



Analisis Sentimen Ulasan Aplikasi Shopee Menggunakan Algoritma Random Forest, Naïve Bayes, dan Support Vector Machine di Kota Semarang

Vinsent Brilian Adiguna¹, Ryan Arya Pramudya²

¹Program Studi Magister Teknik Informatika, Universitas Dian Nuswantoro, Semarang, Indonesia

²Fakultas Ekonomika dan Bisnis, Universitas 17 Agustus 1945 Semarang, Indonesia

Email : vbrilian27@gmail.com, ryanarya-pramudya@untagsmg.ac.id

Alamat : Jl. Imam Bonjol No.207, Pendrikan Kidul, Kec. Semarang Tengah, Kota Semarang, Jawa Tengah 50131

Jl. Pemuda No.70, Pandansari, Kec. Semarang Tengah, Kota Semarang, Jawa Tengah 50133

Korespondensi penulis : vbrilian27@gmail.com

Abstract

The growth of e-commerce in Indonesia has led to the emergence of various online shopping platforms, with Shopee being one of the most popular in Semarang City. User reviews on the Shopee application serve as a valuable data source for analyzing customer satisfaction levels; however, the large volume of data requires a systematic and accurate analytical approach. This study aims to analyze user review sentiments of the Shopee application using three machine learning algorithms: Random Forest, Naïve Bayes, and Support Vector Machine (SVM), as well as comparing the accuracy of these three algorithms. This research utilized 1000 reviews collected through web scraping from the Play Store, which were categorized into three classifications: positive, neutral, and negative sentiments. The analysis process encompassed pre-processing stages, feature extraction using TF-IDF, and classification using Random Forest, Naïve Bayes, and Support Vector Machine algorithms. The results demonstrated that the Random Forest algorithm achieved the highest accuracy at 96.19%, followed by Support Vector Machine with 95.71% accuracy, and Naïve Bayes with 84.76% accuracy. This research highlights the effectiveness of Random Forest and SVM in classifying user review sentiments towards the Shopee application.

Keywords: Random Forest, TF-IDF, Analisis sentiment, Support Vector Machine, Naive Bayes

Abstrak

Perkembangan e-commerce di Indonesia telah menciptakan berbagai platform belanja online, dengan Shopee sebagai salah satu yang terpopuler di Kota Semarang. Ulasan pengguna pada aplikasi Shopee menjadi sumber data yang berharga untuk menganalisis tingkat kepuasan pelanggan, namun volume data yang besar membutuhkan pendekatan analisis yang sistematis dan akurat. Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna aplikasi Shopee menggunakan tiga algoritma machine learning: Random Forest, Naïve Bayes, dan Support Vector Machine (SVM), serta membandingkan akurasi ketiga algoritma tersebut. Dalam penelitian ini data yang digunakan mencakup 1000 ulasan yang dikumpulkan melalui web scraping dari Play Store, dan dikelompokkan menjadi tiga kategori : sentiment positif, netral, dan negative. Proses analisis mencakup tahapan pre-processing, feature extraction menggunakan TF-IDF, serta di klasifikasikan menggunakan algoritma Random Forest, Naïve Bayes, dan Support Vector Machine. Hasil penelitian ini menunjukkan bahwa algoritma Random Forest memiliki akurasi tertinggi sebesar 96.19%, diikuti oleh Support Vector Machine dengan akurasi 95.71%, dan Naïve Bayes dengan akurasi sebesar 84.76%. Penelitian ini menunjukkan efektivitas Random Forest dan SVM dalam mengklasifikasikan sentimen ulasan pengguna terhadap aplikasi shopee

Kata kunci: Random Forest, TF-IDF, Analisis sentiment, Support Vector Machine, Naive Bayes

Received: November 15, 2024; Revised: December 02, 2024; Accepted: December 06, 2024; Online Available: January 31, 2025;

*Vinsent Brilian Adiguna, vbrilian27@gmail.com

1. LATAR BELAKANG

Electronic Commerce merupakan sebuah proses yang dilakukan antara ke dua belah pihak penjual dan pembeli melalui internet.(Khalaf & Diana, 2022) Banyak pelaku bisnis, baik skala besar maupun ritel, yang melakukan transisi atau pengembangan bisnisnya ke digital, terutama dalam bentuk E-commerce (Artanti et al., 2018). E-commerce mudah diakses kapan saja, di mana saja dalam bentuk website dan aplikasi smartphone, tidak dipungkiri bahwa E-commerce dapat mempermudah transaksi antara penjual dan pembeli.(Widagdo et al., 2020) Pembeli tidak harus langsung mendatangi penjual, daerah tidak menghalangi pembelian, toko online buka 24/7, tidak ada tawar menawar, dan banyak diskon, promosi, bahkan gratis ongkos kirim. Namun dengan kelebihan ini muncul beberapa kelemahan, di antaranya tidak jarang penipuan terjadi selama transaksi, penjual tidak mengirimkan produk yang dipesan, atau produk foto tidak sesuai dengan aslinya, paket rusak, atau warna tidak sesuai, Oleh karena itu, konsumen menggunakan media sosial untuk mengekspresikan perasaan mereka tentang layanan yang ditawarkan dan disediakan oleh perusahaan pasar online (Artanti et al., 2018). Banyak bermunculan E-commerce yang membuat masyarakat resah untuk menentukan E-commerce yang dipercaya, oleh karena itu dibutuhkan sebuah evaluasi yg bisa mengenali tingkatan kepopuleran E-commerce di indonesia bersumber pada analisa afeksi pada sosial alat yang dikatakan pada bacaan yang berbentuk pandangan positif ataupun minus (Widagdo et al., 2020).

Shopee menjadi marketplace yang paling banyak dikunjungi pengunjung saat ini di Indonesia rata-rata pengunjung bulanan shopee mencapai 173,9 M pada 2023 oleh iPrice2023. Shopee adalah aplikasi belanja dan jual beli yang dapat di unduh dari halaman Google Playstore. Halaman google playstore memiliki banyak pilihan fitur-fiturnya untuk peringkat dan ulasan. Peringkat dan fitur ini untuk pengguna menilai aplikasi.(Josua Josen A. Limbong et al., 2022) Komentar adalah teks yang mengomentari ide atau karya orang lain. Dengan bantuan ulasan pengguna lama, informasi tentang produk dan layanan yang baik menjadi semakin efektif dan banyak pengguna Internet mempercayai rekomendasi dan pendapat yang telah digunakan oleh pengguna sebelumnya dalam bentuk produk atau layanan (Cahyaningtyas et al., 2021). Belum ada metode yang sistematis dan tepat dimana suatu tinjauan dapat diklasifikasikan sebagai golongan positif

serta negatif. Informasi keterangan konsumen bersumber pada tipe afeksi mempermudah pengembang dalam memperoleh data afeksi dari para konsumen aplikasi e-commerce (Josua Josen A. Limbong et al., 2022).

Analisis sentiment merupakan bidang studi untuk menganalisis pendapat orang tentang entitas seperti layanan, produk, individu organisasi, masalah, peristiwa, dan topik. Analisis sentiment sendiri berfungsi untuk menemukan sebuah pola teks dan selanjutnya dikelompokkan kedalam sentiment positif atau negative. Analisis sentiment mempunyai tiga tahapan yaitu preprocessing, pembobotan, dan klasifikasi. (Lukmana et al., 2019)

Penelitian terkait analisis sentiment sebelumnya telah dilakukan mengenai opini review produk online menggunakan Naïve Bayes. Pada penelitian tersebut memperoleh nilai akurasi yang ditentukan oleh sistem analisis opini review produk online dengan metode Naïve Bayes menghasilkan nilai akurasi terendah pada pengujian 5 kelas. Menggunakan dataset training 80% dan dataset testing 20% didapatkan sebesar 52.66%, sedangkan pada pengujian 3 kelas menggunakan dataset training 90% dan dataset testing 10% memiliki akurasi tertinggi sebesar 77.78%. (Artanti et al., 2018) kemudian penelitian terhadap sentiment terhadap toko online di social media menggunakan metode klasifikasi Naïve Bayes. Hasil penelitian tersebut menyatakan Naïve Bayes dapat dijadikan metode klasifikasi untuk analisis sentiment dengan keakuratan 93.7%. (Artanti et al., 2018)

Penelitian dalam bidang analisis sentiment ini banyak dilakukan karena topik ini sangat menarik untuk dibahas. Salah satu penelitian di bidang ini adalah penelitian tentang analisis sentiment pelanggan toko online JD.ID menggunakan Naïve Bayes Classifier berbasis ikon emosi. Pada penelitiannya didapatkan nilai akurasi NBC tanpa pembobotan TF-IDF dan konversi ikon emosi memiliki nilai akurasi 98%. (Sari & Wibowo, 2019) Penelitian selanjutnya tentang analisis sentiment tentang pinjaman online menggunakan algoritma SVM mendapatkan nilai akurasi 84.4%. (Utami & Erfina, 2021)

Berdasarkan kasus sebelumnya penelitian ini menggunakan metode Naïve Bayes dan Support Vector Machine untuk melakukan analisis sentiment Algoritma Naïve Bayes dan Support Vector Machine umum dan mudah digunakan dalam proses klasifikasi data opini google playstore yang baru yang tidak memiliki label kelas untuk ditempatkan di kelas opini positif atau negative. Hasil penelitian aplikasi e-commerce shopee diharapkan

dapat memberikan hasil klasifikasi yang sangat baik dan akurat dan informasi yang bermanfaat bagi pihak yang membutuhkan. (SETIADI, 2022)

2. KAJIAN TEORITIS

Tinjauan pustaka ini bertujuan untuk membahas dua aspek penting yang berpengaruh terhadap keberlanjutan e-commerce, yaitu sentiment analysis yang dikenal sebagai opinion mining, dan klasifikasi yang bertujuan untuk mengkategorikan sentiment analysis yang akan diperoleh untuk sentiment netral, positif, dan negatif. Keduanya memiliki peranan signifikan dalam menentukan keberhasilan e-commerce dalam membuat keputusan dan evaluasi platform e-commerce.

1) Sentiment Analysis

Analisis sentimen, yang juga dikenal sebagai opinion mining, adalah cabang dari natural language processing (NLP) yang bertujuan untuk mengidentifikasi dan mengekstraksi pendapat atau perasaan yang terkandung dalam teks (Liu, 2012). Dalam era digital saat ini, analisis sentimen menjadi semakin penting karena banyaknya data teks yang dihasilkan melalui media sosial, ulasan produk, dan berbagai platform online lainnya. Analisis sentimen merupakan proses komputasional untuk mengidentifikasi dan mengkategorikan opini yang diekspresikan dalam teks, terutama untuk menentukan apakah sikap penulis terhadap suatu topik, produk, atau jasa bersifat positif, negatif, atau netral (Zhang et al., 2018). Proses ini melibatkan berbagai teknik pemrosesan bahasa alami dan pembelajaran mesin.

2) Klasifikasi

Klasifikasi merupakan salah satu teknik fundamental dalam pembelajaran mesin (machine learning) yang bertujuan untuk memprediksi label atau kategori dari suatu data berdasarkan fitur-fiturnya (Hierons, 1999). Dalam era big data, teknik klasifikasi menjadi semakin penting untuk berbagai aplikasi, mulai dari pengenalan pola hingga pengambilan keputusan otomatis. Adanya beberapa algoritma klasifikasi sebagai berikut Naive Bayes, Support Vector Machine, Random Forest

a) Naïve Bayes

Naive Bayes merupakan algoritma klasifikasi yang berbasis probabilistik, menggunakan teorema Bayes dengan asumsi independensi antar fitur. Meskipun asumsi ini sederhana, Naive Bayes menunjukkan performa yang baik untuk berbagai kasus, terutama dalam klasifikasi teks (Ji et al., 2011).

b) Support Vector Machine

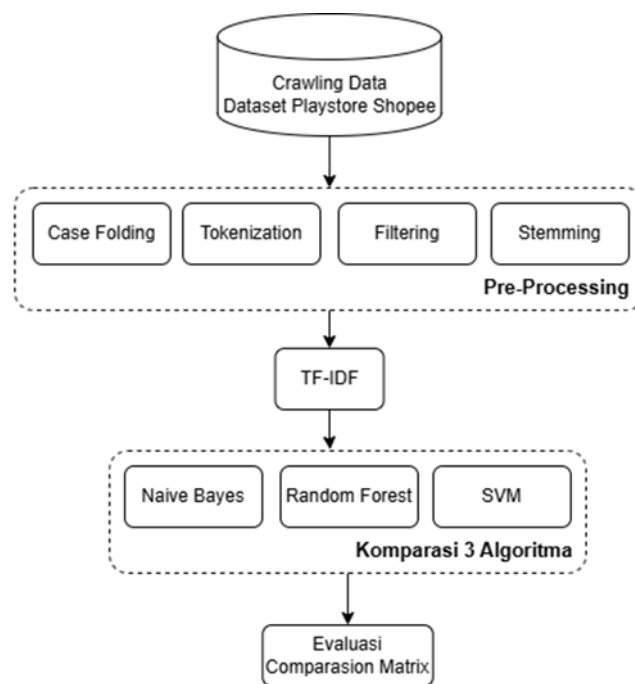
SVM bekerja dengan mencari hyperplane optimal yang memisahkan kelas-kelas data. (Wu et al., 2007) mendemonstrasikan bahwa SVM sangat efektif untuk klasifikasi dengan dimensi tinggi dan memberikan generalisasi yang baik.

c) Random Forest

Random Forest merupakan ensemble learning method yang mengkombinasikan multiple decision trees. (Breiman, 2001) menunjukkan bahwa metode ini dapat mengurangi overfitting dan meningkatkan akurasi klasifikasi.

3. METODE PENELITIAN

Dalam analisis sentimen yang berkaitan dengan ulasan pengguna Shopee tentang fungsinya sebagai platform e-commerce, beberapa tahapan dalam sentiment analysis, pertama dilakukan Crawling Data dari Play Store setelah itu dilakukan pre-processing data seperti case folding, tokenization, filtering, dan stemming. selanjutnya, akan dilakukan komparasi 3 algoritma yaitu Naïve Bayes, Random Forest, dan Support Vector Machine dan tahapan selanjutnya yang berpuncak pada evaluasi penilaian hasil klasifikasi menggunakan confusion matrix.



Gambar 1. Metodologi Penelitian

1) Populasi dan Sampling

Data Sampling ini melibatkan pada seluruh ulasan atau review pengguna yang terdapat pada platform e-commerce shopee. Populasi ini melibatkan 1000 dataset ulasan yang diambil melalui proses crawling data yang digunakan untuk proses analisis menggunakan API Playstore dari aplikasi Shopee. Berdasarkan penelitian terkait ada beberapa penelitian terdahulu yang dapat di jadikan stat of the art

Tabel 1. Penelitian Terkait

Author's	Dataset	Classifier	Best Accuracy
(Widagdo et al., 2020)	Review Produk	Naïve Bayes	77.78%
(Artanti et al., 2018)	Review Toko Online Social Media	Naïve Bayes	93.7%
(Sari & Wibowo, 2019)	Review Aplikasi Jd.ID	SVM	84.4%

2) Data Collection

Tahapan awal ini melibatkan crawling data atau sering disebut pengumpulan data. Crawling data mengacu pada proses sistematis pengumpulan data yang bersumber dari situs web atau platform media sosial melalui penggunaan antarmuka pemrograman aplikasi (API) yang disediakan oleh situs web atau platform media sosial masing-masing. Crawling data mencakup identifikasi dan ekstraksi data yang berkaitan dengan entitas tertentu. Keamanan web biasanya digunakan untuk tujuan pengindeksan dan untuk entitas tertentu berdasarkan struktur grafik tautan web (Bailey et al., 2003). Crawling data digunakan untuk memperoleh dataset yang digunakan untuk proses analisis menggunakan API Playstore. Pada penelitian ini menggunakan 1000 dataset ulasan yang diambil dari aplikasi E-Commerce Shopee pada website google playstore, kemudian dikelompokkan menjadi tiga kelas yaitu ulasan positif, netral, ulasan negative.

3) Pengujian Instrumen Penelitian

a. Data Pre_Processing

Prosedur pra-pemrosesan data memerlukan penghapusan yang cermat dari elemen-elemen asing seperti kalimat, kata, tautan, atau objek yang dianggap tidak perlu untuk pelaksanaan analisis sentimen. Kumpulan data yang diperoleh melalui perayapan web biasanya bermanifestasi sebagai bentuk tekstual yang tidak teratur atau tidak terstruktur, memerlukan prosedur sistematis yang dikenal sebagai prapemrosesan data untuk menghasilkan data yang disanitasi, sehingga meningkatkan keakuratan hasil analisis sentimen. Tahapan yang terlibat dalam prapemrosesan data meliputi Case Folding, Tokenization, Filtering Stopword, Stemming.

- Case Folding

Pada tahapan ini keseluruhan bentuk huruf di seragamkan menjadi lowercase agar lebih memudahkan dalam pencarian. Cara ini memudahkan dalam pencarian, dan disebabkan tidak berubah-ubah dengan huruf kapital (A. S. Rahayu et al., 2022)

- **Tokenization**

Tokenisasi merupakan tahapan pemrosesan teks yang mengurai kalimat ke dalam unit-unit kata individual untuk meningkatkan interpretabilitas makna. Proses ini bertujuan untuk menghasilkan segmentasi kata yang kemudian akan diklasifikasikan sebagai entitas dan dihitung nilai metrik dokumennya untuk keperluan analisis lebih lanjut (I. P. Rahayu et al., 2022).

- **Filtering Stopwords**

Pada tahapan (Wagner, 2010) ini proses membuang kata-kata kurang penting atau menyimpang dari kosa kata yang tidak memiliki arti. Tahapan ini dilakukan untuk mengurangi pemilihan kata dalam dokumen teks dengan menghilangkan kata-kata yang sering muncul namun tidak informatif, yang disebut sebagai stopwords (A. S. Rahayu et al., 2022)

- **Stemming**

Stemming merupakan tahapan esensial yang mentransformasi kata berimbuhan menjadi kata dasar melalui eliminasi afiks, yang mencakup prefiks, infiks, sufiks, serta konfiks (I. P. Rahayu et al., 2022). Keakuratan proses stemming memiliki dampak signifikan terhadap validitas hasil analisis, mengingat perannya sebagai komponen fundamental dalam pengolahan teks.

b. TF_IDF

Pasca tahap preprocessing, ekstraksi fitur diimplementasikan menggunakan TF-IDF, suatu teknik pembobotan kata yang telah teruji dalam ranah pemrosesan teks. Metode ini mengkalkulasi bobot setiap terma berdasarkan frekuensi kemunculannya dalam dokumen dan korpus (Wagner, 2010). Sebagai metode ekstraksi fitur yang telah mapan dalam Natural Language Processing, TF-IDF mengintegrasikan dua komponen fundamental: Term Frequency (TF) yang mengkuantifikasi intensitas kemunculan kata dalam suatu dokumen, dimana frekuensi yang lebih tinggi menghasilkan nilai yang lebih besar, serta Inverse Document Frequency (IDF) yang mengevaluasi signifikansi kata dalam

keseluruhan korpus, dengan premis bahwa kata yang jarang muncul memiliki nilai kepentingan lebih tinggi. Dalam implementasinya, TF-IDF mengalokasikan pembobotan yang lebih substansial pada kata-kata yang memiliki frekuensi tinggi dalam dokumen spesifik namun jarang ditemukan dalam dokumen lain pada korpus yang sama (Salton et al., 1997). Spesifik dalam konteks analisis sentimen ulasan film, implementasi TF-IDF berperan vital dalam identifikasi kata-kata kunci yang berkorelasi kuat dengan sentimen positif atau negatif, sehingga berkontribusi pada peningkatan presisi klasifikasi (Pang et al., 2002).

c. **Klasifikasi**

Salah satu tugas mendasar dalam machine learning melibatkan prediksi label kategoris yang sesuai dengan input tertentu. Dalam analisis sentimen yang berkaitan dengan ulasan film, tujuannya adalah untuk mengklasifikasikan ulasan ini ke dalam kategori sentimen positif atau negatif yang berbeda (Chiu, 2001). Algoritma Naive Bayes merupakan metode klasifikasi probabilistik yang didasarkan pada Teorema Bayes, yang mengemukakan asumsi independensi di antara fitur-fitur; terlepas dari sifat asumsi ini yang agak sederhana, ini sering menunjukkan kemanjuran yang cukup besar dalam aplikasi praktis, terutama di bidang klasifikasi teks (Manning, 2006). Dalam konteks analisis sentimen, pendekatan Naïve Bayes menghitung probabilitas yang terkait dengan setiap kata dalam ulasan untuk memastikan keseluruhan sentimen yang disampaikan (Liu, 2022). Algoritma Support Vector Machine (SVM) diakui karena kemajuannya dalam klasifikasi teks dan analisis sentimen, menunjukkan keserbagunaan dalam penerapan fungsi kernel sambil mengurangi risiko overfitting (Joachims, 1998). Selain itu, SVM mahir dalam mengelola kompleksitas linguistik dan biasanya menghasilkan tingkat akurasi yang tinggi (Joachims, 1998).

Algoritma Random Forest mewujudkan teknik ensemble yang memanfaatkan beberapa pohon keputusan untuk menjalankan tugas klasifikasi. Setiap pohon menghasilkan prediksi, dan kategori yang menerima mayoritas suara ditetapkan sebagai output dari model (Wetteland et al., 2019). Random Forest mampu mengakomodasi banyak fitur tanpa menyerah pada kelebihan pemasangan, dan mahir dalam mengidentifikasi pola rumit dalam data tekstual,

seringkali mencapai kinerja yang terpuji (Tsoumakas & Katakis, 2007). Pada titik ini, data tinjauan film pra-proses akan diklasifikasikan menurut klasifikasi sentimen positif dan negatifnya. Selama proses klasifikasi, tiga algoritma machine learning yang berbeda digunakan, khususnya Naïve Bayes, Support Vector Machine, dan Random Forest.

4. HASIL PENELITIAN DAN PEMBAHASAN

Hasil Penelitian

1. Data Collection

Pengumpulan data dalam penelitian ini memakai metode crawling data dari keterangan aplikasi shopee. Informasi yang diperoleh dalam format .csv dan berjumlah 1000 data. Mengambil hasil penelitian dataset seperti gambar 2.

Gambar 2. Data Collection

reviewid	userName	userImage	content	score	thumbsUpCount	reviewCreatedVersion	at	replyContent	repliedAt
3ad790a4-519a-4157-9e65-82c8912d245	Chan Haruma	https://play-lh.googleusercontent.com/a-/AD_cM...	Notifikasi sering sangat lambat pesanan masuk ...	2	410	3.05.09	2023-07-17 00:31:33	Hi kak maaf ya ats kendalanya terkait kendal...	2023-07-17 01:17:36
24fefeb2-c8de-4400-a286-5c077556eb89	Bocil share	https://play-lh.googleusercontent.com/a-/AD_cM...	Makin kesini makin ga jelas aplikasi nya, kada...	1	346	3.05.09	2023-07-17 19:06:41	Hi kak mohon maaf atas ketidaknyamanya, perih...	2023-07-17 20:59:44
aa5c099b-8b2d-4cc3-8c85-1c345cbb3bdf	Epri Yadi	https://play-lh.googleusercontent.com/a-/AD_cM...	Tolong untuk pihak Shopee.Tolong untuk di eval...	1	3404	3.04.20	2023-06-29 13:39:11	Hi kak, maaf ya Terkait kendala, mimin saran...	2023-06-29 14:18:27
b27a4652-5055-4507-a8c5-18a22e06149a	japhe methe	https://play-lh.googleusercontent.com/a-/AD_cM...	Untuk sekarang ini terlalu banyak tambahan bia...	3	7	3.05.09	2023-07-18 14:28:36	Hi kak, maaf ya udi buat km cemas. Terkait ken...	2022-07-27 21:59:20
8632a011-54ff-46d5-82f1-cdb030e890e7	Ahmad anwar sadad	https://play-lh.googleusercontent.com/a-/AD_cM...	Aplikasinya sebenarnya bagus dengan berbagai p...	2	1318	3.04.20	2023-07-10 17:08:13	Hi ka, mohon maaf buat tdk nyaman, terkait pel...	2023-07-10 18:46:36

2. Pre-Processing

Setelah data berhasil di scrapping menggunakan playstore, tahapan selanjutnya akan di pre-processing dan di seleksi fitur atau atribut yang kurang bermakna, tahapan pre-processing disini akan ada 4 tahapan yaitu case folding, tokenization, tokenizing, filtering stopwords, dan stemming.

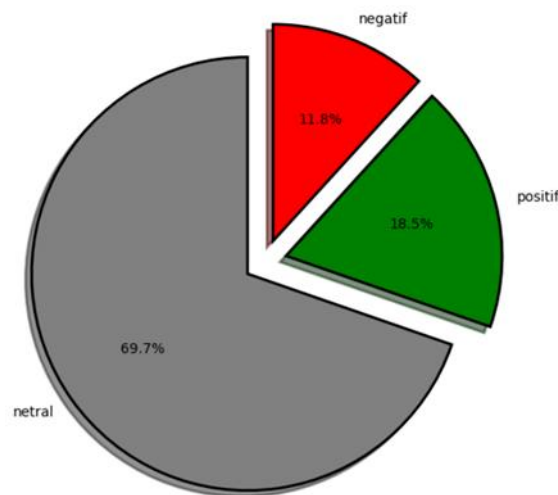
Tabel 2. Hasil Pre-Processing

Proses	Hasil
Data Ulasan	Mudah dan murah, pilihan barangnya juga banyak
Case Folding	mudah dan murah pilihan barangnya juga banyak
Tokenization	"mudah", "dan", "murah", "pilihan", "barangnya", "juga", "banyak".
Filtering Stopwords	"mudah", "murah", "pilihan", "barangnya".
Stemming	"mudah", "murah", "pilih", "barang"

Sumber: Data primer diolah, 2024.

3. Sentiment Analysis

Setelah dilakukan pre-processing proses selanjutnya dapat mengetahui 3 kategori sentiment yaitu sentiment positif, sentiment netral, sentiment negatif dengan hasil sebagai pada gambar 2.



Gambar 3. Sentiment Analysis

Berdasarkan pada gambar 3. Sentiment netral memiliki 69.7% dan sentiment negatif mempunyai 11.8%, dan untuk sentiment positif mendapatkan hasil 18.9% dapat disimpulkan bahwa hasil dari shopee memiliki sentiment netral.

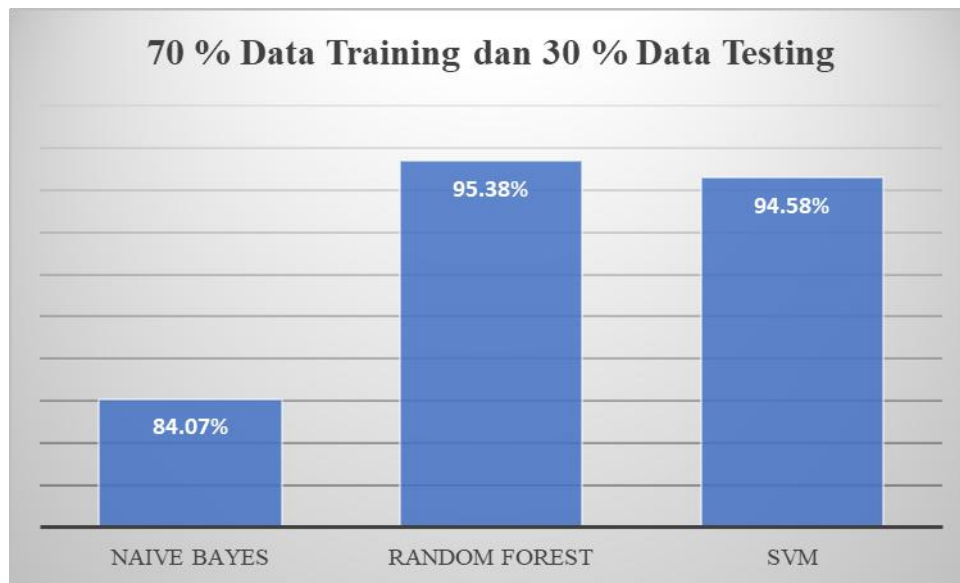
4. Klasifikasi

Setelah sentiment analysis sudah di temukan hasilnya maka akan dilakukan komparasi tiga algoritma klasifikasi menggunakan Naive Bayes, Random Forest, dan Support Vector Machine dengan dataset 1000 dibagi menjadi tiga kali eksperimen yang pertama menggunakan data training 90% dan data testing 10%, data training 80% dan data testing 20%, dan terakhir menggunakan data training 70% dan data testing 30%. dan di validation menggunakan cross-validation dengan k-fold 10.

Tabel 3. Hasil Klasifikasi 70% Data Training, 30% Data Testing

Data Training	Data Testing	<i>Naive Bayes</i>	<i>Random Forest</i>	<i>SVM</i>
70 %	30 %	84.07 %	95.38 %	94.58 %

Sumber: Data primer diolah, 2024.

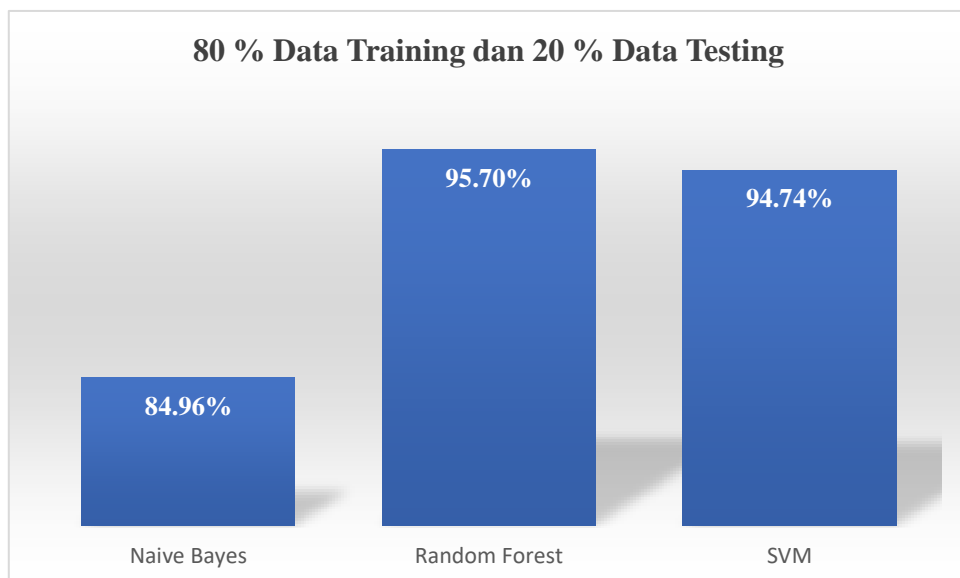


Gambar 4. Hasil Klasifikasi 70% Data Training, 30% Data Testing

Berdasarkan gambar 4. dan tabel 3. menunjukkan bahwa algoritma Random Forest mendapatkan akurasi tertinggi 96.19% dan SVM 95.71% dan naive bayes hanya mendapatkan hasil 84.76% saja.

Tabel 4. Hasil Klasifikasi 80% Data Training, 20% Data Testing

Data Training	Data Testing	Naive Bayes	Random Forest	SVM
80 %	20 %	84.96 %	95.70 %	94.74 %

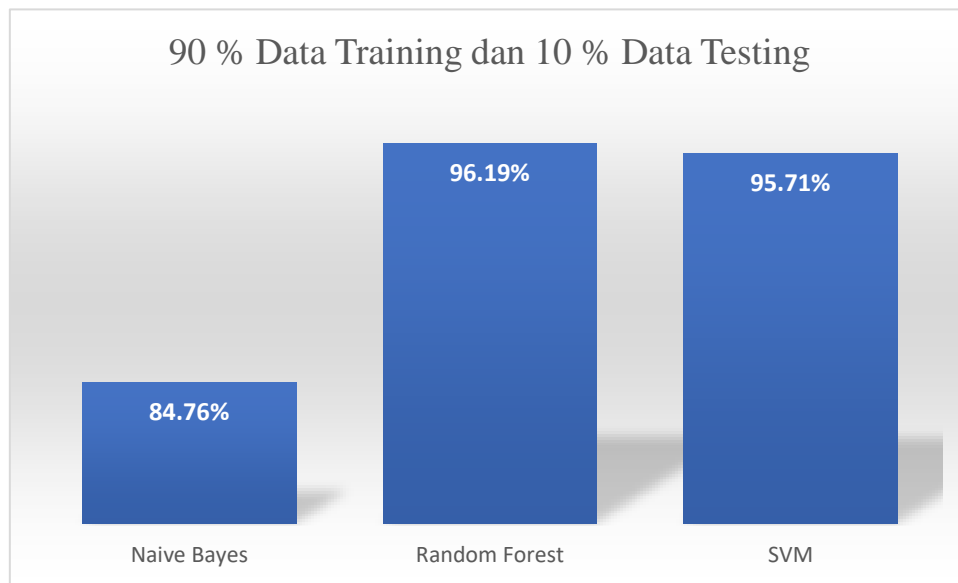


Gambar 5. Hasil Klasifikasi 80% Data Training, 20% Data Testing

Berdasarkan gambar 5. dan tabel 4. menunjukkan bahwa algoritma Random Forest mendapatkan akurasi tertinggi 95.70% dan SVM 94.74% dan Naive Bayes hanya mendapatkan hasil 84.96% saja.

Tabel 5. Hasil Klasifikasi 90% Data Training, 10% Data Testing

Data Training	Data Testing	Naive Bayes	Random Forest	SVM
90 %	10 %	84.76 %	96.19 %	95.71 %



Gambar 6. Hasil Klasifikasi 90% Data Training, 10% Data Testing

Berdasarkan gambar 6. dan tabel 5. menunjukkan bahwa algoritma Random Forest mendapatkan akurasi tertinggi 96.19% dan SVM 95.71% dan Naive Bayes hanya mendapatkan hasil 84.76% saja.

Pembahasan

Berdasarkan hasil penelitian diatas bahwa rekomendasi yang digunakan adalah 90 % data training dan 10% data testing karena mendapatkan hasil akurasi yang tertinggi yaitu algoritma Random Forest 96.17% dan SVM mendapatkan akurasi sebesar 95.71%, dan algoritma Naive Bayes juga mendapatkan hasil yang cukup yaitu 84.76%.

Tabel 6. Purposed Method

Author's	Dataset	Classifier	Best Accuracy
(Widagdo et al., 2020)	Review Produk	Naïve Bayes	77.78%
(Artanti et al., 2018)	Review Toko Online Social Media	Naïve Bayes	93.7%
(Sari & Wibowo, 2019)	Review Aplikasi Jd.ID	SVM	84.4%
Purposed Method	Review Aplikasi Shopee	Naïve Bayes, Random Forest, SVM	84.76%, 96.19%, 95.71%

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian menunjukkan akurasi yang lebih tinggi dibandingkan dengan penelitian sebelumnya. Pada penelitian sebelumnya, metode Naïve Bayes mencapai akurasi terbaik sebesar 93.7%, sedangkan metode yang diusulkan menggunakan kombinasi Naïve Bayes, Random Forest, dan SVM yang menghasilkan akurasi tertinggi sebesar 96.19% Random Forest dan 95.71% SVM. Hal ini menunjukkan bahwa metode yang diusulkan memiliki kinerja yang lebih unggul dalam klasifikasi pada dataset yang digunakan. Hal ini dikarenakan Random Forest memiliki kemampuan yang lebih baik dalam menangkap pola dan hubungan yang kompleks pada data teks, sehingga menghasilkan klasifikasi sentimen yang lebih akurat, namun perlu diingat bahwa pemilihan algoritma terbaik dapat berbeda-beda tergantung dari jenis data dan jumlah data.

Rekomendasi

Bagi peneliti selanjutnya diharapkan dapat menggabungkan beberapa metode atau ensemble learning mungkin dapat meningkatkan akurasi lebih lanjut dan melakukan hyperparameter tuning pada setiap model untuk membantu mencapai kinerja yang optimal, sehingga akan memberikan hasil yang lebih baik untuk dapat memprediksi kinerja akurasi nya yang lebih baik di masa depan.

DAFTAR REFERENSI

- Artanti, D. P., Syukur, A., Prihandono, A., & Setiadi, D. R. I. M. (2018). Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naïve Bayes. 8–9.
- Bailey, P., Craswell, N., & Hawking, D. (2003). Engineering a multi-purpose test collection for Web retrieval experiments. *Information Processing and Management*, 39(6), 853–871. [https://doi.org/10.1016/S0306-4573\(02\)00084-5](https://doi.org/10.1016/S0306-4573(02)00084-5)
- Breiman, L. (2001). No Title. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/a:1010933404324>
- Cahyaningtyas, C., Nataliani, Y., & Wideasari, I. R. (2021). Analisis Sentimen Pada Rating Aplikasi Shopee Menggunakan Metode Decision Tree Berbasis SMOTE. *Aiti*, 18(2), 173–184. <https://doi.org/10.24246/aiti.v18i2.173-184>
- Chiu, D. K. Y. (2001). BOOK REVIEW: “PATTERN CLASSIFICATION”, R. O. DUDA, P. E. HART and D. G. STORK, Second Edition. *International Journal of Computational Intelligence and Applications*, 01(03), 335–339. <https://doi.org/10.1142/s1469026801000251>

- Hierons, R. (1999). *Machine learning*. Tom M. Mitchell. Published by McGraw-Hill, Maidenhead, U.K., International Student Edition, 1997. ISBN: 0-07-115467-1, 414 pages. Price: U.K. £22.99, soft cover. *Software Testing, Verification and Reliability*, 9(3), 191–193. [https://doi.org/10.1002/\(sici\)1099-1689\(199909\)9:3<191::aid-stvr184>3.0.co;2-e](https://doi.org/10.1002/(sici)1099-1689(199909)9:3<191::aid-stvr184>3.0.co;2-e)
- Ji, Y., Yu, S., & Zhang, Y. (2011). A novel Naive Bayes model: Packaged Hidden Naive Bayes. In 2011 6th IEEE Joint International Information Technology and Artificial Intelligence Conference (pp. 484–487). IEEE. <https://doi.org/10.1109/itaic.2011.6030379>
- Joachims, T. (1998). Text categorization with Support Vector Machines: Learning with many relevant features. In *Lecture Notes in Computer Science* (pp. 137–142). Springer Berlin Heidelberg. <https://doi.org/10.1007/bfb0026683>
- Josua Josen A. Limbong, Irwan Sembiring, & Kristoko Dwi Hartomo. (2022). Analisis Klasifikasi Sentimen Ulasan pada E-Commerce Shopee Berbasis Word Cloud Dengan Metode Naive Bayes dan K-Nearest Neighbor. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 9(2), 347–356. <https://doi.org/10.25126/jtiik.202294960>
- Khalaf, R., & Diana, A. (2022). Perancangan Sistem E-Commerce Dengan Menggunakan Business Model Canvas (BMC) Untuk Penjualan Pakaian Pada Exsthrift_. Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI), September, 161–170.
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Synthesis Lectures on Human Language Technologies, 5(1), 1–167. <https://doi.org/10.2200/s00416ed1v01y201204hlt016>
- Liu, B. (2022). *Sentiment analysis and opinion mining*. books.google.com. https://books.google.com/books?hl=en&lr=&id=xYhyEAAAQBAJ&oi=fnd&pg=PP1&dq=sentiment+analysis+text+mining&ots=rk_wPDN2Cz&sig=HuNSIv63bnJHRPznM0RkJ456gDo
- Lukmana, D. T., Subanti, S., & Susanti, Y. (2019). Analisis Sentimen Terhadap Calon Presiden 2019 Dengan Support Vector Machine Di Twitter. *Seminar Nasional Penelitian Pendidikan Matematika (SNP2M) 2019 UMT*, 2002, 154–160.
- Manning, C. D. (2006). *Natural Language Processing, Statistical Approaches to*. In *Encyclopedia of Cognitive Science*. Wiley. <https://doi.org/10.1002/0470018860.s00080>
- Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: sentiment classification using machine learning techniques. *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - EMNLP '02*. <https://doi.org/10.3115/1118693.1118704>
- Rahayu, A. S., Fauzi, A., & Rahmat, R. (2022). Komparasi Algoritma Naïve Bayes Dan Support Vector Machine (SVM) Pada Analisis Sentimen Spotify. *Jurnal Sistem Komputer Dan Informatika (JSON)*, 4(2), 349. <https://doi.org/10.30865/json.v4i2.5398>

- Rahayu, I. P., Fauzi, A., & Indra, J. (2022). Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Naive Bayes Dan Support Vector Machine. *Jurnal Sistem Komputer Dan Informatika (JSON)*, 4(2), 296. <https://doi.org/10.30865/json.v4i2.5381>
- Salton, G., Singhal, A., Mitra, M., & Buckley, C. (1997). Automatic text structuring and summarization. *Information Processing & Management*, 33(2), 193–207. [https://doi.org/10.1016/s0306-4573\(96\)00062-3](https://doi.org/10.1016/s0306-4573(96)00062-3)
- Sari, F. V., & Wibowo, A. (2019). Analisis Sentimen Pelanggan Toko Online Jd.Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi. *Jurnal SIMETRIS*, 10(2), 681–686. <https://jurnal.umk.ac.id/index.php/simet/article/view/3487/1883>
- SETIADI, K. (2022). Analisis Sentimen Pelanggan Terhadap Layanan ShopeeFood Pada Media Sosial Twitter Menggunakan Algoritma Naïve Bayes dan Support Vector Machine (SVM). [https://repository.mercubuana.ac.id/69948/%0Ahttps://repository.mercubuana.ac.id/69948/2/01 Cover.pdf](https://repository.mercubuana.ac.id/69948/%0Ahttps://repository.mercubuana.ac.id/69948/2/01%20Cover.pdf)
- Tsoumakas, G., & Katakis, I. (2007). Multi-Label Classification: An Overview. *International Journal of Data Warehousing and Mining*, 3(3), 1–13. <https://doi.org/10.4018/jdwm.2007070101>
- Utami, D. S., & Erfina, A. (2021). Analisis Sentimen Pinjaman Online di Twitter Menggunakan Algoritma Support Vector Machine (SVM). *SISMATIK (Seminar Nasional Sistem Informasi Dan Manajemen Informatika)*, 1(1), 299–305.
- Wagner, W. (2010). Steven Bird, Ewan Klein and Edward Loper: *Natural Language Processing with Python, Analyzing Text with the Natural Language Toolkit*: O'Reilly Media, Beijing, 2009, ISBN 978-0-596-51649-9. *Language Resources and Evaluation*, 44(4), 421–424. <https://doi.org/10.1007/s10579-010-9124-x>
- Wetteland, R., Engan, K., Eftestøl, T., Kvikstad, V., & Janssen, E. (2019). Multiclass Tissue Classification of Whole-Slide Histological Images using Convolutional Neural Networks. *Proceedings of the 8th International Conference on Pattern Recognition Applications and Methods*. <https://doi.org/10.5220/0007253603200327>
- Widagdo, A. S., W.A, B. S., & Nasiri, A. (2020). Analisis Tingkat Kepopuleran E-Commerce Di Indonesia Berdasarkan Sentimen Sosial Media Menggunakan Metode Naïve Bayes. *Jurnal Informa: Jurnal Penelitian Dan Pengabdian Masyarakat*, 6(1), 1–5. <https://doi.org/10.46808/informa.v6i1.159>
- Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Yu, P. S., Zhou, Z.-H., Steinbach, M., Hand, D. J., & Steinberg, D. (2007). Top 10 algorithms in data mining. *Knowledge and Information Systems*, 14(1), 1–37. <https://doi.org/10.1007/s10115-007-0114-2>
- Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. ... *Mining and Knowledge Discovery*. <https://doi.org/10.1002/widm.1253>