An Ensemble Machine Learning Approach for Sentiment Analysis of Roblox Reviews on Google Play Store
Pendekatan Ensemble Machine Learning untuk Analisis Sentimen Ulasan Roblox di Google Play Store

Annifa Umma'yah Bassiroh[1)], Mochamad Alfan Rosid[2)], Hindarto[3)], Nuril Lutvi Azizah[4)]

[1,2,3,4)] Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia
*Email: alfanrosid@umsida.ac.id

Abstract. *This study aims to classify user sentiment toward Roblox reviews on the Google Play Store using an ensemble machine learning approach. The initial dataset consisted of approximately 11,195 reviews collected using Google Play Scraper. After preprocessing and duplicate removal, 10,937 reviews were used and automatically labeled into positive, neutral, and negative classes using InSet Lexicon. The classification process involved TF-IDF feature extraction, SMOTE–Tomek Links for balancing training data, and four machine learning algorithms: Naive Bayes, Support Vector Machine, K-Nearest Neighbors, and Random Forest. The results show that negative sentiment dominated the dataset at 40.50%, followed by positive sentiment at 31.57% and neutral sentiment at 27.92%. Among individual models, Support Vector Machine achieved the best performance with 93.69% accuracy. Meanwhile, Ensemble Majority Voting produced the highest performance on the 90:10 split, achieving 93.97% accuracy and 93.96% F1-score. These findings indicate that ensemble learning can improve sentiment classification performance on Roblox user reviews.*

Keywords - *sentiment analysis; Roblox reviews; ensemble learning; majority voting; machine learning*

Abstrak. *Penelitian ini bertujuan untuk mengklasifikasikan sentimen pengguna terhadap ulasan Roblox di Google Play Store menggunakan pendekatan ensemble machine learning. Dataset awal berjumlah sekitar 11.195 ulasan yang dikumpulkan menggunakan Google Play Scraper. Setelah melalui tahap preprocessing dan penghapusan data duplikat, jumlah data akhir menjadi 10.937 ulasan yang dilabeli secara otomatis ke dalam kelas positif, netral, dan negatif menggunakan InSet Lexicon. Proses klasifikasi dilakukan melalui ekstraksi fitur TF-IDF, penyeimbangan data menggunakan SMOTE–Tomek Links, serta penerapan algoritma Naive Bayes, Support Vector Machine, K-Nearest Neighbors, dan Random Forest. Hasil penelitian menunjukkan bahwa sentimen negatif mendominasi sebesar 40,50%, diikuti sentimen positif sebesar 31,57% dan sentimen netral sebesar 27,92%. Model individu terbaik diperoleh oleh Support Vector Machine dengan akurasi sebesar 93,69%. Sementara itu, Ensemble Majority Voting menghasilkan performa tertinggi pada rasio 90:10 dengan akurasi sebesar 93,97% dan F1-score sebesar 93,96%. Hasil ini menunjukkan bahwa pendekatan ensemble mampu meningkatkan performa klasifikasi sentimen ulasan Roblox.*

Kata Kunci - *analisis sentimen; ulasan Roblox; ensemble learning; majority voting; machine learning*

I. PENDAHULUAN

Ulasan pengguna semakin penting di platform seperti Google Play Store dalam proses pengambilan keputusan konsumen karena pesatnya pertumbuhan pasar game seluler dan meningkatnya ketergantungan masyarakat terhadap aplikasi digital [1]. Google Play Store adalah toko aplikasi digital berbasis Android dengan beragam pilihan aplikasi, mulai dari belanja online dan transportasi hingga hiburan dan pendidikan. Melalui platform ini, pengguna dapat memberikan penilaian berupa komentar dan rating bintang terhadap aplikasi yang digunakan. Ulasan tersebut mengandung informasi berharga tentang kualitas layanan, performa aplikasi, serta pengalaman pengguna secara langsung [2]. Oleh sebab itu, banyak pengembang aplikasi yang memanfaatkan ulasan di Play Store sebagai masukan untuk meningkatkan kualitas dan kinerja produk mereka.

Untuk mengetahui pola opini pengguna tentang aplikasi tertentu, penelitian sebelumnya telah menunjukkan bahwa data ulasan dari Google Play Store dapat menjadi sumber utama analisis sentimen [2]. Jutaan pengguna dapat berinteraksi, berkreasi, dan berbagi pengalaman daring di Roblox, sebuah platform game sosial dengan basis pengguna yang besar. Ulasan pengguna Roblox di Google Play Store mengungkapkan bagaimana masyarakat umum menilai kualitas layanan, keandalan aplikasi, dan peningkatan fitur yang dilakukan oleh pengembang. Meskipun demikian, analisis manual terbukti tidak efisien dan membutuhkan banyak tenaga kerja karena besarnya volume penilaian dan beragamnya bahasa dan frasa pengguna [3]. Hal ini menunjukkan perlunya metode analisis sentimen otomatis berbasis *supervised learning* untuk mengklasifikasikan opini pengguna ke dalam kategori sentimen positif, negatif, dan netral. Menurut [4] pendekatan ini bertujuan untuk membantu pengembang memahami tingkat kepuasan konsumen dan mengidentifikasi area yang perlu dikembangkan.

Dalam beberapa tahun terakhir, analisis sentimen telah menjadi fokus utama dalam penelitian bidang pengolahan bahasa alami (NLP), khususnya dalam mengklasifikasikan teks berdasarkan emosi atau opini yang terkandung di dalamnya [5]. Salah satu aplikasi fundamental dari analisis sentimen adalah pada ulasan produk dan aplikasi, yang secara signifikan membantu pengembang dan perusahaan dalam memahami persepsi pengguna terhadap produk mereka.

Penelitian terdahulu oleh Alkindi et al. [6] telah membahas analisis sentimen terhadap ulasan pengguna game Roblox di Google Play Store dengan memanfaatkan dua algoritma *supervised learning*, yaitu *SVM* dan *Naive Bayes*. Menurut temuan studi, algoritma *SVM* lebih akurat daripada *Naive Bayes* dalam membagi ulasan menjadi kelompok positif dan negatif [7]. Metode ini, yang hanya menggunakan dua algoritma berbeda, memiliki keterbatasan dalam menghasilkan klasifikasi yang lebih andal dan akurat, terutama ketika menangani data yang tidak seimbang.

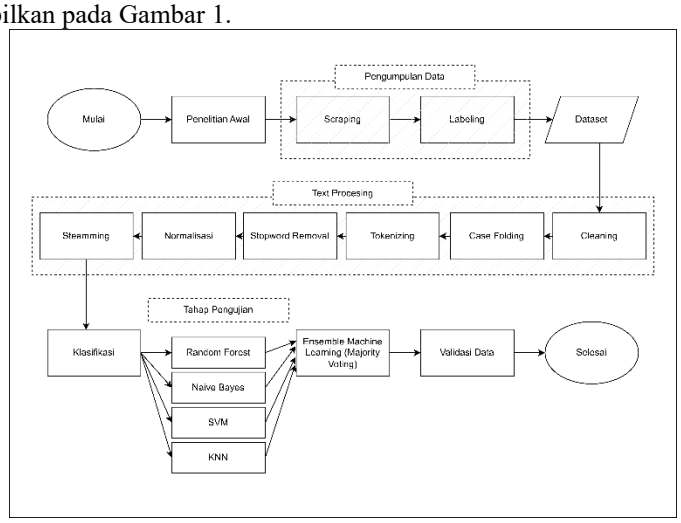
Oleh karena itu, penelitian ini bertujuan untuk memperluas pendekatan sebelumnya dengan mengimplementasikan metode ensemble menggunakan empat algoritma secara simultan, meliputi *Naive Bayes, SVM*, *Random Forest*, dan *KNN*. Pemilihan keempat model ini didasarkan pada karakteristik dan keunggulan yang beragam dalam memproses data teks. Selanjutnya, untuk mengoptimalkan kinerja klasifikasi, penelitian ini mengadopsi metode pembelajaran ensemble melalui teknik *majority voting*. Penelitian ini bertujuan untuk meningkatkan kualitas analisis sentimen evaluasi pengguna aplikasi game, terutama Roblox, dan menawarkan solusi klasifikasi yang lebih fleksibel terhadap beragamnya data ulasan [8].

Dalam analisis sentimen, *supervised learning* merupakan metode yang dominan, di mana suatu model dilatih pada data berlabel untuk memperkirakan kategori sentimen (positif, negatif, atau netral). Algoritma umum yang digunakan dalam *supervised learning* untuk analisis sentimen meliputi *Naive Bayes* , *Support Vector Machine* (*SVM*), *Random Forest*, dan *K-Nearest Neighbors* (*KNN*).

Sebagai metode probabilistik, *Naive Bayes* dikenal karena efisiensinya dalam menangani kumpulan data besar dan kemampuannya untuk mengklasifikasikan data teks dengan cepat berdasarkan frekuensi kemunculan kata-kata tertentu [9]. Di sisi lain, *SVM* cukup baik dalam mengklasifikasikan teks yang kompleks karena dapat menangani data berdimensi tinggi dan membedakan antarkelas dengan jelas. [10]. *Random Forest* bekerja dengan menggabungkan sejumlah pohon keputusan untuk meningkatkan kestabilan hasil prediksi dan menekan risiko *overfitting*, sedangkan *K-Nearest Neighbors* menentukan kelas berdasarkan tingkat kedekatan antar data [11]. Meskipun setiap model memiliki kelebihan dan kekurangannya masing-masing, pembelajaran *ensemble learning* yang mengintegrasikan prediksi dari beberapa model dapat menghasilkan prediksi yang lebih andal dan akurat [12]. Pendekatan *ensemble machine learning* dapat meningkatkan performa sistem dengan menggabungkan kekuatan dari beberapa algoritma pembelajaran mesin sehingga menghasilkan klasifikasi yang lebih kuat dan akurat [13]. Salah satu metodenya adalah pemungutan suara mayoritas, di mana prediksi akhir adalah kelas yang paling sering dipilih oleh setiap model [14].

II. METODE

Penelitian ini menerapkan pendekatan kuantitatif berbasis *supervised learning* sebagai metode dalam mengklasifikasikan sentimen ulasan pengguna aplikasi Roblox di Google Play Store. Adapun alur tahapan penelitian secara keseluruhan ditampilkan pada Gambar 1.



Gambar 1. Alur Penelitian

A. Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan dengan menggunakan Google Play Scraper, yang memungkinkan peneliti untuk mengunduh ulasan pengguna aplikasi game Roblox di Google Play Store. Data yang berhasil dikumpulkan terdiri dari 11,195 ulasan pengguna dalam bentuk teks dan rating bintang, yang disimpan dalam file Dataset.csv. Setiap ulasan kemudian diberi label dengan kategori ulasan positif, netral, dan negatif. Contoh dataset disajikan pada Tabel 1.

Tabel 1. Contoh Dataset

userName	score	at	content
Pengguna Google	4	2026-04-29 15:05:34	gemnya seru tapi kenapa tiba-tiba keluar terus dari permainannya selalu katanya koneksi jaringan buruk menyambung ulang terus padahal sinyal nya bagus jadi tolong di perbaiki itu saja terimakasih
Pengguna Google	1	4/18/2026 12:11:32	Makin lama makin ga seru, sekarang udah gak bisa nge chat in game lagi, terus katanya nanti bakal ada pembagian konten game sesuai usia jadi gak bakal bisa mabar seleluasa nya lagi, apalah 👎
Pengguna Google	3	2026-04-27 15:18:10	game nya asik sih, banyak map dan tmpat gme baru, cuman kadang kadang leg bug juga mau masuk sekarang susah banget padahal jaringan aja for 4G tapi tetap ga bisa masuk, tolong di perbaikin lagi ya

B. *Text Processing*

Tahap *text processing* dilakukan untuk membersihkan dan menyeragamkan data ulasan agar siap digunakan dalam analisis sentimen. Proses ini meliputi *cleaning, case folding, tokenization, stopword removal, normalisasi,* dan *stemming* [15]. Hasil perubahan teks pada setiap tahap ditampilkan pada Tabel 2.

Tabel 2. Contoh *Text Processing*

Tahap	Komentar
Data Awal	Makin lama makin ga seru, sekarang udah gak bisa nge chat in game lagi, terus katanya nanti bakal ada pembagian konten game sesuai usia jadi gak bakal bisa mabar seleluasa nya lagi, apalah 👎
Case Folding	makin lama makin ga seru, sekarang udah gak bisa nge chat in game lagi, terus katanya nanti bakal ada pembagian konten game sesuai usia jadi gak bakal bisa mabar seleluasa nya lagi, apalah 👎
Cleaning	makin lama makin ga seru sekarang udah gak bisa nge chat in game lagi terus katanya nanti bakal ada pembagian konten game sesuai usia jadi gak bakal bisa mabar seleluasa nya lagi apalah
Tokenization	['makin', 'lama', 'makin', 'ga', 'seru', 'sekarang', 'udah', 'gak', 'bisa', 'nge', 'chat', 'in', 'game', 'lagi', 'terus', 'katanya', 'nanti', 'bakal', 'ada', 'pembagian', 'konten', 'game', 'sesuai', 'usia', 'jadi', 'gak', 'bakal', 'bisa', 'mabar', 'seleluasa', 'nya', 'lagi', 'apalah']
Normalization	['semakin', 'lama', 'semakin', 'tidak', 'seru', 'sekarang', 'sudah', 'tidak', 'bisa', 'chat', 'dalam', 'game', 'lagi', 'lalu', 'katanya', 'nanti', 'akan', 'ada', 'pembagian', 'konten', 'game', 'sesuai', 'usia', 'jadi', 'tidak', 'akan', 'bisa', 'bermain', 'bersama', 'seleluasa', 'lagi', 'apa']
Stopword Removal	['lama', 'tidak', 'seru', 'tidak', 'bisa', 'chat', 'game', 'pembagian', 'konten', 'game', 'usia', 'tidak', 'bisa', 'bermain', 'bersama', 'seleluasa']
Stemming	['lama', 'tidak', 'seru', 'tidak', 'bisa', 'chat', 'game', 'bagi', 'konten', 'game', 'usia', 'tidak', 'bisa', 'main', 'sama', 'leluasa']

C. Labeling

Pelabelan sentimen dilakukan menggunakan pendekatan *lexicon-based* dengan kamus *InSet Lexicon* untuk mengklasifikasikan ulasan pengguna Roblox ke dalam tiga kelas, yaitu positif, netral, dan negatif. *InSet Lexicon* bekerja dengan mencocokkan setiap kata hasil *text processing* dengan daftar kata yang terdapat dalam kamus sentimen. Kamus tersebut berisi kata-kata berbahasa Indonesia yang telah memiliki nilai atau bobot polaritas tertentu, baik bernilai positif maupun negatif. Setiap kata yang cocok dengan kamus akan diberikan skor sesuai bobotnya, kemudian seluruh skor kata dalam satu komentar dijumlahkan untuk memperoleh total skor sentimen [16]. Komentar diklasifikasikan sebagai sentimen positif apabila total skor lebih dari 0, netral apabila skor sama dengan 0, dan negatif

apabila skor kurang dari 0. Dari total 10.937 ulasan, diperoleh 3.453 ulasan positif, 3.054 ulasan netral, dan 4.430 ulasan negatif. Sampel hasil pelabelan sentimen dapat dilihat pada Tabel 3.

Tabel 3. Hasil Pelabelan Dataset

Text Processing	**Skor**	**Sentimen**
game ini seru banget grafik nya bagus bisa mabar bareng teman dan permainan nya keren	3	Positif
gamenya sih bagus tapi minusnya saat main game di roblox biasanya keluar sendiri dan saat main aplikasi lain biasa saja	0	Netral
sekarang roblox tidak jelas dikit dikit koneksi terputus padahal koneksi lancar tapi sering banget koneksi terputus	-4	Negatif

D. Pembagian Data

Dataset dibagi menjadi data *training* dan data *testing* menggunakan tiga rasio, yaitu 90:10, 80:20, dan 70:30. Pada rasio 90:10 diperoleh 9.843 data *training* dan 1.094 data *testing* . Pada rasio 80:20 diperoleh 8.749 data *training* dan 2.188 data *testing*. Sementara itu, pada rasio 70:30 diperoleh 7.655 data *training* dan 3.282 data *testing* . Pembagian ini dilakukan untuk mengetahui pengaruh komposisi data terhadap performa model klasifikasi sentimen, sebagaimana ditampilkan pada Tabel 4.

Tabel 4. Pembagian Data

Rasio	**Distribusi Data**		**Total**
	Training	***Testing***	
90 : 10	9.843	1.094	10.937
80 : 20	8.749	2.188	10.937
70 : 30	7.655	3.282	10.937

E. Ekstraksi Fitur *TF-IDF*

Setelah tahap *text preprocessing*, data teks diubah menjadi bentuk numerik menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Metode ini memberikan bobot pada setiap kata berdasarkan frekuensi kemunculan dan tingkat kepentingannya dalam dokumen. Hasil pembobotan TF-IDF kemudian digunakan sebagai masukan pada algoritma *Naive Bayes*, *Support Vector Machine*, *K-Nearest Neighbors*, *Random Forest*, dan *Ensemble Majority Voting.*

F. *SMOTE-Tomek Links*

SMOTE-Tomek Links merupakan metode penyeimbangan data yang menggabungkan teknik *oversampling* dan *undersampling*. *SMOTE* digunakan untuk membentuk data sintetis pada kelas minoritas, sedangkan *Tomek Links* digunakan untuk menghapus data yang berada pada batas antarkelas dan berpotensi menimbulkan ambigu. Dalam penelitian ini, *SMOTE-Tomek Links* digunakan pada data latih untuk menyeimbangkan distribusi kelas sentimen sebelum proses klasifikasi dilakukan.

$$x_{\text{new}} = x_i + \lambda(x_{zi} - x_i) \quad (1)$$

Keterangan: xnew adalah data sintetis baru, xi adalah data minoritas, xzi adalah tetangga terdekat dari xi, dan adalah bilangan acak antara 0 sampai 1.

G. *Naive Bayes*

Naive Bayes merupakan algoritma klasifikasi berbasis probabilitas yang menggunakan Teorema *Bayes*. Algoritma ini sering digunakan dalam klasifikasi teks karena sederhana, cepat, dan efektif untuk data berdimensi tinggi.

$$P(C_i \mid x) = \frac{P(x \mid C_i)P(C_i)}{P(x)} \quad (2)$$

Keterangan: PCi|x adalah probabilitas data x termasuk kelas Ci, Px|Ci adalah probabilitas data x pada kelas Ci, PCi adalah probabilitas awal kelas, dan Px adalah probabilitas data.

H. *Support Vector Machine*

Support Vector Machine (SVM) merupakan algoritma *supervised learning* yang bekerja dengan mencari hyperplane terbaik untuk memisahkan data antarkelas. *SVM* sesuai digunakan pada data teks karena mampu menangani fitur berdimensi tinggi dari hasil pembobotan *TF-IDF*.

$$f(x) = w \cdot x + b \tag{3}$$

Keterangan: w adalah bobot model, x adalah vektor fitur, dan b adalah bias.

I. *K-Nearest Neighbors*

K-Nearest Neighbors (KNN) merupakan algoritma klasifikasi yang menentukan kelas data baru berdasarkan kelas mayoritas dari sejumlah tetangga terdekat[17]. Algoritma ini sederhana, tetapi performanya dipengaruhi oleh pemilihan nilai k dan karakteristik fitur.

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{4}$$

Keterangan: dx,y adalah jarak antara data x dan y, sedangkan xi dan yi adalah nilai fitur ke-i.

J. *Random Forest*

Random Forest merupakan metode *ensemble* yang membangun banyak pohon keputusan. Hasil akhir ditentukan berdasarkan suara mayoritas dari seluruh pohon. Salah satu ukuran yang digunakan dalam pembentukan pohon adalah *Gini Impurity*.

$$\text{Gini} = 1 - \sum_{i=1}^{n} p_i^2 \tag{5}$$

Keterangan: pi adalah proporsi data pada kelas ke-i, sedangkan n adalah jumlah kelas.

K. *Majority voting*

Majority voting merupakan teknik *ensemble learning* yang menggabungkan hasil prediksi dari beberapa algoritma. Prediksi akhir ditentukan berdasarkan kelas yang paling banyak dipilih oleh model individual [18]. Dalam penelitian ini, *majority voting* digunakan untuk menggabungkan hasil klasifikasi *Naive Bayes, SVM, Random Forest,* dan *KNN* agar hasil prediksi sentimen menjadi lebih stabil dan optimal.

L. Evaluasi Model

Confusion matrix digunakan untuk mengevaluasi hasil klasifikasi dengan membandingkan label aktual dan label prediksi. Metrik yang digunakan meliputi *accuracy*, *precision*, *recall*, dan *F1-score*. *Accuracy* digunakan untuk mengukur tingkat ketepatan model dalam mengklasifikasikan seluruh data, baik data positif, netral, maupun negatif.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

Precision digunakan untuk mengukur ketepatan model dalam memprediksi data pada suatu kelas sentimen.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{7}$$

Recall digunakan untuk mengukur kemampuan model dalam menemukan seluruh data yang benar pada suatu kelas.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{8}$$

F1-score digunakan untuk mengukur keseimbangan antara nilai *precision* dan *recall*, sehingga dapat menunjukkan performa model secara lebih menyeluruh.

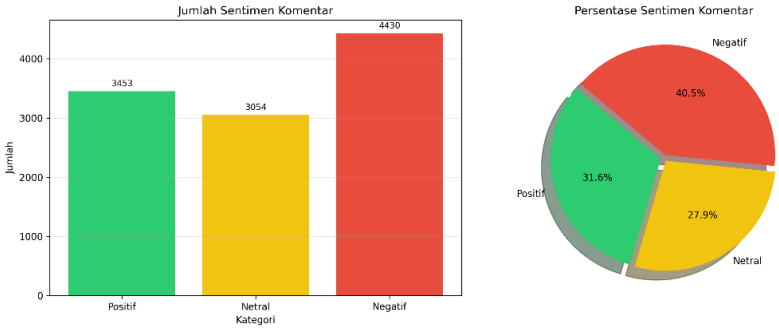
$$F1\text{-}Score = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{9}$$

Keterangan: TP adalah *True Positive*, TN adalah *True Negative*, FP adalah *False Positive*, dan FN adalah *False Negative*.

III. HASIL DAN PEMBAHASAN

A. Distribusi Hasil Pelabelan Sentimen

Dataset awal sebanyak sekitar 11,195 ulasan dikumpulkan melalui Google Play Scraper. Setelah melalui tahap *preprocessing* dan penghapusan data duplikat, jumlah data akhir yang digunakan menjadi 10.937 ulasan.



Gambar 2. Hasil Distribusi Sentimen Ulasan Roblox.

Data tersebut kemudian dilabeli secara otomatis menggunakan *InSet Lexicon*. Hasil pelabelan menunjukkan bahwa sentimen negatif mendominasi sebesar 40,50% atau 4.430 ulasan, diikuti sentimen positif sebesar 31,57% atau 3.453 ulasan, serta sentimen netral sebesar 27,92% atau 3.054 ulasan, sebagaimana ditampilkan pada Gambar 2. Meskipun terjadi pengurangan jumlah data dari dataset awal, proses ini diperlukan untuk memastikan data yang digunakan lebih bersih, tidak berulang, dan siap digunakan pada tahap klasifikasi sentimen.

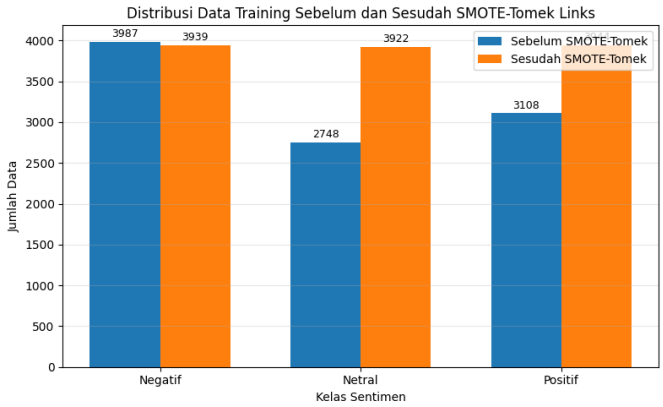
B. Pembagian Data dan Penerapan *SMOTE–Tomek Links*

Penerapan SMOTE–Tomek Links dilakukan pada data training untuk mengatasi ketidakseimbangan kelas sentimen. Pada rasio 90:10, distribusi data sebelum SMOTE–Tomek Links terdiri dari 3.987 data negatif, 2.748 data netral, dan 3.108 data positif. Setelah proses penyeimbangan, jumlah data menjadi lebih proporsional, yaitu 3.939 data negatif, 3.922 data netral, dan 3.944 data positif.

Rasio 90:10 digunakan sebagai visualisasi karena menghasilkan performa terbaik pada model Ensemble Majority Voting, dengan nilai *accuracy* sebesar 0,9397 dan F1-weighted sebesar 0,9396. Hasil ini menunjukkan bahwa SMOTE–Tomek Links mampu memperbaiki ketidakseimbangan distribusi data, khususnya pada kelas netral yang sebelumnya memiliki jumlah paling sedikit.

Tabel 5. Perbandingan Distribusi Data Sebelum dan Sesudah *SMOTE–Tomek Links*

Kelas Sentimen	Sebelum *SMOTE–Tomek Links*	Sesudah *SMOTE–Tomek Links*
Negatif	3.987	3.939
Netral	2.748	3.922
Positif	3.108	3.944
Total	9.843	11.805



Gambar 3. Distribusi Data *Training* Sebelum dan Sesudah *SMOTE–Tomek Links* pada Rasio 90:10

Berdasarkan Tabel 5 dan Gambar 3, terlihat bahwa sebelum proses penyeimbangan, kelas netral memiliki jumlah data paling sedikit dibandingkan kelas negatif dan positif. Setelah diterapkan *SMOTE–Tomek Links*, distribusi data pada ketiga kelas menjadi lebih seimbang. Hal ini menunjukkan bahwa metode tersebut mampu mengurangi ketidakseimbangan data sehingga model dapat mempelajari pola sentimen negatif, netral, dan positif secara lebih proporsional.

C. Hasil Pengujian Model Individu

Pengujian model individu dilakukan menggunakan empat algoritma, yaitu *SVM, Naive Bayes, KNN*, dan *Random Forest*. Pengujian dilakukan pada tiga skenario pembagian data, yaitu 90:10, 80:20, dan 70:30. Evaluasi model menggunakan metrik *accuracy* dan *F1-weighted* untuk mengetahui kemampuan model dalam mengklasifikasikan sentimen ulasan Roblox.

Tabel 6. Hasil Pengujian Model Individu

Rasio	Algoritma	*Accuracy*	*Precision*	*Recall*	F1-Score
90:10	*Naive Bayes*	83,82%	84,44%	83,82%	83,93%
	Random Forest	92,69%	92,74%	92,69%	92,69%
	Support Vector Machine	**93,69%**	**93,71%**	**93,69%**	**93,69%**
	K-Nearest Neighbors	71,30%	75,87%	71,30%	68,66%
80:20	*Naive Bayes*	82,95%	83,95%	82,95%	83,11%
	Random Forest	92,46%	92,55%	92,46%	92,47%
	Support Vector Machine	93,33%	93,35%	93,33%	93,33%
	K-Nearest Neighbors	71,44%	76,70%	71,44%	69,42%
70:30	*Naive Bayes*	83,00%	84,00%	83,00%	83,20%
	Random Forest	92,38%	92,48%	92,38%	92,39%
	Support Vector Machine	92,72%	92,75%	92,72%	92,72%
	K-Nearest Neighbors	70,81%	75,01%	70,81%	68,66%

Berdasarkan hasil pada Tabel 6, *Support Vector Machine* menunjukkan kinerja paling unggul dibandingkan model individu lainnya pada seluruh rasio pembagian data. Capaian tertinggi dihasilkan pada skenario 90:10, yaitu *accuracy* 93,69%, *precision* 93,71%, *recall* 93,69%, dan *F1-score* 93,69%. Sementara itu, *Random Forest* memiliki hasil yang relatif konsisten dengan nilai *accuracy* lebih dari 92% pada setiap skenario pengujian..

Sementara itu, *Naive Bayes* memperoleh performa sedang dengan *accuracy* sekitar 82–83%. K-Nearest *Neighbors* menjadi model dengan performa paling rendah, yaitu sekitar 70–71%. Hal ini menunjukkan bahwa *KNN* kurang optimal pada data teks hasil TF-IDF yang berdimensi tinggi karena proses klasifikasi berbasis jarak menjadi lebih sulit dalam membedakan kedekatan antar dokumen.

D. Hasil Pengujian *Ensemble Majority Voting*

Pengujian selanjutnya menerapkan metode *Ensemble Majority Voting* dengan menggabungkan keputusan dari beberapa algoritma klasifikasi. Penerapan metode ini bertujuan untuk mengevaluasi apakah keputusan mayoritas dari beberapa model dapat menghasilkan klasifikasi sentimen yang lebih optimal. Hasil pengujiannya disajikan pada Tabel 7.

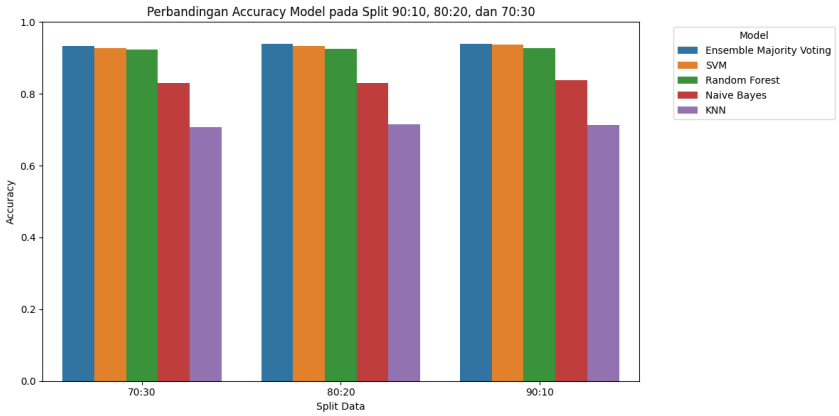
Tabel 7. Hasil Pengujian Model *Ensemble Majority Voting*

Algoritma	*Rasio*	*Accuracy*	*Precision*	*Recall*	F1-Score
Ensemble Majority voting	**90:10**	**93,97%**	**93,96%**	**93,97%**	**93,96%**
	80:20	93,92%	93,94%	93,92%	93,92%
	70:30	93,39%	93,42%	93,39%	93,39%

Tabel 7 menunjukkan bahwa rasio 90:10 menghasilkan kinerja *Ensemble Majority Voting* paling tinggi. Pada skenario tersebut, model mencapai *accuracy* 93,97%, *precision* 93,96%, *recall* 93,97%, dan *F1-score* 93,96%. Adapun rasio 80:20 menghasilkan *accuracy* 93,92%, sedangkan rasio 70:30 memperoleh 93,39%. Meskipun perbedaan hasilnya relatif kecil, rasio 90:10 tetap menjadi skenario dengan performa terbaik.

Meskipun selisih performa antara rasio 90:10 dan 80:20 relatif kecil, rasio 90:10 tetap menjadi skenario terbaik secara numerik. Hasil ini menunjukkan bahwa penggabungan beberapa model melalui majority voting mampu menghasilkan klasifikasi yang stabil dan lebih optimal dibandingkan sebagian besar model individu.

Visualisasi pada Gambar 4 menunjukkan bahwa *Ensemble Majority Voting, Support Vector Machine, dan Random Forest* memiliki performa yang relatif tinggi dan stabil pada seluruh skenario pembagian data. Namun, nilai *accuracy* tertinggi diperoleh oleh *Ensemble Majority Voting* pada rasio 90:10. Sementara itu, *KNN* menunjukkan performa paling rendah dibandingkan model lainnya.



Gambar 4. Perbandingan *Accuracy* Model pada Split 90:10, 80:20, dan 70:30

E. Perbandingan Kinerja Model dan Pembahasan

Berdasarkan hasil pengujian, setiap model menunjukkan performa yang berbeda pada masing-masing skenario pembagian data. Model *Support Vector Machine* menjadi model individu dengan performa terbaik karena mampu menghasilkan nilai *accuracy* dan *F1-score* yang tinggi pada seluruh rasio. Nilai terbaik *SVM* diperoleh pada rasio 90:10 dengan *accuracy* sebesar 93,69% dan *F1-score* sebesar 93,69%. Hal ini menunjukkan bahwa *SVM* efektif digunakan pada data teks hasil ekstraksi fitur *TF-IDF* karena mampu membentuk batas pemisah yang baik antarkelas sentimen.

Model *Random Forest* juga menunjukkan performa yang stabil dengan nilai *accuracy* di atas 92% pada seluruh skenario. Stabilitas ini menunjukkan bahwa *Random Forest* mampu mengenali pola data dengan cukup baik melalui kombinasi beberapa pohon keputusan. Sementara itu, *Naive Bayes* memperoleh hasil yang lebih rendah dibandingkan *SVM* dan *Random Forest* karena model ini bekerja berdasarkan probabilitas kemunculan kata serta mengasumsikan bahwa setiap fitur saling independen.

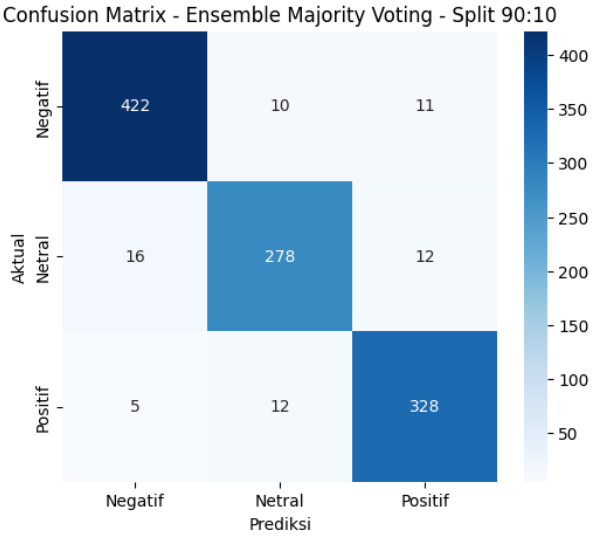
Model *K-Nearest Neighbors* menunjukkan performa paling rendah dibandingkan model lainnya, dengan nilai *accuracy* sekitar 70–71% pada seluruh rasio. Hasil ini menunjukkan bahwa *KNN* kurang optimal dalam mengklasifikasikan data teks hasil *TF-IDF* yang berdimensi tinggi. Hal tersebut dapat terjadi karena *KNN* bekerja berdasarkan jarak antar data, sehingga semakin tinggi dimensi fitur, semakin sulit model menentukan kedekatan antar dokumen secara akurat.

Secara keseluruhan, model *Ensemble Majority Voting* menghasilkan performa terbaik pada rasio 90:10 dengan nilai *accuracy* sebesar 93,97% dan *F1-score* sebesar 93,96%. Hasil ini menunjukkan bahwa penggabungan beberapa model mampu meningkatkan stabilitas prediksi. Dengan menggunakan mekanisme suara mayoritas, kesalahan prediksi dari satu model dapat dikurangi oleh model lainnya. Oleh karena itu, pendekatan *Ensemble Majority Voting* dapat dinilai sebagai metode terbaik dalam klasifikasi sentimen ulasan Roblox pada penelitian ini.

F. Evaluasi Model Terbaik Menggunakan *Confusion Matrix*

Evaluasi lebih lanjut dilakukan menggunakan confusion matrix pada model terbaik, yaitu *Ensemble Majority Voting* dengan rasio 90:10. Berdasarkan confusion matrix, model berhasil mengklasifikasikan 422 data negatif, 278 data netral, dan 328 data positif dengan benar.

Di sisi lain, masih terdapat beberapa Kesalahan klasifikasi pada kelas negatif terdiri atas 10 data yang masuk ke kelas netral dan 11 data ke kelas positif. Pada kelas netral, sebanyak 16 data diklasifikasikan sebagai negatif, sedangkan 12 data lainnya masuk ke kelas positif. Adapun pada kelas positif, terdapat 5 data yang teridentifikasi sebagai negatif dan 12 data sebagai netral. Dari total 1.094 data uji, model mampu mengklasifikasikan 1.028 data secara tepat sehingga memperoleh nilai *accuracy* sebesar 93,97%.



Gambar 5. *Confusion matrix Ensemble Majority Voting* pada Rasio 90:10

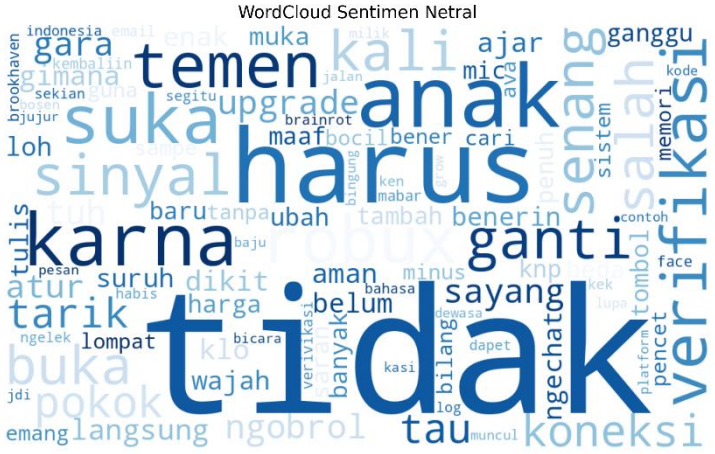
G. Analisis Kata Berdasarkan *WordCloud*

Analisis kata menggunakan wordcloud dilakukan untuk mengetahui kata yang paling sering muncul pada setiap kelas sentimen. Semakin besar ukuran kata, semakin tinggi frekuensi kemunculannya dalam ulasan.

Pada sentimen positif, kata yang dominan muncul antara lain “suka”, “mabar”, “temen”, “gratis”, “baik”, “bareng”, “roblox”, dan “seru”. Kata-kata tersebut menunjukkan bahwa pengguna memberikan respons positif karena Roblox dianggap menyenangkan dan mendukung aktivitas bermain bersama teman.

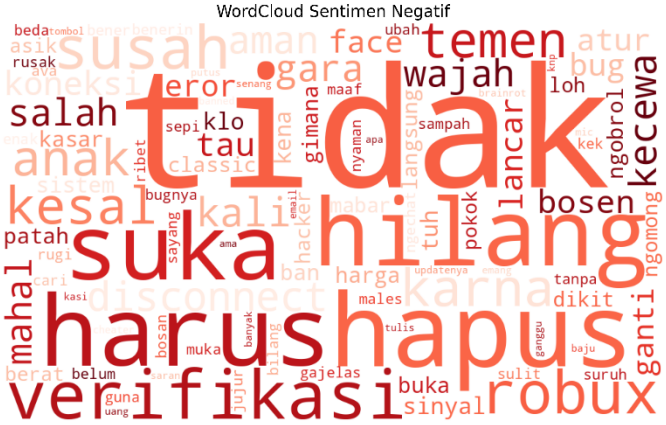
Gambar 6. *WordCloud* Sentimen Positif

Pada sentimen netral, kata yang sering muncul antara lain “tidak”, “harus”, “main”, “game”, “verifikasi”, “robux”, “buka”, dan “koneksi”. Hal ini menunjukkan bahwa ulasan netral cenderung berisi pengalaman biasa atau komentar informatif terkait penggunaan aplikasi.



Gambar 7. *WordCloud* Sentimen Netral

Pada sentimen negatif, kata yang dominan muncul antara lain “tidak”, “hapus”, “susah”, “hilang”, “akun”, “verifikasi”, “robux”, “kecewa”, dan “kesal”. Kata-kata tersebut menunjukkan bahwa ulasan negatif banyak berisi keluhan terkait akun, verifikasi, koneksi, *robux*, dan fitur aplikasi.



Gambar 8. *WordCloud* Sentimen Negatif

Berdasarkan hasil tersebut, dapat disimpulkan bahwa setiap kelas sentimen memiliki karakteristik kata yang berbeda. Sentimen positif didominasi oleh kata yang menunjukkan kepuasan pengguna, sentimen netral berisi kata yang lebih informatif, sedangkan sentimen negatif didominasi oleh kata yang menunjukkan keluhan dan ketidakpuasan pengguna terhadap aplikasi Roblox.

IV. SIMPULAN

Penelitian ini menunjukkan bahwa *Ensemble Majority Voting* memberikan performa terbaik dalam klasifikasi sentimen ulasan Roblox di Google Play Store. Dari 10.937 data yang diberi label secara otomatis menggunakan *InSet Lexicon*, sentimen negatif menjadi kelas paling dominan sebesar 40,50%, diikuti sentimen positif sebesar 31,57% dan sentimen netral sebesar 27,92%. Dominasi sentimen negatif menunjukkan bahwa ulasan pengguna banyak membahas keluhan terkait akun, verifikasi, koneksi, Robux, dan fitur aplikasi. Pada model individu, *Support Vector Machine* memperoleh hasil terbaik dengan *accuracy* sebesar 93,69% pada rasio 90:10, sedangkan performa tertinggi secara keseluruhan dicapai oleh *Ensemble Majority Voting* pada rasio yang sama dengan *accuracy* 93,97% dan *F1-score* 93,96%. Hasil *confusion matrix* juga menunjukkan bahwa model mampu mengklasifikasikan 1.028 dari 1.094 data uji secara tepat. Dengan demikian, kombinasi TF-IDF, SMOTE–Tomek Links, dan *Ensemble Majority Voting* efektif digunakan untuk klasifikasi sentimen ulasan Roblox. Penelitian selanjutnya dapat mengembangkan representasi teks berbasis kontekstual, seperti *word embedding* atau model *transformer*, untuk meningkatkan kemampuan klasifikasi pada data yang lebih kompleks.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Universitas Muhammadiyah Sidoarjo, khususnya Program Studi Informatika Fakultas Sains dan Teknologi, yang telah memberikan dukungan dan fasilitas dalam pelaksanaan penelitian ini. Penulis juga menyampaikan terima kasih kepada orang tua yang senantiasa membersamai, mendoakan, memberikan dukungan, serta motivasi selama proses penyusunan penelitian ini hingga dapat terselesaikan dengan baik.

REFERENSI

[1] A. H. Nurdy, "ANALISIS SENTIMEN ULASAN GAME STUMBLE GUYS PADA PLAYSTORE MENGGUNAKAN ALGORITMA NAÏVE BAYES," 2024.

[2] A. Rahman, E. Utami, and S. Sudarmawan, "Sentimen Analisis Terhadap Aplikasi pada Google Playstore Menggunakan Algoritma Naïve Bayes dan Algoritma Genetika," *J. Komtika (Komputasi dan Inform.*, vol. 5, no. 1, pp. 60–71, 2021, doi: 10.31603/komtika.v5i1.5188.

[3] A. A. Arifiyanti, N. R. Shantika, and A. O. Syafira, "Analisis Sentimen Ulasan Pengguna Bsi Mobile Pada Google Play Dengan Pendekatan Supervised Learning," *J. Inform. Polinema*, vol. 9, no. 3, pp. 283–288, 2023, doi: 10.33795/jip.v9i3.1003.

[4] A. Khan and I. A. , Umair Younis, Alam Sher Kundi, Muhammad Zubair Asghar, Irfan Ullah, Nida Aslam, "Sentiment Classification of User Reviews Using Supervised Learning Techniques with Comparative Opinion Mining Perspective," no. June, 2020, doi: 10.1007/978-3-030-17798-0.

[5] N. Suarna, W. Prihartono, and Nurwanda, "Penerapan NLP (Natural Language Processing) dalam Analisis Sentimen Pengguna Telegram di Playstore," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 1841–1848, 2024.

[6] A. F. Alkindi and N. Nasution, "Analisis Sentimen Ulasan Pengguna Pada Game Roblox Dengan Metode Support Vector Machine Dan Naive Bayes," *J-Com (Journal Comput.*, vol. 4, no. 2, pp. 164–177, 2024, doi: 10.33330/j-com.v4i2.3319.

[7] A. Mustopa, Hermanto, Anna, E. B. Pratama, A. Hendini, and D. Risdiansyah, "Analysis of user reviews for the pedulilindungi application on google play using the support vector machine and naive bayes algorithm based on particle swarm optimization," *2020 5th Int. Conf. Informatics Comput. ICIC 2020*, vol. 2, 2020, doi: 10.1109/ICIC50835.2020.9288655.

[8] D. E. Sondakh, S.Kom, M.T, Ph.D, S. W. Taju, M. G. Tene, and A. E. T. Pangaila, "Sistem Analisis Sentimen Ulasan Aplikasi Belanja Online Menggunakan Metode Ensemble Learning," *CogITo Smart J.*, vol. 9, no. 2, pp. 280–291, 2023, doi: 10.31154/cogito.v9i2.525.280-291.

[9] R. S. Hajjul Ikram, M. Aviciena Hasibuan, Herdi Rianda, Angga Fahriza, Suryadi, "IMPLEMENTASI METODE KLASIFIKASI NAÏVE BAYES PADA ANALISIS SENTIMEN ULASAN APLIKASI MOBILE LEGENDS DI GOOGLE PLAY STORE Hajjul," pp. 1–23, 2024, [Online]. Available: https://jurnal.lldikti13.id/index.php/jd/article/view/33-41

[10] S. I. Nurhafida and F. Sembiring, "Analisis Sentimen Aplikasi Novel Online Di Google Play Store Menggunakan Algoritma Support Vector Machine (SVM)," *J. Sains Komput. Inform. (J-SAKTI*, vol. 6, no. 1, pp. 317–327, 2022.

[11] N. F. Basri and E. Utami, "Penerapan Model Word2Vec dan LSTM dalam Analisis Sentimen Ulasan Pengguna Mobile legends Application of Word2Vec and LSTM Models in Sentiment Analysis of Mobile Legends User Reviews," vol. 14, pp. 856–871, 2025.

[12] Y. A. Mustofa, I. Surya, and K. Idris, "Pendekatan Ensemble pada Analisis Sentimen Ulasan Aplikasi Google Play Store Ensemble Approach to Sentiment Analysis of Google Play Store App Reviews," *Jambura J. Electr. Electron. Eng.*, vol. 6, no. 2, pp. 181–188, 2024.

[13] M. A. K. Mochamad Alfan Rosid, S. Sambada, and F. M. Busono, "Ensemble Machine Learning to Detect Sarcasm in English on Twitter Social Media," vol. 8106, pp. 11–20, 2023, [Online]. Available: https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=ensemble+machine+learning+to+detect+sarcasm+in+english+on+twitter+social+media+&btnG=

[14] B. Samodera, K. Kartini, and M. M. Al Haromainy, "Implementasi Majority Vote Pada Metode Naive Bayes Dan Support Vector Machine(Studi Kasus : Kenaikan Pajak Hiburan)," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, pp. 2525–2535, 2024, doi: 10.23960/jitet.v12i3.4799.

[15] D. Rifaldi, Abdul Fadlil, and Herman, "Teknik Preprocessing Pada Text Mining Menggunakan Data Tweet 'Mental Health,'" *Decod. J. Pendidik. Teknol. Inf.*, vol. 3, no. 2, pp. 161–171, 2023, doi: 10.51454/decode.v3i2.131.

[16] Muhammad Fernanda Naufal Fathoni, Eva Yulia Puspaningrum, and Andreas Nugroho Sihananto, "Perbandingan Performa Labeling Lexicon InSet dan VADER pada Analisa Sentimen Rohingya di Aplikasi X dengan SVM," *Modem J. Inform. dan Sains Teknol.*, vol. 2, no. 3, pp. 62–76, 2024, doi: 10.62951/modem.v2i3.112.
[17] Ratih Puspitasari, Y. Findawati, and M. A. Rosid, "Sentiment Analysis of Post-Covid-19 Inflation Based on Twitter Using the K-Nearest Neighbor and Support Vector Machine Classification Methods," *J. Tek. Inform.*, vol. 4, no. 4, pp. 669–679, 2023, doi: 10.52436/1.jutif.2023.4.4.801.
[18] Z. Yuna Burnama, M. Alfan Rosid, and N. Lutvi Azizah, "Analisis Sentimen Pada Komentar Youtube Dalam Turnamen MPL Season 13 Dengan Metode Ensemble Machine Learning," *TeIKa*, vol. 14, pp. 93–106, 2024.

Conflict of Interest Statement:
The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.