

Analisis Sentimen Review Pelanggan Lazada dengan Sastrawi Stemmer dan SVM-PSO untuk Memahami Respon Pengguna

Abdun Nafi¹, Aris Tri Jaka Harjanta², Bambang Agus Herlambang³, Saeful Fahmi⁴

^{1,2,3,4} Informatika, Universitas PGRI Semarang, Jl. Sidodadi Timur, Kota Semarang, 50232, Jawa Tengah, Indonesia

***Email Korespondensi:**
abdunnafi444@gmail.com

Abstrak

Di era digital, analisis sentimen memainkan peran strategis dalam memahami persepsi pelanggan terhadap produk dan layanan. Penelitian ini bertujuan untuk menganalisis sentimen ulasan pelanggan pada platform Lazada melalui penerapan teknik pemrosesan teks dan algoritma machine learning. Data diambil dari ulasan produk pada platform tersebut, yang kemudian melalui tahap pra-proses, meliputi tokenisasi, penghapusan stopword, dan stemming menggunakan algoritma Sastrawi. Selanjutnya, klasifikasi sentimen dilakukan dengan menggunakan Support Vector Machine yang dioptimalkan melalui metode Particle Swarm Optimization (PSO). Hasil penelitian menunjukkan bahwa kombinasi metode Sastrawi stemmer dan SVM-PSO mampu mencapai akurasi yang signifikan, yaitu sebesar 90,57%, mengalami peningkatan sebesar 6,24% dibandingkan penelitian sebelumnya. Temuan ini memberikan wawasan yang mendalam mengenai persepsi pelanggan, serta menawarkan panduan berharga bagi pengambil keputusan di Lazada dalam meningkatkan kualitas layanan dan kepuasan pelanggan. Studi ini juga menggarisbawahi pentingnya penerapan teknik Natural Language Processing dan algoritma machine learning dalam analisis sentimen pada platform e-commerce, yang terbukti mampu menghasilkan keluaran yang lebih akurat.

Kata Kunci : Analisis sentiment; Sastrawi stemmer; SVM-PSO

Abstract

In the digital era, sentiment analysis plays a strategic role in understanding customer perceptions of products and services. This research aims to analyze customer review sentiment on the Lazada platform through the application of text processing techniques and machine learning algorithms. Data is taken from product reviews on the platform, which then undergoes a preprocessing stage, including tokenization, stopword removal, and stemming using the Sastrawi algorithm. Next, sentiment classification is performed using a Support Vector Machine optimized through the Particle Swarm Optimization (PSO) method. The research results showed that the combination of the Sastrawi stemmer method and SVM-PSO was able to achieve significant accuracy, namely 90.57%, an increase of 6.24% compared to previous research. These findings provide deep insights into customer perceptions and offer valuable guidance for decision-makers at Lazada in improving service quality and customer satisfaction. This study also underscores the importance of applying Natural Language Processing techniques and machine learning algorithms in sentiment analysis on e-commerce platforms, which have proven to produce more accurate outputs.

Keywords: Sentiment Analysis; Sastrawi Stemmer; SVM-PSO

1. Pendahuluan

Analisis sentimen menjadi sangat penting di era digital yang terus berkembang, karena memungkinkan perusahaan untuk memahami secara mendalam bagaimana pelanggan merespon berbagai produk dan layanan (Choirul Rahmadan et al., 2020). Di platform e-commerce seperti Lazada, analisis sentimen dapat memberikan wawasan yang sangat berharga dari ulasan dan komentar pelanggan. Dengan menerapkan teknik analisis sentimen, perusahaan dapat dengan mudah mengidentifikasi kekuatan dan kelemahan produk serta layanan mereka, menggali informasi lebih dalam mengenai keinginan dan kebutuhan pelanggan, dan mengembangkan strategi yang lebih efektif untuk meningkatkan pengalaman pelanggan (Hassan & Islam, 2021). Hal ini tidak hanya membantu dalam perbaikan produk dan layanan yang ada tetapi juga memungkinkan perusahaan untuk

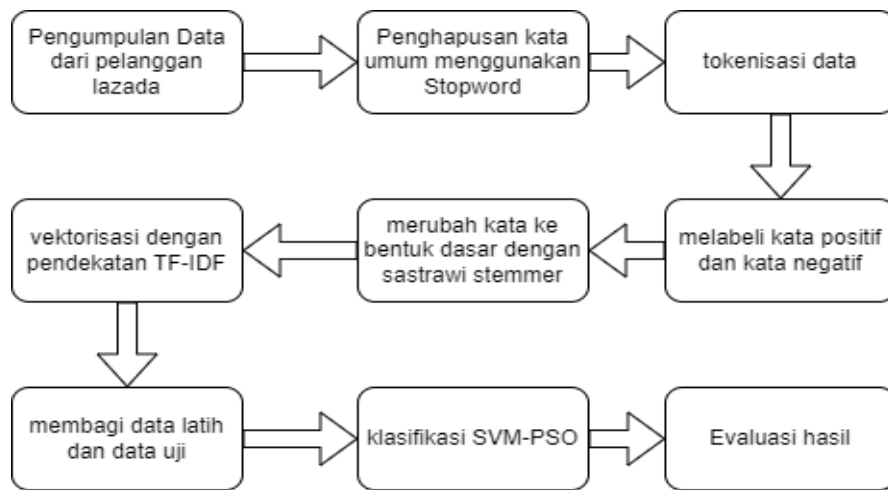
merancang inovasi baru yang lebih sesuai dengan harapan pelanggan, sehingga meningkatkan kepuasan dan loyalitas pelanggan secara signifikan (Saputra et al., 2020).

Support Vector Machines (SVM) adalah algoritma machine learning yang umum digunakan untuk klasifikasi, termasuk analisis sentimen. Algoritma ini berfungsi dengan mencari hyperplane optimal untuk memisahkan kelas sentimen positif dan negatif dalam ruang fitur berdimensi tinggi. Kelebihan SVM adalah kemampuannya dalam menangani data berdimensi besar dan masalah ketidakseimbangan kelas (Hermanto et al., 2024). Sementara itu, *Particle Swarm Optimization* (PSO) adalah algoritma optimasi berbasis populasi yang terinspirasi dari perilaku kawanan hewan, seperti burung dan ikan, dalam mencari solusi optimal. PSO digunakan untuk mengoptimalkan parameter SVM guna meningkatkan akurasi klasifikasi sentimen. PSO bekerja dengan menggerakkan solusi secara iteratif menuju hasil terbaik berdasarkan pengalaman partikel lainnya (LADJAL et al., 2020). *Natural Language Processing* (NLP), cabang ilmu komputer yang mengkaji interaksi antara komputer dan bahasa manusia, digunakan untuk mengekstraksi informasi dari ulasan pelanggan. Proses NLP dalam penelitian ini mencakup tokenisasi, penghapusan stopword, dan stemming (Yadav & Saleena, 2020). Stemming, khususnya dengan algoritma Sastrawi untuk bahasa Indonesia, menghilangkan imbuhan kata untuk mengembalikan kata ke bentuk dasarnya, sehingga menyederhanakan data teks dan meningkatkan akurasi model klasifikasi. Fokus penelitian ini adalah mengevaluasi persepsi konsumen terhadap ulasan yang dibuat di platform *e-commerce* Lazada (Pramita & Nugraha, 2024). Penelitian ini bertujuan untuk mengkategorikan sentimen ulasan pelanggan Menempatkan ke dalam kategori positif dan negatif dengan menerapkan teknik pemrosesan NLP dan algoritma *machine learning* (Djatkiko et al., 2019). Pada penelitian ini, stemmer Sastrawi digunakan untuk memulai proses pra-pemrosesan teks, yang bertujuan untuk meningkatkan akurasi analisis sentimen dengan mengurangi kata menjadi bentuk dasarnya. Proses klasifikasi sentimen dalam penelitian ini dilakukan menggunakan SVM yang di-tuning dengan teknik PSO (Fahmi et al., 2020). Dataset yang digunakan terdiri dari ribuan ulasan konsumen Lazada yang dikumpulkan dari berbagai kategori produk. Tahapan pra-pemrosesan teks mencakup beberapa langkah penting, seperti tokenisasi, penghapusan stopword, dan stemming menggunakan Sastrawi (LADJAL et al., 2020). Hasil dari pra- pemrosesan ini kemudian digunakan sebagai masukan untuk model SVM yang dilatih dan diuji guna mengevaluasi kinerjanya dalam mengkategorikan sentimen ulasan pelanggan. Melalui pendekatan ini, penelitian bertujuan untuk mendapatkan wawasan yang lebih mendalam mengenai persepsi konsumen dan meningkatkan strategi bisnis berdasarkan analisis sentimen yang akurat.

Hasil penelitian menunjukkan bahwa kombinasi stemmer Sastrawi dan algoritma SVM-PSO sangat akurat dalam mengkategorikan sentimen ulasan pelanggan (Mustopa et al., 2020). Temuan ini tidak hanya memberikan pemahaman yang lebih mendalam tentang perilaku pelanggan, tetapi juga memberi para pengambil keputusan di Lazada wawasan berharga untuk meningkatkan layanan dan kepuasan pelanggan. Penelitian ini menekankan pentingnya penggunaan teknik pemrosesan NLP dan machine learning secara bersamaan dalam analisis sentimen untuk platform *e-commerce* (Yadav & Saleena, 2020). Gabungan kedua teknik ini terbukti menghasilkan hasil yang lebih akurat dan informatif. Dengan menerapkan teknik-teknik ini, Lazada dapat enjangkau pemahaman yang lebih komprehensif tentang pola perilaku dan respons pelanggan terhadap produk. Informasi ini vital untuk mengembangkan strategi bisnis yang lebih baik dan meningkatkan posisi kompetitif di pasar *e-commerce* yang semakin kompetitif (Phan et al., 2020). Lebih jauh lagi, pendekatan ini memungkinkan Lazada untuk merancang inovasi produk dan layanan yang lebih sesuai dengan harapan pelanggan, meningkatkan interaksi dan loyalitas pelanggan, serta memperkuat posisinya di pasar global. Dengan demikian, penelitian ini tidak hanya memperkaya teori analisis sentimen, tetapi juga memberikan kontribusi praktis yang signifikan bagi pengembangan strategi bisnis yang berbasis data. Penelitian ini berupaya mengisi kekurangan penelitian sebelumnya dengan mengimplementasikan stemming menggunakan algoritma Sastrawi yang secara spesifik dirancang untuk bahasa Indonesia. Selain itu, penggunaan kombinasi SVM dan PSO sebagai pendekatan optimasi juga menawarkan kontribusi baru dalam meningkatkan performa klasifikasi sentimen. SVM dikenal memiliki kemampuan yang baik dalam menangani data berdimensi tinggi, sementara PSO digunakan untuk mengoptimalkan parameter SVM sehingga dapat memberikan hasil yang lebih akurat dibandingkan pendekatan tradisional.

2. Metode Penelitian

Untuk tujuan klasifikasi sentimen ulasan aplikasi Lazada dalam penelitian ini, digunakan algoritma SVM- PSO. Proses klasifikasi ini melibatkan beberapa langkah kunci, yang dimulai dengan pengumpulan data ulasan pelanggan dari platform Lazada sebagai bagian dari prosedur utama. Setelah data terkumpul, dilakukan beberapa tahap praproses, termasuk penghapusan stopword untuk menghilangkan kata-kata umum yang tidak berkontribusi pada analisis, dan tokenisasi untuk memecah teks menjadi satuan kata atau frasa yang lebih kecil. Selanjutnya, data diberi label sentimen (positif atau negatif) sesuai dengan konteks ulasan, sebelum diterapkan proses stemming menggunakan algoritma Sastrawi, yang berfungsi mengubah kata-kata menjadi bentuk dasar dalam bahasa Indonesia. Setelah praproses selesai, data dibagi menjadi dua kelompok: data latih dan data uji, dengan proporsi 70:30, untuk melatih model dan mengevaluasi kinerjanya. Pada tahap klasifikasi, digunakan algoritma SVM yang dioptimasi dengan PSO untuk memaksimalkan akurasi model. Evaluasi performa model dilakukan dengan menggunakan metrik akurasi, presisi, recall, dan F1-score, sebagaimana digambarkan pada Gambar 1. Prosedur ini dirancang untuk memastikan bahwa model dapat mengklasifikasikan sentimen ulasan secara efektif dan dapat memberikan hasil yang optimal.



Gambar 1. Metode Klasifikasi

Dataset ini dikumpulkan dari berbagai opini konsumen Lazada, mencakup ulasan positif yang memuji kualitas produk serta ulasan negatif yang mengkritik layanan pelanggan. Terdapat total 7.494 Masukan data yang terkait dengan ulasan dari pengguna. Meskipun setiap masukan data memiliki informasi yang beragam, penelitian ini berpusat pada dua kolom: 'ReviewContent' dan 'Rating', seperti yang ditunjukkan pada Tabel 1. Kolom 'reviewcontent' berisi teks ulasan dari pengguna, sedangkan kolom 'rating' Menampilkan penilaian angka yang diberikan oleh pengguna terhadap aplikasi tersebut.

Tabel 1. Dataset

| ReviewContent | Rating |
|---|--------|
| Suka sekali sama pelayanan Lazada. Pesan kemarin pagi, dan pagi ini jam 11 langsung sampai dengan packing yang rapi, ada buble nya sama kardus dari Lazada. Isi pun sesuai dengan apa yang di deskripsikan dan barangnya segel Asus | 5 |
| Kenapa barang barang pesanan saya belum di kirim juga tanggal 13 desember 2022. Lama banget mana nih kenapa Lazada diam saja | 1 |

Stopword merujuk pada kata-kata yang sering muncul dan bersifat umum dalam bahasa alami, seperti "dan", "atau", "yang", "pada", dan sebagainya (J. Liu et al., 2020). Meskipun memiliki frekuensi kemunculan yang tinggi, kata-kata Kurang memberikan informasi yang berarti dalam analisis teks. Dengan Menyingkirkan kata-kata yang tidak diperlukan, kita dapat mengurangi dimensi ruang vektor yang dihasilkan dari proses vektorisasi, yang pada

©2024 J-Intech. Published by LPPM STIKI Malang
This is an open access article under the CC BY SA license. (<https://creativecommons.org/licenses/by-sa/4.0/>)

gilirannya meningkatkan efisiensi komputasi dan akurasi model. Fokus Kepada kata-kata yang memberikan informasi lebih memungkinkan kita agar meningkatkan relevansi karakteristik yang digunakan dalam model pembelajaran mesin, sehingga model dapat lebih efektif dalam menangkap makna dari teks yang dianalisis. Contoh dataset yang sudah melewati proses penghapusan stopwords pada tabel 2.

Tabel 2. Penghapusan Stopword

| ReviewContent | Setelah stopwords |
|---|---|
| Suka sekali sama pelayanan Lazada. Pesan kemarin pagi, dan pagi ini jam 11 langsung sampai dengan packing yang rapi, ada bublehnya sama kardus dari Lazada. Isi pun sesuai dengan apa yang di deskripsikan dan barangnya segel Asus | Suka sekali pelayanan Lazada. Pesan kemarin pagi, pagi ini jam 11 langsung sampai packing rapi, ada bublehnya kardus Lazada. Isi sesuai deskripsikan barang segel Asus. |
| Kenapa barang barang pesanan saya belum di kirim juga tanggal 13 desember 2022. lama banget mana nih kenapa Lazada diam saja | Kenapa barang pesanan belum kirim tanggal 13 Desember 2022. Lama banget Lazada diam saja. |

Dalam tokenisasi, teks dipecah menjadi unit-unit kecil atau kata-kata terpisah yang disebut token (Huan et al., 2020). Ini adalah tahap awal yang sangat penting untuk pemrosesan NLP. Tokenisasi memungkinkan komputer untuk memahami dan menganalisis teks dengan lebih efektif, karena setiap token dapat diperlakukan sebagai unit data independen yang dapat diproses lebih lanjut. Proses ini membantu dalam mengidentifikasi kata-kata kunci, menghilangkan kata-kata yang tidak relevan, dan mempersiapkan teks untuk langkah-langkah NLP berikutnya seperti pelabelan, stemming, dan vektorisasi. Tanpa tokenisasi yang tepat, analisis teks akan menjadi kurang akurat dan efisien.

Proses label dalam pembelajaran mesin dan analisis sentimen melibatkan penetapan label atau kategori pada setiap data berdasarkan fitur atau atribut yang ada (Cruz & Amado, 2020). label digunakan untuk menghubungkan fitur data dengan label yang tepat, sehingga model dapat belajar mengenali pola dan tren dalam data tersebut. Dalam penelitian ini, angka 1 digunakan untuk melabeli sentimen positif dan angka 0 digunakan untuk melabeli sentimen negatif. Secara spesifik, jika rating > 4, sentimen tersebut dilabeli sebagai '1' (positif), sedangkan jika rating < 4, sentimen dilabeli sebagai '0' (negatif). Pendekatan ini memberikan struktur yang jelas untuk data sentimen, memungkinkan model untuk lebih mudah membedakan antara ulasan positif dan negatif. Contoh label terlampir pada tabel 3.

Tabel 3. Pelabelan

| ReviewContent | Rating | label |
|---|--------|-------|
| Suka sekali pelayanan Lazada. Pesan kemarin pagi, pagi ini jam 11 langsung sampai packing rapi, ada bublehnya kardus Lazada. Isi sesuai deskripsikan barang segel Asus. | 5 | 1 |
| Kenapa barang pesanan belum kirim tanggal 13 Desember 2022. Lama banget Lazada diam saja. | 1 | 0 |

Stemming merujuk pada proses mengonversi kata ke bentuk dasar dengan menghapus elemen tambahan seperti awalan, sisipan, dan akhiran (Li et al., 2020). Proses stemming bertujuan untuk menyederhanakan variasi kata sehingga analisis kata menjadi lebih efisien dan konsisten. Dengan mengubah kata-kata ke bentuk dasarnya, stemming membantu dalam mengurangi redundansi dan meningkatkan kualitas data yang digunakan dalam pemrosesan bahasa alami. Hal ini memungkinkan algoritma untuk fokus pada inti makna kata, sehingga meningkatkan akurasi dan efisiensi dalam analisis teks serta dalam membangun model pembelajaran mesin. Hasil stemming menggunakan sastrawi terlampir pada tabel 4.

Tabel 4. Sastrawi Stemming

| ReviewContent | Stemming sastrawi |
|--|--|
| Suka sekali pelayanan Lazada. Pesan kemarin pagi, pagi ini jam 11 langsung sampai packing rapi, ada bublenya kardus Lazada. Isi sesuai deskripsikan barang segel Asus. | Suka sekali layan Lazada. Pesan kemarin pagi, pagi ini jam 11 langsung sampai packing rapi, ada buble kardus Lazada. Isi sesuai deskripsi barang segel Asus. |
| Kenapa barang pesanan belum kirim tanggal 13 Desember 2022. Lama banget Lazada diam saja. | Kenapa barang pesan belum kirim tanggal 13 Desember 2022. Lama banget Lazada diam saja. |

Proses vektorisasi mengubah kata menjadi data numerik yang dapat digunakan oleh algoritma pembelajaran mesin (L. Liu et al., 2020). Pendekatan *Term Frequency-Inverse Document Frequency* (TF-IDF) adalah salah satu metode vektorisasi yang populer yang mengukur pentingnya suatu kata dalam dokumen relatif terhadap seluruh dokumen dalam korpus. Untuk mengonversi teks di kolom prapemrosesan menjadi format numerik, penelitian ini menggunakan library *TfidfVectorizer* dari *sklearn.feature_extraction.text*. Dengan menggunakan TF-IDF, setiap kata diubah menjadi vektor numerik berdasarkan frekuensi kemunculannya dalam dokumen dan pentingnya dalam keseluruhan korpus, sehingga memungkinkan model untuk lebih baik dalam menangkap informasi relevan dari teks.

Pembagian data merupakan langkah penting berikutnya dalam proses pembangunan model pembelajaran mesin (Zhang et al., 2020). Tujuan utama dari pembagian data adalah untuk membagi dataset menjadi dua bagian utama: Data Latih, yang digunakan untuk melatih model agar dapat belajar pola dan hubungan antara fitur dan target, dan Data Uji, yang digunakan untuk menguji dan mengevaluasi kinerja model setelah proses pelatihan. Pada penelitian ini, data dibagi menjadi dua kelompok, yaitu 70% untuk data latih dan 30% untuk data uji. Pembagian ini mengikuti praktik umum dalam pembelajaran mesin, di mana mayoritas data digunakan untuk melatih model agar mampu mengenali pola dan karakteristik sentimen, sedangkan sisanya digunakan untuk mengevaluasi performa model. Proporsi 70:30 ini dipilih untuk memastikan keseimbangan yang optimal antara

jumlah data yang digunakan untuk pelatihan dan validasi model, sehingga dapat mengurangi risiko overfitting atau underfitting pada hasil klasifikasi sentimen. Dengan memisahkan data ini, kita dapat memastikan bahwa model dilatih pada satu subset data dan diuji pada subset data lainnya, sehingga evaluasi kinerja model menjadi lebih objektif dan valid.

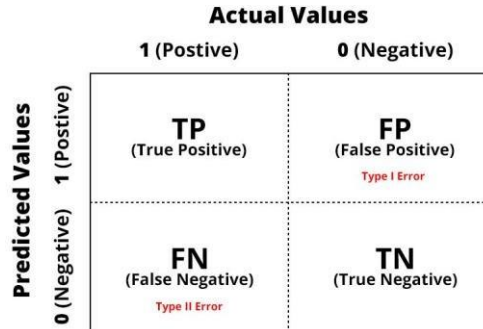
SVM merupakan teknik untuk klasifikasi yang sangat efektif, namun performanya sangat bergantung pada pemilihan parameter yang optimal (1), seperti parameter regularisasi (C) dan parameter kernel (γ untuk kernel RBF) (Punitha & Jeyakarthic, 2020). Dalam penelitian ini, kami menerapkan PSO untuk menentukan kombinasi parameter yang paling optimal bagi SVM (2), dengan harapan dapat meningkatkan akurasi dan efisiensi model klasifikasi yang dihasilkan.

$$f(x) = \sum \alpha_i K(x_i, x) + b \quad (1)$$

$$f(x, y) = (1 - x)^2 + 100(y - x^2)^2 \quad (2)$$

Pengukuran kinerja klasifikasi menggunakan SVM sering kali melibatkan berbagai metrik penting seperti *confusion matrix*, Akurasi, Presisi, Recall, dan Skor F1. *Confusion matrix* adalah tabel yang digunakan untuk merangkum hasil prediksi model klasifikasi dengan memperlihatkan jumlah prediksi benar dan salah untuk setiap kelas (Shuran & Yian, 2020). Matriks ini terdiri dari empat komponen: TP (*True Positive*), TN (*True Negative*), FP (*False Positive*), dan FN (*False Negative*) seperti pada gambar

2. Dengan memanfaatkan metrik-metrik ini, kita dapat memperoleh gambaran menyeluruh tentang kinerja model SVM dalam menangani tugas klasifikasi, sehingga mempermudah dalam mengidentifikasi kekuatan dan kelemahan model serta melakukan perbaikan jika diperlukan.



Gambar 2. Confusion Matrix

3. Hasil

Penelitian ini bertujuan untuk mengevaluasi dampak penerapan Sastrawi Stemmer dan optimasi terhadap keakuratan metode klasifikasi SVM yang digunakan dalam analisis sentimen. Penelitian ini menghadapi tantangan signifikan terkait dengan dataset yang digunakan, yang banyak mengandung singkatan serta penggunaan bahasa yang tidak standar. Langkah pertama yang dilakukan dalam penelitian ini adalah *text preprocessing*, yang meliputi pembersihan data, tokenisasi, pelabelan, penghapusan stopwords, dan stemming. Dalam penelitian ini, penghapusan stopwords dan proses stemming dilakukan dengan menggunakan pustaka Sastrawi. Setelah pembersihan data didapatkan hasil data positif dan negatif seperti pada gambar 3.



Gambar 3. kata positif kata negatif

Langkah selanjutnya adalah melakukan klasifikasi menggunakan beberapa metode klasifikasi yang sering digunakan dalam penelitian analisis sentimen. Metode ini dipilih untuk menentukan mana yang lebih cocok dengan dataset dan memberikan hasil terbaik. Skenario klasifikasi dan evaluasi menggunakan metode klasifikasi (SVM, SVM-PSO, Naïve Bayes, Random Forest dan K-NN). Dalam proses validasi, kami menggunakan cross-validation. Berikut adalah hasil klasifikasi yang kami dapatkan pada tabel 5.

Tabel 5. Hasil Klasifikasi

| Metode klasifikasi | Hasil akurasi |
|--------------------|---------------|
| SVM | 89,46% |
| SVM-PSO | 90,57% |
| Naïve Bayes | 83,50% |
| Random Forest | 85,99% |
| K-NN | 64,07% |

4. Pembahasan

Pada percobaan klasifikasi menggunakan SVM, diperoleh akurasi sebesar 89,46%. Angka akurasi ini menunjukkan bahwa metode SVM memberikan hasil yang cukup tinggi dalam mengklasifikasikan data dengan tepat. Untuk memberikan pemahaman lebih lanjut mengenai performa model, Tabel 6 menunjukkan matriks konfusi yang merinci distribusi prediksi dan label asli selama proses pengujian. Matriks konfusi ini membantu dalam menganalisis bagaimana model SVM menangani setiap kelas, termasuk jumlah prediksi benar dan salah untuk masing-masing kategori. Analisis mendalam terhadap matriks konfusi dapat memberikan wawasan lebih lanjut mengenai kekuatan dan kelemahan model dalam konteks klasifikasi spesifik yang dilakukan.

Tabel 6. Confusion Matrix SVM

| | <i>True.1</i> | <i>True.0</i> |
|---------------|---------------|---------------|
| <i>Pred.1</i> | 970 | 154 |
| <i>Pred.0</i> | 83 | 1042 |

Pada percobaan klasifikasi yang menggunakan metode SVM yang dioptimalkan dengan PSO, diperoleh akurasi sebesar 90,57%. Penerapan optimasi PSO secara signifikan meningkatkan tingkat akurasi dibandingkan dengan model SVM yang tidak menggunakan PSO, menunjukkan bahwa teknik ini efektif dalam meningkatkan performa klasifikasi. PSO membantu dalam menemukan kombinasi hyperparameter yang lebih optimal untuk model SVM, yang berkontribusi pada peningkatan akurasi. Untuk memahami lebih dalam tentang hasil dari uji coba ini, Tabel 7 menyajikan matriks konfusi yang memberikan rincian tentang distribusi prediksi.

Tabel 7. Confusion Matrix SVM-PSO

| | <i>True.1</i> | <i>True.0</i> |
|---------------|---------------|---------------|
| <i>Pred.1</i> | 1030 | 94 |
| <i>Pred.0</i> | 118 | 1007 |

Pada percobaan klasifikasi yang menggunakan metode Naive Bayes dengan penerapan stemmer Sastrawi, diperoleh akurasi sebesar 83,50%. Meskipun Naive Bayes memberikan hasil yang memadai, tingkat akurasi ini lebih rendah dibandingkan dengan model SVM pada dataset yang sama. Penurunan akurasi ini menunjukkan bahwa Naive Bayes mungkin kurang efektif dalam menangani kompleksitas atau karakteristik khusus dari data dibandingkan dengan SVM. Analisis lebih lanjut melalui matriks konfusi yang disajikan dalam Tabel 8.

Tabel 8. Confusion Matrix Naive Bayes

| | <i>True.1</i> | <i>True.0</i> |
|---------------|---------------|---------------|
| <i>Pred.1</i> | 970 | 154 |
| <i>Pred.0</i> | 217 | 908 |

Pada percobaan klasifikasi yang menggunakan metode Random Forest dengan penerapan stemmer Sastrawi, diperoleh akurasi sebesar 85,99%. Hasil ini menunjukkan peningkatan akurasi dibandingkan dengan model Naive Bayes yang mencapai 83,50%, tetapi masih sedikit lebih rendah dibandingkan dengan akurasi yang dicapai oleh model SVM yang mencapai 89,46%. Tabel 9 menyajikan matriks konfusi yang memberikan rincian tentang distribusi prediksi.

Tabel 9. Confusion Matrix Random Forest

| | <i>True.1</i> | <i>True.0</i> |
|---------------|---------------|---------------|
| <i>Pred.1</i> | 1019 | 105 |
| <i>Pred.0</i> | 210 | 915 |

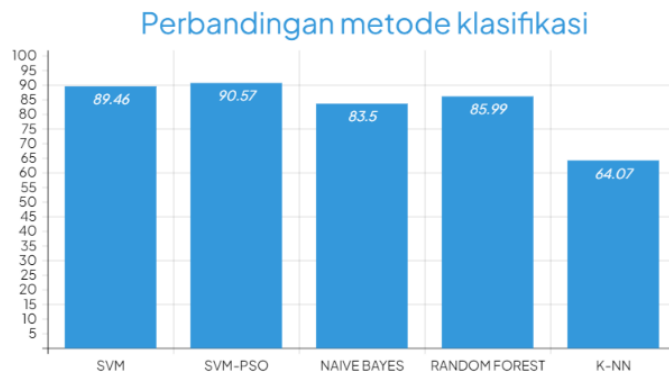
Pada percobaan klasifikasi yang menggunakan metode K-Nearest Neighbor (K-NN) dengan penerapan stemmer Sastrawi, diperoleh akurasi sebesar 64,07%. Hasil akurasi ini menunjukkan bahwa K-NN memiliki performa yang relatif lebih rendah dibandingkan dengan membandingkan metode klasifikasi lain seperti Naive Bayes, Random Forest, dan SVM. Analisis lebih lanjut melalui matriks konfusi yang disajikan dalam Tabel 10 mengungkapkan bahwa metode K-NN mengalami kesulitan signifikan dalam mengklasifikasikan beberapa kelas. Secara khusus,

model ini sering kali salah mengidentifikasi kelas yang seharusnya dikategorikan sebagai positif menjadi negatif. Matriks konfusi memberikan rincian tentang distribusi prediksi benar dan salah untuk masing-masing kelas, menyoroti bahwa K-NN mungkin tidak cukup efektif dalam menangani variasi data atau perbedaan kelas yang ada dalam dataset.

Tabel 10. Confusion Matrix K-NN

| | True.1 | True.0 |
|--------|--------|--------|
| Pred.1 | 1074 | 50 |
| Pred.0 | 762 | 363 |

Gambar 4 menunjukkan bahwa hasil akurasi terbaik untuk klasifikasi analisis sentimen adalah 90,57% dengan menggunakan metode SVM yang dioptimalkan dengan SVM-PSO. Penerapan Sastrawi stemmer memberikan peningkatan akurasi sebesar 6,24% dibandingkan dengan penelitian sebelumnya yang tidak menggunakan Sastrawi stemmer dan PSO.



Gambar 4. Perbandingan Metode Klasifikasi

Dalam kajian yang terkait yang dilakukan oleh Alvina Gusti Pramita dan Fajar Nugraha, penggunaan algoritma K-Means berhasil mencapai akurasi sebesar 87% (Rusyanti & Sembiring, 2022). Penelitian ini menunjukkan bahwa k-means memiliki akurasi yang bagus dan menjadi alat yang efektif dalam analisis sentimen, terutama ketika diterapkan pada dataset yang sesuai dan diproses dengan baik. Di sisi lain, penelitian lain yang dilakukan oleh Vava Alessandro Riyanto dan Dwi Budi Santoso menggunakan algoritma SVM tanpa melakukan proses stemming dan optimasi menggunakan PSO. Hasil penelitian ini menunjukkan akurasi sebesar 84,33% (Riyanto & Santoso, 2024). Meskipun SVM merupakan algoritma yang kuat dalam klasifikasi, hasil ini menyoroti pentingnya proses pra-pemrosesan teks seperti stemming dan optimasi parameter untuk meningkatkan kinerja algoritma. Kedua penelitian ini memberikan wawasan berharga tentang pengaruh berbagai teknik dan metode dalam analisis sentimen, serta menunjukkan bagaimana kombinasi yang tepat dari algoritma dan pra- pemrosesan data dapat menghasilkan peningkatan signifikan dalam akurasi model.

Peningkatan akurasi ini mengindikasikan bahwa stemmer Sastrawi berperan penting dalam meningkatkan kualitas pemrosesan teks, terutama dalam mengurangi variasi kata yang tidak perlu dan meningkatkan konsistensi dalam representasi data. Dengan menerapkan Sastrawi stemmer, model SVM-PSO tidak hanya lebih efektif dalam menangani variasi linguistik dalam data, tetapi juga mampu memberikan hasil klasifikasi yang lebih akurat. Peningkatan ini menunjukkan bahwa teknik pra-pemrosesan teks seperti stemming dapat memberikan kontribusi signifikan terhadap performa model klasifikasi, terutama dalam konteks analisis sentimen di mana ketepatan interpretasi bahasa sangat penting.

5. Kesimpulan

Berdasarkan hasil percobaan, diperoleh akurasi tertinggi dalam klasifikasi sentimen pada dataset ulasan pelanggan Lazada dengan menggunakan SVM yang dioptimasi melalui PSO, serta memanfaatkan stemmer Sastrawi dalam proses stemming. Metode SVM-PSO, dengan akurasi mencapai 90,57%, menunjukkan performa yang lebih baik dibandingkan metode klasifikasi lainnya yang diuji. Penerapan stemmer Sastrawi juga

memberikan kontribusi signifikan terhadap peningkatan akurasi, yaitu sebesar 6,24% dibandingkan dengan model yang tidak menggunakan stemming. Implikasi praktis dari penelitian ini adalah bahwa Lazada, serta

platform e-commerce lainnya, dapat memanfaatkan metode SVM-PSO ini untuk meningkatkan sistem analisis sentimen otomatis pada ulasan pelanggan. Dengan hasil klasifikasi yang lebih akurat, Lazada dapat mengidentifikasi sentimen pelanggan secara lebih efektif, yang dapat mendukung pengambilan keputusan yang lebih cepat dan tepat dalam menanggapi kebutuhan dan keluhan pelanggan. Penerapan model ini dapat membantu Lazada dalam meningkatkan strategi customer engagement, memperbaiki layanan pelanggan, serta mengidentifikasi produk-produk yang memerlukan perbaikan berdasarkan ulasan negatif. Selain itu, model yang diusulkan juga dapat diterapkan untuk mendeteksi tren sentimen secara real-time, sehingga platform dapat menyesuaikan strategi pemasaran dan manajemen produk secara lebih responsif. Harapannya, penelitian selanjutnya dapat meningkatkan akurasi yang lebih tinggi dengan metode optimasi dan teknik praproses yang lebih lanjut, serta menerapkan model ini pada dataset yang lebih besar untuk memperluas generalisasi temuan.

Referensi

- Choirul Rahmadan, M., Nizar Hidayanto, A., Swadani Ekasari, D., Purwandari, B., & Theresiawati. (2020). Sentiment Analysis and Topic Modelling Using the LDA Method related to the Flood Disaster in Jakarta on Twitter. *2020 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, 126–130. <https://doi.org/10.1109/ICIMCIS51567.2020.9354320>
- Cruz, J. C. Dela, & Amado, T. M. (2020). Development of Machine Learning-based Predictive Models for Wireless Indoor Localization Application with Feature Ranking via Recursive Feature Elimination Algorithm. *2020 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, 1–6. <https://doi.org/10.1109/ICSPCC50002.2020.9259526>
- Djatmiko, F., Ferdiana, R., & Faris, M. (2019). A Review of Sentiment Analysis for Non-English Language. *2019 International Conference of Artificial Intelligence and Information Technology (ICAIIIT)*, 448–451. <https://doi.org/10.1109/ICAIIIT.2019.8834552>
- Fahmi, S., Purnamawati, L., Shidik, G. F., Muljono, M., & Fanani, A. Z. (2020). Sentiment Analysis of Student Review in Learning Management System Based on Sastrawi Stemmer and SVM-PSO. *2020 International Seminar on Application for Technology of Information and Communication (ISemantic)*, 643–648. <https://doi.org/10.1109/iSemantic50169.2020.9234291>
- Hassan, R., & Islam, Md. R. (2021). Impact of Sentiment Analysis in Fake Online Review Detection. *2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)*, 21–24. <https://doi.org/10.1109/ICICT4SD50815.2021.9396899>
- Hermanto, H., Fahlapi, R., Kuntoro, A. Y., & Asra, T. (2024). Perbandingan Algoritma Klasifikasi Analisis Sentimen Pengguna Aplikasi Getcontact Dalam Pencegahan Penipuan Online. *J-INTECH*, 12(1), 158–167. <https://doi.org/10.32664/j-intech.v12i1.1262>
- Huan, L., Jingqi, M., & Xingjian, L. (2020). Target tracking method based on the fusion of structured SVM and KCF algorithm. *2020 Chinese Control And Decision Conference (CCDC)*, 1174–1178. <https://doi.org/10.1109/CCDC49329.2020.9164843>
- LADJAL, M., OUALI, M. A., & LASS, M. D. (2020). optimization of SVM parameters with hybrid PCA-PSO methods for water quality monitoring. *2020 International Conference on Electrical Engineering (ICEE)*, 1–6. <https://doi.org/10.1109/ICEE49691.2020.9249881>
- Li, J., Zhao, Y., Gao, X., & Zhao, W. (2020). Feature Extraction and SVM Parameters Optimization Based on ICQPSO. *2020 International Conference on Computer Network, Electronic and Automation (ICCNEA)*, 329–332. <https://doi.org/10.1109/ICCNEA50255.2020.00074>
- Liu, J., Zhu, X., & Zhang, Y. (2020). Application of DE-GWO-SVM Algorithm in Business Order Prediction Model. *2020 IEEE 11th International Conference on Software Engineering and Service Science (ICSESS)*, 432–435. <https://doi.org/10.1109/ICSESS49938.2020.9237714>
- Liu, L., Sha, Y., & Zhang, Q. (2020). Bearing fault diagnosis analysis based on improved adaptive search algorithms and SVMs. *2020 5th International Conference on Automation, Control and Robotics Engineering (CACRE)*, 680–684. <https://doi.org/10.1109/CACRE50138.2020.9230327>
- Mustopa, A., Hermanto, Anna, Pratama, E. B., Hendini, A., & Risdiansyah, D. (2020). Analysis of User Reviews for the PeduliLindungi Application on Google Play Using the Support Vector Machine and Naive Bayes Algorithm

- Based on Particle Swarm Optimization. *2020 Fifth International Conference on Informatics and Computing (ICIC)*, 1–7. <https://doi.org/10.1109/ICIC50835.2020.9288655>
- Phan, H. T., Tran, V. C., Nguyen, N. T., & Hwang, D. (2020). Improving the Performance of Sentiment Analysis of Tweets Containing Fuzzy Sentiment Using the Feature Ensemble Model. *IEEE Access*, 8, 14630–14641. <https://doi.org/10.1109/ACCESS.2019.2963702>
- Pramita, A. G., & Nugraha, F. (2024). Sistem Analisis Sentimen Produk Pada Aplikasi Lazada Menggunakan Metode Naïve Bayes. *Jurnal Digit*, 14(1), 23. <https://doi.org/10.51920/jd.v14i1.362>
- Punitha, S., & Jeyakarthic, M. (2020). Particle Swarm Optimization based Classification Algorithm for Expert Prediction Systems. *2020 International Conference on Inventive Computation Technologies (ICICT)*, 671–675. <https://doi.org/10.1109/ICICT48043.2020.9112392>
- Riyanto, V. A., & Santoso, D. B. (2024). Penerapan Model Support Vector Machine Pada Klasifikasi Sentimen Ulasan Aplikasi Lazada. *JURASIK (Jurnal Riset Sistem Informasi Dan Teknik Informatika)*, 9(1).
- Rusyanti, D., & Sembiring, F. (2022). Implementasi Algoritma K Means Untuk Sentimen Analisis Review Produk Pada E-Commerce Lazada. *SEMINAR NASIONAL SISTEM INFORMASI & MANAJEMEN INFORMATIKA (SISMATIK 2024)*, 2.
- Saputra, F. T., Wijaya, S. H., Nurhadryani, Y., & Defina. (2020). Lexicon Addition Effect on Lexicon-Based of Indonesian Sentiment Analysis on Twitter. *2020 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, 136–141. <https://doi.org/10.1109/ICIMCIS51567.2020.9354269>
- Shuran, C., & Yian, L. (2020). Breast cancer diagnosis and prediction model based on improved PSO-SVM based on gray relational analysis. *2020 19th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES)*, 231–234. <https://doi.org/10.1109/DCABES50732.2020.00067>
- Yadav, S., & Saleena, N. (2020). Sentiment Analysis Of Reviews Using an Augmented Dictionary Approach. *2020 5th International Conference on Computing, Communication and Security (ICCCS)*, 1–5. <https://doi.org/10.1109/ICCCS49678.2020.9277094>
- Zhang, X., Han, P., Xu, L., Zhang, F., Wang, Y., & Gao, L. (2020). Research on Bearing Fault Diagnosis of Wind Turbine Gearbox Based on 1DCNN-PSO-SVM. *IEEE Access*, 8, 192248–192258. <https://doi.org/10.1109/ACCESS.2020.3032719>