

KLASIFIKASI SENTIMEN ULASAN COFFEE SHOP DI PONOROGO BERBASIS NAÏVE BAYES

Faisal Reza Pradhana^{1*}, Eko Prasetyo Widhi², Niken Ratna Sari³, Muhammad Amrico Putra Nurcahyo⁴

^{1,2,3,4}Program Studi Teknik Informatika, Fakultas Sains & Teknologi, Universitas Darussalam Gontor

E-mail: faisalrezapradhana@unida.gontor.ac.id^{1*}, ekoprasetyowidhi@unida.gontor.ac.id²,
ratna.sari@unida.gontor.ac.id³, amriconurcahyo96@gmail.com⁴

INFO ARTIKEL

Sejarah Artikel

Diterima : 30/11/2025

Direvisi : 12/12/2025

Diterbitkan : 15/12/2025

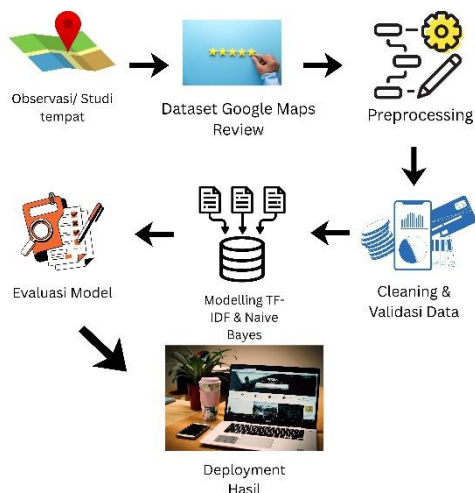
*Corresponding author

faisalrezapradhana@unida.gontor.ac.id

DOI: 10.70247/jumistik.v4i2.239

<https://ojs.amiklps.ac.id>

GRAPHICAL ABSTRACT



ABSTRACT

This study aims to analyze customer sentiment toward three coffee shops with the highest number of reviews in Ponorogo, namely Gayeng Billiard & Naïve, Kedai Sor Sawo, and Origin Lab Coffee. Customer review data were collected through web scraping from Google Maps and processed using the CRISP-DM framework, which includes text cleaning, normalization, tokenization, and feature extraction using the TF-IDF method. The Naïve Bayes algorithm was employed as the primary classification model to categorize sentiment into positive, negative, and neutral classes. Additionally, this research compares sentiment tendencies across the three coffee shops to identify differences in customer perceptions based on the available reviews. The analysis results show that most reviews exhibit positive sentiment, with Variations in customer satisfaction across the three coffee shops. The Naïve Bayes model achieved a satisfactory level of accuracy in classifying sentiments. These findings are expected to provide practical insights for coffee shop owners in understanding customer needs, evaluating service quality, and formulating strategies to enhance both operational performance and customer experience.

Keywords: Sentiment Analysis, Naïve Bayes, Coffee Shops, Costumer Reviews

ABSTRAK

Penelitian ini bertujuan untuk menganalisis sentiment pelanggan terhadap tiga kedai kopi yang memiliki jumlah ulasan terbanyak di Ponorogo, yaitu Gayeng Billiard & Naïve, Kedai Sor Sawo, dan Origin Lab Coffee. Data ulasan diperoleh melalui proses web scraping dari platform Google Maps, kemudian diolah menggunakan tahapan CRISP-DM yang mencakup pembersihan teks, normalisasi, tokenisasi, serta ekstraksi fitur menggunakan metode TF-IDF. Algoritma Naïve Bayes digunakan sebagai model klasifikasi utama untuk mengelompokkan sentiment menjadi positif, sentimen, dan netral. Penelitian ini juga membandingkan kecenderungan sentiment antar ketiga kedai kopi untuk melihat perbedaan persepsi pelanggan berdasarkan ulasan yang ada. Hasil analisis menunjukkan bahwa mayoritas ulasan cenderung positif dan terdapat perbedaan sentiment kepuasan pelanggan pada masing-masing kedai. Model Naïve Bayes yang digunakan menghasilkan sentiment akurasi yang cukup baik dalam mengklasifikasikan sentiment secara keseluruhan. Temuan ini diharapkan dapat memberikan manfaat praktis bagi pemilik usaha coffee shop dalam memahami kebutuhan pelanggan, melakukan evaluasi layanan, serta merumuskan strategi peningkatan kualitas dari sisi operasional maupun pengalaman pelanggan.

Kata kunci: Analisis Sentimen, Naïve Bayes, Kedai Kopi, Ulasan Pelanggan

© 2025 Penerbit STMIK Amika Soppeng. All rights reserved

PENDAHULUAN

Perkembangan teknologi informasi di era digital telah mengubah cara konsumen memberikan evaluasi terhadap layanan maupun produk. Ulasan pelanggan (*customer review*) yang tersebar melalui berbagai platform digital, khususnya *Google Maps*, kini menjadi salah satu sumber informasi penting yang memengaruhi reputasi serta daya saing sebuah bisnis [1]. Ulasan tersebut tidak hanya mencerminkan pengalaman pelanggan, tetapi juga memberikan sentimen mengenai kepuasan dan kualitas layanan dari suatu usaha. Pada kuliner, khususnya *coffee shop*, fenomena ini semakin menonjol seiring meningkatnya popularitas tempat nongkrong sebagai ruang bagi Masyarakat terutama kalangan muda [2].

Di Kabupaten Ponorogo, pertumbuhan *coffee shop* meningkat secara signifikan, ditandai dengan lebih dari 120 kedai kopi yang beroperasi dan aktif menerima ulasan melalui *Google Maps*. Namun, permasalahan utama yang sering ditemukan adalah adanya ketidaksesuaian antara nilai rating dan isi ulasan. Banyak pelanggan memberikan bintang lima tetapi menyampaikan kritik dalam teks ulasannya, atau sebaliknya memberikan rating rendah namun isi komentar bersifat positif. Kondisi ini menunjukkan bahwa rating numerik tidak cukup untuk menggambarkan persepsi pelanggan secara akurat sehingga dibutuhkan analisis terhadap isi teks ulasan secara lebih mendalam [3].

Analisis sentimen merupakan salah satu metode dalam *Natural Language Processing (NLP)* yang bertujuan mengekstraksi opini dan mengelompokkan sentiment pelanggan ke dalam kategori positif, negatif, atau netral [4]. Berbagai metode *machine learning* telah digunakan dalam analisis sentimen, seperti *Support Vector Machine (SVM)*, *K-Nearest Neighbor (K-NN)*, dan *Naïve Bayes*. Di antara metode tersebut, *Naïve Bayes Classifier* banyak digunakan karena sederhana, efisien, dan mampu memberikan performa klasifikasi yang kompetitif pada data teks berdimensi tinggi. Berbagai penelitian sebelumnya telah membuktikan keefektifan metode ini, di antaranya pada ulasan *e-commerce*, aplikasi digital, hingga layanan restoran. Namun, penelitian terkait analisis sentimen khusus untuk *coffee shop* di Ponorogo belum pernah dilakukan, terutama dengan objek lebih dari satu kedai kopi.

Permasalahan tersebut melatarbelakangi dilakukannya penelitian ini, yaitu perlunya pemahaman berbasis data terhadap persepsi pelanggan terhadap layanan *coffee shop* di Ponorogo. Ketersediaan data ulasan yang melimpah pada *Google Maps* dapat dimanfaatkan sebagai sumber informasi strategis untuk mengevaluasi kekuatan dan kelemahan suatu kedai kopi. Penelitian ini secara khusus berfokus pada tiga *coffee shop* dengan jumlah ulasan terbanyak, yaitu *Gayeng*

Billiard & Café, *Kedai Sor Sawo*, dan *Origin Coffee Lab*, sehingga analisis yang dilakukan dapat memberikan gambaran sentimen yang lebih representatif dan komparatif.

Penelitian ini bertujuan untuk mengklasifikasikan sentimen pelanggan berdasarkan ulasan *Google Maps* menggunakan metode *Naïve Bayes*, serta untuk membandingkan hasil sentimen pada tiga *coffee shop* tersebut. Selain itu, penelitian ini diharapkan dapat memberikan informasi yang bermanfaat bagi pemilik usaha dalam meningkatkan kualitas layanan, memahami kebutuhan pelanggan, dan menyusun strategi pengembangan bisnis yang berorientasi pada pengalaman konsumen. Secara akademik, penelitian ini berkontribusi dalam memperkaya kajian analisis sentimen pada sektor kuliner yang masih jarang diteliti, khususnya di wilayah Ponorogo.

TINJAUAN PUSTAKA

Analisis Sentimen merupakan salah satu bidang penting dalam *Natural Language Processing (NLP)* yang berfokus pada pemahaman persepsi, opini, serta emosi pengguna melalui teks yang dipublikasikan secara daring. Seiring meningkatnya penggunaan platform digital sebagai sarana evaluasi layanan, analisis sentimen banyak dimanfaatkan untuk menilai kualitas pelayanan bisnis, termasuk pada sektor kuliner seperti *coffee shop*. Berbagai penelitian sebelumnya telah mengkaji penggunaan algoritma *machine learning* untuk mengklasifikasikan teks ulasan, baik untuk kepentingan evaluasi layanan maupun peningkatan strategi bisnis. Penelitian-penelitian tersebut menjadi landasan penting bagi penelitian ini, terutama terkait penggunaan metode *Naïve Bayes* dalam mengolah ulasan pelanggan dari berbagai platform.

Salah satu penelitian terdahulu dilakukan oleh Prasetyo (2021) yang berjudul Analisis Sentimen dan Kategorisasi Review Pelanggan pada *Naïve Kopi Paste* dengan Metode *Naïve Bayes* dan *KNN* [5]. Penelitian ini bertujuan untuk menganalisis persepsi pelanggan serta mengelompokkan kategori ulasan berdasarkan produk tertentu. Penelitian ini menggunakan dua algoritma, yaitu *Naïve Bayes Classifier* untuk mengklasifikasikan sentimen positif dan sentimen negatif, serta *K-Nearest Neighbor (KNN)* untuk mengelompokkan topik ulasan. Hasil penelitian menunjukkan bahwa *Naïve Bayes* memiliki akurasi lebih tinggi dari *KNN*, yakni 93,33% berbanding 77%. Studi ini relevan dengan penelitian yang dilakukan saat ini karena sama-sama memanfaatkan *Naïve Bayes* untuk klasifikasi sentimen. Perbedaannya terletak pada ruang lingkup: Prasetyo hanya meneliti satu kedai kopi, sedangkan penelitian ini menganalisis beberapa *coffee shop* sekaligus. Selain itu, penelitian saat ini menambah kategori sentimen netral agar klasifikasi lebih merepresentasikan variasi opini pelanggan.

Penelitian lain dilakukan oleh Tia Anggita Sari et al. (2021) dengan judul Analisis Sentimen Ulasan Pelanggan pada Aplikasi Fore Coffee Menggunakan Metode Naïve Bayes[6]. Penelitian ini menggunakan 1801 ulasan dari aplikasi Fore Coffee dan mengklasifikasikan sentimen menjadi positif dan negatif. Dengan pendekatan *supervised learning*, penelitian ini menunjukkan bahwa Naïve Bayes mampu menghasilkan akurasi 74,28%. Studi ini menunjukkan bahwa Naïve Bayes tetap efektif meskipun digunakan pada volume data besar dan konteks brand nasional. Relevansinya dengan penelitian saat ini terletak pada penggunaan metode klasifikasi yang sama dan tujuan yang serupa, yaitu memahami persepsi pelanggan terhadap layanan *coffee shop*. Namun, penelitian Fore Coffee mengambil data dari aplikasi internal, sementara penelitian ini menggunakan Google Maps yang memiliki variasi ulasan yang lebih bebas dan tidak terkontrol.

Selain itu, penelitian oleh Salsabila et al. (2022) mengenai analisis sentimen pelanggan Tokopedia menggunakan Naïve Bayes juga memberikan temuan penting mengenai efektivitas metode ini dalam mengklasifikasikan opini pengguna[7]. Meskipun tidak berfokus pada sektor kuliner, penelitian ini menegaskan bahwa Naïve Bayes sangat cocok digunakan untuk teks pendek yang ditulis secara spontan oleh pengguna, seperti ulasan Google Maps. Temuan ini memperkuat dasar pemilihan metode untuk penelitian ini, mengingat data ulasan pelanggan *coffee shop* juga bersifat singkat, informatif, dan sering kali emosional.

Pada penelitian lain oleh Muslim et al. (2023) yang membandingkan metode Naïve Bayes dan KNN dalam analisis sentimen ulasan aplikasi CapCut, ditemukan bahwa Naïve Bayes memberikan kinerja yang lebih baik[8]. Studi ini memperkuat bukti bahwa Naïve Bayes unggul pada konteks teks ulasan yang memiliki variasi sentimen dan kata kunci emosional. Penelitian ini memiliki relevansi metodologis karena mendukung pemilihan Naïve Bayes sebagai algoritma utama dalam penelitian ini, yang juga menggunakan pendekatan klasifikasi berbasis data teks tidak terstruktur.

Dalam konteks penelitian yang secara khusus melibatkan lebih dari satu objek, penelitian oleh Iskandar dan Nataliani (2021)[9] memberikan 3 sentimen yang cukup mendekati. Meskipun objek penelitian mereka bukan *coffee shop*, namun mereka membandingkan sentimen terhadap beberapa produk *gadget* yang berbeda. Penelitian tersebut menunjukkan bahwa komparasi hasil sentimen antar objek dapat memberikan wawasan strategis yang lebih luas dibandingkan hanya meneliti satu objek saja. Hal ini selaras dengan penelitian ini, yang menganalisis tiga *coffee shop* sekaligus: *Gayeng Billiard & Cafe*, *Origin Coffee Lab*, dan *Kedai Sor Sawo*. Dengan menganalisis beberapa kedai, penelitian ini

dapat menampilkan sentimen menyeluruh mengenai preferensi konsumen di Ponorogo berdasarkan uraian ulasan yang ada di Google Maps.

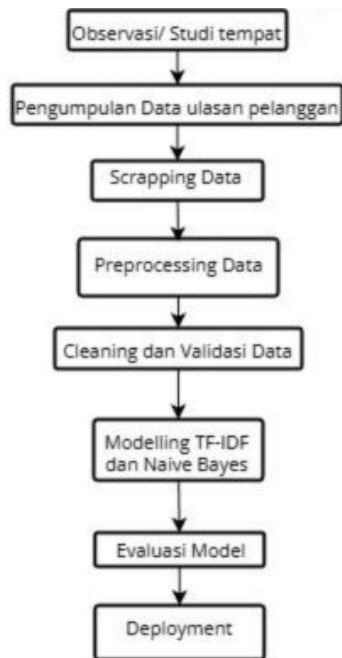
Dari keseluruhan penelitian terdahulu tersebut, dapat disimpulkan bahwa metode Naïve Bayes secara konsisten memberikan performa yang baik dalam analisis sentimen, terutama pada data teks pendek dan informal seperti ulasan pelanggan pada platform digital. Perbedaan konteks, platform data, jumlah objek penelitian, dan kategori sentimen menjadi pembeda utama antara penelitian terdahulu dan penelitian ini. Penelitian saat ini mengambil posisi unik karena memfokuskan analisis pada tiga *coffee shop* dengan jumlah ulasan terbanyak di Kabupaten Ponorogo, menggunakan data dari Google Maps yang memiliki keragaman sentimen dan dinamika opini yang tinggi. Selain itu, penambahan kategori sentimen netral menjadikan penelitian ini lebih komprehensif dibandingkan penelitian terdahulu yang hanya menggunakan dua kategori.

Dengan demikian, penelitian ini tidak hanya melanjutkan temuan dari studi-studi sebelumnya, namun juga memberikan kontribusi baru dalam konteks sentiment kuliner di Ponorogo, mengingat masih terbatasnya penelitian yang membandingkan beberapa *coffee shop* sekaligus menggunakan metode Naïve Bayes. Kajian sentimen ini menjadi landasan teoritis yang kuat bagi penelitian yang dilakukan, baik dari sisi metodologi *data mining* maupun penerapan analisis sentimen dalam bidang kuliner.

METODOLOGI PENELITIAN

Tahapan Penelitian

Penelitian ini dilaksanakan melalui serangkaian tahapan yang terstruktur sebagaimana ditunjukkan pada Gambar 1. Proses dimulai dengan observasi dan studi awal terhadap tiga kedai kopi yang dipilih, kemudian dilanjutkan dengan pengumpulan data ulasan dari Google Maps melalui metode *web scraping*. Data yang diperoleh selanjutnya melalui tahap *preprocessing*, meliputi pembersihan teks, normalisasi, dan validasi kualitas data. Setelah data bersih dan siap digunakan, dilakukan proses pemodelan menggunakan representasi fitur *TF-IDF* dan algoritma Naïve Bayes untuk klasifikasi sentimen. Model yang terbentuk kemudian diuji dan dievaluasi menggunakan beberapa metrik evaluasi dan akurasinya guna memastikan performanya. Tahap akhir penelitian adalah *deployment*, yaitu menyajikan hasil analisis dalam bentuk *dashboard* web interaktif agar dapat dimanfaatkan oleh pemilik kedai kopi maupun pihak terkait.



Gambar1. Alur Penelitian
Sumber: Dokumentasi Pribadi

a. Observasi/ Studi tempat

Metodologi penelitian ini disusun secara sistematis mengikuti rangkaian tahapan yang menggambarkan proses penelitian dari tahap awal hingga tahap akhir. Tahapan ini dirancang untuk memastikan data yang diperoleh valid, proses analisis berjalan optimal, dan hasil penelitian dapat dipertanggungjawabkan secara ilmiah. Penelitian diawali dengan kegiatan observasi atau studi tempat yang dilakukan untuk memahami konteks penelitian secara langsung. Observasi ini bertujuan untuk mengidentifikasi kondisi nyata yang terjadi di lapangan, khususnya terkait aktivitas operasional, karakteristik pelanggan, serta jenis layanan yang ditawarkan oleh tiga coffee shop di Ponorogo yaitu:

1. Gayeng Billiard & Café
2. Kedai Sor Sawo
3. Origin Coffee Lab

Observasi ini memberikan sentimen awal mengenai alasan tingginya jumlah ulasan pelanggan serta bagaimana perbedaan karakteristik antar kedai dapat memengaruhi sentimen yang muncul pada platform digital.

b. Pengumpulan Data

Setelah proses observasi dilakukan, penelitian dilanjutkan dengan tahap pengumpulan data ulasan pelanggan. Data ini diperoleh dari platform Google Maps karena platform tersebut merupakan sumber informasi terbuka yang banyak digunakan oleh pelanggan dalam memberikan penilaian terhadap layanan suatu bisnis. Pengumpulan data dilakukan secara eksplisit terhadap ulasan pelanggan yang

tersedia untuk ketiga kedai kopi tersebut. Pemilihan Google Maps dilakukan karena kelengkapan informasinya, kemudahan akses, serta relevansinya dalam mencerminkan opini pelanggan secara langsung. Data yang dikumpulkan mencakup teks ulasan dan rating yang diberikan oleh pengguna.

c. Scrapping Data

Proses web scrapping dilakukan menggunakan Python dengan alat bantu seperti Selenium atau BeautifulSoup, karena Google Maps tidak menyediakan API resmi untuk data ulasan. Scrapping dilakukan dengan mengakses halaman ulasan masing-masing coffee shop, kemudian seluruh konten HTML dibaca dan dikonversi menjadi dataset mentah berformat Excel atau CSV.

Dataset hasil scrapping biasanya masih sangat besar dan berisi banyak kolom yang tidak relevan, seperti ID elemen halaman, metadata bawaan HTML, kode struktur, tanda sentimen, maupun elemen non-teks. Oleh karena itu, dataset berupa ribuan baris ini masih perlu diproses lebih lanjut sebelum dapat dianalisis lebih jauh.

d. Preprocessing Data

Pada penelitian ini, preprocessing tidak langsung dilakukan terhadap teks komentar, tetapi dimulai dengan membersihkan dataset hasil scrapping terlebih dahulu. Dataset awal berisi ribuan baris termasuk elemen-elemen yang sebenarnya bukan ulasan pelanggan, misalnya baris navigasi, konten kosong, baris duplikat, karakter yang tidak terdeteksi, maupun elemen bawaan struktur Google Maps.

Tahap preprocessing ini menghasilkan data mentahan yang benar-benar hanya berisi komentar pelanggan, sehingga jumlah baris dataset yang semula ribuan menyusut menjadi ratusan. Proses ini meliputi:

1. Penghapusan kolom tidak relevan seperti ID DOM HTML, kode struktur, dan atribut otomatis.
2. Penyaringan baris valid sehingga hanya tersisa komentar asli yang ditulis pelanggan.
3. Penghapusan duplikasi jika ada ulasan yang terekstrak dua kali.
4. Eliminasi baris kosong atau baris yang hanya berisi sentimen.

Setelah memperoleh dataset mentahan yang lebih bersih, barulah dilakukan tahapan pembersihan teks, seperti case folding, tokenizing, stopword removal, dan stemming menggunakan library Sastrawi. Hasil akhir preprocessing berupa dataset komentar yang siap dilabeli sentimen dan dimodelkan lebih lanjut.

e. Cleaning dan Validasi Data

Setelah *preprocessing* dilakukan, penelitian memasuki tahap *cleaning* dan validasi data. Pada tahap ini, data dicek untuk memastikan tidak ada duplikasi ulasan, data kosong, atau entri yang tidak valid. Data juga divalidasi untuk memastikan kesesuaian antara kelas sentimen (positif, negatif, netral) dengan ulasan yang telah diberi label melalui proses anotasi manual. Validasi ini penting agar model tidak dilatih menggunakan data yang salah, yang dapat menurunkan akurasi model secara keseluruhan.

f. Modelling TF-IDF dan Naïve Bayes

Setelah data *dipreprocessing*, penelitian memasuki tahap *Modelling*. Fitur teks ulasan diubah menjadi representasi numerik menggunakan metode *Term Frequency-Inverse Document Frequency (TF-IDF)*. Teknik *TF-IDF* digunakan untuk memberikan bobot pada kata berdasarkan pentingnya kata dalam dokumen dan kelangkaannya dalam korpus, sehingga kata-kata yang sering muncul dalam banyak dokumen diberi bobot lebih rendah, sedangkan kata-kata yang khas untuk ulasan tertentu mendapat bobot lebih tinggi. Pendekatan *TF-IDF* terbukti efektif dalam banyak penelitian klasifikasi teks karena mampu meningkatkan ketajaman pemisahan fitur berdasarkan konteks.

Setelah ekstraksi fitur, data dibagi menjadi dua subset: data latih (*training set*) dan data uji (*test set*). Kemudian model *Naïve Bayes Classifier* dibangun menggunakan data latih. Algoritma ini dipilih karena kecepatan komputasi, kesederhanaan implementasi, serta kinerja yang solid pada tugas klasifikasi teks terutama sentimen *dataset* memiliki banyak fitur akibat representasi *TF-IDF*. Kombinasi *TF-IDF* dan *Naïve Bayes* telah banyak digunakan dalam literatur untuk analisis sentimen dengan hasil akurasi dan efisiensi komputasi yang baik.

g. Evaluasi Model

Tahap selanjutnya adalah evaluasi model. Pada tahap ini, kinerja model dinilai menggunakan metrik evaluasi seperti akurasi, presisi, *recall*, dan *F1-score*. Evaluasi dilakukan untuk mengetahui sejauh mana model mampu mengklasifikasikan sentimen ulasan secara tepat. Selain itu, *confusion matrix* juga digunakan untuk melihat persebaran prediksi benar dan salah pada setiap kelas sentimen, sehingga dapat diketahui bagian mana yang memerlukan perbaikan.

h. Deployment

Pada tahap ini, hasil analisis sentimen diimplementasikan ke dalam bentuk aplikasi web menggunakan *framework Streamlit*. *Deployment* dilakukan agar hasil penelitian dapat digunakan secara langsung oleh pemilik *coffee shop* ataupun pihak lain yang membutuhkan. Aplikasi web ini

menampilkan visualisasi seperti grafik persebaran sentimen, *comparison chart* antar *coffee shop*, *word cloud*, serta hasil klasifikasi yang mudah dipahami pengguna non-teknis. Dengan demikian, penelitian ini tidak hanya menghasilkan model prediktif, namun juga menyediakan *platform* yang dapat digunakan secara praktis untuk pengambilan sentimen bisnis.

HASIL DAN PEMBAHASAN

Algoritma *Naïve Bayes* diterapkan dalam penelitian ini untuk melakukan klasifikasi sentimen pada ulasan pelanggan dari tiga kedai kopi dengan jumlah ulasan terbanyak di Ponorogo, yaitu Kedai *Gayeng Billiard & Café*, Kedai Sor Sawo, dan *Origin Coffee Lab*. Algoritma ini dipilih karena dikenal efisien dalam memproses data teks berukuran besar serta memiliki kinerja yang stabil pada berbagai penelitian analisis sentimen. Setelah melalui tahapan *scrapping*, *preprocessing*, dan pembersihan data, model dibangun menggunakan representasi fitur *Term Frequency-Inverse Document Frequency (TF-IDF)* untuk mengekstraksi bobot kata secara optimal sebelum akhirnya diklasifikasikan oleh *Naïve Bayes*. Bagian ini memaparkan hasil pemodelan, distribusi sentimen pada masing-masing kedai, performa akurasi model, serta interpretasi temuan yang diperoleh dari data ulasan pelanggan.

1. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini diperoleh dari *platform Google Maps*, yang menyediakan sentimen ulasan pelanggan terkait layanan dan pengalaman di berbagai kedai kopi. Penelitian ini secara khusus mengambil data ulasan dari tiga kedai kopi dengan jumlah review terbanyak di Ponorogo, yaitu Kedai *Gayeng Billiard & Café*, Kedai Sor Sawo, dan *Origin Coffee Lab*. Ketiga kedai ini dipilih karena memiliki aktivitas ulasan yang tinggi, sehingga menghasilkan *dataset* yang cukup representatif untuk dianalisis menggunakan metode analisis sentimen.

Tabel 1. Data jumlah ulasan

No.	Tempat Objek Penelitian	Jumlah Ulasan
1	Gayeng Billiard and Cafe	1.863 ulasan
2	Kedai Sor Sawo	1.557 ulasan
3	Origin Coffee Lab	1.278 ulasan

Data diperoleh secara langsung melalui proses *web scrapping* menggunakan *Python*, mengingat *Google Maps* tidak menyediakan *dataset* ulasan secara terbuka dalam bentuk terstruktur. Proses *scrapping* dilakukan dengan mengekstraksi elemen-elemen penting seperti teks ulasan, rating bintang, tanggal ulasan, dan nama pengguna yang muncul di halaman review masing-masing kedai[10]. *Dataset* awal yang dihasilkan berjumlah ribuan baris, namun masih mengandung elemen *HTML*, data tidak relevan,

dan duplikasi. Setelah melalui tahap pemfilteran awal, jumlah ulasan berkurang signifikan dan tersisa hanya komentar pelanggan yang valid untuk dianalisis lebih lanjut.

Data diperoleh dalam *format* CSV yang kemudian diproses pada tahap pra-pemrosesan, sehingga memudahkan implementasi algoritma *Naïve Bayes* pada proses klasifikasi sentimen. Proses perolehan data ini dilakukan secara sistematis untuk memastikan seluruh ulasan publik yang relevan berhasil diambil dan siap digunakan pada tahapan *modelling*.

2. Preprocessing Data

Untuk memastikan bahwa *dataset* yang digunakan benar-benar bersih dan layak untuk dianalisis, tahap *preprocessing* merupakan langkah penting dalam penelitian ini[11]. Setelah ribuan baris data *Google Maps* diekstraksi melalui proses *scrapping*, kumpulan data tetap mengandung sejumlah elemen yang tidak relevan, seperti struktur *HTML*, simbol-simbol sistem, teks navigasi, metadada visual, dan beberapa ulasan duplikat yang diekstraksi berulang kali. Oleh karena itu, *preprocessing* dilakukan menggunakan *Python* dengan tujuan mengekstraksi hanya bagian yang diperlukan, yaitu komentar pelanggan asli, yang merupakan dasar analisis sentimen.

Proses *preprocessing* dimulai dengan memilih kolom yang diperlukan, seperti teks komentar, nilai bintang, dan tanggal ulasan, dan menghapus kolom lain yang muncul karena struktur *HTML* atau mekanisme *rendering* halaman *Google Maps*. Setelah itu, penyaringan baris dilakukan untuk menghilangkan entri tanpa komentar, komentar kosong, hasil parsing yang keliru, dan ulasan yang hanya mengandung karakter non-teks seperti emoji atau simbol teknis. Jumlah *dataset* yang secara signifikan berkurang dari ribuan baris menjadi ratusan ulasan murni yang ditulis oleh pelanggan, yang membuatnya lebih representatif untuk keperluan analisis.

3. Cleaning text dan Validasi Data

Tahap *cleaning text* merupakan proses lanjutan setelah *preprocessing* awal untuk memastikan bahwa setiap ulasan berada dalam kondisi yang konsisten, bersih, dan siap digunakan dalam analisis sentimen[12]. Proses ini dilakukan dengan membersihkan teks dari elemen-elemen yang dianggap sebagai *noise*, seperti tanda baca berlebihan, karakter non-alfabet, emotikon, serta simbol-simbol visual yang tidak memiliki nilai semantik dalam konteks analisis. Selain itu, dilakukan juga normalisasi bahasa dengan mengonversi kata-kata tidak baku menjadi bentuk baku, menghapus kata berulang, serta menstandarisasi pola penulisan agar setiap teks memiliki struktur yang seragam. Langkah ini penting karena kualitas data teks yang tidak konsisten

dapat mempengaruhi proses ekstraksi fitur maupun kinerja model klasifikasi. Pendekatan pembersihan teks seperti ini banyak digunakan dalam penelitian analisis sentimen karena terbukti meningkatkan kualitas representasi data dan akurasi model

Tabel 2. Cleaning Text

Data mentah	Hasil Cleaning
Pelayanan lamaaaaaa (75 menit baru keluar maknanya)Klo pas lagi laper gk recommended sih dijamin keburu masuk angin, asam lambung naik...Harga ya murah banget sepadan sama menunya...Kopinya oke tp langsung drop liat wadah gulanya persis tempat saus kecap plastik di abang² bakso...Yang ane lagi menu steaknya spechlesh...Bayarnya harus CASH gak bisa QRIS ataupun DEBIT...Trus mbak²nya manis didepan tp ngedumel dibelakang. Mungkin klo emang konsepnya bar OKE LAH tp klo konsep resto enggak banget	Pelayanannya sangat lama (sekitar 75 menit hingga makanan keluar). Jika sedang lapar, tempat ini kurang direkomendasikan karena bisa keburu masuk angin atau asam lambung naik. Harganya memang sangat murah dan sesuai dengan menunya. Kopinya cukup baik, namun sedikit mengecewakan ketika melihat wadah gulanya yang mirip tempat saus atau kecap plastik di pedagang bakso. Untuk menu steak, rasanya kurang memuaskan. Pembayaran hanya dapat dilakukan secara tunai, tidak bisa menggunakan QRIS maupun debit. Selain itu, pelayanan dari pegawainya kurang konsisten; ramah di depan tetapi terdengar mengeluh di belakang. Jika konsepnya bar mungkin masih sesuai, tetapi untuk konsep restoran hal ini kurang tepat.
Mantap minuman murah bakaran juga murah",cocok untuk keluarga anak muda dan juga untuk rapat atau kumpul temen",rekomendasi banget sih ini cafe gayeng,ada live music nya juga bisa sambil karaoke 🎤	Mantap; minumannya murah, pilihan bakanannya juga terjangkau. Tempat ini cocok untuk keluarga, anak muda, maupun untuk rapat atau berkumpul bersama teman. Sangat direkomendasikan, suasananya ramai dan menyenangkan, serta tersedia live music yang dapat dinikmati sambil karaoke.

Setelah data dibersihkan, dilakukan proses validasi data berupa pemberian label sentimen pada setiap ulasan melalui teknik *manual labeling* atau *semi-automated labeling*. Pada tahap ini, setiap komentar pelanggan dianalisis untuk menentukan apakah ulasan tersebut bernada positif, negatif, atau netral. Validasi dilakukan dengan mempertimbangkan konteks kalimat, makna subjektif, serta ekspresi kepuasan atau ketidakpuasan yang muncul dalam teks. Tahapan pelabelan ini sangat penting karena kualitas label akan berdampak langsung pada proses pelatihan model *Naive Bayes*. Jika label tidak akurat, maka model akan belajar pola yang salah sehingga menurunkan performa klasifikasi. Beberapa penelitian sebelumnya menekankan bahwa proses validasi dan pelabelan sentimen merupakan unsur krusial dalam pembangunan model analisis sentimen berbasis pembelajaran mesin.

Tabel 3. Sentimen Ulasan

Ulasan Komentar Pelanggan	Hasil Sentimen
Tempat parkir luas, tempat nya enak dan nyaman, banyak pilihan menu	Positif
Makanannya kurang mantap , dagingnya terlalu keras , lalapannya jg kurang	Negatif
Suasana santai persawahan	Netral

4. Modelling TF-IDF dan Naïve Bayes

Tahap pemodelan dimulai dengan representasi teks menggunakan metode *Term Frequency-Inverse Document Frequency (TF-IDF)*. Metode ini bertujuan untuk mengubah data ulasan menjadi fitur numerik sehingga dapat diproses oleh model klasifikasi[13]. *TF-IDF* tidak hanya menghitung frekuensi kemunculan kata dalam suatu dokumen, namun juga memperhitungkan signifikansi relatif kata tersebut terhadap seluruh korpus. Dengan demikian, kata yang

sering muncul tetapi tidak memberikan informasi sentimen seperti “kopi”, “kedai”, atau nama lokasi cenderung memiliki nilai bobot lebih rendah dibanding kata yang bersifat evaluatif seperti “nyaman”, “bagus”, atau “mahal”. Pada penelitian ini, konfigurasi *TF-IDF* menggunakan rentang *n-gram* (1,2), sehingga model mampu menangkap konteks kata tunggal serta frasa dua kata. Penggunaan bigram sangat penting dalam domain ulasan, karena sentimen sering muncul dalam bentuk frasa seperti “layanan lama”, “makan enak”, atau “tempat nyaman”, yang memiliki makna berbeda dibanding kata tunggalnya.

Selain konfigurasi *n-gram*, penyesuaian lain dilakukan pada parameter *min_df=2* dan *max_df=0.9*. Nilai *min_df* memastikan bahwa kata yang muncul hanya satu kali di seluruh dataset tidak dipertimbangkan sebagai fitur, karena cenderung bersifat *noise* atau tidak memiliki makna *generalizable*. Sementara itu, *max_df* membatasi kata yang terlalu sering muncul di hampir seluruh dokumen, sehingga tidak mendominasi *matriks* fitur dan tidak mengurangi sensitivitas model terhadap kata-kata bernuansa emosional. Penelitian ini juga menerapkan *domain-specific stopwords*, yaitu daftar kata yang relevan dengan konteks kedai kopi tetapi tidak bersifat sentimen, seperti “cafe”, “warung”, “ponorogo”, “tempat”, atau kata fungsional seperti “nya”, “yg”, dan “sangat”. Strategi tersebut membuat model fokus pada kata yang mengandung opini atau pengalaman pengguna, bukan sekadar deskripsi lokasi.

```
# TF-IDF ulang
from sklearn.feature_extraction.text import TfidfVectorizer
vectorizer = TfidfVectorizer(ngram_range=(1,2), min_df=2, max_df=0.9, stop_words=[
    "kopi", "cafe", "kedai", "warung", "ponorogo", "tempat", "nya", "yg",
    "sangat", "baik", "buat", "utk", "rasa", "minum", "minuman"])
X = vectorizer.fit_transform(df_balanced['review_cleaned'].dropna())
y = df_balanced['Hasil Validasi'].dropna()
```

Gambar 2. Kode Implementasi *TF-IDF*
Sumber: Hasil Olah Data

Model klasifikasi sentimen dibangun menggunakan algoritma *Multinomial Naïve Bayes*, yang merupakan salah satu metode paling umum dalam pemodelan teks. Model ini mengasumsikan independensi antar fitur, sehingga probabilitas suatu dokumen termasuk ke dalam kelas tertentu ditentukan oleh kombinasi probabilitas setiap kata dalam dokumen tersebut. *Multinomial Naïve Bayes* bekerja dengan baik pada data berbasis hitungan atau bobot seperti hasil *TF-IDF*, karena model menghitung $P(\text{kata}|\text{kelas})$ dan mengakumulasi *log-probabilitas* untuk menentukan kelas yang memiliki skor tertinggi[14]. Karakteristik ini memungkinkan model mempelajari hubungan khusus antara kata dan sentimen, misalnya bahwa kemunculan kata seperti “mahal”, “lama”, atau “buruk” secara signifikan meningkatkan peluang ulasan dikategorikan sebagai sentiment negatif.

Pemodelan *Naïve Bayes* dalam penelitian ini memperlihatkan pola kosakata yang konsisten dengan karakteristik sentimen pada data ulasan. Kata-kata dengan bobot probabilitas tertinggi pada kelas sentimen negatif antara lain “mahal”, “layanan”, “lama”, “kurang”, dan “buruk”. Kata ini umumnya muncul dalam konteks keluhan konsumen mengenai harga, kualitas pelayanan, atau pengalaman yang mengecewakan. Sebaliknya, pada kelas positif, model memprioritaskan kata seperti “nyaman”, “bagus”, “enak”, “mantap”, dan “cocok”, yang lazim digunakan konsumen untuk menyatakan apresiasi atau kepuasan terhadap produk dan suasana kedai kopi. Adapun kelas netral menampilkan kata-kata yang bersifat deskriptif, seperti “ramai”, “angkring”, “luas”, “sepi”, maupun “sawah”, yang menggambarkan kondisi lingkungan tanpa menunjukkan kecenderungan emosional tertentu.

```
# Training model
from sklearn.naive_bayes import MultinomialNB
model = MultinomialNB()
model.fit(X_train, y_train)
```

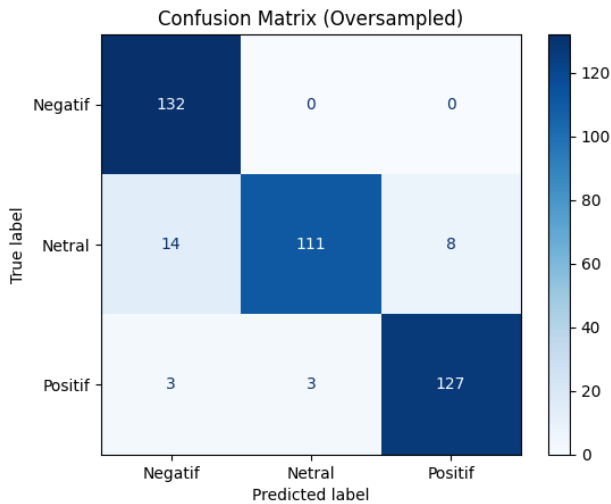
Gambar 3. Kode Implementasi *Naïve Bayes*
Sumber: Hasil Olah Data

Secara keseluruhan, kombinasi *TF-IDF* dan *Multinomial Naïve Bayes* menghasilkan model yang tidak hanya mampu mengklasifikasikan teks berdasarkan sentimen, tetapi juga memberikan interpretabilitas terhadap pola kosakata yang membentuk keputusan model. *TF-IDF* berperan menyaring kata yang representatif terhadap makna sentimen, sementara *Naïve Bayes* mengidentifikasi probabilitas kata tersebut dalam setiap kelas. Dengan demikian, perbedaan semantik antara ulasan bernada positif, negatif, dan netral dapat dipetakan dengan jelas melalui bobot fitur yang dihasilkan. Pendekatan ini memberikan pemahaman yang lebih mendalam mengenai persepsi pelanggan terhadap kedai kopi yang menjadi objek penelitian, sekaligus membuktikan efektivitas model probabilistik dalam domain analisis sentimen berbasis teks konsumen.

5. Evaluasi Model

Evaluasi awal dilakukan menggunakan data uji yang telah diseimbangkan melalui proses *oversampling* untuk menanggulangi ketidakseimbangan distribusi kelas. Dari 398 sampel uji, model menghasilkan akurasi sebesar 93%, yang menunjukkan kinerja prediksi yang kuat pada ketiga kelas sentimen. *Confusion matrix oversampled* memperlihatkan bahwa model mampu mengenali ulasan negatif dan positif dengan sangat baik. Sebanyak 132 ulasan negatif diklasifikasikan tepat sebagai negatif tanpa kesalahan prediksi, dan 127 ulasan positif berhasil dipetakan dengan benar. Kesalahan klasifikasi utama terjadi pada kelas netral, di mana 22 data salah dikategorikan ke kelas negatif (14 ulasan) maupun positif (8 ulasan). Hal ini

menunjukkan bahwa model lebih mudah mengidentifikasi sentimen eksplisit dibanding sentimen deskriptif yang tidak menampilkan evaluasi langsung, sehingga ulasan yang bersifat informatif cenderung terdorong menjadi positif atau negatif.



Gambar 4. Confusion Matrix (Oversampled)
Sumber: Hasil Olah Data

Pengujian lanjutan dilakukan terhadap keseluruhan *dataset* tanpa modifikasi struktur distribusi kelas guna menggambarkan performa model dalam kondisi nyata. Pada skenario ini, model memperoleh akurasi keseluruhan sebesar 94,34%, menunjukkan stabilitas performa meskipun dominasi ulasan positif jauh lebih tinggi dibandingkan kelas lain. *Confusion matrix* menunjukkan 105 ulasan negatif berhasil diklasifikasikan seluruhnya sebagai negatif, menandakan sensitivitas model yang kuat terhadap kata-kata bernada keluhan. Sementara itu, 626 dari 662 ulasan positif berhasil dikenali dengan benar, sedangkan sebagian kecil (24 data) salah dipetakan ke kelas negatif, dan 12 data ke kelas netral. Kelas netral menjadi kelas dengan tingkat ambiguitas tertinggi, di mana 86 ulasan teridentifikasi tepat, namun 5 data diprediksi sebagai negatif dan 8 sebagai positif. Pola ini mengindikasikan bahwa kecenderungan emosi yang tidak tegas dalam ulasan netral menjadi faktor utama yang memicu kesalahan model.

```

*** Akurasi model pada seluruh dataset: 94.34%

Confusion Matrix:
[[105  0  0]
 [  5 86  8]
 [ 24 12 626]]

Classification Report:
              precision    recall  f1-score   support

   Negatif      0.78      1.00      0.88       105
    Netral      0.88      0.87      0.87        99
    Positif      0.99      0.95      0.97       662

 accuracy      0.94      0.94      0.94      866
 macro avg      0.88      0.94      0.91      866
 weighted avg      0.95      0.94      0.94      866

```

Gambar 5. Akurasi Model
Sumber: Hasil Olah Data

Secara umum, kombinasi evaluasi pada *dataset* seimbang dan *dataset* asli menunjukkan konsistensi kinerja algoritma. Model sangat akurat dalam memetakan ulasan dengan ekspresi sentimen yang jelas, baik positif maupun negatif, namun masih memiliki tantangan dalam mengklasifikasikan ulasan yang berada pada spektrum abu-abu sentimen. Meskipun demikian, stabilitas akurasi di atas 90% pada seluruh skenario pengujian, serta kemampuan model mengenali sentimen secara konsisten, memperlihatkan bahwa pendekatan *Multinomial Naïve Bayes* berbasis *TF-IDF* efektif untuk menganalisis kecenderungan opini pelanggan pada kedai kopi yang menjadi objek penelitian.

6. Deployment

Tahap *deployment* pada penelitian ini dilakukan dengan membangun sebuah aplikasi web sederhana menggunakan *Streamlit* yang menampilkan hasil analisis sentimen dari tiga kedai kopi berbeda di Ponorogo[15]. Aplikasi ini memuat antarmuka yang mudah digunakan untuk menampilkan prediksi sentimen berdasarkan model *Naïve Bayes*, grafik perbandingan sentimen antar kedai, serta ringkasan kinerja model. Setiap kedai memiliki tampilan hasil yang terpisah sehingga pengguna dapat melihat perbedaan kecenderungan ulasan pelanggan secara langsung dan lebih terstruktur.

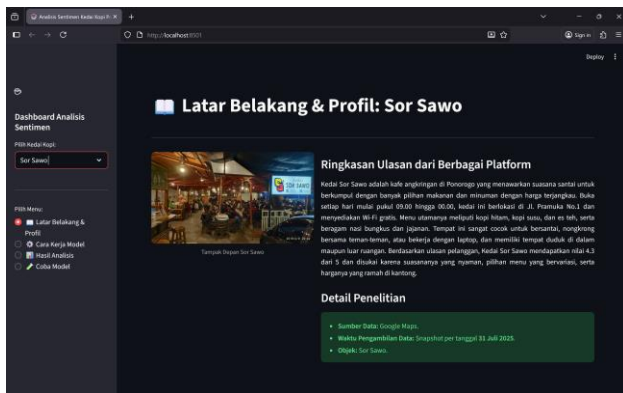
Selain itu, aplikasi ini juga menampilkan teks ulasan asli, hasil *preprocessing*, serta klasifikasi sentimen agar pengguna dapat menelusuri data dengan lebih transparan. Dengan adanya *deployment* ini, hasil penelitian tidak hanya berhenti pada model yang akurat, tetapi juga dapat digunakan secara praktis oleh pemilik kedai atau peneliti lain untuk memahami persepsi pelanggan. Aplikasi ini dapat diakses secara lokal maupun diunggah ke *platform* seperti *Streamlit Cloud* sehingga memudahkan proses demonstrasi pada pihak terkait.



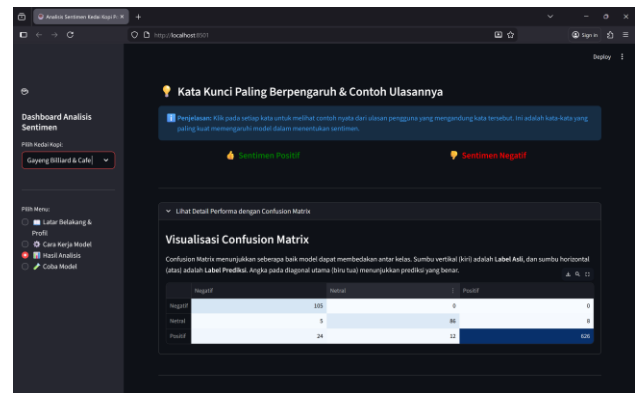
Gambar 6. Tampilan Web halaman Gayeng
Sumber: Hasil Pengembangan Sistem



Gambar 9. Tampilan Web halaman Hasil Analisis
Sumber: Hasil Pengembangan Sistem



Gambar 7. Tampilan Web halaman Sor Sawo
Sumber: Hasil Pengembangan Sistem



Gambar 10. Tampilan Web halaman Analisis Confusion
Matrix
Sumber: Hasil Pengembangan Sistem



Gambar 8. Tampilan Web halaman Sor Sawo
Sumber: Hasil Pengembangan Sistem

KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan hasil penelitian ini, dapat disimpulkan bahwa model yang dibangun mampu mengklasifikasikan ulasan pelanggan ke dalam kategori sentimen positif dan negatif secara cukup efektif. Tahap *preprocessing*, yang meliputi pembersihan teks, normalisasi, penghapusan karakter tidak relevan, serta penggunaan *TF-IDF*, terbukti meningkatkan kualitas data sehingga model dapat belajar dengan lebih optimal. Hasil evaluasi menunjukkan bahwa sebagian besar ulasan terhadap tiga kedai kopi yang dianalisis cenderung bernada positif, sementara ulasan negatif umumnya terkait pelayanan, waktu penyajian, dan kualitas menu tertentu.

Penelitian ini juga menghasilkan aplikasi *deployment* berbasis *Streamlit* yang mampu menampilkan hasil klasifikasi secara interaktif, sehingga dapat dimanfaatkan oleh pemilik kedai maupun peneliti lain untuk memahami persepsi pelanggan. Meskipun demikian, model *Naive Bayes* masih memiliki kelemahan dalam menangani kalimat yang bersifat sarkasme, ambigu, atau memiliki konteks yang kompleks. Dengan demikian, hasil penelitian ini

menjawab tujuan utama, yakni mengetahui kecenderungan sentimen pelanggan dan menyediakan alat bantu analisis yang praktis.

Saran

Penelitian selanjutnya disarankan untuk menggunakan dataset yang lebih besar dan beragam agar hasil analisis menjadi lebih representatif. Metode klasifikasi lain seperti SVM, *Random Forest*, atau model berbasis *transformer* (misalnya *IndoBERT*) dapat dipertimbangkan untuk meningkatkan akurasi, khususnya dalam menangani kalimat kompleks dan sarkasme. Selain itu, analisis sentimen dapat diperluas dengan menggabungkan data dari platform lain seperti *GoFood*, *GrabFood*, dan Instagram untuk memperoleh perspektif yang lebih luas.

Pengembangan aplikasi juga dapat ditingkatkan dengan menambahkan fitur visualisasi lebih rinci, seperti tren sentimen berdasarkan waktu, analisis topik, serta rekomendasi otomatis bagi pemilik kedai terkait aspek yang perlu diperbaiki. Penelitian lanjutan diharapkan mampu mengatasi beberapa keterbatasan penelitian ini dan memberikan kontribusi yang lebih mendalam terhadap pengembangan strategi bisnis berbasis data pada industri coffee shop.

UCAPAN TERIMA KASIH

Peneliti mengucapkan terima kasih yang sebesar-besarnya kepada para pemilik dan pengelola kedai kopi yang telah memberikan izin, dukungan, serta akses data selama proses penelitian berlangsung. Penghargaan juga disampaikan kepada seluruh pihak yang turut membantu, termasuk dosen pembimbing, rekan mahasiswa, serta semua individu yang berkontribusi dalam penyusunan, analisis, dan pengembangan aplikasi penelitian ini. Tanpa bantuan dan kerja sama berbagai pihak tersebut, penelitian ini tidak dapat terselesaikan dengan baik.

DAFTAR PUSTAKA

- [1] W. Parasati, F. Abdurrachman Bachtiar, dan N. Y. Setiawan, "Analisis Sentimen Berbasis Aspek pada Ulasan Pelanggan Restoran Bakso President Malang dengan Metode Naïve Bayes Classifier," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 4, hal. 1–10, 2020, [Daring]. Tersedia pada: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/7134>
- [2] Solikaturun, "Art ikel," vol. 4, no. April, hal. 60–74, 2015.
- [3] D. Bounie, "Pengaruh Ulasan Pelanggan Online terhadap Keputusan Pembelian: Kasus Pengaruh Ulasan Pelanggan Online terhadap Keputusan Pembelian: Kasus Video Game," no. 1, 2025.
- [4] J. Ipmawati, S. Saifulloh, dan K. Kusnawi, "Analisis Sentimen Tempat Wisata Berdasarkan Ulasan pada Google Maps Menggunakan Algoritma Support Vector Machine," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 1, hal. 247–256, 2024, doi: 10.57152/malcom.v4i1.1066.
- [5] V. R. Prasetyo et al., "Analisis Sentimen Dan Kategorisasi Review Pelanggan Pada Cafe Kopi Paste Dengan Metode Naive Bayes Dan K-Nearest Neighbor Sentiment Analysis and Categorization of Customer Reviews on Kopi Paste Cafe Using Naive Bayes and K-Nearest Neighbor Methods," *J. Ilm. NERO*, vol. 8, no. 1, hal. 2023, 2023.
- [6] T. A. Sari, E. Sinduningrum, dan F. Noor Hasan, "KLIK: Kajian Ilmiah Informatika dan Komputer Analisis Sentimen Ulasan Pelanggan Pada Aplikasi Fore Coffee Menggunakan Metode Naïve Bayes," *Media Online*, vol. 3, no. 6, hal. 773–779, 2023, doi: 10.30865/klik.v3i6.884.
- [7] S. M. Salsabila, A. Alim Murtopo, dan N. Fadhilah, "Analisis Sentimen Pelanggan Tokopedia Menggunakan Metode Naïve Bayes Classifier," *J. Minfo Polgan*, vol. 11, no. 2, hal. 30–35, 2022, doi: 10.33395/jmp.v11i2.11640.
- [8] M. S. Yupi Kuspani Putra, "Perbandingan Algoritma Naive Bayes dan Naive Bayes," *J. Inform. dan Teknol.*, vol. 2, no. 3, hal. 61–69, 2019.
- [9] J. W. Iskandar dan Y. Nataliani, "Perbandingan Naïve Bayes, SVM, dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 6, hal. 1120–1126, 2021, doi: 10.29207/resti.v5i6.3588.
- [10] E. Yuniar, D. S. Utsalinah, dan D. Wahyuningsih, "Implementasi Scrapping Data Untuk Sentiment Analysis Pengguna Dompot Digital dengan Menggunakan Algoritma Machine Learning," *J. Janitra Inform. dan Sist. Inf.*, vol. 2, no. 1, hal. 35–42, 2022, doi: 10.25008/janitra.v2i1.145.
- [11] S. Khairunnisa, A. Adiwijaya, dan S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *J. Media Inform. Budidarma*, vol. 5, no. 2, hal. 406, 2021, doi: 10.30865/mib.v5i2.2835.
- [12] Y. LESTANTO dan R. MUALIFA, "Development of a Data Cleaning System for Consumer Master Data using Sorted Neighborhood and N-Gram Methods," *ELKOMIKA J. Tek. Energi Elektr. Tek. Telekomun. Tek. Elektron.*, vol. 13, no. 1, hal. 57, 2025, doi: 10.26760/elkomika.v13i1.57.
- [13] V. B. Lestari dan C. A. Hutagalung, "Evaluation of TF-IDF Extraction Techniques in Sentiment Analysis of Indonesian-Language Marketplaces Using SVM, Logistic Regression, and Naive Bayes," no. 021, hal. 36–44.
- [14] R. Rahmadani, A. Rahim, dan R. Rudiman, "Analisis Sentimen Ulasan 'Ojol the Game' Di

Google Play Store Menggunakan Algoritma Naive Bayes Dan Model Ekstraksi Fitur Tf-Idf Untuk Meningkatkan Kualitas Game," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, 2024, doi: 10.23960/jitet.v12i3.4988.

- [15] G. Hripcsak, J. J. Cimino, dan S. Sengupta, "WebCIS: large scale deployment of a Web-based clinical information system.," *Proc. AMIA Symp.*, hal. 804–808, 1999.