deployment output can be interpreted as genre interest at the collection level rather than at the individual-user level.*

**Keywords:** *Deep Learning, CRISP-DM, Spotify, K-Nearest Neighbor, Naive Bayes, Genre Interest*

## 1. PENDAHULUAN

Perkembangan layanan streaming telah mengubah cara musik dikonsumsi, dikurasi, dan direkomendasikan. Spotify menyediakan katalog yang luas, playlist tematik, serta audio features dan metadata terstruktur yang dapat dimanfaatkan untuk analisis musik digital [1]-[3]. Namun, klasifikasi genre musik tetap menghadapi persoalan metodologis karena batas antar-genre bersifat dinamis, dipengaruhi proses pelabelan metadata, dan tidak selalu sepenuhnya tercermin oleh fitur numerik [4], [5].

Masalah tersebut menjadi lebih penting ketika tujuan analisis bukan sekadar memberi label pada satu lagu, melainkan menggambarkan kecenderungan minat genre pada suatu kumpulan lagu. Dalam konteks ini, hasil klasifikasi perlu dikelola melalui prosedur yang terstruktur agar hubungan

antara tujuan analitis, kualitas data, pemodelan, evaluasi, dan penerapan keluaran dapat ditelusuri dengan jelas. CRISP-DM memberikan kerangka kerja yang sesuai karena menata proses data mining ke dalam enam tahap: business understanding, data understanding, data preparation, modeling, evaluation, dan deployment [6], [7].

Penelitian ini memposisikan peminat genre pada tingkat agregat kumpulan lagu, bukan pada profil pengguna individual. Posisi tersebut sejalan dengan literatur rekomendasi musik yang menempatkan interaksi pengguna, struktur playlist, dan distribusi item sebagai sinyal preferensi pada level kolektif [8]-[11]. Dengan demikian, keluaran model tidak ditafsirkan sebagai riwayat putar pengguna tertentu, melainkan sebagai gambaran distribusi genre pada himpunan track yang dianalisis.

Berangkat dari kebutuhan tersebut, penelitian ini mengimplementasikan deep learning dalam pengertian operasional sebagai siklus pelatihan, evaluasi, penyesuaian, dan pengujian yang dilaksanakan secara iteratif pada fase modeling-evaluation di dalam CRISP-DM. Dua algoritma klasifikasi digunakan, yaitu k-Nearest Neighbor (k-NN) dan Gaussian Naive Bayes, kemudian keduanya juga dikombinasikan melalui rata-rata probabilitas. Penelitian berfokus pada tiga pertanyaan utama: bagaimana menyusun tahapan klasifikasi genre secara sistematis, bagaimana kinerja masing-masing model pada data validasi dan data uji, dan bagaimana sebaran peminat genre pada tingkat kumpulan lagu setelah deployment model dilakukan pada seluruh dataset.

Kontribusi penelitian ini terletak pada tiga hal. Pertama, penelitian menyediakan alur kerja klasifikasi genre berbasis CRISP-DM yang dapat ditelusuri dari tahap perumusan tujuan sampai deployment. Kedua, penelitian menyajikan evaluasi komparatif antara k-NN, Gaussian Naive Bayes, dan model gabungan pada dataset rekomendasi Spotify. Ketiga, penelitian menawarkan pembacaan peminat genre berbasis proporsi hasil prediksi genre pada tingkat kumpulan lagu, sehingga interpretasi tetap berada dalam batas data yang tersedia.

## 2. METODE PENELITIAN

Penelitian menggunakan desain kuantitatif dengan eksperimen komputasional pada data musik digital. Sumber data berasal dari rekomendasi halaman utama Spotify pada periode pengumpulan tahun 2024. Seluruh track hasil rekomendasi dikompilasi ke satu playlist penelitian dan diekspor menjadi berkas CSV melalui Exportify. Unit analisis penelitian adalah track, bukan pengguna. Dataset awal berjumlah 9.864 entri dan setelah pembersihan diperoleh 7.858 track yang siap diproses.

Kerangka kerja penelitian mengikuti CRISP-DM. Pada business understanding ditetapkan tujuan untuk memperoleh klasifikasi genre yang cukup andal sebagai dasar pembacaan peminat genre pada tingkat kumpulan lagu. Pada data understanding dilakukan pemeriksaan struktur dataset, kualitas atribut, dan distribusi label awal. Pada data preparation dilakukan pembersihan nilai kosong, konversi atribut Explicit menjadi numerik biner, normalisasi Min-Max pada fitur numerik, serta pembentukan label bertingkat melalui alur genre\_awal -> genre\_utama -> meta\_genre.

Pemodelan menggunakan 15 fitur numerik, yaitu Track Duration (ms), Popularity, Danceability, Energy, Key, Loudness, Mode, Speechiness, Acousticness, Instrumentalness, Liveness, Valence, Tempo, Time Signature, dan Explicit. Label sasaran dikelompokkan ke delapan meta\_genre: Rock/Metal/Punk, Pop/Indo/Malay, Other, Urban, Country/Folk, EDM/Dance, Latin, dan Jazz/Classical. Pembagian data dilakukan secara stratified dengan random\_state=42 ke dalam data latih 64%, data validasi 16%, dan data uji 20%, sehingga distribusi kelas antar-subset tetap terjaga [16].

Model pertama adalah k-NN berbasis jarak Euclidean dengan pembobotan distance. Kandidat nilai k yang diuji adalah 3, 5, 7, 9, 11, 15, dan 21. Model kedua adalah Gaussian Naive Bayes yang memodelkan distribusi probabilistik fitur numerik pada tiap kelas [12], [13]. Selain dua model tunggal, dibangun model gabungan dengan menghitung rata-rata probabilitas keluaran k-NN dan Gaussian Naive Bayes untuk setiap kelas, kemudian memilih kelas dengan probabilitas tertinggi sebagai prediksi akhir.

Evaluasi kinerja menggunakan accuracy, precision, recall, dan weighted F1-score yang diturunkan dari confusion matrix [14], [15]. Weighted averaging dipilih karena distribusi kelas tidak sepenuhnya seimbang. Pada tahap evaluation juga dilakukan lima pengulangan stratified split untuk melihat kestabilan kinerja model gabungan terhadap variasi pembagian data. Tahap deployment menerapkan model yang ditetapkan pada seluruh dataset untuk memperoleh distribusi meta\_genre hasil prediksi. Peminat genre kemudian dibaca sebagai proporsi  $p(g)=n_g/N$ , dengan  $n_g$  menyatakan jumlah track pada genre  $g$  dan  $N$  menyatakan total track yang dianalisis.

**Tabel 1.** Ringkasan Dataset Dan Pembagian Data

Komponen	Keterangan	Jumlah
Dataset awal	Track hasil kompilasi rekomendasi Spotify	9.864
Dataset siap model	Track setelah pembersihan dan seleksi fitur	7.858
Fitur numerik	Audio features dan metadata numerik	15
Kelas target	meta_genre hasil pengelompokan	8
Data latih	Stratified split	5.028
Data validasi	Stratified split	1.258
Data uji	Stratified split	1.572

### 3. HASIL DAN PEMBAHASAN

Hasil pada tahap data preparation menunjukkan bahwa pengelompokan label berhasil menurunkan fragmentasi genre dari label awal yang sangat beragam menjadi delapan meta\_genre yang lebih terkelola untuk kebutuhan klasifikasi. Kelas Rock/Metal/Punk merupakan kelas terbesar pada dataset siap model, diikuti Pop/Indo/Malay dan Other. Kondisi ini memperlihatkan adanya ketidakseimbangan kelas, sehingga pemilihan weighted F1-score sebagai metrik utama menjadi relevan.

Pengujian kandidat  $k$  pada model  $k$ -NN menunjukkan bahwa konfigurasi  $k=5$  dipilih sebagai konfigurasi operasional setelah mempertimbangkan nilai weighted F1-score dengan presisi penuh pada sistem. Pada tampilan hasil yang telah dibulatkan,  $k=5$  menghasilkan accuracy 68,36%, precision 63,78%, recall 68,36%, dan weighted F1-score 65,21% pada data validasi. Gaussian Naive Bayes memberikan accuracy 32,59% dan weighted F1-score 39,67%. Sementara itu, model gabungan memberikan weighted precision tertinggi pada validasi, yaitu 67,45%, dengan accuracy 52,62% dan weighted F1-score 55,86%.

**Tabel 2.** Kinerja Model Pada Data Validasi

Model	Accuracy (%)	Precision (%)	Recall (%)	Weighted F1 (%)
k-NN (k=5)	68,36	63,78	68,36	65,21

Gaussian Naive Bayes	32,59	57,00	32,59	39,67
k-NN + Naive Bayes	52,62	67,45	52,62	55,86

Jika dibandingkan pada data uji, k-NN tetap menjadi model dengan kinerja agregat tertinggi. Model ini mencapai accuracy 65,39% dan weighted F1-score 62,61%. Gaussian Naive Bayes berada jauh di bawahnya dengan accuracy 30,85% dan weighted F1-score 38,61%. Model gabungan menghasilkan weighted precision paling tinggi pada data uji, yaitu 66,32%, tetapi accuracy dan weighted F1-score-nya masih berada di bawah k-NN. Hasil ini menunjukkan bahwa penggabungan probabilitas belum otomatis meningkatkan semua metrik, terutama ketika distribusi kelas masih timpang dan kedekatan antarfitur beberapa genre saling beririsan.

Analisis kestabilan memperlihatkan bahwa empat dari lima pengulangan validasi model gabungan berada pada kisaran yang relatif berdekatan, yaitu accuracy 48,49%-52,62% dan weighted F1-score 54,02%-55,86%. Namun, satu pengulangan turun cukup tajam menjadi accuracy 32,75% dan weighted F1-score 41,79%. Temuan ini menunjukkan bahwa model gabungan masih sensitif terhadap variasi subset data, sehingga interpretasi hasil perlu dilakukan secara hati-hati dan tidak hanya bertumpu pada satu pembagian data.

**Tabel 3.** Kinerja Model Pada Data Uji

Model	Accuracy (%)	Precision (%)	Recall (%)	Weighted F1 (%)
k-NN (k=5)	65,39	61,52	65,39	62,61
Gaussian Naive Bayes	30,85	60,39	30,85	38,61
k-NN + Naive Bayes	49,55	66,32	49,55	53,56

Pada tahap deployment, model diterapkan ke seluruh 7.858 track untuk menghasilkan distribusi meta\_genre hasil prediksi. Distribusi tersebut penting karena penelitian ini membaca peminat genre pada tingkat kumpulan lagu. Dengan kata lain, setiap track dipandang merepresentasikan ketertarikan terhadap genre yang dilabelkan oleh model, sehingga proporsi akhir dapat dibaca sebagai gambaran komposisi minat genre pada koleksi lagu yang dianalisis.

Distribusi deployment menunjukkan dominasi kuat kelas Rock/Metal/Punk sebesar 62,4%, disusul Pop/Indo/Malay sebesar 11,7% dan Other sebesar 8,1%. Dominasi ini konsisten dengan distribusi kelas saat tahap penyiapan data, sekaligus menunjukkan bahwa koleksi lagu yang dikompilasi dari rekomendasi halaman utama Spotify pada periode penelitian memiliki orientasi kuat ke genre-genre berbasis rock dan turunannya. Di sisi lain, kelas Urban, Jazz/Classical, Country/Folk, EDM/Dance, dan Latin tetap muncul sebagai bagian minor namun terukur dari komposisi keseluruhan.

**Tabel 4.** Distribusi meta\_genre Hasil Deployment

meta_genre	Jumlah lagu	Proporsi p(g) (%)
Rock/Metal/Punk	4.902	62,4

Pop/Indo/Malay	922	11,7
Other	633	8,1
Jazz/Classical	479	6,1
Urban	346	4,4
Country/Folk	267	3,4
EDM/Dance	167	2,1
Latin	142	1,8
Total	7.858	100,0

Secara substantif, hasil penelitian menegaskan bahwa CRISP-DM efektif sebagai kerangka kerja untuk menjaga keterlacakan proses klasifikasi genre. Tahapan business understanding sampai deployment membuat keputusan penelitian - mulai dari definisi masalah, pembersihan data, pemilihan fitur, penentuan label, evaluasi, sampai interpretasi - dapat dijelaskan secara runtut. Nilai praktis kerangka ini tidak hanya terletak pada keberadaan enam fase, tetapi pada kemampuannya menahan interpretasi agar tetap sejalan dengan batas data yang tersedia.

Dari sisi model, k-NN terbukti lebih sesuai untuk representasi fitur yang digunakan dalam penelitian ini. Karakteristik tersebut dapat dijelaskan oleh sifat k-NN yang memanfaatkan kedekatan lokal antar-track pada ruang fitur yang telah dinormalisasi. Sebaliknya, Gaussian Naive Bayes menggunakan asumsi distribusi dan independensi fitur yang lebih sederhana, sehingga kehilangan sebagian kemampuan membedakan genre-genre dengan karakteristik audio yang saling tumpang tindih. Model gabungan meningkatkan precision, tetapi belum mampu melampaui k-NN secara menyeluruh pada data uji.

Keterbatasan utama penelitian ini adalah batas interpretasi data. Dataset berasal dari keluaran rekomendasi Spotify pada periode tertentu, bukan dari log pemutaran individual yang terukur. Oleh sebab itu, hasil deployment harus dibaca sebagai distribusi genre pada kumpulan lagu yang dianalisis, bukan sebagai klaim perilaku pengguna per orang. Selain itu, pengelompokan label ke delapan meta\_genre membantu kestabilan pemodelan, tetapi sekaligus menyederhanakan keragaman genre yang sebenarnya lebih rinci. Kedua hal tersebut perlu dipertimbangkan ketika hasil penelitian hendak dibandingkan dengan studi lain.

#### 4. KESIMPULAN

Penelitian ini menunjukkan bahwa implementasi deep learning menggunakan kerangka kerja CRISP-DM dapat digunakan untuk menata proses klasifikasi genre musik secara sistematis pada dataset rekomendasi Spotify. Setelah pembersihan data, 7.858 track diproses dengan 15 fitur numerik dan delapan meta\_genre sebagai kelas sasaran. Hasil evaluasi memperlihatkan bahwa k-NN merupakan model dengan kinerja agregat terbaik pada data uji, sedangkan model gabungan memberikan weighted precision tertinggi pada validasi tetapi masih sensitif terhadap variasi pembagian data.

Pada tahap deployment, distribusi hasil prediksi menunjukkan bahwa koleksi lagu yang dianalisis didominasi oleh Rock/Metal/Punk, diikuti Pop/Indo/Malay dan Other. Dengan kerangka interpretasi berbasis proporsi  $p(g)=n_g/N$ , distribusi tersebut dapat dibaca sebagai gambaran peminat genre pada tingkat kumpulan lagu. Penelitian lanjutan dapat memperkaya fitur, memperluas

strategi penalaan model, dan menggunakan data perilaku pemutaran yang lebih terukur agar pembacaan minat genre dapat dilakukan dengan presisi yang lebih tinggi.

## DAFTAR PUSTAKA

- [1] M. Schedl, E. Gomez, and J. Urbano, "Music Information Retrieval: Recent Developments and Applications," *Foundations and Trends in Information Retrieval*, vol. 8, no. 2-3, pp. 127-261, 2014.
- [2] D. Duman, P. Neto, A. Mavrolampados, P. Toiviainen, and G. Luck, "Music we move to: Spotify audio features and reasons for listening," *PLOS ONE*, vol. 17, no. 9, e0275228, 2022.
- [3] F. Khan et al., "Effect of Feature Selection on the Accuracy of Music Popularity Classification Using Machine Learning Algorithms," *Electronics*, vol. 11, no. 21, 3518, 2022.
- [4] J. Ramirez and M. J. Flores, "Machine learning for music genre: Multifaceted review and experimentation with audioset," *Journal of Intelligent Information Systems*, vol. 55, no. 3, pp. 469-499, 2020.
- [5] G. Ru, X. Zhang, J. Wang, N. Cheng, and J. Xiao, "Improving Music Genre Classification from Multi-Modal Properties of Music and Genre Correlations Perspective," *arXiv*, 2023.
- [6] C. Schroer, F. Kruse, and J. M. Gomez, "A Systematic Literature Review on Applying CRISP-DM Process Model," *Procedia Computer Science*, vol. 181, pp. 526-534, 2021.
- [7] R. Wirth and J. Hipp, "CRISP-DM: Towards a Standard Process Model for Data Mining."
- [8] S. Berkovsky, I. Cantador, and D. Tikk, *Collaborative Recommendations: Algorithms, Practical Challenges and Applications*. World Scientific, 2018.
- [9] C. Bauer and M. Schedl, "Global and country-specific mainstreamness measures: Definitions, analysis, and usage for improving personalized music recommendation systems," *PLOS ONE*, vol. 14, no. 6, e0217389, 2019.
- [10] D. Sanchez-Moreno, Y. Zheng, and M. N. Moreno-Garcia, "Time-Aware Music Recommender Systems: Modeling the Evolution of Implicit User Preferences and User Listening Habits in a Collaborative Filtering Approach," *Applied Sciences*, vol. 10, no. 15, 5324, 2020.
- [11] A. Caronongan and R. Cabredo, "Modeling User Music Preference Through Usage Scoring and User Listening Behavior for Generating Preferred Playlists," in *Lecture Notes in Computer Science*, vol. 9935, p. 73, 2016.
- [12] T. Pratiwi, A. Sunyoto, and D. Ariatmanto, "Music Genre Classification Using K-Nearest Neighbor and Mel-Frequency Cepstral Coefficients," *Sinkron*, vol. 8, no. 2, pp. 861-867, 2024.
- [13] A. Syah, B. Yuliadi, and R. Sahara, "Music Genre Classification using Naive Bayes Algorithm," *International Journal of Computer Trends and Technology*, vol. 62, no. 1, pp. 50-57, 2018.
- [14] I. M. De Diego, A. R. Redondo, R. R. Fernandez, J. Navarro, and J. M. Moguerza, "General Performance Score for classification problems," *Applied Intelligence*, vol. 52, no. 10, pp. 12049-12063, 2022.
- [15] D. J. Hand, P. Christen, and N. Kirielle, "F\*: An interpretable transformation of the F-measure," *Machine Learning*, vol. 110, no. 3, pp. 451-456, 2021.
- [16] B. Vrigazova, "The Proportion for Splitting Data into Training and Test Set for the Bootstrap in Classification Problems," *Business Systems Research Journal*, vol. 12, no. 1, pp. 228-242, 2021.