

Pemanfaatan Algoritma *Support Vector Machine* dalam Analisis Sentimen Ulasan Pengguna Aplikasi POLRI Presisi

Yopi Handrianto¹, Omar Pahlevi², Mateus Hantoro Winarsa³

^{1,2,3}Universitas Bina Sarana Informatika

e-mail: ¹yopi.yph@bsi.ac.id, ²omar.opi@bsi.ac.id, ³mateus.hantoro@gmail.com

Diterima	Direvisi	Disetujui
27-09-2025	09-11-2025	22-12-2025

Abstrak - Kepolisian Negara Republik Indonesia (Polri) merupakan institusi penting dalam pelayanan publik, termasuk pengurusan administrasi SIM dan SKCK. Untuk mendukung transformasi digital, Polri merilis aplikasi POLRI Presisi. Penelitian ini menganalisis sentimen ulasan pengguna terhadap fitur layanan SIM dan SKCK pada aplikasi tersebut. Penelitian ini menghimpun 4.000 ulasan dari *Google Play Store* dengan metode *web scraping*, yang selanjutnya diberi label sentimen positif, negatif, dan netral menggunakan pendekatan *lexicon-based*. Data ulasan terlebih dahulu melalui tahap prapengolahan teks yang meliputi *cleaning*, *case folding*, normalisasi kata, tokenisasi, penanganan negasi dan *stopword removal* serta proses *stemming*. Selanjutnya, fitur teks direpresentasikan dalam bentuk vektor menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) dan diintegrasikan dengan fitur leksikal untuk memperkaya informasi yang digunakan pada tahap klasifikasi. *Synthetic Minority Oversampling Technique* atau SMOTE diterapkan untuk menyeimbangkan data latih. Model klasifikasi dibangun menggunakan *Support Vector Machine* atau SVM dan dioptimalkan melalui *GridSearchCV*. Hasil analisis terlihat bahwa sentimen negatif mendominasi dengan proporsi 76,36%, diikuti sentimen netral mencapai 12,23% dan sentimen positif mencapai 11,41%. Model *Support Vector Machine* mencapai akurasi dengan proporsi 91,50% dengan nilai *macro F1-score* 0,82 dan *macro ROC-AUC* 0,9693. Dari angka tersebut dapat disimpulkan algoritma *Support Vector Machine* memiliki kinerja yang efektif dalam mengklasifikasikan sentimen ulasan pengguna aplikasi POLRI Presisi dengan tingkat ketepatan yang tinggi.

Kata Kunci: Sentimen Ulasan, SVM, POLRI Presisi

Abstract - The Indonesian National Police (Polri) is a key institution in the provision of public services, including the administration of driving licenses (SIM) and police clearance certificates (SKCK). To support digital transformation, Polri launched the POLRI Presisi mobile application. This study analyzes user sentiment toward the SIM and SKCK service features of the application based on user reviews. A total of 4,000 reviews were collected from the Google Play Store using web scraping techniques and subsequently labeled into positive, negative, and neutral sentiment categories through a lexicon-based approach. The review data underwent several text preprocessing stages, including cleaning, case folding, word normalization, tokenization, negation handling, stopword removal, and stemming. Text features were then represented as vectors using the Term Frequency-Inverse Document Frequency (TF-IDF) method and integrated with lexical features to enrich the information used in the classification process. The Synthetic Minority Oversampling Technique (SMOTE) was applied to balance the training data. The classification model was developed using the Support Vector Machine (SVM) algorithm and optimized through GridSearchCV. The analysis results show that negative sentiment dominates the reviews with a proportion of 76.36%, followed by neutral sentiment at 12.23% and positive sentiment at 11.41%. The Support Vector Machine model achieved an accuracy of 91.50%, with a macro F1-score of 0.82 and a macro ROC-AUC value of 0.9693. These results indicate that the Support Vector Machine algorithm performs effectively in accurately classifying user sentiment toward the POLRI Presisi application.

Keywords: Sentiment Analysis, SVM, POLRI Presisi

PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi menjadi faktor utama yang mempercepat perubahan sistem pelayanan publik di Indonesia melalui penerapan *e-government*, dengan tujuan

meningkatkan efisiensi, transparansi, serta efektivitas layanan (Rizky et al., 2025). Salah satu implementasi nyata dari transformasi ini adalah aplikasi POLRI Presisi, yang dirilis oleh Kepolisian Negara Republik Indonesia untuk mempermudah pengurusan administrasi seperti perpanjangan SIM dan

penerbitan SKCK. Namun, meskipun aplikasi ini diharapkan memberikan kemudahan, ulasan pengguna di *Google Play Store* menunjukkan adanya keluhan terkait kinerja aplikasi, stabilitas fitur, dan kepuasan layanan, yang menandakan perlunya evaluasi berbasis data terhadap pengalaman pengguna.

Pendekatan analisis sentimen terhadap ulasan pengguna digunakan untuk memperoleh gambaran objektif mengenai opini masyarakat serta mendukung proses evaluasi dan pengembangan layanan publik digital (Anggraini & Alita, 2024). *Support Vector Machine* banyak digunakan pada klasifikasi teks dengan dimensi fitur yang besar, serta menunjukkan kemampuan yang baik dalam menghasilkan akurasi tinggi pada analisis sentimen (Iqrom et al., 2025). Studi lain juga mendukung efektivitas SVM untuk pemrosesan opini publik, khususnya dalam konteks ulasan aplikasi dan media sosial (Sevilla & Cahyono, 2025).

Untuk mencapai kinerja model yang optimal, diperlukan prapemrosesan teks yang memadai, mencakup *cleaning*, *case folding*, normalisasi, tokenisasi, *stopword removal*, dan *stemming* guna menjamin kualitas data input (Findawati & Rosid, 2020). Representasi teks tidak hanya mengandalkan TF-IDF, tetapi juga diperkuat dengan fitur leksikal berupa jumlah kata positif, negatif, dan polaritas sentimen untuk meningkatkan kemampuan model klasifikasi (Pratama et al., 2023).

Ketidakeimbangan kelas data menjadi salah satu kendala dalam analisis sentimen karena menyebabkan distribusi sentimen yang tidak proporsional dan berpengaruh pada performa model. Pendekatan SMOTE dipakai sebagai *balancing* data dengan menghasilkan contoh sintetis pada kelas minoritas (Ridwan et al., 2024). Pendekatan yang diterapkan ini mendukung model SVM dalam mempelajari distribusi pola pada sentimen positif, negatif, dan netral secara lebih seimbang.

Studi ini difokuskan pada pengkajian sentimen ulasan pengguna terhadap fitur layanan SIM dan SKCK pada aplikasi POLRI Presisi dengan memanfaatkan algoritma SVM. Riset ini memanfaatkan 4.000 ulasan pengguna aplikasi POLRI Presisi yang diperoleh dari *Google Play Store* melalui pengambilan data otomatis. Ulasan tersebut diklasifikasikan ke dalam sentimen positif, negatif, dan netral menggunakan pendekatan leksikal, kemudian direpresentasikan melalui kombinasi TF-IDF dan fitur leksikal. Ketidakeimbangan data ditangani dengan SMOTE, sementara optimasi model klasifikasi dilakukan menggunakan GridSearchCV. Dengan demikian, penelitian ini diharapkan dapat memberikan gambaran empiris mengenai kinerja algoritma *Support Vector Machine* dalam analisis sentimen, sekaligus menyumbangkan dasar informasi yang relevan bagi upaya peningkatan kualitas layanan publik digital di masa mendatang dan memberikan kontribusi praktis dalam mendukung transformasi digital POLRI sekaligus memperkaya literatur

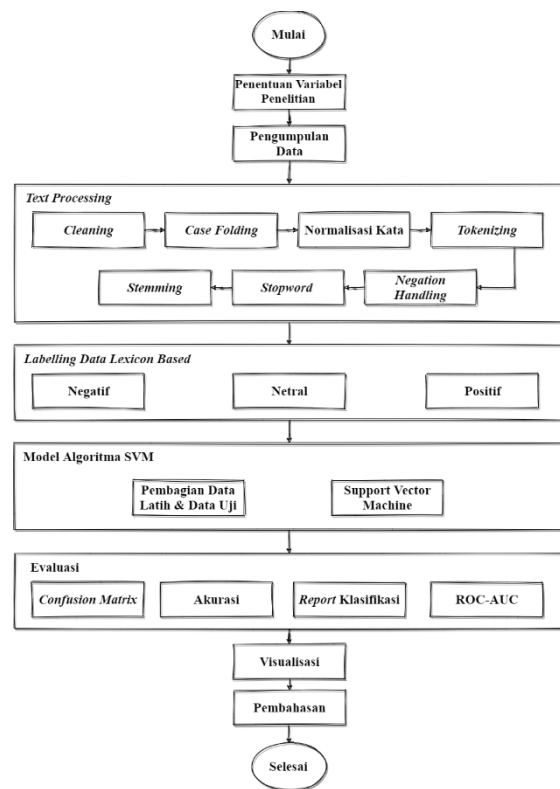
analisis sentimen berbasis SVM pada aplikasi layanan publik di Indonesia.

METODE PENELITIAN

Analisis sentimen dalam penelitian ini dilakukan dengan pendekatan kuantitatif deskriptif berbasis komputasi terhadap ulasan pengguna aplikasi POLRI Presisi terkait layanan SIM dan SKCK. Pendekatan kuantitatif memungkinkan pengolahan data yang telah direpresentasikan secara numerik dilakukan secara sistematis dan terukur (Tampubolon, 2023).

1. Desain Penelitian

Gambar 1 ini merupakan pola penelitian yang dilakukan,



Sumber: Penelitian (2025)

Gambar 1. Pola Penelitian SVM

2. Pengumpulan Data

Pengumpulan data menggunakan *web scraping*. Rentang waktu pengambilan data adalah 1 September 2022 hingga 1 Juni 2025. Dari hasil *web scraping* diperoleh 4000 ulasan berbahasa Indonesia yang secara khusus menyebut layanan SIM dan SKCK. Ulasan ini kemudian dijadikan sampel penelitian.

3. Text Processing (Pra-Pemrosesan Teks)

Tahapan *Text Processing* dilakukan dengan langkah :

- a. *Cleaning*, menghilangkan elemen yang tidak sesuai seperti angka, tanda baca, URL, emoji, dan karakter khusus.
- b. *Case Folding*, mengkonversi semua teks menjadi huruf kecil agar terjaganya konsistensi penulisan.
- c. Normalisasi kata, mengubah kata tidak baku menjadi bentuk baku sesuai kaidah bahasa.
- d. *Tokenisasi*, memecah teks menjadi unit-unit kata.
- e. *Stopword Removal*, menghapus kata-kata umum yang tidak memiliki pengaruh signifikan terhadap makna teks.
- f. *Negation Handling*, yaitu menangani kata negasi agar makna kalimat tetap terjaga dengan tepat.
- g. *Stemming*, yakni mengembalikan kata ke bentuk fundamental untuk menggabungkan variasi kata berimbuhan.

4. Pelabelan Sentimen.

Proses pelabelan sentimen dilakukan dengan menerapkan metode lexicon-based yang mengacu pada kamus sentimen Bahasa Indonesia. Setiap ulasan diklasifikasikan ke dalam kategori positif, negatif, atau netral berdasarkan nilai polaritas dari kata-kata yang terkandung di dalamnya (Putra & Setiawan, 2024).

5. Ekstraksi Fitur.

Ulasan yang telah diberi label diubah menjadi representasi numerik menggunakan TF-IDF untuk kalkulasi nilai bobot kata, serta ditambah fitur leksikal seperti jumlah kata positif, kata negatif, netral, polaritas, dan rasio polaritas (Gifari et al., 2022). Penelitian ini menggunakan SMOTE sebagai solusi terhadap data yang tidak *balance* dengan cara menghasilkan sampel sintetis pada kelas minoritas agar distribusi lebih merata dan model tidak bias terhadap kelas mayoritas. Penelitian ini sesuai dengan riset yang dilakukan (Ridwan et al., 2024), dimana pendekatan yang dilakukan menjelaskan bahwa SMOTE efektif untuk mengatasi data imbalanced dalam klasifikasi teks.

6. Klasifikasi dengan *Support Vector Machine*.

Algoritma SVM memiliki kemampuan bekerja baik pada data linier maupun non-linier (R. F. Putra et al., 2024). Parameter SVM dioptimalkan menggunakan GridSearchCV untuk memperoleh kombinasi terbaik dari nilai C dan gamma.

7. Evaluasi Model.

Model dievaluasi menggunakan beberapa metrik, antara lain:

- a. *Confusion Matrix* untuk membandingkan hasil prediksi dengan data aktual.
- b. Akurasi untuk mengukur proporsi prediksi yang benar.

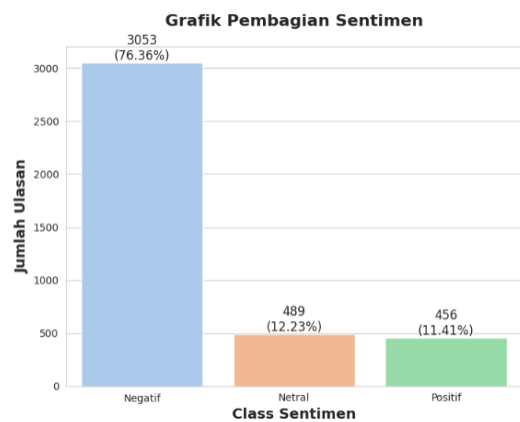
- c. *Precision*, *Recall*, dan *F1-Score* untuk mengukur performa per kelas.
- d. ROC-AUC untuk menilai kemampuan model dalam membedakan kelas sentimen (Muttaqin et al., 2023).

HASIL DAN PEMBAHASAN

1. Hasil Klasifikasi Sentimen.

Dari total 4000 ulasan yang diperoleh, hasil klasifikasi menunjukkan distribusi sebagai berikut:

- a. Negatif: 3.053 ulasan (76,36%)
- b. Positif: 489 ulasan (12,23%)
- c. Netral: 456 ulasan (11,41%)

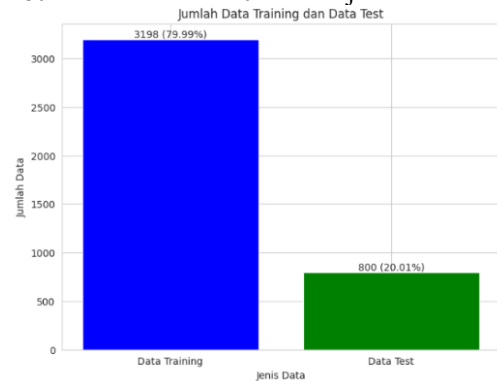


Sumber: Hasil Olah Data (2025)

Gambar 2. Distribusi Sentimen Hasil Klasifikasi

2. Pembagian Data Latih dan Uji.

Selanjutnya, pembagian data menjadi 2 proporsi, yakni 80% data latih dan 20% data uji.



Sumber: Hasil Olah Data (2025)

Gambar 3. Proporsi Distribusi Data

Solusi atas ketidakseimbangan kelas, terapkan SMOTE yang hanya pada data latih saja. Tujuannya supaya model belajar dari distribusi data yang *balance*, sementara data uji tetap mempertahankan kondisi aslinya agar evaluasi tetap objektif.

Sentimen	Sebelum SMOTE	Sesudah SMOTE
Negatif	2442	2442
Netral	391	2442
Positif	365	2442

Sumber: Hasil Olah Data (2025)
Gambar 4. Distribusi Data Latih Sebelum dan Sesudah SMOTE

Dengan pendekatan ini, model dapat belajar lebih baik mengenali kelas minoritas tanpa mengurangi validitas hasil pengujian.

3. Evaluasi Model SVM.

Model klasifikasi menggunakan *Support Vector Machine* (SVM) dengan parameter terbaik hasil GridSearchCV (kernel RBF, C=10, gamma=scale). Hasil evaluasi menunjukkan performa sebagai berikut:

- Akurasi: 91,50%
- Macro F1-score: 0,82
- ROC-AUC Makro: 0,9693

SVM Accuracy: 0.915

SVM Accuracy Percentage: 91.50%

SVM Classification Report:				
	precision	recall	f1-score	support
Negatif	0.95	0.98	0.96	611
Netral	0.72	0.59	0.65	98
Positif	0.85	0.86	0.85	91
accuracy			0.92	800
macro avg	0.84	0.81	0.82	800
weighted avg	0.91	0.92	0.91	800

Sumber: Hasil Olah Data (2025)
Gambar 5. Hasil Evaluasi Model SVM.

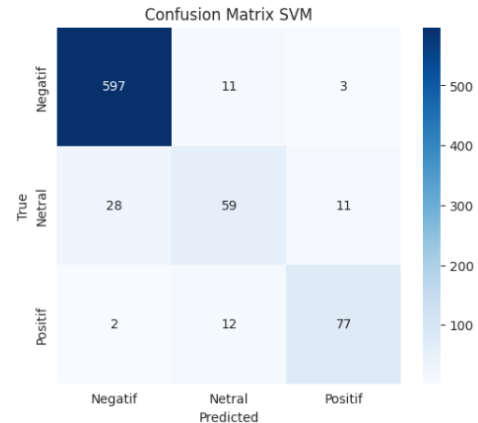
Detail per kelas diperoleh sebagai berikut,

- Negatif → Precision 0,95, Recall 0,98, F1-score 0,96
- Positif → Precision 0,85, Recall 0,86, F1-score 0,85
- Netral → Precision 0,72, Recall 0,59, F1-score 0,65

Hasil diatas terlihat dimana model mampu mengenali ulasan negatif dengan tingkat akurasi paling optimal, sedangkan ulasan netral masih menjadi tantangan karena sering bercampur dengan ekspresi positif maupun negatif.

4. Visualisasi Hasil Klasifikasi

Untuk memperjelas hasil, dilakukan beberapa bentuk visualisasi:

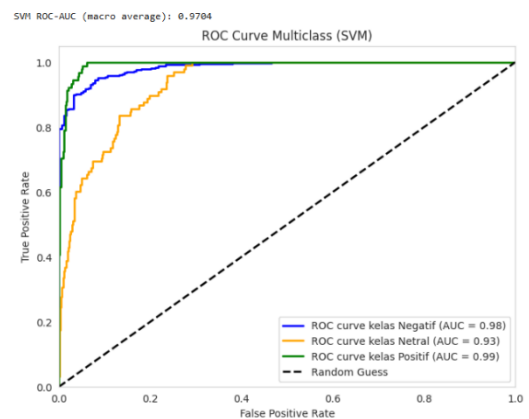


Sumber: Hasil Olah Data (2025)
Gambar 6. Hasil Klasifikasi SVM.

Pada Gambar 6 diatas, memperlihatkan detail prediksi model terhadap tiap kelas:

- Negatif: dari 611 data, sebanyak 597 terklasifikasi benar, hanya 14 salah (11 ke netral, 3 ke positif).
- Netral: dari 98 data, hanya 59 terklasifikasi benar, sementara 39 salah (28 ke negatif, 11 ke positif).
- Positif: dari 91 data, sebanyak 77 terklasifikasi benar, sisanya 14 salah (2 ke negatif, 12 ke netral).

Dari hasil ini terlihat bahwa model sangat baik mengenali kelas negatif dan positif, namun masih cukup banyak kesalahan pada kelas netral. Hal ini selaras dengan nilai recall netral yang rendah (0,59).



Sumber: Hasil Olah Data (2025)
Gambar 7. Kurva ROC-AUC Tiap Kelas

Pada Gambar 7 Kurva ROC menunjukkan kemampuan model dalam membedakan tiap kelas sentimen. Nilai AUC yang diperoleh adalah:

- Negatif: 0,98
- Netral: 0,93
- Positif: 0,99
- Macro Average: 0,9693

Nilai AUC diatas 0,90 pada semua kelas menunjukkan bahwa model memiliki kemampuan diskriminasi yang sangat baik. Kinerja paling kuat ada pada kelas positif (AUC 0,99), diikuti negatif

disarankan untuk mengoptimalkan pengolahan kelas netral yang masih sulit dibedakan, serta mempertimbangkan pendekatan hibrid yang menggabungkan machine learning tradisional dengan model *deep learning* agar hasil klasifikasi semakin presisi.

REFERENSI

- Angraini, J., & Alita, D. (2024). Implementasi Metode SVM Pada Sentimen Analisis Terhadap Pemilihan Presiden (Pilpres) 2024 Di Twitter. *Jurnal Informatika: Jurnal Pengembangan IT*, 9(2), 102–111. <https://doi.org/10.30591/jpit.v9i2.6560>
- Findawati, Y., & Rosid, M. A. (2020). *Buku Ajar Text Mining*. <https://press.umsida.ac.id/index.php/umsidapress/article/view/978-623-6833-19-3/770>
- Gifari, O. I., Adha, M., Rifky Hendrawan, I., Freddy, F., & Durrand, S. (2022). Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine. *JIFOTECH (JOURNAL OF INFORMATION TECHNOLOGY)*, 2(1).
- Iqrom, M., Afdal, Novita, R., Rahmawita, M., & Ahsyar, T. K. (2025). *Analisis Sentimen Aplikasi Gojek, Grab, Dan Maxim Menggunakan Algoritma Support Vector Machine*. 10(1), 2025. <https://jurnal.polbeng.ac.id/index.php/ISI/article/view/353/152>
- Muttaqin, Widiyanto, W. W., Munsarif, M., Mandias, G. F., Pungus, S. R., Widarman, A., Hapsari, W. J., Hardiyanti, S. A., Fatkhudin. Aslam, Pasnur, Bisono, E. F., Anshiro, M., Suryani, & Saputra, N. (2023). *Pengenalan Data Mining*. <https://repository.upy.ac.id/4946/1/FullBook%20Pengenalan%20Data%20Mining.pdf>
- Pratama, M. R., Ramadhan, Y. R., & Komara, M. A. (2023). *Analisis Sentimen BRImo dan BCA Mobile Menggunakan Support Vector Machine dan Lexicon Based*. <https://ojs.stmik-banjarbaru.ac.id/index.php/jutisi/article/viewFile/1431/892>
- Putra, M. B. D., & Setiawan, E. (2024). Metode Lexicon Based Untuk Analisis Sentimen Pengguna Twitter Terhadap Kinerja Isp (Studi Kasus: Indihome, Biznet, MyRepublic). In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Issue 6). <https://ejournal.itn.ac.id/index.php/jati/article/view/11765/6641>
- Putra, R. F., Mukhlis, I. R., Datya, A. I., Pipin, S. J., Reba, F., Al-Husaini, M., Mandowen, S. A., Zain, N. N. L. E., & Judijanto, L. (2024). *Algoritma Pembelajaran Mesin (Dasar, Teknik, Dan Aplikasi)*. <https://www.researchgate.net/publication/379479664>
- Ridwan, Hermaliani, E. H., & Ernawati, M. (2024). Penerapan Metode SMOTE Untuk Mengatasi Imbalanced Data Pada Klasifikasi Ujaran Kebencian. In *Computer Science (CO-SCIENCE)* (Vol. 4, Issue 1). <http://jurnal.bsi.ac.id/index.php/co-science>
- Rizky, A. M., Pratiwi, M. P., Chairunnisa, A., Aiko, I. A., & Ariesmansyah, A. (2025). E-Government: Meningkatkan Efisiensi dan Efektivitas Pelayanan Publik di Indonesia. *Andre Ariesmansyah INNOVATIVE: Journal Of Social Science Research*, 5. <https://j-innovative.org/index.php/Innovative/article/view/17827/11968>
- Sevilla, A., & Cahyono, N. (2025). Sentimen Analisis Pengguna Twitter Terhadap Provider XL Axiata Menggunakan Metode Support Vector Machine. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 9, Issue 3). <https://ejournal.itn.ac.id/index.php/jati/article/view/13647/7602>
- Tampubolon, M. (2023). *Metode Penelitian*. <http://repository.uki.ac.id/11609/1/MetodePenelitian.pdf>