Accredited SINTA 5
72/E/KPT/2024

Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)
Vol. 9 No. 3, Juni 2026
P-ISSN 2620-8342
E-ISSN 2621-3052

Perbandingan Algoritma Machine Learning Berbasis TF-IDF dan SMOTE untuk Analisis Sentimen Ulasan Aming Coffee

Emi Wulandari[1*], Riski Annisa[2], Muhammad Fahmi Julianto[3]
[1-3]Program Studi Informatika, Kampus Kota Pontianak, Universitas Bina Sarana Informatika
[1-3]Jl. Abdul Rahman Saleh No.18, Bangka Belitung Laut, Kec. Pontianak Tenggara,
Kota Pontianak, Kalimantan Barat 78124
E-mail: emiulandaripon11@gmail.com[1*]

Submitted Date : 10 Juni 2026 ***Accepted Date : 25 Juni 2026***

Abstrak - Ulasan pelanggan pada Google Maps dapat dimanfaatkan untuk mengetahui persepsi pelanggan terhadap kualitas produk dan layanan suatu perusahaan. Namun, banyaknya jumlah ulasan dan data yang tidak terstruktur menyebabkan proses analisis secara manual menjadi kurang efektif. Penelitian ini bertujuan untuk melakukan analisis sentimen pada ulasan pelanggan Aming Coffee menggunakan metode pembobotan Term Frequency–Inverse Document Frequency (TF-IDF) dan algoritma machine learning. Data penelitian diperoleh melalui proses scraping sebanyak 4.000 ulasan pelanggan dari Google Maps Aming Coffee. Tahapan penelitian meliputi preprocessing data yang terdiri atas cleaning, case folding, tokenisasi, normalisasi, stopword removal, dan stemming, kemudian dilakukan pembobotan TF-IDF, pembagian data, klasifikasi menggunakan algoritma Naïve Bayes, Support Vector Machine (SVM), dan Decision Tree C4.5, serta penerapan Synthetic Minority Over-sampling Technique (SMOTE) pada data latih untuk menangani ketidakseimbangan kelas. Evaluasi model dilakukan menggunakan accuracy, precision, recall, F1-score, dan confusion matrix. Hasil penelitian menunjukkan bahwa model tanpa SMOTE memperoleh akurasi tertinggi, yaitu Naïve Bayes sebesar 90,63%, diikuti SVM sebesar 90,50%, dan C4.5 sebesar 88,25%. Setelah penerapan SMOTE, akurasi model mengalami penurunan, namun nilai precision meningkat sehingga model menjadi lebih mampu memperhatikan kelas sentimen minoritas. Hasil Exploratory Data Analysis (EDA) menunjukkan bahwa sentimen positif mendominasi ulasan pelanggan Aming Coffee dengan kata yang paling sering muncul antara lain “kopi”, “coffee”, “enak”, “mantap”, dan “ramai”. Penelitian ini menunjukkan bahwa kombinasi TF-IDF, machine learning, dan SMOTE dapat digunakan untuk analisis sentimen ulasan pelanggan serta memberikan informasi yang bermanfaat dalam evaluasi kualitas layanan dan pengambilan keputusan bisnis.
Kata kunci : Analisis Sentimen; TF-IDF; Naïve Bayes; Support Vector Machine; Decision Tree C4.5; SMOTE;

Abstract - *Customer reviews on Google Maps can be utilized to understand customer perceptions of a company's products and services. However, the large volume of unstructured reviews makes manual analysis less effective. This study aims to perform sentiment analysis on Aming Coffee customer reviews using the Term Frequency–Inverse Document Frequency (TF-IDF) weighting method and machine learning algorithms. Research data were collected through scraping 4,000 customer reviews from Google Maps. The research stages included data preprocessing consisting of cleaning, case folding, tokenization, normalization, stopword removal, and stemming, followed by TF-IDF weighting, data splitting, sentiment classification using Naïve Bayes, Support Vector Machine (SVM), and Decision Tree C4.5 algorithms, as well as the implementation of Synthetic Minority Over-sampling Technique (SMOTE) on training data to address class imbalance. Model evaluation was conducted using accuracy, precision, recall, F1-score, and confusion matrix. The results showed that models without SMOTE achieved the highest accuracy, with Naïve Bayes reaching 90.63%, followed by SVM at 90.50% and C4.5 at 88.25%. After applying SMOTE, model accuracy decreased, while precision increased, indicating improved attention to minority sentiment classes. Exploratory Data Analysis (EDA) results revealed that positive sentiment dominated Aming Coffee customer reviews, with frequently occurring words including “kopi,” “coffee,” “enak,” “mantap,” and “ramai.” This study demonstrates that the combination of TF-IDF, machine learning algorithms, and SMOTE can be effectively applied to sentiment analysis and provide useful insights for service quality evaluation and business decision-making.*
Keywords: Sentiment Analysis; TF-IDF; Naïve Bayes; Support Vector Machine; Decision Tree C4.5; SMOTE;

1. Pendahuluan

Ulasan yang diposting oleh pelanggan di platform online seperti Google Reviews dan media sosial dianggap sangat penting dalam membentuk persepsi dan memengaruhi keputusan pelanggan untuk bisnis, termasuk coffee shop[1]. Seiring meningkatnya penggunaan teknologi internet, platform seperti Google Maps dan GoFood telah menjadi wadah bagi konsumen untuk menyampaikan opini secara bebas dan spontan. Ulasan tersebut berisi informasi berharga mengenai pengalaman pelanggan, persepsi terhadap layanan, hingga kualitas

Copyright © Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)

produk yang ditawarkan. Sayangnya, data ini sering kali tidak dimanfaatkan secara optimal karena jumlahnya besar dan tidak terstruktur, sehingga memerlukan metode analisis yang tepat untuk mengekstrak wawasan bermakna[2].

Aming Coffee merupakan salah satu coffee shop yang cukup dikenal di Kota Pontianak dan telah menjadi bagian dari budaya masyarakat dalam menikmati kopi khas daerah. Sebagai salah satu coffee shop yang memiliki banyak pelanggan, Aming Coffee menerima berbagai ulasan dari pengunjung melalui platform Google Maps. Ulasan tersebut mencerminkan pengalaman pelanggan terhadap berbagai aspek, seperti kualitas produk, pelayanan, harga, fasilitas, dan suasana tempat. Beragamnya opini yang diberikan pelanggan menjadikan ulasan tersebut sebagai ulasan pelanggan pada platform Google Maps merupakan sumber informasi yang berharga untuk mengetahui tingkat kepuasan serta persepsi konsumen terhadap layanan yang diberikan. Namun, jumlah ulasan yang terus meningkat dan kontennya yang beragam menyebabkan proses pembacaan dan analisis secara manual menjadi kurang efisien karena memerlukan waktu dan tenaga yang besar. Selain itu, perbedaan opini dalam setiap ulasan menyulitkan pengelola untuk memperoleh gambaran umum sentimen pelanggan secara cepat dan akurat. Oleh sebab itu, diperlukan suatu pendekatan otomatis yang mampu mengolah dan menganalisis ulasan tersebut sehingga informasi yang terkandung di dalamnya dapat dimanfaatkan sebagai dasar evaluasi dan pengambilan keputusan.

Untuk mengatasi permasalahan banyaknya ulasan yang tidak terstruktur, digunakan metode yang mampu mengekstraksi informasi secara otomatis, yaitu analisis sentimen. Analisis sentimen, sebagai bagian dari bidang Natural Language Processing (NLP), bertujuan mengidentifikasi dan mengevaluasi emosi atau opini yang terkandung dalam teks, kemudian mengelompokkannya ke dalam kategori sentimen positif, netral, atau negatif. Teknik ini telah banyak diaplikasikan dan terbukti efektif dalam memahami persepsi publik terhadap berbagai produk dan layanan[3].

Dalam penelitian ini, *Term Frequency–Inverse Document Frequency* (TF-IDF) digunakan sebagai metode pembobotan kata untuk mendukung proses analisis sentimen. TF-IDF merupakan metode statistika yang digunakan untuk mengukur tingkat kepentingan suatu kata dalam sebuah dokumen dibandingkan dengan keseluruhan dokumen dalam satu korpus. Melalui pendekatan ini, kata-kata yang paling merepresentasikan isi ulasan akan memperoleh bobot yang lebih tinggi sehingga dapat meningkatkan efektivitas proses klasifikasi sentimen[4].

Setelah proses pembobotan kata menggunakan TF-IDF, tahap selanjutnya adalah melakukan klasifikasi sentimen dengan algoritma *machine learning*. *Naïve Bayes* merupakan algoritma klasifikasi berbasis probabilitas yang banyak digunakan dalam analisis teks karena strukturnya yang sederhana dan efisiensi komputasinya yang tinggi.[5]. Support Vector Machine (SVM) dikenal memiliki kinerja yang baik dalam menangani data berdimensi tinggi serta menghasilkan performa yang stabil pada berbagai permasalahan klasifikasi teks[6]. Sementara itu, Decision Tree (C4.5) bekerja dengan membagi data berdasarkan atribut-atribut tertentu hingga membentuk pohon keputusan yang mudah dipahami dan diinterpretasikan[7]. Ketiga algoritma tersebut dipilih untuk dibandingkan kinerjanya dalam mengklasifikasikan sentimen ulasan pelanggan Aming Coffee.

2. Tinjauan Pustaka

2.1. Analisis Sentimen

Analisis sentimen merupakan salah satu pendekatan dalam *Natural Language Processing* (NLP) yang digunakan untuk mengidentifikasi, memahami, dan mengklasifikasi opini atau emosi pada data teks ke dalam kategori positif, netral, dan negatif[8]. Teknik ini banyak digunakan pada analisis ulasan produk, media sosisal, serta evaluasi kepuasan pelanggan. Analisis secara manual membutuhkan waktu dan kurang efektif untuk data dalam jumlah besar, sehingga digunakan pendekatan *text mining* dan *machine learning* untuk melakukan klasifikasi secara otomatis[9]. Pendekatan berbasis *machine learning* digunakan untuk mengotomatisasi proses pengelompokan opini tesebut. Penerapan model pembelajaran terawasi (*supervised learning*) tidak hanya meningkatkan efisiensi dalam pengolahan data ulasan yang jumlahnya besar, tetapi juga menghasilkan analisis yang lebih konsisten dan objektif. Hasil ekstraksi sentimen ini dapat dimanfaatkan untuk memahami persepsi pelanggan terhadap suatu produk atau layanan, sehingga dapat menjadi dasar dalam pengambilan keputusan dan peningkatan kualitas layanan bisnis[10]. Dalam konteks penelitian ini, implementasi tersebut diterapkan untuk membedakan polaritas ulasan konsumen pada Aming Coffee.

2.2. Pembobotan TF-IDF

Term Frequency–Inverse Document Frequency (TF-IDF) merupakan metode pembobotan kata untuk mengukur tingkat kepentingan suatu kata pada dokumen berdasarkan frekuensi kemunculan dan penyebarannya pada seluruh dokumen[11].

Secara matematis, bobot akhir suatu kata *t* pada dokumen *d* diperoleh dari hasil perkalian antara nilai *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF), yang dirumuskan pada Persamaan (1) :

$$W_{t,d} = TF(t,d) \times IDF(t) \qquad (1)$$

Copyright © Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)

Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)
Vol. 9 No. 3, Juni 2026
P-ISSN 2620-8342
E-ISSN 2621-3052

Nilai IDF(t) dihitung menggunakan rumus logaritma perbandingan jumlah seluruh dokumen dengan jumlah dokumen yang mengandung kata tertentu, seperti ditunjukkan pada Persamaan (2) :

$$IDF(t) = log_{10}\left(\frac{N}{df(t)}\right) \quad (2)$$

Keterangan simbol matematika pada rumus diatas adalah :

$W_{t,d}$	=	Bobot akhir kata *t* pada dokumen *d*
$TF(t,d)$	=	Frekuensi kemunculan kata *t* di dalam dokumen *d*
N	=	Total keseluruhan dokumen (ulasan) yang ada di dalam korpus
$df(t)$	=	Jumlah dokumen (ulasan) yang mengandung kata *t* di dalamnya

Dalam penelitian ini, metode TF-IDF digunakan untuk memberikan bobot pada kata-kata dalam ulasan pelanggan Aming Coffee sehingga proses klasifikasi sentimen menggunakan algoritma *machine learning* dapat dilakukan dengan lebih efektif dan menghasilkan analisis yang lebih akurat.

2.3. Algoritma *Machine Learning*

2.3.1. Naïve Bayes

Naïve Bayes adalah salah satu algoritma klasifikasi yang bersifat probabilistik dan banyak dimanfaatkan dalam studi sentimen. Algoritma ini beroperasi mengikuti Teorema Bayes dengan cara mengukur peluang suatu data terhadap kategori tertentu. Dalam konteks analisis sentimen, Naïve Bayes berfungsi untuk mengelompokkan pandangan menjadi tiga kategori: positif, netral, dan negatif, berdasarkan kalimat pada teks[12]. Algoritma ini populer karena perhitungan yang mudah dan dapat menghasilkan klasifikasi yang akurat pada teks. Berdasarkan penelitian oleh Harpizon, algoritma Naïve Bayes menunjukkan performa yang sangat baik dalam menganalisis sentimen komentar di YouTube, dengan akurasi mencapai 87%, precision 91%, dan recall 97%[13]. Temuan ini mengindikasikan bahwa algoritma ini efektif dalam mengategorikan data teks berdasarkan sentimen yang ada. Dalam kajian ini, Naïve Bayes juga diterapkan untuk mengklasifikasikan sentimen ulasan pelanggan Aming Coffee dengan menggunakan pembobotan kata melalui metode TF-IDF.

2.3.2. Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah sebuah algoritma pembelajaran mesin yang sering diterapkan dalam analisis sentimen serta pengelompokan teks. Algoritma ini berfungsi untuk menemukan hyperplane terbaik yang bisa memisahkan data ke dalam beberapa kategori dengan cara yang paling efisien. SVM terkenal karena kemampuannya yang unggul dalam mengolah data dengan dimensi tinggi dan dapat memberikan hasil klasifikasi yang konsisten di berbagai situasi analisis teks.[14]. Penelitian telah menunjukkan bahwa SVM sangat akurat dalam proses klasifikasi teks. Penelitian yang dilakukan oleh Josi dkk. menunjukkan bahwa kemampuan algoritma SVM untuk menemukan batas pemisah (hyperplane) yang ideal antar kelas memungkinkannya menghasilkan klasifikasi yang sangat baik pada data teks. SVM banyak digunakan untuk klasifikasi dokumen dan analisis sentimen karena kemampuan ini[15]. Dalam analisis sentimen, SVM digunakan untuk mengklasifikasikan opini menjadi kategori positif, netral, dan negatif berdasarkan fitur yang diperoleh dari pembobotan kata pada teks. Algoritma ini populer karena mampu memberikan akurasi yang baik pada proses klasifikasi teks.

2.3.3. Decision Tree C4.5

Decision Tree C4.5 merupakan algoritma klasifikasi yang bekerja dengan membentuk struktur pohon keputusan berdasarkan atribut tertentu pada data. Algoritma ini digunakan untuk memastikan kelas sentimen berdasarkan pola data teks yang diperoleh dari proses pembobotan kata. Dalam analisis sentimen, C4.5 mampu digunakan untuk mengklasifikasikan opini ke dalam kategori positif, netral, maupun negatif sehingga banyak diterapkan pada penelitian klasifikasi teks[16], [17]. Menurut penelitian, algoritma C4.5 dapat membuat aturan klasifikasi yang mudah dipahami dan diinterpretasikan. Struktur pohon keputusan yang dibuat juga dapat membantu memahami pola hubungan antar atribut yang digunakan untuk klasifikasi. Algoritma C4.5 banyak digunakan dalam berbagai penelitian data mining dan klasifikasi teks karena kemampuan ini[18]. Studi ini menggunakan algoritma C4.5 untuk mengklasifikasikan sentimen ulasan pelanggan Aming Coffee berdasarkan fitur hasil pembobotan TF-IDF. Hasil klasifikasi ini kemudian dibandingkan dengan algoritma Naive Bayes dan Support Vector Machine (SVM) untuk menentukan model yang paling efektif dalam mengklasifikasikan sentimen ulasan pelanggan.

2.3.4. Synthetic Minority Over-sampling Technique (SMOTE)

Untuk mengatasi masalah ketidakseimbangan kelas (class imbalance) dalam proses klasifikasi, teknik penyeimbangan data sintetis minoritas (SMOTE) digunakan. Sehingga distribusi data menjadi lebih seimbang, SMOTE menghasilkan data sintetis pada kelas minoritas berdasarkan kedekatan antar sampel. Dengan menggunakan SMOTE, model pembelajaran mesin dapat lebih baik mengenali pola di kelas minoritas dan mengurangi bias terhadap kelas mayoritas. Studi yang dilakukan oleh Cahyaningtyas dkk. menunjukkan bahwa penerapan SMOTE pada analisis sentimen rating aplikasi Shopee menggunakan algoritma Decision Tree meningkatkan nilai akurasi dan AUC dibandingkan dengan model tanpa SMOTE. Hasil penelitian

Copyright © Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)

Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)
Vol. 9 No. 3, Juni 2026
P-ISSN 2620-8342
E-ISSN 2621-3052

menunjukkan bahwa metode SMOTE efektif meningkatkan kinerja klasifikasi pada data yang tidak seimbang[19].

3. Metode Penelitian

Pertama, data dikumpulkan. Untuk analisis sentimen, dataset utama digunakan dari ulasan pelanggan Aming Coffee yang dikumpulkan melalui scraping data dari Google Maps. Proses penelitian digambarkan pada Gambar1 berikut ini.

Gambar 1. Alur Penelitian

Gambar 1 menggambarkan langkah-langkah dalam penelitian yang mencakup pengumpulan data dari ulasan Google Maps, tahap persiapan data, penandaan sentimen, analisis data eksploratif (EDA), pemberian bobot TF-IDF, pembagian dataset, pemodelan pembelajaran mesin, evaluasi model, dan penafsiran hasil analisis

Sebelum mengumpulkan data untuk analisis, fase persiapan data dilakukan untuk membersihkan dan menyiapkan informasi. Proses ini terdiri dari cleaning, case folding, tokenizing, penghapusan stopword, dan stemming. Dalam analisis sentimen, fase persiapan data sangat krusial karena dapat meningkatkan kualitas teks dan memengaruhi cara kerja model klasifikasi yang akan dibuat.[20].

- ✓ **Cleaning**, yaitu penghapusan karakter-karakter yang tidak perlu seperti tanda baca, angka, URL, emoji, dan simbol-simbol lain.
- ✓ **Case folding**, yaitu mengkonversi semua huruf menjadi huruf kecil.
- ✓ **Tokenizing**, yaitu memecah kalimat menjadi kata-kata atau token-token.
- ✓ **Stopword**, yaitu menghapus kata-kata yang tidak memiliki arti signifikan dalam analisis sentimen.
- ✓ **Stemming**, yaitu mengubah kata-kata yang memiliki imbuhan ke bentuk dasarnya.

Setelah preprocessing selesai, data ulasan dilabeli ke dalam kelas sentimen positif, netral, dan negatif. Setiap ulasan dilabelkan untuk menentukan kelasnya, yang digunakan sebagai target dalam proses pembelajaran model pembelajaran mesin. Baik metode manual maupun otomatis berbasis kamus sentimen (lexicon-based) dapat digunakan untuk melabelkan sentimen, dan bagaimana model klasifikasi yang dihasilkan dipengaruhi oleh kualitas pelabelan[21].

Dalam tahap selanjutnya, analisis data eksploratif (EDA) dilakukan dengan tujuan untuk memahami karakteristik data yang dilabelkan dan mendapatkan informasi awal tentang pola sentimen pelanggan terhadap Aming Coffee. EDA mencakup distribusi jumlah data pada setiap kelas sentimen, visualisasi persentase sentimen, analisis frekuensi kata yang sering muncul, pembuatan cloud kata, analisis panjang ulasan, dan penemuan ketidakseimbangan data (*class imbalance*), hasil EDA memberikan gambaran umum dataset.

Setelah proses EDA selesai, teks diubah menjadi format numerik menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF). Metode ini memberikan nilai pada setiap kata berdasarkan seberapa sering kata tersebut muncul dalam dokumen tertentu dan di seluruh korpus, sehingga dapat mengidentifikasi kata-kata yang dianggap signifikan dalam proses klasifikasi teks. TF-IDF merupakan salah satu teknik ekstraksi fitur yang paling umum digunakan dalam studi analisis teks.[22].

$$TF - IDF(t,d) = TF(t,d) \times IDF(t) \quad (3)$$

Selanjutnya, dataset dipisahkan menjadi 20% untuk data uji dan 80% untuk data latih. Data latih dimanfaatkan untuk membangun model pembelajaran mesin, sementara data uji digunakan untuk menilai kemampuan model dalam mengklasifikasikan informasi yang belum pernah ditemukan sebelumnya.

Setelah itu, data yang telah diberi bobot TF-IDF digunakan untuk menciptakan model pembelajaran mesin. Algoritma yang diaplikasikan dalam penelitian ini adalah Naïve Bayes, Support Vector Machine (SVM), dan Decision Tree C4.5. Untuk menghasilkan model klasifikasi yang efisien, kombinasi preprocessing, TF-IDF, dan algoritma pembelajaran mesin telah banyak diterapkan dalam riset analisis sentimen.[22].

Copyright © Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)

Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)
Vol. 9 No. 3, Juni 2026
P-ISSN 2620-8342
E-ISSN 2621-3052

Model yang telah dilatih selanjutnya akan diuji menggunakan data uji untuk mengevaluasi seberapa efektif performanya dalam mengklasifikasikan sentimen. Proses evaluasi model dilakukan dengan menggunakan beberapa metrik, antara lain:

✓ **Accuracy**, untuk menilai ketepatan prediksi yang dikeluarkan oleh model.
✓ **Precision**, untuk menilai sejauh mana model bisa memprediksi dengan benar setiap kategori sentimen.
✓ **Recall**, untuk mengevaluasi kemampuan model dalam mendeteksi seluruh data pada masing-masing kategori sentimen.
✓ **F1-Score**, untuk mengukur kinerja model dalam mengidentifikasi semua data di setiap kategori sentimen.
✓ **Confusion Matrix**, untuk melihat secara detail hasil klasifikasi untuk setiap jenis sentimen.

Analisis hasil evaluasi dilakukan guna memahami seberapa berhasil model dalam mengklasifikasikan sentimen dari ulasan pelanggan Aming Coffee. Di akhir penelitian, kesimpulan diambil dari hasil pengujian tersebut dan saran-saran disusun untuk pengembangan penelitian yang akan datang.

4. Hasil dan Pembahasan

4.1. Hasil Pengumpulan Data

Data dikumpulkan dengan menggunakan teknik web scraping untuk mengumpulkan ulasan pelanggan Aming Coffee yang ada di Google Maps. Tools scraping digunakan untuk mengumpulkan informasi dari halaman ulasan, seperti teks ulasan (review), rating, nama pengguna, URL ulasan, dan informasi tambahan lainnya yang tersedia di Google Maps.

Proses scraping menghasilkan 4.000 data ulasan pelanggan Aming Coffee. Selanjutnya, untuk memudahkan pengolahan dan analisis, data disimpan dalam format CSV. Selain itu, pemeriksaan kualitas dilakukan untuk memastikan bahwa data tidak mengandung duplikasi, data tidak relevan, maupun data kosong.

Setelah proses seleksi data, data ulasan dikumpulkan dan siap digunakan dalam tahap preprocessing. Sebelum digunakan untuk analisis sentimen, data diproses melalui proses pembersihan, case folding, tokenisasi, normalisasi, stopword, dan stemming. Selanjutnya, untuk klasifikasi sentimen dan pembobotan TF-IDF, algoritma Naive Bayes, Support Vector Machine (SVM), dan Decision Tree C4.5 digunakan sebagai sumber data utama.

Tabel 1. Ringkasan Data Hasil Pengumpulan

Keterangan	Jumlah
Total Data Hasil Scraping	4.000
Sumber Data	Google Maps
Objek Penelitian	Aming Coffee
Teknik Pengumpulan Data	Web Scraping
Format Data	CSV

Pada Tabel 1, menunjukkan bahwa penelitian menggunakan data ulasan pelanggan Aming Coffee yang diperoleh melalui teknik web scraping dari Google Maps. Dataset penelitian terdiri dari 4.000 data yang berhasil dikumpulkan dan digunakan. Analisis sentimen bergantung pada data ini untuk mengetahui persepsi pelanggan terhadap produk dan layanan Aming Coffee.

4.2. Hasil Preprocessing Data

Sebelum data ulasan pelanggan Aming Coffee digunakan untuk analisis sentimen, proses preprocessing dilakukan untuk menyediakan karakter yang tidak diperlukan seperti tanda baca, angka, simbol, URL, dan variasi penulisan kata. Karakter-karakter ini dapat memengaruhi kualitas analisis. Oleh karena itu, beberapa proses preprocessing dilakukan, seperti pembersihan, case folding, tokenisasi, normalisasi, stopword, dan stemming. Tahap ini bertujuan untuk menghasilkan data teks yang lebih bersih, terorganisir, dan konsisten. Dengan demikian, algoritma pengajaran mesin dapat meningkatkan proses pembobotan TF-IDF dan klasifikasi sentimen.

1. Cleaning

Proses *cleaning* dilakukan dengan menghapus karakter seperti tanda baca, angka, URL, emoji, dan simbol lainnya. Tujuan dari tahap ini adalah mengurai *noise* pada data sehingga informasi yang digunakan hanya berupa kata-kata yang relevan dengan ulasan pelanggan.

Tabel 2. Hasil Cleaning

Sebelum Cleaning	Sesudah Cleaning
Kopi legendaris khas pontianak. Selalu ramai 😆 Patut dicobain 👍 Yukkk mampir	kopi legendaris khas pontianak selalu ramai patut dicobain yukkk mampir

Copyright © Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)

Aku ngopi di aming coffe yang di cibinong city mall..... nyaman sekali ngopi di sini.... ada ruang free smooking dan area smooking nya di luar.... sip pokoknyahh 👌 ✨ ✨	aku ngopi di aming coffe yang di cibinong city mall nyaman sekali ngopi di sini ada ruang free smooking dan area smooking nya di luar sip pokoknyahh
Kopi Aming > Kopi Asiang	kopi aming kopi asiang
Bau kopinya seger dan kopinya enak 👍	bau kopinya seger dan kopinya enak
Aming coffee banyak tersebar di Kota Pontianak dan luar kota seperti di Jakarta. Meskipun demikian, vibes nongkrongnya berbeda 😆	aming coffee banyak tersebar di kota pontianak dan luar kota seperti di jakarta meskipun demikian vibes nongkrongnya berbeda

Sumber: Hasil pengolahan data, 2026

Berdasarkan Tabel 2, proses cleaning berhasil menghilangkan karakter yang tidak diperlukan sehingga data menjadi lebih bersih dan mudah diproses pada tahap selanjutnya.

2. Case Folding

Setelah cleaning, proses case folding digunakan untuk mengubah seluruh teks menjadi huruf kecil (lowercase). Pada tahap ini, sistem menyeragamkan bentuk penulisan kata sehingga kata-kata yang memiliki arti yang sama tetapi ditulis dengan huruf besar atau kecil tidak dianggap berbeda.

Tabel 3. Hasil Case Folding

Sebelum Case Folding	Sesudah Case Folding
Time for Coffee HVM	time for coffee hvm
Kopinya mantap	kopinya mantap
Kopi Mantap Snack Pisang Oke	kopi mantap snack pisang oke
Kopi Aming terbaik	kopi aming terbaik
Jelas Sangat Memuaskan	jelas sangat memuaskan

Pada Tabel 3, seluruh huruf kapitas berhasil diubah menjadi huruf kecil sehingga menghasilkan format teks yang lebih konsisten.

3. Tokenizing

Tokenisasi adalah pembagian kalimat menjadi kumpulan kata atau token, yang memungkinkan setiap kata diproses secara terpisah pada langkah berikutnya. Daftar kata yang dihasilkan dari proses tokenisasi berfungsi sebagai representasi isi ulasan pelanggan.

Tabel 4. Hasil Tokenizing

Sebelum Tokenizing	Sesudah Tokenizing
kopi enak suasana lumayan ada makanan ringan tradisional	kopi,enak,suasana,lumayan,ada,makanan,ringan,tradisional
indomienya enak	indomienya,enak
tempat ngopi sejak lama	tempat,ngopi,sejak,lama
jelas sangat memuaskan	jelas, sangat, memuaskan
kopinya enak	kopinya,enak

Pada tabel 4, setiap kalimat berhasil dipecah menjadi token-token kata sehingga memudahkan proses analisis teks pada tahap selanjutnya.

4. Normalisasi

Normalisasi adalah proses yang digunakan untuk mengubah singkatan, kata tidak baku, bahasa gaul, dan kesalahan penulisan menjadi kata baku. Proses ini sangat penting karena ulasan pelanggan sering mengandung variasi penulisan yang berbeda dengan arti yang sama. Kata-kata dapat diseragamkan dengan normalisasi, yang mengurangi jumlah fitur yang tidak penting.

Tabel 5. Hasil Normalisasi

Sebelum Normalisasi	Sesudah Normalisasi
bgt	banget
yg	yang
dgn	dengan

Pada tabel diatas, kata-kata tidak baku berhasil diubah menjadi bentuk baku sehingga data menjadi lebih konsisten dan mudah dianalisis.

5. Stopword

Copyright © Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)

Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)
Vol. 9 No. 3, Juni 2026
P-ISSN 2620-8342
E-ISSN 2621-3052

Dalam tahap stopword removal, kata-kata umum yang sering muncul dalam teks tetapi tidak memberikan informasi yang diperlukan untuk analisis sentimen dihilangkan. Kata-kata seperti "dan", "yang", "di", dan kata-kata umum lainnya dihilangkan agar data lebih berkonsentrasi pada kata-kata yang mengandung informasi atau pendapat pelanggan.

Tabel 6. Hasil Stopword

Sebelum Stopword	Sesudah Stopword
kopi enak suasana lumayan ada makanan ringan tradiosional	kopi,enak,suasana,lumayan,makanan,ringan,tradiosional
memang benar legend dan harganya normal normal saja	legend,harganya,normal,normal
kopi legendaris khas pontianak selalu ramai patut dicobain yuk mampir	kopi,legendaris,khas,pontianak,ramai,patut,dicobain,yuk,mampir

Berdasarkan Tabel 6, stopword berhasil menghilangkan kata-kata umum yang tidak memengaruhi analisis sentimen. Proses ini membuat data teks lebih ringkas dan berkonsentrasi pata kata-kata yang mengandung informasi penting.

6. Stemming

Kata yang memiliki imbuhan diubah menjadi kata dasar selama proses stemming. Tujuan dari langkah ini adalah untuk mengurangi variasi kata dengan makna yang sama, sehingga jumlah fitur yang digunakan dalam proses klasifikasi menjadi lebih efektif. Pustaka Sastrawi untuk bahasa Indonesia digunakan untuk proses stemming dalam penelitian ini.

Tabel 7. Hasil Stemming

Sebelum Stemming	Sesudah Stemming
Menikmati	nikmat
pelayanan	layan

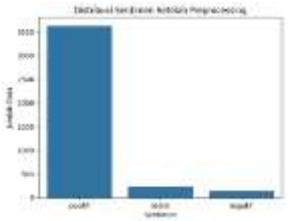
Berdasarkan Tabel 7, proses stemming berhasil mengubah kata berimbuhan menjadi kata dasar. Ini membantu mengurangi variasi kata yang memiliki makna yang sama. Akibatnya, data teks menjadi lebih terstruktur dan dapat digunakan dengan lebih efektif untuk menganalisis sentimen.

4.3. Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) merupakan tahap eksplorasi data yang dilakukan untuk memahami karakteristik dataset sebelum proses pemodelan. Pada tahap ini dilakukan analisis terhadap distribusi data, frekuensi kemunculan kata, serta visualisasi data untuk memperoleh gambaran awal mengenai pola dan karakteristik data. EDA bertujuan membantu peneliti memahami kondisi dataset sehingga proses analisis sentimen dapat dilakukan dengan lebih optimal.

1. Distribusi Sentimen

Visualisasi distribusi sentimen dilakukan untuk mengetahui sebaran jumlah data pada masing-masing kategori sentimen.



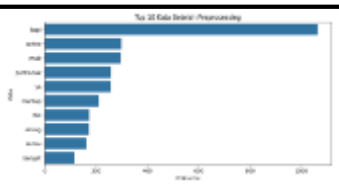
Gambar 2. Distribusi Sentimen Ulasan Pelanggan Aming Coffee

Gambar 2, terlihat bahwa sentimen positif mendominasi dataset dibandingkan sentimen netral dan negatif. Hasil ini menunjukkan bahwa sebagian besar pelanggan memberikan tanggapan yang baik terhadap produk dan layanan Aming Coffee.

2. Frekuensi Kata Terbanyak

Analisis frekuensi kata dan Wordcloud dilakukan setelah preprocessing untuk mengetahui kata yang paling sering muncul dalam ulasan pelanggan.

Copyright © Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)



Gambar 3. Frekuensi Kata Terbanyak pada Ulasan Pelanggan Aming Coffee

Gambar 4. WordCloud Ulasan Pelanggan Aming Coffee Setelah Preprocessing

Berdasarkan Gambar 3 dan Gambar 4, kata yang paling sering muncul antara lain “kopi”, “enak”, “pontianak”, “mantap”, dan “ramai”. Hasil tersebut menunjukkan bahwa pelanggan lebih banyak membahas kualitas kopi, cita rasa produk, serta pengalaman berkunjung ke Aming Coffee. Dominasi kata bernuasa positif juga menunjukkan kecenderungan pelanggan memberikan ulasan yang baik.

4.4. Pembagian Data

Setelah tahap EDA, dataset dibagi menjadi training data dan testing data dengan rasio 80:20. Pembagian ini bertujuan agar model dapat mempelajari pola data pada data pelatihan dan dievaluasi menggunakan data yang belum pernah digunakan sebelumnya.

Tabel 8. Pembagian Dataset

Dataset	Jumlah Data	Persentase
Training Data	3.200	80%
Testing Data	800	20%
Total Data	4.000	100%

Berdasarkan Tabel 8, sebanyak 3.200 data digunakan sebagai training data dan 800 data digunakan sebagai testing data. Pembagian dilakukan secara proporsional untuk mendukung proses pelatihan dan evaluasi model klasifikasi sentimen.

4.5. Pembobotan TF-IDF

Setelah proses distribusi data selesai, langkah selanjutnya ialah pemberian bobot pada kata-kata. Ini dilakukan dengan penerapan metode Frekuensi Istilah-Invers Dokumen (TF-IDF) untuk mengubah teks menjadi bentuk numerik yang dapat diolah oleh algoritma pembelajaran mesin. Metode TF-IDF dianggap lebih efektif dalam membedakan konten dokumen karena memberikan nilai lebih tinggi pada kata-kata yang sering muncul dalam satu dokumen namun jarang terlihat di dokumen lainnya.

Selama tahap preprocessing, proses pembobotan TF-IDF dilakukan terhadap data pelatihan (training data) dan pengujian. Hasil pembobotan menghasilkan matriks TF-IDF berukuran 3200, 2137 untuk data pelatihan dan 800, 2137 untuk data pengujian. Total fitur atau kata unik yang terbentuk dari seluruh dataset setelah proses preprocessing ditunjukkan oleh angka 2.137.

Algoritma Naïve Bayes, Support Vector Machine (SVM), dan Decision Tree C4.5 digunakan sebagai fitur masukan pada proses klasifikasi sentimen setelah hasil pembobotan TF-IDF digunakan.

Tabel 9. Hasil Pembobotan TF-IDF

Dataset	Jumlah Dokumen	Jumlah Fitur
Training Data	3.200	2.137
Testing Data	800	2.137

Pada Tabel 9, proses pembobotan TF-IDF menghasilkan 2.137 fitur yang mewakili kata-kata yang berbeda dalam dataset, seperti yang ditunjukkan dalam Tabel 4.X. Fitur-fitur ini digunakan sebagai representasi numerik dari ulasan pelanggan, sehingga algoritma klasifikasi sentimen dapat digunakan pada tahap berikutnya.

Copyright © Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)

Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)
Vol. 9 No. 3, Juni 2026
P-ISSN 2620-8342
E-ISSN 2621-3052

4.6. Pemodelan Machine Learning

4.6.1. Naïve Bayes + SMOTE

Penelitian ini menggunakan Algoritma Naïve Bayes untuk mengklasifikasikan sentimen ulasan pelanggan Aming Coffee berdasarkan fitur hasil pembobotan TF-IDF. Untuk mengatasi ketidakseimbangan distribusi data sentimen, dilakukan pengujian pada dua skenario, yaitu tanpa SMOTE dan dengan SMOTE. Hasil evaluasi ditampilkan pada tabel berikut.

Tabel 10. Hasil Evaluasi Algoritma Naïve Bayes

Model	Accuracy	Precision	Recall	F1-Score
Naïve Bayes	90,63%	82,13%	90,63%	86,17%
Naïve Bayes + SMOTE	37,38%	86,60%	37,38%	49,62%

Berdasarkan Tabel 10, penerapan SMOTE menyebabkan akurasi model menurun dari 90,63% menjaadi 37,38%, namun nilai Precision meningkat dari 82,13% menjadi 86,60%. Hasil ini menunjukkan bahwa SMOTE membantu model lebih baik dalam mengenali kelas sentimen minoritas meskipun performa keseluruhan menurun.

4.6.2. Hasil Klasifikasi Menggunakan Support Vector Machine (SVM) + SMOTE

Algoritma SVM digunakan untuk mengklasifikasikan sentimen ulasan pelanggan berdasarkan hasil pembobotan TF-IDF. Pengujian dilakukan pada model tanpa SMOTE dan dengan SMOTE untuk melihat pengaruh penyeimbangan data terhadap hasil klasifikasi.

Tabel 11. Hasil Evaluasi Support Vector Machine

Model	Accuracy	Precision	Recall	F1-Score
SVM	90,50%	82,12%	90,50%	86,11%
SVM + SMOTE	38,88%	85,56%	38,88%	51,05%

Berdasarkan Tabel 11, accuracy SVM menurun dari 90,50% menjadi 38,88% setelah penerapan SMOTE, tetapi precision meningkat dari 82,12% menjadi 95,56%. Hasil ini menunjukkan bahwa SMOTE membantu model lebih memperhatikan kelas sentimen yang jumlah datanya lebih sedikit.

4.6.3. Hasil Klasifikasi Menggunakan Decision Tree C4.5 + SMOTE

Algoritma Decision Tree C4.5 ditetapkan untuk mengklasifikasikan sentimen ulasan pelanggan berdasarkan fitur TF-IDF. Pengujian dilakukan untuk membandingkan performa model sebelum dan sesudah penerapan SMOTE.

Tabel 12. Hasil Evaluasi Decision Tree C4.5

Model	Accuracy	Precision	Recall	F1-Score
C4.5	88,25%	83,15%	88,25%	85,36%
C4.5 + SMOTE	37,25%	85,51%	37,25%	49,09%

Berdasarkan Tabel 12, penerapan SMOTE menurunkan accuracy model dari 88,25% menjadi 37,25%, tetapi meningkatkan precision dari 83,15% menjadi 85,51%. Hal ini menunjukkan bahwa SMOTE membantu model mengenali kelas sentimen minoritas dengan lebih baik dibandingkan sebelum dilakukan penyeimbangan data.

4.7. Perbandingan Kinerja Model

Setelah menyelesaikan proses klasifikasi dengan menggunakan algoritma Naive Bayes, Support Vector Machine (SVM), dan Decision Tree C4. 5, dilakukan perbandingan kinerja model menggunakan mematrix Accuracy, Precision, Recall, dan F1-Score untuk menentukan model yang paling efektif dalam mengklasifikasikan sentimen ulasan pelanggan Aming Coffee.

Tabel 13. Perbandingan Hasil Evaluasi Model

Algoritma	Accuracy	Precision	Recall	F1-Score
Naïve Bayes	90,63%	82,13%	90,63%	86,17%
SVM	90,50%	82,11%	90,50%	86,11%
C4.5	88,25%	83,15%	88,25%	85,36%
Naïve Bayes + SMOTE	37,38%	86,61%	37,38%	49,62%
SVM + SMOTE	38,88%	85,56%	38,88%	51,05%
C4.5 + SMOTE	37,25%	85,51%	37,25%	49,09%

Berdasarkan data yang disajikan pada Tabel 13, terlihat adanya kontras performa yang sangat signifikan antara model klasifikasi yang dilatih tanpa menggunakan teknik penanganan ketidakseimbangan data (class

Copyright © Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)

Accredited SINTA 5
72/E/KPT/2024

Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)
Vol. 9 No. 3, Juni 2026
P-ISSN 2620-8342
E-ISSN 2621-3052

imbalance) dengan model yang telah diintegrasikan dengan Synthetic Minority Over-sampling Technique (SMOTE).

Pada eksperimen awal tanpa SMOTE, ketiga algoritma—Naïve Bayes, SVM, dan Pohon Keputusan C4.5—menghasilkan tingkat akurasi global yang sangat tinggi, yaitu berturut-turut sebesar 90,63%, 90,50%, dan 88,25%. Namun, pengujian lebih lanjut melalui analisis confusion matrix membuktikan bahwa tingginya nilai akurasi tersebut merupakan sebuah pseudo-performa atau ilusi optik klasifikasi yang disebabkan oleh dominasi ekstrem data sentimen positif (kelas mayoritas) yang mencapai 725 dari total 800 data uji. Model mengalami major class bias, di mana sistem cenderung menebak seluruh dokumen ke dalam kelas positif dan gagal total dalam mengidentifikasi ulasan yang bersifat negatif maupun netral (menghasilkan nilai recall dan f1-score sebesar 0,00 pada kelas minoritas).

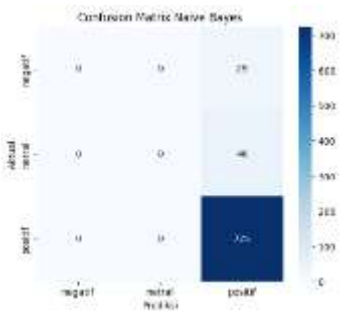
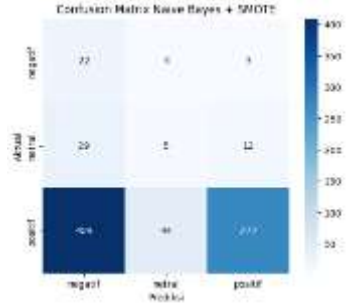
Untuk mengatasi kecacatan logis tersebut, penerapan SMOTE dilakukan pada data pelatihan untuk menyintesis data tiruan pada kelas negatif dan netral hingga mencapai proporsi yang seimbang. Hasil rekonstruksi model menunjukkan terjadinya penurunan akurasi global yang drastis pada seluruh algoritma, di mana Naïve Bayes + SMOTE turun menjadi 37,38%, SVM + SMOTE menjadi 38,88%, dan C4.5 + SMOTE menjadi 37,25%.

Penurunan nilai akurasi ini merupakan sebuah fenomena ilmiah yang wajar dan aman secara metodologis dalam data science. Ketika data dibuat seimbang, model dipaksa untuk benar-benar mempelajari karakteristik semantik dan variasi kata dari setiap kelas sentimen. Mengingat ulasan pelanggan pada platform Google Maps Aming Coffee ditulis dalam bahasa alami yang santai, sering kali terjadi tumpang tindih fitur (overlapping vocabulary) antara ulasan netral dan negatif (misalnya penggunaan kata-kata ambigu atau konteks penulisan yang mirip). Kondisi ini meningkatkan sensitivitas model terhadap kelas minoritas secara agresif (over-generalization), yang tercermin pada Gambar Confusion Matrix SVM + SMOTE di mana model berhasil mengenali 22 data aktual negatif dengan benar, namun di sisi lain ikut mengklasifikasikan sebagian besar data aktual positif ke dalam kelas negatif.

Meskipun akurasi global mengalami reduksi, parameter Precision mengalami peningkatan yang konsisten pada semua model setelah penerapan SMOTE, dengan nilai tertinggi dicapai oleh algoritma Naïve Bayes + SMOTE sebesar 86,60%, diikuti oleh SVM + SMOTE sebesar 85,56%. Kenaikan nilai precision ini menjadi indikator vital bahwa tingkat kepercayaan (confidence rate) model dalam memvalidasi ulasan sentimen meningkat. Dalam konteks implementasi bisnis dan evaluasi kualitas layanan Aming Coffee, kemampuan model dalam mendeteksi dan mengisolasi ulasan negatif (komplain pelanggan) secara presisi jauh lebih bernilai strategis daripada sekadar mengejar nilai akurasi global yang tinggi tetapi bias.

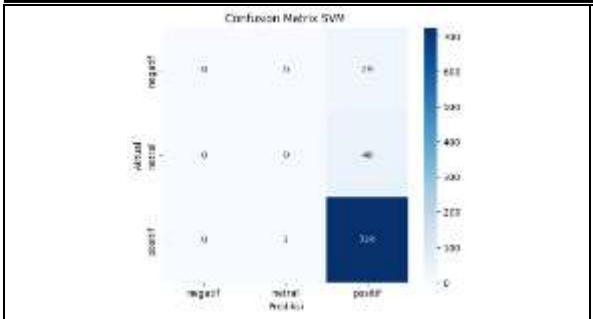
Secara keseluruhan, integrasi metode pembobotan TF-IDF, SMOTE, dan algoritma machine learning (khususnya SVM + SMOTE yang mencatatkan nilai F1-Score tertinggi sebesar 0.510469) berhasil menghasilkan model analisis sentimen yang lebih objektif, sahih, jujur, serta memiliki kemampuan generalisasi yang riil dalam memetakan polaritas opini pelanggan Aming Coffee.

Untuk memperkuat hasil evaluasi pada Tabel13, dilakukan analisis menggunakan Confusion matrix digunakan untuk melihat kemampuan model dalam mengklasifikasi masing-masing kelas sentimen (positif, netral, dan negatif) secara lebih rincu, sehingga tidak hanya berfokus pada nilai akurasi secara keseluruhan.

Confusion Matrix tanpa SMOTE	Confusion Matrix menggunakan SMOTE
	
Gambar 5. Confusion Matrix Naïve Bayes	Gambar 6. Confusion Matrix Naïve Bayes + SMOTE

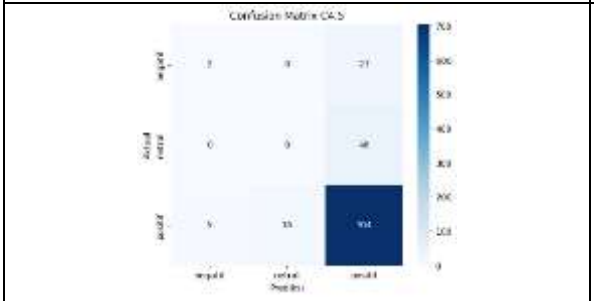
Copyright © Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)

Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)
Vol. 9 No. 3, Juni 2026
P-ISSN 2620-8342
E-ISSN 2621-3052

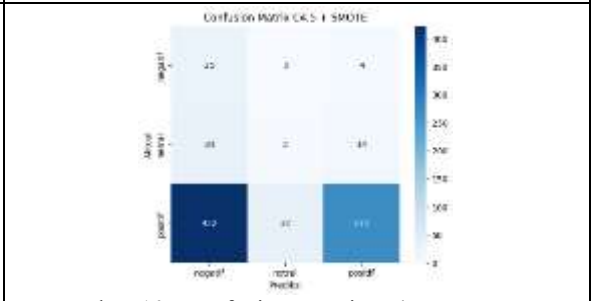


Gambar 7. Confusion Matrix SVM

Gambar 8. Confusion Matrix SVM + SMOTE



Gambar 9. Confusion Matrix C4.5



Gambar 10. Confusion Matrix C4.5 + SMOTE

Berdasrkan Gambar 5-10, model tanpa SMOTE menunjukkan kecenderungan memprediksi sebagain besar data ke kelas positif sehingga kemampuan mengenali sentimen negatif dan netral masih rendah. Setelah penerapan SMOTE, distribusi prediksi menjadi lebih merata dan model mulai mampu mengenali kelas minoritas dengan lebih baik. Meskipun demikian, peningkatan kemampuan pada kelas minoritas menyebabkan penurunan nilai akurasi secara keseluruhan. Dari seluruh model yang diuji, SVM + SMOTE memperoleh nilai F1-Score tertinggi sebesar 51,05%, sedangkan Naïve Bayes tanpa SMOTE tetap menghasilkan akurasi tertinggi sebesar 90,63%.

4.8. Analisis Hasil Penelitian

Hasil analisis sentimen menunjukkan bahwa sebagian besar ulasan pelanggan Aming Coffee memiliki sentimen positif, seperti yang ditunjukkan oleh distribusi data yang didominasi oleh sentimen positif dibandingkan dengan sentimen netral dan negatif. Dominasi sentimen positif menunjukkan bahwa sebagian besar pelanggan memiliki tanggapan yang baik terhadap produk dan layanan yang ditawarkan oleh Aming Coffee.

Hasil analisis frekuensi kata menunjukkan bahwa kata yang paling sering muncul dalam ulasan pelanggan adalah "kopi", diikuti oleh "coffee", "enak", "mantap", dan "ramai". Ini menunjukkan bahwa pelanggan banyak memberikan ulasan tentang kualitas kopi, cita rasa produk, dan pengalaman berkunjung ke Aming Coffee.

Selain itu, visualisasi *WordCloud* menunjukkan bahwa kata-kata positif seperti "kopi", "enak", "mantap", dan "banget" adalah kata-kata yang paling banyak digunakan. Temuan ini menunjukkan bahwa pelanggan lebih cenderung puas dengan produk yang diberikan. Berbicara tentang "ramai" menunjukkan bahwa Aming Coffee telah menjadi salah satu tempat yang cukup diminati oleh pelanggan untuk berkumpul dan bersantai.

Secara keseluruhan, hasil analisis sentimen menunjukkan bahwa pelanggan memiliki persepsi yang positif terhadap Aming Coffee. Temuan ini dapat membantu Aming Coffee mempertahankan kualitas produk dan layanannya yang telah mendapatkan tanggapan positif dari pelanggan.

5. Kesimpulan

Studi ini berhasil melaksanakan analisis sentimen terhadap ulasan pelanggan Aming Coffee yang diambil dari Google Maps dengan menerapkan metode Term Frequency-Inverse Document Frequency (TF-IDF) serta algoritma Naive Bayes, Support Vector Machine (SVM), dan Decision Tree C4. 5. Sebelum melakukan klasifikasi, data melalui proses preprocessing yang meliputi perbaikan, case folding, tokenisasi, normalisasi, penghilangan stopword, dan stemming. Langkah ini bertujuan untuk menghasilkan data yang lebih bersih, terstruktur, dan siap digunakan dalam proses klasifikasi. Hasil dari analisis eksploratif data (EDA) menunjukkan bahwa ulasan positif mengambil porsi terbesar dalam distribusi. Melalui analisis frekuensi kata dan visualisasi WordCloud, istilah-istilah seperti "kopi", "enak", "mantap", dan "ramai" muncul paling sering dalam ulasan dari pelanggan. Temuan ini menandakan bahwa pelanggan banyak memberikan tanggapan mengenai kualitas kopi, rasa produk, dan pengalaman mereka ketika berkunjung ke Aming Coffee. Setelah proses pembobotan TF-IDF selesai, terdapat 2. 137 fitur yang diidentifikasi dan digunakan sebagai masukan

Copyright © Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)

untuk klasifikasi. Hasil pengujian mengindikasikan bahwa algoritma Naive Bayes menunjukkan kinerja terbaik dengan tingkat akurasi sebesar 90,63%, diikuti oleh SVM dengan 90,50%, dan C4. 5 pada 88,25%. Meski ketiga model menunjukkan tingkat akurasi yang cukup tinggi, hasil klasifikasi dan Analisis confusion matrix menunjukkan bahwa model masih mengalami kesulitan dalam mengidentifikasi sentimen negatif dan netral. Secara keseluruhan, analisis menunjukkan bahwa mayoritas pelanggan memberikan ulasan positif terhadap Aming Coffee. Hal ini mengindikasikan bahwa mutu produk dan layanan yang disediakan oleh Aming Coffee memperoleh tanggapan positif dari para pelanggan. Namun, salah satu faktor yang memengaruhi kemampuan model untuk mendeteksi kelas sentimen minoritas adalah ketidakseimbangan distribusi data.

Daftar Pustaka

[1] J. Jurnal and T. Informasi, "ANALISIS SENTIMEN ULASAN KONSUMEN MENGGUNAKAN ALGORITMA TF-ID UNTUK (STUDI KASUS : GUNTHEM PREMIUM COFFEE)," vol. 16, no. 2, pp. 8–16, 2025.
[2] J. Jurnal and T. Informasi, "ANALISIS SENTIMEN KEPUASAN PELANGGAN PADA KOPISHOP TELON COFFEE MENGGUNAKAN METODE TF-IDF," vol. 17, no. 1, pp. 49–59, 2026.
[3] A. Khusna and Y. Yamasari, "Analisis Sentimen Ulasan Google Maps Menggunakan Long Short-Term Memory (LSTM) (Studi Kasus : Kafe di Surabaya)," vol. 07, pp. 446–454, 2025.
[4] J. A. Putra, A. Dharmawan, and J. Gondohanindijo, "Sentimen analisis aplikasi digitalent mobile menggunakan naïve bayes dan svm dengan ekstraksi fitur tf-idf sentimen analysis digitalent mobile application using naïve bayes and svm with tf-idf fitur extraction," vol. 7, 2024.
[5] S. Lestanti and S. N. Budiman, "Analisis Sentimen Berdasarkan Hasil Review Lokasi Google Map Menggunakan Natural Language Toolkit TextBlob dan Naïve Bayes," vol. 5, no. December, pp. 114–126, 2024.
[6] U. Khultsum, E. Rahmawati, A. Purnamawati, and B. R. Annajib, "Implementasi Support Vector Machine dan Resampling dalam Analisis Ulasan Pengguna Google Maps," vol. 9, no. 2, pp. 158–169, 2025.
[7] I. Zulfahmi, "Analisis Sentimen Aplikasi PLN Mobile Menggunakan Metode Decision Tree," vol. 3, no. 1, 2024.
[8] P. A. Permatasari and L. Jasa, "Survei Tentang Analisis Sentimen Pada Media Sosial," vol. 20, no. 2, 2021.
[9] S. Informasi, F. I. Komputer, U. Duta, and B. Surakarta, "PADA ULASAN GOOGLE MAPS RESTORAN AL-GHIFF STEAK MENGGUNAKAN MODEL INDOBERT," vol. 10, no. 2, pp. 336–343, 2025.
[10] A. R. Satria and S. Adinugroho, "Analisis Sentimen Ulasan Aplikasi Mobile menggunakan Algoritma Gabungan Naïve Bayes dan C4 . 5 berbasis Normalisasi Kata Levenshtein," vol. 4, no. 11, pp. 4154–4163, 2020.
[11] R. Ramadhan, Y. A. Sari, and P. P. Adikara, "Perbandingan Pembobotan Term Frequency-Inverse Document Frequency dan Term Frequency-Relevance Frequency terhadap Fitur N-Gram pada Analisis Sentimen," vol. 5, no. 11, pp. 5075–5079, 2021.
[12] S. Nurul, J. Fitriyyah, N. Safriadi, and E. E. Pratama, "Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes," vol. 5, no. 3, pp. 279–285, 2019.
[13] H. Al, R. Harpizon, R. Kurniawan, and I. Iskandar, "Analisis Sentimen Komentar Di YouTube Tentang Ceramah Ustadz Abdul Somad Menggunakan Algoritma Naïve Bayes," vol. 5, no. 1, pp. 131–140, 2022.
[14] H. Himawan, W. Kaswidjanti, A. Sentimen, M. Sosial, and L. Based, "METODE LEXICON BASED DAN SUPPORT VECTOR MACHINE UNTUK MENGANALISIS SENTIMEN PADA MEDIA SOSIAL SEBAGAI REKOMENDASI OLEH-OLEH FAVORIT," vol. 2018, no. November, pp. 235–244, 2018.
[15] R. Nanda, E. Haerani, S. K. Gusti, and S. Ramadhani, "Klasifikasi Berita Menggunakan Metode Support Vector Machine," vol. 5, no. 2, pp. 269–278, 2022.
[16] J. T. Terapan, "Jurnal Teknologi Terapan & Sains 4.0," vol. 5, no. 2, 2024.
[17] R. Hidayat, A. B. Hakim, and R. Nugraha, "Perbandingan Metode Naïve Bayes Dan Decision Tree C4 . 5 untuk Analisis Sentimen Produk Es Teh Indonesia di Media Sosial Twitter," 2024.
[18] A. Prasetio, M. H. Hasibuan, and P. Sitompul, "Simulasi Penerapan Metode Decision Tree (C4 . 5) Pada Penentuan Status Gizi Balita," vol. 4, no. 3, pp. 209–214, 2021.
[19] C. Cahyaningtyas, Y. Nataliani, I. R. Widiasari, F. T. Informasi, U. Kristen, and S. Wacana, "Analisis sentimen pada rating aplikasi Shopee menggunakan metode Decision Tree berbasis SMOTE," vol. 18, no. 2, pp. 173–184, 2021.
[20] B. Hakim, "ANALISA SENTIMEN DATA TEXT PREPROCESSING PADA DATA MINING DENGAN MENGGUNAKAN MACHINE LEARNING DATA TEXT PRE-PROCESSING SENTIMENT ANALYSIS IN DATA MINING USING MACHINE LEARNING School of Computer

Copyright © Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)

Science and Technology , Harbin Institute of Technology," vol. 4, no. 2, pp. 16–22, 2021.
[21] P. Indobert, "Sentiment Analysis of Coretax : A Comparison of Manual , Transformers- Based , and Lexicon-Based Data Labeling on IndoBERT Performance Analisis Sentimen Coretax : Perbandingan Pelabelan Data Manual ," vol. 5, no. July, pp. 1037–1048, 2025.
[22] F. Alzami, E. D. Udayanti, D. P. Prabowo, and R. A. Megantara, "Document preprocessing with TF-IDF to improve the polarity classification performance of unstructured sentiment analysis," vol. 4, no. 3, 2020.

Copyright © Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)