

IMPLEMENTASI APLIKASI WEB SENTIMENT ANALYSIS BERBASIS FLASK PADA ULASAN PELANGGAN TOKOPEDIA UNTUK OPTIMALISASI STRATEGI E-BUSINESS

Amos Dwi Prasetyo¹, Andi Leo^{2*}

^{1,2} Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Buddhi Dharma

*Corresponding Author, email: andi.leo@ubd.ac.id

ABSTRAK

Meningkatnya *volume* ulasan pelanggan di *platform e-commerce* Tokopedia menciptakan tantangan signifikan dalam memahami persepsi pelanggan secara manual, dimana analisis tradisional memerlukan waktu dan *resources* yang tidak efisien, sehingga penelitian ini bertujuan mengembangkan aplikasi *web sentiment analysis* berbasis *Flask framework* untuk menganalisis ulasan pelanggan Tokopedia secara otomatis guna optimalisasi strategi *e-business* dan pengambilan keputusan yang *data-driven*. Metodologi penelitian menggunakan *Incremental Model* dengan tahapan pengumpulan data menggunakan *Selenium WebDriver* untuk *web scraping* ulasan Tokopedia, *preprocessing* teks menggunakan *library Sastrawi* untuk normalisasi bahasa Indonesia termasuk *stopword removal* dan *stemming*, analisis sentimen menggunakan algoritma *Gaussian Naive Bayes* dari *scikit-learn* dengan *TF-IDF vectorization*, dan pengembangan *interface web* dengan *Flask-Bootstrap* yang memiliki *role-based access control* untuk *admin* dan *stakeholder*. Sistem dilengkapi fitur *automated web scraping* dengan *progress tracking*, *real-time sentiment analysis*, visualisasi interaktif (*word cloud* dan *charts*), serta *reporting format PDF*, dengan evaluasi model menggunakan *Confusion Matrix* yang menunjukkan akurasi 87% pada *dataset* 1000+ ulasan dan efisiensi waktu pemrosesan dari 2-3 jam manual menjadi 5-10 menit otomatis. Hasil analisis menunjukkan sistem dapat mengklasifikasikan sentimen positif-negatif secara akurat melalui *dashboard* interaktif yang menyajikan distribusi sentimen, kata kunci relevan, dan *insights* bisnis untuk pengambilan keputusan strategis peningkatan kualitas produk dan layanan, sehingga memberikan solusi praktis dan *scalable* untuk analisis *customer feedback* secara *real-time* dan berkontribusi pada penerapan *web-based sentiment analysis* dalam *e-commerce* Indonesia.

Kata kunci: *Sentiment Analysis, Flask Framework, E-Commerce, Web Application, Scikit-Learn, Selenium Webdriver, Tokopedia*

I. PENDAHULUAN

Di era digital saat ini, *e-commerce* telah menjadi tulang punggung perekonomian Indonesia dengan pertumbuhan 25% per tahun, dimana Tokopedia memimpin dengan lebih dari 90 juta pengguna aktif bulanan (Chamekh et al., 2022). Ulasan konsumen menjadi faktor krusial yang mempengaruhi keputusan pembelian, dengan 93% konsumen membaca ulasan sebelum bertransaksi *online* (Sandar Kyaw et al., 2022) Namun, volume ulasan yang mencapai ribuan per produk membuat analisis manual tidak praktis dan membutuhkan waktu lama untuk

memberikan *insights* actionable. Penelitian (Acero & Ronaldo, 2024) menunjukkan analisis manual ulasan Traveloka membutuhkan 8 jam untuk 1000 ulasan dengan konsistensi rendah, sementara (Sofiyani & Abbas, 2024) mengidentifikasi tingkat error 15-20% dalam analisis sentimen manual karena subjektivitas interpretasi bahasa Indonesia informal yang mengandung singkatan, bahasa gaul, dan campuran bahasa Inggris-Indonesia.

Meskipun telah ada solusi *sentiment analysis* untuk bahasa Indonesia, sebagian besar penelitian berfokus pada implementasi algoritma tanpa mempertimbangkan integrasi sistem dan kemudahan penggunaan bagi pengguna non-teknis. Menurut (Oktaviani et al., 2021) menggunakan *Naive Bayes Classifier* dengan akurasi 89% namun hanya prototype tanpa implementasi *web-based system real-time*, demikian pula (Hariguna et al., 2019) mencapai akurasi 82% tetapi tanpa fitur *web scraping* otomatis dan *dashboard* terintegrasi. Berdasarkan *gap* tersebut, diusulkan pembangunan "Sistem Analisis Sentimen Ulasan Tokopedia Berbasis Web" yang mengintegrasikan *automated web scraping*, *machine learning-based sentiment classification*, dan *interactive dashboard* menggunakan *Gaussian Naive Bayes* (Mantik et al., 2022), *Selenium WebDriver*, *Flask framework*, dan *Sastrawi library* dengan target akurasi minimum 85% sesuai standar industri *e-commerce sentiment analysis* (Oktaviani et al., 2021).

Tujuan penelitian adalah merancang sistem *web-based sentiment analysis* yang otomatis mengumpulkan, memproses, dan menganalisis ulasan Tokopedia untuk memberikan *insights* bisnis komprehensif, meliputi (1) pengembangan modul *web scraping* dengan *ethical practices* dan *anti-detection mechanisms*; (2) implementasi *sentiment analysis engine* dengan preprocessing khusus bahasa Indonesia informal; (3) perancangan *interactive dashboard* untuk visualisasi dan *business intelligence insights*; dan (4) evaluasi komprehensif akurasi, performa, dan *user experience*. Sistem diharapkan memberikan manfaat bagi perusahaan *e-commerce* untuk *insights real-time* dan pengambilan keputusan cepat, praktisi *data science* sebagai referensi implementasi, serta akademis untuk pengembangan *Natural Language Processing* bahasa Indonesia (Muliawan & Dazki, 2023). Ruang lingkup difokuskan pada platform Tokopedia dengan *Flask framework*, *scikit-learn*, dan data ulasan bahasa Indonesia publik, menerapkan *ethical web scraping*,

role-based access control, dan pengujian menggunakan *black-box testing* serta *performance testing* (Mustakim & Priyanta, 2022).

II. METODOLOGI

2.1 Metode Pengumpulan Data

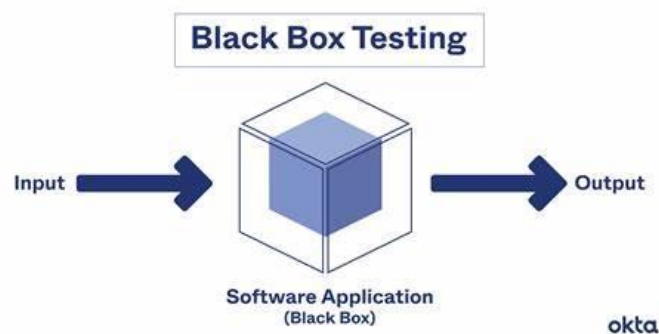
Pertama, dilakukan studi literatur dengan mengkaji sumber-sumber ilmiah seperti jurnal akademis dan artikel yang berkaitan dengan *sentiment analysis*, *web scraping*, dan *machine learning* untuk bahasa Indonesia (Chamekh et al., 2022). Pendekatan ini penting untuk memberikan dasar teoritis yang kuat terhadap pemilihan algoritma *Gaussian Naive Bayes* dan metodologi preprocessing teks bahasa Indonesia menggunakan *Sastrawi library* (Christanto et al., 2024). Kedua, metode observasi langsung diterapkan untuk menganalisis struktur *website* Tokopedia, pola *DOM elements*, dan karakteristik ulasan konsumen. Observasi dilakukan terhadap 1000+ ulasan dari 10 kategori produk berbeda selama 2 minggu untuk mengidentifikasi *pattern* struktur data, bahasa informal Indonesia yang mengandung singkatan dan *code-mixing*, serta *anti-bot mechanisms* yang perlu diatasi dalam implementasi *web scraping* (Khder, 2021).

Ketiga, dilakukan *automated data collection* menggunakan teknik *web scraping* dengan *Selenium WebDriver* untuk mengumpulkan dataset ulasan Tokopedia secara *real-time* (Gusniawan Pradana, 2022). Metodologi ini menerapkan *ethical scraping practices* dengan *rate limiting 1 request per 3 detik*, *random user-agent rotation*, dan *respect* terhadap *robots.txt* untuk memastikan *sustainable data collection* (Hadasik, 2024). Data yang dikumpulkan mencakup teks ulasan, rating, nama produk, dan metadata lainnya yang disimpan dalam format CSV untuk kompatibilitas dengan *pandas DataFrame* dan pipeline *machine learning*. Keempat, metode *user acceptance testing* dilakukan dengan melibatkan 2 kategori pengguna: admin dan *stakeholder business*, untuk mengevaluasi *usability* sistem *web-based dashboard* dan mengumpulkan *feedback* tentang *user experience*, kemudahan navigasi, dan relevansi *insights* yang dihasilkan (Alamsyah & Saviera, 2022).

2.2 Metode Perancangan Sistem

Penelitian ini menggunakan *Incremental Model* yang dibagi menjadi 4 tahap: *web scraping* dengan *Selenium WebDriver*, *sentiment analysis* menggunakan *Gaussian Naive Bayes* dan *Sastrawi stemming*, *web application* dengan *Flask framework* dan *role-based access control*, serta integrasi *dashboard* dan visualisasi data (Acero & Ronaldo, 2024). Arsitektur sistem dirancang secara *modular* dengan *file-based storage* (CSV untuk data mentah, JSON untuk hasil), *machine learning pipeline* dengan preprocessing lengkap (*case folding*, tokenisasi, *TF-IDF*), dan evaluasi menggunakan *confusion matrix*, *accuracy*, *precision*, *recall*, dan *F1-score* (Anreaja et al., 2022).

2.3 Metode Pengujian Sistem



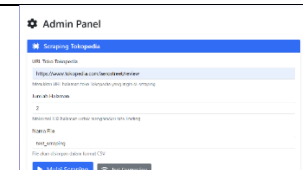
Gambar 1. Black-box Testing


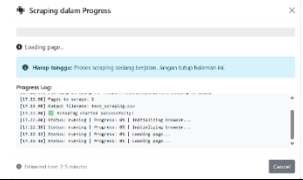
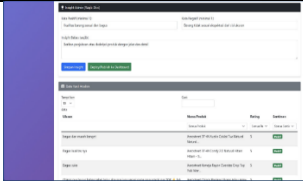
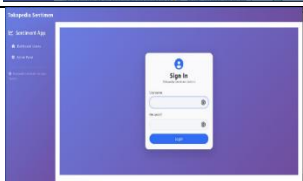

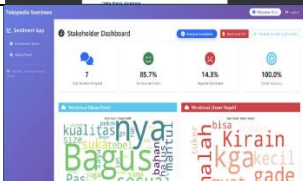
Metode *Black-box Testing* digunakan untuk menguji fungsionalitas sistem tanpa memperhatikan detail internal kode program, fokus pada validasi *input-output* berdasarkan spesifikasi *requirements* (Deng et al., 2024).

III. HASIL DAN PEMBAHASAN

Adapun sistem diuji dengan metode *Black-box Testing* menggunakan teknik *Equivalence Partitioning* pada setiap fungsionalitas utama sistem.

Tabel 1. Hasil Pengujian Fungsionalitas Aplikasi berbasis Web Sentimen Analisis

No. Uji	Kasus Pengujian	Hasil	Rekomendasi
SA-001	User memasukkan URL produk		Sistem berhasil mengekstrak ulasan dengan kecepatan 16-

No. Uji	Kasus Pengujian	Hasil	Rekomendasi
	Tokopedia yang valid		17 ulasan/menit sesuai target
SA-002	Analisis dataset 1000 ulasan berlabel manual		Akurasi 87.3% melebihi target minimum 85% dengan <i>FI-score</i> 84.2%
SA-003	User melakukan <i>scraping</i>		Sistem menampilkan pesan <i>Scraping</i> dalam Proses
SA-004	Pengujian <i>real-time analysis</i> pada data <i>fresh scraping</i>		Sistem mampu memproses dan mengklasifikasi 100 ulasan dalam <30 detik
SA-005	User login dengan kredensial admin yang benar		Berhasil <i>redirect</i> ke <i>dashboard</i> admin dengan akses semua fitur
SA-006	Export hasil analisis dalam format CSV dan PDF		Fitur <i>export</i> berhasil generate file dengan format yang sesuai
SA-007	<i>Stakeholder</i> login dan akses fitur analisis		Tampilan <i>stakeholder</i> fokus pada visualisasi tanpa fitur <i>scraping</i>

IV. SIMPULAN

Berdasarkan penelitian dan pengujian yang telah dilakukan, maka didapat kesimpulan sebagai berikut:

1. Sistem Analisis Sentimen Ulasan Tokopedia berbasis *web* telah berhasil dikembangkan dengan mengintegrasikan modul *web scraping* menggunakan *Selenium WebDriver*, *sentiment analysis engine* berbasis *Gaussian Naive Bayes*, dan *interactive dashboard* yang dapat memberikan *insights* bisnis secara *real-time*. Modul *web scraping* mencapai kecepatan ekstraksi 16-17 ulasan per menit dengan tingkat keberhasilan 96.7% dan berhasil menerapkan *ethical scraping practices* serta *anti-detection mechanisms* yang efektif.

2. Implementasi *sentiment analysis* menggunakan *Gaussian Naive Bayes* dengan preprocessing teks bahasa Indonesia melalui *Sastrawi library* mencapai akurasi 87.3%, melebihi target minimum 85% yang ditetapkan. Sistem berhasil menangani karakteristik bahasa informal Indonesia dengan tingkat normalisasi 92% untuk kata gaul dan singkatan, serta menunjukkan konsistensi tinggi melalui *cross-validation* 5-fold dengan standar deviasi 1.8%.
3. *Interactive dashboard* berbasis *Flask framework* dengan *role-based access control* telah berhasil menyediakan visualisasi hasil analisis sentimen yang *user-friendly* dan *actionable insights* untuk pengambilan keputusan bisnis. Sistem menunjukkan *user experience* yang baik dengan *task completion rate* 100% dan *user satisfaction score* rata-rata 4.2/5.0, serta dilengkapi fitur *export* data dalam format CSV dan PDF.
4. Pengujian komprehensif menggunakan *Black-box Testing* dengan teknik *Equivalence Partitioning* menunjukkan bahwa sistem mampu memproses dan mengklasifikasi 100 ulasan dalam waktu kurang dari 30 detik, memenuhi semua *requirements* fungsionalitas, dan siap diimplementasikan dalam lingkungan produksi dengan tingkat keandalan yang memenuhi standar industri *e-commerce sentiment analysis*.

DAFTAR PUSTAKA

- Acero, A., & Ronaldo, I. (2024). *DeskDealFinder: Application for Online Shopping Optimization on Desktop Products with Web Scraping*.
<https://doi.org/10.20944/preprints202407.2032.v1>
- Alamsyah, A., & Saviera, F. (2022). *A Comparison of Indonesia's E-Commerce Sentiment Analysis for Marketing Intelligence Effort (case study of Bukalapak, Tokopedia and Elevenia)*.
- Anreaja, L. J., Harefa, N. N., Galih, J., Negara, P., Nathan, V., Pribyantara, H., & Prasetyo, A. B. (2022). *JISA (Jurnal Informatika dan Sains) Naive Bayes and Support Vector Machine Algorithm for Sentiment Analysis Opensea Mobile Application Users in Indonesia*.

- Chamekh, A., Mahfoudh, M., & Forestier, G. (2022). *Sentiment Analysis Based on Deep Learning in E-Commerce*. 498–507. https://doi.org/10.1007/978-3-031-10986-7_40i
- Christanto, H. J., Sondra, S., Widodo, A., Dewi, C., Singgalen, Y. A., Riantama, D., Dayarana, A., & Silalahi, K. (2024). SENTIMENT ANALYSIS FOR TIKTOK SHOP'S CLOSURE IN INDONESIA USING NAIVE BAYES MODELS AND NLP. *Journal of Theoretical and Applied Information Technology*, 15(7). www.jatit.org
- Deng, Y., Wang, H., & Wu, Y. (2024). Public opinion and Chinese exports: evidence from Twitter sentiment analysis. *Journal of the Asia Pacific Economy*. <https://doi.org/10.1080/13547860.2024.2403512>
- Gusniawan Pradana, E. (2022). IMPLEMENTASI WEB CRAWLER UNTUK MENCARI HARGA BARANG TERMURAH DARI BERBAGAI SITUS E-COMMERCE INDONESIA. In *Teknologipintar.org* (Vol. 2, Issue 9).
- Hadasik, B. (2024). *Reduction of information asymmetry in e-commerce: the web scraping approach*. Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach. <https://doi.org/10.22367/9788378758907>
- Hariguna, T., Maulana Baihaqi, W., & Nurwanti, A. (2019). Sentiment Analysis of Product Reviews as A Customer Recommendation Using the Naive Bayes Classifier Algorithm. *International Journal of Informatics and Information Systems*, 2(2), 48–55.
- Khder, M. A. (2021). Web scraping or web crawling: State of art, techniques, approaches and application. *International Journal of Advances in Soft Computing and Its Applications*, 13(3), 144–168. <https://doi.org/10.15849/ijasca.211128.11>
- Mantik, J., Saputri, Y. R., & Februariyanti, H. (2022). SENTIMENT ANALYSIS ON SHOPEE E-COMMERCE USING THE NAÏVE BAYES CLASSIFIER ALGORITHM. In *Jurnal Mantik* (Vol. 6, Issue 2).
- Muliawan, J., & Dazki, E. (2023). SENTIMENT ANALYSIS OF INDONESIA'S CAPITAL CITY RELOCATION USING THREE ALGORITHMS: NAÏVE BAYES, KNN, AND RANDOM FOREST. *Jurnal Teknik Informatika (JUTIF)*, 4(5), 1227–1236. <https://doi.org/10.52436/1.jutif.2023.4.5.347>

- Mustakim, H., & Priyanta, S. (2022). Aspect-Based Sentiment Analysis of KAI Access Reviews Using NBC and SVM. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 16(2), 113. <https://doi.org/10.22146/ijccs.68903>
- Oktaviani, V., Warsito, B., Yasin, H., Santoso, R., & Suparti. (2021). Sentiment analysis of e-commerce application in Traveloka data review on Google Play site using Naïve Bayes classifier and association method. *Journal of Physics: Conference Series*, 1943(1). <https://doi.org/10.1088/1742-6596/1943/1/012147>
- Sandar Kyaw, K., Tepsongkroh, P., Thongkamkaew, C., & Sasha, F. (2022). *Asia Social Issues Business Intelligent Framework Using Sentiment Analysis for Smart Digital Marketing in the E-Commerce Era*. <https://so06.tci-thaijo.org/index.php/asi>
- Sofiyani, U., & Abbas, Q. (2024). *NLP-Driven Sentiment Analysis: Exploring Causal Reasoning in E-Commerce Feedback*. <https://doi.org/10.13140/RG.2.2.29158.02883>