

# Analisis Sentimen Ulasan Game Roblox Google Play Store Metode Lexicon-Based dan Random Forest

Ahmad Rifqi Maulana<sup>1</sup>, Hadwitya Handayani Kusumawardhani<sup>2</sup>, Ahmad Khambali<sup>3</sup>

1. Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Muhammadiyah Pekajangan Pekalongan, Indonesia
2. Manajemen Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Muhammadiyah Pekajangan Pekalongan, Indonesia
3. Manajemen Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Muhammadiyah Pekajangan Pekalongan, Indonesia

## Article Info

**Kata Kunci:** Analisis sentimen, data ulasan aplikasi Roblox, *Lexicon-Based*, *preprocessing*, *Random Forrest*

**Keywords:** *Sentiment analysis, Roblox app review data, Lexicon-Based, preprocessing, Random Forrest*

## Article history:

Received 17 August 2018

Revised 15 February 2019

Accepted 4 April 2019

Available online 4 April 2019

## DOI :

<https://doi.org/10.29100/jipi.v4i1.781>

\* Corresponding author.

Corresponding Author

E-mail address:

[rifqia393@gmail.com](mailto:rifqia393@gmail.com) had-

[witya.hk@gmail.com](mailto:witya.hk@gmail.com)

[ahmadkhambali68@gmail.com](mailto:ahmadkhambali68@gmail.com)

## ABSTRAK

Roblox adalah *platform online* yang menyediakan pengguna untuk bermain game, selain itu pengguna juga dapat membuat game sendiri dengan berbagai genre yang sangat beragam seperti petualangan, simulasi, *role-playing* dan komperatif. Aplikasi roblox kepleranya terus meningkat hingga pada tahun 2022 mencapai 54 juta pengguna aktif, Dengan terus meningkatnya Roblox hingga sekarang menyebabkan banyaknya ulasan dan komentar pengguna yang diupload didalam *Google Play Store* sehingga perlunya dilakukan analisis sentimen untuk mengelola data tersebut agar lebih efisien. Penelitian ini berhasil melakukan analisis sentimen terhadap ulasan pengguna aplikasi Roblox yang diperoleh dari *Google Play Store* melalui *web scraping*, dengan total 4.536 data ulasan setelah penghapusan data duplikat. Proses *preprocessing* dan pelabelan sentimen menggunakan metode *lexicon* mampu menghasilkan data yang siap dianalisis. Algoritma *Random Forest* menunjukkan kinerja yang baik dengan nilai akurasi, *precision*, *recall*, dan *f1-score* masing-masing sebesar 84%. Hasil analisis sentimen menunjukkan bahwa sebagian besar ulasan pengguna termasuk dalam kategori sentimen positif, yang mengindikasikan tingkat kepuasan pengguna yang relatif tinggi terhadap fitur dan pengalaman penggunaan aplikasi Roblox.

## ABSTRACT

*Roblox is an online platform that provides users to play games, in addition users can also create their own games with a variety of very diverse genres such as adventure, simulation, role-playing and comparative. The popularity of the Roblox application continues to increase until in 2022 it reached 54 million active users. With the continued increase of Roblox until now, it has caused a large number of user reviews and comments to be uploaded in the Google Play Store so that the need for sentiment analysis to manage the data more efficiently. This study successfully conducted sentiment analysis on Roblox application user reviews obtained from the Google Play Store through web scraping, with a total of 4,536 review data after removing duplicate data. The preprocessing and sentiment labeling process using the lexicon method was able to produce data that was ready for analysis. The Random Forest algorithm showed good performance with accuracy, precision, recall, and f1-score values of 84% each. The sentiment analysis results show that the majority of user reviews fall into*

*the positive sentiment category, indicating a relatively high level of user satisfaction with the features and user experience of the Roblox app.*

1. PENDAHULUAN

Perkembangan teknologi informasi telah membawa perubahan besar dalam cara manusia berinteraksi, baik di dunia nyata maupun digital. Salah satu wujud perkembangan tersebut adalah hadirnya *platform* permainan daring yang tidak hanya berfungsi sebagai hiburan, tetapi juga sebagai media bersosialisasi dan menyalurkan kreativitas, salah satunya Roblox [1]. Roblox merupakan *platform online* yang memungkinkan pengguna untuk bermain sekaligus membuat game dengan berbagai genre, serta berinteraksi dalam komunitas sosial virtual melalui fitur chat, avatar, dan transaksi ekonomi digital menggunakan mata uang *virtual Robux* [2].

Popularitas Roblox terus meningkat, dengan lebih dari 54 juta pengguna aktif harian pada tahun 2022 [3]. Peningkatan jumlah pengguna ini menghasilkan banyak ulasan di *Google Play Store* yang mencerminkan beragam pengalaman dan opini pengguna, baik dari aspek teknis maupun emosional. Namun, banyaknya data ulasan dalam bentuk teks tidak terstruktur memerlukan metode analisis yang efektif untuk mengolah dan memahami informasi tersebut.

Analisis sentimen merupakan metode yang umum digunakan untuk mengidentifikasi opini pengguna ke dalam kategori positif, negatif, atau netral berdasarkan data teks [4]. Dalam penelitian ini digunakan dua pendekatan analisis sentimen, yaitu metode *Lexicon-Based* dan algoritma *Random Forest*. Metode *Lexicon-Based* memanfaatkan kamus kata untuk mengklasifikasikan sentimen berdasarkan polaritas kata dalam kalimat [5], sedangkan *Random Forest* merupakan algoritma klasifikasi berbasis *ensemble decision tree* yang mampu mengenali pola dan konteks data secara lebih kompleks dan akurat [6].

Penelitian ini membahas proses analisis data ulasan pengguna Roblox yang berjumlah besar dan tidak terstruktur agar dapat diolah secara efektif dan menghasilkan informasi yang akurat. Selain itu, penelitian ini menganalisis hasil pelabelan sentimen ulasan pengguna menggunakan kamus *Lexicon Inset* bahasa Indonesia. Selanjutnya, penelitian ini mengevaluasi kinerja algoritma *Random Forest* dalam mengklasifikasikan sentimen berdasarkan nilai akurasi, *precision*, *recall*, dan *F1-score*.

Penelitian ini bertujuan untuk menganalisis sentimen pengguna terhadap game Roblox di *Google Play Store* serta menguji efektivitas metode *Lexicon-Based* dan *Random Forest* dalam mengklasifikasikan opini pengguna. Hasil penelitian diharapkan dapat memberikan gambaran menyeluruh mengenai persepsi pengguna serta menjadi bahan evaluasi bagi pengembangan dalam meningkatkan kualitas dan pengalaman bermain Roblox

2. METODE PENELITIAN

2.1 Jenis Penelitian

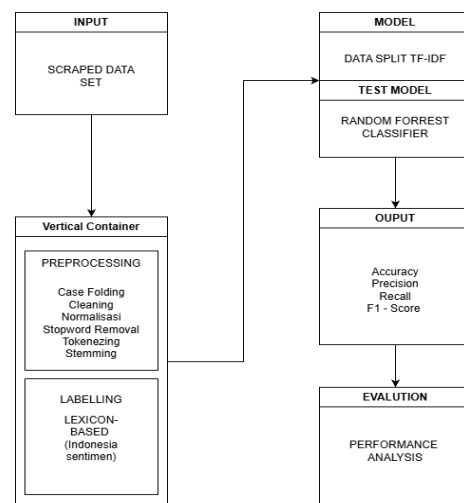
Penelitian kali ini merupakan penelitian kuantitatif dengan pendekatan *text mining*. Penulis memilih jenis penelitian ini karena seluruh proses analisis dilakukan dengan mengolah data berbentuk angka yang merupakan hasil konversi dari teks ulasan pengguna. Pendekatan *text mining* digunakan untuk mengolah data berupa teks tidak terstruktur menjadi informasi yang terstruktur sehingga dapat dianalisis. Melalui pendekatan ini, pendapat pengguna dapat dilihat polanya dengan lebih jelas, objektif, dan teratur.

2.2 Sumber data dan Jenis data

Penelitian ini menggunakan data sekunder yang bersumber dari ulasan pengguna aplikasi roblox di *google play store* yang diperoleh menggunakan metode *web scrapping*.

2.3 Alur Penelitian

Dalam penelitian ini, penulis mengimplementasikan penggabungan metode *Lexicon-Based* dan algoritma *Random Forest* untuk mengolah serta menganalisis ulasan pengguna aplikasi game Roblox. Alur kerja penelitian ini digambarkan pada *Gambar 2.1 Flowchart Alur Penelitian*, yang menunjukkan tahapan penelitian secara keseluruhan mulai dari proses pengumpulan data hingga evaluasi kinerja model analisis sentimen.



Gambar 2. 1 Flowchart Alur Penelitian

Tahap input merupakan proses pengumpulan data ulasan pengguna. Pada penelitian ini, pengambilan data dilakukan dengan memanfaatkan library *Google*

*Play Scraper* untuk memperoleh informasi ulasan dari *Google Play Store* dengan menggunakan *ID* aplikasi *Roblox*, yaitu *com.roblox.client*. Melalui proses *scraping* ini, diperoleh sebanyak 5.000 entri data ulasan. *Dataset* yang telah terkumpul kemudian disimpan dalam format *CSV* agar lebih mudah diproses pada tahap selanjutnya, khususnya pada tahap normalisasi sebelum dilakukan analisis lebih lanjut.

Tahap normalisasi dilakukan untuk menyesuaikan data agar sesuai dengan kebutuhan algoritma *Random Forest* yang digunakan dalam penelitian ini. Tahap ini terdiri dari dua proses utama, yaitu *preprocessing* dan pelabelan sentimen. *Preprocessing* merupakan tahap pembersihan *dataset* agar data yang digunakan lebih terstruktur dan relevan untuk proses pemodelan (Larasati et al., 2022). Pada tahap ini, data ulasan diproses melalui beberapa langkah, yaitu *case folding*, *cleaning*, *normalization*, *stopword removal*, *tokenize*, dan *stemming* [7].

*Case folding* dilakukan dengan mengubah seluruh teks ulasan menjadi huruf kecil untuk menyeragamkan format penulisan. *Cleaning* bertujuan menghilangkan atribut yang tidak berpengaruh terhadap proses klasifikasi, seperti tanda baca, angka, dan simbol. *Normalization* dilakukan untuk memperbaiki kata-kata yang tidak baku agar memiliki makna yang sama. Selanjutnya, *stopword removal* digunakan untuk menghapus kata-kata umum yang tidak memiliki kontribusi signifikan terhadap analisis sentimen. Tahap *tokenize* memecah teks ulasan menjadi unit kata atau token, sedangkan *stemming* bertujuan menghilangkan imbuhan pada kata sehingga diperoleh kata dasar.

Setelah melewati tahap *preprocessing*, data kemudian masuk ke tahap pelabelan sentimen menggunakan pendekatan *Lexicon-Based*. Pendekatan ini merupakan proses penentuan sentimen berdasarkan pencocokan kata-kata dalam ulasan dengan kamus sentimen [5]. Kamus *lexicon* yang digunakan mengandung kategori sentimen positif, negatif, dan netral [8]. Dalam penelitian ini, digunakan kamus *Lexicon Inset* Bahasa Indonesia untuk mengevaluasi polaritas setiap kata dalam kalimat ulasan. Setiap kata dicocokkan dengan daftar kata pada kamus *lexicon* dan diberikan skor sentimen. Skor tersebut kemudian dijumlahkan untuk menentukan sentimen kalimat secara keseluruhan, di mana skor lebih dari nol dikategorikan sebagai sentimen positif, skor sama dengan nol sebagai sentimen netral, dan skor kurang dari nol sebagai sentimen negatif. Kriteria pelabelan sentimen ini ditunjukkan pada *Tabel 2.1 Skor Sentimen*.

Tabel 2.1 Skor Sentimen

Nilai Sentimen	Label
Score > 0	Positif
Score = 0	Netral
Score < 0	Negatif

Setelah proses pelabelan selesai, data masuk ke tahap pemodelan dan pengujian model. Pada tahap ini, *dataset* dibagi menjadi dua bagian, yaitu data latih dan

data uji. Pembagian data dilakukan dengan rasio 80:20, di mana 80% data digunakan sebagai data latih untuk melatih model agar mampu mempelajari pola data, sedangkan 20% data digunakan sebagai data uji untuk mengevaluasi kemampuan model dalam melakukan generalisasi terhadap data yang belum pernah dilihat sebelumnya. Pembagian ini bertujuan untuk memperoleh performa model yang lebih objektif serta meminimalkan terjadinya *overfitting*.

Sebelum dilakukan pelatihan model, data teks diubah ke dalam bentuk numerik menggunakan metode *TF-IDF* (*Term Frequency-Inverse Document Frequency*). Metode ini berfungsi untuk memberikan bobot numerik pada setiap kata berdasarkan frekuensi kemunculannya dalam dokumen dan tingkat keunikannya di seluruh dokumen.

Rumus IDF:

$$idf_t = \log\left(\frac{N}{df_t}\right)$$

Keterangan :

$idf_t$  = Nilai *IDF* dari term  $t$

$N$  = Jumlah dokumen yang digunakan

$df_t$  = Jumlah kemunculan term  $t$  dalam dokumen  $d$

Pembobotan *TF-IDF* akan dihitung melalui persamaan berikut :

$$tf-idf_{td} = tf_{td} \times idf_t$$

Keterangan :

$tf - idf_{td}$  = Bobot *TF-IDF* term  $t$  dalam dokumen  $d$

$tf_{td}$  = Frekuensi muncul term  $t$  dalam dokumen  $d$

$idf_t$  = Nilai *IDF* dalam trem  $t$

Setelah proses pembobotan selesai, data kemudian dilatih menggunakan algoritma *Random Forest*.

Proses *Random Forest* diawali dengan pembagian *dataset* menjadi data latih dan data uji menggunakan variasi rasio tertentu, seperti 10:90 hingga 90:10 [9]. Data latih kemudian diproses menggunakan teknik *bootstrapping* untuk menghasilkan beberapa *subset* data yang akan digunakan dalam pelatihan masing-masing *decision tree*.

Pada setiap *decision tree*, pemilihan atribut dilakukan dengan menghitung nilai *entropy* dan *information gain* untuk menentukan atribut terbaik sebagai *node* pemisah [10]. Rumus menghitung nilai *entropy* adalah sebagai berikut :

$$Entropy(Y) = -\sum p(c|Y) \log^2 p(c|Y)$$

Keterangan:

Y = Himpunan Kasus

$p(c|Y)$  = Proporsi nilai Y terhadap kelas c

Rumus *information gain* dapat dilihat pada persamaan berikut:

$$Information\ gain(Y, a) = Entropy(Y) - \sum_{v \in Values(a)} \frac{|Y_v|}{|Y_a|} Entropy(Y_v)$$

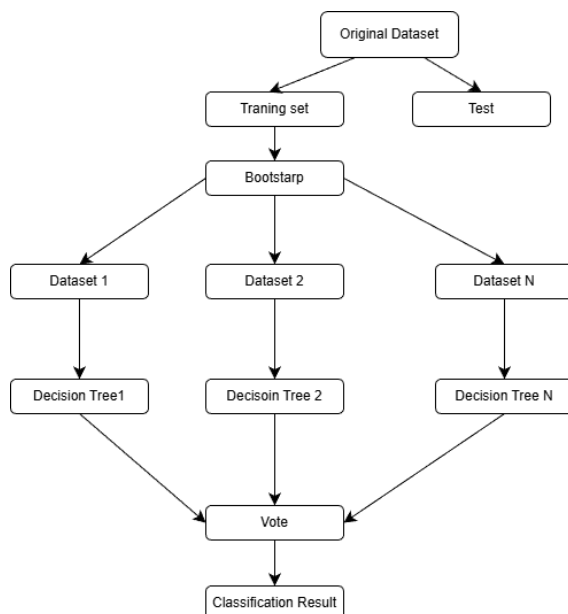
Keterangan:

$Values(a)$  = Nilai yang mungkin dalam himpunan kasus a

$Y_v$  = Subkelas dari kelas Y dengan kelas v yang berhubungan dengan kelas a

$Y_a$  = Semua nilai yang sesuai dengan a

Alur kerja algoritma Random Forest ditunjukkan pada Gambar 2.2 Alur Algoritma Random Forest.



Gambar 2. 1 Alur Algoritma Random Forest

Pada algoritma *Random Forest*, dataset dibagi menjadi data latih dan data uji. Data latih diproses menggunakan metode *bootstrapping*, yaitu teknik pengambilan sampel secara acak dari data original untuk membentuk beberapa dataset baru. *Dataset hasil bootstrapping* tersebut digunakan untuk melatih sejumlah *decision tree*. Setiap *decision tree* menghasilkan prediksi masing-masing, kemudian seluruh hasil prediksi digabungkan menggunakan mekanisme *voting* untuk menentukan hasil klasifikasi akhir yang paling optimal.

Tahap output merupakan tahap untuk melihat hasil kinerja model yang telah dibangun. Pada tahap ini, kemampuan model diukur menggunakan beberapa metrik evaluasi, yaitu akurasi, presisi, *recall*, dan *F1-score*. Metrik-metrik ini memberikan gambaran mengenai tingkat keefektifan model dalam mengklasifikasikan sentimen ulasan pengguna.

Tahap terakhir adalah evaluasi model secara keseluruhan. Evaluasi dilakukan dengan menganalisis hasil klasifikasi pada setiap kategori sentimen menggunakan *confusion matrix* serta metrik evaluasi seperti akurasi, presisi, *recall*, dan *F1-score*. *Confusion matrix* digunakan untuk melihat kemampuan model dalam membedakan sentimen positif, negatif, dan netral secara lebih detail. Melalui evaluasi ini, penulis dapat mengetahui kelebihan dan kekurangan model serta memperoleh gambaran yang lebih jelas mengenai kemampuan algoritma *Random Forest* dalam melakukan klasifikasi sentimen ulasan pengguna aplikasi Roblox.

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Hasil

Tahap *input* diawali dengan pengumpulan data ulasan pengguna aplikasi Roblox menggunakan metode *web scraping*. Proses pengambilan data dilakukan dengan memanfaatkan *library Google Play Scraper* berbasis bahasa pemrograman *Python* yang dijalankan melalui *Google Colab* dengan *ID* aplikasi *com.roblox.client*. Data yang berhasil dikumpulkan berupa 5.000 ulasan terbaru dari *Google Play Store* yang mencakup informasi tanggal ulasan, nama pengguna, isi ulasan, dan rating.

Tabel 3. 1 Dataframe roblox

No	Tanggal	Nama pengguna	Ulasan pengguna	Rating
1	2026-02-02	Pengguna Google	Jaringan bagus, tapi ktanya g ada koneksi,jdi...	1
2	2026-02-02	Pengguna Google	game nya bagus sekali	1
3	2026-02-02	Pengguna Google	udah ga bisa chat	1
4	2026-02-02	Pengguna Google	yang bener lu buat aplikasi umurnya gua udah p	1

Tahap selanjutnya adalah normalisasi data yang terdiri dari *preprocessing* dan pelabelan sentimen. *Preprocessing* merupakan tahap awal pengolahan dataset sebelum dilakukan analisis sentimen, yang bertujuan untuk membersihkan dan menyiapkan data agar lebih terstruktur serta relevan

untuk dianalisis. Proses *preprocessing* meliputi *case folding*, *cleaning*, *normalization*, *tokenize*, *stopword removal*, dan *stemming*.

Pada tahap *case folding*, seluruh teks ulasan diubah menjadi huruf kecil untuk menghilangkan perbedaan penulisan antara huruf kapital dan huruf kecil yang memiliki makna sama. Contoh hasil *case folding* ditampilkan pada Tabel 3.2.

Tabel 3. 2 Hasil Case Folding

No	Sebelum	Case Folding
0	Jaringan bagus, tpi ktanya g ada koneksi, jdi ...	jaringan bagus, tpi ktanya g ada koneksi, jdi ...
1	game nya bagus sekali	game nya bagus sekali
2	udah ga bisa chat	udah ga bisa chat
3	yang bener lu buat aplikasi umurnya gua udah p...	yang bener lu buat aplikasi umurnya gua udah p...
4	kalo udah force close selalu rendernya jadi bl...	kalo udah force close selalu rendernya jadi bl...

Selanjutnya dilakukan *cleaning* untuk menghapus tanda baca, angka, *URL*, emoji, simbol, dan spasi berlebih sehingga teks menjadi lebih bersih dan konsisten. Hasil dari tahap *cleaning* dapat dilihat pada Tabel 3.3.

Tabel 3. 3 Hasil Cleaning

No	Sebelum	Cleaning
0	Jaringan bagus, tpi ktanya g ada koneksi, jdi ...	jaringan bagus tpi ktanya g ada koneksi jdi gb...
1	game nya bagus sekali	game nya bagus sekali
2	udah ga bisa chat	udah ga bisa chat
3	yang bener lu buat aplikasi umurnya gua udah p...	yang bener lu buat aplikasi umurnya gua udah p...
4	kalo udah force close selalu rendernya jadi bl...	kalo udah force close selalu rendernya jadi bl...

Tahap *normalization* bertujuan mengubah kata tidak baku, singkatan, dan kesalahan penulisan menjadi kata baku dengan bantuan kamus *slang* bahasa Indonesia. Contoh hasil *normalization* ditampilkan pada Tabel 3.4.

Tabel 3. 4 Hasil Normalization

No	Sebelum	Normalization
0	jaringan bagus tpi ktanya g ada koneksi jdi gb.	jaringan bagus tetapi ktanya tidak ada koneksi.
1	game nya bagus sekali	game nya bagus sekali
2	udah ga bisa chat	sudah tidak bisa cha
3	yang bener lu buat aplikasi umurnya gua udah p..	yang benar lu buat aplikasi umurnya saya sudah...
4	kalo udah force close selalu rendernya jadi bl...	kalau sudah force close selalu rendernya jadi ...

Setelah itu dilakukan *tokenize*, yaitu proses pemecahan kalimat menjadi unit kata untuk memudahkan analisis lanjutan. Hasil *tokenize* ditampilkan pada Tabel 3.5.

Tabel 3. 5 Hasil Tokenize

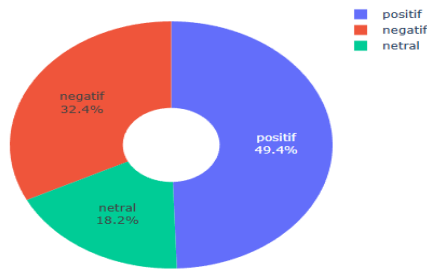
No	Sebelum	Tokenize
0	jaringan bagus tetapi ktanya tidak koneksi ja...	jaringan, bagus, tetapi, ktanya, tidak, koneksi...
1	game nya bagus sekali	game, bagus, sekali
2	sudah tidak bisa chat	sudah, tidak, bisa, chat
3	yang bener lu buat aplikasi umurnya saya sudah..	yang, benar, buat, aplikasi, umurnya, saya, sudah
4	kalau sudah force close selalu rendernya jadi	kalau, sudah, force, close, selalu, rendernya, jadi

Tahap *stopword removal* dilakukan untuk menghilangkan kata-kata umum yang tidak memiliki kontribusi signifikan terhadap analisis sentimen, seperti kata penghubung dan kata depan. Contoh hasil *stopword removal* disajikan pada Tabel 3.6.

Tabel 3. 6 Hasil Stopword Removal

No	Token	Stopword Removal
0	jaringan, bagus, tetapi, ktanya, tidak, konek...	jaringan, bagus, ktanya, koneksi, gbisa, main
1	game, bagus, sekali	game, bagus
2	sudah, tidak, bisa, chat	chat
3	yang, benar, buat, aplikasi, umurnya, saya, s...	aplikasi, umurnya, periksa, umur, dongo, bene...
4	kalau, sudah, force, close, selalu, rendernya...	force, close, rendernya, blur, jangkawannya





Gambar 3. 6 Pie Chart

Tahap pemodelan dilakukan dengan membagi dataset menjadi data latih dan data uji menggunakan rasio 80:20. Hasil pembagian data ditampilkan pada Gambar 3.8, dengan 3.552 data digunakan sebagai data latih dan 889 data sebagai data uji.

Ukuran X\_train: (3552, 3000)  
 Ukuran X\_test: (889, 3000)  
 Ukuran y\_train: (3552,)  
 Ukuran y\_test: (889,)

Gambar 3. 8 Hasil Data Splitting

Selanjutnya dilakukan pembobotan kata menggunakan metode *TF-IDF* untuk mengubah data teks menjadi bentuk numerik agar dapat diproses oleh algoritma *Random Forest*.

Pengujian model dilakukan menggunakan algoritma *Random Forest* dengan menggabungkan beberapa *decision tree* yang dilatih pada data latih. Hasil klasifikasi model ditampilkan pada Gambar 3.9, yang menunjukkan nilai akurasi sebesar 84%. Nilai ini menunjukkan bahwa model memiliki tingkat ketepatan yang cukup baik dalam mengklasifikasikan sentimen ulasan pengguna.

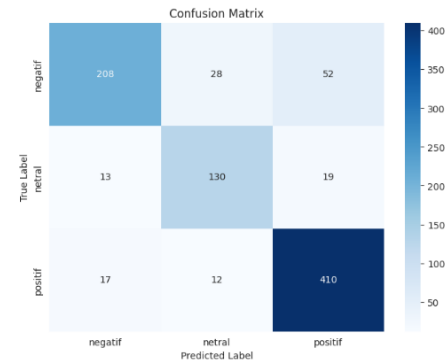
Model RandomForestClassifier berhasil dilatih.

Classification Report:				
	precision	recall	f1-score	support
negatif	0.87	0.72	0.79	288
netral	0.76	0.80	0.78	162
positif	0.85	0.93	0.89	439
accuracy			0.84	889
macro avg	0.83	0.82	0.82	889
weighted avg	0.84	0.84	0.84	889

Accuracy Score: 0.8414

Gambar 3. 9 Hasil Klasifikasi Random Forrest

Evaluasi model dilakukan menggunakan *confusion matrix* dan metrik evaluasi berupa *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil *confusion matrix* ditampilkan pada Gambar 3.10, sedangkan nilai metrik evaluasi disajikan pada Tabel 3.8. Berdasarkan hasil tersebut, diperoleh nilai *accuracy*, *precision*, *recall*, dan *F1-score* masing-masing sebesar 84%, yang menunjukkan performa model yang stabil dan konsisten.



Gambar 3.7 Hasil Confusion Matrix

Tabel 3.8 Hasil Confusion Matrix

Accuracy	Precision	Recall	F1-Score
84%	84 %	84%	84%

### 3.2 PEMBAHASAN

Data ulasan pengguna aplikasi Roblox yang diperoleh dari *Google Play Store* melalui teknik *scraping* berjumlah 5.000 data dan bersifat tidak terstruktur. Setelah dilakukan penghapusan 464 data duplikat, diperoleh 4.536 ulasan *valid* yang kemudian melalui tahap *preprocessing* meliputi *case folding*, *cleaning*, *tokenize*, *stopword removal*, *normalisasi*, dan *stemming* untuk mengurangi *noise* dan menghasilkan data yang lebih terstruktur. Pelabelan sentimen dilakukan menggunakan metode *lexicon-based* dengan kamus *Lexicon Inset Bahasa Indonesia*. Hasilnya, 2.196 ulasan dikategorikan sebagai positif, 1.437 negatif, dan 808 netral, dengan sentimen positif mendominasi.

Proses klasifikasi menggunakan algoritma *Random Forest* menunjukkan hasil evaluasi yang baik dengan nilai *accuracy*, *precision*, *recall*, dan *F1-score* masing-masing sebesar 84%. Hal ini menandakan bahwa model mampu mengklasifikasikan sentimen ulasan secara akurat dan stabil.

### 4. KESIMPULAN

Berdasarkan hasil penelitian analisis sentimen ulasan pengguna aplikasi Roblox menggunakan metode *Lexicon-Based* dan algoritma *Random Forest*, diperoleh beberapa kesimpulan. Dari 5.000 data ulasan yang dikumpulkan melalui *web scraping*, setelah penghapusan 464 data duplikat diperoleh 4.536 data *valid* yang kemudian melalui tahap *preprocessing* untuk meningkatkan kualitas data.

Pelabelan sentimen menggunakan metode *lexicon* berhasil mengklasifikasikan ulasan ke dalam tiga kategori, yaitu 2.196 positif, 1.437 negatif, dan 808 netral, dengan sentimen positif sebagai kategori terbanyak.

Model *Random Forest* menunjukkan performa yang baik dan stabil dengan nilai akurasi, *precision*, *recall*, dan *F1-score* masing-masing sebesar 84%. Secara keseluruhan, hasil penelitian menunjukkan bahwa sebagian besar pengguna memberikan penilaian positif terhadap aplikasi Roblox, yang mencerminkan tingkat kepuasan pengguna yang relatif tinggi.

Berdasarkan hasil penelitian, beberapa rekomendasi dapat dipertimbangkan untuk pengembangan selanjutnya. Proses *preprocessing* dapat ditingkatkan dengan teknik penanganan kata tidak baku, singkatan, dan bahasa gaul agar kualitas data lebih optimal. Evaluasi model juga dapat diperkuat dengan menerapkan *cross validation* untuk menghasilkan pengujian yang lebih stabil dan akurat. Selain itu, hasil analisis sentimen diharapkan dapat dimanfaatkan oleh pengembang aplikasi Roblox sebagai bahan evaluasi dalam meningkatkan kualitas, fitur, dan pengalaman pengguna.

## REFERENSI

- [1] R. R. Suryono, "Sentiment Classification of Indonesian-Language Roblox Reviews Using IndoBERT with SMOTE Optimization," *J. Appl. Informatics Comput.*, vol. 9, no. 4, pp. 1868–1877, 2025.
- [2] U. S. Yahsy and M. Syas, "Komodifikasi Users pada Platform Game Online Roblox," *J. Inter Act*, vol. 2, p. 12, 2022, [Online]. Available: <https://ejournal.atmajaya.ac.id/index.php/interact/article/download/3748/1815/14616>
- [3] N. Kadek, F. Puspita, I. G. I. Sudipa, I. W. Sunarya, N. Wayan, and J. 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, doi: <https://doi.org/10.33395/sinkron.v9i4.15272> e-ISSN: F. A. Larasati, D. E. Ratnawati, and B. T. Hanggara, "Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 9, pp. 4305–4313, 2022, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/11562>
- [5] S. A. Putra and A. Wijaya, "Analisis Sentimen Artificial Intelligence ( AI ) Pada Media Sosial Twitter Menggunakan Metode Lexicon Based," *Anal. Sentimen Artif. Intell. ( AI ) Pada Media Sos. Twitter Menggunakan Metod. Lex. Based*, vol. 7, pp. 21–28, 2023.
- [6] C. G. Indrayanto, D. E. Ratnawati, and B. Rahayudi, "Analisis Sentimen Data Ulasan Pengguna Aplikasi MyPertamina di Indonesia pada Google Play Store menggunakan Metode Random Forest," *J. Pengemb. Teknol. Informasi dan Ilmu Komput.*, vol. 7, no. 3, pp. 1131–1139, 2023, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [7] F. Y. M. Nadhifah, Aini Nur, Kusumawardhani Hadwitya Handayani, "Analisis Sentiment Ulasan Aplikasi Gopay Pada Google Play Store Menggunakan Algoritma Support Vector Machine," *J. Surya Inform.*, vol. 14, no. 1, pp. 1–6, 2024, doi: <https://doi.org/10.48144/suryainformatika.v14i1.1787>.
- [8] O. Manullang, C. Prianto, and N. H. Harani, "Analisis Sentimen Untuk Memprediksi Hasil Calon Pemilu Presiden Menggunakan Lexicon Based dan Random Forest," *J. Ilm. Inform.*, vol. 11, no. 54, p. 11, 2023, [Online]. Available: <https://ejournal.upbatam.ac.id/index.php/jif/article/download/7987/3319>
- [9] A. Syah, "Analisis sentimen aplikasi shopee, tokopedia, lazada dan blibli menggunakan leksikon dan random forest," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, p. 12, 2024, doi: <http://dx.doi.org/10.23960/jitet.v12i3S1.5155>.
- [10] M. Reza, U. Pulungan, D. E. Ratnawati, and B. Rahayudi, "Analisis Sentimen Ulasan Aplikasi PeduliLindungi dengan Metode Random," *Anal. Sentimen Ulas. Apl. PeduliLindungi dengan Metod. Random*, vol. 6, no. 9, pp. 4378–4385, 2022.