

ANALISIS SENTIMEN ULASAN FREE FIRE MENGGUNAKAN NAIVE BAYES LOGISTIC REGRESSION

Aditya Bachtiar ¹⁾, Muhammad Jabbar Hawali ²⁾, Nicolas Clinton Simbolon ³⁾,
Denis Ardan Maulana ⁴⁾, Recha Abriana Anggraini ⁵⁾

^{1,2,3,4,5)} Informatika Universitas Bina Sarana Informatika

15230728@bsi.ac.id ¹⁾, 15230819@bsi.ac.id ²⁾, 15230867@bsi.ac.id ³⁾,
15230823@bsi.ac.id ⁴⁾, recha.rcb@bsi.ac.id ⁵⁾

Abstraksi

Penelitian ini dilatarbelakangi oleh meningkatnya volume ulasan pengguna game Free Fire di Google Play Store yang berpotensi menjadi sumber evaluasi penting bagi pengembang, namun memiliki permasalahan ketidakseimbangan kelas sentimen. Tujuan penelitian adalah membandingkan kinerja algoritma Naïve Bayes dan Logistic Regression dalam analisis sentimen ulasan berbahasa Indonesia. Dataset terdiri dari 12.574 ulasan periode Januari 2024–Juni 2025 yang dikumpulkan melalui web scraping dan diberi label biner berdasarkan rating bintang, dengan dominasi sentimen positif sebesar 86,8%. Proses praproses meliputi case folding, normalisasi emoji dan URL, penghapusan stopword bertema game, stemming bahasa Indonesia, serta ekstraksi fitur TF-IDF n-gram (1,2) yang dioptimalkan menggunakan GridSearchCV 10-fold. Evaluasi difokuskan pada akurasi, presisi, recall, dan F1-score untuk mengakomodasi ketidakseimbangan kelas. Hasil menunjukkan Logistic Regression unggul dengan akurasi 81,63% dan F1-score 87,91%, sedangkan Naïve Bayes memiliki recall tinggi namun akurasi lebih rendah. Temuan ini menegaskan efektivitas Logistic Regression untuk analisis sentimen ulasan game skala besar di Indonesia.

Kata Kunci :

Analisis Sentimen, Free Fire, Naïve Bayes, Logistic Regression, Text Mining

Abstract

This study is motivated by the increasing volume of user reviews of the Free Fire game on the Google Play Store, which has the potential to serve as an important evaluation resource for developers but is affected by sentiment class imbalance. The study aims to compare the performance of Naïve Bayes and Logistic Regression algorithms in sentiment analysis of Indonesian-language user reviews. The dataset consists of 12,574 reviews from January 2024 to June 2025, collected through web scraping and labeled using a binary scheme based on star ratings, with positive sentiment dominating at 86.8%. The preprocessing stages include case folding, emoji and URL normalization, removal of game-related stopwords, Indonesian stemming, and feature extraction using TF-IDF with optimized (1,2) n-grams through 10-fold GridSearchCV. Model evaluation focuses on accuracy, precision, recall, and F1-score to address class imbalance. The results indicate that Logistic Regression outperforms Naïve Bayes with an accuracy of 81.63% and an F1-score of 87.91%, while Naïve Bayes achieves higher recall but lower overall accuracy. These findings confirm the effectiveness of Logistic Regression for large-scale sentiment analysis of game reviews in Indonesia.

Keywords :

Sentiment Analysis, Free Fire, Naïve Bayes, Logistic Regression, Text Mining

Pendahuluan

Industri game mobile di Indonesia terus menunjukkan perkembangan yang signifikan dan menjadi salah satu motor utama pertumbuhan ekonomi digital. Free Fire, sebagai judul battle royale yang konsisten menempati posisi teratas dalam jumlah unduhan dan pemain aktif sejak 2018 hingga pertengahan 2025, menjadi contoh paling mencolok dari fenomena ini. Meski demikian, popularitas tersebut diiringi oleh polarisasi opini yang kuat di Google Play Store. Berbagai keluhan mengenai sistem matchmaking, kinerja teknis, pola monetisasi, hingga kekhawatiran mengenai potensi kecanduan pada kelompok usia

muda menyebabkan munculnya ulasan bernada negatif dalam jumlah signifikan [1]. Situasi ini tidak hanya menggambarkan ketidakpuasan pengguna, tetapi juga menimbulkan risiko penurunan retensi serta memunculkan perhatian regulator terhadap potensi dampak permainan digital terhadap kesehatan mental remaja [2]. Dengan demikian, pemanfaatan analisis sentimen otomatis menjadi krusial untuk membantu pengembang memetakan masalah yang paling mendesak dan menyesuaikan pembaruan game secara lebih tepat.

Sejumlah penelitian sebelumnya sudah menerapkan pendekatan analisis sentimen pada berbagai game

mobile di Indonesia, namun sebagian besar masih terfokus pada judul selain Free Fire dan menggunakan himpunan data yang relatif terbatas. [3] dan [4] misalnya, berhasil mencapai akurasi lebih dari 85% dengan Naïve Bayes dan SVM pada ulasan Mobile Legends dan Roblox, tetapi belum menguji efektivitas algoritma tersebut pada ulasan Free Fire yang cenderung lebih padat slang dan noise. Kajian lainnya oleh [5] serta [6] memperoleh akurasi 82–89% pada Stumble Guys dan eFootball 2024 menggunakan Naïve Bayes, tetapi penelitian dilakukan pada jumlah data yang tidak mencapai 5.000 ulasan dan belum melibatkan perbandingan dengan Logistic Regression yang secara teoretis lebih toleran terhadap ketidakseimbangan kelas.

Keterbatasan menonjol dari penelitian terdahulu terletak pada minimnya studi yang secara spesifik menilai ulasan Free Fire dalam skala besar—mulai dari ratusan ribu hingga jutaan entri—serta kurangnya evaluasi langsung antara Naïve Bayes dan Logistic Regression dengan penyesuaian preprocessing yang dirancang untuk bahasa Indonesia bercampur slang dan emoji. Meskipun [7] telah melakukan analisis aspek pada Mobile Legends dengan Decision Tree, mereka belum memperluas kajian ke algoritma probabilistik klasik yang lebih efisien. Sementara itu, [8] dan [9] berfokus pada pemodelan topik atau platform berbeda seperti Steam dan Twitter sehingga belum menyentuh permasalahan efektivitas model klasik pada konteks Free Fire yang khas dan berjumlah besar.

Kesenjangan pengetahuan tersebut semakin ketara karena sebagian besar penelitian terdahulu menggunakan dataset yang relatif seimbang, padahal distribusi ulasan Free Fire di Indonesia pada 2024–2025 menunjukkan rasio negatif terhadap positif sekitar 3:1 hingga 4:1. Ketidakseimbangan ini membuat model sederhana seperti Naïve Bayes cenderung kurang optimal dibandingkan Logistic Regression yang didukung regularisasi L1/L2. Namun, bukti empiris mengenai performa kedua algoritma tersebut pada data Free Fire aktual dalam bahasa Indonesia masih terbatas, Kondisi ini menunjukkan perlunya evaluasi empiris yang lebih komprehensif terhadap algoritma klasifikasi sentimen yang umum digunakan pada konteks ulasan Free Fire berbahasa Indonesia.

Berdasarkan kondisi tersebut, penelitian ini berupaya mengevaluasi dan membandingkan performa Naïve Bayes dan Logistic Regression dalam mengklasifikasi sentimen pada dari 12.574 ulasan Free Fire yang diambil dari Google Play Store periode Januari 2024–Juni 2025. Penelitian ini mengevaluasi dan membandingkan performa Naïve Bayes dan Logistic Regression berdasarkan akurasi, precision, recall, dan F1-score setelah penerapan preprocessing intensif dan pemodelan fitur TF-IDF. Dua pertanyaan utama yang menjadi fokus

kajian adalah: (1) bagaimana pola distribusi sentimen serta tema-tema dominan yang muncul pada ulasan pengguna Free Fire di Indonesia, dan (2) algoritma mana yang memberikan hasil paling efektif ketika berhadapan dengan data yang tidak seimbang dan kaya noise.

Tinjauan Pustaka

Analisis sentimen merupakan cabang dari Natural Language Processing (NLP) yang berfungsi untuk mengidentifikasi serta menafsirkan opini atau sikap dalam data teks, seperti ulasan pengguna, dengan mengelompokkan hasilnya ke dalam kategori positif, negatif, atau netral. Teknik ini menjadi alat penting untuk memahami persepsi pengguna terhadap suatu produk atau layanan, termasuk dalam konteks penelitian ini yaitu game Free Fire. Penelitian ini menggunakan dua jenis variabel, yakni variabel bebas berupa representasi fitur teks (setelah proses preprocessing dan transformasi seperti TF-IDF) serta variabel terikat berupa kelas sentimen. Algoritma klasifikasi seperti Naïve Bayes dan Logistic Regression digunakan untuk memprediksi kategori sentimen berdasarkan fitur tersebut. Pendekatan ini sejalan dengan penelitian yang diterapkan pada ulasan aplikasi MyPertamina yang memanfaatkan TF-IDF dan membandingkan kedua algoritma tersebut [10].

Naïve Bayes merupakan model probabilistik yang banyak digunakan dalam klasifikasi teks karena kesederhanaannya, kemampuan komputasinya yang cepat, serta efektivitasnya dalam menangani data berdimensi tinggi. Model ini bekerja berdasarkan asumsi independensi antar fitur, sehingga setiap kata memberi kontribusi probabilistik secara terpisah terhadap kelas tertentu; kondisi ini menjadikan NB sering digunakan sebagai model dasar dalam analisis sentimen. Berbagai penelitian telah memanfaatkan NB untuk menganalisis ulasan aplikasi, game, maupun layanan digital [11]. Sementara itu, Logistic Regression menggunakan pendekatan statistik berbasis fungsi logit untuk menghitung probabilitas kelas, sehingga cocok digunakan pada data teks yang telah direpresentasikan ke dalam bentuk numerik melalui TF-IDF. Kombinasi preprocessing, TF-IDF, dan Logistic Regression sering kali menghasilkan kinerja lebih unggul dibanding NB, terutama pada data yang kompleks atau tidak seimbang, sebagaimana dibuktikan dalam studi analisis sentimen pada ulasan produk skin-care [12].

Dalam literatur analisis sentimen, tahap preprocessing seperti pembersihan teks, case folding, penghapusan stopword, tokenisasi, normalisasi kata tidak baku/emoji, serta stemming atau lemmatization dianggap sebagai langkah krusial untuk memastikan kualitas data sebelum digunakan dalam klasifikasi. Setelah itu, representasi fitur seperti TF-IDF atau

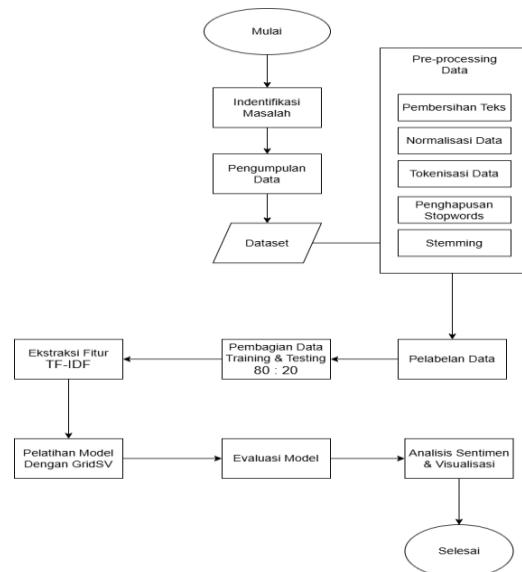
kombinasi N-gram dengan pendekatan berbasis leksikon umum digunakan untuk mengekstraksi informasi penting dari teks. Misalnya, penelitian pada ulasan produk e-commerce menunjukkan bahwa integrasi fitur N-gram dengan Naïve Bayes mampu meningkatkan performa klasifikasi dibandingkan penggunaan pendekatan leksikon saja[13]. Selain itu, evaluasi model sentimen tidak hanya menganalisis akurasi, tetapi juga memerlukan metrik seperti precision, recall, dan F1-score, terutama saat jumlah data tiap kelas tidak seimbang, sebagaimana lazim diterapkan dalam penelitian analisis sentimen terhadap ulasan aplikasi dan game [6].

Beberapa penelitian sebelumnya telah menerapkan Naïve Bayes maupun Logistic Regression pada ulasan aplikasi dan game. Studi pada aplikasi layanan seperti MyPertamina menyatakan bahwa setelah dilakukan preprocessing dan transformasi TF-IDF, Logistic Regression mampu merepresentasikan persepsi pengguna dengan lebih baik[10]. Selain itu, penelitian mengenai ulasan produk skincare di Tokopedia menemukan bahwa Logistic Regression menunjukkan akurasi sekitar 92,3%, lebih tinggi dibandingkan Naïve Bayes yang hanya mencapai sekitar 75,9% [12]. Pada konteks game, penelitian pada ulasan game “OJOL THE GAME” membuktikan bahwa kombinasi NB dan TF-IDF dapat menghasilkan akurasi tinggi, yaitu sekitar 94,12% [14]. Namun, penelitian lainnya pada ulasan game seperti “Mobile Legends” menunjukkan bahwa algoritma lain seperti SVM dapat memberikan performa yang lebih unggul dibanding NB dan Logistic Regression [15].

Meskipun analisis sentimen dengan memanfaatkan algoritma Naïve Bayes (NB) dan Logistic Regression (LR) telah banyak diaplikasikan pada berbagai bidang, namun hasil kajian komparatif yang dilakukan secara sistematis masih menunjukkan variasi yang dipengaruhi oleh karakteristik data serta pendekatan metodologis yang digunakan. Studi pada ulasan aplikasi perpajakan mengungkapkan bahwa Logistic Regression memberikan kinerja yang lebih baik dibandingkan Naïve Bayes setelah melalui tahapan prapemrosesan teks dan pembobotan fitur menggunakan TF-IDF, khususnya pada indikator akurasi dan F1-score, meskipun ruang lingkup penelitian tersebut masih terbatas pada layanan publik[16]. Temuan sejalan juga dilaporkan dalam penelitian analisis sentimen terhadap kebijakan nasional, di mana Logistic Regression menunjukkan performa yang lebih konsisten dibandingkan Naïve Bayes; namun demikian, penelitian tersebut berfokus pada isu kebijakan publik dan belum merepresentasikan karakteristik bahasa informal yang lazim muncul dalam ulasan game daring[17]. Di sisi lain, penelitian yang membandingkan kedua algoritma pada klasifikasi emosi media sosial berbahasa Indo-

nesia menunjukkan bahwa Logistic Regression memiliki kemampuan adaptasi yang lebih baik terhadap variasi bahasa, termasuk penggunaan slang dan ekspresi emosional, terutama setelah diterapkan prapemrosesan dan ekstraksi fitur berbasis TF-IDF[18]. Berdasarkan keseluruhan temuan tersebut, dapat ditegaskan bahwa meskipun Logistic Regression cenderung menunjukkan keunggulan kinerja, masih diperlukan penelitian lanjutan yang secara khusus membandingkan Naïve Bayes dan Logistic Regression pada domain game populer seperti *Free Fire* dengan metodologi yang seragam, sehingga diperoleh pemahaman yang lebih relevan, kontekstual, dan aplikatif bagi analisis sentimen pada game daring.

Metode Penelitian



Gambar 1. Metode dan Langkah Penelitian

Penelitian ini bertujuan untuk membandingkan performa dua algoritma klasifikasi *Naïve Bayes* dan *Logistic Regression* dalam melakukan analisis sentimen terhadap ulasan pengguna *game Free Fire* di *Google Play Store*. Ulasan diklasifikasikan ke dalam dua kategori sentimen, yaitu positif dan negatif.

Pengumpulan Data

Data yang digunakan dalam penelitian ini diperoleh secara langsung melalui proses *web scraping* terhadap ulasan pengguna di *Google Play Store*. Dataset yang berhasil dikumpulkan terdiri dari 12.574 ulasan pengguna *game Free Fire*, yang kemudian digunakan sebagai dasar dalam proses analisis sentimen.

Pra-pemrosesan Data

Tahap pra-pemrosesan dilakukan untuk menyiapkan teks sebelum masuk ke proses analisis. Langkah-langkah yang diterapkan meliputi *case folding*, normalisasi emotikon dan URL, tokenisasi, penghapusan *stopwords*, serta *stemming*. Penghapusan *stopwords* menggunakan kombinasi daftar umum dari pustaka *nltk.corpus* dan daftar *stopwords* bahasa Indonesia dari Sastrawi, yang kemudian diperkaya dengan kosakata khusus dunia gim seperti *afk*, *noob*, *lag*, *garena*, *ff*, dan *skin*. Proses *stemming* dilakukan menggunakan pustaka *StemmerFactory* untuk memastikan setiap kata direduksi ke bentuk dasarnya.

Pelabelan Sentimen

Pelabelan sentimen dilakukan menggunakan pendekatan klasifikasi biner berbasis *rating-based*, di mana ulasan dengan penilaian bintang 1–2 dikategorikan sebagai sentimen negatif, sedangkan ulasan dengan rating bintang 4–5 dikategorikan sebagai sentimen positif. Ulasan dengan rating bintang 3 tidak disertakan dalam penelitian karena bersifat ambigu dan tidak merepresentasikan kecenderungan sentimen yang jelas, sehingga berpotensi menambah *label noise*. Pada skenario biner evaluasi multi-kelas (termasuk netral) disarankan sebagai pekerjaan lanjut.

Pembagian Data

Data dibagi menjadi 80% sebagai data latih dan 20% sebagai data uji menggunakan *train_test_split* dengan metode *stratified sampling*.

Ekstraksi Fitur

TF-IDF digunakan untuk mentransformasikan data teks menjadi representasi numerik. Beberapa parameter diterapkan untuk menjaga konteks frasa penting dan mengurangi noise, dengan konfigurasi sebagai berikut: *lowercase = True*, *strip_accents = 'unicoded'*, *min_df = 3*, *max_df = 0.9*, *ngram_range = (1, 2)*, *sublinear_tf = True*, dan *norm = 'l2'*. Rumus dari TF-IDF

$$W_{t,d} = tf_{t,d} \times idf_t = tf_{t,d} \times \log \frac{N}{df_t} \quad (1)$$

Pelatihan Model

Dua algoritma—Multinomial Naive Bayes dan Logistic Regression dilatih menggunakan data hasil ekstraksi. Proses tuning hyperparameter dilakukan menggunakan GridSearchCV dengan cross-validation sebanyak 10 kali untuk mengoptimalkan performa.

Evaluasi Model

Evaluasi dilakukan dengan metrik *precision*, *recall*, *accuracy*, *F1-score*, dan *confusion matrix* untuk menilai efektivitas masing-masing model dalam mengklasifikasikan sentimen. Metrik Evaluasi dari *Confusion matrix* :

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

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

$$F1-Score = 2 \times \frac{Precision \times Recall}{Recall+Precision} \quad (5)$$

Analisis dan Visualisasi

Analisis lanjutan dilakukan terhadap hasil klasifikasi sentimen untuk mengidentifikasi kecenderungan ulasan pengguna, baik positif maupun negatif. Visualisasi data kemudian digunakan untuk memperjelas pola sentimen yang muncul serta mempermudah interpretasi hasil perhitungan.

Hasil dan Pembahasan

Penelitian ini menerapkan dua algoritma klasifikasi, yaitu *Naive Bayes* dan *Logistic Regression*, untuk menganalisis sentimen ulasan pengguna terhadap permainan Free Fire. Dataset yang digunakan terdiri dari 12.574 ulasan yang diperoleh melalui proses *scraping* pada platform Google Play Store. Tahapan analisis meliputi pra-pemrosesan teks, pelabelan sentimen, pembagian data pelatihan dan pengujian, pelatihan model, serta evaluasi performa kedua algoritma tersebut. Hasil penelitian ini juga menyajikan perbandingan kinerja antara *Naive Bayes* dan *Logistic Regression* dalam melakukan klasifikasi sentimen pada ulasan pengguna.

Hasil Pra-pemrosesan dan Pelabelan

Tahapan pra-pemrosesan mencakup proses pembersihan teks, normalisasi, *tokenization*, penghapusan *stopwords*, serta *stemming* untuk menghasilkan representasi teks yang lebih terstruktur dan siap untuk dianalisis.

Tabel 1. Hasil *preprocessing*

No	Ulasan	Sebelum dilakukan <i>Preprocessing</i>	Ulasan Setelah dilakukan <i>Preprocessing</i>
1	Tolong Garena cara garena ganti nama gi ganti nama nya gi mana soalnya nama saya jelek mohon garena	garena ganti nama gi mana nama jelek garena	

2	Garena Indonesia tolong kok aku layar nya hitam ya ff tolong dibantu ya	garena indonesia aku layar hitam ff bantu
3	karna tidak ada diamon geratis 120 DM masak harus topup dulu baru geratis DM ff	diamon gratis topup gratis ff

Pelabelan dan Pembagian Data Training dan Testing

Data ulasan yang telah melalui tahap *preprocessing* kemudian diberikan label sentimen berdasarkan nilai ulasan yang tersedia. Setelah proses pelabelan selesai, dataset dibagi menjadi dua bagian dengan rasio 80:20 untuk keperluan pelatihan dan pengujian. Pembagian data dilakukan menggunakan metode *stratified split* agar proporsi setiap kelas sentimen tetap seimbang pada kedua subset. Selanjutnya, fitur teks diekstraksi menggunakan metode *Term Frequency–Inverse Document Frequency (TF-IDF)* sehingga setiap ulasan dapat direpresentasikan dalam bentuk numerik dan siap digunakan pada proses pelatihan model klasifikasi.

Tabel 2. Hasil Pelabelan Data

Positif	Negatif
9888	1506

Tabel 3. Hasil Pembagian Data

Positif	Negatif
<i>Training (80%)</i>	
7910	1205
<i>Testing (20%)</i>	
1978	301

Pelatihan Model dan Evaluasi

Dua model klasifikasi dilatih menggunakan fitur teks yang telah direpresentasikan dalam bentuk *TF-IDF*. Proses pelatihan dilengkapi dengan optimasi *hyperparameter* melalui *GridSearchCV* menggunakan skema *cross-validation* 10-fold untuk memastikan performa model yang lebih stabil. Setelah proses tersebut, evaluasi dilakukan dengan menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score* sebagai indikator kinerja masing-masing model.

Tabel 4. Hasil Akurasi dengan <i>GridSearchCV</i> dan <i>CrossValidation</i>	
<i>Naive Bayes</i>	<i>Logistic Regression</i>
93.29%	95.75%

Hasil evaluasi menunjukkan bahwa kedua model memberikan performa tinggi dan stabil. *Logistic Regression* memperoleh hasil terbaik dengan akurasi 95.75%

Visualisasi dan Analisis Sentimen

Perhitungan kata dilakukan menggunakan *Document Frequency* (DF), *Term Frequency* (TF), *Inverse Document Frequency* (IDF), serta kombinasi *Term Frequency–Inverse Document Frequency (TF-IDF)*. Proses ini bertujuan mengidentifikasi frekuensi kemunculan kata dalam korpus dan menentukan bobot kata yang paling berpengaruh dalam analisis sentimen. Dengan metode ini, kata-kata yang sering muncul namun tidak bersifat umum memperoleh bobot lebih tinggi sehingga dapat membantu mengungkap pola dan kecenderungan ulasan pengguna game *Free Fire*.

Tabel 5. Frekuensi Kata

No	Kata	n
1.	Tolong	8371
2.	Seru	7485
3.	Nya	6001
4.	Main	4856
5.	Garena	2712
6.	Game	2649
7.	Gak	2618
8.	Ff	2468
9.	Banget	696
10.	Bagus	585



Gambar 2. Frekuensi Kata

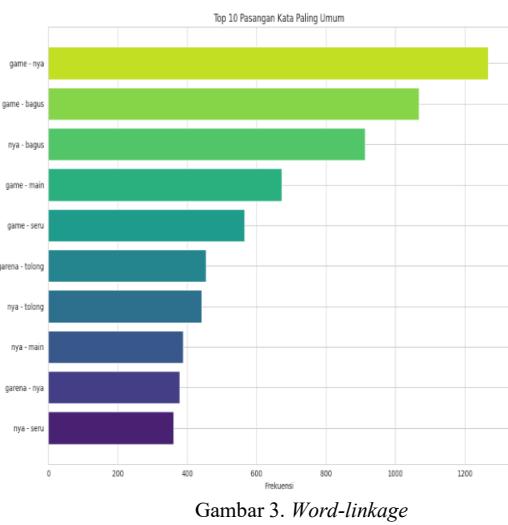
Tahap selanjutnya adalah penerapan metode *TF-IDF* (*Term Frequency–Inverse Document Frequency*) untuk pembobotan kata. Hasil dari pembobotan *TF-IDF* disajikan pada Tabel 7. Proses selanjutnya adalah penentuan keterkaitan antarkata (*word-linkages*), dan hasilnya ditunjukkan pada Tabel 8. Selain itu, hasil visualisasi *Wordcloud* dari ulasan sentimen disajikan pada Gambar 4.

Tabel 6. *TF-IDF*

No	Kata	N	TF	DF	IDF	TF-IDF
1.	tolong	8371	1427	1204	2.345152	230.789102
2.	seru	7485	1618	1495	2.128836	487.442078
3.	nya	6001	3395	2594	1.578045	554.516715
4.	main	4856	2455	1991	1.842492	445.088049
5.	garena	2712	1555	1323	2.250974	278.552620
6.	game	2649	4207	3445	1.294417	746.993400
7.	gak	2618	1040	834	2.711955	206.247569
8.	ff	2468	1682	1384	2.205931	391.558070
9.	banget	696	987	903	2.632557	285.404104
10.	bagus	585	3046	2722	1.529897	754.914789

Tabel 7. *Word-linkage*

No	Kata	n
1.	game-nya	1267
2.	game-bagus	1068
3.	nya-bagus	913
4.	game-main	672
5.	game-seru	565
6.	garena-tolong	454
7.	nya-tolong	442
8.	nya-main	389
9.	garena-nya	379
10.	nya-seru	361



Gambar 3. *Word-linkage*

Visualisasi Wordcloud menunjukkan kata-kata yang dominan muncul dalam masing-masing kategori sentimen. Untuk ulasan positif, kata seperti “bagus”, “seru”, dan “baik” sering muncul, sedangkan untuk ulasan negatif kata seperti “burik”, “bug”, dan “update” mendominasi.

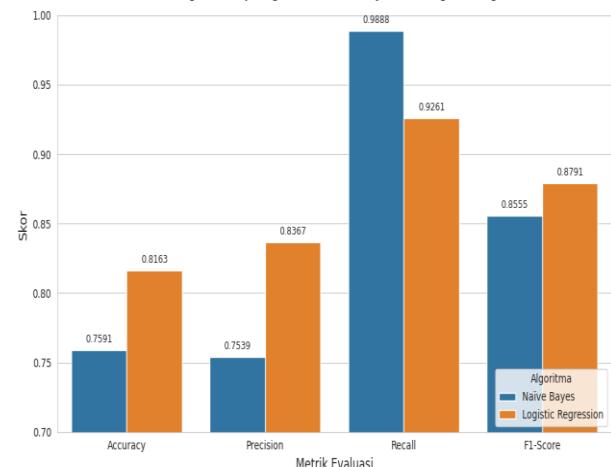


Gambar 4 Visualisasi Wordcloud Positif & Negatif

Perbandingan Pengujian Kinerja Model

Sub-bab ini bertujuan untuk membandingkan performa model *Naïve Bayes* dan *Logistic Regression* dalam klasifikasi sentimen ulasan. Perbandingan dilakukan berdasarkan metrik evaluasi yang disajikan pada Gambar 5 dan data rinci dari *Confusion Matrix* (Tabel X dan Tabel Y).

Perbandingan Kinerja Algoritma Naïve Bayes dan Logistic Regression



Gambar 5. Hasil Visualisasi Perbandingan Kedua Algoritma

Secara keseluruhan, hasil perbandingan kinerja model, yang tercermin pada Gambar 5, menunjukkan bahwa *Logistic Regression (LR)* unggul pada sebagian besar metrik evaluasi. *LR* mencatat nilai Akurasi 81.63% dan Precision 83.67% yang lebih tinggi dibandingkan *Naïve Bayes* 75.91% dan 75.39%, yang menandakan kemampuan prediksi yang lebih akurat secara keseluruhan dan minimnya *False Positives*. Meskipun *Naïve Bayes* menunjukkan *Recall* yang sangat tinggi 98.88% yang berarti model ini efektif dalam mengidentifikasi hampir semua ulasan positif yang sebenarnya (*True Positives*)—tingginya nilai *recall* tersebut diimbangi oleh *Precision* yang lebih rendah. Kinerja paling kritis, *F1-Score*, yang merupakan ukuran keseimbangan model pada data yang tidak seimbang, dikuasai oleh *Logistic Regression* 87.91%. Oleh karena itu, *LR* dinilai sebagai model yang paling stabil dan optimal karena memberikan keseimbangan terbaik antara *Precision* dan *Recall* dalam kasus klasifikasi sentimen ini.

Tabel 8. *Confusion Matrix Naïve Bayes (NB)*

Tabel 3. Confusion Matrix Naive Bayes (2021)			
Prediksi	Negatif	Positif	Total Sebenarnya
Negatif	108(TN)	550 (FP)	658
Sebenarnya			
Positif	19 (FN)	1685 (TP)	1704
Sebenarnya			

Analisis terhadap *Confusion Matrix* model *Naïve Bayes* menunjukkan bahwa model ini sangat unggul dalam mengidentifikasi kelas Positif. Jumlah *True Positives (TP)* yang dicapai adalah 1.685, yang berkontribusi langsung pada nilai *Recall* yang sangat tinggi, yaitu 98,88%. Artinya, *Naïve Bayes* berhasil

menangkap hampir semua ulasan Positif yang ada di dalam data uji. Namun, keunggulan ini disertai dengan kelemahan signifikan pada metrik Precision. Model NB cenderung terlalu optimis dalam prediksinya, yang terbukti dari tingginya jumlah *False Positives (FP)* sebesar 550. Nilai FP ini mewakili ulasan Negatif yang salah diklasifikasikan sebagai Positif, dan tingginya angka ini menjadi penyebab utama rendahnya nilai *Precision* 75.39% model *Naïve Bayes*.

Tabel 9. Confusion Matrix Logistic Regression (LR)

Prediksi	Negatif	Positif	Total Sebenarnya
Negatif	350	308 (FP)	658
Sebenarnya (TN)			
Positif	126 (FN)	1578 (TP)	1704
Sebenarnya			

Analisis terhadap *Confusion Matrix* model *Logistic Regression (LR)* menunjukkan peningkatan kinerja yang signifikan dalam penanganan kelas minoritas. Model *LR* berhasil mengidentifikasi 350 *True Negatives (TN)*, yang jauh lebih tinggi dibandingkan hanya 108 yang dicapai oleh *Naïve Bayes*. Kemampuan *LR* yang lebih baik dalam mengklasifikasikan ulasan Negatif dengan benar ini secara langsung menyebabkan jumlah *False Positives (FP)* yang jauh lebih rendah, yaitu hanya 308 (berbanding 550 pada *NB*). Rendahnya *False Positives* inilah yang menjadi alasan utama mengapa nilai *Precision Logistic Regression* (0.8367) secara substansial lebih tinggi, karena probabilitas prediksi Positifnya jauh lebih dapat diandalkan.

Kesimpulan dan Saran

Berdasarkan hasil penelitian, dapat disimpulkan bahwa ulasan pengguna game Free Fire di Google Play Store Indonesia memiliki ketidakseimbangan kelas yang signifikan, dengan sentimen positif mendominasi sekitar 86,8% dari total data. Hasil analisis sentimen menunjukkan bahwa ulasan positif umumnya berkaitan dengan aspek keseruan dan kualitas permainan, sedangkan ulasan negatif didominasi oleh permasalahan teknis dan pengalaman pengguna, seperti bug, pembaruan, dan performa sistem. Dalam pengujian kinerja model, Logistic Regression terbukti memberikan performa yang lebih optimal dan seimbang dibandingkan *Naïve Bayes*, khususnya pada metrik precision dan F1-score, yang sangat penting dalam kondisi data tidak seimbang. Meskipun *Naïve Bayes* menghasilkan nilai recall yang sangat tinggi, model ini cenderung menghasilkan false positives dalam jumlah besar, sehingga menurunkan ketepatan klasifikasi. Dengan demikian, penelitian ini menegaskan bahwa Logistic Regression merupakan metode yang lebih andal untuk analisis sentimen ulasan Free Fire berbahasa Indonesia dalam skala besar.

Saran

Penelitian selanjutnya dapat dikembangkan dalam beberapa aspek. Dari sisi pelabelan data, klasifikasi sentimen dapat diperluas menjadi tiga kelas dengan menambahkan kategori netral atau mengombinasikan pendekatan rating-based dan lexicon-based guna mengurangi noise anotasi. Dalam menangani ketidakseimbangan kelas, penerapan teknik seperti SMOTE, NearMiss, atau penyesuaian bobot kelas pada Logistic Regression berpotensi meningkatkan kinerja model pada kelas minoritas. Selain itu, proses pra-pemrosesan dapat diperkuat dengan kamus normalisasi yang lebih kaya terhadap istilah slang dan jargon khas game Free Fire agar representasi teks menjadi lebih akurat. Penelitian lanjutan juga disarankan untuk membandingkan metode klasik dengan pendekatan deep learning seperti Bi-LSTM, BERT multilingual, atau IndoBERT pada dataset yang sama. Terakhir, penerapan analisis sentimen berbasis aspek dapat memberikan wawasan yang lebih spesifik bagi pengembang dalam mengevaluasi elemen tertentu, seperti sistem matchmaking, grafis, atau pembaruan game.

Daftar Pustaka

- [1] R. P. Setiawan, B. Irawan, and W. P. Prihartono, "Analisis Sentimen Ulasan Growtopia Di Google Play Store Menggunakan Naïve Bayes Classifier Untuk Identifikasi Kebutuhan Pengguna," *J. Inform. dan Tek. Elektro Terap.*, vol. 13, no. 2, 2025, doi: 10.23960/jitet.v13i2.6415.
- [2] M. K. Alamsyah and N. Pratiwi, "Analisis Sentimen Terkait Opini Masyarakat Terhadap Perkembangan E-Sport Mobile Di Indonesia Menggunakan K Nearest Neighbor," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 9, no. 1, pp. 349–359, 2024, doi: 10.29100/jipi.v9i1.4927.
- [3] P. G. Yehova, J. Jessica, and M. I. Jambak, "Perbandingan Metode Naïve Bayes dan SVM pada Ulasan Google Playstore Mobile Legends Bang Bang," *Device*, vol. 14, no. 1, pp. 140–149, 2024, doi: 10.32699/device.v14i1.7052.
- [4] N. Kadek Feby Puspita Dewi, I. Gede Iwan Sudipa, I. Wayan Sunarya, N. Wayan Jeri Kusuma Dewi, and A. Suryanti Kusuma, "Sentiment Analysis of Roblox Game Reviews on Google Play Store Using Lexicon-SVM Integration," *J. dan Penelit. Tek. Inform.*, vol. 9, no. 4, pp. 1863–1876, 2025, [Online]. Available: <https://doi.org/10.33395/sinkron.v9i4.15272>
- [5] A. H. Nurdy, A. Rahim, and Arbansyah, "Analisis Sentimen Ulasan Game Stumble Guys Pada Playstore Menggunakan

- [6] Algoritma Naïve Bayes,” *Teknika*, vol. 13, no. 3, pp. 388–395, 2024, doi: 10.34148/teknika.v13i3.993.
- [7] W. J. P. Rohim Nur Rahman, Abdul Rahim, “Analisis Sentimen Ulasan Game eFootball 2024 pada Playstore Menggunakan Algoritma Naïve Bayes,” *J. Ilm. Inform.*, vol. 01, no. 3, pp. 39–44, 2025, doi: Jurnal Ilmiah Informatika.
- [8] S. Mujilahwati and M. A. Ubaydillah, “Analisis Sentimen Berbasis Aspek pada Ulasan Game Mobile Legends dengan Pendekatan Decision Tree untuk Evaluasi Pengalaman Pengguna Teknik Informatika , Fakultas Teknik , Universitas Islam Lamongan , Indonesia Aspect-Based Sentiment Analysis on Mobile Leg,” vol. 4, no. 11, pp. 325–333, 2024.
- [9] M. Y. Febrianta, S. Widyanesti, and S. R. Ramadhan, “Analisis Ulasan Indie Video Game Lokal pada Steam Menggunakan Analisis Sentimen dan Pemodelan Topik Berbasis Latent Dirichlet Allocation,” *J. Animat. Games Stud.*, vol. 7, no. 2, pp. 117–144, 2021, doi: 10.24821/jags.v7i2.5162.
- [10] D. Kusnanda and A. Permana, “Implementation of Naïve Bayes Classifier (NBC) for Sentiment Analysis on Twitter in Mobile Legends,” *Int. J. Sci. Technol. Manag.*, vol. 4, no. 5, pp. 1132–1138, 2023, doi: 10.46729/ijstm.v4i5.935.
- [11] D. Y. Saraswati, M. R. Handayani, K. Umam, and M. I. Mustofa, “Sentiment Classification of MyPertamina Reviews Using Naïve Bayes and Logistic Regression,” *J. Appl. Informatics Comput.*, vol. 9, no. 4, pp. 1319–1325, 2025, doi: 10.30871/jaic.v9i4.9723.
- [12] M. Rizky Hanafi and R. Kurniawan, “Sentiment Analysis on Sirekap App Reviews on Google Play Using Naïve Bayes Algorithm,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. October, pp. 1578–1586, 2024.
- [13] J. Informatika, I. Dikusuma, and Y. Widjaja, “Perbandingan Naïve Bayes dan Logistic Regression untuk Analisis Sentimen Ulasan Produk Skincare di Tokopedia,” 2025.
- [14] N. Muhammad, A. Ghazali, Y. Sibaroni, and N. M. Al-Ghazali, “JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika) Journal homepage: <https://jurnal.stkipgritulungagung.ac.id/index.php/jipi> 1257 Sentiment Classification in E-Commerce using Naïve Bayes and Combined Lexicon-N-Gram Features SENTIMENT CLASSIFICATION IN E-COMMERCE USING NAÏVE BAYES AND COMBINED LEXICON-N-GRAM FEATURES,” vol. 10, no. 2, pp. 1257–1271, 2025, [Online]. Available: <https://doi.org/10.29100/jipi.v10i2.6157>
- [15] R. Rahmadani, A. Rahim, and R. Rudiman, “Analisis Sentimen Ulasan ‘Ojol the Game’ Di Google Play Store Menggunakan Algoritma Naïve Bayes Dan Model Ekstraksi Fitur Tf-Idf Untuk Meningkatkan Kualitas Game,” *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, 2024, doi: 10.23960/jitet.v12i3.4988.
- [16] Nevita Cahaya Ramadani, “Analisis Sentimen Untuk Mengukur Ulasan Pengguna Aplikasi Mobile Legend Menggunakan Algoritma Naïve Bayes, SVM, Random Fores, Decision Tree, dan Logistic Regression Nevita,” *JSI J. Sist. Inf.*, vol. 16, no. 1, pp. 123–138, 2024.
- [17] R. Nauli *et al.*, “Analisis Sentimen Pengguna X Terhadap Coretax Menggunakan Naïve Bayes dan Logistic Regression Pendahuluan Metode Penelitian,” vol. 24, no. September, pp. 327–336, 2025.
- [18] Y. R. Z. Mustopo and Afiyati, “Analisis Sentimen Proyek Strategis Nasional Food Estate Menggunakan Algoritma Naïve Bayes, Logistic Regression dan Support Vector Machine,” vol. 9, no. June, pp. 485–494, 2025.
- V. No, G. S. Rasyad, and W. Maharani, “Edumatic : Jurnal Pendidikan Informatika Logistic Regression and Naïve Bayes Comparison in Classifying Emotions on Indonesian X Social Media,” vol. 9, no. 1, pp. 31–40, 2025, doi: 10.29408/edumatic.v9i1.29120.