



Department of Digital Business

Journal of Artificial Intelligence and Digital Business (RIGGS)

Homepage: <https://journal.ilmudata.co.id/index.php/RIGGS>

Vol. 5 No. 1 (2026) pp: 308-315

P-ISSN: 2963-9298, e-ISSN: 2963-914X

Analisis Performa Naïve Bayes, K-Neares Neighbor, dan Random Forest dalam Klasifikasi Sentimen Ulasan Aplikasi Netflix

Yasinta Nurbaya¹, Rossiana Sihombing², Harjanti Bhatara I³, Dian Sifa Salsabila⁴

^{1,2,3,4} Sistem Informasi, Fakultas Teknik Dan Informatika, Universitas Bina Sarana Informatika

vasintanurbaya6@gmail.com¹, rossianasihombing8@gmail.com², harjantibhataraibrahim03@gmail.com³, hurridian8@gmail.com⁴

Abstrak

Dengan jutaan pengguna aktif di Indonesia, Netflix menjadi salah satu platform streaming digital terdepan yang terus berupaya menjaga kualitas layanan dan kepuasan penggunanya. Ulasan pengguna yang tersedia pada Google Play Store merepresentasikan sumber data penting yang mencerminkan persepsi, pengalaman, serta tingkat kepuasan konsumen terhadap layanan yang diberikan. Namun, besarnya jumlah ulasan dan bentuknya yang tidak terstruktur menuntut adanya pendekatan otomatis untuk menganalisis opini pengguna secara sistematis dan objektif. Penelitian ini bertujuan untuk mengkategorikan sentimen ulasan pengguna Netflix ke dalam tiga kelas, yaitu positif, negatif, dan netral, serta membandingkan kinerja algoritma klasifikasi machine learning yang digunakan. Penelitian ini mengimplementasikan tiga algoritma klasifikasi, yaitu Naïve Bayes, K-Nearest Neighbor (KNN), dan Random Forest, dengan memanfaatkan perangkat lunak Orange Data Mining. Proses penelitian meliputi tahap preprocessing data teks, ekstraksi fitur menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF), serta evaluasi model menggunakan teknik 10-fold stratified cross-validation. Kinerja model dievaluasi berdasarkan metrik akurasi, Area Under Curve (AUC), F1-score, precision, recall, dan Matthews Correlation Coefficient (MCC). Hasil penelitian menunjukkan bahwa algoritma Naïve Bayes menghasilkan performa terbaik dengan tingkat akurasi sebesar 99,2% dan nilai AUC sebesar 1,000, diikuti oleh Random Forest dengan akurasi 98,3% dan AUC 0,998. Sebaliknya, algoritma KNN menunjukkan performa terendah dengan akurasi 36,2% dan AUC 0,500. Analisis statistik mengonfirmasi adanya perbedaan performa yang signifikan antara Naïve Bayes dan KNN ($p < 0,05$). Temuan ini menunjukkan bahwa algoritma probabilistik berbasis Naïve Bayes lebih efektif dalam menangani data teks berdimensi tinggi dengan representasi TF-IDF. Secara keseluruhan, penelitian ini menegaskan pentingnya kesesuaian antara karakteristik data dan algoritma dalam klasifikasi sentimen ulasan aplikasi streaming digital seperti Netflix.

Kata kunci: Anaisis Sentimen, Naïve Bayes, K-Nearest Neighbor, Random Forest, Netflix.

1. Latar Belakang

Perkembangan industri layanan streaming digital mengalami pertumbuhan pesat seiring meningkatnya adopsi teknologi dan perubahan perilaku konsumsi masyarakat (Xie, 2024). Perubahan pola konsumsi hiburan berbasis digital mendorong platform streaming untuk semakin memperhatikan kualitas layanan, konten, serta pengalaman pengguna secara menyeluruh (Velankar, 2025). Dalam konteks ini, pemahaman terhadap persepsi dan kepuasan pengguna menjadi faktor strategis dalam menjaga keberlanjutan dan daya saing layanan streaming digital (Ak et al., 2024).

Netflix merupakan salah satu platfrom layanan streaming digital terbesar di dunia yang berdiri pada tahun 1997 di California, Amerika Serikat, dan mulai beroperasi di Indonesia sejak tahun 2016 (Panji, 2016). Selama masa pandemi COVID-19 pada tahun 2020, jumlah pelanggan Netflix meningkat secara signifikan karena meningkatkan kebutuhan masyarakat terhadap hiburan di rumah. Namun setelah pandemi mulai mereda pada tahun 2021, pertumbuhan pelanggan mengalami perlambatan akibat penundaan produksi konten hiburan, yang menyebabkan penurunan minat pelanggan baru terhadap layanan berlangganan (Riyanto, 2021). Kondisi tersebut menunjukkan pentingnya evaluasi berkelanjutan terhadap kualitas layanan dan pengalaman pengguna.

Salah satu sumber informasi penting yang merepresentasikan persepsi dan kepuasan pengguna adalah ulasan pengguna (*user review*) yang tersedia pada platform distribusi aplikasi digital seperti Google Play Store. Google Play Store menyediakan berbagai produk digital, termasuk aplikasi Netflix, serta fitur ulasan pengguna yang dapat

Analisis Performa Naïve Bayes, K-Neares Neighbor, dan Random Forest dalam Klasifikasi Sentimen Ulasan Aplikasi Netflix

dimanfaatkan untuk menilai tingkat kepuasan, mendeteksi permasalahan layanan, dan memahami persepsi pengguna terhadap aplikasi tertentu (Riyanto, 2021). Ulasan pengguna telah menjadi sumber data penting untuk memahami kualitas layanan digital dan preferensi konsumen dalam era digital (Mathayomchan & Taecharungoj, 2020). Dalam konteks aplikasi mobile, analisis dapat memberikan insights berharga tentang fitur yang telah diperbaiki dan ekspektasi pengguna (Haque et al., 2022).

Selain berfungsi sebagai sarana penyampaian opini, ulasan pengguna juga merepresentasikan bentuk komunikasi tidak langsung antara konsumen dan penyedia layanan digital. Ulasan tersebut sering kali memuat pengalaman subjektif, keluhan, apresiasi, serta ekspektasi pengguna terhadap kualitas layanan yang diterima. Namun, jumlah ulasan yang sangat besar dan disajikan dalam bentuk teks tidak terstruktur menjadikan proses analisis manual tidak lagi efisien dan rentan terhadap bias subjektivitas. Oleh karena itu, diperlukan pendekatan otomatis yang mampu mengolah dan mengklasifikasikan opini pengguna secara sistematis dan konsisten.

Pendekatan berbasis *machine learning* menjadi solusi yang relevan dalam menghadapi kompleksitas data teks dalam jumlah besar. Dengan memanfaatkan algoritma klasifikasi, opini pengguna dapat dipetakan ke dalam kategori sentimen tertentu secara cepat dan akurat. Pendekatan ini memungkinkan pengembang aplikasi untuk memperoleh gambaran umum mengenai persepsi pengguna tanpa harus membaca ulasan satu per satu, sehingga proses evaluasi layanan dapat dilakukan secara lebih efektif dan berbasis data.

Analisis sentimen merupakan cabang dari text mining yang berfokus pada pengolahan opini atau pernyataan pengguna untuk mengklasifikasi secara otomatis ke dalam kategori positif, negatif, atau netral secara otomatis (Birjali et al., 2021). Metode ini telah banyak digunakan dalam berbagai domain, termasuk e-commerce, media sosial, dan aplikasi digital (Wankhade et al., 2002). Dalam konteks aplikasi streaming, analisis sentimen berperan penting untuk membantu pengembang mengidentifikasi aspek yang perlu di perbaiki dan memahami ekpetasi pengguna secara mendalam (Xu et al., 2021). Penelitian terkini menunjukkan bahwa analisis sentimen dapat meningkatkan kepuasan pelanggan dengan memberikan feedback loop yang cepat dan akurat kepada pengembang (Hasaan & Mahmood, 2023).

Keberhasilan analisis sentimen sangat dipengaruhi oleh pemilihan algoritma klasifikasi yang digunakan. Naïve Bayes Classifier merupakan algoritma berbasis probabilitas yang memiliki keunggulan dalam kesederhanaan, kecepatan komputasi, serta performa yang baik pada data berdimensi tinggi (Kowsari et al., 2020). Sementara itu *K-Nearest-Neighbor* (KNN) mengelompokkan data berdasarkan kedekatan antar instance, namun cenderung kurang optimal pada dataset dengan jumlah fitur yang besar (Zhang et al., 2022). Adapun Random Forest merupakan metode ensemble learning yang menggabungkan prediksi dari multiple decision trees untuk menghasilkan klasifikasi yang robust dan akurat (Breiman, 2001). Penelitian ini menunjukkan bahwa Random Forest memiliki performa konsisten pada berbagai jenis dataset teks (Aldhyani & Alkahtani, 2023).

Pemilihan ketiga algoritma dalam penelitian ini didasarkan pada perbedaan karakteristik pendekatan yang digunakan. Naïve Bayes merepresentasikan algoritma probabilistik sederhana, KNN mewakili algoritma berbasis jarak, sedangkan Random Forest merupakan algoritma *ensemble learning*. Perbandingan ketiga pendekatan ini memungkinkan evaluasi yang lebih komprehensif terhadap pengaruh perbedaan prinsip kerja algoritma terhadap performa klasifikasi sentimen pada data teks berdimensi tinggi. Dengan demikian, penelitian ini tidak hanya berfokus pada pencapaian akurasi tertinggi, tetapi juga pada kesesuaian algoritma terhadap karakteristik data ulasan aplikasi streaming.

Dengan membandingkan algoritma yang memiliki kompleksitas dan pendekatan berbeda, penelitian ini tidak hanya berfokus pada pencapaian akurasi tertinggi, tetapi juga memberikan pemahaman mengenai kesesuaian algoritma terhadap karakteristik data teks ulasan aplikasi streaming. Hal ini penting karena pemilihan algoritma yang tidak tepat dapat menghasilkan performa klasifikasi yang rendah meskipun data dan preprocessing telah dilakukan dengan baik.

Selain algoritma, representasi fitur juga memainkan peran krusial dalam klasifikasi teks. Term Frequency–Inverse Document Frequency (TF-IDF) merupakan metode yang banyak digunakan untuk mengonversi teks menjadi representasi numerik yang dapat diproses oleh algoritma *machine learning* (Rahutomo et al., 2023). Metode ini memberikan bobot pada kata berdasarkan frekuensi kemunculan dan keunikannya, sehingga lebih informatif dibanding metode Bag of Words sederhana (Kaya et al., 2020).

Sejumlah penelitian sebelumnya menunjukkan bahwa performa algoritma klasifikasi dapat berbeda-beda tergantung pada domain dan karakteristik data yang digunakan (Yadav & Vishwakarma, 2020). Studi komparatif telah dilakukan pada berbagai domain, seperti ulasan produk e-commerce (Sing et al., 2021), media sosial (Sailunaz et al., 2021), dan aplikasi mobile (Hassan et al., 2022). Masukan akademik dari Jefina Tri Kumalasari, M.Kom juga menguatkan bahwa perbedaan hasil klasifikasi dipengaruhi oleh kualitas preprocessing dan representasi fitur yang digunakan. Namun demikian, penelitian yang secara khusus membandingkan kinerja Naïve Bayes, KNN, dan Random Forest dalam klasifikasi sentimen ulasan aplikasi streaming seperti Netflix masih terbatas.

Selain itu, karakteristik ulasan pengguna pada layanan streaming memiliki perbedaan dibandingkan domain e-commerce maupun media sosial, baik dari segi konteks, panjang teks, maupun fokus evaluasi pengguna. Oleh karena itu, diperlukan kajian khusus yang menelaah performa algoritma klasifikasi pada domain aplikasi streaming digital agar diperoleh hasil yang lebih kontekstual dan akurat.

Berdasarkan uraian tersebut, penelitian ini berfokus untuk menganalisis dan membandingkan performa algoritma Naïve Bayes, KNN, dan Random Forest dalam mengklasifikasikan sentimen ulasan pengguna Netflix, serta mengidentifikasi perbedaan performa yang dihasilkan. Tujuan penelitian ini adalah mengevaluasi dan membandingkan performa ketiga algoritma berdasarkan metrik AUC, akurasi, F1-score, precision, recall, dan MCC, serta memberikan rekomendasi algoritma yang paling optimal untuk implementasi sistem analisis sentimen pada aplikasi streaming digital. Secara teoretis, penelitian ini diharapkan dapat memberikan kontribusi terhadap pengembangan ilmu pengetahuan di bidang *text mining* dan analisis sentimen. Secara praktis, hasil penelitian ini diharapkan dapat menjadi acuan bagi pengembang aplikasi streaming dalam merancang sistem pemantauan sentimen pengguna yang efektif dan memiliki tingkat akurasi tinggi.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan eksperimental dengan tujuan membandingkan kinerja tiga algoritma klasifikasi machine learning, yaitu Naïve Bayes, K-Nearest Neighbor (KNN), dan Random Forest, dalam analisis sentimen ulasan aplikasi Netflix yang tersedia pada Google Play Store. Pendekatan eksperimental dipilih karena memungkinkan peneliti untuk mengevaluasi performa masing-masing algoritma secara objektif berdasarkan metrik evaluasi yang terukur dan konsisten.

Seluruh tahapan analisis, mulai dari pengolahan data hingga evaluasi model, dilakukan menggunakan perangkat lunak Orange Data Mining versi 3.36. Pemilihan Orange Data Mining didasarkan pada kemampuannya dalam menyediakan modul pemrosesan teks, ekstraksi fitur, klasifikasi, serta evaluasi performa model secara terintegrasi, sehingga mendukung proses eksperimen yang sistematis dan terstruktur.

2.1 Pengumpulan Data

Dataset ulasan Netflix diperoleh dari Google Play Store periode September-November 2025, mencakup kolom Review (teks ulasan), Username (identitas pengguna), dan sentiment (label kategori). Kolom lain seperti ID, Helpful_Count, Tanggal, dan Judul film diabaikan karena tidak relevan dengan tujuan klasifikasi.

2.2 Preprocessing Data

Tahap preprocessing dilakukan menggunakan modul Preprocess Text di Orange Data Mining dengan tiga tahapan utama (1) Transformation mengkonversi seluruh teks menjadi huruf kecil (lowercase) untuk menghilangkan perbedaan kapitalisasi; (2) Tokenization memisahkan teks menjadi token individual menggunakan pattern Regexp `\w+` yang menghasilkan 1000 tokens dengan 998 types unik; (3) Filtering menghapus stopwords Bahasa Inggris seperti “the”, “is”, “and” untuk mengurangi dimensi fitur dan meningkatkan fokus pada kata-kata informatif.

2.3 Ekstraksi Fitur

Representasi numerik data teks dilakukan menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) melalui modul *Bag of Words*. Konfigurasi yang digunakan: *Term Frequency (Count)*, *Document Frequency (IDF)*, dan tanpa *regularization*. TF-IDF memberikan bobot lebih tinggi pada kata yang sering muncul dalam satu dokumen namun jarang di dokumen lain, sehingga lebih informatif untuk klasifikasi.

2.4 Klasifikasi dan Evaluasi

Tiga algoritma yang dibandingkan adalah *Naïve Bayes*, *K-Nearest Neighbor* (KNN), dan *Random Forest*. Evaluasi model yang menggunakan *10-fold stratified cross-validation* untuk memastikan setiap fold memiliki proporsi kelas

yang seimbang. Metrik evaluasi meliputi: *Accuracy (CA)*, *Area Under Curve (AUC)*, *F1-Score*, *Precision*, *Reccal*, dan *Matthews Correlation Coefficient (MCC)*. Analisis statistik kompratif dilakukan untuk menguji signikansi perbedaan performa antar algoritma, dan visualisasi *ROC Curve* digunakan untuk membandingkan kemampuan dikriminasi masing- masing model.

Penggunaan teknik 10-fold stratified cross-validation bertujuan untuk memperoleh estimasi performa model yang lebih stabil dan representatif. Teknik ini memastikan bahwa setiap kelas sentimen terdistribusi secara proporsional pada setiap fold, sehingga hasil evaluasi tidak bergantung pada satu pembagian data tertentu. Dengan demikian, performa model yang dihasilkan mencerminkan kemampuan generalisasi algoritma terhadap data yang belum pernah dilihat sebelumnya.

Selain akurasi, penelitian ini menggunakan metrik evaluasi lain seperti AUC dan Matthews Correlation Coefficient (MCC) untuk memberikan gambaran performa model yang lebih komprehensif. Penggunaan MCC menjadi penting karena metrik ini mampu mengevaluasi kualitas klasifikasi secara seimbang pada kondisi distribusi kelas yang tidak sepenuhnya identik, sehingga hasil evaluasi menjadi lebih objektif dan robust.

3. Hasil dan Diskusi

Bagian ini menyajikan hasil penelitian yang diperoleh dari proses pengolahan data, pengujian model, serta evaluasi performa algoritma klasifikasi sentimen yang digunakan. Seluruh temuan empiris dianalisis secara terpadu dengan pembahasan konseptual untuk menjelaskan makna hasil, perbandingan performa antar algoritma, serta implikasinya dalam konteks klasifikasi sentimen ulasan aplikasi Netflix berbasis data teks. Pembahasan dilakukan secara sistematis mulai dari karakteristik dataset, performa klasifikasi, analisis statistik, visualisasi performa, hingga diskusi konseptual, implikasi, dan keterbatasan penelitian.

3.1 Karakteristik Dataset dan Preprocessing

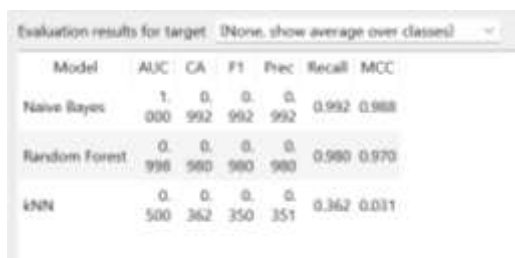
Dataset penelitian terdiri dari 1000 ulasan aplikasi Netflix yang diperoleh dari Google Play Store dengan distribusi sentimen yang relatif seimbang antara kategori positif, negatif, dan netral. Keseimbangan distribusi kelas ini penting untuk memastikan bahwa model klasifikasi tidak bias terhadap salah satu kategori sentimen tertentu dan mampu mempelajari pola sentimen secara proporsional.

Kolom *Sentiment* berfungsi sebagai variabel target, sedangkan kolom *Review* digunakan sebagai variabel independen yang merepresentasikan data teks mentah. Untuk mengubah data teks menjadi format numerik yang dapat diproses oleh algoritma *machine learning*, digunakan metode ekstraksi fitur Term Frequency–Inverse Document Frequency (TF-IDF).

Tahapan preprocessing data mencakup tiga proses utama, yaitu *case folding*, tokenisasi, dan *stopwords removal*. Proses *case folding* dilakukan untuk menyeragamkan seluruh karakter menjadi huruf kecil guna menghindari perbedaan fitur akibat variasi kapitalisasi. Tokenisasi dilakukan menggunakan pola `\W+` sehingga menghasilkan sebanyak 998 kata unik dari total 1000 token. Selanjutnya, *stopwords removal* diterapkan untuk menghilangkan kata-kata umum yang tidak memiliki kontribusi signifikan terhadap pembentukan sentimen. Tahapan preprocessing ini bertujuan untuk menghasilkan representasi data yang lebih bersih, ringkas, dan informatif, sehingga dapat meningkatkan performa algoritma klasifikasi.

3.2 Performa Klasifikasi

Evaluasi performa model dilakukan menggunakan teknik *10-fold stratified cross-validation* untuk memastikan bahwa setiap *fold* memiliki proporsi kelas yang seimbang. Teknik ini dipilih karena mampu memberikan estimasi performa model yang lebih stabil serta mengurangi potensi *overfitting*.



Model	AUC	CA	F1	Prec	Recall	MCC
Naive Bayes	1.000	0.992	0.992	0.992	0.992	0.988
Random Forest	0.998	0.980	0.980	0.980	0.980	0.980
kNN	0.500	0.362	0.350	0.351	0.362	0.031

Gambar 1 Hasil Evaluasi Performa Algoritma Klasifikasi

Sumber: Dokumentasi Peneliti November 2025

Berdasarkan hasil evaluasi, algoritma Naïve Bayes mencapai performa optimal dengan tingkat akurasi sebesar 99,2% dan nilai Area Under Curve (AUC) sempurna sebesar 1,000. Nilai F1-Score, Precision, dan Recall yang konsisten pada angka 0,992 menunjukkan keseimbangan yang sangat baik antara ketepatan prediksi dan kemampuan model dalam mendeteksi seluruh kategori sentimen. Keunggulan algoritma ini menunjukkan efektivitas pendekatan probabilistik dalam menangani data teks berdimensi tinggi yang direpresentasikan menggunakan TF-IDF.

Algoritma Random Forest menempati posisi kedua dengan tingkat akurasi sebesar 98,3% dan nilai AUC sebesar 0,998. Meskipun performanya sedikit berada di bawah Naïve Bayes dengan selisih sekitar 1,2%, algoritma *ensemble learning* ini tetap menunjukkan hasil yang sangat kompetitif. Konsistensi nilai metrik evaluasi di sekitar 0,980 mengindikasikan stabilitas model yang baik serta ketahanannya terhadap *overfitting* melalui penggabungan prediksi dari multiple *decision trees*.

Sebaliknya, algoritma K-Nearest Neighbor (KNN) menunjukkan performa yang sangat rendah dengan tingkat akurasi sebesar 36,2% dan nilai AUC sebesar 0,500, yang setara dengan prediksi acak. Nilai Matthews Correlation Coefficient (MCC) sebesar 0,031 mengonfirmasi tidak adanya korelasi signifikan antara hasil prediksi model dengan label aktual. Rendahnya performa ini mengindikasikan bahwa algoritma berbasis jarak kurang efektif digunakan pada data teks dengan representasi fitur berdimensi tinggi.

Secara keseluruhan, perbedaan performa yang signifikan antar algoritma menunjukkan bahwa keberhasilan klasifikasi sentimen tidak hanya ditentukan oleh jumlah data, tetapi juga oleh kesesuaian algoritma terhadap representasi fitur yang digunakan. Algoritma yang mampu menangani kompleksitas dan sparsity data TF-IDF, seperti Naïve Bayes dan Random Forest, terbukti lebih unggul dibandingkan algoritma berbasis jarak seperti KNN.

3.3 Analisis Statistik dan Visualisasi

Untuk memperkuat hasil evaluasi performa, dilakukan analisis statistik berbasis Area Under ROC Curve (AUC) guna menguji signifikansi perbedaan performa antar algoritma.



Gambar 2 Probabilitas Perbedaan Performa Antar Algoritma

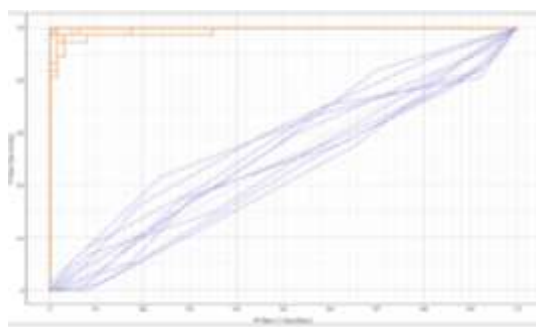
Sumber: Dokumentasi Peneliti November 2025

Hasil analisis statistik menunjukkan adanya perbedaan performa yang signifikan antara algoritma Naïve Bayes dan KNN dengan nilai $p < 0,05$. Probabilitas sebesar 0,965 pada perbandingan antara Naïve Bayes dan Random Forest mengindikasikan bahwa Naïve Bayes memiliki keunggulan dengan tingkat kepercayaan sebesar 96,5%, meskipun perbedaannya relatif kecil. Temuan ini menunjukkan bahwa kedua algoritma tersebut memiliki performa yang sama-sama tinggi, namun Naïve Bayes tetap sedikit lebih unggul secara statistik.

Visualisasi performa menggunakan kurva ROC juga digunakan untuk memvalidasi hasil numerik secara visual. Kurva ROC memberikan gambaran kemampuan model dalam membedakan kelas positif dan negatif melalui perbandingan *true positive rate* dan *false positive rate*.

3.4 Visualisasi ROC Curve

Untuk memberikan gambaran yang lebih komprehensif mengenai kemampuan diskriminasi masing-masing algoritma klasifikasi, analisis visual dilakukan menggunakan kurva Receiver Operating Characteristic (ROC). Kurva ROC digunakan untuk mengevaluasi performa model dalam membedakan kelas sentimen dengan membandingkan nilai True Positive Rate dan False Positive Rate pada berbagai ambang keputusan. Pendekatan ini memungkinkan penilaian performa model secara lebih menyeluruh, tidak hanya berdasarkan nilai akurasi, tetapi juga berdasarkan kemampuan model dalam mempertahankan tingkat sensitivitas yang tinggi dengan kesalahan klasifikasi yang minimal. Visualisasi kurva ROC selanjutnya digunakan untuk mengonfirmasi dan memperkuat temuan numerik yang diperoleh dari metrik evaluasi sebelumnya.



Gambar 3 Kurva ROC Pebandingan Performa Algoritma
Sumber: Dokumentasi Peneliti November 2025

Gambar 3 menampilkan kurva ROC (*Receiver Operating Characteristic*) untuk ketiga algoritma yang diuji. Kurva ROC merupakan representasi grafis dari performa *classifier* yang menggambarkan *trade-off* antara *True Positive Rate (Sensitivity)* pada sumbu Y dan *False Positive Rate (1-Specificity)* pada sumbu X. Untuk menilai performa diskriminatif setiap algoritma secara lebih mendalam, dilakukan analisis kurva ROC sebagaimana diuraikan berikut:

- Naïve Bayes* (Garis Hijau/Teal): Kurva ROC untuk *Naïve Bayes* menunjukkan pola yang hampir sempurna, mendekati sudut kiri atas grafik. Ini mengindikasikan bahwa model mampu mencapai *True Positive Rate* yang sangat tinggi dengan *False Positive Rate* yang minimal. Bentuk kurva yang mendekati ideal ini sesuai dengan nilai AUC 1.000 yang diperoleh.
- Random Forest* (Garis Orange): Kurva ROC untuk *Random Forest* juga menunjukkan performa yang sangat baik, meskipun sedikit di bawah *Naïve Bayes*. Kurva masih sangat dekat dengan sudut kiri atas, yang sesuai dengan nilai AUC 0.998. Perbedaan visual dengan *Naïve Bayes* sangat kecil, menunjukkan bahwa kedua algoritma memiliki kemampuan diskriminasi yang hampir setara.
- KNN (Garis Ungu/Biru): Kurva ROC untuk KNN menunjukkan pola yang sangat berbeda. *Multiple lines* yang terlihat mengindikasikan variasi performa yang tinggi antar fold dalam *cross-validation*. Kurva-kurva tersebut tersebar di sekitar diagonal (garis putus-putus), yang merupakan baseline untuk random *classifier*. Ini memvalidasi nilai AUC 0.500 dan mengkonfirmasi bahwa kNN lebih baik dari prediksi acak pada dataset ini.

Visualisasi ROC Curve ini secara visual memvalidasi hasil numerik yang telah dipaparkan sebelumnya, menunjukkan superioritas *Naïve Bayes* dan *Random Forest* dibandingkan kNN dalam tugas klasifikasi *sentiment* ulasan Netflix.

Secara keseluruhan, hasil penelitian membuktikan superioritas algoritma berbasis probabilitas (*Naïve Bayes*) untuk klasifikasi sentimen teks dengan representasi TF-IDF. Asumsi independensi antar fitur yang menjadi basis *Naïve Bayes*, meskipun tidak sepenuhnya valid secara teoritis, tidak mengurangi efektivitas praktisnya dan justru memberikan keuntungan dalam generalisasi model.

Performa komplementif *Random Forest* (98.3%) menunjukkan bahwa peningkatan kompleksitas model tidak selalu menghasilkan peningkatan akurasi signifikan pada dataset dengan karakteristik tertentu. Sementara itu, kegagalan KNN menekankan pentingnya pertimbangan karakteristik data (terutama dimensionalitas) dalam pemilihan algoritma klasifikasi.

Implikasi praktis penelitian ini adalah kemungkinan implementasi sistem analisis *sentiment real-time* dengan akurasi sangat tinggi (>98%) untuk aplikasi *streaming* seperti Netflix, yang dapat digunakan untuk monitoring kepuasan pengguna, *automated response system*, dan *competitive intelligence*. Namun, penelitian lanjutan dengan dataset lebih besar, *preprocessing bilingual*, dan eksplorasi representasi fitur alternatif (*Word2Vec*, *BERT*) tetap diperlukan untuk validasi dan peningkatan performa lebih lanjut.

3.5 Diskusi Kinerja Algoritma Berdasarkan Karakteristik Data

Keunggulan *Naïve Bayes* dalam penelitian ini dapat dijelaskan melalui karakteristik data teks yang direpresentasikan menggunakan TF-IDF. Representasi TF-IDF menghasilkan ruang fitur berdimensi tinggi dan bersifat *sparse*, di mana sebagian besar nilai fitur bernilai nol. Dalam kondisi tersebut, asumsi independensi antar fitur yang menjadi dasar *Naïve Bayes* justru memberikan keuntungan praktis karena model tidak bergantung pada hubungan kompleks antar kata.

Sebaliknya, algoritma KNN sangat bergantung pada perhitungan jarak antar data dalam ruang fitur. Pada ruang berdimensi tinggi, fenomena *curse of dimensionality* menyebabkan jarak antar titik data menjadi relatif seragam, sehingga konsep kedekatan kehilangan makna diskriminatifnya. Kondisi ini menjelaskan rendahnya performa KNN dalam penelitian ini dan menunjukkan bahwa algoritma berbasis jarak kurang sesuai untuk klasifikasi sentimen berbasis TF-IDF.

Temuan ini menegaskan bahwa efektivitas algoritma klasifikasi sangat bergantung pada kesesuaian antara karakteristik data dan prinsip kerja algoritma yang digunakan. Oleh karena itu, pemilihan algoritma seharusnya tidak hanya didasarkan pada kompleksitas model, tetapi juga pada pemahaman mendalam terhadap struktur dan sifat data teks yang dianalisis.

3.6 Implikasi Teoretis dan Praktis

Secara teoretis, hasil penelitian ini menegaskan bahwa algoritma probabilistik sederhana seperti Naïve Bayes masih sangat relevan dan efektif untuk tugas klasifikasi sentimen modern, khususnya pada data teks berdimensi tinggi. Temuan ini menunjukkan bahwa peningkatan kompleksitas model tidak selalu berbanding lurus dengan peningkatan performa klasifikasi.

Secara praktis, hasil penelitian ini membuka peluang implementasi sistem analisis sentimen real-time dengan tingkat akurasi yang sangat tinggi untuk aplikasi streaming seperti Netflix. Sistem analisis sentimen berbasis machine learning memungkinkan pengembang aplikasi melakukan pemantauan opini pengguna secara berkelanjutan, mendeteksi permasalahan layanan secara cepat, serta mendukung pengambilan keputusan strategis berbasis data dalam pengembangan layanan digital.

3.7 Keterbatasan Penelitian dan Arah Penelitian Selanjutnya

Meskipun menghasilkan performa yang sangat tinggi, penelitian ini memiliki beberapa keterbatasan. Dataset yang digunakan hanya berasal dari satu aplikasi dan satu platform, sehingga generalisasi hasil penelitian masih terbatas. Selain itu, penelitian ini hanya menggunakan representasi fitur TF-IDF dan belum mengeksplorasi pendekatan berbasis word embedding atau deep learning.

Penelitian selanjutnya disarankan untuk menggunakan dataset yang lebih besar dan beragam, menerapkan preprocessing multibahasa, serta membandingkan performa algoritma klasik dengan model berbasis transformer seperti BERT. Selain itu, eksplorasi representasi fitur yang mampu menangkap aspek semantik dan kontekstual teks secara lebih mendalam diharapkan dapat meningkatkan performa dan kedalaman analisis sentimen.

4. Kesimpulan

Berdasarkan hasil penelitian dan analisis klasifikasi sentimen ulasan aplikasi Netflix menggunakan tiga algoritma *machine learning*, dapat disimpulkan bahwa Algoritma Naïve Bayes menunjukkan performa terbaik dengan akurasi 99.2%, AUC 1.000, dan nilai metrik evaluasi lainnya (*F1-Score*, *Precision*, *Recall*) yang konsisten di 0.992, serta MCC 0.988. Keunggulan *Naïve Bayes* terletak pada kesederhanaan model dan efektivitasnya dalam menangani data teks berdimensi tinggi yang dihasilkan dari representasi TF-IDF. Algoritma *Random Forest* menempati posisi kedua dengan performa yang sangat kompetitif, mencapai akurasi di 0.980. Meskipun sedikit di bawah *Naïve Bayes*, perbedaan 1.2% ini relative kecil dan menunjukkan bahwa metode *ensemble learning* tetap efektif untuk klasifikasi sentimen. Sebaliknya, algoritma, *K-Nearest Neighbor* (KNN) menunjukkan performa yang sangat rendah dengan akurasi hanya 36.2% dan AUC 0.500 yang setara dengan prediksi acak. Hal ini disebabkan oleh fenomena "*curse of dimensionality*" pada data teks dengan representasi TF-IDF yang berdimensi tinggi, membuktikan bahwa algoritma berbasis jarak tidak cocok untuk tugas klasifikasi *sentiment*. Hasil analisis statistik menunjukkan perbedaan signifikan secara statistik antara *performa Naïve Bayes* dengan KNN ($p < 0.05$), serta *Random Forest* dengan KNN ($p < 0.05$), sementara perbedaan antara *Naïve Bayes* dan *Random Forest* relative kecil meskipun *Naïve Bayes* sedikit lebih unggul. Selain itu, kombinasi *preprocessing* (*case folding*, *tokenization*, *stopword removal*) dan ekstraksi fitur TF-IDF terbukti efektif dalam menghasilkan representasi data optimal untuk klasifikasi *sentiment*, yang memungkinkan pencapaian akurasi sangat tinggi (diatas 98%) pada algoritma *Naïve Bayes* dan *Random Forest*. Secara keseluruhan, penelitian ini membuktikan bahwa algoritma *Naïve Bayes* dengan representasi fitur TF-IDF merupakan kombinasi paling optimal untuk klasifikasi *sentiment* ulasan aplikasi streaming digital seperti Netflix.

Referensi

1. Abdillah, T., Khaira, U., & Hutabarat, B. F. (2024). Komparasi Metode Naive Bayes dan K-Nearest Neighbors terhadap Analisis Sentimen Pengguna Aplikasi Zenius. *Jurnal PROCESSOR*, 19(1), 32–44, <https://doi.org/10.33998/processor.2024.19.1.1596>

2. Ak, A., Callet, P. Le, Gera, A., Tmar, H., Noyes, D., & Katsavounidis, I. (2024). Sustainable Video Streaming using Acceptability and Annoyance Paradigm. 2024 32nd European Signal Processing Conference (EUSIPCO), 952–956. <https://doi.org/10.23919/EUSIPCO63174.2024.10715438>
3. Aldhyani, T. H. H., & Alkahtani, H. (2023). Cyber Security For Detecting Distributed Denial Of Service Attacks in Agriculture 4.0: Deep Learning Model. *Mathematics*, 11(1), 1–19. <https://doi.org/10.3390/math11010233>
4. Breiman, L. (2001). Random Forests. *Machine Learning* 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
5. Budiman, A., Yaputera, R. A., Achmad, S., & Kurniawan, A. (2023). Student attendance with face recognition (LBPH or CNN): Systematic literature review. *Procedia Computer Science*, 216, 31–38. <https://doi.org/10.1016/j.procs.2022.12.108>
6. Chicco, D., & Jurman, G. (2020). The Advantages Of The Matthews Correlation Coefficient (MCC) Over F1 Score And Accuracy In Binary Classification Evaluation. *BMC Genomics*, 21(1), 6. <https://doi.org/10.1186/s12864-019-6413-7>
7. Fan, X., & Meng, D. (2021). Optimal Action Strategy Approach For Desert Crossing Based On Dynamic Programming. *Procedia Computer Science*, 188, 144–154. <https://doi.org/10.1016/j.procs.2021.05.063>
8. N. Khoirunnisaa, K. Nabila Nastiti Kesuma, S. Setiawan, and A. Yunizar Pratama Yusuf. Klasifikasi Teks Ulasan Aplikasi Netflix Pada Google Play Store Menggunakan Algoritma Naïve Bayes Dan Svm. *SKANIKA*, 7(1), 64–73. <https://doi.org/10.36080/skanika.v7i1.3138>
9. Kumalasari, J. T. (2025). Masukan Akademik Terkait Pendekatan Klasifikasi Teks Dan Analisis Sentimen. Universitas BSI.
10. Mathayomchan, B., & Taecharungroj, V. (2020). “How Was Your Meal?” Examining Customer Experience Using Google Maps Reviews. *International Journal of Hospitality Management*, 90, 102641. <https://doi.org/10.1016/j.ijhm.2020.102641>
11. Nahrowi, D., & Hariyanto, D. (2022). Development And Evaluation Of Portable Microcontroller Training System. *Journal of Physics: Conference Series*, 2406(1), 012012. [10.1088/1742-6596/2406/1/012012](https://doi.org/10.1088/1742-6596/2406/1/012012)
12. Pranata, A. B., Abdillah, A. R., & Irwiensyah, F. (2024). Analisis Sentimen Ulasan Pengguna Aplikasi Netflix Pada Google Play Menggunakan Algoritma Naïve Bayes. *KLIK: Kajian Ilmiah Informatika dan Komputer*, 4(6), 3091–3098. <https://doi.org/10.30865/klik.v4i6.1964>
13. Roy, P. R., et al. (2022). A study on paper and author ranking. In *International Conference on Innovations in Science, Engineering and Technology (ICISSET)*, 545–549. [10.1109/ICISSET54810.2022.9775821](https://doi.org/10.1109/ICISSET54810.2022.9775821)
14. Sailunaz, K., & Alhajj, R. (2019). Emotion And Sentiment Analysis From Twitter Text. *Journal Of Computational Science*, 36, 101003.
15. Suhendra, T., Intan, B., & Martadinata, A. T. (2024). Analisis Sentimen Pengguna Aplikasi Netflix Pada Ulasan Google Playstore Menggunakan Metode Naïve Bayes. In *ESCaF 3rd Conference*.
16. Perbandingan Naïve Bayes dan K-Nearest Neighbor untuk Analisis Sentimen Aplikasi Gapura UB Berdasarkan Ulasan Pengguna pada Playstore. (2023). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 7(1), 229-236. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/12142>
17. Velankar, S. (2025). Decoding Millennial Preferences: Drivers Of Ott Platform Adoption In The Digital Era. *International Journal of Advanced Research*, 13(08), 480–484. <https://doi.org/10.21474/IJAR01/21537>
18. Xie, D. (2024). From Incipience To Status Quo: Development From Theatre To Streaming. *Scholarly Review Journal, SR Online: Showcase (Equinox 2024)*, 1–8. <https://doi.org/10.70121/001c.124871>
19. Yadav, A., & Vishwakarma, D. K. (2020). Sentiment analysis using deep learning architectures: A review. *Artificial Intelligence Review*, 53(6), 4335–4385.
20. Yang, X.-P. (2020). Leximax Minimum Solution Of Addition-Min Fuzzy Relation Inequalities. *Information Sciences*, 524, 184–198.
21. Zulkarnain, Z., Mutia, R., Ariani, J. A., Barik, Z. A., & Azmi, H. (2024). Performance Comparison K-Nearest Neighbor, Naive Bayes, And Decision Tree Algorithms For Netflix Rating Classification. *IJATIS*, 1(1), 16–22.