



Research article

Klasifikasi Sentimen Ulasan Pengguna Aplikasi Shopee di Google Play Store Dengan Metode Naive Bayes

Classification of Sentiment in User Reviews of the Shopee App on Google Play Store Using the Naive Bayes Method

Meida Amanda¹, Tasya Parmi²

^{1,2} Program Studi Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Islam Indragiri
email: ¹meidaamanda113@gmail.com, ²tsyaparmi@gmail.com

ARTICLE INFO

Article history:

Received November 12, 2025

Revised November 27, 2025

Accepted November 29, 2025

Available online December 22, 2025

Keywords:

Klasifikasi

Naive Bayes

Play Store

Sentimen

Shopee

ABSTRACT

Belum ada metode yang relevan dan sistematis yang dapat mengelompokkan sebuah ulasan menjadi kelompok positif dan negatif pada aplikasi belanja *online Shopee*. Penelitian ini bertujuan untuk mengklasifikasi ulasan agar dapat memberikan solusi untuk aplikasi Shopee dalam meningkatkan layanan dan kepuasan pelanggan. Analisis ini merupakan klasifikasi sentimen dengan mengandalkan metode Naive Bayes yang terdiri dari 1.000 ulasan yang berasal dari Google Playstore. Setelah dilakukan pelatihan dan pengujian data dengan pembagian 80% data latih dan 20% data uji, model menghasilkan nilai akurasi sebesar 0,85, precision sebesar 0,79, recall sebesar 0,95, dan f1-score sebesar 0,86. Nilai tersebut menunjukkan bahwa model mampu mengklasifikasikan sentimen ulasan dengan tingkat keakuratan yang tinggi dan keseimbangan performa yang baik antara *precision* dan *recall*. Model terbukti lebih unggul dalam mengenali ulasan dengan sentimen Negatif dibandingkan Positif, di mana recall untuk kelas Negatif mencapai 0,96, sedangkan untuk kelas Positif sebesar 0,75. Studi ini memberikan wawasan tentang pandangan pengguna terhadap Shopee, dan membantu aplikasi dalam meningkatkan pengalaman pengguna pada platform tersebut

There is currently no relevant and systematic method that can classify reviews into positive and negative groups on the Shopee online shopping application. This study aims to classify reviews in order to provide solutions for the Shopee application in improving its services and customer satisfaction. This analysis is a sentiment classification relying on the Naive Bayes method, consisting of 1,000 reviews from Google Playstore. After training and testing the data with an 80% training data and 20% test data split, the model produced an accuracy value of 0.85, precision of 0.79, recall of 0.95, and an F1-score of 0.86. These values indicate that the model is capable of classifying review sentiment with a high degree of accuracy and a good balance between precision and recall. The model proved to be superior in recognizing reviews with negative sentiment compared to positive sentiment, where the recall for the negative class reached 0.96, while for the positive class it was 0.75. This study provides insight into user perceptions of Shopee and helps the application improve the user experience on the platform.

1. Pendahuluan

Pesatnya perkembangan dunia teknologi dan informasi dalam beberapa tahun terakhir, menjadikan internet sebagai alat komunikasi yang banyak diminati oleh masyarakat [13]. Pada zaman modern ini, sentimen atau opini masyarakat semakin bertambah luas dan bebas diungkapkan di berbagai media. Sentimen dapat menjadi potensi besar bagi perusahaan yang ingin mengetahui umpan balik (*feedback*) dari masyarakat. Jumlah pengguna (*user*) aktif dalam komunikasi *Online* memiliki jumlah data yang sangat banyak [1]. Salah satu diantaranya adalah belanja online, belanja online memudahkan masyarakat untuk membeli sesuatu sesuai dengan kebutuhan yang diinginkan. Belanja online disukai oleh masyarakat karena begitu sederhana dan harga barang relatif lebih murah. Munculnya

teknologi baru belanja online ini dapat digunakan setiap orang dalam membuat konsumen untuk lebih sering berkomunikasi dengan penyedia layanan hingga tingginya ketertarikan pada aplikasi tersebut[2]

Shopee merupakan salah satu marketplace yang populer di Indonesia. Shopee adalah aplikasi belanja online yang dapat diunduh di situs Google Playstore, shopee menyediakan berbagai layanan menarik, salah satunya adalah ulasan untuk mengetahui kualitas barang yang dijual [3] Ulasan produk merupakan salah satu sumber informasi tentang kualitas produk yang sangat berpengaruh konsumen. Dalam kegiatan pembelian barang di marketplace. Pembeli dapat memberikan ulasan setelah menerima barang yang dibeli. Ulasan pembelian produk terdiri dari bintang dan isi komentar ulasan yang berisi tanggapan, apresiasi maupun kritik dan masukan pada produk yang telah dibeli tersebut. Ulasan pembelian produk memiliki pengaruh yang signifikan terhadap minat beli dari pembeli lain[4]. Setiap aplikasi memiliki kelebihan dan kekurangan yang mungkin menimbulkan reaksi berbeda dari pengguna, seperti kepuasan atau ketidakpuasan. Pengguna Shopee sering membagikan pengalaman mereka baik positif maupun negatif dibagian komentar Playstore. Analisis sentimen ulasan aplikasi media membantu bisnis dalam memahami permintaan dan preferensi pelanggan mereka serta dalam menemukan masalah yang dapat merusak reputasi mereka[5]

Berdasarkan ulasan-ulasan yang diberikan oleh pengguna aplikasi Shopee diharapkan adanya umpan balik bagi pihak developer untuk meningkatkan kualitas layanan. Belum ada metode yang relevan dan sistematis yang dapat mengelompokkan sebuah ulasan menjadi kelompok positif dan negatif. Dengan adanya klasifikasi data sentimen, akan mempermudah pihak developer dalam mendapatkan informasi sentimen dari para pengguna aplikasi e-commerce. [6]

Penelitian ini dilakukan untuk memberikan pemahaman mendalam tentang cara menerapkan analisis sentimen berbasis aspek pada ulasan dalam pengguna aplikasi Shopee agar dapat memahami umpan balik pengguna dengan baik. Manfaat lain adalah dapat membantu aplikasi Shopee dalam memberikan solusi yang mudah dan perusahaan e-commerce lainnya yang ada di Indonesia dalam meningkatkan pengalaman pengguna dan efisiensi dalam menangani ulasan para pengguna serta meningkatkan kualitas layanan dan kepuasan pelanggan, efisiensi dalam pengelolaan ulasan, juga keunggulan kompetitif.

2. Tinjauan literatur

Analisis sentimen merupakan hasil proses analisis data berupa pandangan dan opini guna menarik kesimpulan dari berbagai opini yang ada. Hasil analisis sentimen dapat berupa persentase sentimen positif, negatif, atau netral. Analisis sentimen berguna untuk berbagai masalah yang menjadi perhatian para pakar dan peneliti interaksi manusia-komputer, serta bidang-bidang seperti sosiologi, pemasaran dan periklanan, psikologi, ekonomi, dan ilmu politik. Salah satu media sosial yang biasa digunakan masyarakat untuk menyampaikan pendapatnya adalah melalui aplikasi Shopee secara langsung.[7] Shopee merupakan salah satu perusahaan yang bergerak dibidang aplikasi e-commerce dan website secara online. Aplikasi Shopee ini dapat memfasilitasi penjual untuk berjualan dengan mudah dan aman dalam proses pembayaran serta pengaturan logistik yang terintegrasi.[18]

Penelitian yang dilakukan oleh Agustina et al. [9] membahas analisis sentimen terhadap ulasan pengguna aplikasi Shopee yang diambil dari Google Play Store. Proses penelitian dimulai dengan pengumpulan data menggunakan teknik *web scraping* berbasis Python melalui library *google-play-scraper* untuk memperoleh 1.000 data ulasan pengguna. Data yang diperoleh kemudian melalui beberapa tahapan *text preprocessing* seperti *case folding*, *tokenizing*, *filtering*, dan *stemming* menggunakan library Sastrawi untuk bahasa Indonesia. Selanjutnya, proses pembobotan teks dilakukan menggunakan metode TF-IDF (Term Frequency-Inverse Document Frequency) untuk menghitung bobot kata berdasarkan tingkat frekuensinya pada dokumen. Model dikembangkan menggunakan algoritma Naïve Bayes, dan hasil evaluasi menunjukkan bahwa metode *Hold-Out* (80% data latih dan 20% data uji) memberikan hasil terbaik dengan nilai akurasi sebesar 83%, *precision* 83%, *recall* 100%, dan *f1-score* 91%. Hasil tersebut menunjukkan bahwa algoritma Naïve Bayes mampu mengklasifikasikan sentimen positif dengan sangat baik.

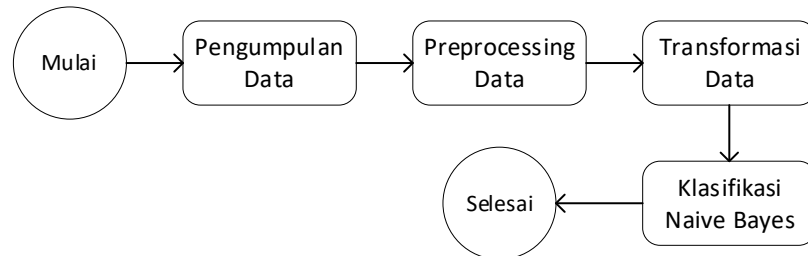
Selanjutnya, penelitian oleh Simanjuntak et al. [7] juga mengimplementasikan algoritma Multinomial Naïve Bayes pada analisis sentimen ulasan pengguna Shopee di Google Play Store dengan jumlah data sebanyak 1.000 ulasan. Penelitian ini menyoroti efisiensi algoritma Naïve Bayes dalam menangani data teks berskala besar dengan waktu komputasi yang singkat. Tahapan penelitian meliputi *preprocessing* data (*case folding*, *stopword removal*, *tokenizing*, dan *stemming*), transformasi fitur dengan metode TF-IDF, serta pembagian data menggunakan teknik *Hold-Out* dengan rasio 80:20 antara data latih dan data uji. Evaluasi model dilakukan menggunakan *confusion matrix* dan menghasilkan nilai akurasi sebesar 84%, *precision* 84%, *recall* 97%, dan *f1-score* 90%. Berdasarkan hasil tersebut, disimpulkan bahwa sebagian besar ulasan pengguna Shopee cenderung bersentimen negatif, yang menggambarkan adanya ketidakpuasan terhadap beberapa aspek layanan aplikasi.

Dari kedua penelitian tersebut dapat disimpulkan bahwa algoritma Naïve Bayes, khususnya varian Multinomial Naïve Bayes, menunjukkan performa yang konsisten dan akurat dalam mengklasifikasikan sentimen ulasan pengguna. Perbedaan hasil klasifikasi antara penelitian Agustina et al. [9] dan Simanjuntak et al. [7] dapat disebabkan oleh perbedaan periode pengambilan data, jumlah sampel, serta konteks pengalaman pengguna terhadap aplikasi Shopee pada waktu yang berbeda. Kedua penelitian tersebut juga menegaskan pentingnya tahapan *preprocessing* dan metode TF-IDF sebagai langkah penting dalam meningkatkan performa model klasifikasi teks.

Secara umum, hasil dari tinjauan literatur ini memperkuat bahwa Naïve Bayes merupakan metode yang efektif, sederhana, dan cepat untuk analisis sentimen teks berbahasa Indonesia. Algoritma ini dapat digunakan sebagai dasar bagi penelitian lanjutan yang membandingkan performanya dengan algoritma lain seperti Support Vector Machine (SVM), K-Nearest Neighbor (KNN), atau Decision Tree untuk meningkatkan akurasi dan stabilitas klasifikasi pada data ulasan pengguna e-commerce.

3. Metode Penelitian

Penelitian ini memanfaatkan 1.000 data ulasan yang diperoleh dari aplikasi e-commerce Shopee melalui situs Google Play Store. Data tersebut kemudian diklasifikasikan ke dalam dua kategori, yaitu ulasan positif dan ulasan negatif. Tahapan lengkap proses penelitian dapat dilihat pada Gambar 1.



Gambar 1. Alur Penelitian

Gambar 1 menunjukkan alur tahapan penelitian yang dimulai dari proses proses yang meliputi pengumpulan data, preprocessing data, transformasi data berupa pembagian dan pembobotan TF-IDF, klasifikasi Naïve Bayes yang didalamnya terdapat evaluasi dari hasil.

3.1 Pengumpulan data

Pada tahap awal ini adalah pengumpulan data dan pelabelan data. Data yang diambil menggunakan teknik web scrapping dengan menggunakan pemrograman python. Data sampel yang diambil merupakan ulasan pengguna aplikasi E-Commerce pada Google Play Store[7]. Proses scraping dilakukan secara otomatis di Google Colab menggunakan bahasa pemrograman Python, dengan hasil data disimpan dalam format CSV sebelum diproses lebih lanjut.[10] Selain itu pada tahap ini juga dilakukan pelabelan data dan cleaning. Data yang telah diperoleh kemudian melalui tahap pelabelan (labeling) dengan mengklasifikasikan ulasan berdasarkan skor bintang menjadi dua kategori, yaitu positif dan negatif. Selanjutnya dilakukan pembersihan data (data cleaning) untuk memastikan dataset yang digunakan bersih dan layak diolah dengan menghapus data kosong, duplikat, serta baris yang tidak memiliki label. Setelah proses ini selesai, dataset hanya berisi dua kolom utama, yaitu content (teks ulasan) dan Label (kategori sentimen).

3.2 Preprocessing data

Preprocessing adalah tahap yang dilakukan untuk menyiapkan data sebelum proses pemodelan dilakukan. Tahap ini termasuk salah satu teknik dalam data mining yang berfungsi untuk mengubah data mentah menjadi format yang lebih terstruktur dan mudah dipahami.[9]

3.1.1 Case Folding

Tahap case folding dilakukan dengan mengubah seluruh huruf pada teks menjadi huruf kecil (lowercase). Langkah ini bertujuan untuk menyeragamkan penulisan huruf agar sistem tidak membedakan antara huruf besar dan huruf kecil. Dengan begitu, proses pencarian dan analisis kata menjadi lebih mudah dan konsisten, karena teks seperti “Shopee” dan “shopee” akan dianggap sama.[11]

3.1.2 Stopword Removal

Filtering dilakukan untuk menghapus komponen yang tidak penting dan hanya mempertahankan komponen yang dianggap penting, proses ini menggunakan stopwords removal. Stopword removal berfungsi untuk menghapus kata yang tidak memiliki bobot atau kata tidak penting, seperti “dan”, “saya”, atau “kamu”.[12] Kata-kata tersebut disebut *stopword* karena sering muncul tetapi tidak memberikan informasi deskriptif terhadap isi dokumen. Dengan menghapus kata-kata ini menggunakan daftar *stopword*, data teks menjadi lebih ringkas tanpa mengubah makna utama dari ulasan yang akan diproses pada tahap berikutnya [18]

3.1.3 Tokenizing

Pada tahap ini diperlukan untuk memecah suatu kalimat menjadi potongan-potongan kata atau *term* yang dipisahkan berdasarkan spasi atau tanda baca. Proses ini bertujuan agar setiap kata dapat diidentifikasi dan dianalisis secara terpisah, sehingga memudahkan sistem dalam memahami struktur teks dan menghitung frekuensi kemunculan kata pada tahap analisis selanjutnya.[10]

3.1.4 Stemming

Tahapan terakhir dalam proses preprocessing yaitu stemming, dimana proses ini bertujuan untuk menghilangkan sufiks dan prefiks dan mengubahnya ke dalam bentuk dasarnya.[14] Proses tersebut dilakukan karena kata berimbuhan memiliki makna yang sama dengan kata dasar[15]

3.3 Transformasi data

Transformasi data menggunakan metode TF-IDF untuk mengubah teks menjadi bentuk numerik dengan menghitung bobot setiap kata berdasarkan frekuensinya. Sebelum proses pembobotan, data dibagi menjadi 80% data latih dan 20% data uji. [16]

3.4 Klasifikasi Naïve Bayes

Multinomial Naive Bayes digunakan untuk melakukan klasifikasi pada data berbentuk teks. Algoritma ini bekerja dengan mengasumsikan bahwa setiap kata dalam dokumen bersifat saling independen, artinya kemunculan suatu kata tidak bergantung pada kata lainnya. Karena itu, setiap dokumen dianalisis secara terpisah sehingga hasil klasifikasinya benar-benar berasal dari isi dokumen tersebut saja.[9] Misalnya, jika terdapat tiga fitur yang mendeskripsikan suatu objek, maka ketiga fitur tersebut secara independen memberikan kontribusi terhadap probabilitas yang mendeskripsikan objek tersebut, yang disebut “Naïve”. [17] Selain itu, metode Naive Bayes memiliki waktu klasifikasi yang singkat mempercepat proses sistem analisis sentimen [11] Bentuk umum teorema bayes adalah sebagai berikut:

$$P(C|X) = P(X|C) \frac{P(C)}{P(X)}$$

Keterangan:

$P(C|X)$ adalah probabilitas kelas C diberikan fitur X (posterior probability)

$P(X|C)$ adalah probabilitas fitur X diberikan kelas C (likelihood)

$P(C)$ adalah probabilitas awal dari kelas C (prior probability)

$P(X)$ adalah probabilitas dari fitur X (evidence), yang bisa diabaikan dalam klasifikasi karena sama untuk semua kelas.

Dalam Multinomial Naive Bayes, setiap dokumen diwakili sebagai vektor dari frekuensi kata-kata, dan probabilitas suatu kata muncul dalam kelas tertentu dihitung dengan persamaan (1) berikut:

$$P(W|C) = \frac{\text{count}(w,C) + a}{\sum_{w \in V} \text{count}(w,C) + aV}$$

Di mana:

$\text{count}(w,C)$ adalah jumlah kemunculan data w dalam kelas C

a adalah nilai smoothing (Laplace Smoothing, biasanya $a=1$)

V adalah jumlah total kata unik dalam seluruh dokumen

Pada tahap ini juga dilakukan evaluasi untuk mengevaluasi hasil dari algoritma yang digunakan. Metode evaluasi yang digunakan Confusion Matrix Metode ini cukup membantu untuk proses analisis kualitas classifier. Setelah Confusion matrix dilakukan maka nilai bisa dilihat dan dihitung nilai accuracy, recall, precision dan f1-score yang akan ditampilkan dalam bentuk presentase.[11]

4. Hasil dan Pembahasan

4.1 Pengumpulan data

Proses web scraping dilakukan menggunakan *Google Colaboratory* dan bahasa pemrograman Python untuk mengumpulkan 1.000 ulasan pengguna aplikasi Shopee dari Google Play Store. Pengambilan data memanfaatkan library *google-play-scraper* dengan memasukkan *ID* aplikasi Shopee, kemudian hasilnya diekspor ke dalam format Excel (.xlsx) untuk digunakan pada tahap pengolahan data selanjutnya.

reviewId	userName	userImage	content	score	thumbsUpCount	reviewCreatedVersion	at	replyContent	repliedAt	app
0	S285a3b3-604-44d5-fbca-9e2c30d2c17	Sayidina Almasi Reza	https://play- th.googleusercontent.com/va-ALV-U...	Ini kenapa ya aplikasi selama di update kok j...	1	117	3.61.28 11-04 13:49:54	Hi kak, maaf buat kecewa Saat ini sistem Shop...	2022-10- 10 04:04:01	
1	c4f0945d1-766c-4139-121c-d54230a5ab	Eddie ...	https://play- th.googleusercontent.com/va-ALV-U...	Sudah berubah shopee aplikasi layanan tidak la...	1	88	3.61.28 11-05 07:25:09	Hi kak, Eddie ... maaf ya udah buat kaka ga ny...	2025-11- 05 09:13:04	
2	42548d2e-1d20-47f5-8847-1a967da179dc	Sudiroja	https://play- th.googleusercontent.com/va-ALV-U...	balasan chat AI sungguh SANGAT tidak berguna, ...	1	113	3.61.28 11-03 03:50:35	Hi kak, maaf Sudiroja, maaf buat kamu ga nyaman ter...	2025-11- 05 07:17:02	
3	28a3ba6b-57a2-4452-804e-a7b7209152bb	Shanum	https://play- th.googleusercontent.com/va/ACgloc...	sebenarnya shopee udah cukup bagus hanya saja ...	4	41	3.60.30 11-03 23:15:31	None	NaT	
4	9c3587cd-2c8f-4bdf-9776-c0632054439d	Arisi Dina	https://play- th.googleusercontent.com/va-ALV-U...	Saya penjual Agak games ama permainan chat, ...	3	56	3.61.28 11-04 00:28:31	None	NaT	

Gambar 2. Web Scrapping

Terlihat pada Gambar 2 pengumpulan data dengan web scrapping menghasilkan beberapa kolom yang berisi reviewId, userName, userImage, content, score, thumbsUpCount, reviewCreatedVersion, at, replyContent, repliedAt, appversion. Dikarenakan terlalu banyak kolom dan yang dibutuhkan hanya dua, datanya menjadi sebagai berikut.

	content	score
76	Buruk sekali shopee ini, kalian masih menahan ...	1
1	Sudah berubah shopee aplikasi layanan tidak la...	1
44	Aplikasinya sudah bagus, mungkin lebih bagus d...	5
52	Untuk keseluruhannya sudah baik. Shopee pay li...	5
206	shopee makin ngaco pengiriman instan waktu pen...	1
51	untuk sistem pembelian mudah dan senang di m...	4
32	sangat kecewa dengan shopee karena melegalkan ...	3

Gambar 3. Web Scrapping filter

Pada Gambar 3 terlihat bahwa datanya menjadi lebih mudah untuk diproses selanjutnya, pada data tersebut hanya ada content dan score yang di maksud dengan content yaitu 'review', kalau score di maksud bintang, seperti bintang 1 dan 2. Selanjutnya dilakukan pelabelan data, proses ini bertujuan membagi data menjadi tiga kelas sentimen, yaitu sentimen positif, sentimen negatif, dan sentimen none, namun dalam penelitian ini yang digunakan hanyalah sentimen positif dan negatif.[10]

	content	score	Label
76	Buruk sekali shopee ini, kalian masih menahan ...	1	Negatif
1	Sudah berubah shopee aplikasi layanan tidak la...	1	Negatif
44	Aplikasinya sudah bagus, mungkin lebih bagus d...	5	Positif
52	Untuk keseluruhannya sudah baik. Shopee pay li...	5	Positif
206	shopee makin ngaco pengiriman instan waktu pen...	1	Negatif
51	untuk sistem pembelian mudah dan senang di m...	4	Positif
32	sangat kecewa dengan shopee karena melegalkan ...	3	None
539	ini gimana sih kok mentok di jasa kirim untuk ...	1	Negatif
47	shopee kembalikan cod jangan pakai paylater, d...	3	None
467	Untuk belanja di Shopee tidak perlu diragukan ...	5	Positif

Gambar 4. Pelabelan data

Gambar 4 menunjukkan jika if score kurang dari < 3 maka "negatif" kemudian elif, jika score nya = 4 maka "positif", elif score nya juga sama = 5 maka hasilnya sama juga "positif". disini if score 3 tidak dimasukkan atau kita buang. Kenapa karna score 3 atau bintang 3 itu seperti ambigu, seperti ada yang masih komentar positif dan negatif. Dapat dilihat pada gambar 4, pelabelan kelas sentimen berdasarkan kata positif dan negatif yang terkandung dalam shopee. Pada teks komentar "Buruk sekali shopee ini, kalian masih menahan..." terdapat kata negatif dan tidak terdapat kata positif sehingga skor sentimennya bernilai -1 yang berlabel negatif. Setelah dilakukan pelabelan data, selanjutnya akan dilakukakn cleaning sebagai berikut.

	content	score	Label
0	Buruk sekali shopee ini, kalian masih menahan ...	1	Negatif
1	Sudah berubah shopee aplikasi layanan tidak la...	1	Negatif
2	Aplikasinya sudah bagus, mungkin lebih bagus d...	5	Positif
3	Untuk keseluruhannya sudah baik. Shopee pay li...	5	Positif
4	shopee makin ngaco pengiriman instan waktu pen...	1	Negatif
5	untuk sistem pembelian mudah dan senang di m...	4	Positif
7	ini gimana sih kok mentok di jasa kirim untuk ...	1	Negatif
9	Untuk belanja di Shopee tidak perlu diragukan ...	5	Positif
10	pengiriman shopee sekarang cepat dan pengembal...	5	Positif
11	Ini kenapa ya aplikasi setelah di update kok j...	1	Negatif
12	kurirnya payah, di hubungi susah, suda di hubu...	1	Negatif
13	pengiriman sepihak dicancel padahal isinya ser...	1	Negatif
14	pengiriman sering terlambat. shopee aplikasi l...	1	Negatif
15	spx hemat sangat mengecewakan.. paket saya ser...	1	Negatif

Gambar 5. Cleaning

Gambar 5 dilakukan tahap cleaning yang berfungsi untuk menghilangkan nilai kosong pada data. Label yang semula memiliki sentimen positif, negatif dan none menjadi hanya sentimen positif dan negatif saja. Label kosong yang dimaksud adalah score 3 karena sebelumnya score tersebut dinilai sebagai sentimen none.

4.2 Preprocessing data

4.2.1 Case folding

Proses ini dilakukan untuk mentransformasikan semua huruf menjadi huruf kecil guna menyeragamkan hasil teks.[9]

	content	score	Label	text_clean
0	Buruk sekali shopee ini, kalian masih menahan ...	1	Negatif	buruk sekali shopee ini kalian masih menahan u...
1	Sudah berubah shopee aplikasi layanan tidak la...	1	Negatif	sudah berubah shopee aplikasi layanan tidak la...
2	Aplikasinya sudah bagus, mungkin lebih bagus d...	5	Positif	aplikasinya sudah bagus mungkin lebih bagus da...
3	Untuk keseluruhannya sudah baik. Shopee pay li...	5	Positif	untuk keseluruhannya sudah baik shopee pay lim...
4	shopee makin ngaco pengiriman instan waktu pen...	1	Negatif	shopee makin ngaco pengiriman instan waktu pen...
5	untuk sistem pembelajaran mudah dan senang di m...	4	Positif	untuk sistem pembelajaran mudah dan senang di m...
7	ini gimana sih kok mentok di jasa kirim untuk ...	1	Negatif	ini gimana sih kok mentok di jasa kirim untuk ...
9	Untuk belanja di Shopee tidak perlu diragukan ...	5	Positif	untuk belanja di shopee tidak perlu diragukan ...
10	pengiriman shopee sekarang cepat dan pengembal...	5	Positif	pengiriman shopee sekarang cepat dan pengembal...
11	Ini kenapa ya aplikasi setelah di update kok j...	1	Negatif	ini kenapa ya aplikasi setelah di update kok j...

Gambar 6. Case folding

Gambar 6 terlihat hasil dari proses case folding “Buruk” menjadi “buruk”. Begitu juga dengan data yang lain dari huruf besar menjadi huruf kecil.

4.2.2 Stopword removal

Penghapusan stopwords adalah proses menghapus kata-kata dan karakter yang terdapat dalam kategori stopwords dan tanda baca. Stopword sendiri adalah kata-kata yang sering digunakan tetapi tidak banyak mempengaruhi emosi yang terkandung dalam sebuah kalimat [17]

	content	score	Label	text_clean	text_stopword
0	Buruk sekali shopee ini, kalian masih menahan ...	1	Negatif	buruk sekali shopee ini kalian masih menahan u...	buruk shopee menahan uangku shopeepay tp login...
1	Sudah berubah shopee aplikasi layanan tidak la...	1	Negatif	sudah berubah shopee aplikasi layanan tidak la...	berubah shopee aplikasi layanan ekspedisi resp...
2	Aplikasinya sudah bagus, mungkin lebih bagus d...	5	Positif	aplikasinya sudah bagus mungkin lebih bagus da...	aplikasinya bagus bagus lengkap shopee menyedi...
3	Untuk keseluruhannya sudah baik. Shopee pay li...	5	Positif	untuk keseluruhannya sudah baik shopee pay lim...	shopee pay limit extra s pinjam dipakai limitn...
4	shopee makin ngaco pengiriman instan waktu pen...	1	Negatif	shopee makin ngaco pengiriman instan waktu pen...	shopee ngaco pengiriman instan pengemasannya j...
5	untuk sistem pembelajaran mudah dan senang di m...	4	Positif	untuk sistem pembelajaran mudah dan senang di m...	sistem pembelajaran mudah senang mengerti cuman...
7	ini gimana sih kok mentok di jasa kirim untuk ...	1	Negatif	ini gimana sih kok mentok di jasa kirim untuk ...	gimana sih mentok jasa kirim proses nya banget...
9	Untuk belanja di Shopee tidak perlu diragukan ...	5	Positif	untuk belanja di shopee tidak perlu diragukan ...	belanja shopee diragukan pengiriman aplikasi n...
10	pengiriman shopee sekarang cepat dan pengembal...	5	Positif	pengiriman shopee sekarang cepat dan pengembal...	pengiriman shopee cepat pengembalian retur bar...

Gambar 7. Stopword removal

Dapat dilihat pada Gambar 7 kalimat pada kolom content “Buruk sekali shopee ini, kalian masih menahan uangku di shopeepay tp saat aku ingin login lagi ke akunku di shopee aku sempat di persulit untuk masuk ke akunku lagi seolah olah akunku itu sudah kadarluasa terus kalau login ke shopee pasti ngebug dan bnyk jebakan batmannya.”, setelah dilakukan stopwords removal menjadi “buruk shopee menahan uangku shopeepay tp login akunku shopee persulit masuk akunku olah akunku kadarluasa login shopee ngebug bnyk jebakan batmannya”

4.2.3 Tokenizing

Proses ini dilakukan untuk memecah atau memotong suatu kalimat menjadi term-term yang dipisahkan berdasarkan spasi[7]

	content	score	Label	text_clean	text_stopword	text_tokens
0	Buruk sekali shopee ini, kalian masih menahan ...	1	Negatif	buruk sekali shopee ini kalian masih menahan u...	buruk shopee menahan uangku shopeepay tp login...	[buruk, shopee, menahan, uangku, shopeepay, tp, login, akunku, shopee, persulit, masuk, akunku, olah, akunku, kadarluasa, login, shopee, ngebug, bnyk, jebakan, batmannya]
1	Sudah berubah shopee aplikasi layanan tidak la...	1	Negatif	sudah berubah shopee aplikasi layanan tidak la...	berubah shopee aplikasi layanan ekspedisi resp...	[berubah, shopee, aplikasi, layanan, ekspedisi, resp...]
2	Aplikasinya sudah bagus, mungkin lebih bagus d...	5	Positif	aplikasinya sudah bagus mungkin lebih bagus da...	aplikasinya bagus bagus lengkap shopee menyedi...	[aplikasinya, bagus, bagus, lengkap, shopee, menyedi...]
3	Untuk keseluruhannya sudah baik. Shopee pay li...	5	Positif	untuk keseluruhannya sudah baik shopee pay lim...	shopee pay limit extra s pinjam dipakai limitn...	[shopee, pay, limit, extra, s, pinjam, dipakai, limitn...]
4	shopee makin ngaco pengiriman instan waktu pen...	1	Negatif	shopee makin ngaco pengiriman instan waktu pen...	shopee ngaco pengiriman instan pengemasannya j...	[shopee, ngaco, pengiriman, instan, pengemasannya, j...]

Gambar 8. Tokenizing

Pada Gambar 8 dapat dilihat kalimat berubah menjadi terpisah “[‘buruk’, ‘shopee’, ‘menahan’, ‘uangku’, ‘shopeepay’, ‘tp’, ‘login’, ‘akunku’, ‘shopee’, ‘persulit’, ‘masuk’, ‘akunku’, ‘olah’, ‘akunku’, ‘kadarluasa’, ‘login’, ‘shopee’, ‘ngebug’, ‘bnyk’, ‘jebakan’, ‘batmannya’]”

4.2.4 Stemming

Tahap stemming berfungsi untuk merubah kata-kata menjadi bentuk dasar mereka. Proses tersebut dilakukan karena kata berimbuhan memiliki makna yang sama dengan kata dasar[17]


```

4633
-----
1 : buruk : buruk
2 : shopee : shopee
3 : menahan : tahan
4 : uangku : uang
5 : shopeepay : shopeepay
6 : tp : tp
7 : login : login
8 : akunku : akun
9 : persulit : sulit
10 : masuk : masuk
11 : olah : olah
12 : kadarluasa : kadarluasa
13 : ngebug : ngebug
14 : bnyk : bnyk
15 : jebakan : jebak

```

Gambar 9. Stemming

Gambar 9 menunjukkan hasil dari proses stemming dapat dilihat. Untuk kata seperti menahan maka nanti diganti menjadi tahan, lalu misalkan ada kata akunku maka diubah menjadi akun.

4.3 Transfromasi data

ada tahap ini dilakukan pemecahan data yang disebut dengan splitting data dengan memecah data test 20% dari keseluruhan data. Ini menghasilkan data testing sebesar 20% dan data training sebesar 80%. Lalu setelah splitting data dilakukan transformasi data menggunakan metode TF-IDF untuk mengubah teks menjadi bentuk numerik dengan menghitung bobot setiap kata berdasarkan frekuensinya

```

print(x_train.shape)
print(y_train.shape)
print(x_test.shape)
print(y_test.shape)

(727, 4211)
(727,)
(182, 4211)
(182,)

```

Gambar 10. Splitting data

Dari hasil Gambar 10 proses splitting, diperoleh total 909 data komentar, yang dibagi menjadi 727 data training (80%) dan 182 data testing (20%). Selanjutnya, data teks tersebut ditransformasikan menjadi matriks TF-IDF dengan ukuran (727, 4211) untuk data training dan (182, 4211) untuk data testing. Angka 727 dan 182 menunjukkan jumlah komentar yang digunakan pada masing-masing bagian, sedangkan 4211 menunjukkan jumlah kata unik (unique terms) yang dihasilkan dari proses ekstraksi fitur.

Selain itu, label sentimen juga dipisahkan menjadi y_{train} (727,) dan y_{test} (182,), dengan masing-masing berisi dua kategori sentimen, yaitu positif dan negatif. Hasil transformasi ini kemudian digunakan sebagai input pada tahap pelatihan dan evaluasi model klasifikasi sentimen.

4.4 Klasifikasi Naïve Bayes

Setelah model Naive Bayes dilatih menggunakan dataset yang telah melalui tahap preprocessing, evaluasi dilakukan untuk mengukur performa model dalam mengklasifikasikan sentimen pengguna terhadap aplikasi Info Shopee. Hasil metrik evaluasi Naïve Bayes ditampilkan pada Tabel 1 berikut:

confusion_matrix:				
[[86 4]				
[23 69]]				
=====				
	precision	recall	f1-score	support
Negatif	0.79	0.96	0.86	90
Positif	0.95	0.75	0.84	92
accuracy			0.85	182
macro avg	0.87	0.85	0.85	182
weighted avg	0.87	0.85	0.85	182

Gambar 11. Klasifikasi Naïve Bayes dan Confusion matrix

Berdasarkan Gambar 11 *confusion matrix* yang dihasilkan, terdapat 86 data Negatif yang terklasifikasi benar (True Negative), 69 data Positif yang terklasifikasi benar (True Positive), 4 data Negatif yang salah diklasifikasikan sebagai Positif (False Positive), serta 23 data Positif yang salah diklasifikasikan sebagai Negatif (False Negative). Nilai *precision* yang diperoleh untuk kelas Negatif sebesar 0,79 menunjukkan bahwa 79% dari data yang diprediksi

sebagai Negatif benar-benar termasuk ke dalam kelas tersebut, sedangkan nilai *recall* sebesar 0,96 mengindikasikan bahwa model mampu mengenali hampir seluruh data Negatif dengan baik. Untuk kelas Positif, nilai *precision* sebesar 0,95 dan *recall* sebesar 0,75 menunjukkan bahwa sebagian besar data yang diprediksi sebagai Positif benar, namun masih terdapat beberapa data Positif yang tidak terdeteksi dengan baik. Nilai *f1-score* untuk masing-masing kelas adalah 0,86 untuk Negatif dan 0,84 untuk Positif, dengan rata-rata keseluruhan (*macro average*) sebesar 0,85.

```
MultinomialNB Accuracy: 0.8516483516483516
MultinomialNB Precision: 0.7889908256880734
MultinomialNB Recall: 0.9555555555555556
MultinomialNB f1_score: 0.864321608040201
```

Gambar 11. Performa Naïve Bayes

Pada Gambar 11 menggambarkan performa model secara keseluruhan terhadap kelas target. Nilai accuracy sebesar 0,85, precision sebesar 0,79, recall sebesar 0,95, dan f1-score sebesar 0,86 menunjukkan bahwa model Multinomial Naive Bayes memiliki performa klasifikasi yang baik. Model mampu mengidentifikasi data Negatif dengan tingkat ketepatan yang tinggi dan hanya sedikit kesalahan dalam mengenali kelas tersebut. Meskipun demikian, masih terdapat kemungkinan kecil terjadinya kesalahan prediksi terhadap beberapa data Positif.

5. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa penerapan metode Multinomial Naive Bayes dalam analisis sentimen ulasan pengguna aplikasi Shopee di Google Play Store mampu memberikan hasil klasifikasi yang cukup baik. Proses penelitian diawali dengan pengumpulan 1.000 data ulasan pengguna melalui teknik *web scraping* menggunakan bahasa pemrograman Python di Google Colaboratory. Data kemudian melalui tahapan *preprocessing* yang meliputi *case folding*, *stopword removal*, *tokenizing*, dan *stemming*, serta proses transformasi menggunakan metode TF-IDF untuk mengubah teks menjadi representasi numerik yang siap digunakan dalam pemodelan. Setelah dilakukan pelatihan dan pengujian data dengan pembagian 80% data latih dan 20% data uji, model menghasilkan nilai akurasi sebesar 0,85, precision sebesar 0,79, recall sebesar 0,95, dan f1-score sebesar 0,86.

Nilai tersebut menunjukkan bahwa model mampu mengklasifikasikan sentimen ulasan dengan tingkat keakuratan yang tinggi dan keseimbangan performa yang baik antara *precision* dan *recall*. Model terbukti lebih unggul dalam mengenali ulasan dengan sentimen Negatif dibandingkan Positif, di mana recall untuk kelas Negatif mencapai 0,96, sedangkan untuk kelas Positif sebesar 0,75. Hal ini menandakan bahwa model masih cenderung berhati-hati dalam memprediksi ulasan Positif dan cenderung melewatkan sebagian data yang seharusnya termasuk dalam kelas tersebut. Secara keseluruhan, metode Multinomial Naive Bayes terbukti efektif dalam melakukan klasifikasi sentimen pada teks ulasan aplikasi Shopee.

References

- [1] J. Edukasi, B. Gunawan, H. S. Pratiwi, and E. E. Pratama, "Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes," vol. 4, no. 2, pp. 113–118, 2018.
- [2] E. Subowo, F. A. Artanto, I. Putri, and W. Umaedi, "BLTSM untuk analisis sentimen berbasis aspek pada aplikasi belanja online dengan cicilan," vol. XII, no. Ii, pp. 132–140, 2022.
- [3] C. Cahyaningtyas, Y. Nataliani, I. R. Widiyari, F. T. Informasi, U. Kristen, and S. Wacana, "Analisis sentimen pada rating aplikasi Shopee menggunakan metode Decision Tree berbasis SMOTE," vol. 18, no. 2, pp. 173–184, 2021.
- [4] E. H. Muktafin and E. T. Luthfi, "Analisis Sentimen pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing," pp. 32–42, 2020, doi: 10.30864/eksplora.v10i1.390.
- [5] J. J. A. Limbong, I. Sembiring, K. D. Hartomo, U. Kristen, and S. Wacana, "Analisis Klasifikasi Sentimen Ulasan pada E-Commerce Shopee Berbasis Word Cloud dengan Metode Naive Bayes dan K-Nearest Neighbor Analysis of Review Sentiment Classification on E-Commerce Shopee Word Cloud Based with Naïve Bayes and K-Nearest Neighbor Methods," vol. 9, no. 2, pp. 347–356, 2022, doi: 10.25126/jtiik.202294960.
- [6] M. Saifurridho, Martanto, and U. Hayati, "Analisis Algoritma K-Nearest Neighbor terhadap Sentimen Pengguna Aplikasi Shopee," vol. 10, no. 1, pp. 21–26, 2024. doi:https://doi.org/10.54914/jit.v10i1.1054
- [7] R. L. Simanjuntak, "Analisis Sentimen Ulasan Pada Aplikasi E-Commerce Shopee Dengan Menggunakan Algoritma Naïve Bayes," vol. 3, no. 3, 2023. doi:https://doi.org/10.55606/teknik.v3i3.2411
- [8] A. Nurusyifa, D. N. W. Ramadhani, E. D. Trisnadewi, E. Octareny, F. N. Afifah, G. N. Fadillah, L. T. Salsabiella, M. Indriayu, M. F. K. Ishaq, N. A. Pramudiasari, and R. Qurniawan, "Membangun toko kelontong berbasis digital di palur kulon dengan Loyverse POS," pp. 76–87, 2024. doi:xxxxx
- [9] N. A. C. A, D. H. Citra, and W. Purnama, "The Implementation of Naive Bayes Algorithm for Sentiment Analysis of Shopee Reviews on Google Play Store Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store," vol. 2, no. April, pp. 47–54, 2022. doi: https://doi.org/10.57152/malcom.v2i1.195
- [10] N. Aditiya, P. Setiaji, and S. Supriyono, "Analisis Sentimen Kepuasan Masyarakat terhadap Aplikasi ‘

- INFO BMKG ' menggunakan Naive Bayes , SVM , dan KNN Sentiment Analysis of Public Satisfaction with the ' INFO BMKG ' ,” vol. 14, pp. 1418–1432, 2025. doi:<https://doi.org/10.32520/stmsi.v14i3.5223>
- [11] T. Puspa, R. Sanjaya, A. Fauzi, A. Fitri, and N. Masruriyah, “Analisis sentimen ulasan pada e-commerce shopee menggunakan algoritma naive bayes dan support vector machine Analysis of review sentiment on shopee e-commerce using the naive bayes algorithm and support vector machine,” vol. 4, pp. 16–26, 2023, doi: 10.37373/infotech.v4i1.422.
- [12] V. No, “Edumatic : Jurnal Pendidikan Informatika,” vol. 5, no. 2, pp. 233–242, 2021, doi: 10.29408/edumatic.v5i2.4089.
- [13] E. Febriyanti, Z. Rifai, S. N. Hasanah, D. F. Wijayanti, L. Suhening, and L. Frastika, “Pemanfaatan media sosial facebook sebagai media pemasaran produk,” *Conf. Inf. Technol. Inf. Syst. Electr. Eng.*, pp. 254–257, 2017.
- [14] K. S. Putri, I. R. Setiawan, A. Pambudi, A. Sentimen, and N. B. Classifier, “‘Technologia’ Vol 14, No. 3, Juli 2023 227 Analisis sentimen terhadap brand skincare lokal menggunakan Naïve Bayes Classifier,” vol. 14, no. 3, pp. 227–232, 2023.
- [15] N. Putu, G. Naraswati, D. C. Rosmilda, D. Desinta, P. D. Statistika, and P. S. Stis, “Analisis Sentimen Publik dari Twitter Tentang Kebijakan Penanganan Covid-19 di Indonesia dengan Naive Bayes Classification,” vol. 10, pp. 228–238, 2021.
- [16] B. Z. Ramadhan, I. Riza, and I. Maulana, “Analisis Sentimen Ulasan Pada Aplikasi E-Commerce Dengan Menggunakan Algoritma Naïve Bayes,” vol. 6, no. 2, pp. 220–225, 2022.
- [17] F. P. Saputra and O. Suria, “Penggabungan Model SVM dan Naive Bayes dengan Pendekatan Soft Voting untuk Analisis Sentimen Tong Tji Tea House Combining SVM and Naive Bayes Models using a Soft Voting Approach for Sentiment Analysis of Tong Tji Tea House,” vol. 14, pp. 2479–2491, 2025.
- [18] Z. Ayu Nuansa Gumilang, “Implementasi Naïve Bayes Classifier dan Asosiasi untuk Analisis Sentimen Data Ulasan Aplikasi E-Commerce Shopee pada Situs Google Play,” 2018.