

Optimasi Sistem Rekomendasi Musik Berbasis *Naïve Bayes*: Studi Kasus pada Pengguna Musik di *Spotify*

Sri Lestari¹, Shabrina Sukma Wardana^{2*}

^{1,2*} Program Studi Sistem Informasi, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, Kota Jakarta Timur, Daerah Khusus Ibukota Jakarta, Indonesia.

Email: sri.lestari1203@gmail.com¹, shabisukma@gmail.com^{2*}

Histori Artikel:

Dikirim 30 Juli 2025; *Diterima dalam bentuk revisi* 15 Agustus 2025; *Diterima* 30 Agustus 2025; *Diterbitkan* 10 September 2025. Semua hak dilindungi oleh Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) STMIK Indonesia Banda Aceh.

Abstrak

Penelitian ini bertujuan untuk mengembangkan sistem rekomendasi musik berbasis algoritma *Naïve Bayes* dengan studi kasus pada pengguna *Spotify* di Indonesia. Data yang digunakan merupakan data primer hasil penyebaran kuesioner kepada 473 responden, mencakup variabel jenis kelamin, usia, dan frekuensi penggunaan *Spotify*. Genre musik dikelompokkan ke dalam dua kategori, yaitu Favorit Mayoritas (Pop, K-Pop, Jazz, R&B, Indie) dan Favorit Minoritas (Hip Hop, Rock, Religi, Dangdut, EDM, Daerah). Proses penelitian meliputi tahap pembersihan dan transformasi data, pembagian data menjadi 80% data latih dan 20% data uji, penerapan algoritma *Naïve Bayes*, serta evaluasi model menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil pengujian menunjukkan bahwa model mencapai akurasi 95,82%, dengan presisi 100% pada kategori Favorit Mayoritas dan 85,51% pada Favorit Minoritas, serta recall masing-masing sebesar 94,44% dan 100%. Nilai rata-rata F1-score berada pada kategori sangat baik, yang mengindikasikan model mampu memprediksi preferensi genre musik secara andal. Temuan ini menunjukkan bahwa sistem rekomendasi yang dihasilkan layak diimplementasikan untuk membantu pengguna menemukan musik sesuai karakteristik demografis dan perilaku penggunaan, sekaligus memberikan kontribusi terhadap pengembangan sistem rekomendasi berbasis data primer.

Kata Kunci: Algoritma *Naïve Bayes*; Preferensi Musik; Sistem Rekomendasi; *Spotify*; Klasifikasi Genre Musik.

Abstract

This study aims to develop a music recommendation system based on the *Naïve Bayes* algorithm, using *Spotify* users in Indonesia as a case study. The dataset was obtained from a questionnaire involving 473 respondents, covering variables such as gender, age, and frequency of *Spotify* usage. Music genres were grouped into two categories: Majority Favorites (Pop, K-Pop, Jazz, R&B, Indie) and Minority Favorites (Hip Hop, Rock, Religious, Dangdut, EDM, Regional). The research process included data cleaning and transformation, splitting the dataset into 80% training and 20% testing, applying the *Naïve Bayes* algorithm, and evaluating the model using accuracy, precision, recall, and F1-score metrics. The experimental results showed that the model achieved an accuracy of 95.82%, with 100% precision for the Majority Favorites category and 85.51% for the Minority Favorites category, along with recall values of 94.44% and 100%, respectively. The average F1-score was in the “very good” category, indicating that the model can reliably predict music genre preferences. These findings suggest that the resulting recommendation system is suitable for implementation to help users discover music aligned with their demographic characteristics and listening behavior, while also contributing to the development of recommendation systems based on primary data.

Keyword: *Naïve Bayes* Algorithm; Music Preference; Recommender System; *Spotify*; Music Genre Classification.

1. Pendahuluan

Musik telah menjadi bagian penting dalam kehidupan sehari-hari, terutama dengan adanya teknologi digital yang memudahkan akses ke berbagai lagu dari seluruh dunia. Platform streaming seperti *Spotify* memberikan pengguna kebebasan untuk mengakses jutaan lagu dari berbagai genre dan negara. Meskipun layanan ini menawarkan pilihan yang sangat beragam, banyaknya pilihan yang tersedia justru menciptakan tantangan bagi pengguna dalam menemukan musik yang sesuai dengan selera pribadi mereka. Hal ini tidak hanya menambah beban bagi pengguna, tetapi juga dapat membuat pengalaman mendengarkan musik menjadi kurang menyenangkan, karena mereka mungkin merasa kesulitan untuk menemukan lagu yang mereka sukai di tengah lautan pilihan yang ada. Dalam menghadapi tantangan ini, sistem rekomendasi menjadi sangat penting untuk membantu pengguna menemukan musik yang lebih sesuai dengan preferensi mereka. Tujuan dari sistem rekomendasi adalah untuk menyarankan lagu atau genre yang relevan berdasarkan data yang dikumpulkan dari kebiasaan atau karakteristik pengguna. Dengan menggunakan pendekatan ini, pengguna dapat lebih mudah menemukan musik yang mereka sukai, yang dapat meningkatkan kepuasan mereka terhadap platform streaming tersebut. Namun, membangun sistem rekomendasi yang efektif dan akurat bukanlah hal yang mudah, terutama ketika mempertimbangkan jumlah data yang sangat besar dan beragam yang perlu diproses untuk mencocokkan preferensi setiap individu. Sistem rekomendasi musik umumnya menganalisis sejumlah faktor, seperti riwayat pemutaran lagu, genre yang sering dipilih, durasi mendengarkan, serta kecenderungan pengguna untuk mengeksplorasi musik baru. Namun, meskipun sistem rekomendasi yang ada saat ini sudah menggunakan berbagai metode untuk menghasilkan saran yang relevan, sering kali hasil yang diberikan masih belum sesuai dengan harapan. Salah satu penyebabnya adalah ketidakmampuan sistem untuk memanfaatkan data yang lebih mendalam mengenai karakteristik individu pengguna, seperti usia, jenis kelamin, dan kebiasaan penggunaan, yang berperan penting dalam menentukan preferensi musik mereka. Oleh karena itu, diperlukan pendekatan yang lebih personal dan berdasarkan data yang lebih lengkap untuk meningkatkan akurasi rekomendasi yang diberikan kepada pengguna.

Penelitian ini bertujuan untuk mengembangkan sistem rekomendasi musik yang lebih akurat dan relevan dengan menggunakan algoritma *Naïve Bayes*. Algoritma ini dipilih karena kesederhanaannya, kemampuannya untuk menangani data dalam jumlah besar, serta efisiensinya dalam menghasilkan prediksi yang cukup akurat. *Naïve Bayes* mengandalkan prinsip *Teorema Bayes* yang mengasumsikan independensi antara fitur-fitur yang ada dalam data. Dengan prinsip ini, algoritma dapat dengan cepat menghasilkan prediksi berdasarkan data yang ada, meskipun data yang digunakan sangat besar dan kompleks. Penggunaan algoritma ini diharapkan dapat menghasilkan sistem rekomendasi yang lebih tepat, yang dapat memprediksi preferensi musik pengguna dengan lebih baik, berdasarkan data yang diperoleh langsung dari mereka. Berbeda dengan kebanyakan penelitian sebelumnya yang lebih banyak menggunakan dataset publik atau metode *collaborative filtering*, penelitian ini memanfaatkan data primer yang dikumpulkan langsung dari pengguna *Spotify* di Indonesia. Data yang dikumpulkan berupa kuesioner yang diisi oleh 473 responden, mencakup informasi mengenai usia, jenis kelamin, frekuensi penggunaan *Spotify*, serta genre musik yang mereka sukai. Berdasarkan data tersebut, musik kemudian dikelompokkan menjadi dua kategori utama: *Favorit Mayoritas*, yang mencakup genre musik populer seperti Pop, K-Pop, Jazz, R&B, dan Indie, serta *Favorit Minoritas*, yang meliputi genre yang lebih sedikit diminati seperti Hip Hop, Rock, Religi, Dangdut, EDM, dan musik Daerah (Senika *et al.*, 2022). Kategori-kategori ini dipilih berdasarkan jumlah responden yang memilih genre tertentu, sehingga dapat lebih menggambarkan preferensi musik pengguna di Indonesia. Hasil observasi awal menunjukkan bahwa sistem rekomendasi yang ada di *Spotify* belum sepenuhnya dapat mencocokkan rekomendasi musik dengan preferensi pengguna. Banyak dari rekomendasi yang muncul tidak sesuai dengan kebiasaan mendengarkan atau selera pribadi pengguna. Hal ini sering terjadi ketika informasi demografis, seperti jenis kelamin, usia, dan pola frekuensi penggunaan, tidak dimanfaatkan dengan optimal oleh sistem. Sebagai contoh, pengguna yang lebih sering mendengarkan musik dari genre tertentu mungkin tidak mendapatkan rekomendasi yang sesuai, meskipun data tersebut seharusnya

bisa memberikan petunjuk yang lebih baik mengenai preferensi mereka. Di sisi lain, sebagian besar sistem rekomendasi yang ada juga belum memanfaatkan data primer yang lebih menggambarkan kebiasaan mendengarkan musik pengguna secara langsung, yang tentu saja lebih representatif dan relevan dalam menghasilkan rekomendasi yang lebih personal. Fenomena ini membuka peluang untuk penelitian ini dalam mengembangkan model klasifikasi yang lebih akurat dengan memanfaatkan data demografis dan perilaku penggunaan yang lebih personal. Pendekatan berbasis data primer ini, yang langsung mengacu pada preferensi pengguna, dapat menghasilkan rekomendasi musik yang lebih tepat dan lebih relevan dengan karakteristik masing-masing individu. Dalam penelitian ini, sistem rekomendasi yang dikembangkan menggunakan algoritma *Naïve Bayes* diharapkan dapat meningkatkan akurasi rekomendasi musik dan memberikan pengalaman yang lebih memuaskan bagi pengguna. Dengan memanfaatkan informasi yang lebih lengkap dan data yang lebih mendalam tentang preferensi pengguna, model yang dikembangkan dalam penelitian ini bertujuan untuk memberikan saran musik yang lebih akurat dan relevan, serta meningkatkan kualitas pengalaman pengguna di *Spotify*. Penelitian ini juga diharapkan dapat memberikan kontribusi dalam pengembangan sistem rekomendasi musik di platform streaming lainnya, yang lebih mengutamakan personalisasi dan relevansi dalam setiap rekomendasi yang diberikan.

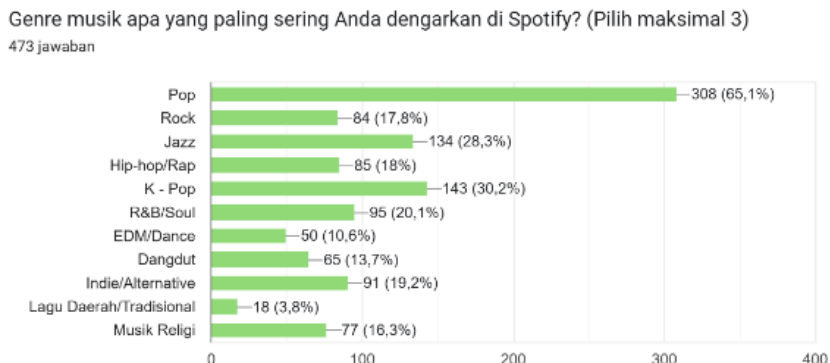
2. Metode Penelitian

Penelitian ini mengadopsi pendekatan kuantitatif dengan memanfaatkan data primer yang diperoleh melalui kuesioner daring. Kuesioner ini disebarluaskan kepada 473 pengguna aktif *Spotify* yang berada di Indonesia. Data yang dikumpulkan mencakup berbagai variabel demografis dan perilaku penggunaan, antara lain jenis kelamin, usia, frekuensi penggunaan *Spotify*, serta genre musik favorit yang didengarkan oleh responden. Genre musik yang dimasukkan dalam kuesioner terdiri dari beberapa kategori, yaitu Dangdut, EDM, HipHop, Indie, Jazz, KPop, Pop, RnB, Religi, Rock, dan musik Daerah, dengan opsi multi-label untuk memberikan keleluasaan kepada responden dalam memilih lebih dari satu genre musik yang mereka sukai. Variabel-variabel yang terkumpul tersebut digunakan untuk menganalisis preferensi musik individu, yang kemudian dijadikan dasar dalam membangun model rekomendasi berbasis algoritma *Naïve Bayes*. Pendekatan ini dipilih karena kemampuan *Naïve Bayes* dalam menangani data besar dengan efisien, serta kemampuannya untuk memprediksi kategori (genre musik) berdasarkan atribut-atribut yang ada, seperti jenis kelamin, usia, dan frekuensi penggunaan. Proses pengumpulan data ini bertujuan untuk memastikan bahwa model yang dibangun dapat memberikan rekomendasi yang lebih tepat dan relevan dengan kebiasaan serta preferensi musik yang dimiliki oleh pengguna *Spotify* di Indonesia.

Tabel 1. Atribut dan Deskripsi Dataset Penelitian

Atribut	Kategori	Keterangan
Jenis Kelamin	Perempuan, Laki-laki	Kategori gender responden
Usia	Muda: <18 dan 18-24 tahun Tua: 25-34, 35-44, dan >44 tahun	Digunakan sebagai fitur klasifikasi dalam system rekomendasi musik
Frekuensi Spotify	Sering: setiap hari, beberapa kali/minggu Jarang: beberapa kali/bulan, jarang	Menggambarkan intensitas penggunaan Spotify responden.
Favorit Mayoritas	Pop, K-Pop, Jazz, R&B, Indie	Genre populer yang banyak dipilih responden.
Favorit Minoritas	HipHop, Rock, Religi, Dangdut, EDM, Daerah	Genre dengan frekuensi pemilihan rendah dalam kuesioner.

Penelitian ini memanfaatkan data primer yang diperoleh dari 473 pengguna aktif *Spotify* di Indonesia, yang dikumpulkan melalui kuesioner daring. Responden diminta untuk mengisi informasi terkait dengan jenis kelamin, kategori usia, frekuensi penggunaan *Spotify*, serta genre musik favorit mereka. Data yang terkumpul kemudian melalui beberapa tahap pemrosesan, termasuk pembersihan data, pengkodean, dan transformasi agar dapat siap digunakan dalam proses analisis lebih lanjut. Proses pembersihan ini melibatkan penghapusan data yang tidak lengkap atau tidak valid, sementara pengkodean dilakukan untuk mengubah variabel kategorikal menjadi format yang dapat diproses oleh algoritma. Setelah data dipersiapkan, dataset kemudian dibagi menjadi dua bagian: 80% digunakan untuk pelatihan model dan 20% untuk pengujian. Pembagian ini bertujuan untuk memastikan bahwa model yang dibangun dapat diuji secara valid dengan data yang belum pernah dilihat sebelumnya, sehingga dapat dievaluasi kinerjanya secara objektif. Pemodelan dilakukan dengan menggunakan algoritma *Naïve Bayes* yang diterapkan melalui perangkat lunak *RapidMiner*. Algoritma ini dipilih karena kemampuannya untuk menangani data dalam jumlah besar secara efisien serta memberikan hasil prediksi yang akurat berdasarkan karakteristik demografis dan perilaku pengguna. Dalam proses pemodelan ini, *Naïve Bayes* digunakan untuk memprediksi genre musik favorit berdasarkan atribut-atribut responden, seperti jenis kelamin, usia, dan frekuensi penggunaan *Spotify*. Untuk mengevaluasi kinerja model yang dibangun, digunakan berbagai metrik, antara lain akurasi, presisi, recall, dan F1-score. Metrik-metrik ini digunakan untuk mengukur seberapa baik model dapat memprediksi genre musik yang sesuai dengan preferensi pengguna. Selain itu, *confusion matrix* juga digunakan untuk menganalisis lebih mendalam tentang distribusi prediksi yang benar dan salah dalam setiap kategori genre. Evaluasi menggunakan metrik-metrik ini bertujuan untuk memastikan bahwa model yang dihasilkan tidak hanya akurat, tetapi juga dapat diandalkan dalam memberikan rekomendasi musik yang lebih personal dan relevan bagi pengguna.



Gambar 1. Diagram Genre Favorit Pengguna Spotify

Pengujian dilakukan untuk menilai efektivitas algoritma *Naïve Bayes* dalam merekomendasikan genre musik berdasarkan data pengguna *Spotify*. Proses pengujian ini meliputi beberapa tahapan penting, dimulai dengan pengumpulan data primer yang telah dijelaskan sebelumnya. Selanjutnya, data yang terkumpul menjalani tahap prapemrosesan, yang mencakup pembersihan data, pengkodean, dan transformasi agar siap untuk diproses lebih lanjut dalam model. Setelah prapemrosesan selesai, data dibagi menjadi dua bagian utama: 80% digunakan sebagai data latih untuk membangun model, sementara 20% sisanya digunakan sebagai data uji untuk mengevaluasi performa model. Pembagian data ini penting untuk memastikan bahwa model yang dikembangkan dapat diuji secara objektif menggunakan data yang belum pernah dilihat sebelumnya, guna menghindari overfitting dan memberikan gambaran yang lebih akurat tentang kemampuannya dalam memprediksi genre musik berdasarkan data pengguna yang baru.

Untuk mengevaluasi kinerja model, digunakan beberapa metrik evaluasi yang umum digunakan dalam pengujian model klasifikasi. Metrik-metrik tersebut meliputi akurasi, presisi, recall, F1-score, serta *confusion matrix*. Akurasi digunakan untuk mengukur sejauh mana model dapat memberikan prediksi yang benar secara keseluruhan, sementara presisi dan recall mengukur seberapa baik model dalam mengidentifikasi genre musik yang relevan dan meminimalisir kesalahan klasifikasi, terutama dalam kategori yang kurang sering dipilih. Nilai F1-score, yang merupakan rata-rata harmonis antara presisi dan recall, digunakan untuk memberikan gambaran umum tentang keseimbangan kinerja model. Selain itu, *confusion matrix* digunakan untuk memvisualisasikan distribusi kesalahan antara prediksi yang benar dan salah pada setiap kategori genre musik, yang membantu dalam menganalisis area yang perlu perbaikan. Keberhasilan model diukur berdasarkan beberapa kriteria, antara lain tingginya akurasi yang dicapai, relevansi fitur yang digunakan dalam model, serta konsistensi performa model dalam menghasilkan rekomendasi yang tepat untuk berbagai kategori genre. Evaluasi ini akan memastikan bahwa model yang dibangun tidak hanya memiliki akurasi yang tinggi, tetapi juga dapat memberikan rekomendasi musik yang sesuai dengan preferensi pengguna secara lebih personal, berdasarkan karakteristik demografis dan kebiasaan mendengarkan mereka.

3. Hasil dan Pembahasan

3.1 Hasil

Penggunaan alat yang tepat sangat penting untuk memastikan kelancaran proses penelitian serta akurasi hasil yang diperoleh. Dalam penelitian ini, berbagai perangkat keras dan perangkat lunak digunakan untuk mendukung setiap tahap analisis, mulai dari pengumpulan data hingga pemodelan dengan algoritma *Naïve Bayes*.

Tabel 2. Tabel Spesifikasi Alat Penelitian

Jenis	Spesifikasi / Fungsi
Hardware	Laptop Lenovo IdeaPad 3, Intel Core i3-10110U, 12GB RAM, Windows 10 64-bit
Software	a. Google Forms (kuisisioner online) b. Google Spreadsheet (penyimpanan & rekap data real-time) c. Microsoft Excel (pembersihan & rekap data real-time) d. RapidMiner Studio V.9.10 (analisis & klasifikasi) e. Windows 10 (system operasi utama)

3.1.1 Implementasi dan Pengujian

Implementasi yang efektif dan pengujian yang mendalam sangat penting untuk memastikan bahwa model yang dikembangkan dapat memberikan rekomendasi musik yang akurat dan relevan sesuai dengan karakteristik pengguna.

1) Pengumpulan Data

Data penelitian diperoleh dari kuisisioner online yang mengumpulkan preferensi musik pengguna Spotify (jenis kelamin, usia, frekuensi penggunaan) dengan total 473 responden untuk membangun sistem rekomendasi berbasis *Naïve Bayes*.

Item	Age	Gender	Frequency	Genre	Preference
Tina Turner	25-34 Tahun	Perempuan	Sering	Pop, Jazz, EDM	Meningkatkan Mood, Menemani Belajar/Bekerja
Angela Widada	18-24 Tahun	Laki-laki	Sering	Pop, Jazz, EDM	Meningkatkan Mood, Menemani Belajar/Bekerja
Wika Laras	25-34 Tahun	Perempuan	Sering	Pop, Jazz, EDM	Meningkatkan Mood, Menemani Belajar/Bekerja
Silvan NIT	18-24 Tahun	Laki-laki	Sering	Pop, Jazz, EDM	Meningkatkan Mood, Menemani Belajar/Bekerja
Tanti	25-34 Tahun	Perempuan	Sering	Pop, Jazz, EDM	Meningkatkan Mood, Menemani Belajar/Bekerja
Raj	18-24 Tahun	Laki-laki	Sering	Pop, Jazz, EDM	Meningkatkan Mood, Menemani Belajar/Bekerja
Rud Adip Almal	18-24 Tahun	Laki-laki	Sering	Pop, Jazz, EDM	Meningkatkan Mood, Menemani Belajar/Bekerja
Chandra Dewi	35-44 Tahun	Perempuan	Sering	Pop, Jazz, EDM	Meningkatkan Mood, Menemani Belajar/Bekerja
Silvan NIT	18-24 Tahun	Laki-laki	Sering	Pop, Jazz, EDM	Meningkatkan Mood, Menemani Belajar/Bekerja
sahifa afa	25-34 Tahun	Perempuan	Sering	Pop, Jazz, EDM	Meningkatkan Mood, Menemani Belajar/Bekerja
Tuti Alwanjari	18-24 Tahun	Laki-laki	Sering	Pop, Jazz, EDM	Meningkatkan Mood, Menemani Belajar/Bekerja

Gambar 2. Hasil Kuisisioner Pengguna Spotify

2) Proses Preprocessing Data

Data melalui tahap preprocessing dimulai dengan pembersihan, yaitu menghapus 13 data kosong dan mengubah data genre agar setiap baris hanya mewakili satu genre. Selain itu, hanya atribut jenis kelamin, usia, frekuensi penggunaan Spotify, dan genre musik yang digunakan. Selanjutnya, dilakukan transformasi data dengan mengelompokkan usia menjadi dua kategori (muda dan tua), frekuensi penggunaan Spotify (sering dan jarang), serta genre musik berdasarkan popularitas menjadi favorit mayoritas dan minoritas. Transformasi ini bertujuan agar data lebih terstruktur dan model klasifikasi dapat bekerja lebih optimal.

No	Jenis_Kelamin	Usia	Frekuensi_Spotify	Genre_Musik
1	Perempuan	Muda	Jarang	Favorit Minoritas
2	Perempuan	Muda	Jarang	Favorit Minoritas
3	Laki-laki	Muda	Sering	Favorit Minoritas
4	Laki-laki	Muda	Jarang	Favorit Mayoritas
5	Laki-laki	Muda	Sering	Favorit Minoritas
6	Laki-laki	Muda	Sering	Favorit Minoritas
7	Laki-laki	Muda	Sering	Favorit Minoritas
8	Perempuan	Muda	Sering	Favorit Mayoritas
9	Perempuan	Muda	Jarang	Favorit Mayoritas
10	Laki-laki	Tua	Jarang	Favorit Minoritas
11	Perempuan	Muda	Sering	Favorit Minoritas
12	Laki-laki	Muda	Sering	Favorit Mayoritas
13	Perempuan	Muda	Sering	Favorit Minoritas
14	Perempuan	Tua	Sering	Favorit Mayoritas
15	Laki-laki	Muda	Sering	Favorit Mayoritas
1185	Laki-laki	Muda	Sering	Favorit Minoritas
1186	Laki-laki	Muda	Sering	Favorit Mayoritas
1187	Perempuan	Muda	Sering	Favorit Minoritas
1188	Laki-laki	Muda	Sering	Favorit Mayoritas
1189	Perempuan	Muda	Sering	Favorit Mayoritas
1190	Laki-laki	Muda	Sering	Favorit Minoritas
1191	Laki-laki	Muda	Sering	Favorit Mayoritas
1192	Laki-laki	Muda	Sering	Favorit Mayoritas

Gambar 3. Dataset yang Sudah di Cleaning dan Transformasi

Setelah pembersihan dan transformasi, dataset dibagi menjadi dua bagian, yaitu data latih sebanyak 80% (953 data) yang digunakan untuk membangun model, dan data uji sebanyak 20% (239 data) untuk mengevaluasi performa model terhadap data baru. Pembagian ini memastikan model dapat mengenali pola dengan baik dan diuji secara valid sebelum digunakan secara luas.

The image shows two side-by-side screenshots of a data analysis software interface. The left screenshot is titled 'ExampleSet (/Local Repository/Data Lath)' and shows a table with 15 rows of data. The right screenshot is titled 'ExampleSet (/Local Repository/Data Uji)' and shows a table with 15 rows of data. Both tables have columns for Row No., Genre_Musik, Jenis_Kelamin, Usia, and Frekuensi_Spotify. The data in both tables is consistent with the data shown in Gambar 3.

Gambar 4. Data Latih dan Data Uji

3) Penerapan Algoritma Naïve Bayes

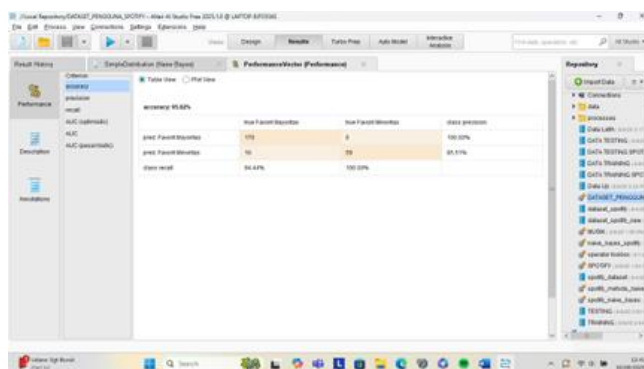
Setelah data selesai dipreprocessing dan dibagi menjadi data latih (80%) dan data uji (20%) menggunakan Microsoft Excel, kedua file tersebut diimpor ke *RapidMiner*. Pada *RapidMiner*, data latih digunakan dengan algoritma Naïve Bayes untuk membangun model klasifikasi berdasarkan label favorit mayoritas dan minoritas. Model yang dibuat kemudian diterapkan pada data uji menggunakan operator *Apply Model*, dan hasilnya dievaluasi dengan operator Performance untuk mengukur akurasi dan kinerja model dalam memprediksi data baru yang belum pernah dilatih sebelumnya.



Gambar 5. Proses Implementasi Algoritma Naive Bayes di RapidMiner

4) Evaluasi Model

Evaluasi model merupakan langkah krusial dalam mengukur sejauh mana sistem rekomendasi yang dibangun mampu berfungsi sesuai dengan tujuan yang telah ditetapkan, yaitu memberikan prediksi yang akurat mengenai preferensi musik pengguna. Pada tahap ini, hasil yang diperoleh dari penerapan algoritma *Naïve Bayes* dievaluasi menggunakan berbagai metrik untuk menilai kinerja model dalam memprediksi genre musik yang relevan.

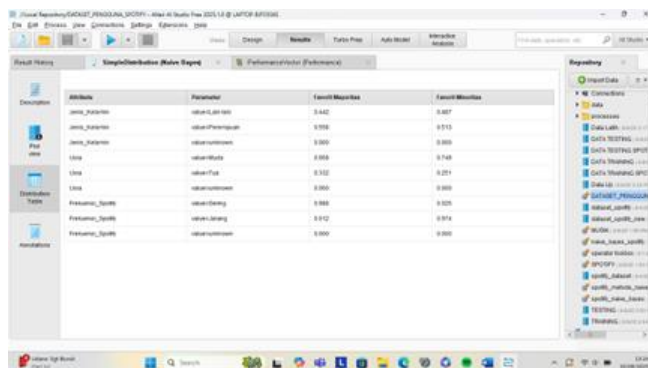


Gambar 6. Hasil Implementasi RapidMiner

Hasil implementasi algoritma *Naïve Bayes* pada dataset penelitian menunjukkan nilai akurasi sebesar 95,82%, yang berarti model mampu mengklasifikasikan data dengan tingkat ketepatan yang sangat tinggi. Pada kategori Favorit Mayoritas, model berhasil mengklasifikasikan seluruh 170 data secara benar, menghasilkan *class precision* sebesar 100% dan *class recall* sebesar 94,44%. Sementara pada kategori Favorit Minoritas, model berhasil mengklasifikasikan 59 data dengan benar dari total data uji, menghasilkan *class precision* sebesar 85,51% dan *class recall* sebesar 100%. Hasil ini menunjukkan bahwa algoritma *Naïve Bayes* memiliki kemampuan yang sangat baik dalam membedakan kedua kategori label, dengan tingkat kesalahan prediksi yang rendah.

5) Hasil Akhir Pengujian

Hasil akhir pengujian merupakan tahap yang menunjukkan sejauh mana model rekomendasi yang dikembangkan dapat memenuhi tujuan yang telah ditetapkan, yakni memberikan prediksi yang akurat dan relevan mengenai preferensi genre musik pengguna. Setelah melalui proses pelatihan dan pengujian menggunakan algoritma *Naïve Bayes*, hasil yang diperoleh menunjukkan kinerja model dalam memprediksi genre musik berdasarkan data yang tersedia.



Gambar 7. Distribusi Probabilitas Atribut pada Model Naive Bayes untuk Prediksi Genre Musik

Berdasarkan hasil pengujian algoritma Naive Bayes, distribusi probabilitas atribut terhadap dua kelas, Favorit Mayoritas dan Favorit Minoritas, menunjukkan variasi pada setiap atribut. Untuk jenis kelamin, pengguna perempuan memiliki probabilitas sedikit lebih tinggi untuk Favorit Mayoritas (0,558) dibanding laki-laki (0,442), sedangkan untuk Favorit Minoritas, laki-laki memiliki probabilitas 0,487 dan perempuan 0,513. Tidak ada data dengan jenis kelamin tidak diketahui. Pada atribut usia, pengguna muda cenderung memiliki probabilitas lebih besar untuk kedua kelas dibanding pengguna tua, dengan probabilitas 0,668 dan 0,749 untuk Favorit Mayoritas dan Minoritas pada usia muda, sementara usia tua lebih rendah pada kedua kelas. Atribut frekuensi penggunaan Spotify menjadi faktor paling signifikan, di mana pengguna yang sering menggunakan Spotify sangat besar kemungkinannya masuk Favorit Mayoritas (0,988), dan pengguna yang jarang menggunakan Spotify hampir pasti masuk Favorit Minoritas (0,974).

3.2 Pembahasan

Penelitian bertujuan mengembangkan sistem rekomendasi musik yang lebih personal dengan algoritma *Naive Bayes* pada pengguna *Spotify* di Indonesia. Hasil pengujian menunjukkan performa model yang memuaskan dengan akurasi mencapai 95,82% serta presisi tinggi pada genre musik mayoritas. Hal ini mengindikasikan bahwa *Naive Bayes*, yang terkenal karena efisiensinya dalam mengolah data besar, sangat efektif untuk klasifikasi genre musik berdasarkan data demografis dan perilaku pengguna (Akbar *et al.*, 2025; Hendri *et al.*, 2024). Frekuensi penggunaan *Spotify* terbukti memiliki pengaruh besar terhadap genre musik yang dipilih. Pengguna yang sering mengakses platform ini lebih cenderung memilih genre yang lebih populer, sementara mereka yang jarang menggunakan *Spotify* lebih memilih genre yang kurang dominan. Temuan tersebut sejalan dengan penelitian sebelumnya yang mengaitkan pola penggunaan aplikasi dengan pilihan konten (Fauzi, n.d.; Ramadlan As'ari *et al.*, 2025). Selain itu, data demografis seperti jenis kelamin dan usia juga mempengaruhi preferensi genre musik. Pengguna perempuan lebih banyak memilih genre mainstream seperti *Pop* dan *K-Pop*, sementara pria cenderung memiliki pilihan genre yang lebih beragam, termasuk *Hip Hop* dan *Rock*. Temuan ini sesuai dengan studi yang menunjukkan bahwa faktor demografis berperan penting dalam prediksi preferensi musik (Hendri *et al.*, 2024). Penggunaan data primer dalam penelitian memberikan keunggulan dengan menyesuaikan sistem rekomendasi pada karakteristik pengguna *Indonesia*, yang memberikan hasil lebih relevan. Dibandingkan dengan banyak penelitian sebelumnya yang mengandalkan dataset publik atau metode *collaborative filtering*, data yang diperoleh langsung dari pengguna *Spotify* di Indonesia memberikan gambaran lebih akurat mengenai preferensi lokal (Luqman *et al.*, 2021; Tannady *et al.*, n.d.). Namun, terdapat beberapa keterbatasan dalam penelitian ini, seperti penggunaan data hanya dari pengguna *Indonesia*, yang mungkin tidak sepenuhnya mewakili preferensi global. Selain itu, model hanya mempertimbangkan data demografis dan frekuensi penggunaan, padahal faktor lain seperti lokasi, mood, atau aktivitas pengguna saat mendengarkan musik dapat memperkaya analisis (Febrian *et al.*, 2025; Rizki Rinaldi *et al.*, 2024).

Oleh karena itu, penelitian lebih lanjut yang melibatkan variabel tambahan serta penggunaan algoritma lain seperti *Random Forest* atau *Deep Learning* akan meningkatkan akurasi dan ketepatan sistem rekomendasi (Maulana *et al.*, 2025; Surapati *et al.*, 2023). Temuan dari penelitian ini membuka peluang bagi pengembangan sistem rekomendasi musik yang lebih relevan dan terpersonalisasi, dengan memanfaatkan data yang lebih beragam dan lengkap. Pengembangan lebih lanjut dari teknologi ini dapat meningkatkan pengalaman pengguna *Spotify* dalam menemukan musik yang sesuai dengan preferensi mereka.

4. Kesimpulan

Model sistem rekomendasi musik yang dikembangkan dengan algoritma *Naïve Bayes* berhasil mengklasifikasikan preferensi genre musik ke dalam dua kategori utama, yaitu Favorit Mayoritas dan Favorit Minoritas, berdasarkan data demografis pengguna seperti jenis kelamin, usia, dan frekuensi penggunaan *Spotify*. Preferensi genre musik terbukti bervariasi sesuai dengan karakteristik pengguna, di mana jenis kelamin, usia, dan frekuensi penggunaan memainkan peran yang signifikan dalam menentukan kategori genre yang diminati. Model ini menunjukkan performa yang sangat baik dengan akurasi 95,82%, presisi tinggi pada kedua kategori genre, serta recall yang optimal. Hasil tersebut membuktikan bahwa *Naïve Bayes* efektif dalam membangun sistem rekomendasi musik yang memanfaatkan data demografis dan perilaku pengguna. Dengan demikian, model ini layak diimplementasikan dalam aplikasi nyata untuk memberikan rekomendasi musik yang lebih sesuai dengan preferensi pribadi pengguna. Untuk penelitian lanjutan, disarankan untuk menambah variabel lain seperti lokasi, mood, dan aktivitas saat mendengarkan musik, yang dapat memperkaya rekomendasi. Selain itu, penelitian lebih lanjut bisa mencakup pengujian algoritma lain seperti *Random Forest* atau *Deep Learning* untuk membandingkan hasil dan meningkatkan akurasi. Penggunaan dataset yang lebih besar dan lebih beragam, termasuk data dari *Spotify API*, juga dapat membantu meningkatkan kualitas model. Selanjutnya, pengembangan model yang mampu memprediksi genre musik secara lebih spesifik akan memberikan rekomendasi yang lebih detail dan terpersonalisasi.

5. Ucapan Terima Kasih

Segala puji syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat, taufik, dan hidayah-Nya sehingga jurnal penelitian ini dapat terselesaikan dengan baik. Penulis menyampaikan terima kasih yang sebesar-besarnya kepada kedua orang tua tercinta atas doa, dukungan, dan kasih sayang yang tiada henti. Ucapan terima kasih juga penulis sampaikan kepada dosen pembimbing yang telah memberikan arahan, bimbingan, dan ilmu yang sangat berharga selama proses penelitian ini. Tak lupa, terima kasih kepada keluarga serta teman-teman yang senantiasa memberikan semangat, dukungan, dan bantuan dalam berbagai bentuk. Penulis menyadari bahwa banyak pihak yang telah membantu namun tidak dapat disebutkan satu per satu, semoga Allah SWT membalas segala kebaikan dan ketulusan mereka dengan pahala yang berlipat ganda.

6. Daftar Pustaka

- Amaia, E., Izzah, A. N., & Risal, A. A. N. Klasifikasi Penyalahgunaan Pesan singkat Menggunakan Algoritma Naïve Bayes.
- Andry, J. F., Wandy, W., & Lee, F. S. (2025). Analisis Musik Terhadap Kesehatan Mental Menggunakan Teknik Classification. *INFORMASI (Jurnal Informatika dan Sistem Informasi)*, 17(1), 40-52. <https://doi.org/10.37424/informasi.v17i1.334>.
- Arifin, A. A. A., Handoko, W., & Efendi, Z. (2022). Implementasi Metode Naive Bayes Untuk Klasifikasi Penerima Program Keluarga Harapan. *J-Com (Journal of Computer)*, 2(1), 21-26.
- As' ari, W. R., Arifin, M., Fithri, D. L., & Setiaji, P. (2025). ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI SPOTIFY DI GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA NAIVE BAYES. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(2), 3600-3607.
- Ginabila, G., & Fauzi, A. (2023). Analisis Sentimen Terhadap Pemutar Musik Online Spotify Dengan Algoritma Naive Bayes dan Support Vector Machine. *Jurnal Ilmiah ILKOMINFO-Ilmu Komputer & Informatika*, 6(2), 111-122. <https://doi.org/10.47324/ilkominfo.v6i2.180>.
- Hendri, D., Nadha, D., Basri, F. K., Wajdi, M. F., & Nadhirah, N. (2024). Comparison of Decision Tree Algorithm, Naive Bayes, K-Nearest Neighbors on Spotify Music Genre. *IJATIS: Indonesian Journal of Applied Technology and Innovation Science*, 1(1), 47-53.
- Jatsiah, S., & Usman, M. S. (2025). Preferensi Genre Music Menggunakan Algoritma Naive Bayes. *TEKNO: Jurnal Penelitian Teknologi dan Peradilan*, 3(1), 15-21.
- Julianto, M. R., Akbar, Y., & Wahyudi, T. (2024). Analisis sentimen respon publik terhadap program internet gratis di platform X melalui pendekatan algoritma Naïve Bayes. *Jurnal Indonesia: Manajemen Informatika dan Komunikasi*, 5(3), 2940-2950.
- Maulana, B., Febrian, D., Fachrezi, I. R., & Zeen, M. F. (2025). Comparison of Support Vector Machine, Random Forest, and C4. 5 Algorithms for Customer Loss Prediction. *IJATIS: Indonesian Journal of Applied Technology and Innovation Science*, 2(1), 1-6. <https://doi.org/10.57152/ijatiss.v2i1.1102>.
- Navisa, S., Hakim, L., & Nabilah, A. (2021). Komparasi Algoritma Klasifikasi Genre Musik pada Spotify Menggunakan CRISP-DM. *Jurnal Sistem Cerdas*, 4(2), 114-125. <https://doi.org/10.37396/jsc.v4i2.162>.
- Pamungkas, A. (2022). *Analisis performansi jaringan Fiber to The Home berbasis Gigabit Capable Passive Optical di IZI Network Sekeloa Utara* (Doctoral dissertation, Universitas Komputer Indonesia).
- Putro, H. F., Vlandari, R. T., & Saptomo, W. L. Y. (2020). Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan. *Jurnal Teknologi Informasi dan Komunikasi (TIKoSIN)*, 8(2). <http://dx.doi.org/10.30646/tikomsin.v8i2.500>.
- Ramdhani, M. R. F., & Lhaksana, K. M. (2023). Analisis Sentimen Menggunakan Metode Naive Bayes dan Support Vector Machine pada Ulasan Aplikasi Spotify. *eProceedings of Engineering*, 10(3).

- Riskawati, R., Fatihanursari, F., Iin, I., & Rinaldi, A. R. (2024). Penerapan Metode Naïve Bayes Classifier Pada Analisis Sentimen Aplikasi Gopay. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 346-353.
- Senika, A., Rasiban, R., & Iskandar, D. (2022). Implementasi metode Naïve Bayes dalam penilaian kinerja sales marketing pada PT. Pachira Distrinusa. *Jurnal Media Informatika Budidarma*, 6(1), 701.
- Surapati, U., & Zulkarnain, A. Y. (2023). Implementasi metode Naïve Bayes untuk mendeteksi hate speech pada Twitter. *INTECOMS J. Inf. Technol. Comput. Sci*, 6(2), 830-837.
- Tannady, H., Andry, J. F., & Honni, H. (2024). Analisis Big Data Spotify dengan Metode Data Mining. *JBASE-Journal of Business and Audit Information Systems*, 7(2).
- Triyono, A., & Faqih, A. (2025). Implementation of the Naive Bayes Method in Sentiment Analysis of Spotify Application Reviews. *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, 4(2), 1091-1097. <https://doi.org/10.59934/jaiea.v4i2.824>.