

Analisis Sentimen Ulasan Aplikasi Roblox di *Google Play Store* Menggunakan Algoritma SVM

Resti Amelia Nur Hidayah¹, Ramdan Hermawan², Iksan Rifki Maulana³, Siti Maesaroh⁴
Informatika, Fakultas Teknik, Universitas Mayasari Bakti, Tasimalaya/Jawa Barat, Indonesia^{1,2,3,4}

*Email Korespodensi: restiamelianh29@gmail.com

Sejarah Artikel:

Diterima 31-12-2025
Disetujui 11-01-2026
Diterbitkan 13-01-2026

ABSTRACT

This study aims to analyze the sentiment of Roblox app user reviews on the Google Play Store using the Support Vector Machine (SVM) algorithm with a comparison of TF-IDF and Word2Vec feature representations. The data used are 1,000 Indonesian language reviews collected automatically through web scraping techniques. Sentiment labeling is carried out based on user ratings into three classes: positive, neutral, and negative. Preprocessing stages include case folding, text cleaning, tokenization, stopword removal, and stemming. The experimental results show that the SVM model with TF-IDF produces an accuracy of 80.2%, while the SVM with Word2Vec achieves an accuracy of 79.19%. Both models show the best performance in the positive sentiment class, but experience a decrease in performance in the neutral and negative classes due to data imbalance. The results show that SVM is effective for sentiment analysis of game app reviews, with performance affected by class distribution and feature representation.

Keywords: Sentiment analysis; Roblox; SVM; TF-IDF; Word2Vec

ABSTRAK

Penelitian ini bertujuan menganalisis sentimen ulasan pengguna aplikasi Roblox di Google Play Store menggunakan algoritma Support Vector Machine (SVM) dengan perbandingan representasi fitur TF-IDF dan Word2Vec. Data yang digunakan berupa 1.000 ulasan berbahasa Indonesia yang dikumpulkan secara otomatis melalui teknik web scraping. Pelabelan sentimen dilakukan berdasarkan rating pengguna ke dalam tiga kelas, yaitu positif, netral, dan negatif. Tahapan prapemrosesan meliputi case folding, pembersihan teks, tokenization, stopword removal, dan stemming. Hasil eksperimen menunjukkan bahwa model SVM dengan TF-IDF menghasilkan akurasi sebesar 80,2%, sedangkan SVM dengan Word2Vec memperoleh akurasi 79,19%. Kedua model menunjukkan performa terbaik pada kelas sentimen positif, namun mengalami penurunan kinerja pada kelas netral dan negatif akibat ketidakseimbangan data. Hasil penelitian menunjukkan bahwa SVM efektif untuk analisis sentimen ulasan aplikasi game, dengan kinerja yang dipengaruhi oleh distribusi kelas dan representasi fitur.

Katakunci: Analisis sentimen; Roblox; SVM; TF-IDF; Word2Vec

Bagaimana Cara Sitasi Artikel ini:

Nur Hidayah, R. A., Hermawan, R., Maulana, I. R., & Maesaroh, S. (2026). Analisis Sentimen Ulasan Aplikasi Roblox di Google Play Store Menggunakan Algoritma SVM. Jejak Digital: Jurnal Ilmiah Multidisiplin, 2(1), 1785-1794. <https://doi.org/10.63822/nkmqv561>

PENDAHULUAN

Perkembangan teknologi informasi telah mendorong peningkatan signifikan dalam penggunaan aplikasi digital di berbagai sektor, termasuk hiburan dan pendidikan. Laporan Digital 2025 Indonesia menunjukkan bahwa jumlah pengguna internet di Indonesia telah mencapai lebih dari 212 juta jiwa, dengan Google Play Store sebagai salah satu platform distribusi aplikasi terbesar (*Digital 2025 Indonesia*, 2025). Kondisi ini menyebabkan meningkatnya volume ulasan pengguna yang mengandung opini terhadap kualitas aplikasi.

Roblox merupakan salah satu aplikasi permainan berbasis sosial yang mengalami pertumbuhan pesat, khususnya di kalangan generasi muda. Ulasan pengguna pada Google Play Store mencerminkan tingkat kepuasan serta permasalahan teknis yang dialami pengguna. Namun, jumlah ulasan yang sangat besar dan penggunaan bahasa informal membuat analisis sentimen secara manual menjadi tidak efisien (Pang et al., n.d.).

Analisis sentimen merupakan bagian dari text mining dan natural language processing (NLP) yang bertujuan untuk mengidentifikasi opini atau emosi dalam teks secara otomatis. Salah satu algoritma yang banyak digunakan dalam klasifikasi sentimen adalah Support Vector Machine (SVM), karena kemampuannya dalam menangani data berdimensi tinggi dan menghasilkan batas klasifikasi yang optimal (Joachims, n.d.).

Penelitian sebelumnya menunjukkan bahwa SVM memiliki performa yang baik dalam analisis sentimen ulasan aplikasi digital. (Fide et al., 2021) berhasil mencapai akurasi tinggi pada ulasan aplikasi TikTok, selanjutnya (Wibowo et al., 2025) menunjukkan bahwa kombinasi SVM dan Word2Vec mampu meningkatkan kinerja klasifikasi. Sementara itu, (Dewi et al., 2025) mengintegrasikan SVM dengan pendekatan lexicon-based menggunakan InSet pada ulasan aplikasi Roblox dan berhasil mencapai akurasi sebesar 87,03%, yang menunjukkan potensi besar penggunaan SVM pada domain hiburan digital. Namun, penelitian yang secara khusus membahas ulasan aplikasi permainan seperti Roblox masih terbatas, terutama yang membandingkan TF-IDF dan Word2Vec.

Berdasarkan hal tersebut, penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna aplikasi Roblox menggunakan algoritma SVM serta membandingkan kinerja representasi fitur TF-IDF dan Word2Vec.

METODE PELAKSANAAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen komputasional (Prof. Dr. Sugiyono, n.d.). Data penelitian dikumpulkan dari Google Play Store menggunakan teknik web scraping, menghasilkan 1.000 ulasan pengguna berbahasa Indonesia.

Tahapan prapemrosesan teks meliputi case folding, cleansing, tokenization, stopword removal, dan stemming untuk meningkatkan kualitas data teks (Suharman & Kamayani Sulaeman, 2025). Representasi fitur dilakukan menggunakan TF-IDF dan Word2Vec. TF-IDF digunakan untuk memberikan bobot kata berdasarkan frekuensi dan distribusinya dalam dokumen, sedangkan Word2Vec menggunakan pendekatan skip-gram untuk menangkap hubungan semantik antar kata (Mikolov et al., 2013).

Klasifikasi sentimen dilakukan menggunakan algoritma SVM dengan kernel Radial Basis Function (RBF). Evaluasi model dilakukan menggunakan metrik accuracy, precision, recall, dan F1-score (Putra, n.d.2025).

HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna aplikasi ROBLOX di Google Play Store menggunakan algoritma Support Vector Machine. Proses pengujian dilakukan setelah seluruh tahapan metode penelitian, mulai dari prapemrosesan teks hingga pembentukan model klasifikasi, berhasil diterapkan secara menyeluruh.

Hasil Pengumpulan Data

Data penelitian diperoleh secara otomatis melalui scraping ulasan aplikasi Roblox pada platform Google Play Store menggunakan pustaka google-play-scraper. Sebanyak 1.000 ulasan pengguna berbahasa Indonesia berhasil dikumpulkan. Setiap ulasan diberi label sentimen berdasarkan skor rating yang diberikan pengguna dengan ketentuan :

- Rating ≥ 4 dikategorikan sebagai sentimen positif
- Rating =3 sebagai sentimen netral
- Rating ≤ 2 sebagai sentimen negatif

Distribusi sentimen dari hasil labeling adalah sebagai berikut seperti di gambar 1:

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Unzipping corpora/stopwords.zip.
📥 Mengambil 1000 ulasan terbaru aplikasi ROBLOX dari Google Play...
Jumlah data diambil: 1000

Distribusi Sentimen:
label
positif    797
negatif    160
netral      43
Name: count, dtype: int64
```

Gambar 1. Hasil Distribusi Sentimen
(Sumber: Hasil pengolahan Data di google collab, 2026)

Distribusi sentimen yang diperoleh dari data awal menunjukkan bahwa mayoritas pengguna memberikan ulasan positif sebesar 79,7%, sedangkan ulasan negatif sebanyak 16,0%, dan netral sebesar 4,3%. Hal ini menunjukkan bahwa mayoritas pengguna memiliki pengalaman yang cukup baik terhadap aplikasi Roblox, meskipun sebagian pengguna masih melaporkan masalah teknis seperti lag, crash, dan bug dalam permainan.

Hasil Preprocessing dan Representasi Fitur

Sebelum pelatihan model, dilakukan proses text preprocessing yang meliputi case folding, tokenization, stopword removal, dan stemming menggunakan algoritma Nazief-Adriani. Proses ini bertujuan untuk menghilangkan kata tidak relevan dan mengubah kata menjadi bentuk dasar, sehingga memperbaiki kualitas data teks.

Dua metode representasi fitur digunakan dalam penelitian ini :

1. TF-IDF (Term Frequency-Inverse Document Frequency) menghitung bobot kata berdasarkan frekuensi kemunculan dan distribusi antar dokumen.
2. Word2Vec (Skip-Gram, vektor 300 dimensi)

Menghasilkan representasi vektor kata yang mempertahankan makna semantik antar kata. Pendekatan ganda ini memungkinkan analisis perbandingan efektivitas model SVM terhadap dua bentuk representasi teks yang berbeda(Ramadhan, 2025).

Contoh disini hasil pembersih teks pada gambar 2

```

Contoh data bersih:
      ulasan      rating      label \
0      iklan tempekkkk      1      negatif
1      bagus 🍌      5      positif
2      saya menyukai Roblox karena dia teman satu² ku...      5      positif
3      seru      3      netral
4      game nya bagus banyak map yang seru banget bua...      5      positif

      clean_text
0      iklan tempekkkk
1      bagus
2      suka roblox teman ku teman pas bosan akunn ska...
3      seru
4      game nya bagus map seru banget cobain deh
    
```

Gambar 2. Hasil Pembersihan Teks

(Sumber: Hasil pengolahan Data di google collab, 2026)

Hasil menunjukkan bahwa proses pembersihan data berhasil menghilangkan berbagai elemen non - teks seperti emoji, angka , tanda baca, serta kata-kata yang tidak memiliki makna. Langkah ini sangat penting untuk mempersiapkan korpus teks yang lebih efisien dalam proses representasi fitur dan klasifikasi sentimen.

Hasil Model TF-IDF dan SVM

Representasi teks pertama menggunakan Term Frequency-Inverse Document Frequency (TF-IDF) dengan batas maksimum 5.000 fitur kata. Model SVM dengan kernel RBF (Radial Basis Function) kemudian dilatih dan diuji menggunakan data uji sebanyak 20% dari total dataset. Hasil evaluasi model TF-IDF adalah sebagai berikut :

```

=== HASIL MODEL TF-IDF ===
Accuracy: 80.2 %

      precision      recall      f1-score      support
negatif      0.67      0.06      0.11      32
netral      0.00      0.00      0.00      9
positif      0.80      1.00      0.89      156

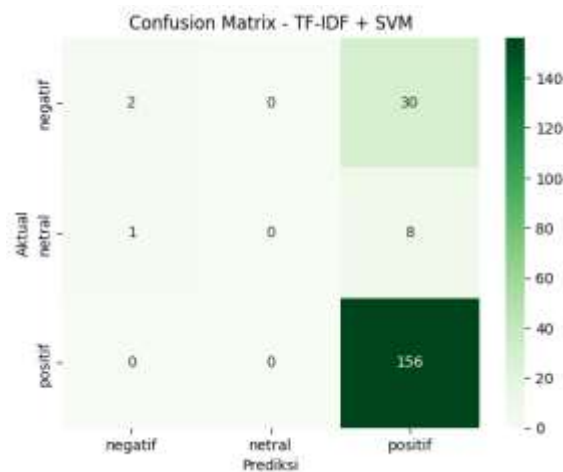
accuracy      0.80      197
macro avg      0.49      0.35      0.34      197
weighted avg      0.75      0.80      0.72      197
    
```

Gambar 3. Hasil Model TF-IDF

(Sumber: Hasil pengolahan Data di google collab, 2026)

Hasil menunjukkan bahwa model TF-IDF dan SVM memiliki performa yang cukup baik dalam mengenali sentimen positif, namun cenderung kurang efektif dalam mengenali ulasan netral dan negatif (Dani et al., n.d.). Hal ini disebabkan oleh ketidakseimbangan jumlah data antar kelas (dominasi ulasan positif sebesar hampir 80%).

Confusion matrix model TF-IDF memperlihatkan bahwa sebagian besar prediksi cenderung mengarah ke kelas positif, yang menunjukkan bias model terhadap kelas dengan jumlah data terbesar seperti pada gambar 4. dibawah ini.



Gambar 4. Confusion Matrix TF-IDF dan SVM
 (Sumber: Hasil pengolahan Data di google collab, 2026)

Hasil Model Word2Vec dan SVM

Representasi teks kedua menggunakan Word2Vec (Skip-Gram) dengan vektor berukuran 300 dimensi dan window size = 5. Model SVM dengan parameter sama (kernel RBF) kemudian dilatih menggunakan data hasil vektorisasi ini. Hasil evaluasi model Word2Vec adalah sebagai berikut :

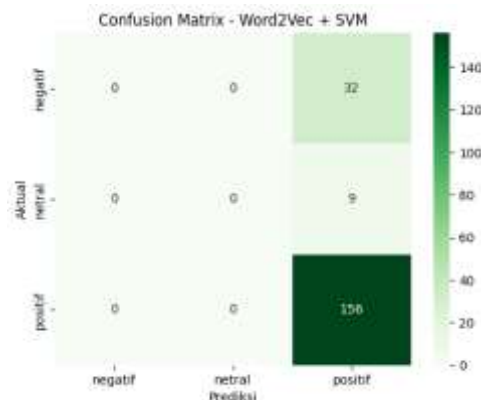
```

=== HASIL MODEL WORD2VEC ===
Accuracy: 79.19 %

```

	precision	recall	f1-score	support
negatif	0.00	0.00	0.00	32
netral	0.00	0.00	0.00	9
positif	0.79	1.00	0.88	156
accuracy			0.79	197
macro avg	0.26	0.33	0.29	197
weighted avg	0.63	0.79	0.70	197

Gambar 5. Hasil Model Word2Vec
 (Sumber: Hasil pengolahan Data di google collab, 2026)



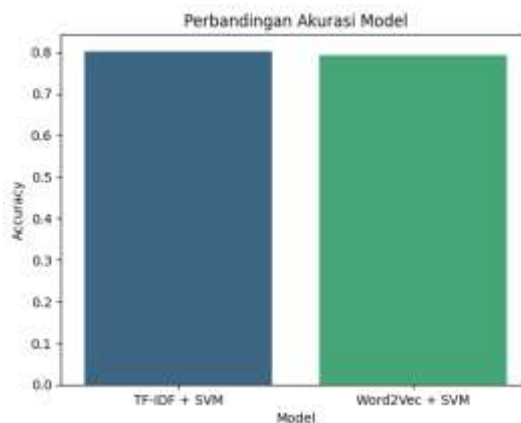
Gambar 6. Confusion Matrix Word2Vec dan SVM
 (Sumber: Hasil pengolahan Data di google collab, 2026)

Dari hasil di atas, terlihat bahwa Word2Vec + SVM memberikan akurasi sedikit lebih rendah dibandingkan TF-IDF, yaitu 79,19%. Sama seperti model TF-IDF, model ini juga menunjukkan bias terhadap kelas positif karena dominasi jumlah datanya (Rifaldy et al., 2025).

Namun demikian, Word2Vec memiliki keunggulan dalam menangkap konteks semantik antar kata, sehingga pada dataset yang lebih seimbang performanya berpotensi meningkat.

Perbandingan Kinerja Model

Perbandingan kedua model divisualisasikan menggunakan diagram batang (bar plot). Dari hasil tersebut, terlihat bahwa :



Gambar 7. Hasil Perbandingan Akurasi Model
(Sumber: Hasil pengolahan Data di google collab, 2026)

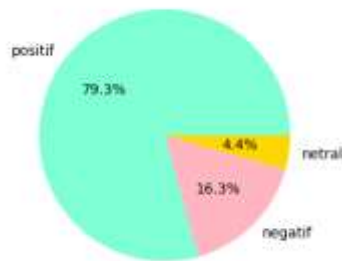
Dari gambar diagram batang diperoleh akurasi TF-IDF+SVM sebesar 80,2% sedangkan akurasi yang diperoleh Word2Vec sebesar 79,19%. Perbedaan akurasi antara kedua model relatif kecil (sekitar 1%), tetapi TF-IDF sedikit lebih unggul. Hal ini menunjukkan bahwa pada data berbahasa Indonesia dengan ekspresi informal dan dominasi opini positif, pendekatan berbasis bobot kata (TF-IDF) masih mampu bekerja lebih stabil dibandingkan pendekatan semantik seperti Word2Vec (Yanti & Utami, 2024).

Analisis Distribusi Sentimen

Hasil klasifikasi menunjukkan dominasi sentimen positif (79,3%), diikuti oleh negatif (16,3%), dan netral (4,4%). Temuan ini menunjukkan bahwa sebagian besar pengguna Roblox di Indonesia memiliki persepsi baik terhadap aplikasi, terutama terkait fitur sosial, kreativitas, dan pengalaman bermain bersama teman. Sementara itu, ulasan negatif cenderung menyoroti masalah teknis seperti lag, bug, dan server error, sedangkan ulasan netral umumnya berupa saran atau evaluasi tanpa emosi yang kuat.

Visualisasi pie chart dari hasil ini memperlihatkan proporsi besar warna hijau (positif), yang menunjukkan tingkat kepuasan pengguna yang cukup tinggi terhadap aplikasi Roblox.

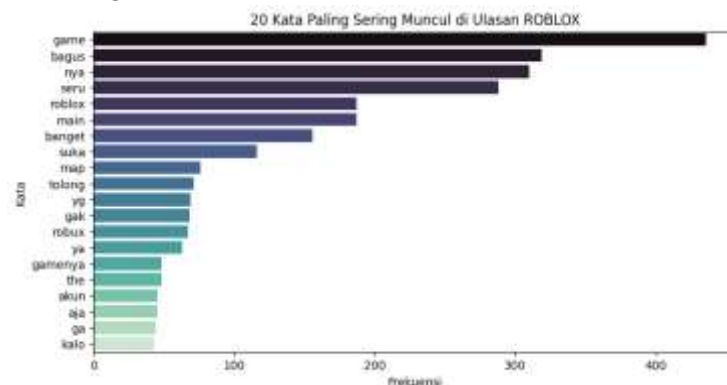
Distribusi Sentimen Pengguna ROBLOX di Google Play



Gambar 8. Hasil Distribusi Sentimen pengguna Roblox di Google Play Store
(Sumber: Hasil pengolahan Data di google collab, 2026)

Analisis Kata Kunci

Analisis frekuensi kata menunjukkan kata paling sering muncul dalam korpus ulasan adalah: “seru”, “bagus”, “keren”, “main”, dan “asik” untuk sentimen positif, serta “lag”, “error”, “lemot”, “bug”, dan “iklan” untuk sentimen negatif.



Gambar 9. Gambar Hasil Kata yang Sering Muncul di Ulasan Roblox
(Sumber: Hasil pengolahan Data di google collab, 2026)

Visualisasi WordCloud memperkuat hasil tersebut, dengan kata “seru” mendominasi di bagian tengah awan kata berwarna biru muda. Pola ini memperlihatkan bahwa persepsi pengguna terhadap Roblox sangat dipengaruhi oleh aspek keseruan dan pengalaman interaktif, sedangkan kendala teknis menjadi sumber keluhan utama (Suryono, 1868).



Gambar 10. Hasil Visualisasi WordCloud ulasan pada Pengguna Roblox
(Sumber: Hasil pengolahan Data di google collab, 2026)

KESIMPULAN

Berdasarkan hasil penelitian analisis sentimen terhadap 1.000 ulasan aplikasi Roblox di Google Play Store menggunakan algoritma Support Vector Machine (SVM) dengan dua pendekatan representasi fitur teks, yaitu TF-IDF dan Word2Vec, diperoleh beberapa kesimpulan sebagai berikut:

1. Model SVM dengan representasi TF-IDF memperoleh akurasi sebesar 80,2%, sedangkan model SVM dengan representasi Word2Vec menghasilkan akurasi 79,19%. Perbedaan performa ini menunjukkan bahwa pendekatan berbasis bobot kata seperti TF-IDF masih lebih efektif untuk dataset berbahasa Indonesia dengan dominasi opini positif dan ekspresi informal.
2. Distribusi sentimen menunjukkan bahwa mayoritas pengguna memberikan ulasan positif sebesar 79,7%, sedangkan ulasan negatif 16% dan netral 4,3%. Hasil ini menggambarkan bahwa aplikasi Roblox memiliki tingkat kepuasan pengguna yang tinggi, khususnya dari segi kreativitas, interaksi sosial, dan pengalaman bermain bersama.
3. Analisis kata kunci memperlihatkan bahwa kata “seru”, “bagus”, “keren”, “main”, dan “asik” mendominasi sentimen positif, sedangkan kata “lag”, “error”, “lemot”, “bug”, dan “iklan” mendominasi ulasan negatif. Temuan ini menunjukkan bahwa kepuasan pengguna dipengaruhi oleh faktor keseruan permainan dan stabilitas teknis aplikasi.
4. Hasil perbandingan dua model memperlihatkan bahwa SVM masih menjadi algoritma yang handal untuk analisis sentimen ulasan aplikasi, dengan performa yang stabil dan interpretasi hasil yang mudah dilakukan.
5. Secara keseluruhan, penelitian ini membuktikan bahwa kombinasi SVM dan TF-IDF efektif dalam menganalisis opini pengguna berbahasa Indonesia, serta dapat dijadikan dasar bagi pengembangan aplikasi untuk memonitor persepsi pengguna secara otomatis.

DAFTAR PUSTAKA

- Dani, A. H., Puspaningrum, E. Y., & Mumpuni, R. (n.d.). *Studi Performa TF-IDF dan Word2Vec Pada Analisis Sentimen Cyberbullying*.
- Dewi, N. K. F. P., Sudipa, I. G. I., Sunarya, I. W., Kusuma Dewi, N. W. J., & Kusuma, A. S. (2025). Sentiment Analysis of Roblox Game Reviews Using Support Vector Machine Method. *Sinkron*, 9(4), 1863–1876. <https://doi.org/10.33395/sinkron.v9i4.15272>
- Digital 2025 Indonesia: Statistik, Tren, dan Peluang di Era Internet & Media Sosial - Sidik Cyber Media. (2025, September 24). <https://sidikcybermedia.com/digital-2025-indonesia-statistik-tren-dan-peluang-di-era-internet-media-sosial/>
- Fide, S., Suparti, S., & Sudarno, S. (2021). ANALISIS SENTIMEN ULASAN APLIKASI TIKTOK DI GOOGLE PLAY MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM) DAN ASOSIASI. *Jurnal Gaussian*, 10(3), 346–358.
- Joachims, T. (n.d.). *Making large-scale SVM learning practical*.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient Estimation of Word Representations in Vector Space* (No. arXiv:1301.3781). arXiv. <https://doi.org/10.48550/arXiv.1301.3781>
- Pang, B., Lee, L., & Bigg, J. (n.d.). *Opinion mining and sentiment analysis*.
- Prof. Dr. Sugiyono. (n.d.). *METODE PENELITIAN KUANTITATIF, KUALITATIF, DAN R&D*. Retrieved January 10, 2026, from <http://archive.org/details/buku-metode-penelitian-sugiyono>

- Ramadhan, M. F. (2025). Klasifikasi Topik dan Sentimen Judul Berita dengan Augmentasi dan TF-IDF. *RIGGS: Journal of Artificial Intelligence and Digital Business*, 4(2), 6732–6741. <https://doi.org/10.31004/riggs.v4i2.1692>
- Rifaldy, F., Sibaroni, Y., & Prasetyowati, S. S. (2025). Effectiveness of Word2Vec and TF-IDF in Sentiment Classification on Online Investment Platforms Using Support Vector Machine. *JIPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 10(2), 863–874. <https://doi.org/10.29100/jipi.v10i2.6055>
- Suharman, A., & Kamayani Sulaeman, M. (2025). Analisis Sentimen Pengguna Aplikasi Livin' by Mandiri Menggunakan Metode Support Vector Machine (SVM) dengan Ekstraksi Fitur TF-IDF dan Word2Vec. *Jurnal Pendidikan dan Teknologi Indonesia*, 5(8), 2201–2212. <https://doi.org/10.52436/1.jpti.941>
- Suryono, R. R. (1868). *Sentiment Classification of Indonesian-Language Roblox Reviews Using IndoBERT with SMOTE Optimization*. 9(4).
- Tampilan Analisis Sentimen Berbasis Aspek Terhadap Ulasan Aplikasi Qasir Mobile Pada Google Playstore Menggunakan Metode Support Vector Machine (SVM)*. (n.d.). Retrieved January 10, 2026, from <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/15001/6661>
- Wibowo, R. S. W., Rakan, M. R. W. R., Miko, J. B. N., & Arifin, M. (2025). Analisis Sentimen Menggunakan Support Vector Machine dan Naive Bayes Pengguna Game Roblox. *JUKOMIKA (Jurnal Ilmu Komputer Dan Informatika)*, 8(2), 106–112. <https://doi.org/10.54650/jukomika.v8i2.618>
- Yanti, I., & Utami, E. (2024). SENTIMENT ANALYSIS OF INDONESIA'S CAPITAL RELOCATION USING WORD2VEC AND LONG SHORT-TERM MEMORY METHOD. *Jurnal Teknik Informatika (Jutif)*, 6(1), 149–158. <https://doi.org/10.52436/1.jutif.2025.6.1.2712>