

Eksperimen Komparatif 10 Algoritma Machine Learning untuk Analisis Sentimen Ulasan Tokopedia

Comparative Experiment of 10 Machine Learning Algorithms for Sentiment Analysis of Tokopedia

Kevin Risky Abadi^{*1}, Rayhan Ferdiansyah², Adam Risqi Alfarisi³, Muhammad Faisal⁴

^{1,2,3,4} Program Studi Sistem Informasi, Fakultas Teknik & Ilmu Komputer, Universitas

Nusantara PGRI Kediri

e-mail: *1kevinrisky18@gmail.com

Abstrak - Penelitian ini bertujuan untuk mengevaluasi dan membandingkan kinerja sepuluh algoritma machine learning dalam tugas klasifikasi sentimen terhadap ulasan produk Tokopedia. Data yang digunakan berasal dari platform Kaggle dan terdiri dari 500 ulasan pengguna yang telah dilabeli secara biner berdasarkan rating. Metode yang diterapkan meliputi tahap praproses teks, representasi fitur menggunakan TF-IDF dan CountVectorizer, serta pelatihan model menggunakan berbagai algoritma seperti Naïve Bayes, SVM, Logistic Regression, Random Forest, dan MLPClassifier. Evaluasi dilakukan menggunakan metrik Accuracy, Precision, Recall, dan F1 Score. Hasil penelitian menunjukkan bahwa model Naïve Bayes dengan CountVectorizer mencapai F1 Score tertinggi sebesar 0,9225, diikuti oleh SVM dengan bigram dan MLPClassifier. Sementara itu, Logistic Regression menunjukkan performa yang stabil meskipun ditambahkan teknik regularisasi L2 dan reduksi dimensi PCA. Temuan ini menunjukkan bahwa kombinasi representasi fitur dan algoritma yang tepat dapat menghasilkan performa yang tinggi, bahkan pada model yang sederhana. Penelitian ini memberikan kontribusi dalam penerapan analisis sentimen berbasis machine learning pada domain e-commerce Indonesia.

Kata kunci – analisis sentimen; klasifikasi teks; machine learning; Tokopedia; ulasan produk

Abstract - This study aimed to evaluate and compare the performance of ten machine learning algorithms in the task of sentiment classification on Tokopedia product reviews. The dataset used was obtained from Kaggle and consisted of 500 user reviews labeled into binary sentiment classes based on rating scores. The methods applied included text preprocessing, feature representation using TF-IDF and CountVectorizer, and model training using various algorithms such as Naïve Bayes, SVM, Logistic Regression, Random Forest, and MLPClassifier. The evaluation used Accuracy, Precision, Recall, and F1 Score as performance metrics. The results showed that the Naïve Bayes model with CountVectorizer achieved the highest F1 Score of 0.9225, followed by SVM with bigram and MLPClassifier. Logistic Regression demonstrated stable performance across configurations including L2 regularization and PCA-based dimensionality reduction. These findings indicate that an appropriate combination of feature representation and algorithm selection can yield high performance, even with simple models. This research contributes to the application of machine learning-based sentiment analysis in the Indonesian e-commerce domain.

Keywords – machine learning; product review; sentiment analysis; text classification; Tokopedia

I. PENDAHULUAN

Dalam era ekonomi digital saat ini, transaksi daring telah menjadi bagian penting dalam perilaku konsumsi masyarakat. E-commerce sebagai bentuk transformasi digital menghadirkan kemudahan dan efisiensi dalam proses jual beli, namun juga menghadirkan tantangan baru dalam pengambilan keputusan konsumen. Salah satu elemen yang memiliki pengaruh signifikan dalam keputusan pembelian adalah ulasan pengguna atau review.

Platform Tokopedia, sebagai salah satu marketplace terbesar di Indonesia, memberikan fitur ulasan produk sebagai indikator reputasi dan kualitas barang yang ditawarkan. Menurut DataReportal (2022), lebih dari 75% konsumen Indonesia menyatakan bahwa mereka mempertimbangkan ulasan online sebelum melakukan pembelian, dan hampir setengahnya menempatkan ulasan online setara pentingnya dengan rekomendasi pribadi [1].

Ulasan yang diberikan oleh pengguna Tokopedia umumnya bersifat teks bebas (free-text), yang mencerminkan persepsi, pengalaman, maupun penilaian subjektif terhadap suatu produk. Semakin meningkatnya jumlah ulasan seiring tingginya transaksi membuat proses analisis secara manual menjadi tidak efektif, lambat, dan rentan terhadap bias interpretasi. Oleh karena itu, diperlukan pendekatan otomatis yang dapat mengklasifikasikan ulasan berdasarkan polaritas sentimen, seperti positif atau negatif, dengan akurasi dan efisiensi yang tinggi.

Analisis sentimen merupakan bagian dari bidang pemrosesan bahasa alami (Natural Language Processing/NLP) [2] yang bertujuan untuk mengidentifikasi opini, sikap, atau emosi seseorang dari suatu teks. Dalam konteks e-commerce, analisis sentimen berperan penting dalam memahami persepsi konsumen terhadap produk maupun layanan, serta dapat dimanfaatkan oleh pelaku bisnis untuk meningkatkan strategi pemasaran dan pengelolaan reputasi produk. Salah satu pendekatan yang umum digunakan dalam analisis sentimen adalah machine learning, yang memungkinkan sistem belajar dari data latih untuk kemudian memprediksi sentimen teks baru secara otomatis.

Penelitian mengenai analisis sentimen berbasis machine learning di Indonesia telah banyak dilakukan, terutama dengan fokus pada penggunaan algoritma Naïve Bayes. Misalnya, Salsabila et al. (2022) melakukan klasifikasi sentimen ulasan Tokopedia menggunakan Naïve Bayes dan melaporkan akurasi sebesar 95,10% terhadap data komentar aplikasi di Google Play Store [3]. Sementara itu, Sera et al. (2023) menerapkan metode serupa pada platform Bukalapak dengan akurasi 88%, menggunakan pendekatan TF-IDF untuk representasi fitur [4]. Kedua studi tersebut menunjukkan bahwa Naïve Bayes cukup efektif dalam mengklasifikasikan teks pendek dalam Bahasa Indonesia.

Namun, meskipun algoritma Naïve Bayes memiliki keunggulan dalam kecepatan dan efisiensi komputasi, pendekatan ini memiliki keterbatasan. Salah satunya adalah asumsi independensi antar fitur yang jarang terpenuhi dalam bahasa alami. Selain itu, belum banyak penelitian lokal yang secara sistematis membandingkan performa berbagai algoritma machine learning lainnya seperti Support Vector Machine (SVM), Logistic Regression, Random Forest, maupun Neural Network (MLPClassifier) dalam konteks analisis sentimen ulasan produk di platform Tokopedia. Putri dan Cahyono (2024) membandingkan SVM dan Naïve Bayes pada komentar publik terhadap pelayanan DKI Jakarta dan menemukan bahwa SVM sedikit lebih unggul [5]. Penelitian lain oleh Daniati dan Utama (2023) menggabungkan Word2Vec dan

ensemble learning untuk meningkatkan akurasi klasifikasi sentimen pada data Twitter [6].

Dengan demikian, terlihat adanya kesenjangan (gap analysis) dalam studi analisis sentimen di Indonesia, khususnya yang terkait dengan perbandingan komprehensif berbagai algoritma machine learning. Sebagian besar penelitian terdahulu cenderung terbatas pada satu atau dua model, tanpa membandingkan berbagai pendekatan secara menyeluruh. Selain itu, mayoritas penelitian berfokus pada data media sosial seperti Twitter atau komentar umum, bukan secara khusus pada ulasan e-commerce seperti Tokopedia yang memiliki karakteristik domain spesifik.

Penelitian ini hadir untuk mengisi kesenjangan tersebut dengan melakukan eksperimen komparatif terhadap sepuluh algoritma machine learning untuk klasifikasi sentimen pada ulasan Tokopedia. Algoritma yang diuji meliputi Naïve Bayes, Logistic Regression, SVM, Random Forest, MLPClassifier, serta beberapa variasi konfigurasi fitur dan preprocessing seperti bigram dan stopword removal. Setiap model dievaluasi menggunakan metrik akurasi, precision, recall, dan F1-score untuk menilai performa klasifikasi secara menyeluruh.

Tujuan umum dari penelitian ini adalah untuk mengevaluasi dan membandingkan kinerja berbagai algoritma machine learning dalam tugas klasifikasi sentimen Bahasa Indonesia terhadap ulasan produk Tokopedia. Tujuan khususnya adalah: (1) mengidentifikasi model dengan performa terbaik berdasarkan F1-score; (2) mengevaluasi pengaruh variasi preprocessing terhadap hasil klasifikasi; dan (3) memberikan rekomendasi model yang optimal untuk digunakan dalam sistem analitik e-commerce di Indonesia.

Penelitian ini memberikan kontribusi kebaruan (novelty) dalam bentuk eksperimen sistematis terhadap berbagai model machine learning pada domain ulasan e-commerce Indonesia, dengan menggunakan dataset yang representatif dan pendekatan evaluasi yang komprehensif. Hasil dari penelitian ini diharapkan dapat memberikan acuan bagi pengembang sistem rekomendasi, analitik bisnis, maupun akademisi yang tertarik dalam bidang text mining dan machine learning berbasis Bahasa Indonesia.

II. TINJAUAN TEORI

2.1 Analisis Sentimen

Analisis sentimen atau opinion mining merupakan salah satu cabang dari pemrosesan bahasa alami (Natural Language Processing/NLP) yang bertujuan untuk mengekstrak, mengidentifikasi, dan mengklasifikasikan opini atau sikap seseorang terhadap suatu entitas berdasarkan teks [2]. Dalam konteks e-commerce, analisis sentimen digunakan untuk menilai persepsi pengguna terhadap suatu produk atau layanan melalui ulasan yang dituliskan secara daring. Proses ini penting karena dapat membantu pelaku bisnis dalam memahami kepuasan dan ekspektasi konsumen.

Tugas utama dalam analisis sentimen adalah mengklasifikasikan teks ke dalam kategori sentimen tertentu, seperti positif, negatif, atau netral. Pendekatan yang umum digunakan adalah metode berbasis kamus (lexicon-based) dan metode berbasis pembelajaran mesin (machine learning). Metode pembelajaran mesin menjadi pilihan utama dalam skala besar karena memiliki kemampuan adaptif terhadap variasi bahasa dan dapat dilatih secara otomatis dari data historis.

2.2 Pemrosesan Bahasa Alami (Natural Language Processing)

NLP merupakan bidang interdisipliner antara linguistik dan kecerdasan buatan yang berfokus pada kemampuan mesin dalam memahami, menginterpretasikan, dan menghasilkan bahasa manusia [7]. Dalam tugas analisis sentimen, proses NLP mencakup tahap praproses data seperti cleansing (pembersihan karakter khusus dan angka), tokenisasi (pemisahan kata), stopword removal (penghapusan kata umum), stemming (pengembalian ke bentuk dasar), dan normalisasi teks.

Tahapan ini krusial untuk memastikan bahwa input ke algoritma machine learning merepresentasikan informasi yang esensial dalam format numerik. Umumnya, fitur diekstraksi menggunakan metode representasi teks seperti Bag of Words (BoW) atau Term Frequency-Inverse Document Frequency (TF-IDF).

2.3 Representasi Fitur: TF-IDF

TF-IDF adalah salah satu teknik pembobotan kata yang paling populer dalam pengolahan teks [8]. TF (Term Frequency) mengukur seberapa sering suatu kata muncul dalam sebuah dokumen, sedangkan IDF (Inverse Document Frequency) mengukur seberapa penting kata tersebut di seluruh kumpulan dokumen. Kombinasi TF dan IDF memungkinkan algoritma untuk lebih fokus pada kata-kata penting dan mengabaikan kata-kata umum.

Rumus dasar TF dan IDF:

- $TF(t) = \text{Jumlah kemunculan kata } t \text{ dalam dokumen} / \text{jumlah total kata dalam dokumen}$
- $IDF(t) = \log(N / df(t))$, di mana N adalah jumlah total dokumen dan $df(t)$ adalah jumlah dokumen yang mengandung kata t .

2.4 Algoritma Machine Learning untuk Analisis Sentimen

Berbagai algoritma pembelajaran mesin telah digunakan untuk tugas klasifikasi sentimen. Setiap algoritma memiliki keunggulan dan kelemahan berdasarkan struktur data dan jenis fitur. Beberapa algoritma yang sering digunakan antara lain:

- Naïve Bayes: Merupakan metode probabilistik yang sederhana dan cepat, berdasarkan Teorema Bayes dengan asumsi independensi antar fitur [4]. Cocok untuk teks pendek namun kurang efektif jika korelasi antar kata signifikan.
- Support Vector Machine (SVM): Merupakan algoritma klasifikasi margin maksimum yang sangat efektif untuk data berdimensi tinggi. SVM mampu membentuk hyperplane pemisah optimal antara kelas [5].
- Logistic Regression: Merupakan model linier yang digunakan untuk prediksi probabilitas kelas. Logistic Regression sering digunakan dalam klasifikasi biner dengan kinerja yang cukup stabil.
- Random Forest: Merupakan metode ensemble yang terdiri dari banyak pohon keputusan. Random Forest mampu mengurangi overfitting dan bekerja baik untuk data non-linier.
- Multilayer Perceptron (MLPClassifier): Merupakan jenis jaringan saraf tiruan (artificial neural network) yang terdiri dari beberapa lapisan neuron dan mampu menangani hubungan kompleks antar fitur.

2.5 Variasi Teknik Ekstraksi Fitur dan Pratraining Model

Dalam analisis sentimen berbasis teks, kualitas hasil klasifikasi tidak hanya dipengaruhi oleh pemilihan algoritma, tetapi juga oleh teknik ekstraksi fitur dan praproses data yang digunakan. Representasi teks yang baik akan mempermudah algoritma machine learning dalam membedakan pola dan polaritas sentimen. Oleh karena itu, pada penelitian ini dilakukan beberapa variasi pada teknik ekstraksi fitur dan pratraining model sebagai bagian dari eksperimen klasifikasi.

Salah satu teknik representasi teks yang paling umum digunakan adalah Term Frequency-Inverse Document Frequency (TF-IDF). TF-IDF mengukur pentingnya sebuah kata dalam suatu dokumen relatif terhadap seluruh dokumen dalam korpus. Semakin sering suatu kata muncul dalam satu dokumen, dan semakin jarang muncul di dokumen lain, maka bobotnya akan semakin tinggi. TF-IDF dinilai lebih efektif dibanding CountVectorizer murni karena mampu mereduksi dominasi kata-kata umum [8]. Sebagai perbandingan, CountVectorizer hanya menghitung frekuensi absolut kata dalam dokumen, tanpa mempertimbangkan keberadaannya secara global, sehingga dapat menyebabkan fitur yang terlalu besar dan kurang informatif.

Selain teknik pembobotan, salah satu modifikasi penting dalam praproses teks adalah penghapusan stopwords. Stopword merupakan kata-kata umum seperti “dan”, “yang”, “di”, “dengan”, dan sebagainya, yang secara semantik tidak menambah makna sentimen [7]. Menghapus stopwords dapat mengurangi dimensi fitur dan membantu model fokus pada kata-kata yang lebih bermakna dalam menentukan polaritas suatu ulasan.

Teknik lain yang digunakan adalah n-gram, yaitu representasi teks berdasarkan urutan kata. Unigram merepresentasikan setiap kata secara independen, sedangkan bigram mempertimbangkan pasangan dua kata secara berurutan. Penggunaan bigram dapat meningkatkan konteks pemahaman, seperti membedakan antara “tidak baik” dan “baik sekali”, yang secara unigram dapat disalahartikan. Namun, penambahan n-gram juga meningkatkan dimensi fitur dan dapat menimbulkan sparsity data.

Selain teknik praproses berbasis teks, penelitian ini juga menerapkan metode reduksi dimensi, yaitu Principal Component Analysis (PCA) [9]. PCA bertujuan mengurangi jumlah fitur yang sangat besar (high dimensional) dengan memproyeksikannya ke ruang baru yang mempertahankan variansi terbesar. Dengan demikian, PCA dapat membantu meningkatkan efisiensi pemrosesan dan mencegah overfitting, terutama pada model dengan kompleksitas tinggi seperti Logistic Regression.

Untuk menghindari overfitting dan menjaga generalisasi model, digunakan pula teknik regularisasi, khususnya L2 regularization pada model Logistic Regression [10]. L2 regularization memberikan penalti terhadap bobot koefisien yang terlalu besar, sehingga model tidak terlalu menyesuaikan diri terhadap data latih dan lebih mampu menggeneralisasi data baru. Dalam konteks klasifikasi teks, regularisasi sangat penting karena jumlah fitur sangat besar dan bersifat spars.

Berbagai kombinasi dari teknik-teknik ini digunakan dalam sepuluh eksperimen yang dilakukan dalam penelitian, dengan tujuan mengevaluasi pengaruhnya terhadap kinerja model secara menyeluruh.

2.6 Tinjauan Pustaka Penelitian Sebelumnya

Penelitian mengenai klasifikasi sentimen ulasan di platform e-commerce Indonesia telah banyak dilakukan. Salsabila et al. [3] menggunakan metode Naïve Bayes pada ulasan Tokopedia dan memperoleh akurasi hingga 95,10%. Sera et al. [4] menerapkan pendekatan serupa di platform Bukalapak dan mencatat akurasi sebesar 88%.

Putri dan Cahyono [5] membandingkan algoritma SVM dan Naïve Bayes pada komentar media sosial tentang pelayanan publik dan menemukan bahwa SVM menunjukkan hasil yang sedikit lebih tinggi dalam hal akurasi dan F1-score. Penelitian lain oleh Daniati dan Utama [6] memanfaatkan teknik ensemble learning dan representasi Word2Vec pada data Twitter, dan memperoleh peningkatan akurasi yang signifikan. Terakhir, Adisti et al. [11] menunjukkan bahwa SVM efektif digunakan dalam deteksi ujaran kebencian melalui media sosial, mendukung potensi algoritma ini untuk tugas klasifikasi berbasis teks dalam bahasa Indonesia

III. METODE

3.1 Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif eksperimental. Fokus utama penelitian adalah mengevaluasi performa berbagai algoritma klasifikasi teks dalam tugas analisis sentimen terhadap ulasan produk di Tokopedia [2]. Eksperimen dilakukan secara komputasional menggunakan Python dan Google Colab.

3.2 Dataset dan Sumber Data

Data yang digunakan dalam penelitian ini diperoleh dari dataset publik yang tersedia di platform Kaggle berjudul “Analisa Sentimen Ulasan di Tokopedia” [12]. Dataset ini berisi kumpulan ulasan produk dari pengguna Tokopedia dalam bahasa Indonesia, termasuk label rating (1–5) yang digunakan untuk penentuan polaritas sentimen. Label diklasifikasikan secara biner: ulasan dengan rating 4 dan 5 dikategorikan sebagai sentimen positif (label 1), sementara rating 1–3 dikategorikan sebagai sentimen negatif (label 0).

Dataset awal berisi 500 baris data ulasan dengan format kolom: ulasan_clean (teks ulasan), ulasan_label (angka), dan informasi tambahan seperti tanggal dan rating. Data kemudian melalui proses seleksi dan praproses sebelum dimasukkan ke dalam model klasifikasi.

3.3 Proses Praproses Teks

Tahapan pra-pemrosesan teks meliputi:

- Fillna: Mengganti nilai null dengan “tidak ada komentar”
- Cleansing: Menghapus karakter spesial, angka, emoji
- Tokenisasi dan lowercasing
- Stopword removal (pada beberapa eksperimen)
- Representasi teks: CountVectorizer dan/atau TF-IDF [7],[8]

3.4 Pembagian Data Latih dan Uji

Data dibagi menjadi 70% data latih dan 30% data uji menggunakan `train_test_split` dari `scikit-learn`, dengan stratifikasi label dan `random_state = 30`.

3.5 Desain Eksperimen dan Model

Sebanyak sepuluh eksperimen dilakukan menggunakan kombinasi algoritma klasifikasi dan teknik representasi fitur teks [2],[7]. Setiap model dirancang dalam bentuk pipeline yang terdiri atas tiga komponen utama, yaitu ekstraksi fitur teks (vectorizer), teknik pembobotan (TF-IDF), dan algoritma klasifikasi [8]. Beberapa model juga memanfaatkan modifikasi praproses atau teknik optimasi tambahan seperti bigram, stopword removal, regularisasi L2, dan reduksi dimensi menggunakan PCA [9], [10].

Eksperimen ini bertujuan untuk membandingkan kinerja berbagai konfigurasi model machine learning dalam tugas klasifikasi sentimen ulasan Tokopedia. Tabel 1 berikut merangkum seluruh konfigurasi eksperimen:

Tabel 1. Desain Eksperimen: Kombinasi Algoritma, Representasi Fitur, dan Modifikasi

Model	Algoritma	Fitur	Modifikasi
M1	Naïve Bayes	TF-IDF	0.0
M2	Logistic Regression	TF-IDF	0.0
M3	SVM (LinearSVC)	TF-IDF	0.0
M4	Random Forest	TF-IDF	0.0
M5	MLPClassifier	TF-IDF	<code>max_iter = 300</code>
M6	Naïve Bayes	CountVectorizer	tanpa TF-IDF
M7	SVM (LinearSVC)	TF-IDF	<code>ngram_range= (1,2)</code> (unigram + bigram)
M8	Logistic Regression	TF-IDF	L2 Regularization
M9	SVM (LinearSVC)	TF-IDF	Stopword removal (English)
M10	Logistic Regression	TF-IDF + PCA	Reduksi dimensi via TruncatedSVD (n=100)

3.6 Evaluasi Kinerja

Setiap model dievaluasi berdasarkan empat metrik evaluasi klasifikasi: Accuracy, Precision, Recall, dan F1 Score. Dalam penelitian ini, F1 Score dipilih sebagai metrik utama karena mampu merepresentasikan keseimbangan antara Precision dan Recall, terutama dalam kasus data teks di mana distribusi kelas bisa tidak seimbang [2].

Evaluasi dilakukan pada data uji (`y_test`) dengan prediksi model (`y_pred`), dan hasilnya disajikan dalam bentuk DataFrame dan visualisasi. Teknik cross-validation tidak diterapkan secara menyeluruh, kecuali pada tuning awal algoritma Naïve Bayes (parameter `var_smoothing`) menggunakan `GridSearchCV` dan `RepeatedStratifiedKfold`.

3.7 Implementasi dan Tools

Seluruh eksperimen dalam penelitian ini diimplementasikan menggunakan bahasa pemrograman Python dengan bantuan sejumlah pustaka (library) pemrosesan data dan machine learning. Proses eksperimen dilakukan secara cloud-based melalui platform Google Colaboratory agar lebih efisien dan mudah direplikasi.

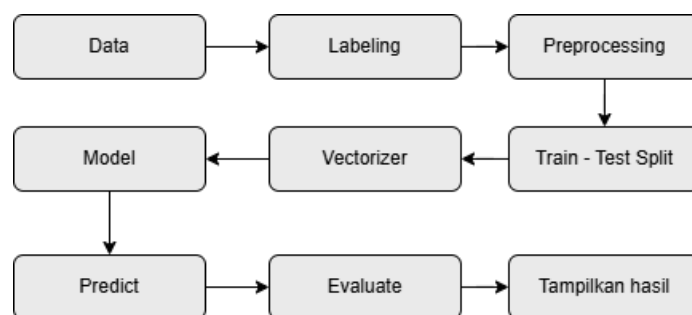
Adapun alat bantu (tools) yang digunakan adalah sebagai berikut:

- Bahasa: Python
- Platform: Google Colaboratory
- Library utama: scikit-learn, pandas, matplotlib, seaborn, wordcloud
- Dataset: Analisa Sentimen Ulasan di Tokopedia (diunduh dari Kaggle: <https://www.kaggle.com/code/raniaprillyap/analisa-sentimen-ulasan-di-tokopedia-1>), disimpan sebagai file lokal tokped_text.xlsx
- Tools Visualisasi: Confusion Matrix, Wordcloud, Bar Chart, Pie Chart

Implementasi dilakukan secara modular menggunakan pipeline untuk menjaga keterurutan proses mulai dari vektorisasi, pemodelan, hingga evaluasi. Semua tahapan tercatat dalam bentuk skrip yang dapat dijalankan ulang secara konsisten.

3.8 Alur Eksperimen

Diagram alir yang mencerminkan alur sebagai berikut:



Gambar 1. Diagram Alur Eksperimen

IV. HASIL DAN PEMBAHASAN

4.1 Hasil Evaluasi Model

Sepuluh model klasifikasi telah diujicobakan untuk menganalisis sentimen ulasan produk Tokopedia. Setiap model dievaluasi menggunakan empat metrik utama, yaitu Accuracy, Precision, Recall, dan F1 Score. F1 Score dipilih sebagai metrik utama karena mampu mencerminkan keseimbangan antara presisi dan sensitivitas dalam klasifikasi data teks. Hasil evaluasi ditampilkan pada Tabel 2 berikut:

Tabel 2. Hasil Evaluasi Sepuluh Model Klasifikasi

No	Model	Accuracy	Precision	Recall	F1 Score
1	NB + CountVectorizer	0.8676	0.8881	0.9597	0.9225
2	MLPClassifier + TF-IDF	0.8609	0.8931	0.9435	0.9176
3	SVM + TF-IDF + Bigram	0.8543	0.8592	0.9839	0.9173
4	SVM + TF-IDF + Stopwords	0.8477	0.8633	0.9677	0.9125
5	SVM + TF-IDF	0.8411	0.8623	0.9597	0.9084
6	Random Forest + TF-IDF	0.8411	0.9032	0.9032	0.9032
7	Logistic Regression + TF-IDF	0.8212	0.8212	1.0000	0.9018
8	Naïve Bayes + TF-IDF	0.8212	0.8212	1.0000	0.9018
9	Logistic Regression + TF-IDF + L2	0.8212	0.8212	1.0000	0.9018
10	Logistic Regression + TF-IDF + PCA (n=100)	0.8212	0.8212	1.0000	0.9018

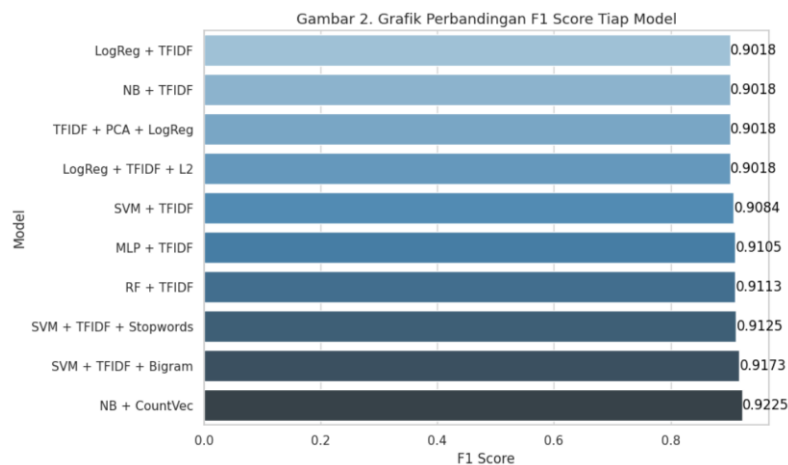
4.2 Pembahasan

Model dengan performa terbaik ditinjau dari F1 Score adalah NB + CountVectorizer dengan skor F1 sebesar 0.9225, diikuti oleh MLPClassifier + TF-IDF (0.9176) dan SVM + TF-IDF + Bigram (0.9173). Hal ini menunjukkan bahwa meskipun TF-IDF merupakan representasi fitur yang dominan, CountVectorizer tetap dapat memberikan performa tinggi ketika dipadukan dengan algoritma Naïve Bayes.

Model SVM secara umum menunjukkan hasil yang konsisten tinggi, khususnya saat dikombinasikan dengan bigram dan stopwords removal. Hal ini sejalan dengan karakteristik SVM yang sangat efektif dalam mengklasifikasikan data berdimensi tinggi seperti teks. Sementara itu, MLPClassifier juga menunjukkan kinerja tinggi meskipun memiliki waktu pelatihan yang lebih panjang karena arsitektur jaringannya.

Logistic Regression, baik dalam versi standar maupun dengan L2 Regularization dan PCA, memberikan hasil yang seragam (F1 Score 0.9018), menandakan stabilitas model ini terhadap variasi preproses dan teknik optimasi.

Menariknya, model Random Forest yang biasanya kuat pada data tabular, hanya menunjukkan performa menengah (F1 Score 0.9032), sedangkan model Naïve Bayes berbasis TF-IDF menghasilkan skor yang setara dengan Logistic Regression. Ini menunjukkan bahwa pemilihan kombinasi algoritma dan teknik representasi fitur memiliki pengaruh signifikan terhadap hasil klasifikasi sentimen.



Gambar 2. Grafik Perbandingan F1 Score Tiap Model

4.3 Relevansi Hasil

Hasil penelitian ini memberikan wawasan penting terkait strategi pemodelan sentimen dalam domain e-commerce Indonesia. Meskipun algoritma yang lebih kompleks seperti MLP dan SVM menunjukkan performa unggul, hasil tertinggi justru dicapai oleh model sederhana: Naïve Bayes dengan representasi CountVectorizer.

Temuan ini menunjukkan bahwa model ringan dengan konfigurasi yang tepat tetap dapat bersaing dalam performa, yang penting bagi aplikasi industri yang mengutamakan efisiensi dan kecepatan. Selain itu, kombinasi bigram dan stopwords removal pada SVM terbukti

meningkatkan performa, yang dapat menjadi strategi tambahan dalam pengembangan sistem analitik berbasis teks.

Secara keseluruhan, penelitian ini menunjukkan bahwa klasifikasi sentimen tidak hanya bergantung pada kompleksitas algoritma, melainkan juga pada kecocokan antara representasi fitur dan karakteristik data.

V. KESIMPULAN DAN SARAN

Penelitian ini telah mengevaluasi sepuluh model klasifikasi machine learning untuk analisis sentimen terhadap ulasan produk di Tokopedia. Berdasarkan eksperimen yang dilakukan, dapat disimpulkan bahwa pemilihan algoritma dan teknik representasi fitur memiliki pengaruh signifikan terhadap performa klasifikasi. Model Naïve Bayes dengan representasi CountVectorizer menunjukkan performa terbaik secara keseluruhan, bahkan mampu mengungguli model kompleks seperti Support Vector Machine dan MLPClassifier. Temuan ini mengindikasikan bahwa model sederhana dapat tetap unggul apabila didukung oleh konfigurasi fitur yang tepat. Selain itu, teknik praproses seperti penggunaan bigram, stopword removal, serta pendekatan reduksi dimensi dan regularisasi turut memberikan kontribusi terhadap peningkatan akurasi dan kestabilan model. Hasil penelitian ini menegaskan bahwa analisis sentimen berbasis machine learning dapat diimplementasikan secara efektif untuk mendukung pengambilan keputusan bisnis di sektor e-commerce Indonesia.

Saran untuk penelitian selanjutnya adalah agar eksperimen diperluas dengan menambahkan variasi dataset yang lebih besar dan lebih seimbang dari segi distribusi sentimen. Penggunaan pendekatan deep learning seperti LSTM atau BERT juga dapat menjadi alternatif untuk mengevaluasi apakah model berbasis representasi kontekstual mampu memberikan hasil yang lebih baik. Selain itu, penelitian lanjutan dapat mengkaji performa model dalam lingkungan produksi secara real-time, serta mengintegrasikan analisis sentimen ke dalam sistem rekomendasi produk untuk meningkatkan pengalaman pengguna secara menyeluruh.

VI. DAFTAR PUSTAKA

- [1] Simon Kemp, "Digital 2022: Indonesia." Accessed: Jun. 19, 2025. [Online]. Available: <https://datareportal.com/reports/digital-2022-indonesia>
- [2] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," *Foundations and Trends® in Information Retrieval*, vol. 2, no. 1–2, pp. 1–135, 2008, doi: 10.1561/15000000011.
- [3] S. M. Salsabila, A. A. Murtopo, and N. Fadhillah, "Analisis Sentimen Pelanggan Tokopedia Menggunakan Metode Naïve Bayes Classifier," Sep. 2022, Accessed: Jun. 19, 2025. [Online]. Available: <https://jurnal.polgan.ac.id/index.php/jmp/article/view/11640>
- [4] E. Sera, S. Komputer, H. Makassar, and B. Riset dan Inovasi, "Analisis Sentimen Ulasan Produk di E-Commerce Bukalapak Menggunakan Natural Language Processing Sentiment Analysis of Product Reviews on E-Commerce's Bukalapak Using Natural Language Processing." [Online]. Available: <https://www.kaggle.com/>.
- [5] R. R. Putri and N. Cahyono, "ANALISIS SENTIMEN KOMENTAR MASYARAKAT TERHADAP PELAYANAN PUBLIK PEMERINTAH DKI JAKARTA DENGAN ALGORITMA SUPER VECTOR MACHINE DAN NAIVE BAYES," 2024.
- [6] E. Daniati and H. Utama, "ANALISIS SENTIMEN DENGAN PENDEKATAN ENSEMBLE LEARNING DAN WORD EMBEDDING PADA TWITTER," 2023.

- [7] Daniel Jurafsky; James H. Martin, *Speech and Language Processing*, 2nd ed. Upper Saddle River, NJ: Prentice Hall, 2009. Accessed: Jun. 19, 2025. [Online]. Available: <https://web.stanford.edu/~jurafsky/slp3/>
- [8] K. SPARCK JONES, "A STATISTICAL INTERPRETATION OF TERM SPECIFICITY AND ITS APPLICATION IN RETRIEVAL," *Journal of Documentation*, vol. 28, no. 1, pp. 11–21, Jan. 1972, doi: 10.1108/eb026526.
- [9] Ian T. Jolliffe, *Principal Component Analysis*, 2nd ed. in Springer Series in Statistics. New York: Springer-Verlag, 2002. doi: 10.1007/b98835.
- [10] Christopher M. Bishop, *Pattern Recognition and Machine Learning*, 1st ed. New York: Springer, 2006. Accessed: Jun. 19, 2025. [Online]. Available: <https://link.springer.com/book/10.1007/978-0-387-45528-0>
- [11] T. C. Adisti, E. Daniati, and A. Ristiyawan, "ANALISIS SENTIMEN UJARAN KEBENCIAN PADA TWEET DI TWITTER," 2025.
- [12] Rania Aprillya, "Analisa Sentimen Ulasan di Tokopedia." Accessed: Jun. 19, 2025. [Online]. Available: <https://www.kaggle.com/code/raniaprillyap/analisa-sentimen-ulasan-di-tokopedia-1>

