

Analisis Sentimen Terhadap Ulasan pada Aplikasi Astro Menggunakan Algoritma Support Vector Machine

Masriah^{1*}, Wahyu Tisno Atmojo²

^{1,2}Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Pradita

*email: masriah@student.pradita.ac.id

DOI: <https://doi.org/10.31603/komtika.v9i1.13129>

Received: 03-02-2025, Revised: 17-03-2025, Accepted: 18-03-2025

ABSTRACT

The Astro application is a quick commerce application that focuses on providing online shopping services with groceries category products to household needs with delivery in minutes. The Astro application was released in 2021 and is still operating today, so it is important for Astro to know the response or sentiment from customers regarding the services that have been provided so that it can be an evaluation for Astro to maintain user satisfaction. To find out the number of positive and negative sentiments from users, a study was conducted on sentiment analysis of user reviews on the Astro application on the Google Playstore using the support vector machine (SVM) algorithm. The data used in this study is as many as 1.000 reviews from web scraping results, then the data goes through the stages of preprocessing, data labeling, data sharing, word weighting with TF-IDF, classification using support vector machine (SVM) and linear kernel to evaluation. The results of this study were obtained from 696 reviews with positive sentiments and 304 reviews with negative sentiments, with the evaluation using the confusion matrix obtained an accuracy value of 84%, precision for the positive class of 93%, precision for the negative class of 69%, recall for the positive class of 83%, recall for the negative class of 85%, F-1 score for the positive class of 88%, and the F-1 score for the class is negative by 76%. From the results of the analysis, a good accuracy value was obtained with a larger number of positive sentiments, so it can be concluded that the Astro Application provides a positive experience to users.

Keywords: Application, Astro, Sentiment Analysis, Support Vector Machine

ABSTRAK

Aplikasi Astro merupakan aplikasi *quick commerce* yang berfokus memberikan layanan belanja *online* dengan produk kategori *groceries* hingga kebutuhan rumah tangga dengan pengiriman dalam waktu hitungan menit. Aplikasi Astro dirilis pada tahun 2021 dan masih beroperasi sampai saat ini, maka dari itu penting bagi Astro untuk mengetahui tanggapan atau sentimen dari pelanggan terkait pelayanan yang telah diberikan agar dapat menjadi evaluasi bagi Astro supaya tetap menjaga kepuasan pengguna. Untuk mengetahui banyaknya sentimen bernada positif dan negatif dari pengguna, dilakukan penelitian mengenai analisis sentimen terhadap ulasan pengguna pada aplikasi Astro di Google Playstore menggunakan algoritma support vector machine (SVM). Data yang digunakan dalam penelitian ini yaitu sebanyak 1.000 ulasan dari hasil *web scraping*, kemudian data tersebut melalui tahap *preprocessing*, pelabelan data, pembagian data, pembobotan kata dengan TF-IDF, klasifikasi menggunakan support vector machine (SVM) dan kernel linear hingga evaluasi. Hasil dari penelitian ini yaitu diperoleh 696 ulasan bersentimen positif dan 304 ulasan bersentimen negatif, dengan evaluasi menggunakan confusion matrix diperoleh nilai akurasi sebesar 84%, presisi untuk kelas positif sebesar 93%, presisi untuk kelas negatif sebesar 69%, *recall* untuk kelas positif sebesar 83%, *recall* untuk kelas negatif sebesar 85%, *f-1 score* untuk kelas positif sebesar 88%, dan *f-1 score* untuk kelas negatif sebesar 76%. Dari hasil analisis diperoleh nilai akurasi yang baik dengan jumlah sentimen positif yang lebih besar sehingga dapat disimpulkan bahwa Aplikasi Astro memberikan pengalaman yang positif kepada pengguna.

Keywords: Aplikasi, Astro, Analisis Sentimen, Support Vector Machine

PENDAHULUAN

Seiring perkembangan teknologi yang signifikan, hampir semua aktivitas manusia saat ini tidak lepas dari peran teknologi. Salah satunya yaitu aktivitas berbelanja *online* di *e-commerce*. Menurut data dari statista yang diolah PDSI (Pusat Data dan Sistem Informasi) kementerian perdagangan, jumlah pengguna *e-commerce* di Indonesia terus mengalami kenaikan sejak tahun 2020, diperkirakan jumlah pengguna *e-commerce* akan terus mengalami peningkatan hingga tahun 2029 yaitu mencapai 99,1 juta pengguna [1]. Sementara itu, Bank Indonesia (BI) juga sebelumnya memperkirakan bahwa nilai transaksi *e-commerce* diprediksi tumbuh 2,8% menjadi Rp487 triliun pada tahun 2024 dan 3,3% menjadi Rp503 triliun pada tahun 2025 [2].

Di era yang kompetitif ini, muncul *quick commerce* menjadi konsep bisnis baru yang menawarkan kecepatan dalam pengiriman barang yang dipesan pelanggan secara *online* dalam hitungan menit atau maksimal dalam beberapa jam. *Quick commerce* berfokus kepada penjualan kebutuhan pokok sehari-hari seperti bahan makanan, dan keperluan rumah tangga [3]. Menurut statista, pasar *quick commerce* di Indonesia diprediksi akan mengalami peningkatan pendapatan yang substansial, mencapai US\$3,15 miliar pada tahun 2025. Hal ini diperkirakan akan menghasilkan tingkat pertumbuhan tahunan sebesar 6,35%, yang mengarah pada volume pasar yang diproyeksikan sebesar US\$4,03 miliar pada tahun 2029. Selain itu, jumlah pengguna di pasar *quick commerce* diperkirakan mencapai 26,3 juta pengguna pada tahun 2029 [4].

Saat ini Astro merupakan salah satu aplikasi *quick commerce* yang berfokus memberikan layanan belanja *online* yang menyediakan lebih dari 1.000 produk dengan kategori *groceries*, kebutuhan rumah tangga hingga obat-obatan dengan pengiriman dalam waktu hitungan menit [5]. Astro juga menawarkan keunggulan jaminan garansi tepat waktu dengan pengembalian biaya ongkos kirim 100% jika pesanan yang diantar melebihi estimasi waktu yang tertera di aplikasi. Saat ini, aplikasi Astro telah diunduh lebih dari 1 juta pengunduh di Google Playstore dengan rata-rata rating 4,9 dan memiliki 31 ribu ulasan pengguna [6].

Astro merupakan salah satu layanan *quick commerce* yang masih bertahan saat ini ditengah persaingan yang ketat diantara layanan *quick commerce* lainnya. Menurut survei yang dilakukan oleh Populix, sebanyak 87% responden aktif berbelanja melalui aplikasi *quick commerce*, terutama mereka yang berusia 26-45 tahun di wilayah Pulau Jawa. Baik responden pria maupun wanita cenderung memilih layanan *quick commerce* yang terintegrasi dalam ekosistem *super apps*, seperti GoMart (60%), Tokopedia Now! (47%), dan GrabMart Kilat (47%). Sementara itu, lima aplikasi *quick commerce* yang paling sering digunakan oleh responden antara lain Segari (16%), AlloFresh (13%), TaniHub (12%), Sayur Kilat (8%), dan Astro (6%) [7]. Berdasarkan hasil survei Populix diperoleh informasi sebanyak 6% responden yang memilih layanan Astro sebagai aplikasi *quick commerce* untuk berbelanja, maka penting bagi Astro untuk mengetahui tanggapan atau sentimen dari pelanggan terkait pelayanan yang telah diberikan agar menjadi bahan evaluasi sehingga tetap menjaga kepuasan pelanggan dan menjadi layanan *quick commerce* yang utama digunakan. Analisis sentimen merupakan kegiatan menganalisis teks digital yang berupa opini seseorang di media sosial atau ulasan di internet tentang suatu topik yang bernada positif, negatif maupun netral [8]. Analisis sentimen dapat membantu perusahaan untuk memahami pelanggan, meningkatkan reputasi merek, meningkatkan kualitas layanan pelanggan, mengetahui apa yang dibicarakan orang tentang

produk dan mereknya, hingga mendapatkan masukan untuk pengembangan produk dan mereknya.

Urgensi dari penelitian ini terletak pada pentingnya memahami pengalaman dan kepuasan pengguna terhadap aplikasi Astro sebagai salah satu layanan *quick commerce* di Indonesia. Beberapa alasan yang mendasari perlunya penelitian ini antara lain:

1. Peningkatan Kualitas Layanan

Analisis sentimen terhadap ulasan pengguna dapat mengidentifikasi aspek-aspek yang perlu diperbaiki dalam aplikasi Astro, seperti fitur, kecepatan layanan, atau kendala teknis lainnya.

2. Pengambilan Keputusan Berbasis Data

Dengan menerapkan algoritma support vector machine (SVM) dalam analisis sentimen, perusahaan dapat memperoleh wawasan yang akurat mengenai opini pelanggan, sehingga keputusan strategis dapat didasarkan pada data yang objektif.

3. Persaingan di Industri Quick Commerce

Quick commerce merupakan industri yang berkembang pesat dengan banyaknya kompetitor. Pemahaman terhadap sentimen pengguna dapat membantu Astro dalam meningkatkan daya saing.

4. Meningkatkan Loyalitas Pengguna

Dengan mengetahui sentimen pelanggan, Astro dapat merancang strategi yang lebih efektif untuk meningkatkan kepuasan dan loyalitas pengguna, sehingga dapat mempertahankan serta menarik lebih banyak pelanggan.

5. Identifikasi Tren dan Preferensi Konsumen

Penelitian ini juga membantu dalam memahami perubahan tren belanja dan preferensi konsumen, sehingga perusahaan dapat menyesuaikan layanan agar lebih relevan dengan kebutuhan pasar.

Penelitian ini akan berfokus pada analisis sentimen terhadap ulasan pengguna pada aplikasi Astro di Google Playstore menggunakan algoritma support vector machine (SVM). Algoritma SVM sebagai kelanjutan dari penelitian sebelumnya yang telah dilakukan oleh Prasetyo dkk. mengenai analisis sentimen terhadap penggunaan aplikasi Astro menggunakan algoritma naïve bayes dengan memperoleh hasil klasifikasi sebesar 74% ulasan dengan sentimen positif dan 26% ulasan dengan sentimen negatif dan akurasi sebesar 63% [9]. Selain itu, terdapat beberapa penelitian sebelumnya yang melakukan penelitian terkait analisis sentimen seperti yang dilakukan oleh Salsabilla dkk. yaitu penerapan support vector machine untuk analisis sentimen pada aplikasi X (Twitter) mengenai obat penyebab gagal ginjal akut pada anak dengan hasil sentimen positif sebanyak 194 data dan sentimen negatif sebanyak 934 data dari 1128 *tweets* dengan *accuracy* 91%, *precision* 91%, *recall* 51% dan *f1-score* 66% [10]. Penggunaan metode yang sama juga dilakukan pada penelitian Darwis dkk. dengan judul penerapan algoritma support vector machine (SVM) untuk analisis sentimen pada data Twitter Komisi Pemberantasan Korupsi (KPK) Republik Indonesia yang memperoleh hasil kelas positif sebanyak 8%, kelas negatif sebanyak 77%, dan kelas netral sebanyak 15%, serta *accuracy* sebesar 82%, *precision* sebesar 90%, serta *recall* sebesar 88% dan *f1-score* sebesar 89% [11].

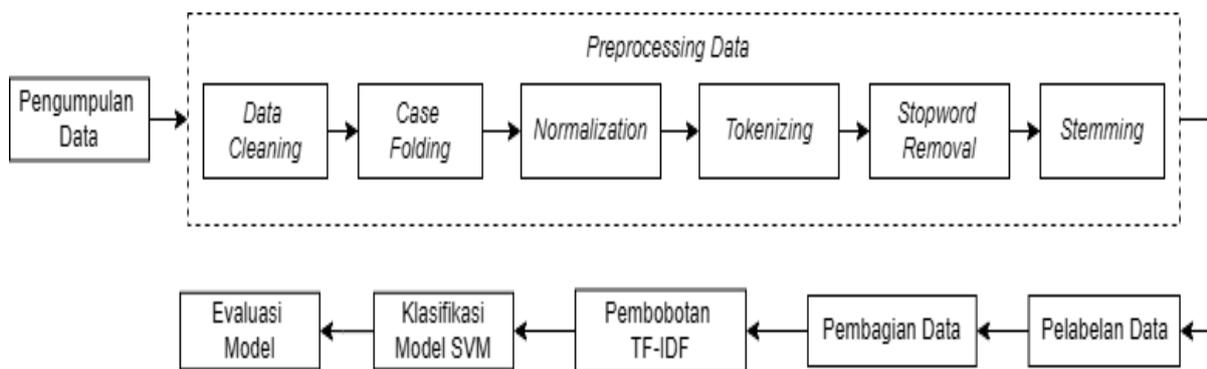
Penelitian berikutnya dilakukan oleh Kusuma dkk. terhadap ulasan pada aplikasi Indodax di Google Playstore menggunakan metode Support Vector Machine dengan memperoleh rata-rata *accuracy* sebesar 85%, *precision* sebesar 94%, *recall* sebesar 78% dan *f1-score* sebesar

85% [12]. Selain itu, penelitian dengan menggunakan dua metode yaitu support vector machine (SVM) dan naïve bayes telah dilakukan oleh ESKIYATURROFIKOH dkk. pada ulasan aplikasi X dengan memperoleh hasil akurasi sebelum menerapkan teknik SMOTE (Synthetic Minority Oversampling Technique) pada metode SVM adalah 75,5%, sedangkan pada algoritma naïve bayes adalah 75%. Setelah penerapan SMOTE, akurasi meningkat menjadi 81% untuk SVM dan tetap 75,5% untuk algoritma naïve bayes [13].

Penelitian ini bertujuan untuk memberikan analisis sentimen terhadap ulasan pengguna pada aplikasi Astro dengan menggunakan algoritma yang berbeda dari penelitian sebelumnya yaitu dengan menggunakan algoritma SVM dengan jumlah dataset 1.000 dengan rentang waktu ulasan dari tahun 2021-2025, serta rasio data latih dan data uji yaitu 80:20 yang diharapkan dapat memberikan informasi terkait penilaian dari pengguna sehingga dapat dijadikan evaluasi untuk pengembangan atau perbaikan layanan aplikasi.

METODE

Penelitian ini terdiri dari tujuh tahap, yaitu pengumpulan data, *preprocessing* data, pelabelan data, pembagian data, pembobotan kata menggunakan TF-IDF, klasifikasi model SVM, dan evaluasi model. Gambar 1 merupakan ilustrasi alur penelitian ini.



Gambar 1. Tahapan Penelitian

Pengumpulan Data

Proses pengumpulan data dilakukan dengan cara menarik data ulasan dari Google Playstore dengan menggunakan metode *scraping* di Google Colab dengan menggunakan bahasa pemrograman python. Data ulasan yang diambil sebanyak 1.000 data yang kemudian di ekspor dalam format csv.

Preprocessing Data

Pada tahap *preprocessing*, dataset yang telah dikumpulkan akan dibersihkan dengan tujuan untuk mencegah data yang kurang konsisten dan menyiapkan data agar lebih mudah dan akurat dalam proses analisis [14]. *Preprocessing* yang dilakukan melalui beberapa proses diantaranya yaitu *data cleaning*, *case folding*, *normalization*, *tokenizing*, *stopword removal*, dan *stemming*.

Pelabelan Data

Pada tahap pelabelan, dataset yang telah melalui tahap *preprocessing* akan diberi label berupa sentimen positif dan negatif. Pelabelan ini dilakukan secara otomatis menggunakan kamus indonesia *sentiment lexicon* [15]. Indonesia *sentiment lexicon* merupakan kamus leksikon dalam Bahasa Indonesia yang dirancang untuk mengidentifikasi sentimen atau opini dalam teks serta mengkategorikannya sebagai positif atau negatif. Kamus ini terdiri dari 3.609 kata positif dan 6.609 kata negatif, masing-masing memiliki skor yang berkisar antara -5 hingga +5 [16].

Pembagian Data

Setelah dilakukan pelabelan terhadap dataset, langkah selanjutnya yaitu membagi dataset dengan rasio 80:20, yaitu 80% sebagai data latih dan 20% sebagai data uji. Data latih akan digunakan untuk melatih model, sedangkan data uji akan digunakan untuk mengevaluasi performa model [17].

Pembobotan Kata Menggunakan TF-IDF (*Term Frequency-Inverse Document Frequency*)

Setelah dilakukan pelabelan, maka tahap selanjutnya yaitu pembobotan dengan metode TF-IDF dengan cara menghitung bobot setiap kata untuk memberikan skor yang mencerminkan tingkat keunikannya. Pembobotan kata menggunakan TF-IDF bertujuan untuk mengidentifikasi kata-kata yang sering muncul sekaligus menentukan seberapa penting kata tersebut dalam dokumen. Tahapan pembobotan ini meliputi perhitungan nilai TF sesuai Persamaan (1), perhitungan IDF berdasarkan Persamaan (2), dan perhitungan nilai akhir TF-IDF menggunakan Persamaan (3) [12].

$$TF_{(t,d)} = \frac{\text{jumlah kemunculan } t \text{ dalam } d}{\text{total } t \text{ dalam } d} \quad (1)$$

$$IDF_{(t)} = \log\left(\frac{\text{jumlah } d}{\text{jumlah } d \text{ yang memuat } t}\right) \quad (2)$$

$$TF-IDF = TF_{(t,d)} \times IDF_{(t)} \quad (3)$$

Keterangan:

t = Kata

d = Dokumen

$TF_{(t,d)}$ = Frekuensi jumlah t dalam d

$IDF_{(t)}$ = Inverse Document Frequency t

Klasifikasi Model Menggunakan Algoritma Support Vector Machine (SVM)

Pada tahap ini, dataset yang sudah diberi label akan dipisahkan menggunakan algoritma support vector machine (SVM). Algoritma SVM merupakan algoritma *machine learning* berbasis *supervised learning* yang digunakan untuk proses klasifikasi dan regresi [18]. SVM bekerja dengan cara menemukan *hyperplane* (garis pemisah) terbaik yang dapat memisahkan data ke dalam dua kelas atau lebih. *Hyperplane* terbaik adalah *hyperplane* yang memiliki margin maksimum, yaitu jarak maksimum antara titik data dari setiap kelas terhadap garis

pemisah. Secara umum, semakin besar margin *hyperplane*, semakin baik model dalam melakukan klasifikasi. *Hyperplane* terbaik dapat ditemukan dengan mengukur margin *hyperplane* dan menentukan titik maksimum margin tersebut [19]. Ketika data tidak dapat dipisahkan secara linear dalam ruang input SVM, maka SVM kesulitan untuk menemukan *hyperplane* terbaik untuk memisahkan kelas. Untuk mengatasi masalah ini, kernel *trick* digunakan untuk mengubah data input ke dalam dimensi yang lebih tinggi (ruang kernel) agar lebih mudah dipisahkan. Pada umumnya terdapat empat jenis fungsi kernel yang dapat digunakan, yaitu linear, polynomial, RBF (radial basis function), dan sigmoid [20].

Evaluasi Menggunakan Confusion Matrix

Evaluasi pada penelitian ini menggunakan confusion matrix. Confusion matrix merupakan matriks yang digunakan untuk mengevaluasi performa model klasifikasi. Confusion matrix berisi jumlah prediksi yang benar dan salah untuk setiap kelas dengan istilah seperti, TP (*true positive*) yaitu model memprediksi data sebagai kelas positif dan memang benar data tersebut berlabel positif, FP (*false positive*) yaitu model memprediksi data sebagai kelas positif dan kenyataannya data berlabel negatif, FN (*false negative*) yaitu model memprediksi data sebagai kelas negatif dan kenyataannya data berlabel positif, TN (*true negative*) yaitu model memprediksi data sebagai kelas negatif dan memang benar data tersebut berlabel negatif. Confusion matrix juga melakukan perhitungan berupa akurasi, *presisi*, *recall* dan *f1-score* dengan rumus sebagai berikut [21].

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100\% \quad (6)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\% \quad (7)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\% \quad (8)$$

$$\text{F1 - Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (9)$$

HASIL DAN PEMBAHASAN

Pengumpulan Data

Pengumpulan data dilakukan dengan cara *web scraping* menggunakan *library* Google Play Scraper di Google Colab untuk mengambil data ulasan pengguna aplikasi Astro dari Google Playstore. *Web scraping* adalah proses otomatis mengumpulkan atau mengekstraksi data dari halaman web untuk mengambil informasi yang diperlukan [22]. Data yang diambil berupa ulasan pengguna aplikasi Astro dari tahun 2021 hingga 2025 dengan jumlah sebanyak 1.000 dan disimpan dalam format csv. Berikut data ulasan yang diambil dari hasil *web scraping* ditunjukkan dalam Tabel 1.

Tabel 1. Hasil *Web Scraping*

No	Date	Username	Rating	Review Text
1	2024-12-30 23:23:01	Ega Yusfita	1	Aplikasinya jelek sekarang. Tiba tiba orderan di batalkan sistem padahal sudah bayar mana sudah isi saldo astro 900 rb buat belanja .eh ga bisa di tarik lg parah bngt. Kupon gratis ongkir pun tiba tiba ga bisa dipake, kupon diskon juga ga bisa dpke. JELEK!
2	2024-12-08 07:46:02	Langgeng Irma Salugiasih	5	Mantap lah ini tu astro. Cepet beneran less 30 menit, buat pengguna baru ada gratis ongkir pula walopun belanja 49k (i mean ngga harus 50k). Barangnya juga oke punya, papper bag tebal guys!! Harga ya so so, ada yg lebih murah ada yg sama aja/lebih mahal.

Preprocessing Data

Preprocessing data adalah tahap awal dalam pengolahan data dengan tujuan meningkatkan kualitas data, mengurangi *noise*, dan memastikan data sesuai dengan format yang dibutuhkan oleh algoritma yang digunakan. Berikut tahapan yang dilakukan dalam *preprocessing* data.

1. *Data Cleaning*: proses ini dilakukan untuk menghapus URL, emoji, karakter atau simbol dan angka yang tidak diperlukan. Berikut adalah tabel perbandingan data sebelum dan sesudah proses *cleaning* yang ditunjukkan pada Tabel 2.

Tabel 2. Hasil *Data Cleaning*

Sebelum <i>Cleaning</i>	Sesudah <i>Cleaning</i>
Aplikasinya jelek sekarang. Tiba tiba orderan di batalkan sistem padahal sudah bayar mana sudah isi saldo astro 900 rb buat belanja.eh ga bisa di tarik lg parah bngt. Kupon gratis ongkir pun tiba tiba ga bisa dipak, kupon diskon juga ga bisa dpke . JELEK!	Aplikasinya jelek sekarang Tiba tiba orderan di batalkan sistem padahal sudah bayar mana sudah isi saldo astro rb buat belanja eh ga bisa di tarik lg parah bngt Kupon gratis ongkir pun tiba tiba ga bisa dipake kupon diskon juga ga bisa dpke JELEK

2. *Case Folding*: proses ini dilakukan untuk mengubah semua huruf pada teks menjadi huruf kecil (*lowercase*), dengan tujuan adalah untuk memastikan konsistensi data dengan menghilangkan perbedaan antara huruf besar (*uppercase*) dan huruf kecil (*lowercase*). Berikut adalah tabel perbandingan data sebelum dan sesudah proses *case folding* yang ditunjukkan pada Tabel 3.

Tabel 3. Hasil *Case Folding*

Sebelum <i>Case Folding</i>	Sesudah <i>Case Folding</i>
Aplikasinya jelek sekarang Tiba tiba orderan di batalkan sistem padahal sudah bayar mana sudah isi saldo astro rb buat belanja eh ga bisa di tarik lg parah bngt Kupon gratis ongkir pun tiba tiba ga bisa dipake kupon diskon juga ga bisa dpke JELEK	aplikasinya jelek sekarang tiba tiba orderan di batalkan sistem padahal sudah bayar mana sudah isi saldo astro rb buat belanja eh ga bisa di tarik lg parah bngt kupon gratis ongkir pun tiba tiba ga bisa dipake kupon diskon juga ga bisa dpke jelek

3. *Normalization*: proses ini dilakukan untuk mengubah kata tidak baku menjadi baku menggunakan kamus kata baku yang di dapat dari *website* kaggle. Berikut adalah tabel perbandingan data sebelum dan sesudah normalisasi yang ditunjukkan pada Tabel 4.

Tabel 4. Hasil Normalisasi

Sebelum Normalisasi	Sesudah Normalisasi
aplikasinya jelek sekarang tiba tiba orderan di batalkan sistem padahal sudah bayar mana sudah isi saldo astro rb buat belanja eh ga bisa di tarik lg parah bngt kupon gratis ongkir pun tiba tiba ga bisa dipake kupon diskon juga ga bisa dpke jelek	aplikasinya jelek sekarang tiba tiba orderan di batalkan sistem padahal sudah bayar mana sudah isi saldo astro ribu buat belanja eh tidak bisa di tarik lagi parah banget kupon gratis ongkir pun tiba tiba tidak bisa dipakai kupon diskon juga tidak bisa dipakai jelek

4. **Tokenizing:** proses ini dilakukan untuk memisahkan kalimat menjadi potongan kata. Berikut adalah tabel perbandingan data sebelum dan sesudah tokenisasi yang ditunjukkan pada Tabel 5.

Tabel 5. Hasil Tokenisasi

Sebelum Tokenisasi	Sesudah Tokenisasi
aplikasinya jelek sekarang tiba tiba orderan di batalkan sistem padahal sudah bayar mana sudah isi saldo astro ribu buat belanja eh tidak bisa di tarik lagi parah banget kupon gratis ongkir pun tiba tiba tidak bisa dipakai kupon diskon juga tidak bisa dipakai jelek	['aplikasinya', 'jelek', 'sekarang', 'tiba', 'tiba', 'orderan', 'di', 'batalkan', 'sistem', 'padahal', 'sudah', 'bayar', 'mana', 'sudah', 'isi', 'saldo', 'astro', 'ribu', 'buat', 'belanja', 'eh', 'tidak', 'bisa', 'di', 'tarik', 'lagi', 'parah', 'banget', 'kupon', 'gratis', 'ongkir', 'pun', 'tiba', 'tiba', 'tidak', 'bisa', 'dipakai', 'kupon', 'diskon', 'juga', 'tidak', 'bisa', 'dipakai', 'jelek']

5. **Stopword Removal:** proses ini dilakukan untuk menghapus kata-kata yang dianggap tidak memiliki makna signifikan, sehingga tidak memberikan kontribusi penting dalam analisis. Kata-kata ini biasanya adalah kata-kata umum seperti "dan", "atau", "yang", "di", "untuk" dan sebagainya. Berikut adalah tabel perbandingan data sebelum dan sesudah proses *stopword removal* yang ditunjukkan pada Tabel 6.

Tabel 6. Hasil *Stopword Removal*

Sebelum <i>Stopword Removal</i>	Sesudah <i>Stopword Removal</i>
['aplikasinya', 'jelek', 'sekarang', 'tiba', 'tiba', 'orderan', 'di', 'batalkan', 'sistem', 'padahal', 'sudah', 'bayar', 'mana', 'sudah', 'isi', 'saldo', 'astro', 'ribu', 'buat', 'belanja', 'eh', 'tidak', 'bisa', 'di', 'tarik', 'lagi', 'parah', 'banget', 'kupon', 'gratis', 'ongkir', 'pun', 'tiba', 'tiba', 'tidak', 'bisa', 'dipakai', 'kupon', 'diskon', 'juga', 'tidak', 'bisa', 'dipakai', 'jelek']	['aplikasinya', 'jelek', 'orderan', 'batalkan', 'sistem', 'bayar', 'isi', 'saldo', 'astro', 'ribu', 'belanja', 'eh', 'tidak', 'tarik', 'parah', 'banget', 'kupon', 'gratis', 'ongkir', 'tidak', 'dipakai', 'kupon', 'diskon', 'tidak', 'dipakai', 'jelek']

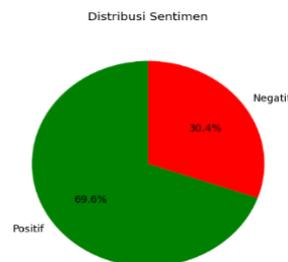
6. **Stemming:** proses ini dilakukan untuk mengubah kata-kata ke dalam bentuk kata dasar, sehingga menghilangkan imbuhan seperti me-, ke-, di-, dan lain-lain. Proses ini menggunakan *library* sastrawi, yang berfungsi untuk melakukan perubahan kata ke dalam bentuk dasar sesuai dengan kata baku yang terdapat dalam kamus Bahasa Indonesia. Berikut adalah tabel perbandingan data sebelum dan sesudah *stemming* yang ditunjukkan pada Tabel 7.

Tabel 7. Hasil *Stemming*

Sebelum <i>Stemming</i>	Sesudah <i>Stemming</i>
['aplikasinya', 'jelek', 'orderan', 'batalkan', 'sistem', 'bayar', 'isi', 'saldo', 'astro', 'ribu', 'belanja', 'eh', 'tidak', 'tarik', 'parah', 'banget', 'kupon', 'gratis', 'ongkir', 'tidak', 'dipakai', 'kupon', 'diskon', 'tidak', 'dipakai', 'jelek']	aplikasi jelek order batal sistem bayar isi saldo astro ribu belanja eh tidak tarik parah banget kupon gratis ongkir tidak pakai kupon diskon tidak pakai jelek

Pelabelan Data

Setelah data melalui tahap *preprocessing*, dataset yang berjumlah 1.000 akan diberi label menjadi kelas berlabel positif dan kelas berlabel negatif yang dilakukan secara otomatis menggunakan kamus indonesia *sentiment lexicon* [15]. Proses pelabelan diawali dengan memecah teks menjadi kata-kata, kemudian akan dihitung jumlah kata positif dan negatif serta mencocokkannya dengan kata yang terdapat pada kamus leksikon positif dan negatif yang setiap katanya memiliki skor dari -5 hingga +5. Jika diperoleh jumlah kata positif lebih besar daripada negatif, maka hasil sentimen yang diberikan adalah positif dan jika diperoleh jumlah kata negatif lebih besar, maka hasil sentimen yang diberikan adalah negatif. Setelah dilakukan pelabelan pada dataset maka diperoleh hasil dengan sentimen positif sebanyak 69,6% atau 696 data dan sentimen negatif sebanyak 30,4% atau 304 data. Berikut adalah ilustrasi perbandingan antara sentimen positif dan negatif yang ditunjukkan pada Gambar 2.



Gambar 2. Visualisasi Perbandingan Sentimen Positif dan Negatif

Pembagian Data

Dataset yang berjumlah 1.000 dan sudah diberi label akan dibagi menjadi data latih dan data uji dengan rasio perbandingan 80:20. Sehingga diperoleh data latih sebanyak 800 data dan data uji sebanyak 200 data.

Pembobotan TF-IDF

Setelah pembagian dataset, maka tahap selanjutnya yaitu pembobotan TF-IDF. Pembobotan TF-IDF dilakukan dengan memberikan bobot pada setiap kata dalam sebuah kalimat dengan cara mengalikan nilai TF dengan IDF.

Klasifikasi Model Menggunakan Algoritma Support Vector Machine (SVM)

Pada tahap klasifikasi model, algoritma yang digunakan adalah support vector machine (SVM) dengan menggunakan kernel linear dilengkapi dengan parameter $C=1$ dan *class weight* = *balanced*. Parameter $C=1$ dipilih setelah dilakukan pencarian untuk menemukan kombinasi parameter yang terbaik dengan menggunakan fungsi GridSearchCV. *Class weight* digunakan untuk menangani ketidakseimbangan kelas dengan memberikan bobot lebih pada kelas

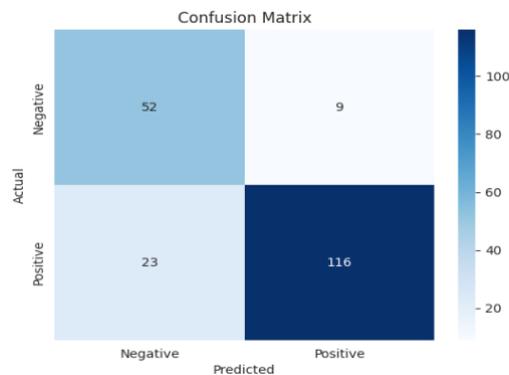
minoritas, dikarenakan pada penelitian ini kelas sentimen positif lebih banyak jumlahnya dibandingkan kelas negatif. Berikut adalah proses pendeklarasian untuk klasifikasi model menggunakan algoritma SVM ditunjukkan pada Gambar 3.

```
17 # Create and train the SVM model
18 svm_linear = SVC(kernel='linear', class_weight='balanced', C=1)
19 # Fit the model with the vectorized data
20 svm_linear.fit(X_train_vectorized, y_train)
```

Gambar 3. Pendeklarasian Klasifikasi Model SVM

Evaluasi

Setelah dilakukan klasifikasi model menggunakan algoritma SVM, maka dilakukan evaluasi terhadap kinerja model dengan menggunakan confusion matrix. Setelah algoritma SVM menghasilkan prediksi untuk data uji, hasil prediksi ini akan dibandingkan dengan label asli dari data uji (yang sudah diberikan oleh kamus sentimen) untuk mengukur akurasi dan performa model. Berikut adalah hasil confusion matrix yang ditunjukkan pada Gambar 4.



Gambar 4. Visualisasi Confusion Matrix

Pada gambar 4 diperoleh *true negative* (TN) sebanyak 52, *false negative* (FN) sebanyak 23, *false positive* (FP) sebanyak 9 dan *true positive* (TP) sebanyak 116. Setelah diperoleh nilai TN, FN, FP dan TP dilanjutkan dengan menghitung nilai akurasi, presisi, *recall* dan *f1-score* yang hasilnya ditunjukkan pada Gambar 5.

```
Best parameters: {'C': 1}
Accuracy with Best Linear Kernel: 0.84
Classification Report:
```

	precision	recall	f1-score	support
Negative	0.69	0.85	0.76	61
Positive	0.93	0.83	0.88	139
accuracy	0.81	0.84	0.84	200
macro avg	0.81	0.84	0.82	200
weighted avg	0.86	0.84	0.84	200

Gambar 5. Hasil Evaluasi

Pada gambar 5 tertera hasil evaluasi yang menghitung nilai akurasi, presisi, *recall* dan *f-1 score*. Berikut hasil perhitungan manual akurasi, presisi, *recall* dan *f-1 score*.

a. Perhitungan Akurasi:

$$\text{Akurasi} = \frac{116 + 52}{116 + 52 + 9 + 23} = \frac{168}{200} \times 100\% = 84\%$$

b. Perhitungan Presisi:

$$\text{Presisi kelas positif} = \frac{116}{116 + 9} = \frac{116}{125} \times 100\% = 92,8\%$$

$$\text{Presisi kelas negatif} = \frac{52}{52 + 23} = \frac{52}{75} \times 100\% = 69,3\%$$

c. Perhitungan *Recall*:

$$\text{Recall kelas positif} = \frac{116}{116 + 23} = \frac{116}{139} \times 100\% = 83,5 \%$$

$$\text{Recall kelas negatif} = \frac{52}{52 + 9} = \frac{52}{61} \times 100\% = 85,2\%$$

d. Perhitungan F1-Score:

$$F1 - \text{Score kelas positif} = 2 \times \frac{0,928 \times 0,835}{0,928 + 0,835} = 2 \times \frac{0,775}{1,763} = 0,878 \times 100 = 87,8\%$$

$$F1 - \text{Score kelas negatif} = 2 \times \frac{0,693 \times 0,852}{0,693 + 0,852} = 2 \times \frac{0,590}{1,545} = 0,764 \times 10 = 76,4\%$$

Setelah dilakukan perhitungan manual diperoleh hasil yang sama dengan hasil perhitungan menggunakan *library* sklearn.metrics yaitu diperoleh nilai akurasi sebesar 84% yang berarti model berhasil memprediksi dengan benar sebesar 84% dari total 200 data uji, presisi untuk kelas positif sebesar 93% menunjukkan bahwa model jarang melakukan kesalahan dalam memprediksi kelas positif, sedangkan presisi untuk kelas negatif sebesar 69% menunjukkan bahwa model masih melakukan beberapa kesalahan dalam mengklasifikasikan data ke dalam kelas negatif, lalu nilai recall untuk kelas positif sebesar 83% dan *recall* untuk kelas negatif sebesar 85% menunjukkan bahwa model lebih mampu mengidentifikasi dengan baik sebagian besar data yang memang benar termasuk ke dalam kelas negatif daripada kelas positif, kemudian nilai *f-1 score* untuk kelas positif sebesar 88%, dan *f-1 score* untuk kelas negatif sebesar 76% menunjukkan bahwa model lebih baik dalam mengenali kelas positif dibandingkan kelas negatif.

KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, yaitu analisis sentimen terhadap ulasan pengguna pada aplikasi Astro didapatkan beberapa kesimpulan yaitu hasil perbandingan sentimen yang didapat dari 1.000 ulasan di Google Playstore terdapat 696 ulasan bersentimen positif dan 304 ulasan bersentimen negatif. Dengan ulasan bersentimen positif lebih banyak jumlahnya, hal ini menyatakan bahwa pengguna cenderung memiliki pengalaman yang positif dalam menggunakan aplikasi Astro, sehingga aplikasi Astro dapat menjadi pilihan tempat berbelanja kebutuhan sehari-hari secara *online*. Dari pembagian data dengan rasio 80:20 dan diklasifikasi menggunakan algoritma SVM dengan kernel linear menghasilkan kinerja model yang cukup baik untuk menganalisis sentimen pada ulasan aplikasi Astro di Google Playstore dengan akurasi sebesar 84%, presisi 93%, *recall* 83% dan *f1-score* 88%.

DAFTAR PUSTAKA

- [1] P. K. Perdagangan, “Perdagangan Digital (E-Commerce) Indonesia Periode 2023,” 2024. [Online]. Available: <https://satudata.kemendag.go.id/ringkasan/produk/perdagangan-digital-e-commerce-indonesia-periode-2023>
- [2] K. P. RI, “Kemendag Ramal Transaksi E-Commerce di RI Tembus Rp533 Triliun - Kementerian Perdagangan Republik Indonesia,” 2024. <https://www.kemendag.go.id/berita/pojok-media/kemendag-ramal-transaksi-e-commerce-di-ri-tembus-rp533-triliun> (accessed Feb. 03, 2025).
- [3] A. E. Setiyono, C. Chandrawatisma, R. H. S. Fanandi, and P. Heriyati, “Antecedents of E-Loyalty as Research for the Quick Commerce Industry,” *Interdiscip. Soc. Stud.*, vol. 2, no. 8, pp. 2287–2299, 2023, doi: 10.55324/iss.v2i8.456.
- [4] “Quick Commerce - Indonesia | Statista Market Forecast,” *statista.com*, 2024. <https://www.statista.com/outlook/emo/online-food-delivery/grocery-delivery/quick-commerce/indonesia> (accessed Feb. 03, 2025).
- [5] “Tentang Astro,” <https://www.astronauts.id/>, 2025. <https://www.astronauts.id/sites/tentang-astro> (accessed Feb. 03, 2025).
- [6] “ASTRO - Groceries in Minutes - Aplikasi di Google Play,” <https://play.google.com/>, 2025. <https://play.google.com/store/apps/details?id=com.astro.shop&hl=id&pli=1> (accessed Feb. 03, 2025).
- [7] “Survei Populix: 87% Masyarakat RI Paling Aktif Belanja Lewat Quick Commerce | Infobanknews,” *infobanknews.com*, 2022. https://infobanknews.com/survei-populix-87-masyarakat-ri-paling-aktif-belanja-lewat-quick-commerce/#google_vignette (accessed Mar. 10, 2025).
- [8] M. R. Fahlevvi, “Analisis Sentimen Terhadap Ulasan Aplikasi Pejabat Pengelola Informasi Dan Dokumentasi Kementerian Dalam Negeri Republik Indonesia Di Google Playstore Menggunakan Metode Support Vector Machine,” *J. Teknol. dan Komun. Pemerintah.*, vol. 4, no. 1, pp. 1–13, 2022, doi: 10.33701/jtkp.v4i1.2701.
- [9] S. A. Prasetyo and W. T. Atmojo, “Analisis Sentimen Terhadap Penggunaan Aplikasi Astro Menggunakan Algoritma Naïve Bayes,” *J. Media Inform. Budidarma*, vol. 7, no. 4, p. 2149, 2023, doi: 10.30865/mib.v7i4.6750.
- [10] T. R. Salsabilla and N. Pratiwi, “Jurnal Teknik Informatika dan Komputer Penerapan Support Vector Machine Untuk Analisis Sentimen pada X (Twitter) Mengenai Obat Penyebab Gagal Ginjal Akut pada Anak,” pp. 67–74, 2024, doi: 10.22236/jutikom.v3i2.16892.
- [11] D. Darwis, E. S. Pratiwi, and A. F. O. Pasaribu, “Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia,” *Edutic - Sci. J. Informatics Educ.*, vol. 7, no. 1, pp. 1–11, 2020, doi: 10.21107/edutic.v7i1.8779.
- [12] A. Kusuma, E. Ermatita, and H. N. Irmanda, “Analisis Sentimen Pada Ulasan Aplikasi Indodax di Google Play Store Menggunakan Metode Support Vector Machine,” *Pros. Semin. Nas. Mhs. Bid. Ilmu Komput. dan Apl.*, vol. 3, no. 2, pp. 773–782, 2022.
- [13] “Eskiyaturrofikoh” and R. R. ’Suryono, “Analisis Sentimen Aplikasi X Pada Google Play Store Menggunakan Algoritma Naïve Bayes Dan Support Vector Machine (Svm),” *JUPI(Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 9, no. 3, pp. 1408–1419, 2024,

doi: 10.29100/jipi.v9i3.5392.

- [14] A. Agung, A. Daniswara, I. Kadek, and D. Nuryana, “Data Preprocessing Pola Pada Penilaian Mahasiswa Program Profesi Guru,” *J. Informatics Comput. Sci.*, vol. 05, pp. 97–100, 2023, doi: 10.26740/jinacs.v5n01.p97-100.
- [15] F. Koto and G. Y. Rahmaningtyas, “Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs,” *Proc. 2017 Int. Conf. Asian Lang. Process. IALP 2017*, vol. 2018-Janua, no. December, pp. 391–394, 2017, doi: 10.1109/IALP.2017.8300625.
- [16] V. Vinardo and I. Wasito, “Two-Stage Sentiment Analysis on Indonesian Online News Using Lexicon-Based,” *Sinkron*, vol. 8, no. 4, pp. 2109–2119, 2023, doi: 10.33395/sinkron.v8i4.12769.
- [17] W. A. Naseer, S. Sarwido, and B. B. Wahono, “Gradient Boosting Optimization with Pruning Technique for Prediction of Bmt Al-hikmah Permata Customer Data,” *Jinteks*, vol. 6, no. 3, pp. 719–727, 2024, doi: 10.51401/jinteks.v6i3.4702.
- [18] Tania Puspa Rahayu Sanjaya, Ahmad Fauzi, and Anis Fitri Nur Masruriyah, “Analisis sentimen ulasan pada e-commerce shopee menggunakan algoritma naive bayes dan support vector machine,” *INFOTECH J. Inform. Teknol.*, vol. 4, no. 1, pp. 16–26, 2023, doi: 10.37373/infotech.v4i1.422.
- [19] F. Romadoni, Y. Umaidah, and B. N. Sari, “Text Mining Untuk Analisis Sentimen Pelanggan Terhadap Layanan Uang Elektronik Menggunakan Algoritma Support Vector Machine,” *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 9, no. 2, pp. 247–253, 2020, doi: 10.32736/sisfokom.v9i2.903.
- [20] D. Diandra Audiansyah, D. Eka Ratnawati, and B. Trias Hanggara, “Analisis Sentimen Aplikasi MyXL menggunakan Metode Support VectorMachine berdasarkan Ulasan Pengguna di Google Play Store,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 8, pp. 3987–3994, 2022, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [21] Fitri Wulandari, Elin Haerani, Muhammad Fikry, and Elvia Budianita, “Analisis sentimen larangan penggunaan obat sirup menggunakan algoritma naive bayes classifier,” *J. CoSciTech (Computer Sci. Inf. Technol.*, vol. 4, no. 1, pp. 88–96, 2023, doi: 10.37859/coscitech.v4i1.4781.
- [22] R. Wahyudi and G. Kusumawardana, “Analisis Sentimen pada Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine,” *J. Inform.*, vol. 8, no. 2, pp. 200–207, 2021, doi: 10.31294/ji.v8i2.9681.



This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/)
