

Perbandingan Kinerja IndoBERT dan KNN Dalam Analisis Sentimen Ulasan Tokopedia Google Playstore

Gunawan^{*1}, Arief Ichwani²

^{1,2}Universitas Esa Unggul; Jl. Arjuna Utara No.9, Duri Kepa, Kec. Kb. Jeruk, Kota Jakarta Barat, Daerah Khusus Ibukota Jakarta 11510, Indonesia, +62 21 39529950

^{1,2}Jurusan Teknik Informatika, Universitas Esa Unggul, Jakarta

e-mail: ^{*1}gunawanlumbangaol77@gmail.com, ²arief.ichwani@esaunggul.ac.id

Abstrak

Perkembangan aplikasi e-commerce di Indonesia mendorong meningkatnya jumlah ulasan pengguna pada platform Google Playstore. Ulasan tersebut mengandung informasi penting mengenai persepsi dan tingkat kepuasan pengguna terhadap layanan aplikasi. Namun, besarnya volume data ulasan menyebabkan analisis manual menjadi tidak efisien. Oleh karena itu, diperlukan pendekatan otomatis berbasis Natural Language Processing (NLP) untuk mengklasifikasikan sentimen ulasan pengguna. Penelitian ini bertujuan untuk membandingkan kinerja model IndoBERT dan K-Nearest Neighbor (KNN) dalam analisis sentimen ulasan aplikasi Tokopedia di Google Playstore. Dataset yang digunakan berjumlah 12.172 ulasan pengguna yang diklasifikasikan ke dalam tiga kelas sentimen, yaitu positif, netral, dan negatif. Proses pelabelan sentimen dilakukan menggunakan kamus InSet. Evaluasi performa model dilakukan menggunakan metrik akurasi, precision, recall, dan F1-score, serta dianalisis menggunakan confusion matrix. Hasil penelitian menunjukkan bahwa model IndoBERT memiliki performa yang lebih unggul dibandingkan KNN dalam seluruh metrik evaluasi. Temuan ini menunjukkan bahwa model berbasis transformer lebih efektif dalam memahami konteks bahasa Indonesia pada data ulasan aplikasi.

Kata kunci—Analisis Sentimen, IndoBERT, K-Nearest Neighbor, NLP, Google Playstore

1. PENDAHULUAN

Perkembangan teknologi informasi telah mendorong transformasi signifikan pada sektor perdagangan digital di Indonesia. Aplikasi e-commerce menjadi sarana utama masyarakat dalam melakukan transaksi jual beli secara daring. Salah satu aplikasi e-commerce yang memiliki jumlah pengguna terbesar di Indonesia adalah Tokopedia. Tingginya jumlah pengguna Tokopedia berbanding lurus dengan banyaknya ulasan yang diberikan pada platform Google Playstore.

Ulasan pengguna pada Google Playstore merepresentasikan pengalaman, kepuasan, serta keluhan pengguna terhadap aplikasi. Informasi ini memiliki nilai strategis bagi pengembang aplikasi untuk mengevaluasi kualitas layanan dan melakukan perbaikan berkelanjutan. Namun, jumlah ulasan yang sangat besar menyebabkan proses analisis secara manual menjadi tidak efektif dan membutuhkan waktu yang lama. Oleh karena itu, diperlukan metode otomatis untuk menganalisis sentimen ulasan pengguna secara efisien.

Analisis sentimen merupakan salah satu cabang NLP yang bertujuan untuk mengklasifikasikan opini atau pendapat ke dalam kategori sentimen tertentu, umumnya positif, netral, dan negatif. Berbagai metode telah digunakan dalam penelitian sebelumnya, mulai dari metode machine learning klasik seperti Naive Bayes dan KNN hingga model deep learning.

Namun, metode klasik memiliki keterbatasan dalam memahami konteks semantik bahasa, khususnya pada bahasa Indonesia yang memiliki struktur linguistik kompleks.

Perkembangan model transformer seperti BERT menawarkan pendekatan yang lebih efektif karena mampu memahami konteks kata secara dua arah. IndoBERT merupakan model BERT yang telah dilatih khusus untuk bahasa Indonesia. Penelitian ini membandingkan kinerja IndoBERT dengan KNN sebagai representasi metode klasik dalam analisis sentimen ulasan aplikasi Tokopedia di Google Playstore. Novelty penelitian ini terletak pada penggunaan dataset aktual berskala besar serta analisis komparatif mendalam antara model transformer dan metode konvensional pada konteks aplikasi e-commerce di Indonesia.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan tahapan pengumpulan data, pra-pemrosesan, pelabelan sentimen, pemodelan, dan evaluasi.

2.1. Pengumpulan Data

Data penelitian berupa 12.172 ulasan pengguna aplikasi Tokopedia yang diambil dari Google Playstore menggunakan teknik web scraping dengan bantuan library Python google-play-scraper sesuai dengan periode pada skripsi. Data ini mencerminkan pengalaman nyata pengguna terhadap aplikasi.

2.2. Pra-pemrosesan Data

Tahapan pra-pemrosesan meliputi pembersihan teks, penghapusan karakter tidak relevan, case folding, tokenisasi, normalisasi kata tidak baku, serta penghapusan stopword bahasa Indonesia. Tahap ini bertujuan untuk meningkatkan kualitas data teks sebelum dilakukan pemodelan.

2.3. Pelabelan Sentimen

Pelabelan sentimen dilakukan menggunakan kamus InSet. Setiap ulasan diberikan skor sentimen berdasarkan kata-kata yang terkandung di dalamnya dan diklasifikasikan ke dalam kelas positif, netral, atau negatif.

2.4. Pemodelan

Model IndoBERT digunakan untuk merepresentasikan pendekatan deep learning berbasis transformer yang mampu memahami konteks semantik bahasa Indonesia. Sebagai pembanding, digunakan algoritma K-Nearest Neighbor (KNN) yang merupakan metode machine learning klasik berbasis kedekatan vektor fitur.

2.5. Evaluasi Model

Evaluasi performa model dilakukan menggunakan metrik akurasi, precision, recall, dan F1-score. Selain itu, confusion matrix digunakan untuk menganalisis pola kesalahan klasifikasi pada setiap kelas sentimen.

3. HASIL DAN PEMBAHASAN

3.1. Distribusi Data Sentimen

Tabel 1. Distribusi Label Sentimen

Label Sentimen	Jumlah	Persentase	Interpretasi
Positif	5.533	45,5%	Mayoritas puas dengan layanan
Netral	3.340	27,4%	Opini informatif tanpa emosi kuat
Negatif	3.299	27,1%	Kekecewaan terhadap layanan
Total	12.172	100%	Dataset seimbang

Tabel 1 menunjukkan distribusi data sentimen hasil pelabelan. Mayoritas ulasan termasuk dalam kategori sentimen positif, yang mengindikasikan bahwa sebagian besar pengguna memberikan respons yang baik terhadap aplikasi Tokopedia. Sentimen netral dan negatif muncul pada ulasan yang bersifat informatif atau mengandung keluhan pengguna.

3.2. Visualisasi WordCloud



Gambar 1. Visualisasi Perbandingan WordCloud

Gambar 1 menunjukkan perbandingan visual WordCloud sebelum dan sesudah dilakukan tahap pra-pemrosesan data. Pada WordCloud sebelum preprocessing, terlihat bahwa kata-kata yang sering muncul masih didominasi oleh stopword dan kata umum seperti “di”, “yang”, dan “dan”, yang tidak memiliki kontribusi signifikan terhadap analisis sentimen. Kondisi ini menunjukkan bahwa data mentah masih mengandung banyak noise yang dapat mengganggu proses pemodelan.

Setelah dilakukan preprocessing, WordCloud memperlihatkan perubahan yang signifikan, di mana kata-kata yang muncul didominasi oleh istilah yang lebih bermakna seperti “tokopedia”, “aplikasi”, “cepat”, dan “bagus”. Perubahan ini mengindikasikan bahwa proses preprocessing berhasil menghilangkan kata-kata tidak relevan dan meningkatkan kualitas representasi teks. Dengan demikian, data yang digunakan pada tahap pemodelan telah lebih bersih dan informatif, sehingga berpotensi meningkatkan performa model analisis sentimen.



Gambar 2. WordCloud Sentimen Positif

Gambar 2 menampilkan WordCloud untuk kelas sentimen positif. Kata-kata yang muncul dengan ukuran dominan antara lain “cepat”, “puas”, “mantap”, “murah”, dan “promo”. Dominasi kata-kata tersebut menunjukkan bahwa pengguna aplikasi Tokopedia memberikan apresiasi terhadap kecepatan layanan, kemudahan penggunaan aplikasi, serta keuntungan promosi yang ditawarkan.

Hasil visualisasi ini memperkuat temuan kuantitatif dari distribusi sentimen yang menunjukkan dominasi sentimen positif. WordCloud sentimen positif memberikan gambaran kualitatif mengenai faktor-faktor utama yang memengaruhi kepuasan pengguna, sehingga dapat menjadi masukan penting bagi pengembang aplikasi dalam mempertahankan dan meningkatkan kualitas layanan.



Gambar 3. WordCloud Gabungan Seluruh Sentimen

Gambar 3 menampilkan WordCloud gabungan dari seluruh kelas sentimen. Kata-kata yang paling sering muncul mencerminkan fokus utama pengguna dalam memberikan ulasan, seperti “aplikasi”, “transaksi”, “pengiriman”, dan “layanan”. Hal ini menunjukkan bahwa ulasan pengguna tidak hanya berfokus pada pengalaman belanja, tetapi juga pada performa aplikasi secara keseluruhan.

Visualisasi WordCloud gabungan ini memberikan gambaran umum mengenai topik-topik utama yang menjadi perhatian pengguna Tokopedia di Google Playstore. Informasi ini penting sebagai konteks pendukung dalam memahami hasil analisis sentimen dan sebagai dasar untuk mengidentifikasi aspek layanan yang perlu dipertahankan maupun ditingkatkan.

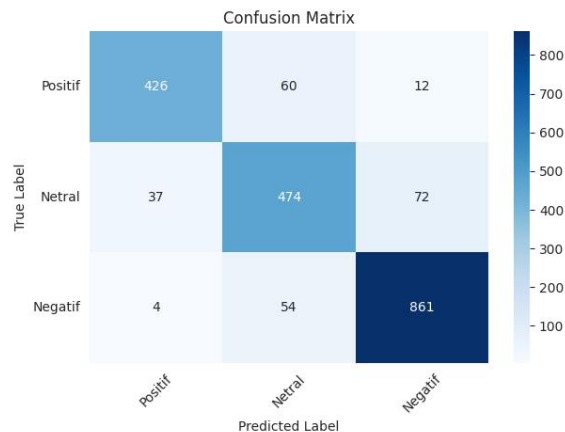
3.3. Hasil Evaluasi Model IndoBERT

Evaluasi performa model IndoBERT dilakukan untuk mengukur kemampuan model dalam melakukan klasifikasi sentimen terhadap ulasan aplikasi Tokopedia di Google Playstore. Pengujian dilakukan menggunakan data hasil preprocessing sebanyak 12.172 ulasan, yang telah melalui proses pelabelan sentimen. Metode evaluasi yang digunakan meliputi metrik akurasi, precision, recall, dan F1-score, serta visualisasi confusion matrix untuk memberikan gambaran kinerja model secara lebih komprehensif.

Tabel 2. Hasil Evaluasi Metrik IndoBERT

Sentimen	Precision	Recall	F1-score	Support
Negatif	0,912	0,855	0,883	498
Netral	0,806	0,813	0,810	583
Positif	0,911	0,937	0,924	919
Accuracy	0,880	0,880	0,880	0,880
Macro Average	0,876	0,868	0,872	2000
Weighted Average	0,881	0,880	0,880	2000

Berdasarkan hasil pengujian yang ditampilkan pada Tabel 2, model IndoBERT memperoleh nilai akurasi sebesar 88,05%. Nilai tersebut menunjukkan bahwa IndoBERT mampu mengklasifikasikan sentimen ulasan dengan tingkat ketepatan yang tinggi. Selain akurasi, nilai precision dan recall pada masing-masing kelas sentimen juga menunjukkan performa yang stabil dan seimbang, yang tercermin dari nilai F1-score yang relatif tinggi.



Gambar 4. Confusion Matrix IndoBERT

Confusion matrix pada Gambar 4 menunjukkan bahwa model IndoBERT mampu mengklasifikasikan sebagian besar data uji dengan tepat pada masing-masing kelas sentimen. Jumlah prediksi benar pada diagonal utama terlihat dominan, yang mengindikasikan tingkat kesalahan klasifikasi yang relatif rendah.

Kesalahan klasifikasi yang terjadi umumnya berada pada kelas sentimen yang memiliki karakteristik teks yang ambigu, seperti ulasan yang mengandung pujian dan kritik secara bersamaan. Meskipun demikian, IndoBERT tetap mampu mempertahankan keseimbangan antara precision dan recall pada setiap kelas sentimen, yang tercermin dari nilai F1-score yang tinggi.

Hasil ini menunjukkan bahwa IndoBERT memiliki kemampuan yang baik dalam memahami konteks semantik bahasa Indonesia, termasuk variasi bahasa informal dan ekspresi subjektif yang umum digunakan pada ulasan aplikasi. Dengan demikian, confusion matrix ini memperkuat hasil evaluasi kuantitatif yang menunjukkan keunggulan performa IndoBERT dalam tugas analisis sentimen.

3.4. Hasil Evaluasi Model K-Nearest Neighbor (KNN)

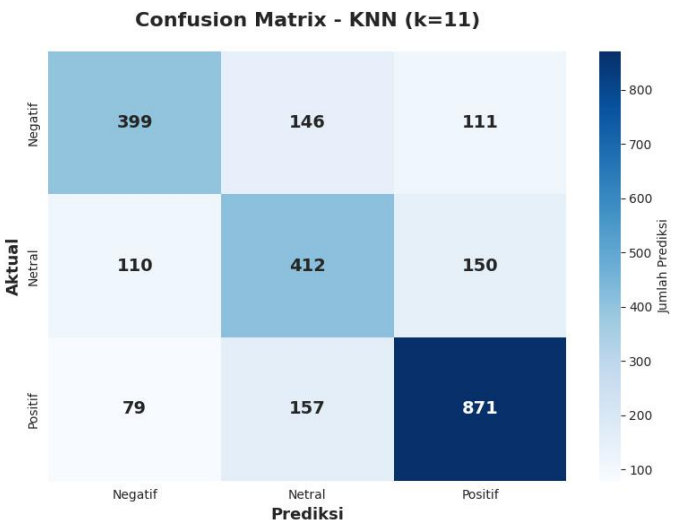
Evaluasi model K-Nearest Neighbor (KNN) dilakukan dengan tujuan membandingkan performanya terhadap model berbasis deep learning, yaitu IndoBERT. Model KNN diuji menggunakan dataset yang sama, yaitu 12.172 ulasan hasil preprocessing, sehingga perbandingan performa dapat dilakukan secara adil dan objektif.

Tabel 3. Hasil Evaluasi Metrik IndoBERT

Sentimen	Precision	Recall	F1-Score	Support
Negatif	0.679	0.608	0.641	656
Netral	0.576	0.613	0.594	672
Positif	0.769	0.787	0.778	1107
Accuracy	0.691	0.691	0.691	2435
Macro Avg	0.675	0.669	0.671	2435
Weighted Avg	0.692	0.691	0.690	2435

Hasil evaluasi model KNN yang ditampilkan pada Tabel 3 menunjukkan bahwa model ini memperoleh nilai akurasi sebesar 69,08%. Nilai ini secara signifikan lebih rendah dibandingkan dengan IndoBERT. Selain itu, nilai precision, recall, dan F1-score pada masing-

masing kelas sentimen juga menunjukkan performa yang kurang optimal, terutama pada kelas sentimen yang memiliki variasi kata dan konteks yang lebih kompleks.



Gambar 5. Confusion Matrix KNN

Confusion matrix pada Gambar 5 memperlihatkan bahwa model KNN menghasilkan jumlah kesalahan klasifikasi yang lebih tinggi dibandingkan dengan IndoBERT. Hal ini ditunjukkan oleh banyaknya data yang terklasifikasi ke kelas sentimen yang tidak sesuai, terutama pada kelas yang memiliki kemiripan kata antar sentimen.

Kesalahan klasifikasi tersebut menunjukkan keterbatasan KNN dalam menangkap konteks dan makna kalimat secara menyeluruh. Model ini cenderung mengandalkan kemiripan jarak antar fitur tanpa mempertimbangkan hubungan semantik antar kata, sehingga kurang efektif dalam menangani teks ulasan yang bersifat kompleks dan kontekstual.

Meskipun KNN masih mampu mengklasifikasikan sebagian data dengan benar, hasil confusion matrix menunjukkan bahwa performanya belum optimal untuk analisis sentimen ulasan aplikasi dengan variasi bahasa yang tinggi. Temuan ini sejalan dengan nilai akurasi dan F1-score yang lebih rendah dibandingkan dengan model IndoBERT.

3.5. Perbandingan Evaluasi Model

Tabel 4. Perbandingan Performa Model IndoBERT dan KNN

Metrik	IndoBERT	KNN	Selisih	Keterangan
Accuracy	88,0%	69,1%	+18,9%	IndoBERT unggul signifikan dalam akurasi keseluruhan klasifikasi tiga kelas.
Precision (Weighted)	0,881	0,692	+0,189	Ketepatan prediksi IndoBERT lebih tinggi dengan tingkat false positive lebih rendah.
Recall (Weighted)	0,880	0,691	+0,189	Kemampuan IndoBERT lebih baik dalam menjangkau instance sebenarnya yang harus terprediksi.
F1-Score (Weighted)	0,880	0,690	+0,190	Keseimbangan presisi dan recall IndoBERT lebih unggul secara konsisten di semua kelas.
F1 Positif	0,924	0,778	+0,146	IndoBERT jauh lebih efektif mengidentifikasi ulasan

F1 Netral	0,810	0,594	+0,216	positif, memahami nuansa apresiatif. Keunggulan terbesar kedua; IndoBERT lebih mampu menangani ambiguitas bahasa netral.
F1 Negatif	0,883	0,641	+0,242	Keunggulan terbesar; IndoBERT sangat superior mendeteksi keluhan dan sentimen negatif.

Perbandingan performa antara model IndoBERT dan KNN ditunjukkan pada Tabel 4, yang merangkum nilai akurasi, precision, recall, dan F1-score dari kedua model. Hasil perbandingan menunjukkan bahwa IndoBERT secara konsisten unggul pada seluruh metrik evaluasi dibandingkan dengan KNN.

IndoBERT mencapai nilai akurasi 88,05%, sedangkan KNN hanya mencapai 69,08%. Selisih akurasi yang cukup besar ini menunjukkan bahwa model berbasis transformer lebih efektif dalam menangani permasalahan analisis sentimen pada teks berbahasa Indonesia. Keunggulan IndoBERT terutama terlihat pada kemampuannya memahami konteks kalimat dan hubungan semantik antar kata, yang tidak dapat dicapai oleh metode berbasis jarak seperti KNN.

Hasil ini sejalan dengan penelitian-penelitian sebelumnya yang menyatakan bahwa model berbasis deep learning, khususnya transformer, memiliki performa yang lebih baik dibandingkan metode machine learning klasik dalam tugas klasifikasi teks. Dengan demikian, penggunaan IndoBERT dalam analisis sentimen ulasan aplikasi Tokopedia terbukti relevan dan tepat secara metodologis.

Secara keseluruhan, hasil perbandingan ini menegaskan bahwa IndoBERT merupakan model yang lebih direkomendasikan untuk analisis sentimen ulasan aplikasi, terutama pada data dengan jumlah besar dan variasi bahasa yang tinggi. Temuan ini juga memperkuat kontribusi penelitian dalam memberikan bukti empiris terkait efektivitas IndoBERT dibandingkan KNN pada konteks ulasan aplikasi di Google Playstore.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa IndoBERT memiliki kinerja yang lebih unggul dibandingkan KNN dalam analisis sentimen ulasan aplikasi Tokopedia di Google Playstore. Keunggulan IndoBERT terletak pada kemampuannya memahami konteks semantik bahasa Indonesia secara lebih mendalam.

5. SARAN

Penelitian selanjutnya dapat mengeksplorasi penggunaan model transformer lain atau pendekatan pelabelan manual untuk meningkatkan kualitas analisis sentimen. Selain itu, objek penelitian dapat diperluas pada aplikasi e-commerce lain untuk memperoleh hasil yang lebih general.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada **Bapak Arief Ichwani, ST, MT** yang telah memberi **“dukungan”** terhadap penelitian ini.

DAFTAR PUSTAKA

- [1] Aji, B. P., & Aditya, R. (2025). Implementasi IndoBERT untuk Klasifikasi Sentimen Ulasan E-Commerce Indonesia. *Jurnal Teknologi Informasi Dan Komputer*, 15(1), 23–35.
- [2] Budiharto, W. (2016). *Machine Learning untuk Klasifikasi Data Mining*. Penerbit Andi.
- [3] Chevalier, J. A., & Mayzlin, D. (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research*, 43(3), 345–354.
- [4] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *ArXiv Preprint ArXiv:1810.04805*.
- [5] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2020). BERT: Bidirectional Encoder Representations from Transformers. *Journal of Machine Learning Research*, 21(1), 1–15.
- [6] Feldman, R., & Sanger, J. (2007). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press.
- [7] Group, G. (2023). *Annual Report 2023: Tokopedia Performance and Market Reach*. GoTo Group Indonesia.
- [8] Han, J., Kamber, M., & Pei, J. (2022). *Data Mining: Concepts and Techniques* (4 (ed.)). Morgan Kaufmann Publishers.
- [9] Kasim Aidid, M., Dewi Putri, A., & Rais, Z. (2025). Penerapan Algoritma K-Nearest Neighbor (K-NN) untuk Analisis Sentimen Ulasan Aplikasi Lazada. *Jurnal Teknologi Informasi*, 13(1), 45–58.
- [10] Kotler, P., & Keller, K. L. (2016). *Marketing Management* (15 (ed.)). Pearson Education Limited.
- [11] Koto, F., Rahimi, A., Lau, J. H., & Baldwin, T. (2020). IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP. *Proceedings of the 28th International Conference on Computational Linguistics*, 757–770.
- [12] Koto, F., & Rahmaningtyas, G. Y. (2017). InSet Lexicon: Evaluation of a Word List for Indonesian Sentiment Analysis in Microblogs BT - 2017 International Conference on Asian Language Processing (IALP).
- [13] Kusuma, I. H., & Cahyono, N. (2023). Analisis Sentimen Masyarakat Terhadap Penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor dan Natural Language Processing. *Jurnal Sistem Informasi Bisnis*, 13(2), 156–168.
- [14] Lillah, R. R., Maylawati, D. S. S., Budiawan, W., Zulfikar, W. B., Uriawan, W., & Wahana, A. (2023). Implementasi Algoritma K-Nearest Neighbor (KNN) untuk Analisis Sentimen Pengguna Aplikasi Tokopedia. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 7(4), 892–899.
- [15] Limbong, J. J. A., Sembiring, I., & Hartomo, K. D. (2022). Analisis Klasifikasi Sentimen Ulasan Pada E-commerce Shopee Berbasis Word Cloud Dengan Metode Naïve Bayes Dan K-Nearest Neighbor. *Jurnal Media Informatika Budidarma*, 6(3), 1725–1734.
- [16] Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.
- [17] Luqyana, W. A., Cholissodin, I., & Perdana, R. S. (2018). Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(11), 4704–4713.
- [18] Mas, R., Panca, R. W., Atmaja1, K., & Yustanti2, W. (2021). Analisis Sentimen Customer Review Aplikasi Ruang Guru Dengan Metode BERT (Bidirectional Encoder Representations from Transformers). *Journal of Emerging Information Systems and Business Intelligence (JEISBI)*, 2(3), 2021. <https://doi.org/10.26740/JEISBI.V2I3.41567>

- [19] Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment Analysis Algorithms and Applications: A Survey. *Ain Shams Engineering Journal*, 5(4), 1093–1113.
- [20] Park, D. H., Lee, J., & Han, I. (2007). The Effect of On-Line Consumer Reviews on Consumer Purchasing Intention: The Moderating Role of Involvement. *International Journal of Electronic Commerce*, 11(4), 125–148.
- [21] Powers, D. M. W. (2020). Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation. *Journal of Machine Learning Technologies*, 2(1), 37–63.
- [22] Rahman, M., & Ikhsan, M. F. (2025). Sentiment Analysis pada Platform E-Commerce: Pendekatan Berbasis Deep Learning. *Jurnal Kecerdasan Buatan*, 8(2), 89–101.
- [23] Rahman, R. D., Setiawan, N. Y., & Bachtiar, F. A. (2025). Analisis Sentimen Berbasis Aspek Menggunakan BERT dan KMeans Clustering pada Review Aplikasi Blibli. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 9(2), 234–245.
- [24] Rizqi, R. M., & Yustanti, W. (2021). Analisis Sentimen Customer Review Aplikasi Ruang Guru dengan Metode BERT (Bidirectional Encoder Representations from Transformers). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 5(9), 3657–3663.
- [25] Saifurridho, M., Martanto, & Hayati, U. (2023). Analisis Sentimen Ulasan Pengguna Aplikasi Shopee di Google Playstore Menggunakan K-Nearest Neighbor (K-NN). *Jurnal Sistem Informasi Dan Teknologi*, 11(3), 234–242.
- [26] Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a Distilled Version of BERT: Smaller, Faster, Cheaper and Lighter. *ArXiv Preprint ArXiv:1910.01108*.
- [27] Sarkar, D., Bali, R., & Sharma, T. (2018). *Practical Machine Learning with Python: A Problem-Solver's Guide to Building Real-World Intelligent Systems*. Apress.
- [28] Schindler, R. M., & Bickart, B. (2005). Published Word of Mouth: Referable, Consumer-Generated Information on the Internet. In C. P. Haugtvedt, K. A. Machleit, & R. F. Yalch (Eds.), *Online Consumer Psychology: Understanding and Influencing Consumer Behavior in the Virtual World* (pp. 35–61). Lawrence Erlbaum Associates.
- [29] Sharda, R., Delen, D., & Turban, E. (2021). *Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support* (11 (ed.)). Pearson Education Limited.
- [30] Sharma, A., Kumar, R., & Singh, P. (2021). Google Colaboratory for Deep Learning: A Comprehensive Study. *International Journal of Computer Science and Mobile Computing*, 10(5), 89–98.
- [31] Sofjan, A. (2017). *Manajemen Pemasaran: Dasar, Konsep, dan Strategi*. PT RajaGrafindo Persada.
- [32] Stanton, W. J., Etzel, M. J., & Walker, B. J. (1994). *Fundamentals of Marketing* (10 (ed.)). McGraw-Hill.
- [33] Sutopo, W., Kusumaningsih, A., & Nurlaeli, L. (2020). E-Marketplace untuk UMKM dalam Meningkatkan Daya Saing Produk. *Jurnal Manajemen Teknologi*, 19(2), 156–170.
- [34] Turban, E., Strauss, J., & Lai, L. (2015). *Social Commerce: Marketing, Technology and Management*. Springer International Publishing.
- [35] Tutuko, B., Fajri, F., & Sukemi, S. (2022). Membandingkan Nilai Akurasi BERT dan DistilBERT pada Dataset Twitter tentang COVID-19. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 9(4), 789–798.
- [36] Urva, G., Sharma, M., & Singh, R. (2023). Multi-level Sentiment Analysis: Document, Sentence and Aspect-based Approaches. *Journal of Natural Language Processing*, 18(2), 145–162.
- [37] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2020). Attention is All You Need. *Advances in Neural Information Processing Systems*, 30, 5998–6008.

- [38] Zhang, Y., Li, M., Wang, S., & Liu, B. (2020). Distance Metrics in K-Nearest Neighbor Algorithm: A Comprehensive Review. *Journal of Data Science and Analytics*, 12(3), 234–248.