

# ANALISIS SENTIMEN MEMBANDINGKAN PENGGUNA APLIKASI *E-COMMERCE* TOKOPEDIA DAN SHOPEE MENGGUNAKAN ALGORITMA *NAIVE BAYES*

Steven<sup>1\*</sup>, Indah Fenriana<sup>2</sup>

<sup>1,2</sup>Program Studi Teknik Informatika, Fakultas Sains dan Teknologi,  
Universitas Buddhi Dharma  
Email: [stevenhou61@gmail.com](mailto:stevenhou61@gmail.com)\*

## Abstrak

Perkembangan pengguna internet di Indonesia, telah mengubah cara masyarakat dalam hal melakukan transaksi jual beli. Aplikasi *E-Commerce* telah menjadi solusi dalam memenuhi kebutuhan konsumen yang menginginkan kemudahan berbelanja tanpa harus keluar rumah. Menurut laporan dari statistika pada penelitian jumlah pengguna *E-commerce* di Indonesia diprediksi akan terus meningkat, Namun meningkatnya penggunaan aplikasi *e-commerce* juga memunculkan berbagai tantangan. Masalah utama meliputi persaingan ketat antar-platform, isu keamanan data. Oleh karena itu, penting bagi mereka untuk melakukan penelitian dan membaca ulasan aplikasi. Dalam konteks ini, penelitian ini difokuskan pada model klasifikasi untuk analisis sentimen ulasan aplikasi menggunakan teknik pembobotan kata. Dengan algoritma *Naive Bayes* sebagai klasifikasi utama dengan penerapan dua teknik pembobotan kata, yaitu *Term Frequency-Inverse Document Frequency* (TF-IDF) dan *Count Vectorizer* mencapai akurasi tinggi, yaitu 80%. Dalam penelitian ini memberikan informasi yang berharga dalam penggunaan teknik pembobotan kata dalam analisis sentiment, yang nantinya membantu meningkatkan keefektifan sistem analisis sentiment untuk membandingkan aplikasi *e-commerce*. Dengan adanya penelitian ini dapat memberikan kontribusi dalam pengembangan sistem analisis sentiment untuk membandingkan aplikasi *e-commerce* di Indonesia.

**Kata kunci:** *Naive Bayes*, TF-IDF, *Count Vectorizer*

## Pendahuluan

Perkembangan teknologi informasi dan komunikasi, khususnya internet, telah mengubah cara masyarakat dalam hal melakukan transaksi jual beli. *E-Commerce*, atau aplikasi jual beli, telah menjadi solusi dalam memenuhi kebutuhan konsumen yang menginginkan kemudahan berbelanja tanpa harus keluar rumah. Di Indonesia, berbagai platform *E-Commerce* seperti tokopedia dan shopee telah muncul sebagai pemain utama dalam industri ini, menciptakan persaingan yang kompetitif dalam menawarkan beragam produk dan layanan (Bintang Zulfikar Ramadhan et al., 2023). Mengingat jumlah penggunaan *E-Commerce di indonesia* diprediksi akan terus meningkat, mencerminkan adopsi yang sangat luas dari belanja online di kalangan konsumen penduduk di Indonesia (Ulhaq et al., 2024) dengan adanya

*Google play store* konsumen kini dapat dengan mudah memberikan feedback melalui ulasan dan penilaian di berbagai platform aplikasi pada *Google play store*. Ulasan ini membuat sumber informasi yang berharga bagi perusahaan dalam memahami pandangan dan tingkat kepuasan pengguna terhadap layanan yang mereka tawarkan (Fransiska & Irham Gufroni, 2020). Untuk mengolah ulasan konsumen yang tersedia di *Google play store*, diperlukan metode yang mampu menangani data teks dalam jumlah besar secara sistematis. Salah satu pendekatan yang umum digunakan adalah *text mining*.

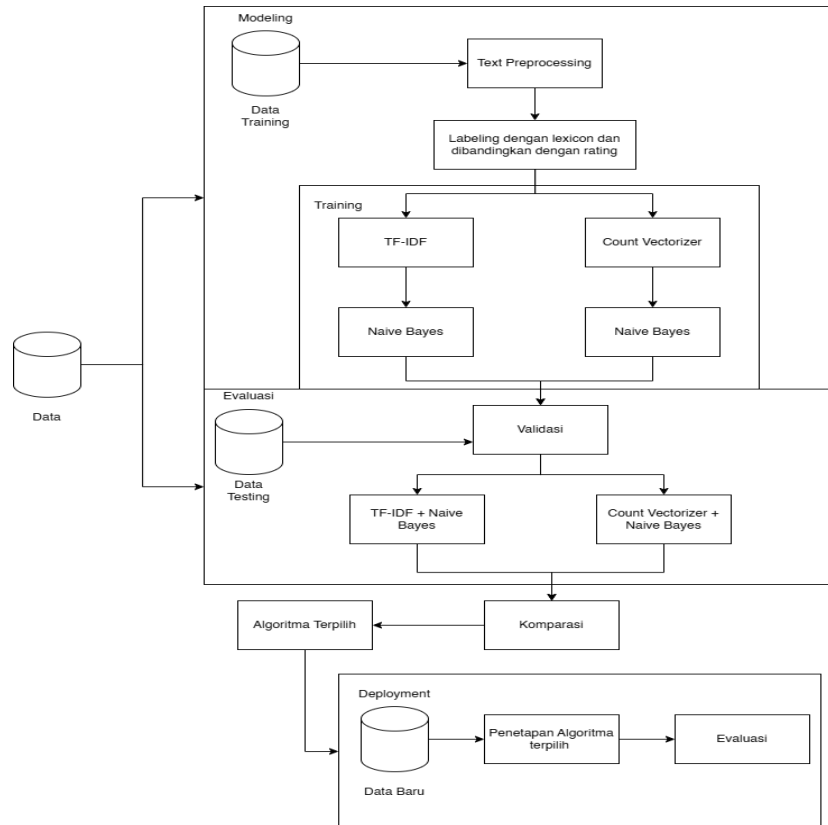
*Text mining* merupakan cara memproses ekstraksi informasi yang berguna dari teks melibatkan identifikasi pola, hubungan, dan informasi penting lainnya dari dokumen teks. Dalam konteks analisis sentimen, *text mining* digunakan untuk melakukan pengekstrakan dan menganalisis opini, sentimen, dan sikap dari teks yang dihasilkan oleh pengguna (Birjali et al., 2021). Teknik ini banyak diterapkan dalam berbagai bidang, termasuk pemasaran, dan penelitian perilaku pengguna (Tabinda Kokab et al., 2022)

Studi kasus menggunakan *feature extraction Term Frequency-Inverse Document Frequency (TF-IDF)* hasil penelitian dengan menunjukkan *Naive Bayes* sebagai salah satu algoritma mendapatkan akurasi tertinggi sebesar 98.17% (Huang et al., 2023). Mengklasifikasi sentimen pada aplikasi *E-Commerce* shopee, algoritma *Naive bayes* berhasil mencapai akurasi sebesar 96,67% dengan menggunakan *feature extraction TF-IDF* untuk pembobotan kata (Pratmanto et al., 2020). Mengklasifikasikan ulasan aplikasi *E-Commerce* Amazon Shopping menggunakan algoritma *Naive bayes* menghasilkan akurasi 82.15% dengan *feature extraction TF-IDF* dalam proses pembobotan kata (Ernianti Hasibuan & Elmo Allistair Heriyanto, 2022). Metode lexicon untuk pelabelan *feature extraction TF-IDF* dan algoritma *Multinomial Naive Bayes* mendapatkan akurasi sebesar 90% (Juanita et al., 2022). Algoritma *Naive Bayes* dan *feature extraction TF-IDF* mendapatkan akurasi sebesar 87.6% (Cahyono & Anggista Oktavia Praneswara, 2023).

Meski banyak penelitian menganalisis sentimen dengan *Naive Bayes* dan *TF-IDF*, perbandingan langsung antara *TF-IDF* dan *Count Vectorizer* masih jarang

dilakukan. Penelitian ini bertujuan membandingkan performa kedua metode tersebut dalam menganalisis ulasan aplikasi *E-Commerce* di Google Play Store.

## Metodologi



**Gambar 1. Flowchart Kerangka Kerja**

### 1. Pengumpulan Data

Jenis data yang dikumpulkan merupakan data primer. Dikumpulkan melalui proses web scraping dari situs Google Play Store, mencakup 14.406 ulasan pengguna Shopee dan 13.853 ulasan Tokopedia.

### 2. Preprocessing Data

Data yang dikumpulkan diproses melalui tahapan preprocessing untuk mempersiapkan teks sebelum analisis. Langkah-langkah yang dilakukan meliputi *Remove emoji*, *Case Folding*, *Duplicate letter removal*, *Slag removal*, *Tokenization*, *Normalisasi*, *Stopword removal*, *Stemming* (Hakim, 2021).

### 3. Labeling Data

Data dilabeli menggunakan metode lexicon dan hasilnya dibandingkan dengan rating. Label sentimen dikategorikan menjadi positif (rating 4-5), negatif (rating 1-2), dan netral (rating 3) (Demircan et al., 2021).

**Tabel 1. Tabel Perhitungan *Polarity Score* dan Rating**

<i>Stopword Removal</i>	<i>Polarity Score</i>	<i>Label dari polarity score</i>	<b>Rating</b>	<b>Label Akhir</b>
mantap	5	Positif	5	Positif
maaf, ngeluh, video, menyampah,	-3	Negatif	2	Negatif

Data yang diperoleh digabungkan hingga total 28.249 data, kemudian dibandingkan menggunakan pembobotan TF-IDF dan *Count Vectorizer*. Data dibagi dengan komposisi 80:20 untuk training dan testing.

**Tabel 2. Distribusi Pembagian Data Shopee dan Tokopedia**

<b>Kelas</b>	<b>Training</b>	<b>Testing</b>	<b>Total</b>
Positif	8.329	2.083	10.412
Negatif	8.448	2.112	10.560
Netral	5.820	1.457	7.277
Total	22.607	5.652	28.249

#### 4. Modeling

Metode *Term Frequency-Inverse Document Frequency* (TF-IDF) yang merupakan alat statistik yang berguna untuk menentukan seberapa penting sebuah kata atau kelompok kata tertentu dalam sebuah dokumen (Suryani & Edy, 2020) dan *Count Vectorizer* teknik dasar *Natural Language Processing* (NLP) yang digunakan untuk ekstraksi fitur dari data teks. Mengubah teks tak terstruktur, menjadi format terstruktur yang sesuai untuk *machine learning* (Turki & Roy, 2022), Meskipun sederhana, metode ini mampu memberikan hasil dengan akurasi tinggi dalam banyak kasus, terutama pada data teks (Putri et al., 2020).

#### 5. Confusion Matrix

*Confusion matrix* digunakan untuk mengevaluasi performa model klasifikasi, baik untuk klasifikasi biner maupun multiklas. Matriks ini memberikan gambaran tentang prediksi model dibandingkan dengan label sebenarnya, dengan empat komponen utama (Valero-Carreras et al., 2023).

**Tabel 3. Confusion Matrix sumber (Valero-Carreras)**

<b>Predicted</b>	<b>Actual values</b>	
	+	-
	<i>True positive</i> (TP)	<i>False positive</i> (FP)
	<i>False negative</i> (FN)	<i>True negative</i> (TN)

*Confusion matrix* digunakan untuk mengevaluasi kinerja pengklasifikasi melalui metrik seperti akurasi, presisi, recall, dan f-score, yang dihitung menggunakan rumus berikut:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

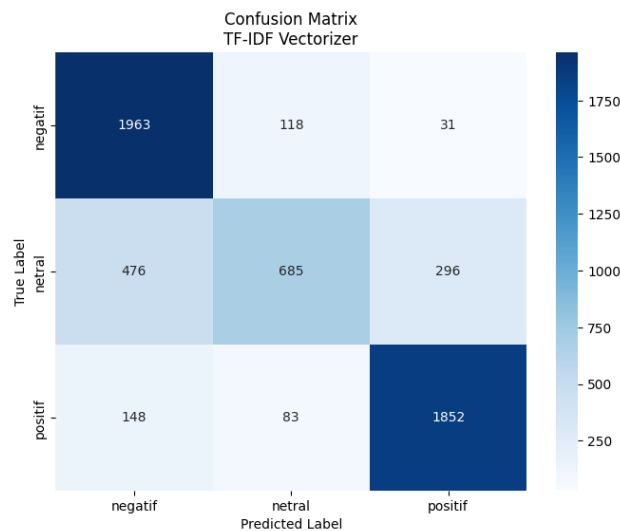
$$Presisi = \frac{TP}{TP + FP} \quad (2)$$

$$Recal = \frac{TP}{TP + FN} \quad (3)$$

$$F - Measure = \frac{2 * Recall * Presisi}{Recall + Presisi} \quad (4)$$

### Hasil Dan Pembahasan

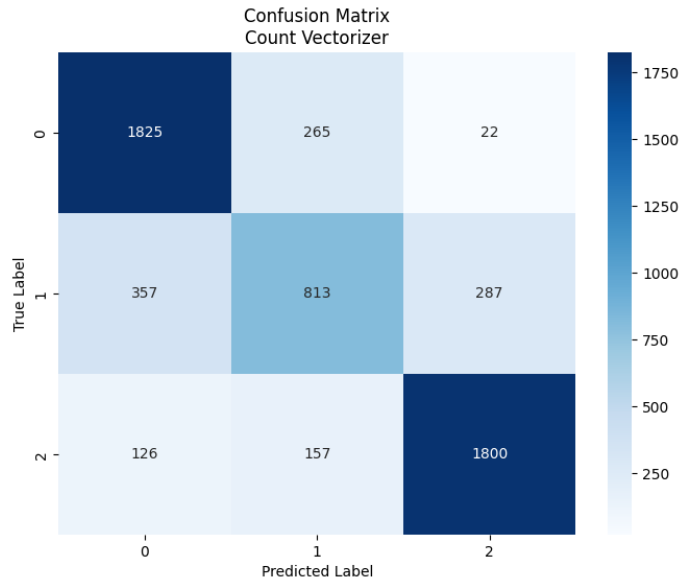
Evaluasi dilakukan menggunakan confusion matrix pada dua model: Naïve Bayes–TF-IDF dan Naïve Bayes–Count Vectorizer. Dataset dibagi menjadi 80% untuk data training dan 20% untuk testing. Hasil evaluasi model Naïve Bayes–TF-IDF menunjukkan: *Confusion matrix* memiliki nilai diatas 75%. Nilai *accuracy* sebesar 0.796%, *precision* 79%, *recall* 76%, dan *F1-score* 76%.



**Gambar 2. Confusion Matrix - TF-IDF**

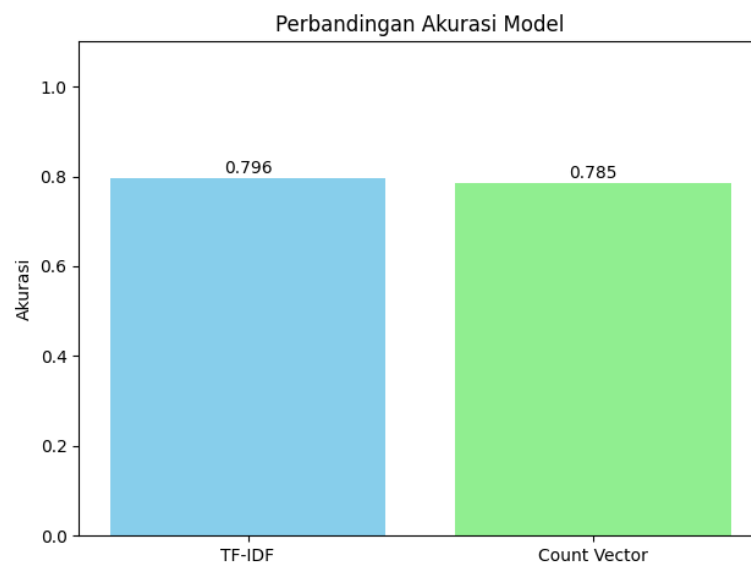
Model berhasil mengidentifikasi 1963 sentimen negatif, 685 netral, dan 1852 positif dengan precision masing-masing 76%, 77%, dan 85%. Nilai recall mencapai 93%, 47%, dan 89%, sedangkan F1-Score tercatat 84% (negatif), 58% (netral), dan 87% (positif). Hasil evaluasi Naïve Bayes Data – *Count Vectorizer* menunjukkan

*Confusion matrix* pada model ini mendapatkan nilai diatas 75%. Nilai *accuracy* 0.785%, *precision* 79%, *recall* 76%, dan *F1-score* 76%. Hasil *confusion matrix* untuk model *Naïve Bayes – Count Vectorizer*.



**Gambar 3. Confusion Matrix – Count Vectorizer**

Model tersebut berhasil mengidentifikasi 1825 data sentimen “negatif”, 813 data sentimen "netral", 1800 sentimen "positif" secara akurat, dengan precision masing-masing sebesar 76%, 77%, dan 85%, nilai 93%, 47%, dan 89%. F-1 Score 84% untuk Negatif, 58% untuk Netral dan 87% untuk positif.



**Gambar 4. Perbandingan Akurasi Model TF-IDF dan CV**

Berdasarkan grafik yang ditunjukkan oleh gambar 4 terlihat bahwa model TF-IDF mencapai akurasi tinggi sebesar 0.796% sementara *Count Vectorizer*

mendapatkan akurasi 0.785% Penggunaan TF-IDF pada model terbukti mendapatkan akurasi tertinggi.

## Simpulan

Penerapan kombinasi teknik pembobotan kata *TF-IDF* dan *Count Vectorizer* pada algoritma *Naïve Bayes* mampu meningkatkan akurasi pada model yang diusulkan, dengan hasil baik diperoleh menggunakan teknik *TF-IDF Naïve Bayes*. Model *Naïve Bayes* – *TF-IDF* adalah model yang mendapatkan akurasi tinggi mencapai 0.796% dengan presisi, recall dan f1-score masing-masing mencapai 79%, 76% dan 76%.

## Daftar Pustaka

- Bintang Zulfikar Ramadhan, Ibnu Riza, & Iqbal Maulana. (2023). Analisis Sentimen Ulasan Pada Aplikasi E-Commerce Shopee Dengan Menggunakan Algoritma Naïve Bayes. *Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 3(3), 23–39. <https://doi.org/10.55606/teknik.v3i3.2411>
- Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems*, 226, 107134. <https://doi.org/10.1016/j.knosys.2021.107134>
- Cahyono, N., & Anggista Oktavia Praneswara. (2023). Analisis Sentimen Ulasan Aplikasi TikTok Shop Seller Center di Google Playstore Menggunakan Algoritma Naive Bayes. *Indonesian Journal of Computer Science*, 12(6), 3925–3940. <https://doi.org/10.33022/ijcs.v12i6.3473>
- Demircan, M., Seller, A., Abut, F., & Akay, M. F. (2021). Developing Turkish sentiment analysis models using machine learning and e-commerce data. *International Journal of Cognitive Computing in Engineering*, 2(July), 202–207. <https://doi.org/10.1016/j.ijcce.2021.11.003>
- Ernianti Hasibuan, & Elmo Allistair Heriyanto. (2022). Analisis Sentimen Pada Ulasan Aplikasi Amazon Shopping Di Google Play Store Menggunakan Naive Bayes Classifier. *Jurnal Teknik Dan Science*, 1(3), 13–24. <https://doi.org/10.56127/jts.v1i3.434>
- Fransiska, S., & Irham Gufroni, A. (2020). Sentiment Analysis Provider by.U on Google Play Store Reviews with TF-IDF and Support Vector Machine (SVM) Method. *Scientific Journal of Informatics*, 7(2), 2407–7658. <http://journal.unnes.ac.id/nju/index.php/sji>
- Hakim, B. (2021). Analisa Sentimen Data Text Preprocessing Pada Data Mining Dengan Menggunakan Machine Learning. *JBASE - Journal of Business and Audit Information Systems*, 4(2). <https://doi.org/10.30813/jbase.v4i2.3000>
- Huang, H., Zavareh, A. A., & Mustafa, M. B. (2023). Sentiment Analysis in E-Commerce Platforms: A Review of Current Techniques and Future Directions. *IEEE Access*, 11(August), 90367–90382. <https://doi.org/10.1109/ACCESS.2023.3307308>
- Juanita, S., Adiyarta, K., & Syafrullah, M. (2022). Sentiment analysis on E-Marketplace User Opinions Using Lexicon-Based and Naïve Bayes Model.

- International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, 2022-October(October 2022), 379–382. <https://doi.org/10.23919/EECSI56542.2022.9946537>
- Nur'aeni, N. N., Ainulyaqin, M., & Edy, S. (2024). Dampak Fenomena E-Commerce Pada Tingkat Penjualan Di Pasar Tradisional Ditinjau Dari Psikologi Dan Ekonomi Islam. *Jurnal Ilmiah Ekonomi Islam*, 10(1), 270. <https://doi.org/10.29040/jiei.v10i1.12146>
- Pratmanto, D., Rousyati, R., Wati, F. F., Widodo, A. E., Suleman, S., & Wijianto, R. (2020). App Review Sentiment Analysis Shopee Application in Google Play Store Using Naive Bayes Algorithm. *Journal of Physics: Conference Series*, 1641(1). <https://doi.org/10.1088/1742-6596/1641/1/012043>
- Tabinda Kokab, S., Asghar, S., & Naz, S. (2022). Transformer-based deep learning models for the sentiment analysis of social media data. *Array*, 14(April), 100157. <https://doi.org/10.1016/j.array.2022.100157>
- Turki, T., & Roy, S. S. (2022). Novel Hate Speech Detection Using Word Cloud Visualization and Ensemble Learning Coupled with Count Vectorizer. *Applied Sciences (Switzerland)*, 12(13). <https://doi.org/10.3390/app12136611>
- Ulhaq, D., Masa, A. P. A., & Setyadi, H. J. (2024). Analisis Perbandingan User Experience Aplikasi E-Commerce Shopee Dan Tokopedia Menggunakan Metode User Experience Questionnaire (UEQ). *Adopsi Teknologi Dan Sistem Informasi (ATASI)*, 3(1), 37–44. <https://doi.org/10.30872/atasi.v3i1.1183>
- Valero-Carreras, D., Alcaraz, J., & Landete, M. (2023). Comparing two SVM models through different metrics based on the confusion matrix. *Computers and Operations Research*, 152(December 2022), 106131. <https://doi.org/10.1016/j.cor.2022.106131>