Router: Jurnal Teknik Informatika dan Terapan
Volume. 4 Nomor. 2 Juni 2026
e-ISSN: 3032-3312; p-ISSN: 3026-3611, Hal. 98-109
DOI: https://doi.org/10.62951/router.v4i2.936
Tersedia: https://journal.aptii.or.id/index.php/Router

Optimasi *Naïve Bayes* dengan *Feature Selection* untuk Analisis Sentimen *GTA V Roleplay*

Mohammad Robi[1*], Ali Akbar Rismayadi[2]
[1-2]Teknik Informatika, Universitas Adhirajasa Reswara Sanjaya, Indonesia
Email: mohammadrobi2605@gmail.com[1], ali@ars.ac.id[2]
**Penulis Korespondensi: mohammadrobi2605@gmail.com*

Abstract*. Sentiment analysis is used to identify and classify opinions expressed in text. In GTA V Roleplay player reviews, high-dimensional textual features and noisy data can reduce classification Accuracy and increase computational complexity. This study addresses these challenges by optimizing the Multinomial Naïve Bayes classifier using Chi-Square and Information Gain Feature Selection techniques. The novelty of this research lies in evaluating these methods on Indonesian-language GTA V Roleplay reviews, which contain informal language and gaming-specific vocabulary, a domain that has received limited attention in previous Indonesian studies. The dataset consists of 604 preprocessed reviews represented using TF-IDF and evaluated through 10-fold stratified cross-validation. The results show that Chi-Square achieved the best performance, with 83% Accuracy, 81% Precision, 96% recall, and an F1-score of 88%, outperforming Information Gain in recognizing negative sentiment. These findings provide useful insights for GTA V Roleplay server developers and contribute to the development of Indonesian-language sentiment analysis in the online gaming domain.*

Keywords*: Feature Selection; GTA V Roleplay; Machine Learning; Naïve Bayes; Sentiment Analysis.*

Abstrak. Analisis sentimen digunakan untuk mengidentifikasi dan mengklasifikasikan opini yang terdapat dalam teks. Pada ulasan pemain *GTA V Roleplay*, tingginya dimensi fitur teks serta keberadaan data yang tidak relevan (*noise*) dapat menurunkan akurasi klasifikasi dan meningkatkan kompleksitas komputasi. Penelitian ini bertujuan mengatasi permasalahan tersebut dengan mengoptimalkan algoritma *Multinomial Naïve Bayes* menggunakan teknik seleksi fitur *Chi-Square* dan *Information Gain*. Kebaruan penelitian ini terletak pada evaluasi kedua metode tersebut pada ulasan *GTA V Roleplay* berbahasa Indonesia yang memiliki karakteristik bahasa informal dan kosakata khusus komunitas gim, yang masih jarang diteliti pada studi sebelumnya di Indonesia. Dataset yang digunakan terdiri atas 604 ulasan yang telah melalui tahap prapemrosesan, direpresentasikan menggunakan TF-IDF, dan dievaluasi dengan 10-*Fold* Stratified *Cross Validation*. Hasil penelitian menunjukkan bahwa metode *Chi-Square* memberikan kinerja terbaik dengan akurasi sebesar 83%, presisi 81%, *recall* 96%, dan *F1-score* 88%, serta lebih unggul dibandingkan *Information Gain* dalam mengenali sentimen negatif. Temuan ini memberikan wawasan bagi pengembang server *GTA V Roleplay* dalam memahami persepsi pemain serta berkontribusi pada pengembangan analisis sentimen berbahasa Indonesia di domain gim daring.

Kata Kunci: Analisis Sentimen; *Feature Selection*; *GTA V Roleplay*; *Machine Learning*; *Naïve Bayes*.

1. LATAR BELAKANG

Industri *game* online mengalami perkembangan yang pesat seiring dengan meningkatnya konektivitas internet serta kemajuan teknologi grafis dan komputasi (Arsini et al., 2023). Perkembangan tersebut menjadikan *game* tidak hanya sebagai sarana hiburan, tetapi juga sebagai media interaksi sosial melalui komunitas virtual (NATALIA, 2022). Salah satu bentuk inovasi yang berkembang adalah *Grand Theft Auto* V (*GTA V*) *Roleplay*, yang memungkinkan pemain berperan sebagai karakter dengan berbagai profesi dalam dunia virtual (Hariyono & Arviani, 2024). Interaksi antar pemain menghasilkan banyak ulasan yang berisi opini mengenai kualitas server, pengalaman bermain, maupun aspek teknis lainnya (Legowo, 2022). Ulasan tersebut merupakan sumber informasi yang penting bagi pengembang, namun

Naskah Masuk: 02 Maret 2026; Revisi: 04 April 2026; Diterima: 10 Mei 2026; Terbit: 30 Juni 2026

umumnya berbentuk teks tidak terstruktur sehingga memerlukan analisis sentimen untuk mengolah dan menginterpretasikan opini pengguna secara sistematis (Murtiningsih, 2021).

Salah satu metode yang banyak digunakan dalam analisis sentimen adalah algoritma *Multinomial Naïve Bayes* karena memiliki proses pelatihan yang sederhana, cepat, dan efektif untuk klasifikasi teks (Asnawi et al., 2021). Namun, data ulasan umumnya memiliki jumlah kata unik yang sangat besar (*high dimensionality*), sehingga meningkatkan kompleksitas komputasi dan dapat menurunkan akurasi akibat banyaknya fitur yang tidak relevan atau bersifat *noise* (Zahra & Hasan, 2024). Dua metode yang banyak digunakan untuk tujuan tersebut adalah *Chi-Square* dan *Information Gain* (Ernayanti et al., 2023). Beberapa penelitian sebelumnya menunjukkan bahwa algoritma *Naïve Bayes* mampu menghasilkan performa yang baik pada analisis sentimen di berbagai domain, seperti ulasan *game*, *e-commerce*, maupun media sosial (Jatmiko et al., 2022). Namun, sebagian besar penelitian hanya menerapkan satu teknik *Feature Selection* atau bahkan tidak menggunakannya sama sekali (Rahmantyo, 2023). Selain itu, penelitian analisis sentimen di Indonesia masih didominasi oleh domain *e-commerce* dan media sosial, sedangkan penerapannya pada ulasan *game* online, khususnya *GTA V Roleplay* berbahasa Indonesia, masih sangat terbatas (Ramadhan & Nurasiah, 2025).

Berdasarkan kondisi tersebut, terdapat research gap berupa belum adanya penelitian yang secara komprehensif membandingkan metode *Chi-Square* dan *Information Gain* untuk mengoptimalkan algoritma *Multinomial Naïve Bayes* pada analisis sentimen ulasan *GTA V Roleplay* berbahasa Indonesia yang memiliki karakteristik bahasa informal dan kosakata khas komunitas *game*r. Kebaruan penelitian ini terletak pada penerapan serta perbandingan kedua teknik *Feature Selection* tersebut untuk meningkatkan performa klasifikasi pada data teks yang berdimensi tinggi dan tidak terstruktur. Penelitian ini bertujuan untuk mengoptimalkan algoritma *Multinomial Naïve Bayes* menggunakan teknik *Feature Selection Chi-Square* dan *Information Gain*, membandingkan kinerja kedua metode berdasarkan nilai *Accuracy*, *Precision, recall*, dan *F1-score*, serta menentukan teknik *Feature Selection* yang paling efektif dalam mengklasifikasikan sentimen ulasan *GTA V Roleplay*. Hasil penelitian diharapkan dapat menjadi masukan bagi pengembang server *GTA V Roleplay* dalam memahami persepsi pemain dan memberikan kontribusi terhadap pengembangan metode analisis sentimen berbahasa Indonesia pada domain *game* online.

2. KAJIAN TEORITIS

Analisis sentimen merupakan salah satu penerapan *Natural Language Processing* (NLP) yang digunakan untuk mengidentifikasi dan mengklasifikasikan opini dalam bentuk teks ke dalam kategori sentimen, seperti positif dan negatif (Wardhani et al., 2024). Teknik ini banyak dimanfaatkan untuk menganalisis ulasan pengguna pada berbagai platform digital, termasuk ulasan permainan daring, sehingga dapat memberikan informasi mengenai tingkat kepuasan dan persepsi pengguna terhadap suatu produk atau layanan (Andriyani et al., 2024). Salah satu algoritma yang banyak digunakan dalam analisis sentimen adalah *Naïve Bayes* karena memiliki proses komputasi yang cepat, sederhana, dan mampu menangani data teks berdimensi tinggi (Firmansyah et al., 2024). Namun, performa algoritma ini sangat dipengaruhi oleh kualitas fitur yang digunakan. Oleh karena itu, diperlukan proses *Feature Selection* untuk mengurangi fitur yang tidak relevan sehingga dapat meningkatkan akurasi klasifikasi sekaligus menekan kompleksitas model (Wulandari et al., 2024).

Metode *Feature Selection* yang umum digunakan adalah *Chi-Square* dan *Information Gain*. Kedua metode tersebut bekerja dengan memilih fitur yang memiliki hubungan paling kuat dengan kelas sentimen sehingga hanya kata-kata yang paling informatif yang digunakan pada proses klasifikasi (Septiana et al., 2021). Sebelum proses seleksi fitur dilakukan, data teks terlebih dahulu melalui tahapan *Preprocessing* yang meliputi case *fold*ing, cleaning, tokenizing, stopword removal, dan stemming, kemudian direpresentasikan menggunakan pembobotan TF-IDF (ADLINA, 2022). Beberapa penelitian terdahulu menunjukkan bahwa penerapan *Feature Selection* mampu meningkatkan performa algoritma *Naïve Bayes* pada berbagai kasus analisis sentimen, seperti ulasan *e-commerce*, media sosial, maupun permainan digital (Setiawan et al., 2025). Namun, penelitian mengenai analisis sentimen pada ulasan *GTA V Roleplay*, khususnya dengan membandingkan metode *Chi-Square* dan *Information Gain* sebagai teknik optimasi *Naïve Bayes*, masih relatif terbatas (Jusia et al., 2025). Oleh karena itu, penelitian ini mengkaji efektivitas kedua metode tersebut dalam meningkatkan kinerja klasifikasi sentimen terhadap ulasan pemain *GTA V Roleplay* (Dana et al., 2024).

3. METODE PENELITIAN

Metode penelitian yang digunakan adalah pendekatan kuantitatif dengan metode eksperimen untuk menganalisis sentimen ulasan pemain *GTA V Roleplay* menggunakan algoritma *Multinomial Naïve Bayes* yang dioptimasi melalui teknik *Feature Selection*. Penelitian ini bertujuan membandingkan kinerja metode *Chi-Square* dan *Information Gain* dalam meningkatkan performa klasifikasi sentimen. Seluruh proses pengolahan data dilakukan

menggunakan bahasa pemrograman Python pada lingkungan Google Colab dengan memanfaatkan beberapa *library*, seperti Pandas, Scikit-learn, Sastrawi, dan *Imbalanced-learn*.

Gambar 1. Desain Penelitian.

Tahapan penelitian diawali dengan pengumpulan data ulasan pemain *GTA V Roleplay* yang diperoleh dari berbagai platform daring, yaitu Discord, Facebook, Steam, TikTok, dan forum komunitas. Data yang diperoleh kemudian diseleksi untuk menghilangkan data duplikat, spam, dan ulasan yang tidak relevan. Selanjutnya setiap ulasan diberi label sentimen positif atau negatif secara manual berdasarkan isi opini sehingga diperoleh dataset yang siap digunakan dalam proses klasifikasi. Data yang telah terkumpul kemudian melalui tahap *Preprocessing* untuk meningkatkan kualitas data teks. Tahapan *Preprocessing* meliputi case *fold*ing, cleaning, tokenizing, stopword removal, dan stemming. Case *fold*ing dilakukan dengan mengubah seluruh huruf menjadi huruf kecil, sedangkan cleaning bertujuan menghapus tanda baca, angka, URL, emoji, dan karakter lain yang tidak diperlukan. *Tokenizing* digunakan untuk memisahkan kalimat menjadi kumpulan kata, kemudian stopword removal menghilangkan kata-kata umum yang tidak memiliki pengaruh terhadap sentimen. Tahap terakhir adalah stemming menggunakan *library* Sastrawi untuk mengubah setiap kata menjadi bentuk dasarnya sehingga proses analisis menjadi lebih efektif.

Setelah *Preprocessing* selesai, data direpresentasikan ke dalam bentuk numerik menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF). Metode ini memberikan bobot pada setiap kata berdasarkan tingkat kemunculannya dalam dokumen dan keseluruhan korpus sehingga kata-kata yang memiliki informasi penting memperoleh bobot lebih tinggi. Selanjutnya diterapkan dua metode *Feature Selection*, yaitu *Chi-Square* dan *Information Gain*, untuk memilih fitur yang paling relevan terhadap kelas sentimen. Penerapan *Feature Selection* bertujuan mengurangi dimensi data, menghilangkan fitur yang tidak relevan, serta meningkatkan efisiensi dan akurasi model klasifikasi. Sebelum proses pelatihan model dilakukan, distribusi data diperiksa untuk mengetahui adanya ketidakseimbangan jumlah data pada masing-masing kelas. Untuk mengatasi permasalahan tersebut digunakan metode *Synthetic Minority Over-sampling Technique* (SMOTE) sehingga distribusi data menjadi lebih seimbang. Selanjutnya evaluasi model dilakukan menggunakan teknik 10-*fold cross-validation*, di mana dataset dibagi menjadi sepuluh bagian.

Proses klasifikasi dilakukan menggunakan algoritma *Multinomial Naïve Bayes* dengan tiga skenario pengujian, yaitu *Naïve Bayes* tanpa *Feature Selection*, *Naïve Bayes* dengan *Feature Selection Chi-Square*, dan *Naïve Bayes* dengan *Feature Selection Information Gain*. Kinerja model dievaluasi menggunakan *Confusion Matrix* dengan metrik *Accuracy*, *Precision*, *recall*, dan *F1-score*.

4. HASIL DAN PEMBAHASAN

Dalam penelitian ini, data dikumpulkan dari berbagai sumber seperti media sosial, forum komunitas pemain, dan formulir survei daring yang disebarluaskan secara publik. Data berupa ulasan dari pemain maupun penonton *game GTA V Roleplay* yang kemudian diberi label sentimen berdasarkan isi ulasannya. Label sentimen dibagi menjadi dua kategori, yaitu positif dan negatif, yang merepresentasikan persepsi informan terhadap pengalaman mereka di dalam server *Roleplay*. Jumlah total data yang berhasil dikumpulkan adalah 604 ulasan, terdiri dari: 375 ulasan positif (62.1%); 229 ulasan negatif (37.9%).

Komposisi data ini menunjukkan bahwa mayoritas pemain atau penonton memiliki kecenderungan menyampaikan kesan positif terhadap server dan pengalaman bermain, meskipun tidak sedikit pula yang mengemukakan kritik dan kekecewaan mereka. Ketidakseimbangan proporsi ini menjadi perhatian dalam tahap pelatihan model, karena dapat mempengaruhi performa klasifikasi jika tidak diimbangi dengan teknik seperti resampling atau penyesuaian bobot. Adapun contoh data yang digunakan dalam penelitian ini dapat dilihat pada Tabel 1.

Tabel 1. Contoh Komposisi data.

No	Ulasan	Sentimen
595	Pernah jadi korban *meta-gaming* tapi admin ga tanggepin	Negatif
596	*RP* di kota ini punya event bulanan yang bikin semangat main terus. Ga monoton	Negatif
597	Mekanik *RP* keren, bisa buka bengkel, *hire* karyawan, sampe buka cabang	Positif
598	Ada *RP* persidangan besar, dan gue berperan sebagai jaksa. Tegang tapi seru	Positif
599	Player *toxic* dibiarkan bebas	Negatif
600	Map-nya terlalu luas tapi minim aktivitas. Jadi banyak zona mati	Negatif
601	Gue salut sama admin server Nusa V, fiturnya lengkap	Positif
602	*Motionlife Roleplay* punya fitur pengelolaan bisnis yang interaktif.	Positif
603	Server Kota Kita punya fitur animasi ekspresi wajah dan gesture tangan. Bikin karakter jadi lebih hidup waktu ngobrol IC	Positif
604	Banyak fitur bagus di server Kisah Tanah Air, tapi dokumentasinya minim.	Negatif

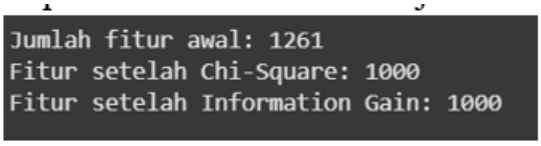
Preprocessing merupakan tahap krusial dalam pengolahan data teks, yang bertujuan untuk menyederhanakan dan membersihkan data dari elemen-elemen yang tidak relevan, agar dapat diproses lebih efektif oleh algoritma machine learning.

Tabel 2. Tahapan *Preprocessing* Data Teks Ulasan *GTA V Roleplay*.

No	Tahap *Preprocessing*	Deskripsi	Contoh Sebelum	Contoh Sesudah
1	*Case Folding*	Mengubah semua huruf menjadi huruf kecil (*lowercase*)	Admin Server Nusa V Hebat	admin server nusa v hebat
2	*Cleaning*	Menghapus angka, tanda baca, emoji, dan karakter khusus	Gue salut sama admin server Nusa V, fiturnya lengkap	Gue salut sama admin server Nusa V fitur lengkap
3	*Tokenisasi*	Memecah teks menjadi satuan kata (token)	fiturnya sangat lengkap	[‘fiturnya’, ‘sangat’, ‘lengkap’]
4	*Stopword Removal*	Menghapus kata-kata umum yang tidak membawa makna penting	fiturnya sangat lengkap dan bagus	fiturnya lengkap bagus
5	*Stemming*	Mengubah kata ke bentuk dasar (*root word*)	memainkan, bermain, permainannya	main

Proses pelabelan sentimen dilakukan secara manual oleh dua anotator independen yang memiliki pengalaman dalam bermain *GTA V Roleplay* dan memahami konteks bahasa komunitas *game*r. Untuk memastikan konsistensi hasil pelabelan, digunakan metrik kesepakatan *Cohen’s Kappa*, yang menghasilkan nilai sebesar 0,87, menunjukkan tingkat kesepakatan yang sangat baik. Perbedaan label yang muncul diselesaikan melalui diskusi hingga tercapai kesepakatan final.

Setelah tahap *Preprocessing* dan representasi fitur menggunakan metode TF-IDF, penelitian ini melanjutkan ke tahap *Feature Selection*. Tujuan utama dari *Feature Selection* adalah untuk mengurangi dimensi data, menghilangkan fitur yang tidak relevan atau berlebihan, serta meningkatkan akurasi dan efisiensi model klasifikasi. Dengan kata lain, proses ini membantu model untuk lebih fokus hanya pada kata-kata atau fitur yang paling berpengaruh terhadap penentuan sentimen. Dataset yang digunakan berjumlah 604 data ulasan, terdiri dari 375 ulasan positif (62,1%) dan 229 ulasan negatif (37,9%). Jumlah kata unik setelah proses TF-IDF mencapai lebih dari 1261 fitur. Oleh karena itu, dilakukan *Feature Selection* menggunakan dua metode statistik yang umum digunakan dalam text classification, yaitu *Chi-Square* dan *Information Gain*.



Gambar 2. Jumlah Fitur Awal dan Setelah Proses *Feature Selection.*

Setelah *Prepocessing* Data lanjut ekstraksi fitur menggunakan TF-IDF: Jumlah fitur yang dihasilkan sebanyak 1.000 fitur (berdasarkan hasil *Feature Selection Chi-Square* atau *Information Gain*). Setiap ulasan direpresentasikan sebagai vektor berdimensi 1000, di mana setiap nilai menunjukkan bobot relatif dari kata tertentu di ulasan tersebut. Nilai-nilai dalam matriks TF-IDF bersifat sparse (banyak nol), karena setiap ulasan umumnya hanya mengandung sebagian kecil dari keseluruhan fitur.

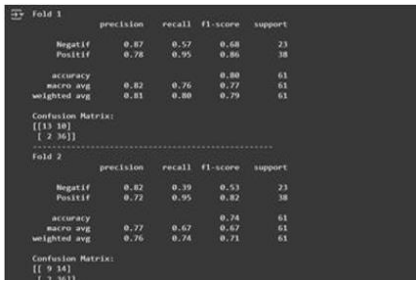
index	abal	abis	absensi	absurd	abuse	acara	acung	ada	adaptasi	addicted	adem	adil	admin	adminnya	adopsi
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Gambar 3. Representasi Data Ulasan Menggunakan TF-IDF.

Setelah dilakukan proses ekstraksi fitur, yakni transformasi data teks mentah menjadi representasi numerik yang dapat dipahami oleh algoritma pembelajaran mesin, langkah selanjutnya dalam alur penelitian ini adalah melakukan pembagian data (data splitting). Ekstraksi fitur ini menjadi tahap krusial karena kualitas representasi fitur sangat menentukan performa model klasifikasi pada tahap-tahap berikutnya, terutama ketika model harus membedakan antara sentimen positif dan negatif secara akurat. Pada penelitian ini, pembagian data dilakukan dengan rasio 80:20, dengan 80% dari keseluruhan data digunakan untuk pelatihan dan 20% untuk pengujian. Teknik pembagian dilakukan secara *stratified split*, yaitu dengan memastikan bahwa proporsi kelas sentimen (positif dan negatif) pada data pelatihan dan pengujian tetap konsisten dengan distribusi aslinya. Teknik stratified split digunakan agar proporsi sentimen positif dan negatif pada kedua subset tetap konsisten, mengingat dataset memiliki ketidakseimbangan kelas (positif 62,1% vs negatif 37,9%).

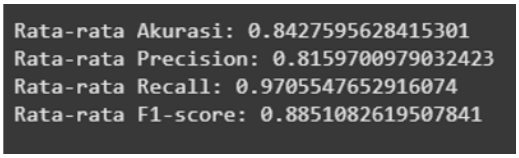
Pendekatan ini sangat penting mengingat dataset yang digunakan memiliki ketidakseimbangan kelas (*class imbalance*), di mana sentimen positif lebih dominan dibandingkan sentimen negatif. Tanpa adanya stratifikasi, pembagian data secara acak berisiko menghasilkan subset data yang tidak representatif, seperti hanya terdiri dari satu jenis kelas sentimen saja, sehingga dapat menyebabkan hasil evaluasi yang bias dan tidak akurat. Selain itu, untuk meningkatkan kehandalan hasil evaluasi, penelitian ini juga menerapkan teknik *cross-validation*, khususnya Validasi silang hingga sepuluh kali lipat. Teknik ini memisahkan data menjadi sepuluh bagian yang sama (*fold*), kemudian bergantian antara menggunakan sembilan *fold* untuk pelatihan dan satu *fold* untuk pengujian, hingga seluruh data terlibat dalam proses pelatihan dan pengujian. Hasil evaluasi akhir merupakan rata-rata dari semua pengujian tersebut.

Dengan demikian, kombinasi antara pembagian data secara stratified dan penggunaan *cross-validation* menjadikan proses evaluasi pada penelitian ini bersifat menyeluruh, objektif, dan representatif terhadap performa sesungguhnya dari model klasifikasi yang dibangun.



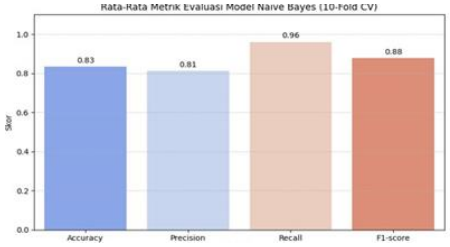
Gambar 4. Hasil Pengujian Model dengan *10-Fold Cross Validation* (*Fold* 1 dan *Fold 2*).

Setelah seluruh data dibagi dan model dilatih menggunakan pendekatan Stratified *K-Fold Cross Validation* dengan nilai k = 10, tahap selanjutnya adalah melakukan evaluasi performa model klasifikasi. Dalam penelitian ini, evaluasi dilakukan terhadap model *Naïve Bayes* yang telah dioptimasi melalui proses *Feature Selection* (*Chi-Square* dan *Information Gain*), serta melalui penerapan teknik SMOTE untuk penyeimbangan data. Evaluasi dilakukan menggunakan metode Stratified *K-Fold Cross Validation* dengan k = 10 membagi data menjadi sepuluh bagian dan mengulangi proses pelatihan dan pengujian sebanyak sepuluh kali. Berdasarkan hasil pengujian terhadap model *Naïve Bayes* menggunakan data yang telah diproses, diekstraksi dengan TF-IDF, dan diseleksi fitur-fiturnya, berikut adalah rata-rata metrik hasil 10-*fold cross-validation*:



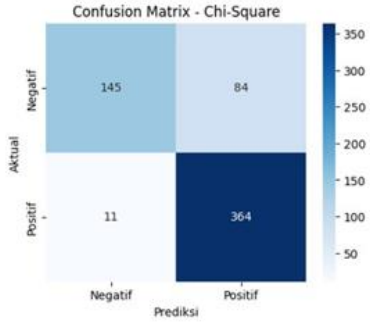
Gambar 5. Rata-Rata Hasil Evaluasi Model.

Gambar 5 menunjukkan rata-rata akurasi, *Precision*, *recall*, dan *F1-score* untuk kedua metode *Feature Selection*. *Chi-Square* terlihat unggul di semua metrik, terutama pada *Precision* dan *F1-score*. Untuk memperkuat pemahaman, ditampilkan visualisasi bar chart dari keempat metrik rata-rata hasil *cross-validation*:



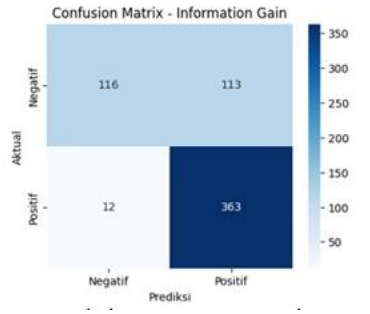
Gambar 6. Rata-Rata Metrik Evaluasi Model *Naïve Bayes*.

Gambar 6 Menampilkan bar chart perbandingan metrik. Batang milik *Chi-Square* lebih konsisten tinggi di semua metrik, mengindikasikan kestabilan performa pada kedua kelas sentimen. *Confusion Matrix* memberikan gambaran yang lebih rinci mengenai jumlah prediksi yang benar dan salah untuk masing-masing kelas, yaitu sentimen positif dan sentimen negatif dalam konteks penelitian ini.



Gambar 7. *Confusion Matrix* Model *Naïve Bayes* dengan *Feature Selection Chi-Square.*

Gambar 7 *Confusion Matrix Chi-Square* menunjukkan nilai True Negative lebih besar dan False Positive lebih kecil dibanding *Information Gain*, yang berarti model lebih akurat mengklasifikasikan ulasan negatif.



Gambar 8. *Confusion Matrix* Model *Naïve Bayes* dengan *Feature Selection Information Gain.*

Gambar 8 memperlihatkan distribusi bobot fitur pada *Information Gain*, di mana beberapa kata kunci sentimen negatif memiliki bobot lebih rendah sehingga berpotensi diabaikan, menjelaskan mengapa kinerjanya lebih rendah dari *Chi-Square.* Berdasarkan hasil evaluasi rata-rata dari *10-fold cross-validation*, model *Naïve Bayes* menunjukkan performa yang cukup tinggi dan stabil, dengan rincian metrik sebagai berikut: *Accuracy*: 83%*; Precision*: 81%*; Recall*: 96%*; F1-score*: 88%. Tingginya nilai *recall* mengindikasikan bahwa model sangat efektif dalam mengenali ulasan dengan sentimen positif. Ini menjadi penting mengingat banyaknya ulasan positif yang ada dalam data awal, dan menunjukkan bahwa model tidak terlalu banyak kehilangan (*False Negative*) pada ulasan yang benar-benar positif.

Penelitian ini menggunakan dua teknik *Feature Selection*: *Chi-Square* dan *Information Gain*, dengan jumlah fitur akhir yang sama yaitu 1000 fitur dari total 1261 fitur awal. Keduanya diuji dalam kondisi data yang sama dan telah diseimbangkan dengan metode SMOTE. Hasil *Confusion Matrix – Information Gain* TP = 363, TN = 116, FP = 113, FN = 12. Model ini memiliki kekuatan tinggi dalam mengenali data positif, namun masih cukup banyak melakukan kesalahan klasifikasi terhadap data negatif (FP tinggi). Hasil *Confusion Matrix – Chi-Square* TP = 364, TN = 145, FP = 84, FN = 11. Model ini menunjukkan kinerja yang lebih seimbang antara prediksi terhadap sentimen positif dan negatif. Nilai *True Negative* meningkat dan False Positive menurun dibandingkan dengan *Information Gain*.

Untuk memvalidasi klaim peningkatan performa akibat SMOTE, dilakukan perbandingan dengan model baseline tanpa penyeimbangan data. Pada data asli, metode *Chi-Square* menghasilkan akurasi 81%, *Precision* 78%, *recall* 94%, dan *F1-score* 85%. Setelah SMOTE, performa meningkat menjadi akurasi 83%, *Precision* 81%, *recall* 96%, dan *F1-score* 88%. Peningkatan terbesar terlihat pada *Precision* kelas negatif yang naik 3% dan penurunan *false positive rate* sebesar 4,6%. Hal ini menunjukkan bahwa SMOTE berhasil mengurangi bias model terhadap kelas mayoritas (positif) dan memperbaiki kemampuan model dalam mengenali ulasan negatif. Secara umum, model *Naïve Bayes* yang telah dioptimasi melalui *Chi-Square Feature Selection* dan SMOTE balancing menunjukkan performa paling optimal dalam klasifikasi sentimen ulasan *GTA V Roleplay.*

5. KESIMPULAN DAN SARAN

Penelitian ini menunjukkan bahwa algoritma *Multinomial Naïve Ba*yes mampu mengklasifikasikan sentimen ulasan *GTA V* Roleplay secara otomatis dengan performa yang baik melalui tahapan *Preprocessing*, ekstraksi fitur TF-IDF, dan evaluasi menggunakan 10-Fold Stratified Cross Validation. Model terbaik diperoleh dengan penerapan feature selection menggunakan metode Chi-Square, yang menghasilkan akurasi sebesar 83%, precision 81%, recall 96%, dan F1-score 88%, serta memberikan performa yang lebih baik dibandingkan Information Gain berdasarkan hasil confusion matrix. Selain itu, penerapan SMOTE terbukti meningkatkan kemampuan model dalam mengenali sentimen negatif dengan mengurangi bias terhadap kelas mayoritas.

DAFTAR REFERENSI

Adlina, F. (2022). *Analisis sentimen online customer review pada toko smartphone Daerah Istimewa Yogyakarta di e-marketplace Shopee menggunakan lexicon based dan word cloud*.

Andriyani, W., Natsir, F., Lubis, H., Tyas, S. H. Y., Meidelfi, D., Faizah, S., Nurlaida, N., Kurniawan, H., Wahyuningtyas, I., & Hasan, F. N. (2024). *Perangkat lunak data mining*. Penerbit Widina.

Arsini, Y., Yani, R., & Nazwa, S. (2023). Pengaruh game online terhadap perilaku sosial pada siswa. *Dewantara: Jurnal Pendidikan Sosial Humaniora, 2*(3), 190–199. https://doi.org/10.30640/dewantara.v2i3.1370

Asnawi, M. H., Firmansyah, I., Novian, R., & Pontoh, R. S. (2021). Perbandingan algoritma naïve bayes, k-NN, dan SVM dalam pengklasifikasian sentimen media sosial. *E-Prosiding Seminar Nasional Statistika | Departemen Statistika FMIPA Universitas Padjadjaran, 10*, 20.

Dana, A. R., Kristananda, R. V., Wibowo, M. B. S., & Prasetya, D. A. (2024). Perbandingan algoritma decision tree dan random forest dengan hyperparameter tuning dalam mendeteksi penyakit stroke. *Prosiding Seminar Nasional Informatika Bela Negara (SANTIKA), 4*, 66–75.

Ernayanti, T., Mustafid, M., Rusgiyono, A., & Hakim, A. R. (2023). Penggunaan seleksi fitur Chi-Square dan algoritma Multinomial Naïve Bayes untuk analisis sentimen pelanggan Tokopedia. *Jurnal Gaussian, 11*(4), 562–571. https://doi.org/10.14710/j.gauss.11.4.562-571

Firmansyah, Y., Kurniawan, R., & Wijaya, Y. A. (2024). Analisis data sentimen pemain game Role-Playing Game (RPG) Honkai Star Rail dengan algoritma Naive Bayes. *Jurnal Informatika dan Rekayasa Perangkat Lunak, 6*(1), 127–135.

Hariyono, M. B., & Arviani, H. (2024). Konstruksi identitas virtual warga Kota Indopride dalam dunia virtual Grand Theft Auto V Roleplay. *JIIP-Jurnal Ilmiah Ilmu Pendidikan, 7*(10), 12060–12069. https://doi.org/10.54371/jiip.v7i10.6141

Jatmiko, H. B., Kurniadi, N. T., & Maulana, D. (2022). Optimasi Naïve Bayes dengan Particle Swarm Optimization untuk analisis sentimen Formula E-Jakarta. *Journal Automation Computer Information System, 2*(1), 22–30. https://doi.org/10.47134/jacis.v2i1.35

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. https://doi.org/10.47065/bulletincsr.v5i4.532

Legowo, Y. A. S. (2022). Gamifikasi dalam pembelajaran di sekolah dasar. *JISPE Journal of Islamic Primary Education, 3*(1), 13–30. https://doi.org/10.51875/jispe.v3i1.43

Murtiningsih, S. (2021). *Filsafat pendidikan video games: Kajian tentang struktur realitas dan hiperealitas permainan digital*. UGM PRESS.

Natalia, R. (2022). *Faktor-faktor yang berhubungan dengan perilaku gangguan kecanduan game online di Desa Batu Horpak Kabupaten Tapanuli Selatan Tahun 2022*.

Rahmantyo, L. E. (2023). Exploring E-Sports: Studi kasus game Pro Evolution Soccer di Indonesia. *Satwika: Kajian Ilmu Budaya dan Perubahan Sosial, 7*(1), 223–236. https://doi.org/10.22219/satwika.v7i1.25496

Ramadhan, M. A. G., & Nurasiah, S. P. (2025). Analisis manajemen risiko terhadap server GTA San Andreas Multiplayer. *Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research, 2*(1b), 1937–1947.

Septiana, R. D., Susanto, A. B., & Tukiyat, T. (2021). Analisis sentimen vaksinasi Covid-19 pada Twitter menggunakan Naive Bayes Classifier dengan feature selection Chi-Squared Statistic dan Particle Swarm Optimization. *Jurnal SISKOM-KB (Sistem Komputer dan Kecerdasan Buatan, 5*(1), 49–56. https://doi.org/10.47970/siskom-kb.v5i1.228

Setiawan, R. P., Irawan, B., & Prihartono, W. P. (2025). Analisis sentimen ulasan Growtopia di Google Play Store menggunakan Naïve Bayes classifier untuk identifikasi kebutuhan pengguna. *Jurnal Informatika dan Teknik Elektro Terapan, 13*(2). https://doi.org/10.23960/jitet.v13i2.6415

Wardhani, D., Astuti, R., & Saputra, D. D. (2024). Optimasi feature selection text mining: Stemming dan stopword untuk sentimen analisis aplikasi SatuSehat. *Innovative: Journal of Social Science Research, 4*(1), 7537–7548.

Wulandari, K. A., Nugraha, A., Luthfiarta, A., & Nisa, L. R. (2024). Peningkatan akurasi deteksi dini penyakit Parkinson melalui pendekatan ensemble learning dan seleksi fitur optimal. *Edumatic: Jurnal Pendidikan Informatika, 8*(2), 575–584. https://doi.org/10.29408/edumatic.v8i2.27788

Zahra, K. H. A., & Hasan, F. N. (2024). Evaluasi user experience pada game GTA V Roleplay Server Indopride menggunakan metode Enhanced Cognitive Walkthrough. *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika, 9*(4), 2292–2302. https://doi.org/10.29100/jipi.v9i4.5659