



Analisis Sentimen E-Commerce Shopee Menggunakan Metode Support Vector Machine (SVM) untuk Mengevaluasi Kualitas Layanan Penjual

Aidil Ramadani¹, Nur Azizah^{2*}, Rahmat Shofan Razaqi³

^{1,2,3} Program Studi Pendidikan Teknologi Informasi, STKIP PGRI Situbondo, Situbondo, Indonesia

¹aidilramadani02@gmail.com, ^{2*}azizah0606@gmail.com, ³fanslaught@gmail.com

^{*}nazizah0606@gmail.com

Abstrak-Pertumbuhan pesat e-commerce menghasilkan volume besar ulasan pelanggan Shopee, sehingga evaluasi manual menjadi tidak efisien dan tidak konsisten. Penelitian ini menjawab kebutuhan akan klasifikasi teks otomatis untuk mengidentifikasi polaritas ulasan dan mendukung evaluasi layanan penjual. Sebuah dataset berlabel yang terdiri dari 800 ulasan Shopee (600 positif, 200 negatif) dikumpulkan dari Kaggle. Teks-teks tersebut diproses terlebih dahulu menggunakan case folding, tokenisasi, penghapusan stopword, dan stemming untuk mengurangi noise dan menstandarkan penulisan. Ulasan dikonversi menjadi vektor fitur TF-IDF (max_features=5.000; rentang n-gram 1-2) dan diklasifikasikan menggunakan Support Vector Machine (SVM) dengan kernel RBF dan penyeimbangan bobot kelas. Data dibagi dengan sampling terstratifikasi (80:20) untuk mempertahankan proporsi label pada set pelatihan dan pengujian. Kinerja model diukur menggunakan akurasi, presisi, recall, skor F1, dan matriks kebingungan. Klasifikasi mencapai akurasi 83,75% dan menghasilkan hasil yang lebih baik untuk ulasan positif, sementara ulasan negatif lebih sering salah diklasifikasikan akibat ketidakseimbangan kelas dan variasi ekspresi keluhan. Sebagai dokumentasi pendukung, sebagian ulasan disiapkan dalam bentuk tangkapan layar dan diberi anotasi di Roboflow, sementara evaluasi sentimen utama tetap didasarkan pada ekstraksi fitur TF-IDF dan klasifikasi SVM. Secara keseluruhan, kombinasi TF-IDF dengan SVM memberikan pendekatan praktis untuk memantau sentimen pelanggan dan memprioritaskan perbaikan layanan penjual seiring waktu.

Kata Kunci: Analisis sentimen; Shopee, Tf-idf, Support vector machine, Klasifikasi text

Abstract-The rapid growth of e-commerce generates a large volume of Shopee customer reviews, making manual evaluation inefficient and inconsistent. This study addresses the need for automated text classification to identify review polarity and support seller service evaluation. A labeled dataset of 800 Shopee reviews (600 positive, 200 negative) was collected from Kaggle. The texts were preprocessed using case folding, tokenization, stopword removal, and stemming to reduce noise and standardize writing. Reviews were converted into TF-IDF feature vectors (max_features=5,000; n-gram range 1-2) and classified using a Support Vector Machine (SVM) with an RBF kernel and class-weight balancing. The data were split with stratified sampling (80:20) to preserve label proportions in training and testing sets. Model performance was measured using accuracy, precision, recall, F1-score, and a confusion matrix. The classifier achieved 83.75% accuracy and produced stronger results for positive reviews, while negative reviews were more frequently misclassified due to class imbalance and diverse complaint expressions. As supporting documentation, a subset of reviews was prepared as screenshots and annotated in Roboflow, while the main sentiment evaluation remained based on TF-IDF feature extraction and SVM classification. Overall, TF-IDF combined with SVM provides a practical approach for monitoring customer sentiment and prioritizing seller service improvements over time.

Keywords: Sentiment analysis, Shopee, Tf-idf, Support vector machine, Text classification

1. PENDAHULUAN

Di era digital saat ini, analisis perilaku konsumen dengan menggunakan teknik data mining dan machine learning telah menjadi sangat penting dalam mengidentifikasi preferensi dan tren pembelian, terutama untuk produk halal dan platform e-commerce seperti Shopee. Hasil yang diharapkan dari penelitian ini adalah mengidentifikasi pola preferensi konsumen serta mengklasifikasikan konsumen yang membeli produk halal berdasarkan pola-pola tersebut [1].

Perkembangan pesat teknologi digital, ditambah dengan semakin meluasnya akses internet, telah mendorong perubahan besar dalam kebiasaan berbelanja masyarakat. Transaksi yang sebelumnya dilakukan secara



langsung kini sebagian besar beralih ke e-commerce karena lebih praktis, menghemat waktu, dan menawarkan pilihan produk serta penjual yang lebih beragam [2]. E-commerce menawarkan berbagai keuntungan, terutama dalam mendukung transaksi jual beli yang cepat dan tanpa batas [3].

Shopee adalah platform e-commerce multinasional dengan basis pengguna yang besar di berbagai negara. Hal ini menyebabkan peningkatan jumlah ulasan pelanggan yang diposting setiap hari. Ulasan-ulasan tersebut mencatat pengalaman pengguna, termasuk penilaian positif maupun negatif terhadap produk dan layanan penjual. Dari sudut pandang bisnis, penguatan saluran penjualan di pasar daring juga berkaitan dengan kemudahan akses terhadap layanan dan informasi, yang dapat meningkatkan kepuasan pengguna [4].

Ulasan pelanggan memiliki nilai strategis karena sering kali berisi komentar mengenai kualitas produk, harga, kecepatan pengiriman, dan layanan penjual. Informasi semacam ini dapat digunakan untuk mengevaluasi kinerja dan meningkatkan layanan. Namun, banyaknya ulasan membuat penilaian manual sulit dilakukan secara efisien dan konsisten [2]. Oleh karena itu, analisis sentimen berbasis NLP diperlukan agar pendapat pelanggan dapat diproses secara otomatis dan diukur dalam skala besar [5].

Analisis sentimen merupakan bagian dari penambangan teks yang bertujuan untuk mengenali dan mengklasifikasikan polaritas opini—seperti positif dan negatif—berdasarkan teks ulasan. Dalam konteks e-commerce (termasuk Shopee), pendekatan ini membantu merangkul persepsi pelanggan secara terstruktur melalui pemrosesan data ulasan dalam jumlah besar [6]. Dalam analisis sentimen berbasis teks, pemilihan algoritma klasifikasi sangat menentukan kualitas prediksi. Support Vector Machine (SVM) sering digunakan untuk klasifikasi teks karena algoritma ini mampu menangani fitur berdimensi tinggi secara efektif dan membentuk batas pemisahan yang optimal antar kelas. Penelitian sebelumnya juga menunjukkan bahwa SVM cenderung stabil dalam mengklasifikasikan sentimen ulasan e-commerce berbahasa Indonesia, terutama bila dikombinasikan dengan ekstraksi fitur TF-IDF [7].

Sejumlah penelitian terdahulu telah mengkaji penerapan analisis sentimen pada ulasan e-commerce dengan memanfaatkan metode Support Vector Machine (SVM). Hasil penelitian menunjukkan bahwa SVM memiliki kemampuan yang baik dalam mengelompokkan opini pengguna ke dalam kategori sentimen positif maupun negatif karena mampu menangani data teks yang memiliki dimensi fitur tinggi. Selain itu, penggunaan pembobotan TF-IDF sebagai teknik ekstraksi fitur juga terbukti dapat meningkatkan representasi informasi penting dalam teks sehingga membantu proses klasifikasi menjadi lebih akurat. Berbagai temuan tersebut menunjukkan bahwa kombinasi TF-IDF dan SVM merupakan pendekatan yang efektif untuk analisis sentimen pada ulasan pelanggan di platform e-commerce.

Meskipun demikian, penelitian sebelumnya umumnya berfokus pada pengukuran performa model klasifikasi sentimen tanpa mengaitkannya secara spesifik dengan evaluasi kualitas layanan penjual. Oleh karena itu, penelitian ini mengimplementasikan metode TF-IDF dan SVM untuk menganalisis sentimen ulasan pelanggan Shopee dengan tujuan memperoleh gambaran mengenai kualitas layanan penjual berdasarkan opini yang disampaikan oleh pengguna.

Selain tersedia dalam bentuk teks terstruktur, ulasan dalam kondisi nyata juga dapat muncul dalam bentuk tangkapan layar. Untuk mengatasi hal ini, beberapa ulasan disiapkan sebagai gambar pendukung dan diberi anotasi di Roboflow guna memudahkan pengorganisasian data ulasan visual. Anotasi tersebut membantu mengidentifikasi area yang relevan dalam gambar dan memperbaiki struktur dokumentasi pendukung [8]. Dalam konteks ini, kotak pembatas berguna untuk menandai area teks sehingga data visual dapat dikelola dengan lebih jelas dan sistematis [9].

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk menganalisis sentimen ulasan pelanggan Shopee menggunakan metode Support Vector Machine (SVM) dengan representasi fitur TF-IDF, mengklasifikasikan ulasan ke dalam sentimen positif dan negatif, serta mengevaluasi kinerja model menggunakan metrik accuracy, precision, recall, F1-score, dan confusion matrix untuk memperoleh gambaran mengenai kualitas layanan penjual berdasarkan ulasan pelanggan.

2. METODOLOGI PENELITIAN

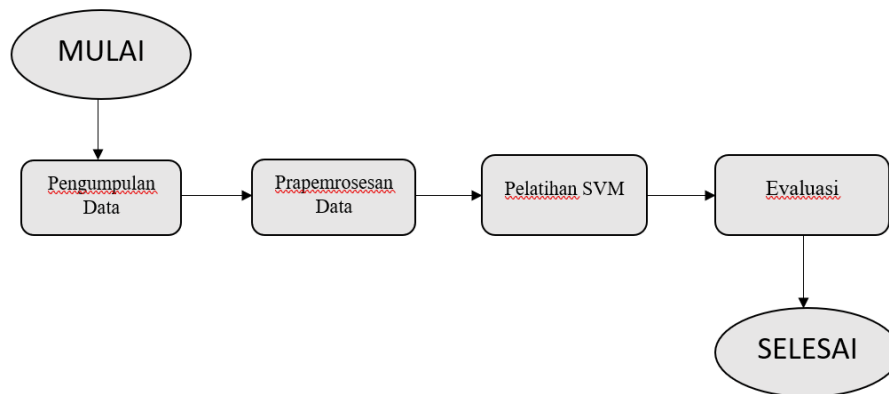
Penelitian ini menggunakan metode analisis sentimen dengan algoritma Support Vector Machine (SVM) untuk mengklasifikasikan ulasan pelanggan Shopee ke dalam kategori sentimen positif dan negatif. Proses penelitian meliputi pengumpulan data, prapemrosesan data, pelatihan SVM, serta evaluasi model menggunakan accuracy, precision, recall, F1-score, dan confusion matrix.

2.1 Diagram Penelitian

Proses penelitian dimulai dengan pengumpulan data melalui pengunduhan dataset CSV berisi ulasan produk Shopee dari Kaggle. Dataset ini berisi teks ulasan pelanggan yang telah diberi label berdasarkan sentimen. Langkah

selanjutnya adalah prapemrosesan untuk membersihkan dan menstandarkan teks, kemudian mengubahnya menjadi fitur numerik menggunakan Term Frequency-Inverse Document Frequency (TF-IDF). Fitur TF-IDF kemudian digunakan dalam melatih model klasifikasi. Pada tahap ini, SVM dilatih untuk membedakan antara sentimen positif dan negatif dengan membangun hiperplane pemisah yang optimal.

Langkah selanjutnya adalah evaluasi model, yang dilakukan dengan menggunakan data uji. Evaluasi dilakukan dengan menghitung metrik kinerja seperti akurasi, presisi, recall, dan F1-score, serta dianalisis lebih lanjut menggunakan confusion matriks. Tahap terakhir dari penelitian ini adalah menarik kesimpulan, yang menandai selesainya proses penelitian berdasarkan hasil evaluasi kinerja model klasifikasi sentimen.



Gambar 1. Diagram Penelitian.

2.2 Populasi

Populasi penelitian mencakup ulasan produk di Shopee yang mencerminkan pendapat pelanggan mengenai kualitas produk, harga, pengiriman, dan layanan penjual. Shopee dipilih karena memiliki jumlah ulasan yang tinggi dan kumpulan data yang memadai untuk eksperimen analisis sentimen.

2.3 Sampel

Sampel diambil dengan menggunakan metode pengambilan sampel purposif berdasarkan kriteria berikut:

1. Sampel dalam penelitian ini terdiri dari 800 ulasan, dengan 600 ulasan positif dan 200 ulasan negatif.
2. Ulasan-ulasan tersebut berbentuk teks dan layak untuk diproses (tidak kosong/tidak terdapat duplikat yang berlebihan).
3. Beberapa ulasan dapat ditampilkan dalam bentuk tangkapan layar dengan kualitas visual yang jelas untuk mendukung kebutuhan anotasi.

Pendekatan ini bertujuan untuk memastikan bahwa data yang digunakan relevan dan memadai untuk proses analisis sentimen.

2.4 Pengumpulan Data

Data penelitian dikumpulkan menggunakan data sekunder yang bersumber dari platform Kaggle. Kumpulan data yang digunakan terdiri dari ulasan produk e-commerce Shopee dalam format teks (CSV) yang berisi teks ulasan dan label sentimen. Prosedur pengumpulan data dilakukan melalui tahapan-tahapan berikut:

1. Pengunduhan Kumpulan Data: Kumpulan data ulasan produk Shopee diunduh dari Kaggle dalam format CSV. Data ini digunakan sebagai sumber utama dalam proses analisis sentimen berbasis teks dengan menggunakan metode Support Vector Machine (SVM).
2. Dokumentasi Visual (Tangkapan Layar): Sebagian data tinjauan dalam format teks diubah menjadi bentuk gambar melalui tangkapan layar. Tahap ini dimaksudkan untuk mendukung dokumentasi dan pemberian anotasi pada tampilan tinjauan menggunakan Roboflow.
3. Penyimpanan Data: Sebanyak 40 tangkapan layar telah disimpan, dengan setiap gambar berisi sekitar 20 ulasan. Data gambar tersebut hanya digunakan sebagai dokumentasi pendukung dan bahan anotasi, sedangkan data utama tetap berupa teks ulasan dalam format CSV untuk pelatihan dan evaluasi model SVM.

2.5 Prapemrosesan Data

Tahap prapemrosesan dilakukan untuk mempersiapkan data agar lebih terstruktur dan siap digunakan dalam proses pemodelan. Dalam penelitian ini, prapemrosesan mencakup persiapan data ulasan dalam format teks (CSV)

sebagai data utama untuk analisis sentimen, serta persiapan data pendukung berupa tangkapan layar ulasan untuk keperluan anotasi. Pada data teks, prapemrosesan bertujuan untuk mengurangi noise dan menstandarkan format teks sehingga fitur yang dihasilkan lebih representatif. Langkah-langkah umum meliputi case folding (mengubah huruf menjadi huruf kecil), tokenisasi (memecah kalimat menjadi token/kata), penghapusan stopwords (menghapus kata-kata umum yang kurang informatif), dan stemming (mengubah kata ke bentuk dasarnya) [10]. Penghapusan kata-kata pengisi diperlukan agar teks menjadi lebih representatif, karena analisis sentimen - yang penting untuk memahami opini publik - memerlukan pemrosesan awal teks dengan cara menghapus kata-kata pengisi yang tidak signifikan [11].

Selain prapemrosesan data teks, penelitian ini juga mencakup pelabelan/anotasi sebagai bagian dari prapemrosesan karena bertujuan untuk mempersiapkan data sebelum proses pelatihan model berbasis gambar (support). Anotasi dilakukan dengan menggunakan teknik bounding box untuk menandai area teks yang akan dianalisis dalam gambar dan menetapkan kelas positif dan negatif sesuai dengan label sentimen. Secara umum, anotasi kotak batas pada data gambar memang merupakan tugas yang memakan waktu dan memerlukan ketelitian, sehingga penggunaan pendekatan/alat anotasi membantu membuat proses pelabelan menjadi lebih efisien dan terstandarisasi [12]. Selain prapemrosesan teks, penelitian ini juga mencakup proses anotasi sebagai tahap persiapan pendukung untuk dokumentasi tinjauan berbasis tangkapan layar. Anotasi dilakukan di Roboflow untuk menandai area tinjauan pada gambar dan mengatur data visual pendukung. Sementara itu, proses analisis sentimen utama tetap menggunakan klasifikasi teks berbasis TF-IDF + SVM.

2.6 Pelatihan SVM

Selama tahap pelatihan SVM, data ulasan Shopee yang telah melalui proses prapemrosesan digunakan sebagai masukan utama untuk membangun model klasifikasi sentimen positif dan negatif berbasis teks. Tangkapan layar pendukung yang diberi anotasi di Roboflow hanya digunakan untuk keperluan dokumentasi dan tidak terlibat dalam proses pelatihan model utama.

1. Representasi Text (TF - IDF)

Agar SVM dapat memproses data teks, ulasan direpresentasikan sebagai vektor numerik menggunakan Term Frequency-Inverse Document Frequency (TF-IDF). Pemilihan algoritma SVM dalam penelitian ini didukung oleh literatur karena algoritma Support Vector Machine telah banyak diterapkan dan terbukti efektif untuk analisis sentimen dan tugas-tugas penambangan teks [13]. Salah satu pendekatan yang paling sering digunakan dalam seleksi fitur adalah Term Frequency/Inverse Document Frequency (TF/IDF) [14].

TF-IDF adalah skema pembobotan yang menonjolkan kata-kata informatif dalam sebuah dokumen dan mengurangi pengaruh kata-kata yang terlalu umum, sehingga efektif untuk mewakili teks dalam tugas klasifikasi [15].

Secara umum, bobot TF-IDF dapat ditulis sebagai:

$$TF - IDF(t, d) = TF(t, d) \times \log\left(\frac{N}{df(t)}\right)$$

Dimana:

1. $TF(t, d)$ menunjukkan frekuensi kemunculan istilah t dalam dokumen d ,
2. N adalah jumlah total dokumen dalam kumpulan data,
3. $df(t)$ menunjukkan jumlah dokumen yang mengandung istilah t .

Dalam penerapannya, bobot TF-IDF dihitung menggunakan Tf-idf Vectorizer dengan pengaturan `max_features=5000` untuk membatasi jumlah fitur dan `ngram_range=(1,2)` agar model dapat menangkap pola unigram dan bigram yang sering kali mengandung sinyal sentimen.

2. Pelatihan Model SVM

Setelah fitur TF-IDF terbentuk, model dilatih menggunakan SVC (Support Vector Classification). SVM membangun batas keputusan dengan margin maksimum sehingga cocok untuk data berdimensi tinggi seperti vektor TF-IDF. Kernel yang digunakan adalah RBF, sehingga model mampu membentuk batas keputusan non-linear ketika pemisahan kelas tidak sepenuhnya linear di ruang fitur. Karena distribusi label tidak seimbang (600 positif dan 200 negatif), proses pelatihan menggunakan `class_weight='balanced'` sehingga bobot kelas secara otomatis disesuaikan berdasarkan frekuensi kelas, guna memastikan bahwa pelatihan tidak terlalu condong ke kelas mayoritas. Validasi dilakukan dengan membagi data menjadi data pelatihan dan data uji menggunakan fungsi `'train_test_split'` dengan parameter `'test_size=0.2'` dan `'stratify=y'`. Penggunaan parameter `'stratify'` bertujuan

untuk menjaga proporsi kelas pada data uji agar konsisten dengan distribusi label dalam dataset, sehingga menghasilkan evaluasi yang lebih representatif.

Penelitian ini menggunakan kumpulan data yang terdiri dari 800 ulasan (600 ulasan positif dan 200 ulasan negatif). Dengan pembagian 80:20, diperoleh 640 data pelatihan dan 160 data uji, dengan distribusi data uji yang tetap, yaitu 120 ulasan positif dan 40 ulasan negatif. Hasil prediksi pada data uji kemudian dirangkum melalui laporan klasifikasi dan confusion matriks untuk menghitung metrik akurasi, presisi, recall, dan F1-score, karena metrik-metrik ini memberikan gambaran umum tentang kinerja per kelas dan lebih informatif daripada hanya mengandalkan Akurasi [16].

2.7 Evaluasi Model

Evaluasi model dilakukan untuk menilai kemampuan klasifikasi SVM dalam menggeneralisasi sentimen ulasan pada data uji yang tidak digunakan selama proses pelatihan. Dalam evaluasi klasifikasi teks, merupakan praktik umum untuk menggunakan kumpulan data uji yang disisihkan dan melaporkan metrik yang relevan agar kinerja sistem dapat dibandingkan secara objektif pada tugas yang sama [16]. Untuk menentukan algoritma mana yang lebih baik, digunakan beberapa ukuran penilaian, seperti akurasi, presisi, recall, F1-score, dan AUC [17].

Pengukuran kinerja dimulai dengan akurasi untuk melihat proporsi keseluruhan prediksi yang benar. Namun, dalam kondisi distribusi kelas yang tidak seimbang, akurasi bisa terlihat tinggi meskipun model tersebut lemah dalam kelas minoritas. Dalam kondisi seperti itu, akurasi tidak cukup dijadikan satu-satunya tolok ukur karena penggunaan metode ini dapat menimbulkan bias dan bahkan menghasilkan kesimpulan yang menyesatkan [18]. Oleh karena itu, evaluasi perlu dilengkapi dengan metrik yang didasarkan pada confusion matriks agar gambaran kinerja menjadi lebih representatif [19]. Metrik presisi, recall, dan F1-score dihitung berdasarkan komponen confusion matriks (TP, TN, FP, FN). Secara umum:

- $\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$
- $\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$
- $\text{F1-score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$

Perhitungan ini digunakan untuk menilai ketepatan prediksi dalam suatu kelas (presisi) dan kemampuan model untuk mengidentifikasi semua anggota kelas tersebut (recall), sedangkan skor F1 merangkum keduanya sebagai rata-rata harmonik. Selain nilai per kelas, ringkasan kinerja juga dapat disajikan dalam bentuk rata-rata makro dan rata-rata tertimbang. Rata-rata makro menghitung nilai rata-rata metrik antar kelas tanpa pembobotan, sehingga setiap kelas diperlakukan sama, sedangkan rata-rata tertimbang menghitung nilai rata-rata metrik dengan pembobotan berdasarkan jumlah contoh yang benar (support) di setiap kelas, sehingga nilai-nilai tersebut cenderung lebih dipengaruhi oleh kelas mayoritas [20].

Sebagai dasar untuk evaluasi klasifikasi, confusion matriks dapat digunakan karena confusion matriks merupakan metode yang umum digunakan untuk menghitung akurasi dari keseluruhan data yang diketahui guna menentukan suatu keputusan [21]. Untuk menganalisis pola kesalahan, digunakan confusion matriks karena matriks ini menunjukkan jumlah prediksi yang benar dan salah di setiap kelas (TP, TN, FP, FN), sehingga memudahkan interpretasi jenis kesalahan yang dominan - misalnya, jumlah FN di kelas negatif atau FP di kelas positif [22]. Oleh karena itu, evaluasi utama analisis sentimen dalam penelitian ini mengacu pada metrik klasifikasi teks yang dihasilkan dari model TF-IDF + SVM, termasuk akurasi, presisi, recall, skor F1, dan confusion matriks. Anotasi visual pendukung yang disiapkan di Roboflow tidak digunakan sebagai tolak ukur evaluasi utama.

3. HASIL DAN PEMBAHASAN

3.1 Deskripsi Data Penelitian

Kumpulan data yang digunakan terdiri dari ulasan pelanggan Shopee yang telah diberi label sebagai sentimen positif dan negatif.

```
... Total subset: 800
Distribusi subset:
sentiment
positive      600
negative      200
Name: count, dtype: int64
```

Gambar 2. Distribusi data penelitian.

Berdasarkan distribusi label pada Gambar 2, jumlah ulasan positif lebih dominan dibandingkan ulasan negatif. Kondisi ini menunjukkan bahwa dataset tersebut tidak seimbang, sehingga model berpotensi untuk mempelajari pola pada kelas mayoritas dengan lebih mudah. Oleh karena itu, penilaian kinerja model tidak hanya menggunakan akurasi, tetapi juga metrik khusus kelas (presisi, recall, dan F1-score) agar kemampuan model dalam mengenali sentimen negatif dapat dievaluasi secara lebih adil.

3.2 Pembagian Data Pelatihan dan Data Uji

Untuk mengevaluasi kemampuan generalisasi model, data tersebut dibagi menjadi data pelatihan dan data uji dengan perbandingan 80:20.

```
Train size: 640 Test size: 160
Distribusi test:
sentiment
positive      120
negative      40
```

Gambar 3. Pembagian data pelatihan dan data uji, serta distribusi label dalam data uji.

Gambar 3 menunjukkan proses pembagian kumpulan data menjadi data pelatihan dan data pengujian untuk mengukur kemampuan generalisasi model. Dengan skema 80:20, 640 ulasan digunakan pada tahap pelatihan dan 160 ulasan digunakan pada tahap pengujian. Pembagian dilakukan dengan menggunakan stratifikasi sehingga proporsi kelas dalam data uji tidak berubah dari distribusi awal, memastikan bahwa komposisi data uji tetap 120 positif dan 40 negatif. Strategi ini membantu memastikan bahwa evaluasi kinerja model lebih representatif dan tidak bias akibat perbedaan distribusi label antara data pelatihan dan data uji.

3.3 Hasil Kinerja Model

Setelah model TF-IDF + SVM dilatih menggunakan data pelatihan, model tersebut diuji pada data uji untuk memperoleh metrik kinerja. Hasil evaluasi ditampilkan pada Gambar 4, yang memuat nilai akurasi serta laporan klasifikasi (presisi, recall, F1-score, dan support) untuk kelas positif dan negatif. Penyajian metrik per kelas diperlukan agar kinerja model pada kelas minoritas (negatif) dapat terlihat dengan jelas.

```
Accuracy: 0.8375

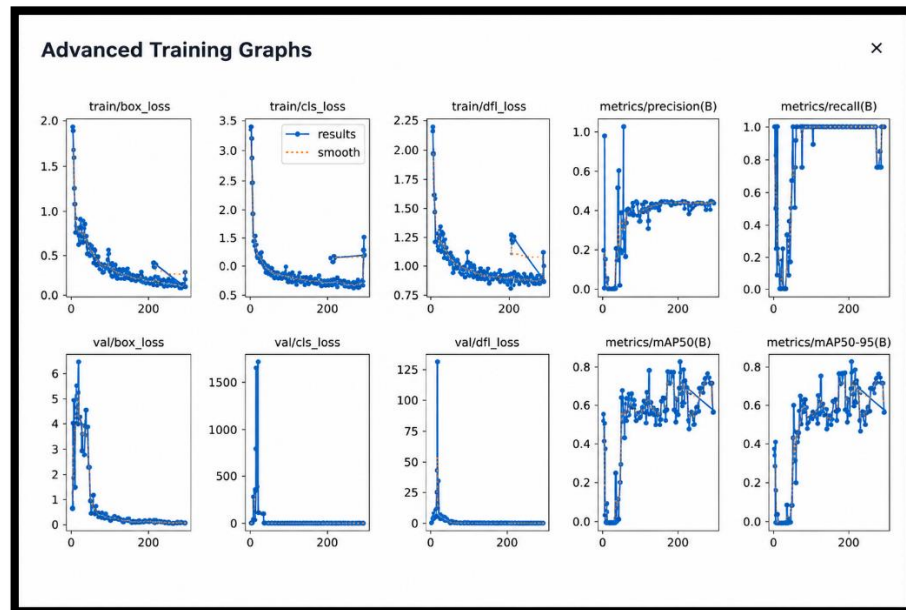
Classification report:
              precision    recall  f1-score   support

negative      0.85        0.42        0.57         40
positive      0.84        0.97        0.90        120

accuracy              0.84              160
macro avg      0.84        0.70        0.73        160
weighted avg   0.84        0.84        0.82        160
```

Gambar 4. Laporan akurasi dan klasifikasi model TF-IDF + SVM pada data uji.

3.4 Dokumen Pendukung dari Roboflow



Gambar 5. Kurva pelatihan dan validasi model pada Roboflow.

Gambar 5 menunjukkan kurva pelatihan dan validasi selama proses pelatihan model di platform Roboflow dengan menggunakan data tangkapan layar ulasan sebagai dokumen pendukung. Kurva train/box_loss dan val/box_loss menggambarkan perubahan dalam kesalahan prediksi posisi kotak pembatas relatif terhadap anotasi, sedangkan kurva train/cls_loss dan val/cls_loss menunjukkan kesalahan klasifikasi kelas objek yang terdeteksi. Kurva train/dfl_loss dan val/dfl_loss mewakili komponen kerugian tambahan yang terkait dengan kualitas regresi kotak pembatas. Secara umum, tren penurunan loss pada tahap awal menunjukkan bahwa model mulai mempelajari pola anotasi; selanjutnya, kurva cenderung stabil seiring proses pelatihan mendekati konvergensi. Hasil ini digunakan untuk mendukung proses anotasi visual, sementara evaluasi utama studi ini tetap didasarkan pada klasifikasi teks menggunakan TF-IDF dan SVM. Selain nilai loss, grafik metrics/precision(B) dan metrics/recall(B) menunjukkan perkembangan akurasi dan kelengkapan deteksi objek selama pelatihan, sedangkan metrics/mAP50(B) dan metrics/mAP50-95(B) merangkum kinerja deteksi pada ambang batas IoU yang berbeda. Nilai mAP yang meningkat dan kemudian stabil menunjukkan peningkatan kinerja deteksi dari epoch ke epoch. Grafik ini digunakan sebagai bukti pendukung untuk proses anotasi/visualisasi berbasis gambar, sementara evaluasi utama analisis sentimen dalam studi ini tetap mengacu pada kinerja klasifikasi teks menggunakan TF-IDF + SVM. Sebagai pendukung proses dokumentasi dan visualisasi, ulasan yang telah dikumpulkan juga disajikan dalam bentuk gambar untuk keperluan anotasi berbasis gambar. Proses ini dilakukan melalui Roboflow dengan tujuan memperoleh ringkasan pelatihan berupa grafik epoch dan metrik deteksi pada data gambar. Perlu ditekankan bahwa metrik yang dihasilkan oleh Roboflow (misalnya, mAP@50-95) mewakili kinerja model pada tugas berbasis gambar, sehingga metrik tersebut tidak dapat dibandingkan secara langsung dengan metrik klasifikasi sentimen berbasis teks pada TF-IDF + SVM.

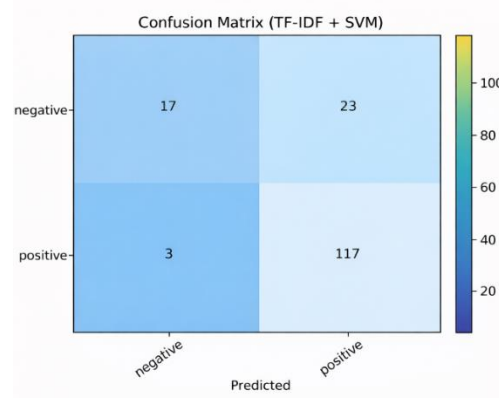
MODEL ACCURACY RESULTS (Best Weights)		
Metric	Value	Percentage
Accuracy (mAP@50-95)	0.8464	84.64%
Precision	0.6591	65.91%
Recall	0.9447	94.47%
F1-Score	0.7764	77.64%

Gambar 6. Ringkasan hasil pelatihan model pada Roboflow.

Berdasarkan Gambar 6, model Roboflow menghasilkan nilai $mAP@50-95$ sebesar 0,8464, presisi 0,6591, recall 0,9447, dan F1-score 0,7764. Nilai-nilai ini digunakan sebagai ringkasan hasil pelatihan pada data gambar dan berfungsi sebagai bukti pendukung untuk proses anotasi/visualisasi. Evaluasi utama analisis sentimen dalam studi ini masih mengacu pada hasil klasifikasi teks menggunakan TF-IDF + SVM sebagaimana dijelaskan pada subbab sebelumnya.

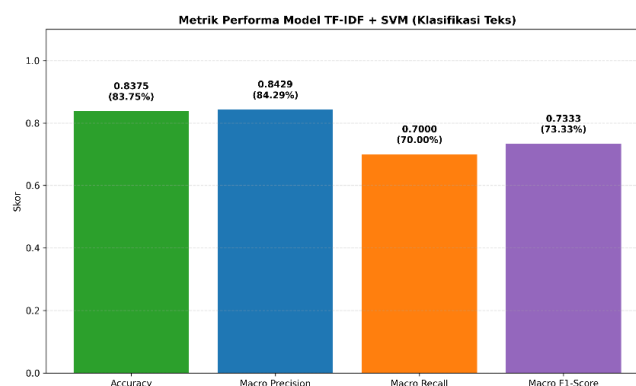
3.5 Confusion Matriks

Selain laporan klasifikasi, evaluasi model juga dilakukan dengan menggunakan confusion matriks untuk melihat rincian jumlah prediksi yang benar dan salah pada setiap kelas. Confusion matriks membantu mengidentifikasi arah kesalahan klasifikasi, misalnya ketika ulasan negatif diprediksi sebagai positif, sehingga perbedaan kinerja antar kelas dapat dijelaskan secara lebih konkret.



Gambar 7. Visualisasi confusion matriks heatmap model TF-IDF + SVM.

Berdasarkan Gambar 7, dari 40 ulasan negatif dalam data uji, 17 di antaranya diprediksi dengan benar sebagai negatif, sedangkan 23 ulasan negatif diprediksi secara keliru sebagai positif. Sementara itu, dari 120 ulasan positif, 117 di antaranya diprediksi dengan benar sebagai positif dan 3 ulasan positif diprediksi secara salah sebagai negatif. Pola ini menunjukkan bahwa kesalahan yang paling dominan terjadi pada kelas negatif, yang cenderung bergeser ke prediksi positif, sehingga membuat kelas negatif lebih sulit dikenali oleh model.



Gambar 8. Ringkasan metrik kinerja model TF-IDF + SVM pada data uji.

Berdasarkan Gambar 8, kinerja model pada data uji ditampilkan melalui nilai akurasi dan rata-rata makro (presisi, recall, dan F1-score). Penyajian rata-rata makro membantu memberikan gambaran yang lebih seimbang antar kelas, terutama dalam kondisi di mana distribusi label tidak seimbang. Nilai recall makro yang relatif lebih rendah menunjukkan bahwa masih ada kelas-kelas yang lebih sulit dikenali secara konsisten, sejalan dengan pola kesalahan yang terlihat pada confusion matriks.

4. KESIMPULAN

Penelitian ini memaparkan penerapan analisis sentimen pada ulasan pelanggan Shopee melalui pendekatan klasifikasi teks. Ulasan-ulasan tersebut, yang semula berupa data teks, diubah menjadi fitur numerik menggunakan TF-IDF, kemudian dianalisis dengan algoritma Support Vector Machine (SVM) untuk membedakan antara sentimen positif dan negatif. Sebanyak 800 ulasan digunakan, yang terdiri dari 600 ulasan positif dan 200 ulasan negatif, sehingga menghasilkan distribusi kelas yang tidak seimbang. Evaluasi dilakukan dengan memisahkan data pelatihan dan data uji menggunakan pembagian bertingkat (80:20). Berdasarkan hasil pengujian, model TF-IDF + SVM mencapai akurasi sebesar 0,8375 (83,75%). Metrik per kelas menunjukkan bahwa model ini lebih konsisten dalam mengenali ulasan positif, sementara mendeteksi ulasan negatif masih relatif lebih menantang. Pola kesalahan ini juga terlihat pada matriks kebingungan, di mana beberapa ulasan negatif masih diklasifikasikan sebagai positif. Kondisi ini menunjukkan bahwa dominasi kelas positif dan variasi ekspresi keluhan dalam kelas negatif dapat memengaruhi kemampuan model untuk menangkap sinyal sentimen negatif secara komprehensif. Sebagai dokumen pendukung, beberapa ulasan juga disiapkan dalam bentuk tangkapan layar dan diberi anotasi menggunakan Roboflow. Namun, penilaian utama dalam penelitian ini tetap berfokus pada kinerja klasifikasi teks menggunakan TF-IDF + SVM. Secara praktis, hasil analisis sentimen dapat membantu penjual mengidentifikasi tren kepuasan pelanggan sekaligus memetakan indikasi ketidakpuasan yang perlu diprioritaskan untuk diperbaiki, sehingga mendukung kualitas layanan dan kinerja penjual di Shopee.

REFERENSI

- [1] I. Iskandar and A. Nazir, "Data Mining for Analyzing Consumer Segmentation : Identifying Consumer Preference Patterns Using the Fuzzy C-Means Clustering on Halal Products," vol. 11, no. 2, pp. 134–141, 2025, doi: 10.24014/coreit.v11i2.38608.
- [2] N. Nur and A. Sjaif, "A Survey on Sentiment Analysis Approaches in," vol. 12, no. 10, pp. 674–679, 2021.
- [3] A. D. Larasati and A. Diana, "Optimization Of Digital Marketing Utilization Based on E-Commerce to Enhance Sales and Marketing," vol. 15, pp. 144–150, 2026.
- [4] T. Azizah, "RANCANG BANGUN SISTEM INFORMASI PENJUALAN BERBASIS WEB (E-COMMERCE) PADA TOKO RUMAH POPOK KINAN STKIP PGRI Situbondo , Indonesia PENDAHULUAN Laju informasi di era digital ini berkembang dengan sangat pesat di seluruh dunia . Segala bentuk informasi yang," vol. 10, no. 1, pp. 154–170, 2023.
- [5] U. Singh, A. Saraswat, H. K. Azad, K. Abhishek, and S. Shitharth, "Towards improving e - commerce customer review analysis for sentiment detection," *Sci. Rep.*, no. 2022, pp. 1–15, 2022, doi: 10.1038/s41598-022-26432-3.
- [6] M. Xanderina *et al.*, "J-ENSISTEC (Journal of Engineering and Sustainable Technology) Vol. 10|No. 02, June 2024 ANALISIS SENTIMEN ULASAN E-COMMERCE SHOPEE PADA GOOGLE PLAY STORE MENGGUNAKAN MACHINE LEARNING," vol. 10, no. 02, pp. 990–998, 2024.
- [7] K. Hantoro, D. Handayani, and S. Setiawati, "A Implementation of Text Mining In Sentiment Analysis of Shopee Indonesia Using SVM," vol. 3, no. 2, pp. 115–120, 2022.
- [8] R. Sapkota and M. Flores-calero, "YOLO11 to Its Genesis : A Decadal and Comprehensive Review of The You Only Look Once (YOLO) Series," 2024.
- [9] N. Azizah, "Car Vehicle Image Object Detection Using You Only Live Once (YOLO)," 2020.
- [10] H. Ma, A. P. Wibawa, and M. I. Akbar, "Klasifikasi artikel ilmiah dengan berbagai skenario preprocessing," vol. 2, no. 2, pp. 70–78, 2020.
- [11] T. Wahyuni and D. A. Salsabila, "AKURASI ANALISIS SENTIMEN TEKS DENGAN MENGGUNAKAN TF-IDF STUDI KASUS NLP ANALYSIS AND COMPARISON OF STOPWORDS ON TEXT SENTIMENT," vol. 10, pp. 1–5, 2025.
- [12] B. Adhikari, "Iterative Bounding Box Annotation for Object Detection," 2020.
- [13] F. I. Komputer, U. S. Karawang, J. H. S. Ronggowaluyo, T. Timur, and J. Barat, "PENERAPAN ALGORITMA SUPPORT VECTOR MACHINE TERHADAP ANALISIS SENTIMEN PADA ULASAN," vol. 13, no. 3, pp. 1908–1917, 2024.



- [14] R. Hakim and D. Rolliawati, "Topic Modeling Pada Abstrak Skripsi Menggunakan Metode Latent Semantic Analysis," vol. 11, pp. 83–90, 2022.
- [15] F. Carvalho and G. P. Guedes, "TF-IDFC-RF : A Novel Supervised Term Weighting Scheme for Sentiment Analysis," pp. 1–28, 2020.
- [16] R. Yacoub, "Probabilistic Extension of Precision , Recall , and F1 Score for More Thorough Evaluation of Classification Models," pp. 79–91, 2020.
- [17] T. Sela, A. Sonita, T. Informatika, and U. M. Bengkulu, "Comparison of Naive Bayes and Support Vector Machine Algorithms in Sentiment Analysis," pp. 973–984, 2025.
- [18] C. Anam and N. Rusdiana, "Analisis Peningkatan Kualitas Klasifikasi Pada Dataset Tidak Seimbang," vol. 5, no. 1, pp. 38–44, 2020.
- [19] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," pp. 1–13, 2020.
- [20] T. P. Pagano *et al.*, "Context-Based Patterns in Machine Learning Bias and Fairness Metrics : A Sensitive Attributes-Based Approach," no. ML, pp. 1–18, 2023.
- [21] N. Wijaya, "Perbandingan Tingkat Akurasi Pada Jenis Kedelai Berdasarkan Citra Kedelai Menggunakan Backpropagation Comparison of Accuracy Rate for Soybean Type Using Backpropagation," pp. 23–32, 2021, doi: 10.30818/jpkm.2021.2060204.
- [22] S. Sathyanarayanan and B. R. Tantri, "Confusion Matrix-Based Performance Evaluation Metrics," vol. 27, no. 4, 2024
- [23] Singh, U., Saraswat, A., Azad, H. K., Abhishek, K., & Shitharth, S. (2022). Towards improving e-commerce customer review analysis for sentiment detection. *Scientific Reports*, 12, 21983 /doi.org/10.1038/s41598-022-26432-3.
- [24] Hussain, A., Cambria, E., & Schuller, B. W. (2022). Sentiment analysis and opinion mining: Recent advances and applications. *IEEE Transactions on Affective Computing*.
- [25] Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J. (2021). Deep Learning Based Text Classification: A Comprehensive Review. *ACM Computing Surveys*, 54(3), 1–40. doi.org/10.1145/3439726
- [26] Wahyuni, T., & Salsabila, D. A. (2025). Akurasi Analisis Sentimen Teks dengan Menggunakan TF-IDF: Studi Kasus NLP Analysis and Comparison of Stopwords on Text Sentiment. *Jurnal Teknologi Informasi dan Komputer*.
- [27] Hantoro, K., Handayani, D., & Setiawati, S. (2022). Implementation of Text Mining in Sentiment Analysis of Shopee Indonesia Using SVM. *Journal of Information Technology Research*, 3(2), 115–120.
- [28] Sari, R., & Wibowo, A. (2023). Sentiment Analysis of Marketplace Customer Reviews Using Support Vector Machine and Naïve Bayes Algorithm. *International Journal of Information System and Computer Science*.
- [29] Kowsari, K., Jafari Meimandi, K., Heidarysafa, M., Mendu, S., Barnes, L., & Brown, D. (2021). Text Classification Algorithms: A Survey. *Information*, 12(4), 150.
- [30] Chicco, D., & Jurman, G. (2020). The Advantages of the Matthews Correlation Coefficient (MCC) Over F1 Score and Accuracy in Binary Classification Evaluation. *BMC Genomics*, 21, 6.