

Penerapan Algoritma *Naive Bayes* pada Analisis Sentimen Aplikasi Traveloka pada Platform Playstore

Eka Ardiya Putri*, Berlilana

Fakultas Ilmu Komputer, Program Studi Informatika, Universitas Amikom Purwokerto, Indonesia

Email: ¹putriardiya8@gmail.com, ²berli_amikom@yahoo.co.id

Email Penulis Korespondensi: putriardiya8@gmail.com

Submitted: 25/10/2024; Accepted: 01/12/2024; Published: 03/12/2024

Abstrak—Angka pengguna internet yang ada di negara Indonesia kian meningkat dalam setiap tahunnya, menjadikannya negara dengan kenaikan jumlah pemakai internet tercepat di dunia, di urutan berikutnya setelah Tiongkok, India, serta Amerika Serikat. Pada tahun 2017, di Indonesia pada sektor ekonomi digital menghadirkan dampak yang cukup tinggi terhadap PDB, menunjukkan angka 7,3%, sementara perkembangan ekonomi total hanya mencapai 5,1%. Traveloka muncul pada tahun 2012 dan telah berkembang pesat tergolong sebagai aplikasi travel paling unggul di Asia Tenggara. Sebagaimana yang diterapkan oleh aplikasi Traveloka mengaplikasikan data scraping guna mengumpulkan 5000 data ulasan dari platform yang dituju. Dengan meningkatnya ulasan pengguna aplikasi Traveloka di Playstore, tantangan utamanya adalah mengklasifikasikan sentimen ulasan secara otomatis dan akurat. Tujuan dari penelitian ini adalah untuk mengetahui sejauh mana penilaian pengguna terhadap aplikasi Traveloka. Hasil penelitian menunjukan model tersebut memiliki *Accuracy* sebesar 0.91, mengindikasikan bahwa 91% dari total data berhasil diprediksi dengan benar. *F1 Score* model yang mencapai 0.90 mencerminkan keseimbangan optimal antara *Precision* dan *Recall*, menunjukkan bahwa model tidak hanya tepat dalam memprediksi hasil positif, tetapi juga mampu menangkap hampir semua contoh positif. *Precision* sebesar 0.92 menunjukkan tingkat akurasi tinggi dalam prediksi positif, sedangkan *Recall* yang mencapai 0.88 menunjukkan bahwa kemampuan model yang tergolong sangat baik dalam mendeteksi semua data positif yang ada. Dalam analisis ini, dari 940 data yang digunakan, ditemukan 250 *True Positive* (TP), 18 *False Positive* (FP), 608 *True Negative* (TN) dan 64 *False Negative* (FN), dengan pembagian data 80:20. Temuan ini menunjukkan bahwa model dapat memprediksi sebagian besar data dengan akurat, meskipun terdapat beberapa kesalahan dalam klasifikasi positif dan negatif. Hasil ini mengindikasikan bahwa model memiliki efektivitas tinggi dalam identifikasi dan prediksi data positif, memberikan dasar yang kuat untuk aplikasi lebih lanjut dalam analisis data.

Kata Kunci: Analisis Sentimen; Naive Bayes; Playstore; Traveloka; Ulasan Pengguna

Abstract—The number of internet users in Indonesia is increasing every year, making it the fastest-growing country in the world, next only to China, India and the United States. In 2017, in Indonesia, the digital economy sector had a high impact on GDP, showing a figure of 7.3%, while the total economic development only reached 5.1%. Traveloka appeared in 2012 and has grown rapidly to be classified as the most superior travel application in Southeast Asia. As applied by the Traveloka application, it applies data scraping to collect 5000 review data from the intended platform. With the increase of Traveloka app user reviews on Playstore, the main challenge is to classify the sentiment of the reviews automatically and accurately. The purpose of this research is to find out the extent of user assessment of the Traveloka application. The results show that the model has an *Accuracy* of 0.91, indicating that 91% of the total data was predicted correctly. The model's *F1 Score* of 0.90 reflects the optimal balance between *Precision* and *Recall*, indicating that the model is not only correct in predicting positive results, but also able to capture almost all positive examples. *Precision* of 0.92 indicates a high level of accuracy in positive predictions, while *Recall* of 0.88 indicates that the model's ability to detect all positive data is very good. In this analysis, out of the 940 data used, 250 *True Positive* (TP), 18 *False Positive* (FP), 608 *True Negative* (TN) and 64 *False Negative* (FN) were found, with an 80:20 data split. The findings show that the model can predict most of the data accurately, despite some errors in positive and negative classification. These results indicate that the model has high effectiveness in the identification and prediction of positive data, providing a strong basis for further applications in data analysis.

Keywords: Sentiment Analysis; Naive Bayes; Playstore; Traveloka; User Reviews

1. PENDAHULUAN

Dengan semakin berkembangnya teknologi informasi, Aplikasi mobile sudah berubah menjadi alat yang sangat dibutuhkan dalam aktivitas sehari-hari. Aplikasi mobile kini tidak hanya dimanfaatkan untuk berkomunikasi, tetapi juga untuk berbagai keperluan seperti belanja, hiburan, pendidikan, dan perjalanan satu di antaranya aplikasi yang terkenal di Indonesia yaitu Traveloka, aplikasi tersebut menawarkan beragam layanan. mulai dari pemesanan tiket pesawat, hotel, hingga aktivitas wisata. Traveloka mempermudah pengguna untuk mengatur perjalanan mereka hanya melalui satu platform, yang mengintegrasikan berbagai layanan dalam satu aplikasi[1]. Kepraktisan dan kemudahan yang ditawarkan oleh Traveloka membuatnya menjadi opsi utama untuk banyak pengguna di Indonesia.

Traveloka di aplikasikan pada tahun 2012 dan sejak itu mulai mengalami pertumbuhan yang pesat menjadi salah satu aplikasi travel terbesar di Asia Tenggara[2]. Aplikasi ini memberikan kesempatan kepada pengguna untuk menemukan dan memesan tiket pesawat serta tempat menginap, tiket kereta, dan bahkan tiket atraksi dan aktivitas. Dengan berbagai fitur canggih seperti integrasi dengan berbagai metode pembayaran, layanan pelanggan yang responsif, serta berbagai promo menarik, Traveloka berhasil menarik jutaan pengguna.

Namun, dengan meningkatnya jumlah pengguna, berbagai ulasan dan feedback juga semakin banyak diterima. Ulasan-ulasan ini dapat berupa kritik, saran, dan apresiasi yang sangat berharga untuk pengembang aplikasi agar

memahami kebutuhan dan keinginan pengguna. Dalam upaya meningkatkan kualitas pelayanan, analisis ulasan pengguna menjadi salah satu cara efektif untuk memahami kepuasan dan kebutuhan pengguna.

Playstore adalah sebuah platform yang telah digunakan secara luas oleh banyak pengguna[3]. Sebagai hasil dari penggunaannya yang meluas, banyak orang memberikan umpan balik berupa ulasan dan penilaian tentang aplikasi yang tersedia. Umpan balik ini mencakup berbagai aspek, mulai dari pengalaman pengguna hingga kinerja aplikasi, yang sangat berharga untuk pengembang aplikasi dalam meningkatkan kualitas produk mereka. Dengan banyaknya ulasan yang ada, ini bisa menjadi bahan untuk evaluasi Traveloka untuk menjadi lebih baik dengan cara melakukan analisis sentimen.

Analisis sentimen merupakan suatu proses yang memanfaatkan pemrosesan bahasa alami (*Natural Language Processing, NLP*) dan analisis teks guna mengenali serta mengekstrak informasi subjektif dari teks[4]. Dalam konteks ini, penilaian sentimen pada ulasan pengguna aplikasi Traveloka di *Playstore* dapat memberikan informasi yang penting untuk pengembang aplikasi untuk melakukan perbaikan serta peningkatan layanan. Ulasan pengguna di *Playstore* biasanya mengandung opini yang mencerminkan pengalaman nyata pengguna terhadap aplikasi, sehingga analisis terhadap ulasan tersebut dapat memberikan paparan yang rinci jelas mengenai keunggulan dan kelemahan aplikasi.

Pendekatan *Naïve Bayes* merupakan metode yang biasa dipakai pada analisis sentimen karena kesederhanaan dan efektivitasnya. Metode ini berlandaskan *Teorema Bayes* dengan anggapan keterhubungan diantara fitur-fitur yang tersedia. Metode *Naïve Bayes* mengklasifikasikan teks berdasarkan probabilitas, yang dihitung dari frekuensi kata-kata dalam kelompok (positif dan negatif). Keunggulan metode ini adalah memiliki kemampuan dalam menangani data dengan dimensi tinggi dan performa yang baik meskipun dengan data latih yang terbatas [5].

Di samping itu, analisis sentimen dengan *Naïve Bayes* dapat memberikan bantuan dalam menentukan tren sentimen dari waktu ke waktu. Misalnya, perubahan dalam kebijakan perusahaan atau fitur baru dalam aplikasi dapat tercermin dalam perubahan sentimen ulasan pengguna. Dengan memahami tren ini, pengembang aplikasi dapat menghasilkan keputusan yang lebih optimal dan berdasarkan data yang akurat dalam perencanaan dan pengembangan produk.

Penelitian sebelumnya menggambarkan wawasan berharga berkaitan dengan penggunaan algoritma *Naïve Bayes* dalam penelitian tentang analisis sentimen. Sebagai contoh, penelitian terdahulu yang dilakukan oleh Fenny Novianti dan Kiky Rizky Nova Wardani pada tahun 2023, memberikan gambaran mendalam tentang aplikasi algoritma ini. Penelitian tersebut menggunakan pembobotan *TF-IDF* untuk mengukur relevansi kata pada ulasan dan membagi data menjadi data latih dan data uji dengan perbandingan rasio 80:20. Hasil analisis menunjukkan akurasi yang di hasilkan algoritma *Naïve Bayes* sebesar 75%, *Recall* 67%, *Precision* 86%, dan *F1-score* 76%. Dari hasil didapatkan menandakan keberhasilan metode *Naïve Bayes* dalam klasifikasi sentimen, namun juga menggarisbawahi adanya ruang untuk perbaikan dalam hal *Recall* yang lebih rendah. Penelitian ini menginspirasi penelitian ini untuk mengeksplorasi penerapan yang sama dalam konteks ulasan aplikasi Traveloka di *Playstore*, dengan tujuan menyediakan wawasan yang lebih mendalam tentang sentimen pengguna dan faktor-faktor yang mempengaruhinya[6].

Penelitian yang dilakukan oleh Deni Wijaya, Rizki Adi Saputra, dan Faldy Irwiensyah, pada penelitian ini berfokus pada penerapan algoritma *Naïve Bayes* guna mengklasifikasikan ulasan pengguna yang berfokus pada ulasan melalui platform *Playstore*. Dengan total data 2000 komentar, evaluasi menggambarkan model *Naïve Bayes* mampu mencapai tingkat akurasi sebesar 63,61%, dengan 307 *True Positive*, 74 *True Negative*, 26 *False Positive*, dan 192 *Precision* model tercatat sebesar 92,19%, sementara *Recall* berada pada angka 61,52%. Penelitian ini menggambarkan performa pada algoritma *Naïve Bayes* yang efisien pada saat memprediksi sentimen positif, namun masih terdapat ruang untuk perbaikan pada prediksi sentimen negative [7].

Penelitian selanjutnya yang dilaksanakan oleh Irni Di Estetika, Irfan Darmawan, dan Oktariani Nurul Pratiwi, menunjukkan hasil yang signifikan dalam menganalisis sentimen ulasan pengguna. Algoritma *Naïve Bayes* yang diterapkan mencatat akurasi yang didapatkan sebesar 83%, dengan angka *Precision* 82%, *Recall* 80,33%, dan *F1-score* 80,66%. Hasil dari penelitian tersebut mengungkapkan bahwa ulasan dengan sentimen positif lebih dominan dibandingkan dengan ulasan negatif, memberikan wawasan yang berharga bagi pengembang dalam meningkatkan layanan platform [8].

Penelitian oleh Nova Rosalina Siahaan dan rekan-rekan, menunjukkan efektivitas metode *Naïve Bayes Classifier* dalam melakukan analisis sentimen ulasan aplikasi WhatsApp. Penelitian ini mengungkapkan, ulasan produk diklasifikasikan ke dalam tiga kategori, yaitu positif, negatif, dan netral. Proses pelabelan dilakukan pada dataset dengan kriteria tertentu: rating di bawah 3 dikategorikan sebagai sentimen negatif, rating sama dengan 3 sebagai sentimen netral, dan rating di atas 3 sebagai sentimen positif. Penelitian ini juga menambahkan kolom 'Label' yang mencakup kategori-kategori sentimen tersebut. Dengan demikian, penelitian tersebut menyediakan uraian yang jelas mengenai sentimen pengguna terhadap aplikasi WhatsApp [9].

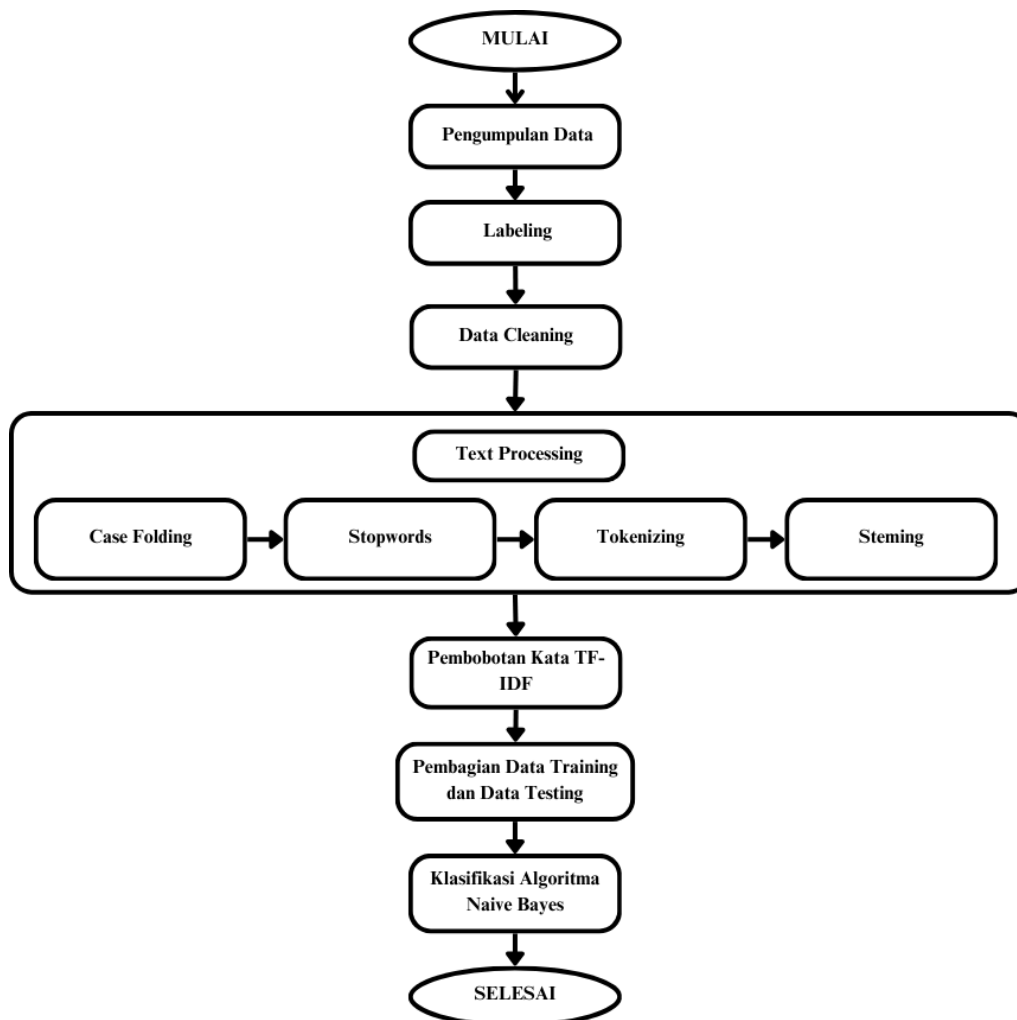
Meskipun analisis sentimen telah banyak dilakukan pada berbagai aplikasi dan platform, terdapat satu kesenjangan utama dalam penelitian ini. Sebagian besar penelitian sebelumnya lebih banyak fokus pada Twitter dan Facebook, sementara analisis ulasan pengguna di *Playstore* masih relatif kurang dieksplorasi. Padahal, ulasan di *Playstore* memberikan informasi langsung dari pengguna aplikasi yang memiliki pengalaman nyata dengan aplikasi tersebut. Studi ini bertujuan untuk menjembatani kesenjangan tersebut melalui analisis sentimen dari ulasan pengguna aplikasi Traveloka di *Playstore* menggunakan algoritma *Naïve Bayes*.

Melalui studi ini, diharapkan dapat menyajikan pemahaman yang mendalam tentang sentimen positif dan negatif yang dominan serta faktor-faktor yang berdampak pada kepuasan pengguna. Informasi tersebut akan sangat berarti bagi pengembang aplikasi Traveloka guna mengembangkan kualitas pada layanan serta memenuhi harapan pengguna dengan lebih baik.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Pada penelitian ini menganalisis sentimen untuk mengolah data ulasan pengguna aplikasi Traveloka. Proses penelitian mencakup beberapa tahapan, mulai dari pengumpulan data, pra-pemrosesan, hingga penerapan algoritma klasifikasi untuk menganalisis sentimen ulasan. Algoritma utama yang digunakan dalam penelitian ini adalah Naïve Bayes, yang telah terbukti efektif dalam klasifikasi teks. Tahapan penelitian yang digunakan ditunjukkan dalam Gambar 1 berikut.



Gambar 1. Tahapan Penelitian

2.2 Pengumpulan Data

Proses pengumpulan data diterapkan dengan memanfaatkan Google Play Scraper, sebuah alat yang dirancang untuk mengunduh data ulasan dari platform Playstore. Dalam penelitian ini, sebanyak 5.000 ulasan berhasil dikumpulkan, dengan perhatian khusus pada ulasan yang ditulis dalam bahasa Indonesia. Fokus pada ulasan berbahasa Indonesia dilakukan untuk memastikan bahwa data yang diolah sesuai dengan konteks wilayah pengguna dan lebih relevan dalam analisis sentimen. Dengan demikian, pengumpulan data ini membantu dalam memperoleh informasi yang lebih akurat dan tepat sasaran untuk penelitian yang berhubungan dengan perilaku dan pengalaman pengguna aplikasi di Indonesia.

2.3. Labeling

Labeling adalah proses memberi label atau tanda pada data sehingga dapat digunakan untuk tujuan analisis, pelatihan model, atau aplikasi lain dalam bidang ilmu data dan pembelajaran mesin [10]. Labeling sangat penting dalam berbagai

konteks, terutama dalam pembelajaran mesin, di mana data berlabel di gunakan pada saat melatih model agar dapat membuat prediksi sentimen atau keputusan sentiment yang tepat.

2.4 Data Cleaning

Data cleaning merupakan sebuah proses identifikasi serta perbaikan atau penghapusan data yang korup, tidak akurat, atau tidak lengkap dari dataset [11]. Langkah-langkah dalam data cleaning meliputi penghapusan data duplikat, penanganan data yang hilang dengan imputasi atau penghapusan, koreksi kesalahan data seperti kesalahan pengetikan, penghapusan data yang tidak relevan, standarisasi format data, konversi tipe data yang sesuai, validasi aturan data, dan deteksi serta penanganan outlier [12]. Proses ini penting untuk memastikan kualitas, efisiensi, dan akurasi data dalam analisis, serta mendukung pengambilan keputusan yang lebih baik berbasis data yang valid dan dapat diandalkan.

2.5 Text Processing

Text processing adalah proses mengolah teks tidak terstruktur untuk analisis lebih lanjut [13]. Langkah-langkahnya diantaranya adalah tokenisasi (mengubah kata menjadi lebih kecil / unit-unit yang lebih kecil), sedangkan normalisasi (mengubah teks ke format konsisten), penghapusan stopwords (kata yang tidak penting), stemming dan lemmatization (mengubah kata ke bentuk dasar), serta ekstraksi fitur (mengubah teks menjadi representasi numerik seperti *TF-IDF*). Proses ini penting untuk aplikasi seperti analisis sentimen, ekstraksi informasi, dan pemodelan bahasa, memungkinkan komputer memahami dan mengolah teks manusia secara efektif.

a. Case Folding

Case folding merupakan salah satu proses yang memiliki tujuan guna mengganti semua karakter huruf ke bentuk huruf kecil (lowercase) [14]. Proses ini penting karena membantu dalam konsistensi data, memungkinkan analisis teks yang lebih akurat dengan menghindari perbedaan yang disebabkan oleh huruf besar dan kecil. Misalnya, kata "Traveloka", "traveloka", dan "TRAVELOKA" akan diperlakukan sama setelah case folding. Ini sangat berguna dalam aplikasi seperti pencarian teks dan analisis sentimen, di mana perbedaan huruf besar dan kecil tidak relevan terhadap makna kata.

b. Stopwords Removal

Stopwords removal adalah sebuah proses mengeliminasi kata-kata umum pada saat pre-processing yang kerap muncul dalam tulisan namun memiliki sedikit atau tidak ada nilai informatif untuk analisis lebih lanjut [15]. Kata-kata ini mencakup antara lain "dan," "atau," "adalah," dan "itu" dalam bahasa Indonesia, atau "and," "or," "is," dan "the" dalam bahasa Inggris. Stopwords berperan dalam mengurangi dimensi data serta memusatkan perhatian pada kata-kata penting. dan memiliki makna dalam yang terdapat pada teks. Dalam analisis teks seperti pemrosesan bahasa alami (*NLP*) dan penambangan teks, stopwords removal meningkatkan efisiensi dan akurasi algoritma dengan mengurangi kebisingan dalam data.

c. Tokenizing

Tokenizing adalah proses dalam text processing yang membagi teks dipecah menjadi unit-unit lebih kecil, seperti kata, frasa, atau kalimat, yang disebut token [16]. Proses ini penting untuk analisis teks karena memungkinkan pemrosesan lebih lanjut pada level yang lebih detail. Token dapat berupa kata dalam sebuah kalimat, misalnya, kalimat "Saya suka belajar pemrograman" akan dipecah menjadi token: ["Saya", "suka", "belajar", "pemrograman"]. Tokenizing memungkinkan aplikasi *NLP* untuk bekerja dengan data teks, seperti analisis sentimen, ekstraksi informasi, dan pemodelan bahasa, dengan lebih efektif dan akurat.

d. Stemming

Stemming adalah proses dalam text processing yang mengubah kata-kata ke bentuk aslinya atau dasarnya dengan menghilangkan akhiran atau prefix [17]. Tujuannya ialah menyederhanakan variasi kata yang memiliki makna sama menjadi bentuk dasar yang konsisten. misalnya, kata-kata seperti "belajar," "belajarannya," dan "belajarnya" dapat dipotong menjadi bentuk dasar "belajar." Metode ini membantu dalam analisis teks dengan mengurangi dimensi data dan meningkatkan relevansi hasil analisis, seperti dalam pencarian atau analisis sentimen, karena variasi kata yang berbeda diperlakukan sebagai satu entitas yang sama.

2.6 Pembobotan *TF-IDF*

Term Frequency-Inverse Document Frequency (TF-IDF) merupakan sistem pembobotan yang memberikan bobot pada setiap kata dalam dokumen berdasarkan frekuensi term (*tf*) dan frekuensi resiprokal dokumen (*tf-idf*) [18].

a. *Term Frequency (TF)* menghitung banyaknya kemunculan suatu kata yang ada pada dokumen menunjukkan seberapa sering kata tersebut muncul. Angka kemunculan kata itu tinggi, maka semakin tinggi nilai *Term Frequency (TF)* yang didapatkan

b. *Inverse Document Frequency (IDF)* merupakan ukuran yang mengilustrasikan jarangness suatu kata muncul di dokumen. Semakin memiliki jumlah besar pada kata-kata yang kerap terlihat di banyak dokumen maka akan memiliki nilai *IDF* yang rendah, akan tetapi kata-kata yang jarang muncul akan memiliki angka *IDF* cukup tinggi. Hasil dari perkalian *TF* dan *IDF* memberikan skor *TF-IDF* pada setiap kata mencerminkan signifikansi kata tersebut dalam konteks dokumen spesifik dibandingkan dengan seluruh koleksi dokumen. Teknik ini membantu



dalam ekstraksi fitur dan meningkatkan relevansi dalam aplikasi seperti pencarian informasi dan analisis teks dengan menghasilkan bobot yang cukup tinggi pada kata-kata yang unik dan relevan dalam konteks dokumen yang tertuju.

$$W_{ij} = \frac{n_{ij}}{\sum_{j=1}^p n_{j,i}} \log 2 \frac{D}{d_j} \tag{1}$$

Rumus (1) diaplikasikan guna menghitung pada bobot *TF-IDF* dari istilah *jjj* yang ada dalam dokumen *iii*, dengan bobot ini mencakup 2 komponen utama yang saling terkait, yaitu *Term Frequency (TF)* dan *Inverse Document Frequency (IDF)*. *Term Frequency (TF)* mengukur seberapa sering istilah tersebut muncul dalam dokumen tertentu, sementara *Inverse Document Frequency (IDF)* menilai seberapa jarang istilah tersebut muncul di seluruh koleksi dokumen. Dengan menggabungkan kedua komponen ini, bobot *TF-IDF* dapat memberikan gambaran yang lebih jelas tentang relevansi istilah *jjj* dalam konteks dokumen *iii*, sehingga membantu dalam analisis dan pemrosesan teks.

2.7 Pembagian Data Train dan Testing

Pembagian ini merujuk pada proses membagi dataset menjadi dua bagian yang berbeda yaitu data train dan data test untuk tujuan pelatihan dan evaluasi model *machine learning*[19]. Tujuan utama dari pembagian ini adalah guna memvalidasi model yang dikembangkan mampu menggeneralisasi dengan baik dalam data yang sebelumnya dilihat

- a. Data Train (80%): Sebagian besar data diterapkan guna melatih model. Dengan penjelasan tersebut berarti model akan "belajar" dari data tersebut, menyesuaikan parameter-parameter untuk memprediksi hasil atau klasifikasi dengan lebih baik.
- b. Data Test (20%): Sisanya di aplikasikan untuk menguji model yang sudah dilatih sebelumnya. *Data test* tidak digunakan selama proses pelatihan dan memberikan gambaran objektif tentang seberapa baik model mampu menghasilkan prediksi atau klasifikasi data yang tidak dilihat sebelumnya.

Pembagian data ini membantu dalam membuktikan model tidak hanya menghafal data pelatihan (*overfitting*), tetapi juga mampu menangani data yang belum pernah terlihat sebelumnya secara efektif.

2.8 Klasifikasi Algoritma Naïve Bayes

Naïve Bayes merupakan sebuah algoritma klasifikasi yang menggunakan pendekatan probabilitas, yang didasarkan dalam *Teorema Bayes* dan mengasumsikan bahwa setiap fitur saling independent [20]. Walaupun asumsi independensi ini biasanya tidak berlaku akurat dalam praktik, *Naïve Bayes* tetap menjadi metode yang efektif serta efisien untuk banyak aplikasi. Berikut adalah penjelasan lebih mendetail tentang algoritma ini:

$$P(H|X) = \frac{P(H|X)P(H)}{P(X)} \tag{2}$$

Teorema Bayes memperbarui probabilitas hipotesis *H* dengan mempertimbangkan bukti baru *X*. Probabilitas awal *P(H)* merepresentasikan keyakinan awal sebelum adanya bukti tambahan, sedangkan probabilitas *P(X|H)* menggambarkan kemungkinan terjadinya bukti *X* jika hipotesis *H* benar. Probabilitas total *P(X)* berfungsi untuk menormalkan hasil, sehingga probabilitas akhir tetap valid. Teori ini memungkinkan kita untuk memperbarui keyakinan terhadap hipotesis *H* berdasarkan bukti baru *X*.

3. HASIL DAN PEMBAHASAN

Tahapan berikutnya akan mengevaluasi hasil dari analisis serta tujuan penelitian yang berfokus pada pemahaman perspektif umum pengguna mengenai aplikasi Traveloka pada *Play Store*, yaitu dengan cara menerapkan/mengimplementasikan algoritma *Naïve Bayes*. Studi ini bertujuan guna mengeksplorasi bagaimana pengguna aplikasi Traveloka menilai serta memberikan tanggapan terhadap pengalaman mereka dalam menggunakan aplikasi tersebut, terkhusus dalam menetapkan apakah ulasan itu positif atau negatif.

3.1 Pengumpulan Data

Dalam tahapan pengumpulan data, *Google Play Scraper* diaplikasikan agar mencari data secara otomatis serta mengumpulkan informasi dari *Google Play Store*. Metode ini efektif untuk memperoleh detail aplikasi, ulasan pengguna, penilaian, dan jumlah unduhan. Sebanyak 5000 ulasan berhasil dikumpulkan, yang mendukung pembuatan keputusan yang lebih baik serta strategis. Hasil yang didapatkan pada tahap pengumpulan data ditampilkan pada Tabel 1.

Tabel 1. Hasil dari Pengumpulan Data

	username	at	content
0	Pengguna Google	2024-07-11 12:48:29	mau pesen hotel susah kalo bisanya cuma bayar ...
1	Pengguna Google	2024-07-25 06:57:13	Saya pengguna traveloka semenjak aplikasi ini ...
2	Pengguna Google	2024-07-06 02:29:34	Saya suka heran sama algoritma harga di travel...



3	Pengguna Google	2024-06-05 21:05:46	Tidak bisa scroll saat menambah jumlah penumpang...
4	Pengguna Google	2024-07-24 05:43:52	Pembelian tiket by Aplikasi Traveloka terjadi ...

Tabel 1 menunjukkan hasil pengumpulan data ulasan pengguna aplikasi Traveloka di platform Playstore, mencakup jumlah ulasan, skor rating, dan kategori sentimen (positif atau negatif). Data ini memberikan gambaran awal mengenai kepuasan atau ketidakpuasan pengguna dan menjadi dasar bagi analisis sentimen lebih lanjut. Tabel ini membantu memastikan bahwa dataset yang digunakan cukup representatif dan mendukung kualitas hasil penelitian.

3.2 Labeling

Pada tahap selanjutnya, yaitu labeling, dilakukan pemberian label positif dan negatif pada data berdasarkan kriteria tertentu. Proses ini bertujuan untuk mengklasifikasikan data sesuai dengan sentimen yang terkandung dalam teks, sehingga analisis sentimen dapat dilakukan dengan lebih akurat. Seperti yang di tampilkan dalam tabel 2 berikut ini.

Tabel 2. Memberi Label Positif Dan Negatif

	content	score	Label
75	proses cepat, cuma kuponnya buat yg bekerja sa...	5	Positif
106	Traveloka mantap,tiap pulang kampung naik kere...	5	Positif
13	Kecewa karena udah pesan room udah dibayar lun...	1	Negatif
290	bagus kalo lagi tanggung bulan bisa beli tiket...	5	Positif
19	Terbaik CSnya, fast respon padahal lagi hari l...	5	Positif

Pada Tabel 2 menampilkan proses pelabelan pada data ulasan pengguna, di mana setiap ulasan diklasifikasikan ke dalam kategori sentimen positif atau negatif. Proses pelabelan ini bertujuan untuk mempermudah analisis sentimen dengan mengidentifikasi ekspresi kepuasan atau ketidakpuasan pengguna terhadap aplikasi Traveloka. Dengan adanya label sentimen ini, penelitian dapat mengukur proporsi ulasan positif dan negatif secara akurat, yang akan menjadi dasar dalam evaluasi kinerja aplikasi dan pemahaman pola sentimen pengguna.

3.3 Data Cleaning

Tahap pembersihan data dilakukan untuk mengatasi ketidakkonsistenan, kesalahan, dan data yang tidak akurat dalam dataset, mencakup identifikasi dan perbaikan anomali, penghapusan duplikasi, serta penanganan nilai yang hilang atau tidak valid, sehingga dataset yang digunakan dalam analisis atau pemodelan menjadi lebih valid dan bebas dari gangguan yang dapat memengaruhi kualitas serta keandalan hasil penelitian. Seperti yang di tampilkan pada Tabel 3.

Tabel 3. Tahap Data Cleaning

	content	score	Label
0	proses cepat, cuma kuponnya buat yg bekerja sa...	5	Positif
1	Traveloka mantap,tiap pulang kampung naik kere...	5	Positif
2	Kecewa karena udah pesan room udah dibayar lun...	1	Negatif
3	bagus kalo lagi tanggung bulan bisa beli tiket...	5	Positif
4	Terbaik CSnya, fast respon padahal lagi hari l...	5	Positif

Tabel 3 menunjukkan tahap pembersihan data (data cleaning) yang dilakukan untuk memastikan kualitas dataset sebelum analisis lebih lanjut. Pada tahap ini, dilakukan identifikasi dan penghapusan data yang tidak konsisten, duplikat, atau tidak relevan, serta penanganan terhadap nilai yang hilang atau tidak valid.

3.4 Text Processing

Proses text processing dimulai dengan case folding, yaitu mengubah semua teks menjadi huruf kecil untuk konsistensi. Kemudian, dilakukan penghapusan stopwords untuk menghilangkan kata-kata umum yang tidak memiliki makna signifikan, seperti "dan," "yang," atau "atau." Setelah itu, teks dipecah menjadi unit kata atau token melalui proses tokenisasi, dan setiap kata diubah menjadi bentuk dasarnya melalui stemming, seperti "berjalan" menjadi "jalan." Terakhir, teks direpresentasikan dalam TF-IDF, sehingga siap digunakan dalam model analisis atau pemodelan lebih lanjut.

3.4.1 Case Folding

Alur case folding dimulai dengan mengubah seluruh teks menjadi huruf kecil untuk menyamakan format dan mengurangi perbedaan yang tidak perlu akibat kapitalisasi. Pada tahap ini, setiap kata dalam teks yang mengandung huruf besar, seperti "Traveloka" atau "Pengguna," diubah menjadi huruf kecil, menjadi "traveloka" dan "pengguna." Hasil dari perubahan menjadi Case Folding pada Tabel 4

Tabel 4. Case Folding

Label	text_clean
0	Positif proses cepat cuma kuponnya buat yg bekerja sam...

1	Positif	traveloka mantaptiap pulang kampung naik keret...
2	Negatif	kecewa karena udah pesan room udah dibayar lun...
3	Positif	bagus kalo lagi tanggung bulan bisa beli tiket...
4	Positif	terbaik csnya fast respon padahal lagi hari li...

Tabel 4 menunjukkan tahap case folding proses ini dilakukan untuk menghilangkan perbedaan kapitalisasi yang tidak relevan dalam analisis, seperti “Traveloka” dan “traveloka” yang dapat diinterpretasikan sebagai dua entitas berbeda jika tidak distandarkan.

3.4.2 Stopwords Removal

Langkah awal dalam proses ini adalah mengidentifikasi stopwords, yaitu kata-kata umum antara lain, "atau", "saya", "dan" dan "kamu" yang kerap muncul dalam teks. Stopwords ini umumnya tidak memberikan informasi tambahan yang penting bagi analisis. Pada Tabel 5 adalah hasil dari Stopwords Removal.

Tabel 5. Stopwords Removal

	text_clean	text_StopWord
0	proses cepat cuma kuponnya buat yg bekerja sam...	proses cepat kuponnya yg gak perusahaan yg
1	traveloka mantaptiap pulang kampung naik keret...	traveloka mantaptiap pulang kampung kereta pes...
2	kecewa karena udah pesan room udah dibayar lun...	kecewa udah pesan room udah dibayar lunas date...
3	bagus kalo lagi tanggung bulan bisa beli tiket...	bagus kalo tanggung beli tiket aja rekomendasikan
4	terbaik csnya fast respon padahal lagi hari li...	terbaik csnya fast respon libur pengajuan uang...

Tabel 5 menunjukkan tahap penghapusan stopwords dari data ulasan pengguna. Pada tahap ini, kata-kata umum seperti "dan," "yang," dan "atau" dihilangkan dari teks ulasan untuk meningkatkan fokus pada kata-kata yang memiliki makna signifikan bagi analisis sentimen. Secara teknis, penghapusan dilakukan dengan memanfaatkan daftar stopwords, menggunakan pustaka seperti `NLTK` atau `Sastrawi` dalam bahasa pemrograman *Python* untuk proses otomatis. Tahap ini membantu mengurangi "noise" dalam data dan meningkatkan efisiensi serta relevansi fitur yang diekstraksi untuk analisis.

3.4.3 Tokenizing

Pada tahap ini, library *NLTK* digunakan untuk menjalankan proses tokenisasi. Proses ini bertujuan agar hasil normalisasi dapat diubah menjadi token, seperti yang tertera/terkandung pada Tabel 6.

Tabel 6. Tokenizing

	text_StopWord	text_tokens
0	proses cepat kuponnya yg gak perusahaan yg	[proses, cepat, kuponnya, yg, gak, perusahaan,...
1	traveloka mantaptiap pulang kampung kereta pes...	[traveloka, mantaptiap, pulang, kampung, keret...
2	kecewa udah pesan room udah dibayar lunas date...	[kecewa, udah, pesan, room, udah, dibayar, lun...
3	bagus kalo tanggung beli tiket aja rekomendasikan	[bagus, kalo, tanggung, beli, tiket, aja, reko...
4	terbaik csnya fast respon libur pengajuan uang...	[terbaik, csnya, fast, respon, libur, pengajua...

Tabel 6 menunjukkan tahap tokenisasi, yaitu proses membagi teks ulasan pengguna menjadi kata-kata atau token individu. Proses ini dilakukan untuk memudahkan analisis setiap kata secara terpisah. Secara teknis, tokenisasi dilakukan dengan memecah teks berdasarkan spasi atau tanda baca menggunakan pustaka seperti *NLTK* di *Python*. Tahap tokenisasi penting untuk menyiapkan data teks dalam bentuk yang dapat dianalisis lebih lanjut, seperti dalam proses ekstraksi fitur atau analisis frekuensi kata.

3.4.4 Stemming

Proses ini membantu menyamakan kata-kata yang bermakna sama akan tetapi bentuk yang berbeda, sehingga analisis teks menjadi lebih konsisten dan efisien. Dengan menggunakan stemming, kita dapat mengurangi kompleksitas data teks dan fokus pada bentuk dasar kata untuk tujuan pemrosesan lebih lanjut..berikut pada tabel 7 merupakan hasil dari Stemming.

Tabel 7. Stemming

	text_tokens	text_steamindo
0	[proses, cepat, kuponnya, yg, gak, perusahaan,...	proses cepat kupon yg gak usaha yg

1	[traveloka, mantaptiap, pulang, kampung, traveloka mantaptiap pulang kampung kereta sen... keret...]	
2	[kecewa, udah, pesan, room, udah, dibayar, lun...]	kecewa udah pesan room udah bayar lunas dateng...
3	[bagus, kalo, tanggung, beli, tiket, aja, reko...]	bagus kalo tanggung beli tiket aja rekomendasi
4	[terbaik, csnya, fast, respon, libur, pengajua...]	baik csnya fast respon libur aju uang refund a...

Tabel 7 menunjukkan tahap stemming, proses ini dilakukan secara teknis menggunakan pustaka seperti `Sastrawi` di Python, yang secara otomatis mengidentifikasi dan memotong akhiran atau awalan kata. Tahap stemming bertujuan untuk menyederhanakan kata-kata sehingga kata yang memiliki akar yang sama dapat dikelompokkan, membantu meningkatkan akurasi analisis dalam mendeteksi pola sentimen.

3.4 Pembobotan Kata TF-IDF

Proses *Term Frequency-Inverse Document Frequency* (TF-IDF) dimulai dengan menghitung *Term Frequency* (TF), yaitu frekuensi kemunculan setiap kata dalam dokumen. Lalu, dihitung *Inverse Document Frequency* (IDF) untuk menilai keunikan kata tersebut di seluruh dokumen. Hasil perkalian TF dan IDF menghasilkan skor TF-IDF, yang menunjukkan pentingnya kata dalam dokumen dan dataset secara keseluruhan. Matriks TF-IDF ini kemudian siap digunakan untuk analisis lebih lanjut, seperti klasifikasi atau klusterisasi. Gambar 2 merupakan hasil dari TF-IDF dari penelitian ini

```
[0.00060616 0.00058999 0.00051421 ... 0.00052045 0.00050948 0.00046511]
[0.0005324 0.00044201 0.00050909 0.00050576 0.00048514 0.00049077
0.00043733 0.00058308 0.00049488 0.00042858 0.00058418 0.00028793
```

Gambar 2. Hasil TF-IDF

Pada gambar 2 di jelaskan Hasil TF-IDF yang disajikan dalam *array* [0.00060616 0.00058999 0.00051421 ... 0.00052045 0.00050948 0.00046511] merefleksikan bobot setiap kata dalam suatu dokumen atau kumpulan dokumen. Setiap nilai dalam *array* ini menggambarkan tingkat signifikansi suatu kata dalam konteks dokumen yang sedang dianalisis, dengan mempertimbangkan frekuensi kemunculan kata tersebut serta prevalensinya di seluruh koleksi dokumen. Analisis ini penting dalam memahami kontribusi relatif kata-kata terhadap pemahaman konten dokumen secara keseluruhan.

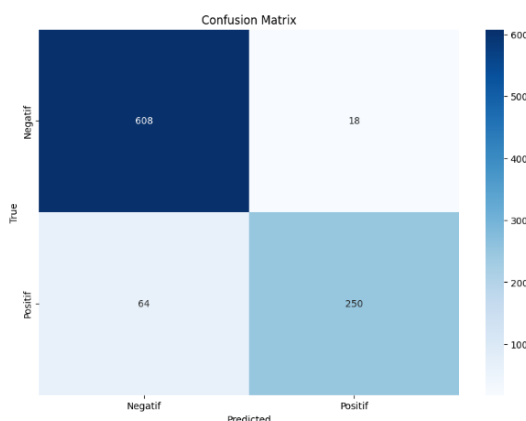
3.5 Pembagian Data Train dan Data Testing

Pada tahap selanjutnya, data dibagi menjadi data latih (training data) dan data uji (testing data) dengan rasio 80:20. Artinya, 80% dari data dipakai pada saat melatih model, sementara 20% sisanya digunakan untuk mengujinya. Pembagian ini memiliki tujuan guna melatih model pada subset data (data latih) dan kemudian mengujinya pada subset data yang terpisah (data uji) untuk mengevaluasi kinerja model. Dengan cara ini, dapat dipastikan bahwa model tidak hanya bekerja baik pada data yang sudah dilatih, tetapi juga mampu menggeneralisasi dengan baik pada data yang belum pernah dilihat sebelumnya. Pembagian data ini penting untuk menghindari *overfitting* dan mendapatkan estimasi kinerja model yang lebih akurat.

```
(3757, )
(3757, )
(940, )
(940, )
```

Gambar 3. Pembagian Data Training dan Data Testing

3.6 Klasifikasi Algoritma Naïve Bayes



Gambar 4. Confusion Matrix

Confusion Matrix pada Gambar 4 memperlihatkan bahwa mayoritas data berhasil diprediksi dengan tepat sesuai dengan nilai aktualnya. Dari total 940 data yang diolah dengan pembagian 80:20, hasil analisis menunjukkan 250 *True Positive* (TP), 18 *False Positive* (FP), 608 *True Negative* (TN) dan 64 *False Negative* (FN). Hasil evaluasi model dari *Confusion Matrix* menunjukkan bahwa terdapat 250 *True Positive* (TP), di mana data yang seharusnya positif diprediksi secara akurat sebagai positif, dan 608 *True Negative* (TN), di mana data negatif diprediksi secara akurat sebagai negatif. Namun, model juga menghasilkan 18 *False Positive* (FP), yaitu data negatif yang keliru diprediksi sebagai positif, serta 64 *False Negative* (FN), di mana data positif salah diprediksi sebagai negatif. Hasil ini memberikan gambaran mengenai performa model dalam mengklasifikasikan data secara tepat, serta mengidentifikasi kesalahan prediksi pada data positif maupun negatif.

```

MultinomialNB Accuracy: 0.9127659574468086
MultinomialNB Precision: 0.9047619047619048
MultinomialNB Recall: 0.9712460063897763
MultinomialNB f1_score: 0.9368258859784283
=====

```

	precision	recall	f1-score	support
Negatif	0.90	0.97	0.94	626
Positif	0.93	0.80	0.86	314
accuracy			0.91	940
macro avg	0.92	0.88	0.90	940
weighted avg	0.91	0.91	0.91	940

Gambar 5. Hasil Klasifikasi Algoritma *Naive Bayes*

Pada Gambar 4 ini menunjukkan metrik evaluasi model dengan hasil sebagai berikut: *Accuracy* model adalah 0.91, yang mencerminkan persentase prediksi yang benar dari total data. *F1 Score* adalah 0.90, menandakan keseimbangan antara *Precision* dan *Recall*. *Precision* mencapai 0.92, menunjukkan akurasi prediksi positif model. *Recall* adalah 0.88, menggambarkan kemampuan model dalam menangkap semua contoh positif yang ada

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat ditarik kesimpulan bahwa model menunjukkan kinerja yang sangat baik dalam evaluasi. *Accuracy* model mencapai 0.91, yang mengindikasikan bahwa 91% dari total data berhasil diprediksi dengan benar. *F1 Score*, yang bernilai 0.92, menegaskan adanya keseimbangan yang optimal antara *Precision* dan *Recall*, yang penting dalam konteks data yang mungkin tidak seimbang. *Precision* model adalah 0.92, yang menunjukkan tingkat akurasi tinggi dalam prediksi positif, yakni seberapa sering prediksi positif model benar-benar positif. Sementara itu, *Recall* yang mencapai 0.88 menunjukkan kemampuan model yang sangat baik dalam mendeteksi semua contoh positif yang ada dalam data, memperlihatkan bahwa model sangat efektif dalam menangkap data positif yang sebenarnya. Dalam analisis ini, dari total 940 data yang digunakan, ditemukan 250 *True Positive* (TP), 18 *False Positive* (FP), 608 *True Negative* (TN) dan 64 *False Negative* (FN). Pembagian data dilakukan dengan perbandingan 80:20, dimana 80% digunakan untuk pelatihan sedangkan 20% untuk pengujian model. Elemen-elemen dalam *confusion Matrix* ini menunjukkan bahwa model dapat memprediksi sebagian besar data secara akurat, meskipun terdapat beberapa kesalahan dalam klasifikasi positif dan negatif. Dengan hasil ini, model tidak hanya menunjukkan akurasi tinggi tetapi juga kemampuan yang sangat baik dalam mengidentifikasi dan memprediksi data positif secara efektif.

REFERENCES

- [1] H. Hanifah, C. Hayati, and A. Sadiqin, "Mapping out model bisnis sharing economy pada unicorn asal Indonesia," *Journal of Management and Digital Business*, vol. 4, no. 2, pp. 216–233, Jul. 2024, doi: 10.53088/jmdb.v4i2.932.
- [2] S. G. Gunawan, "PERTANGGUNGJAWABAN HUKUM TRAVELOKA SEBAGAI PELAKU USAHA DALAM FENOMENA PEMBATALAN TIKET SEPIHAK TERHADAP KONSUMEN," *Bureaucracy Journal: Indonesia Journal of Law and Social-Political Governance*, vol. 3, no. 1, 2023, doi: 10.53363/bureau.v3i1.207.
- [3] N. B. Sidauruk and N. Riza, "SENTIMEN ANALISIS DATA PENGGUNA TERHADAP KAI ACCESS Systematic Literature Review," *Jurnal Mahasiswa Teknik Informatika*, vol. 7, no. 2, pp. 1297–1303, 2023, doi: doi.org/10.36040/jati.v7i2.6764.
- [4] N. Suarna and W. Prihartono, "PENERAPAN NLP (NATURAL LANGUAGE PROCESSING) DALAM ANALISIS SENTIMEN PENGGUNA TELEGRAM DI PLAYSTORE," *Jurnal Mahasiswa Teknik Informatika*, vol. 8, no. 2, p. 1841, 2024, doi: doi.org/10.36040/jati.v8i2.8469.
- [5] A. Supian, B. Tri Revaldo, N. Marhadi, and L. Efrizoni, "Acuan Supian Perbandingan Kinerja Naïve Bayes dan SVM pada Analisis Sentimen Twitter Ibukota Nusantara," *Journal of Informatics Science*, vol. 12, no. 01, pp. 16–21, 2024, doi: 10.33884/jif.v12i01.8721.



- [6] F. Novianti and K. R. N. Wardani, “ANALISIS SENTIMEN MASYARAKAT TERHADAP DATA TWEET TRAVELOKA SELAMA RAPID TEST ANTIGEN MENGGUNAKAN ALGORITMA NAÏVE BAYES,” *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 8, no. 3, pp. 922–933, Aug. 2023, doi: 10.29100/jupi.v8i3.3973.
- [7] D. Wijaya, R. A. Saputra, and F. Irwiensyah, “KLIK: Kajian Ilmiah Informatika dan Komputer Analisis Sentimen Ulasan Aplikasi Samsat Digital Nasional Pada Google Playstore Menggunakan Algoritma Naïve Bayes,” *Media Online*, vol. 4, no. 4, 2024, doi: 10.30865/klik.v4i4.1738.
- [8] I. Darmawan and O. Nurul Pratiwi, “ANALISIS SENTIMEN ULASAN PRODUK TOKO ONLINE RUBYLICIOUS UNTUK PENINGKATAN LAYANAN MENGGUNAKAN ALGORITMA NAIVE BAYES,” 2020.
- [9] N. R. Siahaan *et al.*, “ANALISIS SENTIMEN ULASAN APLIKASI MEDIA SOSIAL WHATSAPP MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER,” *JURNAL ILMIAH BETRIK*, vol. 14, no. 02, 2023, doi: doi.org/10.36050/betrik.v14i02%20AGUSTUS.104.
- [10] A. A. Kurniawan and M. Mustikasari, “Implementasi Deep Learning Menggunakan Metode CNN dan LSTM untuk Menentukan Berita Palsu dalam Bahasa Indonesia,” vol. 5, no. 4, pp. 2622–4615, 2020, doi: 10.32493/informatika.v5i4.7760.
- [11] N. Istiqomah and F. Novika, “Sentiment Analysis Penyedia layanan Asuransi dari Media Sosial Twitter,” *Jurnal Tekno Kompak*, vol. 18, no. 1, pp. 77–89, 2024, doi: 10.33365/jtk.v18i1.3465.
- [12] A. K. Fauziyyah, “ANALISIS SENTIMEN PANDEMI COVID19 PADA STREAMING TWITTER DENGAN TEXT MINING PYTHON,” *Jurnal Ilmiah SINUS*, vol. 18, no. 2, p. 31, Jul. 2020, doi: 10.30646/sinus.v18i2.491.
- [13] Yuyun, Nurul Hidayah, and Supriadi Sahibu, “Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 4, pp. 820–826, Aug. 2021, doi: 10.29207/resti.v5i4.3146.
- [14] Z. A. Nurdiansa and B. Berlilana, “Sentiment Analysis of Reviews on Lazada Apps using Naïve Bayes Algorithm,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 8, no. 1, p. 594, Jan. 2024, doi: 10.30865/mib.v8i1.7255.
- [15] I. Habib Kusuma and N. Cahyono, “Analisis Sentimen Masyarakat Terhadap Penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor,” vol. 8, no. 3, 2023.
- [16] A. Ariansyah and U. Indahyanti, “Fitur Ekstraksi pada Pemodelan Topik Menggunakan Metode Latent Dirichlet Allocation pada Peristiwa Kebocoran Data,” *Indonesian Journal of Applied Technology*, vol. 1, no. 2, 2024, doi: 10.47134/ijat.v1i2.3041.
- [17] P. Gede Surya Cipta Nugraha and N. Wayan Wardani, “Stemming Dokumen Teks Bahasa Bali Dengan Metode Rule Base Approach,” *JATISI*, vol. 7, no. 3, 2020, doi: 10.35957/jatisi.v7i3.538.
- [18] M. Hafizh Mahendra, D. Triantoro Murdiansyah, and K. Muslim Lhaksana, “Analisis Sentimen Tweet COVID-19 Menggunakan Metode K-Nearest Neighbors dengan Ekstraksi Fitur TF-IDF dan CountVectorizer,” *Jurnal Ilmu Multidisiplin*, vol. 1(2), pp. 37–43, 2023, doi: 10.69688/dike.v1i2.35.
- [19] F. Syahro and N. Fitriani, “PERBANDINGAN PERFORMA MODEL MACHINE LEARNING SUPPORT VECTOR MACHINE, NEURAL NETWORK, DAN K-NEAREST NEIGHBORS DALAM PREDIKSI HARGA SAHAM,” *Jar 's*, vol. 2, no. 1, p. 13, 2023, doi: 10.24929/jars.v2i1.2983.
- [20] 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.