

Penerapan Metode *Support Vector Machine (SVM)* dalam Analisis Sentimen Ulasan Aplikasi Tokopedia di *Google Play Store*

Alya' Zhahrina, Ummu Sofiah, Dina Wahyu Apri Andayani, Nurhafizi Khairurrabbani, Fara Saputri

Universitas Bumigora, Mataram, Indonesia

Correspondence : e-mail: alyazhahrina@gmail.com

Abstrak

Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna aplikasi Tokopedia di Google Play Store menggunakan metode Support Vector Machine (SVM). Data dikumpulkan melalui teknik web scraping dan terdiri atas teks ulasan serta rating pengguna. Tahapan penelitian mencakup preprocessing teks (case folding, tokenisasi, normalisasi, stopword removal, dan stemming), pelabelan sentimen berdasarkan rating, serta klasifikasi dengan algoritma SVM. Dataset dibagi dengan rasio 80:20 untuk pelatihan dan pengujian model. Hasil evaluasi menunjukkan bahwa model menghasilkan akurasi sebesar 77,91% dengan nilai precision positif 78%, recall positif 83%, dan F1-score positif 81%. Analisis menunjukkan bahwa aspek layanan, pengiriman, dan kualitas barang menjadi isu utama dalam persepsi pengguna. Model mampu mengklasifikasikan sentimen dengan baik, khususnya pada kelas positif, namun cenderung mengalami kesalahan pada kelas negatif. Temuan ini menunjukkan bahwa algoritma SVM efektif diterapkan dalam analisis sentimen berbahasa Indonesia, serta dapat dimanfaatkan untuk mendukung pengambilan keputusan strategis oleh penyedia layanan aplikasi. Penelitian ini juga merekomendasikan pengembangan lebih lanjut dengan pendekatan deep learning untuk menangkap konteks semantik secara lebih mendalam.

Kata kunci: analisis sentimen, Tokopedia, Google Play Store, SVM, Klasifikasi.

Abstract

This study aims to analyze user sentiment toward the Tokopedia application on Google Play Store using the Support Vector Machine (SVM) method. Data were collected through web scraping and consisted of user review texts and ratings. The research process included text preprocessing (case folding, tokenization, normalization, stopword removal, and stemming), sentiment labeling based on rating values, and classification using SVM with a linear kernel. The dataset was split into 80% training data and 20% testing data. The evaluation results showed that the model achieved an accuracy of 77,91%, precision of 78%, recall of 83%, and an F1-score of 81%. The analysis revealed that "service", "delivery", and "product quality" were the most dominant themes influencing user perception. The model performed well in classifying sentiment, particularly for positive classes, although it showed some misclassification in negative reviews. These findings indicate that the SVM algorithm is effective for sentiment classification in Indonesian-language reviews and can support strategic decision-making by application service providers. This study recommends further development using deep learning approaches to better capture semantic and contextual nuances in user reviews.

Keywords: Sentiment analysis, Tokopedia, Google Play Store, SVM, Classification..

1. Pendahuluan

Dewasa ini tidak dapat dipungkiri bahwa perkembangan globalisasi selalu berjalan beriringan dengan pesatnya peningkatan serta perkembangan di berbagai bidang, terutama di bidang teknologi dan inovasi. Pesatnya perkembangan teknologi dan inovasi ini kemudian tidak hanya memiliki pengaruh signifikan pada perkembangan keilmuan saja, melainkan juga pada industri dan ekonomi. Di Indonesia sendiri, perkembangan teknologi telah mendorong tumbuhnya berbagai industri E-Commerce dengan

sangat pesat. Apabila merujuk pada laporan PPRO *Payments & E-Commerce Report in Asia 2018*, pertumbuhan *e-commerce* telah meningkat hingga mencapai angka 78% di hampir setiap tahunnya [1]. Peningkatan konstan dari angka ini tentunya membuka kesempatan yang sangat baik bagi masing-masing industri untuk memasarkan produknya dengan cara yang mudah, yakni dengan mengoptimasi penggunaan *platform-platform e-commerce* seperti Tokopedia.

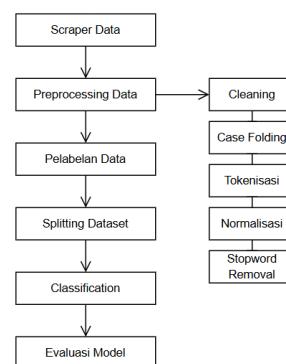
Dengan demikian, untuk dapat membuat suatu *platform* tetap kompetitif di pasar *e-commerce*, maka penting bagi *platform-platform e-commerce* untuk memperhatikan opini-opini konsumen terhadap layanan, jasa, atau bahkan produk yang mereka sediakan [2]. Hal ini tidak hanya penting dilakukan karena prosesnya dapat memperlhatkan animo pasar terkait layanan dari suatu aplikasi *e-commerce* saja, melainkan proses ini juga penting karena hasil opini-opini konsumen dapat dijadikan bahan pengambilan keputusan bagi penyedia layanan untuk mengembangkan atau memperbaiki berbagai kesalahan pada sistem yang terdapat pada aplikasi terkait [3]. Maka dari itu, salah satu elemen utama yang memengaruhi keputusan konsumen di *platform e-commerce* adalah ulasan-ulasan dari pengguna itu sendiri [4].

Dalam hal ini, sebagai salah satu aplikasi *e-commerce* terpopuler di Indonesia, Tokopedia sendiri telah menjadikan interaksi antarpengguna di dalam ulasan-ulasan pelanggan sebagai sumber informasi yang penting bagi pengembangan aplikasinya [5]. Guna melacak opini-opini konsumen terkait produk, jasa, atau layanan ini sendiri, tentu terdapat banyak sekali medium yang dapat digunakan untuk mengumpulkan opini-opini tersebut, misalnya dengan mengambil data teks dari ulasan-ulasan dan interaksi antarpelanggan di berbagai media sosial, atau bahkan mengambil data teks melalui survei penilaian pengguna atau hasil kepuasan pelanggan yang tersedia di laman pihak ketiga seperti pada laman penilaian dari *Google Play Store* [4], [6]. Meskipun demikian, banyaknya penilaian serta ulasan-ulasan pelanggan yang termuat secara acak di berbagai medium yang berbeda-beda pada akhirnya menimbulkan suatu masalah tersendiri bagi penyedia layanan dalam mengumpulkan informasi-informasi ulasan pengguna [6], [7]. Terlebih, ulasan-ulasan yang tersusun secara acak ini juga tentunya membuat penyedia layanan kesulitan untuk mengetahui penilaian-penilaian mana yang positif dengan penilaian mana yang bersifat negatif [6]. Maka dari itu, pada penelitian ini akan diterapkan metode SVM (*Support Vector Machine*) untuk mengklasifikasikan data-data teks dari ulasan pengguna aplikasi Tokopedia di *Google Play Store* hingga kemudian dapat diekstrak menjadi suatu analisis sentimen [3], [6].

Analisis sentimen pengguna aplikasi Tokopedia sendiri sebelumnya pernah dilakukan dalam penelitian [1], [5], [6], [8], [9]. Meskipun pada penelitian [1], [6] metode yang sama, yakni SVM telah digunakan, namun penelitian terdahulu pada analisisnya tidak berfokus pada medium ulasan *Google Play Store* dan menambahkan metode optimasi yang disebut dengan metode PSO (*Particle Swarm Optimization*). Sedangkan, pada penelitian [5], [8], [9] analisis sentimen dilakukan dengan menggunakan metode lain, seperti metode *K-nearest Neighborhood* dan metode *Naive Bayes*. Maka dari itu, tujuan dari disusunnya penelitian ini adalah untuk menerapkan metode SVM dalam analisis sentimen pengguna aplikasi Tokopedia melalui ulasan *Google Play Store*.

2. Metode Penelitian

Pada penyusunan analisis sentimen dalam penelitian, penelitian ini melalui beberapa tahapan utama yang umumnya digunakan dalam analisis sentimen dengan metode klasifikasi SVM (*Support Vector Machine*). Tahapan-tahapan ini meliputi tahapan pengumpulan data, pra-pemprosesan data, tokenisasi, normalisasi, pembobotan, dan pengklasifikasian data menggunakan algoritma SVM [10]. Gambar 1. Berikut merupakan visualisasi dari alur serta proses dari penelitian ini.



Gambar 1. Diagram Alur serta Proses Penelitian

a. Pengumpulan Data

Data pada penelitian ini diperoleh melalui teknik *web scraping* terhadap ulasan aplikasi Tokopedia di *Google Play Store* menggunakan *library Python* seperti *Google Play Scraper*. Data yang diambil pada penelitian ini mencakup teks ulasan, skor *rating*, dan tanggal ulasan. Pengambilan data difokuskan pada ulasan dalam bahasa Indonesia. Kemudian setelah data terkumpul, data akan disimpan dalam format .CSV untuk memudahkan proses analisis selanjutnya.

b. Pra-Pemrosesan Data

Langkah pra-pemrosesan merupakan tahap krusial dalam analisis teks karena menentukan kualitas representasi data yang akan digunakan dalam model pembelajaran mesin. Proses ini pun meliputi:

- 1) Pembersihan teks, yaitu proses menghilangkan tanda baca, angka, simbol-simbol, hingga karakter-karakter khusus lainnya yang terdapat pada teks ulasan [6].
- 2) *Case Folding*, yaitu proses yang dilakukan guna menyeragamkan huruf yang terdapat dalam dataset [10].
- 3) Tokenisasi, yaitu proses memecah kalimat menjadi unit kata [10].
- 4) Normalisasi, yaitu proses untuk mengubah kata-kata informal menjadi kata baku agar variasi kata menjadi konsisten [11]
- 5) *Stopword removal*, yaitu proses menghapus kata-kata umum yang tidak bermakna secara analitis seperti kata “di”, “yang”, “dan” [11].
- 6) *Stemming*, yaitu proses mengubah kata ke bentuk dasarnya. Contohnya, kata “berbelanja” akan diubah menjadi kata “belanja” [11].
- 7) *Labeling Sentimen*, yaitu proses yang dilakukan secara semi-otomatis berdasarkan skor *rating* dengan validasi manual pada subset data untuk meningkatkan akurasi *labeling* [1]. Misalnya, *rating* 4-5 pada *Google Play Store* akan diasumsikan sebagai sentimen positif, kemudian *rating* 1-2 akan diasumsikan sebagai negatif, sedangkan *rating* 3 akan dinilai sebagai netral.

c. Pembagian Data

Pada tahapan ini, dataset dibagi menjadi data latih dan data uji menggunakan fungsi *train test split* dengan parameter *test size*=0.2 dan *random state*=42. Ini berarti 80% data digunakan untuk pelatihan model dan 20% dataset digunakan untuk pengujian performa model [4].

d. Klasifikasi dan Evaluasi Sentimen dengan SVM

Model SVM diterapkan pada penelitian ini untuk mengoptimalkan proses klasifikasi menggunakan kernel linear[6], [11]. Model kemudian diuji untuk memprediksi kategori sentimen pada data uji yang telah dilakukan. Kemudian evaluasi performa model dilakukan dengan menghitung metrik-metrik evaluasi standar klasifikasi, yakni akurasi, *precision*, *recall*, *F1-Score* [2]. Selain itu, *confusion matrix* juga digunakan untuk mengidentifikasi distribusi prediksi benar dan salah dalam masing-masing kelas sentimen.

3. Hasil dan Pembahasan

3.1. Pengumpulan Data

Data yang dikumpulkan untuk digunakan dalam proses analisis sentimen pada penelitian ini terdapat sebanyak 2000 dataset. Proses pembuatan model analisis sentimen menggunakan SVM pada penelitian ini mencakup pembagian data latih dan data uji. Penelitian ini menggunakan rasio 80% : 20% untuk perbandingan data latih dan data uji. Data yang digunakan sebagai data latih yakni sejumlah 1340 dataset, dan 335 dataset untuk data uji. Gambar 2 berikut menunjukkan proses pengumpulan data dari penelitian ini.

| | Review ID | Username | Rating | Review Text | Date |
|---|---------------------------------------|------------------|--------|---|---------------------|
| 0 | 24afeff6b-4dd2-4838-a92c-35e73a025064 | ibim ibrahim | 1 | bertahun tahun pakai aplikasi ini, ko makin ke... | 2025-06-28 06:45:02 |
| 1 | 3b81f5dc-ec25-4d05-bb99-d7040bb6559 | Fatchullah Wahid | 1 | isinya banyak penipuan dan penyelesaian dari p... | 2025-06-28 06:41:14 |
| 2 | 89c496a4-28b5-42c6-abba-cd4dfba26c0b | Suratno | 5 | produk elektroniknya banyak yang murah | 2025-06-28 06:29:54 |
| 3 | d6df3dc7-b616-4465-8943-c0daa2abdf80 | Malik Mc | 1 | punya saldo disini berasa perampasan, mau nari... | 2025-06-28 06:18:23 |
| 4 | 948c8ab5-0aa6-4c97-a72e-e3b6d32440e0 | asus riska | 1 | KEPADA PIHAK TOKOPEDIA. KALAU PEMBELI KOMPLAIN... | 2025-06-28 06:00:02 |

Gambar 2. Hasil Pengumpulan Data

3.2. Pra-Pemprosesan

Tahapan awal dalam penelitian ini adalah proses pembersihan data teks. Data ulasan aplikasi Tokopedia dari *Google Play Store* dikumpulkan dan diproses untuk menghilangkan unsur-unsur yang tidak relevan dengan analisis sentimen, misalnya karakter, angka, tanda baca, dan spasi ganda. Kemudian dilakukan proses *case folding* untuk menghubungkan semua huruf menjadi huruf kecil. Setelah itu, teks diolah dengan tokenisasi, *stopword removal*, dan *stemming* menggunakan pustaka sastrawi.

3.2.1. Pembersihan Teks

Pembersihan teks pada penelitian ini dilakukan dengan membersihkan atau menghilangkan HTML, *username emoticon*, *hashtag (#)*, angka, dan lain sebagainya. Gambar 3 berikut menunjukkan hasil proses pembersihan teks terhadap data ulasan yang telah dikumpulkan penelitian ini.

| | Date | Username | Rating | Review Text | cleaning |
|----|---------------------|------------------|--------|---|---|
| 0 | 2025-06-28 06:45:02 | ibim ibrahim | 1 | bertahun tahun pakai aplikasi ini, ko makin ke... | bertahun tahun pakai aplikasi ini ko makin ke... |
| 1 | 2025-06-28 06:41:14 | Fatchullah Wahid | 1 | isinya banyak penipuan dan penyelesaian dari p... | isinya banyak penipuan dan penyelesaian dari p... |
| 2 | 2025-06-28 06:29:54 | Suratno | 5 | produk elektroniknya banyak yang murah | produk elektroniknya banyak yang murah |
| 3 | 2025-06-28 06:18:23 | Malik Mc | 1 | punya saldo disini berasa perampasan, mau nari... | punya saldo disini berasa perampasan mau nari... |
| 4 | 2025-06-28 06:00:02 | asus riska | 1 | KEPADA PIHAK TOKOPEDIA. KALAU PEMBELI KOMPLAIN... | KEPADA PIHAK TOKOPEDIA KALAU PEMBELI KOMPLAIN... |
| 5 | 2025-06-28 05:53:29 | Nizz Cann | 1 | maaf aja aku kasih rating 1 soalnya gak bisa p... | maaf aja aku kasih rating soalnya gak bisa pi... |
| 6 | 2025-06-28 05:43:32 | sukmo adi | 5 | aplikasinya bagus | aplikasinya bagus |
| 7 | 2025-06-28 05:41:45 | Ahmad Hasanudin | 5 | bagus | bagus |
| 8 | 2025-06-28 05:40:50 | Ekwie Tan | 1 | ndableg ! dari dulu diingatkan supaya diterib... | ndableg dari dulu diingatkan supaya diterib... |
| 10 | 2025-06-28 05:29:34 | Yuka Chan | 1 | aplikasi ribet banget dah padahal tinggal sat ... | aplikasi ribet banget dah padahal tinggal sat ... |

Gambar 3. Hasil Pembersihan Teks pada Dataset

3.2.2 Case Folding

Proses penyeragaman huruf pada penelitian ini dilakukan dengan mengubah setiap kata yang menggunakan huruf kapital menjadi huruf kecil. Berikut merupakan hasil dari proses *Case Folding* terlampir pada gambar 4.

| | Date | Username | Rating | Review Text | cleaning | case_folding |
|---|---------------------|------------------|--------|---|---|---|
| 0 | 2025-06-28 06:45:02 | ibim ibrahim | 1 | bertahun tahun pakai aplikasi ini, ko makin ke... | bertahun tahun pakai aplikasi ini ko makin ke... | bertahun tahun pakai aplikasi ini ko makin ke... |
| 1 | 2025-06-28 06:41:14 | Fatchullah Wahid | 1 | isinya banyak penipuan dan penyelesaian dari p... | isinya banyak penipuan dan penyelesaian dari p... | isinya banyak penipuan dan penyelesaian dari p... |
| 2 | 2025-06-28 06:29:54 | Suratno | 5 | produk elektroniknya banyak yang murah | produk elektroniknya banyak yang murah | produk elektroniknya banyak yang murah |
| 3 | 2025-06-28 06:18:23 | Malik Mc | 1 | punya saldo disini berasa perampasan, mau nari... | punya saldo disini berasa perampasan mau nari... | punya saldo disini berasa perampasan mau nari... |
| 4 | 2025-06-28 06:00:02 | asus riska | 1 | KEPADA PIHAK TOKOPEDIA. KALAU PEMBELI KOMPLAIN... | KEPADA PIHAK TOKOPEDIA KALAU PEMBELI KOMPLAIN... | kepada pihak tokopedia kalau pembeli komplain... |
| 5 | 2025-06-28 05:53:29 | Nizz Cann | 1 | maaf aja aku kasih rating 1 soalnya gak bisa p... | maaf aja aku kasih rating soalnya gak bisa pi... | maaf aja aku kasih rating soalnya gak bisa pi... |
| 6 | 2025-06-28 05:43:32 | sukmo adi | 5 | aplikasinya bagus | aplikasinya bagus | aplikasinya bagus |

Gambar 4. Hasil *Case Folding* pada Dataset

3.2.3 Tokenisasi

Setelah penyeragaman huruf dilakukan pada tahap *Case Folding*, tokenisasi dilakukan untuk memecah unit-unit kalimat menjadi unit-unit kata agar dapat mempermudah proses analisis selanjutnya. Contohnya, apabila input berisi “Aplikasinya bagus dan sangat mudah digunakan”, maka proses tokenisasi pada penelitian ini akan memecah kalimat tersebut menjadi unit-unit kata seperti “aplikasi”, “bagus”, “dan”, “mudah”. Gambar 5 berikut menunjukkan proses tokenisasi pada penelitian ini.

| | Date | Username | Rating | Review Text | cleaning | case_folding | normalisasi | tokenize |
|---|---------------------|------------------|--------|---|---|---|--|----------|
| 0 | 2025-06-28 06:45:02 | ibim ibrahim | 1 | bertahun tahun pakai aplikasi ini, ko makin ke... | bertahun tahun pakai aplikasi ini ko makin ke... | bertahun tahun pakai aplikasi ini kok makin ke... | [bertahun, tahun, pakai, aplikasi, ini, kok, m... | |
| 1 | 2025-06-28 06:41:14 | Fatchullah Wahid | 1 | isinya banyak penipuan dan penyelesaian dari p... | isinya banyak penipuan dan penyelesaian dari p... | isinya banyak penipuan dan penyelesaian dari p... | [isinya, banyak, penipuan, dan, penyelesaian, ...] | |
| 2 | 2025-06-28 06:29:54 | Suratno | 5 | produk elektroniknya banyak yang murah | produk elektroniknya banyak yang murah | produk elektroniknya banyak yang murah | [produk, elektroniknya, banyak, yang, murah] | |
| 3 | 2025-06-28 06:18:23 | Malik Mc | 1 | punya saldo disini berasa perampasan, mau nari... | punya saldo disini berasa perampasan mau nari... | punya saldo disini berasa perampasan mau nari... | [punya, saldo, disini, berasa, perampasan, mau... | |
| 4 | 2025-06-28 06:00:02 | asus riska | 1 | KEPADA PIHAK TOKOPEDIA. KALAU PEMBELI KOMPLAIN... | KEPADA PIHAK TOKOPEDIA KALAU PEMBELI KOMPLAIN... | kepada pihak tokopedia kalau pembeli komplain ... | [kepada, pihak, tokopedia, kalau, pembeli, kom... | |

Gambar 5. Hasil Tokenisasi

3.2.4 Normalisasi

Normalisasi teks pada dataset bertujuan untuk menyamakan variasi penulisan kata agar setiap tulisan pada dataset memiliki bentuk yang konsisten, sebab dalam kasus ulasan pengguna pada *Google Play Store*, sering kali penelitian ini menemukan adanya penggunaan bahasa tidak baku, singkatan, serta bentuk kata informal. Maka dari itu, proses normalisasi penting untuk dilakukan, karena ketidak tepatan penggunaan kata dan bahasa dapat memengaruhi akurasi model klasifikasi. Berikut merupakan hasil normalisasi yang dilakukan penelitian ini.

| | Date | Username | Rating | Review Text | cleaning | case_folding | normalisasi |
|---|---------------------|------------------|--------|---|---|---|---|
| 0 | 2025-06-28 06:45:02 | ibim ibrahim | 1 | bertahun tahun pakai aplikasi ini, ko makin ke... | bertahun tahun pakai aplikasi ini ko makin ke... | bertahun tahun pakai aplikasi ini ko makin ke... | bertahun tahun pakai aplikasi ini ko makin ke... |
| 1 | 2025-06-28 06:41:14 | Fatchullah Wahid | 1 | isinya banyak penipuan dan penyelesaian dari p... |
| 2 | 2025-06-28 06:29:54 | Suratno | 5 | produk elektroniknya banyak yang murah |
| 3 | 2025-06-28 06:18:23 | Malik Mc | 1 | punya saldo disini berasa perampasan, mau nari... | punya saldo disini berasa perampasan mau nari... | punya saldo disini berasa perampasan mau nari... | punya saldo disini berasa perampasan mau nari... |
| 4 | 2025-06-28 06:00:02 | asus riska | 1 | KEPADA PIHAK TOKOPEDIA. KALAU PEMBELI KOMPLAIN... | KEPADA PIHAK TOKOPEDIA KALAU PEMBELI KOMPLAIN... | kepada pihak tokopedia kalau pembeli komplain... | kepada pihak tokopedia kalau pembeli komplain ... |
| 5 | 2025-06-28 05:53:29 | Nizz Cann | 1 | maaf aja aku kasih rating 1 soalnya gak bisa p... | maaf aja aku kasih rating soalnya gak bisa pi... | maaf aja aku kasih rating soalnya gak bisa pi... | maaf saja aku kasih rating soalnya tidak bisa ... |
| 6 | 2025-06-28 05:43:32 | sukmo adi | 5 | aplikasinya bagus | aplikasinya bagus | aplikasinya bagus | aplikasinya bagus |
| 7 | 2025-06-28 05:41:45 | Ahmad Hasanudin | 5 | bagus | bagus | bagus | bagus |

Gambar 6. Hasil Normalisasi Teks pada Dataset

3.2.5 Stopword Removal

Setelah proses normalisasi dilakukan, maka akan ditemukan kumpulan kata-kata yang tidak penting dan tidak relevan dengan penelitian. Adapun hasil *Stopword Removal* pada penelitian ini dapat dilihat pada gambar 7 berikut.

| | Date | Username | Rating | Review Text | cleaning | case_folding | normalisasi | tokenize | stopword removal |
|---|---------------------|------------------|--------|---|---|---|--|---|------------------|
| 0 | 2025-06-28 06:45:02 | ibim ibrahim | 1 | bertahun tahun pakai aplikasi ini, ko makin ke... | bertahun tahun pakai aplikasi ini ko makin ke... | bertahun tahun pakai aplikasi ini ko makin ke... | [bertahun, tahun, pakai, aplikasi, ini, kok, m...] | [bertahun, pakai, aplikasi, kesini, update, nu... | |
| 1 | 2025-06-28 06:41:14 | Fatchullah Wahid | 1 | isinya banyak penipuan dan penyelesaian dari p... | isinya banyak penipuan dan penyelesaian dari p... | isinya banyak penipuan dan penyelesaian dari p... | [isinya, banyak, penipuan, dan, penyelesaian, ...] | [isinya, penipuan, penyelesaian, tokopedia, me... | |
| 2 | 2025-06-28 06:29:54 | Suratno | 5 | produk elektroniknya banyak yang murah | produk elektroniknya banyak yang murah | produk elektroniknya banyak yang murah | [produk, elektroniknya, banyak, yang, murah] | [produk, elektroniknya, murah] | |
| 3 | 2025-06-28 06:18:23 | Malik Mc | 1 | punya saldo disini berasa perampasan, mau nari... | punya saldo disini berasa perampasan mau nari... | punya saldo disini berasa perampasan mau nari... | [punya, saldo, disini, berasa, perampasan, mau...] | [saldo, berasa, perampasan, menarik, duit, rib... | |
| 4 | 2025-06-28 06:00:02 | asus riska | 1 | KEPADA PIHAK TOKOPEDIA. KALAU PEMBELI KOMPLAIN... | KEPADA PIHAK TOKOPEDIA KALAU PEMBELI KOMPLAIN... | kepada pihak tokopedia kalau pembeli komplain... | [kepada, pihak, tokopedia, kalau, pembeli, kom...] | [tokopedia, pembeli, komplain, tokopedia, tolo... | |

Gambar 7. Hasil *Stopword Removal* pada Dataset

3.2.6 Stemming Data

Dalam penelitian ini, proses *stemming* digunakan untuk menyederhanakan token hasil normalisasi agar setiap kata direduksi ke bentuk dasar Bahasa Indonesia. Proses ini dilakukan menggunakan pustaka sastrawi menggunakan *library* yang tersedia pada bahasa pemrograman *python*. Berikut merupakan gambar hasil *stemming* data.

| | Date | Username | Rating | Review Text | cleaning | case_folding | normalisasi | tokenize | stopword removal | stemming_data |
|---|---------------------|------------------|--------|---|---|---|--|---|---|---------------|
| 0 | 2025-06-28 06:45:02 | ibim ibrahim | 1 | bertahun tahun pakai aplikasi ini, ko makin ke... | bertahun tahun pakai aplikasi ini ko makin ke... | bertahun tahun pakai aplikasi ini ko makin ke... | [bertahun, tahun, pakai, aplikasi, ini, kok, m...] | [bertahun, pakai, aplikasi, kesini, update, nu... | tahun pakai aplikasi kesini update nua hancur ... | |
| 1 | 2025-06-28 06:41:14 | Fatchullah Wahid | 1 | isinya banyak penipuan dan penyelesaian dari p... | isinya banyak penipuan dan penyelesaian dari p... | isinya banyak penipuan dan penyelesaian dari p... | [isinya, banyak, penipuan, dan, penyelesaian, ...] | [isinya, penipuan, penyelesaian, tokopedia, me... | isi tipu selesai tokopedia bantu fitur tokoped... | |
| 2 | 2025-06-28 06:29:54 | Suratno | 5 | produk elektroniknya banyak yang murah | produk elektroniknya banyak yang murah | produk elektroniknya banyak yang murah | [produk, elektroniknya, banyak, yang, murah] | [produk, elektroniknya, murah] | produk elektronik murah | |
| 3 | 2025-06-28 06:18:23 | Malik Mc | 1 | punya saldo disini berasa perampasan, mau nari... | punya saldo disini berasa perampasan mau nari... | punya saldo disini berasa perampasan mau nari... | [punya, saldo, disini, berasa, perampasan, mau...] | [saldo, berasa, perampasan, menarik, duit, rib... | saldo asa ampas tank duit ribet sesuai mood a... | |
| 4 | 2025-06-28 06:00:02 | asus riska | 1 | KEPADA PIHAK TOKOPEDIA. KALAU PEMBELI KOMPLAIN... | KEPADA PIHAK TOKOPEDIA KALAU PEMBELI KOMPLAIN... | kepada pihak tokopedia kalau pembeli komplain... | [kepada, pihak, tokopedia, kalau, pembeli, kom...] | [tokopedia, pembeli, komplain, tokopedia, tolo... | tokopedia beli komplain tokopedia tolong denga... | |

Gambar 8. Hasil Proses *Stemming* Data

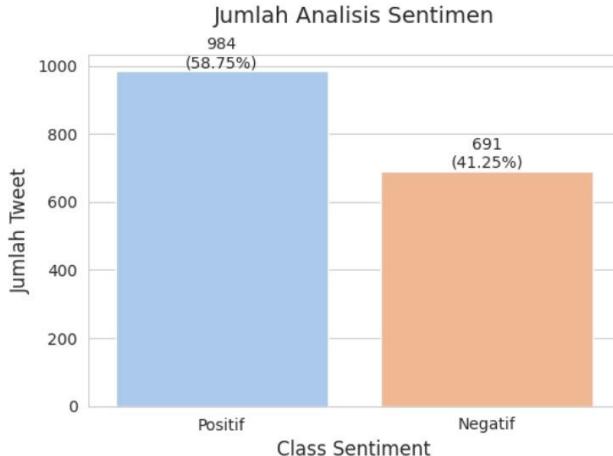
3.2.7 Labeling Sentimen

Pada proses *labeling* data pada penelitian ini, kategori label sentimen tidak dibuat menjadi tiga kategori, misalnya kategori sentimen negatif, positif, dan netral. Namun pada penelitian ini, kategori label sentimen hanya dipecah menjadi dua kategori umum saja, yakni kategori sentimen negatif dan kategori sentimen dengan label positif. Pada penelitian ini, proses *labeling* sentimen dilakukan dengan merujuk pada skor *rating* yang terdapat pada ulasan pengguna aplikasi dan juga merujuk pada hasil *stemming* data sebelumnya. Maka dari itu, Gambar 9. berikut menunjukkan hasil *labeling* sentimen penelitian ini.

| | Date | Username | Rating | Review Text | cleaning | case_folding | normalisasi | tokenize | stopword removal | stemming_data | Sentiment |
|---|---------------------|------------------|--------|---|---|---|---|---|---|---------------|-----------|
| 0 | 2025-06-28 06:45:02 | ibim ibrahim | 1 | berthun tahun pakai aplikasi ini, ko makin ke... | berthun tahun pakai aplikasi ini ko makin ke... | berthun tahun pakai aplikasi ini ko makin ke... | berthun tahun pakai aplikasi ini kok makin ke... | ['berthun', 'tahun', 'pakai', 'aplikasi', 'kesin', 'in...'] | tahun pakai aplikasi kesini update nua hancur ... | | Positif |
| 1 | 2025-06-28 06:41:14 | Fatchullah Wahid | 1 | isinya banyak penipuan dan penyelesaian dari p... | ['isinya', 'banyak', 'penipuan', 'penyelesaian', 'tokope...'] | isi tipu selesai tokopedia bantu fitur tokoped... | | Positif |
| 2 | 2025-06-28 06:29:54 | Suratno | 5 | produk elektroniknya banyak yang murah | ['produk', 'elektroniknya', 'banyak', 'yang', 'murah'] | produk elektronik murah | | Positif |
| 3 | 2025-06-28 06:18:23 | Malik Mc | 1 | punya saldo disini berasa perampasan, mau nar... | punya saldo disini berasa perampasan mau nar... | punya saldo disini berasa perampasan mau nar... | punya saldo disini berasa perampasan mau nar... | ['punya', 'saldo', 'disini', 'berasa', 'perampas...', 'menar...'] | saldo asa ampas tarik duit ribet sesuai mood a... | | Positif |
| 4 | 2025-06-28 06:00:02 | asus riska | 1 | KEPADA PIHAK TOKOPEDIA. KALAU PEMBELI KOMPLAIN... | KEPADA PIHAK TOKOPEDIA KALAU PEMBELI KOMPLAIN... | kepada pihak tokopedia kalau pembeli komplain... | kepada pihak tokopedia kalau pembeli komplain ... | ['kepada', 'pihak', 'tokopedia', 'kalau', 'pembeli', 'komplain', '...'] | tokopedia boli komplain tokopedia tolong denga... | | Positif |

Gambar 9. Hasil Labeling Sentimen

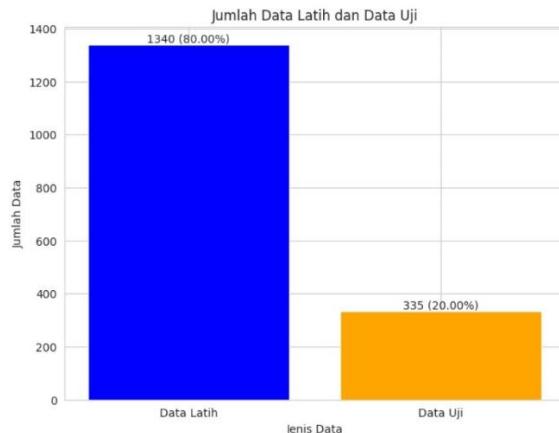
Setelah seluruh data ulasan telah diberi label, maka penelitian ini telah memperoleh jumlah hasil dari data-data ulasan tersebut berdasarkan labelnya. Gambar 10 menunjukkan hasil pelabelan sentimen di masing-masing data.



Gambar 10. Jumlah Analisis Sentimen

3.3 Pembagian Data

Setelah proses *labeling* data dilakukan, tahapan selanjutnya sebelum dilakukan proses klasifikasi menggunakan metode SVM adalah membagi data yang telah diberi label menjadi dua, yakni menjadi data latih dan data uji. Pada penelitian ini digunakan perbandingan sebesar 80% : 20%. Berikut merupakan hasil pembagian data.

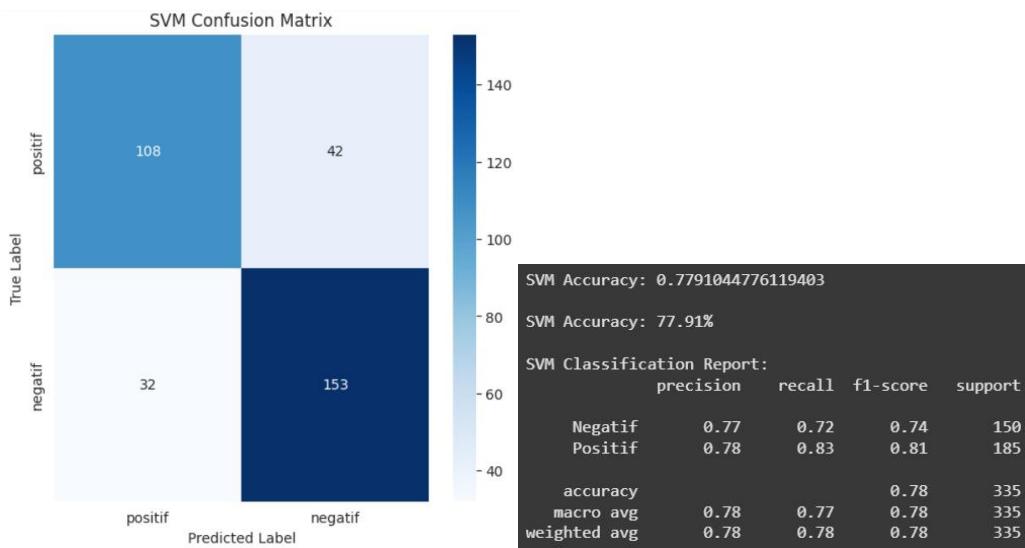


Gambar 11. Hasil Pembagian Data menjadi Data Latih dan Data Uji

3.4 Klasifikasi dan Evaluasi Sentimen dengan Metode SVM

Setelah data dibagi, kemudian metode SVM diterapkan dalam penelitian ini untuk mendeteksi ciri-ciri data pada masing-masing kelas yang dibentuk dalam data latih. Hasil dari pelatihan data menggunakan

metode SVM kemudian dievaluasi menggunakan data uji dengan mengukur hasil *confusion matrix* guna mencari nilai akurasi, presisi, dan lain sebagainya. Gambar 12 menunjukkan hasil dari proses klasifikasi dan evaluasi menggunakan metode SVM.



Gambar 12. Hasil Klasifikasi dan Evaluasi dengan Metode SVM

4. Kesimpulan

Penelitian ini memiliki tujuan untuk menerapkan metode SVM dalam mengklasifikasikan sentimen ulasan pengguna terhadap aplikasi Tokopedia di *Google Play Store*. Hal ini dilakukan guna dapat dijadikan acuan dalam pengambilan keputusan dan kebijakan dari industri terkait. Melalui tahapan komprehensif yang mencakup pengumpulan data, pra-pemrosesan, pelabelan sentimen, serta pembobotan menggunakan metode SVM, hasil analisis penelitian ini menunjukkan bahwa metode ini cukup efektif dalam mengelola data teks ulasan pengguna aplikasi tersebut. Dari total data ulasan yang dikumpulkan, proses analisis berdasarkan hasil pelabelan sentimen, hasil analisis mampu menghasilkan distribusi data yang cukup representatif untuk masing-masing kategori sentimen. Bahkan, klasifikasi menggunakan metode ini dapat menghasilkan nilai akurasi sebesar 77,91% dengan nilai presisi negatif 77% dan nilai presisi positif sebesar 78%, nilai *recall* negatif sebesar 72% serta nilai *recall* positif sebesar 83%, kemudian F1-score positif sebesar 81% dan F1-score negatif sebesar 74%.

Selain itu, hasil analisis juga menunjukkan bahwa aspek seperti layanan pengiriman, harga, dan kualitas produk menjadi topik yang paling dominan dalam membentuk sentimen pengguna. Kata-kata kunci seperti kata “cepat”, “kurir”, dan “barang” menjadi kata dengan bobot tertinggi, menandakan bahwa aspek logistik merupakan elemen paling kuat yang membentuk persepsi pengguna terhadap aplikasi Tokopedia. Analisis juga menemukan bahwa klasifikasi ulasan positif cenderung memiliki akurasi yang lebih tinggi dibandingkan klasifikasi lain. Hal ini dapat terjadi karena klasifikasi sentimen negatif cenderung memiliki ekspresi teks informal yang ambigu.

Daftar Pustaka

- [1] R. Nurfitriana Handayani, “Optimasi Algoritma Support Vector Machine Untuk Analisis Sentimen Pada Ulasan Produk Tokopedia Menggunakan Pso,” *Media Inform.*, vol. 20, no. 2, pp. 97–108, 2021.
- [2] M. Ahmad, S. Aftab, M. S. Bashir, N. Hameed, I. Ali, and Z. Nawaz, “SVM optimization for sentiment analysis,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 4, pp. 393–398, 2018, doi: 10.14569/IJACSA.2018.090455.
- [3] R. Obiedat *et al.*, “Sentiment Analysis of Customers’ Reviews Using a Hybrid Evolutionary SVM-Based Approach in an Imbalanced Data Distribution,” *IEEE Access*, vol. 10, pp. 22260–22273, 2022, doi: 10.1109/ACCESS.2022.3149482.
- [4] R. S. Putra and R. Dijaya, “Analisis Sentimen Pada Ulasan Aplikasi Ferizy Menggunakan Metode Support Vector Machine,” *Archive.Umsida.Ac.Id*, pp. 1–10, 2023, [Online]. Available: <https://archive.umsida.ac.id/index.php/archive/preprint/view/3441>
- [5] D. Pajri, Y. Umaidah, and T. N. Padilah, “K-Nearest Neighbor Berbasis Particle Swarm

- Optimization untuk Analisis Sentimen Terhadap Tokopedia,” *J. Tek. Inform. dan Sist. Inf.*, vol. 6, no. 2, pp. 242–253, 2020, doi: 10.28932/jutisi.v6i2.2658.
- [6] I. S. Milal, M. H. M. Hasanudin, M. A. Nur Azhari, R. A. Nugraha, N. Agustina, and S. E. Damayanti, “Klasifikasi Teks Review Pada E-Commerce Tokopedia Menggunakan Algoritma Svm,” *Naratif J. Nas. Riset, Apl. dan Tek. Inform.*, vol. 5, no. 1, pp. 34–45, 2023, doi: 10.53580/naratif.v5i1.191.
- [7] M. I. Putri and I. Kharisudin, “Penerapan Synthetic Minority Oversampling Technique (SMOTE) Terhadap Analisis Sentimen Data Review Pengguna Aplikasi Marketplace Tokopedia,” *Prism. Pros. Semin. Nas. Mat.*, vol. 5, pp. 759–766, 2022, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/>
- [8] R. Apriani and D. Gustian, “Analisis Sentimen Dengan Naïve Bayes Terhadap Komentar Aplikasi Tokopedia,” *J. Rekayasa Teknol. Nusa Putra*, vol. 6, no. 1, pp. 54–62, 2019, doi: 10.52005/rekayasa.v6i1.86.
- [9] S. M. Salsabila, A. Alim Murtopo, and N. Fadhilah, “Analisis Sentimen Pelanggan Tokopedia Menggunakan Metode Naïve Bayes Classifier,” *J. Minfo Polgan*, vol. 11, no. 2, pp. 30–35, 2022, doi: 10.33395/jmp.v11i2.11640.
- [10] R. Tineges, A. Triayudi, and I. D. Sholihat, “Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM),” *J. Media Inform. Budidarma*, vol. 4, no. 3, p. 650, 2020, doi: 10.30865/mib.v4i3.2181.
- [11] J. Alga, C. Wulandari, and B. Intan, “Analisis Sentimen Aplikasi Youtube di Google Play Store Menggunakan Machine Learning,” vol. 4, no. 4, pp. 408–416, 2024.