

Analisis Sentimen Pengguna Aplikasi Bukalapak di Platform Playstore Menggunakan Metode *Naïve Bayes*

Muhammad Subhan Mahendrasyah^{1,*}, Taqwa Hariguna²

¹ Fakultas Ilmu Komputer, Program Studi Informatika, Universitas Amikom Purwokerto, Jawa Tengah, Indonesia

² Fakultas Ilmu Komputer, Program Studi Sistem Informasi, Universitas Amikom Purwokerto, Jawa Tengah, Indonesia

Email: ¹andraahay00@gmail.com, ²taqwa@amikompurwokerto.ac.id

Email Penulis Korespondensi: andraahay00@gmail.com

Submitted: 09/07/2024; Accepted: 28/07/2024; Published: 09/09/2024

Abstrak—Indonesia, sebagai negara dengan pertumbuhan pengguna internet yang signifikan, mencatat kontribusi ekonomi digital 7,3% terhadap PDB pada tahun 2017, melampaui pertumbuhan ekonomi keseluruhan sebesar 5,1%. Salah satu tantangan utama adalah efisiensi dalam mengelola ulasan pengguna untuk meningkatkan layanan, seperti yang dilakukan oleh aplikasi Bukalapak menggunakan data scraping untuk mengumpulkan 5000 ulasan. Studi ini menggunakan algoritma *Naïve Bayes* untuk menganalisis sentimen ulasan pengguna aplikasi Bukalapak, dengan fokus pada identifikasi pola sentimen positif dan negatif. Tujuannya adalah memperdalam pemahaman tentang persepsi pengguna terhadap layanan Bukalapak dan memberikan dasar untuk pengambilan keputusan strategis dalam meningkatkan pengalaman pengguna dan layanan aplikasi. Algoritma *Naïve Bayes* dalam penelitian ini mencapai tingkat akurasi 67,9%, dengan 13,3% ulasan ditemukan positif dan 86,7% ulasan negatif. Hasil analisis menyoroti pentingnya perbaikan dalam aspek-aspek tertentu layanan Bukalapak, yang dapat mengarah pada pengembangan lebih lanjut untuk meningkatkan kepuasan pengguna. Mayoritas ulasan Bukalapak menunjukkan kekurangan atau kritik terhadap layanannya, yang menyoroti pentingnya peningkatan dalam aspek-aspek tertentu. Model *Naïve Bayes* memberikan pemahaman yang jelas mengenai sentimen pengguna, yang menjadi kunci dalam pengambilan keputusan strategis dan upaya peningkatan pengalaman pengguna di platform Bukalapak. Dengan demikian, penelitian ini memberikan kontribusi penting dalam mengarahkan langkah-langkah perbaikan dan pengembangan lebih lanjut dalam meningkatkan layanan aplikasi Bukalapak serta memahami persepsi pengguna dengan lebih baik.

Kata Kunci: Bukalapak, Playstore, Analisis Sentimen, Ulasan, *Naïve Bayes*

Abstract—Indonesia, as a country with significant growth in internet users, recorded a 7.3% digital economy contribution to GDP in 2017, surpassing the overall economic growth of 5.1%. One of the main challenges is efficiency in managing user reviews to improve services, as done by Bukalapak app using data scraping to collect 5000 reviews. This study uses the *Naïve Bayes* algorithm to analyze the sentiment of Bukalapak app user reviews, focusing on identifying positive and negative sentiment patterns. The goal is to deepen the understanding of user perceptions of Bukalapak services and provide a basis for strategic decision-making in improving user experience and application services. The *Naïve Bayes* algorithm in this study achieved an accuracy rate of 67.9%, with 13.3% of reviews found to be positive and 86.7% of reviews negative. The analysis results highlight the importance of improvements in certain aspects of Bukalapak's services, which can lead to further development to increase user satisfaction. The majority of Bukalapak reviews indicate shortcomings or criticism of its services, which highlights the importance of improvement in certain aspects. The *Naïve Bayes* model provides a clear understanding of user sentiment, which is key in strategic decision-making and efforts to improve user experience on the Bukalapak platform. Thus, this research makes an important contribution in directing further improvement and development steps in enhancing Bukalapak app services as well as better understanding user perceptions.

Keywords: Bukalapak, Playstore, Sentiment Analysis, Reviews, *Naïve Bayes*

1. PENDAHULUAN

Jumlah individu yang menggunakan internet di Indonesia terus meningkat per tahun. Saat ini, di Indonesia berada di deretan ke-empat di dunia dalam hal pertumbuhan populasi pengguna internet, setelah negara China, India, dan Amerika Serikat. Pada tahun 2017, ekonomi digital di Indonesia memberikan kontribusi signifikan terhadap PDB, mencapai 7,3%, meskipun pertumbuhan ekonomi secara keseluruhan hanya sebesar 5,1% [1]. Jumlah pengguna internet yang besar ini menciptakan peluang besar bagi perkembangan ekonomi digital di Indonesia.

Mayoritas pengguna internet di Indonesia sering berbelanja online, dan Bukalapak adalah salah satu marketplace yang paling populer. Banyak pengguna berbagi pengalaman mereka saat bertransaksi di Bukalapak, sehingga feedback dan keluhan dapat disampaikan dengan cepat dan mudah.

Semakin banyak orang berbelanja di Bukalapak, semakin penting peran umpan balik dari pengguna. Mereka sering berbagi pengalaman positif maupun negatif tentang transaksi di sana. Ini membantu pengguna lain dalam membuat keputusan belanja dan memberi Bukalapak masukan berharga untuk meningkatkan layanan mereka[2]. Interaksi ini menciptakan lingkungan yang dinamis dan responsif, di mana suara konsumen memiliki pengaruh besar terhadap perkembangan Bukalapak.

Penelitian ini akan menggunakan teknik data crawling untuk mengumpulkan data dengan tujuan menganalisis pandangan publik terhadap Bukalapak dan mengategorikan opini tersebut menjadi 2 golongan: positif dan negatif. Hasil analisis sentimen diharapkan bisa menghasilkan perspektif yang bernilai bagi pengembang untuk pengembangan dan peningkatan aplikasi [3]. Penelitian ini mempunyai harapan dapat berkontribusi positif dengan meningkatkan mutu layanan, memperbaiki keahlian user, serta meneruskan wawasan yang penting buat beragam pihak terkait pada lingkungan Bukalapak dan Google Play Store [4].

Salah satu isu yang terlihat merupakan banyaknya analisis dan evaluasi dari user sampai menjadi susah untuk mengelompokkan beserta melakukan analisis terhadap kupasan tersebut [5]. Sentimen adalah opini, ulasan, atau interpretasi yang diberikan oleh seseorang terhadap suatu objek [6], berdasarkan observasi mereka dan menunjukkan pandangan pribadi. Ulasan kini dapat membuat keterangan yang penting untuk pengelola maupun pengguna layanan[7].

Penelitian yang dilakukan oleh Irni Di Estetika, Irfan Darmawan, dan Oktariani Nurul Pratiwi pada tahun 2021 melalui judul "Analisis Sentimen Ulasan Pengguna untuk Peningkatan Layanan Menggunakan Algoritma *Naïve Bayes*" menunjukkan bahwa sentimen pengguna Bukalapak dapat terbagi sebagai 2 kelompok: positif dan negatif. Hasil studi ini menunjukkan akurasi sebesar 83%, *Precision* sebesar 82%, *Recall* setara 80,33%, dan F1-score sebesar 80,66%. Data ini selanjutnya ditunjukkan memakai word cloud untuk mengidentifikasi kata-kata yang sering muncul dalam setiap kategori sentimen. Sentimen positif asal pengguna terhadap Bukalapak harus dijaga, sementara sentimen negatif yang menunjukkan ketidakpuasan pelanggan terkait transaksi, pengiriman, dan layanan customer service perlu ditingkatkan oleh Bukalapak.

Penelitian yang dilakukan oleh Deni Wijaya, Rizki Adi Saputra, dan Faldy Irwiensyah pada tahun 2024 dengan topik "Analisis Sentimen Ulasan Aplikasi Samsat Digital Nasional pada Google Playstore Menggunakan Algoritma *Naïve Bayes*" menampilkan bahwa sentimen penggunaan aplikasi SIGNAL di platform Playstore mengkategorikan 2.000 ulasan ke dalam kategori sentimen positif dan negatif. Dari jumlah tersebut, 1.665 ulasan dikategorikan sebagai bersih setelah proses pembersihan dan transformasi data, dan kemudian diberi label secara manual. Evaluasi model *Naïve Bayes* menunjukkan konsistensi setara 63,61%, dengan 307 *True Positive*, 74 *True Negative*, 26 *False Positive*, dan 192 *False Negative*. *Precision* mencapai 92,19% dan *Recall* setara 61,52%. Analisis semuanya memperlihatkan bahwasanya keterangan user cenderung lebih positif terhadap aplikasi SIGNAL, meskipun ada beberapa kutipan negatif. Pada intinya menunjukkan bahwa user lebih suka aplikasi tersebut [9].

Penelitian yang dilaksanakan oleh Nova Rosalina Siahaan, Rosita Yolanda Tiffany, Shandy Roland Evansius Sinaga, Elsa Vio Nauli Br Naibaho, dan Mohammad Irfan Fahmi pada tahun 2023 yang berjudul "Analisis Sentimen Ulasan Aplikasi Media Sosial WhatsApp Menggunakan Metode *Naïve Bayes Classifier*" menampilkan bahwa metode *Naïve Bayes Classifier* efektif dalam menganalisis sentimen kutipan aplikasi media sosial WhatsApp. Dalam pelabelan, ulasan produk diklasifikasikan ke dalam kelompok positif, negatif, dan netral. Pada studi ini, regu peneliti telah melaksanakan pelabelan pada dataset. Pelabelan di dataset dilakukan sebagai berikut: rating < 3 dikategorikan sebagai sentimen negatif, rating = 3 dikategorikan sebagai sentimen netral, dan rating > 3 dikategorikan sebagai sentimen positif. Selain itu, ditambahkan kolom 'Label' yang berisikan sentimen positif, negatif, dan netral. [10].

Penelitian yang dilakukan oleh Irwansyah Saputra, Hanafi Eko Daron, Fachri Amsury, Muhammad Rizki Fahdia, Benni Ramadhan, Anggie Ardiansyah. Yang berjudul "Analisis Sentimen Pengguna Marketplace Bukalapak dan Tokopedia di Twitter Menggunakan Machine Learning" serta mengaplikasikan tiga algoritma antara lain Decision Tree, K-NN, dan *Naïve Bayes*. Tiga metode untuk memperkirakan label dataset. Hasilnya accuracy dari Decision Tree, K-NN, dan *Naïve Bayes* yaitu 82%, 80%, dan 79%. Hasil precision dari Decision Tree, K-NN, dan *Naïve Bayes* menunjukkan Decision Tree 81,95%, K-NN 87,68%, dan *Naïve Bayes* 89,50%. Hasilnya menunjukkan Decision Tree 86%, K-NN 72%, dan *Naïve Bayes* 66%. Maka dapat di tarik kesimpulan algoritma Decision Tree merupakan classifier memiliki performa yang baik untuk diaplikasikan di perhitungan yang memakai dataset media sosial lantaran membuahkan prediksi yang menghasilkan hasil yang lebih tepat dan akurat.

Analisis sentimen adalah metode dalam pengolahan bahasa alami yang bertujuan untuk mengidentifikasi dan menganalisis pendapat atau emosi yang terdapat pada teks. Teknik ini diaplikasikan untuk menentukan apakah suatu pernyataan atau dokumen memiliki muatan sentimen yang positif dan negatif. Dalam analisis sentimen, biasanya diterapkan teknik pemrosesan bahasa alami, pembelajaran mesin, dan algoritma klasifikasi untuk mengidentifikasi kata kunci, frasa, atau konteks yang mengindikasikan sentimen tertentu[12].

Analisis sentimen dapat membantu perusahaan seperti Bukalapak untuk memahami tren dan pola dalam ulasan pengguna dengan menggunakan algoritma pemrosesan bahasa alami *Naïve Bayes*[13], analisis sentimen dapat menguraikan teks ulasan dan menentukan sentimen yang terkandung di dalamnya. Analisis sentimen adalah proses menemukan pendapat pengguna tentang beberapa topik atau teks yang disampaikan. Proses ini menentukan apakah sepotong tulisan bermakna positif dan negatif [14]. Namun faktanya masih ada beberapa Analisis sentimen yang tidak hanya menggunakan metode algoritma *Naïve Bayes* pada penelitian, hal ini dikarenakan ada kesenjangan pada data algoritma yang lain. Pada studi ini, analisis sentimen kutipan user akan dilakukan menerapkan Algoritma *Naïve Bayes*, yaitu salah satu algoritma populer di klasifikasi teks untuk membuktikan apakah algoritma *Naïve Bayes* lebih efektif dari algoritma yang lain atau malah sebaliknya [15].

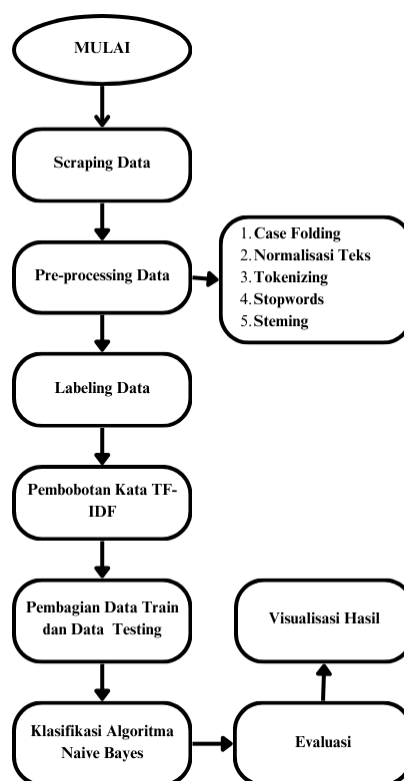
Algoritma *Naïve Bayes* bekerja dengan menggunakan probabilitas untuk menentukan kategori sentimen dari suatu ulasan[16]. Algoritma ini mengasumsikan bahwa fitur-fitur dalam ulasan (seperti kata-kata) adalah independen satu sama lain, sehingga memudahkan dalam perhitungan probabilitas. Meskipun asumsi ini mungkin tidak selalu akurat, *Naïve Bayes* tetap menjadi salah satu algoritma yang efektif dan efisien dalam analisis sentimen. Selain itu algoritma *Naïve Bayes* untuk penanganan sentimen yang berubah dalam waktu nyata karena sentimen dapat berubah seiring waktu terutang respons peristiwa tertentu. Dan penelitian ini diperlukan untuk mengembangkan algoritma yang dapat mengikuti mengantisipasi perubahan.

Dengan penerapan analisis sentimen ini, Bukalapak dapat mengoptimalkan strategi peningkatan layanan berdasarkan masukan langsung dari pengguna, sehingga dapat meningkatkan kepuasan dan loyalitas pelanggan. Selain itu, analisis sentimen juga dapat membantu Bukalapak dalam mengidentifikasi tren dan perubahan dalam persepsi pengguna dalam jangka waktu tertentu, yang dapat dimanfaatkan untuk merumuskan strategi bisnis yang lebih adaptif dan responsif terhadap kebutuhan pasar.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Di studi kali ini, akan dibuat sebuah sistem untuk membeberkan sentiment user Bukalapak. Sistem ini akan menggolongkan kutipan user tentang performa Bukalapak ke dalam 2 kategori, yaitu, negatif dan positif. Data yang diaplikasikan ialah kutipan aplikasi Bukalapak, yang mencakup kutipan positif dan negatif. Konsep penelitian yang digunakan ditunjukkan dalam Gambar 1 berikut.



Gambar 1. Konsep Penelitian

Setelah Gambar 1, terdapat penjelasan yang lebih rinci mengenai setiap tahapan dalam penelitian ini. Oleh karena itu, penjelasan tersebut akan disampaikan dibawah ini:

2.1.1 Scraping Data

Pada tahap awal, dilakukan tahapan scraping data. Scraping data merupakan proses yang dilakukan untuk mengumpulkan data-data ulasan dari Google Play Store dengan menggunakan bahasa pemrograman Python [17]. Pada pencarian data kali ini menggunakan Google Play Scraper untuk mengumpulkan ulasan tentang aplikasi Bukalapak melalui ID yang terdapat pada platform Google Play Store. Program ini dirancang untuk mengumpulkan ulasan berdasarkan bahasa Indonesia dan dari pengguna di Indonesia. Jumlah maksimum ulasan yang diambil adalah sebanyak 5000 ulasan. Pendekatan ini memungkinkan pengumpulan data yang lebih terfokus dan relevan untuk analisis sentimen berdasarkan pengguna di wilayah dan bahasa tertentu.

2.1.2 Pre-processing

Tahap pra-pemrosesan dilakukan untuk mendapatkan data yang bersih guna mempermudah tahap selanjutnya dan menghasilkan analisis yang lebih akurat. Proses ini mencakup berbagai langkah penting yang bertujuan untuk mengganti data mentah menggambarkan format yang terstruktur dan bebas dari ketidakakuratan atau anomali yang dapat mengganggu analisis. prediktif bekerja lebih efektif dan memberikan hasil yang lebih andal. Tahap ini sangat



krusial dalam alur kerja analitik data karena data yang bersih dan terorganisir merupakan dasar dari setiap analisis yang sukses[18]. Langkah-langkah tersebut meliputi:

- a. Pembersihan Data: Menghilangkan nilai yang hilang atau tidak konsisten.
- b. Transformasi Data: Memastikan keseragaman dalam bentuk dan unit pengukuran.
- c. Penghilangan Elemen yang Tidak Relevan atau Redundan: Membuat data lebih fokus dan relevan.

Dengan melakukan pra-pemrosesan ini, kualitas data meningkat secara signifikan. Hal ini memungkinkan algoritma analisis dan model prediktif bekerja lebih efektif dan memberikan hasil yang lebih andal. Tahap ini sangat krusial dalam alur kerja analitik data karena data yang bersih dan terorganisir merupakan dasar dari setiap analisis yang sukses. Berikut merupakan tahapan text Processing :

- a. Case Folding

Case folding ialah suatu tahapan di pengolahan teks yang bertujuan untuk merubah bentuk kata-kata agar seragam dalam hal kapitalisasi, dengan cara mengubah semua huruf menjadi huruf kecil (lower case). Ini penting dalam analisis teks karena memastikan bahwa kata-kata yang sebenarnya sama tidak dianggap berbeda karena perbedaan kapitalisasi. Misalnya, kata "Data" dan "data" dianggap sama setelah proses case folding. Proses ini menjadi langkah awal yang kritis sebelum melakukan tokenisasi dan analisis lebih lanjut pada teks, memastikan konsistensi dan akurasi dalam pengolahan data teks. Case folding adalah salah satu metode yang sederhana namun efektif dalam pra-pemrosesan teks, meskipun sering kali terabaikan. Tujuannya adalah untuk mengubah semua huruf dalam dokumen menjadi huruf kecil [1].

- b. Normalisasi Teks

Normalisasi diaplikasikan untuk mengganti kata non-baku atau slang menjadi bentuk kata baku yang standar. Proses ini memanfaatkan kamus slang untuk mengidentifikasi dan menggantikan kata-kata informal dengan kata-kata yang lebih umum digunakan atau baku. Contohnya, mengubah kata "yg" menjadi "yang" atau "gimana" menjadi "bagaimana". Tujuan normalisasi adalah untuk menyederhanakan teks sehingga lebih mudah diproses dan dipahami oleh algoritma analisis teks, serta menghasilkan hasil yang lebih konsisten dan akurat dalam analisis data [20].

- c. Tokenisasi

Tokenisasi (Tokenizing) adalah langkah dalam pemrosesan teks yang melibatkan pemisahan kalimat atau teks menjadi bentuk-bentuk yang lebih kecil, seperti kata atau simbol tertentu. Proses ini vital dalam analisis teks karena mengharuskan komputer untuk mengerti serta memanipulasi lebih efisien. Setiap unit yang dihasilkan dari tokenisasi disebut sebagai token, yang merupakan entitas terkecil yang dapat diolah dalam analisis teks, seperti kata, angka, atau tanda baca [21].

- d. Stopwords

Setelah proses tokenisasi, langkah berikutnya adalah melakukan filter stopwords. Pada langkah ini, stopwords dihapus guna menghilangkan sebagian kata yang tidak relevan dalam teks [9]. Langkah pertama adalah mengidentifikasi kata-kata yang dianggap sebagai stopwords, yang umumnya mencakup kata-kata umum seperti "dan", "atau", "saya", "kamu", dan sebagainya. Setelah itu, stopwords tersebut dihapus dari teks atau dokumen yang sedang diproses. Dengan menghilangkan stopwords, teks menjadi lebih tertuju terhadap kata kunci atau makna utama dengan relevan dalam menganalisis sebuah teks.

- e. Stemming

Stemming adalah proses menghilangkan imbuhan dari kata sehingga hanya sisa kata dasarnya. Tujuannya adalah untuk mengurangi variasi kata menjadi bentuk dasarnya, yang membantu menghindari ambiguitas atau kemungkinan kata dengan lebih dari satu makna. Proses ini sering digunakan dalam analisis teks dan pemrosesan bahasa alami untuk meningkatkan akurasi dan efisiensi dalam pengolahan teks yang lebih lanjut. Stemming memainkan peran penting dalam mempersiapkan teks untuk analisis lebih lanjut, seperti klasifikasi teks atau pengelompokan dokumen berdasarkan topik [22].

2.1.3 Labeling

Setelah melewati proses pre-processing, langkah berikutnya adalah melakukan pelabelan pada data melalui 2 golongan, ialah positif dan negatif. Labeling dilakukan dengan mencocokkan kalimat dengan kata-kata positif maupun negatif berdasarkan aturan dari KBBI [23]. Proses pelabelan data dijalankan secara otomatis dengan mengaplikasikan kamus kosakata lexicon untuk menghitung nilai poin sentimen. Evaluasi sentimen positif atau negatif pada data merujuk pada kamus kosakata lexicon yang mencakup daftar kata-kata positif dan negatif dalam bahasa Indonesia. Jumlah data yang digunakan dalam lexicon positif adalah 3609, sementara lexicon negatif mencakup 6609 data. Proses ini dilakukan secara otomatis, dan hasil evaluasi akan mengklasifikasikan kalimat ke dalam kategori positif jika skornya > 0 jika tidak kalimat tersebut masuk ke dalam kategori negative.

2.1.4 Pembobotan Kata *TF-IDF*

Langkah berikutnya dalam proses ini adalah pemberian bobot *TF-IDF* (*Term Frequency-Inverse Document Frequency*) kepada kata-kata di dokumen. Langkah ini melibatkan penghitungan frekuensi kemunculan setiap kata dalam dokumen (*Term Frequency*) dan mengalikan hasil tersebut dengan nilai *IDF* (*Inverse Document Frequency*)

yang sesuai. Hal ini membantu mengevaluasi kepentingan relatif dari kata-kata dalam sebuah dokumen terhadap keseluruhan korpus teks, memungkinkan pengidentifikasi kata-kata yang lebih penting atau representatif untuk klasifikasi atau analisis lebih lanjut. *TF-IDF* sering digunakan dalam pengambilan informasi dan sistem temu kembali informasi untuk meningkatkan ketepatan dan relevansi hasil pencarian [24].

$$W_{ij} = \frac{n_{ij}}{\sum_{j=1}^p n_{j,i}} \log_2 \frac{D}{d_j} \quad (1)$$

Rumus (1) ini digunakan untuk menghitung bobot TF-IDF dari istilah ij dalam dokumen iii . Bobot ini menggabungkan dua komponen utama: Term Frequency (TF) dan Inverse Document Frequency (IDF).

2.1.5 Klasifikasi Algoritma *Naïve Bayes*

Pada tahap selanjutnya, menggunakan algoritma *Naïve Bayes*, yang didasarkan pada asumsi kuat tentang independensi setiap fitur atau kata dalam kelas yang diberikan [25]. 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 [11]. Algoritma ini memungkinkan sistem untuk secara efektif memahami dan memprediksi sentimen dari ulasan berdasarkan analisis teks, memberikan dasar yang kuat untuk evaluasi kualitas layanan aplikasi dan respons pengguna.

$$P(H|X) = \frac{P(H) \prod_{i=1}^n P(x_i|H)}{P(X)} \quad (2)$$

Teorema Bayes memperbarui probabilitas hipotesis (H) berdasarkan bukti baru (X) . Probabilitas awal $(P(H))$ mencerminkan keyakinan awal kita. Probabilitas (X) terjadi jika (H) benar diwakili oleh $(P(X|H))$. Probabilitas total $(P(X))$ menormalkan hasil, memastikan validitas probabilitas akhir. Teorema ini membantu memperbarui keyakinan terhadap (H) berdasarkan bukti (X) .

2.1.6 Pembagian Data Train serta Data Testing

Langkah berikutnya, dataset dibagi menjadi 2, antara lain data latih serta data uji, dengan perbandingan 80:20. Data latih digunakan untuk melatih algoritma klasifikasi, yang berfungsi sebagai basis Pengetahuan ini diaplikasikan guna memprediksi kelas data baru yang belum pernah dikenali sebelumnya. Pada Data uji diaplikasikan pada saat menilai keberhasilan klasifikasi oleh classifier serta mengevaluasi performa dan akurasi model yang telah dilatih [26]. Setelah pembagian ini dilakukan, langkah selanjutnya adalah menjalankan proses klasifikasi menggunakan algoritma yang telah dipilih, seperti *Naïve Bayes Multinomial* atau *Naïve Bayes Bernoulli*. Proses ini bertujuan untuk mengidentifikasi pola dan karakteristik dari data latih sehingga model dapat mengklasifikasikan data uji dengan akurasi yang optimal.

2.1.7 Evaluasi

Evaluasi data merupakan langkah akhir dalam proses klasifikasi melalui algoritma *Naïve Bayes*. Di langkah ini, dataset telah dikelompokkan berdasarkan kategori yang telah ditetapkan dengan menerapkan algoritma *Naïve Bayes*, dilakukan evaluasi guna menjadi tolak ukur kualitas hasilnya. Evaluasi ini melibatkan pengukuran berbagai metrik seperti akurasi, presisi, *Recall*, dan *F1-score* diaplikasikan sebagai penilaian guna mengukur performa. Seberapa baik algoritma dapat mengkategorikan ulasan ke dalam golongan positif dan negatif. Hasil evaluasi ini sangat penting guna mendalami kinerja dan kehandalan algoritma dalam memprediksi sentimen ulasan aplikasi Bukalapak berdasarkan data ulasan yang telah diproses.

2.1.8 Visualisasi Hasil

Setelah proses klasifikasi sentimen dari data yang telah diproses, hasil klasifikasi ini akan disajikan. Data hasil klasifikasi dapat digunakan untuk menganalisis ulasan pengguna Aplikasi Bukalapak melalui platform Play Store. Tujuan analisis ini guna memahami persepsi umum user terhadap aplikasi tersebut berdasarkan evaluasi sentimen yang telah dilakukan. Dengan demikian, informasi dari hasil klasifikasi sentimen dapat memberikan wawasan berharga dalam meningkatkan kualitas layanan dan respons terhadap pengguna aplikasi.

3. HASIL DAN PEMBAHASAN

Bab selanjutnya mengeksplorasi hasil dari analisis serta tujuan untuk mengetahui perspektif umum pengguna terhadap aplikasi tersebut tersedia di platform Play Store, melalui metode *Naïve Bayes*. Tujuan studi ini adalah untuk memahami bagaimana pengguna menilai dan merespons pengalaman mereka menggunakan aplikasi Bukalapak, khususnya dalam mengklasifikasikan ulasan-ulasan tersebut sebagai positif atau negatif.



3.1 Scraping Data

Menggunakan alat Google Play Scraper untuk mengambil data secara otomatis dengan melalui scraping data dari Google Play Store merupakan metode yang efektif dan efisien untuk mengumpulkan berbagai jenis informasi aplikasi, seperti detail aplikasi, ulasan pengguna, penilaian, jumlah unduhan, dan informasi penting lainnya. Dengan memanfaatkan alat ini, dapat dengan mudah mengakses dan mengekstrak data dalam jumlah besar secara otomatis, yang kemudian dapat dianalisis untuk memperoleh wawasan berharga mengenai kinerja aplikasi, tren pengguna, dan umpan balik konsumen. Hasil yang di dapatkan dari scraping data pada kali ini adaah 5000 data. Pendekatan ini memungkinkan pengembang, peneliti, dan analis pasar untuk membuat keputusan yang lebih informatif dan strategis berdasarkan data terkini dari salah satu platform distribusi aplikasi. Berikut merupakan hasil dari scraping data seperti pada tabel 1.

Tabel 1. Hasil Scraping Data

| No | Review ID | Komentar | at |
|----|--------------------------------------|--|------------------------|
| 1 | 3846a0dc-bb8c-4ec3-9b03-400802870dc5 | KEREN... belanja gampang ga ribet dan mudah tinggal pilih barang yang kita inginkan,selain jadi pembeli akun kita juga bisa jadi pelapak/penjual dengan muda | 2024-06-02 06:38:49 |
| 2 | cb83375d-2050-4562-99ae-fece0d36ff51 | Lapak seller tiba2 dibekukan, paylater ga bisa digunakan lg. Apk udah ga ada yg bisa diandalkan lg. UNINSTAL! | 2024-06-21 18:48:54 |
| 3 | db99d126-4b32-4b23-953a-68595b41ca4b | Aplikasi yang sangat bagus,mencari barang atau pelatihan sangat mudah dan proses transaksi sangat cepat | 2024-06-04 14:12:53 |
| 4 | 627ef64c-15c7-4e90-b2a7-453b55fa7bfb | Kendala sistem playletter selalu gangguan dalam sistem. Entah sampai baru terselesaikan waktu perbaikanga | 2024-06-01 16:44:09 |
| 5 | 59695c93-4a27-4ba8-97f5-5d8b4b8a2acf | Buat buka lapak trima kasih. selama berbelanja saya merasa lebih mudah dan dimudahkan.sekali lagi trima kasih. | 2024-06-02 16:05:27 |

3.2 Pre-processing

Sebelum memulai proses pengolahan data, langkah awalnya adalah mengidentifikasi dan menghapus kolom-kolom yang dianggap tidak relevan atau tidak diperlukan dalam dataset. Tindakan ini dilakukan untuk membersihkan dataset dari informasi yang tidak akan memberikan kontribusi dalam analisis atau pemodelan yang akan dilakukan selanjutnya. Tujuannya adalah untuk memastikan bahwa data yang akan diproses hanya terdiri dari informasi yang relevan dan sesuai dengan kebutuhan penelitian.

3.2.1 Case Folding

Pada langkah berikutnya, seluruh data komentar yang masih berhuruf kapital akan diubah menjadi huruf kecil untuk memastikan keseragaman dalam data yang hendak diolah. Tahap ini dikenal sebagai case folding, yang merupakan salah satu langkah penting dalam preprocessing data teks. Dalam implementasinya, kami menggunakan *Library Natural Language Toolkit (NLTK)* dan *NLTK* dalam Python. *Library NLTK* digunakan untuk membantu dalam analisis teks, sementara *NLTK* efektif untuk manipulasi data dan membaca dataset. Setiap baris data, yang masing-masing merepresentasikan satu tweet, akan dibaca dan dianalisis. Jika dalam baris data tersebut terdapat huruf kapital, oleh karena itu huruf tersebut akan dirubah jadi huruf kecil. Berikut pada tabel 2 merupakan hasil dari perubahan dari data mentah hasil scraping data ke case folding

Tabel 2. Perubahan dari Data ke Case Folding

| No | Komentar | Case Folding |
|----|--|--|
| 0 | KEREN... belanja gampang ga ribet dan mudah tinggal pilih barang yang kita inginkan,selain jadi pembeli akun kita juga bisa jadi pelapak/penjual dengan muda | keren.. belanja gampang ga ribet dan mudah tinggal pilih barang yang kita inginkan,selain jadi pembeli akun kita juga bisa jadi pelapak/penjual dengan mudah |
| 1 | Lapak seller tiba2 dibekukan, paylater ga bisa digunakan lg. Apk udah ga ada yg bisa diandalkan lg. UNINSTAL! | lapak seller tiba2 dibekukan, paylater ga bisa digunakan lg. apk udah ga ada yg bisa diandalkan lg. uninstal! |



| No | Komentar | Case Folding |
|----|--|--|
| 2 | Aplikasi yang sangat bagus,mencari barang atau pelatihan sangat mudah dan proses transaksi sangat cepat | aplikasi yang sangat bagus,mencari barang atau pelatihan sangat mudah dan proses transaksi sangat cepat |
| 3 | Kendala sistem playletter selalu gangguan dalam sistem. Entah sampai baru terselesaikan waktu perbaikanga | kendala sistem playletter selalu gangguan dalam sistem. entah sampai baru terselesaikan waktu perbaikanga |
| 4 | Buat buka lapak trima kasih. selama berbelanja saya merasa lebih mudah dan dimudahkan.sekali lagi trima kasih. | buat buka lapak trima kasih. selama berbelanja saya merasa lebih mudah dan dimudahkan.sekali lagi trima kasih. |

3.2.2 Normalisasi

Normalisasi teks adalah proses mengubah teks agar memiliki format yang konsisten dan standar sesuai dengan kebutuhan tertentu. Langkah-langkah dalam normalisasi teks termasuk menghapus karakter tidak diinginkan, mengganti kata-kata non-baku dengan kata-kata baku, dan menggabungkan kata-kata yang terpisah. Tujuan dari normalisasi teks adalah untuk meningkatkan konsistensi dan akurasi data serta mempermudah proses analisis data, seperti yang di jelaskan pada Tabel 3.

Tabel 3. Perubahan Dari Case Folding ke Normalisasi

| No | Case Folding | Normalisasi |
|----|--|---|
| 0 | keren.. belanja gampang ga ribet dan mudah tinggal pilih barang yang kita inginkan,selain jadi pembeli akun kita juga bisa jadi pelapak/penjual dengan mudah | keren belanja gampang tidak rumit dan mudah tinggal pilih barang yang kita inginkan selain jadi pembeli akun kita juga bisa jadi pelapak/penjual dengan mudah |
| 1 | lapak seller tiba2 dibekukan, paylater ga bisa digunakan lg. apk udah ga ada yg bisa diandalkan lg. uninstal! | lapak seller tiba2 dibekukan paylater tidak bisa digunakan lagi aplikasi sudah tidak ada yang bisa diandalkan lagi hapus instalasi |
| 2 | aplikasi yang sangat bagus,mencari barang atau pelatihan sangat mudah dan proses transaksi sangat cepat | aplikasi yang sangat bagus mencari barang atau pelatihan sangat mudah dan proses transaksi sangat cepat |
| 3 | kendala sistem playletter selalu gangguan dalam sistem. entah sampai baru terselesaikan waktu perbaikanga | kendala sistem playletter selalu gangguan dalam sistem entah sampai baru terselesaikan waktu perbaikanga |
| 4 | buat buka lapak trima kasih. selama berbelanja saya merasa lebih mudah dan dimudahkan.sekali lagi trima kasih. | buat buka lapak terima kasih selama berbelanja saya merasa lebih mudah dan dimudahkan.sekali lagi terima kasih |

3.2.3 Tokenizing

Tahap ini bertujuan untuk menguraikan teks menjadi segmen-segmen yang spesifik berdasarkan kata-kata, yang dikenal dengan istilah tokenisasi. Proses tokenisasi dilakukan untuk memisahkan string atau membagi kalimat menjadi kata-kata yang bermakna. Dalam langkah ini, penulis memanfaatkan perpustakaan *NLTK* untuk menjalankan proses tokenisasi. Berikut adalah uraian dari normalisasi kemudian di ubah ke tokenizing seperti pada tabel 4

Tabel 4. Perubahan dari Normalisasi ke Tokenizing

| No | Normalisasi | Tokenizing |
|----|---|---|
| 0 | keren belanja gampang tidak rumit dan mudah tinggal pilih barang yang kita inginkan selain jadi pembeli akun kita juga bisa jadi pelapak/penjual dengan mudah | ['keren', 'belanja', 'gampang', 'tidak', 'rumit', 'mudah', 'tinggal', 'pilih', 'barang', 'yang', 'kita', 'inginkan', 'selain', 'jadi', 'pembeli', 'akun', 'kita', 'juga', 'bisa', 'jadi', 'pelapak/penjual', 'dengan', 'mudah'] |
| 1 | lapak seller tiba2 dibekukan paylater tidak bisa digunakan lagi aplikasi sudah tidak ada yang bisa diandalkan lagi hapus instalasi | ['lapak', 'seller', 'tiba2', 'dibekukan', 'paylater', 'tidak', 'bisa', 'digunakan', 'lagi', 'aplikasi', 'sudah', 'yang', 'bisa', 'diandalkan', 'lagi', 'hapus', 'instalasi'] |



| No | Normalisasi | Tokenizing |
|----|--|--|
| | | 'tidak', 'yang', 'bisa', 'diandalkan', 'lagi', 'hapus', 'instalan'] |
| 2 | aplikasi yang sangat bagus mencari barang atau pelatihan sangat mudah dan proses transaksi sangat cepat | ['aplikasi', 'yang', 'sangat', 'bagus', 'mencari', 'barang', 'atau', 'pelatihan', 'sangat', 'mudah', 'proses', 'transaksi', 'sangat', 'cepat'] |
| 3 | kendala sistem playletter selalu gangguan dalam sistem entah sampai baru terselesaikan waktu perbaikanga | ['kendala', 'sistem', 'playletter', 'selalu', 'gangguan', 'dalam', 'sistem', 'entah', 'sampai', 'baru', 'terselesaikan', 'waktu', 'perbaikanga'] |
| 4 | buat buka lapak terima kasih selama berbelanja saya merasa lebih mudah dan dimudahkan.sekali lagi terima kasih | ['buat', 'buka', 'lapak', 'terima', 'kasih', 'selama', 'berbelanja', 'saya', 'merasa', 'lebih', 'mudah', 'dimudahkan.sekali', 'lagi', 'terima', 'kasih'] |

3.2.4 Stopwords

Pada proses ini, tujuannya adalah untuk meningkatkan konsentrasi pada kata-kata yang lebih informatif dalam proses analisis atau pemrosesan teks. Langkah pertama adalah mengidentifikasi kata-kata yang dianggap sebagai Stopwords, yang umumnya termasuk kata-kata umum seperti "dan", "atau", "saya", "kamu", dan sebagainya. Setelah itu, Stopwords dihapus dari teks atau dokumen yang sedang diolah.

Tabel 5. Pengubahan dari Tokenizing ke Stopwords

| No | Tokenizing | Stopwords |
|----|---|--|
| 0 | ['keren', 'belanja', 'gampang', 'tidak', 'rumit', 'mudah', 'tinggal', 'pilih', 'barang', 'yang', 'kita', 'inginkan', 'selain', 'jadi', 'pembeli', 'akun', 'kita', 'juga', 'bisa', 'jadi', 'pelapak/penjual', 'dengan', 'mudah'] | keren,belanja,gampang,tidak,rumit,mudah, tinggal,pilih,barang,yang,kita,inginkan, selain,jadi,pembeli,akun,kita,juga,bisa,jadi, pelapak/penjual,dengan,mudah |
| 1 | ['lapak', 'seller', 'tiba2', 'dibekukan', 'paylater', 'tidak', 'bisa', 'digunakan', 'lagi', 'aplikasi', 'sudah', 'tidak', 'yang', 'bisa', 'diandalkan', 'lagi', 'hapus', 'instalan'] | lapak,seller,tiba2,dibekukan,paylater, tidak,bisa,digunakan,lagi,aplikasi,sudah, tidak,yang,bisa,diandalkan,lagi,hapus,instalan |
| 2 | ['aplikasi', 'yang', 'sangat', 'bagus', 'mencari', 'barang', 'atau', 'pelatihan', 'sangat', 'mudah', 'proses', 'transaksi', 'sangat', 'cepat'] | aplikasi,bagus,mencari,barang,pelatihan, mudah,proses,transaksi,cepat |
| 3 | ['kendala', 'sistem', 'playletter', 'selalu', 'gangguan', 'dalam', 'sistem', 'entah', 'sampai', 'baru', 'terselesaikan', 'waktu', 'perbaikanga'] | kendala,sistem,playletter,gangguan,sistem terselesaikan,perbaikanga |
| 4 | ['buat', 'buka', 'lapak', 'terima', 'kasih', 'selama', 'berbelanja', 'saya', 'merasa', 'lebih', 'mudah', 'dimudahkan.sekali', 'lagi', 'terima', 'kasih'] | buka,lapak,terima,kasih,berbelanja,mudah, dimudahkan.sekali,terima,kasih |

3.2.5 Stemming

Tahapan Stemming merupakan proses penghapusan imbuhan dari suatu kata, sehingga hanya kata dasarnya yang tersisa. Tujuan dari Stemming adalah untuk mengurangi variasi kata menjadi bentuk dasarnya, yang dapat membantu mengurangi ambiguitas atau kemungkinan adanya kata yang memiliki lebih dari satu makna. Pada Tabel 6 merupakan hasil pengubahan stopwords kemudian di ubah ke stemming.

Tabel 6. Pengubahan dari Stopwords ke Stemming

| No | Stopwords | Stemming |
|----|---|---|
| 0 | keren,belanja,gampang,tidak,rumit,mudah, tinggal, pilih,barang,yang,kita,inginkan, selain,jadi,pembeli,akun,kita,juga,bisa,jadi, pelapak/penjual,dengan,mudah | keren belanja gampang rumit mudah tinggal pilih barang beli akun lapak jual mudah |
| 1 | lapak,seller,tiba2,dibekukan,paylater, tidak,bisa,digunakan,lagi,aplikasi,sudah, tidak,yang,bisa,diandalkan,lagi,hapus,instalan | lapak seller tiba2 beku paylater aplikasi andal hapus instaln |

| No | Stopwords | Stemming |
|----|---|--|
| 2 | aplikasi,bagus,mencari,barang,pelatihan, mudah,proses,transaksi,cepat | aplikasi bagus cari barang latih mudah proses transaksi cepat |
| 3 | kendala,sistem,playletter,gangguan,sistem terselesaikan,perbaikanga | kendala sistem playletter ganggu sistem selesai perbaikanga |
| 4 | buka,lapak,terima,kasih,berbelanja,mudah, dimudahkan.sekali,terima,kasih | buka lapak terima kasih berbelanja mudah dimudahkan sekali terima,kasih |

3.3 Labeling

Setelah melewati proses pre-processing, langkah berikutnya adalah melakukan pelabelan pada data dengan 2 kategori, yaitu positif dan negatif. Proses pelabelan data dilakukan secara otomatis dengan menggunakan kamus kosakata *lexicon* untuk menghitung nilai poin. Evaluasi sentimen positif atau negatif pada data merujuk pada kamus kosakata *lexicon* yang mencakup daftar kata-kata positif dan negatif dalam bahasa Indonesia. Jumlah data yang digunakan dalam *Lexicon* positif adalah 3609, sementara *lexicon* negatif mencakup 6609 data. Berikut pada tabel 7 terdapat uraian labeling data

Tabel 7. Labeling Data

| No | Stemming | Labeling |
|----|---|----------|
| 0 | keren belanja gampang rumit mudah tinggal pilih barang beli akun lapak jual mudah | Positif |
| 1 | lapak seller tiba2 beku paylater aplikasi andal hapus instalasi | Negatif |
| 2 | aplikasi bagus cari barang latih mudah proses transaksi cepat | Positif |
| 3 | kendala sistem playletter ganggu sistem selesai perbaikanga | Negatif |
| 4 | buka lapak terima kasih berbelanja mudah dimudahkan sekali terima,kasih | Positif |



Gambar 2. Worldcloud Ulasan Paositif



Gambar 3. Worldcloud Ulasan Negative

Pada Gambar 2 dan Gambar 3 merupakan cara visual untuk menampilkan kata-kata yang sering muncul dalam sebuah teks. Dalam analisis sentimen, word cloud membantu kita melihat kata-kata yang sering digunakan dalam ulasan atau komentar. Kata-kata yang lebih sering muncul akan ditampilkan lebih besar, sehingga kita bisa dengan mudah melihat tema atau perasaan dominan dalam teks tersebut. Misalnya, jika banyak orang mengatakan "aplikasi" atau "terima kasih" kata-kata tersebut akan tampak lebih besar di word cloud, menunjukkan sentimen positif. Sebaliknya, jika banyak yang menggunakan kata "aplikasi" atau "komplain," kata-kata itu akan lebih besar, menunjukkan sentimen negatif.

3.4 Pembobotan TF-IDF

Output ini menampilkan hasil perhitungan TF-IDF (Term Frequency-Inverse Document Frequency) untuk sebuah dokumen. Pustaka Python diaplikasikan untuk mengkalkulasi nilai IDF (Inverse Document Frequency) dari setiap

kata di dokumen tersebut, dan hasilnya disajikan dalam bentuk array. Nilai *IDF* sangat penting dalam mengevaluasi signifikansi sebuah kata di dokumen, terutama di konteks seluruh kumpulan dokumen atau korpus.

```
[8.82424599 8.82424599 8.82424599 ... 8.82424599 8.82424599 8.82424599]
(0, 7011)    0.12269512624030293
(0, 6804)    0.10127503176470962
(0, 6621)    0.1813771971550581
```

Gambar 4. Pembobotan Kata *TF-IDF*

3.5 Pembagian Data latih serta Data Uji

Studi ini melibatkan pembagian dataset menjadi 2 kelompok, yaitu data Training serta data Testing. Dalam konteks ini, penulis memilih perbandingan 80:20 untuk pembagian data training serta testing. Pemilihan rasio ini didasarkan pada ukuran dataset yang besar dalam penelitian ini dan bertujuan untuk mengurangi risiko overfitting, dimana model pembelajaran mesin menjadi terlalu khusus dalam mempelajari data training sehingga performa model kurang optimal saat diuji dengan data baru. Seperti seperti yang telah di paparkan pada gambar 5 merupakan pembagian Data Train 4000 dan Data Test 1000.

```
Jumlah x train adalah 4000
Jumlah x test adalah 1000
Jumlah y train adalah 4000
Jumlah y tes adalah 1000
```

Gambar 5. Pembagian Data Train dan Data Testing

Pernyataan tersebut menjelaskan jumlah data yang terdapat dalam set pelatihan dan pengujian, serta memastikan bahwa jumlah targetnya sesuai. Secara spesifik, terdapat 4000 sampel dalam set pelatihan dan 1000 sampel dalam set pengujian, untuk ke2 variabel X dan Y. Ini menunjukkan bahwa ukuran dataset telah memadai dan siap digunakan untuk proses pelatihan dan evaluasi model.

3.6 Klasifikasi Algoritma *Naïve Bayes*

```
nb = MultinomialNB()
nb.fit(X_train,y_train)

nb_prediction = nb.predict(X_test)
accuracy_score(nb_prediction,y_test)

0.679
```

Gambar 6. Akurasi Algoritma *Naïve Bayes*

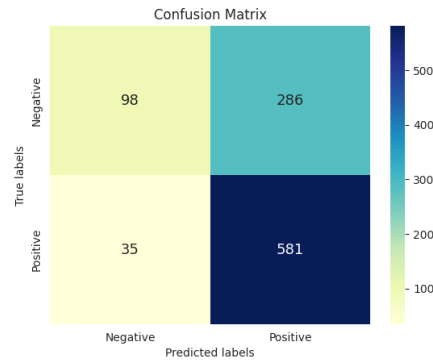
Pada Gambar 6 menunjukkan nilai klasifikasi dari algoritma *Naïve Bayes* tercatat sebesar 0,679, yang mencerminkan tingkat akurasi sebesar 71,14%. Model ini mampu mengklasifikasikan data dengan tepat sekitar 67,9% dari waktu, sesuai dengan nilai yang diperoleh dari fungsi *accuracy_score*. Dan Pada Gambar 7 di jelaskan bahwa hasil klasifikasi algoritma menunjukkan bahwa sentimen positif sebesar 13,3%, sementara sentimen negatif mencapai 86,7%. Selain itu, terdapat selisih sebesar -73,4% antara ke2nya

```
Number of sentiment after modelling:
percentage of positive sentiment: 13.3 %
percentage of negative sentiment: 86.7 %
difference between percentage positive and negative: -73.4 %
```

Gambar 7. Presentase Hasil Sentiment dengan NBC

3.7 Evaluasi

Evaluasi ini bertujuan untuk memahami skenario dan proses optimal dalam kinerja algoritma. Kinerja model dievaluasi dengan menggunakan matriks kebingungan. Hasil eksekusi model diimplementasikan dan dipresentasikan melalui heatmap, tabel, serta langkah-langkah perhitungan. Berikut ini adalah penilaian terhadap klasifikasi algoritma *Naïve Bayes*. Ditahap klasifikasi menggunakan algoritma *Naïve Bayes*, penulis menyusun scenario beserta penggunaan sampel data yang terbagi menjadi 80% data latih serta 20% data uji. Total jumlah data uji yang diaplikasikan adalah 1000 data.



Gambar 8. Confusion Matrix

Pada Gambar 8 merupakan hasil dari Confusion Matrix, hasil prediksi menunjukkan bahwa sebagian besar data telah diprediksi dengan akurat sesuai dengan nilai aktualnya. Berikut adalah penjelasan dari confusion matrix di atas. Ditemukan 98 *True Negative* (TN), 581 *True Positive* (TP), 286 *False Positive* (FP), dan 35 *False Negative* (FN) dalam analisis ini, dengan total 402 data, menggunakan perbandingan 80:20. Elemen-elemen dalam matriks kebingungan dapat dijelaskan sebagai berikut. *True Negative* (TN) adalah jumlah data yang sebenarnya negatif dan diprediksi sebagai negatif. *True Positive* (TP) adalah jumlah data yang sebenarnya positif dan diprediksi sebagai positif. *False Positive* (FP) adalah jumlah data yang sebenarnya negatif melainkan diprediksi sebagai positif. *False Negative* (FN) adalah jumlah data yang sebenarnya positif melainkan diprediksi sebagai negatif.

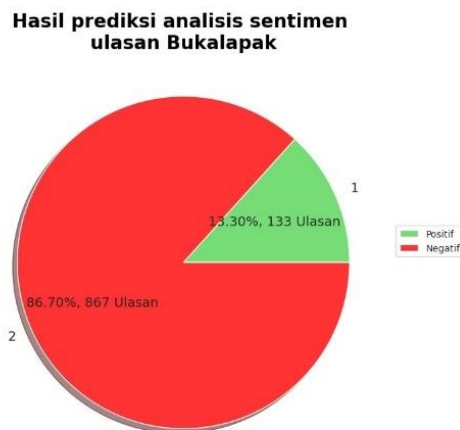
```
Accuracy model: 0.679
Recall model: 0.679
Precision model: 0.696
```

Gambar 9. Hasil Model Klasifikasi Algoritma NBC

Pada Gambar 9 merupakan hasil di mana Accuracy Model, Recall, dan presisi dihitung dengan memanfaatkan fungsi `accuracy_score`, `Recall_score`, dan `Precision_score` dari pustaka sklearn. Fungsi-fungsi tersebut memerlukan 2 parameter sebagai masukan, yakni label sebenarnya (`y_test`) dan hasil prediksi (`nb_prediction`). Setelahnya, skor akurasi, Recall, dan presisi akan dicetak ke konsol dengan 2 angka di belakang koma. Dan hasilnya yaitu Accuracy Model 0.679%; Recall Model 0.679%; Dan Precision Model 0.696%.

3.8 Visualisasi Hasil

Langkah akhir dalam penelitian ini adalah memperlihatkan hasil klasifikasi sentimen yang telah dilakukan menggunakan metode algoritma, menunjukkan 133 ulasan positif (13,30%) dan 867 (86,70%) ulasan negative. Berikut adalah hasilnya.



Gambar 10. Chart Visualisasi Sentimen NBC

Pada Gambar 10 merupakan hasil dari penelitian yang dilakukan dan didapatkan hasil kesimpulan, bahwa penulis menganalisis sentiment ulasan pada platform Playstore menggunakan algoritma Naive Bayes. Didapatkan hasil sentiment sebanyak 133 ulasan positif (13,30%) dan 867 (86,70%).

4. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan, dapat disimpulkan bahwa penerapan analisis sentimen dengan model *Naïve Bayes* efektif dalam menginterpretasikan dan memahami ulasan aplikasi Bukalapak. Algoritma *Naïve Bayes* yang diterapkan dalam studi ini berhasil mencapai tingkat akurasi sebesar 67,9%. Dari hasil analisis sentimen, ditemukan bahwa dari total ulasan yang dianalisis, terdapat 133 ulasan positif, yang menyumbang 13,3% dari keseluruhan ulasan. Sementara itu, ulasan negatif lebih banyak, yaitu sebanyak 867 ulasan atau 86,7% dari total ulasan yang diolah. Hal ini menunjukkan bahwa sentimen negatif lebih dominan dalam data yang diteliti, dengan selisih sebesar 73,4% antara ulasan positif dan negatif. Temuan ini mengindikasikan bahwa mayoritas pengguna aplikasi Bukalapak memberikan ulasan yang kurang memuaskan atau kritis terhadap layanan yang mereka terima. Analisis ini memberikan wawasan yang berharga bagi pengembang aplikasi Bukalapak untuk memahami persepsi pengguna dan mengidentifikasi area yang memerlukan perbaikan.

REFERENCES

- [1] A. A. Tanjung, M. Syafii, S. B. Tarigan, and W. G. Harahap, "Analisis Pengaruh Ekonomi Digital Terhadap Pertumbuhan Ekonomi di Indonesia: Model Data Panel," *ekuitas*, vol. 4, no. 2, pp. 567–575, Dec. 2022, doi: 10.47065/ekuitas.v4i2.2223.
- [2] E. M. Asih, "Analisis pada Shopee sebagai E-Commerce Terpopuler di Indonesia," *jeba*, vol. 2, no. 1, pp. 73–79, Jun. 2024.
- [3] I. Verawati and S. N. Jaelani, "Analisis Sentimen Pengguna Twitter Terhadap Bus Listrik Menggunakan Naïve Bayes," *MIB*, vol. 8, no. 2, 2024, doi: 10.30865/mib.v8i2.7030.
- [4] D. S. Sayogo, B. Irawan, and A. Bahtiar, "ANALISIS SENTIMEN ULASAN BUKALAPAK DI GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA NAÏVE BAYES," *JNANALOKA*, vol. 7, no. 6, 2023. doi: 10.36802/jnanaloka
- [5] I. Darmawan and O. Nurul Pratiwi, "ANALISIS SENTIMEN ULASAN PRODUK TOKO ONLINE RUBYLICIOUS UNTUK PENINGKATAN LAYANAN MENGGUNAKAN ALGORITMA NAIVE BAYES," vol. 7, no. 2, 2020. Availabel: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/12679>
- [6] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, "Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM)," *JJEEE*, vol. 5, no. 1, pp. 32–35, Jan. 2023, doi: 10.37905/jjee.v5i1.16830.
- [7] Z. A. Nurdiansah, "Sentiment Analysis of Reviews on Lazada Apps using Naïve Bayes Algorithm," vol. 8, 2024, <http://dx.doi.org/10.30865/mib.v8i1.7255>.
- [8] I. Darmawan and O. Nurul Pratiwi, "ANALISIS SENTIMEN ULASAN PRODUK TOKO ONLINE RUBYLICIOUS UNTUK PENINGKATAN LAYANAN MENGGUNAKAN ALGORITMA NAIVE BAYES (BUKALAPAK)," vol. 7, no. 2, 2020. Availabel: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/12679>
- [9] D. Wijaya, R. A. Saputra, and F. Irwiensyah, "Analisis Sentimen Ulasan Aplikasi Samsat Digital Nasional Pada Google Playstore Menggunakan Algoritma Naïve Bayes," *KLIK*, vol. 4, no.4, 2024, doi: <https://doi.org/10.30865/klik.v4i4.1738>.
- [10] N. R. Siahaan, R. Y. Tiffany, S. R. E. Sinaga, V. N. B. Naibaho, and M. I. Fahmi, "ANALISIS SENTIMEN ULASAN APLIKASI MEDIA SOSIAL WHATSAPP MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER," *BETRIK*, vol. 14, no. 22, 2023, <https://doi.org/10.36050/betrik.v14i02%20AGUSTUS.104>.
- [11] K. Anwar, "Analisa sentimen Pengguna Bukalapak Di Indonesia Pada Review Smartphone Menggunakan Naïve Bayes," *KLIK*, vol. 2, no. 4, pp. 148–155, Feb. 2022, doi: 10.30865/klik.v2i4.315.
- [12] V. Alviani, S. Alam, and I. Kurniawan, "ANALISIS SENTIMEN REVIEW APLIKASI WETV PADA PLATFORM TWITTER MENGGUNAKAN SUPPORT VECTOR MACHINE," *STORAGE: Jurnal Ilmiah Teknik dan Ilmu Komputer*, vol. 2, no. 3, pp. 143–149, Aug. 2023, doi: 10.55123/storage.v2i3.2351.
- [13] T. A. Azzahra *et al.*, "Perbandingan Efektivitas Naïve Bayes dan SVM dalam Menganalisis Sentimen Kebencanaan di Youtube," *MIB*, vol. 8 no. 1, 2024, doi: 10.30865/mib.v8i1.7186.
- [14] S. M. Siroj, I. Arwani, and D. E. Ratnawati, "Analisis Sentimen Opini Publik pada Twitter terhadap Efek Pembelajaran Daring di Universitas Brawijaya menggunakan Metode K-Nearest Neighbor," *JPTIUK*, vol. 5, no. 7, 2021. Available: <http://j-ptiik.ub.ac.id>.
- [15] T. Taslim, S. Handayani, and F. Fajrizal, "Kinerja Komparatif Optimasi Algoritma Naïve Bayes dalam Klasifikasi Teks untuk Uji Klinis Kanker," *eksplora*, vol. 13, no. 1, pp. 113–123, Sep. 2023, doi: 10.30864/eksplora.v13i1.994.
- [16] D. Darwis, N. Siskawati, and Z. Abidin, "PENERAPAN ALGORITMA NAÏVE BAYES UNTUK ANALISIS SENTIMEN REVIEW DATA TWITTER BMKG NASIONAL," *JTK*, vol. 15, no. 1, p. 131, Feb. 2021, doi: 10.33365/jtk.v15i1.744.
- [17] N. Agustina, D. H. Citra, W. Purnama, C. Nisa, and A. R. Kurnia, "Implementasi Algoritma Naïve Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store: The Implementation of Naïve Bayes Algorithm for Sentiment Analysis of Shopee Reviews On Google Play Store," *MALCOM*, vol. 2, no. 1, pp. 47–54, Apr. 2022, doi: 10.57152/malcom.v2i1.195.
- [18] M. S. Simanjuntak and N. Damanik, "Performance Analysis Of Support Vector Machine In Identifying Comments And Ratings On E-Commerce," *International Journal of Basic and Applied Science*, vol. 11 no. 1, 2022. doi.org/10.35335/ijobas.v11i1.79
- [19] M. I. Syafii, "Sentimen Analisis Pada Media Sosial Menggunakan Metode Naïve Bayes Classifier," *Jurnal Teknologi Pintar*, vol. 3, no. 2, 2023.
- [20] A. M. Simarmata, A. Z. Putra, and A. M. Husein, "Penerapan Metode Computer Vision Dalam Klasifikasi Buah Jeruk Menggunakan Teknik Image Pre-Processing," *DSI*, vol. 3, no. 2, 2023, doi: 10.47709/dsi.v3i2.4010.
- [21] T. Jamaluddin, M. A. Bijaksana, and I. Asror, "Perbandingan Algoritma Sentencepiece BPE dan Unigram Pada Tokenisasi Artikel Bahasa Indonesia," *eProceedings of Engineering*, vol. 7, no. 2, 2020.
- [22] W. Rifai and E. Winarko, "Modification of Stemming Algorithm Using A Non Deterministic Approach To Indonesian Text," *Indonesian J. Comput. Cybern. Syst.*, vol. 13, no. 4, p. 379, Oct. 2019, doi: 10.22146/ijccs.49072.



- [23] B. Wijaya Rauf, “Sentimen Analisis Pertambangan Di Konawe Utara Dengan Metode Naïve Bayes,” *Prosiding Seminar Nasional Pemanfaatan Sains Dan Teknologi Informasi*, vol. 1, no. 1, 2023, pp. 1–5. [Online]. Available: <https://t.co/fSdh2dCADm>.
- [24] R. Situmorang, U. M. Husni Tamyis, and L. S. Andar Muni, “Analisis Sentimen Destinasi Wisata di JawaBarat Pada Twitter Menggunakan Algoritma Naïve Bayes Classifier,” *Simtek: j. sist. inf. dan Teknik kompûter.*, vol. 8, no. 2, pp. 339–342, Oct. 2023, doi: 10.51876/simtek.v8i2.287.
- [25] Tania Puspa Rahayu Sanjaya, Ahmad Fauzi, and Anis Fitri Nur Masruriyah, “Analisis sentimen ulasan pada e-commerce shopee menggunakan algoritma Naïve Bayes dan support vector machine,” *infotech*, vol. 4, no. 1, pp. 16–26, Jun. 2023, doi: 10.37373/infotech.v4i1.422.
- [26] G. Gumelar, Q. Ain, R. Marsuciati, S. A. Bambang, A. Sunyoto, and S. Mustafa, “Kombinasi Algoritma Sampling dengan Algoritma Klasifikasi untuk Meningkatkan Performa Klasifikasi Dataset Imbalance,” *Prosiding SISFOTEK*, vol. 5, no. 1, 2021. Available: <https://seminar.iaii.or.id/index.php/SISFOTEK/article/view/295>.