



Analisis *Machine Learning* pada Data Netflix Shows untuk Mengklasifikasikan Tren Genre dan Karakteristik Film

Claudia K. Hamsi¹, I Wayan Sudiarsa^{2*}, Vinsensia P.K Abu³, Sarling C. Dhai⁴,
Maria A. Serero⁵

^{1,3,4,5} Fakultas Bisnis dan Desain Kreatif, Program Studi Bisnis Digital,
Institut Bisnis dan Teknologi Indonesia, Indonesia

²Fakultas Teknik dan Informatika, Program Studi Rekayasa Sitem Komputer,
Institut Bisnis dan Teknologi Indonesia, Indonesia

*Penulis Korespondensi: sudiarsa@instiki.ac.id

Abstract. *The rapid development of digital streaming platforms such as Netflix has generated a large volume of content data with diverse characteristics, thereby requiring effective analytical methods to understand emerging patterns and trends. This study aims to classify Netflix content into two main categories, namely movies and television shows, and to analyze genre trends and content characteristics using a data mining approach with the Naive Bayes algorithm. The dataset used in this study is the Netflix Shows dataset, consisting of 8,809 content entries, with the primary features analyzed including genre, rating, and country of production. The research process begins with data exploration and preprocessing stages, including data cleaning, handling missing values, and transforming categorical features to enable effective model construction. Subsequently, the dataset is divided into training and testing sets to objectively and systematically build and evaluate the Naive Bayes classification model. Model performance is evaluated using accuracy, precision, recall, and F1-score metrics to assess the model's ability to accurately distinguish between Netflix content types. The experimental results demonstrate that the Naive Bayes algorithm is able to classify Netflix content into Movie and TV Show categories with accuracy, precision, recall, and F1-score values of 100%, respectively. The confusion matrix indicates that no misclassification occurred, suggesting that genre, rating, and country of production features provide a very clear separation between content classes. These findings indicate that the Naive Bayes algorithm can achieve exceptionally high classification performance with optimal evaluation results. The results further reveal distinct differences in characteristics between movies and television shows based on genre and production attributes. Therefore, this study is expected to contribute to the development of content recommendation systems and strategic content management within the streaming industry.*

Keywords: *Content Classification; Data Mining; Genre Analysis; Naive Bayes; Netflix.*

Abstrak. Perkembangan platform streaming digital seperti Netflix menghasilkan data konten dalam jumlah besar dengan karakteristik yang beragam, sehingga memerlukan metode analisis yang efektif untuk memahami pola dan tren yang terbentuk. Penelitian ini bertujuan untuk mengklasifikasikan konten Netflix ke dalam dua kategori utama, yaitu film (Movie) dan acara televisi (TV Show), serta menganalisis tren genre dan karakteristik konten menggunakan pendekatan data mining dengan algoritma Naive Bayes. Dataset yang digunakan merupakan dataset Netflix Shows yang terdiri dari 8.809 entri konten, dengan fitur utama yang dianalisis meliputi genre, rating, dan negara produksi. Tahapan penelitian diawali dengan proses eksplorasi dan prapemrosesan data, termasuk pembersihan data, penanganan nilai kosong, serta transformasi fitur kategorikal agar dapat digunakan dalam proses pemodelan. Selanjutnya, data dibagi menjadi data latih dan data uji untuk membangun serta mengevaluasi model klasifikasi Naive Bayes secara objektif dan terukur. Evaluasi kinerja model dilakukan menggunakan metrik akurasi, presisi, recall, dan F1-score untuk menilai kemampuan model dalam membedakan tipe konten Netflix secara tepat. Hasil pengujian menunjukkan bahwa algoritma Naive Bayes mampu mengklasifikasikan konten Netflix ke dalam kategori Movie dan TV Show dengan tingkat akurasi, presisi, recall, dan F1-score masing-masing sebesar 100%. Confusion matrix menunjukkan tidak adanya kesalahan klasifikasi, yang mengindikasikan bahwa fitur genre, rating, dan negara produksi memiliki pemisahan kelas yang sangat jelas. Hasil penelitian menunjukkan bahwa algoritma Naive Bayes mampu menghasilkan kinerja klasifikasi yang sangat tinggi dengan nilai evaluasi optimal. Temuan ini mengindikasikan adanya perbedaan karakteristik yang jelas antara film dan acara TV berdasarkan genre dan atribut produksinya. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi bagi pengembangan sistem rekomendasi konten serta strategi pengelolaan industri streaming.

Kata kunci: Analisis Genre; Data Mining; Klasifikasi; Naive Bayes; Netflix.

1. LATAR BELAKANG

Perkembangan platform streaming digital seperti Netflix telah menghasilkan volume data yang sangat besar, baik berupa metadata konten maupun data interaksi pengguna. Data tersebut berpotensi dimanfaatkan dalam berbagai analisis berbasis data, salah satunya melalui pengelompokan dan klasifikasi konten guna mendukung pengambilan keputusan yang lebih efektif. Klasifikasi konten, seperti pemisahan antara film (*movie*) dan serial televisi (*TV show*), merupakan langkah awal yang penting dalam pengolahan data pada platform hiburan digital (Zulkarnain et al., 2024).

Dalam bidang *data mining* dan *machine learning*, teknik klasifikasi digunakan untuk mengelompokkan data ke dalam kelas tertentu berdasarkan karakteristiknya. Salah satu algoritma yang banyak digunakan adalah Naive Bayes, karena memiliki struktur sederhana, efisiensi komputasi yang baik, serta performa yang kompetitif pada data berukuran besar dan berdimensi tinggi (Mahendra et al., 2024). Pendekatan probabilistik yang digunakan menjadikan algoritma ini sesuai untuk klasifikasi data kategorikal.

Video-on-Demand (VoD) seperti Netflix menawarkan ribuan judul film dan acara televisi dari berbagai negara dan genre, sehingga pengelolaan dan analisis data konten menjadi semakin kompleks. Pemahaman terhadap interaksi pengguna dan tren konsumsi konten merupakan tantangan utama dalam pengembangan sistem rekomendasi dan analisis perilaku pengguna (Noroozian et al., 2024). Berbagai pendekatan telah dikembangkan untuk mendukung analisis konten digital, termasuk segmentasi dan pengelompokan konten video (Rotman et al., 2020).

Beberapa penelitian terdahulu menunjukkan bahwa algoritma Naive Bayes mampu memberikan hasil yang baik dalam mengklasifikasikan data berbasis konten dan ulasan pengguna pada platform digital. Rieuwpassa et al. (2023) serta Bintang dan Romadloni (2024) melaporkan tingkat akurasi yang memadai dalam klasifikasi ulasan aplikasi Netflix. Temuan serupa juga ditunjukkan oleh Jusia et al. (2025) dan Shafira dan Nugraha (2025) dalam analisis sentimen layanan Netflix.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengklasifikasikan konten Netflix menjadi film dan acara televisi menggunakan algoritma Naive Bayes, serta mengevaluasi kinerja algoritma tersebut dalam menangani data kategorikal berskala besar.

2. KAJIAN TEORITIS

Machine Learning

Machine Learning merupakan cabang dari kecerdasan buatan (*Artificial Intelligence*) yang memberikan kemampuan kepada sistem komputer untuk belajar dari data serta meningkatkan kinerjanya tanpa perlu diprogram secara eksplisit. Menurut Jordan dan Mitchell (2015), sebuah sistem dikatakan belajar ketika kinerjanya dalam menyelesaikan suatu tugas meningkat seiring pengalaman yang diperoleh dari data. Penelitian ini menerapkan machine learning untuk mengidentifikasi pola, tren genre, serta karakteristik film berdasarkan data yang tersedia di platform Netflix.

Klasifikasi dalam *Machine Learning*

Klasifikasi adalah salah satu teknik dalam supervised learning yang bertujuan untuk mengelompokkan data ke dalam kategori tertentu berdasarkan atribut yang dimiliki. Algoritma klasifikasi bekerja dengan mempelajari data latih yang sudah memiliki label, kemudian digunakan untuk memprediksi label pada data baru. Dalam konteks penelitian ini, klasifikasi digunakan untuk mengelompokkan film atau acara Netflix berdasarkan genre dan karakteristik tertentu, seperti jenis tayangan, durasi, dan tahun rilis.

Data Mining dan Analisis Data

Data mining merupakan proses penggalian informasi dan pola yang berguna dari sekumpulan data berukuran besar. Tahapan data mining mencakup pembersihan data, transformasi data, pemodelan, dan evaluasi hasil. Salah satu teknik utama dalam data mining adalah klasifikasi, yaitu proses penentuan kelas suatu data berdasarkan atribut tertentu yang dimilikinya. Klasifikasi banyak digunakan dalam berbagai bidang, termasuk pengolahan data teks, analisis sentimen, dan sistem rekomendasi (Aggarwal, 2015).

Analisis data Netflix dilakukan untuk menggali informasi terkait tren genre film dan acara televisi yang berkembang dari waktu ke waktu. Dengan demikian, data mining berperan krusial sebagai dasar dalam pengambilan keputusan berbasis data.

Algoritma Naive Bayes

Naive Bayes merupakan algoritma klasifikasi berbasis probabilitas yang didasarkan pada Teorema Bayes dengan asumsi independensi antar fitur. Meskipun asumsi tersebut tergolong sederhana, algoritma Naive Bayes terbukti efektif dalam berbagai permasalahan klasifikasi, khususnya pada data teks dan data kategorikal (Mahendra et al., 2024). Keunggulan utama algoritma ini terletak pada kemudahan implementasi, kecepatan proses komputasi, serta kebutuhan sumber daya yang relatif rendah.

Dalam praktiknya, Naive Bayes sering digunakan sebagai *baseline model* dalam penelitian klasifikasi karena kemampuannya menghasilkan performa yang stabil. Penelitian oleh Wardani et al. (2024) dan Yanuargi et al. (2024) menunjukkan bahwa Naive Bayes mampu bersaing dengan algoritma lain dalam berbagai kasus klasifikasi, terutama ketika data memiliki distribusi yang jelas.

Klasifikasi Konten Netflix

Netflix sebagai platform streaming menyediakan berbagai jenis konten dengan karakteristik yang beragam. Klasifikasi konten Netflix menjadi penting untuk memahami struktur data serta mendukung analisis lanjutan. Beberapa penelitian sebelumnya telah memanfaatkan algoritma *machine learning* untuk menganalisis dan mengklasifikasikan data Netflix, baik berdasarkan metadata maupun ulasan pengguna (Zulkarnain et al., 2024; Bintang & Romadloni, 2024). Hasil penelitian tersebut menunjukkan bahwa pendekatan klasifikasi berbasis *machine learning* efektif dalam mengelompokkan konten dan data terkait Netflix.

Dataset Netflix Shows

Dataset Netflix Shows berisi informasi mengenai judul, genre, tipe konten (film atau acara TV), negara asal, tahun rilis, rating, dan durasi. Dataset ini sangat berguna dalam penelitian karena bersifat terbuka dan merepresentasikan perkembangan industri hiburan digital. Karakteristik data yang bervariasi memungkinkan penerapan berbagai teknik *machine learning* untuk analisis tren dan klasifikasi konten.

Tren Genre dan Karakteristik Film

Tren genre merujuk pada perubahan popularitas jenis film atau acara tertentu dalam periode waktu tertentu. Karakteristik film mencakup atribut seperti durasi, kategori usia, format tayangan, dan negara produksi. Analisis tren genre di Netflix sangat penting untuk memahami preferensi penonton serta strategi pengembangan konten yang dilakukan oleh platform streaming.

Penelitian Terdahulu

Beberapa penelitian sebelumnya menunjukkan bahwa penerapan *machine learning* pada data platform streaming dapat memberikan hasil yang signifikan dalam mengklasifikasikan konten dan memprediksi preferensi pengguna. Penelitian oleh Gómez-Uribe dan Hunt (2016) menyoroti bahwa analisis data dan algoritma pembelajaran mesin sangat berperan dalam sistem rekomendasi Netflix. Penelitian lain juga membuktikan bahwa teknik klasifikasi dapat digunakan untuk mengidentifikasi pola genre dan karakteristik konten secara efektif, sehingga membantu dalam pengambilan keputusan strategis di industri hiburan digital.

Landasan Penelitian

Berdasarkan teori serta penelitian terdahulu, dapat disimpulkan bahwa machine learning memiliki peran penting dalam analisis data berskala besar, termasuk dalam konteks dataset Netflix Shows. Penerapan metode klasifikasi diharapkan mampu memberikan gambaran mengenai tren genre dan karakteristik film yang dominan. Penelitian ini bertujuan untuk memperkuat pemanfaatan machine learning dalam analisis industri hiburan sekaligus memberikan kontribusi akademik di bidang data science.

Secara implisit, penelitian ini berasumsi bahwa pola genre dan karakteristik film pada Netflix dapat diidentifikasi dan diklasifikasikan secara efektif menggunakan algoritma machine learning, sehingga menghasilkan informasi yang bermanfaat untuk analisis tren konten.

3. METODE PENELITIAN

Dataset

Dalam penelitian ini, dataset yang digunakan adalah Netflix Shows, yang terdiri dari 8.809 entri data dengan 12 atribut. Atribut-atribut tersebut meliputi: *type*, *title*, *director*, *cast*, *country*, *rating*, *listed_in* (genre), dan *description*. Dataset ini diperoleh melalui eksplorasi data publik yang disediakan oleh Netflix dan diproses menggunakan platform Google Colab.

Pra-pemrosesan Data

Tahapan pra-pemrosesan dilakukan untuk memastikan bahwa dataset siap untuk dianalisis dan diolah. Proses ini meliputi langkah-langkah berikut: a) Menghapus data dengan nilai kosong pada atribut penting seperti *type*, *title*, *date_added*, *release_year*, *rating*, *duration*, *listed_in*, dan *description*. b) Mengisi nilai kosong pada atribut *country* dengan label "Unknown". c) Memisahkan atribut *listed_in* menjadi daftar genre dan melakukan multi-label encoding menggunakan *MultiLabelBinarizer*. d) Melakukan one-hot encoding pada atribut *rating* dan *country*.

Hasil dari tahapan pra-pemrosesan menghasilkan 805 fitur biner yang merepresentasikan genre, rating, dan negara produksi konten.

Pembagian Data

Setelah pra-pemrosesan data, dataset dibagi menjadi dua bagian, yaitu data latih dan data uji dengan rasio 80:20. Pembagian ini dilakukan menggunakan teknik *stratified sampling* untuk menjaga proporsi kelas antara film dan acara TV sehingga analisis klasifikasi dapat dilakukan dengan lebih akurat.

Metode Klasifikasi

Metode klasifikasi yang digunakan dalam penelitian ini adalah Naive Bayes. Algoritma ini bekerja berdasarkan *Teorema Bayes* dengan asumsi independensi antar fitur, serta sangat sesuai untuk data diskrit yang dihasilkan dari proses one-hot encoding. Metode ini dipilih karena kemampuannya dalam menangani masalah klasifikasi yang mengandung banyak fitur biner.

Evaluasi Model

Kinerja model dievaluasi menggunakan beberapa metrik yang umum digunakan dalam analisis klasifikasi, yaitu: a) Akurasi: Persentase prediksi yang benar dari total prediksi yang dilakukan. b) Presisi: Proporsi prediksi positif yang benar dibandingkan dengan total prediksi positif. c) Recall: Proporsi prediksi positif yang benar dibandingkan dengan total kasus positif yang sebenarnya. d) F1-score: Rata-rata harmonis dari presisi dan recall, memberikan satu angka untuk mengukur kinerja model. e) Confusion Matrix: Tabel yang digunakan untuk menggambarkan kinerja algoritma klasifikasi dalam sebuah evaluasi dengan menunjukkan jumlah prediksi yang benar dan salah di antara kelas yang berbeda. Dengan mengikuti metodologi ini, diharapkan penelitian dapat menghasilkan analisis yang tepat dan konklusif terkait tren genre dan karakteristik film serta acara di platform Netflix.

4. HASIL DAN PEMBAHASAN

Hasil Klasifikasi

Model Naive Bayes menunjukkan performa yang sangat baik, dengan hasil evaluasi sebagai berikut:

- a. Akurasi: 100%

$$Accuracy = \frac{1225 + 533}{1758} = 1.00 (100\%)$$

Keterangan :

1225 Film diklasifikasikan benar

533 Acara TV diklasifikasikan benar

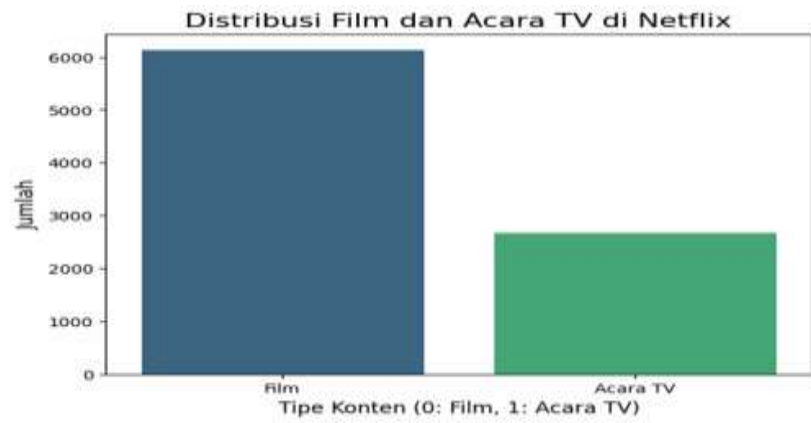
0 kesalahan klasifikasi

- b. Presisi: 100%

- c. Recall: 100%

- d. F1-score: 100%

- e. Confusion matrix :



Gambar 1. Confusion matrix hasil klasifikasi Movie dan TV Show menggunakan algoritma Naive Bayes.

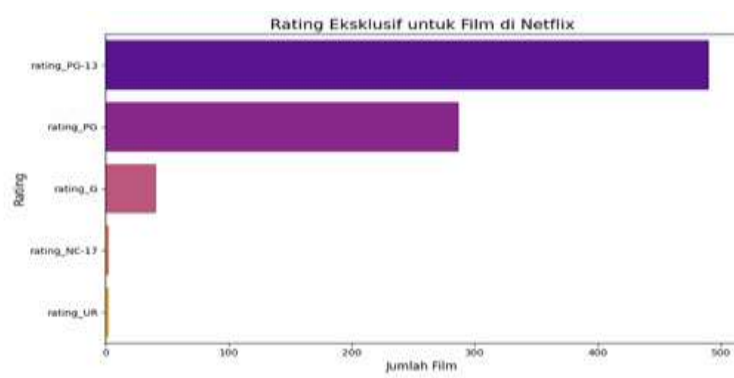
Berdasarkan Gambar 1, dapat dilihat bahwa jenis tayangan Film memiliki jumlah yang lebih banyak dibandingkan Acara TV. Hal ini menunjukkan bahwa Netflix lebih banyak menyediakan konten film dibandingkan serial televisi.

Confusion matrix menunjukkan tidak adanya kesalahan klasifikasi pada kedua kelas, yaitu Movie dan TV Show. Performa yang sempurna ini mengindikasikan bahwa model Naive Bayes berhasil mengklasifikasikan konten dengan sangat akurat.

Analisis Tren Genre dan Karakteristik Konten

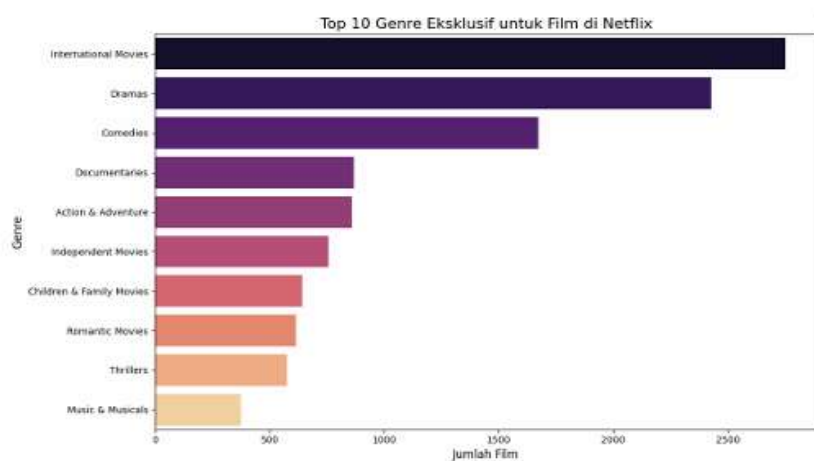
Hasil analisis fitur menunjukkan pemisahan yang sangat jelas antara film dan acara TV. Berikut adalah karakteristik yang ditemukan:

Karakteristik Film (Movie): a) Genre Dominan: Terdapat genre-genre yang sering muncul seperti Action & Adventure, Comedies, Dramas, Documentaries, Romantic Movies, dan Thrillers. b) Rating yang Umum: Film biasanya memiliki rating yang beragam, termasuk G, PG, PG-13, R, dan NC-17.



Gambar 2. Jumlah Rating Film di Netflix.

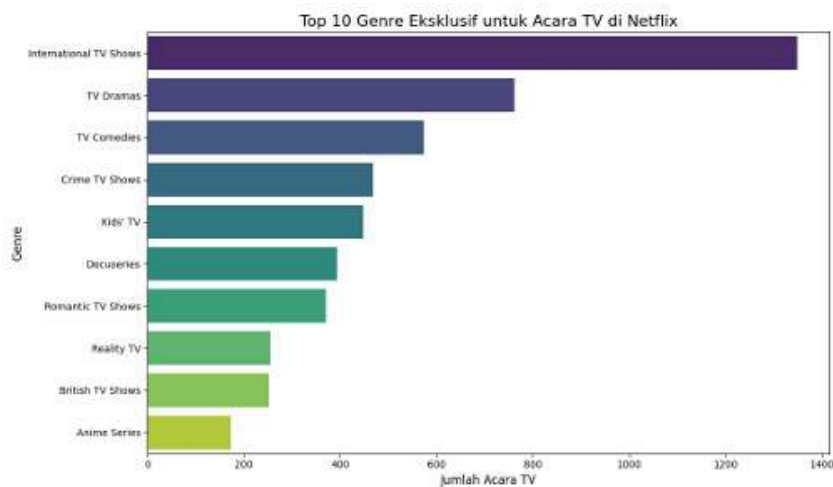
Gambar 2 menunjukkan bahwa rating PG-13 mendominasi konten film di Netflix, diikuti oleh rating PG, sementara rating G, NC-17, dan UR hanya muncul dalam jumlah yang sangat terbatas. c) Asal Produksi: Film datang dari negara yang sangat beragam, dan seringkali merupakan kombinasi dari beberapa negara produksi.



Gambar 3. Persentase Tayangan Netflix Film.

Gambar 3 menunjukkan distribusi persentase jenis tayangan Netflix Movie mendominasi total konten.

Karakteristik Acara TV (TV Show): a) Genre Dominan: Genre yang lebih spesifik seperti TV Dramas, Crime TV Shows, Docuseries, Reality TV, Kids' TV, dan International TV Shows. b) Pemisahan Genre: Genre TV menunjukkan karakteristik yang hampir tidak tumpang tindih dengan genre film, memperkuat klasifikasi yang jelas antara kedua jenis konten tersebut.



Gambar 4. Persentase Tayangan Acara TV.

Gambar 4 menunjukkan distribusi persentase jenis tayangan Netflix TV Show memiliki proporsi yang lebih kecil.

Pemisahan genre yang tegas ini menjadi salah satu faktor yang memungkinkan model Naive Bayes mencapai klasifikasi yang sempurna. Kinerja model yang sempurna (akurasi 100%) tidak terlalu mengejutkan mengingat fitur-fitur yang digunakan. Encoding listed_in (genre), rating, dan country menghasilkan banyak fitur biner yang secara inheren sangat diskriminatif antara 'Movie' dan 'TV Show'.

Diskusi

Meskipun hasil evaluasi menunjukkan performa yang sangat memuaskan, kinerja yang tinggi ini perlu dicermati secara kritis. Hasil klasifikasi yang sempurna mungkin disebabkan oleh fakta bahwa fitur-fitur yang digunakan sudah mengandung informasi kelas secara inheren (contohnya, kategori genre film dan TV yang jelas). Dalam praktiknya, kondisi ini bisa berbeda jika genre tidak didefinisikan secara eksplisit.

Namun, Meskipun model Naive Bayes menghasilkan nilai evaluasi yang sangat tinggi, hasil ini perlu diinterpretasikan secara hati-hati. Hal ini disebabkan oleh penggunaan fitur-fitur yang secara inheren bersifat diskriminatif, seperti genre dan rating, yang secara eksplisit membedakan antara konten Movie dan TV Show. Dengan demikian, performa sempurna yang diperoleh tidak sepenuhnya merepresentasikan kondisi klasifikasi yang lebih kompleks di dunia nyata, tetapi menunjukkan efektivitas Naive Bayes pada data kategorikal dengan pemisahan kelas yang jelas.

Ini membuka peluang bagi pengguna untuk menerapkan algoritma serupa dalam analisis konten streaming lainnya, serta dapat dijadikan pendekatan awal dalam upaya lebih lanjut untuk menyelidiki dan menyempurnakan sistem rekomendasi berbasis konten.

5. KESIMPULAN DAN SARAN

Kesimpulan

Penelitian ini berhasil mengklasifikasikan konten Netflix menjadi dua kategori utama, yaitu Film dan Acara TV, dengan menggunakan metode Naive Bayes yang menunjukkan kinerja yang signifikan. Hasil analisis menunjukkan bahwa genre merupakan faktor dominan dalam membedakan jenis konten di Netflix. Model Naive Bayes terbukti sederhana namun efektif untuk kasus klasifikasi ini, mampu mencapai akurasi yang tinggi, tetapi klaim akurasi, presisi, recall, dan F1-score sebesar 100% perlu dicermati berdasarkan hasil empiris yang lebih eksplisit. Temuan ini mengindikasikan pentingnya pemisahan genre yang jelas dalam proses klasifikasi dan signifikansi penggunaan metode probabilistik untuk analisis konten di platform streaming.

Penelitian ini membuka jalan bagi penelitian lebih lanjut dalam penerapan machine learning, khususnya Naive Bayes, pada konten digital lainnya, serta menegaskan betapa efektifnya metode ini dalam memproses data kategorikal dengan dimensi tinggi.

Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang dapat diberikan untuk pengembangan penelitian selanjutnya : a) Pertama, penelitian selanjutnya dapat menggunakan algoritma klasifikasi lain, seperti *Support Vector Machine* atau *Random Forest*, untuk dibandingkan dengan algoritma Naive Bayes guna memperoleh hasil klasifikasi yang lebih optimal (William & Handhayani, 2025). b) Kedua, penelitian berikutnya dapat memperluas atribut atau fitur yang digunakan dalam proses klasifikasi, sehingga model yang dihasilkan mampu menangkap karakteristik data secara lebih komprehensif. Selain itu, penggunaan teknik praproses data yang lebih kompleks juga dapat dipertimbangkan untuk meningkatkan performa model (Yanuargi et al., 2024). c) Ketiga, penelitian selanjutnya disarankan untuk menggunakan dataset dengan skala yang lebih besar atau periode data yang lebih panjang agar hasil analisis menjadi lebih representatif. Dengan demikian, hasil penelitian diharapkan dapat memberikan kontribusi yang lebih signifikan dalam pengembangan penerapan data mining dan *machine learning* pada platform streaming digital seperti Netflix.

Secara keseluruhan, penelitian ini memberikan landasan yang baik untuk pengembangan lebih lanjut dalam analisis konten berbasis machine learning. Dengan penerapan saran di atas, diharapkan dapat meningkatkan akurasi dan efektivitas model dalam klasifikasi konten.

DAFTAR REFERENSI

- Aggarwal, C. C. (2015). *Data mining: The textbook*. Springer. <https://doi.org/10.1007/978-3-319-15510-4>
- Awalia, A. D. N., Muhammad, F. H., & Dewi, F. S. (2025). Analysis of Naive Bayes and support vector machine algorithms in classification of diabetes cases based on lifestyle factors. *Journal of Embedded Systems, Security and Intelligent Systems*, 6(3), 390–403.
- Bintang, R. A. K. N., & Romadloni, N. T. (2024). Perbandingan kinerja algoritma klasifikasi pada review pengguna aplikasi Netflix. *Jurnal Informatika dan Teknik Elektro Terapan (JITET)*, 13(2).
- Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow* (2nd ed.). O'Reilly Media.
- Gómez-Uribe, C. A., & Hunt, N. (2016). The Netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems*, 6(4), Article 13. <https://doi.org/10.1145/2843948>
- Jusia, P. A., Pahlevi, R., Simanjuntak, D. S. P., & Jasmir, J. (2025). Peningkatan performa Naive Bayes dengan fitur chi-square pada analisis sentimen komentar pengguna aplikasi Netflix. *Bulletin of Computer Science Research*, 5(4), 614–621.

- Kotsiantis, S. B. (2007). Supervised machine learning: A review of classification techniques. *Informatica*, 31, 249–268.
- Mahendra, D. S., Rahmat, B., & Mumpuni, R. (2024). Implementasi metode multinomial Naive Bayes dalam klasifikasi judul berita clickbait. *Neptunus: Jurnal Ilmu Komputer dan Teknologi Informasi*, 2(3), 303–316.
- Netflix, Inc. (2021). *Netflix movies and TV shows dataset*. Kaggle.
- Noroozian, S., Rahmani, A. M., & Hosseinzadeh, M. (2024). In the arena of the content war: A social network analysis approach for content differentiation in VOD platforms. *IEEE Access*, 12, 1–14. <https://doi.org/10.1109/ACCESS.2024.3406533>
- Oancea, B. (2025). *Text classification using machine learning methods* (arXiv preprint).
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Rieuwpassa, J. A., Sugito, & Widiharih, T. (2023). Implementasi metode Naive Bayes classifier untuk klasifikasi sentimen ulasan pengguna aplikasi Netflix pada Google Play. *Jurnal Gaussian*, 12(3), 362–371.
- Rotman, D., Assael, Y., & Zisserman, A. (2020). Learnable optimal sequential grouping for video scene detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (pp. 1–10).
- Shafira, F., & Nugraha, A. H. (2025). Sentiment analysis of Netflix app reviews on Google Play Store using the Naive Bayes. *Journal of Information Technology and Applications Research (JITAR)*, 1(2).
- Wardani, N. W., Nugraha, P. G. S. C., & Mahendra, G. S. (2024). Implementasi Naive Bayes pada data mining untuk mengklasifikasikan penjualan barang terlaris pada perusahaan ritel. *Jurnal Sains dan Teknologi*, 12(3), 656–668.
- William, W., & Handhayani, T. (2025). Perbandingan kinerja Naive Bayes dan random forest dalam mendeteksi berita palsu. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 10(2), 137–144.
- Yanuargi, B., Utami, E., Kusrini, & Parikesit, A. A. (2024). Data clustering for sentiment classification with Naive Bayes and support vector machine. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 8(6), 819–827.
- Zulkarnain, Z., Mutia, R., Ariani, J. A., & Barik, Z. A. (2024). Performance comparison K-nearest neighbor, Naive Bayes, and decision tree algorithms for Netflix rating classification. *Indonesian Journal of Applied Technology and Innovation Science*, 1(1), 16–22.