

# Pengaruh Preprocessing Data Teks untuk Analisis Ulasan Aplikasi E-Wallet pada Google Play Store

Oleh:

**Arrizqi Fauzy Aufar (201080200101)**

Progam Studi Informatika

Universitas Muhammadiyah Sidoarjo

2024

# Pendahuluan

Dalam era digitalisasi yang berkembang pesat ini, layanan **E-Wallet** telah menjadi bagian yang tak terpisahkan dari kehidupan sehari-hari pengguna di berbagai negara. Kemajuan teknologi keuangan, kenyamanan, dan kemudahan akses yang ditawarkan oleh E-Wallet telah memicu adopsi yang signifikan dari solusi pembayaran ini.

Perkembangan era digital telah menghasilkan dinamika baru dalam kaitannya dengan Keterkaitan antara penyedia layanan E-Wallet dan pengguna. Pengguna **E-Wallet Dana**, seperti pengguna layanan serupa, semakin sering menyuarakan pengalaman mereka melalui ulasan dan komentar di platform distribusi seperti **Google Play Store**. Ulasan ini bukan hanya menjadi sumber berharga bagi calon pengguna yang ingin mengevaluasi aplikasi Dana, tetapi juga merupakan informasi berharga bagi perusahaan itu sendiri.

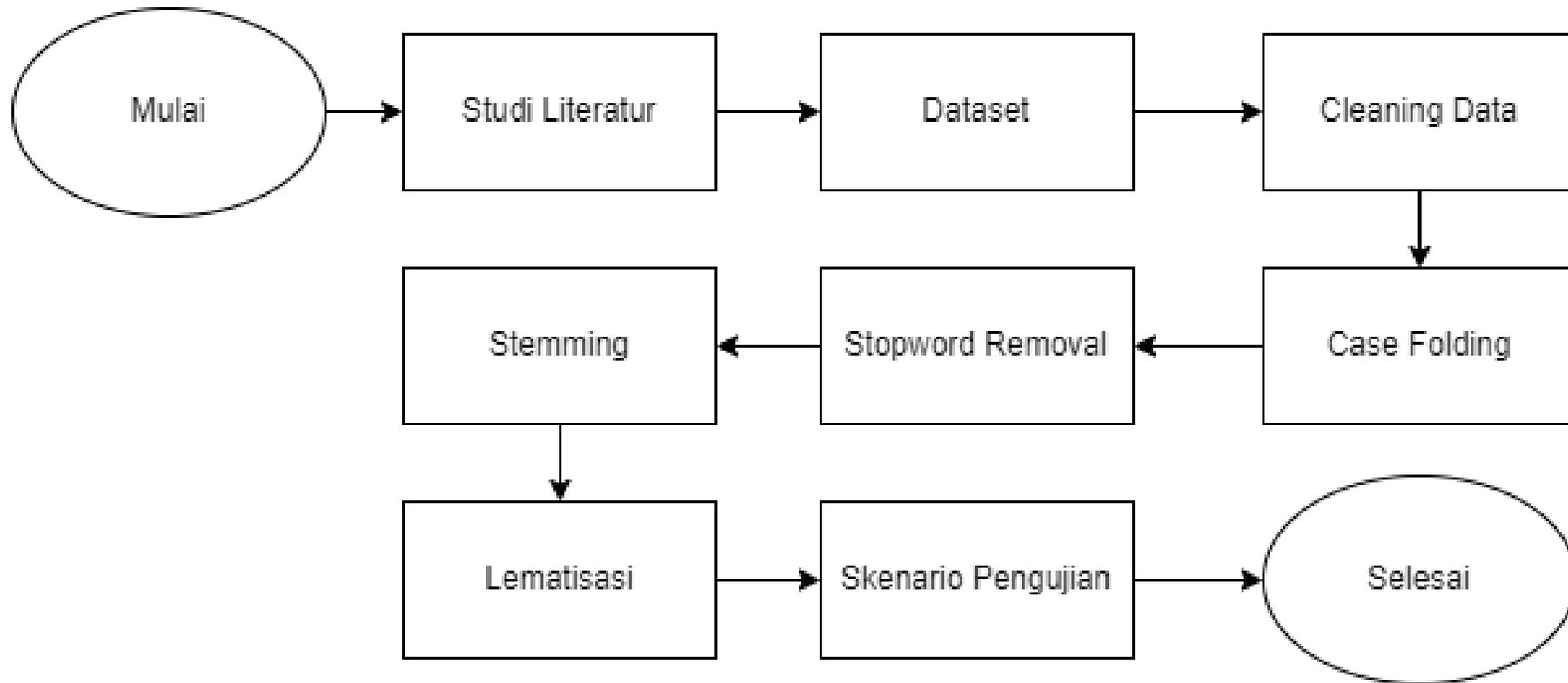
# Pendahuluan

Jurnal ini diharapkan dapat meningkatkan akurasi dalam menangkap sentimen pengguna terhadap aplikasi Dana dengan mengatasi masalah-masalah preprocessing yang ada, sehingga memberikan wawasan yang lebih akurat kepada penyedia layanan. Hasil penelitian ini akan memiliki dampak positif pada pengembangan produk dan layanan Dana, membantu mereka dalam meningkatkan kepuasan pelanggan dan menjaga kompetitivitas di pasar digital.

# Pertanyaan Penelitian (Rumusan Masalah)

Rumusan masalah akan lebih memfokuskan permasalahan utama pada penelitian ini **“Apa saja teknik preprocessing yang paling sesuai untuk membersihkan dan mempersiapkan data ulasan aplikasi Dana agar dapat dianalisis sentimennya dengan akurat?”**

# Metode



# Skenario Pengujian

- Studi akan mengadopsi tujuh skenario yang berbeda, dimana setiap skenario mengimplementasikan variasi dalam tahap preprocessing. Setiap skenario akan diuji dengan kombinasi dan urutan teknik preprocessing yang beragam.

Skenario	<i>Preprocessing</i>
1	<i>case folding, cleaning, tokenizing, stopword removal, stemming (sastrawi)</i>
2	<i>case folding, cleaning, tokenizing, stemming (sastrawi), lemmatization</i>
3	<i>case folding, cleaning, tokenizing, stopword removal, lemmatization</i>
4	<i>case folding, cleaning, tokenizing, stopword removal</i>
5	<i>case folding, cleaning, tokenizing, stemming (sastrawi)</i>
6	<i>case folding, cleaning, tokenizing, lemmatization</i>
7	<i>case folding, cleaning, tokenizing, stopword removal, stemming (sastrawi), lemmatization</i>

# Hasil

	skenario	akurasi (SVM)			
		Linear	RBF	Polynomial	Sigmoid
1	case folding, cleaning, tokenizing, stopword removal, stemming (sastrawi)	100%	100%	99,42%	94,25%
2	case folding, cleaning, tokenizing, stemming (sastrawi), lemmatization	100%	100%	99,42%	98,85%
3	case folding, cleaning, tokenizing, stopword removal, lemmatization	100%	100%	98%	96,50%
4	case folding, cleaning, tokenizing, stopword removal	100%	100%	99,42%	97,12%
5	case folding, cleaning, tokenizing, stemming (sastrawi)	100%	100%	99,42%	98,27%
6	case folding, cleaning, tokenizing, lemmatization	100%	100%	99%	99,50%
7	case folding, cleaning, tokenizing, stopword removal, stemming (sastrawi), lemmatization	98,1%	98,1%	94,25%	93,30%

# Pembahasan

Secara keseluruhan, teknik preprocessing yang menggabungkan lemmatization dan stemming (sastrawi) memberikan hasil terbaik dalam hal akurasi, terutama dengan kernel Linear dan RBF yang menunjukkan konsistensi tinggi. Kernel Polynomial dan Sigmoid memperlihatkan variasi akurasi yang lebih besar tergantung pada teknik preprocessing yang digunakan, dengan skenario keenam menonjol sebagai hasil terbaik di antara semua pengujian.

# Temuan Penting Penelitian

Berdasarkan hasil dari beberapa skenario pengujian yang telah dilakukan, pengaruh preprocessing terhadap berbagai model teknik ditentukan untuk menghasilkan tingkat akurasi yang lebih tinggi. Dari penelitian tersebut dapat disimpulkan bahwa system kinerja terbaik dihasilkan ketika menggunakan kombinasi case folding, cleaning, tokenizing, dan lemmatization yang menghasilkan tertinggi diantara teknik lain. Skenario pengujian diuji menggunakan metode SVM menggunakan keempat kernel yang menghasilkan 100% pada kernel Linear, 100% pada kernel RBF, 99% pada kernel Polynomial, 99.50% pada kernel Sigmoid. Peneliti menemukan bahwa dengan menggunakan seluruh teknik preprocessing pada skenario pengujian terakhir belum tentu meningkatkan keakurasian.

# Manfaat Penelitian

- ✓ Menambah referensi khususnya yang berhubungan dengan sentimen analisis terkait studi kasus penggunaan E-Wallet Dana.
- ✓ Membantu penyedia layanan untuk mengetahui respon dari pengguna aplikasi untuk meningkatkan kepuasan pengguna.
- ✓ Menambah wawasan dan pengetahuan baru dalam merancang preprocessing teks pada analisis sentimen studi kasus aplikasi E-Wallet Dana.

# Referensi

- N. D. Abrilia and S. Tri, “Pengaruh Persepsi Kemudahan Dan Fitur Layanan Terhadap Minat Menggunakan E-Wallet Pada Aplikasi Dana Di Surabaya,” *J. Pendidik. Tata Niaga*, vol. 8, no. 3, pp. 1006–1012, 2020.
- S. P. Anggraini and S. Suaidah, “Sistem Informasi Sentral Pelayanan Publik dan Administrasi Kependudukan Terpadu dalam Peningkatan Kualitas Pelayanan Kepada Masyarakat Berbasis Website ...,” *J. Teknol. Dan Sist. Inf.*, vol. 3, no. 1, pp. 12–19, 2022, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/sisteminformasi/article/view/1658%0Ahttp://jim.teknokrat.ac.id/index.php/sisteminformasi/article/viewFile/1658/579>
- J. W. Iskandar and Y. Nataliani, “Perbandingan Naïve Bayes, SVM, dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 6, pp. 1120–1126, Dec. 2021, doi: 10.29207/resti.v5i6.3588.
- M. A. Rosid, A. S. Fitriani, I. R. I. Astutik, N. I. Mulloh, and H. A. Gozali, “Improving Text Preprocessing for Student Complaint Document Classification Using Sastrawi,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 874, no. 1, 2020, doi: 10.1088/1757-899X/874/1/012017.
- R. Ulgasesa, A. B. P. Negara, and T. Tursina, “Pengaruh Stemming Terhadap Performa Klasifikasi Sentimen Masyarakat Tentang Kebijakan New Normal,” *J. Sist. dan Teknol. Inf.*, vol. 10, no. 3, p. 286, 2022, doi: 10.26418/justin.v10i3.53880.
- Z. R. N. S. Prasetija, A. Romadhony, and E. B. Setiawan, “Analisis Pengaruh Normalisasi Teks pada Klasifikasi Sentimen Ulasan Produk Kecantikan,” *e-Proceeding Eng.*, vol. 9, no. 3, pp. 1769–1775, 2022, [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/18184/17795>
- H. H. Mubaroroh, H. Yasin, and A. Rusgiyono, “Analisis Sentimen Data Ulasan Aplikasi Ruangguru Pada Situs Google Play Menggunakan Algoritma Naïve Bayes Classifier Dengan Normalisasi Kata Levenshtein Distance,” *J. Gaussian*, vol. 11, no. 2, pp. 248–257, 2022, doi: 10.14710/j.gauss.v11i2.35472.

- G. A. BUNTORO, R. ARIFIN, G. N. SYAIFUDDIIN, A. SELAMAT, O. KREJCAR, and H. FUJITA, “Implementation of a Machine Learning Algorithm for Sentiment Analysis of Indonesia’s 2019 Presidential Election,” *IIUM Eng. J.*, vol. 22, no. 1, pp. 78–92, 2021, doi: 10.31436/IIUMEJ.V22I1.1532.
- D. Duei Putri, G. F. Nama, and W. E. Sulistiono, “Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier,” *J. Inform. dan Tek. Elektro Terap.*, vol. 10, no. 1, pp. 34–40, 2022, doi: 10.23960/jitet.v10i1.2262.
- A. B. Putra Negara, “The Influence Of Applying Stopword Removal And Smote On Indonesian Sentiment Classification,” *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 14, no. 3, p. 172, 2023, doi: 10.24843/lkjiti.2023.v14.i03.p05.
- S. J. Angelina, A. Bijaksana, P. Negara, and H. Muhandi, “Analisis Pengaruh Penerapan Stopword Removal Pada Performa Klasifikasi Sentimen Tweet Bahasa Indonesia,” vol. 02, no. 1, pp. 165–173, 2023, doi: 10.26418/juara.v2i1.69680.
- H. A. Almuzaini and A. M. Azmi, “Impact of Stemming and Word Embedding on Deep Learning-Based Arabic Text Categorization,” *IEEE Access*, vol. 8, pp. 127913–127928, 2020, doi: 10.1109/ACCESS.2020.3009217.
- O. Manullang, C. Prianto, and N. H. Harani, “Analisis Sentimen Untuk Memprediksi Hasil Calon Pemilu Presiden Menggunakan Lexicon Based Dan Random Forest,” *J. Ilm. Inform.*, vol. 11, no. 02, pp. 159–169, 2023, doi: 10.33884/jif.v11i02.7987.
- F. Noer Azzahra et al., “Penerapan Metode Naive Bayes Dalam Klasifikasi Spam SMS Menggunakan Fitur Teks Untuk Mengatasi Ancaman Pada Pengguna,” *J. Inf. Syst. Res.*, vol. 5, no. 3, p. 880, 2024, doi: 10.47065/josh.v5i3.5070.
- D. A. Vonega, A. Fadila, and D. E. Kurniawan, “Analisis Sentimen Twitter Terhadap Opini Publik Atas Isu Pencalonan Puan Maharani dalam PILPRES 2024,” *J. Appl. Informatics Comput.*, vol. 6, no. 2, pp. 129–135, 2022, doi: 10.30871/jaic.v6i2.4300.

