

PayPal Usage in Indonesia with K-Nearest Neighbor Algorithm

Amannia Zeze Barki, Muhammad Ravi Azzaki, dan Dodi Vionanda*

Departemen Statistika, Universitas Negeri Padang, Padang, Indonesia

*Corresponding author: dodi_vionanda@fmipa.unp.ac.id

Submitted : 11 Juli 2025

Revised : 14 Agustus 2025

Accepted : 20 Agustus 2025

ABSTRACT

The development of information and digital technology has had a significant impact on the financial sector. In Indonesia, digital payment technologies such as PayPal, Gopay, Shopeepay, OVO, and DANA have become an integral part of the modern payment system. Since the implementation of the national electronic clearing system, RTGS, and ATMs in 2005, transactions have become increasingly convenient. This study analyzes user sentiment toward PayPal in Indonesia to understand user experience and provide insights for service development, marketing strategies, and brand reputation management. Review data from the PayPal app was collected from Google Plat via web scrapping and processed to yield 597 clean data points. Initial sentiment was categorized into positive, neutral, and negative, wordcloud visualization displayed positive and negative sentiment, while neutral sentiment was analyzed numerically. Automatic labeling was performed using the NLTK library based on rating values, above 3 positive, below 3 negative, and exactly 3 neutral. The results showed 146 positive reviews, 451 negative reviews, and a few neutral reviews. Sentiment classification using the K-Nearest Neighbor (K-NN) method yielded adequate accuracy, indicating that PayPal's acceptance in Indonesia is largely influenced by users' negative experiences. These findings provide a foundation for developing strategies to improve service quality and update PayPal's operational policies in the Indonesian market.

Keywords: Digital Payments, Indonesia, K-Nearest Neighbor, PayPal, Sentiment Analysis.



This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

I. PENDAHULUAN

Perkembangan teknologi informasi dan digital telah mengubah banyak aspek dalam kehidupan sehari – hari, termasuk dalam sektor keuangan. Di Indonesia, teknologi pembayaran digital telah memberikan dampak signifikan. Pada tahun 2005, sistem kliring elektronik secara nasional mulai diterapkan, memungkinkan proses transfer dana dilakukan secara elektronik di seluruh negeri (Bank Indonesia, 2005). Selain itu, munculnya layanan *real-time gross settlement* (RTGS) dan penggunaan mesin ATM sejak pertengahan 1980-an telah memudahkan masyarakat dalam melakukan berbagai jenis pembayaran. Penggunaan teknologi *Fintech* yang dimulai sekitar tahun 2017 semakin mempermudah transaksi jual beli. Layanan seperti uang elektronik (*e-money*) dan dompet digital (*e-wallet*) telah menjadi bagian penting dari sistem pembayaran modern. Contoh *e-wallet* yang populer di Indonesia adalah PayPal, Gopay, Shopeepay, OVO, dan DANA (Sedik, 2023).

Analisis sentimen terhadap PayPal dapat memberikan wawasan mendalam tentang pengalaman pengguna (IBM, 2024). Dengan mengidentifikasi aspek positif dan negatif dari layanan PayPal, perusahaan dapat mengambil langkah – langkah untuk meningkatkan kualitas layanan dan memenuhi kebutuhan pelanggan di setiap negara dengan lebih baik. Dengan mengetahui sentimen, pengguna dapat membantu PayPal dalam merumuskan strategi pemasaran yang lebih efektif di masing – masing negara. Sentimen positif dapat dimanfaatkan dalam kampanye pemasaran, sementara sentimen negatif dapat diatasi dengan komunikasi yang lebih baik dan perbaikan layanan.

Data sentimen dapat digunakan oleh manajemen PayPal untuk membuat keputusan bisnis yang lebih informasional dan berbasis data. Misalnya, jika banyak pengguna mengeluhkan tentang masalah tersebut untuk mencegah dampak negatif yang lebih luas. Dengan melakukan perbandingan analisis sentimen antara PayPal dan kompetitornya, perusahaan dapat mengidentifikasi kekuatan dan kelemahan relatifnya (Ivosights, 2023). Hal ini penting untuk tetap kompetitif di pasar yang sangat dinamis dan selalu berubah.

Analisis *sentiment* yang konsisten dan berkelanjutan dapat membantu dalam perbaikan terus – menerus layanan yang ditawarkan oleh PayPal. *Feedback* dari pengguna melalui analisis sentimen dapat diterapkan dalam pengembangan produk dan layanan baru. Sentimen pengguna yang diekspresikan di media sosial dan *platform online* lainnya sangat mempengaruhi reputasi merek. PayPal, dapat menggunakan analisis sentimen untuk memantau dan mengelola reputasi mereka, merespon masalah dengan cepat, dan menjaga hubungan baik dengan pelanggan.

Relevansi strategi ini dapat dilihat pada penelitian terdahulu yang menerapkan analisis *sentiment* pada layanan keuangan digital. Salah satunya adalah penelitian oleh Riskawati dkk, 2024 menggunakan algoritma naïve bayes untuk menganalisis persepsi pengguna terhadap aplikasi GoPay. Dengan menggunakan data ulasan dari *Google Play Store* sebanyak 1000 data yang dikumpulkan melalui *web scrapping*. Penelitian ini berhasil mengkategorikan ulasan menjadi kategori positif dan negatif. Hasil evaluasi menunjukkan akurasi 85%, presisi 84%, recall 96% dan skor f1 90%. Hasil ini memberi pengembang Gopay ide bagus untuk meningkatkan layanan mereka berdasarkan umpan balik pengguna dan menjadi dasar untuk metode analisis sentimen untuk layanan keuangan digital.

Dalam penelitian ini, metode yang digunakan adalah algoritma K-Nearest Neighbor (K-NN) untuk melakukan klasifikasi sentimen pada ulasan pengguna PayPal di Indonesia. Pemilihan metode ini didasarkan pada kemampuannya dalam mengelompokkan data berdasarkan jumlah tetangga terdekat, sehingga sesuai untuk memproses data teks yang telah melalui tahapan *preprocessing*. Beberapa penelitian sebelumnya juga telah berhasil menggunakan KNN untuk analisis sentimen. Penelitian yang dilakukan oleh Masturoh & Pohan, 2021 yang menerapkan KNN pada data ulasan aplikasi DANA dan memperoleh hasil akurasi 86,64% pada kelas ulasan yang terdiri dari penilaian bintang 1 dan 5 dengan $k = 1$. Hal ini menunjukkan bahwa KNN dapat memebrikan hasil yang kompetitif dan relevan untuk mengidentifikasi pola sentimen pengguna terhadap layanan digital.

Kasus ini membandingkan penggunaan PayPal di Indonesia dari perspektif pengguna, dengan fokus kepada analisis sentimen. Dengan memahami bagaimana pengalaman penggunaan PayPal di negara Indonesia, kita dapat lebih memperluas wawasan terkait faktor – faktor yang mempengaruhi penerimaan platform ini di pasar yang berbeda. Melalui penelitian ini, diharapkan dapat ditemukan hal – hal yang bermafaat sebagai pertimbangan untk perkembangan dan memperbaiki kebijakan layanan platform ini sendiri. Pada bagian selanjutnya, penelitian ini akan membahas landasan teoritis terkait pembayaran digital, perkembangan PayPal di Indonesia, serta kerangka analisis *sentiment* yang digunakan pada kasus ini.

II. METODE PENELITIAN

Pada penelitian ini teknik pengumpulan data akan dilakukan secara langsung menggunakan ulasan dari *Google Play* pada aplikasi “PayPal”. Data akan dianalisis penggunaan *python scrapper* yang dijalankan di *google collaboratory*. K-Nearest Neighbor digunakan sebagai suatu metode di dalam penelitian ini.

A. Sumber Data

Kumpulan ulasan diperoleh dengan menggunakan Teknik *Scrapping data* dari API yang disediakan oleh *google playstore* dan dibuat menjadi *python library* bernama *google-play-scrapper*. *Scrapping* dilakukan pada tanggal 13 Maret 2024, menggunakan *scrapping data*. Total jumlah ulasan yang penulis dapatkan dari hasil *Scrapping data* adalah 597 ulasan.

B. K-Nearest Neighbor

Algoritma K-Nearest Neighbor adalah metode klasifikasi objek baru berdasar tetangga terdekat dan kelas yang paling sering muncul sebagai hasil klasifikasi (Rivki & Bachtiar, 2017). K-Nearest Neighbor (K-NN) merupakan algoritma klasifikasi yang paling efektif, paling sederhana, paling cepat, dan paling mudah dipakai. Prinsip kerja K-NN adalah dengan mencari k data terdekat yang merupakan tetangga bagi data baru yang ingin diklasifikasikan atau diprediksi, kemudian menggunakan kelas atau nilai dari tetangga-tetangga itu untuk menentukan kelas atau nilai data baru. Metode ini digunakan adalah untuk mengetahui nilai akurasi yang dihasilkan itu. Algoritma K-NN adalah merupakan suatu teknik klasifikasi objek. Teknik ini didasarkan pada data pelatihan yang paling dekat dengan. Algoritma K-Nearest Neighbor (K-NN) adalah algoritma pembelajaran mesin yang sederhana sifatnya. Algoritma ini dapat pula digunakan untuk menyelesaikan masalah-masalah regresi dan klasifikasi. Algoritma merupakan jenis pembelajaran di bawah pengawasan. Metode non-parametrik tidak memperkirakan distribusi data yang mendasari itu.

Nilai yang lebih optimal akan dihasilkan pada klasifikasi teks dengan memakai metode K-NN jika rumus dari cosine similarity dipakai untuk pembobotan pada tiap kata di data yang diproses. Dalam proses pembobotan kata, tahapan tf , df , idf , $tf-idf$ harus dilakukan sebelum menghitung nilai *cosine similarity*.

C. *Scrapping*

Langkah awal dalam analisis sentimen terhadap aplikasi PayPal dimulai dengan proses *scrapping*, yaitu teknik pengambilan data ulasan dari situs Google Play Store menggunakan metode *web scrapping*. Data yang diperoleh kemudian disimpan dalam format file CSV. Dataset tersebut berupa data teks yang berisi opini, pendapat, maupun ekspresi emosi pengguna, yang selanjutnya akan dianalisis untuk memperoleh nilai sentimen dari masing-masing ulasan.

Istilah *scrapping* umum digunakan dalam bidang teknologi informasi dan pengembangan perangkat lunak (Rizaldi & Putranto, 2017). Secara umum, *scrapping* mengacu pada proses pengumpulan data secara otomatis dari suatu situs web. Proses ini sering dimanfaatkan untuk berbagai keperluan, seperti analisis bisnis atau pengembangan aplikasi yang berbasis data dari sumber daring. Dalam konteks klasifikasi dengan algoritma *K-Nearest Neighbor*, data hasil *scrapping* digunakan sebagai objek analisis, di mana algoritma akan mengidentifikasi k tetangga terdekat berdasarkan metrik jarak tertentu, seperti jarak geometris atau cosine similarity.

D. *Text Preprocessing*

Dalam proses *preprocessing*, dataset akan melewati sejumlah tahapan untuk dipersiapkan sebelum dianalisis. Tahap awal yang dilakukan adalah *case folding*, yaitu proses standarisasi dengan mengubah seluruh teks menjadi huruf kecil (lowercase). Selanjutnya, dilakukan *tokenizing*, yaitu pemecahan teks menjadi unit-unit kata tunggal atau token. Setelah itu, dilakukan proses *stopword removal* untuk menghapus kata-kata yang tidak memiliki kontribusi penting dalam analisis sentimen. Tahap akhir dari pra-pemrosesan ini adalah *stemming* atau normalisasi, yang bertujuan untuk menghilangkan imbuhan pada kata dan mengembalikannya ke bentuk dasar.

Tahapan pra-pemrosesan teks bertujuan untuk mengubah data mentah menjadi format data terstruktur guna mempermudah proses analisis sentimen maupun penerapan algoritma machine learning. Secara umum, proses ini mencakup tiga aktivitas utama, yaitu pembersihan, normalisasi, dan transformasi data. Menurut Furqan dkk. (2022), langkah-langkah pra-pemrosesan meliputi:

- 1) *Cleaning* yaitu menghapus elemen-elemen tidak relevan dalam data seperti tanda baca, angka, URL, nama pengguna, mention, hashtag, emotikon, serta simbol lainnya yang dapat mengganggu analisis.
- 2) *Case Folding* yakni proses mengonversi semua huruf dalam data menjadi huruf kecil (lowercase) agar konsistensi kata tetap terjaga.
- 3) *Stemming* yaitu mengubah kata berimbuhan menjadi bentuk dasarnya dengan menghapus prefix, sufiks, atau konfiks yang melekat pada kata tersebut.
- 4) *Tokenizing* yaitu memisahkan kalimat atau string menjadi satuan kata (token) sehingga dapat diproses lebih lanjut.
- 5) *Stopword Removal* atau *Filtering* yaitu penyaringan kata-kata umum yang tidak memberikan nilai penting dalam analisis, seperti kata sambung, preposisi, atau artikel.

E. *Pelabelan Sentimen*

Pelabelan ini akan memproses data secara otomatis yaitu melakukan perhitungan nilai skor menggunakan library NLTK (*Natural Language Toolkit*). Library NLTK ini menyediakan beberapa modul dan fungsi yang berguna untuk melakukan analisis sentiment atau pelabelan sentiment, diantaranya tersedia *lexicon sentiment* NLTK yang berisi daftar-daftar kata yang telah diberi label sentiment (positif atau negatif). Menghitung nilai skor pelabelan sentimen dengan metode perhitungan jumlah skor.

III. HASIL DAN PEMBAHASAN

A. *Analisis Pembagian Data*

Pada tahap ini, peneliti berupaya untuk mengidentifikasi serta permasalahan dalam penelitian. Peneliti ingin mengetahui sejauh mana efektivitas penggunaan algoritma *k-nearest neighbors* (K-NN) dalam mengklasifikasikan sentimen dari ulasan pengguna terhadap aplikasi dompet digital, khususnya PayPal.

Penelitian ini memanfaatkan data ulasan pengguna aplikasi PayPal yang diperoleh melalui tautan *Google Play*, dengan total data sebanyak 597 ulasan yang terdiri dari 451 review negatif dan 146 review positif.

Dibawah ini adalah contoh ulasan positif dan negatif dari aplikasi PayPal:

1. *Review Positif*
Sangat berguna untuk pembayaran luar maupun dalam bagi saya. Good setelah di update Terimakasih PayPal semoga kamu makin kaya raya
2. *Review Negatif*

Jujur bikin kesel pakai PayPal, sangat disayangkan aplikasi ini bermasalah, gara-gara aplikasi ini bermasalah saya jadi susah bertransaksi dengan orang luar negeri, Saya dirugikan! Ketika mau daftar akun malah GK bisa, tolong di perbaiki aplikasinya developer!!!!!!

Tabel 1. Pembagian Data Berdasarkan Sentimen

Sentimen	Jumlah data
Negatif	451
Netral	33
Positif	113
Total Data	597

Hasil klasifikasi sentimen yang ditampilkan pada Tabel 1 menunjukkan total 597 data ulasan, yang terdiri dari 113 ulasan positif, 451 ulasan negatif, dan 33 ulasan netral. Berdasarkan pembagian tersebut, diketahui bahwa 76% ulasan bersentimen negatif, 19% ulasan bersentimen positif, dan 5% ulasan bersentimen netral, yang dapat dilihat pada Gambar 1. Mayoritas pengguna PayPal memberikan ulasan yang didominasi oleh sentiment negatif, yang jumlahnya lebih tinggi dibandingkan sentiment positif maupun netral.

B. Pengolahan Data Awal

Variabel k dalam algoritma *K-Nearest Neighbor* (K-NN) merujuk pada jumlah tetangga terdekat yang digunakan dalam proses klasifikasi (Adhi Putra, 2021). Sebelum model dijalankan, tahap awal pengolahan data dilakukan melalui proses seleksi, yang meliputi pembersihan dan transformasi data ke dalam format yang sesuai. Seluruh data ulasan, yang berjumlah 597, disimpan dalam format CSV.

Selanjutnya, data dianalisis menggunakan metode K-NN dengan membagi dataset menjadi data latih (*training data*) dan data uji (*testing data*). Pada tahap awal, algoritma ini diuji menggunakan beberapa variasi nilai k , yaitu dari $k = 2$ hingga $k = 10$, guna menemukan nilai k yang menghasilkan tingkat akurasi tertinggi dalam klasifikasi.

C. Preprocessing Data

Pada tahap ini, peneliti menerapkan proses *preprocessing* terhadap data yang telah dikumpulkan. Adapun tahapan yang dilakukan mencakup:

1. *Remove duplicates*, langkah ini bertujuan untuk mengeliminasi data ulasan yang muncul lebih dari satu kali agar tidak terjadi pengulangan informasi.
2. *Case folding*, seluruh teks dalam ulasan dikonversi menjadi huruf kecil (lowercase) menggunakan fitur *transform cases* dalam perangkat lunak RapidMiner, agar konsistensi penulisan kata tetap terjaga.
3. *Tokenizing*, kalimat dalam ulasan dipecah menjadi kata per kata (token), sekaligus menghapus karakter pemisah, angka, dan tanda baca yang tidak relevan.
4. *Stopwords*, bertujuan untuk menyaring dan menghapus kata-kata yang sering muncul namun tidak memberikan kontribusi bermakna terhadap proses analisis, seperti kata sambung atau kata umum lainnya
5. *Stemming*, proses ini digunakan untuk menelusuri akar kata dari setiap token dengan menghapus imbuhan, baik awalan maupun akhiran, menggunakan file khusus yang telah disiapkan sebelumnya.

Tabel 2. Preprocessing Data

Tahapan	Hasil
Data Review Asli	Administrasi nya sangat mahal. Dan untuk transfer gagal akan kena denda RP.50.000 sangat mahal padahal biaya transfer di bawah 1 juta sudah Rp.16.000
Case Folding	administrasi nya sangat mahal dan untuk transfer gagal akan kena denda rp sangat mahal padahal biaya transfer di bawah juta sudah rp
Tokenizing	['administrasi', 'nya', 'mahal', 'transfer', 'gagal', 'kena', 'denda', 'rp', 'mahal', 'biaya', 'transfer', 'juta', 'rp']
Stopwords	administrasi nya mahal transfer gagal kena denda rp mahal biaya transfer juta rp

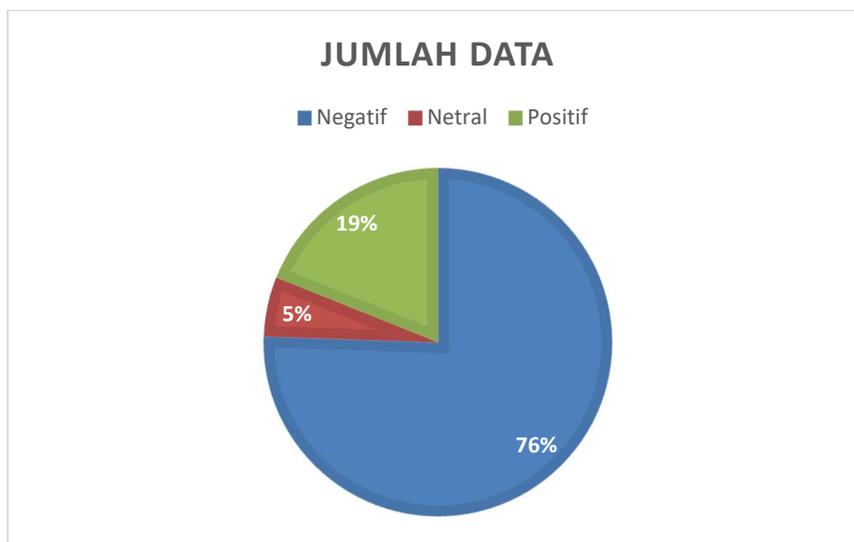
Stemming	['administrasi nya mahal transfer gagal kena denda rp mahal biaya transfer juta rp']
----------	--

Tabel 2 diatas menjelaskan tahapan dari preprocessing dari data review asli dan perubahannya jika dilakukan langkah preprocessing. Hasil ini yang akan dianalisis selanjutnya.

D. Pelabelan Sentimen

Penentuan skor sentimen dilakukan dengan cara menghitung selisih antara skor sentimen positif dan negatif pada setiap komentar pengguna. Dalam penelitian ini, klasifikasi sentimen dibagi ke dalam tiga kategori, yaitu positif, netral, dan negatif. Proses pelabelan dilakukan secara otomatis menggunakan pustaka NLTK yang dijalankan melalui platform *Google Colaboratory*.

Data yang diperoleh dari proses *scraping* kemudian dipilah dan melalui sejumlah tahapan *preprocessing* hingga menghasilkan total 597 data ulasan. Pada tahap pelabelan, ulasan dengan nilai rating di atas 3 diberi label positif, sedangkan ulasan dengan rating di bawah 3 diklasifikasikan sebagai negatif. Berdasarkan kriteria tersebut, diperoleh 146 ulasan dengan sentimen positif dan 451 ulasan dengan sentimen negatif.



Gambar 1. Pembagian Data Berdasarkan Sentimen



Gambar 2. Visualisasi Wordcloud Sentimen a. Negatif dan b. Positif

Berdasarkan Gambar 1, diketahui bahwa proporsi ulasan positif terhadap aplikasi PayPal sebesar 19%, sedangkan ulasan negatif mencapai 76%. Temuan ini menunjukkan bahwa persepsi publik terhadap penggunaan PayPal di Indonesia cenderung bernuansa negatif. Dengan kata lain, sebagian besar pengguna menyampaikan pengalaman yang kurang memuaskan saat menggunakan layanan tersebut. Visualisasi persebaran sentimen tersebut dapat dilihat lebih lanjut pada Gambar 2.

E. Hasil Pengujian K-Nearest Neighbor

Penelitian ini menerapkan algoritma K-Nearest Neighbor dalam proses klasifikasi terhadap 597 data ulasan yang telah dikumpulkan.

Tabel 2. Pengujian *Accuracy* berdasarkan Nilai K

Nilai K	<i>Accuracy</i>
2	0.78
3	0.70
4	0.78
5	0.75
6	0.78
7	0.76
8	0.78
9	0.78
10	0.79

Dari tabel diatas, dapat disimpulkan bahwa. Nilai *accuracy* tertinggi dari setiap pengujian nilai K yang berbeda berada pada Nilai K ke 10 dengan nilai *accuracy* sebesar 0,79. Dapat diartikan bahwa penggunaan aplikasi PayPal pada layanan *google play* memiliki akurasi sebesar 79% di negara Indonesia.

IV. KESIMPULAN

Penelitian ini melakukan klasifikasi terhadap ulasan pengguna aplikasi PayPal dengan menggunakan algoritma K-Nearest Neighbor (K-NN). Sebanyak 597 data ulasan dianalisis, terdiri atas 451 ulasan bernuansa negatif dan 146 ulasan bernada positif. Sentimen-sentimen tersebut berkaitan dengan ekspresi seperti "bagus", "cepat", "kecewa", "bodoh", dan "tidak layak". Hasil analisis menunjukkan bahwa nilai akurasi tertinggi dalam klasifikasi dicapai saat nilai parameter K berada pada angka 10, dengan tingkat akurasi sebesar 79%.

Meskipun studi ini telah menyelesaikan analisis terhadap sentimen pengguna terhadap PayPal, masih terdapat ruang untuk pengembangan lebih lanjut. Penelitian lanjutan dapat mempertimbangkan penggunaan metode klasifikasi yang lebih mutakhir, pendekatan pelabelan yang lebih akurat, dataset yang lebih seimbang, serta perluasan variabel penelitian lainnya. Oleh karena itu, temuan dari studi ini diharapkan dapat menjadi rujukan bagi penelitian selanjutnya yang mengangkat topik serupa.

DAFTAR PUSTAKA

- Furqan, M., Sriani, S., & Sari, S. M. (2022). Analisis Sentimen Menggunakan K-Nearest Neighbor Terhadap New Normal Masa Covid-19 Di Indonesia. *Techno.Com*, 21(1), 51–60.
- Bank Indonesia. (2005, 22 Juli). *Peraturan Bank Indonesia Nomor 7/18/PBI/2005 tentang Sistem Kliring Nasional Bank Indonesia*. Peraturan BPK.
- Riskawati, Fatihanursari, Iin, & Rizki Rinaldi, A. (2024). Penerapan Metode Naïve Bayes Classifier Pada Analisis Sentimen Aplikasi Gopay. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 346–353.
- Masturoh, I., & Pohan, F. (2021). Analisis sentimen aplikasi dompet digital DANA menggunakan metode *K-Nearest Neighbor*. *Pilar Nusa Mandiri*, 17(1), 53–60.
- Rizaldi, T., & Putranto, H. A. (2017). Perbandingan Metode Web Scraping Menggunakan CSS Selector dan Xpath Selector. *Teknika*, 6(1), 43–46.
- Rivki, M., & Bachtiar, A. M. (2017). Implementasi Algoritma K-Nearest Neighbor dalam Pengklasifikasian Follower Twitter yang menggunakan Bahasa Indonesia. *Jurnal Sistem Informasi*, 13(1), 31.
- Sedik, C. E. (2023). *Daftar 7 E-Wallet Terpopuler di Indonesia*. Finansial Bisnis.

- IBM. (2024). *Bagaimana analisis sentimen dapat digunakan untuk meningkatkan pengalaman pelanggan?*. Diakses dari <https://www.ibm.com/id-id/think/insights/how-can-sentiment-analysis-be-used-to-improve-customer-experience>
- Ivosights. (2023). *Optimalkan Keputusan Bisnis dengan Sentiment Analysis Tools*.
- Adhi Putra, A. D. (2021). Analisis Sentimen pada Ulasan pengguna Aplikasi Bibit Dan Bareksa dengan Algoritma KNN. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 8(2), 636–646.
- Arham, A. (2023). *Labeling Sentimen Bahasa Indonesia Secara Otomatis*. Medium.