Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 830−841
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10141

Analisis Sentimen Ulasan Berbahasa Inggris Apex Legends di Steam Menggunakan TF-IDF N-Gram dan Multinomial Naive Bayes

M. Akbar Zidane[1], Yuli Praptomo Pamungkas Hari Sungkowo[2,*]

[1]Teknik Informatika, Sekolah Tinggi Manajemen Informatika dan Komputer El Rahma Yogyakarta, Yogyakarta
Jl. Sisingamangaraja Jl. Karangkajen No.76, Brontokusuman, Kec. Mergangsan, Kota Yogyakarta, Daerah Istimewa Yogyakarta, Indonesia
[2]Sistem Informasi, Sekolah Tinggi Manajemen Informatika dan Komputer El Rahma Yogyakarta
Jl. Sisingamangaraja Jl. Karangkajen No.76, Brontokusuman, Kec. Mergangsan, Kota Yogyakarta, Daerah Istimewa Yogyakarta, Indonesia
Email: [1]akbarzidane12@gmail.com, [2,*]y.praptomo@gmail.com
Email Penulis Korespondensi: y.praptomo@gmail.com
Submitted: **02/06/2026**; Accepted: **04/07/2026**; Published: **05/07/2026**

Abstrak− Pengguna game online Apex Legends terus meningkat seiring dengan komunitas yang selalu aktif dan juga membuat jumlah ulasan pengguna juga ikut meningkat. Dalam kondisi ini, membuat analisis ulasan secara manual menjadi tidak efektif, apalagi karena banyaknya ulasan yang ditulis menggunakan bahasa inggris yang informal, mengandung kata negasi, dan juga menunjukkan distribusi kelas sentimen yang masih tidak seimbang. Dalam penelitian ini tujuannya adalah untuk mengklasifikasikan ulasan dari pengguna Apex Legends di platform Steam kedalam sentimen positif dan negatif ddengan menggunakan algoritma Multinomial Naive Bayes dengan pembobotan TF-IDF berdassarkan fitur N-Gram dengan kombinasi Unigram dan Bigram. Dataset diperoleh melalui web scraping dari platform Steam dengan jumlah total 9.000 ulasan, kemudian dilakukan preprocessing yang menghasilkan 8.981 ulasan yang valid. Akana tetapi data masih menunjukkan ketidakseimbangan kelas, Proses random undersampling kemudian diterapkan untuk memperoleh 5.512 data yang seimbang. Hasil pengujian menunjukkan bahwa model bisa mencapai akurasi sebesar 0,8132 atau 81,32%. Untuk kelas negatif, model memperoleh presisi 0,79, recall 0,85, dan f1-score 0,82, sementara kelas positif memperoleh presisi 0,83, recall 0,78, dan f1-score 0,81. Model yang dilatih juga diterapkan ke dalam dasbor berbasis Streamlit untuk mendukung visualisasi dan prediksi sentimen ulasan baru. Kontribusi penelitian ini adalah penerapan kombinasi fitur N-Gram (unigram dan bigram) pada Multinomial Naive Bayes untuk menangani konteks negasi dan bahasa informal, penggunaan random undersampling untuk mengatasi ketidakseimbangan kelas, serta implementasi model ke dalam dasbor berbasis Streamlit yang memungkinkan visualisasi dan prediksi sentimen ulasan baru secara langsung.

Kata Kunci: Analisis Sentimen; Apex Legends; Multinomial Naïve Bayes; Unigram-Bigram; Steam Reviews; Dashboard;

Abstract−The number of users of the online game Apex Legends continues to increase along with the always active community, which also leads to an increase in the number of user reviews. In this condition, conducting manual review analysis becomes ineffective, especially due to the numerous reviews written in informal English, containing negation words, and also showing an imbalanced sentiment class distribution. In this study, the aim is to classify reviews from Apex Legends users on the Steam platform into positive and negative sentiments using the Multinomial Naive Bayes algorithm with TF-IDF weighting based on N-Gram features with a combination of Unigram and Bigram. The dataset was obtained through web scraping from the Steam platform with a total of 9,000 reviews, followed by preprocessing which resulted in 8,981 valid reviews. However, the data still showed class imbalance. The random undersampling process was then applied to obtain 5,512 balanced data points. The test results show that the model can achieve an accuracy of 0.8132 or 81.32%. For the negative class, the model obtained a precision of 0.79, recall of 0.85, and f1-score of 0.82, while the positive class obtained a precision of 0.83, recall of 0.78, and f1-score of 0.81. The trained model is also applied to a Streamlit based dashboard to support the visualization and prediction of new review sentiments. The contributions of this study are the application of combined N-Gram features (unigram and bigram) to Multinomial Naive Bayes for handling negation context and informal language, the use of random undersampling to address class imbalance, and the deployment of the trained model into a Streamlit-based dashboard that enables direct visualization and sentiment prediction of new reviews.

Keywords: Sentiment Analysis; Apex Legends; Multinomial Naïve Bayes; Unigram-Bigram; Steam Reviews; Dashboard;

1. PENDAHULUAN

Dalam beberapa tahun ini perkembangan game online bisa dilihat dari meningkatnya pengguna, serta meningkatnya ulasan terhadap pengalaman bermain oleh pengguna, Ulasan yang dimaksud berisi tentang banyaknya pendapat mengenai kualitas dalam permainan, kestabilan server, system perbandingan, pembaruan fitur, performa permainan, dan tingkat kepuasan dalam bermain. Pada hal ini Steam adalah platform game yang menyediakan ruang supaya pengguna dari suatu game bisa merekomendasikan game masuk dalam kategori layak atau tidak berdasarkan dari pengalaman bermain yang telah pengguna miliki [1]. Karena hal itu, ulasan pengguna dapat dipandang sebagai sumber data yang sangat bernilai karena langsung memuat opini, dari pemain terhadap suatu game.

Game Apex Legends adalah salah satu game daring kompetitif yang masih memiliki banyak pemain dan juga komunitas yang masih sangat aktif juga. Pada platform Steam, jumlah ulasan pengguna selalu bertambah setiap harinya.Dalam penelitian ini, ulasan yang dianalisis adalah ulasan dari pengguna yang menggunakan Bahasa Inggris yang di peroleh dari game Apex Legends pada platform Steam. Dengan jumlah ulasan yang nilainya banyak, ini juga dapat membantu dalam pemahaman persepsi dari setiap pemain, namun jika dilihat dari sisi lain

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 830−841
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10141

juga masih banyak menimbulkan beberapa permasalahan jika proses analisis masih dilakukan secara manual karena hal ini membutuhkan waktu yang cukup lama, terlebih ketika jumlah ulasan tersebut mencapai ribuan ulasan atau bahkan lebih. Oleh karena itu, jika dilakukan secara manual penilaian dapat bersifat subjektif karena ketergantungan dari pemahaman stiap pembaca mengenai isi dari sebuah ulasan. Maka sangat diperlukan pendekatan yang bisa mengolah data ulasan secara otomatis agar pola sentimen pengguna ddapat diketahui secara lebih baik, lebih cepat dan lebih terukur.

Menjadi salah satu dari banyaknya game daring yang populer pada platform Steam yang memiliki komunitas yang ramai dan masih sangat aktif dengan banyaknya pemain hingga saat ini juga dalam konteks ini Apex Legends adalah game yang dimaksud. Analisis sentimen adalah salah satu dari banyaknya metode yang bisa digunakan sebagai proses pengklasifikasian ulasan dalam bentuk teks. Di dalam kajian tentang pemrosesan bahasa alami, analisis sentimen juga termasuk bagian dari teks klasifikasi yang bertujuan untuk mengenali kecenderungan dalama sebuah opini atay polaritas di dalam data teks.[2],[3]. Hasil dari klasifikasi merupakan ringkasan tentang seberapa baik pengguna menerima sebuah game yang di tawarkan untuk di berikan ulasan. Steam juga sampai saat ini masih memberikan kebebasan akses ke data ulasan melalui endpoint tertentu kepada pengguna. Ini juga mencakup fitur voted_up, yang menggunakan data ulasan tekstual dan rekomendasi dari pengguna untuk menilai apakah suatu perasaan itu positif atau negatif [4]. Karena hal ini juga, ketika membuat model klasifikasi sentimen berbasis mesin pembelajaran, ulasan pengguna Steam untuk game online Apex Legends menyediakan bahan studi yang fantastis.

Namun , tidaklah mudah untuk menyaring ulasan dari pengguna. Data yang yang digunakan dalam penelitian ini adalah ulasan pengguna berbahasa inggris untuk game Apex legends dari platform Steam. Ulasan yang telah diambil tersebut juga tidak semuanya menggunakan bahasa yang formal, banyak dari ulasan pengguna menggunakan bahasa gaul, akronim, bahasa emosi, kosakata khusus di dunia game, dan juga frasa yang tidak selalu mengikuti aturan bahasa baku. Fitur-fitur ini bisa mencegah model dalam memahami makna sebenarnya dari sebuah ulasan. Misalnya, ulasaan yang panjang bisa saja sulit dikategorikan jika didalamnya tergabung pujian dan kritik. Oleh karena itu, tahapan preprocessing sangat penting dilakukan sebelum data dapat digunakan dalam proses pelatihan model, karena data yang berbahasa inggris juga teknik preprocesing harus disesuaikan dengan fitur-fitur tulisan bahasa inggris. Teknik preprocessing juga sangat penting dilakukan supaya data yang telah diambil diproses terlebih dahulu untuk dibersihkan, di normalisasi, dan juga di tata ulang agar algoritma bisa di proses dengan lebih efisien[5]. Selain itu, pemilihan tahapan preprocessing perlu disesuaikan dengan karakteristik data dan model yang digunakan, karena preprocessing dapat memengaruhi performa klasifikasi teks, termasuk pada model pembelajaran mesin konvensional [6],[7].

Salah satu masalah yang sangat harus di pertimbangkan dalam analisis sentimen adalah dengan adanya istilah negatif. Seperti kata-kata "no", "never", dan "not" dapat membuat makna kata dari sebuah pernyataan berubah secara signifikan. Misalnya, frasa "not good" juga tidak dapat diartikan sebagai sentimen yang positif hanya dengan menyertakan kata "good" saja. Jika kata negatif dihapus pada tahap penghapusan (stopwords), maka dari itu makna asli dari sebuah pernyataan dapat berubah dan akan menyebabkan kesalahan pada klasifikasi. Dalam hal ini negasi memiliki peranan penting dalam prosses penentuan polaritas sentimen karena dapat membalikkan makna sebuah opini dalam sebuah pernyataan menurut beberapa penelitian [8], [9]. Kajian terbaru juga menunjukkan bahwa negation handling menjadi salah satu tantangan penting dalam analisis sentimen karena kata negasi seperti not dan never dapat mengubah orientasi sentimen serta berpotensi menyebabkan kesalahan klasifikasi apabila tidak ditangani dengan tepat [10]. Dikarenakan hal ini, kata negasi tetap di pertahankan dalam preprocessing pada tahap penelitian ini untuk menjaga konteks keaslian sentimen dari sebuah ulasan.

Selain itu, variasi jumlah data setiap kelasnya membuat klasifikasi sentimen juga menjadi hal yang menantang, di samping masalah bahasa kasual dan negasi. Jumlah ulasan positif dan negatif sangat tidak seimbang dalam data ulasan. Jika dibiarkan begitu saja, model akan cenderung lebih memahami kelas yang mayoritas daripada kelas minoritas. Maka dari itu secara keseluruhan hasil klasifikasi cukup baik secara umum akan tetapi tidak sebaik ketika mengidentifikasi sentimen untuk kelas dengan data yang lebih sedikit. Metode yang harus dilakukan untuk mengatasi masalah ini adalah dengan proses random undersampling. Proses ini terdiri dari mengurangi data di kelas mayoritas dan menyeimbangkannya dengan kelas minoritas [11]. Pendekatan random undersampling juga banyak digunakan dalam penelitian klasifikasi karena ketidakseimbangan kelas dapat menyebabkan kelas minoritas lebih mudah salah diklasifikasikan [12],[13]. Metode ini digunakan dalam penelitian ini untuk mencapai struktur data yang lebih seimbang sebelum melakukan pelatihan model.

Penelitian sebelumnya meneliti tentang analisis sentimen untuk ulasan game dan ulasan aplikasi dengan beberape metode yang berbeda. Dan juga dalam penelitian tersebut data yang digunakan diambil dari platform Steam dan Metacritic. Dari hasil penelitian tersebut, disimpulkan bahwa metode algoritma machine learning bisa digunakan untuk klasifikasi ulasan pengguna di dalam industri game [14]. Penelitian lain pada domain game juga menunjukkan bahwa ulasan pengguna dari platform seperti Steam dapat digunakan untuk memahami persepsi pemain melalui tahapan preprocessing, ekstraksi fitur, klasifikasi, dan evaluasi model [15],[16]. Penelitian lain tentang ulasan game dari platform steam juga pernah dilakukan tetapi menggunakan metode yang berbeda yaitu metode Random Forest. Pembobotan TF-IDF juga dilakukan menggunakan basis Bigram dan Trigram. Dalam penelitian tersebut menunjukkan bahwa representasi frasa juga digunakan untuk mengidentifikasi pola sentimen pada ulasan game [17]. Ada juga penelitian tentang ulasan aplikasi menunjukkan bahwa menggunakan kombinasi

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 830−841
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10141

dari Unigram dan Bigram didalam Multinomial Naive Bayes juga dapat meningkatkan hasil dari klasifikasi daripada menggunakan fitur tunggal [18]. Selain itu, penelitian tentang komentar dan ulasan online juga menunjukkan bahwa Multinomial Naive Bayes dapat dikombinasikan dengan TF-IDF dan N-Gram untuk meningkatkan representasi fitur pada klasifikasi sentimen [19]. Pada konteks Steam, penelitian lain juga menerapkan Naive Bayes dan TF-IDF untuk mengklasifikasikan opini pengguna, tetapi objek kajiannya masih berfokus pada aplikasi Steam, bukan ulasan game Apex Legends secara khusus [20].

Sebelumnya ada penelitian yang secara khusus juga membahas tentang ulasan game Apex Legends pada Steam akan tetapi penelitian tersebut menggunakan metode Naïve Bayes. Hasil yang di peroleh dari penelitian tersebut yaitu dengan skor akurasi sebesar 75%, dengan precision 83%, recall 82%, dan F1-Score 83% [21]. Hasil tersebut juga menunjukkan bahwa metode Naïve Bayes memiliki potensi untuk digunakan dalam klasifikasi sentimen ulasan game Apex Legend. Namun, masih terdapat ruang pengembangan, terutama dalam sisi jumlah data, penanganan ketidakseimbangan kelas, serta juga pemanfaatan fitur frasa yang juga penting untuk menangkap konteks kata yang saling berdekatan. Perbedaan penelitian ini tidak hanya terletak pada objek dan jumlah data, tetapi juga pada rancangan eksperimen yang digunakan. Penelitian ini menggunakan dataset ulasan terbaru tahun 2026, mempertahankan kata negasi pada tahap preprocessing, menyeimbangkan kelas menggunakan random undersampling, serta menerapkan TF-IDF N-Gram melalui kombinasi Unigram dan Bigram [10],[13],[18],[19]. Oleh sebab itu, penelitian ini menggunakan dataset ulasan terbaru tahun 2026 dengaj jumlah 9.000 data ulasan awal dan terdapat 19 data kosong, sehingga data bersih yang digunakan berjumlah 8.981 ulasan valid. Data tersebut terdiri dari 2.756 ulasan negatif dan 6.225 ulasan positif. Penelitian ini menerapkan random undersampling sehingga diperoleh data seimbang sebanyak 2.756 ulasan positif dan 2.756 ulasan negatif. Selain itu, penelitian ini juga menggunakan pembobotan TF-IDF dengan optimasi fitur N-Gram melalui kombinasi Unigram dan Bigram. Dari sisi hasil, model dalam penelitian ini memperoleh accuracy sebesar 0,8132 atau 81,32%, sehingga menunjukkan peningkatan dibandingkan penelitian Apex Legends sebelumnya yang memperoleh akurasi 75%. Namun, perbandingan tersebut tetap perlu dipahami secara hati-hati karena perbedaan yang ada pada jumlah data, komposisi kelas, tahapan preprocessing, konfigurasi fitur dan algoritma yang digunakan merupakan varian dari Naïve Bayes itu sendiri.

Berdasarkan uraian tersebut, permasalahan utama dalam penelitian ini dirumuskan pada bagaimana mengklasifikasikan sentimen ulasan pengguna game Apex Legends di platform Steam ke dalam dua kategori, yaitu positif dan negatif, dengan tetap mempertimbangkan karakteristik teks ulasan yang ditulis dalam bahasa Inggris informal, kemunculan kata negasi, serta ketidakseimbangan jumlah data pada masing-masing kelas sentimen. Untuk menjawab permasalahan tersebut, penelitian ini menerapkan algoritma Multinomial Naive Bayes yang dioptimalkan melalui ekstraksi fitur TF-IDF N-Gram dengan kombinasi Unigram dan Bigram. Tahap preprocessing dirancang tanpa menghapus kata negasi agar konteks sentimen pada setiap ulasan tetap terjaga, sedangkan ketidakseimbangan kelas ditangani menggunakan teknik random undersampling.

Kontribusi penelitian ini dapat diuraikan ke dalam tiga aspek utama yang sejalan dengan pendekatan yang digunakan. Pertama, penelitian ini memadukan pembobotan TF-IDF dengan fitur N-Gram melalui kombinasi Unigram dan Bigram pada algoritma Multinomial Naive Bayes untuk menangani konteks negasi sekaligus bahasa Inggris informal, sehingga model tidak hanya mengenali kata tunggal, tetapi juga frasa pendek yang membentuk makna sentimen. Kedua, penelitian ini menerapkan teknik random undersampling untuk mengatasi ketidakseimbangan kelas, yaitu dengan menyeimbangkan 8.981 ulasan valid hasil pembersihan dari 9.000 data awal menjadi 5.512 data yang proporsional antara sentimen positif dan negatif. Ketiga, model yang telah dilatih diimplementasikan ke dalam dashboard berbasis Streamlit sebagai media visualisasi hasil analisis sekaligus sarana prediksi sentimen terhadap ulasan baru.

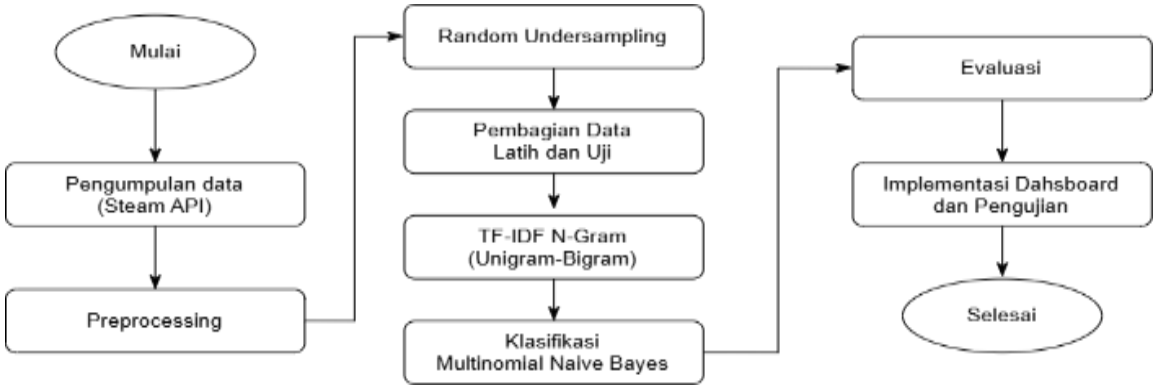
Berdasarkan rumusan tersebut, penelitian ini bertujuan membangun model klasifikasi sentimen pada ulasan pengguna game Apex Legends di platform Steam menggunakan algoritma Multinomial Naive Bayes yang dioptimalkan dengan fitur TF-IDF N-Gram melalui kombinasi Unigram dan Bigram. Kinerja model dievaluasi menggunakan accuracy, precision, recall, f1-score, dan confusion matrix sebagai metrik yang umum digunakan dalam pembelajaran mesin [22]. Hasil klasifikasi selanjutnya diimplementasikan ke dalam dashboard berbasis Streamlit yang menampilkan distribusi sentimen, hasil evaluasi model, serta prediksi sentimen terhadap masukan ulasan baru. Pemilihan Streamlit didasari kemampuannya sebagai framework Python untuk membangun aplikasi data interaktif, sehingga hasil penelitian tidak hanya disajikan dalam bentuk tabel evaluasi, tetapi juga melalui media visualisasi yang lebih komunikatif[23]. Penelitian ini juga dapat menjadi sumber dan bahan untuk penelitian selanjutnya mengenai analisis sentimen, terutama pada ulasan pengguna dalam industri game. Selain itu, hasil penelitian ini dapat memberikan gambaran kepada pemain dan calon pemain terhadap persepsi pengguna pada game Apex Legends.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen pembelajaran mesin (machine learning) untuk mengklasifikasikan sentimen ulasan pengguna game Apex Legends di platform Steam. Metode utama yang diterapkan adalah algoritma Multinomial Naive Bayes dengan ekstraksi fitur TF-IDF N-Gram melalui kombinasi Unigram dan Bigram. Pemilihan metode ini didasarkan pada kemampuannya dalam menangani

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 830−841
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10141

klasifikasi teks berbasis probabilitas serta efektivitasnya dalam merepresentasikan kata tunggal maupun frasa pendek yang membentuk konteks sentimen. Untuk menjaga keaslian makna ulasan, kata negasi tetap dipertahankan pada tahap preprocessing, sedangkan ketidakseimbangan jumlah data antarkelas ditangani menggunakan teknik random undersampling. Secara keseluruhan, penelitian ini dilaksanakan melalui serangkaian tahapan yang saling berurutan, mulai dari pengumpulan data hingga implementasi model ke dalam dashboard. Gambaran umum alur penelitian tersebut ditampilkan pada Gambar 1.



Gambar 1. Alur penelitian

Berdasarkan Gambar 1, alur penelitian disusun untuk menggambarkan keseluruhan proses klasifikasi sentimen ulasan pengguna Apex Legends. Proses diawali dengan pengumpulan data ulasan dari platform Steam, kemudian dilanjutkan dengan preprocessing, random undersampling, ekstraksi fitur TF-IDF N-Gram, pembagian data, klasifikasi menggunakan Multinomial Naive Bayes, evaluasi model, serta implementasi model ke dalam dashboard sebagai media visualisasi dan pengujian ulasan baru. Masing-masing tahapan tersebut diuraikan secara lebih rinci pada subbagian berikut.

2.1 Tahapan Penelitian

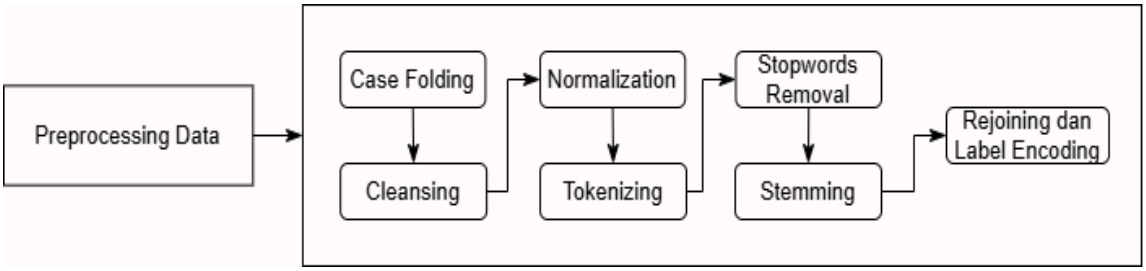
Tahapan penelitian ini dirancang secara berurutan dan saling bergantung, sehingga keluaran dari satu tahap menjadi masukan bagi tahap berikutnya. Rancangan tersebut bertujuan menjaga konsistensi proses sejak pengumpulan data hingga implementasi model ke dalam dashboard. Uraian rinci dari setiap tahap dijelaskan pada subbagian 2.2 sampai 2.9 berikut.

2.2 Pengumpulan Data

Data penelitian diperoleh dari ulasan pengguna Apex Legends pada platform Steam melalui proses web scraping. Atribut yang digunakan meliputi teks ulasan, rekomendasi pengguna, label sentimen, dan waktu pembuatan ulasan. Pada tahap ini, data yang kosong atau tidak sesuai kebutuhan penelitian tidak dimasukkan ke dalam dataset akhir agar data yang digunakan pada tahap berikutnya benar-benar valid.

2.3 Preprocessing Data

Preprocessing dilakukan untuk mengubah teks ulasan mentah menjadi data yang lebih bersih dan siap digunakan dalam pemodelan. Tahapan yang dilakukan meliputi case folding, cleansing, normalization, tokenizing, stopwords removal, stemming, rejoining, dan label encoding. Kata negasi seperti not, no, dan never tetap dipertahankan karena dapat memengaruhi makna sentimen dalam ulasan. Alur tahapan preprocessing data pada penelitian ini ditampilkan pada Gambar 2.



Gambar 2. Alur Tahapan Preprocessing Data

Berdasarkan Gambar 2, tahapan preprocessing dilakukan secara berurutan, dimulai dari case folding untuk menyeragamkan huruf, cleansing untuk menghapus elemen tidak relevan, normalization, tokenizing, stopwords removal, stemming, rejoining, hingga label encoding. Pada tahap stopwords removal, kata negasi seperti not, no,

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 830−841
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10141

dan never sengaja tidak dihapus agar konteks sentimen pada ulasan tetap terjaga. Hasil akhir tahapan ini berupa teks bersih pada kolom final_text yang selanjutnya digunakan pada proses ekstraksi fitur.

2.4 Random Undersampling

Random undersampling digunakan untuk mengatasi ketidakseimbangan jumlah data antara kelas sentimen positif dan negatif. Teknik ini dilakukan dengan mengurangi jumlah data pada kelas mayoritas secara acak hingga seimbang dengan kelas minoritas. Dengan data yang lebih seimbang, model diharapkan tidak terlalu condong mengenali salah satu kelas saja.

2.5 Pembagian Data

Dataset dibagi menjadi data latih dan data uji dengan rasio 80:20 menggunakan fungsi train_test_split. Data latih digunakan untuk membangun model, sedangkan data uji digunakan untuk mengukur kemampuan model dalam memprediksi data baru. Nilai random_state digunakan agar hasil pembagian data tetap konsisten saat program dijalankan ulang.

2.6 Ekstraksi TF-IDF N-Gram

Ekstraksi fitur dilakukan menggunakan metode TF-IDF N-Gram dengan kombinasi Unigram dan Bigram. TF-IDF digunakan untuk mengubah teks menjadi bobot numerik berdasarkan tingkat kepentingan kata dalam dokumen, sedangkan kombinasi Unigram dan Bigram membantu model mengenali kata tunggal maupun pasangan kata yang memiliki makna sentimen. Proses pembobotan dilakukan secara otomatis menggunakan program, sehingga perhitungan bobot TF-IDF tidak dihitung secara manual satu per satu. Rumus pembobotan TF-IDF ditunjukkan pada persamaaan berikut ini.

$$\mathrm{TF-IDF}\,(t,d) = \mathrm{TF}(t,d) \times \mathrm{IDF}(t) \quad (1)$$

$$\mathrm{IDF}\,(t) = \log\left(\frac{N}{\mathrm{DF}(t)}\right) \quad (2)$$

Pada persamaan tersebut, TF(t,d) merupakan frekuensi kemunculan term t dalam dokumen d, IDF(t) menunjukkan bobot kepentingan term terhadap seluruh dokumen, N merupakan jumlah seluruh dokumen, dan DF(t) merupakan jumlah dokumen yang mengandung term tersebut.

2.7 Klasifikasi Multinomial Naïve Bayes

Klasifikasi dilakukan menggunakan algoritma Multinomial Naive Bayes untuk menentukan apakah ulasan termasuk ke dalam sentimen positif atau negatif. Algoritma ini menghitung peluang suatu ulasan masuk ke kelas tertentu berdasarkan fitur teks yang diperoleh dari hasil pembobotan TF-IDF. Proses pelatihan model dan perhitungan probabilitas dilakukan melalui program, sehingga perhitungan manual tidak dilakukan satu per satu pada setiap data ulasan. Adapun dasar perhitungan probabilitas pada algoritma Naive Bayes ditunjukkan pada persamaan di berikut ini.

$$P(C|D) = \frac{P(C|D)\times P(C)}{P(D)} \quad (3)$$

Pada persamaan tersebut, P(C|D) merupakan peluang kelas C berdasarkan dokumen D, P(D|C) merupakan peluang dokumen D muncul pada kelas C, P(C) merupakan peluang awal suatu kelas, dan P(D) merupakan peluang kemunculan dokumen secara umum. Model kemudian membandingkan nilai probabilitas pada setiap kelas sentimen, lalu memilih kelas dengan probabilitas tertinggi sebagai hasil klasifikasi.

2.8 Evaluasi Model

Evaluasi model dilakukan untuk mengetahui kemampuan model dalam mengklasifikasikan sentimen. Metrik yang digunakan meliputi confusion matrix, accuracy, precision, recall, dan f1-score. Nilai evaluasi diperoleh dari perbandingan antara label aktual dan hasil prediksi model. Seluruh nilai evaluasi dihitung secara otomatis oleh program berdasarkan perbandingan antara label aktual dan label hasil prediksi. Persamaan yang digunakan dalam evaluasi model ditunjukkan sebagai berikut.

$$\mathrm{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

$$\mathrm{Precision} = \frac{TP}{TP+FP} \quad (5)$$

$$\mathrm{Recall} = \frac{TP}{TP+FN} \quad (6)$$

$$\mathrm{F1-Score} = 2 \times \frac{\mathrm{Precision}\times\mathrm{Recall}}{\mathrm{Precision}+\mathrm{Recall}} \quad (7)$$

Pada persamaan tersebut, TP merupakan data positif yang diprediksi benar, TN merupakan data negatif yang diprediksi benar, FP merupakan data negatif yang salah diprediksi sebagai positif, dan FN merupakan data

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 830−841
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10141

positif yang salah diprediksi sebagai negatif. Hasil evaluasi dari metrik tersebut digunakan sebagai dasar untuk menilai performa model klasifikasi sentimen.

2.9 Implementasi Dashboard dan Pengujian Ulasan Baru

Dashboard digunakan sebagai media pendukung untuk menampilkan hasil analisis sentimen dalam bentuk visual. Melalui dashboard, pengguna dapat melihat distribusi sentimen, visualisasi data, hasil evaluasi model, serta melakukan pengujian prediksi terhadap ulasan baru. Dashboard tidak menjadi metode utama klasifikasi, tetapi berfungsi sebagai bentuk implementasi dan penyajian hasil penelitian.

3. HASIL DAN PEMBAHASAN

Bagian ini menjelaskan hasil dari setiap tahapan penelitian yang telah dilakukan, mulai dari pengumpulan data, preprocessing, penyeimbangan data, ekstraksi fitur, pembagian data, pelatihan model, evaluasi, hingga implementasi dashboard. Hasil yang ditampilkan disesuaikan dengan proses penelitian pada analisis sentimen ulasan berbahasa Inggris game Apex Legends menggunakan TF-IDF N-Gram dan Multinomial Naive Bayes.

3.1 Hasil Pengumpulan dan Processing Data

Data penelitian diperoleh melalui proses web scraping ulasan pengguna game Apex Legends pada platform Steam. Jumlah data awal yang dikumpulkan sebanyak 9.000 ulasan. Setelah melalui proses preprocessing dan pengecekan ulang pada kolom final_text, ditemukan 19 data kosong yang tidak dapat digunakan pada tahap ekstraksi fitur dan pemodelan. Dengan demikian, jumlah data bersih yang digunakan dalam penelitian ini adalah 8.981 ulasan valid. Rincian hasil seleksi data tersebut ditampilkan pada Tabel 1.

Tabel 1. Hasil Seleksi Data.

No	Tahapan Data	Jumlah Data
1	Data awal hasil web scraping	9000
2	Data final_text kosong	19
3	Data valid akhir (final_text)	8981

Berdasarkan Tabel 1, dapat diketahui bahwa proses seleksi data menghasilkan 8.981 ulasan valid dari total 9.000 data awal. Sebanyak 19 data kosong pada kolom final_text tidak digunakan karena tidak memiliki fitur teks yang dapat diolah pada tahap ekstraksi fitur. Data valid tersebut kemudian digunakan pada tahapan berikutnya, yaitu case folding, cleansing, normalization, tokenizing, stopwords removal, stemming, rejoining, dan label encoding. Untuk memberikan gambaran yang lebih jelas terhadap perubahan teks setelah melalui tahapan preprocessing, contoh hasil sebelum dan sesudah preprocessing ditampilkan pada Tabel 2.

Tabel 2. Contoh Hasil Teks Sebelum dan Sesudah Preprocessing Data.

No	Sebelum Preprocessing Data	Sesudah Preprocessing Data (final_text)	Label
1	Better Stay Away From This Game It's Not Good For Your Mental Health	better stay away game not_good mental health	Negative
2	I know skill issue, but the sounds are still terrible and	know skill issu but sound still terribl never ragey moment game	Negative
3	good game with amazing movement to play with friends. quiet enjoyable	good game amaz movement play friend quiet enjoy	Positive
4	Dont ever spend money on this game things can be a expensive and you never know if your gonna lose your account or not. Like me a lot of people i have met . Also just not worth it	dont ever spend money game thing expens never know gonna lose account not like lot peopl met also not_worth	Negative
5	Used to play this game all the time, not anymore, but the new updates look interesting enough to get back in	use play game time not anymor but new updat look interest enough get back	Positive

Berdasarkan Tabel 2, dapat diketahui bahwa teks hasil preprocessing menjadi lebih ringkas karena elemen yang kurang relevan, seperti huruf kapital, tanda baca, dan kata yang tidak berpengaruh terhadap makna sentimen telah dihilangkan. Namun, kata-kata penting tetap dipertahankan, terutama kata negasi dan kata yang memiliki nilai sentimen. Beberapa contoh yang terlihat pada tabel adalah not good, never, not, not like, not worth, dan but. Kata-kata tersebut dipertahankan karena dapat memengaruhi makna ulasan, misalnya frasa good akan memiliki makna berbeda ketika muncul dalam bentuk not good. Hasil akhir preprocessing berupa kolom final_text kemudian digunakan pada tahap ekstraksi fitur dan klasifikasi sentimen. Selain ditampilkan dalam bentuk contoh

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 830−841
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10141

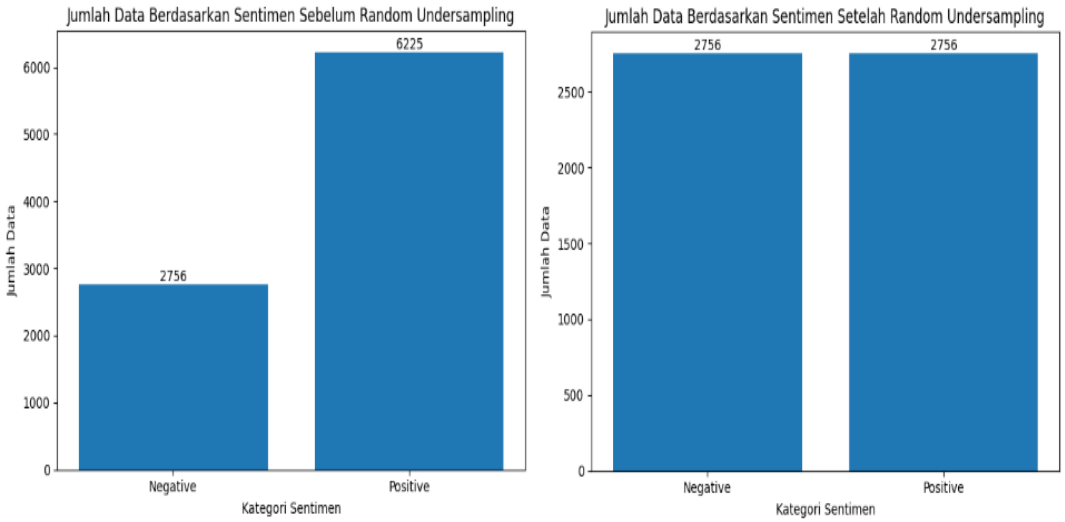
teks pada Tabel 2, hasil preprocessing juga divisualisasikan dalam bentuk wordcloud untuk melihat kata-kata yang paling sering muncul pada kolom final_text. Visualisasi tersebut ditampilkan pada Gambar 3.

Gambar 3. Wordcloud Hasil Preprocessing Ulasan Apex Legends

Gambar 3 menampilkan wordcloud dari hasil preprocessing pada kolom final_text. Visualisasi ini digunakan untuk memberikan gambaran umum mengenai kata-kata yang sering muncul dalam ulasan pengguna setelah teks dibersihkan. Kata dengan ukuran lebih besar menunjukkan frekuensi kemunculan yang lebih tinggi dalam dataset. Hasil ini membantu melihat pola awal pada data ulasan sebelum dilakukan pembobotan fitur menggunakan TF-IDF N-Gram.

3.2 Hasil Random Undersampling

Setelah data bersih diperoleh, dilakukan pengecekan jumlah data berdasarkan kelas sentimen. Hasilnya menunjukkan bahwa data masih belum seimbang, dengan 2.756 ulasan negatif dan 6.225 ulasan positif. Jumlah tersebut memperlihatkan bahwa kelas positif menjadi kelas mayoritas. Untuk mengurangi risiko model terlalu condong mengenali kelas mayoritas, dilakukan proses random undersampling pada kelas positif hingga jumlahnya sama dengan kelas negatif. Perbandingan jumlah data sentimen sebelum dan setelah proses Random Undersampling ditampilkan pada Gambar 4.



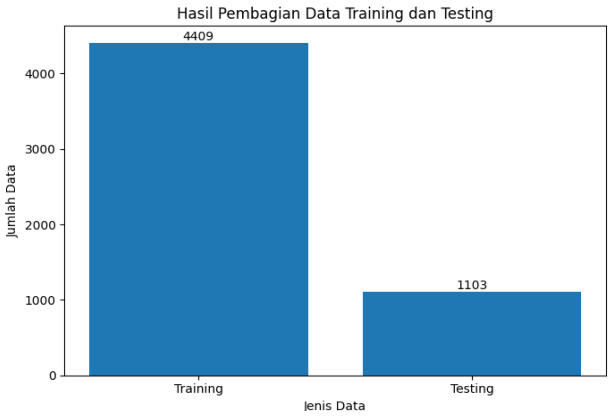
Gambar 4. Perbandingan Jumlah Data Sentimen Sebelum dan Setelah Random Undersampling

Pada Gambar 4, dapat diketahui bahwa sebelum proses balancing, jumlah data pada kelas positif lebih dominan dibandingkan kelas negatif. Kelas positif memiliki 6.225 data, sedangkan kelas negatif memiliki 2.756 data. Setelah dilakukan Random Undersampling, jumlah data pada kelas positif dikurangi hingga menjadi 2.756 data, sehingga jumlah kedua kelas menjadi seimbang. Hasil balancing ini menghasilkan total 5.512 data yang siap digunakan pada tahap pembagian data latih dan data uji.

3.3 Hasil Pembagian Data Latih dan Data Uji

Setelah proses Random Undersampling, dataset dibagi menjadi data latih dan data uji dengan rasio 80:20. Data latih digunakan untuk membangun model, sedangkan data uji digunakan untuk mengukur performa model terhadap data baru. Hasil pembagian data latih dan data uji ditampilkan pada Gambar 5.

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 830−841
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10141



Gambar 5. Hasil Pembagian Data Training dan Testing

Berdasarkan Gambar 5, data hasil balancing sebanyak 5.512 data dibagi menjadi 4.409 data latih dan 1.103 data uji. Pembagian ini menunjukkan bahwa 80% data digunakan untuk proses pelatihan model, sedangkan 20% data digunakan untuk proses pengujian. Dengan pembagian tersebut, model dapat dilatih menggunakan data mayoritas dan diuji menggunakan data yang belum digunakan pada tahap pelatihan.

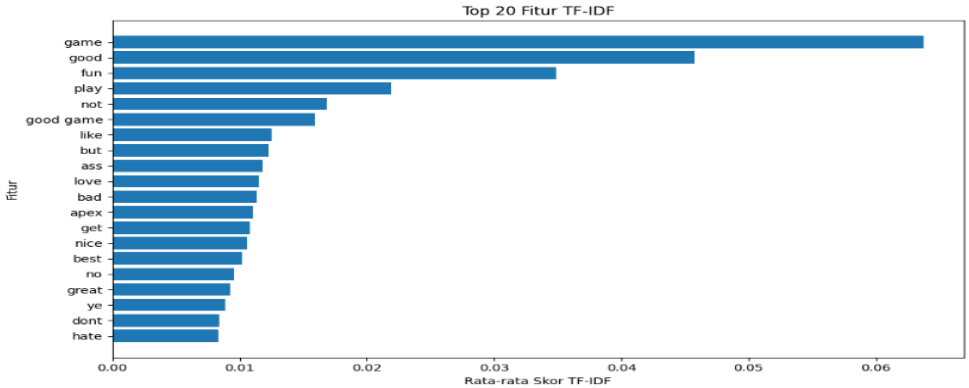
3.4 Hasil Ekstraksi Fitur TF-IDF N-Gram

Setelah data dibagi menjadi data latih dan data uji, tahap berikutnya adalah ekstraksi fitur menggunakan TF-IDF N-Gram. Pada tahap ini, teks ulasan pada kolom final_text diubah menjadi bentuk numerik agar dapat diproses oleh algoritma Multinomial Naive Bayes. Proses TF-IDF dilakukan dengan menerapkan fit_transform pada data latih dan transform pada data uji, sehingga data uji tetap berperan sebagai data baru yang belum dipelajari oleh model. Pendekatan ini juga digunakan untuk menghindari terjadinya data leakage pada proses evaluasi. Konfigurasi ekstraksi fitur TF-IDF N-Gram yang digunakan dalam penelitian ini ditampilkan pada Tabel 3.

Tabel 3. Konfigurasi Ekstraksi Fitur TF-IDF N-Gram Kombinasi Unigram dan Bigram

No	Aspek	Keterangan
1	Sumber data fitur	Kolom final_text
2	Metode Pembobotan	TF-IDF
3	Fungsi yang digunakan	TfidfVectorizer
4	Skema N-Gram	Unigram dan Bigram
5	Transformasi data latih	fit_transform
6	Transformasi data uji	transform
7	Output proses	Matriks fitur numerik
8	Tahap lanjutan	KlasifikasiMultinomial Naïve Bayes

Berdasarkan Tabel 3, penggunaan TF-IDF N-Gram bertujuan untuk memberikan bobot pada kata atau pasangan kata berdasarkan tingkat kepentingannya dalam dokumen. Kombinasi Unigram dan Bigram digunakan agar model tidak hanya mengenali kata tunggal, tetapi juga frasa pendek yang dapat memengaruhi makna sentimen. Misalnya, frasa seperti not good atau no lag memiliki makna yang berbeda jika hanya dilihat sebagai kata terpisah. Oleh karena itu, penggunaan Bigram membantu model menangkap konteks sederhana dalam ulasan pengguna. Selain konfigurasi ekstraksi fitur, penelitian ini juga menampilkan fitur dengan nilai TF-IDF tertinggi untuk mengetahui kata atau pasangan kata yang paling dominan setelah proses pembobotan. Hasil visualisasi fitur TF-IDF tertinggi ditampilkan pada Gambar 6.



Gambar 6. Top 20 Fitur TF-IDF N-Gram dengan kombinasi Unigram dan Bigram

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 830−841
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10141

Berdasarkan Gambar 6, dapat diketahui bahwa beberapa fitur seperti game, good, fun, play, dan not good memiliki nilai bobot TF-IDF yang cukup dominan. Fitur tersebut menunjukkan kata atau pasangan kata yang sering muncul dan memiliki peran dalam membentuk representasi sentimen pada ulasan. Keberadaan fitur Bigram seperti not good juga menunjukkan bahwa proses ekstraksi fitur mampu menangkap konteks negasi, sehingga makna sentimen tidak hanya dilihat dari kata tunggal, tetapi juga dari pasangan kata yang membentuk konteks tertentu.

3.5 Evaluasi Model Multinomial Naive Bayes

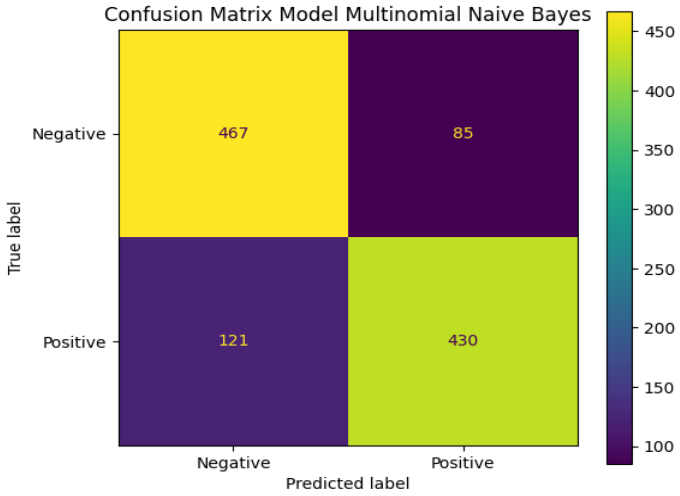
Setelah model Multinomial Naive Bayes dilatih menggunakan data latih, tahap berikutnya adalah pengujian model menggunakan data uji. Evaluasi dilakukan untuk mengetahui kemampuan model dalam mengklasifikasikan ulasan ke dalam kelas sentimen negatif dan positif. Pada penelitian ini, evaluasi model menggunakan confusion matrix, accuracy, precision, recall, dan f1-score. Ringkasan hasil evaluasi model ditampilkan pada Tabel 4.

Tabel 4. Hasil Evaluasi Model Multinomial Naive Bayes

	Precision	Recall	F1-Score	Support
Negative	0.794218	0.846014	0.819298	552.000000
Positive	0.834951	0.780399	0.806754	551.000000
Accuracy	0.813237	0.813237	0.813237	0.813237
Macro avg	0.814585	0.813207	0.813026	1103.000000
Weighted avg	0.814566	0.813237	0.813032	1103.000000

Berdasarkan Tabel 4, model memperoleh nilai akurasi sebesar 0,8132 atau 81,32%. Nilai tersebut menunjukkan bahwa model mampu mengklasifikasikan sebagian besar data uji dengan benar. Pada kelas negatif, model memperoleh nilai precision sebesar 0,79, recall sebesar 0,85, dan F1-score sebesar 0,82. Sementara itu, pada kelas positif, model memperoleh nilai precision sebesar 0,83, recall sebesar 0,78, dan F1-score sebesar 0,81. Nilai F1-score pada kedua kelas yang relatif seimbang menunjukkan bahwa model memiliki performa yang cukup baik dalam mengenali sentimen negatif maupun positif.

Untuk melihat pola kesalahan dan keberhasilan prediksi model secara lebih rinci, hasil klasifikasi juga ditampilkan dalam bentuk confusion matrix pada Gambar 7.



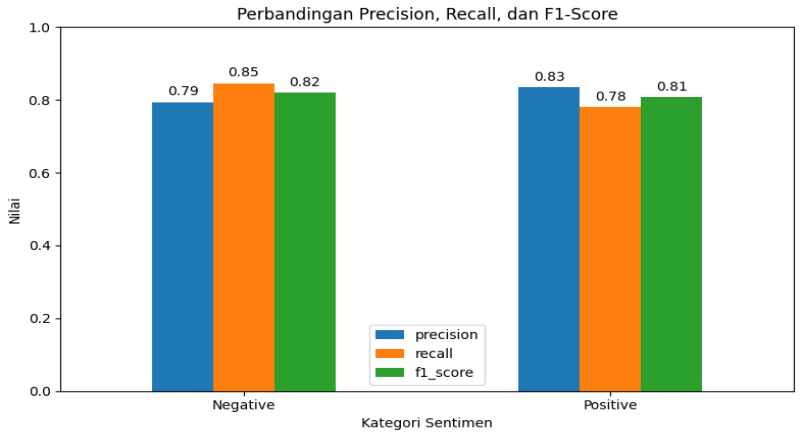
Gambar 7. Confusion Matrix Hasil Klasifikasi

Berdasarkan Gambar 7, model berhasil mengklasifikasikan 467 ulasan negatif dengan benar dan 430 ulasan positif dengan benar. Selain itu, terdapat 85 ulasan negatif yang salah diklasifikasikan sebagai positif dan 121 ulasan positif yang salah diklasifikasikan sebagai negatif. Kesalahan prediksi tersebut dapat terjadi karena beberapa ulasan memiliki struktur kalimat informal, penggunaan kata negasi, atau konteks sentimen yang tidak selalu dapat ditangkap hanya berdasarkan pola kata. Secara umum, hasil confusion matrix menunjukkan bahwa model memiliki kemampuan klasifikasi yang cukup baik, meskipun masih terdapat ruang perbaikan terutama pada pengenalan kelas positif. Selain menggunakan confusion matrix, performa model juga divisualisasikan melalui perbandingan nilai precision, recall, dan F1-score pada masing-masing kelas sentimen. Visualisasi tersebut ditampilkan pada Gambar 8.

Berdasarkan Gambar 8, dapat diketahui bahwa nilai evaluasi pada kedua kelas sentimen relatif seimbang. Pada kelas negatif, model memperoleh nilai precision sebesar 0,79, recall sebesar 0,85, dan F1-score sebesar 0,82. Nilai recall yang lebih tinggi pada kelas negatif menunjukkan bahwa model cukup baik dalam mengenali ulasan negatif yang sebenarnya. Sementara itu, pada kelas positif, model memperoleh nilai precision sebesar 0,83, recall sebesar 0,78, dan F1-score sebesar 0,81. Nilai precision yang lebih tinggi pada kelas positif menunjukkan bahwa

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 830−841
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10141

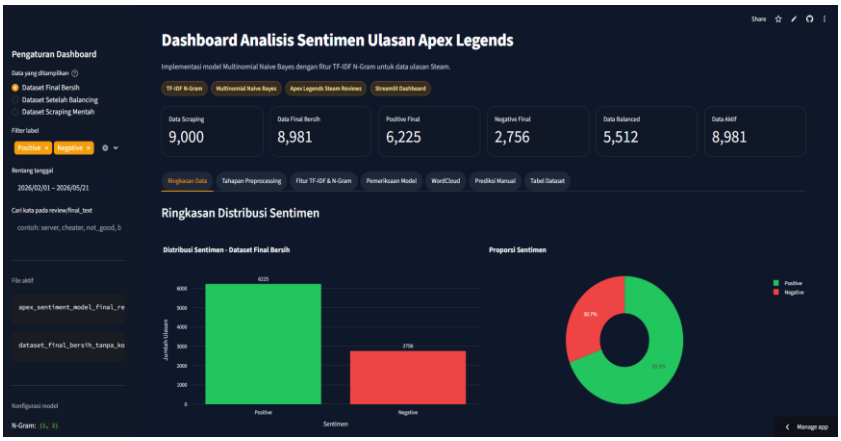
prediksi model terhadap ulasan positif cukup tepat. Secara keseluruhan, nilai F1-score pada kedua kelas yang berada di kisaran 0,81 sampai 0,82 menunjukkan bahwa model memiliki performa klasifikasi yang cukup stabil dalam membedakan sentimen negatif dan positif.



Gambar 8. Perbandingan Precision, Recall, dan F1-Score

3.6 Implementasi Dashboard dan Pengujian Ulasan Baru

Model Multinomial Naive Bayes yang telah dilatih diimplementasikan ke dalam dashboard berbasis Streamlit. Dashboard digunakan untuk menampilkan distribusi sentimen, wordcloud, visualisasi data, hasil evaluasi model, serta pengujian prediksi terhadap ulasan baru. Pada fitur pengujian, teks ulasan yang dimasukkan pengguna diproses melalui tahap preprocessing dan ekstraksi fitur TF-IDF sebelum diklasifikasikan menjadi sentimen positif atau negatif. Tampilan hasil implementasi dashboard ditunjukkan pada Gambar 9.



Gambar 9. Tampilan Dashboard Analisis Sentimen Ulasan Apex Legends

Berdasarkan Gambar 9, dashboard berfungsi sebagai media pendukung untuk menyajikan hasil analisis secara lebih visual dan interaktif. Dashboard ini tidak melakukan pelatihan ulang model, tetapi menggunakan model dan vectorizer yang telah disimpan sebagai dasar dalam proses prediksi.

3.7 Pembahasan

Hasil penelitian ini menunjukkan bahwa model Multinomial Naive Bayes dengan pembobotan TF-IDF N-Gram melalui kombinasi Unigram dan Bigram memperoleh akurasi sebesar 81,32%. Hasil tersebut menunjukkan bahwa metode yang digunakan mampu mengklasifikasikan sentimen ulasan berbahasa Inggris pada game Apex Legends di platform Steam dengan cukup baik. Jika dibandingkan dengan penelitian terdahulu [21] yang juga meneliti ulasan game Apex Legends di Steam menggunakan Naïve Bayes Classifier, penelitian tersebut memperoleh akurasi sebesar 75%, precision sebesar 83%, recall sebesar 82%, dan F1-score sebesar 83%. Dengan demikian, hasil penelitian ini menunjukkan peningkatan dari sisi akurasi dibandingkan penelitian terdahulu, meskipun perbandingan pada metrik precision, recall, dan F1-score tetap perlu dilihat secara hati-hati karena adanya perbedaan rancangan eksperimen.

Perbedaan hasil tersebut dipengaruhi oleh beberapa aspek, terutama pada jumlah dataset, tahapan preprocessing, penanganan ketidakseimbangan kelas, serta metode ekstraksi fitur yang digunakan. Penelitian ini menggunakan 9.000 ulasan awal dan menghasilkan 8.981 ulasan valid setelah preprocessing. Karena data masih tidak seimbang, dilakukan random undersampling sehingga diperoleh 5.512 data seimbang, terdiri dari 2.756

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 830−841
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10141

ulasan negatif dan 2.756 ulasan positif. Selain itu, penelitian ini mempertahankan kata negasi seperti not, no, dan never agar makna sentimen tidak berubah pada tahap preprocessing. Hal ini penting karena ulasan game sering menggunakan bahasa informal, ekspresi emosional, serta frasa negasi yang dapat memengaruhi polaritas sentimen.

Selain itu, penggunaan TF-IDF N-Gram dengan kombinasi Unigram dan Bigram menjadi pembeda utama dalam penelitian ini. Kombinasi tersebut memungkinkan model tidak hanya mengenali kata tunggal, tetapi juga pasangan kata yang memiliki makna sentimen tertentu, seperti not good dan not worth. Dengan pendekatan tersebut, model dapat menangkap konteks sederhana yang mungkin tidak terbaca jika hanya menggunakan fitur kata tunggal. Oleh karena itu, peningkatan akurasi pada penelitian ini dapat dikaitkan dengan penggunaan dataset yang lebih besar, penyeimbangan data, pemertahanan kata negasi, serta pemanfaatan fitur Unigram dan Bigram pada Multinomial Naive Bayes. Namun, penelitian ini tetap memiliki keterbatasan dalam memahami konteks kalimat informal dan ulasan yang mengandung campuran kritik serta pujian, sehingga penelitian selanjutnya dapat membandingkan metode ini dengan algoritma lain atau pendekatan berbasis deep learning.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa analisis sentimen ulasan berbahasa Inggris game Apex Legends pada platform Steam dapat dilakukan dengan menerapkan kombinasi preprocessing teks, random undersampling, ekstraksi fitur TF-IDF N-Gram, dan algoritma Multinomial Naive Bayes. Dari 9.000 data awal hasil web scraping, terdapat 19 data kosong pada kolom final_text yang tidak digunakan, sehingga data valid yang diproses berjumlah 8.981 ulasan. Data tersebut memiliki ketidakseimbangan kelas, yaitu 2.756 ulasan negatif dan 6.225 ulasan positif, sehingga dilakukan random undersampling hingga diperoleh data seimbang sebanyak 5.512 ulasan. Salah satu aspek penting dalam penelitian ini adalah dipertahankannya kata negasi seperti not, no, never, dan bentuk frasa seperti not good atau not worth pada tahap preprocessing. Hal ini dilakukan karena kata negasi dapat mengubah makna sentimen secara langsung, sehingga apabila dihapus sebagai stopwords, konteks asli ulasan berpotensi hilang dan dapat memengaruhi hasil klasifikasi. Penggunaan TF-IDF dengan kombinasi Unigram dan Bigram juga membantu model dalam mengenali kata tunggal maupun pasangan kata yang memiliki makna sentimen tertentu. Hasil pengujian menunjukkan bahwa model Multinomial Naive Bayes memperoleh accuracy sebesar 0,8132 dengan nilai precision, recall, dan f1-score yang relatif seimbang pada kelas positif dan negatif. Model juga telah diimplementasikan ke dalam dashboard berbasis Streamlit sebagai media visualisasi dan pengujian ulasan baru. Meskipun demikian, penelitian ini masih memiliki keterbatasan, terutama pada kemampuan model dalam memahami konteks kalimat informal, campuran kritik dan pujian dalam satu ulasan, serta variasi ungkapan pemain yang tidak selalu dapat ditangkap oleh pendekatan berbasis fitur kata. Oleh karena itu, penelitian selanjutnya dapat mempertimbangkan penambahan jumlah data, pengembangan teknik penanganan negasi yang lebih mendalam, serta perbandingan dengan algoritma lain agar hasil klasifikasi sentimen menjadi lebih optimal.

REFERENCES

[1] Valve Corporation, “User Reviews,” Steamworks Documentation, 2026. Accessed: May 15, 2026. [Online]. Available: https://partner.steamgames.com/doc/store/reviews
[2] D. Jurafsky and J. H. Martin, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models, 3rd ed., 2026. [Online]. Available: https://web.stanford.edu/~jurafsky/slp3/
[3] Y. Mao, Q. Liu, and Y. Zhang, “Sentiment Analysis Methods, Applications, and Challenges: A Systematic Literature Review,” Journal of King Saud University - Computer and Information Sciences, vol. 36, no. 4, 2024, doi: 10.1016/j.jksuci.2024.102048.
[4] Valve Corporation, “User Reviews - Get List,” Steamworks Documentation, 2026. Accessed: May 15, 2026. [Online]. Available: https://partner.steamgames.com/doc/store/getreviews
[5] M. A. Palomino and F. Aider, “Evaluating the Effectiveness of Text Pre-Processing in Sentiment Analysis,” Applied Sciences, vol. 12, no. 17, 2022, doi: 10.3390/app12178765.
[6] J. R. Jim, M. A. R. Talukder, P. Malakar, M. M. Kabir, K. Nur, and M. F. Mridha, “Recent Advancements and Challenges of NLP-Based Sentiment Analysis: A State-of-the-Art Review,” Natural Language Processing Journal, vol. 6, Art. no. 100059, 2024, doi: 10.1016/j.nlp.2024.100059.
[7] M. Siino, I. Tinnirello, and M. La Cascia, “Is Text Preprocessing Still Worth the Time? A Comparative Survey on the Influence of Popular Preprocessing Methods on Transformers and Traditional Classifiers,” Information Systems, vol. 121, Art. no. 102342, 2024, doi: 10.1016/j.is.2023.102342.
[8] L. B. Ilmawan, Muladi, and D. D. Prasetya, “Negation Handling for Sentiment Analysis Task: Approaches and Performance Analysis,” International Journal of Electrical and Computer Engineering (IJECE), vol. 14, no. 3, pp. 3382-3393, 2024, doi: 10.11591/ijece.v14i3.pp3382-3393.
[9] I. N. O. Darmayasa, N. A. Sanjaya ER, I. G. A. G. A. Kadyanan, and A. A. I. N. E. Karyawati, “Pengaruh Teknik Penanganan Negasi dalam Analisis Sentimen,” Jurnal Teknologi Informasi dan Ilmu Komputer, vol. 12, no. 2, pp. 275-282, 2025, doi: 10.25126/jtiik.2025129079.
[10] N. Punetha and G. Jain, “Advancing Sentiment Analysis by Addressing Negation Handling Challenge via Unsupervised Mathematical Approach,” Social Network Analysis and Mining, vol. 15, no. 1, Art. no. 20, 2025, doi: 10.1007/s13278-025-01416-z.

Journal of Information System Research (JOSH)
Volume 7, No. 4, Juli 2026, pp 830−841
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v7i4.10141

[11] S. N. Almuayqil, M. Humayun, N. Z. Jhanjhi, M. F. Almufareh, and N. A. Khan, "Enhancing Sentiment Analysis via Random Majority Under-Sampling with Reduced Time Complexity for Classifying Tweet Reviews," Electronics, vol. 11, no. 21, 2022, doi: 10.3390/electronics11213624.
[12] N. Habbat, H. Nouri, H. Anoun, and L. Hassouni, "Sentiment Analysis of Imbalanced Datasets Using BERT and Ensemble Stacking for Deep Learning," Engineering Applications of Artificial Intelligence, vol. 126, Art. no. 106999, 2023, doi: 10.1016/j.engappai.2023.106999.
[13] C. Kaope and Y. Pristyanto, "The Effect of Class Imbalance Handling on Datasets Toward Classification Algorithm Performance," MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer, vol. 22, no. 2, pp. 227-238, 2023, doi: 10.30812/matrik.v22i2.2515.
[14] J. Y. Tan, A. S. K. Chow, and C. W. Tan, "A Comparative Study of Machine Learning Algorithms for Sentiment Analysis of Game Reviews," IEM Journal, vol. 82, no. 3, pp. 63-68, 2022, doi: 10.54552/v82i3.101.
[15] A. K. Saputra, M. R. Handayani, N. C. H. Wibowo, and K. Umam, "Sentiment Analysis of User Reviews on the Game GTA V Using Support Vector Machine," Jurnal SISFOKOM (Sistem Informasi dan Komputer), vol. 14, no. 3, pp. 284-290, 2025, doi: 10.32736/sisfokom.v14i3.2368.
[16] S. Aura and D. Novianto, "Comparative Sentiment Analysis of TheoTown Reviews on Steam and Google Play Store Using Support Vector Machine," MALCOM: Indonesian Journal of Machine Learning and Computer Science, vol. 6, no. 2, pp. 886-896, 2026, doi: 10.57152/malcom.v6i2.2661.
[17] M. D. Purbolaksono, U. K. Dewi, R. L. Wicaksono, and A. P. Wibowo, "Sentiment Analysis of Game Review Using Random Forest," International Journal on Information and Communication Technology (IJoICT), vol. 10, no. 2, pp. 161-169, 2024, doi: 10.21108/ijoict.v10i2.1007.
[18] A. D. Adyatma, L. Afuan, and E. Maryanto, "The Effect of Unigram and Bigram in the Naïve Bayes Multinomial for Analyzing of Comment Sentiment of Gojek Application in Google Play Store," Jurnal Teknik Informatika (JUTIF), vol. 4, no. 6, pp. 1535-1540, 2023, doi: 10.52436/1.jutif.2023.4.6.1310.
[19] A. Gerliandeva, Y. H. Chrisnanto, and H. Ashaury, "Optimasi Klasifikasi Sentimen pada Komentar Online Menggunakan Multinomial Naïve Bayes dan Ekstraksi Fitur TF-IDF serta N-Grams," Jurnal Pekommas, vol. 9, no. 2, pp. 259-272, 2024, doi: 10.56873/jpkm.v9i2.5585.
[20] A. Kho and F. F. Tampinongkol, "Analisis Sentimen Pengguna Aplikasi Steam dengan Algoritma Naive Bayes," Ranah Research: Journal of Multidisciplinary Research and Development, vol. 7, no. 6, pp. 4620-4627, 2025, doi: 10.38035/rrj.v7i6.1803.
[21] B. J. Rizqullah, L. Afuan, and N. Chasanah, "Sentiment Analysis of Apex Legends Game Reviews on Steam Using Naïve Bayes Classifier," Jurnal Ilmiah Teknik Komputer, vol. 1, no. 1, pp. 41-50, 2026, doi: 10.20884/1.jitk.1.1.33.
[22] A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 3rd ed. Sebastopol, CA: O'Reilly Media, 2022. [Online]. Available:https://www.oreilly.com/library/view/hands-on-machine-learning/9781098125967/
[23] Streamlit, "Streamlit Documentation," Streamlit Documentation, 2026. Accessed: May 24, 2026. [Online]. Available: https://docs.streamlit.io/