

Klasifikasi Sentimen Ulasan Game Roblox di *Google Play Store* Menggunakan *TF-IDF Weighted Word2Vec* dan *XGBoost*

Mohamad Sahrul Bilalea^{1*}, Dewi Rahmawaty Isa¹, Siti Nurmardia Abdussamad¹

¹ Program Studi Statistika, Universitas Negeri Gorontalo, Indonesia

* Corresponding author: sahrul.bilalea02@gmail.com

Sitasi: M. S. Bilalea, D. R. Isa, and S. N. Abdussamad, "Klasifikasi Sentimen Ulasan Game Roblox di *Google Play Store* Menggunakan *TF-IDF Weighted Word2Vec* dan *XGBoost*," *Jurnal Teknologi Informasi dan Multimedia*, vol. 8, no. 4, pp. 912–925, 2026.

Diterima : 08-07-2026
Direvisi : 09-09-2026
Disetujui : 14-09-2026
Diterbitkan : 21-09-2026

Copyright: © 2026 oleh para penulis. Karya ini dilisensikan di bawah Creative Commons Attribution-Share Alike 4.0 International License.

Abstrak. Ulasan pengguna pada *Google Play Store* memuat beragam opini mengenai pengalaman menggunakan game Roblox, baik positif maupun negatif. Banyaknya ulasan menyebabkan identifikasi sentimen secara manual kurang efisien, sehingga diperlukan metode analisis sentimen yang mampu mengolah ulasan secara sistematis. Penelitian ini bertujuan mengklasifikasikan sentimen ulasan game Roblox menggunakan *TF-IDF Weighted Word2Vec* sebagai representasi fitur dan *Extreme Gradient Boosting (XGBoost)* sebagai algoritma klasifikasi. Sebanyak 1.179 ulasan dikumpulkan melalui *web scraping* selama November 2025, kemudian diberi label secara manual dan divalidasi oleh ahli bahasa Indonesia. Tahapan penelitian meliputi *pre-processing* berupa *case folding*, *cleaning*, *tokenizing*, *stopword removal*, dan *stemming*, dilanjutkan representasi fitur menggunakan *TF-IDF Weighted Word2Vec* dengan vektor 200 dimensi. Dataset dibagi dengan rasio 80:20 untuk data latih dan uji, sedangkan optimasi model dilakukan menggunakan *Grid Search* dengan *5-fold Cross Validation*. Hasil pelabelan terdiri atas 786 ulasan positif dan 393 ulasan negatif. Konfigurasi terbaik menghasilkan rata-rata akurasi validasi sebesar 89,5%, sedangkan data uji memperoleh *accuracy* sebesar 0,89. Kelas positif memperoleh *precision*, *recall*, dan *F1-score* masing-masing 0,91, 0,92, dan 0,91, sementara kelas negatif memperoleh 0,83, 0,82, dan 0,83. Hasil tersebut menunjukkan bahwa kombinasi *TF-IDF Weighted Word2Vec* dan *XGBoost* mampu mengklasifikasikan sentimen ulasan game Roblox dengan baik.

Kata kunci: Analisis Sentimen; *Extreme Gradient Boosting*; *Google Play Store*; Roblox; *Term Frequency-Inverse Document Frequency Weighted Word2Vec*.

Abstract. User reviews on the *Google Play Store* contain a wide range of opinions regarding the experience of using the Roblox game, both positive and negative. The large number of reviews makes manual sentiment identification inefficient, necessitating a sentiment analysis method capable of processing reviews systematically. This study aims to classify the sentiment of Roblox game reviews using *TF-IDF Weighted Word2Vec* as the feature representation and *Extreme Gradient Boosting (XGBoost)* as the classification algorithm. A total of 1,179 reviews were collected via web scraping during November 2025, then manually labeled and validated by an Indonesian language expert. The research stages included preprocessing—such as case folding, cleaning, tokenization, stopwords removal, and stemming—followed by feature representation using *TF-IDF Weighted Word2Vec* with 200-dimensional vectors. The dataset was split in an 80:20 ratio for training and testing, while model optimization was performed using *Grid Search* with *5-fold Cross Validation*. The labeling results consisted of 786 positive reviews and 393 negative reviews. The best configuration yielded an average validation accuracy of 89.5%, while the test data achieved an accuracy of 0.89. The positive class achieved precision, recall, and *F1-score* of 0.91, 0.92, and 0.91, respectively, while the negative class achieved 0.83, 0.82, and 0.83. These results indicate that the combination of *TF-IDF Weighted Word2Vec* and *XGBoost* is capable of effectively classifying the sentiment of Roblox game reviews.

Keywords: Sentiment Analysis; *Extreme Gradient Boosting*; *Google Play Store*; Roblox; *Term Frequency-Inverse Document Frequency Weighted Word2Vec*.

1. Pendahuluan

Kemajuan Ieknologi informasi dan Komunikasi (TIK) telah memengaruhi pola masyarakat dalam berkomunikasi, memperoleh, serta menyebarkan informasi[1]. Perkembangan akses internet, perangkat digital, dan media sosial turut mengubah posisi pengguna dari sekadar penerima informasi menjadi pihak yang aktif menghasilkan sekaligus membagikan informasi melalui berbagai media digital[2, 3]. Salah satu bentuk aktivitas tersebut dapat ditemukan pada penggunaan fitur ulasan aplikasi, karena pengguna dapat menyampaikan pengalaman maupun penilaiannya setelah menggunakan suatu layanan digital. Bagi pengembang, kumpulan ulasan tersebut dapat dimanfaatkan sebagai bahan evaluasi untuk mengetahui kualitas dan pengalaman pengguna terhadap aplikasi yang dikembangkan[4, 5].

Google Play Store menjadi salah satu media distribusi aplikasi yang menyediakan ruang bagi pengguna untuk memberikan *rating* dan menyampaikan ulasan mengenai aplikasi yang telah digunakan[6]. Salah satu aplikasi yang memiliki jumlah pengguna dan ulasan yang beragam adalah Roblox, sebuah permainan berbasis komunitas yang memungkinkan penggunanya menciptakan, memainkan, serta berinteraksi dalam lingkungan virtual. Ulasan pengguna Roblox memuat beragam opini, mulai dari apresiasi terhadap kreativitas hingga keluhan terkait *bug* dan kinerja aplikasi[7]. Kondisi tersebut menunjukkan bahwa ulasan pengguna dapat menjadi sumber informasi untuk mengetahui persepsi pengguna terhadap Roblox. Namun, semakin beragamnya ulasan menyebabkan proses identifikasi dan pengelompokan sentimen secara manual menjadi kurang efisien. Oleh karena itu, diperlukan analisis sentimen yang mampu mengolah ulasan pengguna secara sistematis untuk mengidentifikasi kecenderungan sentimen secara otomatis.

Dalam perkembangan *machine learning*, berbagai algoritma telah dimanfaatkan untuk melakukan klasifikasi terhadap data teks. Salah satunya adalah *Extreme Gradient Boosting (XGBoost)*, yaitu algoritma berbasis *gradient boosting* yang membangun model secara bertahap dengan memanfaatkan sejumlah *decision tree* untuk memperoleh prediksi yang lebih baik[8]. *XGBoost* memiliki mekanisme regularisasi yang membantu mengendalikan kompleksitas model dan mengurangi risiko *overfitting*. Penggunaan *XGBoost* dalam klasifikasi teks juga menunjukkan kemampuan yang baik dalam membedakan kategori sentiment[9]. Selain itu, penelitian sebelumnya menunjukkan bahwa *XGBoost* dapat memberikan kinerja yang kompetitif dibandingkan algoritma klasifikasi lainnya pada data ulasan aplikasi digital[10]. Oleh karena itu, *XGBoost* digunakan sebagai algoritma klasifikasi dalam penelitian ini.

Selain pemilihan algoritma klasifikasi, bentuk representasi data teks juga menjadi faktor yang berpengaruh terhadap kemampuan model dalam mengenali pola sentimen. *TF-IDF Weighted Word2Vec* merupakan salah satu pendekatan yang memadukan informasi bobot kata melalui *TF-IDF* dengan representasi kata berbasis *Word2Vec* [11]. Berbeda dengan *TF-IDF* yang berfokus pada bobot kata dan *Word2Vec* yang merepresentasikan hubungan semantik antarkata, kombinasi keduanya memungkinkan kedua informasi tersebut digunakan secara bersamaan dalam membentuk representasi teks. Penelitian Rifaldy dkk.[11]. juga membandingkan *TF-IDF*, *Word2Vec*, dan *TF-IDF Weighted Word2Vec* dalam klasifikasi sentimen dan menunjukkan bahwa penggunaan *TF-IDF Weighted Word2Vec* memberikan representasi fitur yang lebih baik dibandingkan penggunaan *TF-IDF* atau *Word2Vec* secara terpisah. Oleh karena itu, pendekatan ini dipilih untuk merepresentasikan ulasan pengguna Roblox sebelum dilakukan proses klasifikasi sentiment.

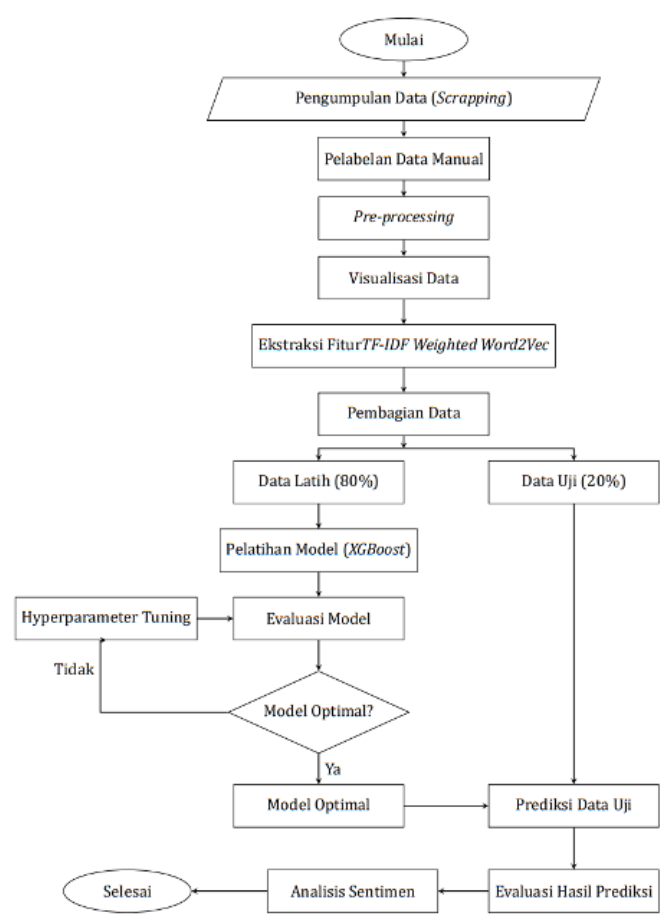
Pemilihan *TF-IDF Weighted Word2Vec* juga didukung oleh beberapa penelitian terdahulu. Rifaldy dkk.[11] menerapkan dan membandingkan *TF-IDF*, *Word2Vec*, dan *TF-IDF Weighted Word2Vec* untuk klasifikasi sentimen ulasan pengguna, sehingga menunjukkan relevansi kombinasi tersebut dalam representasi data

teks. Khan dan Rekha [12] juga menerapkan *TF-IDF Weighted Word2Vec* pada klasifikasi berita palsu dengan pendekatan ensemble dan memperoleh hasil klasifikasi yang baik, yang menunjukkan bahwa representasi tersebut dapat diterapkan pada berbagai tugas klasifikasi teks. Selain itu, Dewi dkk. [13] melakukan analisis sentimen terhadap ulasan game Roblox di *Google Play Store* menggunakan pendekatan *Lexicon-SVM*. Penelitian tersebut menunjukkan bahwa ulasan Roblox dapat digunakan sebagai sumber data untuk mengidentifikasi sentimen pengguna, sedangkan penelitian ini menggunakan *TF-IDF Weighted Word2Vec* dan *XGBoost* untuk melakukan klasifikasi pada objek yang sama. Berdasarkan penelitian terdahulu tersebut, penggunaan *TF-IDF Weighted Word2Vec* sebagai representasi fitur dan *XGBoost* sebagai algoritma klasifikasi dipilih untuk menganalisis sentimen ulasan game Roblox di *Google Play Store*.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk mengklasifikasikan sentimen positif dan negatif pada ulasan game Roblox di *Google Play Store* menggunakan kombinasi *TF-IDF Weighted Word2Vec* dan *XGBoost*. Kontribusi penelitian ini terletak pada penerapan kombinasi tersebut pada ulasan game Roblox serta evaluasi kinerja model dalam mengidentifikasi sentimen pengguna.

2. Bahan dan Metode

Penelitian ini dilaksanakan melalui beberapa tahapan yang dilakukan secara sistematis untuk menghasilkan model klasifikasi sentimen yang optimal. Tahapan penelitian meliputi proses pengumpulan data, prapemrosesan teks, ekstraksi fitur menggunakan *TF-IDF Weighted Word2Vec*, pelatihan model *XGBoost*, serta evaluasi kinerja model. Alur tahapan penelitian disajikan pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.1. Pengumpulan Data

Penelitian ini menggunakan data sekunder yang bersumber dari ulasan pengguna game Roblox pada *Google Play Store* dan dikumpulkan selama November 2025. Proses pengambilan data dilakukan dengan teknik *web scraping* menggunakan bahasa pemrograman *Python* beserta beberapa pustaka pendukung. Sebanyak 1.179 ulasan pengguna diperoleh dan digunakan sebagai dataset penelitian. Data yang diperoleh mencakup isi ulasan, *rating*, serta waktu publikasi ulasan. Contoh data ulasan yang digunakan dalam penelitian disajikan pada [Tabel 1](#). Selanjutnya, data tersebut menjadi dasar dalam penelitian dan diproses melalui tahapan *preprocessing*, pembentukan fitur menggunakan *TF-IDF Weighted Word2Vec*, serta klasifikasi sentimen dengan algoritma *Extreme Gradient Boosting (XGBoost)*.

Tabel 1. Contoh Data Ulasan Game Roblox

Index	Ulasan
1	Gamenya bagus dan ramah dan baik tidak ngeleg bagus banget bisa mabar bersama teman-teman bagus mantap game bagus semoga game ini lancar terus amin
2	Sangat menyenangkan bisa dapat teman dan kreativitas lainnya ini sangat seru aku kasih bintang
3	Ngeleg-nya minta ampun. Aku hanya bermain game yang saya suka, mengapa dibikin ngeleg seperti Chrome saat nonton bokep
4	Game kotak-kotak buruk
5	Seru dan menyenangkan

2.2. Analisis Sentimen

Dalam bidang *Natural Language Processing (NLP)*, analisis sentimen digunakan untuk memperoleh informasi mengenai opini, emosi, atau sikap yang terdapat dalam sebuah teks terhadap objek tertentu, seperti produk, layanan, maupun topik yang sedang dibahas[14]. Pendekatan ini memungkinkan isi teks dikelompokkan berdasarkan polaritas sentimennya, yang pada umumnya terdiri atas sentimen positif, negatif, dan netral[9]. Selain menentukan polaritas, analisis sentimen juga dapat digunakan untuk memperoleh informasi subjektif dari kumpulan teks secara otomatis sehingga kecenderungan opini pengguna dapat diketahui. Dalam penelitian ini, pendekatan tersebut digunakan untuk menentukan sentimen pada ulasan game Roblox di *Google Play Store* dengan memanfaatkan fitur *TF-IDF Weighted Word2Vec* dan model *Extreme Gradient Boosting (XGBoost)*.

2.3. Pre-Processing

Tahap *preprocessing* dilakukan sebelum data teks digunakan dalam proses pemodelan dengan tujuan menghasilkan data yang lebih terstruktur dan sesuai untuk dianalisis menggunakan *machine learning*[15]. Pembersihan dan penyeragaman teks diperlukan karena data ulasan yang diperoleh dari pengguna dapat memiliki bentuk yang beragam. Proses tersebut dapat membantu meningkatkan kualitas data yang menjadi masukan bagi model klasifikasi[16]. Dalam penelitian ini, *preprocessing* dilakukan melalui lima tahapan, yaitu data *case folding*, data *cleaning*, *tokenization*, *stopword removal*, dan *stemming* [17].

1. Case Folding

Case folding digunakan untuk menyeragamkan bentuk penulisan dalam data teks dengan mengonversi karakter huruf kapital menjadi huruf kecil (lower-case)[18]. Penyeragaman tersebut dilakukan agar teks memiliki format yang konsisten ketika dilanjutkan ke tahapan *preprocessing* berikutnya.

2. Data Cleaning

Data *cleaning* dilakukan untuk menghilangkan berbagai unsur yang tidak diperlukan dalam data ulasan sebelum memasuki tahap analisis. Unsur tersebut dapat berupa karakter khusus, *URL*, *username*, emoji, maupun bagian data yang tidak konsisten[19]. Dengan proses ini, teks yang digunakan pada

tahap berikutnya menjadi lebih terstruktur dan sesuai dengan kebutuhan analisis.

3. *Tokenization*

Tokenization dilakukan dengan membagi teks menjadi bagian-bagian yang lebih kecil berupa token, terutama unit kata[20]. Hasil pemecahan tersebut kemudian menjadi bentuk masukan yang digunakan dalam tahapan pengolahan teks selanjutnya.

4. *Stopword Removal*

Pada tahap *stopword removal*, kata-kata yang dianggap kurang memberikan informasi terhadap proses klasifikasi dihilangkan dari teks[21]. Kata yang termasuk dalam kategori tersebut, misalnya kata hubung atau kata yang memiliki frekuensi kemunculan tinggi, sehingga model dapat lebih diarahkan pada kata yang memiliki informasi relevan terhadap sentimen.

5. *Stemming*

Stemming digunakan untuk mengembalikan kata yang memiliki imbuhan ke bentuk dasarnya[22]. Proses ini bertujuan mengurangi variasi bentuk kata yang memiliki makna dasar serupa sehingga representasi teks yang digunakan dalam analisis menjadi lebih seragam.

2.4. *Term Frequency-Inverse Document Frequency Weighted Word2Vec*

TF-IDF Weighted Word2Vec merupakan pendekatan representasi teks yang memanfaatkan bobot *TF-IDF* bersama dengan representasi kata dari *Word2Vec*. *TF-IDF* memberikan bobot berdasarkan tingkat kepentingan suatu kata di dalam dokumen, sedangkan *Word2Vec* mengubah kata menjadi vektor numerik yang dapat merepresentasikan keterkaitan semantik berdasarkan konteks penggunaannya[23, 24]. Penggabungan kedua pendekatan tersebut memungkinkan representasi dokumen mempertimbangkan dua aspek sekaligus, yaitu bobot kepentingan kata dan informasi semantik yang terkandung di dalamnya. Dengan demikian, fitur yang dihasilkan dapat digunakan sebagai representasi numerik dokumen pada proses klasifikasi[10]. Dalam penelitian ini, pembentukan representasi dokumen menggunakan *TF-IDF Weighted Word2Vec* dilakukan melalui beberapa tahapan berikut.

1. Menghitung nilai *Term Frequency (TF)* untuk setiap kata dalam dokumen guna merepresentasikan tingkat kemunculan kata tersebut.
2. Menghitung nilai *Inverse Document Frequency (IDF)* untuk setiap kata berdasarkan distribusinya pada seluruh dokumen, sehingga kata yang bersifat umum memperoleh bobot lebih rendah.
3. Menghitung bobot *TF-IDF* dengan mengalikan nilai *TF* dan *IDF* untuk memperoleh tingkat kepentingan setiap kata dalam dokumen.
4. Melatih atau menggunakan model *Word2Vec* untuk memetakan setiap kata ke dalam vektor berdimensi tetap yang merepresentasikan makna semantik kata berdasarkan konteks kemunculannya.
5. Menjumlahkan semua vektor kata berbobot dalam dokumen dan membaginya dengan jumlah total bobot untuk mendapatkan representasi dokumen berdimensi tetap.

2.5. *Extreme Gradient Boosting*

Extreme Gradient Boosting (XGBoost) adalah algoritma *machine learning* yang menggunakan pendekatan *gradient boosting* dengan membentuk sejumlah *decision tree* secara berurutan untuk menghasilkan model prediksi yang semakin baik[8]. Pada prosesnya, model baru dibangun dengan mempertimbangkan kesalahan dari model sebelumnya sehingga kesalahan prediksi dapat diminimalkan secara bertahap. *XGBoost* juga menerapkan beberapa mekanisme optimasi, antara lain *parallel processing*, *tree pruning*, dan regularisasi, yang mendukung efisiensi proses pembelajaran serta membantu mengendalikan kemungkinan terjadinya *overfitting* [8]. Proses pembentukan model *XGBoost* dalam penelitian ini dijelaskan melalui tahapan berikut.

1. Mengoptimalkan fungsi objektif untuk meminimalkan kesalahan prediksi sekaligus mengendalikan kompleksitas model.
2. Menerapkan fungsi regularisasi guna mencegah terjadinya kompleksitas model yang berlebihan.
3. Melakukan pendekatan terhadap fungsi objektif menggunakan ekspansi Taylor orde dua agar proses perhitungan menjadi lebih efisien, dengan mencari nilai gradien dan hessian.
4. Menentukan bobot optimal pada setiap simpul daun (*leaf*) berdasarkan nilai gradien dan hessian.
5. Menghitung nilai gain untuk menentukan titik pemisahan (*splitting*) terbaik dalam struktur pohon.
6. Menerapkan mekanisme shrinkage dengan mengalikan kontribusi setiap pohon baru dengan **learning rate** (η) guna mencegah terjadinya *overfitting*, sehingga pembaruan prediksi dilakukan secara bertahap.
7. Mengonversi keluaran model menjadi nilai probabilitas melalui fungsi aktivasi logistik (*sigmoid*).
8. Menghitung nilai fungsi kerugian pada klasifikasi biner.

2.6. Hyperparameter Tuning

Hyperparameter tuning dilakukan untuk menentukan kombinasi nilai *hyperparameter* yang dapat memberikan kinerja terbaik pada model *machine learning* [25]. Dalam penelitian ini, pencarian konfigurasi parameter dilakukan menggunakan metode *Grid Search*. Metode tersebut bekerja dengan mengevaluasi seluruh kombinasi nilai *hyperparameter* yang telah ditetapkan dalam ruang pencarian untuk memperoleh konfigurasi yang paling sesuai [26]. Penerapan *tuning* dilakukan pada model *XGBoost* dengan melibatkan beberapa *hyperparameter*, yaitu `n_estimators`, `max_depth`, `min_child_weight`, `learning_rate` (η), `gamma`, `subsample`, dan `colsample_bylevel` [27].

2.7. Confusion Matrix

Confusion Matrix digunakan untuk mengevaluasi kemampuan model klasifikasi dengan melihat kesesuaian antara hasil prediksi dan kelas aktual pada data [28]. Pada klasifikasi biner, hasil prediksi dikelompokkan ke dalam empat kondisi, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) [29]. Keempat kondisi tersebut digunakan sebagai komponen dalam perhitungan *accuracy*, *precision*, *recall*, dan *F1-score* untuk menilai kinerja model.

1. Accuracy

Accuracy digunakan untuk menunjukkan tingkat ketepatan model berdasarkan perbandingan antara jumlah prediksi yang benar dan keseluruhan data yang dievaluasi [28]. Nilai *accuracy* dihitung menggunakan Persamaan (1) berikut.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

2. Precision

Precision menunjukkan tingkat ketepatan model dalam mengidentifikasi kelas positif dengan membandingkan jumlah prediksi positif yang tepat terhadap keseluruhan data yang dikategorikan sebagai positif [28]. Nilai *precision* dihitung menggunakan Persamaan (2) berikut.

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

3. Recall

Recall digunakan untuk mengetahui kemampuan model dalam menemukan data yang sebenarnya termasuk ke dalam kelas positif [28]. Nilai *recall*

diperoleh melalui perhitungan pada [Persamaan \(3\)](#) berikut.

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

4. *F1-score*

F1-score merupakan ukuran evaluasi yang menggabungkan nilai *precision* dan *recall* melalui perhitungan rata-rata harmonik. Metrik ini dapat memberikan gambaran kinerja yang lebih seimbang, khususnya ketika data memiliki distribusi kelas yang tidak sama[30]. Perhitungan *F1-score* disajikan pada [Persamaan \(4\)](#) berikut.

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

3. Hasil

3.1. *Pre-processing Data*

Dataset penelitian diperoleh dari ulasan pengguna Game Roblox pada *Google Play Store* yang dikumpulkan melalui teknik *web scraping* selama periode 1–30 November 2025. Periode tersebut dipilih untuk membatasi ruang lingkup pengumpulan data pada ulasan yang tersedia dalam satu periode pengamatan sehingga data yang digunakan memiliki rentang waktu yang konsisten. Dari proses pengumpulan tersebut diperoleh sebanyak 1.179 ulasan yang selanjutnya digunakan sebagai data penelitian. Setiap ulasan kemudian dikategorikan secara manual ke dalam dua kelas sentimen, yaitu positif dan negatif, dengan mempertimbangkan makna serta konteks yang terdapat dalam masing-masing ulasan. Untuk memastikan hasil anotasi lebih valid, proses pelabelan yang telah dilakukan selanjutnya diverifikasi oleh ahli bahasa Indonesia. Berdasarkan hasil pelabelan dan verifikasi tersebut, diperoleh 786 ulasan dengan sentimen positif dan 393 ulasan dengan sentimen negatif. Sebelum dilakukan *pre-processing*, sebanyak 1.179 ulasan digunakan sebagai data masukan. Seluruh ulasan tersebut kemudian melalui tahapan *case folding*, *cleaning*, *tokenizing*, *stopword removal*, dan *stemming*. Setelah proses *pre-processing* selesai, jumlah ulasan tetap sebanyak 1.179 data karena tahapan tersebut hanya melakukan perubahan dan penyesuaian pada bentuk teks tanpa menghilangkan dokumen yang telah ditetapkan sebagai data penelitian. Dengan demikian, seluruh data tetap dipertahankan agar informasi dari setiap ulasan yang telah diberi label dapat digunakan pada tahap pembentukan fitur dan klasifikasi. Tahapan *pre-processing* tersebut bertujuan untuk mempersiapkan teks agar lebih terstruktur dan sesuai dengan kebutuhan analisis. Contoh hasil dari masing-masing tahapan *pre-processing* disajikan pada [Tabel 2](#).

Tabel 2. Tahapan *Pre-processing Data*

Proses	Input	Output
<i>Case Folding</i>	Semenjak update mau login aja lama bener padahal sinyal aman aja lancar main game pubg/ml/ghensin + sering Disconnect/DC, udh gitu gk bisa lagi tuh pake apk lain yg mana buat afk biar hp gk panas, soalnya keditec apk apakah itu taik emang developer nya update gk pake otak bikin rate hancur ulah kau sendiri	semenjak update mau login aja lama bener padahal sinyal aman aja lancar main game pubg/ml/ghensin + sering disconnect/dc, udh gitu gk bisa lagi tuh pake apk lain yg mana buat afk biar hp gk panas, soalnya keditec apk apakah itu taik emang developer nya update gk pake otak bikin rate hancur ulah kau sendiri

3.2. Ekstraksi Fitur *TF-IDF Weighted Word2Vec*

Data teks yang telah melewati tahap *pre-processing* selanjutnya diubah menjadi bentuk numerik melalui ekstraksi fitur menggunakan *TF-IDF Weighted Word2Vec*. Pendekatan ini memanfaatkan bobot *TF-IDF* untuk menunjukkan tingkat kepentingan kata dalam dokumen, kemudian menggabungkannya dengan representasi *Word2Vec* yang menangkap hubungan semantik antar kata. Bobot *TF-IDF* yang diperoleh untuk setiap kata digunakan sebagai pembobot pada vektor kata dari model *Word2Vec*. Pada penelitian ini, model *Word2Vec* menggunakan *vector size* 200, *window size* 5, *min_count* 2, dan *epochs* sebanyak 15. Vektor kata yang telah diberi bobot kemudian dirataratakan sehingga diperoleh representasi numerik berdimensi tetap untuk setiap dokumen.

Hasil proses ekstraksi menghasilkan matriks fitur berukuran 1179×200 . Ukuran tersebut menunjukkan bahwa sebanyak 1.179 ulasan direpresentasikan ke dalam 200 fitur numerik yang diperoleh melalui kombinasi *TF-IDF* dan *Word2Vec*. Matriks yang dihasilkan selanjutnya digunakan sebagai *input* dalam proses pelatihan model *XGBoost*. Contoh nilai fitur hasil ekstraksi ditampilkan pada Tabel 3.

Tabel 3. Contoh Matriks *TF-IDF Weighted Word2Vec*

Ulasan	w2v_1	w2v_2	w2v_3	...	w2v_198	w2v_199	w2v_200
1	0,188	0,117	0,040	...	0,154	0,023	0,182
2	0,155	0,096	0,030	...	0,125	0,021	0,151
3	0,164	0,105	0,037	...	0,138	0,021	0,162
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1177	0,274	0,169	0,058	...	0,224	0,025	0,263
1178	0,172	0,108	0,038	...	0,141	0,021	0,167
1179	0,105	0,064	0,023	...	0,089	0,014	0,103

3.3. Pelatihan dan Evaluasi Model *XGBoost*

Dataset terlebih dahulu dibagi menjadi data latih dan data uji menggunakan rasio 80:20, dengan 943 data sebagai data latih (80%) dan 236 data sebagai data uji (20%). Rasio 80:20 dipilih karena memberikan proporsi data latih yang cukup untuk membangun model, sekaligus menyediakan data uji yang cukup untuk mengevaluasi kinerja model. Penggunaan rasio ini mempertimbangkan bahwa rasio 70:30 akan mengurangi jumlah data yang digunakan untuk pelatihan, sedangkan rasio 90:10 akan mengurangi jumlah data yang digunakan untuk pengujian. Oleh karena itu, rasio 80:20 digunakan untuk menjaga keseimbangan antara kebutuhan data latih dan data uji. Data latih digunakan dalam proses pembentukan model klasifikasi, sedangkan data uji digunakan untuk mengukur kinerja model yang telah dibangun. Model klasifikasi dikembangkan menggunakan algoritma *XGBoost* dengan masukan berupa fitur hasil ekstraksi *TF-IDF Weighted Word2Vec*. Selanjutnya, untuk memperoleh konfigurasi model yang optimal, dilakukan *hyperparameter tuning* menggunakan *Grid Search* dengan *5-fold Cross Validation*. Proses tersebut dilakukan dengan menguji berbagai kombinasi parameter secara sistematis untuk memperoleh konfigurasi terbaik. Parameter yang digunakan dalam pencarian ditampilkan pada Tabel 4.

Tabel 4. Parameter *XGBoost*

Parameter	Nilai
n_estimators	(150, 300, 450)
max_depth	(4, 6, 8)
learning_rate	(0,05; 0,1; 0,2)
subsample	(0,8; 1,0)
colsample_bytree	(0,8; 1,0)
gamma	(0; 0,2)
min_child_weight	(1, 3, 5)

Berdasarkan ruang pencarian parameter pada Tabel 4, *hyperparameter tuning* menghasilkan 648 kombinasi yang kemudian dievaluasi menggunakan *Grid Search* dengan *5-fold Cross Validation*. Hasil pengujian menunjukkan bahwa konfigurasi terbaik diperoleh dengan `n_estimators = 450`, `max_depth = 6`, `learning_rate = 0,2`, `subsample = 0,8`, `colsample_bytree = 1,0`, `gamma = 0`, dan `min_child_weight = 3`. Konfigurasi tersebut menghasilkan rata-rata akurasi validasi sebesar 89,5%. Parameter terbaik tersebut selanjutnya digunakan untuk membangun model *XGBoost*, kemudian kinerjanya diuji menggunakan data uji melalui *confusion matrix* dan metrik evaluasi. Hasil *confusion matrix* model *XGBoost* disajikan pada Tabel 5.

Tabel 5. *Confusion Matrix XGBoost*

Kelas Aktual	Kelas Prediksi	
	Negatif	Positif
Negatif	65	14
Positif	13	144

Berdasarkan *confusion matrix* pada Tabel 5, model mampu memberikan prediksi yang tepat terhadap 65 ulasan dengan kelas negatif dan 144 ulasan dengan kelas positif. Sementara itu, sebanyak 14 ulasan negatif dikategorikan sebagai positif dan 13 ulasan positif dikategorikan sebagai negatif. Nilai metrik yang diperoleh dari hasil klasifikasi model *XGBoost* selanjutnya disajikan pada Tabel 6.

Tabel 6. Hasil Evaluasi Model

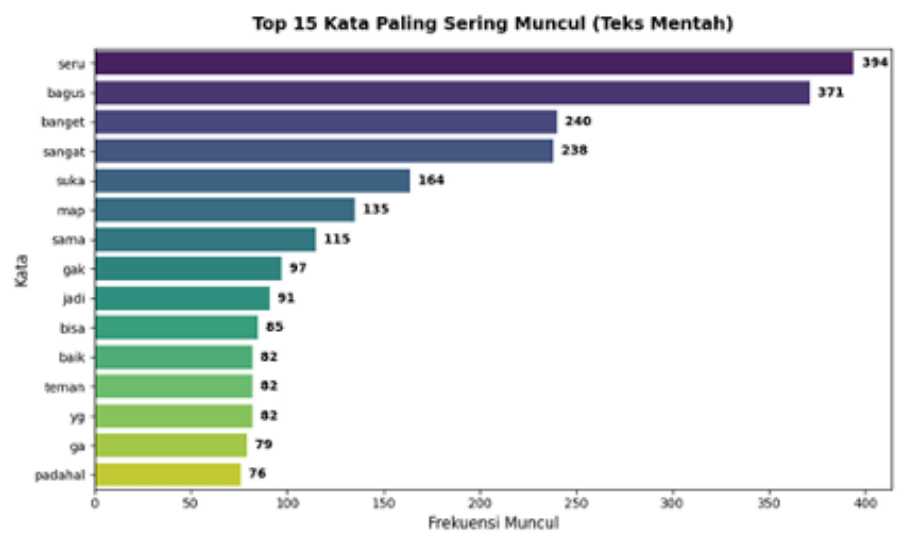
Kelas	Precision	Recall	F1-Score	Accuracy
0 (Negatif)	0,83	0,82	0,83	0,89
1 (Positif)	0,91	0,92	0,91	0,89

Berdasarkan hasil pada Tabel 6, model *XGBoost* menghasilkan *accuracy* sebesar 0,89 atau mampu memberikan prediksi yang benar terhadap 89% data uji. Untuk kelas negatif, nilai *precision* yang diperoleh adalah 0,83, *recall* sebesar 0,82, dan *f1-score* sebesar 0,83. Pada kelas positif, model menghasilkan *precision* sebesar 0,91, *recall* sebesar 0,92, serta *f1-score* sebesar 0,91. Hasil tersebut memperlihatkan bahwa model mampu melakukan klasifikasi pada kedua kategori sentimen dengan baik, meskipun nilai evaluasi pada kelas positif lebih tinggi dibandingkan kelas negatif. Secara keseluruhan, penggunaan *TF-IDF Weighted Word2Vec* sebagai representasi fitur yang dikombinasikan dengan *XGBoost* menghasilkan kinerja klasifikasi yang baik dalam membedakan ulasan positif dan negatif.

3.4. Analisis Sentimen

Analisis *feature importance* dilakukan untuk mengetahui fitur yang memiliki kontribusi paling besar terhadap proses prediksi pada model *XGBoost*. Hasil analisis tersebut dapat digunakan untuk melihat fitur yang berperan dalam membedakan sentimen positif dan negatif serta memberikan gambaran mengenai karakteristik data yang dipertimbangkan oleh model dalam menghasilkan prediksi. Hasil *feature importance* pada model *XGBoost* ditampilkan pada Gambar 3.

Berdasarkan Gambar 3, beberapa kata dengan frekuensi kemunculan tertinggi dalam ulasan Roblox meliputi seru (394), bagus (371), banget (240), sangat (238), suka (164), dan map (135). Tingginya kemunculan kata seru, bagus, dan suka memperlihatkan kecenderungan penilaian positif dari pengguna terhadap pengalaman bermain Roblox. Sebaliknya, kata seperti gak, ga, dan padahal menunjukkan adanya ungkapan yang berkaitan dengan keluhan atau sentimen negatif pada sebagian ulasan. Dengan demikian, pola frekuensi kata yang terdapat dalam dataset dapat memberikan gambaran mengenai karakteristik sentimen pengguna dan mendukung proses klasifikasi menggunakan model *XGBoost*.



Gambar 3. *Feature Importance Model XGBoost*

4. Pembahasan

Hasil penelitian memperlihatkan bahwa penggunaan *TF-IDF Weighted Word2Vec* bersama *XGBoost* menghasilkan model klasifikasi sentimen dengan tingkat akurasi sebesar 89%. *TF-IDF* berperan dalam memberikan bobot terhadap kata berdasarkan tingkat kepentingannya dalam dokumen, sedangkan *Word2Vec* memberikan representasi yang mempertimbangkan hubungan semantik antarkata. Penggabungan kedua pendekatan tersebut menghasilkan fitur yang dapat digunakan model untuk membedakan ulasan dengan sentimen positif dan negatif. Selain itu, tingginya frekuensi kata seru, bagus, dan suka juga memperlihatkan bahwa ulasan dalam dataset cenderung memberikan penilaian positif terhadap game Roblox.

Berdasarkan hasil *confusion matrix*, masih terdapat 14 ulasan negatif yang diprediksi sebagai positif dan 13 ulasan positif yang diprediksi sebagai negatif. Kesalahan klasifikasi tersebut dapat terjadi karena ulasan pengguna cenderung menggunakan bahasa informal, singkatan, kata tidak baku, serta ungkapan yang maknanya bergantung pada konteks. Ulasan negatif yang mengandung kata-kata dengan kecenderungan positif dapat menyebabkan model memprediksinya sebagai positif apabila konteks keseluruhan ulasan tidak terwakili secara optimal oleh fitur. Sebaliknya, ulasan positif yang mengandung kata atau ungkapan bernuansa negatif dapat menyebabkan model memprediksinya sebagai negatif. Hal ini menunjukkan bahwa meskipun *TF-IDF Weighted Word2Vec* mempertimbangkan bobot kata dan hubungan semantik, pendekatan tersebut masih memiliki keterbatasan dalam menangkap konteks sentimen secara menyeluruh pada ulasan yang kompleks.

Temuan penelitian ini memiliki kesesuaian dengan hasil penelitian Rifaldy dkk.[11], yang memperlihatkan bahwa *TF-IDF Weighted Word2Vec* mampu memberikan representasi fitur yang lebih informatif dibandingkan penggunaan *TF-IDF* atau *Word2Vec* secara terpisah. Penelitian Dewi dkk.[7], turut mendukung temuan tersebut melalui hasil analisis sentimen pada ulasan Roblox yang dapat memberikan informasi mengenai persepsi pengguna. Sementara itu, penelitian ini menggunakan *XGBoost* dengan pendekatan *boosting* sebagai algoritma klasifikasi. Penggunaan pendekatan tersebut juga sejalan dengan penelitian Ruhyana dkk.[31], yang menunjukkan bahwa algoritma berbasis *boosting* memiliki kemampuan dalam mengolah data teks dengan dimensi yang tinggi.

Meskipun memberikan hasil klasifikasi yang baik, penelitian ini masih memiliki keterbatasan karena hanya menggunakan dua kategori sentimen, yaitu positif

dan negatif. Penelitian berikutnya dapat mempertimbangkan penggunaan dataset dengan jumlah dan variasi yang lebih besar serta menambahkan kategori sentimen netral. Selain itu, kinerja *XGBoost* dapat dibandingkan dengan pendekatan *deep learning*, seperti BERT atau IndoBERT, untuk mengetahui kemungkinan peningkatan performa klasifikasi.

5. Kesimpulan

Berdasarkan hasil penelitian, *TF-IDF Weighted Word2Vec* yang dikombinasikan dengan *XGBoost* dapat digunakan untuk mengidentifikasi sentimen pada ulasan Game Roblox dari *Google Play Store*. Pengujian terhadap data menghasilkan *accuracy* sebesar 89%. Pada kategori negatif, diperoleh *precision* 0,83, *recall* 0,82, dan *f1-score* 0,83, sedangkan kategori positif menghasilkan *precision* 0,91, *recall* 0,92, dan *f1-score* 0,91. Perbedaan nilai evaluasi tersebut tetap menunjukkan kemampuan model dalam mengenali kedua kategori sentimen, dengan hasil yang lebih tinggi pada kategori positif. Dari sisi karakteristik ulasan, kata seru, bagus, dan suka merupakan kata yang dominan dan memperlihatkan kecenderungan sentimen positif pengguna. Dengan demikian, pendekatan *TF-IDF Weighted Word2Vec* bersama *XGBoost* dapat dimanfaatkan untuk mengklasifikasi sentimen pada ulasan pengguna Game Roblox di *Google Play Store*. Penelitian ini memiliki keterbatasan pada penggunaan dua kategori sentimen, yaitu positif dan negatif, sehingga belum dapat menggambarkan ulasan dengan sentimen netral. Selain itu, penelitian ini menggunakan data ulasan dari satu sumber, yaitu *Google Play Store*, sehingga hasil yang diperoleh belum tentu menggambarkan sentimen pengguna dari sumber lainnya. Penelitian selanjutnya dapat mempertimbangkan penambahan kategori sentimen netral, penggunaan dataset yang lebih besar dan beragam, serta sumber data lainnya untuk memperoleh hasil klasifikasi yang lebih komprehensif.

Pernyataan Penggunaan Artificial Intelligence. Penulis menggunakan ChatGPT untuk membantu penyuntingan bahasa, perbaikan tata bahasa, dan peningkatan kejelasan naskah. Seluruh keluaran yang dihasilkan oleh AI telah diperiksa dan diverifikasi oleh penulis. Penulis bertanggung jawab sepenuhnya atas keakuratan, orisinalitas, dan integritas isi artikel.

Pernyataan Kontribusi Penulis. MSB berkontribusi dalam seluruh tahapan penelitian, meliputi konseptualisasi, metodologi, pengumpulan dan pengolahan data, analisis, penyusunan, serta penyelesaian naskah dari awal hingga akhir. DRI dan SNA berkontribusi sebagai pembimbing dalam pelaksanaan penelitian dan penyusunan naskah, termasuk memberikan arahan, masukan, peninjauan, dan evaluasi terhadap penelitian serta artikel. Seluruh penulis telah membaca, meninjau, dan menyetujui versi akhir artikel serta bertanggung jawab atas isi artikel.

Pernyataan Konflik Kepentingan. Penulis menyatakan bahwa tidak terdapat konflik kepentingan yang berkaitan dengan penelitian, kepengarangan, dan publikasi artikel ini.

Pernyataan Pendanaan. Penelitian ini tidak menerima pendanaan khusus dari lembaga pemerintah, komersial, maupun organisasi nirlaba.

Daftar Pustaka

- [1] D. Wiryany, S. Natasha, and R. Kurniawan, "Perkembangan teknologi informasi dan komunikasi terhadap perubahan sistem komunikasi inonesia," *J. Nomosleca*, vol. 8, no. 2, pp. 242–252, 2022. <https://doi.org/10.26905/nomosleca.v8i2.8821>.

- [2] D. D. Fitri, "Perkembangan teknologi informasi dan komunikasi dalam pembelajaran pendidikan agama islam," *Arus J. Psikol. dan Pendidik.*, vol. 2, no. 1, pp. 53–57, 2023. <https://doi.org/10.57250/ajpp.v2i1.166>.
- [3] M. Khairi *et al.*, "Pengaruh teknologi dalam transformasi ekonomi dan bisnis di era digital," *J. Perangkat Lunak*, vol. 7, no. 1, pp. 71–78, 2025. <https://doi.org/10.32520/jupel.v7i1.3947>.
- [4] Y. A. Mustofa and I. S. K. Idris, "Ensemble approach to sentiment analysis of google play store app reviews," *Jambura J. Electr. Electron. Eng.*, vol. 6, no. 2, pp. 181–188, 2024. <https://doi.org/10.37905/jjee.v6i2.25184>.
- [5] R. Kurniawan, H. O. L. Wijaya, and R. P. Aprisusanti, "Sentiment analysis of google play store user reviews on digital population identity app using k-nearest neighbors," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 13, no. 2, pp. 170–178, 2024. <https://doi.org/10.32736/sisfokom.v13i2.2071>.
- [6] M. R. Hadiyasa and S. M. Isa, "Google play store users comment review classification using svm classifier and random forest," *IJISCS (International J. Inf. Syst. Comput. Sci.)*, vol. 7, no. 3, p. 200, 2023. <https://doi.org/10.56327/ijiscs.v7i3.1584>.
- [7] N. K. F. P. Dewi, I. G. I. Sudipa, I. W. Sunarya, N. W. J. K. Dewi, and A. S. Kusuma, "Sentiment analysis of roblox game reviews on google play store using lexicon-svm integration," *Sink. J. dan Penelit. Tek. Inform.*, vol. 9, no. 4, pp. 1863–1867, 2025. <https://doi.org/10.63956/jitar.v1i2.31>.
- [8] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, vol. 22, pp. 785–794, 2016. <https://doi.org/10.1145/2939672.2939785>.
- [9] A. Fölling and J. Lepping, "Knowledge discovery for scheduling in computational grids," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 2, no. 4, pp. 287–297, 2012. <https://doi.org/10.1002/widm.1060>.
- [10] M. J. Setiawan and V. R. S. Nastiti, "Dana app sentiment analysis: Comparison of xgboost, svm, and extra trees," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 13, no. 3, pp. 337–345, 2024. <https://doi.org/10.32736/sisfokom.v13i3.2239>.
- [11] F. Rifaldy, Y. Sibaroni, and S. S. Prasetyowati, "Effectiveness of word2vec and tf-idf in sentiment classification on online investment platforms using support vector machine," *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 10, no. 2, pp. 863–874, 2025. <https://doi.org/10.29100/jupi.v10i2.6055>.
- [12] Z. A. Khan and V. Rekha, "Fake news detection using tf-idf weighted with word2vec: An ensemble approach," *Int. J. Intell. Syst. Appl. Eng.*, vol. 11, no. 3, pp. 1065–1076, 2023. <https://ijisae.org/index.php/IJISAE/article/view/3366>.
- [13] I. G. A. N. Lestari, N. M. R. M. Dewi, K. G. Meiliana, and I. K. A. A. Aryanto, "Effectiveness of adaboost and xgboost algorithms in sentiment analysis of movie reviews," *J. Appl. Informatics Comput.*, vol. 9, no. 2, pp. 258–264, 2025. <https://doi.org/10.30871/jaic.v9i2.9077>.
- [14] S. M. Mohammad, "Sentiment analysis: Detecting valence, emotions, and other affectual states from text," in *Emot. Meas.*, pp. 201–237, 2021. <https://doi.org/10.1016/B978-0-08-100508-8.00009-6>.
- [15] K. Maharana, S. Mondal, and B. Nemade, "A review: Data pre-processing and data augmentation techniques," in *Global Transitions Proceedings*, vol. 3, pp. 91–99, 2022. <https://doi.org/10.1016/j.gltp.2022.04.020>.
- [16] M. A. Palomino, "Evaluating the effectiveness of text pre-processing in sentiment analysis," *Applied Sciences*, vol. 12, no. 17, 2022. <https://doi.org/10.3390/app12178765>.
- [17] Y. Albalawi, J. Buckley, and N. S. Nikolov, "Investigating the impact of pre-processing techniques and pre-trained word embeddings in detecting arabic health information on social media," *J. Big Data*, vol. 8, no. 1, 2021. <https://doi.org/10.1186/s40537-021-00488-w>.
- [18] Gilbert, S. Alam, and M. I. Sulisty, "Analisis sentimen berdasarkan ulasan pengguna aplikasi mypertamina pada google playstore menggunakan metode naïve bayes," *STORAGE J. Ilm. Tek. dan Ilmu Komput.*, vol. 2, no. 3, pp. 100–108, 2023. <https://doi.org/10.55123/storage.v2i3.2333>.
- [19] D. Chicco, L. Oneto, and E. Tavazzi, "Eleven quick tips for data cleaning and feature engineering," *PLoS Comput. Biol.*, vol. 18, no. 12, pp. 1–21, 2022. <https://doi.org/10.1371/journal.pcbi.1010718>.
- [20] V. Zouhar, C. Meister, J. L. Gastaldi, L. Du, M. Sachan, and R. Cotterell, "Tokenization and the noiseless channel," in *Proc. Annu. Meet. Assoc. Comput. Linguist.*, vol. 1, pp. 5184–5207, 2023. <https://doi.org/10.18653/v1/2023.acl-long.284>.
- [21] S. Sarica and J. Luo, "Stopwords in technical language processing," *PLoS One*, vol. 16, no. 8, pp. 1–13, 2021. <https://doi.org/10.1371/journal.pone.0254937>.
- [22] M. U. Albab, Y. K. P., and M. N. Fawaiq, "Optimization of the stemming technique on text preprocessing president 3 periods topic," *J. Transform.*, vol. 20, no. 2, pp. 1–12, 2023. <https://doi.org/10.26623/transformatika.v20i2.5374>.
- [23] Z. Zhan, "Comparative analysis of tf-idf and word2vec in sentiment analysis: A case of food reviews," *ITM Web Conf.*, vol. 70, p. 02013, 2025. <https://doi.org/10.1051/itmconf/20257002013>.

- [24] K. W. Church, “Emerging trends: Word2vec,” *Nat. Lang. Eng.*, vol. 23, no. 1, pp. 155–162, 2017. <https://doi.org/10.1017/S1351324916000334>.
- [25] J. A. Ilemobayo *et al.*, “Hyperparameter tuning in machine learning: A comprehensive review,” *J. Eng. Res. Reports*, vol. 26, no. 6, pp. 388–395, 2024. <https://doi.org/10.9734/jerr/2024/v26i61188>.
- [26] P. Liashchynskiy and P. Liashchynskiy, “Grid search, random search, genetic algorithm: A big comparison for nas,” pp. 1–11, 2019. <http://arxiv.org/abs/1912.06059>.
- [27] S. E. H. Yulianti, O. Soesanto, and Y. Sukmawaty, “Penerapan metode extreme gradient boosting (xgboost) pada klasifikasi nasabah kartu kredit,” *J. Math. Theory Appl.*, vol. 4, no. 1, pp. 21–26, 2022. <https://doi.org/10.31605/jomta.v4i1.1792>.
- [28] S. Sathyanarayanan, “Confusion matrix-based performance evaluation metrics,” *African J. Biomed. Res.*, vol. 27, no. 4, pp. 4023–4031, 2024. <https://doi.org/10.53555/ajbr.v27i4s.4345>.
- [29] G. Zeng, “Invariance properties and evaluation metrics derived from the confusion matrix in multiclass classification,” *Mathematics*, vol. 13, no. 16, 2025. <https://doi.org/10.3390/math13162609>.
- [30] J. L. Rizky and Z. P. Putra, “Analisis perbandingan algoritma pembelajaran mesin untuk meningkatkan akurasi dan klasifikasi tumor otak,” *Indonesian Journal of Applied Informatics*, vol. 9, no. 1, pp. 31–44, 2024. <https://doi.org/10.20961/ijai.v9i1.90101>.
- [31] N. Ruhyana, K. Salsabila, A. Agung, and T. Mardiana, “Sentiment analysis of user reviews bri mobile application with gradient boost method,” *Jurnal Riset Informatika*, vol. 7, no. 230, pp. 48–54, 2023. <https://doi.org/10.34288/jri.v7i2.342>.