Department of Digital Business
Journal of Artificial Intelligence and Digital Business (RIGGS)
Homepage: https://journal.ilmudata.co.id/index.php/RIGGS
Vol. 4 No. 2 (2025) pp: 7940-57945
P-ISSN: 2963-9298, e-ISSN: 2963-914X

Analisis Sentimen Ulasan Shopee di Google Play dengan TF-IDF dan Logistic Regression

Shaifany Fatriana Kadir[1], Ahmad Fairuzabadi[2]
[1,2]Bisnis Digital, Fakultas Ekonomi dan Bisnis, Universitas Widya Gama
[1]shaifanyfk@widyagama.ac.id*, [2]fairuz@widyagama.ac.id

Abstrak

Pertumbuhan aplikasi belanja daring mendorong meningkatnya jumlah ulasan pengguna yang tersedia di Google Play Store. Ulasan ini tidak hanya mencerminkan pengalaman pengguna, tetapi juga berperan sebagai masukan penting bagi pengembang aplikasi. Penelitian ini menganalisis sentimen ulasan aplikasi Shopee Indonesia (package com.shopee.id) periode 2024–2025 dengan menggunakan pendekatan Term Frequency–Inverse Document Frequency (TF-IDF) dan algoritma Logistic Regression. Data sebanyak 5.000 ulasan dikumpulkan melalui pustaka google-play-scraper dan dilabeli otomatis berdasarkan skor rating: 1–2 negatif, 3 netral, dan 4–5 positif. Proses preprocessing meliputi normalisasi teks, penghapusan stopword bahasa Indonesia, serta stemming. Distribusi data yang timpang (4049 positif, 796 negatif, 155 netral) ditangani dengan Random Oversampling pada data latih. Evaluasi dilakukan menggunakan stratified split 80:20 dan 5-fold cross-validation. Hasil menunjukkan Logistic Regression memberikan performa terbaik dengan akurasi 85,11% dan macro-F1 0,58 pada data uji, lebih baik dibandingkan SVM dan Naive Bayes. Confusion matrix memperlihatkan bahwa kelas positif dapat dikenali dengan baik (F1=0,92), sementara kelas netral sulit diprediksi (F1=0,14). Visualisasi WordCloud digunakan untuk menggambarkan kata dominan pada tiap kelas. Studi ini menegaskan efektivitas Logistic Regression untuk analisis ulasan aplikasi berbahasa Indonesia, meskipun tantangan besar masih terdapat pada ketidakseimbangan kelas minoritas.

Kata kunci: Analisis Sentimen, Shopee, Google Play, TF-IDF, Logistic Regression

1. Latar Belakang

Transformasi digital telah mendorong pertumbuhan pesat aplikasi belanja daring di Asia Tenggara. Indonesia, sebagai salah satu pasar terbesar, mencatat peningkatan signifikan penggunaan aplikasi e-commerce, salah satunya Shopee, yang kini tidak hanya menjadi sarana transaksi, tetapi juga sebagai platform interaksi sosial, promosi, dan layanan pembayaran digital[1].

Google Play Store sebagai toko aplikasi utama Android menyediakan ruang bagi pengguna untuk meninggalkan ulasan. Ulasan ini berperan ganda: membantu calon pengguna dalam mengambil keputusan serta memberikan masukan langsung bagi pengembang aplikasi[2]. Sejumlah penelitian menunjukkan bahwa kualitas ulasan aplikasi berkorelasi erat dengan peringkat aplikasi dan jumlah unduhan[3]. Berbeda dengan ulasan produk di dalam marketplace yang menekankan kualitas barang, ulasan aplikasi di Google Play lebih banyak berfokus pada performa teknis aplikasi, stabilitas, desain antarmuka, dan integrasi layanan logistik[4].

Dengan jumlah ulasan yang sangat besar, analisis manual tidak lagi efisien. Oleh karena itu, analisis sentimen berbasis *Natural Language Processing* (NLP) dan *machine learning* menjadi pendekatan yang banyak digunakan untuk mengklasifikasikan opini pengguna ke dalam kategori positif, negatif, atau netral[5]. Teknik ini telah berhasil diterapkan pada berbagai domain, termasuk media sosial, ulasan produk, dan layanan publik[6], [7]. Salah satu metode representasi teks yang paling populer adalah Term Frequency–Inverse Document Frequency (TF-IDF), yang sederhana namun efektif dalam mengolah data teks berdimensi tinggi[8]. Algoritma klasik seperti Logistic Regression dan Support Vector Machine (SVM) juga masih kompetitif, terutama pada teks pendek dan dataset berbahasa lokal[9].

Meski demikian, tantangan utama dalam analisis sentimen adalah ketidakseimbangan kelas. Sebagian besar pengguna aplikasi cenderung memberikan ulasan positif, sementara ulasan negatif dan netral relatif sedikit[10]. Kondisi ini dapat membuat model klasifikasi bias terhadap kelas mayoritas. Selain itu, ulasan netral sering kali bersifat ambigu, mengandung sentimen campuran, atau sekadar deskriptif tanpa ekspresi emosional[11][12].

Penelitian terdahulu menekankan pentingnya teknik penyeimbangan data seperti oversampling untuk mengatasi masalah tersebut, bahkan ada yang menyarankan pendekatan khusus seperti aspect-based sentiment analysis untuk menangani ulasan ambigu[13][14].

Di Indonesia, penelitian terkait analisis sentimen pada e-commerce lebih banyak berfokus pada ulasan produk Shopee. Model berbasis IndoBERT dilaporkan mampu mencapai akurasi tinggi pada ulasan produk, sementara pendekatan klasik seperti Naive Bayes juga cukup efektif. Akan tetapi, mayoritas studi masih menekankan ulasan produk, bukan ulasan aplikasi di Google Play, padahal ulasan aplikasi lebih relevan untuk mengevaluasi aspek teknis, performa, dan pengalaman pengguna[15].

Selain itu, studi internasional menunjukkan bahwa model klasik sering kali lebih stabil dibandingkan model kompleks ketika jumlah data terbatas atau distribusi data timpang[16]. Validasi ketat melalui cross-validation juga ditekankan untuk menghindari klaim performa yang bias.

Berdasarkan konteks tersebut, penelitian ini difokuskan pada analisis sentimen ulasan aplikasi Shopee Indonesia di Google Play periode 2024–2025. Tujuan penelitian adalah: (1) menganalisis pengaruh ketidakseimbangan kelas terhadap performa model, (2) menguji efektivitas Random Oversampling pada data latih, dan (3) membandingkan kinerja Logistic Regression dengan baseline klasik seperti SVM, LinearSVC, dan Naive Bayes. Kontribusi penelitian ini adalah: (1) menyediakan dataset terbaru ulasan aplikasi Shopee 2024–2025, (2) membangun pipeline anti-leakage untuk balancing dan evaluasi, serta (3) memberikan evaluasi empiris baseline model klasik untuk ulasan aplikasi berbahasa Indonesia.

2. Metode Penelitian

2.1. Pengumpulan Data

Data penelitian dikumpulkan dari Google Play Store menggunakan pustaka Python *google-play-scraper*. Target pengambilan adalah aplikasi Shopee Indonesia dengan package com.shopee.id, difilter pada bahasa Indonesia dan negara Indonesia. Ulasan difokuskan pada periode 2024–2025. Setiap ulasan terdiri dari teks, skor rating, dan tanggal unggah. Label sentimen ditentukan secara otomatis: skor 1–2 sebagai negatif, 3 sebagai netral, dan 4–5 sebagai positif. Total data yang terkumpul sebanyak 5.000 ulasan dengan distribusi awal 4049 positif, 796 negatif, dan 155 netral. Scraping dilakukan dengan jeda waktu agar sesuai dengan *rate limiting* Google Play, serta semua data diperlakukan anonim tanpa informasi personal.

```
all_reviews = []
continuation_token = None
batches = 50   # 50 x 100 = ~5000
for i in range(batches):
    result, continuation_token = reviews(
        'com.shopee.id',
        lang='id', country='id',
        count=100,
        continuation_token=continuation_token
    )
    all_reviews.extend(result)
    print(f'Batch {i+1}/{batches} selesai, total ulasan: {len(all_reviews)}')
    time.sleep(2)  # rate limiting

df = pd.DataFrame(all_reviews)

# Keep only columns we need, and filter 2024-2025
df['date'] = pd.to_datetime(df['at'])
df = df[df['date'].dt.year.isin([2024, 2025])]
df = df.dropna(subset=['content', 'score']).copy()
```

Gambar 1. Source Code Pegumpulan Data

Berikut adalah data sampling tabel ulasan aplikasi shopee yang telah difilter. Ulasan terdiri dari kolom tanggal, skor rating, dan teks. Terdapat 15 data sampling.

Tabel 1. Sampling Ulasan Aplikasi Shopee

Tanggal	Skor Rating	Ulasan
2025-09-18	5	bagus orang pergi belanja tingal shopee aja goddd
2025-09-18	5	suka banget apk
2025-09-18	2	blockzi tdk menangkan

DOI: https://doi.org/10.31004/riggs.v4i2.2850
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

2025-09-18	3	tolong buka shopee langsung buka live video risih kadang nge bug ganti tab suara
2025-09-18	5	membantu harga bersahabat
2025-09-18	5	gampang urusannya
2025-09-18	1	tolonglah pisahkan isi nya tu food food aja shop shop aja karna pay udah dipisah trus tolonglah reels live nya yg muncul duluan pas pencet apps nya berat
2025-09-14	1	shopee update mulu gimana orang nggak protes pesanan yg belom sampe yg bahan makanan udah seminggu gmn shopee orang nggak betah aplikasi
2025-09-14	2	hati2 berbelanja sy beli pencuci kendaraan hrg 300rb tp barang tdk bs dipakai alasan macam2 tp tdk dikembalikan dibalikin uangnya nya tp ngomong aja basa basi sy disuruh bengkel diganti ongkos service tp basa basi chat tdk dibalas aplikasi tolong penjual nakal dikasi sanksi jngn hny kasih reward bintang jelek tp mrk bs bikin akun lg
2025-09-14	4	orderan tpt pelayanan td sesuai kembalikan ongkir tg shopis is de bess
2024-12-09	3	shopee bngt barang memuaskan
2024-12-09	2	shoppe gak cod pengiriman banget pengiriman banget beda pulau sampe minggu
2024-12-09	5	memudahkan berbelanja
2024-12-09	5	shopee tp bonus yg sdh titipan belanja teman2ku pulsa murah bonus spaylater mantap berguna tenor pembayaran kredit ponsel 12 pembayaran mantap
2024-12-09	5	shopee mempermudah belanja rumah

2.2. Preprocessing Teks

Tahap *preprocessing* dilakukan untuk mempersiapkan teks ulasan. Seluruh teks dikonversi menjadi huruf kecil, URL, angka, tanda baca, dan karakter khusus dihapus. *Stopword removal* dilakukan menggunakan daftar stopword bahasa Indonesia dari NLTK. Proses *stemming* dengan library Sastrawi digunakan untuk mengubah kata ke bentuk dasar[17]. Ulasan kosong atau duplikat dihapus untuk menjaga konsistensi dataset.

2.3. Ekstraksi Fitur

Ekstraksi fitur dilakukan dengan metode TF-IDF unigram–bigram. Parameter min_df=2 digunakan agar kata yang sangat jarang muncul diabaikan, sedangkan max_df=0.9 untuk mengecualikan kata yang terlalu umum. Representasi TF-IDF dipilih karena sederhana, stabil, dan efektif dalam memproses data teks berdimensi tinggi.

2.4. Penanganan Kelas Tidak Seimbang

Distribusi awal data menunjukkan dominasi kelas positif sebesar 80,98% dan kelas netral hanya 3,1%. Untuk mengatasi ketidakseimbangan tersebut, penelitian ini menggunakan Random Oversampling pada data latih. Teknik ini menduplikasi ulasan minoritas sehingga distribusi menjadi lebih seimbang. Oversampling hanya diterapkan pada data latih di dalam pipeline untuk menghindari *data leakage*. Pendekatan ini dianggap lebih sesuai dibanding SMOTE untuk representasi sparse TF-IDF.

```
Distribusi awal:
 sentiment
positive    4049
negative     796
neutral      155
Name: count, dtype: int64
Train size: 3866, Test size: 967
```

Gambar 2. Distribusi Awal Data Sentimen

2.5. Pemodelan

Beberapa algoritma klasifikasi diuji, meliputi Logistic Regression (solver liblinear), SVM linear, LinearSVC (dengan kalibrasi), dan Multinomial Naive Bayes. Semua model dikemas dalam pipeline yang mencakup TF-IDF,

DOI: https://doi.org/10.31004/riggs.v4i2.2850
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Random Oversampling, dan classifier. Dengan demikian, proses berjalan secara *end-to-end* dari teks mentah hingga prediksi sentimen.

2.6. Validasi dan Evaluasi

Dataset dibagi dengan stratified split 80:20 agar distribusi kelas terjaga pada data latih dan uji. Selain itu, dilakukan 5-fold cross-validation pada data latih dengan metrik macro-F1 untuk mengevaluasi kestabilan model. Evaluasi akhir di data uji menggunakan akurasi, precision, recall, F1-score, confusion matrix, serta macro ROC-AUC. Visualisasi *WordCloud* digunakan untuk menampilkan kata dominan tiap kelas, dan analisis kesalahan dilakukan untuk meninjau ulasan yang salah klasifikasi,

3. Hasil dan Diskusi

3.1 Distribusi Data

Distribusi awal ulasan memperlihatkan ketimpangan yang signifikan dengan dominasi ulasan positif sebesar 80,98%, negatif 15,95%, dan netral 3,1%. Hal ini menunjukkan perlunya penanganan kelas tidak seimbang.

Tabel 2. Distribusi Awal Ulasan

Kelas	Jumlah	Proporsi
Positif	4049	80,98%
Negatif	796	15,92%
Netral	155	3,1%

3.2 Hasil Cross-Validation

Hasil 5-fold CV pada data latih menunjukkan Logistic Regression sebagai model terbaik.

Tabel 3. Hasil Cross-Validation (Macro-F1)

Model	Mean ± Std
Logistic Regression	0.5568 ± 0.0172
LinearSVC Calibrated	0.5449 ± 0.0183
SVM Linear Calibrated	0.5467 ± 0.0226
Multinomial NB	0.5356 ± 0.0125

3.3 Evaluasi di Data Uji

Pada 967 ulasan data uji, Logistic Regression menghasilkan akurasi 85,11% dengan macro-F1 0,58.

Tabel 4. Classification Report

Kelas	Precision	Recall	F1-score	Support
Negatif	0.6387	0.7722	0.6991	158
Netral	0.1200	0.1667	0.1406	30
Positif	0.9574	0.8935	0.9243	779
Akurasi	-	-	0.8511	967
Macro avg	0.5668	0.6108	0.5880	-
Weighted avg	0.8788	0.8511	0.8628	-

Tabel 5. Confusion Matrix

Tue/Pred	Negatif	Netral	Positif
Negatif	122	16	20
Netral	14	5	11
Positif	28	55	696

Hasil menunjukkan ulasan positif terklasifikasi baik, sedangkan ulasan netral sering salah dikategorikan sebagai positif. Nilai macro ROC-AUC 0,8134 menunjukkan kemampuan pemisahan antar kelas cukup baik.

3.4 WordCloud dan Analisis Kesalahan

WordCloud menunjukkan ulasan positif didominasi kata “bagus”, “mantap”, dan “puas”, ulasan negatif memuat kata “tidak”, “kecewa”, dan “lama”, sedangkan ulasan netral banyak mengandung kata ambigu seperti “bisa” dan “aplikasi”. Analisis kesalahan menunjukkan ulasan campuran seperti *“bagus sih, tapi pengiriman lama”* sering menimbulkan salah klasifikasi.

DOI: https://doi.org/10.31004/riggs.v4i2.2850
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Gambar 3. *WordCloud* Ulasan Positif

Gambar 4. *WordCloud* Ulasan Netral

Gambar 5. WordCloud Ulasan Negatif

Diskusi

Hasil penelitian memperlihatkan bahwa Logistic Regression lebih konsisten dibandingkan SVM, LinearSVC, dan Naive Bayes dengan akurasi 85,11% dan macro-F1 0,58. Keunggulan ini sesuai dengan temuan terdahulu bahwa algoritma klasik masih kompetitif pada teks pendek dan dataset lokal. Penerapan Random Oversampling membantu mengurangi dominasi kelas positif dan meningkatkan recall pada kelas negatif, meskipun kinerja kelas netral tetap rendah. Hal ini menegaskan bahwa ulasan netral sulit diprediksi karena sering bersifat ambigu atau campuran, sejalan dengan literatur yang menyebutkan tantangan serupa. Analisis kesalahan menunjukkan bahwa kata-kata positif yang dominan membuat model cenderung mengklasifikasikan ulasan campuran sebagai positif. WordCloud juga menegaskan dominasi kata positif dibandingkan kata netral. Kondisi ini menunjukkan keterbatasan TF-IDF yang hanya berbasis frekuensi kata, tanpa memahami konteks. Oleh karena itu, penelitian lanjutan perlu mengeksplorasi pendekatan yang lebih kontekstual seperti aspect-based sentiment analysis atau embedding berbasis transformer bahasa Indonesia.

4. Kesimpulan

Penelitian ini menganalisis 5.000 ulasan aplikasi Shopee di Google Play periode 2024–2025 dengan TF-IDF dan Logistic Regression. Logistic Regression terbukti menjadi model terbaik dengan akurasi 85,11% dan macro-F1 0,58, sementara balancing dengan Random Oversampling berhasil memperbaiki distribusi kelas meski belum

DOI: https://doi.org/10.31004/riggs.v4i2.2850
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

mampu meningkatkan kinerja pada ulasan netral. Kontribusi utama penelitian ini adalah penyediaan dataset terbaru, pengembangan pipeline bebas *data leakage*, dan evaluasi baseline model klasik untuk ulasan aplikasi berbahasa Indonesia. Implikasi praktisnya, pipeline ini dapat digunakan pengembang aplikasi untuk memantau opini pengguna. Penelitian selanjutnya disarankan mengeksplorasi aspect-based sentiment analysis atau transformer bahasa Indonesia untuk meningkatkan prediksi, khususnya pada kelas minoritas.

Referensi

[1] A. Daza, N. D. González Rueda, M. S. Aguilar Sánchez, W. F. Robles Espíritu, and M. E. Chauca Quiñones, "Sentiment Analysis on E-Commerce Product Reviews Using Machine Learning and Deep Learning Algorithms: A Bibliometric Analysis, Systematic Literature Review, Challenges and Future Works," *International Journal of Information Management Data Insights*, vol. 4, no. 2, p. 100267, Nov. 2024, doi: 10.1016/J.JJIMEI.2024.100267.

[2] P. Vijayaragavan *et al.*, "Sustainable sentiment analysis on E-commerce platforms using a weighted parallel hybrid deep learning approach for smart cities applications," *Sci Rep*, vol. 14, no. 1, pp. 1–19, Dec. 2024, doi: 10.1038/S41598-024-78318-1;SUBJMETA.

[3] Q. Li *et al.*, "A Survey on Text Classification: From Traditional to Deep Learning," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 13, no. 2, p. 31, Apr. 2022, doi: 10.1145/3495162.

[4] C. A. Nurhaliza Agustina, R. Novita, Mustakim, and N. E. Rozanda, "The Implementation of TF-IDF and Word2Vec on Booster Vaccine Sentiment Analysis Using Support Vector Machine Algorithm," *Procedia Comput Sci*, vol. 234, pp. 156–163, Jan. 2024, doi: 10.1016/J.PROCS.2024.02.162.

[5] H. Suroyo and E. J. Pratama, "Comparison of Text Representation Methods for Sentiment Analysis Using Support Vector Machine," *Journal of Advances in Information and Industrial Technology*, vol. 7, no. 1, pp. 21–30, May 2025, doi: 10.52435/JAIIT.V7I1.610.

[6] P. S. Ghatora, S. E. Hosseini, S. Pervez, M. J. Iqbal, and N. Shaukat, "Sentiment Analysis of Product Reviews Using Machine Learning and Pre-Trained LLM," *Big Data and Cognitive Computing 2024, Vol. 8, Page 199*, vol. 8, no. 12, p. 199, Dec. 2024, doi: 10.3390/BDCC8120199.

[7] M. Mujahid *et al.*, "Data oversampling and imbalanced datasets: an investigation of performance for machine learning and feature engineering," *J Big Data*, vol. 11, no. 1, pp. 1–32, Dec. 2024, doi: 10.1186/S40537-024-00943-4/TABLES/15.

[8] A. R. Susanti and E. N. Ilahi, "Sentiment Analysis of User Reviews of E-commerce Applications: Case Study on the Shoppe Platform," *Journal of Social Science*, vol. 5, no. 4, pp. 983–988, Jul. 2024, doi: 10.46799/JSS.V5I4.885.

[9] Y. Mao, Q. Liu, and Y. Zhang, "Sentiment analysis methods, applications, and challenges: A systematic literature review," *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 4, p. 102048, Apr. 2024, doi: 10.1016/J.JKSUCI.2024.102048.

[10] A. Siddiqua, V. Bindumathi, G. Raghu, and Y. S. V. Bhargav, "Aspect-based Sentiment Analysis (ABSA) using Machine Learning Algorithms," *3rd IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics, ICDCECE 2024*, 2024, doi: 10.1109/ICDCECE60827.2024.10549140.

[11] F. Suandi *et al.*, "Enhancing Sentiment Analysis Performance Using SMOTE and Majority Voting in Machine Learning Algorithms," pp. 126–138, Dec. 2024, doi: 10.2991/978-94-6463-620-8_10.

[12] P. Monika, C. Kulkarni, N. Harish Kumar, S. Shruthi, and V. Vani, "Machine learning approaches for sentiment analysis," *Int J Health Sci (Qassim)*, pp. 1286–1300, Apr. 2022, doi: 10.53730/ijhs.v6ns4.6119.

[13] R. Kosasih and A. Alberto, "Sentiment analysis of game product on shopee using the TF-IDF method and naive bayes classifier," *ILKOM Jurnal Ilmiah*, vol. 13, no. 2, pp. 101–109, Aug. 2021, doi: 10.33096/ILKOM.V13I2.721.101-109.

[14] N. P. Arthamevia, Adiwijaya, and M. D. Purbolaksono, "Aspect-Based Sentiment Analysis in Beauty Product Reviews Using TF-IDF and SVM Algorithm," *2021 9th International Conference on Information and Communication Technology, ICoICT 2021*, pp. 197–201, Aug. 2021, doi: 10.1109/ICOICT52021.2021.9527489.

[15] C. Wang, X. Zhu, and L. Yan, "Sentiment Analysis for E-Commerce Reviews Based on Deep Learning Hybrid Model," *ACM International Conference Proceeding Series*, pp. 38–46, Aug. 2022, doi: 10.1145/3556384.3556391.

[16] G. O. Assunção, R. Izbicki, and M. O. Prates, "Is augmentation effective to improve prediction in imbalanced text datasets?," Apr. 2023, Accessed: Sep. 19, 2025. [Online]. Available: https://arxiv.org/pdf/2304.10283

[17] A. H. I. B. A. Azrir, N. Palanichamy, S. C. Haw, and K. W. Ng, "Improving Sentiment Analysis of Shopee Reviews with Informal Language and Slang," *Journal of Logistics, Informatics and Service Science*, vol. 11, no. 3, pp. 151–169, 2024, doi: 10.33168/JLISS.2024.0311.

DOI: https://doi.org/10.31004/riggs.v4i2.2850
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)