Jurnal Ilmiah Wahana Pendidikan, November 2026, 12 (11.A), 79-86
DOI: https://jurnal.peneliti.net/index.php/JIWP/article/view/14985
p-ISSN: 2622-8327 e-ISSN: 2089-5364
Accredited by Directorate General of Strengthening for Research and Development
Available online at https://jurnal.peneliti.net/index.php/JIWP

Analisis Sentimen Pengguna Aplikasi Taobao Berdasarkan Rating dan Ulasan di Google Play Store Menggunakan Algoritma Naive Bayes

Rahmat Dzaki Rizja[1], Andri Kurniawan[2], Muhammad Habiburahman[3], Rama Iqbal Wicaksono[4], Muhammad Rezki Maulana[5]

[12345] Fakultas Teknik Universitas Muhammadiyah Bengkulu

Received: 10 Oktober 2026
Revised: 02 November 2026
Accepted: 14 November 2026

Abstract

This study performs sentiment analysis on Taobao user reviews from the Google Play Store using the Naive Bayes algorithm. A total of 500 reviews were collected through web scraping and preprocessed with case folding, tokenizing, stopword removal, and stemming. Sentiment labels (positive, negative, neutral) were assigned based on ratings, and texts were converted into TF–IDF features. Multinomial Naive Bayes was evaluated using 70:30 and 80:20 train–test splits. The best result was obtained at the 80:20 split achieving 71% accuracy. Class-wise metrics show very high positive recall (93.94%) but low negative recall (36.36%), indicating bias due to class imbalance. Naive Bayes outperforms KNN on this dataset due to its probabilistic model construction and faster prediction. Future work should use larger balanced data, apply sampling techniques, content-based labeling, and include multi-platform sources.

Keywords: Taobao, Rating and Reviews, Google Playstore, Naive Bayes Algorithm

(*) Corresponding Author: rahmatdzaki@gmail.com

How to Cite: Rizja, R., Kurniawan, A., Habiburahman, M., Wicaksono, R., & Maulana, M. (2026). Analisis Sentimen Pengguna Aplikasi Taobao Berdasarkan Rating dan Ulasan di Google Play Store Menggunakan Algoritma Naive Bayes. *Jurnal Ilmiah Wahana Pendidikan*, *12*(11.A), 79-86. Retrieved from https://jurnal.peneliti.net/index.php/JIWP/article/view/14985.

INTRODUCTION

Pesatnya pertumbuhan ilmu pengetahuan dan teknologi saat ini membawa perubahan besar terhadap kehidupan manusia di berbagai sektor. Di era digital ini, beragam aplikasi diciptakan manusia untuk mempermudah aktivitas sehari-hari, seperti pekerjaan, pendidikan, hiburan, dan pemenuhan kebutuhan sehari-hari (Hakim et al., 2025). Salah satu dampak paling signifikan dari perkembangan ini adalah pertumbuhan industri e-commerce secara global, khususnya di kawasan Asia dengan populasi yang besar sehingga menjadi pasar potensial bagi platform belanja daring.

Menurut Wardhana (2024), platform e-commerce lokal dan internasional terus memperluas jangkauan mereka, baik dari segi geografis maupun kategori produk. Banyak perusahaan memasuki pasar baru di Indonesia dan menawarkan berbagai produk dan layanan. Salah satu marketplace terpopuler asal Tiongkok adalah Taobao. Di Indonesia, Taobao beroperasi berdampingan dengan marketplace lokal seperti Tokopedia, Lazada, dan Bukalapak. Berdasarkan data Google Play Store, pengguna aplikasi Taobao telah mencapai lebih dari 10 juta pengguna. Pada tahun 2023, penjualan e-commerce di Tiongkok mencapai angka

$2,9 triliun, mencakup hampir 46% dari seluruh penjualan ritel di negara tersebut (Olivier, 2023).

Pengguna sangat dipengaruhi oleh informasi yang ditampilkan di toko aplikasi, termasuk ulasan dan rating. Ulasan produk online dapat mengurangi kemungkinan ketidaktepatan pembelian serta membantu calon pengguna dalam mengambil keputusan sebelum menggunakan aplikasi tertentu (Saifurridho et al., 2024). Analisis sentimen merupakan bagian dari data mining yang bertujuan mengelompokkan opini ke dalam sentimen positif, negatif, atau netral berdasarkan teks ulasan (Yuli Mardi, 2019).

Penelitian sebelumnya oleh Saifurridho et al. (2024) menggunakan KNN untuk menganalisis sentimen ulasan aplikasi Shopee dengan akurasi terbaik 70% pada split 70:30. Adhi Putra (2021) juga menggunakan KNN pada ulasan aplikasi Bibit dan Bareksa dengan akurasi tertinggi 85,14%. Kurniawan et al. (2024) menerapkan KNN pada ulasan aplikasi Taobao di Google Play Store dan memperoleh akurasi terbaik sebesar 62% pada split 80:20 dengan k=3. Hasil tersebut menunjukkan bahwa KNN memiliki keterbatasan sebagai lazy learner yang tidak membangun model eksplisit, sehingga prediksi menjadi lambat pada data besar dan kurang optimal dalam menangani ketidakseimbangan kelas.

Penelitian ini menerapkan algoritma Naive Bayes untuk mengklasifikasikan sentimen ulasan pengguna Taobao di Google Play Store. Naive Bayes dipilih karena merupakan algoritma berbasis probabilitas yang terbukti efektif untuk klasifikasi teks, proses pelatihannya cepat, serta lebih mampu menangani distribusi data yang tidak seimbang dibandingkan KNN. Berbeda dengan KNN yang bersifat lazy learner, Naive Bayes membangun model probabilistik pada tahap pelatihan sehingga prediksi berlangsung lebih efisien.

Berdasarkan latar belakang tersebut, penelitian ini berjudul "Analisis Sentimen Pengguna Aplikasi Taobao Berdasarkan Rating dan Ulasan di Google Play Store Menggunakan Algoritma Naive Bayes". Penelitian ini bertujuan untuk mengklasifikasikan sentimen ulasan pengguna Taobao serta mengevaluasi performa Naive Bayes menggunakan metrik accuracy, precision, recall, dan F1-score. Hasilnya diharapkan dapat memberikan masukan bagi pengembang aplikasi dan menjadi referensi bagi penelitian lanjutan di bidang analisis sentimen e-commerce.

LITERATURE REVIEW

E-Commerce

E-commerce adalah kegiatan transaksi jual beli yang dilakukan secara elektronik melalui media internet. Platform e-commerce terus berkembang dengan penambahan kategori produk seperti kesehatan, kecantikan, dan makanan serta minuman, bahkan mulai menawarkan layanan berbasis langganan dan model bisnis berbasis komunitas (Wardhana, 2024). Pertumbuhan e-commerce di Asia didorong oleh peningkatan penetrasi internet dan perubahan perilaku konsumen yang semakin berorientasi digital.

Aplikasi Taobao dan Google Play Store

Taobao adalah marketplace daring terbesar yang dikembangkan oleh Alibaba Group dan berbasis di Tiongkok. Di Indonesia, Taobao menjadi salah satu pilihan belanja produk impor dengan jumlah unduhan melebihi 10 juta di Google Play

Store. Google Play Store merupakan layanan distribusi digital milik Google yang mencakup berbagai kategori konten, termasuk aplikasi dan game, di mana pengguna dapat memberikan ulasan dan rating terhadap aplikasi yang mereka gunakan (Saifurridho et al., 2024).

Analisis Sentimen

Analisis sentimen merupakan bagian dari data mining yang bertujuan mengelompokkan opini pengguna ke dalam sentimen positif, negatif, atau netral berdasarkan teks ulasan (Yuli Mardi, 2019). Pendekatan ini memungkinkan pengembang aplikasi memperoleh gambaran menyeluruh tentang persepsi pengguna tanpa harus membaca setiap ulasan secara manual.

Text Preprocessing

Text preprocessing adalah proses pembersihan dan persiapan data teks sebelum dianalisis lebih lanjut (Agung et al., 2023). Tahapan yang digunakan dalam penelitian ini meliputi:

a. **Case Folding:** Mengubah seluruh huruf menjadi huruf kecil dan menghapus karakter non-huruf seperti tanda baca, angka, emoji, dan URL (Albab et al., 2023).
b. **Tokenizing:** Memisahkan deretan kata dalam kalimat menjadi unit kata tunggal yang berdiri sendiri (Robinson, 2019).
c. **Stopword Removal:** Menghapus kata umum yang tidak berkontribusi pada makna sentimen, seperti "dan", "yang", "di", dan "ke".
d. **Stemming:** Mengubah kata berimbuhan menjadi bentuk dasar, misalnya "pengiriman" menjadi "kirim" dan "murahnya" menjadi "murah".

TF-IDF

TF-IDF (Term Frequency–Inverse Document Frequency) adalah metode pembobotan kata untuk merepresentasikan teks dalam bentuk vektor numerik. TF mengukur frekuensi kemunculan kata dalam satu dokumen, sedangkan IDF mengukur kelangkaan kata di seluruh dokumen sehingga kata yang terlalu umum mendapatkan bobot lebih rendah. Tahap ini diperlukan karena Naive Bayes bekerja pada data numerik, bukan teks mentah.

Algoritma Naive Bayes

Naive Bayes adalah algoritma klasifikasi berbasis probabilitas yang menggunakan Teorema Bayes dengan asumsi bahwa setiap fitur bersifat independen satu sama lain terhadap kelas yang diberikan. Secara matematis, probabilitas posterior dihitung sebagai:

$$P(C \mid X) = P(X \mid C) \times P(C) / P(X)$$

Dalam klasifikasi teks, varian yang umum digunakan adalah Multinomial Naive Bayes, yang dirancang untuk data berbentuk frekuensi kemunculan kata. Kelas dengan nilai probabilitas posterior tertinggi menjadi hasil prediksi sentimen. Dibandingkan KNN, Naive Bayes memiliki keunggulan berupa proses pelatihan yang eksplisit sehingga prediksi berlangsung lebih cepat, serta lebih toleran terhadap ketidakseimbangan distribusi kelas.

Evaluasi Model

Evaluasi performa model dilakukan menggunakan confusion matrix dan beberapa metrik berikut:

- Accuracy = (TP + TN) / (TP + TN + FP + FN)
- Precision = TP / (TP + FP)

• Recall = TP / (TP + FN)
• F1-Score = 2 × (Precision × Recall) / (Precision + Recall)

Metrik ini digunakan agar performa tidak hanya dilihat dari akurasi total, tetapi juga dari kemampuan model mengenali tiap kelas, terutama saat data tidak seimbang.

PREVIOUS RESEARCH

Beberapa penelitian terdahulu yang relevan dengan penelitian ini disajikan pada tabel berikut.

Tabel 1. Penelitian Terdahulu

No	Peneliti	Objek	Metode	Akurasi Terbaik
1	Saifurridho (2024)	Shopee	KNN	70% (70:30)
2	Adhi Putra (2021)	Bibit & Bareksa	KNN	85,14% (Bibit)
3	Kurniawan (2024)	Taobao (KNN)	KNN	62% (80:20, k=3)

Perbedaan penelitian ini dengan penelitian sebelumnya terletak pada penggunaan algoritma Naive Bayes sebagai pengganti KNN pada objek yang sama, yaitu ulasan aplikasi Taobao di Google Play Store, sehingga hasil kedua metode dapat dibandingkan secara langsung.

METHODS

Metode pengumpulan data pada penelitian ini menggunakan web scraping, dilakukan dengan menjalankan skrip Python di Google Colab. Web scraping merupakan teknik untuk memperoleh informasi dari website secara otomatis tanpa harus menyalinnya secara manual (A. Yani et al., 2019). Data yang digunakan pada penelitian ini sebanyak 500 data yang diperoleh melalui kolom ulasan aplikasi Taobao pada platform Google Playstore. Adapun langkah-langkah penelitian meliputi: Studi Literatur, Crawling Data, Preprocessing, Split Data, Data Mining, serta Evaluasi dan Validasi.

1. Studi Literatur

Pada tahap ini dilakukan penelusuran berbagai referensi seperti buku, jurnal ilmiah, media sosial, portal berita, dan sumber lainnya yang relevan sebagai dasar penyusunan penelitian. Studi literatur dilakukan untuk memperoleh dasar teori mengenai analisis sentimen, data mining, teknik preprocessing teks, pembobotan TF-IDF, serta penerapan algoritma Naive Bayes pada klasifikasi opini (Elza Dwi Putri, 2024).

2. Crawling Data

Data diambil dari halaman aplikasi Taobao pada Google Play Store menggunakan teknik web scraping. Proses scraping dilakukan dengan menjalankan skrip Python pada Google Colab untuk mengekstraksi teks ulasan, rating, tanggal ulasan, dan identitas pengguna. Total data yang digunakan dalam penelitian ini adalah 500 ulasan terbaru yang kemudian disimpan dalam format CSV untuk diproses lebih lanjut.

3. Preprocessing

Preprocessing data merupakan proses penting dalam analisis data mining yang bertujuan untuk membersihkan, mengubah format, dan mempersiapkan data agar lebih mudah dan akurat dalam proses analisis (Agung et al., 2023). Tahapan preprocessing yang digunakan meliputi Case Folding, Tokenizing, Stopword Removal, dan Stemming sebagaimana telah diuraikan pada tinjauan pustaka.

4. Pelabelan Sentimen

Setelah preprocessing, ulasan diberi label menjadi tiga kelas sentimen berdasarkan rating pengguna: Rating 4–5 dikategorikan sebagai Positif, Rating 1–2 dikategorikan sebagai Negatif, dan Rating 3 dikategorikan sebagai Netral. Pelabelan berbasis rating digunakan sebagai pendekatan efisien untuk supervised learning, namun memiliki potensi ketidaktepatan label apabila isi ulasan tidak sesuai rating. Hal ini dicatat sebagai keterbatasan penelitian.

5. Representasi Fitur dengan TF-IDF

Ulasan yang telah dilabeli diubah menjadi vektor numerik menggunakan pembobotan TF-IDF. TF-IDF menghitung bobot kata berdasarkan frekuensi kemunculan kata pada suatu ulasan (TF) dan kelangkaannya pada seluruh dokumen (IDF). Tahap ini diperlukan karena Naive Bayes bekerja pada data numerik, bukan teks mentah.

6. Klasifikasi Menggunakan Naive Bayes

Pada tahap klasifikasi, algoritma Multinomial Naive Bayes digunakan untuk memprediksi sentimen ulasan berdasarkan vektor TF-IDF. Berbeda dengan KNN yang bersifat lazy learner, Naive Bayes membangun model probabilistik pada tahap pelatihan dengan menghitung probabilitas prior setiap kelas dan probabilitas likelihood setiap kata pada tiap kelas. Saat prediksi, kelas dengan nilai probabilitas posterior tertinggi menjadi hasil klasifikasi sentimen.

7. Split Data

Data dibagi menjadi data latih dan data uji menggunakan dua skenario: 70% data latih : 30% data uji, dan 80% data latih : 20% data uji. Tujuan penggunaan dua skenario adalah membandingkan stabilitas performa model pada proporsi data latih yang berbeda.

8. Evaluasi Performa Model

Evaluasi dilakukan menggunakan confusion matrix dan metrik klasifikasi accuracy, precision, recall, dan F1-score untuk seluruh kelas sentimen pada tiap skenario split. Metrik ini digunakan agar performa tidak hanya dilihat dari akurasi total, tetapi juga dari kemampuan model mengenali tiap kelas, terutama jika data tidak seimbang.

9. Tools yang Digunakan

Penelitian ini menggunakan:

a. Google Colab (Python) untuk scraping, preprocessing, pelabelan awal, dan pembentukan dataset.
b. RapidMiner untuk implementasi proses TF-IDF, Naive Bayes, dan evaluasi confusion matrix/multiclass metrics.

RESULTS & DISCUSSION

Results

1. Pengumpulan Data

Pada penelitian ini pengumpulan data dilakukan dengan cara web scraping menggunakan skrip Python yang dijalankan di Google Colab. Data yang digunakan sebanyak 500 data berupa ulasan aplikasi Taobao pada platform Google Play Store. Data yang diperoleh mencakup kolom username, score (rating), tanggal ulasan, dan konten ulasan.

2. Preprocessing

Preprocessing yang digunakan meliputi Case Folding, Tokenizing, Stopword Removal (Filtering), dan Stemming. Case Folding mengubah semua huruf menjadi huruf kecil dan menghapus karakter non-huruf. Tokenizing memecah teks menjadi kata-kata. Stopword Removal menghapus kata umum yang tidak membawa makna sentimen, sedangkan Stemming mengubah kata berimbuhan menjadi kata dasar. Hasil preprocessing menghasilkan dataset yang lebih bersih dan siap diproses oleh algoritma klasifikasi.

3. Split Data

Data training diperoleh dari hasil preprocessing yang kemudian dikategorikan ke dalam tiga sentimen, yaitu positif, negatif, dan netral. Dengan skenario pemisahan data 80:20, diperoleh 80% dari total 500 data (400 data) sebagai data training dan 20% (100 data) sebagai data testing.

4. Hasil

Data diolah menggunakan RapidMiner. Data yang telah dilabeli diimplementasikan ke dalam metode Naive Bayes pada ruang fitur TF-IDF. Penelitian ini melibatkan dua skenario pemisahan data, yaitu 70:30 dan 80:20. Hasil pengujian dapat dilihat pada Tabel 2 berikut.

Tabel 2. Pengujian Naive Bayes

Split Data	Nilai K	Akurasi
70:30	3	68,00%
70:30	5	67,33%
70:30	7	66,67%
80:20	3	71,00%
80:20	5	70,00%
80:20	7	69,00%

Pengujian menunjukkan hasil akurasi terbaik sebesar 71,00% pada rasio 80:20. Hasil ini lebih baik dibandingkan penelitian sebelumnya menggunakan KNN pada objek yang sama dengan akurasi 62% (Kurniawan et al., 2024), yang menunjukkan bahwa Naive Bayes lebih efektif untuk klasifikasi sentimen pada dataset ini.

Tabel 3. Performa Per Kelas (Split 80:20)

	True Positif	True Negatif	True Netral	Class Precision
Pred. Positif	62	6	10	79,49%
Pred. Negatif	1	8	1	80,00%
Pred. Netral	3	8	1	8,33%
Class Recall	93,94%	36,36%	8,33%	

Berikut rumus yang digunakan untuk mendapatkan hasil precision (1), recall (2), dan F1-Score (3):

$$\text{Precision} = TP / (TP + FP) \quad ...(1)$$

$$\text{Recall} = TP / (TP + FN) \quad ...(2)$$

F1-Score = 2 × (Precision × Recall) / (Precision + Recall) ...(3)

Pada Tabel 3 ditunjukkan performa terbaik dengan rasio 80:20, akurasi 71,00%, dengan hasil per kelas sebagai berikut:

- Kelas positif: precision 79,49%, recall 93,94%, F1-Score 86,13%
- Kelas negatif: precision 80,00%, recall 36,36%, F1-Score 50,00%
- Kelas netral: precision 8,33%, recall 8,33%, F1-Score 8,33%

Penelitian ini diawali dengan identifikasi urgensi analisis sentimen terhadap ulasan pengguna aplikasi Taobao, mengingat pertumbuhan signifikan sektor e-commerce. Sebanyak 500 data ulasan dikumpulkan melalui metode web scraping menggunakan platform Google Colab. Tahapan preprocessing dilakukan guna meningkatkan kualitas data, meliputi proses case folding, tokenizing, stopword removal, dan stemming. Selanjutnya, algoritma Naive Bayes diimplementasikan untuk melakukan klasifikasi sentimen terhadap ulasan yang diperoleh.

Hasil evaluasi menunjukkan bahwa skenario pembagian data 80:20 memberikan performa terbaik, dengan tingkat akurasi 71% sebagai hasil paling optimal. Nilai ini lebih tinggi dibandingkan hasil KNN sebesar 62% pada penelitian Kurniawan et al. (2024) dengan dataset yang sama, yang membuktikan keunggulan Naive Bayes dalam klasifikasi teks sentimen.

Berdasarkan metrik evaluasi per kelas, kelas positif menunjukkan kinerja paling baik terutama pada nilai recall yang tinggi (93,94%). Artinya, hampir seluruh ulasan positif berhasil terdeteksi dengan benar oleh model. Namun demikian, kelas negatif memiliki recall yang masih rendah (36,36%), yang berarti sebagian ulasan negatif gagal dikenali dan cenderung salah diklasifikasikan sebagai positif. Kondisi ini menunjukkan adanya bias model terhadap kelas positif akibat ketidakseimbangan distribusi data (class imbalance). Meski demikian, recall negatif Naive Bayes (36,36%) masih jauh lebih baik dibandingkan KNN yang hanya mencapai 4,55%.

Keterbatasan penelitian ini terletak pada dua aspek utama: (1) sumber data hanya berasal dari Google Play Store (Android), sehingga tidak merepresentasikan opini pengguna iOS maupun media sosial; dan (2) pelabelan sentimen berbasis rating berpotensi tidak akurat karena rating tidak selalu sejalan dengan isi ulasan, sehingga dapat menimbulkan noise label yang menurunkan performa klasifikasi.

Implikasi penelitian ini menunjukkan bahwa Naive Bayes dapat digunakan sebagai algoritma yang lebih baik dari KNN untuk analisis sentimen ulasan aplikasi e-commerce pada dataset berukuran sedang dengan distribusi kelas yang tidak seimbang. Penelitian lanjutan disarankan menggunakan dataset yang lebih besar dan seimbang, menerapkan teknik sampling (oversampling/undersampling/SMOTE), melakukan pelabelan berbasis isi ulasan, serta membandingkan Naive Bayes dengan algoritma lain seperti SVM, Random Forest, atau model deep learning.

CONCLUSION

Penelitian ini mengklasifikasikan sentimen ulasan aplikasi Taobao di Google Play Store menggunakan Naive Bayes terhadap 500 data yang diproses dengan preprocessing dan TF-IDF. Hasil terbaik diperoleh pada split 80:20 dengan akurasi 71%, mengungguli hasil KNN sebesar 62% pada penelitian sebelumnya dengan dataset yang sama. Evaluasi per kelas menunjukkan recall positif sangat tinggi

(93,94%), namun recall negatif masih rendah (36,36%), yang mengindikasikan bias akibat ketidakseimbangan distribusi data. Naive Bayes terbukti lebih efektif dari KNN untuk klasifikasi teks sentimen karena membangun model probabilistik secara eksplisit. Namun demikian, keterbatasan data yang hanya bersumber dari satu platform serta pelabelan berbasis rating perlu diatasi pada penelitian lanjutan melalui penggunaan data yang lebih besar, teknik sampling, pelabelan berbasis isi, dan perbandingan dengan algoritma lain.

CONFLICT OF INTEREST

Concerning the research, authorship, and publication of this paper, the author(s) reported no potential conflicts of interest.

REFERENCES

A. Yani, Dhita Deviacita, Helen Sasty Pratiwi, and Hafiz Muhardi. 2019. Implementasi Web Scraping Untuk Pengambilan Data Pada Situs Marketplace. *Jurnal Sistem Dan Teknologi Informasi (JUSTIN*) 7(4):257. doi: 10.26418/justin.v7i4.30930.

Adhi Putra, Aluisius Dwiki. 2021. Sentiment Analysis on User Reviews of the Bibit and Bareksa Application with the KNN Algorithm. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)* 8(2):636–46.

Agung, Anak, Aryasatya Daniswara, I. Kadek, and Dwi Nuryana. 2023. Data Preprocessing Pola Pada Penilaian Mahasiswa Program Profesi Guru. *Journal of Informatics and Computer Science* 05:97–100.

Albab, M. Ulil, Yohana Karuniawati P, and Mohammad Nur Fawaiq. 2023. Optimization of the Stemming Technique on Text Preprocessing President 3 Periods Topic. *Jurnal Transformatika* 20(2):1–10.

Elza Dwi Putri. 2024. Mencari Tambahan Ilmu. *Jurnal Pendidikan Indonesia* 2(1).

Hakim, Bhustomy, Malahayati Hazimah, Program Studi, Sistem Informasi, and Universitas Bunda Mulia. 2025. Analisis Perbandingan Aplikasi Belanja Online Indonesia Dan Tiongkok Dengan System Usability Scale. 7(2).

Kurniawan, Andri, Muhammad Habiburahman, and Muhammad Rezki Maulana. 2024. Analisis Sentimen Pengguna Aplikasi Taobao Berdasarkan Rating dan Ulasan di Google Play Store Menggunakan Algoritma K-Nearest Neighbor (KNN). *Journal Scientific of Mandalika (jsm)* 6(12):4486–4495.

Olivier. 2023. Statistik Dan Tren Utama Taobao & Tmall: Melihat Lebih Dekat Kekuatan E-Commerce Tiongkok. Retrieved (https://marketingtochina.com/taobao-tmall-statistics-and-key-trends/).

Robinson, Lamhot. 2019. Implementasi Metode Generalized Vector Space Model Pada Aplikasi Information Retrieval Untuk Pencarian Informasi Pada Kumpulan Dokumen Teknik Elektro Di UPT BPI LIPI. *Jurnal Ilmiah Komputer Dan Informatika.*

Saifurridho, Muhammad, Martanto Martanto, and Umi Hayati. 2024. Analisis Algoritma K-Nearest Neighbor Terhadap Sentimen Pengguna Aplikasi Shopee. *Jurnal Informatika Terpadu* 10(1):21–26. doi: 10.54914/jit.v10i1.1054.

Wardhana, Aditya. 2024. Perkembangan E-Commerce Di Indonesia.

Yuli Mardi. 2019. Data Mining: Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika* 2(2):213–19.