

KLASIFIKASI DATA TEKS ULASAN TOKO SEPATU SOVELLA DI SHOPEE MENGGUNAKAN ALGORITMA NAIVE BAYES DALAM PROSES DATA MINING

Zahratu Azfil Yugian, Agustia Hananto, Tukino, Elfina Novalias

Sistem Informasi, Universitas Buana Perjuangan Karawang
Jalan Ronggo Waluyo, Puseur Jaya, Telukjambe Timur, Karawang, Jawa Barat, Indonesia
si23.zahratuyugian@mhs.ubpkarawang.ac.id

ABSTRAK

Peningkatan aktivitas belanja berani melalui platform *e-commerce* seperti Shopee menghasilkan sejumlah besar data ulasan pelanggan yang mengandung informasi penting mengenai persepsi dan kepuasan pengguna. Namun, teks data yang tidak terstruktur dalam jumlah besar susah untuk proses analisis manual secara efektif. Penelitian ini bertujuan untuk mengembangkan model klasifikasi sentimen ulasan pelanggan pada toko sepatu Sovella di Shopee menggunakan algoritma *Naive Bayes*. Metode yang digunakan meliputi pengumpulan 1.020 ulasan data, pra-pemrosesan, ekstraksi fitur menggunakan TF-IDF, dan proses klasifikasi menggunakan *Multinomial Naive Bayes*. Evaluasi dilakukan dengan metrik akurasi, presisi, recall, dan F1-score. Hasil pengujian menunjukkan bahwa model mencapai akurasi sebesar 94,41% dengan rata-rata F1-score 93,48%, menunjukkan kinerja klasifikasi yang sangat baik dalam membedakan empat kategori sentimen: sangat bagus, bagus, tidak bagus, dan sangat tidak bagus. Pendekatan menggunakan *Naive Bayes* efektif untuk analisis sentimen ulasan *e-commerce* dan berpotensi menjadi alat pendukung dalam pengambilan keputusan bisnis berbasis data.

Kata kunci: klasifikasi teks, analisis sentimen, data mining, *Naive Bayes*, ulasan Shopee, *E-Commerce*.

1. PENDAHULUAN

Perkembangan pesat teknologi digital telah membawa perubahan besar dalam perilaku konsumsi masyarakat, terutama dalam cara berbelanja. Situs *e-commerce* seperti Shopee sekarang berfungsi tidak hanya sebagai tempat transaksi jual beli, tetapi juga sebagai tempat interaktif bagi pengguna untuk memberikan komentar tentang produk dan layanan yang mereka terima. Ulasan ini, yang sebagian besar terdiri dari teks yang tidak terstruktur, menyimpan informasi berharga yang menggambarkan pandangan, tingkat kepuasan, dan ekspektasi dari pelanggan. Namun, dari ulasan data yang tersedia bisa menjadi tantangan tersendiri dalam analisis secara manual. Dengan pendekatan probabilistik yang sederhana dan efektif dengan menggunakan, salah satu metode klasifikasi untuk menentukan gambaran persepsi masyarakat di dalam Text Mining adalah metode *Naive Bayes* yang sering disebut dengan *Naive Bayes Classifie* [1].

Salah satu metode yang sering diterapkan dalam analisis data berbasis teks adalah teknik text mining, terutama metode klasifikasi dengan menggunakan algoritma *Naive Bayes*. Algoritma ini terkenal karena pendekatan probabilistiknya yang sederhana, cepat, serta memiliki kinerja yang kompetitif dalam hal klasifikasi teks. Dalam ulasan ini, hal utama yang dibahas dalam permasalahan ini adalah bagaimana melakukan klasifikasi ulasan pelanggan pada toko Sovella di Shopee ke dalam kategori sentimen yang memiliki arti (yaitu: sangat baik, baik, kurang baik, dan sangat kurang baik). Analisis data sentimen adalah metode untuk mengidentifikasi, mengekstraksi, atau menemukan informasi emosional dari teks, terutama dalam konteks opini atau emosi pengguna. Analisis ini

terutama bertujuan untuk memahami perasaan atau pendapat yang diungkapkan pengguna melalui ulasan di Shopee [2].

Tujuan dari penelitian ini adalah untuk menerapkan algoritma *Naive Bayes* dalam mengklasifikasikan ulasan pelanggan di toko Sovella, menciptakan model klasifikasi sentimen yang dapat beradaptasi dengan data *e-commerce*, serta melakukan evaluasi kinerja model berdasarkan metrik seperti akurasi, presisi, recall, dan f1-score. Selain itu, hasil klasifikasi yang dihasilkan diharapkan dapat memberikan kontribusi berarti dalam pengambilan keputusan strategi yang berbasis data, terutama bagi pelaku bisnis yang ingin meningkatkan kualitas layanan, kepercayaan dari konsumen, dan efisiensi operasional mereka.

Untuk mencapai tujuan tersebut, penelitian ini disusun melalui langkah-langkah sistematis yang dimulai dengan pengumpulan data ulasan pelanggan secara representatif dari platform Shopee, diikuti dengan tahap preprocessing untuk membersihkan dan merestrukturisasi teks agar siap untuk dianalisis. Selanjutnya, ekstraksi fitur dilakukan dengan pendekatan TF-IDF untuk merepresentasikan dokumen dalam format numerik, dan kemudian dilanjutkan dengan melakukan klasifikasi menggunakan algoritma *Naive Bayes*. Tahapan akhir adalah model evaluasi yang dilakukan dengan mengukur kinerja klasifikasi sesuai dengan metrik kuantitatif yang relevan, sambil menyusun interpretasi hasil sebagai dasar untuk rekomendasi yang praktis. Melalui pendekatan ini, Penelitian ini diharapkan dapat memberikan kontribusi penting terhadap analisis sentimen pada aplikasi seluler dan memberikan

pemahaman yang lebih mendalam mengenai persepsi pengguna terhadap aplikasi toko Shopee [3].

2. TINJAUAN PUSTAKA

Dalam era digital saat ini, platform e-commerce seperti Shopee telah menjadi salah satu media utama untuk aktivitas jual beli secara online. Platform e-commerce tidak hanya terbatas untuk transaksi, tetapi juga menghasilkan sejumlah data yang tidak terstruktur berupa ulasan atau review pengguna. Data teks ini mengandung informasi yang baik untuk peningkatan layanan maupun analisis kepuasan pelanggan. Oleh karena itu, diperlukan pendekatan yang tepat untuk mengolah dan menginterpretasikan data. Salah satu pendekatan yang digunakan untuk mengelola data teks adalah proses data mining, khususnya teknik klasifikasi teks. Dalam penelitian ini, algoritma Naive Bayes sering menjadi pilihan utama dengan kesederhanaannya dan efisiensinya.

Penelitian pertama dari Penelitian ketiga dari [4] dalam jurnal yang berjudul “Penerapan Algoritma Naive Bayes pada Analisis Sentimen Ulasan Aplikasi Whoosh – Kereta Cepat Di Google Play Store”. Penelitian ini fokus pada penerapan Algoritma Naive Bayes untuk menganalisis sentimen ulasan pengguna aplikasi WHOOSH di Google Play Store. Algoritma Naive Bayes dipilih karena efisiensi dan implementasinya yang mudah. Menggunakan dataset 500 ulasan yang sudah dibersihkan dan dilabeli (positif atau negatif), model dilatih dan mencapai akurasi 81.25%. Tingginya precision untuk kelas positif (90%) menunjukkan kemampuan model mengidentifikasi ulasan positif secara tepat.

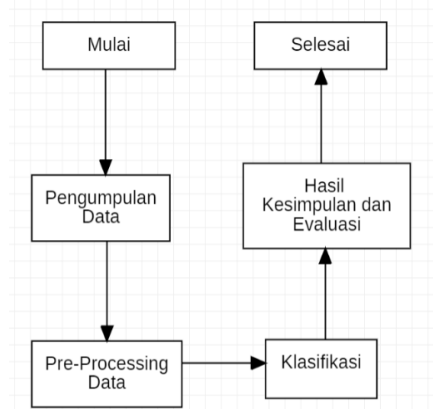
Penelitian kedua dari [5] dalam jurnal yang berjudul Analisis Sentimen Ulasan Pelanggan Terhadap Layanan Grab Indonesia menggunakan Multinomial Naïve Bayes Classifier” Peneliti menggunakan metode naïve bayes untuk mengklasifikasikan pendapat tentang layanan grab Indonesia. Analisis sentimen, yang juga dikenal sebagai penggalian opini, adalah studi tentang pikiran, sentimen, penilaian, sikap, dan emosi seseorang mengenai suatu objek.

Penelitian ketiga dari [6] dalam penelitian yang berjudul “Analisis Data Sentimen Ulasan Aplikasi Dana di Google Play Store Menggunakan Algoritma Naïve Bayes” Studi ini menggunakan teknik klasifikasi Naive Bayes, yang terkenal karena kecepatan dan keakuratan pemrosesan datanya yang luar biasa, untuk menganalisis data sentimen dari ulasan Aplikasi Dana di Google Play Store.

Dengan menjadi satu teknik data mining dan algoritma Naive Bayes, penelitian ini bertujuan untuk menghasilkan model klasifikasi yang akurat dalam mengidentifikasi sentimen pelanggan terhadap produk dan layanan toko sepatu Sovella, dan sehingga dapat meningkatkan kepuasan pelanggan.

3. METODE PENELITIAN

Penelitian ini menggunakan model model klasifikasi data teks ulasan pada platform Shopee menggunakan algoritma *Naive Bayes* dalam konteks *data mining*. Metode ini digunakan karena mampu mengidentifikasi pola sentimen dalam data teks secara sistematis dan terukur. Seluruh proses dilakukan melalui tahapan *preprocessing* data, ekstraksi fitur, pembangunan model klasifikasi, serta evaluasi performa model menggunakan metrik yang umum dalam analisis klasifikasi, seperti akurasi, presisi, recall, dan F1-score.



Gambar 1. Tahapan Penelitian

3.1. Pengumpulan Data

Langkah selanjutnya penelitian ini yaitu “Pengumpulan Data”, peneliti mengambil objek penelitian yaitu : Toko sepatu Sovella di aplikasi shopee. Peneliti ini mencari beberapa informasi lalu di kumpulkan menjadi sebuah data yang berupa data ulasan pelanggan tentang Toko sepatu Sovella di Shopee, dari ulasan pelanggan lalu dibuat untuk dianalisis sentimen dari pelanggan. Oleh karena itu, kualitas dan kuantitas sangat menentukan efektivitas model klasifikasi yang akan di buat dan proses pengumpulan data harus dilakukan secara cermat dan terstruktur.

3.2. Pre-Processing Data

Setelah data sudah terkumpul, tahapan berikutnya adalah “Pre-Processing Data”. Tahapan ini sangat penting dalam analisis teks yang bersifat tidak terstruktur. Proses pre-processing ini melibatkan berbagai teknik yang digunakan seperti mengubah seluruh huruf menjadi huruf kecil, menghapus kata-kata umum yang tidak bermakna, mengubah kata ke bentuk dasarnya. Tujuan dari tahapan ini adalah menghilangkan noise dan menyajikan data yang sudah bersih, sehingga dapat meningkatkan akurasi model.

3.3. Klasifikasi

Setelah data diproses, dilakukan tahap “Klasifikasi”, yaitu proses dimana tahapan ini sudah menggunakan algoritma seperti *Naive Bayes* digunakan untuk mengklasifikasikan data ke dalam sentimen (sangat bagus, bagus, tidak bagus, dan sangat tidak bagus). Algoritma ini menghitung

probabilitas berdasarkan distribusi kata dalam setiap kelas nya. Meskipun Naive Bayes ini sederhana tetapi terbukti dari penelitian sebelum nya, bahwa Naive Bayes sangat efektif dalam menangani kalsifikasi data teks karena efisiensi dan kecepatan proses nya terutama dalam dataset nya.

3.4. Hasil Kesimpulan dan Evaluasi (evaluasi model, confusion matriks)

Selanjutnya, hasil klasifikasi dianalisis pada tahap “Hasil Kesimpulan dan Evaluasi”. Evaluasi model dilakukan menggunakan berbagai matriks seperti confusion matriks, akurasi, presisi, recall, dan F1-score. Melalui evaluasi ini, peneliti dapat menilai sejauh mana model berhasil mengklasifikasikan data secara akurat, serta dapat mengidentifikasi kesalahan melalui perbaikan. Tahap ini juga menghasilkan kesimpulan dari data yang sudah dilalui berbagai tahap, serta memberikan rekomendasi kepada peneliti lain untuk pengembangan penelitian selanjutnya.

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

1	Ulasan	Sentimen
2	Bahan terasa cukup kokoh meskipun belum diuji untuk jangka panjang.	Bagus
3	Pilihan yang sangat tepat, melebihi ekspektasi dalam hal kenyamanan dan tampilan.	Sangat Bagus
4	Sangat mengecewakan, kualitas jauh di bawah standar.	Sangat Tidak Bagus
5	Pengiriman tepat waktu dan kondisi produk baik saat diterima.	Bagus
6	Desain terlihat bagus di gambar, namun realitanya kurang menarik.	Tidak Bagus
7	Warna produk berbeda dengan gambar yang tertera di etalase toko.	Tidak Bagus
8	Pengiriman tepat waktu dan kondisi produk baik saat diterima.	Bagus
9	Sepatu menyebabkan luka dan rasa tidak nyaman parah.	Sangat Tidak Bagus
10	Desain menarik dan cocok untuk digunakan dalam berbagai suasana.	Bagus
1003	Layanan pelanggan cukup membantu saat terjadi kendala kecil.	Bagus
1004	Penjual tidak merespon komplain dan tidak memberikan solusi.	Sangat Tidak Bagus
1005	Bahan terasa cukup kokoh meskipun belum diuji untuk jangka panjang.	Bagus
1006	Pengiriman memakan waktu lebih lama dari yang dijanjikan.	Tidak Bagus
1007	Warna produk berbeda dengan gambar yang tertera di etalase toko.	Tidak Bagus
1008	Sepatu nyaman dipakai, bahan berkualitas tinggi dan pengiriman sangat cepat.	Sangat Bagus
1009	Sepatu menyebabkan luka dan rasa tidak nyaman parah.	Sangat Tidak Bagus
1010	Penjual tidak merespon komplain dan tidak memberikan solusi.	Sangat Tidak Bagus
1011	Ukuran sepatu sesuai dan tidak menimbulkan rasa tidak nyaman.	Bagus
1012	Sepatu kurang nyaman dipakai dan terasa sempit pada bagian depan.	Tidak Bagus
1013	Barang diterima dalam kondisi sangat baik dan pengemasan sangat rapi.	Sangat Bagus
1014	Produk datang dalam kondisi rusak dan sangat mengecewakan.	Sangat Tidak Bagus
1015	Layanan pelanggan cukup membantu saat terjadi kendala kecil.	Bagus
1016	Produk sama sekali tidak sesuai dengan informasi yang diberikan.	Sangat Tidak Bagus
1017	Bahan terasa panas di kaki dan kurang fleksibel untuk aktivitas panjang.	Tidak Bagus
1018	Kualitas premium dan sangat sesuai dengan deskripsi. Layanan pelanggan pun responsif.	Sangat Bagus
1019	Desain terlihat bagus di gambar, namun realitanya kurang menarik.	Tidak Bagus
1020	Produk sangat memuaskan, kualitas luar biasa dan sesuai dengan ekspektasi.	Sangat Bagus
1021	Produk datang dalam kondisi rusak dan sangat mengecewakan.	Sangat Tidak Bagus

Gambar 2. Pengumpulan Data

Penelitian ini menggunakan data berupa ulasan pelanggan dari toko sepatu Sovella di platform Shopee. Sebanyak 1020 ulasan dikumpulkan, yang merepresentasikan berbagai persepsi konsumen terhadap produk, mencakup aspek seperti kenyamanan, ukuran, kualitas produk, dan kesesuaian dengan deskripsi. Setiap ulasan diberi label sentimen ke dalam empat kategori yang dimana label ini digunakan sebagai target klasifikasi pada tahapan berikutnya Label tersebut, yaitu:

- Sangat Bagus
- Bagus
- Tidak Bagus
- Sangat Tidak Bagus

4.2. Pre-processing Data

Langkah pre-processing bertujuan untuk membersihkan data teks ulasan dari unsur-unsur yang tidak relevan dan menyusunnya dalam format yang tidak relevan dan menyusunnya dalam format yang dapat dianalisis secara akurat oleh model. Tahapan pre-processing meliputi:

- Case folding: mengubah semua teks menjadi huruf kecil.
- Tokenizing: memisahkan teks ke dalam satuan kata.
- Stopword removal: menghapus kata-kata umum yang tidak bermakna (seperti “dan”, “yang”, “dengan”).
- Stemming: mengubah kata ke bentuk dasar (contoh: “membeli” menjadi “beli” atau kata dikembalikan ke bentuk dasarnya menggunakan algoritma NLP.
- Filtering: menyaring simbol, angka, dan karakter khusus yang tidak digunakan.

Tahapan ini penting untuk mengurangi *noise* dalam data, serta meningkatkan kualitas informasi yang akan digunakan dalam proses klasifikasi.

4.3. Ekstraksi Fitur dengan TF-IDF

Setelah data dibersihkan, tahap berikutnya adalah melakukan ekstraksi fitur menggunakan TF-IDF (Term Frequency-Inverse Document Frequency). TF-IDF digunakan untuk mengukur pentingnya suatu kata dalam sebuah dokumen relatif terhadap seluruh kumpulan dokumen [7]

Output dari proses ini adalah representasi numerik dari ulasan teks yang siap digunakan oleh algoritma klasifikasi. Semakin besar nilai TF-IDF suatu kata, semakin penting kata tersebut dalam dokumen tersebut dibandingkan dengan dokumen lain. Contoh kata dengan nilai TF-IDF tinggi dalam ulasan:

- “Nyaman”
- “Cepat”
- “Sesuai”
- “Cacat”
- “Ukuran”

4.4. Klasifikasi dengan Algoritma Naive Bayes

Setelah data siap, dilakukan klasifikasi menggunakan algoritma *Naive Bayes* khususnya varian *Multinomial Naive Bayes*, yang cocok untuk data teks. *Naive Bayes* dipilih karena sifatnya yang sederhana, cepat, dan terbukti efektif untuk klasifikasi teks berbasis probabilistik [8].

Prosesnya meliputi:

- Data dibagi menjadi 80% data latih (816 data) dan 20% data uji (204 data).
- Model dilatih menggunakan data latih untuk mengenali pola distribusi kata terhadap masing-masing kelas sentimen.
- Kemudian model diuji dengan data uji untuk melihat akurasi dan kemampuan generalisasi.

4.5. Evaluasi Model

Untuk mengevaluasi performa model klasifikasi, digunakan beberapa metrik yaitu:

- Akurasi

Mengukur persentase prediksi yang benar dari keseluruhan data uji.

$$\text{Akurasi} = \frac{\text{Jumlah prediksi benar}}{\text{Total Data Uji}} = \frac{963}{1020}$$

Hasil = 94.41%

Berdasarkan hasil pengujian model klasifikasi teks menggunakan algoritma Naive Bayes, di peroleh nilai akurasi sebesar 94,41% yang diperoleh dari 963 prediksi benar terhadap total 1.020 data di uji. Capaian akurasi ini menunjukkan bahwa model yang dibangun mampu melakukan klasifikasi ulasan pelanggan terhadap produk toko sepatu Sovella di shopee dengan tingkat ketepatan yang sangat tinggi.

b. Precision, Recall dan F1-Score

Tabel 1. Precision, Recall dan F1-Score

Kelas	Precision	Recall	F1-Score
Sangat Bagus	94,33%	8,13%	96,19%
Bagus	91,27%	92,00%	91,63%
Tidak Bagus	89,80%	91,66%	90,72%
Sangat Tidak Bagus	95,42%	95,38%	95,40%

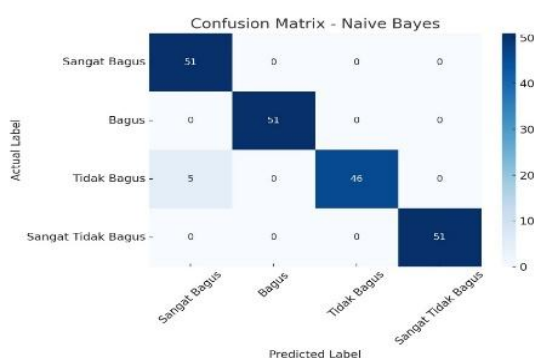
Nilai F1-Score rata-rata 93,48%

c. Confusion Matrix

Classification Report

	precision	recall	f1-score	support
Sangat Bagus	0.91	1.00	0.95	51
Bagus	1.00	1.00	1.00	51
Tidak Bagus	1.00	0.90	0.95	51
Sangat Tidak Bagus	1.00	1.00	1.00	51
accuracy			0.98	204
macro avg	0.98	0.98	0.98	204
weighted avg	0.98	0.98	0.98	204

Gambar 3. Hasil precision, recal, F1-Score, Support



Gambar 4. Confusion Matrix – Naive Bayes

Confusion matrix hasil klasifikasi terhadap 1020 data ulasan memberikan gambaran visual seberapa sering model mengklasifikasikan data ke kelas yang benar atau salah. Pada hasil diatas terlihat bahwa :

- Akurasi tinggi (98%) menunjukkan bahwa model bekerja sangat baik.
- Kelas “Tidak Bagus” mengalami sedikit kesalahan klasifikasi ke “Sangat Bagus”

sebanyak 5 data.

- Precision dan recall sempurna pada kelas Bagus dan Sangat Tidak Bagus, menunjukkan klasifikasi yang sangat presisi.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma Naive Bayes mampu melakukan klasifikasi terhadap data teks ulasan pelanggan toko sepatu Sovella di Shopee dengan tingkat efektivitas yang baik. Pendekatan ini terbukti efisien dalam mengidentifikasi sentimen pelanggan ke dalam empat kategori, yaitu sangat bagus, bagus, tidak bagus, dan sangat tidak bagus, dengan waktu komputasi yang relatif singkat dan akurasi klasifikasi yang kompetitif. Penerapan data mining terhadap ulasan pelanggan tidak hanya memberikan pemahaman mendalam terhadap persepsi konsumen, namun juga berpotensi besar sebagai dasar pengambilan keputusan strategi bagi pelaku bisnis. Dalam konteks toko Sovella, hasil klasifikasi dapat dimanfaatkan untuk memancarkan kualitas produk, pelayanan, serta aspek logistik yang menjadi perhatian pelanggan.

Model Naive Bayes menunjukkan performa yang baik dalam mengklasifikasikan ulasan Shopee dengan 4 kelas sentimen. Hal ini disebabkan oleh: Kualitas data latih yang representatif, Proses pre-processing dan TF-IDF yang optimal. Kesesuaian algoritma *Naive Bayes* dengan data teks.

Meskipun demikian, terdapat sedikit kesalahan prediksi antar kelas yang mirip makna. Untuk peningkatan ke depan, bisa dilakukan: Penambahan data latih, Penyesuaian threshold probabilitas, Penggabungan dengan algoritma lain (*ensemble*)

DAFTAR PUSTAKA

- [1] M. Darwis, S. S. S. Huda, and A. Hananto, “Penerapan Algoritma Naive Bayes pada Analisis Sentimen,” *Jurnal Teknologi Informasi*, vol. 6, no. 1, pp. 45–52, 2021.
- [2] R. Kurniawan and Y. Arie Wijaya, “Analisis Data Sentimen Ulasan Pengguna Aplikasi Shopee di Google Play Store dengan Klasifikasi Algoritma Naive Bayes,” vol. 6, no. 1, 2024.
- [3] A. Nurian, M. S. Ma’arif, I. N. Amalia, and C. Rozikin, “ANALISIS SENTIMEN PENGGUNA APLIKASI SHOPEE PADA SITUS GOOGLE PLAY MENGGUNAKAN NAIVE BAYES CLASSIFIER,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 1, Jan. 2024, doi: 10.23960/jitet.v12i1.3631
- [4] Hartati, T., Sohadi, R. T., Tohidi, E., & Wahyudin, E. (2024). *Penerapan Algoritma Naive Bayes pada Analisis Sentimen Ulasan Aplikasi Whoosh – Kereta Cepat Di Google Play Store*. 6(1), 244–249.
- [5] Dwijayanti, M., Noor Hasan, F., & Zein Adam, R. (2022). Analisis Sentimen Pada Ulasan Pelanggan Menggunakan Metode Naive Bayes

- Classifier (Studi Kasus: Grab Indonesia). *Prosiding Seminar Nasional Teknoka*, 6(January), 93–99. <https://doi.org/10.22236/teknoka.v6i1.441>
- [6] Surya Sayogo, D., Irawan, B., & Bahtiar, A. (2024). Analisis Data Sentimen Ulasan Aplikasi Dana di Google Play Store Menggunakan Algoritma Naïve Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(6), 3314–3319. <https://doi.org/10.36040/jati.v7i6.8178>
- [7] Rahmadani, R., Rahim, A., & Rudiman, R. (2024). Analisis Sentimen Ulasan “Ojol The Game” Di Google Play Store Menggunakan Algoritma Naive Bayes Dan Model Ekstraksi Fitur Tf-Idf Untuk Meningkatkan Kualitas Game. *Jurnal Informatika dan Teknik Elektro Terapan*, 12(3).
- [8] Asri, Y. S. T., Kuswardani, D., & Kom, M. (2024). *Machine Learning & Deep Learning: Analisis Sentimen Menggunakan Ulasan Pengguna Aplikasi*. Uwais Inspirasi Indonesia.\