



## Analisis Sentimen Ulasan Penggunaan Aplikasi Maxim Pada Google Play Store Menggunakan Algoritma *Naive Bayes*, SVM, *CatBoost* Berbasis NLP

Nanda Mediya Sari<sup>1\*</sup>, Jasmir<sup>2</sup>, Elvi Yanti<sup>3</sup>

Teknik Informatika, Ilmu Komputer, Universitas Dinamika Bangsa, Indonesia

Email: [nandamediya07@gmail.com](mailto:nandamediya07@gmail.com)<sup>1\*</sup>, [ijayjasmir@yahoo.com](mailto:ijayjasmir@yahoo.com)<sup>2</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: [nandamediya07@gmail.com](mailto:nandamediya07@gmail.com)

**Abstract.** *Sentiment analysis is a technique in Natural Language Processing (NLP) used to identify user opinion tendencies based on textual reviews. This study analyzer user reviews of the Maxim application on the Google Play Store and compares three Machine Learning algorithms-Naïve Bayes, Support Vector Machine (SVM), and CatBoost-in classifying sentiment. The research stages include data collection, text preprocessing, feature extraction using TF-IDF and Chi-Square, class balancing using SMOTE, and performance evaluation through Accuracy, Precision, Recall, and F1-Score. ANOVA is used to examine the influence of feature selection on model performance. The results show that each model exhibits different performance level across the tested feature combinations. The CatBoost achieved the highest accuracy of 99,26% and demonstrating the most stable performance. Meanwhile, the Naïve Bayes and SVM models experienced performance decreases experiments, especially after applying SMOTE. These findings indicate that the choice of algorithm, feature extraction method, and class balancing technique significantly affects classification outcomes. Overall, CatBoost is identified as the best-performing model, providing more consistent classification result in accordance with the characteristics of the user reviews.*

**Keywords:** *CatBoost; Naive Bayes; Sentiment Analisis; Support Vector Machine;*

**Abstrak.** Analisis sentimen merupakan teknik dalam Natural Language Processing (NLP) yang digunakan untuk mengidentifikasi kecenderungan opini pengguna berdasarkan ulasan. Penelitian ini menganalisis ulasan pengguna aplikasi Maxim pada Google Play Store dan membandingkan tiga algoritma Machine Learning, yaitu Naive Bayes, Support Vector Machine (SVM), dan CatBoost dalam mengklasifikasikan sentimen. Tahapan penelitian meliputi pengambilan data, preprocessing teks, ekstraksi fitur menggunakan TF-IDF dan Chi-Square, penyeimbangan kelas menggunakan SMOTE, serta evaluasi performa melalui metrik Accuracy, Precision, Recall, dan F1-Score. Selain itu, ANOVA digunakan untuk melihat pengaruh seleksi fitur terhadap kinerja model. Hasil penelitian menunjukkan bahwa setiap algoritma memiliki performa berbeda pada kombinasi fitur yang diuji. CatBoost memperoleh akurasi tertinggi sebesar 99,26% dan menunjukkan performa paling stabil. Sementara itu, model Naive Bayes dan SVM mengalami penurunan performa pada beberapa eksperimen, terutama setelah menggunakan SMOTE. Hal ini menunjukkan bahwa pemilihan algoritma, metode ekstraksi fitur, dan teknik penyeimbangan data memiliki pengaruh yang signifikan terhadap hasil klasifikasi. Secara keseluruhan, CatBoost menjadi model terbaik yang mampu memberikan hasil klasifikasi yang lebih konsisten sesuai karakteristik ulasan pengguna.

**Kata kunci:** *CatBoost, Naive Bayes, Analisis Sentimen, Support Vector Machine.*

### 1. LATAR BELAKANG

Perkembangan teknologi digital yang pesat saat ini berpengaruh terhadap berbagai aspek kehidupan manusia, termasuk sektor transportasi. Kehadiran layanan transportasi *online* telah menjadi solusi praktis bagi masyarakat untuk memenuhi kebutuhan mobilitas sehari-hari. Salah satu penyedia layanan transportasi *online* yang cukup populer di Indonesia adalah aplikasi *Maxim*, yang menawarkan layanan transportasi dengan tarif kompetitif dan fitur yang mudah diakses melalui smartphone (Februara, 2024).

Aplikasi *Maxim* adalah sebuah aplikasi untuk memesan layanan sehari-hari, seperti ojek atau taksi *online* (Wananda Purna Maulidi et al., 2024). *Maxim* didirikan pada tahun 2003 di Shadrinsk, Rusia, oleh seorang insinyur muda yang ahli di bidang teknologi komputer. Awalnya, layanan yang ditawarkan hanya berupa taksi berskala kecil di kota tersebut. Seiring perkembangan teknologi dan meningkatnya kebutuhan transportasi, *Maxim* mengembangkan sistem pemesanan yang lebih modern, aman, dan mudah digunakan. Perusahaan pun memperluas operasionalnya ke berbagai kota di Rusia serta menambahkan layanan transportasi *online*, termasuk ojek dan mobil (Oktapianti et al., 2023). Pada tahun 2014, *Maxim* mulai memperluas ke beberapa negara, masuk di Indonesia pada tahun 2018, dengan tujuan menghadirkan layanan transportasi *online* yang praktis dan terjangkau bagi masyarakat (MAXIM, 2022).

Seiring meningkatnya jumlah pengguna, banyak ulasan atau *review* yang diberikan pada *Google Play Store* terkait pengalaman mereka dalam menggunakan aplikasi *Maxim*. Ulasan-ulasan tersebut beragam, mulai dari pujian atas tarif yang terjangkau dan kemudahan penggunaan, hingga keluhan mengenai kualitas layanan, keterlambatan *driver*, atau masalah teknis aplikasi. Banyaknya ulasan yang masuk membuat pengembang kesulitan untuk meninjau satu per satu secara manual guna memahami persepsi pengguna terhadap kualitas layanan (Husain et al., 2024).

Kondisi ini menimbulkan tantangan bagi pihak pengembang untuk dapat menganalisis umpan balik pengguna secara cepat dan akurat. Jika proses analisis dilakukan secara manual, maka proses perbaikan layanan akan memakan waktu cukup lama dan dapat menurunkan tingkat kepuasan pelanggan. Untuk itu dibutuhkan suatu metode otomatis yang mampu mengelompokkan opini pengguna berdasarkan kecenderungan sentimennya, apakah bersifat positif, negatif, atau *netral* (Planlar et al., 2025).

Analisis sentimen adalah salah satu bidang yang mempelajari bagaimana seseorang mengekspresikan pendapat, perasaan, penilaian, maupun emosinya terhadap suatu produk atau layanan melalui teks, seperti postingan, komentar, atau ulasan. Teknik ini berfungsi untuk secara otomatis untuk mengetahui pendapat pengguna (Thomas et al., 2021).

Untuk melakukan analisis sentimen, terdapat beberapa algoritma yang dapat digunakan untuk mengklasifikasikan teks, yaitu *Naive Bayes*, *Support Vector Machine (SVM)*, *Categorical Boosting (CatBoost)* (Mubaroroh et al., 2022). Tetapi metode ini juga dibandingkan dengan dua metode *representasi teks*, yaitu *TF-IDF (Term Frequency-Inverse*

*Document Frequency*) dan *Chi-Square*. Perbandingan ini diharapkan mampu menunjukkan metode representasi mana yang lebih optimal dalam mendukung klasifikasi (Septiani & Isabela, 2022).

Namun, data ulasan pengguna dari *Google Play Store* sering kali tidak seimbang, di mana satu kelas sentimen lebih banyak dibandingkan kelas lainnya. Ketidakseimbangan ini dapat menyebabkan performa model menjadi kurang optimal. Oleh karena itu, penelitian ini menggunakan metode *SMOTE* (*Synthetic Minority Over-Sampling Technique*) untuk menyeimbangkan jumlah data pada masing-masing kelas agar model dapat belajar dengan lebih baik (Hermaliani & Ernawati, 2024).

Selain itu, penelitian ini juga menggunakan metode *ANOVA* (*Analysis of Variance*) sebagai teknik seleksi fitur untuk menentukan fitur yang paling berpengaruh terhadap hasil sentimen. Dengan *ANOVA*, fitur yang tidak relevan dapat dihilangkan, sehingga akurasi model diharapkan meningkat (Hariyanti et al., 2024).

Ada beberapa penelitian yang pernah dilakukan sebelumnya oleh Nizar Fawwazun yang dilakukan pada tahun 2024 Analisis Sentimen Terhadap Aplikasi *Tiktok* Dari Ulasan Pada *Google Play store* Menggunakan Metode *Naïve Bayes*, menghasilkan akurasi 83,66% (Hilmi & Irwiensyah, 2024). Selanjutnya, juga dilakukan oleh Ayu Basirotul Muzayyanah di lakukan pada tahun 2024 tentang Analisis Sentimen Pada Ulasan Aplikasi Ehadrah di *Google Play Store* Menggunakan *Support Vector Machine* (*SVM*), menghasilkan akurasi 94,15% (Muzayyanah et al., 2024). Ada juga penelitian lain di lakukan oleh Faizal Riza dilakukan pada tahun 2024 melakukan Analisis Sentimen Pengguna *Mobile Banking Livin' By Mandiri* Pada *Google Play Store* menggunakan metode *SVM*, *LGBM*, *CatBoost* menghasilkan akurasi 85.06% (Riza & Kurniawan, 2024). Selanjutnya dilakukan juga oleh Kharistianus Darmanto pada tahun 2025 melakukan Analisis Sentimen Ulasan Aplikasi *Jconnect Mobile* di *Google Play Store* Menggunakan Metode *Naïve Bayes* dan *Support Vector Machine* menghasilkan akurasi 87, 07% pada *Naïve Bayes* dan 96,20% pada *SVM* (Darmanto & Sugiono, 2025). Penelitian lain juga dilakukan oleh Gilbert Darmawan pada tahun 2023 melakukan Analisis Sentimen Berdasarkan Ulasan Aplikasi *Mypertamina* Pada *Google Play Store* Menggunakan Metode *Naïve Bayes* menghasilkan akurasi 91,06% (Gilbert et al., 2023).

Berdasarkan latar belakang yang telah di jelaskan, peneliti bermaksud melakukan penelitian yang bertujuan untuk mengetahui persepsi pengguna terhadap aplikasi *Maxim* melalui analisis sentimen ulasan pada *Google Play Store* menggunakan algoritma *Naive Bayes*,

*Support Vector Machine (SVM)* dan *Categorical Boosting (CatBoost)*. Penelitian ini diharapkan dapat memberikan gambaran umum mengenai kepuasan dan keluhan pengguna, sehingga dapat menjadi bahan masukan bagi pengembang aplikasi *Maxim* untuk meningkatkan kualitas layanan. Adapun judul dari penelitian ini adalah “**Analisis Sentimen Ulasan Pengguna Aplikasi Maxim pada Google Play Store Menggunakan Algoritma Naive Bayes, Support Vector Machine (SVM), dan CatBoost Berbasis NLP**”.

## **2. KAJIAN TEORITIS**

### **Analisis Sentimen**

Analisis sentimen adalah teknik analisis data berbasis teks yang menggunakan metode *Natural Language Processing (NLP)* dan *Machine Learning (ML)* untuk mengidentifikasi dan mengklasifikasi makna emosional dalam suatu teks. melalui pendekatan ini, suatu sistem dapat menentukan apakah pernyataan tersebut bersifat positif, negatif atau netral (Tan et al., 2023). Menurut Sajid, analisis sentimen merupakan kegiatan yang bergantung pada konteks, di mana hasil klasifikasi sentimen akan mendapatkan data yang bervariasi. Terdapat tantangan utamanya adalah perbedaan kata-kata opini dan struktur kalimat yang digunakan untuk menggambarkan suatu peristiwa (Kurnia et al., 2023)

### ***Natural Language Processing (NLP)***

*Natural Language processing (NLP)* merupakan salah satu bagian dari kecerdasan buatan yang berperan dalam menghubungkan komputer dengan bahasa manusia. Melalui *NLP*, komputer bisa memahami, menganalisis, *Natural Language processing (NLP)* merupakan salah satu bagian dari kecerdasan buatan yang berperan dalam menghubungkan komputer dengan bahasa manusia. Melalui *NLP*, komputer bisa memahami, menganalisis, mengolah dan memberikan respon terhadap bahasa alami yang digunakan oleh manusia. Tujuan utamanya adalah untuk menciptakan komunikasi yang lebih mudah antara manusia dan mesin dengan memanfaatkan bahasa sehari-hari (Widiantoro et al., 2024).

### ***Machine Learning***

Menurut Alzubi, *Machine Learning* merupakan bagian dari kecerdasan buatan (*Artificial Intelligence/AI*) yang membuat komputer mampu berpikir serta belajar secara mandiri (Ginabila & Fauzi, 2023). Tujuan dari pembelajaran mesin adalah meningkatkan dalam pengolahan data. Dalam praktiknya, manusia sering menghadapi kesulitan dalam

mengelola data yang berukuran besar. Oleh sebab itu, penerapan pembelajaran mesin menjadi salah satu solusi yang efektif untuk mengatasi permasalahan tersebut (Ewen Hokijuliandy, Herlina Napitupulu, 2023)

### ***Data Mining***

Menurut Larose, *Data Mining* merupakan salah satu teknik untuk menggali atau menambang pengetahuan dari sekumpulan data. *Data Mining* adalah suatu analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dengan sebelumnya yang dapat dipahami dan bermanfaat bagi pemilik data (Widaningsih, 2019).

### ***Naive Bayes***

Menurut Yulita, *Naive Bayes* adalah metode yang didasarkan pada perhitungan probabilitas dengan memanfaatkan prinsip *teorema Bayes*. Dalam penerapannya, algoritma ini menggabungkan probabilitas prior dan probabilitas bersyarat ke dalam suatu persamaan yang berfungsi menghitung kemungkinan setiap kelas atau kategori yang akan diklasifikasi (Ginabila & Fauzi, 2023).

### ***Support Vector Machine (SVM)***

Algoritma *Support Vector Machine (SVM)* merupakan metode yang paling sering digunakan dalam analisis sentimen, karena memiliki kemampuan yang baik dalam mengolah data berukuran besar dan dengan tingkat kompleksitas yang tinggi (Rahmasari et al., 2025). Menurut Laksmi, *Support Vector Machine (SVM)* adalah algoritma (*supervised learning*) yang digunakan untuk melakukan analisis sebuah data, baik dalam bentuk klasifikasi ataupun regresi (Ginabila & Fauzi, 2023).

### ***Categorical Boosting (CatBoost)***

Menurut Hancock, *Category Boosting (CatBoost)* merupakan algoritma yang bersifat *open source* dan dikembangkan berdasarkan konsep *gradient boosting* berbasis pohon keputusan. Algoritma ini memiliki kemampuan untuk mengolah fitur yang berkategori atau *string*, jadi dapat digunakan secara efisien pada berbagai jenis data (Syukur et al., 2025).

### ***Term Frequency-Inverse Document Frequency (TF-IDF)***

*Term frequency (TF)* adalah ukuran yang menunjukkan besar frekuensi kemunculan suatu kata (*t*) di dalam sebuah dokumen (*d*). Konsep dasarnya adalah bahwa semakin sering kata tersebut muncul, maka semakin tinggi pula bobotnya (Nico et al., 2022). *Inverse*

*Document Frequency (IDF)* digunakan untuk menentukan tingkat kepentingan suatu kata dalam kumpulan dokumen. Kata yang jarang muncul di dokumen akan memiliki nilai *IDF* yang tinggi, nilai *IDF* dihitung dengan membagi jumlah total dokumen yang memuat kata tersebut, kemudian dilakukan transformasi logaritmik guna menormalkan skala perhitungannya (Septiani & Isabela, 2022).

### ***Chi-Square***

*Chi-Square* adalah metode statistik non-parametrik yang paling sering digunakan dalam penelitian. Metode ini berfungsi untuk membandingkan dua atau lebih kelompok berdasarkan data yang sudah dikategorikan. *Chi-Square* dapat diterapkan untuk analisis satu kelompok dengan data berskala interval atau rasio (Heryana, 2020).

### ***Synthetic Minority Over-Sampling (SMOTE)***

*Synthetic Minority Over-Sampling (SMOTE)* merupakan salah satu algoritma preprocessing yang paling umum dipakai untuk menangani masalah ketidakseimbangan data. Teknik ini bekerja dengan menyesuaikan distribusi data yang tidak seimbang melalui pembuatan sampel sintesis pada kelas minoritas, sehingga performa metode klasifikasi dapat meningkat (Syukron, 2023).

### ***Analysis Of Variance (ANOVA)***

Menurut Setiawan, *Analysis of Variance (ANOVA)* merupakan suatu metode statistik yang digunakan untuk membandingkan nilai rata-rata dari beberapa kelompok data sekaligus. Teknik ini merupakan pengembangan dari uji t, sehingga dapat digunakan untuk menguji perbedaan tiga atau lebih rata-rata populasi dalam satu waktu. Tujuan utama analisis ini adalah untuk mengetahui apakah perbedaan nilai pada variable terkait (Y) yang dipengaruhi oleh variable pada variable bebas (X) (Aziza et al., 2024).

### ***Maxim***

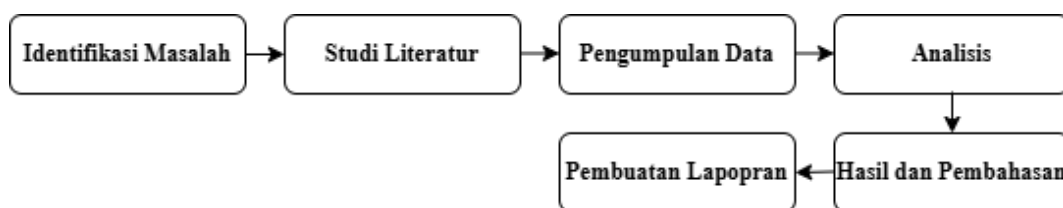
*Maxim* adalah sebuah layanan transportasi *online* yang mulai beroperasi di Indonesia pada bulan Juli 2018 dan dikelola oleh perusahaan bernama PT. Teknologi Indonesia (Afdillah Dwivi, 2023). *Maxim* merupakan perusahaan teknologi berskala internasional yang menyediakan sebuah layanan transportasi *online*, serta berbagai layanan tambahan seperti pengantaran makanan, pengiriman barang, kargo, dan layanan-layanan lainnya (Saripah Aini Pohan et al., 2024).

## **Google Play Store**

*Google Play Store* adalah *platform* untuk perangkat android yang berfungsi sebagai tempat mengunduh dan distribusi aplikasi secara *online*. Melalui ini, pengguna dapat mengunduh berbagai aplikasi-aplikasi dan dapat memberikan rating atau ulasan terkait aplikasi yang mereka gunakan (Kurniawan, 2022).

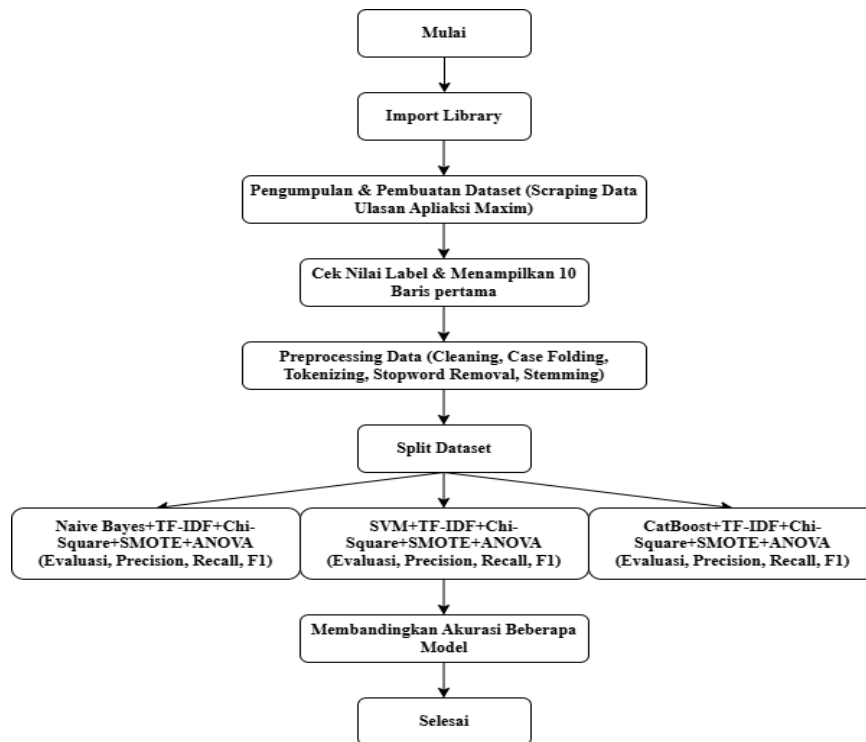
### **3. METODE PENELITIAN**

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen untuk menganalisis sentimen ulasan pengguna aplikasi *Maxim* pada *Google Play Store*. Penelitian bertujuan mengotomatisasi proses pengelompokan sentimen ulasan pengguna yang sulit dilakukan secara manual akibat jumlah data yang besar. Data penelitian berupa ulasan pengguna yang dikumpulkan melalui teknik *web scraping*, kemudian disimpan dalam format *CSV* untuk selanjutnya diolah. Tahapan penelitian meliputi identifikasi masalah, studi literatur, pengumpulan data, analisis data, serta penyusunan laporan akhir. Pada tahap analisis, data teks diproses menggunakan teknik preprocessing, lalu dengan eksperimen fitur menggunakan *TF-IDF* dan *Chi-Square*, serta penyeimbangan kelas menggunakan *SMOTE*. Proses klasifikasi menggunakan algoritma Machine Learning, yaitu *Naive Bayes*, *SVM*, dan *CatBoost*. Evaluasi kinerja model menggunakan metrik *Accuracy*, *Precision*, *Recall*, dan *F1-Score* untuk menilai efektivitas masing-masing algoritma dalam mengklasifikasi sentimen ulasan ke dalam kategori positif, negatif, dan netral. Adapun tahap kerangka kerja penelitian dapat dilihat pada gambar 1:



**Gambar 1.** Kerangka Kerja Penelitian

Selanjutnya, kerangka proses eksperimen dilakukan melalui beberapa tahapan yang tersusun secara sistematis untuk memperoleh hasil yang sesuai dengan tujuan penelitian. Alur eksperimen dapat dilihat pada gambar 2:



Gambar 2. Alur Eksperimen

#### 4. HASIL DAN PEMBAHASAN

Pada tahap penelitian ini, proses analisis dilakukan melalui tahap *preprocessing* data untuk memastikan data ulasan pengguna siap diolah. Selanjutnya, diterapkan metode klasifikasi berbasis *Machine Learning*, yaitu *Naive Bayes*, *Support Vector Machine*, dan *CatBoost* untuk mengklasifikasikan sentimen ulasan pengguna aplikasi *Maxim* pada *Google Play Store*. Metode dikombinasikan dengan *SMOTE* untuk penyeimbangan kelas serta seleksi fitur menggunakan *TF-IDF* dan *Chi-Square*, dengan *ANOVA* digunakan untuk mengevaluasi seleksi fitur terhadap kinerja model. Hasil analisis kemudian diinterpretasikan menjadi informasi sentimen pengguna terhadap layanan aplikasi *Maxim*, baik dalam kategori positif, negatif, dan netral.

##### Pengambilan Data

Data yang digunakan dalam penelitian ini berasal dari ulasan pengguna aplikasi *Maxim* pada *Google Play Store* yang dikumpulkan melalui teknik *web scraping* pada 8 Oktober 2025. Data tersebut mencakup tanggal ulasan, teks komentar, dan label sentimen yang terdiri dari positif, negatif, dan netral.

Hasil *scraping* data *Maxim* dapat dilihat pada tabel 1.

**Tabel 1.** Hasil Scraping Data *Maxim*

1	tanggal	comment	sentiment
2	10/11/2025 8:29	Nggak bisa login	Netral
3	10/11/2025 7:40	Tidak bisa mendaftar	Netral
4	10/11/2025 7:04	Aplikasi jelek. Topup via bank BCA gak masuk-masuk uangnya udah lama bgt	Negatif
5	10/11/2025 6:34	Sebenarnya oke aja nggak ribet tapi pas order makanan lama cari driver nya terus gak bisa dibatalin	Netral
6	10/11/2025 6:23	Maxim keren, murah,cepat, selamanya tetap promo dan drivernya good	Positif
7	10/11/2025 6:19	Nyaman dan cpt sampai	Positif
8	10/11/2025 6:09	Murah, aman, nyaman serta ramah2 dan tertib berlalulintas drivernya	Positif
9	10/11/2025 5:46	MAXIM DI TOPUP IN KASPRO MALAH ERROR, DUITKU ILANG NJING!!	Negatif
10	10/11/2025 5:28	Tolong kalo ada yang pesan makanan siapkan uang talangannya supaya tida ribet ba qris	Netral

### Instalasi Library Sastrawi

Penggunaan sastrawi bertujuan untuk menyederhanakan kata sehingga mempermudah proses analisis pada tahap preprocessing data.

### Import Library

Tahap ini, mengimport *library python* untuk pembersihan dan pengolahan teks (*text preprocessing*). *Library* digunakan untuk menghapus tanda baca, memisahkan kata, menghilangkan kata tidak penting, serta melakukan stemming pada teks berbahasa Indonesia.

### Load Dataset

Pada tahap ini, peneliti memuat dataset hasil pengumpulan ulasan aplikasi *Maxim* dari file CSV. Selanjutnya melakukan standarisasi nama kolom agar lebih seragam dan memudahkan proses pemodelan selanjutnya.

### Cek Nilai Label dan Menampilkan 10 Baris Pertama

Pada tahap ini, peneliti menampilkan jumlah atau distribusi data pada setiap kategori (positif, negatif,netral) untuk mengetahui seberapa seimbang data yang digunakan. Serta menampilkan 10 baris pertama.

## Pre-Processing Data

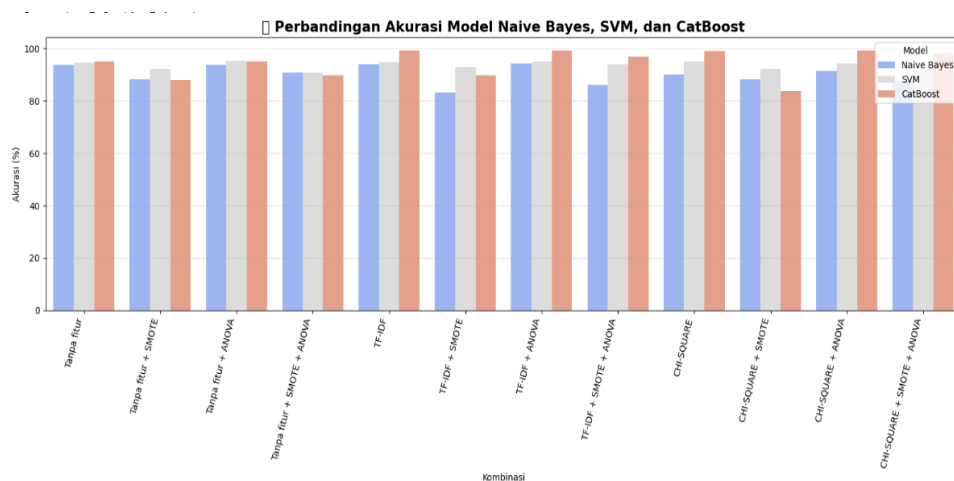
Tahap *pre-processing* data dilakukan sebelum proses analisis pemodelan untuk memastikan kualitas data ulasan pengguna aplikasi *Maxim* agar siap digunakan dalam klasifikasi sentimen. Pada tahap ini, dataset hasil *scraping* dibersihkan dan distandarkan melalui beberapa proses, meliputi *cleaning* untuk menghapus yang tidak relevan, *case folding* untuk menyeragamkan huruf menjadi kecil, *tokenizing* untuk memecahkan teks menjadi kata, normalisasi kata tidak baku, *stopword removal*, serta *stemming* menggunakan library sastrawi. Proses ini bertujuan menghasilkan data yang lebih terstruktur, konsisten, dan fokus pada informasi penting.

## Split Data

Pada tahap ini, dataset dibagi menjadi data latih, validasi, dan uji dengan proporsi 70%, 10%, dan 20%. Pembagian ini bertujuan untuk melatih model, memantau performa selama pelatihan serta mengevaluasi akurasi model secara objektif dengan menjaga keseimbangan distribusi label setiap subset.

## Evaluasi Perbandingan Akurasi Model

Berdasarkan perbandingan akurasi, *CatBoost* dengan kombinasi Chi-Square dan *ANOVA* memperoleh hasil tertinggi sebesar 99,26%. Hasil ini menunjukkan bahwa penggunaan seleksi fitur yang tepat mampu meningkatkan kinerja *CatBoost* secara signifikan dalam klasifikasi sentimen. Adapun evaluasi perbandingan perbandingan akurasi model dapat dilihat pada gambar 3:



Gambar 3. Diagram Evaluasi Perbandingan Model

## 5. KESIMPULAN DAN SARAN

Hasil penelitian menunjukkan bahwa analisis sentimen ulasan pengguna aplikasi Maxim pada Google Play Store berhasil dilakukan dengan mengelompokkan ulasan ke dalam kategori positif, negatif, netral. Penerapan algoritma *Naive Bayes*, *Support Vector Machine*, dan *CatBoost* menghasilkan performa yang berbeda, dengan *Catboost* sebagai algoritma paling akurat dan stabil, khususnya setelah diterapkan teknik SMOTE dan seleksi fitur ANOVA. Selain itu, penggunaan TF-IDF dan *Chi-Square* memberikan pengaruh signifikan terhadap kinerja model, dimana TF-IDF menghasilkan performa yang lebih konsisten, sementara *Chi-Square* efektif sebagai metode seleksi fitur pada beberapa skenario. Penelitian selanjutnya disarankan untuk memperluas sumber dan jumlah data, menguji algoritma lain termasuk pendekatan *deep learning*, serta mengembangkan teknik preprocessing penyeimbangan data, dan seleksi fitur guna meningkatkan kualitas dan akurasi analisis sentimen.

## DAFTAR REFERENSI

- Afdillah Dwivi. (2023). Analisis Sentimen Ulasan Aplikasi Maxim Untuk Peningkatan Layanan Menggunakan Algoritma Naïve Bayes. *EProceedings ...*, 11(4), 1–2.
- Aziza, N., Maarif, U., & Latif, H. (2024). *PENGANTAR STATISTIK : Analisis Varian ( ANOVA )* (Issue February).
- Darmanto, K., & Sugiono, S. (2025). Analisis Sentimen Ulasan Aplikasi Jconnect Mobile Di Google Play Story Menggunakan Algoritma Multinomial Naïve Bayes Dan Support Vectore Machine. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(5), 8357–8364. <https://doi.org/10.36040/jati.v9i5.15083>
- Ewen Hokijuliandy, Herlina Napitupulu, F. F. (2023). *SisInfo Analisis Sentimen Menggunakan Metode Klasifikasi Support Vector Machine ( SVM )* *SisInfo*. 5(2), 40–49.
- Februara, M. S. (2024). Perkembangan Transportasi Ojek Online dan Pengaruhnya. *Jurnal UPI*, 14(1), 31–42.
- Gilbert, Syariful Alam, & M. Imam Sulisty. (2023). Analisis Sentimen Berdasarkan Ulasan Pengguna Aplikasi Mypertamina Pada Google Playstore Menggunakan Metode Naïve Bayes. *STORAGE: Jurnal Ilmiah Teknik Dan Ilmu Komputer*, 2(3), 100–108. <https://doi.org/10.55123/storage.v2i3.2333>
- Ginabila, G., & Fauzi, A. (2023). Analisis Sentimen Terhadap Pemutar Musik Online Spotify Dengan Algoritma Naive Bayes dan Support Vector Machine. *Jurnal Ilmiah ILKOMINFO - Ilmu Komputer & Informatika*, 6(2), 111–122. <https://doi.org/10.47324/ilkominfo.v6i2.180>

- Hariyanti, E., Hostiadi, D. P., Atmojo, Y. P., & Darma, I. M. (2024). *Analisis Perbandingan Metode Seleksi Fitur pada Model Klasifikasi Decision Tree untuk Deteksi Serangan di Jaringan Komputer*. 208–217.
- Hermaliani, E. H., & Ernawati, M. (2024). *Penerapan Metode SMOTE Untuk Mengatasi Imbalanced Data Pada Klasifikasi Ujaran Kebencian*. 4(1).
- Heryana, A. (2020). Jumlah kelompok Fungsi Syarat data. *Universitas Esa Unggul, May*, 1–20. <https://doi.org/10.13140/RG.2.2.23266.15047>
- Hilmi, N. F., & Irwiensyah, F. (2024). Analisis Sentimen Terhadap Aplikasi Tiktok Dari Ulasan Pada Google Playstore Menggunakan Metode Naive Bayes. *Smatika Jurnal*, 14(01), 146–156. <https://doi.org/10.32664/smatika.v14i01.1210>
- Husain, N. P., Sukirman, S., & SAJIAH, S. (2024). Analisis Sentimen Ulasan Pengguna Tiktok pada Google Play Store Berbasis TF-IDF dan Support Vector Machine. *Journal of System and Computer Engineering (JSCE)*, 5(1), 91–102. <https://doi.org/10.61628/jsce.v5i1.1105>
- Kurnia, Purnamasari, I., & Saputra, D. D. (2023). Analisis Sentimen Dengan Metode Naive Bayes, SMOTE Dan Adaboost Pada Twitter Bank BTN. *Jurnal JTIC (Jurnal Teknologi Informasi Dan Komunikasi)*, 7(2), 235–242. <https://doi.org/10.35870/jtik.v7i3.707>
- Kurniawan, S. (2022). *ANALISIS DATASET GOOGLE PLAY STORE MENGGUNAKAN METODE EXPLORATORY DATA ANALYSIS (EDA) Analysis of Google Play Store Datasets Using the Exploratory Data Analysis (EDA) Method*. December. <https://doi.org/10.13140/RG.2.2.14192.12808>
- MAXIM. (2022). Visi & Misi Maxim. *Taximaxim.Com*, 56–78.
- Mubaroroh, H. H., Yasin, H., & Rusgiyono, A. (2022). Analisis Sentimen Data Ulasan Aplikasi Ruangguru Pada Situs Google Play Menggunakan Algoritma Naive Bayes Classifier Dengan Normalisasi Kata Levenshtein Distance. *Jurnal Gaussian*, 11(2), 248–257. <https://doi.org/10.14710/j.gauss.v11i2.35472>
- Muzayyanah, A. B., Pawening, R. E., & Arifin, Z. (2024). Analisis Sentimen Pada Ulasan Aplikasi Ehadrah Di Google Playstore Menggunakan Support Vector Machine (Svm). *IDEALIS: InDonEsiA Journal Information System*, 7(2), 258–266. <https://doi.org/10.36080/idealism.v7i2.3250>
- Nico, Budiyanto, U., & Fatimah, T. (2022). Implementasi Algoritma Pembobotan TF-IDF dan Cosine Similarity untuk Penetapan Kategori Artikel pada Website Universitas Budi Luhur. *Jurnal TICOM: Technology of Information and Communication*, 10(3), 218–222.
- Oktapianti, I., Epty Hidayaty, D., Tuhagana, A., Studi Manajemen, P., Ekonomi dan Bisnis, F., & Buana Perjuangan Karawang, U. (2023). The Influence Of Maxim Online Transportation Service Rates On Digital Financial Transactions On Buana Perjuangan Karawang University Students. *Management Studies and Entrepreneurship Journal*, 4(5), 5736–5742.

- Planlar, E., Kamu, H. K., Konur, C., No, S., Planlari, E., & Kiral, H. (2025). *İrem ÇAĞLAR 1* , *İzzet ARI 2* , *Halis KIRAL 3*. 8(12), 197–212.
- Rahmasari, F., Rahaningsih, N., Dana, R. D., & Rohmat, C. L. (2025). Optimasi Analisis Sentimen Aplikasi Glints Menggunakan Algoritma Support Vector Machine (Svm). *Jurnal Informatika Dan Teknik Elektro Terapan*, 13(1). <https://doi.org/10.23960/jitet.v13i1.5681>
- Riza, F., & Kurniawan, B. (2024). Perbandingan Performa Algoritma Klasifikasi Machine Learning Untuk Analisis Sentimen Pengguna Mobile Banking Livin' by Mandiri Pada Google Play Store. *Ismetek*, 18(2), 116–122.
- Saripah Aini Pohan, Samsudin, & Fathiya Hasyifah Sibarani. (2024). Analisis Sentimen Terhadap Aplikasi Maxim Menggunakan Algoritma Random Forest. *Journal of Science and Social Research*, 7(3), 1201–1208.
- Septiani, D., & Isabela, I. (2022). Term Frequency Inverse Document Frequency (Tf-Idf) Analysis in Information Retrieval in Text Documents. *Jurnal Sistem Dan Teknologi Informasi Indonesia(SINTESIA)*, 1(2), 81–88.
- Syukron, A. (2023). Penerapan Metode Smote Untuk Mengatasi Ketidakseimbangan Kelas Pada Prediksi Gagal Jantung. 10(1), 47–50.
- Syukur, M. A. A., Suhartono, S., & Chamidy, T. (2025). Prediksi Kualitas Udara Menggunakan Metode CatBoost. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 10(2), 249–258. <https://doi.org/10.14421/jiska.2025.10.2.249-258>
- Tan, K. L., Lee, C. P., & Lim, K. M. (2023). A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research. *Applied Sciences (Switzerland)*, 13(7). <https://doi.org/10.3390/app13074550>
- Thomas, S., Yuliana, & Noviyanti. P. (2021). Study Analisis Metode Analisis Sentimen pada YouTube. *Journal of Information Technology*, 1(1), 1–7. <https://doi.org/10.46229/jifotech.v1i1.201>
- Wananda Purna Maulidi, Hani Fatu Zahra, & Syamsul Hidayat. (2024). Analisis Persaingan Ojek Online Pada Aplikasi Gojek, Maxim Dan Grab. *Intellektika: Jurnal Ilmiah Mahasiswa*, 2(2), 70–79. <https://doi.org/10.59841/intellektika.v2i2.950>
- Widaningsih, S. (2019). Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm. *Jurnal Tekno Insentif*, 13(1), 16–25. <https://doi.org/10.36787/jti.v13i1.78>
- Widiantoro, A., Dwiyoga Mustafid Sanjaya, & Ridwan. (2024). *Pengantar NLP Dan Topik Model*.