

Sentiment Analysis of Shoe Product Reviews on Indonesian E-Commerce Platform Using Lexicon Based and Support Vector Machine

Fitriani Muttakin^{*1}, Nadila Andrika², Salsabila³

^{1,2,3}Information System, Faculty of Science and Technology, Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia

Email: fitrianimuttakin@uin-suska.ac.id

Received : Sep 22, 2024; Revised : Oct 19, 2024; Accepted : Nov 5, 2024; Published : Apr 26, 2025

Abstract

The rapid development of e-commerce has encouraged people, especially young people, to switch from offline shopping to online platforms such as Shopee that offer fashion products, including shoes, at affordable prices and a wide selection. This phenomenon creates great opportunities for sellers, but also poses challenges related to analyzing product quality contained in customer reviews. The large number of scattered and unstructured reviews makes it difficult for potential buyers to accurately assess products. Therefore, this study aims to analyze the sentiment of 10,323 shoe product reviews on Shopee using the Support Vector Machine (SVM) algorithm and the Lexicon-Based method. SVM was chosen because of its advantage in achieving high accuracy in text classification, with accuracy results reaching 92.62%. The Lexicon-Based method is used to detect specific sentiment words, which provides deeper insight into consumer opinions on shoe products. The analysis results show that shoe product reviews are dominated by positive sentiments, reflecting a high level of customer satisfaction. The findings not only provide guidance for sellers in designing more effective marketing strategies, but also help potential buyers in making better decisions based on objective sentiment analysis. In addition, this study contributes to the literature related to sentiment analysis with SVM in the e-commerce domain, especially for fashion shoes. Thus, the combined use of SVM and Lexicon-Based methods shows great potential in providing valuable insights into consumer preferences as well as increasing customer confidence in choosing shoe products in the e-commerce.

Keywords : *E-Commerce, Lexicon-Based, Sentiment Analysis, Shoe Fashion, Support Vector Machine.*

This work is an open access article and licensed under a Creative Commons Attribution-Non Commercial 4.0 International License



1. PENDAHULUAN

Perkembangan dan popularitas teknologi e-commerce yang cepat menyebabkan semakin banyak orang yang tertarik untuk berbelanja melalui berbagai platform e-commerce daripada metode offline. Saat berbelanja di toko fisik, pelanggan dapat berbelanja kapan saja dan di mana saja tanpa perlu menunggu hingga akhir pekan untuk berbelanja sehingga dapat menghemat waktu dan tenaga para pelanggan [1].

Shopee merupakan salah satu e-commerce yang paling populer dan banyak dikunjungi masyarakat Indonesia. Menurut Sea Group, pada tahun 2022 terdapat 27.000 Usaha Mikro Kecil dan Menengah (UMKM) yang berasal dari Indonesia, Malaysia, dan beberapa negara lain [2]. Dengan demikian, terdapat banyak toko fashion lokal dan impor di Shopee dengan berbagai jenis produk yang ditawarkan seperti baju, sepatu, tas, dan sebagainya [3] [4].

Produk fashion sepatu semakin diminati masyarakat, khususnya kalangan anak muda, karena pilihan model dan merk yang beragam. Pada platform Shopee, produk sepatu ditawarkan dengan harga terjangkau namun tetap memiliki kualitas yang kompetitif jika dibandingkan dengan produk sepatu yang dijual di toko offline. Fenomena ini menyebabkan meningkatnya jumlah pembeli yang beralih ke platform online untuk memenuhi kebutuhan fashion mereka, termasuk kebutuhan alas kaki berupa

sepatu. Seiring dengan semakin banyaknya produk yang tersedia, ulasan atau review pelanggan menjadi penting bagi calon pembeli untuk menilai kualitas produk tersebut. Review pelanggan membantu dalam membentuk persepsi umum mengenai produk sepatu yang dijual di e-commerce.

Namun, tantangan muncul karena informasi yang tersedia dari review seringkali tersebar dan sulit untuk dianalisis secara menyeluruh. Meskipun penelitian terdahulu telah melakukan analisis sentimen pada berbagai produk fashion, masih sedikit penelitian yang secara spesifik berfokus pada produk sepatu di e-commerce seperti Shopee. Oleh karena itu, penelitian ini berupaya untuk mengisi kekosongan/gap tersebut dengan fokus pada analisis sentimen terhadap ulasan produk sepatu di e-commerce, sehingga dapat memberikan wawasan mengenai kelebihan dan kekurangan produk berdasarkan pengalaman konsumen. Melalui analisis sentimen, penelitian ini akan membantu para penjual dan calon pembeli dalam memahami persepsi pelanggan terhadap produk sepatu yang ditawarkan, sehingga dapat meningkatkan kepercayaan pelanggan dan mendukung pengambilan keputusan yang lebih baik dalam memilih produk fashion sepatu di e-commerce [5].

Tujuan analisis sentimen ini untuk mengelompokkan polaritas yang ada dalam teks, seperti dokumen, kalimat, atau bahkan aspek tertentu. Analisis sentimen dapat dilakukan dengan menggunakan data dari review customer untuk menentukan apakah pendapat yang terkandung dalam teks bersifat positif maupun negatif [6]. Jumlah pengunjung yang mengunjungi Shopee berubah dari tahun ke tahun dengan berbagai jenis review yang mereka berikan. Karena jumlah data review pengunjung Shopee yang terus bertambah, penjual akan membutuhkan waktu yang lama untuk membaca dan menyimpulkan review secara menyeluruh pada halaman toko yang berada di Shopee. Selain itu, peringkat yang diberikan tidak dapat menggambarkan pendapat pembeli tentang sepatu lokal tersebut. Persepsi merk yang baik akan membentuk pendapat pembeli tentang barang dan jasa yang diharapkan akan mendorong mereka untuk membeli barang dan jasa tersebut.

Hal ini mendorong peneliti untuk menganalisis pendapat pengguna agar mengetahui persepsi pelanggan terhadap produk fashion sepatu yang tersedia di toko-toko Shopee dengan berbasis metode Lexicon-Based untuk pelabelan, dan algoritma Support Vector Machine [7] sebagai modelnya.

Beberapa penelitian terkait yang telah dilakukan sebelumnya antara lain Riyanto dkk menyimpulkan model Support Vector Machine (SVM) lebih unggul dibandingkan Random Forest dan Decision Tree dalam hal akurasi prediksi kepuasan mahasiswa, dengan akurasi yang tinggi karena penggunaan kernel RBF yang mampu menangkap hubungan non-linear [8]. Algoritma Support Vector Machine (SVM) memberikan hasil lebih unggul dibandingkan Random Forest dalam menganalisis sentimen komentar YouTube boygroup BTS [9]. SVM dengan kernel RBF dan Polynomial menunjukkan akurasi training tertinggi sebesar 96%.

Algoritma Support Vector Machine (SVM) lebih unggul dibandingkan K-Nearest Neighbor (K-NN) dalam mengklasifikasikan sentimen masyarakat terhadap kebijakan kenaikan harga pasca pandemi

COVID-19 [10], dengan menunjukkan SVM mencapai akurasi sebesar 79%, sementara K-NN hanya mencapai 54%. Support Vector Machine (SVM) memberikan performa yang lebih baik dibandingkan dengan Naïve Bayes dalam analisis sentimen produk fashion sepatu [11]. Penelitian analisis sentimen yang menggunakan Algoritma Support Vector Machine disimpulkan sebagai algoritma klasifikasi yang baik dengan nilai akurasi sebesar 98% dan skor F1 sebesar 0.98 terhadap penggunaan aplikasi shopee [12]. Selain itu, Algoritma SVM mengungguli Naïve Bayes digunakan untuk menganalisis sentimen konsumen terhadap produk Amazon, dengan hasil SVM mencapai 84% dan Naive Bayes 82,87% [13].

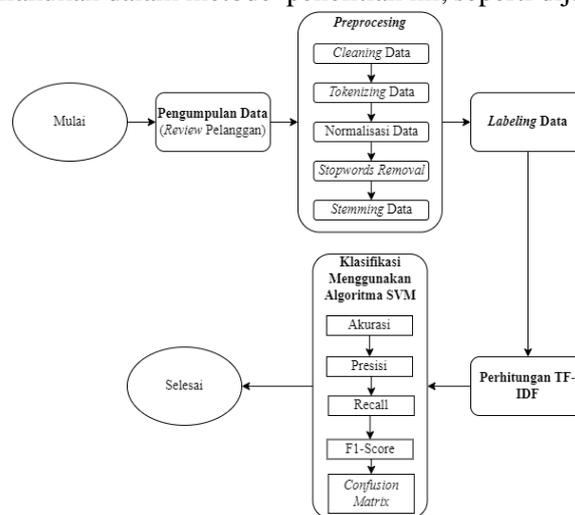
Penelitian terkait selanjutnya dilakukan oleh Watmah dkk pada e-commerce dengan membandingkan klasifikasi menggunakan 3 metode yaitu K-Nearest Neighbor, Support Vector Machine Dan Random Forest Dengan Perolehan Nilai Akurasi Masing-Masing Sebesar KNN 89.0%, Random Forest 83.0%, Dan SVM 89.4%. Hal ini membuktikan bahwa algoritma svm memiliki nilai akurasi lebih baik dibandingkan algoritma knn dan random forest [14].

Selanjutnya, pada Algoritma Multiclass (One Vs Rest) Support Vector Machine berhasil mengklasifikasikan tweet dari akun Shopee, Tokopedia, Bukalapak, dan JDId Online Shop ke dalam tiga kelas (negatif, netral, dan positif). Hal ini menunjukkan bahwa SVM dapat menjadi pilihan yang efektif dalam analisis sentimen tweet dari berbagai akun toko online [15].

Berdasarkan analisis permasalahan dan penelitian terkait, penelitian ini menggunakan Algoritma Klasifikasi Support Vector Machine dan menggunakan lexicon-based karena memiliki kemampuan untuk mengkategorikan komentar pengguna pada aplikasi Shopee secara positif atau negatif dan menghasilkan akurasi model yang baik.

2. METODE PENELITIAN

Beberapa tahapan dilakukan dalam metode penelitian ini, seperti dijabarkan pada gambar 1.



Gambar 1. Metode Penelitian

2.1. Cleaning

Cleaning adalah proses menghapus semua karakter bukan abjad dari review pelanggan untuk menemukan karakter atau simbol yang tidak dipahami atau tidak relevan untuk analisis kalimat [16].

2.2. Preprocessing

Preprocessing dilakukan untuk memastikan data akurat dan bersih. Tahapan ini mencakup proses Cleaning, Case Folding, Tokenizing, Transforming, Stemming, dan Stopword Removal.

2.2.1. Cleaning

Cleaning adalah proses menghapus semua karakter bukan abjad dari review pelanggan untuk menemukan karakter atau simbol yang tidak dipahami atau tidak relevan untuk analisis kalimat [16].

2.2.2. Case Folding

Case Folding merupakan proses penulisan huruf dengan mengubah setiap huruf menjadi lowercase (huruf kecil). Ini hanya dapat menulis ulang huruf A hingga Z dan semua karakter lainnya akan hilang [16].

2.2.3. Tokenizing

Tokenizing adalah cara untuk mengkodekan string tertentu berdasarkan kata-kata yang digunakan. Tahap tokenizing adalah hasil dari cleaning dan case folding, di mana kalimat evaluasi review pelanggan diubah menjadi kata per kata [16]. Tokenisasi adalah proses mengubah judul dokumen menjadi kata kunci atau karakter yang sesuai dengan situasi. Selain itu, semua kata dan kalimat harus ditulis ulang sesuai spasi dokumen [17].

2.2.4. Normalisasi

Normalisasi merupakan langkah yang tepat untuk memanfaatkan secara efektif sejumlah besar teks informal. Normalisasi teks adalah proses mengubah teks yang tidak biasa menjadi format yang lazim. Ini memungkinkan penggunaan teknik pemrosesan bahasa natural (NLP) konvensional bersama dengan metode pemrosesan biasa. Selain itu, menggunakan data yang telah distandarisasi daripada yang belum distandarisasi meningkatkan pembangunan model klasifikasi [18].

2.2.5. Stopword Removal

Stopwords removal merupakan metode untuk menghilangkan kata-kata umum yang tidak berguna dan tidak bermakna. Stopwords yang sering muncul, seperti kata-kata seperti "dan", "atau", "tapi", dan lain-lain, sangat umum. Tujuan stopwords removal adalah untuk menghilangkan kata-kata yang tidak berguna bagi sistem dan tidak memiliki arti, sehingga meningkatkan kinerja dan kecepatan proses. "Penentu", "konjungsi", "proposisi", dan "sejenisnya" adalah beberapa kata henti (Sulistio & Handoyo, 2022). Selain itu, fungsi regex dapat digunakan untuk menghilangkan tanda baca, seperti ".|,|:|!|?|()|'|\"|...|-|_|\$|&|*|>|<\" [19].

2.2.6. Stemming

Stemming merupakan suatu proses untuk mengembalikan kata ke bentuk dasarnya dengan menghapus imbuhan [20]. Tujuan stemming adalah untuk membuat data yang memiliki kata-kata dasar yang sama seragam [21].

2.3. Labeling

Dasar utama data mining dengan Metode Lexicon Based adalah labeling. Data yang diperoleh harus diuraikan lebih lanjut sebagai kalimat yang mengandung nilai positif atau negatif. Setiap ulasan pelanggan harus dilabeli dengan menunjukkan apakah ulasan tersebut positif atau negatif. [21]. Lexicon InSet telah terbukti efektif dalam menganalisis sentimen pada data berbahasa Indonesia. InSet Lexicon, atau Indonesia Sentiment Lexicon, terdiri dari 3.609 kata positif dan 6.609 kata negatif dalam bahasa Indonesia, di mana setiap kata memiliki bobot nilai atau skor polaritas yang berkisar antara -5 hingga +5. Tujuan InSet Lexicon adalah untuk menemukan dan mengkategorikan pendapat tertulis menjadi pendapat positif atau negatif, yang kemudian dapat digunakan untuk mengevaluasi bagaimana masyarakat melihat hal-hal tertentu..[22]. Perhitungan Inset Lexicon adalah dapat dilihat pada Persamaan 1.

$$Nilai Sentimen = \sum(Bobot Kata \times Frekuensi Kata) \tag{1}$$

Keterangan:

Bobot kata : Nilai sentimen kata InSet Lexicon.

Frekuensi kata : Jumlah kemunculan kata dalam teks

2.4. Perhitungan TF-IDF

Metode yang dikenal sebagai Term Frequency-Inverse Document Frequency (TF-IDF) digunakan untuk menentukan jumlah bobotan yang terkait dengan setiap kata yang telah diekstrak. Dalam model pembobotan TF-IDF, frekuensi kata (TF) dan frekuensi dokumen balik (IDF) diintegrasikan untuk menghitung jumlah kata yang sering digunakan dalam proses pencarian data. Frekuensi kata (TF) menghitung jumlah kata yang muncul dalam satu dokumen, dan frekuensi dokumen balik (IDF) menghitung jumlah kata yang muncul di beberapa dokumen (komentar), yang dianggap sebagai kata umum yang tidak penting untuk dinilai. [23]. Metode TF-IDF menggabungkan dua gagasan : Pertama, nilai TF untuk setiap kata dihitung dengan bobot awal 1, dan kemudian dengan menggunakan Persamaan 2 untuk menghitung nilai IDF.

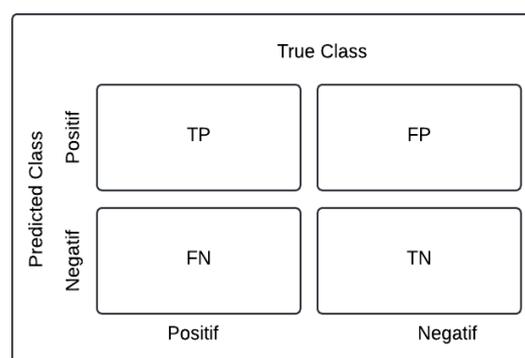
$$IDF(Word) = \log \left(\frac{tf}{df} \right) \tag{2}$$

Metrik IDF (word) digunakan untuk setiap kata yang dicari, di mana td adalah jumlah total dokumen yang ada dan df adalah jumlah kata yang muncul di seluruh dokumen [24].

2.5. Klasifikasi Menggunakan SVM

Pada tahap ini, comment dikumpulkan dalam dua kategori: negatif dan positif. Dalam analisis ini, Support Vector Machine (SVM) adalah algoritma yang digunakan. Setelah data diklasifikasikan, kualitas model yang dibuat dengan mempertahankan akurasi yang diperlukan akan dievaluasi [5]. Menurut Putri dkk metode yang paling umum untuk meregresi atau mengklasifikasikan data berdasarkan data sebelumnya adalah dengan menggunakan Support Vector Machines, dan pekerjaan pemodelan diawasi secara menyeluruh. Support Vector Machines (SVM) adalah contoh pengklasifikasi linier, non-probabilistik, dan biner [19]. SVM menggunakan batas keputusan (Decision Boundaries) untuk membangun model linier atau hiper-bidang yang paling efektif untuk mengklasifikasikan data. Ini menghasilkan klasifikasi dari perbandingan data-data.

Gambar di bawah menunjukkan dasar awal confusion matrix yang dilakukan setelah menghitung metric [25].



Gambar 2. Ilustrasi Predictive Class

Gambar 2 menunjukkan dasar awal confusion matrix yang dilakukan setelah menghitung metric. Dimana TP merupakan True Positive yang bertujuan untuk menghitung jumlah prediksi benar pada kelas positif. Kemudian, TN merupakan True Negative yang bertujuan untuk menghitung jumlah prediksi benar pada kelas negatif. FN merupakan False Negative yang bertujuan untuk menghitung kelas asli positif yang diprediksi negatif. Terakhir, FP merupakan False Positive yang bertujuan untuk menghitung kelas asli negatif yang diprediksi positif.

a. Accuracy

Indikator jumlah kelas yang diprediksi dengan benar oleh model. Model ini adalah yang paling umum digunakan, tetapi dapat menyebabkan interpretasi kinerja yang salah ketika memiliki data yang tidak seimbang. Persamaan untuk metrik ini adalah sebagai berikut:

$$Akurasi = \frac{TN+TP}{TN+FP+TP+FN} \quad (3)$$

b. Precision

Untuk mengukur seberapa akurat model mengidentifikasi kelas positif sebagai kelas yang benar-benar positif, statistik ini tidak dapat menunjukkan dengan jelas jumlah kelas positif aktual yang memiliki hasil prediksi yang akurat. Persamaan untuk metrik ini adalah sebagai berikut:

$$Presisi = \frac{TP}{TP+FP} \quad (4)$$

c. Recall (True Positive Rate (TPR))

Statistik menunjukkan bahwa kelas positif (aktual) diramalkan sebagai kelas positif; namun, statistik ini tidak menangkap kemampuan model untuk meramalkan kelas positif sebagai kelas positif. Persamaan yang digunakan untuk menilai metrik ini adalah sebagai berikut:

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

d. F1 Score

Sebuah ukuran yang menghitung kekurangan dari recall dan akurasi dalam penilaian kinerja kelas positif dengan menghitung rata-rata harmonik dari keduanya. Namun, karena dua ukuran pertama hanya berkaitan dengan kinerja kelas positif. Namun, menerapkan versi berbobot yang memperhitungkan kelas lengkap dan distribusinya menyelesaikan masalah ini. Persamaan berikut berlaku untuk metrik:

$$F1\ Score = 2 \times \frac{Presisi \times Recall}{Presisi + Recall} = 2 \times \frac{2TP}{2TP+FP+FN} \quad (6)$$

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Langkah pertama dalam penelitian ini adalah mengumpulkan data yang diperlukan. Data yang digunakan untuk penelitian ini berjumlah 10.323 data yang diambil secara manual dari komentar pelanggan pada masing-masing toko di Shopee. Tampilan data awal yang akan digunakan untuk penelitian ini disajikan pada Tabel 1.

Tabel 1. Tampilan Data Awal

| No | ID Barang | Date | Comment | Rating |
|-------|------------------------|-----------------------------|--|-----------|
| 1 | Geofmax 20 | 14/01/2020 14:27 | Packing ajieb pengiriman super duper on time good .next klo udah jebol bakal balik ksini lagi | 5 |
| 2 | Geofmax 20 | 16/01/2020 16:56 | Bagusssssssss banggettttttttttttttttttttttttttttttt | 5 |
| 10322 | Sepatu Puma | 08/03/2024 20:23 | Terbaik, fast respon | 5 |
| 10323 | Sepatu Puma | 04/04/2024 14:44 | Aip mantap keren euy | 5 |

3.2. Cleaning Data

Pada fase ini, proses case folding berfungsi untuk mengubah semua huruf kapital dalam kalimat menjadi huruf kecil. Ini dilakukan untuk memastikan pemrosesan data yang konsisten agar mengurangi duplikasi dan redundansi kata.

Pada tahap ini, karakter dan tanda baca yang tidak berguna dihapus dari kata-kata. Tabel 2 menunjukkan hasil dari proses cleaning data ini.

Tabel 2. Cleansing Data

| No | Comment | Cleaning Text |
|------|---|--|
| 1 | Packing ajieb pengiriman super duper on time good .next klo udah jebol bakal balik ksini lagi | packing ajieb pengiriman super duper on time good next klo udah jebol bakal balik ksini lagi |
| 2 | Bagusssssssss banggettttttttttttttttttttttttttttttt | bagus banget |
| 9483 | Terbaik, fast respon | terbaik fast respon |
| 9484 | Aip mantap keren euy | aip mantap keren euy |

3.3. Tokenizing Data

Tokenize data yaitu proses membagi kalimat menjadi kata perkata. Hal ini juga dapat membantu dalam mengidentifikasi kata-kata kunci, menghitung frekuensi kata yang muncul, dan memberikan wawasan tentang sentiment yang terkandung dalam kalimat. Hasil dari proses tokenisasi dapat dilihat pada Tabel 3.

Tabel 3. Tokenizing Data

| No | Cleaning Text | Tokenizing Text |
|------|--|---|
| 1 | packing ajieb pengiriman super duper on time good next klo udah jebol bakal balik ksini lagi | ['packing', 'ajieb', 'pengiriman', 'super', 'duper', 'on', 'time', 'good', 'next', 'klo', 'udah', 'jebol', 'bakal', 'balik', 'ksini', 'lagi'] |
| 2 | bagus banget | [bagus, banget] |
| 9483 | terbaik fast respon | [terbaik, fast, respon] |
| 9484 | aip mantap keren euy | [aip, mantap, keren, euy] |

3.4. Normalisasi Data

Normalisasi data adalah untuk memastikan bahwa setiap elemen data memberikan kontribusi yang seimbang untuk analisis atau pembuatan model, tanpa satu elemen yang mendominasi yang lain. Hasil dari tahap normalisasi data dapat dilihat pada Tabel 4.

Tabel 4. Normalisasi Data

| No | Tokenizing Text | Normalisasi Text |
|------|---|---|
| 1 | ['packing', 'ajieb', 'pengiriman', 'super', 'duper', 'on', 'time', 'good', 'next', 'klo', 'udah', 'jebol', 'bakal', 'balik', 'ksini', 'lagi'] | ['packing', 'ajieb', 'pengiriman', 'super', 'duper', 'on', 'time', 'good', 'next', 'kalo', 'udah', 'jebol', 'bakal', 'balik', 'kesini', 'lagi'] |
| 2 | [bagus], [banget] | [bagus, banget] |
| | | |
| 9483 | [terbaik, fast, respon] | [terbaik, fast, respon] |
| 9484 | [aip, mantap, keren, euy] | [aip, mantap, keren, euy] |

3.5. Stopwords Removal Data

Stopwords removal dilakukan untuk menghilangkan kata hubung seperti "dan", "atau", "yang", dan sebagainya yang dianggap tidak memberikan kontribusi yang signifikan terhadap makna kalimat sentimen. Tabel 5 menunjukkan hasil dari proses stopwords removal.

Tabel 5. Stopwords Removal Data

| No | Normalisasi Text | Stopwords Removal Text |
|------|---|--|
| 1 | ['packing', 'ajieb', 'pengiriman', 'super', 'duper', 'on', 'time', 'good', 'next', 'klo', 'udah', 'jebol', 'bakal', 'balik', 'ksini', 'lagi'] | ['packing', 'ajieb', 'pengiriman', 'super', 'duper', 'on', 'time', 'good', 'next', 'udah', 'jebol', 'ksini'] |
| 2 | [bagus], [banget] | [bagus, banget] |
| | | |
| 9483 | [terbaik, fast, respon] | [terbaik, fast, respon] |
| 9484 | [aip, mantap, keren, euy] | [aip, mantap, keren, euy] |

3.6. Stemming Data

Stemming mengubah setiap kata ke bentuk dasarnya. Ini memungkinkan pemahaman yang lebih baik tentang variasi kata yang memiliki akar kata yang sama. Dengan demikian, meskipun kata-kata muncul dalam bentuk yang berbeda dalam sebuah kalimat, analisis sentimen dapat dilakukan dengan lebih akurat. Tabel 6 menunjukkan hasil pekerjaan yang telah dilakukan pada tahap stemming.

Tabel 6. Stemming Data

| No | Stopwords Removal Text | Stemming Text |
|------|--|---|
| 1 | ['packing', 'ajieb', 'pengiriman', 'super', 'duper', 'on', 'time', 'good', 'next', 'udah', 'jebol', 'ksini'] | ['packing', 'ajieb', 'kirim', 'super', 'duper', 'on', 'time', 'good', 'next', 'udah', 'jebol', 'ksini'] |
| 2 | [bagus], [banget] | [bagus, banget] |
| | | |
| 9483 | [terbaik, fast, respon] | [baik, fast, respon] |

3.7. Labeling Data

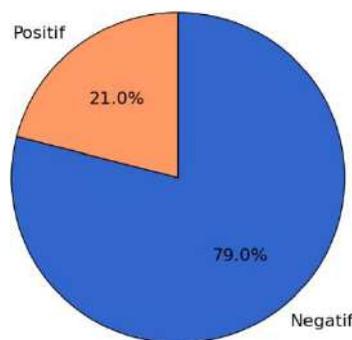
Setelah tahap preprocessing selesai, setiap data dilabeli. Pelabelan ini dilakukan untuk mengetahui nilai yang terkandung dalam setiap kalimat. Ini dilakukan dengan menggunakan metode analisis berbasis lexicon, yang memberikan nilai positif dan negatif untuk setiap kata yang ada dalam kalimat. Tabel 7 menampilkan hasil pelabelan tersebut.

Tabel 7. Labeling Data

| No | Stemming Text | Skor Sentimen | Sentimen |
|------|---|---------------|----------|
| 1 | ['packing', 'ajieb', 'kirim', 'super', 'duper', 'on', 'time', 'good', 'next', 'udah', 'jebol', 'ksini'] | 3.0 | Positif |
| | | | |
| 9483 | ['baik', 'fast', 'respon'] | 4.0 | Positif |
| 9484 | ['aip', 'mantap', 'keren', 'euy'] | 5.0 | Positif |

3.8. Perhitungan Nilai Positif dan Negatif

Setelah mendapatkan data clean pada tahapan sebelumnya, langkah berikutnya adalah menghitung jumlah komentar positif dan negatif pada data tersebut. Perhitungan ini dapat dilakukan dengan menggunakan persamaan counter pada masing-masing data, dan hasilnya adalah 5.991 komentar positif dan 1.596 komentar negatif dengan total data setelah pelabelan 7.587. Gambar 2 memberikan hasil analisis sentimen terhadap produk fashion sepatu menunjukkan bahwa dari total review pelanggan yang dianalisis, 79% menunjukkan sentimen positif dan 21% menunjukkan sentimen negatif.



Gambar 2. Persentase Perhitungan Sentimen Analisis

Persentase yang signifikan dari sentimen positif menunjukkan bahwa mayoritas konsumen memiliki pandangan yang baik terhadap produk tersebut yang dapat mencakup aspek seperti kualitas, kenyamanan, desain, dan harga. Sebaliknya, 21% dari respons yang menunjukkan sentimen negatif perlu mendapatkan perhatian yang lebih besar. Temuan ini menunjukkan bahwa produk secara keseluruhan diterima dengan baik di pasaran, tetapi ada beberapa area yang perlu diperbaiki untuk meningkatkan kepuasan konsumen.

Perhitungan TF-IDF menggabungkan dua ide untuk menghitung bobot frekuensi kata dalam dokumen tertentu dan frekuensi kebalikan dokumen yang mengandung kata. Hasil perhitungan TF-IDF dengan menggunakan Persamaan 2 yang ditunjukkan pada Tabel 8.

Tabel 8. Perhitungan TF-IDF

| Kata | TF IDF | TF |
|---------|----------|----|
| remake | 0,326657 | 0 |
| details | 0,295031 | 0 |
| price | 0,288739 | 0 |
| ikat | 0,000000 | 0 |
| gokill | 0,000000 | 0 |
| inceran | 0,000000 | 0 |

Dalam penelitian ini, dokumen remake, details, dan price merupakan kata-kata dengan nilai TF-IDF tinggi yang merupakan tema utama dokumen. Dalam ulasan produk, kata "price" dan "details" menunjukkan bahwa harga dan rincian produk adalah komponen penting yang mempengaruhi sentimen ulasan. Remake juga bisa merujuk pada versi terbaru atau perbaikan produk yang dibahas dalam ulasan.

3.11. Klasifikasi Menggunakan Support Vector Machine

Pada tahap ini, pengklasifikasian data dilakukan menggunakan Algoritma SVM dengan dataset 9.484. Data akan dibagi menjadi data latih dan data uji dengan perbandingan 80:20, dimana rasio yang lebih umum digunakan akan menghasilkan hasil yang mewakili data asli yang memungkinkan pengaturan parameter model yang lebih baik. Dengan menggunakan nilai negatif dan positif, model SVM akan mengkategorikan teks ke dalam kategori sentimen yang relevan. Tabel 9 menunjukkan hasil pembagian data pelatihan dan pengujian menggunakan Algoritma SVM.

Tabel 9. Hasil Prediksi Data Latih Menggunakan Algoritma SVM

| Metode Evaluasi | Nilai |
|-----------------|----------|
| Akurasi | 0,997199 |
| Presisi | 0,997209 |
| Recall | 0,997199 |
| F1-Score | 0,997192 |

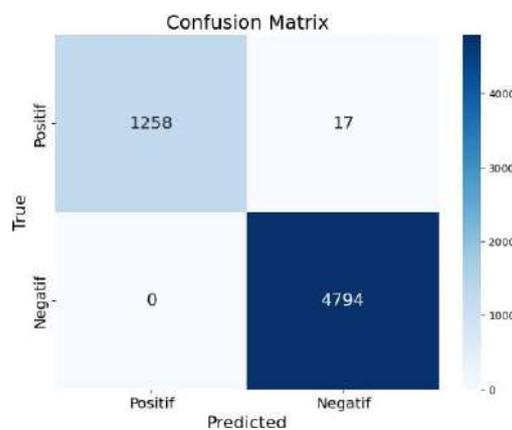
Berdasarkan Tabel 9 menunjukkan bahwa evaluasi model SVM dengan dataset 7.587 digunakan sebagai data latih, menunjukkan bahwa model ini mempunyai performa lebih baik dengan skor F1 99,72 sehingga secara keseluruhan, model ini membantu meningkatkan efisiensi operasional dan mendukung keputusan yang lebih akurat berdasarkan data.

Tabel 10. Hasil Prediksi Data Uji Menggunakan Algoritma SVM

| Metode Evaluasi | Nilai |
|-----------------|----------|
| Akurasi | 0,926219 |
| Presisi | 0,926682 |
| Recall | 0,926219 |
| F1-Score | 0,922160 |

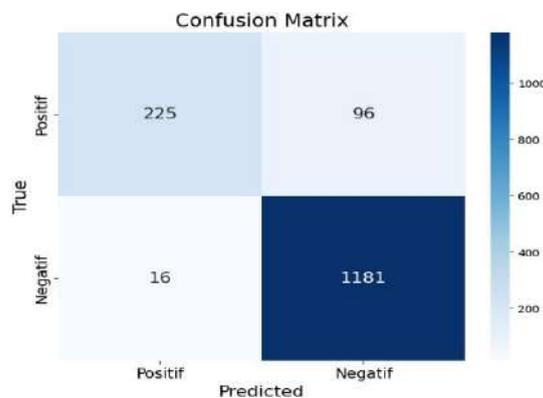
Pada Tabel 10, model SVM menunjukkan kinerja yang baik pada data tes dengan skor akurasi, presisi, recall, dan F1-nya yang semuanya di atas 92%, meskipun sedikit lebih rendah dibandingkan data latih. Dengan akurasi sebesar 92,62%, model mampu memprediksi sekitar 92,62% dari total data tes dengan benar, menunjukkan keandalannya dalam memproses data baru. Hasil yang lebih dari 92% akurasi dengan jumlah dataset sebanyak 1.897 ini menunjukkan bahwa model dapat diandalkan untuk membuat keputusan strategis seperti pengelolaan risiko, deteksi penipuan, atau klasifikasi pelanggan.

Pengklasifikasian menggunakan Algoritma SVM ini menghasilkan Confusion Matrix. Gambar 5 dan Gambar 6 menunjukkan hasil penggunaan Algoritma SVM pada data latih dan data uji.



Gambar 5. Confusion Matrix SVM Pada Data Latih

Berdasarkan Gambar 5, Confusion Matrix dengan nilai True Positive (1258), False Positive (17), False Negative (0), dan True Negative (4794), model menunjukkan performa yang baik. Model ini seimbang dalam kemampuan mendeteksi kasus positif, menjadikannya cukup andal. Dengan demikian, model ini memberikan informasi yang akurat dan seimbang mengenai respons pelanggan, sehingga dapat digunakan untuk mengidentifikasi peluang perbaikan produk, menyusun strategi pemasaran, serta meningkatkan kepuasan pelanggan.



Gambar 6. Confusion Matrix SVM Pada Data Uji

Gambar 6 menunjukkan tingkat keberhasilan yang tinggi ditunjukkan oleh model yang efektif dalam menyesuaikan layanan atau produk dengan kebutuhan pelanggan. Hal ini didukung oleh sebagian besar interaksi atau transaksi dengan pelanggan berjalan dengan baik dan memenuhi ekspektasinya.

Pelanggan memiliki pengalaman yang baik, yang membuat mereka puas dan percaya pada produk sepatu lokal. Pengalaman positif ini mendorong pelanggan untuk menjadi setia dan kembali membeli produk ini.

4. DISKUSI

Pada bagian ini membahas seberapa baik SVM melakukan pemodelan analisis sentimen pelanggan terhadap fashion sepatu. Perbandingan dilakukan terhadap dua metode lainnya yaitu K-NN dan Random Forest. Evaluasi kinerja diukur menggunakan empat metrik : Akurasi, Presisi, Recall, dan F1-Score. Tabel berikut menjabarkan perbandingan hasil prediksi dari masing-masing metode.

Tabel 11. Perbandingan Hasil Prediksi Data Uji Menggunakan SVM, K-NN dan RF

| Metode Evaluasi | SVM | K-NN | RF |
|-----------------|----------|---------|---------|
| Akurasi | 0,926219 | 0.90656 | 0.82902 |
| Presisi | 0.926682 | 0.90551 | 0.81543 |
| <i>Recall</i> | 0.926219 | 0.90656 | 0.82902 |
| <i>F1-Score</i> | 0.922160 | 0.90095 | 0.81718 |

Pada tabel 11, dapat dilihat SVM memberikan kinerja terbaik pada data uji, selanjutnya diikuti oleh K-NN, sementara Random Forest (RF) menunjukkan penurunan kinerja yang cukup signifikan. Hal ini menunjukkan bahwa SVM lebih baik dalam menggeneralisasi pada data baru, dibandingkan K-NN dan RF yang berada dibawah SVM.

Hasil analisis sentimen menunjukkan dominasi opini positif dengan akurasi metode SVM sebesar 92,62% menunjukkan algoritma ini lebih unggul dibandingkan dengan K-NN dan RF dalam analisis sentimen produk fashion sepatu. Hasil ini menegaskan bahwa penggunaan SVM efektif dalam meningkatkan akurasi analisis sentimen dan mendukung pengambilan keputusan yang lebih baik.

Analisis sentimen yang akurat juga memperkuat hubungan dengan pelanggan memiliki dampak signifikan terhadap strategi bisnis, teknologi informasi dan hubungan antar pelanggan. Dalam konteks strategi bisnis, dominasi sentimen positif ini menunjukkan bahwa produk atau layanan yang ditawarkan oleh perusahaan berhasil memenuhi ekspektasi pelanggan, yang merupakan indikator keberhasilan dalam strategi pemasaran dan pengelolaan merek.

5. KESIMPULAN

Penelitian ini memberikan hasil analisis bahwa opini pelanggan terhadap produk fashion sepatu cenderung positif. Hal ini menunjukkan bahwa toko merek sepatu di platform e-commerce berhasil memenuhi ekspektasi pelanggan, memperkuat citra merek, serta meningkatkan loyalitas dan penjualan. Penggunaan analisis sentimen otomatis membantu penjual dalam mengelola ulasan secara efektif, dan memungkinkan respons cepat terhadap umpan balik pelanggan. Peningkatan interaksi dengan pelanggan dan pemanfaatan ulasan positif sebagai alat promosi lebih lanjut memperkuat hubungan dengan pelanggan dan menarik pelanggan baru, membantu toko mencapai keunggulan kompetitif di pasar.

Hasil keseluruhan dari analisis sentimen ini dapat membantu konsumen lain dalam menilai kualitas produk fashion sepatu di e-commerce Indonesia. Hal ini dikarenakan Algoritma SVM dalam penelitian ini menunjukkan performa yang baik dengan tingkat akurasi sebesar 92,62%. Hasil analisis sentimen meningkatkan pengambilan keputusan strategis, meningkatkan kepercayaan pelanggan, dan memberikan dasar untuk pengembangan penelitian sistem informasi yang lebih canggih di masa depan. Dengan demikian, perusahaan memiliki landasan yang kuat untuk meningkatkan produk, layanan, dan hubungan pelanggan dengan menggunakan Algoritma SVM.

UCAPAN TERIMA KASIH

Peneliti mengucapkan terimakasih yang sebesar-besarnya kepada semua pihak yang telah memberikan dukungan, semoga hasil penelitian ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan, khususnya dalam bidang analisis sentimen produk fashion.

DAFTAR PUSTAKA

- [1] L. Yang, Y. Li, J. Wang, and R. S. Sherratt, "Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning," *IEEE Access*, vol. 8, pp. 23522–23530, 2020, doi: 10.1109/ACCESS.2020.2969854.
- [2] J. lee and Nguyen, "ASEAN Digital Generation Report: Digital Financial Inclusion," *Word Economic Forum*, Dec. 2022. [Online]. Available: <https://www.sea.com/sustainability>
- [3] N. P. Rahmayanti, "Pengaruh Marketplace dan Pembayaran Digital Terhadap Tingkat Penjualan UMKM Di Kota Banjarmasin," *Al-KALAM J. Komun. BISNIS DAN Manaj.*, vol. 10, no. 1, p. 28, Jan. 2023, doi: 10.31602/al-kalam.v10i1.9095.
- [4] L. Gunawan, M. S. Anggreainy, L. Wihan, Santy, G. Y. Lesmana, and S. Yusuf, "Support vector machine based emotional analysis of restaurant reviews," *Procedia Comput. Sci.*, vol. 216, pp. 479–484, 2023, doi: 10.1016/j.procs.2022.12.160.
- [5] W. Agustiono and A. K. Darmawan, "E-Marketplace Selection for Indonesian MSMEs: A Sentiment Analysis Approach Combining Lexicon-Based and Support Vector Machine," in *2023 IEEE 9th Information Technology International Seminar (ITIS)*, Batu Malang, Indonesia: IEEE, Oct. 2023, pp. 1–5. doi: 10.1109/ITIS59651.2023.10420382.
- [6] F. A. Nugraha, N. H. Harani, and R. Habibi, "Analisis Sentimen Terhadap Pembatasan Sosial Menggunakan Deep Learning. Kreatif Industri Nusantara.," in 1. , *Kreatif Industri Nusantara*, 2020.
- [7] I. Hidayah, A. E. Permanasari, and N. Woro Wijayanti, "Sentiment Analysis on Product Review using Support Vector Machine (SVM)," in *2019 5th International Conference on Science and Technology (ICST)*, Yogyakarta, Indonesia: IEEE, Jul. 2019, pp. 1–4. doi: 10.1109/ICST47872.2019.9166189.
- [8] A. D. Riyanto, A. M. Wahid, and A. A. Pratiwi, "ANALYSIS OF FACTORS DETERMINING STUDENT SATISFACTION USING DECISION TREE, RANDOM FOREST, SVM, AND NEURAL NETWORKS: A COMPARATIVE STUDY".
- [9] A. N. Syafia, M. F. Hidayattullah, and W. Sutеды, "Studi Komparasi Algoritma SVM Dan Random Forest Pada Analisis Sentimen Komentar Youtube BTS," *J. Inform. J. Pengemb. IT*, vol. 8, no. 3, pp. 207–212, Sep. 2023, doi: 10.30591/jpit.v8i3.5064.
- [10] Ratih Puspitasari, Y. Findawati, and M. A. Rosid, "SENTIMENT ANALYSIS OF POST-COVID-19 INFLATION BASED ON TWITTER USING THE K-NEAREST NEIGHBOR AND SUPPORT VECTOR MACHINE CLASSIFICATION METHODS," *J. Tek. Inform. Jutif*, vol. 4, no. 4, pp. 669–679, Aug. 2023, doi: 10.52436/1.jutif.2023.4.4.801.
- [11] I. Yunanto and S. Yulianto, "TWITTER SENTIMENT ANALYSIS PEDULILINDUNGI APPLICATION USING NAÏVE BAYES AND SUPPORT VECTOR MACHINE," *J. Tek. Inform. Jutif*, vol. 3, no. 4, pp. 807–814, Aug. 2022, doi: 10.20884/1.jutif.2022.3.4.292.
- [12] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, "Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM)," *Jambura J. Electr. Electron. Eng.*, vol. 5, no. 1, pp. 32–35, Jan. 2023, doi: 10.37905/jjee.v5i1.16830.
- [13] S. Dey, S. Wasif, D. S. Tonmoy, S. Sultana, J. Sarkar, and M. Dey, "A Comparative Study of Support Vector Machine and Naive Bayes Classifier for Sentiment Analysis on Amazon Product Reviews," in *2020 International Conference on Contemporary Computing and Applications (IC3A)*, Lucknow, India: IEEE, Feb. 2020, pp. 217–220. doi: 10.1109/IC3A48958.2020.233300.
- [14] S. Watmah, S. Suryanto, and M. Martias, "Komparasi Metode K-NN, Support Vector Machine Dan Random Forest Pada E-Commerce Shopee," *INSANtek*, vol. 2, no. 1, pp. 15–21, Jun. 2021, doi: 10.31294/instk.v2i1.419.
- [15] H. Syahputra, "Sentiment Analysis of Community Opinion on Online Store in Indonesia on Twitter using Support Vector Machine Algorithm (SVM)," *J. Phys. Conf. Ser.*, vol. 1819, no. 1, p. 012030, Mar. 2021, doi: 10.1088/1742-6596/1819/1/012030.
- [16] H. Syah and A. Witanti, "ANALISIS SENTIMEN MASYARAKAT TERHADAP VAKSINASI COVID-19 PADA MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)," *J. Sist. Inf. Dan Inform. Simika*, vol. 5, no. 1, pp. 59–67, Feb. 2022, doi: 10.47080/simika.v5i1.1411.

-
- [17] H. G. Sulistio and A. Handoyo, "Aspect-Based Sentiment Analysis pada Ulasan E- Commerce dengan Metode Support Vector Machine untuk Mendapatkan Informasi Sentimen dari Beberapa Aspek," vol. 10, no. 2, 2022.
- [18] K. Mehmood, D. Essam, K. Shafi, and M. K. Malik, "An unsupervised lexical normalization for Roman Hindi and Urdu sentiment analysis," *Inf. Process. Manag.*, vol. 57, no. 6, p. 102368, Nov. 2020, doi: 10.1016/j.ipm.2020.102368.
- [19] A. J. Putri, A. S. Syafira, M. E. Purbaya, and D. Purnomo, "Analisis Sentimen E-Commerce Lazada pada Jejaring Sosial Twitter Menggunakan Algoritma Support Vector Machine," *J. TRINISTIK J. Tek. Ind. Bisnis Digit. Dan Tek. Logist.*, vol. 1, no. 1, pp. 16–21, Mar. 2022, doi: 10.20895/trinistik.v1i1.447.
- [20] R. Ulgasesa, A. B. P. Negara, and T. Tursina, "Pengaruh Stemming Terhadap Performa Klasifikasi Sentimen Masyarakat Tentang Kebijakan New Normal," *J. Sist. Dan Teknol. Inf. JustIN*, vol. 10, no. 3, p. 286, Sep. 2022, doi: 10.26418/justin.v10i3.53880.
- [21] H. T. Ismet, T. Mustaqim, and D. Purwitasari, "Aspect Based Sentiment Analysis of Product Review Using Memory Network," *Sci. J. Inform.*, vol. 9, no. 1, pp. 73–83, May 2022, doi: 10.15294/sji.v9i1.34094.
- [22] D. Musfiroh, U. Khaira, P. E. P. Utomo, and T. Suratno, "Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon: Sentiment Analysis of Online Lectures in Indonesia from Twitter Dataset Using InSet Lexicon," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 24–33, Mar. 2021, doi: 10.57152/malcom.v1i1.20.
- [23] W. A. Luqyana, I. Cholissodin, and R. S. Perdana, "Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine," vol. 2, pp. 4704–4713, Nov. 2018.
- [24] A. Deolika, K. Kusriani, and E. T. Luthfi, "ANALISIS PEMBOBOTAN KATA PADA KLASIFIKASI TEXT MINING," *J. Teknol. Inf.*, vol. 3, no. 2, p. 179, Dec. 2019, doi: 10.36294/jurti.v3i2.1077.
- [25] H. Nuraliza, O. N. Pratiwi, and F. Hamami, "Analisis Sentimen IMDb Film Review Dataset Menggunakan Support Vector Machine (SVM) dan Seleksi Feature Importance," no. 1, 2022.

