
Sentiment Analysis of Pasar Dayak Samarinda Visitors Using the Naive Bayes Algorithm

Julia Emakulata Buring¹⁾, Nursobah²⁾, dan Rizky Zakariyya Rasyad³⁾

^{1,2}Program Studi Teknik Informatika, STMIK Widya Cipta Dharma, ³Program Studi Bisnis Digital, STMIK Widya Cipta Dharma

^{1,2,3}Jl. M. Yamin No. 25, Gunung Kelua, Samarinda Ulu, Kota Samarinda, Kalimantan Timur 75123

E-mail: 2243098@wicida.ac.id¹⁾, nursobah@wicida.ac.id²⁾, rizky@wicida.ac.id³⁾

ABSTRACT

Visitor reviews on Google Maps contain valuable information about public perceptions of Pasar Dayak Samarinda, but the growing number of reviews makes manual analysis inefficient. This study aims to classify visitor reviews into positive, negative, and neutral sentiments using the Multinomial Naive Bayes algorithm and to implement the resulting model in a Flask-based website. The research follows the Cross Industry Standard Process for Data Mining (CRISP-DM), consisting of business understanding, data understanding, data preparation, modeling, evaluation, and deployment. The collected reviews were processed through case folding, cleaning, repetition removal, slang normalization, tokenization, stopword removal, and stemming. The final dataset used for modeling consisted of 199 reviews, divided into 159 training data and 40 testing data. RandomOverSampler was applied to balance the training classes, while TF-IDF with unigram and bigram features was used to convert text into numerical representations. Model evaluation showed an accuracy of 75.00%, precision of 71.27%, recall of 75.00%, and F1-score of 70.41%. The positive class achieved the best performance, with an F1-score of 0.89, while the neutral class remained the most difficult to identify. The model was deployed through a Flask-based website that supports manual and batch sentiment prediction and displays evaluation results. These findings indicate that Multinomial Naive Bayes can provide a practical overview of visitor sentiment, although additional balanced and varied data are required to improve minority-class recognition.

Keywords: Sentiment Analysis, Naive Bayes, TF-IDF, Google Maps, Pasar Dayak Samarinda

Analisis Sentimen Pengunjung Pasar Dayak Samarinda Menggunakan Algoritma Naive Bayes

ABSTRAK

Ulasan pengunjung pada Google Maps memuat informasi penting mengenai persepsi masyarakat terhadap Pasar Dayak Samarinda, tetapi jumlah ulasan yang terus bertambah menyebabkan analisis manual menjadi kurang efektif. Penelitian ini bertujuan mengklasifikasikan ulasan pengunjung ke dalam sentimen positif, negatif, dan netral menggunakan algoritma Multinomial Naive Bayes serta mengimplementasikan model ke dalam website berbasis Flask. Penelitian menerapkan metode Cross Industry Standard Process for Data Mining (CRISP-DM), yang mencakup business understanding, data understanding, data preparation, modeling, evaluation, dan deployment. Data ulasan diproses melalui case folding, cleaning, repetition removal, normalisasi slang, tokenizing, stopword removal, dan stemming. Dataset akhir yang digunakan pada pemodelan berjumlah 199 ulasan, terdiri atas 159 data latih dan 40 data uji. RandomOverSampler digunakan untuk menyeimbangkan kelas pada data latih, sedangkan TF-IDF dengan fitur unigram dan bigram digunakan untuk mengubah teks menjadi representasi numerik. Hasil evaluasi menunjukkan Accuracy sebesar 75,00%, Precision sebesar 71,27%, Recall sebesar 75,00%, dan F1-Score sebesar 70,41%. Kelas positif memperoleh performa terbaik dengan F1-Score 0,89, sedangkan kelas netral masih paling sulit dikenali. Model kemudian diterapkan pada website berbasis Flask yang mendukung prediksi sentimen secara manual dan batch serta menampilkan hasil evaluasi. Hasil ini menunjukkan bahwa Multinomial Naive Bayes dapat memberikan gambaran praktis mengenai sentimen pengunjung, meskipun penambahan data yang lebih seimbang dan beragam masih diperlukan untuk meningkatkan pengenalan kelas minoritas.

Kata Kunci: Analisis Sentimen, Naive Bayes, TF-IDF, Google Maps, Pasar Dayak Samarinda

1. PENDAHULUAN

Pasar tradisional memiliki peran penting sebagai pusat kegiatan ekonomi dan interaksi sosial masyarakat. Salah satu pasar tradisional di Kota Samarinda adalah Pasar

Dayak Samarinda yang menyediakan berbagai kebutuhan pokok sekaligus mencerminkan aktivitas budaya masyarakat lokal. Perkembangan teknologi informasi mengubah cara pengunjung menyampaikan pengalaman

terhadap suatu tempat. Melalui Google Maps, pengunjung dapat memberikan penilaian dan ulasan mengenai kebersihan, pelayanan, fasilitas, kenyamanan, harga, serta kualitas produk yang dijual.

Jumlah ulasan yang terus bertambah menyebabkan pengelola sulit memahami keseluruhan opini pengunjung apabila dianalisis secara manual. Analisis sentimen dapat digunakan untuk mengolah data teks dan mengelompokkan opini menjadi sentimen positif, negatif, atau netral. Naive Bayes dipilih karena memiliki proses komputasi yang sederhana, cepat, serta banyak digunakan untuk klasifikasi teks (Budianto & Anwar, 2022; Azizah dkk., 2024).

Penelitian sebelumnya menunjukkan bahwa Naive Bayes mampu diterapkan pada ulasan objek wisata, e-commerce, tempat makan, dan opini publik. Fatikhah dkk. (2024) melaporkan peningkatan akurasi setelah penyeimbangan data, sedangkan Apriliansyah dkk. (2025) memperoleh akurasi 78% pada ulasan pengunjung Pantai Kejawan. Namun, kajian mengenai pasar tradisional, khususnya Pasar Dayak Samarinda, masih terbatas. Kebaruan penelitian ini terletak pada objek kajian berupa pasar tradisional berciri budaya Dayak, penggunaan tiga kelas sentimen, penyeimbangan data dengan RandomOverSampler, serta implementasi model pada website berbasis Flask.

Tujuan penelitian ini adalah menganalisis sentimen pengunjung Pasar Dayak Samarinda berdasarkan ulasan Google Maps menggunakan Multinomial Naive Bayes serta menyediakan media yang memudahkan pengguna melakukan prediksi sentimen dan melihat hasil analisis.

2. RUANG LINGKUP

Penelitian ini berfokus pada analisis sentimen terhadap ulasan pengunjung Pasar Dayak Samarinda yang diperoleh dari platform Google Maps. Ulasan yang digunakan berupa data teks yang diklasifikasikan ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral. Proses klasifikasi dilakukan menggunakan metode ekstraksi fitur Term Frequency–Inverse Document Frequency (TF-IDF), penyeimbangan data dengan RandomOverSampler, serta algoritma Multinomial Naive Bayes. Hasil penelitian berupa model klasifikasi sentimen yang diimplementasikan ke dalam website berbasis Flask, sehingga pengguna dapat melakukan prediksi sentimen secara manual melalui input teks maupun secara batch menggunakan file yang berisi kumpulan ulasan. Ruang lingkup penelitian ini dibatasi pada analisis sentimen berdasarkan data ulasan Google Maps dan tidak membahas aspek kebijakan maupun pengelolaan Pasar Dayak Samarinda secara langsung.

3. BAHAN DAN METODE

3.1 Analisis Sentimen dan Naive Bayes

Analisis sentimen merupakan teknik text mining untuk mengidentifikasi kecenderungan opini yang terkandung dalam teks menjadi kategori tertentu (Sugiarta dkk., 2023). Pada penelitian ini, kategori yang digunakan

adalah positif, negatif, dan netral. *Multinomial Naive Bayes* menghitung probabilitas suatu dokumen termasuk dalam sebuah kelas berdasarkan kemunculan fitur kata. Metode ini sesuai untuk data teks karena bekerja baik pada fitur berbentuk frekuensi atau bobot kata.

Persamaan untuk algoritma *Naive Bayes* secara umum dapat dilihat pada persamaan (1) berikut

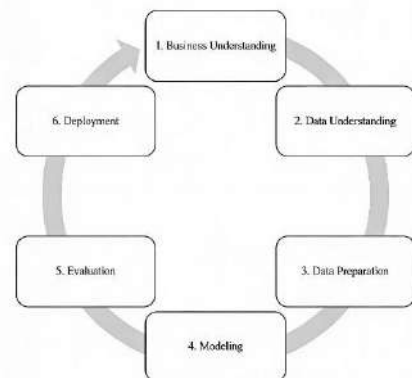
$$P(H | X) = \frac{P(X | H) \times P(H)}{P(X)} \quad (1)$$

3.2 TF-IDF dan Penyeimbangan Data

TF-IDF digunakan untuk mengubah teks menjadi representasi numerik dengan memberi bobot lebih tinggi pada kata yang penting dalam suatu dokumen dan relatif jarang muncul pada seluruh dokumen (Alhaqq dkk., 2022). Ketidakseimbangan kelas ditangani menggunakan RandomOverSampler, yaitu menambahkan sampel kelas minoritas secara acak sampai jumlah setiap kelas pada data latih menjadi seimbang.

3.3 Tahapan CRISP-DM

Metode penelitian mengikuti CRISP-DM yang terdiri atas enam tahap. Business understanding digunakan untuk memahami masalah dan tujuan analisis. Data understanding mencakup pengumpulan ulasan, pemeriksaan struktur data, dan pelabelan manual. Data preparation meliputi pembersihan serta transformasi teks. Modeling mencakup pembagian data, oversampling, TF-IDF, dan pelatihan Multinomial Naive Bayes. Evaluation menggunakan confusion matrix, accuracy, precision, recall, dan F1-score. Deployment dilakukan dengan mengintegrasikan model ke dalam website Flask.



Gambar 1. Tahapan CRISP-DM
Figure 1. CRISP-DM Process

Gambar 1., menunjukkan tahapan CRISP-DM yang digunakan dalam penelitian ini, meliputi Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment. Tahapan tersebut dilakukan secara sistematis untuk mengolah data ulasan pengunjung Pasar Dayak Samarinda hingga menghasilkan model klasifikasi sentimen menggunakan algoritma Multinomial Naive Bayes.

3.4 Pengumpulan dan Pelabelan Data

Data dikumpulkan dari ulasan Google Maps melalui web scraping dan disimpan dalam format CSV. Pelabelan dilakukan secara manual oleh peneliti bersama tiga orang rekan. Apabila terdapat perbedaan penilaian, label ditentukan melalui diskusi hingga tercapai kesepakatan. Prosedur ini digunakan untuk mengurangi subjektivitas dan meningkatkan konsistensi label.

3.5 Data Preparation

Tahapan preprocessing terdiri atas case folding, cleaning, repetition removal, normalisasi kata slang, tokenizing, stopword removal, dan stemming. Case folding menyeragamkan seluruh huruf menjadi huruf kecil. Cleaning menghapus URL, angka, tanda baca, karakter selain huruf, dan spasi berlebih. Repetition removal menghapus pengulangan kata berurutan. Normalisasi slang mengubah kata tidak baku menjadi kata baku. Tokenizing memecah kalimat menjadi kata, stopword removal menghapus kata yang kurang informatif, sedangkan stemming mengubah kata berimbuhan menjadi kata dasar menggunakan Sastrawi.

3.6 Pemodelan dan Evaluasi

Dataset akhir dibagi menggunakan `train_test_split` dengan rasio 80:20, stratify, dan `random_state=42`. Data latih diseimbangkan menggunakan `RandomOverSampler`. Pipeline pemodelan menggabungkan `TfidfVectorizer` dan `MultinomialNB`. Parameter TF-IDF meliputi `analyzer=word`, `ngram_range=(1,2)`, `max_features=5000`, `min_df=2`, `max_df=0.95`, dan `sublinear_tf=True`. `Multinomial Naive Bayes` menggunakan `alpha=1.0`. Kestabilan model diuji menggunakan Stratified 5-Fold Cross Validation dengan metrik F1-weighted.

4. PEMBAHASAN

4.1 Karakteristik dan Persiapan Data

Dataset awal berjumlah 245 ulasan. Setelah cleaning dan penghapusan data yang tidak memenuhi kriteria, data menjadi lebih bersih dan siap diproses. Hasil penggabungan data preprocessing dengan label manual menghasilkan 199 data yang digunakan untuk pemodelan. Distribusi data menunjukkan kelas positif lebih dominan dibandingkan kelas netral dan negatif sehingga diperlukan penyeimbangan kelas pada data latih.

Tabel 1. Komposisi Data Pemodelan

Table 1. Modeling Data Composition

Jenis Data	Jumlah	Persentase
Data latih	159	79,90%
Data uji	40	20,10%
Total	199	100%

4.2 RandomOverSampler

Sebelum oversampling, data latih terdiri atas 109 data positif, 32 data netral, dan 18 data negatif. Ketimpangan tersebut berpotensi membuat model lebih cenderung

mengenali kelas positif. `RandomOverSampler` menambah sampel kelas minoritas hingga setiap kelas memiliki 109 data. Dengan demikian, jumlah data latih setelah oversampling menjadi 327 sampel dan setiap kelas memiliki peluang belajar yang lebih seimbang.

Tabel 2. Distribusi Data Latih Sebelum dan Sesudah Oversampling

Table 2. Training Data Distribution Before and After Oversampling

Kelas	Sebelum	Sesudah
Positif	109	109
Netral	32	109
Negatif	18	109

4.3 Hasil Cross Validation

Stratified 5-Fold Cross Validation menghasilkan F1-weighted sebesar 0,6295; 0,5765; 0,5523; 0,5440; dan 0,5664. Nilai rata-rata adalah 0,5737 dengan standar deviasi 0,0301. Variasi yang relatif kecil menunjukkan performa model cukup konsisten pada tiap fold, meskipun tingkat performanya masih dapat ditingkatkan.

4.4 Hasil Evaluasi Model

Evaluasi dilakukan terhadap 40 data uji yang tidak digunakan pada proses pelatihan. Model memperoleh accuracy 75,00%, precision 71,27%, recall 75,00%, dan F1-score 70,41%. Hasil tersebut menunjukkan bahwa 30 dari 40 data uji berhasil diklasifikasikan dengan benar secara keseluruhan.

Tabel 3. Hasil Evaluasi Model

Table 3. Model Evaluation Results

Metrik	Nilai
Accuracy	75,00%
Precision	71,27%
Recall	75,00%
F1-Score	70,41%

Berdasarkan classification report, kelas positif memperoleh precision 0,82, recall 0,96, dan F1-score 0,89. Kelas negatif memperoleh precision 0,40, recall 0,50, dan F1-score 0,44. Kelas netral memperoleh precision 0,50, recall 0,12, dan F1-score 0,20. Model mampu mengenali komentar positif dengan sangat baik, tetapi masih kesulitan membedakan komentar netral karena pola katanya dapat beririsan dengan kelas positif maupun negatif dan jumlah data asalnya lebih sedikit.

Tabel 4. Classification Report Tiap Kelas

Table 4. Classification Report for Each Class

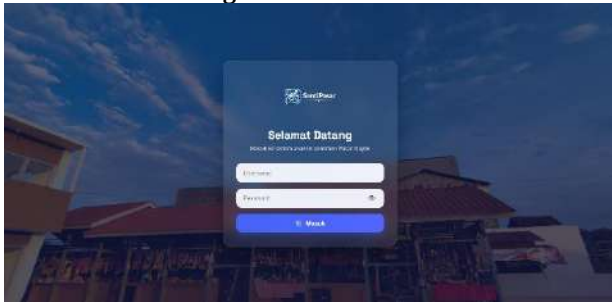
Kelas	Precision	Recall	F1-Score
Positif	0,82	0,96	0,89
Netral	0,50	0,12	0,20
Negatif	0,40	0,50	0,44

4.5 Deployment Website

Model yang telah dilatih diimplementasikan dalam website berbasis Flask. Sistem menyediakan halaman login, dashboard, dataset, analisis, hasil, dan informasi tentang sistem. Analisis dapat dilakukan secara manual

melalui input teks dan secara batch melalui file CSV. Halaman hasil menampilkan label prediksi, ringkasan distribusi sentimen, confusion matrix, serta nilai accuracy, precision, recall, dan F1-score. Implementasi ini memudahkan pengguna memanfaatkan model tanpa menjalankan kode secara langsung di Google Colab.

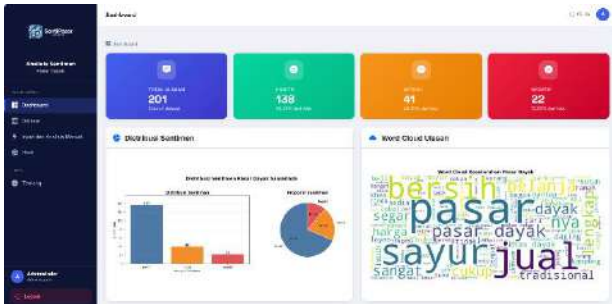
4.5.1 Halaman Login



Gambar 2. Halaman Login
Figure 2. Login Page

Halaman login merupakan tampilan awal website yang digunakan untuk proses autentikasi pengguna. Pengguna harus memasukkan username dan password yang valid sebelum dapat mengakses seluruh fitur pada sistem analisis sentimen.

4.5.2 Halaman Dashboard



Gambar 3. Halaman Dashboard
Figure 3. Dashboard Page

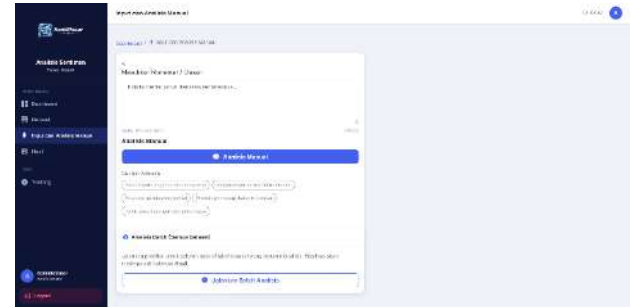
Halaman dashboard menampilkan informasi utama serta menu navigasi sistem. Melalui halaman ini, pengguna dapat mengakses fitur dataset, analisis sentimen, hasil analisis, dan informasi mengenai sistem.

4.5.3 Halaman Dataset

Gambar 4. Halaman Dataset
Figure 4. Dataset Page

Halaman dataset menampilkan kumpulan data ulasan pengunjung Pasar Dayak Samarinda yang diperoleh dari Google Maps. Data tersebut menjadi sumber utama yang digunakan dalam proses pelatihan dan pengujian model klasifikasi sentimen.

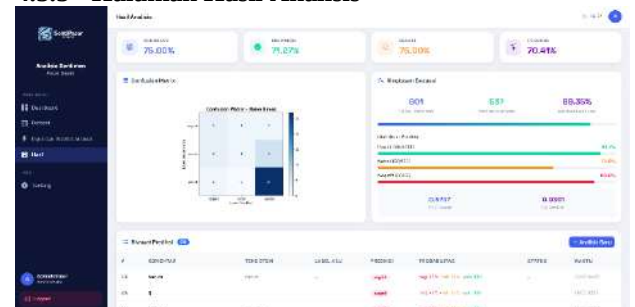
4.5.4 Halaman Analisis Sentimen



Gambar 5. Halaman Analisis Sentimen
Figure 5. Sentiment Analysis Page

Halaman analisis sentimen menyediakan dua metode prediksi, yaitu analisis manual dengan memasukkan satu komentar secara langsung dan analisis batch dengan mengunggah file berformat CSV yang berisi banyak ulasan. Sistem kemudian memproses data menggunakan model Multinomial Naive Bayes dan menampilkan hasil klasifikasi sentimen secara otomatis.

4.5.5 Halaman Hasil Analisis



Gambar 6. Halaman Hasil Analisis
Figure 6. Analysis Results Page

Halaman hasil analisis menampilkan hasil prediksi sentimen beserta informasi pendukung, seperti distribusi sentimen, confusion matrix, serta nilai Accuracy, Precision, Recall, dan F1-Score. Halaman ini memudahkan pengguna dalam mengevaluasi kinerja model yang telah dibangun.

4.6 Analisis Hasil

Dominasi performa kelas positif menunjukkan bahwa model memperoleh pola kata positif yang lebih kuat dan lebih banyak pada data awal. Walaupun RandomOverSampler menyeimbangkan jumlah sampel, metode tersebut menggandakan sampel kelas minoritas dan tidak selalu menambah variasi linguistik baru. Karena itu, penambahan ulasan negatif dan netral yang lebih beragam diperkirakan lebih efektif untuk meningkatkan

kemampuan generalisasi daripada hanya menambah duplikasi data.

Nilai accuracy 75% menunjukkan model telah mampu memberikan gambaran umum sentimen pengunjung. Namun, nilai F1 kelas netral yang rendah menandakan hasil prediksi kelas tersebut perlu ditafsirkan secara hati-hati. Model lebih sesuai digunakan sebagai alat bantu analisis awal, bukan sebagai satu-satunya dasar pengambilan keputusan pengelolaan pasar.

5. KESIMPULAN

Penelitian berhasil menerapkan CRISP-DM untuk mengolah ulasan Google Maps Pasar Dayak Samarinda dan membangun model analisis sentimen menggunakan TF-IDF, RandomOverSampler, serta Multinomial Naive Bayes. Dataset akhir berjumlah 199 data, yang dibagi menjadi 159 data latih dan 40 data uji. Model memperoleh Accuracy 75,00%, Precision 71,27%, Recall 75,00%, dan F1-Score 70,41%. Kelas positif memiliki performa terbaik, sedangkan kelas netral masih sulit dikenali. Model berhasil diimplementasikan pada website Flask sehingga dapat digunakan untuk prediksi sentimen manual maupun batch serta menampilkan hasil evaluasi.

6. SARAN

Penelitian selanjutnya disarankan menambah jumlah ulasan, terutama pada kelas negatif dan netral, menggunakan kamus slang yang lebih lengkap, menangani emoji dan bahasa daerah, serta membandingkan Naive Bayes dengan Support Vector Machine, Random Forest, atau deep learning. Website juga dapat dikembangkan agar mampu mengambil ulasan secara otomatis dan menyediakan visualisasi yang lebih interaktif.

7. REFERENSI

- Alhaqq, R. I., Putra, I. M. K., & Ruldeviyani, Y. (2022). *Analisis sentimen terhadap penggunaan aplikasi MySAPK BKN di Google Play Store*, 11(2), 105–113.
- Apriliansyah, R. D. R., Astuti, R., Prihartono, W., & Hamonangan, R. (2025). *Penerapan algoritma Naive Bayes untuk analisis sentimen pengunjung di Pantai Kejawan*, 13(1).
- Azizah, H., Syuhada, F., & Sa'adati, Y. (2024). *Sentiment analysis of tourist destinations based on Google Maps reviews using the Naive Bayes method: Case study of Bukit Merese*, 7, 467–476.
- Budianto, I., & Anwar, S. N. (2022). *Analisis sentimen pengguna Twitter mengenai program vaksinasi COVID-19 menggunakan algoritma Naive Bayes*, 6(1).
- Doloksaribu, H. P., & Samuel, Y. T. (2022). *Komparasi algoritma data mining untuk analisis sentimen aplikasi PeduliLindungi*, 16(1), 1–11.
- Fatikhah, T. Z., Muhtarom, M., & Oktaviani, I. (2024). *Analisis sentimen ulasan Mie Gacoan Solo Veteran di Google Maps*, 6(3), 649–658.
- Gondowastradmojo, A., & Kusumastuti, R. (2024). *Analisa sentimen ulasan di Tokopedia dengan metode Naive Bayes*, 369–382.
- Krisdiyanto, T., & Nurharyanto, E. M. O. (2021). *Analisis sentimen opini masyarakat Indonesia terhadap kebijakan PPKM pada media sosial Twitter menggunakan Naive Bayes classifiers*, 7(1), 32–37.
- Prasetyo, S. D., Hilabi, S. S., & Nurapriani, F. (2023). *Analisis sentimen relokasi Ibu Kota Nusantara menggunakan Naive Bayes*. *Jurnal KomtekInfo*, 10, 1–7. <https://doi.org/10.35134/komtekinfo.v10i1.330>
- Rifa'i, A., Sujaini, H., & Prawira, D. (2021). *Sentiment analysis objek wisata Kalimantan Barat pada Google Maps menggunakan metode Naive Bayes*, 7(3), 400–407.
- Sugiarta, I. K. A., Dewi, P. A. C., & Utami, N. W. (2023). *Analisa sentimen mahasiswa terhadap layanan STMIK Primakara menggunakan algoritma Naive Bayes dan K-Nearest Neighbor*, 5(3), 364–372.