

Komparasi TF-IDF dan BoW pada Analisis Sentimen Shopee-Tokopedia

The Influence of TF-IDF and BoW on Shopee-Tokopedia Sentiment Analysis

Jihan Salsabila¹, Silvia Meida², Efelien Anindya Shifani³, Hana Mar'atul Afifah⁴, Hidayat^{5*}

^{1,2,3,4}Program Studi Teknik Informatika, Universitas Komputer Indonesia, Bandung, Indonesia

⁵Program Studi Teknik Komputer, Universitas Komputer Indonesia, Bandung, Indonesia

*E-mail: hidayat@email.unikom.ac.id

Abstrak

Perkembangan e-commerce di Indonesia telah memicu peningkatan interaksi digital berupa ulasan dan opini pengguna terhadap layanan maupun produk. Data teks dari ulasan tersebut menyimpan informasi penting yang dapat diolah melalui analisis sentimen untuk memahami persepsi pengguna. Penelitian ini bertujuan untuk membandingkan efektivitas metode ekstraksi fitur Term Frequency–Inverse Document Frequency (TF-IDF) dan Bag of Words (BoW) dalam mengklasifikasikan sentimen ulasan pengguna, serta mengevaluasi performa algoritma Support Vector Machine (SVM) dan Random Forest (RF) pada platform e-commerce Shopee dan Tokopedia. Sebanyak 5.000 ulasan pengguna dianalisis melalui tahapan pra-pemrosesan teks, pelabelan sentimen berbasis leksikon, penerapan metode ekstraksi fitur TF-IDF dan BoW, pelatihan model dengan algoritma SVM dan RF, serta evaluasi kinerja menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil pengujian menunjukkan kombinasi BoW dan SVM memberikan akurasi tertinggi sebesar 90% pada ulasan Shopee, menjadikannya konfigurasi paling optimal dalam studi ini. Selain itu, pada ulasan Tokopedia, konfigurasi yang sama (BoW dan SVM) juga menghasilkan akurasi kuat sebesar 88%. Secara umum, algoritma SVM menunjukkan performa lebih stabil dibandingkan RF, sementara metode BoW terbukti lebih efektif (terukur hingga 90% akurasi) dalam merepresentasikan data ulasan e-commerce berbahasa Indonesia ini. Temuan ini memberikan kontribusi penting bagi pengembangan sistem analisis sentimen yang efisien dan akurat bagi platform e-commerce.

Kata kunci: Analisis Sentimen, TF-IDF, Bag of Words, Support Vector Machine, Random Forest

Abstract

The rapid growth of e-commerce in Indonesia has led to an increase in user interactions in the form of reviews and opinions on services and products. These textual data contain valuable information that can be processed through sentiment analysis to better understand user perceptions. This study aims to compare the effectiveness of Term Frequency–Inverse Document Frequency (TF-IDF) and Bag of Words (BoW) feature extraction methods in classifying user sentiments, as well as to evaluate the performance of Support Vector Machine (SVM) and Random Forest (RF) algorithms on Shopee and Tokopedia platforms. A total of 5,000 user reviews were analyzed through text preprocessing, lexicon-based sentiment labeling, application of TF-IDF and BoW feature extraction methods, model training using SVM and RF algorithms, and performance evaluation using accuracy, precision, recall, and F1-score metrics. The experimental results show the combination of BoW and SVM achieved the highest accuracy of 90% on Shopee reviews, making it the most optimal configuration in this study. Additionally, in Tokopedia reviews, the same configuration (BoW and SVM) also produced a strong accuracy of 88%. In general, the SVM algorithm showed more stable performance than RF, while the BoW method proved to be more effective (measured at up to 90% accuracy) in representing this Indonesian-language e-commerce review data. These findings contribute to the development of more accurate sentiment analysis systems in the local e-commerce domain.

Keywords: Sentiment Analysis, TF-IDF, Bag of Words, Support Vector Machine, Random Forest

Naskah diterima 19 Ags. 2025; direvisi 28 Nov. 2025; dipublikasikan 30 Apr. 2026.

JAMIKA is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.



I. PENDAHULUAN

Di era digital yang semakin berkembang pesat, interaksi manusia banyak berlangsung secara daring melalui platform media sosial dan e-commerce. Aktivitas pengguna di ruang digital menghasilkan berbagai

opini dalam bentuk komentar, ulasan, hingga tanggapan tekstual. Setiap harinya, miliaran data teks tercipta dari interaksi ini, yang mencerminkan pandangan, pengalaman, serta perasaan pengguna terhadap produk atau layanan. Data tekstual yang melimpah ini menyimpan potensi strategis apabila dimanfaatkan secara optimal melalui teknik analisis sentimen, yakni proses untuk mengidentifikasi opini positif, negatif, atau netral dalam suatu pernyataan guna mendukung pengambilan keputusan berbasis data yang lebih akurat [1].

Analisis sentimen sendiri menjadi salah satu pendekatan penting dalam memahami persepsi publik, terutama dalam konteks bisnis digital. Dalam praktiknya, analisis ini membutuhkan transformasi teks mentah menjadi representasi numerik agar dapat diproses oleh algoritma *machine learning*. Salah satu tahap krusial dalam proses ini adalah ekstraksi fitur, yaitu mengubah kata-kata menjadi bentuk vektor numerik. Dua metode yang paling umum digunakan untuk ekstraksi fitur dalam *Natural Language Processing* (NLP) adalah *Bag of Words* (BoW) dan *Term Frequency-Inverse Document Frequency* (TF-IDF). Kedua metode ini memiliki pendekatan berbeda dalam merepresentasikan pentingnya kata dalam suatu dokumen, yang kemudian mempengaruhi performa model analisis sentimen [2].

BoW merepresentasikan dokumen sebagai kumpulan kata-kata unik dari seluruh korpus tanpa mempertimbangkan urutan atau makna kata, dengan mengandalkan frekuensi kemunculan kata sebagai bobot. Meski sederhana dan banyak digunakan, pendekatan ini memiliki kelemahan karena tidak mampu menangkap konteks atau tingkat kepentingan kata dalam keseluruhan dokumen [2]. Sebaliknya, TF-IDF menawarkan pendekatan yang lebih cerdas dengan tidak hanya menghitung frekuensi kata dalam satu dokumen, tetapi juga mempertimbangkan seberapa jarang kata tersebut muncul dalam keseluruhan korpus. Hal ini membuat TF-IDF lebih efektif dalam menangkap kata-kata yang relevan terhadap konteks spesifik, dan secara empiris cenderung memberikan hasil klasifikasi yang lebih baik [2].

Sejumlah penelitian mendukung keunggulan TF-IDF dibandingkan BoW dalam tugas klasifikasi sentimen. Penelitian oleh Omar et al. [3] menunjukkan bahwa kombinasi TF-IDF dan algoritma *Support Vector Machine* (SVM) menghasilkan akurasi sebesar 87,6%, lebih tinggi dibandingkan BoW yang hanya mencapai 81,2%. Penelitian lain oleh Suryaningrum [4] juga membandingkan performa TF-IDF dengan *Count Vectorizer* (varian BoW) dalam klasifikasi ujaran kebencian menggunakan SVM, dan menemukan bahwa TF-IDF memberikan akurasi sebesar 88,8%, sementara *Count Vectorizer* hanya mencapai 83,5%.

Penelitian oleh Putra et al. [2] memperkuat temuan tersebut dengan membandingkan TF-IDF dan BoW menggunakan SVM pada data ulasan layanan pengiriman. TF-IDF mencatat akurasi 86%, sedikit lebih unggul dari BoW yang memperoleh 85%, meskipun BoW lebih baik dalam aspek *recall*. Terakhir, Suryanti dan Prasetyaningrum [5] menggunakan algoritma *Naïve Bayes* untuk menganalisis sentimen terhadap diet kopi *Americano* di Twitter. Hasilnya menunjukkan bahwa TF-IDF menghasilkan akurasi 85%, jauh lebih tinggi dibandingkan BoW yang hanya mencapai 64%. Keempat penelitian tersebut menunjukkan bahwa TF-IDF lebih konsisten dalam menghasilkan performa klasifikasi yang unggul dibandingkan BoW, baik dengan pendekatan SVM.

Meskipun mayoritas studi literatur menunjukkan superioritas TF-IDF dalam tugas klasifikasi teks berkat kemampuannya menyeimbangkan frekuensi kata dan kekhasannya dalam dokumen, namun variasi hasil dapat terjadi pada korpus spesifik. Beberapa studi menunjukkan bahwa pada teks yang pendek, seperti ulasan atau komentar media sosial, BoW yang lebih sederhana justru dapat memberikan hasil yang kompetitif atau bahkan superior, terutama ketika didukung oleh algoritma klasifikasi yang kuat. Inilah yang menjadi *gap* penelitian ini, yaitu diperlukan investigasi komparatif yang ketat antara BoW dan TF-IDF pada korpus ulasan *e-commerce* (Shopee dan Tokopedia) berbahasa Indonesia, yang secara inheren memiliki karakteristik informal dan pendek, menggunakan algoritma kontemporer, seperti SVM dan *Random Forest* (RF) untuk mengidentifikasi arsitektur optimal yang relevan dengan konteks lokal.

Dalam klasifikasi sentimen, algoritma yang umum digunakan antara lain SVM dan RF. SVM bekerja dengan mencari bidang pemisah terbaik (*hyperplane*) untuk membedakan dua kategori data berdasarkan margin maksimal, sehingga sangat efektif dalam menangani data berdimensi tinggi dan linier. Di sisi lain, RF merupakan algoritma berbasis *ensemble learning* yang membentuk sejumlah pohon keputusan secara acak, kemudian menggabungkan hasilnya melalui proses voting mayoritas untuk menghasilkan prediksi akhir yang lebih konsisten dan akurat. Pendekatan kolektif ini secara signifikan meningkatkan stabilitas model dan mengurangi risiko *overfitting*, menghasilkan hasil prediksi yang lebih konsisten dan akurat dalam tugas klasifikasi teks. Keduanya telah terbukti memberikan performa yang kompetitif dalam berbagai studi analisis sentimen, tergantung pada karakteristik data dan teknik ekstraksi fitur yang digunakan [6].

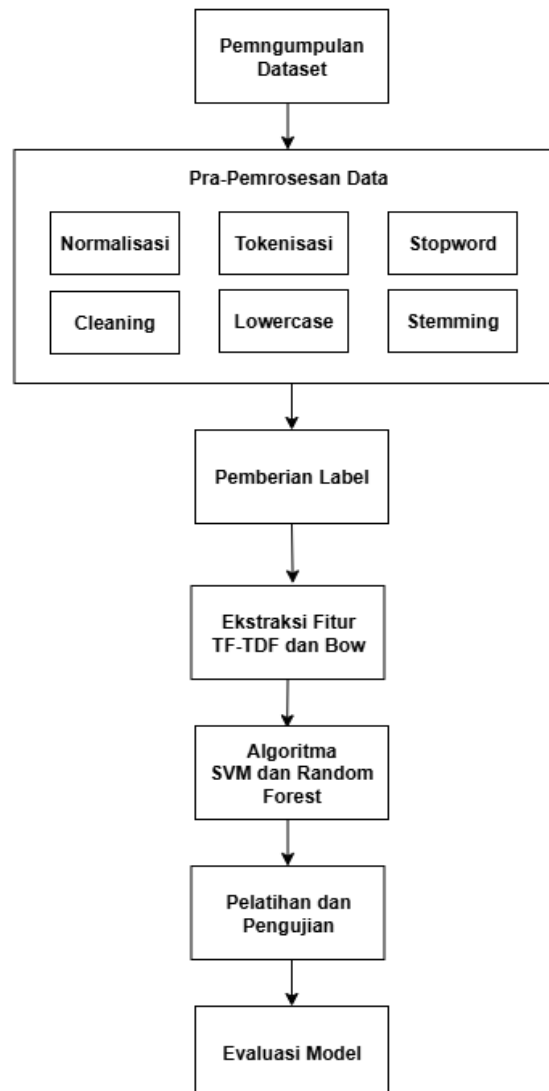
Penelitian ini bertujuan untuk menganalisis pengaruh penggunaan teknik ekstraksi fitur TF-IDF dan BoW terhadap performa algoritma SVM dan RF dalam mengklasifikasikan sentimen ulasan pengguna terhadap

aplikasi Shopee dan Tokopedia. Secara khusus, penelitian ini bertujuan untuk mengidentifikasi metode ekstraksi fitur dan algoritma klasifikasi yang paling efektif dalam konteks analisis sentimen *e-commerce* di Indonesia.

Kontribusi utama penelitian ini adalah menetapkan dan membuktikan konfigurasi BoW-SVM sebagai arsitektur optimal untuk analisis sentimen ulasan *e-commerce* Indonesia, sebuah temuan yang secara empiris berbeda dari kecenderungan umum literatur yang mendukung TF-IDF. Temuan ini memberikan rekomendasi berbasis bukti untuk pemilihan teknik ekstraksi fitur yang paling efisien, yang sangat relevan bagi pengembang yang berupaya membangun sistem analisis sentimen *real-time* dengan sumber daya komputasi yang efisien bagi sektor *e-commerce*.

II. METODE PENELITIAN

Penelitian ini melalui beberapa tahapan seperti ditunjukkan pada Gambar 1. Tahapan dimulai dari pengumpulan dataset, pra-pemrosesan data, ekstraksi fitur, klasifikasi algoritma SVM dan RF, pelatihan dan pengujian data, dan evaluasi model.



Gambar 1. Diagram Alur Penelitian

Pengumpulan Dataset

Pengumpulan data pada penelitian ini dilakukan dengan teknik web scraping menggunakan library google-play-scraper, yang berfungsi untuk mengambil ulasan aplikasi dari Google Play Store secara otomatis [7]. Dataset yang digunakan terdiri dari 5.000 komentar pengguna, masing-masing 2.500 data komentar dari aplikasi Shopee dan 2.500 dari Tokopedia. Data yang diambil merupakan komentar dalam bahasa Indonesia yang memuat opini atau ekspresi sentimen pengguna terhadap produk atau layanan dari masing-masing platform. Secara teknis, proses *crawling* dimulai dengan menentukan App ID dari masing-masing aplikasi, kemudian menggunakan fungsi *reviews()* dari pustaka google-play-scraper untuk mengambil komentar terbaru (*sort by newest*). Setiap komentar yang diambil mencakup nama pengguna (*username*), skor penilaian (*score*), dan isi komentar (*content*).

Pra-Pemrosesan Data

Pra-Pemrosesan adalah proses melakukan perubahan dari dokumen teks menjadi *term index* dengan tujuan menjadi *set term index* sehingga dapat digunakan sebagai *key word* untuk mengawali sebuah dokumen.

1. Pembersihan Teks

Tahapan ini bertujuan menghapus karakter-karakter yang tidak diperlukan seperti emoji, simbol, tanda baca, dan angka yang tidak memiliki makna kontekstual. Pembersihan dilakukan untuk menghilangkan elemen-elemen yang dapat mengganggu proses analisis atau mengakibatkan makna ganda [8].

Contoh: Teks “Barangnya bagus banget!!! 🍊🍊” → “Barangnya bagus banget”

2. Normalisasi Kata

Normalisasi dilakukan untuk mengubah kata-kata tidak baku, slang, atau kata gaul menjadi bentuk baku yang sesuai dengan kaidah Bahasa Indonesia. Pada penelitian ini, proses normalisasi dilakukan menggunakan kamus slang yang dibuat secara manual berdasarkan pengumpulan data kata tidak baku dari korpus yang digunakan. Contohnya, kata "ga" diubah menjadi "tidak", "bgt" menjadi "banget", dan sebagainya. Langkah ini penting untuk mengurangi keragaman kata yang memiliki arti sama namun bentuk penulisannya berbeda[9].

3. Lowercase

Semua huruf dalam dokumen diubah menjadi huruf kecil untuk menjaga konsistensi data[8]. Ini penting karena dalam sistem komputerisasi, “Shopee” dan “shopee” akan dianggap sebagai dua kata berbeda. Sebagai contoh, kalimat “Shopee Sangat Bagus Sekali” akan diubah menjadi “shopee sangat bagus sekali”.

4. Tokenisasi

Tokenisasi adalah proses pemecahan teks menjadi unit-unit kata (token). Misalnya, kalimat “barang cepat sampai” akan diubah menjadi daftar token ["barang", "cepat", "sampai"]. Tahapan ini memungkinkan analisis teks dilakukan pada level kata dan sangat penting untuk ekstraksi fitur berbasis kata seperti *Bag-of-Words* dan TF-IDF [8].

5. Stopword Removal

Menghapus kata-kata umum yang tidak memberikan informasi penting seperti “yang”, “dan”, “itu” [9].

6. Stemming

Memenggal imbuhan pada suatu kata menjadi bentuk dasar dari kata tersebut. misalnya “menurunkan”, “diturunkan”, “penurunan” → “turun” [9].

Pemberian Label

Labeling dilakukan secara otomatis dengan menggunakan pendekatan berbasis leksikon, yaitu kamus kata yang memiliki nilai sentimen (positif, negatif, netral). Kata dalam teks dibandingkan dengan kamus ini untuk menentukan sentimennya [10]. Pada penelitian ini, pelabelan dilakukan sebelum tahap pelatihan (training), dengan membandingkan setiap kata dalam ulasan Shopee dan Tokopedia terhadap kamus leksikon sentimen Bahasa Indonesia yang disusun secara manual.

Pelabelan ini menggunakan pendekatan berbasis leksikon dengan *Valence Aware Dictionary and sEntiment Reasoner* (VADER) dari NLTK. VADER tidak hanya mengidentifikasi polaritas sentimen, tetapi juga intensitasnya, dengan mempertimbangkan faktor linguistik seperti huruf kapital dan tanda baca. Penggunaan VADER dalam penelitian ini didasari oleh kemampuannya untuk secara efektif menganalisis

sentimen dalam teks media sosial dan informal, yang relevan untuk data ulasan [18]. Hasil pelabelan ditunjukkan pada Tabel 1 yang menghasilkan skor:

1. compound > 0,05 = positif
2. compound < -0,05 = negatif
3. compound antara -0,05 hingga 0,05 = netral

TABEL 1
 HASIL PELABELAN

No	Ulasan	Label Sentimen
1	Pesan sesuai deskripsi foto lapak	Positif
2	Terima kasih belanja toko	Netral
3	Kecewa	Negatif

Ekstraksi Fitur

Dalam proses analisis sentimen berbasis teks, ekstraksi fitur menjadi tahap penting untuk mengubah data teks menjadi representasi numerik yang dapat diproses oleh algoritma klasifikasi. *Term Frequency-Inverse Document Frequency* (TF-IDF) yaitu menghitung bobot suatu kata berdasarkan seberapa sering kata tersebut muncul dalam suatu dokumen dan seberapa jarang kata tersebut muncul di seluruh dokumen [8],[11]. Bobot tinggi menandakan bahwa kata tersebut penting untuk dokumen tertentu. Metode TF-IDF menghitung bobot kata berdasarkan dua komponen: frekuensi kemunculan kata dalam dokumen (*term frequency*) dan tingkat keunikan kata dalam seluruh dokumen (*inverse document frequency*). Kata yang sering muncul di satu dokumen namun jarang muncul di dokumen lain akan memiliki bobot lebih tinggi[12].

Bag Of Word (BoW) yaitu menghitung frekuensi kemunculan kata dalam dokumen tanpa memperhatikan urutan kata [11]. Pada metode BoW, setiap dokumen direpresentasikan sebagai vektor berdasarkan frekuensi kemunculan kata tanpa memperhatikan urutan kata. Langkah awalnya yaitu membentuk *vocabulary* dari seluruh dokumen, lalu menghitung jumlah kemunculan masing-masing kata dalam setiap dokumen. Setelah teks dari ulasan dibersihkan dan diproses (tokenisasi, stopword removal, lematisasi), fitur diekstraksi menggunakan metode *CountVectorizer*. Ini merupakan metode *Bag of Words*, yang mengubah teks menjadi vektor numerik berdasarkan frekuensi kata. Hasil dari kedua metode ini berupa matriks fitur yang digunakan sebagai input untuk klasifikasi algoritma pada tahap selanjutnya.

Algoritma yang Digunakan

Pada penelitian ini, algoritma klasifikasi yang digunakan adalah SVM dan RF. Pemilihan kedua algoritma tersebut didasarkan pada kemampuannya yang baik dalam menangani data teks dan telah terbukti menghasilkan performa tinggi dalam tugas analisis sentimen. SVM adalah algoritma *supervised learning* yang digunakan untuk klasifikasi dengan cara menemukan *hyperplane* terbaik yang memisahkan kelas data. SVM sangat cocok untuk data berdimensi tinggi seperti teks, terutama jika dipadukan dengan ekstraksi fitur TF-IDF [13].

Random Forest adalah metode ensemble learning yang membangun beberapa decision tree dan menggabungkannya untuk meningkatkan akurasi klasifikasi. RF cenderung stabil terhadap data tidak seimbang dan cocok digunakan dengan representasi BoW [14],[15]. secara umum, kedua algoritma ini memberikan pendekatan berbeda: SVM berfokus pada margin optimal antar kelas, sedangkan *Random Forest* menggabungkan kekuatan beberapa model lemah untuk membentuk keputusan yang kuat [19]. Oleh karena itu, pada penelitian ini dilakukan pengujian terhadap keduanya guna mengetahui algoritma mana yang lebih efektif dalam menganalisis sentimen ulasan pengguna Shopee dan Tokopedia.

Pelatihan dan Pengujian

Tahap pelatihan dan pengujian bertujuan untuk membangun serta mengukur kinerja model klasifikasi terhadap data ulasan. Data yang telah melalui proses ekstraksi fitur (TF-IDF atau BoW) dibagi menjadi dua bagian, yaitu data pelatihan dan data pengujian dengan rasio 80:20 [16]. Pada tahap pelatihan, algoritma seperti SVM dan RF dilatih menggunakan data berlabel untuk mempelajari pola hubungan antara fitur dan kelas sentimen [20]. Setelah itu, model diuji menggunakan data pengujian untuk mengevaluasi kemampuannya dalam mengklasifikasikan data baru. Evaluasi dilakukan menggunakan metrik seperti akurasi, presisi, *recall*, dan *F1-score* untuk menilai seberapa baik model bekerja dalam mendeteksi sentimen secara tepat.

Evaluasi Data

Tahap evaluasi dilakukan setelah model klasifikasi dilatih dan diuji untuk menilai performa dalam mengklasifikasikan sentimen. Evaluasi menggunakan beberapa metrik utama, yaitu akurasi, presisi, *recall*, dan F1-score. Masing-masing metrik memberikan sudut pandang yang berbeda terhadap keberhasilan model.

1. Akurasi mengukur proporsi prediksi yang benar dari seluruh data uji.
2. Presisi menunjukkan ketepatan model dalam memprediksi kelas positif, yaitu berapa banyak prediksi positif yang benar.
3. *Recall* mengukur kemampuan model menemukan seluruh data yang benar-benar positif.
4. *F1-score* merupakan rata-rata harmonis antara presisi dan *recall*, digunakan saat distribusi kelas tidak seimbang.

Penggunaan keempat metrik ini penting untuk memberikan evaluasi menyeluruh terhadap performa model, terutama pada data teks yang memiliki karakteristik tidak seimbang [17]. Dalam penelitian tersebut, *F1-score* dianggap paling mencerminkan kemampuan model saat menangani ulasan pendek dengan banyak variasi kata. Evaluasi dilakukan pada hasil klasifikasi data pengujian dari algoritma SVM dan *RF*. Perbandingan nilai evaluasi antar algoritma membantu menentukan metode terbaik dalam analisis sentimen ulasan Shopee dan Tokopedia.

III. HASIL DAN PEMBAHASAN

Analisis Deskriptif

Penelitian ini bertujuan untuk menganalisis sentimen dari ulasan pengguna aplikasi Shopee dan Tokopedia yang dikumpulkan langsung dari Google Play Store. Dataset terdiri dari total 5.000 ulasan, dengan 2.500 ulasan untuk masing-masing aplikasi Shopee dan Tokopedia. Setiap ulasan mencakup teks umpan balik pengguna dan rating bintang dari 1 hingga 5.

Untuk tujuan klasifikasi sentimen, ulasan-ulasan ini telah dikategorikan menjadi tiga kelas sentimen: positif, negatif, dan netral. Pembagian kategori ini memungkinkan model untuk mengidentifikasi tidak hanya polaritas sentimen ekstrim, tetapi juga sentimen yang lebih moderat atau tidak jelas. Pendekatan ini penting untuk mendapatkan pemahaman yang lebih nuansa tentang persepsi pengguna terhadap kedua aplikasi *e-commerce* tersebut.

```
Distribusi Sentimen Shopee:  
sentiment_vader  
Positive    60.601137  
Neutral     29.813160  
Negative     9.585703  
Name: proportion, dtype: float64
```

Gambar 2. Distribusi sentimen, a) Shopee, b) Tokopedia

Distribusi sentimen untuk Shopee (Gambar 2) menunjukkan bahwa sebagian besar sentimen pengguna adalah positif, mencapai 60,60% dari total sentimen yang teridentifikasi. Sentimen netral berada pada angka 29,61%, sementara sentimen negatif merupakan proporsi terkecil, yaitu 9,79%. Dominansi sentimen positif untuk Shopee mengindikasikan persepsi pengguna yang secara umum baik terhadap layanan atau produk yang ditawarkan.

```
Distribusi Sentimen Tokopedia:  
sentiment_vader  
Positive    41.154941  
Neutral     35.583571  
Negative    23.261488  
Name: proportion, dtype: float64
```

Gambar 3. Distribusi sentimen Tokopedia

Distribusi sentimen untuk Tokopedia (Gambar 2) menunjukkan pola yang berbeda dengan shopee. Meskipun sentimen positif masih merupakan kategori terbesar (41.11%), proporsinya secara signifikan lebih rendah dibandingkan dengan Shopee. Sentimen netral untuk Tokopedia sedikit lebih tinggi daripada Shopee, yaitu 35.58%. Yang paling mencolok adalah proporsi sentimen negatif yang lebih tinggi untuk Tokopedia, mencapai 23.30%, menunjukkan adanya area yang mungkin memerlukan perhatian lebih lanjut dari pihak platform.

Prosedur Eksperimen dan Evaluasi Model

Pengujian ini melibatkan eksplorasi dua teknik ekstraksi fitur, yaitu *Bag of Words* (BoW) dan TF-IDF (*Term Frequency-Inverse Document Frequency*), yang diaplikasikan pada data teks yang telah melalui tahap pra-pemrosesan (normalisasi, tokenisasi, penghapusan *stopwords*, dan *stemming*). Fitur-fitur yang dihasilkan kemudian digunakan untuk melatih dan mengevaluasi dua algoritma klasifikasi: SVM dan *RF*. Evaluasi kinerja setiap konfigurasi model dilakukan berdasarkan metrik *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Hasil penelitian ini diharapkan memberikan wawasan mengenai efektivitas berbagai kombinasi ekstraksi fitur dan algoritma dalam tugas analisis sentimen pada ulasan aplikasi berbahasa Indonesia, serta mengidentifikasi konfigurasi model yang paling optimal.

Evaluasi Model SVM - Data Shopee
 Akurasi: 0.90
 Classification Report:

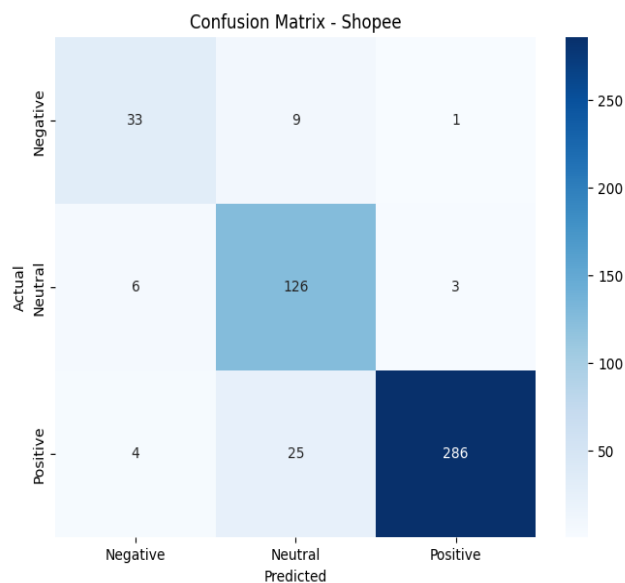
	precision	recall	f1-score	support
Negative	0.77	0.77	0.77	43
Neutral	0.79	0.93	0.85	135
Positive	0.99	0.91	0.95	315
accuracy			0.90	493
macro avg	0.85	0.87	0.86	493
weighted avg	0.91	0.90	0.90	493

Gambar 4. Hasil Pengujian Data Shopee Menggunakan Bow dan SVM

Gambar 4 menampilkan hasil pengujian dengan menggunakan ekstraksi fitur BoW dan model SVM menunjukkan kinerja yang sangat baik dalam klasifikasi sentimen pada data Shopee dengan akurasi keseluruhan 90%. Kinerja terbaik dicapai pada klasifikasi sentimen "Positif" yang menunjukkan presisi dan *recall* yang sangat tinggi. Meskipun kelas "Negatif" memiliki jumlah sampel yang lebih sedikit, model masih menunjukkan kemampuan klasifikasi yang wajar. Hasil ini mengindikasikan bahwa model SVM dapat menjadi alat yang efektif untuk memahami sentimen pengguna terhadap platform *e-commerce* seperti Shopee.

Model SVM telah dievaluasi untuk klasifikasi sentimen (Negatif, Netral, Positif) pada data ulasan Shopee, mencapai akurasi keseluruhan yang sangat baik sebesar 0,90 dari total 493 sampel uji. Berdasarkan *classification report* dan analisis *confusion matrix* (Gambar 4 dan Gambar 5), model menunjukkan kinerja optimal pada identifikasi sentimen "Positif" dengan presisi 0,99 dan *recall* 0,91 (286 dari 315 sampel positif terklasifikasi dengan benar), menunjukkan kemampuan model yang sangat baik dalam mengenali ulasan positif. Untuk sentimen "Netral", model juga berkinerja kuat dengan presisi 0,79 dan *recall* 0,93 (126 dari 135 sampel netral terklasifikasi dengan benar), meskipun terdapat beberapa kesalahan klasifikasi ke sentimen negatif (6 sampel) atau positif (3 sampel).

Sementara itu, untuk sentimen "Negatif", meskipun merupakan kelas minoritas (43 sampel), model masih mencapai presisi dan *recall* 0,77 (33 sampel negatif terklasifikasi dengan benar), namun perlu dicatat adanya beberapa kasus sentimen negatif yang salah diklasifikasikan sebagai netral (9 sampel) atau positif (1 sampel). Secara keseluruhan, hasil ini menegaskan bahwa model SVM efektif dalam klasifikasi sentimen pada data Shopee, dengan area untuk peningkatan potensial pada identifikasi sentimen negatif.



Gambar 5. *Confusion Matrix* Data Shopee Menggunakan Bow dan SVM

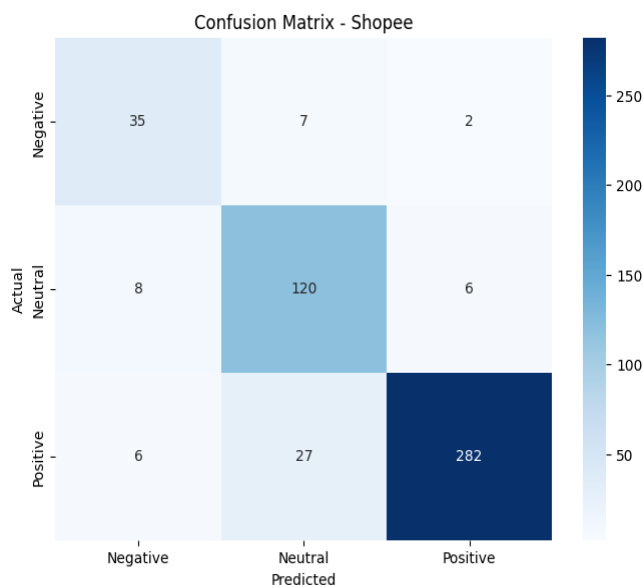
Evaluasi kinerja model SVM untuk klasifikasi sentimen (Negatif, Netral, Positif) pada 493 sampel data ulasan Shopee dengan menggunakan ekstraksi fitur TF-IDF (Gambar 6) menunjukkan akurasi keseluruhan sebesar 0,89. Analisis metrik per-kelas mengungkapkan kinerja yang kuat pada identifikasi sentimen "Positif" dengan presisi 0,97 dan *recall* 0,90. Sentimen "Netral" juga terklasifikasi dengan baik, mencapai presisi 0,78 dan *recall* 0,90. Meskipun demikian, untuk sentimen "Negatif", model menunjukkan presisi 0,71 dan *recall* 0,80. Hasil ini mengindikasikan bahwa model SVM efektif dalam klasifikasi sentimen pada data pengguna Shopee, dengan akurasi tinggi terutama pada kelas dominan "Positif" dan "Netral", namun masih terdapat ruang untuk peningkatan pada presisi klasifikasi sentimen "Negatif".

Evaluasi Model SVM - Data Shopee
Akurasi: 0.89

Classification Report:

	precision	recall	f1-score	support
Negative	0.71	0.80	0.75	44
Neutral	0.78	0.90	0.83	134
Positive	0.97	0.90	0.93	315
accuracy			0.89	493
macro avg	0.82	0.86	0.84	493
weighted avg	0.90	0.89	0.89	493

Gambar 6. Hasil Pengujian Data Shopee Menggunakan TF-IDF dan SVM



Gambar 7. *Confusion Matrix* Hasil Pengujian Data Shopee Menggunakan TF-IDF dan SVM

Berdasarkan *confusion matrix* yang diberikan (Gambar 7), analisis lebih lanjut mengkonfirmasi distribusi kesalahan klasifikasi model. Sebanyak 35 dari 44 sampel "Negatif" terklasifikasi dengan benar, namun 7 di antaranya salah diprediksi sebagai "Netral" dan 2 sebagai "Positif". Untuk sentimen "Netral", 120 dari 134 sampel berhasil diidentifikasi, dengan kesalahan dominan pada prediksi sebagai "Negatif" (8 sampel) atau "Positif" (6 sampel). Sementara itu, model sangat efektif dalam mengidentifikasi sentimen "Positif", dengan 282 dari 315 sampel terklasifikasi secara tepat, dan kesalahan klasifikasi ke sentimen lain relatif minimal (6 sebagai "Negatif", 27 sebagai "Netral"). Data ini secara spesifik menunjukkan area di mana model masih dapat ditingkatkan, terutama dalam membedakan sentimen negatif dari kelas lainnya.

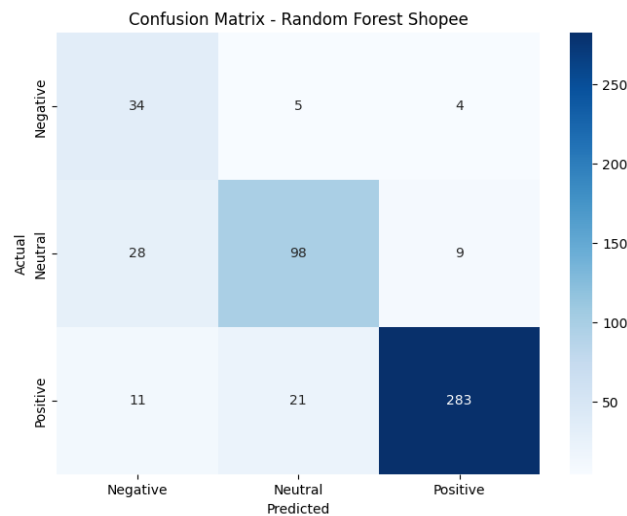
Evaluasi Model Random Forest - Data Shopee
 Akurasi: 0.84
 Classification Report:

	precision	recall	f1-score	support
Negative	0.47	0.79	0.59	43
Neutral	0.79	0.73	0.76	135
Positive	0.96	0.90	0.93	315
accuracy			0.84	493
macro avg	0.74	0.81	0.76	493
weighted avg	0.87	0.84	0.85	493

Gambar 8. Hasil Pengujian Data Shopee Menggunakan BoW dan *Random Forest*

Evaluasi model *Random Forest* pada data sentimen Shopee menghasilkan akurasi 0,84 (Gambar 8), sedikit di bawah model SVM sebelumnya (0,90). Model ini menunjukkan kinerja setara dengan SVM dalam mengklasifikasikan sentimen "Positif" (presisi 0,96, *recall* 0,90). Namun, *Random Forest* berkinerja lebih rendah untuk sentimen "Netral" (presisi 0,79, *recall* 0,73) dan khususnya "Negatif" (presisi 0,47, *recall* 0,79),

mengindikasikan bahwa SVM lebih efektif dalam penanganan kelas minoritas dan distensi sentimen netral/negatif pada dataset ini.



Gambar 9. Confusion Matrix Hasil Pengujian Data Shopee Menggunakan BoW dan Random Forest

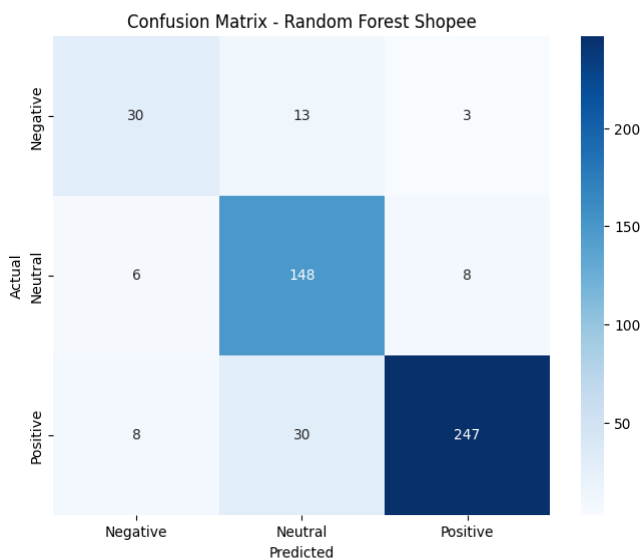
Analisis *Confusion matrix* model *Random Forest* (Gambar 9) menunjukkan kinerja kuat pada sentimen "Positif", dengan 283 dari 315 sampel terklasifikasi tepat. Namun, model menunjukkan tantangan signifikan pada sentimen "Negatif", di mana hanya 34 dari 43 sampel aktual yang teridentifikasi benar, dengan kesalahan dominan berupa misklasifikasi sebagai "Netral" (5) dan terutama oleh *false positive* dari kelas "Netral" (28) dan "Positif" (11). Untuk sentimen "Netral", 98 dari 135 sampel terklasifikasi akurat, namun kesalahan ke "Negatif" (28) perlu diperhatikan. Data ini secara empiris menyoroti defisiensi model dalam membedakan sentimen minoritas secara presisi.

Evaluasi Model Random Forest - Data Shopee
Akurasi: 0.84
Classification Report:

	precision	recall	f1-score	support
Negative	0.47	0.79	0.59	43
Neutral	0.79	0.73	0.76	135
Positive	0.96	0.90	0.93	315
accuracy			0.84	493
macro avg	0.74	0.81	0.76	493
weighted avg	0.87	0.84	0.85	493

Gambar 10. Hasil Pengujian Data Shopee Menggunakan TF-IDF dan Random Forest

Evaluasi komparatif model SVM dan RF dengan fitur TF-IDF pada data sentimen Shopee menunjukkan model SVM (0,89) sedikit mengungguli *Random Forest* (0,84) dalam akurasi keseluruhan (Gambar 10). Keduanya berkinerja kuat pada sentimen "Positif", namun *Random Forest* menunjukkan presisi yang lebih rendah pada sentimen "Netral" dan secara signifikan lebih lemah pada kelas minoritas "Negatif" (presisi 0,68, *recall* 0,65) dibandingkan SVM, mengindikasikan efektivitas SVM yang lebih baik dalam menangani kompleksitas dan ketidakseimbangan data pada klasifikasi sentimen ini.



Gambar 11. *Confusion Matrix* Hasil Pengujian Data Shopee Menggunakan TF-IDF dan *Random Forest*

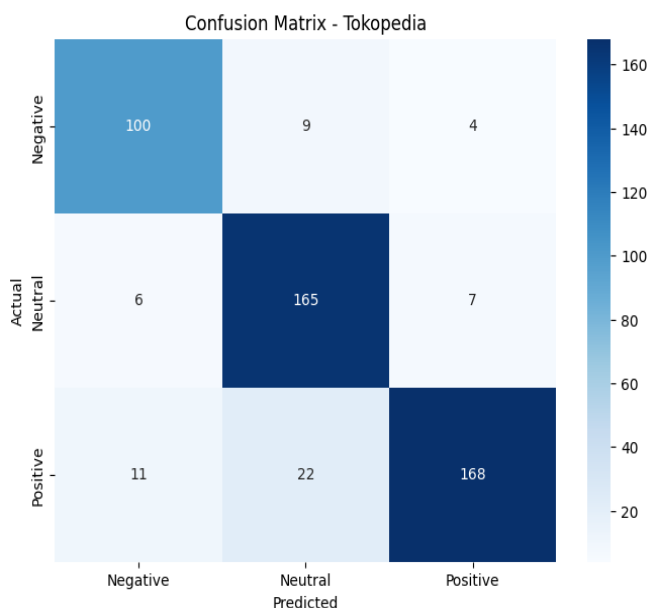
Analisis *Confusion Matrix* dengan ekstraksi fitur TF-IDF dan model *Random Forest* (Gambar 11) menunjukkan kinerja optimal pada sentimen "Positif", dengan 247 dari 285 sampel terklasifikasi tepat, dan misklasifikasi minimal. Untuk sentimen "Netral", 148 dari 162 sampel teridentifikasi akurat, namun terdapat kesalahan ke "Negatif" (6) dan "Positif" (8). Defisiensi terbesar terlihat pada sentimen "Negatif", di mana hanya 30 dari 46 sampel aktual terklasifikasi benar, serta mengalami *false positive* yang substansial dari kelas "Netral" (6) dan "Positif" (8). Hal ini secara empiris menyoroti tantangan model dalam klasifikasi presisi untuk kelas minoritas dan meminimalkan *false positive* pada sentimen "Negatif".

Evaluasi Model SVM - Data Tokopedia
Akurasi: 0.88
Classification Report:

	precision	recall	f1-score	support
Negative	0.85	0.88	0.87	113
Neutral	0.84	0.93	0.88	178
Positive	0.94	0.84	0.88	201
accuracy			0.88	492
macro avg	0.88	0.88	0.88	492
weighted avg	0.88	0.88	0.88	492

Gambar 12. Hasil Pengujian Data Tokopedia Menggunakan Bow dan SVM

Evaluasi model SVM menggunakan ekstraksi fitur *Bag of Words* (BoW) pada 492 sampel data sentimen Tokopedia menunjukkan akurasi keseluruhan 0,88 (Gambar 12). Model ini menunjukkan kinerja yang seimbang dan kuat di ketiga kelas: sentimen "Negatif" mencapai presisi 0,85 dan *recall* 0,88; sentimen "Netral" menunjukkan presisi 0,84 dengan *recall* sangat tinggi (0,93); dan sentimen "Positif" mencapai presisi 0,94 dengan *recall* 0,84. Dengan *f1-score* yang konsisten tinggi di semua kelas (sekitar 0,87-0,88), hasil ini mengindikasikan bahwa SVM dan BoW adalah pendekatan yang sangat efektif dan seimbang untuk klasifikasi sentimen pada data Tokopedia.



Gambar 13. Confusion Matrix Hasil Pengujian Data Tokopedia Menggunakan Bow dan SVM

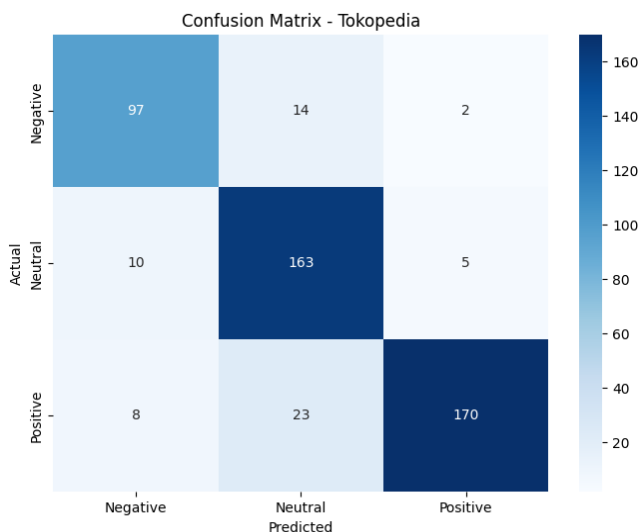
Analisis matriks konfusi model SVM pada data Tokopedia (Gambar 13) lebih lanjut menguatkan kinerja model yang seimbang. Sentimen "Negatif" terklasifikasi dengan akurasi tinggi, dengan 100 dari 113 sampel aktual teridentifikasi benar, dan misklasifikasi minimal ke "Netral" (9) atau "Positif" (4). Untuk sentimen "Netral", model sangat efektif dengan 165 dari 178 sampel aktual terklasifikasi tepat, menunjukkan *recall* yang sangat tinggi meskipun ada beberapa kesalahan ke "Negatif" (6) atau "Positif" (7). Sentimen "Positif" juga menunjukkan kinerja solid, dengan 168 dari 201 sampel aktual terklasifikasi benar, meski terdapat misklasifikasi ke "Negatif" (11) dan "Netral" (22). Pola ini secara empiris menegaskan kapabilitas model dalam klasifikasi sentimen yang akurat dan seimbang di seluruh kategori.

Evaluasi Model SVM - Data Tokopedia
Akurasi: 0.87
Classification Report:

	precision	recall	f1-score	support
Negative	0.84	0.86	0.85	113
Neutral	0.81	0.92	0.86	178
Positive	0.96	0.85	0.90	201
accuracy			0.87	492
macro avg	0.87	0.87	0.87	492
weighted avg	0.88	0.87	0.87	492

Gambar 14. Hasil Pengujian Data Tokopedia Menggunakan TF-IDF dan SVM

Hasil Evaluasi model menggunakan ekstraksi fitur TF-IDF dengan metode SVM pada 492 sampel data sentimen Tokopedia menghasilkan akurasi keseluruhan 0,87 (Gambar 14). Model ini menunjukkan kinerja yang kuat dan seimbang di seluruh kelas: sentimen "Positif" mencapai presisi tertinggi (0,96) dan *f1-score* 0,90, sentimen "Netral" unggul pada *recall* (0,92) dengan presisi 0,81, sementara sentimen "Negatif" menunjukkan presisi 0,84 dan *recall* 0,86. *Macro average* dan *weighted average* yang konsisten (0,87 dan 0,88) menegaskan kapabilitas model SVM dalam klasifikasi sentimen yang efektif pada dataset ini.



Gambar 15. *Confusion Matrix* Hasil Pengujian Data Tokopedia Menggunakan TF-IDF dan SVM

Berdasarkan *confusion matrix* (Gambar 15), model SVM yang dilatih dengan fitur TF-IDF menunjukkan performa klasifikasi yang cukup baik. Kelas Negatif berhasil dikenali sebanyak 97 dari 113 data, dengan 14 data salah sebagai Netral dan 2 sebagai Positif. Untuk kelas Netral, akurasi tinggi terlihat dengan 163 dari 178 data diklasifikasikan benar, hanya 10 salah sebagai Negatif dan 5 ke Positif. Pada kelas Positif, 170 dari 201 data berhasil dikenali, sementara 23 salah diklasifikasikan sebagai Netral dan 8 sebagai Negatif. Model cenderung kuat dalam mengenali Netral dan Positif, namun masih terjadi kebingungan antara Negatif dan Netral. Hal ini menunjukkan bahwa meskipun TF-IDF dan SVM efektif, pemisahan antar kelas Negatif dan Netral masih perlu ditingkatkan

```

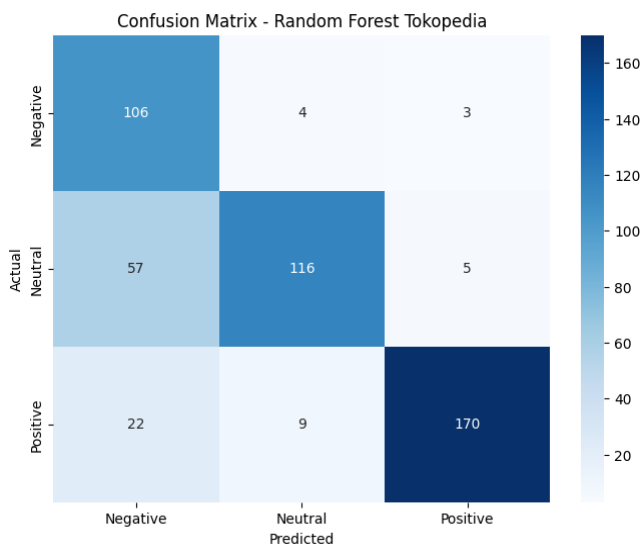
Evaluasi Model Random Forest - Data Tokopedia
Akurasi: 0.80
Classification Report:

```

	precision	recall	f1-score	support
Negative	0.57	0.94	0.71	113
Neutral	0.90	0.65	0.76	178
Positive	0.96	0.85	0.90	201
accuracy			0.80	492
macro avg	0.81	0.81	0.79	492
weighted avg	0.85	0.80	0.80	492

Gambar 16. Hasil Pengujian Data Tokopedia Menggunakan BoW dan *Random Forest*

Evaluasi model *Random Forest* menggunakan ekstraksi fitur *Bag of Words* (BoW) pada 492 sampel data sentimen Tokopedia menunjukkan akurasi keseluruhan 0,80 (Gambar 16). Akurasi ini lebih rendah signifikan dibandingkan dengan model SVM sebelumnya (0,87). Secara spesifik, model *Random Forest* masih menunjukkan kinerja yang baik untuk sentimen "Positif" (presisi 0,96, *recall* 0,85). Namun, untuk sentimen "Netral", meskipun presisinya tinggi (0,90), *recall*-nya sangat rendah (0,65), mengindikasikan banyak ulasan netral yang tidak teridentifikasi. Kinerja paling bermasalah terlihat pada sentimen "Negatif", yang meskipun memiliki *recall* tinggi (0,94), presisinya sangat rendah (0,57), menandakan tingginya proporsi ulasan non-negatif yang keliru diklasifikasi sebagai negatif.



Gambar 17. *Confusion Matrix* Hasil Pengujian Data Tokopedia Menggunakan BoW dan *Random Forest*

Confusion Matrix model *Random Forest* (Gambar 17) lebih lanjut menjelaskan pola klasifikasi yang kurang seimbang tersebut. Teridentifikasi bahwa 106 dari 113 sampel aktual "Negatif" terklasifikasi benar, menunjukkan *recall* yang kuat. Namun, kelemahan utama terletak pada sentimen "Netral", di mana hanya 116 dari 178 sampel aktual teridentifikasi tepat, dengan 57 sampel secara signifikan keliru diprediksi sebagai "Negatif". Ini secara langsung menjelaskan rendahnya *recall* netral dan rendahnya presisi negatif. Sentimen "Positif" terklasifikasi baik dengan 170 dari 201 sampel aktual tepat, meskipun 22 sampel keliru diprediksi sebagai "Negatif". Pola misklasifikasi ini, terutama *false positive* yang tinggi pada kelas "Negatif" dari sampel "Netral" dan "Positif", menjadi penyebab utama penurunan akurasi keseluruhan dibandingkan dengan model SVM.

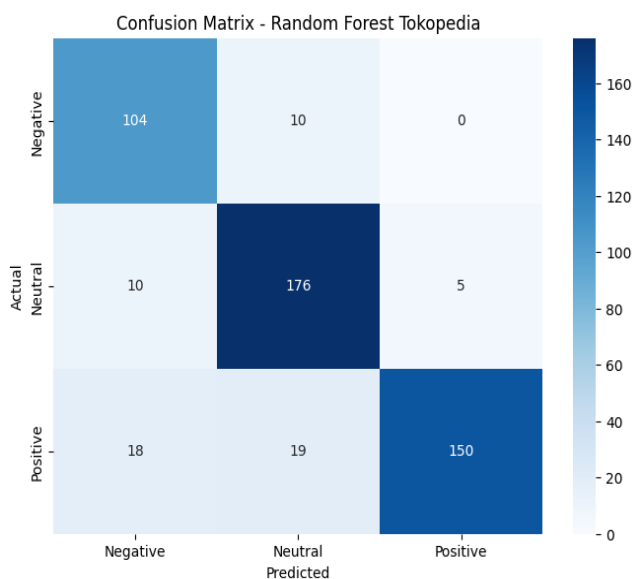
Evaluasi Model Random Forest - Data Tokopedia
Akurasi: 0.87
Classification Report:

	precision	recall	f1-score	support
Negative	0.79	0.91	0.85	114
Neutral	0.86	0.92	0.89	191
Positive	0.97	0.80	0.88	187
accuracy			0.87	492
macro avg	0.87	0.88	0.87	492
weighted avg	0.88	0.87	0.87	492

Gambar 18. Hasil Pengujian Data Tokopedia Menggunakan TF-IDF dan *Random Forest*

Evaluasi model *Random Forest* menggunakan ekstraksi fitur TF-IDF pada 492 sampel data sentimen Tokopedia menghasilkan akurasi keseluruhan 0,87 (Gambar 18). Angka ini setara dengan akurasi yang dicapai oleh model SVM sebelumnya (0,87). Secara rinci, RF menunjukkan presisi sangat tinggi (0,97) pada sentimen "Positif" namun dengan *recall* sedikit lebih rendah (0,80) dibandingkan SVM. Untuk sentimen "Netral", RF memiliki presisi (0,86) dan *recall* (0,92) yang sangat kuat, sedikit lebih baik dalam presisi dan setara dalam *recall* dibandingkan SVM. Pada sentimen "Negatif", RF mencapai *recall* yang sangat tinggi (0,91) namun

dengan presisi (0,79) yang sedikit di bawah SVM. Keseluruhan, kedua model menunjukkan performa yang kompetitif, meskipun dengan *trade-off* yang berbeda dalam distribusi metrik antar kelas.



Gambar 19. *Confusion Matrix* Hasil Pengujian Data Tokopedia Menggunakan TF-IDF dan *Random Forest*

Berdasarkan *confusion matrix* (Gambar 19), model *Random Forest* dengan fitur TF-IDF menunjukkan performa klasifikasi yang cukup baik. Pada kelas Negatif, 104 dari 114 data berhasil diklasifikasikan dengan benar, dengan 10 data salah diprediksi sebagai Netral dan tidak ada yang salah sebagai Positif. Untuk kelas Netral, model berhasil mengklasifikasikan 176 dari 191 data secara akurat, sementara 10 salah sebagai Negatif dan 5 sebagai Positif. Kelas Positif menunjukkan performa yang cukup baik, dengan 150 dari 187 data diklasifikasikan benar, namun masih terjadi kesalahan ke Negatif (18) dan Netral (19). Secara keseluruhan, model memiliki akurasi tinggi pada kelas Netral, namun masih perlu ditingkatkan dalam membedakan antara Positif dan dua kelas lainnya.

Hasil dan Analisis Komparatif

Berdasarkan hasil pengujian komprehensif yang telah disajikan, analisis performa berbagai kombinasi metode ekstraksi fitur dan algoritma klasifikasi telah dilakukan pada dataset ulasan aplikasi Shopee dan Tokopedia. Dari observasi awal, dataset ulasan aplikasi Shopee secara konsisten menunjukkan kinerja model yang lebih optimal dibandingkan dengan dataset Tokopedia di sebagian besar konfigurasi yang diuji, tercermin dari nilai akurasi dan metrik evaluasi lainnya yang lebih tinggi. Fenomena ini dapat mengindikasikan adanya karakteristik linguistik, distribusi sentimen, atau tingkat kejelasan ekspresi pengguna yang lebih distingtif pada ulasan Shopee, yang memungkinkan model untuk belajar dan mengklasifikasikan sentimen secara lebih efektif. Oleh karena itu, pembahasan komparatif selanjutnya akan difokuskan secara mendalam pada hasil pengujian yang diperoleh dari dataset ulasan aplikasi Shopee, mengingat dataset ini merepresentasikan performa model yang paling superior dan ilustratif dalam konteks penelitian ini

Secara keseluruhan, algoritma SVM menunjukkan kinerja yang lebih unggul dibandingkan *Random Forest* dalam klasifikasi sentimen pada dataset ini, terlepas dari metode ekstraksi fitur yang digunakan. Ketika dikombinasikan dengan *Bag of Words* (BoW), SVM mencapai akurasi 0,90, dan dengan TF-IDF, akurasi 0,89. Bandingkan dengan *Random Forest* yang hanya mencapai akurasi 0,84 baik dengan BoW maupun TF-IDF. Perbedaan signifikan ini mengindikasikan bahwa SVM lebih efektif dalam membangun hyperplane pemisah yang optimal untuk data teks berdimensi tinggi yang dihasilkan oleh kedua metode ekstraksi fitur.

Dalam hal metode ekstraksi fitur, *Bag of Words* (BoW) secara umum menghasilkan akurasi yang sedikit lebih tinggi atau setara dibandingkan TF-IDF. Pada pasangan dengan SVM, BoW mengungguli TF-IDF (0,90 vs 0,89). Sementara itu, pada pasangan dengan *Random Forest*, kedua metode ekstraksi fitur menunjukkan akurasi yang identik (0,84). Hal ini menunjukkan bahwa, untuk karakteristik dataset ulasan ini, representasi

frekuensi kemunculan kata sederhana yang dihasilkan oleh BoW sudah cukup informatif, bahkan dalam beberapa kasus sedikit lebih unggul, dibandingkan dengan representasi yang mempertimbangkan bobot *invers* frekuensi dokumen.

Dari keempat kombinasi yang diuji, kombinasi *Bag of Words* (BoW) dengan SVM terbukti menjadi konfigurasi paling optimal, mencapai akurasi tertinggi sebesar 0,90 (90%). Performa ini juga didukung oleh nilai *F1-Score* yang kuat pada setiap kelas sentimen, menunjukkan keseimbangan yang baik antara *precision* dan *recall* dalam mengklasifikasikan sentimen positif, negatif, dan netral.

Rekomendasi

Hasil penelitian merekomendasikan agar Shopee mempertahankan kualitas layanan karena dominasi sentimen positif sebesar 60,6% menunjukkan kepuasan pengguna, terutama pada aspek kecepatan pengiriman, keaslian produk, dan harga kompetitif. Meski ulasan negatif relatif kecil (9,79%), Shopee perlu memperkuat penanganan keluhan dengan customer service yang lebih responsif, termasuk integrasi chatbot cerdas untuk merespons keluhan umum. Selain itu, dengan hasil terbaik BoW + SVM (akurasi 90%), Shopee disarankan mengimplementasikan sistem real-time sentiment analysis guna memantau opini pelanggan secara otomatis sehingga permasalahan dapat segera diatasi. Untuk menjaga loyalitas, Shopee juga dapat memperluas program engagement seperti *voucher*, *cashback*, dan *flash sale* agar kepuasan serta keterikatan pengguna semakin meningkat.

Sementara itu, rekomendasi pada Tokopedia adalah memperbaiki area yang banyak dikritisi, mengingat proporsi sentimen negatif cukup tinggi (23,3%), khususnya terkait lambatnya pengiriman, kualitas produk dari pihak ketiga, dan layanan refund sehingga diperlukan strategi perbaikan yang konkret. Sebagai *marketplace* murni, Tokopedia juga perlu meningkatkan transparansi dan kepercayaan dengan memperketat sistem verifikasi penjual serta memastikan informasi produk lebih jelas agar kualitas layanan lebih terjamin. Selain itu, hasil analisis yang menunjukkan akurasi SVM cukup seimbang (87–88%) dapat dimanfaatkan untuk membangun algoritma pemantauan sentimen otomatis sehingga tren ketidakpuasan pengguna dapat terdeteksi sejak dini. Untuk menjaga loyalitas, Tokopedia dapat meluncurkan program *customer recovery* berupa strategi *win-back* seperti pemberian *voucher* kompensasi atau *free* ongkir kepada pelanggan yang memberikan ulasan negatif agar persepsi mereka berbalik positif.

IV. KESIMPULAN

Penelitian ini telah membandingkan secara komprehensif efektivitas BoW dan TF-IDF dalam menganalisis sentimen ulasan Shopee dan Tokopedia menggunakan algoritma SVM dan RF. Hasil studi berhasil mengisi gap literatur dengan membuktikan bahwa konfigurasi BoW yang dipadukan dengan SVM adalah arsitektur paling optimal, mencapai akurasi 90% pada ulasan Shopee dan 88% pada ulasan Tokopedia. Kontribusi terpenting adalah penemuan ini menyediakan rekomendasi berbasis data yang spesifik untuk korpus ulasan *e-commerce* Indonesia, menunjukkan bahwa metode representasi fitur yang lebih sederhana pun dapat menghasilkan performa superior, yang berimplikasi pada efisiensi sistem analisis sentimen *real-time*.

DAFTAR PUSTAKA

- [1] H. Huang, A. A. Zavareh, and M. B. Mustafa, "Sentiment Analysis in E-Commerce Platforms: A Review of Current Techniques and Future Directions," *IEEE Access*, vol. 11, no. August, pp. 90367–90382, 2023, doi: 10.1109/ACCESS.2023.3307308.
- [2] K. Tri Putra, M. Amin Hariyadi, and C. Crysdiyan, "Perbandingan Feature Extraction Tf-Idf Dan Bow Untuk Analisis Sentimen Berbasis SVM," *J. Cahaya MAndalika*, p. 1449, 2023.
- [3] Mohamed Omar, "Comparative Analysis of Feature Selection Methods for Twitter Sentiment Classification," *J. Inf. Syst. Eng. Manag.*, vol. 10, no. 21s, pp. 330–338, Mar. 2025, doi: 10.52783/jisem.v10i21s.3331.
- [4] K. M. Suryaningrum, "Comparison of the TF-IDF Method with the Count Vectorizer to Classify Hate Speech," *Eng. Math. Comput. Sci. J.*, vol. 5, no. 2, pp. 79–83, May 2023, doi: 10.21512/emacsjournal.v5i2.9978.
- [5] R. Suryanti and T. Prasetyaningrum, "Perbandingan Metode TF-IDF dan Bag of Words dalam Analisis Sentimen Diet Kopi Americano di Media Sosial Twitter Menggunakan Naïve Bayes," *Technol. Sci.*, vol. 7, no. 1, 2025, doi: 10.47065/bits.v7i1.7244.

- [6] I. Septiana and D. Alita, "Perbandingan Random Forest dan SVM dalam Analisis Sentimen Quick Count Pemilu 2024," *J. Inform. J. Pengemb. IT*, vol. 9, no. 3, pp. 224–233, 2024, doi: 10.30591/jpit.v9i3.6640.
- [7] A. Gupta and D. Kamthania, "Study of Sentiment on Google Play Store Applications," *SSRN Electron. J.*, pp. 1–4, 2021, doi: 10.2139/ssrn.3833926.
- [8] Siti Mutmainah, Fathir, and Erin Eka Citra, "Improving the Accuracy of Social Media Sentiment Classification with the Combination of TF-IDF Method and Random Forest Algorithm," *Journix J. Informatics Comput.*, vol. 1, no. 1, pp. 30–40, 2025, doi: 10.63866/journix.v1i1.2.
- [9] F. Irwannia and A. H. Lubis, "Analisis Sentimen Produk Berdasarkan Review Pelanggan Shopee Menggunakan KNN Product Sentiment Analysis Based on Shopee Customer Reviews Using KNN," vol. 5, no. 2, pp. 239–249, 2025, doi: 10.34007/incoding.v5i2.865.
- [10] A. Hasan, Y. R. Ramadhan, and M. Minarto, "Sentiment Analysis of Telemedicine Applications on Twitter Using Lexicon-Based and Naive Bayes Classifier Methods," *J. Ris. Inform.*, vol. 5, no. 4, pp. 481–490, 2023, doi: 10.34288/jri.v5i4.244.
- [11] J. E. Br Sinulingga and H. C. K. Sitorus, "Analisis Sentimen Opini Masyarakat terhadap Film Horor Indonesia Menggunakan Metode SVM dan TF-IDF," *J. Manaj. Inform.*, vol. 14, no. 1, pp. 42–53, 2024, doi: 10.34010/jamika.v14i1.11946.
- [12] K. P. Harmandini and K. M. L., "Analysis of TF-IDF and TF-RF Feature Extraction on Product Review Sentiment," *Sinkron*, vol. 8, no. 2, pp. 929–937, 2024, doi: 10.33395/sinkron.v8i2.13376.
- [13] M. Zainottah, R. Saputra, Y. Servanda, and I. Rosita, "Critical Sentiment Analysis of Tokopedia Electronic Products Using SVM-Logistic & TF-IDF Ensemble Methods," *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, vol. 4, no. 3, 2025.
- [14] A. Ananta Firdaus, A. Id Hadiana, and A. Kania Ningsih, "Klasifikasi Sentimen pada Aplikasi Shopee Menggunakan Fitur Bag of Word dan Algoritma Random Forest," *Ranah Res. J. Multidiscip. Res. Dev.*, vol. 6, no. 5, pp. 1678–1683, 2024, doi: 10.38035/rrj.v6i5.994.
- [15] A. Syah, F. Nurdiansyah, and A. Y. Rahman, "Analisis Sentimen Aplikasi Shopee, Tokopedia, Lazada Dan Blibli Menggunakan Leksikon Dan Random Forest," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3S1, 2024, doi: 10.23960/jitet.v12i3s1.5155.
- [16] G. Kanugrahan, V. H. C. Putra, and Y. Ramdhani, "Analisis Sentimen Aplikasi Gojek Menggunakan SVM, Random Forest dan Decision Tree," *J. Infortech*, vol. 6, no. 2, pp. 171–178, 2024, doi: 10.31294/infortech.v6i2.24594.
- [17] T. Wahyudi *et al.*, "Klasifikasi Sentimen X-Twitter Perihal Pindahan Ibu Kota Indonesia Menggunakan Ekstraksi Fitur Tf-Idf Dan Metode Support Vector Machine (SVM)," *J. Teknol. Inf.*, vol. 18, no. 2, pp. 185–199, 2024, [Online]. Available: <https://doi.org/10.47111/JTIAvailableonlineathttps://e-journal.upr.ac.id/index.php/JTI>
- [18] M. Arief and N. A. Samsudin, "Hybrid Approach with VADER and Multinomial Logistic Regression for Multiclass Sentiment Analysis in Online Customer Review," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 12, pp. 311–320, 2023, doi: 10.14569/IJACSA.2023.0141232.
- [19] D. A. Fitri and D. Damayanti, "Komparasi Algoritma Random Forest Classifier Dan Support Vector Machine Untuk Sentimen Masyarakat Terhadap Pinjaman Online Di Media Sosial," *JIPi (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 9, no. 4, pp. 2018–2029, 2024, doi: 10.29100/jipi.v9i4.5608.
- [20] S. A. S. Mola, D. L. B. Baun, I. O. Nunes, and M. M. A. R. Sani, "Analisis Sentimen Aplikasi Halo Bca Di Google Play Store Menggunakan Metode Naive Bayes, Support Vector Machine Dan Random Forest," *HOAQ (High Educ. Organ. Arch. Qual. J. Teknol. Inf.)*, vol. 15, no. 2, pp. 69–79, 2024, doi: 10.52972/hoaq.vol15no2.p69-79.