
PERBANDINGAN KINERJA SUPPORT VECTOR MACHINE DAN RANDOM FOREST UNTUK ULASAN ROBLOX DI PLAY STORE

Deni¹⁾, Hannie, S.Kom., MMSI.²⁾, Billy Ibrahim Hasbi, M.Kom.³⁾

Program Studi Sistem Informasi, Fakultas Ilmu Komputer

Universitas Singaperbangsa Karawang

Jl. HS.Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat 41361

e-mail: 2210631250008@student.unsika.ac.id¹⁾, emailaddress2²⁾

Abstrak

Game Roblox merupakan salah satu platform permainan daring dengan jumlah pengguna yang sangat besar dan ulasan pengguna yang beragam di Google Play Store. Ulasan tersebut mengandung informasi penting mengenai persepsi, kepuasan, serta keluhan pengguna terhadap kualitas aplikasi. Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna Roblox berbahasa Indonesia serta membandingkan kinerja algoritma *Support Vector Machine* (SVM) dan *Random Forest* (RF) dalam mengklasifikasikan sentimen positif dan negatif. Metodologi yang digunakan adalah SEMMA, yang meliputi tahap pengumpulan data melalui *web scraping*, eksplorasi data, preprocessing teks, pemodelan, dan evaluasi. Data yang digunakan berjumlah 5.174 ulasan dengan periode 1 Juni hingga 1 Oktober 2025. Pelabelan sentimen dilakukan secara otomatis berdasarkan rating bintang dan divalidasi oleh ahli bahasa. Proses preprocessing meliputi *case folding*, *cleaning*, normalisasi, tokenisasi, *stopword removal*, dan *stemming*, sedangkan ekstraksi fitur menggunakan metode TF-IDF. Pemodelan dilakukan dengan pembagian data 80% data latih dan 20% data uji. Hasil penelitian menunjukkan bahwa algoritma SVM menghasilkan kinerja yang lebih baik dibandingkan Random Forest dengan nilai akurasi sebesar 84,35%, sedangkan Random Forest memperoleh akurasi sebesar 83,19%. Berdasarkan hasil evaluasi tersebut, SVM dinilai lebih optimal untuk analisis sentimen ulasan pengguna game Roblox di Google Play Store.

Kata kunci : analisis sentimen; roblox; support vector machine; random forest

1. Pendahuluan

Kemajuan teknologi informasi dan komunikasi telah mendorong perubahan signifikan dalam pola kehidupan masyarakat, di mana perangkat mobile dan internet menjadi bagian integral dari aktivitas sehari-hari. Perkembangan ini memungkinkan masyarakat untuk berinteraksi, belajar, bekerja, serta menikmati hiburan secara daring, salah satunya melalui *game online*. *Game online* tidak lagi sekadar berfungsi sebagai sarana hiburan, tetapi juga berkembang menjadi media sosial, ruang kreativitas, dan bahkan bagian dari ekosistem ekonomi digital [1]. Seiring dengan pertumbuhan tersebut, industri *game online* menghadapi berbagai tantangan, seperti peningkatan kualitas pengalaman pengguna, keamanan, monetisasi, serta kepuasan pengguna, sekaligus membuka peluang besar dalam pemanfaatan data interaksi pengguna berskala besar. Dalam konteks ini, ulasan pengguna (*user reviews*) menjadi sumber data penting yang merefleksikan persepsi, kepuasan, kritik, dan harapan pengguna terhadap suatu aplikasi *game* [2].

Salah satu platform *game online* yang mengalami pertumbuhan pesat dalam beberapa tahun terakhir adalah Roblox. Berbeda dengan *game* konvensional, Roblox merupakan sebuah platform berbasis *user-generated content* yang memungkinkan pengguna tidak hanya bermain, tetapi juga menciptakan dan membagikan *game* mereka sendiri [3]. Karakteristik ini menjadikan Roblox sebagai fenomena global yang menjangkau berbagai kelompok usia dan wilayah. Berdasarkan data DemandSage, pada kuartal II tahun 2025 Roblox tercatat memiliki sekitar 111,8 juta pengguna aktif harian secara global, meningkat sebesar 41% dibandingkan tahun sebelumnya, yang menunjukkan tingginya tingkat keterlibatan pengguna [4]. Selain itu, laporan AiTectonic menyebutkan bahwa jumlah pengguna aktif bulanan Roblox mencapai sekitar 380 juta pengguna pada tahun yang sama, menegaskan posisinya sebagai salah satu platform *game* sosial dengan pertumbuhan tercepat di dunia [5].

Dari sisi demografi, pengguna Roblox berasal dari latar belakang usia yang beragam, mulai dari anak-anak hingga dewasa muda. Data Statista pada kuartal II tahun 2025 menunjukkan bahwa pengguna berusia di bawah 13 tahun mencapai sekitar 39,7 juta pengguna aktif harian, sementara pengguna berusia di atas 13 tahun mencapai lebih dari 72 juta pengguna aktif [5]. Keberagaman ini menyebabkan ulasan pengguna di platform seperti Google Play Store memiliki variasi emosi, gaya bahasa, serta topik pembahasan yang sangat luas. Oleh karena itu, analisis sentimen terhadap ulasan pengguna menjadi penting untuk memahami kecenderungan persepsi dan tingkat kepuasan pengguna terhadap performa dan pengalaman bermain Roblox, serta sebagai bahan evaluasi bagi pengembang dalam meningkatkan kualitas aplikasi.

Analisis sentimen merupakan bagian dari *text mining* yang bertujuan untuk mengklasifikasikan opini pengguna ke dalam kategori tertentu, seperti positif dan negatif, dengan memanfaatkan teknik *machine learning*. Berbagai algoritma telah digunakan dalam analisis sentimen, di antaranya *Support Vector Machine* (SVM) dan *Random Forest* (RF). SVM dikenal efektif dalam menangani data berdimensi tinggi dan menghasilkan batas klasifikasi yang optimal, sedangkan *Random Forest* sebagai metode ensemble berbasis *decision tree* memiliki keunggulan dalam mengatasi variabilitas data dan mengurangi risiko *overfitting* [6]. Pemilihan dan perbandingan kedua algoritma ini menjadi relevan dalam konteks ulasan pengguna *game* yang bersifat besar, kompleks, dan tidak terstruktur, seperti ulasan Roblox di Google Play Store.

Beberapa penelitian terdahulu menunjukkan keunggulan algoritma SVM dalam analisis sentimen. Penelitian yang membandingkan SVM dan *Naive Bayes* pada ulasan pengguna Roblox menemukan bahwa SVM menghasilkan tingkat akurasi tertinggi, mencapai 90% pada skema pembagian data 90:10 [7]. Studi lain yang membandingkan SVM dan *Random Forest* pada ulasan aplikasi KAI Access di Google Play Store juga menunjukkan bahwa SVM memperoleh akurasi lebih tinggi, yaitu 97%, dibandingkan *Random Forest* sebesar 93% [8]. Selain itu, penelitian terhadap ulasan aplikasi Threads di Google Play Store menunjukkan bahwa algoritma SVM dan *Multinomial Naive Bayes* memiliki performa yang relatif seimbang dengan akurasi sekitar 81%, yang menegaskan efektivitas pendekatan *machine learning* dalam analisis sentimen ulasan aplikasi [9].

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk membandingkan kinerja algoritma *Support Vector Machine* dan *Random Forest* dalam menganalisis sentimen ulasan pengguna game Roblox berbahasa Indonesia di Google Play Store. Hasil penelitian ini diharapkan dapat memberikan kontribusi secara teoretis dalam pengembangan kajian analisis sentimen serta memberikan manfaat praktis bagi pengembang dan pemangku kepentingan dalam meningkatkan kualitas dan kepuasan pengguna aplikasi game Roblox.

2. Landasan Teori

Text Mining dan Analisis Sentimen

Text mining merupakan proses sistematis untuk mengekstraksi dan menganalisis data teks tidak terstruktur dalam jumlah besar guna menemukan pola, konsep, topik, serta informasi bermakna lainnya [10]. Proses ini mencakup tahapan pengumpulan data, pembersihan teks, transformasi, ekstraksi fitur, hingga analisis menggunakan pendekatan statistik atau *machine learning* [11].

Salah satu penerapan utama *text mining* adalah analisis sentimen, yaitu teknik untuk mengidentifikasi dan mengklasifikasikan opini atau emosi dalam teks ke dalam kategori tertentu, seperti positif dan negatif [12]. Analisis sentimen memiliki peran penting dalam memahami persepsi pengguna terhadap aplikasi atau *game* digital, karena ulasan pengguna merefleksikan tingkat kepuasan, kritik, dan harapan terhadap kualitas layanan [13]. Pendekatan *machine learning* banyak digunakan dalam analisis sentimen karena mampu menangani data berskala besar dan kompleks dengan tingkat akurasi yang baik.

Metode Pembobotan Fitur TF-IDF

Term Frequency–Inverse Document Frequency (TF-IDF) merupakan metode pembobotan kata yang digunakan untuk merepresentasikan dokumen teks ke dalam bentuk numerik [14]. TF-IDF mengukur tingkat kepentingan suatu kata berdasarkan frekuensi kemunculannya dalam sebuah dokumen (*Term Frequency*) dan tingkat kelangkaannya dalam seluruh korpus dokumen (*Inverse Document Frequency*).

Secara matematis, bobot TF-IDF dirumuskan sebagai berikut:

$$TF-IDF(w_i, d) = TF(w_i, d) \times IDF(w_i) \quad (1)$$

$$IDF(w_i) = \log\left(\frac{N}{DF_i} + 1\right) \quad (2)$$

Dengan keterangan sebagai berikut:

w_i	: kata ke- i (term) dalam dokumen
d	: dokumen ke- d
$TF(w_i, d)$: frekuensi kemunculan kata w_i dalam dokumen d
$IDF(w_i, d)$: nilai inverse document frequency dari kata w_i
N	: jumlah seluruh dokumen dalam dataset
DF_i	: jumlah dokumen yang mengandung kata w_i

Metode TF-IDF efektif digunakan pada data ulasan karena mampu menonjolkan kata-kata yang bersifat informatif, serta mengurangi pengaruh kata umum yang kurang bermakna [6].

Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan algoritma klasifikasi berbasis optimasi yang bertujuan menentukan *hyperplane* terbaik sebagai batas pemisah antar kelas data [15]. SVM bekerja dengan memaksimalkan margin antara *hyperplane* dan data terdekat yang disebut *support vector*. Keunggulan utama SVM terletak pada kemampuannya menangani data berdimensi tinggi, sehingga sangat efektif digunakan pada klasifikasi teks berbasis TF-IDF.

Random Forest (RF)

Random Forest adalah algoritma klasifikasi berbasis *ensemble learning* yang menggabungkan sejumlah *Decision Tree* untuk menghasilkan keputusan akhir melalui mekanisme *majority voting* [16]. Setiap pohon dibangun menggunakan subset data dan fitur yang berbeda, sehingga meningkatkan stabilitas model dan mengurangi risiko *overfitting*. *Random*

Forest digunakan sebagai algoritma pembanding karena kemampuannya menangani variabilitas data dan kompleksitas pola dalam teks ulasan.

SEMMA

Sample, Explore, Modify, Model, Assess (SEMMA) merupakan metodologi analisis data yang dikembangkan oleh SAS Institute untuk mendukung proses data mining secara terstruktur dan sistematis. Tahap *Sample* digunakan untuk pengumpulan data ulasan, *Explore* untuk memahami karakteristik data, *Modify* untuk proses pra-pemrosesan teks, *Model* untuk penerapan algoritma klasifikasi, serta *Assess* untuk mengevaluasi kinerja model. Metodologi SEMMA sesuai digunakan dalam analisis sentimen karena memberikan alur kerja yang jelas dari tahap awal hingga evaluasi hasil [17].

Confusion Matrix

Confusion Matrix merupakan metode evaluasi kinerja model klasifikasi dengan membandingkan hasil prediksi terhadap label sebenarnya pada data uji [18]. Matriks ini terdiri dari empat komponen utama, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN).

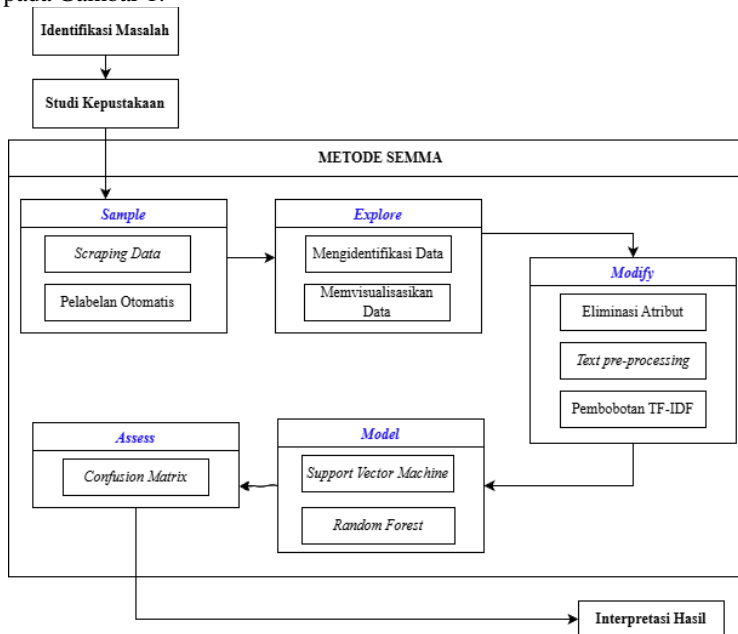
Tabel 1. *Confusion Matrix*

Predicted Class	Observed	
	True	False
True	True Positive (TP)	False Positive (FP)
False	False Negative (FN)	True Negative (TN)

Berdasarkan *confusion matrix*, kinerja model dievaluasi menggunakan beberapa metrik, antara lain *accuracy*, *precision*, *recall*, dan *F1-score* [18]. *Accuracy* menunjukkan tingkat ketepatan model secara keseluruhan, *precision* mengukur ketepatan prediksi positif, *recall* menunjukkan kemampuan model mengenali data positif, sedangkan *F1-score* merupakan harmonisasi antara *precision* dan *recall*. Evaluasi ini digunakan untuk menilai efektivitas algoritma dalam mengklasifikasikan sentimen ulasan pengguna [19].

3. Metode Penelitian

Pelaksanaan penelitian ini mengikuti tahapan metodologi penelitian dengan menggunakan pendekatan SEMMA. Tahapan proses SEMMA disajikan pada Gambar 1.



Gambar 1. Bagan Alur Penelitian

Identifikasi Masalah

Tahap awal penelitian dimulai dengan mengidentifikasi permasalahan yang terjadi secara nyata, kemudian dilakukan perumusan masalah untuk menentukan arah serta solusi yang paling tepat dalam penyelesaiannya.

Studi Kepustakaan

Selanjutnya akan dilakukan riset dan pengumpulan informasi yang relevan dengan topik penelitian berdasarkan beberapa sumber seperti jurnal, website, dan karya ilmiah lainnya.

Sample

Penelitian ini menggunakan data ulasan pengguna *game* Roblox yang diperoleh dari Google Play Store melalui proses *web scraping* menggunakan Python dengan periode pengambilan data 1 Juni hingga 1 Oktober 2025. Seluruh data hasil *scraping* digunakan dalam penelitian tanpa perhitungan ukuran sampel statistik. Pelabelan sentimen dilakukan secara otomatis berdasarkan nilai rating pengguna, di mana rating tinggi dikategorikan sebagai sentimen positif dan rating rendah sebagai sentimen negatif, kemudian di validasi oleh ahli bahasa.

Explore

Tahap eksplorasi dilakukan untuk memahami karakteristik data hasil *scraping* melalui analisis deskriptif dan visualisasi. Beberapa teknik visualisasi, seperti *word cloud* dan grafik distribusi data, digunakan untuk mengidentifikasi pola dan kecenderungan sentimen yang terdapat dalam data ulasan.

Modify

Pada tahap ini dilakukan pengolahan dan persiapan data melalui seleksi atribut dan pembersihan teks. Proses text preprocessing meliputi *case folding*, tokenisasi, penghapusan tanda baca, *stopword removal*, dan *stemming*. Selanjutnya, data teks diubah ke dalam bentuk numerik menggunakan metode *Term Frequency–Inverse Document Frequency (TF-IDF)*.

Model

Tahap pemodelan dilakukan dengan menerapkan algoritma *Support Vector Machine* dan *Random Forest* pada data yang telah diproses. Proses pelatihan dan pengujian model dilakukan menggunakan bahasa pemrograman Python pada Google Colaboratory dengan pembagian data latih dan data uji sebesar 80:20.

Assess

Evaluasi kinerja model dilakukan menggunakan *Confusion Matrix* untuk memperoleh nilai akurasi, *precision*, dan *recall*. Hasil evaluasi digunakan untuk membandingkan kinerja masing-masing algoritma dalam mengklasifikasikan sentimen ulasan pengguna.

Interpretasi Hasil

Tahap interpretasi dilakukan untuk menganalisis hasil evaluasi model sehingga dapat ditarik kesimpulan mengenai algoritma yang memiliki kinerja terbaik serta memberikan rekomendasi yang relevan berdasarkan temuan penelitian.

4. Hasil Penelitian

Identifikasi Masalah

Meningkatnya jumlah ulasan pengguna *game* Roblox di Google Play Store menghasilkan data opini yang beragam dan bernilai untuk dianalisis. Data ulasan tersebut dapat dimanfaatkan untuk memahami persepsi serta tingkat kepuasan pengguna. Oleh karena itu, diperlukan analisis sentimen menggunakan algoritma *Support Vector Machine* dan *Random Forest* untuk membandingkan kinerja keduanya dalam mengklasifikasikan sentimen positif dan negatif.

Studi Kepustakaan

Studi kepustakaan dilakukan dengan menelaah berbagai jurnal dan sumber ilmiah yang membahas analisis sentimen dan penerapan algoritma *Support Vector Machine* serta *Random Forest* pada data teks. Hasil kajian digunakan sebagai dasar dalam pemilihan metode penelitian, tahapan analisis, dan indikator evaluasi kinerja model.

Sample

Penelitian ini menggunakan 5.174 ulasan pengguna *game* Roblox yang diperoleh dari Google Play Store melalui web *scraping*. Pelabelan sentimen dilakukan secara otomatis berdasarkan rating, dengan rating 4–5 sebagai sentimen positif dan rating 1–2 sebagai sentimen negatif. Untuk meningkatkan keandalan data, hasil pelabelan divalidasi oleh ahli bahasa sebelum digunakan pada tahap pemodelan. Untuk melihat hasil dari *scraping*, dapat dilihat pada Gambar 2.

	reviewId	userName	userImage	content	score	thumbsUpCount	reviewCreatedVersion	at	replyContent	repliedAt	appVersion
0	0819576-ca62-4a9a-864b-127876c9c94a	Eriana patrisia	lh.googleusercontent.com/a/ACg8oc...	https://play- bagus ya game nya	5	0	2690721.0	30/09/2025 23:55	NaN	NaN	2690721.0
1	4081c19-53e6-4c6d-b4af-9268dda8b59b	Novia Fransiska	lh.googleusercontent.com/a/ALV-U...	https://play- mantap	5	0	2690721.0	30/09/2025 23:49	NaN	NaN	2690721.0
2	70407f4-7314-4709-b329-931cb4965114	Yenny Sahfitri	lh.googleusercontent.com/a/ACg8oc...	https://play- gamenya kayak game kotak kotak anak ngusan	5	0	2692843.0	30/09/2025 23:47	NaN	NaN	2692843.0
3	6113e9f3-1631-45c5-815f-39f7aecd936c	Rachel Ayunda	lh.googleusercontent.com/a/ACg8oc...	https://play- Game nya bagus juga ya untuk gabut. Dan jadinya...	5	0	2690721.0	30/09/2025 23:47	NaN	NaN	2690721.0
4	76e07c29-1308-4891-bd37-d182c269d3c	lusanawidiwati	lh.googleusercontent.com/a/ACg8oc...	https://play- seru banget	5	0	2692843.0	30/09/2025 23:40	NaN	NaN	2692843.0

Gambar 2. Hasil Scraping

Untuk melihat hasil labeling otomatis, dapat dilihat pada Gambar 3.

DataFrame dengan Label Sentimen:

	reviewId	userName	userImage	content	score	thumbsUpCount	reviewCreatedVersion	at	replyContent	repliedAt	appVersion	sentiment_label
0	0819576-ca62-4a9a-864b-127876c9c94a	Eriana patrisia	lh.googleusercontent.com/a/ACg8oc...	https://play- bagus ya game nya	5	0	2690721.0	30/09/2025 23:55	NaN	NaN	2690721.0	positif
1	4081c19-53e6-4c6d-b4af-9268dda8b59b	Novia Fransiska	lh.googleusercontent.com/a/ALV-U...	https://play- mantap	5	0	2690721.0	30/09/2025 23:49	NaN	NaN	2690721.0	positif
2	70407f4-7314-4709-b329-931cb4965114	Yenny Sahfitri	lh.googleusercontent.com/a/ACg8oc...	https://play- gamenya kayak game kotak kotak anak ngusan	5	0	2692843.0	30/09/2025 23:47	NaN	NaN	2692843.0	positif
3	6113e9f3-1631-45c5-815f-39f7aecd936c	Rachel Ayunda	lh.googleusercontent.com/a/ACg8oc...	https://play- Game nya bagus juga ya untuk gabut. Dan jadinya...	5	0	2690721.0	30/09/2025 23:47	NaN	NaN	2690721.0	positif
4	76e07c29-1308-4891-bd37-d182c269d3c	lusanawidiwati	lh.googleusercontent.com/a/ACg8oc...	https://play- seru banget	5	0	2692843.0	30/09/2025 23:40	NaN	NaN	2692843.0	positif

a. Explore

Meingidentifikasi Data

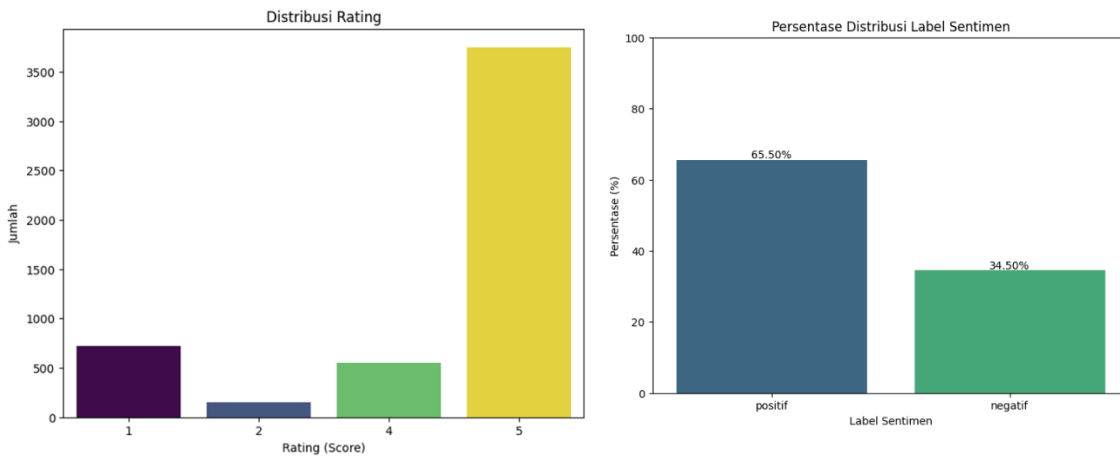
Tahap eksplorasi dilakukan untuk memahami karakteristik awal data ulasan pengguna Roblox yang berjumlah 5.174 ulasan, dengan atribut utama berupa teks ulasan dan rating pengguna. Hasil analisis statistik deskriptif menunjukkan nilai rata-rata rating sebesar 4,25 dengan median 5, yang mengindikasikan bahwa sebagian besar pengguna memberikan penilaian positif terhadap game Roblox. Hasil dari analisis deskriptif bisa dilihat pada Tabel 2.

Tabel 2. Statistika Deskriptif

	Score
count	5174.000000
mean	4.247198
std	1.427332
Min	1.000000
25%	4.000000
50%	5.000000
75%	5.000000
max	5.000000

Visualisasi Data

Hasil visualisasi distribusi rating memperlihatkan dominasi rating 4 dan 5, dengan rating 5 sebagai yang paling banyak diberikan. Berdasarkan pelabelan sentimen, sebanyak 65,50% ulasan tergolong sentimen positif, sedangkan 34,50% termasuk sentimen negatif. Meskipun sentimen positif lebih dominan, proporsi sentimen negatif yang cukup signifikan menunjukkan adanya aspek yang masih perlu dievaluasi. Hasil distribusi rating dan persentase distribusi label sentimen dapat dilihat pada Gambar 4.



Gambar 4. Distribusi Rating dan Persentase Distribusi Label Sentimen

Visualisasi Word Cloud

Visualisasi *word cloud* menunjukkan bahwa sentimen positif didominasi oleh kata-kata yang menggambarkan keseruan dan kepuasan pengguna terhadap *gameplay*, sedangkan sentimen negatif didominasi oleh kata-kata yang berkaitan dengan permasalahan teknis, seperti gangguan performa, kendala *login*, dan *error* aplikasi. Temuan ini mengindikasikan bahwa perbaikan pada stabilitas dan performa sistem berpotensi meningkatkan pengalaman bermain pengguna. Hasil *word cloud* dapat dilihat pada Gambar 5.



Gambar 5. Visualisasi Word Cloud Positif dan Negatif

**b. Modify
Eliminasi Atribut**

Dari seluruh atribut tersebut, yang akan digunakan hanya *content*, *score*, dan validasi sebagai teks utama untuk analisis sentimen karena berisi isi ulasan pengguna. Contoh data hasil eliminasi dapat dilihat pada Tabel 3.

Tabel 3. Hasil Eliminasi data

No	Content	Rating	Validasi
1	bagus ya game nya	5	Positif
2	mantap	5	Positif
3	gamenya kayak game kotak kotak anak ingusan	5	Negatif
...
5174	Sistem verifikasi di laptop ini ngeselin anjir, sampai 9 tahap buat buat login akun, buat apa coba? main di handphone juga rendering nya minta ampun dah	1	Negatif

**Preprocessing
Case Folding**

Tahap preprocessing dilakukan untuk membersihkan dan menyiapkan data teks ulasan agar siap digunakan pada proses pemodelan. Proses ini diawali dengan *case folding*, yaitu mengubah seluruh huruf pada teks menjadi huruf kecil untuk menyeragamkan penulisan kata. Hasil case folding dapat dilihat pada Tabel 4.

Tabel 4. Hasil *Case Folding*

No.	Sebelum <i>Case Folding</i>	Setelah <i>Case Folding</i>
1	ini permainan bagus bgtt sihh bisa mabar tapi yang aku gak puas avanya tapi gapapa aku kasih bintang 5 🌟❤️	ini permainan bagus bgtt sihh bisa mabar tapi yang aku gak puas avanya tapi gapapa aku kasih bintang 5 🌟❤️
2	bagus banget game nya , bisa Mabar sama teman atau BESTie kita puas bgt pas pertama punya Roblox	bagus banget game nya , bisa mabar sama teman atau bestie kita puas bgt pas pertama punya roblox
3	gak ada opsi nomor telepon indonesia di pemulihan akun roblox	gak ada opsi nomor telepon indonesia di pemulihan akun roblox

Cleaning

Tahap berikutnya adalah *cleaning*, digunakan untuk menghapus elemen yang tidak relevan seperti tanda baca, angka, emotikon, URL, dan karakter khusus. Hasil *cleaning* dapat dilihat pada Tabel 5.

Tabel 5. Hasil *Cleaning*

No.	Sebelum <i>Cleaning</i>	Setelah <i>Cleaning</i>
1	ini permainan bagus bgtt sihh bisa mabar tapi yang aku gak puas avanya tapi gapapa aku kasih bintang 5 🌟❤️	ini permainan bagus bgtt sihh bisa mabar tapi yang aku gak puas avanya tapi gapapa aku kasih bintang
2	bagus banget game nya , bisa Mabar sama teman atau BESTie kita puas bgt pas pertama punya Roblox	bagus banget game nya bisa mabar sama teman atau bestie kita puas bgt pas pertama punya roblox
3	gak ada opsi nomor telepon indonesia di pemulihan akun roblox	gak ada opsi nomor telepon indonesia di pemulihan akun roblox

Normalize

Tahap berikutnya *normalize*, yaitu mengubah kata tidak baku dan singkatan ke bentuk baku sesuai kaidah Bahasa Indonesia agar makna teks lebih konsisten. Hasil *normalize* dapat dilihat pada Tabel 6.

Tabel 6. Hasil *Normalize*

No.	Sebelum <i>Normalize</i>	Setelah <i>Normalize</i>
1	ini permainan bagus bgtt sihh bisa mabar tapi yang aku gak puas avanya tapi gapapa aku kasih bintang	ini permainan bagus banget sih bisa mabar tapi yang aku tidak puas avanya tapi gapapa aku kasih bintang
2	bagus banget game nya bisa mabar sama teman atau bestie kita puas bgt pas pertama punya roblox	bagus banget game nya bisa mabar sama teman atau sahabat kita puas banget pas pertama punya roblox
3	gak ada opsi nomor telepon indonesia di pemulihan akun roblox	tidak ada opsi nomor telepon indonesia di pemulihan akun roblox

Tokenizing

Tokenizing dilakukan untuk memecah teks menjadi unit kata, sehingga memudahkan pengolahan secara komputasional. Hasil *tokenizing* dapat dilihat pada Tabel 7.

Tabel 7. Hasil *Tokenizing*

No.	Sebelum <i>Tokenizing</i>	Setelah <i>Tokenizing</i>
1	ini permainan bagus bgtt sihh bisa mabar tapi yang aku gak puas avanya tapi gapapa aku kasih bintang	ini permainan bagus banget sih bisa mabar tapi yang aku tidak puas avanya tapi gapapa aku kasih bintang
2	bagus banget game nya bisa mabar sama teman atau bestie kita puas bgt pas pertama punya roblox	bagus banget game nya bisa mabar sama teman atau sahabat kita puas banget pas pertama punya roblox
3	gak ada opsi nomor telepon indonesia di pemulihan akun roblox	tidak ada opsi nomor telepon indonesia di pemulihan akun roblox

Stopword removal

Tahap berikutnya *stopword removal* untuk menghilangkan kata-kata umum yang tidak memiliki kontribusi signifikan terhadap sentimen, sehingga teks menjadi lebih fokus pada kata bermakna. Hasil *stopword removal* dapat dilihat pada Tabel 8.

Tabel 8. Hasil *Stopword Removal*

No.	Sebelum <i>Stopword removal</i>	Setelah <i>Stopword removal</i>
1	ini permainan bagus banget sih bisa mabar tapi yang aku tidak puas avanya tapi gapapa aku kasih bintang	ini permainan bagus banget sih bisa mabar tapi yang aku tidak puas avanya tapi gapapa aku kasih bintang
2	permainan bagus banget sih mabar yang aku puas avanya gapapa aku kasih bintang	permainan bagus banget sih mabar yang aku puas avanya gapapa aku kasih bintang
3	bagus banget game nya bisa mabar sama teman atau sahabat kita puas banget pas pertama punya roblox	bagus banget game nya bisa mabar sama teman atau sahabat kita puas banget pas pertama punya roblox

Stemming

Tahap terakhir adalah *stemming*, yaitu mengubah kata berimbuhan menjadi kata dasar menggunakan algoritma Sastrawi guna meningkatkan konsistensi representasi kata dalam data. Hasil *stemming* dapat dilihat pada Tabel 9.

Tabel 9. Hasil *Stemming*

No.	Sebelum <i>Stemming</i>	Setelah <i>Stemming</i>
1	ini permainan bagus banget sih bisa mabar tapi yang aku tidak puas avanya tapi gapapa aku kasih bintang	ini permainan bagus banget sih bisa mabar tapi yang aku tidak puas avanya tapi gapapa aku kasih bintang
2	permainan bagus banget sih mabar yang aku puas avanya gapapa aku kasih bintang	permainan bagus banget sih mabar yang aku puas avanya gapapa aku kasih bintang
3	bagus banget game nya bisa mabar sama teman atau sahabat kita puas banget pas pertama punya roblox	bagus banget game nya bisa mabar sama teman atau sahabat kita puas banget pas pertama punya roblox

Pembobotan TF-IDF

Data ulasan yang telah melalui tahap preprocessing diubah ke dalam bentuk numerik menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) dengan pendekatan unigram dan jumlah fitur maksimum sebanyak 300 kata. Metode ini digunakan untuk merepresentasikan tingkat kepentingan kata dalam setiap dokumen ulasan. Hasil pembobotan menunjukkan bahwa kata-kata seperti *game*, *bagus*, *banget*, *sangat*, *main*, *seru*, dan *roblox* memiliki nilai TF-IDF tertinggi. Dominasi kata-kata bernuansa positif tersebut mengindikasikan bahwa sebagian besar ulasan pengguna mengandung opini positif terhadap *game* Roblox. Untuk melihat hasil pembobotan TF-IDF 10 teratas, dapat dilihat pada Tabel 10.

Tabel 10. Top 10 hasil pembobotan kata TF-IDF

Index	Term	TF-IDF
17	game	2.354423
5	bagus	2.087716
8	banget	1.792257
50	sangat	1.491595
32	main	1.265200
56	seru	1.207063
48	roblox	1.134230
34	mantap	1.000000
40	orang	0.984687
2	anak	0.842261

c. Model**Klasifikasi *Support Vector Machine***

Model Support Vector Machine (SVM) dengan kernel linear digunakan untuk melakukan klasifikasi sentimen positif dan negatif pada ulasan pengguna. Pemodelan dilakukan menggunakan skenario pembagian data 80:20, dimana data latih telah melalui proses ekstraksi fitur TF-IDF. Hasil dari akurasi *Support Vector Machine* dalam dilihat pada Gambar 6.

```

=====
HASIL SVM LINEAR
=====
Akurasi: 0.8434782608695652
      precision    recall  f1-score   support

negatif      0.81      0.72      0.76      357
positif      0.86      0.91      0.88      678

accuracy                    0.84      1035
macro avg      0.83      0.81      0.82      1035
weighted avg   0.84      0.84      0.84      1035

```

Gambar 6. Hasil akurasi *Support Vector Machine*

Berdasarkan hasil pengujian *model Support Vector Machine* (SVM) dengan kernel linear, diperoleh nilai akurasi sebesar 84,35%. Pada kelas sentimen negatif, model SVM menghasilkan nilai *precision* sebesar 0,81, *recall* sebesar 0,72, dan *f1-score* sebesar 0,76. Sementara itu, pada kelas sentimen positif, diperoleh nilai *precision* sebesar 0,86, *recall* sebesar 0,91, dan *f1-score* sebesar 0,88.

Klasifikasi *Random Forest*

Sebagai pembandingan terhadap SVM, penelitian ini menggunakan algoritma *Random Forest* yang menggabungkan sejumlah pohon keputusan untuk meningkatkan kinerja dan keandalan hasil klasifikasi. Hasil dari akurasi *Random Forest* dalam dilihat pada Gambar .

```

=====
HASIL RANDOM FOREST
=====
Akurasi: 0.8318840579710145
      precision    recall  f1-score   support

negatif      0.78      0.71      0.74      357
positif      0.85      0.90      0.87      678

accuracy                    0.83      1035
macro avg      0.82      0.80      0.81      1035
weighted avg   0.83      0.83      0.83      1035

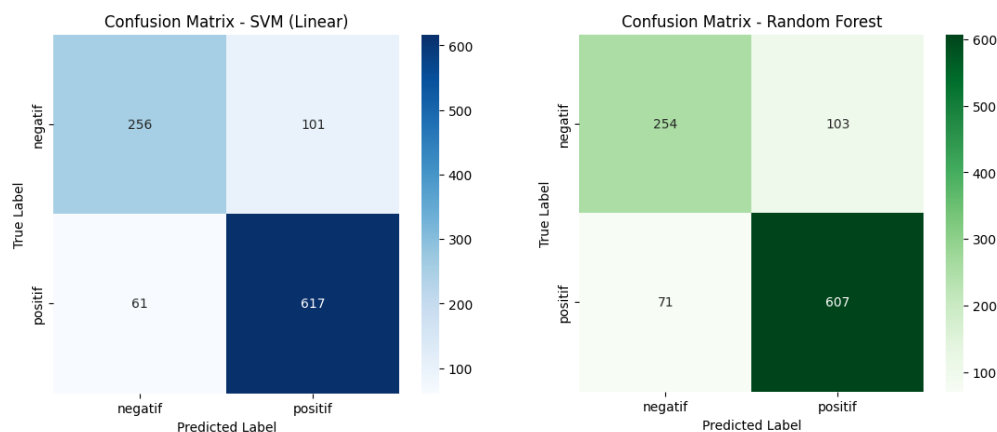
```

Gambar 7. Hasil akurasi *Random Forest*

Model Random Forest menghasilkan nilai akurasi sebesar 83,19%, sedikit lebih rendah dibandingkan dengan model SVM. Pada kelas sentimen negatif, *Random Forest* memperoleh nilai *precision* sebesar 0,78, *recall* sebesar 0,71, dan *f1-score* sebesar 0,74. Sedangkan pada kelas sentimen positif, model ini menghasilkan nilai *precision* sebesar 0,85, *recall* sebesar 0,90, dan *f1-score* sebesar 0,87.

d. Assess

Evaluasi kinerja *model Support Vector Machine* (SVM) dan *Random Forest* (RF) dilakukan menggunakan *confusion matrix* untuk menilai kemampuan masing-masing *model* dalam mengklasifikasikan sentimen positif dan negatif. Hasil evaluasi menunjukkan bahwa kedua model mampu melakukan klasifikasi dengan cukup baik, namun SVM menghasilkan jumlah kesalahan yang lebih rendah dibandingkan *Random Forest*. Untuk *confusion matrix* dapat dilihat pada Gambar 8.



Gambar 8. Confusion Matrix SVM dan RF

Berdasarkan hasil evaluasi menggunakan *confusion matrix*, diperoleh nilai *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)* dari masing-masing model klasifikasi, sebagaimana ditunjukkan pada Tabel 11.

Tabel 11. Hasil *Confusion Matrix* SVM dan RF

<i>Model</i>	TP	FP	FN	TN
SVM	617	101	61	256
<i>Random Forest</i>	607	103	71	254

Berdasarkan Tabel tersebut, *model Support Vector Machine (SVM)* menghasilkan jumlah *True Positive* dan *True Negative* yang lebih tinggi serta *False Positive* dan *False Negative* yang lebih rendah dibandingkan dengan *model Random Forest*. Nilai *False Negative* pada *Random Forest* lebih besar, yang menunjukkan bahwa model tersebut lebih sering keliru dalam mengklasifikasikan sentimen positif sebagai sentimen negatif. Selain itu, nilai *False Positive* *Random Forest* juga sedikit lebih tinggi dibandingkan SVM, yang mengindikasikan kestabilan klasifikasi SVM lebih baik. Dengan demikian, dapat disimpulkan bahwa model SVM memiliki kinerja yang lebih optimal dan seimbang dalam mengklasifikasikan sentimen ulasan pengguna *game* Roblox.

e. Interpretasi Hasil

Mayoritas ulasan pengguna *game* Roblox menunjukkan sentimen positif. Proses preprocessing dan pembobotan TF-IDF mampu merepresentasikan teks secara efektif untuk klasifikasi sentimen. Hasil evaluasi menunjukkan bahwa *Support Vector Machine* memiliki kinerja yang lebih baik dibandingkan *Random Forest*, sehingga lebih direkomendasikan untuk analisis sentimen ulasan Roblox berbahasa Indonesia dengan karakteristik data serupa.

5. Kesimpulan

Berdasarkan hasil penelitian terhadap ulasan pengguna *game* Roblox di Google Play Store menggunakan metodologi SEMMA, dapat disimpulkan bahwa proses pengumpulan data melalui *web scraping* menghasilkan 5.174 ulasan yang kemudian diproses melalui pelabelan sentimen berbasis rating dan divalidasi oleh ahli bahasa. Data yang telah melalui tahap preprocessing dan pembobotan TF-IDF berhasil direpresentasikan dengan baik untuk analisis sentimen. Penerapan algoritma *Support Vector Machine* dan *Random Forest* dengan pembagian data 80% data latih dan 20% data uji menunjukkan bahwa *Support Vector Machine* memiliki kinerja yang lebih unggul dengan nilai akurasi 84,35% dibandingkan *Random Forest* sebesar 83,19%, serta menghasilkan nilai *precision*, *recall*, dan *f1-score* yang lebih baik, terutama pada kelas sentimen negatif. Evaluasi menggunakan *confusion matrix* juga menunjukkan bahwa SVM menghasilkan kesalahan klasifikasi yang lebih rendah, sehingga dapat direkomendasikan sebagai algoritma yang lebih sesuai untuk analisis sentimen ulasan pengguna Roblox berbahasa Indonesia.

Untuk pengembangan penelitian selanjutnya, disarankan untuk mengeksplorasi algoritma klasifikasi lain atau pendekatan *deep learning*, mengombinasikan metode *machine learning* dengan pendekatan *lexicon-based* atau *hybrid*, serta memperluas jumlah dan sumber data agar hasil analisis sentimen memiliki tingkat generalisasi dan keandalan yang lebih baik.

Daftar Pustaka

- [1] S. Harahap and Z. S. Ritonga, "Analisis Statistik Telekomunikasi dan Pendidikan di Indonesia Tahun 2022 dan 2023," *Jurnal Pendidikan dan Kependidikan*, vol. 9, no. 1, pp. 29–35, Jun. 2024.

- [2] M. Fajar, Masyhuri, and Y. Muda, “Kecanduan Game Online pada Remaja,” 2024. doi: <https://doi.org/10.37985/jer.v5i3.1273>.
- [3] U. S. Yahsy and M. Syas, “Some rights reserved Komodifikasi Users pada Platform Game Online Roblox,” *JURNAL INTERACT*, vol. 11, no. 2, 2022, [Online]. Available: <http://ojs.atmajaya.ac.id/index.php/fiabikom/index>
- [4] S. Singh, “How Many People Play Roblox 2025 [Player Count & Downloads],” demandsage.com.
- [5] J. Clement, “Daily active users (DAU) of Roblox games worldwide from 4th quarter 2019 to 2nd quarter 2025, by age group,” [statista.com](https://www.statista.com).
- [6] Daryanti and Tri Widodo, “Analisis Sentimen Ulasan Penumpang Maskapai Penerbangan Indonesia Menggunakan Support Vector Machine, Naive Bayes, dan Random Forest,” *The Indonesian Journal of Computer Science*, vol. 13, no. 5, pp. 8437–8445, Oct. 2024, doi: [10.33022/ijcs.v13i5.4452](https://doi.org/10.33022/ijcs.v13i5.4452).
- [7] A. F. Alkindi, “Analisis Sentimen Ulasan Pengguna Pada Game Roblox Dengan Metode Support Vector Machine Dan Naive Bayes,” Universitas Lancang Timur, Pekanbaru, 2024.
- [8] N. B. Sidauruk, N. Riza, R. Nuraini, and S. Fatonah, “PENGUNAAN METODE SVM DAN RANDOM FOREST UNTUK ANALISIS SENTIMEN ULASAN PENGGUNA TERHADAP KAI ACCESS DI GOOGLE PLAYSTORE,” 2023.
- [9] M. Arya Java, M. Syafrullah, and F. Teknologi, “Analisis Sentimen Ulasan Pengguna Aplikasi Threads pada Google Play Store Menggunakan Multinomial Naive Bayes dan Support Vector Machine,” *Jurnal TICOM: Technology of Information and Communication*, vol. 12, no. 2, 2024, [Online]. Available: <https://github.com/nasalsabila/kamus-alay>
- [10] Runimeirati, A. Muis, and F. Muhammad, “Pelatihan Text Mining Menggunakan Bahasa Pemrograman Python,” *Jurnal Pengabdian Kepada Masyarakat*, vol. 3, pp. 36–46, 2023, [Online]. Available: <https://pusdig.web.id/index.php/abdimas/index>
- [11] T. Ridwansyah, “Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naive Bayes Classifier,” *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 2, no. 5, pp. 178–185, 2022, [Online]. Available: <https://djournals.com/klik>
- [12] Jasmarizal, Rahmaddeni, Junadhi, and M. K. Anam, “Penerapan Metode Support Vector Machine Untuk Analisis Sentimen Terhadap,” *Indonesian Journal of Computer Science Attribution*, vol. 13, no. 1, pp. 1438–1450, Feb. 2024, doi: <https://doi.org/10.33022/ijcs.v13i1.3654>.
- [13] M. R. Qisthiano, “Klasifikasi Terhadap Prediksi Kelulusan Mahasiswa Dengan Menggunakan Metode Support Vector Machine (SVM),” *2 st Proceeding STEKOM*, vol. 2022, pp. 204–8, 2022.
- [14] I. Adriansyah, M. D. Mahendra, E. Rasywir, and Y. Pratama, “Perbandingan Metode Random Forest Classifier dan SVM Pada Klasifikasi Kemampuan Level Beradaptasi Pembelajaran Jarak Jauh Siswa,” *Bulletin of Informatics and Data Science*, vol. 1, no. 2, pp. 98–103, 2022, [Online]. Available: <https://ejournal.pdsi.or.id/index.php/bids/index>
- [15] M. Namira, “Analisis Sentimen Vaksinasi Covid-19 Pada Komentar YouTube Dengan Menggunakan Algoritma Naive Bayes Clasifier (NBC) Dan Support Vector Machine (SVM),” Universitas Isla Negeri Syarif Hidayatullah Jakarta, Jakarta, 2023.
- [16] K. Irfansyah and Zaehol Fatah, “IMPLEMENTASI ALGORITMA CLUSTERING K-MEANS PADA PENGGUNA WARTEL DI PONDOK PESANTREN SALAFIYAH SYAFI’IYAH SUKOREJO,” *Jurnal Ilmiah Multidisiplin Ilmu*, vol. 1, no. 5, pp. 81–86, Dec. 2024, doi: [10.69714/55xet429](https://doi.org/10.69714/55xet429).
- [17] D. Normawati and S. A. Prayogi, “Implementasi Naive Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter,” Sep. 2021.
- [18] B. P. Pratiwi, A. S. Handayani, and Sarjana, “Pengukuran Kinerja Sistem Kualitas Udara Dengan Teknologi WSN Menggunakan Confusion Matrix,” *Jurnal Informatika Upgris*, vol. 6, no. 2, pp. 66–75, Dec. 2020.
- [19] Deni, R. Fatih Musthofa, H. Surya Herfiana, and B. Nurina Sari, “PERBANDINGAN ALGORITMA SUPPORT VECTOR MACHINE DAN RANDOM FOREST UNTUK ANALISIS SENTIMEN ISU IJAZAH PALSU JOKO WIDODO DI MEDIA SOSIAL X,” Karawang, Dec. 2025. [Online]. Available: <http://ejournal.stmik-time.ac.id>