

# Evaluasi Performa Naive Bayes Dan SVM Dalam Memprediksi Sentimen Ulasan Pengguna Aplikasi Pinterest

Oleh:

Indri Arianti Wibowo

Uce Indahyanti

Progam Studi

Universitas Muhammadiyah Sidoarjo

Februari, 2026

# Pendahuluan

Penggunaan Pinterest yang mencapai lebih dari 522 juta pengguna aktif bulanan pada tahun 2024 menghasilkan banyak ulasan yang memengaruhi persepsi terhadap kualitas aplikasi, sehingga diperlukan analisis sentimen untuk mengolah teks secara efisien. Penelitian ini membandingkan performa Naive Bayes dan Support Vector Machine (SVM) dalam mengklasifikasikan sentimen positif dan negatif pada ulasan Pinterest. Meskipun analisis sentimen pada aplikasi mobile telah banyak dilakukan, belum ada penelitian yang secara khusus membandingkan kedua algoritma pada ulasan Pinterest, sehingga penelitian ini dilakukan untuk mengisi celah tersebut. Proses meliputi pre-processing, ekstraksi fitur TF-IDF, penyeimbangan data menggunakan Random Oversampling (ROS), dan evaluasi menggunakan confusion matrix dengan metrik akurasi, presisi, recall, dan F1-score guna mendukung peningkatan kualitas aplikasi.

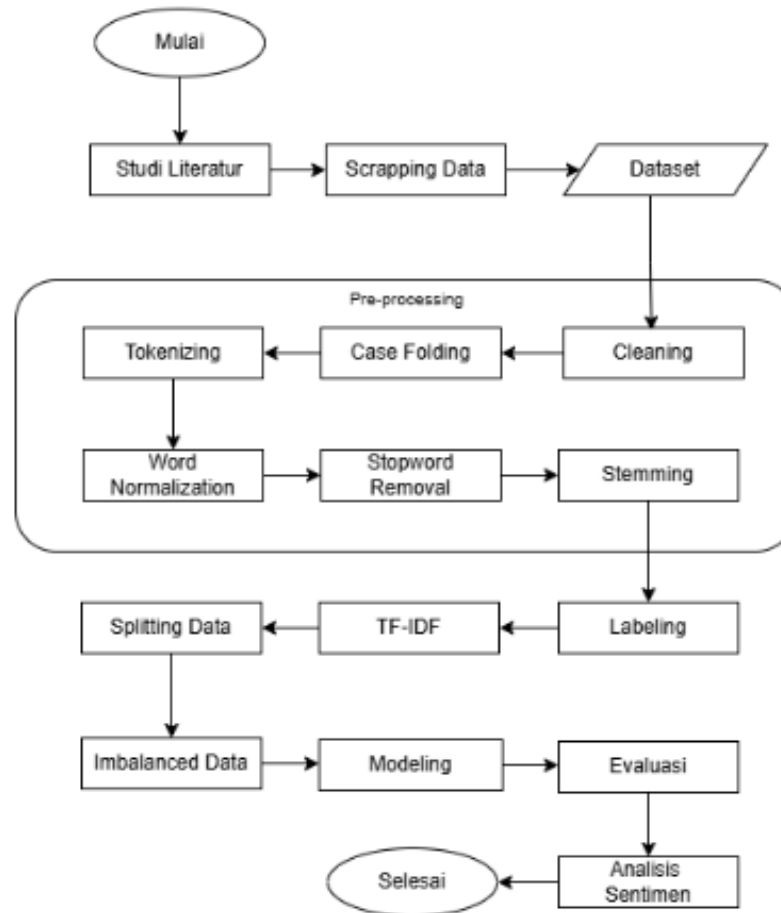
# Pertanyaan Penelitian (Rumusan Masalah)

1. Bagaimana penggunaan algoritma naive bayes dan SVM dalam mengklasifikasikan sentimen terkait ulasan pengguna aplikasi pinterest?
2. Bagaimana perbandingan performa algoritma naive bayes dan SVM dalam menganalisis sentimen ulasan pengguna aplikasi pinterest dengan menggunakan confusion matrix seperti akurasi, precision, recall, dan f1-score?

# Metode

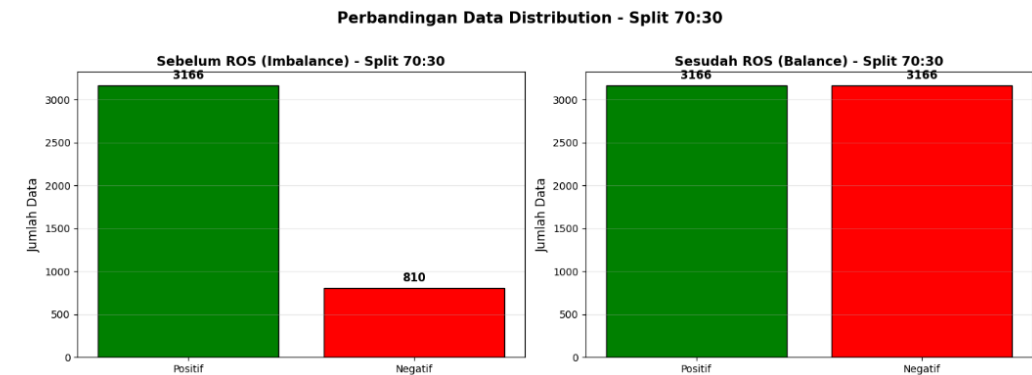
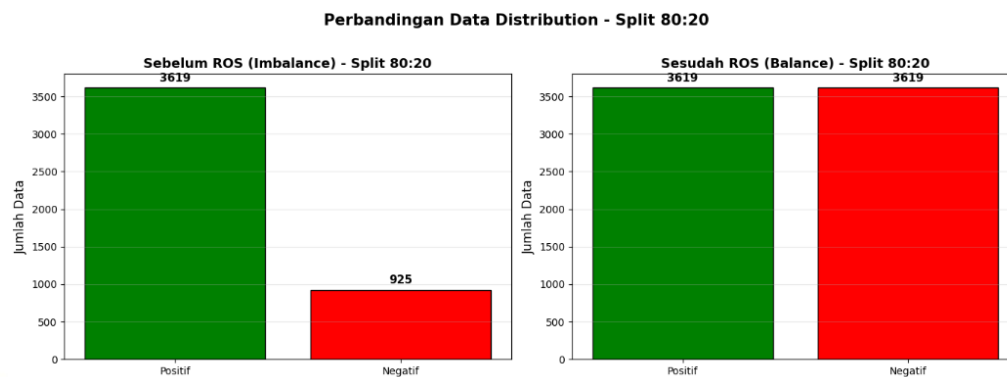
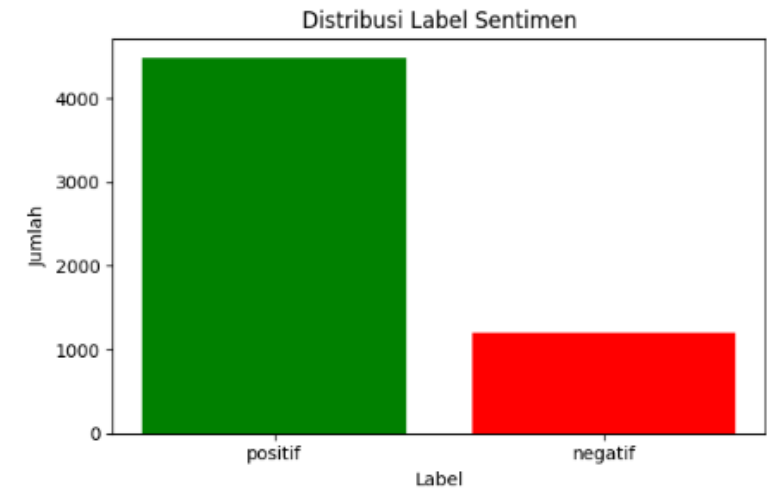
- Dataset: 10.000 ulasan pengguna aplikasi Pinterest pada Google Playstore (Januari 2020-Januari 2025). Setelah tahapan pre-processing menjadi 5.681 (khusus hanya ulasan berbahasa Indonesia)
- Pre-processing: cleaning, case folding, tokenizing, word normalization, stopwords removal, stemming.
- Labeling: rating 1-2 sebagai negatif, rating 4-5 sebagai positif, dan rating 3 digunakan metode lexicon-based.
- Ekstraksi fitur: TF-IDF (metode pembobotan kata untuk mengubah data teks menjadi bentuk numerik).
- Splitting data: rasio 80:20 dan 70:30.
- Imbalanced data: menggunakan penerapan Random Oversampling (ROS).
- Modeling: algoritma Naive Bayes dan Support Vector Machine (SVM)
- Evaluasi: confusion matrix yang meliputi akurasi, precision, recall, dan F1-score

# Metode



# Hasil

Hasil analisis menghasilkan ketidakseimbangan data ulasan aplikasi Pinterest, dengan kelas positif lebih dominan daripada kelas negatif. Sehingga, diterapkan penyeimbangan data dengan Random Oversampling (ROS) pada rasio 80:20 dan 70:30 untuk menyeimbangkan data latih agar lebih seimbang.



# Hasil

Pengujian dilakukan menggunakan dua algoritma yaitu Naive Bayes dan SVM dengan pembagian data menjadi rasio 80:20 dan 70:30 serta kondisi data latih yaitu imbalanced dan balanced. Tujuannya untuk menganalisis dan membandingkan performa Naive Bayes dan SVM dalam klasifikasi sentimen ulasan Pinterest serta mengevaluasi pengaruh pembagian data dan penerapan Random Oversampling (ROS) terhadap performa model, khususnya pada kelas negatif.

Rasio	Algoritma	Kondisi Train	Akurasi	Precision	Recall	F1-Score
80:20	Naive Bayes	Imbalanced	86,72%	86,07%	86,72%	85,30%
		Balanced (ROS)	85,31%	88,44%	85,31%	86,18%
	SVM	Imbalanced	89,01%	88,60%	89,01%	88,71%
		Balanced (ROS)	88,30%	88,89%	88,30%	85,54%
70:30	Naive Bayes	Imbalanced	86,22%	85,45%	86,22%	84,66%
		Balanced (ROS)	86,69%	89,41%	86,69%	87,43%
	SVM	Imbalanced	87,97%	88,59%	88,97%	88,70%
		Balanced (ROS)	88,21%	88,92%	88,21%	88,48%

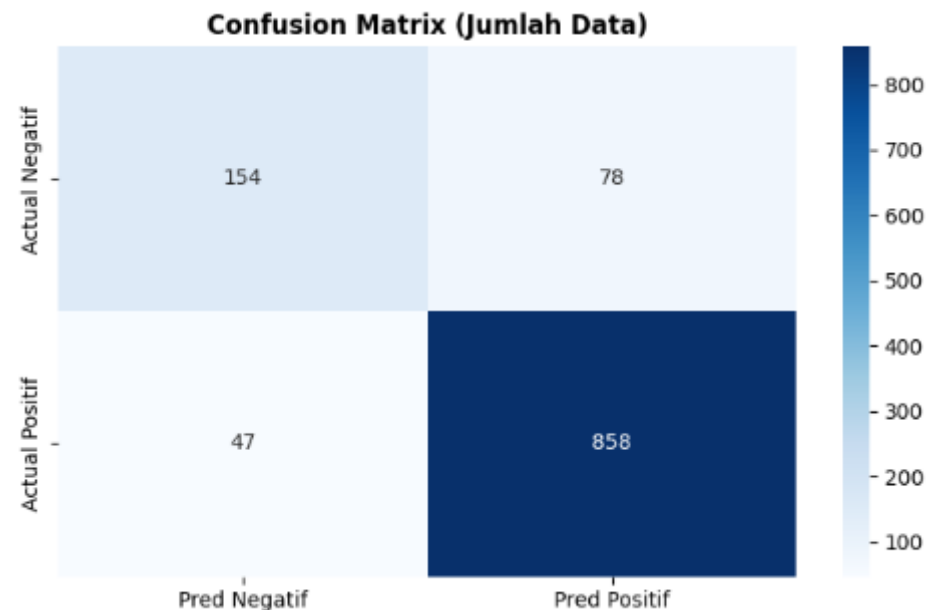
# Hasil

Penerapan Random Oversampling (ROS) meningkatkan recall kelas minoritas pada kedua algoritma, yaitu Naive Bayes dan SVM pada rasio 80:20 dan 70:30 dengan kondisi data imbalanced dan balanced.

Algoritma	Rasio Split	Kondisi Train	Recall	F1-score
Naive Bayes	80:20	Imbalanced	46,55%	58,86%
		Balanced (ROS)	85,34%	70,34%
	70:30	Imbalanced	44,38%	56,72%
		Balanced (ROS)	87,03%	72,68%
SVM	80:20	Imbalanced	66,38%	71,13%
		Balanced (ROS)	77,16%	72,91%
	70:30	Imbalanced	66,86%	77,17%
		Balanced (ROS)	77,81%	72,87%

# Hasil

*Confusion matrix* menggambarkan performa model dalam mengklasifikasikan kelas negatif dan positif. Model berhasil mengidentifikasi 154 data negatif dan 858 data positif secara benar. Namun, tetap ada kesalahan klasifikasi, yaitu 78 data negatif yang diidentifikasi sebagai positif (*false positive*) dan 47 data positif yang diidentifikasi sebagai negatif (*false negative*).



# Pembahasan

Algoritma *Naive Bayes* dan *Support Vector Machine (SVM)* memiliki keunggulan masing-masing dalam menangani data ulasan yang tidak seimbang. *Naive Bayes* dengan penerapan ROS rasio 70:30 menjadi model terbaik untuk deteksi sentimen negatif karena menghasilkan *recall* kelas negatif tertinggi, yang meningkat signifikan yaitu 87,03%, sedangkan *Support Vector Machine (SVM)* meraih akurasi global paling tinggi yaitu 89,01% pada rasio pembagian data 80:20. Namun, memiliki *recall* kelas negatif yang lebih rendah. Kontribusi ini mencakup evaluasi performa berbasis metrik per kelas, analisis dampak ketidakseimbangan data, serta *trade-off* antara akurasi global dan *recall* kelas negatif.

# Temuan Penting Penelitian

- Algoritma SVM pada rasio 80:20 menghasilkan akurasi global tertinggi dengan hasil akurasi 89,01% dan F1-score 88,71%.
- Algoritma Naive Bayes setelah penerapan ROS pada rasio 70:30 menghasilkan kelas recall terbaik untuk deteksi sentimen negatif menghasilkan recall sebesar 87,03% dan F1-score 72,68%.
- Penerapan ROS dapat meningkatkan recall kelas negatif secara signifikan.
- Terjadinya trade-off pada penurunan akurasi global dan meningkatnya deteksi sentiment negatif.
- Membuktikan bahwa ROS penting pada klasifikasi data yang tidak seimbang.

# Manfaat Penelitian

Penelitian ini memberikan gambaran dalam bidang analisis sentimen dengan membandingkan performa algoritma Naive Bayes dan Support Vector Machine (SVM) pada ulasan aplikasi Pinterest di Google Play Store, khususnya pada kondisi data tidak seimbang. Penelitian ini juga menunjukkan pentingnya penanganan ketidakseimbangan data melalui Random Oversampling (ROS) serta menegaskan bahwa evaluasi model perlu memperhatikan metrik per kelas, terutama recall dan F1-score kelas negatif. Selain itu, hasil penelitian ini dapat membantu memahami persepsi dan permasalahan pengguna serta menjadi referensi bagi pengembangan penelitian selanjutnya di bidang analisis sentimen.

# Referensi

1. B. of Apps, "Pinterest Revenue and Usage Statistics," Business of Apps. Accessed: Jan. 02, 2026. [Online]. Available: <https://www.businessofapps.com/data/pinterest-statistics/>
2. B. Uddin, D. D. F. Basam, and E. M. L. Wijayadi, "Efektivitas Pemanfaatan Pinterest Terhadap Kreativitas Pengguna," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 7, no. 4, pp. 678–684, 2024.
3. A. Nurian and B. N. Sari, "Analisis Sentimen Ulasan Pengguna Aplikasi Google Play Menggunakan Naïve Bayes," *J. Inform. dan Tek. Elektro Terap.*, vol. 11, no. 3s1, pp. 829–835, 2023, doi: 10.23960/jitet.v11i3s1.3348.
4. Y. T. R. Sholiha, L. A. M. Nabilah, and Imron, "Analisis Sentimen Aplikasi Liputan6.Com pada Ulasan Pengguna di Google Playstore dengan Menggunakan Algoritma Support Vector Machine (Svm) dan Naïve Bayes," *Saturnus J. Teknol. dan Sist. Inf.*, vol. 3, no. 3, pp. 30–51, 2025, doi: <https://doi.org/10.61132/saturnus.v3i3.867>.
5. P. Anggraini and Winarsih, "KOMPARASI NAÏVE BAYES, SUPPORT VECTOR MACHINE, DAN RANDOM FOREST DALAM ANALISIS SENTIMEN APLIKASI SHOPEE DI GOOGLE PLAY STORE," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 3, pp. 4451–4457, 2025.
6. A. Lowell, A. Lowell, K. Candra, and E. Indra, "Perbandingan Metode Support Vector Machine (SVM) Dan Naive Bayes Pada Analisis Sentimen Ulasan Aplikasi OVO," *J. MEDIA Inform. [ JUMIN ]*, vol. 6, no. 2, pp. 896–905, 2025.
7. F. Damayanti, A. Rahim, and N. A. Verdikha, "ANALISIS SENTIMEN ULASAN PENGGUNA TERHADAP APLIKASI K24KLIK," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 2, pp. 3224–3230, 2025.
8. J. A. Putra, A. Dharmawan, and J. Gondohanindijo, "SENTIMEN ANALISIS APLIKASI DIGITALENT MOBILE MENGGUNAKAN NAIVE BAYES DAN SVM DENGAN EKSTRAKSI FITUR TF-IDF," *J. Inf. Technol. Comput. Sci.*, vol. 7, 2024.

# Referensi

9. A. Nabilla M. P., A. Adrian Y. L., A. Indra J., Y. Arya S., Sumanto, and A. Diah H., "Komparasi Naive Bayes dan SVM untuk Analisis Sentimen Pada E-Commerce Seller Center," *J. Sains Dan Teknol.*, vol. 5, no. 3, pp. 199–208, 2025.
10. M. Aji, A. Maldini, and S. Andryana, "ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI PERBANKAN," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 3, pp. 4098–4105, 2025.
11. M. I. Prayugah, U. Indahyanti, and N. Ariyanti, "ANALISIS SENTIMEN PUBLIK PADA PEMERINTAH DALAM SERANGAN RANSOMWARE DENGAN PENDEKATAN SMOTE," *JOISIE (Journal Inf. Syst. Informatics Eng.*, vol. 8, no. 2, pp. 333–343, 2024.
12. M. D. Qur'ani, H. Setiawan, and I. A. Kautsar, "PENERAPAN METODE SUPPORT VECTOR MACHINE (SVM) UNTUK MEMPREDIKSI PEMILIHAN KARIR BAGI ALUMNI UMSIDA," *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 10, no. 4, pp. 2906–2916, 2025.
13. A. R. Annjani, U. Indahyanti, A. W. Azinar, and R. Dijaya, "PENERAPAN ALGORITMA MACHINE LEARNING UNTUK MEMPREDIKSI SENTIMEN KEPUASAN PENGGUNA WEVERSE," *JATI(Jurnal Mhs. Tek. Inform.*, vol. 10, no. 1, pp. 784–790, 2026.
14. R. A. P. Sari, S. Kacung, and B. Santoso, "ANALISIS SENTIMEN LAYANAN KESEHATAN BPJS MENGGUNAKAN METODE SVM," *J. Inform. Teknol. dan Sains*, vol. 7, no. 2, pp. 878–885, 2025.
15. S. A. S. Mola, D. L. B. Baun, I. O. Nunes, and M. M. A. R. Sani, "ANALISIS SENTIMEN APLIKASI HALO BCA DI GOOGLE PLAY STORE MENGGUNAKAN METODE NAIVE BAYES , SUPPORT VECTOR MACHINE DAN RANDOM FOREST PENDAHULUAN," *HOAQ J. Teknol. Inf.*, vol. 15, no. 2, pp. 69–79, 2024.

# Referensi

16. D. Fitriono, R. Indriati, and A. Ristyawan, "Analisis Sentimen Ulasan Aplikasi Chatgpt Menggunakan Algoritma Support Vector Machine dan Lexicon Based," *JSITIK*, vol. 3, no. 2, pp. 101–111, 2025.
17. D. N. N. Husnina, D. E. Ratnawati, and B. Rahayudi, "Analisis Sentimen Pengguna Aplikasi RedBus berdasarkan Ulasan di Google Play Store menggunakan Metode Naïve Bayes," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 2, 2023.
18. D. W. Bhatara and R. R. Suryono, "ANALISIS SENTIMEN APLIKASI BCA MOBILE MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN SUPPORT VECTOR MACHINE," *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 9, no. 4, pp. 1907–1917, 2024.
19. F. A. Soraya and A. D. Indriyanti, "Perbandingan Algoritma Naïve Bayes dan Support Vector Machine dalam Analisis Sentimen Aplikasi Teman Bus," *JEISBI (Journal Emerg. Inf. Syst. Bus. Intell.*, vol. 05, no. 02, pp. 118–125, 2024.
20. N. D. H. Sadikin and S. Susanti, "Analisis Sentimen Publik Terhadap Kampanye Pengurangan Sampah Plastik Menggunakan Algoritma Naïve Bayes," *J. FASILKOM*, vol. 15, no. 2, pp. 202–212, 2025.
21. A. Y. Lubisa and M. Y. H. Setyawan, "Analisis Sentimen Aplikasi Pospay di PlayStore Menggunakan Algoritma Support Vector Machine & Naive Bayes," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 6, no. 3, pp. 514–521, 2024.
22. Fornieli, "kamus\_slang," Kaggle. Accessed: Feb. 26, 2026. [Online]. Available: <https://www.kaggle.com/datasets/fornigulo/kamus-slag>
23. H. P. Jelita, M. I. Saad, and Wahyuni, "Penerapan Algoritma Naïve Bayes Dalam Analisis Sentimen Masyarakat Terhadap STMIK Widya Cipta Dharma," *Bull. Inf. Technol.*, vol. 6, no. 2, pp. 148–160, 2025, doi: 10.47065/bit.v5i2.2029.

