

Penerapan Metode K-Nearest Neighbors dan Naïve Bayes pada Analisis Sentimen Pengguna Aplikasi Bstation melalui Platform Playstore

Sigit Fathu Amrillah, Dwi Krisbiantoro, Agung Prasetyo*

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

Email: ¹sigitfathuamrillah@gmail.com, ²dwikris@amikompurwokerto.ac.id, ^{3,*}pras@amikompurwokerto.ac.id

Email Penulis Korespondensi: pras@amikompurwokerto.ac.id

Submitted: 29/08/2024; Accepted: 01/12/2024; Published: 03/12/2024

Abstrak—Streaming adalah metode pendistribusian konten digital secara langsung melalui internet, yang memungkinkan pengguna untuk mengakses media tanpa perlu mengunduh file. Bstation adalah platform streaming yang menggabungkan (OGV) dan User-Generated Content (UGC). Penelitian ini menilai efektivitas algoritma K-Nearest Neighbors (KNN) dan Naïve Bayes dalam menganalisis sentimen pada ulasan pengguna terhadap aplikasi Bstation, dengan menggunakan sampel data sebanyak 5.000 ulasan. Permasalahan yang dihadapi adalah banyaknya pengguna aplikasi Bstation, sehingga diperlukan analisis sentimen untuk mengukur dan memahami penilaian masyarakat terhadap aplikasi tersebut secara lebih akurat. Penelitian ini bertujuan menganalisis sentimen pengguna Bstation di Playstore dan membandingkan performa K-Nearest Neighbors (KNN) serta Naïve Bayes untuk menentukan metode terbaik dalam klasifikasi ulasan dan pola sentimen pengguna. Temuan menunjukkan bahwa Naïve Bayes mencapai akurasi 84%, melampaui KNN yang hanya mencapai 68%. Naïve Bayes menunjukkan 86% presisi dan 88% recall untuk sentimen negatif, sementara mencapai 78% presisi dan 76% recall untuk sentimen positif. Sebaliknya, KNN mencapai 80% presisi dan 66% recall untuk sentimen negatif, dan 54% recall untuk sentimen positif. F1-Score untuk Naïve Bayes juga lebih tinggi, mencerminkan keseimbangan yang lebih baik secara keseluruhan. Rata-rata makro dan rata-rata tertimbang rata-rata tertimbang untuk precision, recall, dan F1-score dengan Naïve Bayes masing-masing adalah 82% dan 83%. masing-masing adalah 82% dan 83%, sedangkan KNN mencatat rata-rata makro 0,67. Kesimpulannya, Naïve Bayes lebih efektif dalam analisis sentimen dibandingkan KNN, memberikan performa yang lebih konsisten dan akurat.

Kata Kunci: K-Nearest Neighbors; Naïve Bayes; Analisis Sentimen; Bstation; Playstore

Abstract—Streaming is a method of distributing digital content directly over the internet, which allows users to access media without the need to download files. Bstation is a streaming platform that combines (OGV) and User-Generated Content (UGC). This research assesses the effectiveness of the K-Nearest Neighbors (KNN) and Naïve Bayes algorithms in analyzing sentiment in user reviews of the Bstation application, using a data sample of 5,000 reviews. The problem faced is the large number of users of the Bstation application, so sentiment analysis is needed to measure and understand the public's assessment of the application more accurately. This research aims to analyze the sentiment of Bstation users on Playstore and compare the performance of K-Nearest Neighbors (KNN) and Naïve Bayes to determine the best method for classifying reviews and user sentiment patterns. The findings showed that Naïve Bayes achieved 84% accuracy, surpassing KNN which only achieved 68%. Naïve Bayes showed 86% precision and 88% recall for negative sentiment, while achieving 78% precision and 76% recall for positive sentiment. In contrast, KNN achieved 80% precision and 66% recall for negative sentiments, and 54% recall for positive sentiments. The F1-Score for Naïve Bayes is also higher, reflecting a better balance overall. The macro average and weighted average weighted average for precision, recall, and F1-score with Naïve Bayes were 82% and 83%, respectively, while KNN recorded a macro average of 0.67. In conclusion, Naïve Bayes is more effective in sentiment analysis than KNN, providing more consistent and accurate performance.

Keywords: K-Nearest Neighbors; Naïve Bayes; Sentiment Analysis; Bstation; Playstore

1. PENDAHULUAN

Dalam era digital saat ini, aplikasi streaming telah menjadi salah satu platform hiburan utama bagi masyarakat. Kemajuan teknologi dan penetrasi internet yang semakin luas telah mendorong perubahan signifikan dalam cara orang mengonsumsi media[1]. Aplikasi streaming memungkinkan pengguna untuk mengakses berbagai jenis konten, seperti film, acara televisi, dan video, secara langsung melalui internet tanpa harus menunggu atau mengunduh file[2]. Dengan hanya beberapa klik, pengguna dapat menikmati konten favorit mereka kapan saja dan di mana saja, menjadikan aplikasi ini sangat praktis dan diminati.

Salah satu aplikasi streaming yang sedang berkembang pesat adalah Bstation. Bstation adalah platform gabungan yang mengintegrasikan Video yang Dibuat Secara Profesional (OGV) dan Konten yang Dihasilkan oleh Pengguna (UGC). Melalui penyediaan streaming OGV, penggemar dan pencinta hiburan ACG dapat menikmati berbagai konten yang menarik dan berkualitas tinggi. Sementara itu, para pembuat video memiliki kesempatan untuk menyalurkan kreativitas mereka melalui pembuatan dan berbagi konten UGC. Aplikasi ini telah menarik perhatian pengguna di berbagai kalangan, terutama di kalangan muda yang cenderung lebih tertarik pada konten-konten yang segar, dinamis, dan sesuai dengan tren terkini. Bstation menawarkan berbagai macam konten yang menarik dan beragam, mulai dari animasi, film, serial, hingga video musik, yang semuanya dapat diakses dengan mudah oleh penggunanya. Popularitas Bstation yang terus meningkat tidak lepas dari upaya mereka dalam menghadirkan konten yang berkualitas tinggi dan relevan dengan minat pengguna saat ini.

Dengan semakin meningkatnya jumlah pengguna Bstation, volume ulasan yang diberikan oleh pengguna di Playstore juga meningkat signifikan. Ulasan ini mencerminkan pengalaman pengguna dengan aplikasi tersebut dan

mencakup berbagai aspek, mulai dari kegunaan, kinerja, hingga konten yang disediakan. Ulasan pengguna di Playstore ini tidak hanya penting bagi calon pengguna lainnya sebagai bahan pertimbangan, tetapi juga sangat berharga bagi pengembang aplikasi untuk memahami kekuatan dan kelemahan produk mereka[3].

Analisis sentimen melibatkan penggunaan pemrosesan bahasa alami (NLP), analisis teks, dan metode linguistik komputasi untuk mendeteksi dan mengekstrak informasi subjektif dari sumber teks[4]. Analisis ini bertujuan untuk menentukan sikap atau sentimen pengguna terhadap suatu subjek, apakah bersifat positif dan negatif. Dalam konteks ulasan aplikasi di Playstore, analisis sentimen membantu mengidentifikasi bagaimana perasaan pengguna terhadap berbagai fitur dan aspek aplikasi tersebut[5].

Menganalisis ulasan pengguna di Playstore sangat penting untuk analisis sentimen karena dapat memahami persepsi umum pengguna terhadap aplikasi seperti Bstation. Melalui analisis ini, kita dapat mengekstrak wawasan berharga mengenai aspek-aspek yang disukai maupun yang perlu diperbaiki. Dalam konteks ini, dua metode yang sering di aplikasikan pada saat analisis sentimen ialah K-Nearest Neighbors (KNN) dan Naïve Bayes. KNN adalah metode non-parametrik yang mengklasifikasikan data berdasarkan kedekatannya dengan data lain dalam ruang fitur, sedangkan Naïve Bayes merupakan sebuah metode probabilistik yang menggunakan teorema Bayes dengan asumsi independensi antar fitur yang lainnya[6].

Menurut Tegar Abdillah, Ulfa Khaira dan Benedika Ferdian Hutabarat, dengan judul “Komparasi Metode Naïve Bayes dan K-Nearest Neighbors Terhadap Analisis Sentimen Pengguna Aplikasi Zenius” Pada hasil pengujian skenario pemisahan data menggunakan Split Validation dengan rasio 90:10 antara data training dan testing, Naïve Bayes mencapai akurasi 88,41%, sedangkan KNN mencapai 100%. Dalam analisis ini, KNN mengungguli Naïve Bayes dalam precision, F1-Score dan Recall [7], dan. Dari hasil penelitian dengan menggunakan algoritma KNN dan NBC diperoleh akurasi yang relatif baik sehingga disimpulkan bahwa algoritma KNN dan NBC merupakan algoritma yang tepat untuk digunakan dalam penelitian.

Pada penelitian selanjutnya yang di lakukan oleh Dany Pratmanto dan Fabriyan Fandi Dwi Imaniawan dengan judul “Analisis Sentimen Terhadap Aplikasi Canva Menggunakan Algoritma Naïve Bayes Dan K-Nearest Neighbors” Temuan penelitian menunjukkan bahwa algoritma KNN dan NBC mencapai akurasi yang relatif tinggi, yang menunjukkan bahwa mereka cocok untuk digunakan dalam penelitian ini. Secara khusus, algoritma KNN mencapai akurasi 83,92%, sedangkan Naïve Bayes memiliki akurasi 77,41%. Selain itu, KNN menunjukkan nilai recall dan presisi yang lebih tinggi dibandingkan dengan Naïve Bayes, dengan angka 83,66% dan 84,49%. [8]. Selain itu, algoritma KNN menghasilkan nilai AUC yang lebih tinggi dibandingkan dengan Naïve Bayes, dengan nilai 95,00% berbanding 94,50%.

Penelitian yang di lakukan oleh Syafrizal, M. Afdal, Rice Novita yang berjudul “Analisis Sentimen Ulasan Aplikasi PLN Mobile Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor” penelitian ini menggunakan 3000 data, sedangkan penelitian yang di lakukan menggunakan 5000 data, tidak hanya itu penelitian tersebut juga mengklasifikasikan menjadi 2 klasifikasi yaitu positif netral dan negatif, sedangkan penulis hanya mengklasifikasikan positif dan negatif[9].

Penelitian selanjutnya yang dilakukan oleh Bintang Zulfikar Ramadhan, Ibnu Riza, dan Iqbal Maulana dengan judul “Analisis Sentimen Ulasan Pada Aplikasi E-Commerce Dengan Menggunakan Algoritma Naïve Bayes” Temuan dari penelitian ini melakukan beberapa skenario, dalam skenario 1 (80:20) untuk aplikasi Shopee menggunakan algoritma Naïve Bayes menunjukkan akurasi 92%, presisi 92,13%, recall 98,8%, dan F1-score 95,35%. [10]. Pada penelitian tersebut di buktikan bahwa pada tahap pengujian algoritma melewati 3 tahapan skenario pada saat Pembagian Data Training dan Data Testing mulai dari (70:30), (80:20), dan (90:10), hasil yang paling besar yang di dapatkan adalah pada pembagian data (80:20), maka dari itu penelitian penulis menggunakan perbandingan (80:20).

Meskipun sudah banyak penelitian yang melakukan analisis sentimen terhadap ulasan aplikasi di Playstore, masih terdapat keterbatasan dalam studi yang secara spesifik mengkaji ulasan pengguna aplikasi streaming seperti Bstation di Playstore. Sebagian besar penelitian sebelumnya lebih banyak berfokus pada aplikasi e-commerce atau media sosial. Selain itu, masih minimnya penelitian yang secara langsung membandingkan keefektifan metode KNN serta Naïve Bayes dalam analisis sentimen dengan memanfaatkan ulasan pengguna aplikasi Bstation pada platform Playstore. Oleh karena itu, penelitian ini bertujuan untuk menjembatani kesenjangan ini dengan melakukan analisis sentimen pada ulasan pengguna Playstore untuk aplikasi Bstation menggunakan metode K-Nearest Neighbors dan Naïve Bayes.

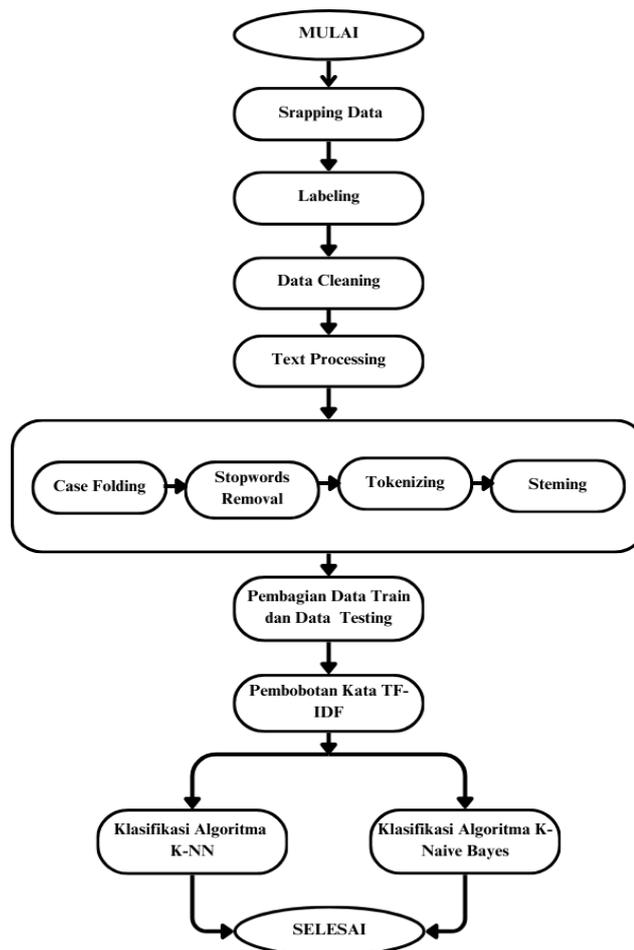
Penelitian ini bertujuan untuk menilai sentimen pengguna terhadap aplikasi Bstation berdasarkan ulasan dari Playstore dan membandingkan efektivitas metode K-Nearest Neighbors dan Naïve Bayes dalam mengklasifikasikan sentimen tersebut. Penelitian ini diharapkan dapat meningkatkan pengembangan aplikasi streaming yang lebih baik dan menawarkan wawasan baru dalam analisis sentimen menggunakan kedua metode tersebut.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini dilaksanakan melalui serangkaian tahapan tertata dan terstruktur dan sistematis untuk memastikan bahwa setiap langkah yang diambil menghasilkan data dan temuan yang akurat serta dapat diandalkan. Tahapan-tahapan ini dirancang dengan cermat untuk menjawab tujuan penelitian secara efektif, dimulai dari pengumpulan data,

pra-pemrosesan, hingga analisis akhir yang bertujuan untuk memberikan wawasan mendalam mengenai sentimen pengguna terhadap aplikasi Bstation



Gambar 1. Tahapan Penelitian

2.2 Scraping Data

Scraping data dari aplikasi Playstore adalah proses pengambilan data secara otomatis dari halaman aplikasi yang ada di Google Playstore[11]. Data yang diambil biasanya mencakup ulasan pengguna, rating, jumlah unduhan, deskripsi aplikasi, dan informasi lain yang tersedia di halaman aplikasi tersebut . Proses ini sering digunakan dalam penelitian atau analisis data untuk memahami sentimen pengguna, tren aplikasi, atau untuk mendapatkan insight lainnya dari data yang tersedia.

2.3 Labeling

Labeling adalah suatu tahapan penting yang melibatkan pemberian label atau kategori pada data mentah yang telah berhasil diperoleh. mengklasifikasikan data ke dalam dua kategori berdasarkan sentimen atau polaritas yang diekspresikan. Dalam konteks analisis sentimen, data seperti ulasan pengguna atau komentar di media sosial sering diberi label sebagai "positif" atau "negatif" berdasarkan nada atau makna dari teks yang dituliskan[12].

2.4 Data Cleaning

Data cleaning adalah proses membersihkan dan mempersiapkan data mentah sebelum digunakan dalam analisis. Proses ini melibatkan beberapa langkah yang dirancang untuk menghapus beberapa kesalahan atau noise, ketidakkonsistenan, dan elemen tidak berguna atau tidak relevan dari data, sehingga meningkatkan akurasi dan keandalan hasil analisis [13].

2.5 Text Processing

Pemrosesan teks melibatkan serangkaian langkah atau teknik dengan bertujuan untuk menyiapkan serta menangani data teks untuk analisis selanjutnya. Dalam bidang analisis sentimen, pembelajaran mesin, atau pemrosesan bahasa alami (NLP), pemrosesan teks sangat penting karena mengubah teks mentah menjadi format terstruktur yang sesuai untuk berbagai tugas analisis [14]. Beberapa tahapan Pada Text Processing



a. Case Folding

Case folding merupakan sebuah proses pengubahan dari semua huruf menjadi huruf yang kecil untuk memberikan perbedaan yang disebabkan oleh huruf besar [15]. Ini memastikan konsistensi dalam analisis, sehingga kata-kata seperti "Hello" dan "hello" diperlakukan sebagai entitas yang sama. Case folding penting untuk menghindari pengaruh kapitalisasi dalam pencarian teks, analisis sentimen, dan aplikasi pemrosesan bahasa alami lainnya.

b. Stopwords Removal

Penghapusan stopwords melibatkan penghilangan kata-kata umum yang kerap muncul dalam sebuah teks namun tidak memberikan kontribusi informasi yang signifikan terhadap analisis. Kata-kata seperti "dan", "atau", "yang", dan istilah serupa lainnya sering dianggap sebagai noise dalam pemrosesan teks karena tidak menambah konteks atau nilai yang berarti bagi analisis [16]. Dengan menghapus stop words, kita dapat fokus terhadap kata-kata yang lebih berarti dan umum sehingga dapat meningkatkan efektivitas analisis dan memperjelas hasil yang diperoleh dari data teks. Proses ini membantu dalam mengurangi dimensi data dan mempercepat pemrosesan, serta memungkinkan algoritma analisis untuk lebih fokus pada informasi yang substansial.

c. Tokenizing

Tokenizing merupakan sebuah proses membagi teks menjadi unit-unit lebih kecil yang biasa disebut token. Token ini bisa dalam bentuk kata, frasa, atau simbol yang merupakan bagian dasar dari teks yang akan dianalisis. Tokenisasi adalah langkah pertama pada pemrosesan bahasa alami (NLP) serta analisis teks yang memungkinkan pemahaman serta pemrosesan teks yang lebih terstruktur [17].

d. Stemming

Stemming melibatkan pengurangan kata menjadi kata dasar atau akarnya dengan cara menghilangkan sufiks atau biasa disebut awalan. [18]. Stemming merupakan sebuah proses perubahan kata menjadi bentuk dasar dengan cara menghilangkan sufiks atau awalan. Stemming membantu dalam pemrosesan bahasa alami (NLP) dan analisis teks dengan mengurangi dimensi data dan meningkatkan kemampuan sistem untuk mengidentifikasi hubungan antar kata yang memiliki makna serupa. Proses ini penting dalam berbagai aplikasi seperti pencarian informasi, pengolahan teks, dan analisis sentimen, di mana konsistensi bentuk kata dapat mempermudah pengolahan dan analisis data teks.

2.6 Pembagian Data Traing dan Data Testing

Proses membagi data menjadi subset pelatihan dan pengujian merupakan sebuah teknik yang di aplikasikan ke dalam pembelajaran mesin guna membagi dataset menjadi dua bagian itu salah satu untuk melatih model dan yang lainnya untuk mengevaluasi kinerjanya [19].

a. Data Train (Training Data) : Ini adalah subset dari dataset yang digunakan untuk proses pelatihan model. Pada fase ini, model machine learning memproses data ini untuk mengidentifikasi pola, hubungan, dan karakteristik yang relevan, yang kemudian digunakan untuk membangun model prediktif. Biasanya, data train mencakup proporsi besar dari keseluruhan dataset, seperti 70% hingga 80%

b. Data Testing (Testing Data) : Data testing mengacu pada sebagian dataset yang diaplikasikan guna menilai dan mengukur pada kinerja model yang sudah dilatih. Tujuan dari data testing yaitu guna mengukur sejauh mana model yang dilatih dapat menyimpulkan atau menerapkan pengetahuan yang diperoleh dari data train pada data baru dan tidak dikenal. Data testing umumnya mencakup sisa dari dataset, misalnya 20% hingga 30%

Proses pembagian ini penting guna menghindari overfitting, yaitu model data terlalu sesuai dengan data pelatihan dan kurang mampu generalisasi pada data yang belum terlihat. Selain itu, pembagian ini memungkinkan penilaian objektif terhadap efektivitas model dan kemampuannya dengan menerapkan pembelajaran pada data yang tidak terlihat sebelumnya.

2.7 Pembobotan TF-IDF

TF-IDF (Term Frequency-Inverse Document Frequency) merupakan sebuah teknik statistik yang dilakukan untuk mengevaluasi signifikansi sebuah kata yang ada di dalam sebuah dokumen dengan mempertimbangkan frekuensinya pada dokumen tersebut dibandingkan dengan kemunculannya di seluruh korpus dokumen. Teknik ini menggabungkan frekuensi istilah (TF) dari kata yang ada pada dokumen dengan frekuensi dokumen kebalikannya (IDF) di seluruh korpus. TF-IDF sering di aplikasikan pada pemrosesan bahasa alami serta pencarian informasi guna mengidentifikasi fitur-fitur teks utama dan menilai relevansi kata-kata dalam dokumen.

2.8 Klasifikasi Algoritma K-Nearest Neighbors (KNN)

Dalam analisis sentimen, algoritma K-Nearest Neighbors (KNN) digunakan untuk mengkategorikan teks ke dalam kelas-kelas sentimen, seperti positif atau negatif [6]. Proses ini dimulai dengan membandingkan teks baru yang ingin dianalisis dengan teks-teks pada data Train yang sudah melalui tahap labeling sentimen [20].

$$dis(x_1, x_2) = \sqrt{\sum_{i=0}^n (x_{1i} - x_{2i})^2} \quad (1)$$

Pada Rumus 1 merupakan sebuah rumus algoritma K-Nearest Neighbors (KNN), jarak antara sampel data (x_1) (data yang akan diklasifikasikan) dan (x_2) (data training) dihitung menggunakan rumus jarak, seperti jarak



Euclidean. Variabel (i) mewakili setiap fitur atau atribut dari data, dan (d) adalah jarak yang dihasilkan dari perhitungan tersebut. Setelah jarak antara (x1) dan semua data training (x2) dihitung, algoritma akan memilih (k) tetangga terdekat berdasarkan jarak terkecil. Kemudian, kelas atau nilai dari data (x1) ditentukan oleh mayoritas dari kelas tetangga terdekat tersebut, sementara (p) mewakili jumlah total data training yang digunakan dalam proses perhitungan.

2.9 Klasifikasi Algoritma Naïve Bayes

Pada bagian selanjutnya, mengaplikasikan algoritma *Naïve Bayes*, yang didasarkan pada asumsi kuat tentang independensi setiap fitur atau kata dalam kelas yang diberikan. Tujuan penerapan *Naïve Bayes* adalah untuk mencapai tingkat akurasi dalam klasifikasi data ulasan komentar mengenai Aplikasi Bukalapak, dengan mengelompokkan kutipan ke dalam 2 golongan pokok positif dan negatif. Algoritma klasifikasi *Naïve Bayes* bertujuan untuk memisahkan data ke dalam kelas-kelas yang spesifik. Kinerja pengklasifikasi dievaluasi berdasarkan akurasi prediksi [21], [22]. Algoritme ini memungkinkan sistem untuk secara akurat memahami dan memprediksi sentimen ulasan melalui analisis teks, menawarkan dasar yang kuat untuk menilai kualitas layanan aplikasi dan umpan balik pengguna. Metodologi penelitian melibatkan perincian masalah yang diteliti setepat mungkin (secara matematis, jika memungkinkan) dan menguraikan pendekatan yang diusulkan. Algoritma *Naïve Bayes* didasarkan pada Teorema Bayes dengan asumsi bahwa setiap fitur bersifat independen. Rumus Teorema Bayes dapat ditulis sebagai berikut:

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

Pada Rumus 2 merupakan sebuah rumus Teorema Bayes dimana rumus yang digunakan untuk menghitung probabilitas terjadinya suatu peristiwa berdasarkan informasi atau bukti yang ada. Dalam rumus tersebut, (P(H|X)) merepresentasikan probabilitas suatu hipotesis (H) benar setelah adanya bukti (X), dikenal sebagai posterior probability. Probabilitas ini dihitung dengan mengalikan (P(X|H)), yaitu peluang bukti (X) muncul jika hipotesis (H) benar (disebut likelihood), dengan (P(H)), yaitu probabilitas awal dari hipotesis(H) sebelum melihat bukti (disebut prior probability). Hasilnya kemudian dibagi dengan (P(X)), probabilitas keseluruhan dari bukti (X) terjadi tanpa memperhatikan hipotesis (H), yang dikenal sebagai marginal likelihood. Dengan teorema ini, kita dapat memperbarui keyakinan terhadap suatu hipotesis seiring dengan adanya bukti baru.

3. HASIL DAN PEMBAHASAN

Pada bagian ini merupakan bagian yang akan dibahas mengenai hasil analisis sentimen yang dilakukan terhadap ulasan pengguna terhadap aplikasi streaming Bstation di Playstore dengan menerapkan algoritma K-Nearest Neighbors (KNN) serta algoritma *Naïve Bayes*. Pada Hasil evaluasi model klasifikasi yang diterapkan akan diuraikan secara mendetail, meliputi performa setiap metode dalam mengidentifikasi sentimen positif, negatif, dan netral. Selain itu, akan dilakukan perbandingan antara kedua metode untuk menilai keakuratan, presisi, dan recall yang dihasilkan, serta interpretasi dari metrik-metrik evaluasi seperti True Negative, True Positive, False Positive, dan False Negative.

3.1 Scraping Data

Data akan dikumpulkan dari sumber data yaitu Google Play Store dengan memanfaatkan alat `google-play-scraper` untuk aplikasi Bstation, yang menghasilkan total 5.000 ulasan pengguna. Proses pengumpulan ini mencakup ekstraksi elemen-elemen penting seperti rating yang diberikan oleh pengguna, komentar dalam ulasan, serta metadata terkait seperti tanggal ulasan dan versi aplikasi yang digunakan. Pengumpulan data secara menyeluruh ini bermaksud guna mendapatkan bayangan yang komprehensif mengenai sentimen ulasan pengguna pada aplikasi Bstation. Dengan volume data yang signifikan, analisis yang akan dilakukan diharapkan dapat mengidentifikasi pola-pola sentimen yang relevan, memberikan wawasan mengenai tingkat kepuasan dan ketidakpuasan pengguna, serta mengungkap area-area yang memerlukan perbaikan dalam aplikasi. Data yang diperoleh akan digunakan sebagai basis untuk analisis sentimen lebih lanjut dengan menerapkan metode analisis data yang canggih seperti K-Nearest Neighbors dan *Naïve Bayes*, guna menghasilkan temuan yang bermanfaat bagi pengembang aplikasi serta pemangku kepentingan lainnya.

Tabel 1. Hasil dari Scraping Data

	userName	score	at	content
0	Pengguna Google	5	2024-07-31 21:05:35	Aku gak suka setiap kali unduh, selalu terjeda..
1	Pengguna Google	1	2024-08-03 14:02:54	Kritik dari saya, iklannya mengganggu banget, ..
2	Pengguna Google	1	2024-07-29 19:07:18	Lebih baik versi sebelumnya min tapi gk ada ik..
3	Pengguna Google	1	2024-08-07 08:27:30	Sungguh mengecewakan setelah update aplikasi j..
4	Pengguna Google	1	2024-08-03 09:49:11	saya sih ga keberatan kalau disuruh premium, t..

Pada Tabel 1 ditampilkan tahapan-tahapan yang dilakukan dalam proses scraping data. Tahapan ini meliputi pengumpulan berbagai informasi, seperti tanggal ulasan, komentar atau opini yang diberikan oleh pengguna, serta data lain yang relevan untuk keperluan analisis lebih lanjut.

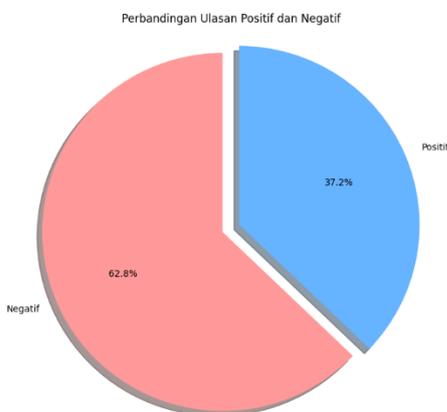
3.2 Labeling Data

Dalam bagian ini, hasil penelitian akan diuraikan secara rinci, diikuti oleh pembahasan mengenai implikasi dari temuan tersebut. Pertama, hasil utama dari analisis data akan disajikan dengan jelas. Kemudian, makna dan relevansi hasil tersebut akan diinterpretasikan dalam konteks penelitian yang lebih luas. Selanjutnya, hubungan temuan dengan teori dan penelitian sebelumnya akan dianalisis, serta potensi dampak praktis atau teoritis akan dikaji. Terakhir, bagian ini akan membahas keterbatasan penelitian dan memberikan rekomendasi untuk penelitian lanjutan berdasarkan hasil yang diperoleh, pada penjelasan di atas di representasikan pada di bawah ini Tabel 2. Pemberian Label Positif dan Negatif.

Tabel 2. Pemberian Label Positif dan Negatif

	content	score	Label
3692	Tolong ya Bstation? Kurang-kurangnya iklan yang..	1	Negatif
2176	Sebenarnya saya mau kasih bintang 5 tapi anime..	4	Positif
708	Semenjak update frame nya drop parah sama seri..	1	Negatif
117	Kadang subtitle gk ad sm klo mau masuk itu sus..	3	None
1325	Bagus sih, tapi bnyk premium sama anime ny g b..	4	Positif

Pada Tabel 2 ditunjukkan proses pemberian label pada setiap ulasan pengguna, di mana ulasan-ulasannya dikategorikan ke dalam dua kelompok, yaitu positif dan negatif. Pemberian label ini dilakukan berdasarkan skor yang diperoleh dari masing-masing ulasan, sehingga setiap ulasan dapat diidentifikasi sesuai dengan sentimen yang diungkapkan oleh pengguna.



Gambar 2. Hasil Angka dari Pelabelan Data

Pada Gambar 2. merupakan hasil analisis sentimen menunjukkan bahwa 37,2% dari ulasan pengguna aplikasi Bstation di Playstore diklasifikasikan menjadi sentimen positif, sementara itu 62,8% dikategorikan menjadi sentimen negatif. Persentase sentimen positif yang relatif lebih rendah menunjukkan bahwa sebagian besar ulasan pengguna cenderung kurang puas dengan aplikasi, mencerminkan adanya kekurangan atau masalah yang dirasakan oleh pengguna. Sebaliknya, persentase sentimen negatif yang lebih tinggi menandakan adanya tren ketidakpuasan yang dominan dalam feedback pengguna, yang bisa menjadi indikasi area-area tertentu dalam aplikasi yang memerlukan perbaikan atau peningkatan.

3.3 Data Cleaning

Data cleaning merupakan suatu tahapan yang penting dalam menganalisis data dengan tujuan untuk memperbaiki dan menyiapkan data agar dapat digunakan dengan akurat dan efektif. Proses ini melibatkan beberapa langkah untuk memastikan bahwa data yang digunakan bebas dari kesalahan, inkonsistensi, atau data yang tidak relevan, seperti yang di tampilkan dalam Tabel 3.

Tabel 3. Tahap Data Cleaning

	Content	score	Label
0	Tolong ya Bstation? Kurang-kurangnya iklan yang..	1	Negatif
1	Sebenarnya saya mau kasih bintang 5 tapi anime..	4	Positif
2	Semenjak update frame nya drop parah sama seri..	1	Negatif
4	Bagus sih, tapi bnyk premium sama anime ny g b..	4	Positif

Pada Tabel 3 dijelaskan tahap data cleaning, yang merupakan proses penting dalam penyiapan data sebelum dianalisis lebih lanjut. Pada tahap ini, data yang tidak relevan atau tidak lengkap dibersihkan untuk memastikan kualitas data yang optimal. Setelah proses data cleaning selesai, diperoleh data sebanyak 4.263 data yang siap untuk dianalisis.

3.4 Text Processing

Text processing, atau pemrosesan teks, adalah serangkaian teknik yang digunakan untuk memanipulasi dan menganalisis data berbasis teks. Teknik ini memainkan peran krusial dalam berbagai aplikasi yang melibatkan interaksi dengan data teks. Pemrosesan teks merupakan komponen penting pada pemrosesan bahasa alami (NLP), yang bertujuan untuk memungkinkan komputer memahami, menganalisis, dan menghasilkan bahasa manusia dengan cara yang berarti. Selain itu, pemrosesan teks juga berperan dalam analisis sentimen, yang di aplikasikan guna memberi menilai dan memahami perasaan maupun pendapat yang ada dalam teks, serta dalam sistem pencarian informasi yang memungkinkan pencarian dan ekstraksi informasi relevan dari koleksi data teks yang besar. Proses pemrosesan teks mencakup berbagai langkah, mulai dari pre-processing untuk membersihkan dan menyiapkan data teks. Langkah-langkah ini meliputi case folding, Stopwords Removal, Tokenizing, Stemming.

a. Case Folding

Case folding merupakan suatu teknik pemrosesan teks yang melibatkan perubahan dari keseluruhan huruf dalam teks beertansformasi huruf kecil. Tujuan utama dari case folding adalah guna mengurangi jenis-jenis dalam data teks oleh karena itu perbandingan dan pencarian menjadi lebih konsisten dan akurat. Ini membantu dalam mengatasi masalah yang disebabkan oleh perbedaan kapitalisasi huruf yang tidak relevan dalam analisis teks. Pada Tabel 4. Ditampilkan hasil dari Case folding.

Tabel 4. Hasil dari Case Folding

	content	score	Label	text_clean
0	Tolong ya Bstation? Kurang-kurangin iklan yang..	1	Negatif	tolong ya bstation kurangkurangin iklan yang s..
1	Sebenarnya saya mau kasih bintang 5 tapi anime..	4	Positif	sebenarnya saya mau kasih bintang tapi anime ..
2	Semenjak update frame nya drop parah sama seri..	1	Negatif	semenjak update frame nya drop parah sama seri..
4	Bagus sih, tapi bnyk premium sama anime ny g b..	4	Positif	bagus sih tapi bnyk premium sama anime ny g be..

Pada Tabel 4 ditampilkan hasil dari proses case folding, yaitu salah satu langkah dalam pra-pemrosesan data yang bertujuan untuk mengubah seluruh huruf pada teks menjadi huruf kecil. Proses ini dilakukan untuk menyamakan format penulisan dan mengurangi variasi yang tidak diperlukan dalam analisis teks.

b. Stopwords Removal

Penghapusan kata henti adalah teknik penting pada pemrosesan teks dimana terdapat penghilangan kata-kata umum dan tidak penting dari teks. Kata-kata henti ini sering di aplikasikan pada bahasa akan tetapi belum memberikan informasi atau konteks yang sangat berarti untuk analisis. Contohnya termasuk “dan”, “atau”, “adalah”, “itu”, dan “dalam”. Proses ini diilustrasikan pada Tabel 5, yang menunjukkan perubahan case folding ke stopwords removal.

Tabel 5. Perubahan dari Case Folding ke Stopwords Removal

	content	score	Label	text_clean	text_StopWord
0	Tolong ya Bstation? Kurang-kurangin iklan yang..	1	Negatif	tolong ya bstation kurangkurangin iklan yang s..	tolong ya bstation kurangkurangin iklan mengga..
1	Sebenarnya saya mau kasih bintang 5 tapi anime..	4	Positif	sebenarnya saya mau kasih bintang tapi anime ..	kasih bintang anime cari kasih bintang deh
2	Semenjak update frame nya drop parah sama seri..	1	Negatif	semenjak update frame nya drop parah sama seri..	semenjak update frame nya drop parah
4	Bagus sih, tapi bnyk premium sama anime ny g b..	4	Positif	bagus sih tapi bnyk premium sama anime ny g be..	bagus sih bnyk premium anime ny g lengkap

Seperti yang ditampilkan pada Tabel 5, setelah proses case folding selesai, langkah berikutnya adalah stopwords removal. Pada tahap ini, kata-kata umum yang tidak memiliki makna signifikan dalam analisis, seperti "dan", "yang", "di", dan lainnya, dihapus dari teks. Proses ini bertujuan untuk memfokuskan analisis hanya pada kata-kata yang lebih relevan terhadap konteks ulasan.



c. Tokenizing

Tokenizing merupakan langkah dasar pada pemrosesan teks dimana melibatkan pemecahan teks panjang yang ada pada dataset menjadi unit-unit yang cenderung lebih kecil yang biasa disebut token. Token dapat berupa kata, frasa, kalimat, mungkin bahkan karakter, tergantung pada tujuan analisis dan teknik yang digunakan. Tokenisasi membantu dalam mengubah teks yang menjadi format lebih terstruktur dan bisa dikelola untuk analisis selanjutnya. Tabel 6. Merupakan hasil dari pengubahan Stopwords ke Tokenizing.

Tabel 6. Hasil Pengubahan Stopwords Removal ke Tokenizing

content	score	Label	text_clean	text_StopWord	text_tokens
0 Tolong ya Bstation? Kurang-kurangnya iklan yang..	1	Negatif	tolong ya bstation kurangkurangin iklan yang s..	tolong ya bstation kurangkurangin iklan mengga..	[tolong, ya, bstation, kurangkurangin, iklan, ..
1 Sebenarnya saya mau kasih bintang 5 tapi anime..	4	Positif	sebenarnya saya mau kasih bintang tapi anime ..	kasih bintang anime cari kasih bintang deh	[kasih, bintang, anime, cari, kasih, bintang, ..
2 Semenjak update frame nya drop parah sama seri..	1	Negatif	semenjak update frame nya drop parah sama seri..	semenjak update frame nya drop parah	[semenjak, update, frame, nya, drop, parah]
4 Bagus sih, tapi bnyk premium sama anime ny g b..	4	Positif	bagus sih tapi bnyk premium sama anime ny g be..	bagus sih bnyk premium anime ny g lengkap	[bagus, sih, bnyk, premium, anime, ny, g, leng..

Tabel 6 menunjukkan hasil dari proses stopwords removal yang dilanjutkan dengan tokenizing, yang merupakan langkah penting dalam pemrosesan teks untuk analisis sentiment setelah itu, teks dipecah menjadi token-token individual pada kolom text_tokens, yang merupakan kumpulan kata-kata yang siap digunakan untuk analisis sentimen lebih lanjut.

d. Stemming

Stemming adalah teknik dalam pemrosesan teks yang digunakan untuk mengurangi kata-kata ke bentuk dasar atau "stem" mereka. Tujuan utama dari stemming adalah untuk mengelompokkan berbagai variasi dari kata yang sama sehingga mereka dapat diperlakukan sebagai entitas yang sama dalam analisis. Misalnya, kata-kata seperti "running", "runner", dan "ran" dapat dikembalikan ke bentuk dasar "run" melalui proses stemming. Pada Tabel 7. adalah hasil dari Stemming

Tabel 7. Hasil dari Proses Tokenizing ke Stemming

content	score	Label	text_clean	text_StopWord	text_tokens	text_steamindo
0 Tolong ya Bstation? Kurang-kurangnya iklan yang..	1	Negatif	tolong ya bstation kurangkurangin iklan yang s..	tolong ya bstation kurangkurangin iklan mengga..	[tolong, ya, bstation, kurangkurangin, iklan, ..	tolong ya bstation kurangkuran gin iklan mengga..
1 Sebenarnya saya mau kasih bintang 5 tapi anime..	4	Positif	sebenarnya saya mau kasih bintang tapi anime ..	kasih bintang anime cari kasih bintang deh	[kasih, bintang, anime, cari, kasih, bintang, ..	kasih bintang anime cari kasih bintang deh
2 Semenjak update frame nya drop parah sama seri..	1	Negatif	semenjak update frame nya drop parah sama seri..	semenjak update frame nya drop parah	[semenjak, update, frame, nya, drop, parah]	semenjak update frame nya drop parah
4 Bagus sih, tapi bnyk premium sama anime ny g b..	4	Positif	bagus sih tapi bnyk premium sama anime ny g be..	bagus sih bnyk premium anime ny g lengkap	[bagus, sih, bnyk, premium, anime, ny, g, leng..	bagus sih bnyk premium anime ny g lengkap

Pada Tabel 7 ditampilkan hasil dari proses stemming, yaitu tahap pemrosesan teks di mana kata-kata dalam ulasan dikembalikan ke bentuk dasarnya. Proses ini dilakukan untuk menyederhanakan variasi kata yang memiliki makna serupa dengan mengubahnya menjadi akar kata. Dengan demikian, kata-kata yang memiliki bentuk berbeda namun

berakar sama, seperti "berlari", "lari", dan "pelari", semuanya akan distandardisasi menjadi "lari", sehingga analisis sentimen menjadi lebih konsisten dan akurat.

3.5 Pembobotan TF-IDF

TF-IDF (Term Frequency-Inverse Document Frequency) merupakan sebuah metode statistik yang di aplikasikan guna menilai pentingnya sebuah kata pada dokumen relatif terhadap koleksi dokumen (corpus). Ini adalah teknik yang sering digunakan dalam analisis teks, termasuk analisis sentimen, untuk mengekstraksi fitur dari data teks.

```
[0.00037659 0.00037257 0.00037855 ... 0.00079068 0.00090295 0.00063563]
[0.0008339 0.00051338 0.0002712 0.00087474 0.0004606 0.0005255
```

Gambar 3. Pembobotan TF-IDF

Pada Gambar 3. merupakan hasil yang di tunjukkan merupakan bagian dari vektor hasil perhitungan TF-IDF (Term Frequency-Inverse Document Frequency) untuk suatu teks atau dokumen. Angka-angka dalam vektor ini ([0.00037659, 0.00037257, ..]) menunjukkan nilai TF-IDF untuk masing-masing kata, dengan nilai rendah seperti 0.00037659 dan 0.00037257 menunjukkan bahwa kata-kata tersebut memiliki frekuensi relatif rendah dalam dokumen atau sering muncul di seluruh dokumen dalam corpus, sehingga tidak terlalu penting atau unik. Sebaliknya, nilai yang lebih tinggi seperti 0.00079068 dan 0.00090295 menunjukkan kata-kata yang lebih unik dan relevan untuk dokumen tersebut dibandingkan dengan dokumen lainnya dalam corpus. Dengan demikian, vektor TF-IDF ini digunakan guna menentukan kata-kata yang paling direkomendasikan dalam sebuah dokumen dan membedakan dokumen berdasarkan kata-kata uniknya

3.6 Pembagian Data Traing dan Data Testing

```
(3410,)
(3410,)
(853,)
(853,)
```

Gambar 4. Pembagian Data Training dan Data Testing

Pada Gambar 4. merupakan pengelompokan data menjadi 2 bagian yaitu data train dan data test merupakan sebuah langkah sangat penting pada proses machine learning guna memastikan model yang dilatih dapat dievaluasi dengan data yang belum pernah sebelumnya diperlihatkan. Pembagian ini membantu menilai bagaimana model akan berperformasi pada data yang tidak dikenali. Berdasarkan informasi yang diberikan:

- (3410,): Ini adalah ukuran dari data train, menunjukkan ada 3410 sampel.
- (3410,): Ini juga ukuran dari data train, kemungkinan menunjukkan bahwa ada 3410 sampel dalam data train.
- (853,): Ini adalah ukuran dari data test, menunjukkan ada 853 sampel.
- (853,): Ini juga ukuran dari data test, kemungkinan menunjukkan bahwa ada 853 sampel dalam data test.

3.7 Klasifikasi Algoritma K-NEAREST NEIGHBORS (KNN)

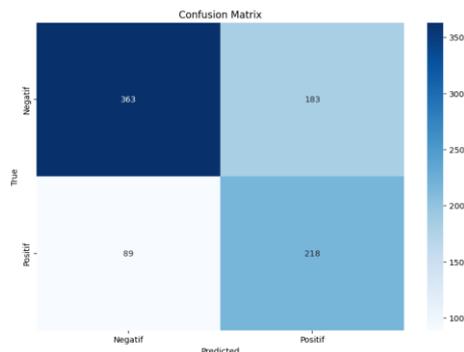
```
Accuracy of KNN Classifier on test set: 0.68
Confusion Matrix:
[[363 183]
 [ 89 218]]
Classification Report:
              precision    recall  f1-score   support

   Negatif      0.80      0.66      0.73      546
   Positif      0.54      0.71      0.62      307

   accuracy          0.68      0.68      0.68      853
  macro avg          0.67      0.69      0.67      853
 weighted avg          0.71      0.68      0.69      853
```

Gambar 5. Hasil dari Algoritma K-Nearst Neighbors (KNN)

Pada Gambar 5. adalah hasil dari algoritma K-Nearest Neighbors (KNN) menunjukkan bahwa model memiliki akurasi keseluruhan sebesar 68% dalam mengklasifikasikan ulasan pengguna aplikasi. Untuk kategori sentimen negatif, model mencapai precision 80%, yang berarti 80% dari prediksi negatif adalah benar, namun recall-nya hanya 66%, menunjukkan bahwa 66% dari ulasan negatif sebenarnya berhasil teridentifikasi. Sementara itu, untuk kategori sentimen positif, precision model adalah 54%, menandakan bahwa setengah dari prediksi positif adalah benar, dengan recall 71%, yang menunjukkan model mampu mengidentifikasi 71% dari ulasan positif dengan benar. F1-score untuk kategori negatif adalah 0.73 dan untuk positif adalah 0.62, menunjukkan keseimbangan antara precision dan recall di kedua kategori. Rata-rata pada hasil makro dari precision, recall, dan f1-score adalah 0.67, sedangkan rata-rata tertimbang adalah 0.69, mencerminkan performa model yang sedikit lebih baik dalam mengklasifikasikan sentimen secara keseluruhan dengan penekanan lebih pada kategori negatif



Gambar 6. Confusion Matrix Algoritma K-Nearest Neighbors (KNN)

Pada Gambar 6. adalah hasil evaluasi model yang menunjukkan bahwa model berhasil mengidentifikasi 363 sampel negatif dengan benar (True Negative) dan 218 sampel positif dengan benar (True Positive). Namun, model juga melakukan kesalahan dengan memprediksi 183 sampel negatif sebagai positif (False Positive) dan gagal mendeteksi 89 sampel positif yang sebenarnya (False Negative). Secara keseluruhan, model lebih baik dalam mengidentifikasi kelas negatif, tetapi masih cenderung overestimasi terhadap kelas positif dan gagal mengenali beberapa sampel positif yang sebenarnya. Hal ini menunjukkan bahwa model perlu ditingkatkan untuk mengurangi kesalahan prediksi dan meningkatkan deteksi kelas positif.

3.8 Klasifikasi Algoritma Naïve Bayes

```

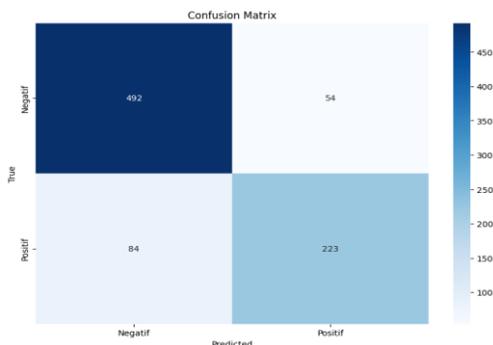
MultinomialNB Accuracy: 0.8382180539273154
MultinomialNB Precision: 0.8541666666666666
MultinomialNB Recall: 0.9010989010989011
MultinomialNB f1_score: 0.8770053475935828
=====

```

	precision	recall	f1-score	support
Negatif	0.85	0.90	0.88	546
Positif	0.81	0.73	0.76	307
accuracy			0.84	853
macro avg	0.83	0.81	0.82	853
weighted avg	0.84	0.84	0.84	853

Gambar 7. Hasil dari Algoritma Naïve Bayes

Pada Gambar 7. Merupakan Hasil evaluasi model yang tunjukkan bahwa pada kelas Negatif, model mencapai presisi sebesar 86%, recall 88%, dan f1-score 87%, memperlihatkan kinerja yang cukup baik dalam mengidentifikasi dan mengklasifikasikan sampel negatif dengan benar. Untuk kelas Positif, presisi model adalah 78%, recall 76%, dan f1-score 77%, yang menunjukkan bahwa model sedikit kurang efektif dalam mendeteksi dan mengklasifikasikan sampel positif dibandingkan dengan sampel negatif. Secara keseluruhan, model mencapai akurasi sebesar 83% memiliki rata-rata presisi, recall, dan f1-score masing-masing sebesar 82% untuk macro average dan 83% untuk weighted average. Kesimpulannya, model ini lebih andal dalam mengklasifikasikan sampel negatif daripada sampel positif, tetapi masih memiliki kinerja yang konsisten dan solid secara keseluruhan. Namun, ada ruang untuk peningkatan dalam mendeteksi dan mengklasifikasikan sampel positif.



Gambar 8. Hasil Confusion Matrix Algoritma Naïve Bayes

Pada Gambar 8. Merupakan Hasil evaluasi model yang menunjukkan bahwa model berhasil mengidentifikasi 492 sampel negatif dengan benar (True Negative) dan 223 sampel positif dengan benar (True Positive), menunjukkan kinerja yang kuat dalam mendeteksi kedua kelas. Namun, model juga salah mengklasifikasikan 84 sampel positif

sebagai negatif (False Negative) dan 54 sampel negatif sebagai positif (False Positive). Secara keseluruhan, model membuktikan kemampuan yang cukup baik mendeteksi kelas negatif pada kesalahan minimal dalam mengklasifikasikan sampel negatif sebagai positif. Namun, ada sedikit kekurangan dalam mendeteksi semua sampel positif, yang tercermin dari jumlah False Negatives yang lebih tinggi. Kesimpulannya, meskipun model cukup andal dalam mengidentifikasi sampel negatif, perbaikan diperlukan untuk meningkatkan deteksi sampel positif agar mengurangi kesalahan klasifikasi yang terjadi.

4. KESIMPULAN

Dalam penelitian ini, perbandingan antara algoritma K-Nearest Neighbors (KNN) dan Naïve Bayes guna menganalisis sentimen mengungkapkan perbedaan kinerja yang signifikan. Hasil evaluasi menunjukkan bahwa Naïve Bayes unggul dengan akurasi 84%, dibandingkan dengan KNN yang mencapai akurasi 68%. Ini menunjukkan bahwa Naïve Bayes secara keseluruhan lebih akurat dalam memprediksi sentimen ulasan pengguna. Untuk kelas sentimen negatif, Naïve Bayes memperoleh presisi 86%, recall 88%, dan f1-score 87%, menandakan kemampuan yang sangat baik dalam mengidentifikasi dan mengklasifikasikan ulasan negatif. Sebaliknya, KNN mencatat presisi 80% dan recall 66% untuk sentimen negatif, yang menunjukkan performa yang lebih rendah dalam mendeteksi ulasan negatif. Pada kelas sentimen positif, Naïve Bayes memiliki presisi 78%, recall 76%, dan f1-score 77%, sedangkan KNN mencatat presisi 54% dan recall 71%. F1-score Naïve Bayes juga lebih tinggi dibandingkan KNN, menunjukkan keseimbangan yang relatif baik antara presisi dan recall. Rata-rata makro dan rata-rata tertimbang dari presisi, recall, dan f1-score pada Naïve Bayes masing-masing adalah 82% dan 83%, sedangkan pada KNN, rata-rata makro dari precision, recall, dan F1-score adalah 0.67. Dengan hasil-hasil ini, dapat disimpulkan bahwa Naïve Bayes lebih efektif dan efisien dalam menganalisis sentimen ulasan dibandingkan dengan KNN, memberikan performa yang lebih konsisten dan akurat dalam identifikasi sentimen.

REFERENCES

- [1] A. Sampurna, F. Ramadhan, S. Al Azhar Sihombing, A. Balqis, and A. Ridha, "Dampak Integrasi Platform Streaming Online dalam Transformasi Broadcasting Kontemporer," *Jurnal Pendidikan Tambusai*, vol. 8(1), pp. 4821-4829, 2024, doi: 10.25139/jkm.v5i1.3637.
- [2] Alya Dwi Yuliani and Oji Kurniadi, "Peranan Media Streaming dalam Menggantikan Televisi Konvensional di Kalangan Masyarakat," *Jurnal Riset Manajemen Komunikasi*, vol. 3, no. 2, pp. 109-114, Dec. 2023, doi: 10.29313/jrmk.v3i2.3140.
- [3] F. Noor Hasan, "Analisis Sentimen Pengguna Aplikasi CapCut Pada Ulasan di Play Store Menggunakan Metode Naïve Bayes," *KLIK*, vol. 4, no. 4, 2024, doi: 10.30865/klik.v4i4.1555.
- [4] A. Mukti, A. D. Hadiyanti, A. Nurlaela, and J. Panjaitan, "Sistem Analisa Sentiment Bakal Calon Presiden 2024 Menggunakan Metode NLP Berbasis Web," *JURIKOM*, vol. 6, no. 1, p. p-ISSN, 2023, doi: 10.30865/jurikom.v9i2.3989.
- [5] M. N. Fahriza and N. Riza, "ANALISIS SENTIMEN PADA ULASAN APLIKASI CHAT GENERATIVE PRE-TRAINED TRANSFORMER GPT MENGGUNAKAN METODE KLASIFIKASI K-NEAREST NEIGHBOR(KNN) Systematic Literature Review," *Jurnal Mahasiswa Teknik Informatika*, vol. 7, no. 2, 2023, doi: 10.36040/jati.v7i2.6767.
- [6] Mulyana D I and AKbar A, "OPTIMASI KLASIFIKASI BATIK BETAWI MENGGUNAKAN DATA AUGMENTASI DENGAN METODE KNN DAN GLCM," *Jurnal Aplikasi Teknologi Informasi dan Manajemen (JATIM)*, vol. 3 no.2, 2022, doi: 10.31102/jatim.v3i2.1577.
- [7] T. Abdillah, U. Khaira, and B. F. Hutabarat, "Komparasi Metode Naive Bayes dan K-Nearest Neighbors Terhadap Analisis Sentimen Pengguna Aplikasi Zenius," *Jurnal PROCESSOR*, vol. 19, no. 1, May 2024, doi: 10.33998/processor.2024.19.1.1596.
- [8] D. Prاتمanto, F. Fandi, D. Imaniawan, and C. Author, "Analisis Sentimen Terhadap Aplikasi Canva Menggunakan Algoritma Naive Bayes Dan K-Nearest Neighbors," *Computer Science (CO-SCIENCE)*, vol. 3, no. 2, p. 52112, 2023, doi: 10.31294/coscience.v3i2.1917.
- [9] S. Syafrizal, M. Afdal, and R. Novita, "Analisis Sentimen Ulasan Aplikasi PLN Mobile Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor," *MALCOM*, vol. 4, no. 1, pp. 10-19, Dec. 2023, doi: 10.57152/malcom.v4i1.983.
- [10] B. Z. Ramadhan, I. Riza, and I. Maulana, "Analisis Sentimen Ulasan Pada Aplikasi E-Commerce Dengan Menggunakan Algoritma Naïve Bayes," *Journal of Applied Informatics and Computing (JAIC)*, vol. 6, no. 2, p. 220, 2022, doi: 10.30871/jaic.v6i2.4725.
- [11] N. C. Agustina, D. Herlina Citra, W. Purnama, C. Nisa, and A. Rozi Kurnia, "Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store," *MALCOM*, vol. 2, pp. 47-54, 2022, doi: 10.57152/malcom.v2i1.195.
- [12] B. Wijaya Rauf, "Sentimen Analisis Pertambangan Di Konawe Utara Dengan Metode Naïve Bayes," in *Prosiding Sempatin*, 2023, pp. 1-5. [Online]. Available: <https://t.co/fSdh2dCADm>
- [13] 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.
- [14] A. V. Sudiantoro and E. Zuliarso, "ANALISIS SENTIMEN TWITTER MENGGUNAKAN TEXT MINING DENGAN ALGORITMA NAÏVE BAYES CLASSIFIER," in *SINTAK SEMINAR NASIONAL*, 2018, pp. 69-73.
- [15] 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.
- [16] I. Habib Kusuma and N. Cahyono, "Analisis Sentimen Masyarakat Terhadap Penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor," vol. 8, no. 3, 2023.



- [17] M. H. Al-Areef and K. Saputra, “Analisis Sentimen Pengguna Twitter Mengenai Calon Presiden Indonesia Tahun 2024 Menggunakan Algoritma LSTM,” *SAINTIKOM*, vol. 22, pp. 270–279, 2023, doi: 10.53513/jis.v22i2.8680.
- [18] A. Guterres, Gunawan, and J. Santoso, “Stemming Bahasa Tetun Menggunakan Pendekatan Rule Based,” *Teknika*, vol. 8, no. 2, pp. 142–147, Oct. 2019, doi: 10.34148/teknika.v8i2.224.
- [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] M. N. Fahriza and N. Riza, “ANALISIS SENTIMEN PADA ULASAN APLIKASI CHAT GENERATIVE PRE-TRAINED TRANSFORMER GPT MENGGUNAKAN METODE KLASIFIKASI K-NEAREST NEIGHBOR(KNN) Sitematic Literature Review,” *Jurnal Mahasiswa Teknik Informatika*, vol. 7, no. 2, 2023, doi: 10.36040/jati.v7i2.6767.
- [21] D. Darwis, N. Siskawati, and Z. Abidin, “Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional,” *Jurnal Tekno Kompak*, vol. 15, no. 1, pp. 131–145, 2021, doi: 10.33365/jtk.v15i1.744.
- [22] D. Rustiana and N. Rahayu, “ANALISIS SENTIMEN PASAR OTOMOTIF MOBIL: TWEET TWITTER MENGGUNAKAN NAÏVE BAYES,” *Jurnal SIMETRIS*, vol. 8, no.1, 2017, doi: 10.24176/SIMET.V8I1.841.