

MODEL PREDIKSI PRESEPSI PENGGUNA DUOLINGO BERDASARKAN KOMENTAR DIGITAL

Oleh:

Brigide Tirenia Loresta

Uce Indahyanti

Progam Studi

Universitas Muhammadiyah Sidoarjo

Februari, 2026

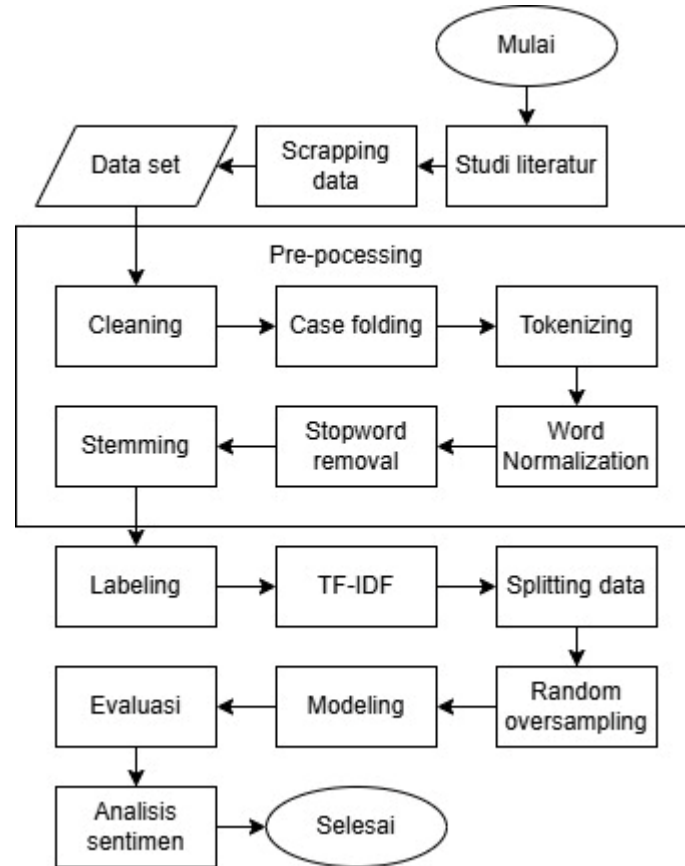
Pendahuluan

Duolingo merupakan aplikasi pembelajaran bahasa populer dengan lebih dari 50 juta pengguna aktif harian dan sekitar 130 juta pengguna aktif bulanan pada tahun 2025, serta menyediakan fitur interaktif yang mendukung proses belajar secara fleksibel dan efektif. Ulasan pengguna di Google Play Store dapat dimanfaatkan melalui analisis sentimen untuk mengetahui persepsi terhadap kualitas aplikasi menggunakan pendekatan machine learning dan NLP. Penelitian ini menggunakan algoritma Support Vector Machine (SVM) dan Decision Tree untuk mengklasifikasikan sentimen positif dan negatif dari data komentar, dengan tahapan preprocessing, TF-IDF, serta penanganan data tidak seimbang menggunakan Random Oversampling. Hasil penelitian diharapkan dapat membandingkan performa kedua algoritma dan memberikan masukan bagi pengembang dalam meningkatkan kualitas aplikasi.

Pertanyaan Penelitian (Rumusan Masalah)

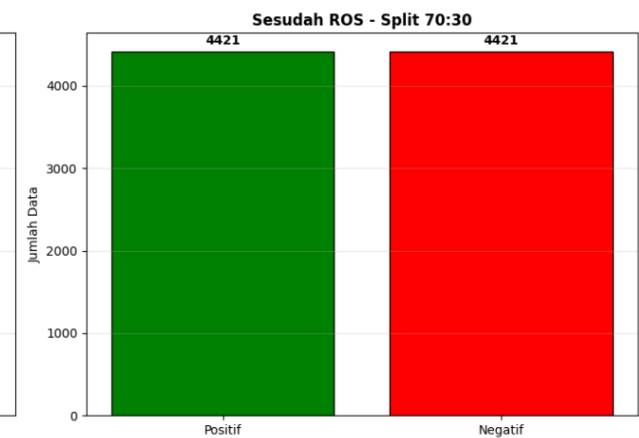
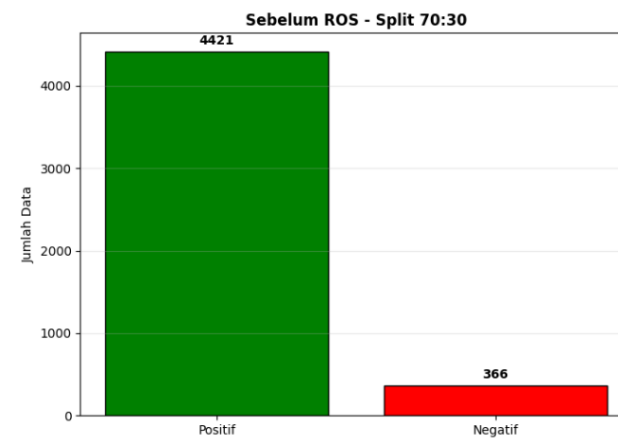
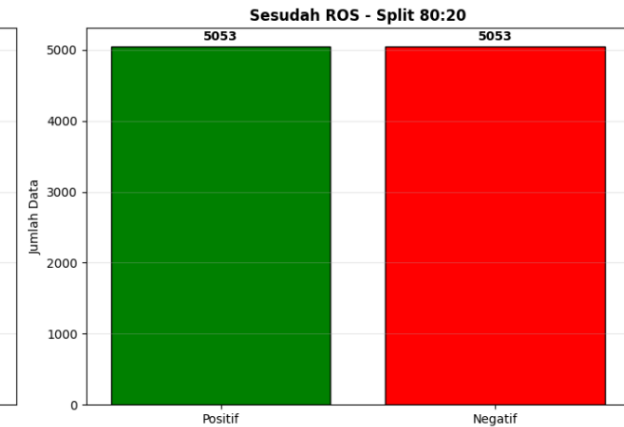
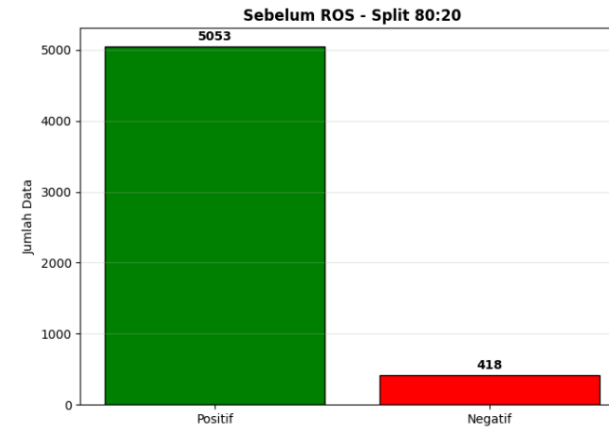
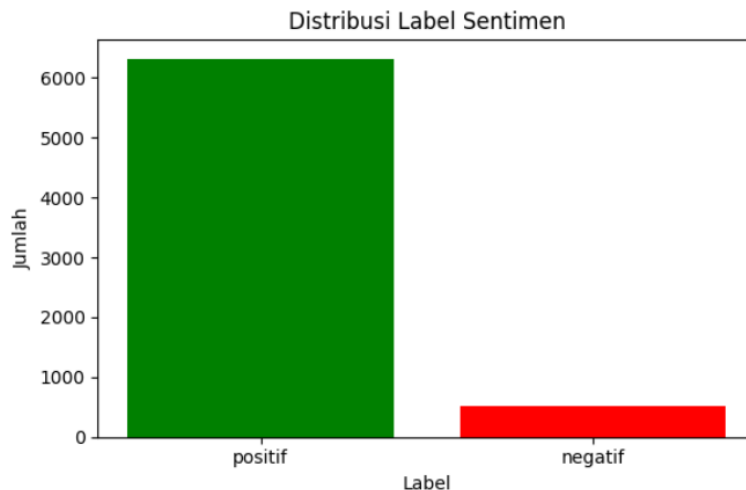
1. Bagaimana menerapkan algoritma SVM dan Decision Tree untuk mengklasifikasikan ulasan pengguna aplikasi duolingo di google playstore?
2. Bagaimana menerapkan confusion matrix untuk mengevaluasi algoritma untuk analisis sentimen diatas?

Metode



Hasil

Hasil analisis menghasilkan ketidakseimbangan data ulasan aplikasi Duolingo, 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.



Pembahasan

Hasil pengujian algoritma dilakukan untuk mengevaluasi kinerja *Support Vector Machine* (SVM) dan *Decision Tree* dalam mengklasifikasikan sentimen ulasan pengguna *Duolingo*. Pengujian dilakukan menggunakan dua skenario rasio pembagian data, yaitu 80:20 dan 70:30, dengan kondisi data latih *imbalance* dan *balance* menggunakan *Random Oversampling* (ROS). Tujuannya adalah untuk membandingkan kinerja SVM dan *Decision Tree*.

Rasio	Algoritma	Kondisi Train	Akurasi	Precision	Recal	F1-Score
80:20	SVM	Imbalanced	93.57%	92.52%	93.57%	92.11%
		Balanced (ROS)	92.76%	92.99%	92.76%	92.87%
	Decision Tree	Imbalanced	91.30%	90.17%	91.30%	90.65%
		Balanced (ROS)	90.35%	90.85%	90.35%	90.59%
70:30	SVM	Imbalanced	93.13%	91.69%	93.13%	91.72%
		Balanced (ROS)	92.74%	92.97%	92.74%	92.85%
	Decision Tree	Imbalanced	91.86%	91.36%	91.86%	91.59%
		Balanced (ROS)	90.25%	91.30%	90.25%	90.73%

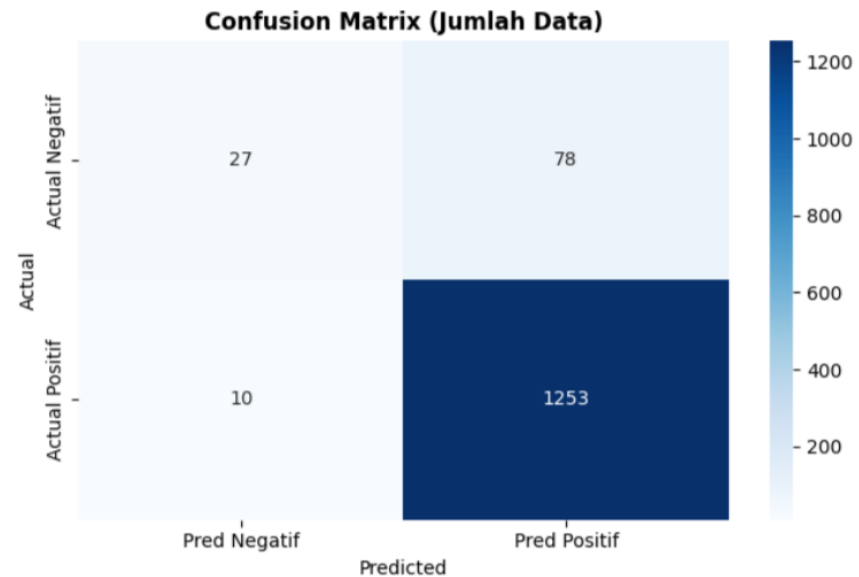
Hasil

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
SVM	80:20	Imbalanced	25.71%	38.03%
		Balanced (ROS)	56.19%	54.38%
	70:30	Imbalanced	24.84%	35.62%
		Balanced (ROS)	56.05%	54.15%
Decision Tree	80:20	Imbalanced	29.52%	34.25%
		Balanced (ROS)	42.86%	40.54%
	70:30	Imbalanced	40.13%	43.00%
		Balanced (ROS)	48.41%	43.18%

Hasil

Confusion matrix menggambarkan performa model dalam mengklasifikasikan kelas negatif dan positif. Model berhasil mengidentifikasi 27 data negatif dan 1.253 data positif secara benar. Namun, masih terdapat kesalahan klasifikasi, yaitu 78 data negatif yang diidentifikasi sebagai positif (false positive) dan 10 data positif yang diidentifikasi sebagai negatif (false negative).



Pembahasan

Pembahasan menunjukkan bahwa model Support Vector Machine (SVM) pada pembagian data 80:20 dengan kondisi imbalance menghasilkan akurasi tertinggi sebesar 93,57%, yang dipengaruhi oleh dominasi prediksi benar pada kelas positif. Namun, model masih memiliki kelemahan dalam mengenali sentimen negatif yang ditunjukkan oleh tingginya kesalahan klasifikasi (false positive), sehingga recall pada kelas negatif rendah. Untuk mengatasi hal tersebut, diterapkan Random Oversampling (ROS) guna menyeimbangkan data latih. Meskipun penerapan ROS menurunkan akurasi secara keseluruhan, metode ini terbukti meningkatkan kemampuan model dalam mendeteksi kelas minoritas, khususnya sentimen negatif. Oleh karena itu, SVM dengan ROS dianggap lebih optimal karena mampu memberikan keseimbangan antara akurasi dan sensitivitas terhadap kelas minoritas.

Temuan Penting Penelitian

1. SVM memiliki performa terbaik dibandingkan Decision Tree pada seluruh skenario pengujian.
2. Akurasi tertinggi sebesar 93,57% diperoleh oleh SVM pada rasio data 80:20 dalam kondisi imbalance.
3. Dataset mengalami ketidakseimbangan kelas, dengan dominasi sentimen positif (6.316 data) dibandingkan negatif (523 data).
4. Model tanpa penyeimbangan cenderung bias ke kelas mayoritas, sehingga recall pada kelas negatif rendah.
5. Penerapan Random Oversampling (ROS) berhasil menyeimbangkan data latih dan meningkatkan performa pada kelas minoritas.
6. Recall sentimen negatif meningkat signifikan (dari sekitar 25% menjadi $\pm 56\%$) setelah menggunakan ROS.
7. Terjadi trade-off kinerja, yaitu akurasi sedikit menurun tetapi kemampuan mendeteksi sentimen negatif meningkat.
8. Model terbaik secara fungsional adalah SVM dengan ROS, karena lebih seimbang dalam mengklasifikasikan kedua kelas.
9. Ulasan positif didominasi oleh kemudahan dan efektivitas aplikasi, sedangkan ulasan negatif banyak menyoroti iklan dan penurunan kualitas aplikasi.

Manfaat Penelitian

Penelitian ini memberikan manfaat berupa pengembangan model prediksi sentimen pengguna Duolingo yang tidak hanya akurat tetapi juga seimbang dalam mengklasifikasikan ulasan positif dan negatif. Hasil penelitian ini dapat membantu pengembang dalam memahami persepsi pengguna berdasarkan komentar digital sehingga dapat digunakan sebagai bahan evaluasi untuk meningkatkan kualitas dan fitur aplikasi. Selain itu, penelitian ini menunjukkan efektivitas algoritma Support Vector Machine (SVM) dalam analisis sentimen serta pentingnya penanganan data tidak seimbang menggunakan metode Random Oversampling (ROS). Penelitian ini juga dapat menjadi referensi bagi penelitian selanjutnya serta mendukung pengambilan keputusan berbasis data dari opini pengguna.

Referensi

1. I. Duolingo, “Duolingo surpasses 50 million daily active users and grows revenue in 2025,” Duolingo Investor Relations. [Online]. Available: <https://investors.duolingo.com>
2. Y. Finance, “Duolingo surpasses 50 million daily active users,” Yahoo Finance. [Online]. Available: <https://finance.yahoo.com/news/duolingo-surpasses-50-million-daily-210100511.html>
3. J. Empati, T. Salsabila, N. Nafilah, F. Patangga, S. Zulfa, and N. Listyaningsih, “LITERATURE REVIEW : EFEKTIVITAS PENGGUNAAN APLIKASI DUOLINGO TERHADAP MOTIVASI BELAJAR,” vol. 13, pp. 302–312, 2024.
4. Friska Aditia Indriyani, Ahmad Fauzi, and Sutan Faisal, “Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine,” *TEKNOSAINS J. Sains, Teknol. dan Inform.*, vol. 10, no. 2, pp. 176–184, 2023, doi: 10.37373/tekno.v10i2.419.
5. S. Putri, A. Ati, P. Muharani, and R. Qori, “Sentiment Analysis of Gojek User Reviews using TF-IDF and Machine Learning in Orange Platform,” vol. 6, no. 4, pp. 296–305, 2025.
6. E. F. Laili *et al.*, “KOMPARASI ALGORITMA DECISION TREE DAN SUPPORT VECTOR MACHINE (SVM) DALAM,” vol. 8, no. 1, pp. 67–76, 2025.
7. F. Alifiana, M. F. Asnawi, I. A. Ihsannudin, M. A. M. Baihaqy, and D. Asmarajati, “Analisis Sentimen Aplikasi Duolingo Menggunakan Algoritma Naïve Bayes Dan Support Machine Learning,” *Device*, vol. 13, no. 2, pp. 223–230, 2023, doi: 10.32699/device.v13i2.5905.

Referensi

8. K. A. Rokhman, B. Berlilana, and P. Arsi, “Perbandingan Metode Support Vector Machine Dan Decision Tree Untuk Analisis Sentimen Review Komentar Pada Aplikasi Transportasi Online,” *J. Inf. Syst. Manag.*, vol. 3, no. 1, pp. 1–7, 2021, doi: 10.24076/joism.2021v3i1.341.
9. Z. Abidin, T. Suratno, and M. F. Putri, “PENERAPAN RANDOM OVERSAMPLING DAN PRINCIPAL COMPONENT ANALYSIS UNTUK MENINGKATKAN AKURASI PREDIKSI KEBANGKRUTAN APPLICATION OF RANDOM OVERSAMPLING AND PRINCIPAL COMPONENT ANALYSIS TO ENHANCE THE ACCURACY OF BANKRUPTCY PREDICTION FOR,” vol. 12, no. 5, 2025.
10. T. A. Anastasya, A. Diani, P. Saka, M. Juventus, D. Deke, and A. M. Rizki, “OPTIMASI ALGORITMA SVM DENGAN PSO UNTUK ANALISIS SENTIMEN PADA MEDIA SOSIAL X TERHADAP KINERJA TIMNAS DI ERA SHIN TAE-YONG,” vol. 9, no. 1, pp. 384–391, 2025.
11. M. I. Prayugah, U. Indahyanti, and N. Ariyanti, “Analisis sentimen publik pada pemerintah dalam serangan ransomware dengan pendekatan smote,” vol. 8, no. 2, pp. 333–343, 2024.
12. U. Indahyanti *et al.*, “Perbandingan Algoritma Machine Learning dalam,” vol. 11, no. 2, pp. 96–105, 2025.
13. Z. Y. Burnama, M. A. Rosid, and N. L. Azizah, “Analisis Sentimen Pada Komentar Youtube Dalam Turnamen MPL Season 13 Dengan Metode Ensemble Machine Learning Sentiment Analysis on YouTube Comments in MPL Season 13 Tournament Using Ensemble Machine Learning Method”.

Referensi

14. A. Zhahrina, U. Sofiah, D. Wahyu, A. Andayani, and N. Khairurrabbani, “Penerapan Metode Support Vector Machine (SVM) dalam Analisis Sentimen Ulasan Aplikasi Tokopedia di Google Play Store,” no. September 2025, pp. 84–91.
15. A. Okta, K. Adi, A. Sanjaya, and A. B. Setiawan, “Penerapan Inset Lexicon untuk Analisis Sentimen Penonton Video JKT48 di YouTube,” vol. 9, pp. 1276–1283.
16. J. Penerapan, T. Informasi, D. S. Sarassati, S. Yulianto, J. Prasetyo, and B. Rob, “IT-EXPLORE Analisis sentimen terhadap dampak banjir rob dengan menggunakan metode Support Vector Machine,” vol. 04, pp. 233–244, 2025, doi: 10.24246/itexplore.v4i2.2025.pp233-244.
17. R. A. P. Sari, S. Kacung, and B. Santoso, “ANALISIS SENTIMEN LAYANAN KESEHATAN BPJS MENGGUNAKAN METODE SVM,” pp. 878–885, 2025.
18. D. Fitriono, R. Indriati, and A. Ristyawan, “Analisis Sentimen Ulasan Aplikasi Chatgpt Menggunakan Algoritma Support Vector Machine dan Lexicon Based,” vol. 3, no. 2, pp. 101–111, 2025.
19. S. V. M. Tf-idf, “Analisis Sentimen terhadap RSUD Salatiga Menggunakan Abstrak,” vol. 6, no. 1, pp. 478–489, 2025.
20. F. M. Fadillah, Y. Cahyana, and A. Fauzi, “BULLETIN OF COMPUTER SCIENCE RESEARCH Analisis Sentimen Masyarakat Terhadap Pembatasan BBM Peralite Menggunakan Random Forest dan K-Nearest Neighbor,” vol. 5, no. 4, pp. 340–352, 2025, doi: 10.47065/bulletincsr.v5i4.547.

