



# Penerapan NLP Menggunakan Algoritma *Naive Bayes*, C4.5, XGBoost untuk Analisis Sentimen Ulasan Produk Kecantikan di Tokopedia dan Shopee

Srikandi Alifya<sup>1\*</sup>, Jasmir<sup>2</sup>, Elvi Yanti<sup>3</sup>

Teknik Informatika, Ilmu Komputer, Universitas Dinamika Bangsa, Indonesia

Email: [srikandialifya@gmail.com](mailto:srikandialifya@gmail.com)<sup>1</sup>, [ijay\\_jasmir@yahoo.com](mailto:ijay_jasmir@yahoo.com)<sup>1</sup>, [elvot92@gmail.com](mailto:elvot92@gmail.com)<sup>3</sup>

Alamat: Jl. Kol. M. Kukuh, Pall Merah, Kec. Kota Baru, Kota Jambi, Jambi 36125

\*Penulis Korespondensi: [srikandialifya@gmail.com](mailto:srikandialifya@gmail.com)

**Abstract.** *The growth of e-commerce in Indonesia has led to an increase in product reviews, including for beauty products on Tokopedia and Shopee. These reviews serve as important sources of information to assess consumer satisfaction; however, manually analyzing thousands of reviews daily is impractical. This study applies Natural Language Processing (NLP) with Naive Bayes, C4.5, XGBoost algorithms to classify sentiment in Indonesian-language reviews. The dataset used consists of 76,256 reviews labeled as positive, negative, and neutral. The research stages include text preprocessing, feature representation using BoW and TF-IDF, data balancing through SMOTE, and model performance evaluation based on accuracy, precision, and recall. Differences in results among the algorithms were analyzed using ANOVA. The results show that Naive Bayes achieved the highest accuracy at 67.71%, followed by XGBoost at 65.91%, and C4.5 at 58.39%, with Naive Bayes performing best in identifying positive and negative sentiments, while XGBoost and C4.5 handled more complex data patterns effectively. These findings provide guidance for sentiment analysis in Indonesian and support businesses in obtaining automated insights from customer reviews to improve product quality and services.*

**Keywords:** *Sentiment Analysis; NLP; Naive Bayes; C4.5; XGBoost*

**Abstrak.** Pertumbuhan e-commerce di Indonesia menyebabkan meningkatnya jumlah ulasan produk, termasuk untuk produk kecantikan di Tokopedia dan Shopee. Ulasan ini menjadi sumber informasi penting untuk menilai kepuasan konsumen, namun analisis manual atas ribuan ulasan setiap hari menjadi tidak praktis. Penelitian ini menggunakan *Natural Language Processing* (NLP) dengan algoritma *Naive Bayes*, C4.5, XGBoost untuk mengklasifikasikan sentimen ulasan berbahasa Indonesia. Dataset yang digunakan terdiri dari 76.256 ulasan dengan label positif, negatif, dan netral. Tahapan penelitian meliputi *text preprocessing*, representasi fitur menggunakan BoW dan TF-IDF, penyeimbangan data melalui SMOTE, serta evaluasi performa model berdasarkan akurasi, presisi, dan recall. Perbedaan hasil antar algoritma dianalisis menggunakan ANOVA. Hasil menunjukkan bahwa *Naive Bayes* memiliki akurasi tertinggi 67,71%, diikuti XGBoost 65,91%, dan C4.5 58,39%, dengan *Naive Bayes* unggul dalam mengenali sentimen positif dan negatif, sedangkan XGBoost dan C4.5 mampu menangani pola data yang lebih kompleks. Temuan ini memberikan panduan bagi analisis sentimen berbahasa Indonesia serta mendukung pelaku bisnis dalam memperoleh informasi otomatis dari ulasan konsumen untuk meningkatkan kualitas produk dan layanan.

**Kata kunci:** Analisis Sentimen; NLP; *Naive Bayes*; C4.5; XGBoost

## 1. LATAR BELAKANG

Perkembangan teknologi informasi yang begitu pesat telah membawa perubahan signifikan dalam berbagai aspek kehidupan manusia, terutama di bidang perdagangan elektronik (*e-commerce*). Saat ini, masyarakat tidak lagi terbatas pada aktivitas berbelanja di toko fisik, melainkan beralih ke platform daring seperti *Tokopedia* dan *Shopee*. Kedua platform tersebut telah menjadi pilihan utama masyarakat Indonesia dalam melakukan transaksi

pembelian berbagai produk, termasuk produk kecantikan yang memiliki permintaan tinggi di pasar digital (Nadia Dwi Oktaviani1, 2025).

Dalam ekosistem *e-commerce*, ulasan pengguna (*review*) memiliki peran penting sebagai sumber informasi bagi calon pembeli. Namun, meningkatnya jumlah pengguna aktif juga menyebabkan volume ulasan bertambah secara masif hingga mencapai ribuan setiap harinya. Kondisi tersebut menjadikan analisis manual terhadap seluruh ulasan menjadi tidak efisien dan memakan waktu (Valen et al., 2023).

Selain volume data yang besar, analisis ulasan pengguna juga dihadapkan pada berbagai tantangan kompleks. Banyak pengguna menulis ulasan dengan bahasa informal, ejaan tidak baku, singkatan, serta campuran bahasa Indonesia dan bahasa asing yang sulit ditafsirkan sistem komputer. Selain itu, ketidakseimbangan jumlah ulasan positif dan negatif (*class imbalance*) juga menjadi kendala yang dapat memengaruhi akurasi hasil klasifikasi sentimen.

Masalah lainnya adalah adanya kata tidak baku dan kesalahan ejaan yang sering muncul, sehingga sistem analisis teks kesulitan mengenali kata tersebut. Untuk mengatasi hal ini, diperlukan tahapan awal berupa *text preprocessing*, yaitu serangkaian langkah pembersihan dan penyiapan data teks sebelum diolah oleh algoritma pembelajaran mesin. Tahapan ini meliputi proses *cleaning* untuk menghapus karakter tidak relevan seperti angka, simbol, URL, dan tanda baca berlebih, *case folding* yang mengubah seluruh huruf menjadi huruf kecil agar data lebih konsisten, *tokenizing* untuk memecah kalimat menjadi satuan kata (*token*) sehingga setiap kata dapat dianalisis secara individual, *normalisasi* yang bertujuan mengganti kata tidak baku atau singkatan menjadi bentuk baku, misalnya “gk” menjadi “tidak” dan “bgt” menjadi “banget”, *stopword removal* untuk menghapus kata-kata umum yang tidak memiliki makna penting seperti “dan”, “yang”, dan “dengan”, serta *stemming* yang mengembalikan setiap kata ke bentuk dasarnya, misalnya “membeli”, “pembelian”, dan “dibeli” menjadi “beli”. Tahapan *preprocessing* ini sangat penting agar sistem dapat memahami teks secara lebih baik dan menghasilkan analisis sentimen yang lebih akurat (Albab et al., 2023).

*Naive Bayes* merupakan algoritma klasifikasi berbasis probabilitas yang berlandaskan *Teorema Bayes*. Algoritma ini menghitung peluang kemunculan kata tertentu dalam kategori sentimen (positif, negatif, atau netral) (Arsadhana et al., 2025). *C4.5* adalah pengembangan dari algoritma *Decision Tree* yang menggunakan konsep *entropy* dan *information gain* untuk membentuk pohon keputusan (Haekal et al., 2021). Sementara itu, *XGBoost* (*Extreme Gradient Boosting*) merupakan algoritma di mana setiap iterasi berfokus pada kesalahan model

sebelumnya. (Wati et al., 2025). Selain pemilihan algoritma, representasi data teks juga berperan penting. Dua metode representasi fitur yang digunakan dalam penelitian ini adalah *Bag of Words (BoW)* dan *TF-IDF (Term Frequency–Inverse Document Frequency)*. *BoW* memiliki keterbatasan dalam menangkap makna kontekstual antar kata. Sementara itu, *TF-IDF* merupakan pengembangan dari *BoW* yang memberikan bobot lebih tinggi pada kata yang penting dan jarang muncul di seluruh dokumen. *TF-IDF* menggabungkan dua aspek, yaitu *term frequency* (frekuensi kemunculan kata dalam dokumen) dan *inverse document frequency* (tingkat kelangkaan kata di seluruh kumpulan dokumen). Dengan metode ini, kata umum seperti “dan”, “yang”, atau “itu” akan memiliki bobot rendah, (Studi Informatika et al., 2025).

Sejumlah penelitian sebelumnya telah dilakukan dalam bidang analisis sentimen menggunakan algoritma *Naive Bayes* dan metode representasi teks seperti *TF-IDF* maupun *BoW*. Penelitian yang dilakukan oleh Billy Gunawan, Helen Sasty Pratiwi, dan Enda Esyudha Pratama (2018) pada *Jurnal Edukasi dan Penelitian Informatika (JEPIN)* menerapkan algoritma *Naive Bayes* dengan pembobotan *TF-IDF* pada ulasan produk daring berbahasa Indonesia. Hasil pengujian menunjukkan tingkat akurasi sebesar 77,78%, dengan nilai *recall* 93,33%, *precision* 77,78%, (Gunawan et al., 2023). Selanjutnya, Binti Najibah Agus Ratri dkk. (2021) dari Universitas Brawijaya meneliti *Analisis Sentimen Review Produk Kecantikan* menggunakan algoritma *Naive Bayes* dengan pembobotan *Term Frequency (TF)* dan *Log TF*. Dari hasil uji coba, metode *TF* menghasilkan akurasi 55%, sedangkan *Log TF* menghasilkan 52%, (Najibah Agus Ratri & Arum Sari, 2021). Kemudian penelitian oleh Jeny Angreyani dan Yonky Fernando (2025) dalam *J-Com (Journal of Computer)* melakukan analisis sentimen ulasan pengguna aplikasi *Shopee* di *Google Play Store* dengan menggunakan algoritma *Naive Bayes* dan representasi fitur *TF-IDF*. Penelitian tersebut menghasilkan akurasi 87%, (Oecd, 2025). Selain penelitian-penelitian tersebut, Penelitian oleh Ambar Fisty Setyaningsih, Dian Septiyani, dan Sisilia Rahma Widiyari (2023) dalam *Jurnal Teknologi Informatika dan Komputer MH. Thamrin* berjudul “*Implementasi Algoritma Naïve Bayes untuk Analisis Sentimen Masyarakat pada Twitter mengenai Kepopuleran Produk Skincare di Indonesia*” menunjukkan bahwa algoritma *Naïve Bayes* mampu memberikan hasil analisis yang akurat terhadap opini publik. Dengan menggunakan 200 data tweet dan tahapan *preprocessing* berbasis *software Orange*, penelitian tersebut menghasilkan akurasi sebesar 86%, presisi 84%, dan *recall* 80%. (Setyaningsih et al., 2023). Penelitian oleh Alfin Cahyo Wibisono, Tiara Siti Nadira, dan Tata Sutabri (2025) dalam *Nusantara Journal of Multidisciplinary Science* meneliti *Analisis Sentimen Pelanggan pada Platform Shopee Menggunakan Metode Naive*

*Bayes*. Penelitian ini menggunakan 714 ulasan pelanggan dan berhasil mencapai akurasi sebesar 80%. (Wibisono et al., 2025).

Untuk mengatasi permasalahan tersebut, penelitian ini memanfaatkan teknik SMOTE. SMOTE bekerja dengan menambahkan data sintetis pada kelas minoritas berdasarkan kemiripan antar sampel, sehingga jumlah data di setiap kelas menjadi lebih seimbang.

Penelitian ini juga menggunakan metode statistik ANOVA (*Analysis of Variance*) untuk membandingkan performa algoritma *Naive Bayes*, *C4.5*, dan *XGBoost*. Melalui ANOVA, dapat diketahui apakah perbedaan nilai akurasi yang dihasilkan ketiga algoritma tersebut benar-benar signifikan secara statistik atau hanya terjadi karena variasi data. Dengan adanya uji ANOVA. (Azhima et al., 2023).

Dari berbagai penelitian tersebut dapat disimpulkan bahwa algoritma *Naive Bayes* memberikan performa yang baik dalam analisis sentimen berbahasa Indonesia, terutama bila dikombinasikan dengan pembobotan *TF-IDF*. Namun, peningkatan akurasi masih dapat dicapai melalui integrasi algoritma lain seperti *C4.5* dan *XGBoost*, yang lebih adaptif dalam mengenali pola kompleks serta akurasi klasifikasi. Melalui penelitian ini dan Berdasarkan uraian latar belakang di atas, penelitian ini berjudul **“Penerapan NLP Menggunakan Algoritma Naive Bayes C4.5, XGBoost untuk Analisis Sentimen Ulasan Produk Kecantikan Di Tokopedia Dan Shopee.”**

## 2. KAJIAN TEORITIS

### *NLP (Natural Language Processing )*

*Natural Language Processing* (NLP) atau Pemrosesan Bahasa Alami merupakan salah satu cabang dari kecerdasan buatan (*Artificial Intelligence*) yang mempelajari bagaimana komputer dapat memahami untuk menciptakan sistem yang mampu mengenali dan merespons bahasa alami layaknya manusia. Melalui penerapan NLP, komputer tidak hanya mampu membaca teks, tetapi juga memahami struktur gramatikal, makna semantik, serta konteks dari suatu kalimat sehingga dapat menghasilkan keluaran yang relevan dan bermakna (Dwi Prasetyo & Betty Yel, 2024).

### *Text Preprocessing*

*Text preprocessing* merupakan tahap fundamental dalam pengolahan data teks sebelum data tersebut dapat dianalisis lebih lanjut menggunakan teknik *Natural Language Processing*

(NLP) .Tahap ini bertujuan untuk menyiapkan data mentah agar menjadi bersih, terstruktur, serta siap diolah secara efisien oleh sistem komputer. Data teks yang bersumber dari *platform e-commerce* seperti *Tokopedia* dan *Shopee* (Aufar et al., 2023).

### **Representasi Teks**

Representasi teks adalah proses untuk mengubah data teks yang awalnya tidak terstruktur menjadi bentuk numerik agar dapat dianalisis dan diolah oleh algoritma *machine learning*. Transformasi ini sangat penting karena sebagian besar algoritma hanya mampu memproses data dalam format angka, sehingga teks perlu direpresentasikan dalam bentuk vektor atau matriks agar informasi yang terkandung di dalamnya dapat dimanfaatkan secara efektif (Chaerul et al., 2025).

### ***Naive Bayes***

*Naive Bayes* merupakan algoritma berbasis probabilitas yang memanfaatkan *Teorema Bayes* untuk memperkirakan kemungkinan suatu data termasuk dalam kategori tertentu. Algoritma ini bekerja dengan asumsi bahwa setiap fitur atau atribut dalam data bersifat independen satu sama lain. Meskipun asumsi *independensi* ini jarang sepenuhnya berlaku di dunia nyata, *Naive Bayes* tetap terbukti memberikan hasil yang baik pada banyak kasus, khususnya pada pengolahan teks, penyaringan email spam, dan analisis sentimen (Syafi'i et al., 2024).

### **Algoritma C45**

*C4.5* merupakan salah satu metode pembelajaran mesin yang digunakan untuk membentuk model berupa pohon keputusan (*decision tree*) dari data pelatihan. Dengan pendekatan ini, *C4.5* mampu menghasilkan struktur pohon yang sistematis, mudah dipahami, dan memiliki kemampuan yang baik dalam mengenali pola tersembunyi di dalam *dataset* (Lukhayu Pritalia, 2018).

### ***Extreme Gradient Boosting Xgboost***

*XGBoost*, atau *Extreme Gradient Boosting*, merupakan salah satu algoritma pembelajaran mesin yang termasuk dalam metode *boosting*, di mana prediksi dari beberapa model lemah (*weak learners*) digabungkan untuk membentuk model yang lebih kuat dan akurat. *XGBoost* menggunakan pendekatan *gradient boosting* yang *iteratif*, di mana setiap model baru fokus memperbaiki kesalahan dari model sebelumnya, sehingga kesalahan prediksi dapat diminimalkan secara bertahap (Tribuana et al., 2025).

## Analisis Sentimen

Analisis sentimen, atau *opinion mining*, merupakan salah satu cabang dari *Natural Language Processing* (NLP) yang fokus pada identifikasi, ekstraksi, serta evaluasi opini atau perasaan subjektif yang terkandung dalam teks. Tujuan utama dari analisis sentimen adalah untuk menentukan polaritas suatu teks, apakah mengandung opini positif, negatif, atau netral. Dengan kata lain, analisis ini memungkinkan sistem komputer untuk secara otomatis memahami bagaimana seseorang mengekspresikan pendapat, persepsi, atau tanggapan terhadap suatu topik, produk, layanan, maupun entitas tertentu (Muzaki et al., 2024).

## Tokopedia

Tokopedia merupakan salah satu *platform e-commerce* pertama di Indonesia yang menyediakan kesempatan bagi individu maupun bisnis untuk membuka toko secara *online*. Didirikan pada 6 Februari 2009 oleh William Tanuwijaya, Tokopedia resmi diperkenalkan melalui situs webnya pada 17 Agustus 2009. Kantor pusat Tokopedia berlokasi di kompleks Ruko Graha Handaya, Unit R, S, dan T, di Jalan Raya Perjuangan No.12-A, Kebon Jeruk, Jakarta Barat. Bangunan kantor ini memiliki empat lantai dengan luas total sekitar 600 m<sup>2</sup> (Salsabila et al., 2022).

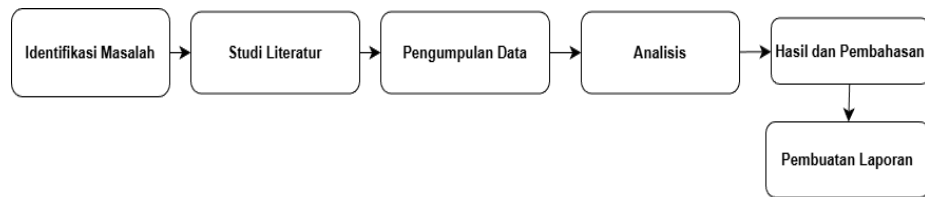
## Shopee

*Shopee* merupakan salah satu *platform e-commerce* yang pertama kali diluncurkan oleh Sea Group pada tahun 2015 di Singapura dan sejak itu berkembang pesat di beberapa negara Asia Tenggara, termasuk Indonesia, Malaysia, Thailand, Taiwan, Vietnam, dan Filipina. Sebagai pendatang baru di industri *e-commerce* Indonesia, *Shopee* berhasil menarik perhatian masyarakat internet dan memperoleh pertumbuhan pengguna aktif yang signifikan, menjadikannya salah satu *marketplace* paling populer di Tanah Air pada tahun 2017 (Agustina et al., 2022).

## 3. METODE PENELITIAN

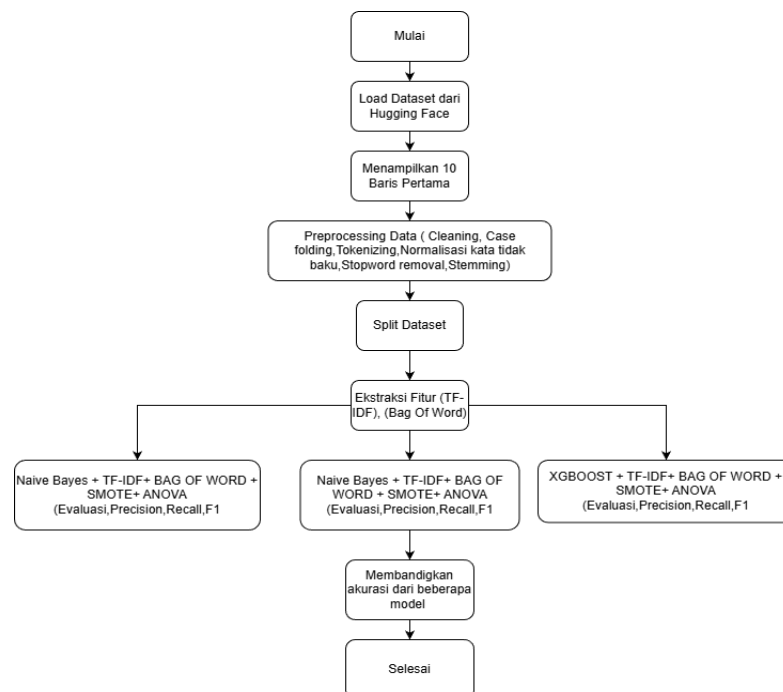
Penelitian ini menggunakan pendekatan terstruktur yang dirancang dalam bentuk kerangka kerja penelitian untuk menjelaskan tahapan pelaksanaan penelitian secara sistematis. *Dataset* yang digunakan diperoleh dari platform *Hugging Face* dalam format CSV, yang berisi ulasan pengguna terhadap produk kecantikan pada platform *Shopee* dan *Tokopedia*. Data

ulasan tersebut dianalisis untuk mengelompokkan sentimen ke dalam tiga kategori, yaitu positif, negatif, dan netral. Kerangka kerja penelitian yang menggambarkan alur penelitian secara umum disajikan pada gambar 1:



**Gambar 1.** Kerangka Kerja Penelitian

Pada tahapan selanjutnya, Kerangka alur eksperimen yang dilakukan pada penelitian berikut yang bisa di lihat pada gambar 2 berikut:



**Gambar 2.** Alur Eksperimen

#### 4. HASIL DAN PEMBAHASAN

Pada tahap penelitian ini, dilakukan analisis terhadap data ulasan produk kecantikan yang telah diproses sebelumnya. membahas hasil eksperimen analisis sentimen dengan menggunakan teknik representasi teks *Bag of Words* (BoW) dan TF-IDF, serta penerapan *anova* untuk seleksi fitur dan *smote* dalam menangani ketidakseimbangan data. Proses klasifikasi sentimen dilakukan dengan menerapkan algoritma *Naive Bayes*, *C4.5*, *XGBoost*.

Kinerja masing-masing model dievaluasi menggunakan metrik akurasi, precision, recall, dan *F1-score*, Serta model dalam mengklasifikasikan sentimen positif, negatif, dan netral.

### Data penelitian

Penelitian ini memanfaatkan data sekunder yang bersumber dari platform *Hugging Face* dengan dataset “*sekarmulyani/ulasan-produk-kecantikan*”. Dataset tersebut memuat 76.256 ulasan pengguna produk kecantikan yang dikumpulkan dari platform *e-commerce Tokopedia* dan *Shopee*, dengan seluruh ulasan ditulis dalam bahasa Indonesia dan tersedia secara terbuka. Berikut tampilan 4 baris awal pertama dataset sehingga struktur dan karakteristik data ulasan yang dianalisis dapat dipahami dengan lebih jelas dan dapat dilihat pada tabel 1.

**Tabel 1.** Dataset Ulasan Produk Kecantikan

No	Review	Bintang 1	Bintang 2	Bintang 3	Bintang 4	Bintang 5	Sentimen	Label
1	pengalaman penggunaan gampang patah	False	True	False	False	False	Negatif	1
2	exp cm sampe bulan sept cpt bgt	False	True	False	False	False	Negatif	1
3	produk bagus hanya saja saat proses pengemasan lama sekali.	False	False	True	False	False	Netral	2
4	kulit langsung dingin dan bersih..mantap	False	False	False	False	True	Positif	3

### Load dataset

Pada tahap ini, dataset penelitian mulai dimuat ke dalam lingkungan kerja menggunakan library *Datasets* dari *Hugging Face*. Dataset yang digunakan merupakan kumpulan ulasan produk kecantikan yang tersedia secara publik di *Hugging Face* dengan nama. Proses pemanggilan dilakukan melalui fungsi `load_dataset()`.

### Pre-Processing

Tahap awal dalam proses *pre-processing* adalah melakukan *setup environment* dengan *menginstal* pustaka yang diperlukan untuk pengolahan teks. Salah satu pustaka yang digunakan

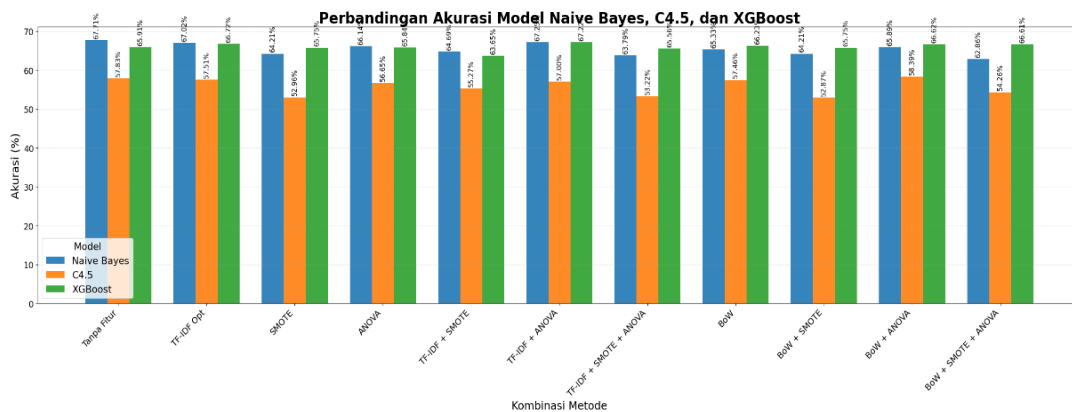
adalah Sastrawi, yaitu *library Python* yang berfungsi untuk melakukan proses *stemming* pada teks berbahasa Indonesia.

### Split Data

Pada tahapan ini Total *dataset* yang digunakan dalam penelitian ini berjumlah 76.256 data. Untuk memperoleh proporsi yang seimbang antara kebutuhan pelatihan dan evaluasi, pembagian dilakukan dengan persentase 75% untuk data latih, 20% untuk data validasi, dan 5% untuk data uji

### Evaluasi Perbandingan Akurasi Model

Pada Bagian ini membahas perbandingan akurasi algoritma *Naive Bayes*, *C4.5*, dan *XGBoost* berdasarkan kombinasi fitur *TF-IDF*, *Bag of Words*, *SMOTE*, dan *ANOVA*. Hasil pengujian menunjukkan bahwa *Naive Bayes* memiliki akurasi tertinggi, diikuti oleh *XGBoost* yang stabil, sementara *C4.5* menghasilkan akurasi terendah. Perbandingan performa ketiga algoritma ditampilkan pada Gambar 3.



**Gambar 3.** Grafik Perbandingan Akurasi Model

## 5. KESIMPULAN DAN SARAN

Hasil penelitian menunjukkan bahwa *Naive Bayes* memiliki performa terbaik, diikuti oleh *XGBoost* yang stabil, sementara *C4.5* menghasilkan akurasi terendah. Representasi fitur dan kualitas *preprocessing* berpengaruh terhadap kinerja model, di mana fitur yang lebih sederhana cenderung memberikan hasil lebih optimal. Penelitian selanjutnya disarankan menggunakan *dataset* yang lebih beragam serta meningkatkan *preprocessing*, penyeimbangan data, dan optimasi parameter untuk memperoleh hasil yang lebih akurat.

## DAFTAR REFERENSI

- Agustina, N., Citra, D. H., Purnama, W., Nisa, C., & Kurnia, A. R. (2022). Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 2(1), 47–54. <https://doi.org/10.57152/malcom.v2i1.195>
- Albab, M. U., P., Y. K., & Fawaiq, M. N. (2023). Optimization of the Stemming Technique on Text Preprocessing President 3 Periods Topic. *Jurnal Transformatika*, 20(2), 1–12. <https://doi.org/10.26623/transformatika.v20i2.5374>
- Arsadhana, M., Efendi, B., & Trihudyatmanto, M. (2025). Analisis Kepuasan Pelanggan Melalui Sentimen Ulasan Menggunakan Algoritma Naive Baye S. *Jurnal Magisma*, XIII(1), 1–8.
- Aufar, A. F., Mochamad Alfian Rosid, Eviyanti, A., & Astutik, I. R. I. (2023). Optimizing Text Preprocessing for Accurate Sentiment Analysis on E-Wallet Reviews. *JICTE (Journal of Information and Computer Technology Education)*, 7(2), 42–50. <https://doi.org/10.21070/jicte.v7i2.1650>
- Azhima, T., Siswa, Y., Informatika, T., Sains, F., Muhammadiyah, U., & Timur, K. (2023). *Komparasi Optimasi Chi-Square , CFS , Information Gain Dan ANOVA Dalam Evaluasi Peningkatan Akurasi Algoritma Klasifikasi Data Performa Akademik Mahasiswa*. 18(1).
- Chaerul, M., Triyono, G., Komputer, M. I., Informasi, F. T., & Luhur, U. B. (2025). Analisis Sentimen Kebijakan Pembatasan Subsidi Bahan Bakar Minyak di Indonesia Tahun 2024 Menggunakan Algoritma Klasifikasi The Sentiment Analysis of the Fuel Subsidy Limitation Policy Using Support Vector Classifier and Random Forest Classifier Algorithm. 5(5), 1471–1484.
- Dwi Prasetyo, A., & Betty Yel, M. (2024). Analisa Sentimen Menggunakan Algoritma C4.5 Dan Naïve Bayes Dengan Topik Artificial Intelligence Pada Media Sosial Twitter (X) Sentiment Analysis Using C4.5 and Naïve Bayes Algorithms With Artificial Intelligence Topics on Social Media Twitter (X). *Journal of Information Technology and Computer Science (INTECOMS)*, 7(5), 1610–1617.
- Gunawan, B., Pratiwi, Sasty, H., & Pratama, Eesyudha, E. (2023). Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes. *Jurnal Edukasi Dan Penelitian Informatika*, 2(1), 95–103.
- Haekal, B. V., Ernawati, I., & Chamidah, N. (2021). Klasifikasi Kepuasan Pengguna Layanan Aplikasi Shopee Menggunakan Metode Decision Tree C4.5. *Informatik : Jurnal Ilmu Komputer*, 17(3), 188. <https://doi.org/10.52958/iftk.v17i3.3648>
- Lukhayu Pritalia, G. (2018). Penerapan Algoritma C4.5 untuk Penentuan Ketersediaan Barang E-commerce. *Indonesian Journal of Information Systems*, 1(1), 47–56. <https://doi.org/10.24002/ijis.v1i1.1727>
- Muzaki, A., Febriana, V., & Cholifah, W. N. (2024). Analisis Sentimen Pada Ulasan Produk di E-Commerce dengan Metode Naive Bayes. *Jurnal Riset Dan Aplikasi Mahasiswa Informatika (JRAMI)*, 5(4), 758–765. <https://doi.org/10.30998/jrami.v5i4.9647>
- Nadia Dwi Oktaviani, R. H. (2025). Pengaruh Kualitas Produk Dan Ewom Terhadap Keputusan Pembelian Skincare Pada E-Commerce Shopee. *Digital Business and Entrepreneurship Journal Vol.*, 3(1), 36–46.

- Najibah Agus Ratri, B., & Arum Sari, Y. (2021). Analisis Sentimen Review Produk Kecantikan menggunakan Metode Naïve Bayes. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 5(12), 2548–2964. <http://j-ptiik.ub.ac.id>
- Oecd, K. (2025). *기후 개발 자원 동향 1) 1*. 5(1), 1–7. <https://doi.org/10.33330/j-com.v5i1.3570>
- Recurrent, M., & Network, N. (2022). *Analisis Sentimen dari Aplikasi Shopee Indonesia Menggunakan*. 5(1), 31–38.
- Salsabila, S. M., Alim Murtopo, A., & Fadhilah, N. (2022). Analisis Sentimen Pelanggan Tokopedia Menggunakan Metode Naïve Bayes Classifier. *Jurnal Minfo Polgan*, 11(2), 30–35. <https://doi.org/10.33395/jmp.v11i2.11640>
- Setyaningsih, A. F., Septiyani, D., & Widiyarsi, S. R. (2023). Implementasi Algoritma Naïve Bayes untuk Analisis Sentimen Masyarakat pada Twitter mengenai Kepopuleran Produk Skincare di Indonesia. *Jurnal Teknologi Informatika Dan Komputer*, 9(1), 224–235. <https://doi.org/10.37012/jtik.v9i1.1409>
- Studi Informatika, P., Matematika dan Ilmu Pengetahuan Alam, F., Raya Kampus UNUD, J., Jimbaran, B., & Selatan, K. (2025). Komparasi Ekstraksi Fitur BoW dan TF-IDF untuk Klasifikasi SMS Menggunakan Naive Bayes I Komang Dwiprayoga a1 , Made Agung Raharja a2. *Jnatia*, 3(2), 247–254.
- Syafi'i, A., Afdal, M., Saputra, E., & Novita, R. (2024). Analisis Sentimen Ulasan Pengguna Aplikasi Penjualan Pulsa Menggunakan Algoritma Naïve Bayes Classifier. *Jurnal Teknologi Sistem Informasi Dan Aplikasi*, 7(3), 1300–1308. <https://doi.org/10.32493/jtsi.v7i3.41364>
- Tribuana, D., Baharuddin, B., & Muhammad Resky, A. (2025). Penerapan Algoritma XGBoost Untuk Prediksi Kepuasan Pelanggan Pada Layanan E-Commerce: Studi Pada Dataset Transaksi Nyata. *Jurnal Teknologi Dan Bisnis Cerdas*, 1(1), 50–59. <https://doi.org/10.64476/jtbc.v1i1.5>
- Valen, R., Bagus, I., & Mahendra, M. (2023). *Membandingkan Analisis Sentimen Review Pelanggan Shopee Dan Tokopedia Menggunakan Google 's NLP API*. 11(4), 655–662.
- Wati, H. L., Anggraeni, N., Kolbiah, S., Hendar, U., & Agustina, N. (2025). *PERBANDINGAN ALGORITMA RANDOM FOREST DAN NARATIF : Jurnal Ilmiah Nasional Riset Aplikasi dan Teknik Informatika*. 07(01), 64–71.
- Wibisono, A. C., Nadira, T. S., & Sutabri, T. (2025). Analisis Sentimen Pelanggan pada Platform Shopee Menggunakan Metode Naive Bayes. *Nusantara Journal of Multidisciplinary Science*, 2(6), 1259–1266. <https://jurnal.intekom.id/index.php/njms>
- Agustina, N., Citra, D. H., Purnama, W., Nisa, C., & Kurnia, A. R. (2022). Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 2(1), 47–54. <https://doi.org/10.57152/malcom.v2i1.195>
- Albab, M. U., P., Y. K., & Fawaiq, M. N. (2023). Optimization of the Stemming Technique on Text Preprocessing President 3 Periods Topic. *Jurnal Transformatika*, 20(2), 1–12. <https://doi.org/10.26623/transformatika.v20i2.5374>
- Arsadhana, M., Efendi, B., & Trihudiyatmanto, M. (2025). Analisis Kepuasan Pelanggan Melalui Sentimen Ulasan Menggunakan Algoritma Naive Baye S. *Jurnal Magisma*, XIII(1), 1–8.

- Aufar, A. F., Mochamad Alfian Rosid, Eviyanti, A., & Astutik, I. R. I. (2023). Optimizing Text Preprocessing for Accurate Sentiment Analysis on E-Wallet Reviews. *JICTE (Journal of Information and Computer Technology Education)*, 7(2), 42–50. <https://doi.org/10.21070/jicte.v7i2.1650>
- Azhima, T., Siswa, Y., Informatika, T., Sains, F., Muhammadiyah, U., & Timur, K. (2023). *Komparasi Optimasi Chi-Square , CFS , Information Gain Dan ANOVA Dalam Evaluasi Peningkatan Akurasi Algoritma Klasifikasi Data Performa Akademik Mahasiswa*. 18(1).
- Chaerul, M., Triyono, G., Komputer, M. I., Informasi, F. T., & Luhur, U. B. (2025). *Analisis Sentimen Kebijakan Pembatasan Subsidi Bahan Bakar Minyak di Indonesia Tahun 2024 Menggunakan Algoritma Klasifikasi The Sentiment Analysis of the Fuel Subsidy Limitation Policy Using Support Vector Classifier and Random Forest Classifier Algorithm*. 5(5), 1471–1484.
- Dwi Prasetyo, A., & Betty Yel, M. (2024). Analisa Sentimen Menggunakan Algoritma C4.5 Dan Naive Bayes Dengan Topik Artificial Intelligence Pada Media Sosial Twitter (X) Sentiment Analysis Using C4.5 and Naive Bayes Algorithms With Artificial Intelligence Topics on Social Media Twitter (X). *Journal of Information Technology and Computer Science (INTECOMS)*, 7(5), 1610–1617.
- Gunawan, B., Pratiwi, Sasty, H., & Pratama, Esyudha, E. (2023). Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes. *Jurnal Edukasi Dan Penelitian Informatika*, 2(1), 95–103.
- Haekal, B. V., Ernawati, I., & Chamidah, N. (2021). Klasifikasi Kepuasan Pengguna Layanan Aplikasi Shopee Menggunakan Metode Decision Tree C4.5. *Informatik : Jurnal Ilmu Komputer*, 17(3), 188. <https://doi.org/10.52958/iftk.v17i3.3648>
- Lukhayu Pritalia, G. (2018). Penerapan Algoritma C4.5 untuk Penentuan Ketersediaan Barang E-commerce. *Indonesian Journal of Information Systems*, 1(1), 47–56. <https://doi.org/10.24002/ijis.v1i1.1727>
- Muzaki, A., Febriana, V., & Cholifah, W. N. (2024). Analisis Sentimen Pada Ulasan Produk di E-Commerce dengan Metode Naive Bayes. *Jurnal Riset Dan Aplikasi Mahasiswa Informatika (JRAMI)*, 5(4), 758–765. <https://doi.org/10.30998/jrami.v5i4.9647>
- Nadia Dwi Oktaviani<sup>1</sup>, R. H. (2025). Pengaruh Kualitas Produk Dan Ewom Terhadap Keputusan Pembelian Skincare Pada E-Commerce Shopee. *Digital Business and Entrepreneurship Journal Vol.*, 3(1), 36–46.
- Najibah Agus Ratri, B., & Arum Sari, Y. (2021). Analisis Sentimen Review Produk Kecantikan menggunakan Metode Naive Bayes. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 5(12), 2548–2964. <http://j-ptiik.ub.ac.id>
- Oecd, K. (2025). *기후 개발 자원 동향 1* 1. 5(1), 1–7. <https://doi.org/10.33330/j-com.v5i1.3570>
- Recurrent, M., & Network, N. (2022). *Analisis Sentimen dari Aplikasi Shopee Indonesia Menggunakan*. 5(1), 31–38.
- Salsabila, S. M., Alim Murtopo, A., & Fadhilah, N. (2022). Analisis Sentimen Pelanggan Tokopedia Menggunakan Metode Naive Bayes Classifier. *Jurnal Minfo Polgan*, 11(2), 30–35. <https://doi.org/10.33395/jmp.v11i2.11640>

- Setyaningsih, A. F., Septiyani, D., & Widiyanti, S. R. (2023). Implementasi Algoritma Naïve Bayes untuk Analisis Sentimen Masyarakat pada Twitter mengenai Kepopuleran Produk Skincare di Indonesia. *Jurnal Teknologi Informatika Dan Komputer*, 9(1), 224–235. <https://doi.org/10.37012/jtik.v9i1.1409>
- Studi Informatika, P., Matematika dan Ilmu Pengetahuan Alam, F., Raya Kampus UNUD, J., Jimbaran, B., & Selatan, K. (2025). Komparasi Ekstraksi Fitur BoW dan TF-IDF untuk Klasifikasi SMS Menggunakan Naive Bayes I Komang Dwiprayoga a1 , Made Agung Rahrja a2. *Jnatia*, 3(2), 247–254.
- Syafi'i, A., Afdal, M., Saputra, E., & Novita, R. (2024). Analisis Sentimen Ulasan Pengguna Aplikasi Penjualan Pulsa Menggunakan Algoritma Naïve Bayes Classifier. *Jurnal Teknologi Sistem Informasi Dan Aplikasi*, 7(3), 1300–1308. <https://doi.org/10.32493/jtsi.v7i3.41364>
- Tribuana, D., Baharuddin, B., & Muhammad Resky, A. (2025). Penerapan Algoritma XGBoost Untuk Prediksi Kepuasan Pelanggan Pada Layanan E-Commerce: Studi Pada Dataset Transaksi Nyata. *Jurnal Teknologi Dan Bisnis Cerdas*, 1(1), 50–59. <https://doi.org/10.64476/jtbc.v1i1.5>
- Valen, R., Bagus, I., & Mahendra, M. (2023). *Membandingkan Analisis Sentimen Review Pelanggan Shopee Dan Tokopedia Menggunakan Google 's NLP API*. 11(4), 655–662.
- Wati, H. L., Anggraeni, N., Kolbiah, S., Hendar, U., & Agustina, N. (2025). *PERBANDINGAN ALGORITMA RANDOM FOREST DAN NARATIF : Jurnal Ilmiah Nasional Riset Aplikasi dan Teknik Informatika*. 07(01), 64–71.
- Wibisono, A. C., Nadira, T. S., & Sutabri, T. (2025). Analisis Sentimen Pelanggan pada Platform Shopee Menggunakan Metode Naive Bayes. *Nusantara Journal of Multidisciplinary Science*, 2(6), 1259–1266. <https://jurnal.intekom.id/index.php/njms>