

Analisis Sentimen Aplikasi Shopee di *Goole Play Store* Menggunakan Klasifikasi Algoritma *Naïve Bayes*

Moch. Rifki Firdaus^{1*}, Nining Rahaningsih², Raditya Danar Dana³

¹ Program Studi Teknik Informatika, STMIK IKMI Cirebon

^{2,3} Program Studi Teknik Informatika, STMIK IKMI Cirebon

mochrifkifirdaus978@gmail.com

Abstrak

Melalui Teknologi Informasi (IT), internet sekarang dimanfaatkan untuk mendorong aktifitas bisnis dan pasar. Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap ulasan aplikasi Shopee di platform Google Play Store dengan menerapkan metode klasifikasi algoritma *Naïve Bayes*. Sebagai salah satu aplikasi *e-commerce* terkemuka, Shopee sangat bergantung pada umpan balik pengguna untuk terus meningkatkan kualitas layanannya. Pemahaman mendalam terhadap sentimen pelanggan yang terungkap dalam ulasan aplikasi dapat menjadi landasan untuk perbaikan dan peningkatan berkelanjutan. Metode klasifikasi algoritma *Naïve Bayes* diimplementasikan untuk mengkategorikan sentimen dalam ulasan pelanggan pada aplikasi Shopee di Google Play Store. Data yang digunakan melibatkan sejumlah besar ulasan yang mencakup berbagai aspek aplikasi. Proses analisis sentimen melibatkan tahap *preprocessing* data, ekstraksi fitur, dan pelatihan mode *Naïve Bayes*. Hasil penelitian ini memberikan gambaran tentang cara pengguna merespons aplikasi Shopee, mengidentifikasi pola sentimen positif, netral dan negatif yang mungkin mempengaruhi reputasi dan popularitas aplikasi. Temuan ini dapat menjadi pedoman bagi pengembang aplikasi untuk fokus pada aspek-aspek tertentu yang mempengaruhi kepuasan pengguna, selain itu, penerapan metode klasifikasi algoritma *Naïve Bayes* membuktikan keandalannya dalam mengidentifikasi sentimen pelanggan secara efisien, memberikan kontribusi pada pemahaman mendalam terkait penggunaan algoritma ini dalam konteks analisis sentimen pada platform aplikasi.

Kata kunci: Analisis Sentimen, Aplikasi Shopee, Google Play Store, Klasifikasi Algoritma, *Naïve Bayes*

Abstract

Through Information Technology (IT), the internet is now used to encourage business and market activities. This research aims to conduct sentiment analysis of Shopee application reviews on the Google Play Store platform by applying the *Naïve Bayes* algorithm classification method. As one of the leading *e-commerce* applications, Shopee relies heavily on user feedback to continuously improve the quality of its services. A deep understanding of customer sentiment revealed in app reviews can be the foundation for continuous improvement and improvement. The *Naïve Bayes* algorithm classification method was implemented to categorize sentiment in customer reviews on the Shopee application on the Google Play Store. The data used involves a large number of reviews covering various aspects of the application. The sentiment analysis process involves data *preprocessing*, feature extraction, and *Naïve Bayes* mode training. The results of this research provide an overview of how users respond to the Shopee application, identifying positive, neutral and negative sentiment patterns that may influence the application's reputation and popularity. These findings can be a guide for application developers to focus on certain aspects that influence user satisfaction. In addition, the application of the *Naïve Bayes* algorithm classification method proves its reliability in identifying customer sentiment efficiently, contributing to a deep understanding regarding the use of this algorithm in the context of analysis sentiment on the application platform.

Keywords: Sentiment Analysis, Shopee Application, Google Play Store, Algorithm Classification, *Naïve Bayes*

PENDAHULUAN

Dengan pesatnya perkembangan teknologi informasi, transformasi digital berdampak pada banyak bidang kehidupan, termasuk bisnis dan *e-commerce*. Sebagai bagian integral dari revolusi ini, aplikasi *e-*

commerce telah menjadi pilar utama yang memfasilitasi interaksi antara konsumen dan penyedia layanan. Salah satu perwakilan terkemuka dalam dunia *e-commerce*, Shopee adalah platform *e-commerce* yang memainkan peran penting dalam ekosistem digital. Dalam konteks ini, penelitian ini bertujuan untuk

menyelidiki sentimen pengguna terhadap aplikasi Shopee, dengan fokus pada ulasan yang dipublikasikan di *Google Play Store*. Sebagai pusat distribusi utama aplikasi seluler, *Google Play Store* adalah saluran utama bagi pengguna untuk mendapatkan wawasan nyata tentang pengalaman aplikasi Shopee. Penelitian ini menerapkan taksonomi algoritma *Naïve Bayes* untuk memahami bagaimana reaksi pengguna terhadap aplikasi Shopee dan bagaimana analisis sentimen membantu dalam perbaikan dan peningkatan berkelanjutan dalam lingkungan kompetitif *e-commerce*. Saya mencoba memahaminya lebih dalam. Oleh karena itu, penelitian ini bertujuan untuk berkontribusi pada pemahaman tentang dinamika antara pengguna dan aplikasi *e-commerce* serta relevansinya mengingat tuntutan konstan akan inovasi dan kepuasan pelanggan.

Mengelola dan memahami sentimen pelanggan menjadi semakin penting karena aplikasi *e-commerce* menjadi perantara utama antara konsumen dan produk. Dalam konteks aplikasi Shopee *Google Play Store*, mengidentifikasi masalah sentimen pelanggan adalah penting untuk meningkatkan kualitas layanan dan menjaga kepuasan pengguna. Permasalahan ini mencakup perbedaan besar dalam ulasan pelanggan, dari tanggapan positif hingga negatif, yang dapat menimbulkan tantangan dalam menganalisis dan memahami preferensi pengguna dengan cepat dan akurat. Selain itu, dengan bertambahnya jumlah pengguna aplikasi *e-commerce*, banyaknya ulasan yang meningkat setiap harinya menjadi sulit untuk diproses secara manual. Tantangan lainnya adalah kemampuan untuk mengenali emosi dan nuansa kompleks dalam ulasan yang sering kali tidak ditangani oleh pendekatan tradisional. Oleh karena itu, penelitian ini menerapkan metode klasifikasi algoritma *Naïve Bayes* untuk memberikan solusi potensial guna meningkatkan efisiensi dan akurasi analisis sentimen, serta mengamati dengan cermat reaksi pengguna terhadap aplikasi digital *all-in-one* Shopee yang selalu berubah. Dengan memahami dan mengatasi permasalahan tersebut, diharapkan penelitian ini dapat memberikan wawasan berharga untuk mengembangkan aplikasi *e-commerce* yang lebih responsif dan memenuhi kebutuhan pengguna.

Meskipun studi sebelumnya di bidang analisis sentimen aplikasi *e-commerce* telah memberikan dasar yang berharga, masih ada celah pengetahuan yang perlu dijelajahi lebih lanjut. Menurut riset yang dilakukan oleh Chen dan Li (2019), yang membahas analisis sentimen terkait aplikasi *commerce* secara umum, belum secara spesifik mengeksplorasi aplikasi Shopee di *Google Play Store*. Penelitian ini mengindikasikan bahwa fluktuasi sentimen pelanggan dapat memberikan bernilai. Namun, kekurangan penelitian ini terletak pada ketidakfokusan pada metode dapat meningkatkan akurasi dan efisiensi analisis.

Tujuan utama dari penelitian ini adalah untuk mengetahui dan menganalisis sentimen pengguna terhadap aplikasi Shopee di *Google Play Store* menggunakan taksonomi Algoritma *Naive Bayes*. Penelitian ini bertujuan untuk memahami dinamika kompleks ulasan pengguna, mengidentifikasi pola sentimen positif dan negatif, serta memberikan wawasan mendalam mengenai preferensi dan kebutuhan pengguna terhadap aplikasi Shopee. Oleh karena itu, penelitian ini bertujuan untuk mengisi kesenjangan literatur yang ada khususnya dalam analisis sentimen aplikasi-aplikasi Shopee, sekaligus berkontribusi pada pemahaman yang lebih luas tentang interaksi antara pengguna dan platform *e-commerce* dalam konteks digital. Signifikansi penelitian ini sangat penting karena memberikan gambaran detail mengenai kepuasan pengguna terhadap aplikasi Shopee. Hal ini dapat digunakan oleh pengembang dan pemangku kepentingan untuk meningkatkan layanan dan pengalaman pengguna kami. Lebih lanjut, penelitian ini dapat berkontribusi pada pengembangan metode klasifikasi algoritma *Naive Bayes* dalam analisis sentimen, sehingga meningkatkan pemahaman kita tentang efektivitas algoritma tersebut dalam konteks aplikasi *e-commerce*. Dengan mempertimbangkan aspek-aspek tersebut, penelitian ini memberikan wawasan praktis dan akademis yang berharga yang tidak hanya memperkaya literatur tetapi juga berkontribusi terhadap pengembangan aplikasi *e-commerce* yang lebih mudah beradaptasi dan sesuai dengan kebutuhan pengguna.

Untuk mencapai tujuan analisis sentimen review aplikasi Shopee di *Google Play Store*, penelitian ini mengadopsi metode klasifikasi

algoritma *Naïve Bayes*. Langkah pertama adalah mengumpulkan *review* aplikasi Shopee dalam jumlah besar di *Google Play Store*. Data validasi kemudian menjalani tahap prapemrosesan yang mencakup pembersihan, normalisasi, dan pembuatan fitur yang relevan. Proses analisis sentimen kemudian menggunakan algoritma klasifikasi *Naïve Bayes* untuk mengklasifikasikan ulasan ke dalam kategori sentimen positif dan negatif. Pendekatan ini diharapkan dapat memberikan hasil yang akurat dan efisien dalam mengidentifikasi dan memahami tanggapan pengguna terhadap aplikasi Shopee. Lebih lanjut, penggunaan metode klasifikasi algoritma *Naïve Bayes* diharapkan dapat memberikan kontribusi pada literatur terkait analisis sentimen dan memahami potensi metode ini dalam mengatasi kompleksitas fluktuasi sentimen pengguna dalam konteks aplikasi *e-commerce*. diperkuat.

Implikasi hasil penelitian ini memiliki dampak yang signifikan terhadap pemahaman kita tentang respons pengguna terhadap aplikasi Shopee di *Google Play Store* dan memberikan landasan yang kuat untuk perkembangan teknologi di bidang *e-commerce*. Dengan berhasil menerapkan metode klasifikasi algoritma *Naïve Bayes* dalam analisis sentimen, penelitian ini dapat memberikan pemahaman mendalam tentang pola-pola sentimen yang mungkin memengaruhi reputasi dan popularitas aplikasi. Temuan ini dapat dijadikan pedoman bagi praktisi *e-commerce* untuk merancang strategi yang lebih terarah dan responsif terhadap kebutuhan pengguna. Para peneliti di bidang Informatika juga dapat memanfaatkan temuan ini sebagai sumber inspirasi untuk pengembangan metode analisis sentimen yang lebih canggih dan efektif. Dengan memahami potensi metode klasifikasi algoritma *Naïve Bayes* dalam menangani kompleksitas variasi sentimen pelanggan, penelitian ini dapat menjadi landasan untuk penelitian lebih lanjut dalam meningkatkan akurasi dan efisiensi algoritma-analgoritma sejenis. Secara lebih luas, hasil penelitian ini dapat memberikan sumbangan pada pemahaman kita tentang hubungan antara pengguna dan aplikasi *e-commerce* di era digital. Implikasi ini tidak hanya relevan bagi pemangku kepentingan di industri *e-commerce*, tetapi juga untuk perkembangan teknologi secara keseluruhan, menggambarkan potensi metode klasifikasi algoritma *Naïve Bayes* sebagai alat yang efektif dalam analisis sentimen di berbagai

konteks aplikasi dan platform. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi yang berarti terhadap pemahaman kita tentang interaksi manusia dan teknologi di era digital ini.

TINJAUAN PUSTAKA

Dalam upaya untuk memahami persepsi masyarakat terhadap aplikasi Shopee, penelitian ini mengadopsi klasifikasi sentimen menggunakan algoritma *Naïve Bayes*. Beberapa penelitian sebelumnya yang mendukung pemilihan serta temuan-temuan penting dalam analisis sentiment diberbagai konteks.

Analisis sentimen terhadap penggunaan paylater dan pandangan masyarakat mengenai paylater. Peneliti menggunakan metode *Naïve Bayes Classifier* dan *TextBlob* untuk menganalisis data yang dikumpulkan melalui Twitter. Evaluasi dilakukan menggunakan confusion matrix, dengan *Naïve Bayes* memiliki akurasi 96% dan *TextBlob* memiliki akurasi 61%. Hasil pengumpulan dataset menunjukkan bahwa masyarakat cenderung merasa kurang baik terhadap penggunaan paylater, dengan analisis sentimen yang lebih banyak mendapatkan sentimen negatif daripada positif (Safira, Masyarakat... and Hasan, 2023). Penelitian kedua membahas mengenai analisis sentimen publik di Twitter terkait COVID-19 menggunakan algoritma *Naïve Bayes* dan KNN. Studi ini membandingkan hasil klasifikasi kedua metode tersebut dan mengidentifikasi tren opini publik di Twitter. Data dikumpulkan dari Twitter dengan kata kunci "COVID-19" dan kemudian diklasifikasikan sebagai positif atau negatif. Data tersebut kemudian dibersihkan dan diproses sebelum dianalisis menggunakan metode *Naïve Bayes* dan KNN. Hasilnya menunjukkan bahwa metode *Naïve Bayes* memiliki tingkat akurasi yang lebih tinggi, dan terdapat kecenderungan opini publik di Twitter cenderung positif terhadap COVID-19 (Syarifuddin, 2020). Penelitian ketiga membahas mengenai analisis sentimen terhadap dompet elektronik di Twitter menggunakan metode *Naïve Bayes Classifier*. Penelitian ini bertujuan untuk mengklasifikasikan tanggapan publik terhadap aplikasi dompet elektronik di Indonesia ke dalam kelas positif dan negatif. Data dikumpulkan dari Twitter menggunakan hashtag #GoPay, #OVO, dan #Dana, dan 100 data poin dianalisis. Hasilnya menunjukkan bahwa dompet elektronik GoPay memiliki

sentimen positif tertinggi, diikuti oleh Dana dan OVO. Jurnal ini memberikan bukti empiris dan rekomendasi bagi masyarakat dalam pemilihan dompet elektronik berdasarkan analisis tersebut. Metodologi penelitian meliputi pengumpulan data, preprocessing, dan analisis menggunakan Naïve Bayes Classifier, dengan RapidMiner digunakan untuk preprocessing dan analisis sentiment (Putra *et al.*, 2020). Penelitian keempat membahas mengenai analisis sentimen pengguna Twitter terhadap layanan BPJS menggunakan metode Lexicon Based dan Naïve Bayes Classifier. Penelitian ini bertujuan untuk membantu BPJS dalam melihat pendapat pengguna Twitter tentang pelayanan mereka. Metode analisis sentimen yang digunakan meliputi tahap perencanaan, analisis, desain, implementasi, dan pengujian, serta pengumpulan data dari Twitter menggunakan Twitter API. Metode Naïve Bayes Classifier adalah teknik prediksi berbasis probabilitas sederhana, sedangkan Lexicon-Based adalah metode analisis sentimen dengan menggunakan kamus sebagai sumber bahasa atau leksikal (Rosadi *et al.*, 2021). Penelitian kelima membahas mengenai analisis sentimen terhadap data Twitter yang berkaitan dengan diskusi tentang ibu kota baru Indonesia. Metode Naïve Bayes digunakan untuk mengklasifikasikan sentimen positif dan negatif dari tweet-tweet tersebut. Artikel ini juga memberikan metodologi penelitian, serta menyoroti relevansi analisis sentimen dalam berbagai bidang seperti ekonomi, politik, dan hukum. Hasil penelitian menunjukkan akurasi, presisi, dan recall yang tinggi dalam menggunakan metode Naïve Bayes untuk klasifikasi sentimen, terutama dalam konteks politik yang sensitif. Rekomendasi untuk penelitian selanjutnya termasuk menambah dataset dan membandingkan dengan model klasifikasi lainnya, serta menyertakan referensi untuk studi lebih lanjut (Aryanti, 2022). Penelitian keenam membahas mengenai analisis sentimen publik terhadap kinerja Dewan Perwakilan Rakyat (DPR) Indonesia yang diekspresikan melalui Twitter. Studi ini menggunakan pengumpulan data, preprocessing, dan klasifikasi menggunakan metode Naïve Bayes Classifier. Hasilnya menunjukkan sentimen positif, dan negatif terhadap DPR, dengan skor akurasi 80%. Jurnal ini juga membahas topik terkait Twitter, text mining, Python, Naive Bayes, dan Jupyter.

Metodologi penelitian melibatkan penggunaan Jupyter Hub untuk menjalankan notebook Jupyter, Design Science Research (DSR) sebagai metode penelitian, dan Analisis Sentimen untuk data mining. Studi ini fokus pada analisis sentimen terhadap tweet yang terkait dengan kinerja DPR Indonesia menggunakan algoritma Naive Bayes. Hasilnya menunjukkan skor akurasi 80% dan memberikan wawasan tentang sentimen dari tweet tersebut. Kesimpulannya menyoroti pengembangan yang berhasil dari sistem analisis sentimen dan menyarankan penggunaan database untuk menyimpan jumlah data yang besar serta menerapkan fitur visualisasi dengan opsi penyaringan. Referensi yang disebutkan mencakup sumber tentang preprocessing teks, social media mining dengan Python, dan workshop pemrograman Python (Duei Putri, Nama and Sulistiono, 2022). Penelitian ketujuh membahas mengenai analisis sentimen opini publik terhadap Undang-Undang Cipta Kerja pada Twitter menggunakan metode Naive Bayes Classifier. Penelitian ini menggunakan algoritma TF-IDF untuk pembobotan kata dan metode Naive Bayes Classifier untuk klasifikasi opini publik. Hasil evaluasi menunjukkan bahwa sentimen negatif lebih tinggi daripada sentimen positif (Nurdiansyah, Rahman and Pandunata, 2021). Penelitian kedelapan membahas mengenai analisis sentimen pelanggan Tokopedia menggunakan metode Naïve Bayes Classifier. Penelitian ini bertujuan untuk mengetahui hasil analisis sentimen opini publik di Tokopedia dan untuk mengetahui hasil akurasi dari pengklasifikasian menggunakan metode Naive Bayes Classifier. Data dikumpulkan dari aplikasi Tokopedia di Google Play Store dan kemudian dilakukan preprocessing, pembagian data training dan data testing, serta analisis menggunakan metode Naïve Bayes Classifier. Studi ini menggunakan metode Naïve Bayes Classifier untuk menganalisis sentimen pelanggan Tokopedia berdasarkan komentar-komentar mereka. Hasil pengujian menunjukkan akurasi sebesar 95,10%. Penulis merekomendasikan penambahan metode lain dan penggunaan data geografis untuk meningkatkan akurasi. Studi ini juga mencantumkan referensi dari berbagai sumber terkait analisis sentimen dan e-commerce di Indonesia (Salsabila, Alim Murtopo and Fadhilah, 2022). Paper membahas mengenai

analisis sentimen terhadap dokumen Twitter mengenai dampak virus Corona menggunakan metode klasifikasi Naive Bayes. Penelitian ini bertujuan untuk menentukan pendapat publik terhadap dampak virus tersebut. Data yang digunakan adalah dokumen tweet dari Twitter tentang dampak virus Corona. Metode Naive Bayes mampu mengklasifikasikan dokumen tweet dengan akurasi 67% dan tingkat kesalahan 33%. Studi ini menyimpulkan bahwa metode Naive Bayes menghasilkan performa yang stabil dan penelitian lebih lanjut dapat dikembangkan dengan memperhatikan fitur semantik dalam dokumen tweet. Selain itu, jurnal ini juga membahas penggunaan metode confusion matrix untuk menghitung akurasi dan tingkat kesalahan dalam analisis sentimen dokumen Twitter serta permasalahan yang ditemukan dalam penelitian tersebut (Astari, Dewa Gede Hendra Divayana and Gede Indrawan, 2020). Paper membahas mengenai analisis sentimen terhadap layanan transportasi online menggunakan data dari Twitter. Studi ini membandingkan kinerja metode Naive Bayes Classifier dan Support Vector Machine (SVM) untuk analisis sentimen terhadap perusahaan GrabId dan GojekIndonesia. Hasil penelitian menunjukkan bahwa metode SVM menghasilkan akurasi yang lebih baik untuk kedua perusahaan tersebut. Jurnal ini juga mencakup detail mengenai pengumpulan data, pra-pemrosesan, klasifikasi, dan evaluasi analisis sentiment (Dwianto and Sadikin, 2021).

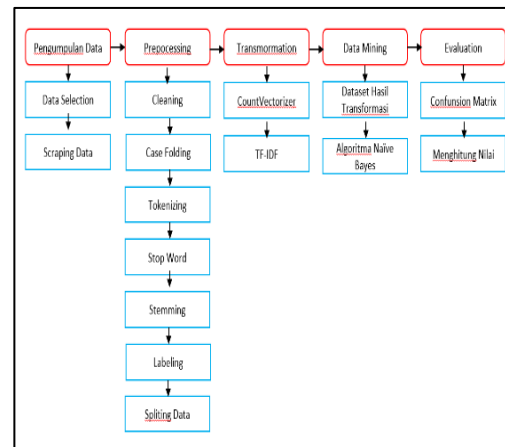
METODE PENELITIAN

Pada penelitian ini, peneliti melakukan pemodelan topik yang ada pada kelompok sentimen pengguna aplikasi Shopee terhadap pelayanan dan performa aplikasi Shopee menggunakan metode klasifikasi algoritma *Naive Bayes* agar dapat mengklasifikasikan komentar positif, netral, dan negatif dari penggunaan aplikasi Shopee di *Google Play Store* dan dapat digunakan untuk menilai kualitas layanan sekaligus sebagai bahan pertimbangan untuk developer aplikasi Shopee.

3.1 Tahapan Metode Penelitian

Pada penelitian ini, peneliti melakukan pemodelan topik yang ada pada kelompok sentimen pengguna aplikasi Shopee terhadap pelayanan dan performa aplikasi Shopee menggunakan metode klasifikasi algoritma *Naive Bayes* agar dapat mengklasifikasikan

komentar positif, netral, dan negatif dari penggunaan aplikasi Shopee di *Google Play Store* dan dapat digunakan untuk menilai kualitas layanan sekaligus sebagai bahan pertimbangan untuk developer aplikasi Shopee.



Gambar 1. Tahapan Metode Penelitian

3.2 Sumber Data

Dalam sentimen analisis yang dilakukan memerlukan data ulasan aplikasi Shopee yang berasal dari *Google Play Store*. Data yang diperlukan merupakan data teks berupa ulasan pelanggan terhadap aplikasi Shopee. Banyak sekali *review* yang diberikan oleh pelanggan aplikasi Shopee, untuk itulah peneliti mempelajari rating *review* publik aplikasi Shopee di *Google Play Store*.

3.3 Teknik Pengumpulan Data

Pengumpulan data dalam penelitian ini akan dilakukan melalui teknik web scraping dan penggunaan API untuk mengakses ulasan dan komentar pengguna di *Google Play Store* yang terkait dengan aplikasi Shopee. Teknik ini akan memungkinkan kami untuk secara otomatis mengumpulkan data dari sumber-sumber tersebut tanpa perlu interaksi langsung dengan pengguna.

3.4 Teknik Analisis Data

Penelitian ini menggunakan teknik analisis data menggunakan metode Naive Bayes untuk mengklasifikasikan sentimen pada ulasan pengguna PUBG Mobile yang dikumpulkan. Ulasan dikategorikan ke dalam sentimen positif, negatif, atau netral berdasarkan nilai rating / score yang diberikan pengguna. Algoritma Naive Bayes dipilih karena menerapkan pendekatan probabilistik dan statistik yang

sesuai untuk kasus klasifikasi teks. Naive Bayes memprediksi kemungkinan sebuah ulasan masuk ke kelas tertentu berdasarkan pola pada data latih sebelumnya. Implementasi Naive Bayes dilakukan dengan bahasa pemrograman Python pada platform Google Colab. Tahapan analisis data meliputi text preprocessing, penghitungan kata, persiapan data latih dan uji, pelatihan model Naive Bayes, klasifikasi sentimen pada data uji, serta evaluasi hasil klasifikasi. Dengan nilai-nilai ini, peneliti dapat menghitung metrik evaluasi seperti akurasi, presisi, recall, dan F1-Score untuk setiap kelas.

HASIL DAN PEMBAHASAN

4.1 Hasil Penelitian

Hasil penelitian ini akan menjelaskan secara rinci langkah-langkah yang dilewati mulai dari proses pengumpulan data, *preprocessing* data, *transformation* data, dan evaluasi data yang dilakukan. Selain itu, analisis data yang mendalam, yang ditunjang melalui berbagai macam jenis visualisasi seperti *statistic*, tabel, dan grafik data, akan memberikan kontribusi signifikan dalam meningkatkan pemahaman pembaca. Temuan yang dihasilkan juga akan menjadi dasar untuk menjadi rekomendasi penelitian selanjutnya.

4.1.1 Data Selection

Sebelum melakukan *scraping* data *review* dari *Google Play Store*, peneliti melakukan data *selection* terlebih dahulu dikarenakan kolom ulasan yang tersedia itu sangat banyak seperti “reviewId, username, userImage, content, score, thumbsUpCount, reviewCreate/Version, at, repliedAt dan appVersion”.

4.1.2 Scraping Data

Untuk melakukan pengambilan data menggunakan teknik web *scrapping* dilakukan dengan mengeksekusi *script python* di *Google Colab*, yaitu mengunjungi alamat web *google collab* di <https://colab.research.google.com/>, kemudian Peneliti melakukan instalasi beberapa *library* yang diperlukan di antaranya yaitu *google-play-scraper*, *pandas*, *numpy*, *matplotlib.pyplot*.

4.1.3 Preprocessing

Langkah-langkah *preprocessing* digunakan untuk membersihkan dan mempersiapkan

dataset sebelum analisis lebih lanjut. Proses ini melibatkan penghapusan informasi yang tidak relevan seperti URL, *mention*, dan *hashtag*, serta penghapusan karakter *non-alphabetic*. Selanjutnya, teks dikonversi menjadi huruf kecil, di-tokenisasi, dan menjalani proses penghapusan *stop word* dan *stemming*. Pada tahap penghapusan *stop word*, digunakan *library* Sastrawi yang menyediakan kamus *stop word* bahasa Indonesia. Dengan menghapus *stop word*, kualitas teks dapat ditingkatkan dan pengaruh kata-kata yang tidak relevan dalam analisis dapat dikurangi. Selanjutnya, proses *stemming* juga dilakukan menggunakan *library* Sastrawi. *Stemming* adalah proses mengubah kata-kata menjadi kata dasar atau akar kata. Dalam bahasa Indonesia, kata-kata sering memiliki variasi bentuk yang sama tetapi dengan akhiran atau awalan yang berbeda. Dengan melakukan *stemming*, variasi kata yang tidak perlu dapat dikurangi dan kata-kata dengan akar yang sama dapat digabungkan untuk menghasilkan representasi yang lebih ringkas dan konsisten. Untuk menunjang penggunaan *library* sastrawi peneliti juga menggunakan *library Swifter* untuk meningkatkan kecepatan eksekusi pada pemrosesan data berukuran besar, dengan menggunakan *Swifter* proses seperti penghapusan *stop word* dan *stemming* pada teks dapat dilakukan secara parallel dan mampu mengurangi waktu eksekusi yang dibutuhkan. Setelah semua proses tersebut selesai kemudian dilakukan *labeling* data untuk mengkategorikan sentimen negatif, positif dan netral. Setelah melewati beberapa langkah tersebut sebelum melakukan transformasi data terlebih dahulu dilakukan *Split* data yaitu membagi dataset menjadi data Latih(*Training*) Dan Data Uji(*Test*).

4.1.4 Transformation

Transformasi data merupakan proses mengubah data dari satu format ke format lain yang lebih sesuai untuk analisis atau penggunaan selanjutnya. Transformasi data teks dilakukan dengan tujuan mengubah data teks menjadi format numerik agar dapat dianalisis menggunakan algoritma data *mining*.

4.1.5 Data Mining

Data *mining* merupakan proses analisis data yang kompleks, terstruktur, dan tidak terstruktur

untuk menemukan pola, atribut, relasi, dan keanekaragaman lain untuk mendapatkan informasi yang diperlukan. Pada tahap data *mining*, akan dilakukan penerapan algoritma *Naive Bayes classifier* untuk melakukan klasifikasi sentimen pada data ulasan pengguna aplikasi Shopee. Algoritma *Naive Bayes* dipilih karena kemampuannya dalam melakukan klasifikasi dengan cepat dan akurat, serta telah terbukti efektif dalam kasus analisis sentimen. Algoritma *Naive Bayes* akan memodelkan peluang kemunculan kata-kata tertentu pada setiap kelas sentimen (positif, negatif, netral). Kemudian peluang tersebut digunakan untuk memprediksi kelas sentimen dari ulasan pengguna baru. Data yang digunakan adalah data ulasan pengguna aplikasi Shopee yang telah melalui *preprocessing* dan transformasi. Data yang sebelumnya telah dibagi menjadi data latih dan data uji, untuk lebih lengkapnya bisa dilihat pada Tabel 1.

Tabel 1. Pembagian Dataset

| DATA LATIH | DATA UJI |
|------------|----------|
| 80% | 20% |

4.1.6 Evaluation

Evaluasi model *machine learning* adalah proses mengukur seberapa baik model yang dilatih dapat melakukan prediksi atau klasifikasi pada data baru. Tujuannya adalah untuk menilai performa model sebelum diimplementasikan, sehingga diketahui apakah model sudah cukup baik atau perlu perbaikan lebih lanjut. Dalam penelitian ini evaluasi dilakukan menggunakan teknik *confusion matrix* yaitu teknik yang digunakan untuk mengevaluasi klasifikasi model untuk memperkirakan objek yang benar atau salah. Rumus yang digunakan untuk menghitung *confusion matrix multiclass* (terdiri lebih dari 2 dataset) tentunya berbeda dengan rumus *confusion matrix* 2 kelas perhitungan confusion matrix dilakukan dengan mencari nilai True Positive (TP), False Positive (FP), True Negative (TN) dan False Negative (FN) untuk setiap kelas (negatif, netral, positif) untuk lebih lengkapnya dapat dilihat pada tabel 2.

Tabel 2. Nilai Hasil *Confusion Matrix*

| KELAS | N. PREDIKSI | Nt. PREDIKSI | P. PREDIKSI |
|---------|----------------|-----------------|----------------|
| NEGATIF | TN | FNt | FP |
| NETRAL | FN | TNt | FP |
| POSITIF | FN | FNt | TP |

True Positive (TP), merupakan nilai kelas yang diprediksi bernilai *positive* dan faktanya *positive*. *False Positive* (FP), merupakan nilai kelas yang diprediksi bernilai *positive* dan faktanya *negative* atau *neutral*. *True Negative* (TN), merupakan nilai kelas yang diprediksi bernilai *negative* dan faktanya *negative*. *False Negative* (FN), merupakan nilai kelas yang diprediksi bernilai *negative* dan faktanya *positive* atau *neutral*. *True Neutral* (TNt) merupakan nilai kelas yang diprediksi bernilai *neutral* dan faktanya *neutral*. *False neutral* (FNt) merupakan nilai yang kelas yang diprediksi bernilai *neutral* dan faktanya *positive* dan *negative*.

4.2 Pembahasan

Performa algoritma *Naive Bayes* dalam mengklasifikasikan sentimen ulasan pengguna aplikasi Shopee pada data latih dan data uji dapat dilihat dari nilai akurasi, presisi, *recall*, dan F1-Score yang dihasilkan dari *confusion matrix* Setiap Model yang sebelumnya telah diproses menjadi representasi vektor menggunakan *CountVectorizer* dan dilakukan pembobotan kata menggunakan TF-IDF. Sebelum digunakan untuk memprediksi data uji, terlebih dahulu dilakukan pelatihan model *Naive Bayes* pada data latih agar model dapat mempelajari pola-pola yang terdapat pada data berupa kumpulan fitur tekstual dari ulasan pengguna. Proses pelatihan ini penting agar model dapat belajar dan memetakan pola yang membedakan sentimen positif dan negatif pada ulasan aplikasi Shopee. Data latih digunakan untuk menyesuaikan bobot dan parameter model sehingga akurasi klasifikasi sentimen oleh model *Naive Bayes* dapat optimal. Data latih yang telah melalui beberapa tahapan *preprocessing text* dan ekstraksi fitur menggunakan *CountVectorizer* dan TF-IDF, tahapan ini bertujuan untuk mengubah data tekstual menjadi representasi numerikal agar dapat diolah oleh model *Naive Bayes*. Representasi vektor TF-IDF mewakili

bobot pentingnya masing-masing kata pada setiap dokumen teks ulasan. Hasil *modeling* data latih bisa dilihat pada Tabel 3

Tabel 3. Hasil Akurasi Data Latih

| AKURASI DATA LATIH |
|--------------------|
| 0.97 |

Akurasi model *Naïve Bayes* pada data latih sebesar 97%. Artinya, dengan *pipeline preprocessing* dan ekstraksi fitur pada model algoritma *Naïve Bayes* mampu mengklasifikasikan sentimen ulasan aplikasi Shopee pada data latih dengan akurasi 97%. Akurasi model *Naïve Bayes* pada data latih yang cukup tinggi ini mengindikasikan bahwa model telah berhasil memetakan pola-pola pada data latih dengan baik sehingga mampu melakukan prediksi sentimen dengan akurat. Model pada data latih ini kemudian dapat digunakan untuk klasifikasi sentimen pada data uji. Performa model pada data latih belum tentu sama dengan performanya pada data uji. Oleh karena itu, perlu dilakukan pengujian lebih lanjut pada data uji untuk mengevaluasi kemampuan generalisasi model dalam memprediksi data baru yang belum pernah dilihat sebelumnya. Oleh karena itu, perlu dilakukan pengujian lebih lanjut pada data uji untuk mengevaluasi kemampuan generalisasi model dalam memprediksi data baru yang belum pernah dilihat sebelumnya. Setelah semua proses selesai nilai akurasi prediksi data latih pada data uji bisa dilihat pada Tabel 4.

Tabel 4. Hasil Akurasi Data Uji

| AKURASI DATA UJI |
|------------------|
| 0.88 |

Berdasarkan Tabel 4., dapat dilihat bahwa akurasi model *Naïve Bayes* pada data uji adalah 88%. Angka akurasi ini lebih rendah dibandingkan akurasi tertinggi model pada data latih yang mencapai 97%. Penurunan akurasi ini mengindikasikan bahwa model *Naïve Bayes* memiliki tantangan dalam menggeneralisasi pola dari data latih ke data uji, dan perlu perhatian lebih lanjut untuk meningkatkan kinerjanya pada data uji. Pada pembahasan sebelumnya telah dijelaskan performa model *Naive Bayes* pada data latih dan data uji berdasarkan nilai akurasi. Selanjutnya akan

dibahas lebih rinci mengenai performa model pada masing-masing kelas sentimen dengan melihat nilai presisi, recall, dan F1-score yang dihitung dari confusion matrix. Dari kelima model, Model pertama memiliki akurasi paling tinggi pada data uji yaitu 97%, sehingga confusion matrix Model pertama yang akan dianalisis lebih lanjut. Dari confusion matrix, dapat dihitung nilai TP, FP, TN, dan FN pada masing-masing kelas yang selanjutnya digunakan untuk menghitung metrik evaluasi. Nilai confusion matrix model pertama bisa dilihat pada Tabel 5.

Tabel 5. Confusion Matrix

| KELAS | N. PREDIKSI | Nt. PREDIKSI | P. PREDIKSI |
|---------|----------------|-----------------|----------------|
| NEGATIF | 12 | 0 | 6 |
| NETRAL | 2 | 0 | 2 |
| POSITIF | 2 | 0 | 76 |

Dari Tabel 5. Menghasilkan Nilai Akurasi, Presisi, Recall Dan F1-Score yang bisa dilihat pada Tabel 6.

Tabel 6. Hasil Akurasi

| MODEL | AKURASI | PRESISI | RECALL | F1- SCORE |
|------------------|---------|---------|--------|--------------|
| MODEL PERTAMA | 80% | 94% | 98% | 193% |

Sebelum mendapatkan nilai seperti yang ditampilkan pada Tabel 6. terlebih dahulu dilakukan perhitungan disetiap kelasnya untuk mencari nilai presisi, recall, dan F1-score disetiap kelasnya.

$$Akurasi = \frac{76 + 12 + 0}{12 + 2 + 2 + 6 + 2 + 76 + 0} \times 100\% = 0.88$$

$$= 88\%$$

Maka Kelas negatif:

$$Precision = \frac{76}{76 + 6} = 0.926$$

$$Recall = \frac{76}{76 + 0} = 1$$

$$F1 = 2x \frac{0.926 \times 1}{0.917 + 1} = 0.961$$

Kelas netral :

$$Precision = \frac{76}{76 + 6} = 0.9268$$

$$Recall = \frac{76}{76 + 2} = 0.9743$$

$$F1 - Score = 2x \frac{0.9268x1}{0.9268 + 1} = 0.9613$$

Kelas Positif :

$$Precision = \frac{76}{76 + 2} = 0.974$$

$$Recall = \frac{76}{76 + 2} = 0.974$$

$$F1 - Score = 2x \frac{0.974 \times 0.974}{0.974 + 0.974} = 0.974$$

Kemudian setelah melakukan perhitungan pada setiap kelasnya kita bisa melakukan perhitungan rata-rata dari setiap kelasnya seperti berikut ini:

$$Precision = \frac{Presisi K.1 + Presisi K.2 + Presisi K.3}{3}$$

$$Precision = \frac{92.6\% + 92.6\% + 97.4\%}{3} = 94.2\%$$

$$Recall = \frac{Recall K.1 + Recall K.2 + Recall K.3}{3}$$

$$Recall = \frac{100\% + 97.4\% + 97.4\%}{3} = 98.27\%$$

$$F1 - Score = 2x \frac{F1 + F2 + F3}{3}$$

$$F1 - Score = 2x \frac{96.1\% + 96.1\% + 97.4\%}{3} = 93.07\%$$

Dari 18 data yang diprediksi sebagai negatif, 12 di antaranya memang bernilai negatif dan 6 merupakan data bukan negatif yang salah diklasifikasikan (*false positive*). Oleh karena adanya *false positive* ini, nilai presisi kelas negatif menjadi lebih rendah yaitu 92.68%, meskipun recall cukup tinggi 97.44%. Pada kelas netral, dari 4 data tidak ada data yang diklasifikasikan dengan benar sebagai netral. Terdapat sebanyak 4 data atau semua data netral salah diprediksi diantaranya 2 data sebagai kelas (*false negative*) dan 2 data sebagai kelas (*false positive*), menyebabkan recall yaitu 97.44%. Dan nilai presisinya yaitu 92.68%. Pada kelas positif, dari 78 data hanya 76 yang benar diklasifikasikan positif dan 2 data salah diprediksi ke kelas lain (*false negative*). Menyebabkan nilai recall sebesar 97.44%. dan nilai presisi kelas positif yaitu 97.44%. Performa algoritma *Naïve Bayes* dalam melakukan klasifikasi sentimen ulasan aplikasi Shopee menunjukkan kemampuan yang cukup baik, dengan akurasi tertinggi 88% pada data uji. Meskipun akurasi ini lebih rendah jika dibandingkan performa pada data latih yang mencapai 97%.

SIMPULAN

Hasil analisis variasi sentimen pada ulasan aplikasi Shopee di Google Play Store, disimpulkan bahwa terdapat beragam sentimen yang tercermin dari pengalaman pengguna. Variasi ini mencakup aspek positif, negatif, dan netral, sentimen ulasan pengguna aplikasi Shopee didominasi oleh sentimen positif sebesar 74.6% atau sebanyak 746 data dari total 1000 data ulasan. Sentimen positif yang tinggi ini mengindikasikan banyaknya pengguna sudah merasa puas terhadap berbagai aspek dalam aplikasi Shopee. Sementara itu, proporsi sentimen negatif hanya sebesar 21.2% atau 212 data dari 1000 data. Sentimen negatif ini kebanyakan mengeluh terhadap masalah teknis seperti lag, bug pada aplikasi. Namun porsinya masih kalah banyak dibandingkan sentimen positif. Sedangkan sentimen netral paling sedikit yaitu 4.2% atau 42 data. Hal ini mengindikasi sedikitnya ulasan objektif tanpa opini kuat terhadap aplikasi Shopee ini.

Untuk penelitian selanjutnya yaitu menyelidiki dan membandingkan performa algoritma klasifikasi sentimen lainnya untuk memastikan pemilihan metode yang paling efektif. Penggunaan ensemble learning atau pendekatan gabungan dari beberapa algoritma mungkin dapat meningkatkan akurasi klasifikasi sentiment.

DAFTAR PUSTAKA

- Aryanti, D. (2022) 'Analisis Sentimen Ibukota Negara Baru Menggunakan Metode Naïve Bayes Classifier', *Journal of Information System Research (JOSH)*, 3(4), pp. 524–531. Available at: <https://doi.org/10.47065/josh.v3i4.1944>.
- Astari, N.M.A.J., Dewa Gede Hendra Divayana and Gede Indrawan (2020) 'Analisis Sentimen Dokumen Twitter Mengenai Dampak Virus Corona Menggunakan Metode Naive Bayes Classifier', *Jurnal Sistem dan Informatika (JSI)*, 15(1), pp. 27–29. Available at: <https://doi.org/10.30864/jsi.v15i1.332>.
- Duei Putri, D., Nama, G.F. and Sulistiono, W.E. (2022) 'Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier', *Jurnal Informatika dan Teknik Elektro Terapan*, 10(1), pp. 34–40. Available at:

<https://doi.org/10.23960/jitet.v10i1.2262>.

- Dwianto, E. and Sadikin, M. (2021) 'Analisis Sentimen Transportasi Online pada Twitter Menggunakan Metode Klasifikasi Naïve Bayes dan Support Vector Machine', *Format : Jurnal Ilmiah Teknik Informatika*, 10(1), p. 94. Available at: <https://doi.org/10.22441/format.2021.v10.i1.009>.
- Nurdiansyah, Y., Rahman, F. and Pandunata, P. (2021) 'Analisis Sentimen Opini Publik Terhadap Undang-Undang Cipta Kerja pada Twitter Menggunakan Metode Naïve Bayes Classifier', *Prosiding Seminar Nasional Sains Teknologi dan Inovasi Indonesia (SENASTINDO)*, 3(November), pp. 201–212. Available at: <https://doi.org/10.54706/senastindo.v3.2021.158>.
- Putra, M.W.A. *et al.* (2020) 'Analisis Sentimen Dompet Elektronik Pada Twitter Menggunakan Metode Naïve Bayes Classifier', *IT Journal Research and Development*, 5(1), pp. 72–86. Available at: [https://doi.org/10.25299/itjrd.2020.vol5\(1\).5159](https://doi.org/10.25299/itjrd.2020.vol5(1).5159).
- Rosadi, A. *et al.* (2021) 'Analisis Sentimen Berdasarkan Opini Pengguna pada Media Twitter Terhadap BPJS Menggunakan Metode Lexicon Based dan Naïve Bayes Classifier Twitter Text Mining', 20, pp. 39–52.
- Safira, A., Masyarakat...□, A.S. and Hasan, F.N. (2023) 'ANALISIS SENTIMEN MASYARAKAT TERHADAP PAYLATER MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER', *Jurnal Sistem Informasi*, 5(1).
- Salsabila, S.M., Alim Murtopo, A. and Fadhillah, N. (2022) 'Analisis Sentimen Pelanggan Tokopedia Menggunakan Metode Naïve Bayes Classifier', *Jurnal Minfo Polgan*, 11(2), pp. 30–35. Available at: <https://doi.org/10.33395/jmp.v11i2.11640>.
- Syarifuddin, M. (2020) 'Analisis Sentimen Opini Publik Mengenai Covid-19 Pada Twitter Menggunakan Metode Naïve Bayes Dan Knn', *INTI Nusa Mandiri*, 15(1), pp. 23–28. Available at: <https://doi.org/10.33480/inti.v15i1.1347>.