Bulletin of Artificial Intelligence
Volume 4, No 2, October 2025 Page: 80-89
ISSN 2962-3944 (media online)
DOI: 10.62866/buai.v4i2.237
https://journal.grahamitra.id/index.php/buai

Analisis Sentimen Masyarakat Terhadap Aplikasi Shopee dengan Menggunakan Algoritma Regresi Logistik

Auralia Izmi*, Haya Atiqah Tampubolon, Nur Rifki TriAji, Heru Satria Tambunan

Teknik Informatika, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia
Email: 1,*auraliaizmi489@gmail.com, 2hayaatiqah608@gmail.com, 3nurrifkitriaji@gmail.com, 4heru@amiktunasbangsa.ac.id
Email Penulis Korespondensi: auraliaizmi489@gmail.com

Abstrak−Penelitian ini menganalisis sentimen masyarakat terhadap aplikasi Shopee dengan menggunakan algoritma regresi logistik. Data review diambil dari Google Play Store dengan menggunakan teknik scraping, kemudian diolah melalui tahapan pra-pemrosesan seperti tokenisasi, normalisasi, stopword removal, dan stemming, untuk memastikan data bersih dan relevan. Metode Term Frekuensi-Invers Dokumen Frekuensi (TF-IDF) digunakan untuk menentukan bobot setiap term dalam dokumen. Data diklasifikasikan menggunakan model regresi logistik, yang diuji melalui tiga skenario distribusi data pelatihan dan pengujian: 80:20, 70:30, dan 60:40. Hasilnya menunjukkan bahwa model mencapai akurasi tertinggi pada skenario 80:20 sebesar 85%, diikuti oleh 82% pada skenario 70:30, dan 80% pada skenario 60:40. Precision dan recall juga menunjukkan kinerja yang baik dalam mengklasifikasikan sentimen positif, negatif dan netral. Penelitian ini memberikan wawasan penting mengenai pola sentimen masyarakat terhadap aplikasi Shopee, termasuk kepuasan pengguna dan tantangan yang dihadapi terhadap layanan tersebut. Penelitian ini membuktikan bahwa algoritma regresi logistik efektif dalam menangani klasifikasi sentimen dan memberikan hasil yang mudah diinterpretasikan. Namun terdapat keterbatasan dalam menangani data dengan distribusi kategori yang tidak seimbang. Penelitian di masa depan dapat menggunakan algoritma lain seperti Random Forest atau Deep Learning untuk meningkatkan akurasi dan performa model dalam menangani data yang lebih kompleks.

Kata Kunci: Sentimen; Regresi Logistik; Shopee; TF-IDF; Analisis Sentimen

Abstract−This study analyzes public sentiment toward the Shopee application using a logistic regression algorithm. Review data were collected from the Google Play Store via web scraping and subsequently processed through several pre-processing stages—including tokenization, normalization, stopword removal, and stemming—to ensure the data were clean and relevant. The Term Frequency-Inverse Document Frequency (TF-IDF) method was employed to assign weights to each term within the documents. The data were classified using a logistic regression model, which was evaluated across three different training-to-testing data distribution scenarios: 80:20, 70:30, and 60:40. The results indicate that the model achieved its highest accuracy—85%—under the 80:20 scenario, followed by 82% for the 70:30 scenario and 80% for the 60:40 scenario. Precision and recall metrics also demonstrated strong performance in classifying positive, negative, and neutral sentiments. This research provides valuable insights into public sentiment patterns regarding the Shopee application, highlighting both user satisfaction and the challenges encountered with the service. The study confirms that the logistic regression algorithm is effective for sentiment classification and yields results that are easily interpretable. However, it faces limitations when dealing with datasets characterized by an imbalanced distribution of categories. Future research could explore other algorithms, such as Random Forest or Deep Learning, to enhance model accuracy and performance when handling more complex datasets.

Keywords: Sentiment; Logistic Regression; Shopee; TF-IDF; Sentiment Analysis

1. PENDAHULUAN

Industri e-commerce di Indonesia mengalami pertumbuhan pesat dalam beberapa tahun terakhir, yang didorong oleh meningkatnya penggunaan internet dan perangkat mobile. Salah satu platform e-commerce yang mendominasi pasar Indonesia adalah Shopee [1]. Aplikasi ini tidak hanya memfasilitasi kebutuhan berbelanja secara daring, tetapi juga mendukung gaya hidup digital masyarakat serta turut membentuk ulang pola perilaku konsumen modern [2]. Berdasarkan laporan dari iPrice, Shopee tercatat sebagai marketplace terpopuler di Indonesia pada kuartal pertama tahun 2020, dengan rata-rata kunjungan bulanan mencapai 71 juta pengguna [3]. Angka ini mencerminkan tingginya keterlibatan masyarakat dengan aplikasi tersebut, sekaligus memberikan peluang besar untuk menganalisis opini dan sentimen yang diungkapkan oleh para pengguna.Salah satu metode untuk memahami opini publik terhadap suatu produk atau layanan adalah melalui analisis sentimen. Analisis sentimen merupakan studi berbasis komputer yang bertujuan memahami opini, sikap, dan perasaan seseorang terhadap suatu entitas melalui data teks [4].

Dalam konteks ini, analisis sentimen sangat berguna untuk mengevaluasi ulasan atau komentar pengguna, yang dapat digunakan untuk menentukan polaritas sentimen positif, negatif, atau netral dalam suatu dokumen atau pernyataan [5].Berbagai algoritma machine learning telah diterapkan dalam analisis sentimen, di antaranya Naïve Bayes dan Support Vector Machine (SVM) [6]. Selain itu, metode seperti K-Nearest Neighbor (KNN), Decision Tree, dan Linear Discriminant Analysis (LDA) juga sering digunakan untuk klasifikasi teks [7]. Setiap metode memiliki kelebihan dan kelemahannya masing-masing. Naïve Bayes, misalnya, adalah metode statistik yang menghitung probabilitas suatu hipotesis dengan mempertimbangkan kemungkinan atribut yang diamati. Metode ini terkenal karena mudah diimplementasikan dan memiliki efisiensi tinggi dalam perhitungan [8], serta mampu menangani data yang tidak lengkap [9]. Namun, kelemahan utamanya terletak pada asumsi bahwa setiap atribut saling independen, yang tidak selalu sesuai dengan data nyata [10]. Di sisi lain, SVM adalah algoritma pembelajaran terawasi yang menawarkan akurasi tinggi dalam klasifikasi. Kekuatan utama SVM terletak pada kemampuannya untuk menentukan hyperplane optimal yang memisahkan data dari berbagai kelas dengan margin maksimal [11]. Namun, SVM dapat mengalami

© 2025 The Author(s). Published by *CV. Graha Mitra Edukasi*. Page 80
This work is licensed under a Creative Commons Attribution 4.0 International License.

Bulletin of Artificial Intelligence
Volume 4, No 2, October 2025 Page: 80-89
ISSN 2962-3944 (media online)
DOI: 10.62866/buai.v4i2.237
https://journal.grahamitra.id/index.php/buai

penurunan performa ketika fitur antar kelas sangat mirip, sehingga membingungkan model dalam melakukan klasifikasi[12][13].

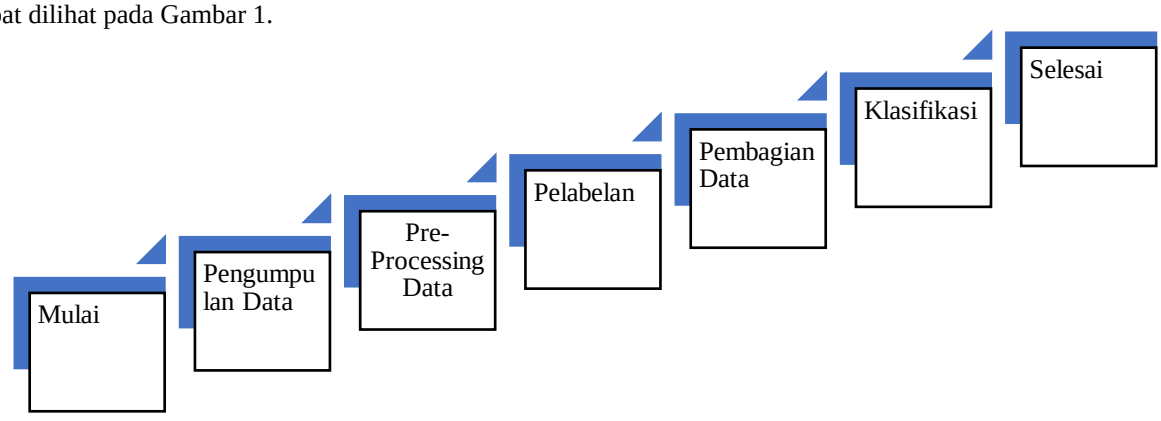
Beberapa penelitian sebelumnya juga telah memanfaatkan data dari Google Play Store maupun media sosial untuk melakukan analisis sentimen[14]. Sebagai contoh, penelitian oleh Dedi Darwis dan tim menggunakan algoritma Naïve Bayes untuk mengklasifikasikan opini masyarakat terhadap BMKG melalui Twitter. Dengan menggunakan Python 3.7.4 dan teknik text mining, mereka melakukan fine-grained sentiment analysis terhadap tweet dan berhasil mencapai akurasi sebesar 69,97% [15]. Penelitian ini menunjukkan bahwa meskipun metode Naïve Bayes sederhana, hasilnya tetap dapat diandalkan dalam konteks tertentu.Algoritma lain yang cukup populer dalam analisis sentimen adalah Regresi Logistik. Teknik ini digunakan untuk menjelaskan hubungan antara satu variabel dependen dengan satu atau lebih variabel independen, baik berupa data kontinu maupun kategorik[16].Regresi logistik terdiri dari tiga jenis, yaitu regresi logistik biner[17], regresi logistik multinomial, dan regresi logistik ordinal [18]. Model regresi logistik biner, misalnya, digunakan untuk menganalisis hubungan antara variabel prediktor dengan variabel respon dikotomis, yaitu nilai 1 untuk menunjukkan adanya karakteristik, dan 0 untuk ketidakhadirannya [19] [20]. Model ini mengikuti distribusi Bernoulli[21] [22] [23]. Logistic Regression menjadi algoritma yang banyak digunakan karena memiliki struktur sederhana dan hasil yang mudah diinterpretasikan [24]. Selain itu, algoritma ini mampu menangani dataset besar dengan jumlah fitur yang banyak tanpa mudah mengalami overfitting. Namun, dalam kasus yang lebih kompleks, model klasifikasi lain seperti Decision Tree, Random Forest, atau Deep Learning dapat memberikan hasil yang lebih akurat tergantung pada struktur dan kompleksitas datanya[25].

Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap ulasan pengguna aplikasi Shopee di Google Play Store dengan menggunakan algoritma Regresi Logistik. Melalui pendekatan ini, penelitian berusaha untuk memahami isi ulasan pengguna, mengidentifikasi pola-pola sentimen yang muncul, dan menyediakan wawasan yang bermanfaat bagi pengembang aplikasi dalam merancang strategi peningkatan layanan. Penelitian ini juga mengisi kekosongan literatur dalam konteks lokal, serta berkontribusi terhadap pengembangan dan pemahaman efektivitas Regresi Logistik dalam analisis sentimen di sektor e-commerce. Hasil dari studi ini diharapkan tidak hanya relevan secara akademik, tetapi juga memberikan manfaat praktis bagi pengambil kebijakan dan pelaku industri digital dalam meningkatkan kepuasan pengguna.

2. METODE PENELITIAN

2.1 Kerangka Dasar Penelitian

Penelitian ini memanfaatkan metode analisis sentimen dengan algoritma Regresi Logistik untuk mengelompokkan sentimen masyarakat Indonesia terhadap aplikasi Shopee. Adapun langkah-langkah yang dilakukan dalam penelitian dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

Pada Gambar 1 kerangka penelitian, berikut langkah-langkah proses kegiatan analisis sentimen:

a. Pengumpulan Data
   Data dikumpulkan dari sumber yang relevan, dalam hal ini adalah ulasan pengguna aplikasi Shopee di Google Play Store. Proses ini melibatkan scraping atau pengambilan data menggunakan tools seperti Python dengan library tertentu (misalnya google-play-scraper atau BeautifulSoup).
b. Pre-Processing Data
   Data mentah dari hasil pengumpulan selanjutnya dibersihkan dan diproses agar siap untuk dianalisis. Tahapan ini meliputi:
   1. Case folding (mengubah teks menjadi huruf kecil semua),
   2. Cleansing (menghapus simbol, angka, dan karakter khusus),
   3. Tokenizing (memecah teks menjadi kata-kata),
   4. Stopword removal (menghapus kata-kata umum yang tidak penting),

© 2025 The Author(s). Published by *CV. Graha Mitra Edukasi*.
This work is licensed under a Creative Commons Attribution 4.0 International License.

Bulletin of Artificial Intelligence
Volume 4, No 2, October 2025 Page: 80-89
ISSN 2962-3944 (media online)
DOI: 10.62866/buai.v4i2.237
https://journal.grahamitra.id/index.php/buai

5. Stemming (mengembalikan kata ke bentuk dasarnya).

c. Pelabelan
Setelah data diproses, setiap data diberi label berdasarkan polaritas sentimen, yaitu:
   a. Positif
   b. Negatif
   c. (Opsional: Netral, tergantung pada skema klasifikasi yang digunakan)

   Pelabelan bisa dilakukan secara manual oleh peneliti atau otomatis jika tersedia dataset yang telah dilabeli sebelumnya.

d. Pembagian Data
Dataset yang sudah dilabeli kemudian dibagi menjadi dua subset utama:
   a. Data Latih (Training Data)
   b. Data Uji (Testing Data)

   Pembagian ini umumnya dilakukan dengan rasio 80:20 atau 70:30.

e. Klasifikasi
Pada tahap ini dilakukan proses klasifikasi menggunakan algoritma Regresi Logistik. Model dilatih menggunakan data latih, lalu diuji menggunakan data uji untuk mengevaluasi performa model. Evaluasi bisa menggunakan metrik seperti:
   a. Akurasi
   b. Presisi
   c. Recall
   d. F1-score

Proses analisis berakhir setelah hasil klasifikasi diperoleh dan evaluasi performa model dilakukan. Hasil ini akan digunakan untuk menarik kesimpulan dan memberikan rekomendasi.

2.3 Data Set

Pada Tabel 1 Data yang kami ambil kemudian diekspor ke dalam format CSV. Data ulasan yang telah diekstraksi kemudian diolah menjadi dataset. Langkah-langkah yang dilakukan mencakup proses tokenisasi, yaitu memecah kalimat menjadi kata-kata individu; normalisasi, yakni mengubah kata-kata tidak baku menjadi bentuk standar; penghapusan stopword untuk menyaring kata-kata umum yang sering muncul namun memiliki nilai informasi rendah; serta stemming, yang bertujuan mengubah kata berimbuhan menjadi bentuk dasarnya. Proses pelabelan data dalam penelitian ini dilakukan dengan memanfaatkan Python. Python berperan dalam mengidentifikasi kata-kata dalam kalimat yang memiliki sifat positif, negatif, atau netral berdasarkan perhitungan nilai polaritasnya.

Tabel 1. Data Set

No	Name	Date	Review
1	Rijuww	9 Desember 2024	bagus bgt banyak fiturnya. mau tf bisa, belanja apalagi. belanja apapun ada, ky token pulsa, bayar tagihan, n byk lagi. kurangnya shopee ga terlalu perhatiin setiap penjualan toko ngejual apa ajaa. kan ada tuh yg jual file buku digital, memangnya itu legal?
2	Riffno Kurniadi	9 Desember 2024	Bagaimana cara nonaktifkan ""Putar video otomatis"" suatu produk di beranda jika memiliki video, dan video ulasan para pembeli. Saya tidak bisa menemukan pengaturan ini dalam setelan. Sebab dengan putar video otomatis sangat menguras kuota disaat kuota minim
3	Jernih Kurnia idawati Lahagu	8 Desember 2024	Udh bertahun tahun Saya sering belanja di aplikasi Shopee, barangnya memang bagus dan sesuai ekspetasi, namun akhir akhir ini, udh kedua kalinya pesanan saya lewat estimasi pengiriman, entah apa kendalanya yg membuat pesanan saya sering terlambat pengiriman, mudah mudahan aplikasi Shopee bisa mengambil ketegasan agar kurir yg ngantar paket bisa mengirim pesanan tepat waktu, selain itu akhir akhir ini, untuk metode pembayaran COD tidak bisa digunakan, entah knp masalahnya
4	Bierdella Alsava	8 Desember 2024	Udh bertahun tahun Saya sering belanja di aplikasi Shopee, barangnya memang bagus dan sesuai ekspetasi, namun akhir akhir ini, udh kedua kalinya pesanan saya lewat estimasi pengiriman, entah apa kendalanya yg membuat pesanan saya sering terlambat pengiriman, mudah mudahan aplikasi Shopee bisa mengambil ketegasan agar kurir yg ngantar paket bisa

© 2025 The Author(s). Published by *CV. Graha Mitra Edukasi*. Page 82
This work is licensed under a Creative Commons Attribution 4.0 International License.

Bulletin of Artificial Intelligence
Volume 4, No 2, October 2025 Page: 80-89
ISSN 2962-3944 (media online)
DOI: 10.62866/buai.v4i2.237
https://journal.grahamitra.id/index.php/buai

No	Name	Date	Review
			mengirim pesanan tepat waktu, selain itu akhir akhir ini, untuk metode pembayaran COD tidak bisa digunakan, entah knp masalahnya
5	Raka Fvnky	8 Desember 2024	Aplikasi shopee ini sudah bagus banget cuma min, ada kekurangannya yaitu di metode pembayaran nya, sekarang kan sudah era modern jadi pembayaran bisa secara online, mohon tambahkan metode pembayaran seperti dana, gopay dll. Sebenarnya ada shopepay jugaa sudah membantu cuma kalau blm ada yang punya KTP atau blm punya rekening susah min, contohnya pelajar mau pesen barang tapi pelajar itu blm punya KTP/ no rek. Jadi susah pembayaran nya tapi kalau ada dana atau gopay apk digital jadi mudah
….	……	…..	……
100.	Accsae	14 Juni 2024	Sekarang shopee banyak biaya admin dll jadinya ga nyaman buat transaksi…. Padahal pake shopee/paylater… tetep aja banyak biaya lain-lain ya

2.4 TF-IDF

Metode Term Frekuensi - Inverse Document Frekuensi (TF-IDF) digunakan untuk menentukan bobot kata-kata yang muncul dalam suatu dokumen[26]. Proses ini terdiri dari dua tahap utama: Term Frekuensi (TF) dan Inverse Document Frekuensi (IDF). Tahap TF mengukur seberapa sering sebuah kata muncul dalam dokumen tertentu, dengan kata yang muncul lebih sering diberi nilai yang lebih tinggi. Sementara itu, tahap IDF menghitung frekuensi kemunculan kata tersebut di seluruh dokumen, memberikan bobot lebih besar pada kata-kata yang lebih jarang digunakan dan unik[27][28]. Adapun persamaan yang digunakan yakni sebagai berikut:

$$TF(t) = \frac{N(d,t)}{T} \quad (1)$$

TF(t,d) merepresentasikan frekuensi kemunculan istilah t dalam dokumen d. Nilai ini dihitung berdasarkan jumlah kemunculan istilah t dalam dokumen d (N(t,d)) dibandingkan dengan total jumlah istilah dalam dokumen tersebut (T). Dengan demikian, setiap dokumen dan kata akan memiliki nilai TF(t,d) yang unik[29]

$$IDF(t) = \frac{\log N}{N\,(t)} \quad (2)$$

IDF(t) adalah inverse document frequency dari istilah t, di mana N merepresentasikan jumlah total dokumen, dan N(t) menunjukkan jumlah dokumen yang mengandung istilah t[29]. Rumus yang digunakan untuk menghitung Term Frequency - Inverse Document Frequency adalah sebagai berikut:.

$$TF\ -\ IDF = TF * IDF \quad (3)$$

Pembagian data dilakukan dengan membagi data menjadi rasio 80:20, 70:30, dan 60:40 untuk data pelatihan dan data pengujian. Tujuan dari pembagian ini adalah untuk mengevaluasi akurasi kinerja algoritma Logistic Regression dalam mengklasifikasikan sentimen masyarakat terhadap aplikasi Shopee.

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Data ulasan atau komentar aplikasi Shopee yang digunakan dalam penelitian ini berjumlah 100 data. Contoh dari data ulasan/komentar Shopee yang telah dikumpulkan dapat dilihat pada Tabel 1.

Tabel 2. Hasil Case Folding

No	Komentar/Ulasan
1.	Sumpah demi apapun gue benci banget kenapa jadi gak bisa milih ekspedisi lagi sih??? Tiba tiba semua pesanan pada pake spx/anter aja??? Benci banget sama spx pengiriman lama banget beda jauh sama jnt. Kalo emang lu gak mau biarin kita nentuin ekspedisi jangan kerja lama gitu. Barang gue udah 2 minggu belum sampe-sampe perkara pake spx. Dan pesanan baru juga gabisa diubah langsung auto pake spx. Dulu lu pada pernah ngubah sistemnya begini dan banyak yang ga terima, dan bagusnya lu pada balikin jadi system lama, terus kenapa dibalikin lagi system kalian yang paling JELEK ini??? STOP BIKIN SISTEM AUTOMATIS PEMILIHAN EKSPEDISI COKKK MERUGIKAN BANGET.

© 2025 The Author(s). Published by *CV. Graha Mitra Edukasi*. Page 83
This work is licensed under a Creative Commons Attribution 4.0 International License.

Bulletin of Artificial Intelligence
Volume 4, No 2, October 2025 Page: 80-89
ISSN 2962-3944 (media online)
DOI: 10.62866/buai.v4i2.237
https://journal.grahamitra.id/index.php/buai

No	Komentar/Ulasan
2.	Kenapa ya akhir-akhir ini saya Cuma bisa buka shopee 5 detik terus layar hitam dan tiba-tiba keluar dari shopee dan ngelag banget loh ini, bukan hp saya yang lag tapi shopee nya. Karna app yang lain tida begitu…. Tolong di benerin ya, Makasi.
3.	Sekrang shopee banyak biaya admin dll jadinya ga nyaman buat transaksi…. Padahal pake shopee/paylater… tetep aja banyak biaya lain-lain ya

3.2 Pre-processing Data

Tabel 3 adalah Tahap awal sebelum melakukan klasifikasi data adalah preprocessing. Proses preprocessing yang dilakukan mencakup case folding, tokenisasi, penghapusan stopword, dan stemming.

Tabel 3. Hasil Case Folding

No	Komentar/Ulasan	Case Folding
1	Sumpah demi apapun gue benci bgt kenapa jd gak bisa milih ekspedisi lagi sih??? Tiba2 semua pesanan pada pake spx/anteraja??? Benci bgt sm spx pengiriman lama bgt beda jauh sama jnt. Kalo emg lu gak mau biarin kita nentuin ekspedisi, ya benerin lah tuh ekpedisi jgn kerja lama gitu. brg gue udh 2 minggu blm sampe2 perkara pake spx. Dan pesanan baru jg gabisa diubah lgsg auto pke spx. dulu lu pada prnh ngubah sistemnya begini dan banyak yg gak terima, dan bagusnya lu pada balikin jd sistem lama.trus kenapa dibalikin lg sistem kalian yg plg JELEK ini??? STOP BIKIN SISTEM AUTOMATIS PEMILIHAN EKSPEDISI COKKKK MERUGIKAN BANGET	sumpah demi apapun gue benci bgt kenapa jd gak bisa milih ekspedisi lagi sih??? Tiba2 semua pesanan pada pake spx/anteraja??? Benci bgt sm spx pengiriman lama bgt beda jauh sama jnt. Kalo emg lu gak mau biarin kita nentuin ekspedisi, ya benerin lah tuh ekpedisi jgn kerja lama gitu. brg gue udh 2 minggu blm sampe2 perkara pake spx. Dan pesanan baru jg gabisa diubah lgsg auto pke spx. dulu lu pada prnh ngubah sistemnya begini dan banyak yg gak terima, dan bagusnya lu pada balikin jd sistem lama.trus kenapa dibalikin lg sistem kalian yg plg jelek ini??? Stop bikin sistem automatis pemilihan ekspedisi cokkkk merugikan banget
2	Kenapa ya akhir akhit ini saya cuma bisa buka shoope 5 detik terus layar hitam dan tiba tiba keluar dari shoope dan ngelg bgt loh ini, bukan hp saya yg ngelag tpi ini shoope ny. Karna app yg lain tidak seperti itu…Tolong di benerin ya.	kenapa ya akhir akhit ini saya cuma bisa buka shoope 5 detik terus layar hitam dan tiba tiba keluar dari shoope dan ngelg bgt loh ini, bukan hp saya yg ngelag tpi ini shoope ny. karna app yg lain tidak seperti itu...tolong di benerin ya.
3	Sekarang shoope banyak biaya admin dll jadinya ga nyaman buat transaksi…padahal pake shoopeepay/paylater…tetap aja banyak biaya lain2nya	sekarang shoope banyak biaya admin dll jadinya ga nyaman buat transaksi…padahal pake shoopeepay/paylater…tetap aja banyak biaya lain2nya

Tabel 4 menunjukkan hasil dari proses case folding, yang bertujuan untuk mengubah semua kata menjadi huruf kecil serta menghapus elemen non-teks seperti tanda baca, simbol, emoji, dan angka dalam kalimat di dokumen. Tahap berikutnya dalam preprocessing adalah tokenisasi. Di Tabel 4, terlihat bahwa proses ini memecah kalimat menjadi token agar dapat diproses pada tahap selanjutnya.

Tabel 4. Hasil Tokenisasi

No	Case Folding	Tokenisasi
1	sumpah demi apapun gue benci bgt kenapa jd gak bisa milih ekspedisi lagi sih??? Tiba2 semua pesanan pada pake spx/anteraja??? Benci bgt sm spx pengiriman lama bgt beda jauh sama jnt. Kalo emg lu gak mau biarin kita nentuin ekspedisi, ya benerin lah tuh ekpedisi jgn kerja lama gitu. brg gue udh 2 minggu blm sampe2 perkara pake spx. Dan pesanan baru jg gabisa diubah lgsg auto pke spx. dulu lu pada prnh ngubah sistemnya begini dan banyak yg gak terima, dan bagusnya lu pada balikin jd sistem lama.trus kenapa dibalikin lg sistem kalian yg plg jelek ini??? Stop bikin sistem automatis pemilihan ekspedisi cokkkk merugikan banget	['sumpah', 'demi', 'apapun', 'gue', 'benci', 'bgt', 'kenapa', 'jd', 'gak', 'bisa', 'milih', 'ekspedisi', 'lagi', 'sih', 'Tiba2', 'semua', 'pesanan', 'pada', 'pake', 'spxanteraja', 'Benci', 'bgt', 'sm', 'spx', 'pengiriman', 'lama', 'bgt', 'beda', 'jauh', 'sama', 'jnt', 'Kalo', 'emg', 'lu', 'gak', 'mau', 'biarin', 'kita', 'nentuin', 'ekspedisi', 'ya', 'benerin', 'lah', 'tuh', 'ekpedisi', 'jgn', 'kerja', 'lama', 'gitu', 'brg', 'gue', 'udh', '2', 'minggu', 'blm', 'sampe2', 'perkara', 'pake', 'spx', 'Dan', 'pesanan', 'baru', 'jg', 'gabisa', 'diubah', 'lgsg', 'auto', 'pke', 'spx', 'dulu', 'lu', 'pada', 'prnh', 'ngubah', 'sistemnya', 'begini', 'dan', 'banyak', 'yg', 'gak', 'terima', 'dan', 'bagusnya', 'lu', 'pada', 'balikin', 'jd', 'sistem', 'lamatrus', 'kenapa', 'dibalikin', 'lg', 'sistem', 'kalian', 'yg', 'plg', 'jelek', 'ini', 'Stop', 'bikin', 'sistem', 'automatis', 'pemilihan', 'ekspedisi',

© 2025 The Author(s). Published by *CV. Graha Mitra Edukasi*.
This work is licensed under a Creative Commons Attribution 4.0 International License.

Bulletin of Artificial Intelligence
Volume 4, No 2, October 2025 Page: 80-89
ISSN 2962-3944 (media online)
DOI: 10.62866/buai.v4i2.237
https://journal.grahamitra.id/index.php/buai

No	Case Folding	Tokenisasi
		'cokkkk', 'merugikan', 'banget']
2	kenapa ya akhir akhit ini saya cuma bisa buka shoope 5 detik terus layar hitam dan tiba tiba keluar dari shoope dan ngelg bgt loh ini, bukan hp saya yg ngelag tpi ini shoope ny. karna app yg lain tidak seperti itu...tolong di benerin ya.	['kenapa', 'ya', 'akhir', 'akhit', 'ini', 'saya', 'cuma', 'bisa', 'buka', 'shoope', '5', 'detik', 'terus', 'layar', 'hitam', 'dan', 'tiba', 'tiba', 'keluar', 'dari', 'shoope', 'dan', 'ngelg', 'bgt', 'loh', 'ini', 'bukan', 'hp', 'saya', 'yg', 'ngelag', 'tpi', 'ini', 'shoope', 'ny', 'karna', 'app', 'yg', 'lain', 'tidak', 'seperti', 'itutolong', 'di', 'benerin', 'ya']
3	sekarang shoope banyak biaya admin dll jadinya ga nyaman buat transaksi…padahal pake shoopeepay/paylater…tetap aja banyak biaya lain2nya	['sekarang', 'shoope', 'banyak', 'biaya', 'admin', 'dll', 'jadinya', 'ga', 'nyaman', 'buat', 'transaksi…padahal', 'pake', 'shoopeepaypaylater…tetap', 'aja', 'banyak', 'biaya', 'lain2nya']

Tabel 5 menunjukkan hasil dari proses normalisasi. Dalam penelitian ini, normalisasi dilakukan dengan mengubah kata-kata non-standar menjadi kata-kata standar yang sesuai dengan Kamus Besar Bahasa Indonesia (KBBI) dengan mencocokkan kata-kata tersebut dengan dataset yang sudah ada (kalimat informal - kalimat formal).

Tabel 5. Hasil Normalisasi

No	Tokenisasi	Normalisasi
1	['sumpah', 'demi', 'apapun', 'gue', 'benci', 'banget', 'kenapa', 'jadi', 'gak', 'bisa', 'milih', 'ekspedisi', 'lagi', 'sih', '?', '?', '?', 'tiba', 'tiba', 'semua', 'pesanan', 'pada', 'pake', 'spx/anter', 'aja', '?', '?', '?', 'benci', 'banget', 'sama', 'spx', 'pengiriman', 'lama', 'banget', 'beda', 'jauh', 'sama', 'jnt', '.', 'kalo', 'emang', 'lu', 'gak', 'mau', 'biarin', 'kita', 'nentuin', 'ekspedisi', 'jangan', 'kerja', 'lama', 'gitu', '.', 'barang', 'gue', 'udah', '2', 'minggu', 'belum', 'sampe-sampe', 'perkara', 'pake', 'spx', '.', 'dan', 'pesanan', 'baru', 'juga', 'gabisa', 'diubah', 'langsung', 'auto', 'pake', 'spx', '.', 'dulu', 'lu', 'pada', 'pernah', 'ngubah', 'sistemnya', 'begini', 'dan', 'banyak', 'yang', 'ga', 'terima', ',', 'dan', 'bagusnya', 'lu', 'pada', 'balikin', 'jadi', 'system', 'lama', ',', 'terus', 'kenapa', 'dibalikin', 'lagi', 'system', 'kalian', 'yang', 'paling', 'jelek', 'ini', '?', '?', '?', 'stop', 'bikin', 'sistem', 'automatis', 'pemilihan', 'ekspedisi', 'cokkk', 'merugikan', 'banget']	['sumpah', 'demi', 'apapun', 'saya', 'benci', 'sangat', 'kenapa', 'jadi', 'tidak', 'bisa', 'milih', 'ekspedisi', 'lagi', 'sih', '?', '?', '?', 'tiba', 'tiba', 'semua', 'pesanan', 'pada', 'pakai', 'spx/anter', 'saja', '?', '?', '?', 'benci', 'sangat', 'sama', 'spx', 'pengiriman', 'lama', 'sangat', 'beda', 'jauh', 'sama', 'jnt', '.', 'kalau', 'emang', 'Anda', 'tidak', 'mau', 'biarin', 'kita', 'nentuin', 'ekspedisi', 'jangan', 'kerja', 'lama', 'gitu', '.', 'barang', 'saya', 'udah', '2', 'minggu', 'belum', 'sampe-sampe', 'perkara', 'pakai', 'spx', '.', 'dan', 'pesanan', 'baru', 'juga', 'gabisa', 'diubah', 'langsung', 'auto', 'pakai', 'spx', '.', 'dulu', 'Anda', 'pada', 'pernah', 'ngubah', 'sistemnya', 'begini', 'dan', 'banyak', 'yang', 'tidak', 'terima', ',', 'dan', 'bagusnya', 'Anda', 'pada', 'balikin', 'jadi', 'system', 'lama', ',', 'terus', 'kenapa', 'dibalikin', 'lagi', 'system', 'kalian', 'yang', 'paling', 'jelek', 'ini', '?', '?', '?', 'stop', 'bikin', 'sistem', 'automatis', 'pemilihan', 'ekspedisi', 'sekali', 'merugikan', 'sangat']
2	['kenapa', 'ya', 'akhir-akhir', 'ini', 'saya', 'cuma', 'bisa', 'buka', 'shopee', '5', 'detik', 'terus', 'layar', 'hitam', 'dan', 'tiba-tiba', 'keluar', 'dari', 'shopee', 'dan', 'ngelag', 'banget', 'loh', 'ini', ',', 'bukan', 'hp', 'saya', 'yang', 'lag', 'tapi', 'shopee', 'nya', '.', 'karna', 'app', 'yang', 'lain', 'tida', 'begitu…', '.', 'tolong', 'di', 'benerin', 'ya', ',', 'makasi']	['kenapa', 'ya', 'akhir-akhir', 'ini', 'saya', 'hanya', 'bisa', 'buka', 'shopee', '5', 'detik', 'terus', 'layar', 'hitam', 'dan', 'tiba-tiba', 'keluar', 'dari', 'shopee', 'dan', 'lambat', 'sangat', 'ya', 'ini', ',', 'bukan', 'hp', 'saya', 'yang', 'lag', 'tapi', 'shopee', 'nya', '.', 'karena', 'app', 'yang', 'lain', 'tidak', 'begitu…', '.', 'tolong', 'di', 'benerin', 'ya', ',', 'terima kasih']
3	['sekarang', 'shopee', 'banyak', 'biaya', 'admin', 'dll', 'jadinya', 'ga', 'nyaman', 'buat', 'transaksi…', '.', 'padahal', 'pake', 'shopee/paylater…', 'tetep', 'aja', 'banyak', 'biaya', 'lain-lain', 'ya']	['sekarang', 'shopee', 'banyak', 'biaya', 'admin', 'dan lain-lain', 'menjadi', 'tidak', 'nyaman', 'buat', 'transaksi…', '.', 'padahal', 'pakai', 'shopee/paylater…', 'tetap', 'saja', 'banyak', 'biaya', 'lain-lain', 'ya']

Tabel 6 menunjukkan hasil dari proses penghapusan stopword. Pada tahap ini, tujuannya adalah untuk mengekstrak kata-kata yang penting dan menghilangkan kata-kata yang kurang signifikan serta tidak memiliki makna.

Tabel 6. Hasil Normalisasi

No	Normalisasi	Stopword
1	['sumpah', 'demi', 'apapun', 'saya', 'benci', 'sangat', 'kenapa', 'jadi', 'tidak', 'bisa', 'milih', 'ekspedisi', 'lagi', 'sih', '?', '?', '?', 'tiba', 'tiba', 'semua', 'pesanan', 'pada', 'pakai', 'spx/anter', 'saja', '?', '?', '?', 'benci', 'sangat', 'sama', 'spx', 'pengiriman', 'lama', 'sangat', 'beda', 'jauh', 'sama', 'jnt', '.', 'kalau', 'emang', 'Anda', 'tidak', 'mau', 'biarin', 'kita', 'nentuin', 'ekspedisi', 'jangan', 'kerja', 'lama', 'gitu',	['sumpah', 'apapun', 'benci', 'milih', 'ekspedisi', 'sih', '?', '?', '?', 'pesanan', 'pakai', 'spx/anter', '?', '?', '?', 'benci', 'spx', 'pengiriman', 'beda', 'jnt', '.', 'emang', 'biarin', 'nentuin', 'ekspedisi', 'kerja', 'gitu', '.', 'barang', 'udah', '2', 'minggu', 'sampe-

© 2025 The Author(s). Published by *CV. Graha Mitra Edukasi*. Page 85
This work is licensed under a Creative Commons Attribution 4.0 International License.

Bulletin of Artificial Intelligence
Volume 4, No 2, October 2025 Page: 80-89
ISSN 2962-3944 (media online)
DOI: 10.62866/buai.v4i2.237
https://journal.grahamitra.id/index.php/buai

No	Normalisasi	Stopword
	'.', 'barang', 'saya', 'udah', '2', 'minggu', 'belum', 'sampe-sampe', 'perkara', 'pakai', 'spx', '.', 'dan', 'pesanan', 'baru', 'juga', 'gabisa', 'diubah', 'langsung', 'auto', 'pakai', 'spx', '.', 'dulu', 'Anda', 'pada', 'pernah', 'ngubah', 'sistemnya', 'begini', 'dan', 'banyak', 'yang', 'tidak', 'terima', ',', 'dan', 'bagusnya', 'Anda', 'pada', 'balikin', 'jadi', 'system', 'lama', ',', 'terus', 'kenapa', 'dibalikin', 'lagi', 'system', 'kalian', 'yang', 'paling', 'jelek', 'ini', '?', '?', '?', 'stop', 'bikin', 'sistem', 'automatis', 'pemilihan', 'ekspedisi', 'sekali', 'merugikan', 'sangat']	sampe', 'perkara', 'pakai', 'spx', '.', 'pesanan', 'gabisa', 'diubah', 'langsung', 'auto', 'pakai', 'spx', '.', 'ngubah', 'sistemnya', 'terima', ',', 'bagusnya', 'balikin', 'system', ',', 'dibalikin', 'system', 'jelek', '?', '?', '?', 'stop', 'bikin', 'sistem', 'automatis', 'pemilihan', 'ekspedisi', 'merugikan']
2	['kenapa', 'ya', 'akhir-akhir', 'ini', 'saya', 'hanya', 'bisa', 'buka', 'shopee', '5', 'detik', 'terus', 'layar', 'hitam', 'dan', 'tiba-tiba', 'keluar', 'dari', 'shopee', 'dan', 'lambat', 'sangat', 'ya', 'ini', ',', 'bukan', 'hp', 'saya', 'yang', 'lag', 'tapi', 'shopee', 'nya', '.', 'karena', 'app', 'yang', 'lain', 'tidak', 'begitu…', '.', 'tolong', 'di', 'benerin', 'ya', ',', 'terima kasih']	['ya', 'akhir-akhir', 'buka', 'shopee', '5', 'detik', 'layar', 'hitam', 'shopee', 'lambat', 'ya', ',', 'hp', 'lag', 'shopee', 'nya', '.', 'app', 'begitu…', '.', 'tolong', 'benerin', 'ya', ',', 'terima kasih']
3	['sekarang', 'shopee', 'banyak', 'biaya', 'admin', 'dan lain-lain', 'menjadi', 'tidak', 'nyaman', 'buat', 'transaksi…', '.', 'padahal', 'pakai', 'shopee/paylater…', 'tetap', 'saja', 'banyak', 'biaya', 'lain-lain', 'ya']	['shopee', 'biaya', 'admin', 'dan lain-lain', 'nyaman', 'transaksi…', '.', 'pakai', 'shopee/paylater…', 'biaya', 'lain-lain', 'ya']

Selanjutnya ditampilkan hasil stemming pada Tabel 7 berikut ini.

Tabel 7. Hasil Stemming

No	Stopword	Stemming
1	['sumpah', 'apapun', 'benci', 'milih', 'ekspedisi', 'sih', '?', '?', '?', 'pesanan', 'pakai', 'spx/anter', '?', '?', '?', 'benci', 'spx', 'pengiriman', 'beda', 'jnt', '.', 'emang', 'biarin', 'nentuin', 'ekspedisi', 'kerja', 'gitu', '.', 'barang', 'udah', '2', 'minggu', 'sampe-sampe', 'perkara', 'pakai', 'spx', '.', 'pesanan', 'gabisa', 'diubah', 'langsung', 'auto', 'pakai', 'spx', '.', 'ngubah', 'sistemnya', 'terima', ',', 'bagusnya', 'balikin', 'system', ',', 'dibalikin', 'system', 'jelek', '?', '?', '?', 'stop', 'bikin', 'sistem', 'automatis', 'pemilihan', 'ekspedisi', 'merugikan']	['[', "'sumpah", "'", ',', "'apapun", "'", ',', "'benci", "'", ',', "'milih", "'", ',', "'ekspedisi", "'", ',', "'sih", "'", ',', "'", '?', "'", ',', "'", '?', "'", ',', "'", '?', "'", ',', "'", "'pesanan", "'", ',', "'pakai", "'", ',', "'spx/anter", "'", ',', "'", '?', "'", ',', "'", "'", ',', "'", '?', "'", ',', "'", '?', "'", ',', "'benci", "'", ',', "'spx", "'", ',', "'pengiriman", "'", ',', "'beda", "'", ',', "'jnt", "'", ',', "'", ',', '.', "'", ',', "'emang", "'", ',', "'biarin", "'", ',', "'nentuin", "'", ',', "'ekspedisi", "'", ',', "'kerja", "'", ',', "'gitu", "'", ',', "'", "'", ',', '.', "'", ',', "'barang", "'", ',', "'udah", "'", ',', "'", '2', "'", ',', "'minggu", "'", ',', "'sampe-samp", "'", ',', "'perkara", "'", ',', "'pakai", "'", ',', "'spx", "'", ',', "'", "'", ',', '.', "'", ',', "'pesanan", "'", ',', "'gabisa", "'", ',', "'diubah", "'", ',', "'langsung", "'", ',', "'auto", "'", ',', "'pakai", "'", ',', "'spx", "'", ',', "'", "'", ',', '.', "'", ',', "'ngubah", "'", ',', "'sistemnya", "'", ',', "'terima", "'", ',', "'", "'", ',', "'", "'", ',', "'bagusnya", "'", ',', "'balikin", "'", ',', "'system", "'", ',', "'", "'", ',', "'dibalikin", "'", ',', "'system", "'", ',', "'jelek", "'", ',', "'", "'", ',', "'", '?', "'", ',', "'", '?', "'", ',', "'", "'", ',', "'", '?', "'", ',', "'", "'stop", "'", ',', "'bikin", "'", ',', "'", "'sistem", "'", ',', "'automati", "'", ',', "'", "'pemilihan", "'", ',', "'", "'ekspedisi", "'", ',', "'", "'merugikan", "'", ',', ']']
2	['ya', 'akhir-akhir', 'buka', 'shopee', '5', 'detik', 'layar', 'hitam', 'shopee', 'lambat', 'ya', ',', 'hp', 'lag', 'shopee', 'nya', '.', 'app', 'begitu…', '.', 'tolong', 'benerin', 'ya', ',', 'terima kasih']	['[', "'ya", "'", ',', "'akhir-akhir", "'", ',', "'buka", "'", ',', "'shope", "'", ',', "'", '5', "'", ',', "'detik", "'", ',', "'layar", "'", ',', "'hitam", "'", ',', "'shope", "'", ',', "'lambat", "'", ',', "'ya", "'", ',', "'", "'", ',', "'", "'", ',', "'hp", "'", ',', "'lag", "'", ',', "'shope", "'", ',', "'nya", "'", ',', "'", "'", ',', '.', "'", ',', "'app", "'", ',', "'begitu…", "'", ',', "'", "'", ',', '.', "'", ',', "'tolong", "'", ',', "'benerin", "'", ',', "'ya", "'", ',', "'", "'", ',', "'", "'", ',', "'terima", 'kasih', "'", ']']
3	['shopee', 'biaya', 'admin', 'dan lain-lain', 'nyaman', 'transaksi…', '.', 'pakai', 'shopee/paylater…', 'biaya', 'lain-lain', 'ya']	['[', "'shope", "'", ',', "'biaya", "'", ',', "'admin", "'", ',', "'dan", 'lain-lain', "'", ',', "'nyaman", "'", ',', "'transaksi…", "'", ',', "'", "'", ',', '.', "'", ',', "'pakai", "'", ',', "'shopee/paylater…", "'", ',', "'biaya", "'", ',', "'lain-lain", "'", ',', "'ya", "'", ']']

Setelah menyelesaikan taahapan Tabel 7, langkah berikutnya adalah menentukan bobot setiap istilah berdasarkan frekuensi kemunculannya dalam dokumen menggunakan metode TF-IDF. Metode TF-IDF adalah alat statistik numerik yang dapat mengidentifikasi kata kunci dari istilah tertentu. Selain itu, TF-IDF dikenal karena efisiensinya, kesederhanaannya, dan keakuratannya. Perhitungan TF-IDF dilakukan menggunakan rumus yang telah dijelaskan dalam metode penelitian.

3.3 Pembagian Data (Data Split)

© 2025 The Author(s). Published by *CV. Graha Mitra Edukasi*. Page 86
This work is licensed under a Creative Commons Attribution 4.0 International License.

Bulletin of Artificial Intelligence
Volume 4, No 2, October 2025 Page: 80-89
ISSN 2962-3944 (media online)
DOI: 10.62866/buai.v4i2.237
https://journal.grahamitra.id/index.php/buai

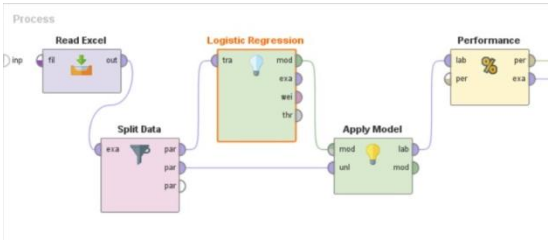
Proses klasifikasi teks komentar atau ulasan aplikasi Shopee menggunakan Logistic Regression dilakukan dengan membagi data menjadi dua bagian: data pelatihan dan data pengujian. Dataset yang digunakan terdiri dari 100 ulasan yang telah diproses dan diberi label sebelumnya. Data ini dibagi dengan rasio 80:20, 70:30, dan 60:40 untuk data pelatihan dan pengujian. Detail pembagian data dapat dilihat pada Tabel 8.

Tabel 8. Pembagian Data (Data Split)

Perbandingan	Training	Testing
80 : 20	80% 40 (positif) 24 (negatif) 1 (netral)	20% 10 (positif) 6 (negatif) 4 (netral)
70 : 30	70% 35 (positif) 2 (negatif) 14 (netral	30% 15 (positif) 9 (negatif) 6 (netral)
60 :40	60% 30 (positif) 18 (negatif) 12 (netral)	40% 20 (positif) 12 (negatif) 8 (netral

3.4 Modeling Regresi Logistik

Dalam upaya menyelesaikan, kami disini menggunakan aplikasi RapidMiner Studio. Dengan antarmuka yang intuitif, kita dapat membaca data dari file Excel, membaginya untuk pelatihan dan pengujian, membangun model regresi logistik, menerapkan model tersebut ke data uji, dan akhirnya mengevaluasi kinerja model untuk mendapatkan hasil yang akurat dan informatif. Untuk gambar pemrosesannya dapat dilihat pada Gambar 2.



Gambar 2. Proses analisis data pada aplikasi RapidMiner

Berdasarkan Gambar 2, dibagian tengah layar ada proses analisis data yang sedang dibuat. Pertama, data diambil dari Excel dan dimasukkan ke dalam program. Selanjutnya, data ini dibagi menjadi dua bagian : satu bagian untuk melatih model dan satunya lagi untuk pengujiannya. Kemudian, kita membangun model regresi logistic berdasarkan data pelatihan ini. Setelah model dibuat, kita menerapkannya pada data uji untuk memprediksi hasil.

3.4 Modeling Regresi Logistik

Proses klasifikasi menggunakan Logistic Regression dalam penelitian ini dilakukan melalui tiga percobaan. Percobaan pertama menggunakan 80% data latih dan 20% data uji. Percobaan kedua menggunakan 70% data latih dan 30% data uji, sementara percobaan ketiga menggunakan 60% data latih dan 40% data uji. Hasil dari masing-masing skema percobaan dapat dilihat pada Tabel 9.

Tabel 9. Hasil percobaan atau Pengujian

Perbandingan	Akurasi	Precision (positif)	Precision (negatif)	Precision (netral)	Recall (positif)	Recall (negatif)	Recall (netral)
80 : 20	85%	0.93	0.89	0.85	0.90	0.91	0.80
70 : 30	82%	0.857	0.889	0.714	0.800	0.889	0.833
60 : 40	80%	0.889	0.883	0.75	0.8	0.833	0.75

3.5 Pembahasan

Berdasarkan hasil pengujian pada Tabel 9, algoritma Regresi Logistik mampu mengklasifikasikan sentimen masyarakat terhadap aplikasi Shopee ke dalam tiga kategori, yaitu positif, negatif, dan netral. Pengujian dilakukan menggunakan tiga skenario pembagian data latih dan data uji, yaitu 80:20, 70:30, dan 60:40. Hasil menunjukkan bahwa skenario

© 2025 The Author(s). Published by *CV. Graha Mitra Edukasi*.
This work is licensed under a Creative Commons Attribution 4.0 International License.

Bulletin of Artificial Intelligence
Volume 4, No 2, October 2025 Page: 80-89
ISSN 2962-3944 (media online)
DOI: 10.62866/buai.v4i2.237
https://journal.grahamitra.id/index.php/buai

80:20 memperoleh akurasi tertinggi sebesar 85%, kemudian skenario 70:30 sebesar 82%, dan skenario 60:40 sebesar 80%. Hasil tersebut menunjukkan bahwa pada dataset yang digunakan, proporsi data latih sebesar 80% memberikan kinerja klasifikasi yang lebih baik dibandingkan dua skenario lainnya.

Pada skenario 80:20, nilai precision untuk sentimen positif, negatif, dan netral masing-masing sebesar 0,93; 0,89; dan 0,85. Sementara itu, nilai recall masing-masing sebesar 0,90; 0,91; dan 0,80. Nilai tersebut memperlihatkan bahwa model memiliki kemampuan yang cukup baik dalam mengenali ketiga kategori sentimen, meskipun kemampuan mengenali sentimen netral masih lebih rendah dibandingkan sentimen positif dan negatif. Kondisi ini menunjukkan bahwa ulasan netral berpotensi memiliki karakteristik kata yang lebih beragam atau beririsan dengan kategori sentimen lainnya.

Jika dibandingkan dengan penelitian Santoso dan Wibowo [4], yang menganalisis sentimen ulasan aplikasi BuzzBreak pada Google Play Store menggunakan metode Naïve Bayes Classifier, penelitian ini memiliki kesamaan pada sumber data, yaitu ulasan pengguna aplikasi dari Google Play Store. Namun, objek penelitian dan algoritma klasifikasi yang digunakan berbeda. Penelitian ini berfokus pada aplikasi Shopee dengan algoritma Regresi Logistik dan menguji tiga skenario pembagian data. Informasi hasil pada penelitian Santoso dan Wibowo yang dicantumkan dalam referensi penelitian ini tidak memuat nilai akurasi, sehingga perbandingan kuantitatif secara langsung belum dapat dilakukan. Meskipun demikian, kedua penelitian menunjukkan bahwa ulasan pada Google Play Store dapat dimanfaatkan sebagai sumber data untuk mengidentifikasi sentimen pengguna terhadap layanan aplikasi.
Perbandingan berikutnya dilakukan dengan penelitian Darwis, Siskawati, dan Abidin [15] tentang analisis sentimen data Twitter BMKG Nasional menggunakan algoritma Naïve Bayes. Penelitian tersebut memperoleh akurasi sebesar 69,97%, sedangkan penelitian ini menghasilkan akurasi tertinggi sebesar 85% pada skenario pembagian data 80:20. Secara numerik, akurasi penelitian ini lebih tinggi 15,03 poin persentase. Namun, hasil tersebut tidak dapat diartikan bahwa Regresi Logistik selalu lebih unggul daripada Naïve Bayes, karena kedua penelitian menggunakan objek, sumber data, karakteristik teks, dan kemungkinan distribusi kelas yang berbeda. Perbandingan ini hanya memberikan gambaran bahwa model Regresi Logistik pada data ulasan Shopee yang digunakan dalam penelitian ini menghasilkan tingkat ketepatan klasifikasi yang baik.

Secara keseluruhan, hasil penelitian memperlihatkan bahwa penggunaan pembobotan TF-IDF dan algoritma Regresi Logistik dapat digunakan untuk mengklasifikasikan sentimen pengguna aplikasi Shopee. Penurunan akurasi dari 85% menjadi 80% ketika proporsi data uji diperbesar menunjukkan bahwa hasil model dipengaruhi oleh komposisi pembagian data. Selain itu, jumlah data yang digunakan masih terbatas, yaitu 100 ulasan, sehingga hasil penelitian perlu ditafsirkan sesuai dengan cakupan dataset tersebut. Penelitian selanjutnya disarankan menggunakan jumlah ulasan yang lebih besar, memastikan pembagian data mempertahankan proporsi setiap kelas, serta melakukan pengujian berulang atau validasi silang agar evaluasi kinerja model lebih stabil dan representatif.

4. KESIMPULAN

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap aplikasi Shopee melalui algoritma regresi logistik. Data review yang diambil dari Google Play Store diproses melalui tahapan pra-pemrosesan yang meliputi tokenisasi, normalisasi, stopword removal, dan stemming. Proses ini memastikan bahwa data yang digunakan untuk analisis bersih dan relevan. Metode Term Frekuensi-Invers Dokumen Frekuensi (TF-IDF) digunakan untuk menentukan bobot setiap kata dalam dokumen, sehingga memungkinkan model untuk memahami pentingnya istilah tertentu dalam analisis sentimen.Proses klasifikasi dilakukan dengan menggunakan tiga skenario pembagian data latih dan data uji: 80:20, 70:30, dan 60:40. Hasilnya menunjukkan bahwa model regresi logistik mencapai akurasi tertinggi pada skenario 80:20 sebesar 85%, sedangkan pada skenario 70:30 dan 60:40 menghasilkan akurasi masing-masing sebesar 82% dan 80%. Precision dan recall untuk sentimen positif, negatif, dan netral menunjukkan kinerja yang baik di semua skenario. Penelitian ini memberikan wawasan penting mengenai pola sentimen masyarakat terhadap aplikasi e-commerce seperti Shopee, termasuk kepuasan pengguna dan tantangan layanan. Penelitian ini menunjukkan bahwa algoritma regresi logistik efektif untuk klasifikasi sentimen, terutama dalam menangani data dengan jumlah fitur yang banyak. Namun terdapat keterbatasan dalam menangani data dengan distribusi kategori yang tidak merata. Penelitian selanjutnya dapat menggunakan algoritma lain seperti Random Forest atau Deep Learning untuk meningkatkan akurasi model dan kemampuan menangani kompleksitas data. Selain itu, pengumpulan data yang lebih besar dapat meningkatkan generalisasi model ke berbagai kategori sentimen.

REFERENCES

[1] D. Ardiansyah, A. Saepudin, R. Aryanti, E. Fitriani, and Royadi, "Analisis Sentimen Review Pada Aplikasi Media Sosial Tiktok Menggunakan Algoritma K-Nn Dan Svm Berbasis Pso," *J. Inform. Kaputama*, vol. 7, no. 2, pp. 233–241, 2023, doi: 10.59697/jik.v7i2.148.
[2] K. A. Majid and L. Rahmawati, "Fenomena dan perilaku belanja online melalui TikTok Shop pada Generasi Z perspektif konsumsi Islam," *J. Ilm. Ekon. Islam*, vol. 10, no. 2, pp. 1796–1806, 2024, doi: 10.29040/jiei.v10i2.13383.
[3] T. Fatiah Rahmadanti, M. Jajuli, and I. Purnamasari, "Klasifikasi Pengguna Shopee Berdasarkan Promosi Menggunakan Naïve Bayes," *Gener. J.*, vol. 5, no. 2, pp. 81–90, 2021, doi: 10.29407/gj.v5i2.15998.
[4] D. P. Santoso and W. Wibowo, "Analisis Sentimen Ulasan Aplikasi Buzzbreak Menggunakan Metode Naïve Bayes Classifier

© 2025 The Author(s). Published by *CV. Graha Mitra Edukasi*.
This work is licensed under a Creative Commons Attribution 4.0 International License.

Volume 4, No 2, October 2025 Page: 80-89
ISSN 2962-3944 (media online)
DOI: 10.62866/buai.v4i2.237
https://journal.grahamitra.id/index.php/buai

pada Situs Google Play Store," *J. Sains dan Seni ITS*, vol. 11, no. 2, 2022, doi: 10.12962/j23373520.v11i2.72534.
[5] V. K. S. Que, A. Iriani, and H. D. Purnomo, "Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 9, no. 2, pp. 162–170, 2020, doi: 10.22146/jnteti.v9i2.102.
[6] H. C. Husada and A. S. Paramita, "Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *Teknika*, vol. 10, no. 1, pp. 18–26, 2021, doi: 10.34148/teknika.v10i1.311.
[7] Normah, B. Rifai, S. Vambudi, and R. Maulana, "Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis SMOTE," *J. Tek. Komput. AMIK BSI*, vol. 8, no. 2, pp. 174–180, 2022, doi: 10.31294/jtk.v4i2.
[8] E. Martantoh and N. Yanih, "Implementasi metode Naive Bayes untuk klasifikasi karakteristik kepribadiaan siswa di sekolah MTS Darussa'adah menggunakan PHP MySQL," *J. Teknol. Sist. Inf.*, vol. 3, no. 2, pp. 166–175, 2022, doi: 10.35957/jtsi.v3i2.2896.
[9] N. P. Nugraha, R. Azim, S. Z. Daffa, and P. S. Ningayu, "Perbandingan Akurasi Metode Naïve Bayes dan Metode KNN untuk Memprediksi Gagal Ginjal Kronis," *J. Rekayasa Elektro Sriwij.*, vol. 5, no. 1, pp. 1–10, 2023, doi: 10.36706/jres.v5i1.63.
[10] A. Pebdika, R. Herdiana, and D. Solihudin, "Klasifikasi Menggunakan Metode Naive Bayes Untuk Menentukan Calon Penerima Pip," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 452–458, 2023, doi: 10.36040/jati.v7i1.6303.
[11] J. A. P. Ginting, R. M. Sari, M. R. D. Siregar, and D. Kiswanto, "Analisis Support Vector Machine (Svm) Untuk Klasifikasi Jenis Kelamin Pada Ikan Cupang Dengan Bantuan Local Binary Pattern (Lbp)," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 6, pp. 12782–12786, 2024, doi: 10.36040/jati.v8i6.12028.
[12] S. Rabbani, D. Safitri, N. Rahmadhani, A. A. F. Sani, and M. K. Anam, "Perbandingan evaluasi kernel SVM untuk klasifikasi sentimen dalam analisis kenaikan harga BBM: Comparative evaluation of SVM kernels for sentiment classification in fuel price increase analysis," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 2, pp. 153–160, 2023, doi: 10.57152/malcom.v3i2.897.
[13] F. Aziz, P. Ishak, and S. Abasa, "Klasifikasi Depresi Menggunakan Support Vector Machine: Pendekatan Berbasis Data Text Mining," *J. Pharm. Appl. Comput. Sci.*, vol. 2, no. 2, pp. 33–38, 2024, doi: 10.59823/jopacs.v2i2.53.
[14] M. M. Maarif and N. Setiyawati, "Analisis Sentimen Review Aplikasi LinkedIn di Google Play Store Menggunakan Support Vector Machine," *Progresif J. Ilm. Komput.*, vol. 20, no. 1, p. 454, 2024, doi: 10.35889/progresif.v20i1.1614.
[15] D. Darwis, N. Siskawati, and Z. Abidin, "Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional," *J. TEKNO KOMPAK*, vol. 15, no. 1, pp. 131–145, 2021, doi: 10.33365/jtk.v15i1.744.
[16] S. P. Rahayu, R. R. Aripin, and I. S. Ahmad, "Pemodelan Regresi Multivariat pada Penentuan Faktor-Faktor yang Berpengaruh terhadap Kesejahteraan di Jawa Tengah," *J. Varian*, vol. 3, no. 2, pp. 125–138, 2020, doi: 10.30812/varian.v3i2.654.
[17] A. Tripena, Y. Lianawati, and A. A. Setyawan, "Analisis Faktor-Faktor yang Mempengaruhi Kemiskinan Rumah Tangga di Desa Kotayasa melalui Pendekatan Regresi Logistik Biner," *Electro Luceat*, vol. 9, no. 2, pp. 43–58, 2023, doi: 10.32531/jelekn.v9i2.708.
[18] S. S. Dewi and R. Resmawan, "Analisis Regresi Logistik Multinomial dengan Metode Bayes untuk Identifikasi Faktor-Faktor Terjadinya Diabetes Melitus," *J. Math. Theory Appl.*, vol. 5, no. 2, pp. 51–60, 2023, doi: 10.31605/jomta.v5i2.2520.
[19] N. Neneng, A. S. Puspaningrum, and A. A. Aldino, "Perbandingan Hasil Klasifikasi Jenis Daging Menggunakan Ekstraksi Ciri Tekstur Gray Level Co-occurrence Matrices (GLCM) Dan Local Binary Pattern (LBP)," *Smatika J.*, vol. 11, no. 01, pp. 48–52, 2021, doi: 10.32664/smatika.v11i01.572.
[20] A. Tantowi, D. Pasha, and A. T. Priandika, "Implementasi Sistem Informasi Pembayaran Berbasis Sms Gateway (Studi Kasus : SMK Negeri 1 Bandar Lampung)," *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 2, no. 2, pp. 188–195, 2021, doi: 10.33365/jatika.v2i2.921.
[21] M. R. Handoko and N. Neneng, "Sistem Pakar Diagnosa Penyakit Selama Kehamilan Menggunakan Metode Naive Bayes Berbasis Web," *J. Teknol. Dan Sist. Inf.*, vol. 2, no. 1, pp. 50–58, 2021.
[22] S. Pramono, I. Ahmad, and R. I. Borman, "Analisis Potensi Dan Strategi Penembaan Ekowisata Daerah Penyanga Taman Nasional Way Kambas," *J. Teknol. dan Sist. Inf.*, vol. 1, no. 1, pp. 57–67, 2020, doi: 10.33365/jtsi.v1i1.208.
[23] P. B. Ramadhanu and A. T. Priandika, "Rancang Bangun Web Service Api Aplikasi Sentralisasi Produk Umkm Pada Uptd Plut Kumkm Provinsi Lampung," *J. Teknol. dan Sist. Inf.*, vol. 2, no. 1, pp. 59–64, 2021, doi: 10.33365/jtsi.v2i1.609.
[24] A. Novantika and S. Sugiman, "Analisis Sentimen Ulasan Pengguna Aplikasi Video Conference Google Meet menggunakan Metode SVM dan Logistic Regression," in *PRISMA, Prosiding Seminar Nasional Matematika*, 2022, pp. 808–813.
[25] G. Honestya, M. Sajida, and A. Ramadhanu, "Klasifikasi Jenis Daun Herbal Menggunakan Metode Logistic Regression dan Decision Tree Classifier Berdasarkan Fitur (Warna dan Bentuk)," *J. Inf. Syst. Educ. Dev.*, vol. 2, no. 1, pp. 52–55, 2024, doi: 10.62386/jised.v2i1.59.
[26] A. Apriani, H. Zakiyudin, and K. Marzuki, "Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF System Penerimaan Mahasiswa Baru pada Kampus Swasta," *J. Bumigora Inf. Technol.*, vol. 3, no. 1, pp. 19–27, 2021, doi: 10.30812/bite.v3i1.1110.
[27] I. S. Wibowo, A. Witanti, and I. Susilawati, "Keyword Extraction Judul Berita Online Di Indonesia Menggunakan Metode TF-IDF," *JATISI*, vol. 11, no. 1, pp. 99–111, 2024, doi: 10.35957/jatisi.v11i1.6718.
[28] J. E. B. Sinulingga and H. C. K. Sitorus, "Analisis Sentimen Opini Masyarakat terhadap Film Horor Indonesia Menggunakan Metode SVM dan TF-IDF," *J. Manaj. Inform.*, vol. 14, no. 1, pp. 42–53, 2024, doi: 10.34010/jamika.v14i1.11946.
[29] F. Rizal, A. Wijaya, and F. Hasyim, "Analisis Sentimen Masyarakat Indonesia Terhadap Aplikasi TikTok Menggunakan Algoritma Logistic Regression," *AKIRATECH J. Comput. Electr. Eng.*, vol. 1, no. 2, pp. 57–65, 2024, doi: 10.63935/akiratech.v1i2.33.

© 2025 The Author(s). Published by *CV. Graha Mitra Edukasi*.
This work is licensed under a Creative Commons Attribution 4.0 International License.