JMI, 13 (2) (2026) pp. 98-104

Published by: Institute of Computer Science (IOCS)

JMI (Jurnal Manajemen Informatika) Medicom

Journal homepage: www. journals.joninstitute.org/index.php/jmi

Optimasi *Preprocessing* dengan SMOTE pada TF-IDF dan LSTM Untuk Analisis Sentimen Ulasan Gopay

Muslimah[1], Novi Wulandari [2], Roni Anagora [3]

[1]Program Studi Ilmu Komputer, STMIK Al Muslim, Bekasi, Indonesia

[2,3]Program Studi Manajemen Informatika, STMIK Al Muslim, Bekasi, Indonesia

A R T I C L E I N F O

Article history:

Received Sep 17, 2026
Revised Oct 15, 2026
Accepted Nov 10, 2026

Keywords:

Analisis Sentimen
GoPay
TF-IDF
LSTM
SMOTE
Streamlit

ABSTRAK

Analisis sentimen pada ulasan aplikasi *GoPay* diperlukan untuk mengetahui tanggapan pengguna terhadap layanan yang digunakan. Ketidakseimbangan jumlah data pada kelas sentimen dapat menyebabkan model lebih dominan mempelajari kelas mayoritas sehingga kemampuan klasifikasi terhadap kelas tertentu menjadi kurang optimal. Penelitian ini bertujuan menerapkan *Synthetic Minority Over-sampling Technique* (SMOTE) serta mengevaluasi penggunaan *Term Frequency-Inverse Document Frequency* (TF-IDF) dan *Long Short-Term Memory* (LSTM) dalam meningkatkan kinerja klasifikasi sentimen ulasan *GoPay*. Evaluasi dilakukan menggunakan dataset GoPay App Reviews Bahasa Indonesia sebanyak 10.000 ulasan yang terdiri dari 7.000 ulasan positif dan 3.000 ulasan negatif. Tahapan penelitian meliputi *preprocessing*, pembagian data dengan perbandingan 80:20, penerapan SMOTE pada data *training*, pembentukan model, dan evaluasi menggunakan *accuracy, precision*, *recall*, dan *F1-score*. Pengujian dilakukan melalui tiga optimasi, yaitu TF-IDF + SMOTE, LSTM + SMOTE, serta TF-IDF + SMOTE + LSTM. Hasil *validation internal* menunjukkan akurasi masing-masing sebesar 98,62%, 98,38%, dan 98,56%. Model terbaik diperoleh pada TF-IDF + SMOTE dengan *Logistic Regression* dan menghasilkan *accuracy* sebesar 98,05% pada *final testing*. Model tersebut kemudian diimplementasikan dalam aplikasi web berbasis Streamlit untuk melakukan klasifikasi sentimen pada ulasan baru. Hasil penelitian menunjukkan bahwa penerapan SMOTE pada TF-IDF mampu menghasilkan kinerja klasifikasi yang tinggi dalam menganalisis sentimen ulasan *GoPay*.

This is an open access article under the CC BY-NC license.

Corresponding Author:

Muslimah
Program Studi Ilmu Komputer, STMIK Al Muslim
Bekasi, Indonesia.
Email: muslimah@almuslim.ac.id

1. PENDAHULUAN

Ulasan pada aplikasi pembayaran digital menjadi salah satu sumber informasi penting dalam mengevaluasi kualitas layanan aplikasi. Pada *platform Google Play Store,* seseorang dapat memberikan berbagai opini, kritik, maupun saran terhadap performa aplikasi yang digunakan. Salah satu aplikasi pembayaran digital dengan jumlah ulasan yang terus meningkat adalah *GoPay.* Ulasan tersebut mencerminkan pengalaman pengguna terhadap berbagai layanan seperti transaksi pembayaran, transfer saldo, atau stabilitas sistem aplikasi. Namun, jumlah ulasan yang sangat besar menyebabkan proses analisis secara manual menjadi sulit dilakukan karena membutuhkan waktu yang lama dan hasil yang kurang konsisten (Husain et al., 2024).

Untuk mengatasi permasalahan tersebut, analisis sentimen menjadi pendekatan yang banyak digunakan dalam pengolahan data teks. Analisis sentimen bertujuan mengelompokkan opini pengguna ke dalam kategori positif, negatif, atau netral secara otomatis. Dalam implementasinya, metode ini mampu membantu perusahaan memahami respon pengguna terhadap kualitas layanan aplikasi secara lebih cepat dan terukur. Proses klasifikasi sentimen pada data ulasan masih memiliki masalah, terutama pada distribusi data yang tidak seimbang (*Imbalanced class*). Pada umumnya, jumlah ulasan positif lebih dominan dibandingkan ulasan negatif menjadikan model klasifikasi cenderung lebih fokus mempelajari kelas mayoritas dan mengalami penurunan performa (Nurrahman et al., 2025).

Kondisi ketidakseimbangan data tersebut menyebabkan model menghasilkan nilai akurasi yang tinggi tetapi kurang optimal dalam mengenali sentimen negatif. Permasalahan ini menjadi tantangan penting dalam analisis sentimen karena opini negatif justru sering memuat informasi utama mengenai keluhan dan kekurangan layanan aplikasi. Oleh karena itu, diperlukan teknik *preprocessing* yang mampu menyeimbangkan distribusi data agar proses klasifikasi dapat berjalan lebih optimal. Salah satu metode yang banyak digunakan untuk menangani permasalahan tersebut adalah SMOTE (*Synthetic Minority Over-sampling Technique*). Metode ini bekerja dengan membangkitkan data sintetis pada kelas minoritas sehingga distribusi data menjadi lebih seimbang dan model dapat mempelajari karakteristik data secara lebih baik (Candra et al., 2024).

Selain permasalahan *imbalance* data, pemilihan metode representasi fitur dan model klasifikasi juga memengaruhi performa analisis sentimen. Salah satu metode representasi fitur teks yang banyak digunakan adalah TF-IDF (*Term Frequency-Inverse Document Frequency*). TF-IDF bekerja dengan memberikan bobot pada kata berdasarkan tingkat kemunculannya pada dokumen sehingga mampu merepresentasikan informasi penting dari suatu teks. Metode ini dikenal memiliki proses komputasi yang ringan dan cukup efektif pada berbagai penelitian analisis sentimen. Penelitian oleh (Harmandini & Muslim, 2024) menunjukkan bahwa metode TF-IDF mampu meningkatkan kualitas representasi fitur pada klasifikasi sentimen data ulasan produk. Selain itu, metode TF-IDF mampu memberikan performa klasifikasi yang cukup baik pada data ulasan aplikasi digital (Kadir & Fairuzabadi, 2025).

Di sisi lain, perkembangan metode *Deep Learning* mendorong penggunaan *Long Short-Term Memory* (LSTM) dalam analisis sentimen. LSTM merupakan pengembangan dari *Recurrent Neural Network* (RNN) yang dirancang untuk memahami hubungan antar kata dalam suatu kalimat. Kemampuan tersebut membuat LSTM dinilai lebih baik dalam menangkap konteks kalimat dibandingkan metode konvensional. Informasi dari penelitian oleh (Mutmainah et al., 2024) menunjukkan bahwa metode LSTM mampu memberikan hasil klasifikasi yang baik terhadap data ulasan pengguna aplikasi. Hasil yang sama juga ditunjukkan pada penelitian (Putri & Suryanto, 2025) dan (Saputra et al., 2026) menjelaskan bahwa metode LSTM mampu memahami pola sentimen pada data teks secara lebih baik.

Beberapa penelitian sebelumnya telah membahas penggunaan TF-IDF maupun LSTM pada analisis sentimen aplikasi digital. Penelitian oleh (Tholib & Arifin, 2024) menunjukkan bahwa kedua metode memiliki karakteristik performa yang berbeda dalam proses klasifikasi. Penelitian dari (Daniati et al., 2026) menjelaskan bahwa perbedaan algoritma klasifikasi dapat memengaruhi tingkat akurasi analisis sentimen. Penelitian lain seperti (Ramayasa et al., 2023) juga menjelaskan bahwa metode representasi fitur memiliki pengaruh terhadap hasil klasifikasi sentimen.

Meskipun berbagai penelitian mengenai analisis sentimen telah banyak dilakukan, sebagian besar penelitian tersebut masih berfokus pada penggunaan satu metode tertentu atau belum membahas pengaruh keseimbangan data menggunakan SMOTE terhadap performa TF-IDF dan LSTM secara bersamaan. Penelitian mengenai analisis sentimen pada ulasan aplikasi pembayaran digital, khususnya *GoPay,* masih relatif terbatas dibandingkan aplikasi e-commerce maupun media sosial. Oleh karena itu, penelitian ini dilakukan untuk membandingkan performa metode TF-IDF dan LSTM pada analisis sentimen ulasan *GoPay* setelah penerapan optimasi *preprocessing* menggunakan SMOTE. Penelitian ini diharapkan dapat mengetahui metode yang memiliki performa terbaik berdasarkan parameter *Accuracy*, *Precision*, *Recall*, dan *F1-Score* sehingga mampu menghasilkan klasifikasi sentimen yang lebih optimal.

2. METODE PENELITIAN

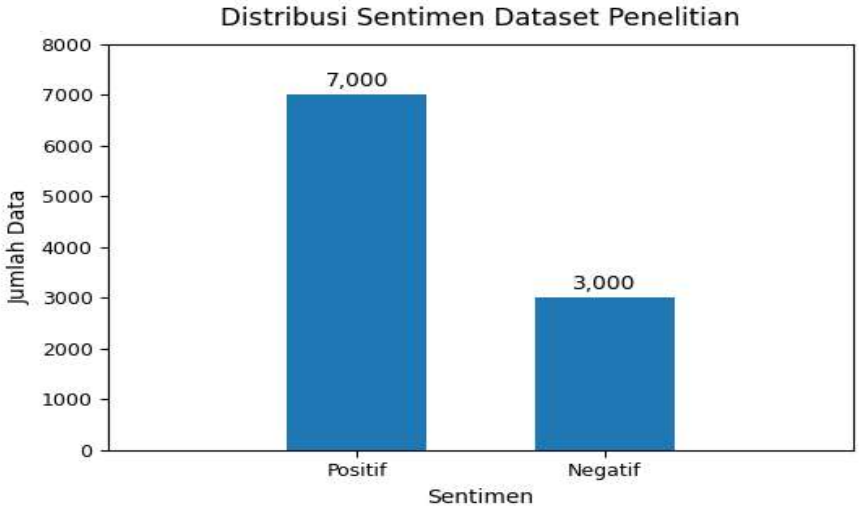
Penelitian ini menggunakan metode penelitian kuantitatif dengan melakukan eksperimen. Pendekatan kuantitatif digunakan karena penelitian berfokus pada pengolahan data ulasan pengguna aplikasi *GoPay* dalam bentuk data numerik yang selanjutnya dianalisis menggunakan metode klasifikasi sentimen. Hasil penelitian diukur berdasarkan nilai parameter evaluasi untuk mengetahui performa model yang digunakan pada proses analisis sentimen (Saputra et al., 2026).

Penelitian dilakukan dengan memanfaatkan data ulasan pengguna aplikasi *GoPay* yang diperoleh dari *Google Play Store*. Data yang telah dikumpulkan kemudian melalui beberapa tahapan pengolahan, yaitu *preprocessing*, penyeimbangan data menggunakan *Synthetic Minority Over-sampling Technique* (SMOTE), representasi fitur menggunakan TF-IDF, dan proses klasifikasi sentimen menggunakan *Long Short-Term Memory* (LSTM). Tahapan tersebut mengacu pada penelitian analisis sentimen yang memanfaatkan TF-IDF sebagai representasi fitur teks dan LSTM sebagai metode klasifikasi untuk memahami pola sentimen pada data ulasan pengguna (Mutmainah et al., 2024).

Pendekatan eksperimen pada penelitian ini dilakukan dengan membandingkan performa metode setelah penerapan SMOTE pada data ulasan *GoPay*. Evaluasi model dilakukan menggunakan *Confusion Matrix* dengan parameter *Accuracy*, *Precision*, *Recall*, dan *F1-Score* untuk mengetahui kemampuan model dalam mengklasifikasikan sentimen secara optimal. Penggunaan SMOTE dipilih karena terbukti mampu membantu mengatasi permasalahan ketidakseimbangan data pada penelitian analisis sentimen sebelumnya (Candra et al., 2024).

3. HASIL DAN PEMBAHASAN

Penelitian menggunakan sebanyak 10.000 ulasan pengguna aplikasi GoPay yang berasal dari Google Play Store. Berdasarkan hasil pelabelan, terdapat 7.000 data dengan sentimen positif dan 3.000 data dengan sentimen negatif. Data tersebut kemudian dibagi menggunakan rasio 80%:20%, sehingga diperoleh 8.000 data *training* dan 2.000 data *testing*. Data *training* selanjutnya dibagi menjadi 6.400 data *training internal* dan 1.600 data *validation internal*.



Gambar 1. Distribusi Sentimen Dataset Penelitian

Distribusi tersebut menunjukkan bahwa jumlah ulasan positif lebih banyak dibandingkan ulasan negatif. Kondisi ini menunjukkan adanya ketidakseimbangan kelas pada dataset yang digunakan. Oleh karena itu, SMOTE diterapkan pada data *training* untuk membantu menyeimbangkan distribusi kelas sebelum proses klasifikasi. Penerapan SMOTE hanya dilakukan pada data *training*, sedangkan data *testing* tetap menggunakan distribusi data asli agar proses evaluasi dapat dilakukan secara objektif.

Penelitian ini melakukan tiga skenario optimasi untuk memperoleh model dengan kinerja klasifikasi terbaik. Optimasi pertama menggunakan TF-IDF, SMOTE, dan *Logistic Regression*, optimasi kedua menggunakan LSTM dan SMOTE, sedangkan optimasi ketiga menggunakan kombinasi TF-IDF, SMOTE, dan LSTM. Ketiga model dievaluasi menggunakan 1.600 data *validation internal.*

Tabel 1. Perbandingan Hasil Optimasi

No	Model	*Accuracy*
1	TF-IDF + SMOTE	98,62%
2	LSTM + SMOTE	98,38%
3	TF-IDF, SMOTE, dan LSTM	98,56%

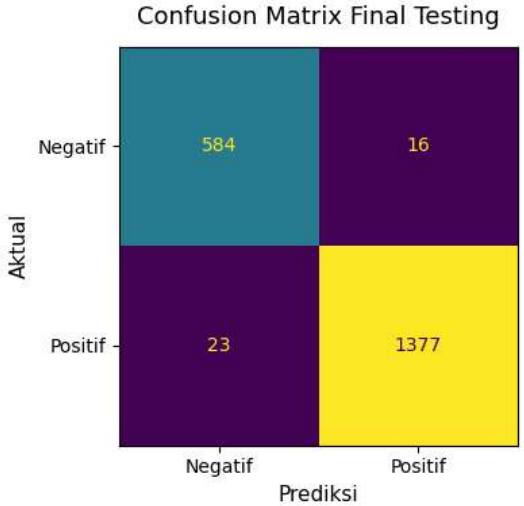
Berdasarkan Tabel 1, model TF-IDF + SMOTE memperoleh *accuracy* sebesar 98,62%, sedangkan model LSTM + SMOTE memperoleh *accuracy* sebesar 98,38% dan kombinasi TF-IDF, SMOTE, dan LSTM memperoleh *accuracy* sebesar 98,56%. Hasil tersebut menunjukkan bahwa ketiga model memiliki kinerja klasifikasi yang tinggi. Namun, model TF-IDF + SMOTE memperoleh nilai *accuracy* tertinggi sehingga dipilih sebagai model terbaik untuk digunakan pada tahap *final testing.*

Model TF-IDF + SMOTE dengan *Logistic Regression* yang memperoleh *accuracy* tertinggi pada *validation internal* selanjutnya diuji menggunakan 2.000 data *testing* yang tidak digunakan selama proses pelatihan, optimasi, maupun pemilihan model. Hasil *final testing* disajikan pada Tabel 2.

Tabel 2. *Classification Report Final Testing* Model Terbaik

Kelas /Metrik	*Precision*	*Recall*	*F1-Score*	*Support*
Negatif (0)	0.9621	0.9733	0.9677	600
Positif (1)	0.9885	0.9836	0.9860	1.400
Accuracy	0.9805	0.9805	0.9805	1
Macro avg	0.9753	0.9785	0.9769	2000
Weighted avg	0.9806	0.9805	0.9805	2000

Berdasarkan Tabel 2, data *testing* terdiri atas 600 data negatif dan 1.400 data positif. Hasil pengujian menunjukkan bahwa model memperoleh *accuracy* sebesar 98,05%, *Precision Macro* sebesar 97,53%, *Recall Macro* sebesar 97,85%, dan *F1-Score Macro* sebesar 97,69%.



Gambar 2. *Confusion Matrix Final Testing*

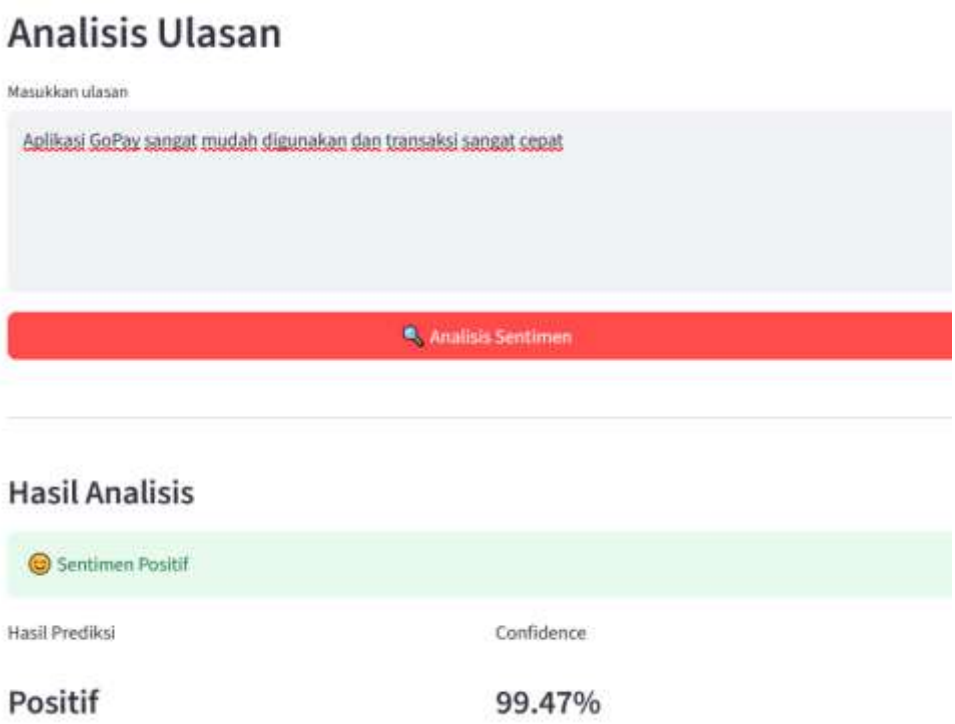
Berdasarkan *confusion matrix* pada Gambar 2, dari 600 data aktual negatif, sebanyak 584 data berhasil diklasifikasikan sebagai negatif dan 16 data diklasifikasikan sebagai positif. Sementara itu, dari 1.400 data aktual positif, sebanyak 1.377 data berhasil diklasifikasikan sebagai positif dan 23 data diklasifikasikan sebagai negatif. Secara keseluruhan, model berhasil melakukan klasifikasi

dengan benar terhadap 1.961 data dari 2.000 data *testing*, sedangkan 39 data mengalami kesalahan klasifikasi. Kesalahan tersebut terdiri atas 16 data negatif yang diprediksi sebagai positif dan 23 data positif yang diprediksi sebagai negatif. Dengan demikian, diperoleh *accuracy* sebesar 98,05%.

Hasil *final testing* menunjukkan bahwa performa model mengalami penurunan sebesar 0,57% dibandingkan *accuracy* pada *validation internal* sebesar 98,62%. Penurunan tersebut masih berada pada tingkat yang relatif kecil dan menunjukkan bahwa model tetap mampu mempertahankan performa klasifikasi yang tinggi ketika digunakan pada data yang tidak terlibat dalam proses pelatihan maupun pemilihan model.

Pada saat sebuah ulasan diberikan, sistem memproses teks tersebut menggunakan tahapan *preprocessing* yang telah ditentukan dalam penelitian. Tahapan tersebut meliputi *cleaning, case folding*, normalisasi, *tokenizing, stopword removal,* dan *stemming*. Setelah proses tersebut selesai, teks diubah menjadi representasi TF-IDF dan digunakan sebagai masukan bagi model *Logistic Regression* hasil Optimasi 1.

Hasil pengolahan kemudian ditampilkan pada halaman aplikasi. Sistem memberikan informasi berupa kategori sentimen dari ulasan, yaitu positif atau negatif, dan nilai *confidence* dari hasil prediksi. Dengan adanya informasi tersebut, pengguna dapat mengetahui hasil klasifikasi sekaligus melihat tingkat keyakinan model terhadap prediksi yang diberikan.



Gambar 3. Tampilan Hasil Analisis Sentimen

Sebagai contoh implementasi, ketika pengguna memasukkan ulasan mengenai layanan *GoPay*, sistem akan memproses teks tersebut dan memberikan hasil klasifikasi secara langsung. Hasil yang ditampilkan berupa kategori sentimen dan nilai *confidence*, sehingga pengguna dapat mengetahui hasil prediksi dengan mudah.

Implementasi ini menunjukkan penerapan model analisis sentimen ke dalam aplikasi berbasis web yang dapat digunakan untuk melakukan klasifikasi terhadap ulasan baru. Aplikasi berfungsi sebagai media penerapan model setelah proses pelatihan dan pengujian selesai dilakukan.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan mengenai analisis sentimen pada ulasan pengguna *GoPay*, dapat ditarik beberapa kesimpulan sebagai berikut:

1. Dataset yang digunakan dalam penelitian berjumlah 10.000 ulasan. Data melalui tahap *preprocessing* yang terdiri dari *cleaning*, *case folding*, normalisasi, *tokenizing*, *stopword removal*, dan s*temming*. Setelah itu, teks diubah menjadi bentuk numerik menggunakan TF-IDF. Data dibagi dengan perbandingan 80%:20%, yaitu 8.000 data *training* dan 2.000 data *testing*. Dari 8.000 data *training*, sebanyak 6.400 data digunakan sebagai *training internal* dan 1.600 data sebagai *validation internal.*
2. Hasil pengujian menunjukkan bahwa Optimasi 1 menggunakan TF-IDF dan SMOTE memperoleh *accuracy* sebesar 98,62%. Optimasi 2 menggunakan LSTM dan SMOTE memperoleh *accuracy* sebesar 98,38%, sedangkan Optimasi 3 menggunakan TF-IDF, SMOTE, dan LSTM memperoleh *accuracy* sebesar 98,56%. Dari ketiga hasil tersebut, Optimasi 1 memperoleh *accuracy* tertinggi pada *validation internal*.
3. Berdasarkan hasil tersebut, model terbaik yang digunakan dalam penelitian adalah TF-IDF dengan SMOTE menggunakan *Logistic Regression*. Model ini memperoleh a*ccuracy* sebesar 98,62% pada *validation internal* dan selanjutnya digunakan pada proses *final testing.*
4. Pada *final testing* menggunakan 2.000 data *testing*, model TF-IDF dengan SMOTE dan *Logistic Regression* memperoleh *accuracy* sebesar 98,05%. Dari 2.000 data tersebut, sebanyak 1.961 data berhasil diprediksi dengan benar dan 39 data mengalami kesalahan prediksi. Hasil ini menunjukkan bahwa model masih mampu memberikan hasil klasifikasi yang baik pada data yang tidak digunakan dalam proses optimasi.
5. Model terbaik diterapkan pada aplikasi berbasis web menggunakan Streamlit. Aplikasi dapat menerima ulasan baru dan memberikan hasil prediksi berupa sentimen positif atau negatif dan nilai *confidence*. Pengujian menggunakan 10 ulasan baru menunjukkan bahwa model dapat digunakan untuk melakukan klasifikasi terhadap ulasan yang berada di luar dataset penelitian.

REFERENSI

Adhiatma, F. D., & Qoiriah, A. (2022). Penerapan Metode TF-IDF dan Deep Neural Network untuk Analisa Sentimen pada Data Ulasan Hotel. *Journal of Informatics and Computer Science (JINACS)*, 183-193. https://doi.org/10.26740/jinacs.v4n02.p183-193

Arasy, A., Agustian, S., Handayani, L., & Iskandar, I. (2025). Klasifikasi Sentimen Menggunakan Metode Multilayer Perceptron dengan Fitur TF-IDF: Sentiment Classification Using Multilayer Perceptron Algorithm with TF-IDF Features. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *5*(3), 908-919. https://doi.org/10.57152/malcom.v5i3.2052

Candra, C., Chandra, K. W., & Irsyad, H. (2024). Efektifitas SMOTE dalam Mengatasi *Imbalanced class* Algoritma K-Nearest Neighbors pada Analisis Sentimen terhadap Starlink. *Jurnal Ilmu Komputer dan Informatika*, *4*(1), 31-42. https://doi.org/10.54082/jiki.132

Daniati, E., Nugroho, A., Ristyawan, A., & Utama, H. (2026). Perbandingan Kinerja Algoritma SVM, LSTM, dan Fine-tuned IndoBERT dalam Analisis Sentimen Opini Masyarakat Indonesia terhadap Mobil Listrik. *The Indonesian Journal of Computer Science Research*, *5*(1), 1-13. https://doi.org/10.59095/ijcsr.v5i1.245

Harmandini, K. P. (2024). Analysis of TF-IDF and TF-RF feature extraction on product review sentiment. *Sinkron: jurnal dan penelitian teknik informatika*, *8*(2), 929-937. https://doi.org/10.33395/v8i2.13376

Husain, N. P., Sukirman, S., & SAJIAH, S. (2024). Analisis Sentimen Ulasan Pengguna Tiktok pada *Google Play Store* Berbasis TF-IDF dan Support Vector Machine. *Journal of System and Computer Engineering*, *5*(1), 91-102. https://doi.org/10.61628/jsce.v5i1.1105

Kadir, S. F., & Fairuzabadi, A. (2025). Analisis Sentimen Ulasan Shopee di Google Play dengan TF-IDF dan Logistic Regression. *RIGGS: Journal of Artificial Intelligence and Digital Business*, *4*(2), 7940-57945. https://doi.org/10.31004/riggs.v4i2.2850

Musfiroh, M., Tholib, A., & Arifin, Z. (2024). Analisis Sentimen Terhadap Ulasan Aplikasi Shopee di *Google Play Store* Menggunakan Metode TF-IDF dan *Long Short-Term Memory*. *J. Electr. Eng. Comput*, *6*(2), 371-381. DOI: 10.33650/jeecom.v4i2

Mutmainah, S. (2024). Metode *Deep Learning* LSTM dalam Analisis Sentimen Aplikasi PeduliLindungi. *Scientific: Journal of Computer Science and Informatics*, *1*(1), 9-19. https://doi.org/10.34304/scientific.v1i1.231

Nurrahman, A. A., Mauladi, M., & Rahman, A. (2025). Analisis Sentimen Masyarakat terhadap Kenaikan Harga Bahan Bakar Minyak Menggunakan Support Vector Machine dan SMOTE. *sudo Jurnal Teknik Informatika*, *4*(2), 50-56. https://doi.org/10.56211/sudo.v4i2.908

Pratiwi, G. I., Nurandini, A., Meizheilla, D., Rinjani, E. N., Apriliani, Z. R., & Widodo, R. (2025). Analisis Sentimen Pengguna Shopee Menggunakan Lstm. *Journal of Informatics and Interactive Technology*, *2*(2), 384-391. https://doi.org/10.63547/jiite.v2i2.91

Putri, D., & Suryanto, A. (2025). Penerapan LSTM untuk Menganalisis Sentimen Review Pengguna Aplikasi Zoom pada Play Store. *IMTechno: Journal of Industrial Management and Technology*, *6*(2), 100-106. https://doi.org/10.31294/imtechno.v6i2.9614

Raihan Hasya. (2022) E Wallet Yang Paling Sering Dipakai Masyarakat Indonesia. https://goodstats.id/article/ini-10-e-wallet-yang-paling-sering-dipakai-masyarakat-indonesia-M4TA4

Ramayasa, I. P., Saryanti, I. G. A. D., & Dharmendra, I. K. (2023). Perbandingan Metode Vektorisasi Pada Analisa Sentiment, Studi Kasus: Cyberbullying Pada Komentar Instagram. *Jurnal Teknologi Informasi Dan Komputer*, *9*(5). https://doi.org/10.36002/jutik.v9i5.2645

Saputra, R. D., Hasanuddin, T., & Ramdaniah, R. (2025). Analisis Review Pengguna Terhadap Fitur Baru Whatsapp Menggunakan LSTM. *LINIER: Literatur Informatika dan Komputer*, *2*(4), 502-509. https://doi.org/10.33096/linier.v2i4.3335

Sutrisno, J., & Putri, J. G. E. (2026). Analisis Sentimen Pada Aplikasi Livin'By Mandiri Menggunakan Metode Tf-Idf Dan Naive Bayes. *JIEMS (Journal of Industrial Engineering and Management Systems)*, *17*(1). http://dx.doi.org/10.30813/jiems.v17i1.8931