



Implementation of the Naïve Bayes Algorithm for Sentiment Analysis of Portos Coffee Shop Customer Reviews Using a Natural Language Processing (NLP) Approach

Yakub Mas Arbi Sitompul^{1*}, Marah Dolly Nst²

Information Systems Study Program, Faculty of Computer Science and Information Technology, Universitas Muhammadiyah Sumatera Utara, Medan, Indonesia

Corresponding Author: Yakub Mas Arbi: yakubsitompul24@gmail.com

ARTICLE INFO

Keywords: Sentiment analysis, Multinomial Naïve Bayes, Natural language processing, TF-IDF, Coffee shop

Received : 27, July

Revised : 20, August

Accepted: 23, September

©2026 Sitompul, Nst (s): This is an open-access article distributed under the terms of the [Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).



ABSTRACT

This study implements the Multinomial Naïve Bayes (MNB) algorithm with a Natural Language Processing (NLP) approach to classify customer sentiment toward Portos Coffee Shop and builds a web-based monitoring system. A total of 586 reviews from Google Maps (525) and Instagram (61), collected in 2026, were preprocessed with Sastrawi stemming, weighted using TF-IDF, and labeled by a hybrid rating-InSet lexicon method. Hold-out testing (80:20) achieved 87.29% accuracy and a 90.20% F1-score, while 5-fold cross-validation averaged 90.79% accuracy. Positive sentiment accounted for 62.12% of reviews. The model was deployed using FastAPI, Next.js, PostgreSQL, and Docker, and passed all unit and black-box tests.

INTRODUCTION

Indonesia plays a strategic role on the global commodity map as one of the world's largest coffee producers. Changing lifestyles have transformed coffee consumption from merely meeting a beverage need into a socio-economic and cultural phenomenon known as coffee culture. This has driven rapid growth of the coffee shop industry across national and regional areas, and business competition in this sector has consequently intensified. The sustainability of a coffee shop depends heavily on how adaptively its management responds to customers' expressions of satisfaction and disappointment.

Portos Coffee Shop, located in Aek Songsongan District, Asahan Regency, North Sumatra, is a local culinary business that continues to develop dynamically. Although it has a relatively stable customer base, marketing interactions and customer opinion channels have shifted from conventional verbal communication to digital channels (user-generated content). Customers actively share their reviews on location-review platforms such as Google Maps and in the comment sections of social media such as Instagram. The two platforms differ in text characteristics and user behavior. Google Maps typically provides semi-formal, structured reviews with relatively long narratives, accompanied by a numeric star rating from 1 to 5. Instagram, in contrast, contains short, informal opinions with a high density of slang and abbreviations, and is often filled with emojis, mentions, and hashtags.

However, the accumulation of abundant unstructured textual data creates an operational challenge. Manual evaluation (human audit) demands considerable time and effort and is highly prone to the evaluator's subjective bias. To overcome these limitations, text mining and Natural Language Processing (NLP) offer automated solutions by transforming natural-language text into structured numerical data that can be classified computationally. Text mining is the extraction of non-trivial, useful knowledge from unstructured text documents, and sentiment analysis (opinion mining) is a specific branch of it that uses NLP techniques to identify and classify the polarity of attitudes, opinions, and emotions expressed toward a particular object (Liu, 2012). At the document level, sentiment analysis is used to extract the polarity of customer opinions, such as positive and negative classes (Varian et al., 2024).

Multinomial Naïve Bayes (MNB) is a probability-based machine learning algorithm that is well established in text classification (Manning et al., 2008). It has been applied to opinion and review data in many domains, including cafe and coffee shop application reviews (Sari et al., 2023; V. R. Prasetyo et al., 2023), Google reviews (Aryandi et al., 2023; Handayanto et al., 2021), Instagram comments (Hidayatullah et al., 2022; Stephanie & Irsyad, 2024), and public opinion on social issues (Darwis et al., 2021; Sianipar et al., 2023). Comparisons with other classifiers have also been reported (Apriliani et al., 2023; Asih et al., 2023; Hilma et al., 2023), and Google Maps reviews of coffee shops have been used for recommendation modeling (Saputra, 2026). MNB was selected in this study because of its very fast computational performance, efficient processing requirements, and accuracy that remains optimal even when applied to small to medium-sized datasets.

This study aims to (1) apply the MNB algorithm with an NLP approach to a multi-channel dataset (Google Maps and Instagram) from Portos Coffee Shop; (2) analyze the overall effect of the text preprocessing stages and Term Frequency-Inverse Document Frequency (TF-IDF) weighting on classification accuracy; and (3) build a web-based sentiment monitoring system (a web-based decision support system) using a modern architecture (FastAPI, Next.js, PostgreSQL, and Docker) so that the trained model can be used functionally by the coffee shop's management.

IMPLEMENTATION AND METHODS

Research Workflow

This study used an experimental quantitative approach combined with a staged information system development method, the Waterfall model (Situmorang et al., 2021). The main stages comprised data collection, text preprocessing, hybrid labeling, TF-IDF feature extraction, training of the Multinomial Naïve Bayes model, evaluation (hold-out and cross-validation), and integrated system deployment.

Data Collection

Data were collected in 2026 using automated web-scraping scripts from two main platforms. From Google Maps, the review text, star rating, username, and timestamp were extracted. From Instagram, user comments were extracted from the official account @portos_idn.

Hybrid Data Labeling

Class labels ($C = \{\text{Positive, Negative}\}$) were assigned using a hybrid criterion. For Google Maps reviews, ratings of 4–5 were labeled Positive and ratings of 1–2 were labeled Negative, whereas ratings of 3 were passed on to scoring with the InSet lexicon. For Instagram comments and 3-star reviews, the Indonesian Sentiment Lexicon (InSet; Koto & Rahmaningtyas, 2017) was used: a total score ≥ 0 was labeled Positive and a score < 0 was labeled Negative. Finally, a manual validation by a domain expert was carried out to correct the bias of automatic labeling on sarcastic or ambiguous sentences.

Text Preprocessing

Processing Indonesian-language text poses particular morphological and lexical challenges because of the high variation of affixed words and the non-standard language used on social media. The standard preprocessing steps applied were as follows. *Cleaning and case folding* remove non-alphabetic characters (digits, punctuation, emojis, and URL symbols) and convert all letters to lowercase. *Tokenizing* splits sentences into single-word tokens. *Word normalization* converts slang, non-standard words, and abbreviations (e.g., “bgt”, “bgst”, “enakkl”) into standard words according to the Indonesian dictionary (Kamus Besar Bahasa Indonesia, KBBI). *Stopword removal* eliminates common function words that carry no contextual sentiment (e.g., “yang”, “di”, “dan”, “ke”). *Stemming* returns affixed words to their root form; this study adopted the Nazief-Adriani algorithm implemented through the Sastrawi library.

TF-IDF Weighting

The Term Frequency–Inverse Document Frequency (TF-IDF) method represents the importance of a term t in a document d within a corpus D (Manning et al., 2008). The weight $W(t,d)$ is calculated using the following equations:

$$TF(t,d) = f(t,d) / \sum f(t',d) \quad (1)$$

$$IDF(t) = \log(N / df(t)) \quad (2)$$

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

where $f(t,d)$ is the frequency of term t in document d , N is the total number of documents in the corpus ($|D|$), and $df(t)$ is the number of documents that contain term t .

Multinomial Naïve Bayes

Naïve Bayes is based on Bayes' theorem with the inherent conditional assumption that all features $X = \{x_1, x_2, \dots, x_n\}$ are absolutely independent given class C . In the Multinomial Naïve Bayes (MNB) variant, the features represent the frequency or weight of term occurrence. The conditional probability of term w_i given class c is calculated with Laplace smoothing ($\alpha = 1$), and the predicted class is the one that maximizes the log posterior (Manning et al., 2008):

$$P(w_i | c) = (\sum f(w_i, d_c) + 1) / (\sum f(w, d_c) + |V|) \quad (4)$$

$$\hat{y} = \operatorname{argmax}_{c \in C} [\log P(c) + \sum f_i \cdot \log P(w_i | c)] \quad (5)$$

where $\hat{y}$ is the predicted class, $P(c)$ is the prior probability of class c , f_i is the weight of term w_i in the document being classified, $\sum f(w_i, d_c)$ is the total weight or frequency of word w_i across all documents of class c , and $|V|$ is the total size of the unique vocabulary in the corpus.

RESULTS AND DISCUSSION

Data Collection and Cleaning Results

The initial collection produced a total of 638 raw reviews. After the filtering stage (removing empty entries, texts consisting solely of a single emoji, promotional bot comments, and duplicates across platforms), a clean dataset of 586 reviews was obtained. The distribution of the data is shown in Table 1.

Table 1. Final Dataset Distribution by Data Source and Sentiment Class

Data Source	Positive	Negative	Total Data	Percentage (%)
Google Maps	332	193	525	89.59
Instagram	32	29	61	10.41
TOTAL	364	222	586	100.00

Overall, positive opinions dominate customers' perceptions of Portos Coffee Shop at 62.12% (364 reviews), while negative opinions account for 37.88% (222 reviews).

Qualitative Analysis of the Text Preprocessing Stages

Applying the NLP pipeline to an actual review confirms a significant reduction of noise and a simplification of word dimensions. Table 2 traces the changes in text structure at each processing stage.

Table 2. Illustration of Sentence Changes at Each NLP Preprocessing Stage

Preprocessing Stage	Processed Text
Raw data	"Tempatnya Nyaman Bgt, Kopi susu-nya ENAK bgt tapi service-nya lambat!! :("
Cleaning and case folding	<i>tempatnya nyaman bgt kopi susunya enak bgt tapi servicenya lambat</i>
Tokenizing	<i>['tempatnya', 'nyaman', 'bgt', 'kopi', 'susunya', 'enak', 'bgt', 'tapi', 'servicenya', 'lambat']</i>
Word normalization	<i>['tempatnya', 'nyaman', 'banget', 'kopi', 'susunya', 'enak', 'banget', 'tapi', 'pelayanan', 'lambat']</i>
Stopword removal	<i>['tempatnya', 'nyaman', 'banget', 'kopi', 'susunya', 'enak', 'banget', 'pelayanan', 'lambat']</i>
Stemming (Sastrawi)	<i>['tempat', 'nyaman', 'banget', 'kopi', 'susu', 'enak', 'banget', 'layan', 'lambat']</i>

Note. The raw review is in Indonesian; in English it reads: "The place is really comfortable, the milk coffee is really delicious but the service is slow!! :(".

The reduction in the number of unique words (vocabulary reduction) after all preprocessing stages reached 48.3%, which substantially reduced the sparsity of the TF-IDF matrix representation.

TF-IDF Weighting

Feature extraction used a combination of unigrams and bigrams with a minimum document frequency threshold ($\text{min_df} = 2$). The features with the highest TF-IDF weights in each sentiment class are presented in Table 3. The positive terms relate to taste, comfort, and friendliness, whereas the negative terms relate to slow service, price, room temperature, and disappointment.

Table 3. Top Text Features by Mean TF-IDF Weight

Rank	Positive Term	TF-IDF Weight	Negative Term	TF-IDF Weight
1	enak (delicious)	0.0842	lambat (slow)	0.0911
2	nyaman (comfortable)	0.0761	mahal (expensive)	0.0825
3	ramah (friendly)	0.0689	panas (hot)	0.0784
4	murni (pure)	0.0543	kecewa (disappointed)	0.0652
5	tempat nyaman (comfortable place)	0.0491	lama layan (long service)	0.0610

Note. Terms are shown in their original (stemmed) Indonesian form, with English glosses in parentheses.

Performance Evaluation of the Multinomial Naïve Bayes Model

The dataset of 586 reviews was divided into 80% training data (468 reviews) and 20% test data (118 reviews) using stratified random sampling. Model performance was assessed with accuracy, precision, recall, and F1-score (Azhari et al., 2021). Testing on the test data produced an accuracy of 87.29%, a precision of 86.25%, a recall of 94.52%, and an F1-score of 90.20%. To ensure the stability of the model against variation in data samples, 5-fold cross-validation was carried out, as summarized in Table 4.

Table 4. Results of the 5-Fold Cross-Validation

Fold	Training Data	Test Data	Accuracy (%)
Fold 1	468	118	91.53
Fold 2	468	118	89.83
Fold 3	469	117	92.31
Fold 4	469	117	88.89
Fold 5	469	117	91.45
AVERAGE	-	-	90.79

The average accuracy of 90.79% shows that the MNB algorithm has a high level of generalization and shows no sign of overfitting.

Ablation Study: Effect of Preprocessing on Performance

To measure how vital each NLP sub-process is, a stepwise scenario test (ablation study) was conducted. The experimental results are compared in Table 5.

Table 5. Effect of NLP Preprocessing Combinations on Model Accuracy

Preprocessing Configuration	Accuracy (%)	F1-Score (%)
No preprocessing (raw text)	71.18	75.40
Cleaning + case folding + tokenizing	78.81	82.15
+ Slang word normalization	83.05	86.10
+ Stopword removal	84.74	87.90
+ Sastrawi stemming (complete)	87.29	90.20

The data show a substantial performance improvement (+16.11 percentage points in accuracy) from the condition without preprocessing to the complete NLP pipeline. The largest single gain came from the cleaning, case folding, and tokenizing stage (+7.63 points), followed by slang normalization (+4.24 points), stemming (+2.55 points), and stopword removal (+1.69 points). This is consistent with studies that apply text preprocessing before Multinomial Naïve Bayes classification (Nurian & Sari, 2023).

System Architecture and Testing

The sentiment monitoring system was built on a Docker-based microservices architecture (Putra et al., 2020), with a client-server separation between the backend and the user interface (Ulumi & Perwitasari, 2024). The backend was developed using FastAPI (Python; Lathkar, 2023), exposes RESTful API endpoints (Adi Pranata et al., 2018), and loads the Scikit-learn, Joblib, and Sastrawi libraries. The visualization interface was built using Next.js (TypeScript; Sofi'ie & Qoiriah, 2023), and PostgreSQL serves as the relational database (Nugroho & Winarko, 2013).

The software testing results were as follows. First, functional (black-box) testing (Dewi et al., 2023) of 14 main functionality scenarios (API endpoints, CSV import/export, real-time prediction, and retraining) achieved a 100% success rate, with all scenarios valid. Second, automated unit testing using Pytest on the backend modules reached a code coverage of 91.04%. Third, the inference process for real-time sentiment prediction of a single review took 0.18 seconds on average.

CONCLUSIONS AND RECOMMENDATIONS

The Multinomial Naïve Bayes (MNB) algorithm proved highly effective in classifying the multi-channel customer reviews of Portos Coffee Shop, with an accuracy of 87.29%, a precision of 86.25%, a recall of 94.52%, an F1-score of 90.20%, and an average 5-fold cross-validation accuracy of 90.79%. Comprehensive natural language processing proved crucial, improving classification accuracy by 16.11 percentage points compared with raw text. The sentiment distribution shows a majority of positive perceptions (62.12%), while the negative aspects (37.88%) are concentrated on service speed and the comfort of the air-conditioning facilities. The monitoring system based on FastAPI, Next.js, and Docker was implemented reliably, with a code coverage of 91.04% and a latency below 0.2 seconds.

The management of Portos Coffee Shop is advised to use the dashboard to prioritize improvements to the air-conditioning facilities and to serving efficiency during peak hours. Future research could develop the handling of sarcastic sentences using deep learning approaches (such as IndoBERT) and extend the analysis to Aspect-Based Sentiment Analysis (ABSA).

REFERENCES

- Adi Pranata, I. A., et al. (2018). Rancang bangun RESTful API sistem informasi geografis [Design and development of a RESTful API for a geographic information system]. *Jurnal Teknologi Informasi dan Komputer*, 4(1), 45–52.
- Apriliani, D., et al. (2023). Analisis sentimen masyarakat terhadap vaksinasi Covid-19 menggunakan algoritma K-Nearest Neighbors [Sentiment analysis of public opinion on Covid-19 vaccination using the K-Nearest Neighbors algorithm]. *Jurnal Informatika dan Rekayasa Perangkat Lunak*, 5(1), 12–20.
- Aryandi, F., et al. (2023). Analisis sentimen ulasan pelanggan pada Google Maps menggunakan text mining [Sentiment analysis of customer reviews on Google Maps using text mining]. *Jurnal Sains dan Teknologi Informasi*, 8(2), 101–109.
- Asih, N. L., et al. (2023). Komparasi algoritma Decision Tree dan Naïve Bayes untuk mengukur kepuasan pelanggan Starbucks [Comparison of the Decision Tree and Naïve Bayes algorithms for measuring Starbucks customer satisfaction]. *Jurnal Sistem Informasi dan Data Science*, 2(1), 33–41.
- Azhari, M., et al. (2021). Evaluasi kinerja algoritma klasifikasi teks dalam pemrosesan bahasa alami [Performance evaluation of text classification algorithms in natural language processing]. *Jurnal Riset Komputer*, 8(3), 142–150.

- Darwis, D., et al. (2021). Penerapan algoritma Naïve Bayes dalam analisis sentimen opini publik [Application of the Naïve Bayes algorithm in public opinion sentiment analysis]. *Jurnal Teknoinfo*, 15(1), 29–35.
- Dewi, S., et al. (2023). Pengujian black box pada sistem informasi manajemen berbasis web [Black-box testing of a web-based management information system]. *Jurnal Rekayasa Perangkat Lunak*, 4(2), 88–95.
- Handayanto, A., et al. (2021). Analisis sentimen Google Review menggunakan Naive Bayes Classifier [Sentiment analysis of Google Reviews using the Naive Bayes Classifier]. *Jurnal Informatika Medika*, 3(2), 67–75.
- Hasanah, U., & Sari, Y. (2024). Klasifikasi sentimen pengguna aplikasi Maxim menggunakan Naïve Bayes Classifier [Sentiment classification of Maxim application users using the Naïve Bayes Classifier]. *Jurnal Media Informatika Budidarma*, 8(1), 205–214.
- Hidayatullah, A., et al. (2022). Analisis sentimen komentar Instagram produk Eiger menggunakan Naïve Bayes Classifier [Sentiment analysis of Instagram comments on Eiger products using the Naïve Bayes Classifier]. *Jurnal Teknologi dan Sistem Komputer*, 10(2), 115–123.
- Hilma, F., et al. (2023). Perbandingan Support Vector Machine dan Naïve Bayes dalam klasifikasi cyberbullying [Comparison of Support Vector Machine and Naïve Bayes in cyberbullying classification]. *Jurnal Edukasi dan Penelitian Informatika*, 9(1), 50–58.
- Irvandi, M., et al. (2023). Analisis sentimen wisata halal Pulau Lombok menggunakan Naïve Bayes dan wordcloud [Sentiment analysis of halal tourism in Lombok Island using Naïve Bayes and wordcloud]. *Jurnal Ilmiah Komputasi*, 22(3), 310–318.
- Koto, F., & Rahmaningtyas, G. Y. (2017). InSet lexicon: Evaluation of a wordlist for Indonesian sentiment analysis. In *Proceedings of the 1st International Conference on Asian Language Processing (IALP)* (pp. 1–6).
- Lathkar, M. (2023). *Python API development with FastAPI*. Apress.
- Lestandy, M., et al. (2021). Analisis sentimen citra merek menggunakan algoritma Naïve Bayes [Sentiment analysis of brand image using the Naïve Bayes algorithm]. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 5(4), 720–727.
- Liu, B. (2012). *Sentiment analysis and opinion mining*. Morgan & Claypool Publishers.

- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press.
- Nugroho, A., & Winarko, E. (2013). *Sistem manajemen basis data relasional* [Relational database management systems]. Andi Offset.
- Nurian, R., & Sari, Y. (2023). Penerapan preprocessing teks dan Multinomial Naïve Bayes pada analisis sentimen [Application of text preprocessing and Multinomial Naïve Bayes in sentiment analysis]. *Jurnal Algoritma*, 20(1), 89–98.
- Prasetyo, S. D., et al. (2023). Komparasi algoritma Naïve Bayes dan KNN pada sentimen relokasi IKN [Comparison of the Naïve Bayes and KNN algorithms on sentiment regarding the relocation of the national capital (IKN)]. *Jurnal Teknologi Informasi*, 7(1), 40–48.
- Prasetyo, V. R., et al. (2023). Analisis sentimen ulasan Cafe Kopi Paste menggunakan Naïve Bayes dan K-Nearest Neighbor [Sentiment analysis of Cafe Kopi Paste reviews using Naïve Bayes and K-Nearest Neighbor]. *Jurnal Diterima (Teknologi Informasi dan Ilmu Komputer)*, 10(3), 511–520.
- Putra, E., et al. (2020). Implementasi containerization menggunakan Docker pada arsitektur microservices [Implementation of containerization using Docker in a microservices architecture]. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 4(8), 2600–2608.
- Saputra, A. (2026). Sistem rekomendasi coffee shop berdasarkan ulasan Google Maps menggunakan Logistic Regression [Coffee shop recommendation system based on Google Maps reviews using Logistic Regression]. *Jurnal Sistem Komputer dan Informatika*, 7(2), 180–189.
- Sari, P., et al. (2023). Analisis sentimen ulasan aplikasi Fore Coffee menggunakan algoritma Naïve Bayes [Sentiment analysis of Fore Coffee application reviews using the Naïve Bayes algorithm]. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 7(1), 112–122.
- Setiaji, P., & Sastra, R. (2021). *Pemodelan perangkat lunak berbasis UML* [UML-based software modeling]. Graha Ilmu.
- Sianipar, N., et al. (2023). Analisis sentimen opini kereta cepat Jakarta-Bandung menggunakan Naïve Bayes [Sentiment analysis of opinions on the Jakarta-Bandung high-speed train using Naïve Bayes]. *Jurnal Eduikasi Informatika*, 6(2), 99–107.
- Situmorang, H., et al. (2021). Penerapan Waterfall model dalam pengembangan sistem informasi [Application of the Waterfall model in information system development]. *Jurnal Tekno Kompak*, 15(2), 130–141.

- Sofi'ie, M., & Qoiriah, A. (2023). Pengembangan aplikasi web menggunakan Next.js dan React [Web application development using Next.js and React]. *Jurnal Manajemen Informatika*, 13(1), 55–64.
- Stephanie, A., & Irsyad, M. (2024). Penerapan Multinomial Naïve Bayes dalam deteksi cyberbullying pada Instagram [Application of Multinomial Naïve Bayes in cyberbullying detection on Instagram]. *Jurnal Media Informatika*, 5(1), 15–24.
- Ulumi, M., & Perwitasari, A. (2024). Arsitektur client-server dalam pemrograman web modern [Client-server architecture in modern web programming]. *Jurnal Informatika Mulawarman*, 19(1), 40–47.
- Varian, M., et al. (2024). Document level sentiment analysis for Indonesian online reviews. *IEEE Access*, 12, 14200–14212.
- Weske, M. (2019). *Business process management: Concepts, languages, architectures*. Springer.
- Wulandari, R., et al. (2022). Analisis kualitas perangkat lunak menggunakan standard ISO/IEC 25010 [Software quality analysis using the ISO/IEC 25010 standard]. *Jurnal Sistem Informasi*, 11(2), 75–84.