



Perbandingan Analisis Sentimen Untuk Prediksi Kepuasan Ulasan Produk Kopi Pada Media Sosial Menggunakan Algoritma Svm Dan Naïve Bayes

Trisena Pramuja^{1*}, Bambang Irawan²

^{1*,2} Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhadi Setiabudi

Email : Senapramuja0905@gmail.com

Abstract:

The development of social media has led to a significant increase in the number of consumer reviews of various types of products, including coffee products. To help manufacturers understand consumer satisfaction levels more efficiently, sentiment analysis is a relevant method because it is able to identify opinions automatically. This study compares the performance of two widely used algorithms, namely Support Vector Machine (SVM) and Multinomial Naive Bayes (MNB), in predicting sentiment on consumer reviews related to coffee products on social media. The dataset was analyzed through the stages of text cleanup, TF-IDF transformation, and label encoding process. Both models are developed using a uniform pipeline with consistent parameters to ensure an objective performance comparison. The results show that SVM algorithms with linear kernels produce the highest accuracy compared to Naive Bayes. In addition, a confusion matrix is applied to evaluate the accuracy of predictions in each sentiment category. These findings confirm that SVM is more effective in short-text-based sentiment analysis tasks, such as product reviews on social media platforms.

Keywords: Sentiment Analysis, SocialMedia, SVM, Naive Bayes, Coffee.

1. PENDAHULUAN

(Kabir Muhammad & Putra, 2025) Analisis sentimen merupakan salah satu topik penting dalam bidang pemrosesan bahasa alami (Natural Language Processing) dan data mining yang terus berkembang pesat di era digital saat ini. Berbagai platform daring seperti e-commerce, media sosial, dan forum diskusi menyediakan ribuan hingga jutaan ulasan produk dan opini publik yang dapat dimanfaatkan sebagai sumber data untuk mengetahui persepsi, kepuasan, atau kritik dari pengguna. Analisis sentimen terhadap ulasan produk e-commerce telah berkembang menjadi domain penelitian yang sangat vital dalam natural language processing, mengingat kemampuannya untuk mengekstrak wawasan bisnis yang actionable dari data tekstual yang tidak terstruktur.

(A & R, 2022) Dalam konteks produk kopi, memahami sentimen konsumen merupakan aspek penting untuk meningkatkan kualitas produk, memperbaiki layanan, serta merumuskan strategi pemasaran yang lebih tepat sasaran. Namun demikian, penelitian yang secara khusus membandingkan performa SVM dan Naive Bayes dalam memprediksi sentimen terkait kepuasan konsumen terhadap produk kopi di media sosial masih terbatas. Oleh sebab itu, penelitian ini bertujuan untuk mengevaluasi kinerja kedua algoritma tersebut secara objektif dengan menggunakan dataset ulasan konsumen yang telah melalui rangkaian proses prapemrosesan teks.

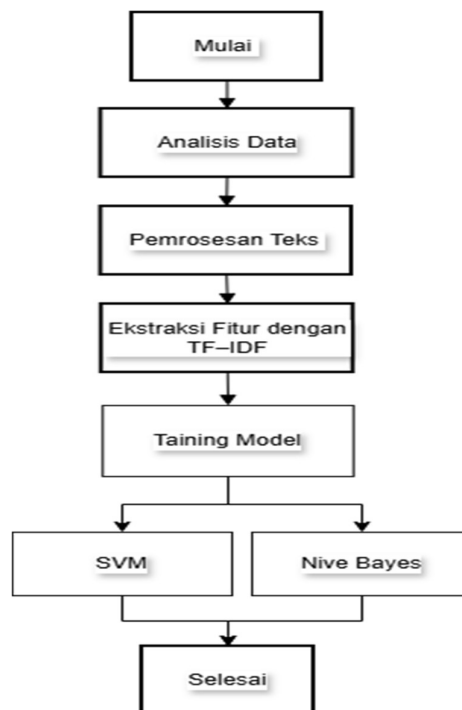
(Y & F, 2023) Penelitian ini diharapkan dapat memberikan kontribusi terhadap pengembangan metode analisis sentimen pada data ulasan produk, khususnya dalam

industri kopi. Selain itu, hasil yang diperoleh dapat menjadi referensi bagi peneliti maupun pelaku industri dalam menentukan algoritma yang paling tepat untuk menganalisis opini konsumen berbasis teks pendek pada platform media sosial.

2. METODOLOGI PENELITIAN

Metode penelitian ini dirancang secara sistematis untuk menjamin bahwa proses analisis sentimen dilakukan secara objektif dan dapat direplikasi. Penelitian ini menerapkan pendekatan *supervised learning* untuk mengklasifikasikan sentimen ulasan produk kopi pada media sosial ke dalam tiga kategori, yaitu positif, negatif, dan netral. Penelitian ini dilakukan melalui serangkaian tahapan sistematis, dimulai dari studi literatur untuk memahami konsep analisis sentimen, (G & T, 2024) algoritma Support Vector Machine (SVM) dan Naïve Bayes, serta implementasi metode pada industri kopi. Data ulasan pelanggan dikumpulkan dari platform e-commerce Tokopedia dan melalui tahap preprocessing untuk memastikan kualitas data sebelum dianalisis. Selanjutnya, dataset dibagi menjadi data latih dan data uji, yang kemudian digunakan untuk membangun serta mengevaluasi model berbasis SVM dan Naïve Bayes.

Data yang digunakan berupa ulasan konsumen dalam bentuk teks pendek yang bersifat tidak terstruktur. Oleh karena itu, tahapan penelitian meliputi proses pra-pemrosesan teks untuk meminimalkan noise, ekstraksi fitur menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF), serta penerapan algoritma Support Vector Machine (SVM) Linear dan Multinomial Naive Bayes sebagai model klasifikasi.



Gambar 1 flowchart

2.1 Analisis Sentimen

Analisis sentimen merupakan bidang dalam *Natural Language Processing* (NLP) yang bertujuan mengidentifikasi opini atau emosi yang terkandung dalam teks. Sentimen

biasanya dikategorikan menjadi positif, negatif, dan netral. Teknologi ini banyak digunakan pada data media sosial, ulasan produk, dan platform digital lainnya, sehingga mampu memberikan wawasan mengenai persepsi dan pengalaman konsumen. Berbagai penelitian terdahulu menunjukkan bahwa analisis sentimen berbasis *supervised learning* mampu mencapai performa tinggi apabila didukung oleh data yang sudah diberi label dan diproses dengan baik. (G & T, 2024)

2.2 Pemrosesan Teks (Text Preprocessing)

Pemrosesan teks merupakan langkah penting dalam analisis sentimen karena teks mentah biasanya mengandung berbagai elemen yang tidak relevan seperti emotikon, tanda baca, huruf kapital tidak teratur, URL, serta duplikasi kata. Menurut beberapa studi, proses normalisasi yang meliputi *lowercasing*, penghapusan karakter non-alfanumerik, pembersihan spasi, dan *stopword removal* dapat meningkatkan kualitas fitur serta mengurangi noise pada data teks. Hasil dari tahapan ini adalah representasi teks yang lebih bersih, konsisten, dan siap digunakan pada proses ekstraksi fitur.

2.3 Ekstraksi Fitur dengan TF-IDF

(R & E, 2023) Salah satu teknik representasi teks yang paling banyak digunakan dalam penelitian analisis sentimen adalah *Term Frequency–Inverse Document Frequency* (TF-IDF). Metode ini mengukur seberapa penting sebuah kata dalam satu dokumen dibandingkan dengan seluruh dokumen dalam corpus.

Secara teoritis, TF-IDF terdiri dari dua komponen utama:

1) Term Frequency (TF)

Menggambarkan seberapa sering sebuah kata muncul dalam dokumen. Semakin sering muncul, semakin besar nilainya.

2) Inverse Document Frequency (IDF)

Mengukur tingkat keunikan kata dalam keseluruhan dokumen. Kata yang umum muncul pada banyak dokumen memiliki nilai IDF kecil, sedangkan kata yang jarang muncul memiliki nilai IDF lebih tinggi:

3) TF-IDF:

Kata yang sering muncul di dokumen tertentu tetapi jarang di dokumen lain menjadi bobot tinggi. Kata umum seperti “dan”, “yang” menjadi bobot rendah.

Kombinasi keduanya memberikan bobot fitur yang mampu menonjolkan kata-kata penting dan menurunkan pengaruh kata umum seperti “dan” atau “yang”. Banyak penelitian menerapkan TF-IDF bersama n-gram (misalnya unigram dan bigram) untuk menangkap konteks kata secara lebih baik dan meningkatkan akurasi klasifikasi. (M & T, 2022)

2.4 Metode Klasifikasi Sentimen

Berbagai algoritma telah banyak digunakan dalam penelitian analisis sentimen. Model tradisional seperti *Logistic Regression*, *Support Vector Machine (SVM)*, dan *Multi-Layer Perceptron (MLP)* terbukti efektif ketika dipadukan dengan representasi teks TF-IDF.

a. Support Vector Machine (SVM Linear)

Algoritma SVM dengan kernel linear digunakan untuk memisahkan kelas sentimen melalui penentuan margin pemisahan yang maksimal. SVM dikenal memiliki kinerja yang baik pada data teks yang bersifat sparse, termasuk data yang direpresentasikan menggunakan TF-IDF. Pada implementasinya, model SVM dilatih menggunakan perintah SVC (kernel='linear', probability=True). Model ini diterapkan dalam sebuah pipeline yang mencakup proses TF-IDF hingga tahap klasifikasi dengan SVM.

b. Multinomial Naïve Bayes (MNB)

Naive Bayes merupakan algoritma berbasis probabilistik yang sangat sesuai untuk data dengan representasi frekuensi kata. Varian Multinomial Naive Bayes secara luas digunakan dalam analisis teks karena kesederhanaannya serta performanya yang konsisten. Dalam implementasinya, MNB digunakan sebagai model pembandingan melalui alur kerja yang sama, yaitu: TF-IDF → MultinomialNB.

Model ini menghasilkan keluaran yang digunakan dalam penyusunan confusion matrix, perhitungan metrik evaluasi.

2.5 Analisis Sentimen pada Ulasan Produk

Analisis sentimen pada ulasan produk digunakan untuk memahami persepsi konsumen terhadap kualitas produk, harga, rasa, kemasan, maupun pengalaman pembelian. Pada industri kopi, ulasan mengenai rasa kopi, konsistensi kualitas, kecepatan layanan, dan harga menjadi indikator penting dalam menilai kepuasan pelanggan. (R & E, 2023) Penggunaan analisis sentimen berbasis SVM dan Naive Bayes dalam penelitian ini memungkinkan evaluasi objektif terhadap model terbaik dalam memprediksi sentimen ulasan produk kopi secara otomatis.

3. HASIL DAN PEMBAHASAN

Hasil penelitian bahwa mayoritas ulasan produk kopi di media sosial didominasi oleh sentimen positif, yang menandakan tingkat kepuasan konsumen yang relatif baik terhadap produk kopi. Sentimen netral dan negatif muncul dalam jumlah lebih sedikit dan umumnya berkaitan dengan komentar informatif serta keluhan terkait cita rasa atau harga produk. Tahapan pra-pemrosesan teks dan ekstraksi fitur menggunakan metode TF-IDF terbukti mampu meningkatkan kualitas data sehingga mendukung proses klasifikasi sentimen. Berdasarkan hasil pelatihan model, algoritma Support Vector Machine (SVM) Linear menghasilkan performa terbaik dengan nilai akurasi, precision, recall, dan F1-score tertinggi dibandingkan dengan Multinomial Naive Bayes.

Analisis confusion matrix menunjukkan bahwa SVM Linear memiliki tingkat kesalahan klasifikasi yang sangat rendah pada seluruh kategori sentimen. Sementara itu, Naive Bayes masih mengalami kesalahan klasifikasi, khususnya pada sentimen net. Berdasarkan perbandingan kinerja kedua algoritma, dapat disimpulkan bahwa SVM Linear secara konsisten unggul dalam seluruh metrik evaluasi dibandingkan dengan Multinomial Naive Bayes. Keunggulan ini menunjukkan bahwa SVM Linear lebih efektif dalam menangani data teks pendek yang direpresentasikan menggunakan TF-IDF, seperti ulasan produk kopi di media sosial. Dengan demikian, algoritma SVM Linear

merupakan model yang paling optimal untuk digunakan dalam analisis sentimen pada penelitian ini.

3.1 Hasil Dataset

Penelitian ini menggunakan dataset berupa komentar pengguna yang telah diberi label sentimen. Data disimpan dalam format CSV dan terdiri atas dua atribut utama, yakni kolom *Komentar* sebagai representasi data teks dan *Sentimen* sebagai kelas target. Pada tahap awal, data dimuat dan diperiksa jumlah keseluruhannya beserta distribusi masing-masing kelas sentimen. Langkah ini dilakukan untuk memastikan bahwa dataset memiliki proporsi yang memadai dan layak digunakan dalam proses klasifikasi.

Tabel 1 Dataset

Komentar	Sentimen	Platform
Cita rasa kopinya kurang kuat dan tidak sesuai.	Negatif	Twitter
Pengemasan rapih, kopi sampai rumah tetap fresh.	Positif	Toko Online
Aplikasi pemesanan mudah digunakan dan update promo terbaru.	Positif	Toko Online

3.2 Hasil Sentimen Per-Kategori

Dominasi hasil sentimen positif menunjukkan bahwa mayoritas pengguna memberikan penilaian yang baik terhadap produk kopi, baik dari aspek rasa, kualitas produk, maupun pelayanan. Sentimen netral mencerminkan komentar yang bersifat informatif atau tidak menunjukkan kecenderungan emosi yang jelas. Sementara itu, sentimen negatif merepresentasikan adanya ketidakpuasan pengguna, yang umumnya berkaitan dengan faktor harga, cita rasa, atau konsistensi produk.

Tabel 2 Hasil Sentimen Per-Kategori

Kategori Sentimen	Total Jumlah
Positif	170
Negatif	67
Netral	78

3.3 Hasil Processing Text

Sebelum memasuki tahap pemodelan, data teks terlebih dahulu melalui proses *preprocessing*. Tahapan ini mencakup konversi seluruh teks menjadi huruf kecil (*case folding*), penghapusan URL, karakter non-alfanumerik, serta pembersihan spasi berlebih yang kerap muncul pada komentar media sosial. Hasil dari proses ini disimpan dalam kolom *clean_text* dan digunakan sebagai input utama pada tahap pembentukan fitur. Secara umum, *preprocessing* bertujuan untuk mengurangi noise yang dapat mengganggu proses pembelajaran model, sekaligus membantu meningkatkan akurasi hasil klasifikasi.

	komentar	clean_text
0	Kopi ini terasa enak dan membantu menambah sem...	kopi ini terasa enak dan membantu menambah sem...
1	Cita rasa kopinya kurang kuat dan kurang sesua...	cita rasa kopinya kurang kuat dan kurang sesua...
2	Pelayanan di toko cukup cepat dan para pegawai...	pelayanan di toko cukup cepat dan para pegawai...
3	Harganya terasa terlalu mahal dibandingkan kua...	harganya terasa terlalu mahal dibandingkan kua...
4	Kopi susu di sini sangat creamy dan rasanya ni...	kopi susu di sini sangat creamy dan rasanya ni...

Gambar 2 Procesing Text

3.4 Hasil Training Model

Pada penelitian ini digunakan dua algoritma klasifikasi, yaitu Support Vector Machine (SVM) Linear dan Naive Bayes Multinomial. Kedua algoritma tersebut dilatih menggunakan data latih yang telah melalui proses preprocessing teks serta ekstraksi fitur menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF).

Tabel 3 Hasil Training Model

Model	SVM	Naïve BAYes
Akurasi	99.2063	
Precision	99.2126	
Recall	99.2063	
F1 Score	99.2072	

seluruh metrik tersebut menunjukkan bahwa model SVM Linear mampu mengklasifikasikan data sentimen secara optimal dengan tingkat kesalahan yang sangat rendah. Sementara itu, model Naive Bayes memperoleh nilai akurasi sebesar 89,29%, precision sebesar 91,07%, recall sebesar 89,29%, dan F1-score sebesar 88,87%. Meskipun performa yang dihasilkan masih tergolong baik, nilai metrik Naive Bayes lebih

rendah dibandingkan dengan SVM Linear. Hal ini menunjukkan bahwa asumsi independensi antar fitur pada algoritma Naive Bayes memiliki keterbatasan dalam menangkap hubungan kompleks antar kata pada data teks.

3.5 Hasil Perbandingan SVM dan Naïve Bayes

a) Confusion Matriks SVM

Sebelum memasuki tahap pemodelan, data teks terlebih dahulu melalui proses *preprocessing*. Tahapan ini mencakup konversi seluruh teks menjadi huruf kecil (*case folding*), penghapusan URL, karakter non-alfanumerik, serta pembersihan spasi berlebih yang kerap muncul pada komentar media sosial. Hasil dari proses ini disimpan dalam kolom *clean_text* dan digunakan sebagai input utama pada tahap pembentukan fitur. Secara umum, *preprocessing* bertujuan untuk mengurangi noise yang dapat mengganggu proses pembelajaran model, sekaligus membantu meningkatkan akurasi hasil klasifikasi.

Confusion Matrix - SVM

Aktual	Negatif	9	2	2
	Netral	5	4	7
	Positif	1	5	28
		Negatif	Netral	Positif
		Prediksi		

Gambar 3 Confusion Matrix SVM

b) Confusion Matrix Naïve Bayes

Algoritma Naive Bayes menunjukkan kecenderungan kuat dalam memprediksi kelas netral, yang tercermin dari jumlah prediksi benar pada kelas tersebut sebanyak 13 data. ini menunjukkan bahwa Naive Bayes memiliki keterbatasan dalam menangkap pola yang lebih kompleks pada sentimen positif, terutama pada data teks dengan variasi kosakata dan konteks yang lebih beragam, sebagaimana umum ditemukan pada komentar pengguna di media sosial.

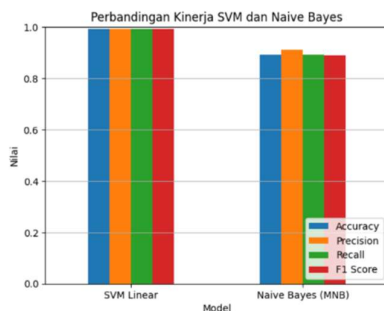
Confusion Matrix - Naive Bayes

Aktual	Negatif	10	3	0
	Netral	3	13	0
	Positif	6	22	6
		Negatif	Netral	Positif
		Prediksi		

Gambar 4 Confusion Matrix Naive Bayes

c) Perbandingan Kinerja Model

Hasil diagram perbandingan kinerja model bahwa SVM Linear secara konsisten unggul pada seluruh metrik evaluasi dibandingkan dengan Naive Bayes Multinomial. Keunggulan tersebut terlihat pada nilai akurasi, precision, recall, dan F1-score yang lebih tinggi dan stabil. Dengan demikian, dapat disimpulkan bahwa model SVM Linear merupakan algoritma yang paling optimal dan memiliki performa terbaik dalam melakukan analisis sentimen terhadap komentar produk kopi pada penelitian ini.



Gambar 5 Diagram Perbandingan Kinerja

4. KESIMPULAN

Penelitian ini menunjukkan keberhasilan penerapan analisis sentimen berbasis *supervised learning* dalam mengklasifikasikan ulasan produk kopi di media sosial ke dalam tiga kategori sentimen, yaitu positif, netral, dan negatif. Hasil analisis mengindikasikan bahwa sentimen positif mendominasi keseluruhan data, yang mencerminkan tingkat kepuasan konsumen terhadap produk kopi yang relatif tinggi. Berdasarkan evaluasi kinerja model, algoritma Support Vector Machine (SVM) Linear terbukti memberikan performa yang lebih baik dibandingkan dengan Multinomial Naive Bayes, ditunjukkan oleh nilai akurasi, *precision*, *recall*, dan *F1-score* yang lebih tinggi serta tingkat kesalahan klasifikasi yang lebih rendah. Keunggulan SVM Linear terletak pada kemampuannya dalam mengolah data teks berdimensi tinggi dan bersifat *sparse* yang direpresentasikan menggunakan metode TF-IDF.

Sementara itu, Multinomial Naive Bayes memiliki kelebihan dalam hal kesederhanaan model dan efisiensi komputasi, namun memiliki keterbatasan dalam menangkap hubungan kontekstual antar kata karena asumsi independensi antar fitur. Meskipun demikian, penelitian ini masih memiliki keterbatasan, khususnya pada cakupan dataset dan jumlah algoritma yang digunakan. Oleh karena itu, pengembangan penelitian selanjutnya dapat dilakukan dengan memperluas sumber data, menerapkan teknik penyeimbangan kelas, serta membandingkan kinerja model dengan metode yang lebih kompleks, seperti *deep learning* atau pendekatan *word embedding*, guna meningkatkan akurasi analisis sentimen.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Universitas Muhadi Setiabudi yang telah memberikan dukungan akademik dan fasilitas selama pelaksanaan penelitian ini. Ucapan terima kasih juga disampaikan kepada dosen pembimbing yang telah memberikan arahan, bimbingan, dan masukan yang konstruktif sehingga penelitian ini dapat diselesaikan dengan baik. Selain itu, penulis mengapresiasi pihak-pihak yang telah membantu dalam penyediaan data serta kontribusi lainnya yang mendukung kelancaran penelitian ini.

REFERENSI

- A, P. W., & R, N. H. (2022). Penerapan Text Mining untuk Analisis Sentimen Ulasan Produk Online. *Jurnal Informatika dan Sistem Informasi*, 4(2), 76-84.
- G, A., & T, E. (2024). Penggunaan Algoritma Naive Bayes dan Support Vector Machine untuk Analisis Sentimen pada Media Sosial. *Jurnal Informatika dan Teknik Elektro Terapan(JITET)*, 11(1), 85-94.
- Kabir Muhammad, F. A., & Putra, R. E. (2025). Perbandingan Analisis Sentimen Untuk Prediksi Kepuasan Pelanggan Kedai Kopi Di Kofind Menggunakan Algoritma SVM Dan Naive Bayes. *Jurnal Informatika dan Ilmu Komputer (JINACS)*, 6(4), 1039-1048.
- M, F. R., & T, H. P. (2022). Analisis Sentimen Ulasan Konsumen Menggunakan Support Vector Machine dengan Pembobotan TF-IDF. *Jurnal Teknologi dan Sistem Komputer*, 10(3), 201-208.
- R, H. S., & E, S. (2023). Analisis Sentimen Ulasan Produk Menggunakan TF-IDF dan Naive Bayes. *Jurnal Informatika Polinema*, 9(2), 98-105.
- Y, A. P., & F, N. (2023). Analisis Sentimen Media Sosial Menggunakan Metode Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 9(1), 1834-1842.