

KLASIFIKASI SENTIMEN ULASAN PENGGUNA APLIKASI ROBLOX DI GOOGLE PLAY STORE MENGGUNAKAN MACHINE LEARNING SUPERVISED

Giri Marullana¹, Nining Rahaningsih², Irfan Ali³, Ade Rizki Rinaldi⁴.

Program Studi Teknik Informatika¹
Program Studi Komputerisasi Akuntansi²
Program Studi Rekayasa Perangkat Lunak^{3,4}

STMIK IKMI Cirebon
<https://ikmi.ac.id/page/18/?lang=de>
marullanagiri@gmail.com

(*) Corresponding Author : marullanagiri@gmail.com
Published : 30 Juni 2026

Abstract—The rapid development of digital technology and the increasing use of mobile applications have made evaluating application quality through user reviews on platforms such as the Google Play Store increasingly important. These reviews contain users' opinions, experiences, and complaints; however, their large volume and the use of informal language make manual analysis inefficient and prone to bias. Therefore, an automated analysis system is needed to process textual data quickly and accurately. This study focuses on the automatic sentiment classification of user reviews for the Roblox application and aims to determine the most effective machine learning algorithm. Roblox, as a popular online gaming platform, generates diverse and unstructured reviews that often include slang, requiring proper preprocessing. The dataset was collected through web scraping using the google-play-scraper library in Python, resulting in approximately 100,000 Indonesian-language reviews, which were then reduced to 9,897 data points after preprocessing. This research employs a hybrid approach by combining a lexicon-based method using the InSet sentiment dictionary for data labeling with supervised machine learning for classification. Feature representation is carried out using TF-IDF, and the dataset is split into 80% training data and 20% testing data. Three algorithms are applied: Naïve Bayes, Support Vector Machine (SVM), and Random Forest. The evaluation results show that SVM achieves the best performance with an accuracy of 0.89, followed by Random Forest at 0.80 and Naïve Bayes at 0.62. In addition, a User Acceptance Test was conducted on the developed web-based application, resulting in an average satisfaction score of 4.35 (categorized as "Very Satisfied").

Keywords: Sentiment Analysis, Roblox, Google Play Store, Naïve Bayes, Support Vector Machine, Random Forest, TF-IDF, Supervised Machine Learning

Abstrak—Perkembangan teknologi digital dan meningkatnya penggunaan aplikasi mobile menjadikan evaluasi kualitas aplikasi melalui ulasan pengguna di platform seperti Google Play Store semakin penting. Ulasan tersebut mengandung opini, pengalaman, dan keluhan pengguna, namun jumlahnya yang besar serta penggunaan bahasa tidak baku membuat analisis manual menjadi tidak efisien dan rentan bias. Oleh karena itu, diperlukan sistem analisis otomatis yang mampu mengolah data teks secara cepat dan akurat. Penelitian ini berfokus pada klasifikasi sentimen ulasan pengguna aplikasi Roblox secara otomatis serta menentukan algoritma machine learning yang paling efektif. Roblox sebagai platform permainan daring populer menghasilkan ulasan yang beragam, tidak terstruktur, dan sering mengandung slang, sehingga membutuhkan preprocessing yang tepat. Dataset diperoleh melalui proses web scraping menggunakan pustaka google-play-scraper pada Python, menghasilkan sekitar 100.000 ulasan berbahasa Indonesia, yang kemudian diseleksi menjadi 9.897 data setelah preprocessing. Penelitian ini menggunakan pendekatan hybrid dengan menggabungkan metode lexicon-based menggunakan kamus sentimen InSet untuk pelabelan data, serta machine learning supervised untuk klasifikasi. Representasi fitur dilakukan menggunakan TF-IDF, dan data dibagi menjadi 80% data latih dan 20% data uji. Tiga algoritma yang digunakan adalah Naïve Bayes, Support Vector Machine (SVM), dan Random Forest. Hasil evaluasi menunjukkan bahwa SVM memiliki performa terbaik dengan akurasi 0,89, diikuti Random Forest sebesar

0,80 dan Naïve Bayes sebesar 0,62. Selain itu, dilakukan User Acceptance Test terhadap aplikasi berbasis web yang dikembangkan, dengan hasil rata-rata kepuasan sebesar 4,35 (kategori sangat puas).

Kata Kunci : Analisis Sentimen, Roblox, Google Play Store, Naïve Bayes, Support Vector Machine, Random Forest, TF-IDF, Machine learning Supervised

INTRODUCTION

Analisis terhadap ulasan semacam ini memerlukan pendekatan analisis sentimen, yaitu proses mengekstraksi opini, sikap, dan emosi dari teks pengguna untuk menentukan apakah ulasan bersifat positif, negatif, atau netral. Dalam konteks aplikasi digital, analisis sentimen berperan penting dalam pengambilan keputusan berbasis data, misalnya untuk memperbaiki fitur yang kurang disukai atau meningkatkan kepuasan pengguna. Penelitian oleh [1] menunjukkan bahwa analisis sentimen berbasis *Natural Language Processing (NLP)* dapat mengungkap pola persepsi publik secara makro, misalnya dalam konteks transportasi selama pandemi COVID-19, di mana mayoritas pengguna menunjukkan sentimen positif terhadap moda transportasi pribadi dan sepeda. Temuan ini menegaskan bahwa pendekatan NLP dan analisis sentimen dapat digunakan untuk memahami perilaku masyarakat terhadap sistem atau layanan tertentu, termasuk aplikasi digital seperti Roblox.

Dalam ranah metode, [2] melalui tinjauan sistematis menunjukkan bahwa pendekatan *lexicon-based* sering kali digunakan sebagai *baseline* karena kesederhanaan dan interpretabilitasnya, tetapi cenderung kalah akurat dibandingkan algoritma *machine learning supervised* seperti *Support Vector Machine (SVM)*, *Naïve Bayes*, dan *Random Forest*, terutama ketika tersedia *dataset* besar dan berlabel. Studi [3] memperkuat pandangan tersebut dengan menunjukkan bahwa pendekatan berbasis pembelajaran mesin secara konsisten memberikan hasil klasifikasi yang lebih baik pada data sosial, seperti Twitter, dibandingkan metode leksikal tradisional. Hal ini menunjukkan perlunya peralihan menuju pendekatan *supervised learning* dalam analisis sentimen berbasis teks pengguna.

Selain itu, [4] dalam tinjauan literturnya terhadap 81 studi analisis sentimen di bidang transportasi menyimpulkan bahwa model *machine learning supervised* seperti *Naïve Bayes*, *SVM*, dan *Random Forest* merupakan tiga algoritma paling umum dan efektif untuk

klasifikasi sentimen dalam bahasa alami. Penggunaan ketiga model ini menawarkan keseimbangan antara akurasi, efisiensi, dan interpretabilitas, menjadikannya pilihan yang tepat untuk penelitian berbasis data ulasan pengguna seperti pada aplikasi Roblox. Sementara itu, [5] menunjukkan bahwa penggabungan antara pendekatan leksikal dan pembelajaran mesin, seperti model *Lexicon-Enhanced BERT (Lebert)*, dapat meningkatkan performa klasifikasi, yang membuka peluang untuk penelitian lanjutan di masa depan.

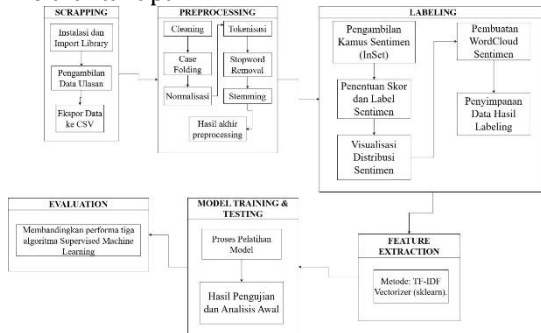
Dari perspektif sosial, penelitian [6] dan [7] menegaskan pentingnya analisis sentimen dalam memantau kesejahteraan atau persepsi masyarakat terhadap layanan digital, khususnya selama kondisi darurat seperti pandemi. Meskipun studi tersebut berfokus pada konteks transportasi publik, metodologi yang digunakan relevan untuk menganalisis data pengguna aplikasi digital lain, termasuk Roblox. Dalam konteks hiburan dan game daring, ulasan pengguna sering kali berisi emosi intens, sarkasme, atau kata tidak baku, yang menjadikannya tantangan tersendiri bagi model analisis sentimen. Oleh karena itu, penerapan algoritma pembelajaran mesin yang mampu mempelajari pola linguistik kompleks menjadi krusial untuk mencapai hasil klasifikasi yang akurat.

Melalui latar belakang tersebut, penelitian ini bertujuan untuk mengimplementasikan dan membandingkan kinerja tiga algoritma *machine learning supervised* — yaitu *Naïve Bayes*, *SVM*, dan *Random Forest* — dalam melakukan klasifikasi sentimen terhadap ulasan pengguna aplikasi Roblox yang diperoleh dari *Google Play Store*. Pendekatan ini diharapkan mampu menghasilkan model terbaik yang dapat memprediksi polaritas opini pengguna secara tepat dan memberikan insight berharga bagi pengembang dalam memahami persepsi publik terhadap platform mereka.

MATERIALS AND METHODS

Penelitian ini menggunakan pendekatan analisis kuantitatif eksperimental berbasis *machine learning supervised*, yang bertujuan untuk mengklasifikasikan sentimen pengguna aplikasi Roblox berdasarkan ulasan mereka di *Google Play*

Store. Analisis dilakukan secara sistematis melalui tahapan:



Gambar 1. Alur penelitian

Alur penelitian pada gambar tersebut dimulai dari tahap scrapping data, yaitu proses pengambilan data ulasan pengguna dari Google Play Store. Pada tahap ini dilakukan instalasi dan impor library yang dibutuhkan, kemudian pengambilan data ulasan secara otomatis, dan hasilnya disimpan dalam format CSV sebagai dataset awal. Selanjutnya, data yang telah diperoleh masuk ke tahap preprocessing, yang bertujuan untuk membersihkan dan menyiapkan data agar siap dianalisis. Proses ini meliputi cleaning (pembersihan data), case folding (mengubah huruf menjadi kecil), normalisasi kata, tokenisasi, penghapusan stopwords, serta stemming. Hasil dari tahap ini adalah data teks yang telah bersih dan terstruktur.

Tahap berikutnya adalah labeling, di mana data diberi label sentimen menggunakan kamus sentimen InSet. Proses ini mencakup pengambilan kamus, penentuan skor dan label sentimen (positif, negatif, atau netral), serta visualisasi distribusi sentimen. Selain itu, dilakukan juga pembuatan wordcloud untuk melihat kata-kata yang dominan, dan seluruh hasil pelabelan disimpan untuk digunakan pada tahap selanjutnya.

Setelah data memiliki label, dilakukan feature extraction menggunakan metode TF-IDF untuk mengubah data teks menjadi representasi numerik yang dapat diproses oleh algoritma machine learning. Kemudian masuk ke tahap model training dan testing, di mana model dilatih menggunakan data latih dan diuji menggunakan data uji untuk melihat performanya. Hasil pengujian kemudian dianalisis sebagai evaluasi awal.

Tahap terakhir adalah evaluation, yaitu membandingkan performa beberapa algoritma supervised machine learning yang digunakan. Pada tahap ini dilakukan analisis metrik evaluasi seperti akurasi dan lainnya untuk menentukan model terbaik dalam klasifikasi sentimen. Secara

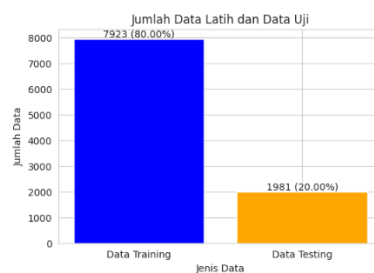
keseluruhan, alur penelitian ini menunjukkan proses terstruktur dari pengumpulan data hingga evaluasi model untuk menghasilkan sistem analisis sentimen yang optimal.

RESULTS AND DISCUSSION

Pada tahap ini dilakukan proses pelatihan (*training*) dan pengujian (*testing*) model *machine learning* menggunakan data ulasan pengguna Roblox yang telah melalui seluruh tahapan *preprocessing* dan *feature extraction*. Tujuan utama dari tahap ini adalah mengukur performa masing-masing algoritma dalam melakukan klasifikasi sentimen terhadap tiga label, yaitu positif, negatif, dan netral. Model yang digunakan pada penelitian ini terdiri dari tiga algoritma *supervised learning* yang umum digunakan dalam klasifikasi teks, yaitu *Naïve Bayes*, *Support Vector Machine*, dan *Random Forest*.

Setelah proses ekstraksi fitur menggunakan *TF-IDF*, tahap selanjutnya adalah pelatihan model (*training*) dan pengujian model (*testing*) untuk mengklasifikasikan sentimen ulasan pengguna aplikasi Roblox.

Pada penelitian ini, dilakukan pengaturan parameter (*hyperparameter*) pada masing-masing algoritma untuk memperoleh performa yang optimal. Pada algoritma *Support Vector Machine (SVM)*, digunakan *kernel linear* yang dipilih karena sesuai dengan karakteristik data teks berdimensi tinggi seperti *TF-IDF*. Pada algoritma *Random Forest*, digunakan jumlah pohon (*n_estimators*) sebanyak 100, yang secara empiris mampu memberikan keseimbangan antara akurasi dan waktu komputasi. Sedangkan pada *Naïve Bayes* digunakan model *MultinomialNB* yang umum digunakan untuk klasifikasi teks. Pemilihan parameter ini didasarkan pada eksperimen awal dan referensi dari penelitian terdahulu, sehingga meskipun tidak menggunakan *Grid Search* secara menyeluruh, konfigurasi parameter yang digunakan sudah cukup *representatif* untuk menghasilkan performa model yang optimal. Proses pelatihan dan pengujian dilakukan menggunakan pustaka *scikit-learn* dengan data latih sebesar 80 % dan data uji sebesar 20 % dari total korpus (7923 data latih dan 1981 data uji).



Gambar 2 Proses Pelatihan Model

Hasil Pengujian dan Analisis Awal

Ketiga model diuji terhadap *dataset* uji yang terdiri atas 1981 ulasan untuk menilai kinerja masing-masing algoritma dalam memprediksi tiga kelas sentimen — Positif, Netral, dan Negatif..

Visualisasi *Confusion matrix* untuk model (SVM), *Naïve Bayes* dan *Random Forest*.

Confusion matrix for SVM:

SVM Confusion Matrix

	Negatif	Netral	Positif
Actual Negatif	291	78	11
Actual Netral	34	625	70
Actual Positif	3	41	827
	Negatif	Netral	Positif

Predicted

Gambar 3 SVM Confusion Matrix

Confusion matrix untuk model *Support Vector Machine (SVM)* dalam mengklasifikasikan sentimen ulasan pengguna Roblox. Berdasarkan visualisasi tersebut, *SVM* mampu menghasilkan prediksi yang sangat akurat pada seluruh kelas sentimen. Model berhasil mengklasifikasikan 291 ulasan negatif, 625 ulasan netral, dan 827 ulasan positif secara benar. Kesalahan prediksi yang muncul relatif kecil, yaitu *misclassification* pada kelas negatif ke netral (78 data) serta kelas netral ke positif (70 data). Secara keseluruhan, distribusi prediksi pada setiap kelas menunjukkan bahwa *SVM* memiliki kemampuan generalisasi yang baik pada data teks yang telah melalui proses *preprocessing* dan *TF-IDF*. Hasil ini kemudian tercermin pada nilai akurasi akhir sebesar 0.92, menjadikan *SVM* sebagai model dengan performa terbaik di antara ketiga algoritma yang diuji. *Confusion matrix* for Naive Bayes:

Naive Bayes Confusion Matrix

	Negatif	Netral	Positif
Actual Negatif	96	134	150
Actual Netral	3	218	508
Actual Positif	1	28	842
	Negatif	Netral	Positif

Predicted

Gambar 4 *Confusion matrix* Naive Bayes

Confusion matrix untuk model *Naïve Bayes*, yang menunjukkan performa klasifikasi sentimen pada *dataset* yang sama. Secara umum, model berhasil mengidentifikasi kelas positif dengan cukup baik (842 prediksi benar), namun mengalami tingkat kesalahan yang lebih tinggi pada kelas negatif dan netral. Terlihat bahwa banyak ulasan negatif dan netral yang salah dipetakan ke kelas positif, masing-masing sebanyak 150 dan 508 data. Hal ini mengindikasikan bahwa *Naïve Bayes* kurang mampu menangkap distribusi kata yang mewakili sentimen negatif dan netral, terutama pada teks berbahasa Indonesia yang memiliki variasi kosakata tidak baku. Meski demikian, model masih menunjukkan performa yang cukup layak dengan akurasi sebesar 0.85, sehingga tetap dapat digunakan sebagai pembandingan untuk algoritma lain.

Confusion matrix for Random Forest:

Random Forest Confusion Matrix

	Negatif	Netral	Positif
Actual Negatif	220	99	61
Actual Netral	31	556	142
Actual Positif	3	37	831
	Negatif	Netral	Positif

Predicted

Gambar 5 *Random Forest* Confusion Matrix

Confusion matrix untuk model *Random Forest*. Model ini menunjukkan kinerja yang lebih stabil dibandingkan *Naïve Bayes*, terutama pada kelas positif dengan 831 data berhasil diklasifikasikan secara benar. Pada kelas netral, model mengidentifikasi 556 data secara akurat, meskipun masih terdapat sejumlah kesalahan prediksi ke kelas positif (142 data). Sementara itu, untuk kelas negatif, *Random Forest* menghasilkan 220 prediksi benar dan beberapa kesalahan prediksi ke kelas netral. Secara keseluruhan, pola distribusi prediksi

menunjukkan bahwa *Random Forest* memiliki kemampuan representasi fitur yang lebih baik dibandingkan *Naïve Bayes*, namun masih berada di bawah *SVM* dalam aspek akurasi dan ketepatan prediksi antar kelas. Nilai akurasi akhir sebesar 0.89 menempatkannya sebagai model dengan performa menengah dalam penelitian ini.

CONCLUSION

Berdasarkan seluruh rangkaian penelitian yang telah dilakukan, mulai dari proses pengumpulan data ulasan pengguna aplikasi Roblox, pra-pemrosesan teks, pelabelan sentimen menggunakan *InSet lexicon*, hingga penerapan tiga algoritma *machine learning supervised* (*SVM*, *Naïve Bayes*, dan *Random Forest*), diperoleh beberapa kesimpulan penting sebagai berikut:

Penelitian berhasil melakukan proses analisis awal terhadap ± 100.000 ulasan pengguna aplikasi Roblox yang diperoleh melalui teknik *web scraping* menggunakan library *google-play-scraper* pada Python. Setelah dilakukan tahapan *preprocessing*—meliputi *cleaning*, *case folding*, *normalisasi*, *tokenisasi*, *stopword removal*, dan *stemming*—diperoleh 9.897 ulasan yang layak untuk dianalisis. Penelitian ini berhasil mengimplementasikan pelabelan sentimen dilakukan menggunakan metode *lexicon-based* dengan kamus sentimen Bahasa Indonesia (*InSet*), sehingga setiap ulasan dapat diklasifikasikan sebagai positif, negatif, atau netral. Proses ini menghasilkan pemahaman yang jelas mengenai persepsi pengguna terhadap aplikasi Roblox serta struktur data yang siap digunakan dalam pelatihan model *machine learning*.

Penelitian ini berhasil mengimplementasikan ketiga algoritma—*Naïve Bayes*, *Support Vector Machine (SVM)*, dan *Random Forest*—berhasil diterapkan menggunakan representasi fitur *TF-IDF* yang menghasilkan 7.250 fitur unik dari seluruh korpus ulasan. *Dataset* dibagi menjadi 80% data latih dan 20% data uji, kemudian model dilatih menggunakan pustaka *scikit-learn* pada Python. Setiap model diuji menggunakan metrik standar klasifikasi meliputi *accuracy*, *precision*, *recall*, dan *F1-score*. Implementasi berjalan dengan baik dan ketiga algoritma mampu memproses teks berbahasa Indonesia secara optimal.

Hasil evaluasi menunjukkan bahwa *SVM* adalah model paling unggul dengan akurasi tertinggi sebesar 0.89 dan nilai *F1-score* yang

paling seimbang antar ketiga kelas sentimen (positif, negatif, netral). Performa *Random Forest* berada di posisi kedua dengan akurasi 0.80, sedangkan *Naïve Bayes* menghasilkan akurasi 0.62. Dominasi *SVM* disebabkan oleh kemampuannya membentuk *hyperplane* optimal pada data berdimensi tinggi seperti matriks *TF-IDF*. Dengan demikian, *SVM* ditetapkan sebagai model terbaik untuk klasifikasi sentimen ulasan aplikasi Roblox di Google Play Store.

REFERENCE

- [1] X. Chen, Z. Wang, and X. Di, "Sentiment Analysis on Multimodal Transportation during the COVID-19 Using Social Media Data," *Information*, vol. 14, no. 2, p. 113, Feb. 2023, doi: 10.3390/info14020113.
- [2] A. Ligthart, C. Catal, and B. Tekinerdogan, "Systematic reviews in sentiment analysis: a tertiary study," *Artificial Intelligence Review*, vol. 54, no. 7, pp. 4997–5053, Oct. 2021, doi: 10.1007/s10462-021-09973-3.
- [3] Y. Qi and Z. Shabrina, "Sentiment analysis using Twitter data: a comparative application of lexicon- and machine-learning-based approach," *Social Network Analysis and Mining*, vol. 13, no. 1, p. 31, Feb. 2023, doi: 10.1007/s13278-023-01030-x.
- [4] E. C. M. Torres and L. G. de Picado-Santos, "Sentiment Analysis and Topic Modeling in Transportation: A Literature Review," *Applied Sciences*, vol. 15, no. 12, p. 6576, Jun. 2025, doi: 10.3390/app15126576.
- [5] J. Mutinda, W. Mwangi, and G. Okeyo, "Sentiment Analysis of Text Reviews Using Lexicon-Enhanced Bert Embedding (LeBERT) Model with Convolutional Neural Network," *Applied Sciences*, vol. 13, no. 3, p. 1445, Jan. 2023, doi: 10.3390/app13031445.
- [6] M. Tran, C. Draeger, X. Wang, and A. Nikbakht, "Monitoring the well-being of vulnerable transit riders using machine learning based sentiment analysis and social media: Lessons from COVID-19," *Environment and Planning B: Urban Analytics and City Science*, vol. 50, no. 1, pp. 60–75, Jan. 2023, doi: 10.1177/23998083221104489.
- [7] T. Fan, H. Wang, P. Wu, C. Ling, and M. T. Ahvanooey, "Multimodal sentiment analysis for social media contents during public emergencies," *Journal of Data and Information Science*, vol. 8, no. 3, pp. 61–87, Jun. 2023, doi: 10.2478/jdis-2023-0012.