

Analysis of Changes in Customer Sentiment in the Period Before and After the Google Review-Based Digital Promotion Program as a Basis for Managerial Decision Making at Kopi Kenangan

Haikal Yudistira Simarmata

Program Studi Sistem Informasi Fakultas Ilmu Komputer Dan Teknologi Informasi
Universitas Muhammadiyah Sumatera Utara

Corresponding Author e-mail: haikalsimarmata@gmail.com

Article History:

Received: 23-09-2026

Revised: 24-09-2026

Accepted: 29-09-2026

Keywords: *Sentiment Analysis, Google Reviews, Kopi Kenangan, Multinomial Naive Bayes, Digital Promotion*

Abstract : *Customer reviews on Google Reviews provide valuable information for understanding consumer responses to products and services. This study aims to analyze customer sentiment before and after a digital promotion program and classify reviews using Multinomial Naive Bayes. The study analyzed 500 reviews from 15 Kopi Kenangan outlets in North Sumatra, consisting of 250 reviews from the pre-promotion period and 250 from the post-promotion period. The data were processed using text preprocessing, TF-IDF with unigram and bigram features, an 80:20 train-test split, and oversampling of the training data. The model achieved 71.00% accuracy and a macro F1-score of 61.48%. Positive sentiment increased from 48.40% to 63.20%, while negative sentiment decreased from 31.60% to 20.40%. These results indicate a shift toward a more positive sentiment distribution after the promotion period. However, the change cannot be interpreted as a direct effect of the promotion because other factors were not controlled. The findings can support management in monitoring customer sentiment and evaluating service and marketing activities.*

How to Cite: Haikal Yudistira Simarmata. (2026). The title. Ambidextrous Journal of Innovation Efficiency and Technology in Organizations. 5(3). Pp 480-492
<https://doi.org/10.61536/ambidextrous.v5i03.893>



<https://doi.org/10.61536/ambidextrous.v5i03.893>

This is an open-access article under the [CC-BY-SA License](https://creativecommons.org/licenses/by-sa/4.0/).



Introduction

The Indonesian coffee industry is experiencing rapid growth with the emergence of various modern coffee shops, including Kopi Kenangan. This development coincides with the increasing use of digital technology in marketing activities and consumer interactions. In this environment, consumers are increasingly utilizing online reviews, particularly Google Reviews, to convey their experiences, assessments, and responses to products and services. Customer reviews can directly reflect perceptions, satisfaction, and dissatisfaction, making them a valuable source of information that companies can utilize to understand consumer response trends. Sentiment analysis of these reviews allows unstructured

information to be transformed into a snapshot of customer opinion trends and can be used to identify areas of service that require attention (Munaspin et al., 2025).

The use of digital promotions is also part of marketing activities that can shape consumer perceptions. Activities through social media platforms like Instagram and TikTok, including the use of brand ambassadors, can increase brand awareness and foster engagement with consumers. In the context of the coffee business, digital marketing activities can be related to shaping consumer experiences, brand image, and perceptions of a brand. Therefore, comparing customer responses before and after digital promotional activities can be used to identify changes in customer sentiment trends. This information can support management in evaluating customer responses and considering future marketing strategies.

The challenges in processing customer reviews lie in the unstructured nature of the data and the diversity of language used to convey experiences. The increasing number of reviews makes manual opinion grouping less efficient, necessitating computational processing to more systematically identify sentiment trends. Sentiment analysis can group reviews into positive, neutral, and negative categories, while the Naive Bayes algorithm can be used to classify text based on word patterns within the reviews. The application of these methods to Kopi Kenangan customer reviews demonstrates that online review data can be processed into more structured information to support managerial evaluation (Munaspin et al., 2025).

Another issue relates to the need to understand changes in customer response trends over different periods. Sentiment analysis results in the coffee industry and other service industries suggest that review data can be used to understand customer perceptions of service quality. However, changes in customer sentiment cannot be directly attributed to digital promotions because customer responses can also be influenced by other factors, such as product quality, service, price, and experience during a visit (Azlan & Rahmi, 2025). Beyond promotions, service quality remains a crucial element in shaping consumer perceptions, as customer reviews can provide information on response trends that can be used to prioritize improvements and evaluate services (Anim & Indiani, 2020).

Based on these conditions, an analysis is needed that can systematically process Google Reviews and compare sentiment distributions across two observation periods. The integration of digital promotions, Google Reviews, and sentiment analysis can provide information on changes in customer opinion trends, which can then be used as considerations in managerial evaluations (Munaspin et al., 2025; Aprilianti et al., 2025). In this study, digital promotions were not positioned as a variable whose influence was directly tested, but rather used as context in determining the division of observation periods. The Pre-Promotion Period covered customer reviews from July to December 2025, while the Post-Promotion Period covered reviews from January to June 2026. The study also did not identify promotional exposure for each customer or analyze each type of promotion separately.

This study aims to determine the distribution of Kopi Kenangan customer sentiment based on Google Review reviews in the period before and after the digital promotion program, applying the Multinomial Naive Bayes method to classify review texts into positive, neutral, and negative categories, and analyzing changes in sentiment distribution in both periods as supporting information in managerial decision making. This research is important because the results of processing customer reviews can provide more structured information regarding consumer response tendencies and assist the managerial evaluation process. The novelty of the study lies in the comparison of the distribution of Kopi Kenangan Google Review review sentiment in the period before and after the digital promotion activity using Multinomial Naive Bayes, so that the analysis results are not only used for text classification, but also presented as supporting information in managerial evaluation.

Method

This study uses sentiment analysis and text mining approaches to process Kopi Kenangan customer reviews obtained from Google Reviews. Text mining is used to extract information from unstructured text data so that it can be processed into more structured information for analysis purposes (Kumar et al., 2021). The research data consists of 500 Google Review reviews from 15 Kopi Kenangan outlets in North Sumatra. The data is divided into two observation periods: July to December 2025 as the Pre-Promotion Period and January to June 2026 as the Post-Promotion Period. Ratings are used as the basis for forming sentiment labels, while review text is used for the classification process using

Multinomial Naive Bayes. The sentiment analysis approach to customer reviews has also been applied in previous research on coffee shop reviews using Naive Bayes (Azhar et al., 2021). Digital promotions in this study are only used as a context for dividing the observation period and are not treated as a variable causing sentiment changes.

The research instrument was the Google Review dataset, which had review_time, review, rating, and period attributes. Data processing was performed using Python and the Flask framework with the assistance of the Pandas, Sastrawi, Scikit-learn, and Openpyxl libraries. The text data first went through the stages of case folding, cleaning, tokenization, normalization, stopword removal, and stemming. These stages were used to clean and standardize the text before the classification process because review data can contain spelling variations, non-standard words, symbols, and other elements that can be noisy (Abidin et al., 2024; Romadhony et al., 2024). After preprocessing, the text was converted into numeric form using TF-IDF weighting with a combination of unigrams and bigrams. TF-IDF is used to assign weights to words based on their occurrence in the document and their level of occurrence throughout the document (Azhar et al., 2021). Next, the data was stratified into 80% training data and 20% testing data, consisting of 400 training data sets and 100 testing data sets. Oversampling was applied only to the training data to reduce class imbalance. The training data was then used to build a Multinomial Naive Bayes model. The combination of TF-IDF and Multinomial Naive Bayes has been applied to classify text data and customer reviews (Anggina et al., 2022). Model performance was evaluated using a confusion matrix, accuracy, macro precision, macro recall, and macro F1-score.

The research data source was customer reviews on Google Reviews from 15 Kopi Kenangan outlets in North Sumatra. The research sample consisted of 500 Google Reviews divided equally based on the observation period: 250 reviews in the Pre-Promotion Period and 250 reviews in the Post-Promotion Period. The Pre-Promotion Period covered reviews from July to December 2025, while the Post-Promotion Period covered reviews from January to June 2026. Each data set included the review time, review text, rating, and period. Sentiment labels were generated based on Google Review ratings, with ratings 1 and 2 categorized as negative sentiment, rating 3 as neutral sentiment, and ratings 4 and 5 as positive sentiment. The Naive Bayes approach to customer review classification has also been used in research on Google Reviews and the food and beverage industry (Sakdiyah, 2023; Sari et al., 2023).

The research procedure was carried out in stages, starting from dataset collection and validation, sentiment label formation, text preprocessing, data division, TF-IDF weighting, oversampling, Multinomial Naive Bayes modeling, model evaluation, and comparison of sentiment distributions between periods. The dataset was first checked to ensure the completeness of the review_time, review, rating, and period columns. Next, the ratings were used to form three sentiment categories: negative, neutral, and positive. The review text was then processed through case folding, cleaning, tokenization, normalization, stopword removal, and stemming. The preprocessed data was represented using TF-IDF with a combination of unigrams and bigrams, then divided stratified with a ratio of 80% training data and 20% testing data. Oversampling was only performed on the training data, while the testing data retained its original distribution. Next, the training data was used to build a Multinomial Naive Bayes model and the model was tested using the testing data. Evaluation was carried out based on accuracy, macro precision, macro recall, macro F1-score, and confusion matrix. The preprocessing stage is crucial in the classification process because variations in text processing techniques can affect the performance of Multinomial Naive Bayes (Holle et al., 2024). After the model was evaluated, the study compared the sentiment distribution in the Pre-Promotion Period and the Post-Promotion Period. This comparison was based on sentiment labels derived from ratings, so changes in sentiment distribution were not directly influenced by model prediction errors. The analysis results were then presented through a web-based application as supporting information for managerial evaluation.

Results and Discussion

System Implementation

This study developed a web-based customer sentiment analysis system for Kopi Kenangan using the Multinomial Naive Bayes algorithm. The system was used to process Google Reviews and compare sentiment distributions between the pre- and post-promotion periods.

The application was developed using Python and the Flask framework, while the interface was created using HTML and CSS. MySQL is used to store user account data during registration and login. The dataset, preprocessing results, models, and evaluation results are stored in Excel, Joblib, and JSON formats.

The research dataset consists of 500 reviews: 250 reviews before and 250 reviews after the promotion. The data contains review_time, review, rating, and period attributes. Before classification, the review texts underwent case folding, cleaning, tokenization, normalization, stopword removal, and stemming.

Sentiment labels are determined based on ratings. Ratings 1 and 2 are categorized as negative, 3 as neutral, and 4 and 5 as positive. The data is then automatically divided into 400 training data sets and 100 testing data sets.

The preprocessed text was converted into TF-IDF features using a combination of unigrams and bigrams. To reduce class imbalance, oversampling was applied only to the training data, increasing the number to 669. The best model used a min_df value of 1 and an alpha of 0.7.

Model results are evaluated using accuracy, macro precision, macro recall, macro F1-score, and confusion matrix. The system also displays a comparison of the number and percentage of sentiments across the two periods to support managerial decision-making.

System Page Implementation

Interface implementation is the stage of applying the design developed in Chapter III to a web-based application. Each page is designed to support a series of sentiment analysis steps, from user authentication, dataset upload, text preprocessing, Multinomial Naive Bayes model training, model performance evaluation, to presenting sentiment comparisons between the pre- and post-promotion periods.

The application interface was developed using HTML and CSS, while data processing and page navigation were managed using the Flask framework. The system consists of Login, Registration, Dashboard, Dataset Review, Preprocessing, Naive Bayes, Model Evaluation, and Promotion Comparison pages. All pages are connected through a navigation menu in the sidebar, allowing users to perform data processing steps in a structured manner.

1. Login Page

The login page serves as a gateway for users before accessing the system's main features. On this page, users must enter their registered username and password. The system then checks this data against the account data stored in the MySQL database.

Before the authentication process begins, the system checks the completeness of the input fields. If the username or password is missing, the system will display a notification that both fields are required. Once the data is complete, the system will search for the account based on the username and password combination provided.

If the user data matches, the system saves the username to the session and redirects the user to the Dashboard page. The session is used to maintain the login status while the user is running the application. If the data does not match, the system displays a message stating that the username or password is incorrect. An error message is also displayed when the system experiences a connection problem with the database.

The login page aims to restrict access to dataset uploads, preprocessing, model training, and analysis results. Users who fail to authenticate will be redirected to the login page when attempting to access key system features. This mechanism aligns with the application's authentication design and implementation.

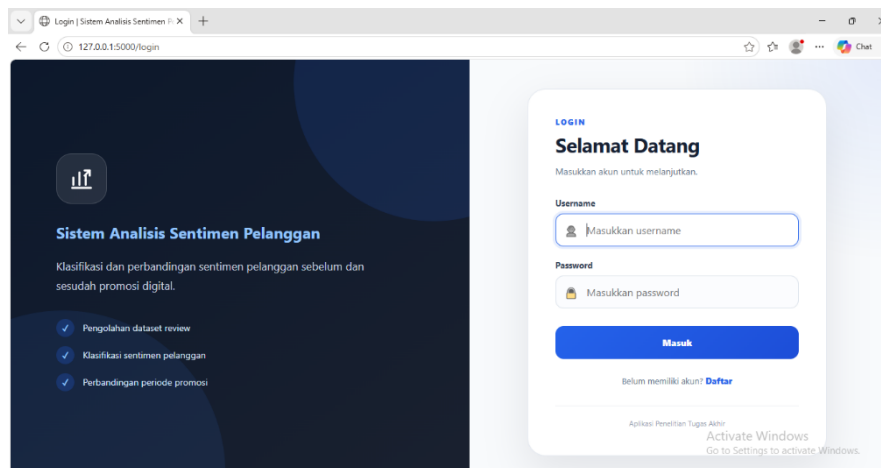


Figure 1. Login Page

2. Registration Page

The registration page is used to create an account before users log in to the system. Users enter a username and password, then the system checks the data for completeness and saves it to the users table in the MySQL database.

If registration is successful, the user is redirected to the login page. If the username is already in use or a save error occurs, the system displays a notification prompting them to change their account details. The created account is used to access all sentiment analysis features.

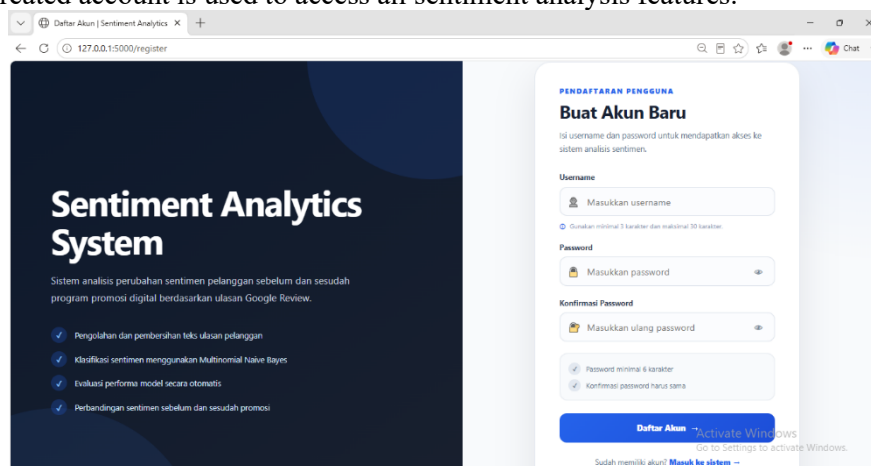


Figure 2. Registration Page

3. Dashboard Page

The dashboard page appears after a user successfully logs in. This page provides general information about the application and a summary of the dataset, including data volume, number of columns, validation status, and review breakdown between the pre- and post-promotion periods.

In addition to displaying a data summary, the dashboard provides navigation menus to the Dataset Review, Preprocessing, Naive Bayes, Model Evaluation, and Promotion Comparison pages. These pages allow users to view the initial state of the dataset and proceed with the analysis process sequentially.

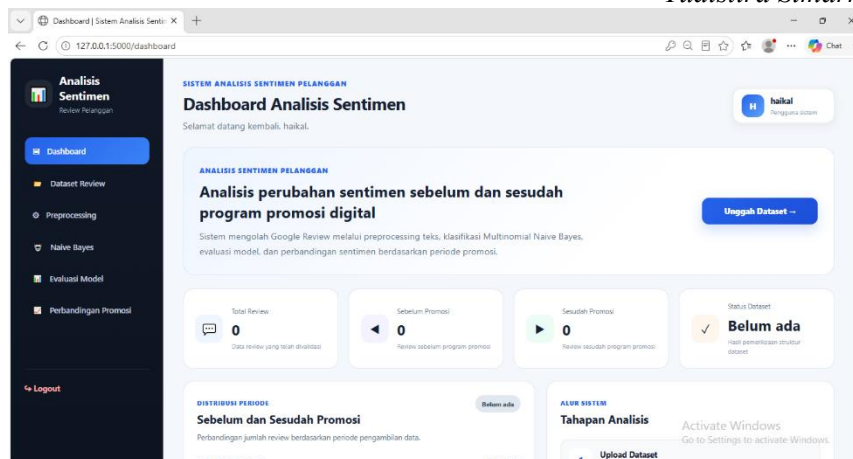


Figure 3. Dashboard page

4. Dataset Review Page

The Review Dataset page allows you to enter customer review datasets into the system in CSV or XLSX format. Once the file is uploaded, the system validates the format and the presence of the review_time, review, rating, and period columns.

If the dataset is declared valid, the system displays information such as the file name, number of data points, number of columns, file size, validation status, and the number of reviews in the pre- and post-promotion periods. The research dataset consists of 500 reviews, evenly divided into 250 Pre-Promotion and 250 Post-Promotion reviews. Some initial data is also displayed as a preview to ensure the dataset is successfully read by the system.

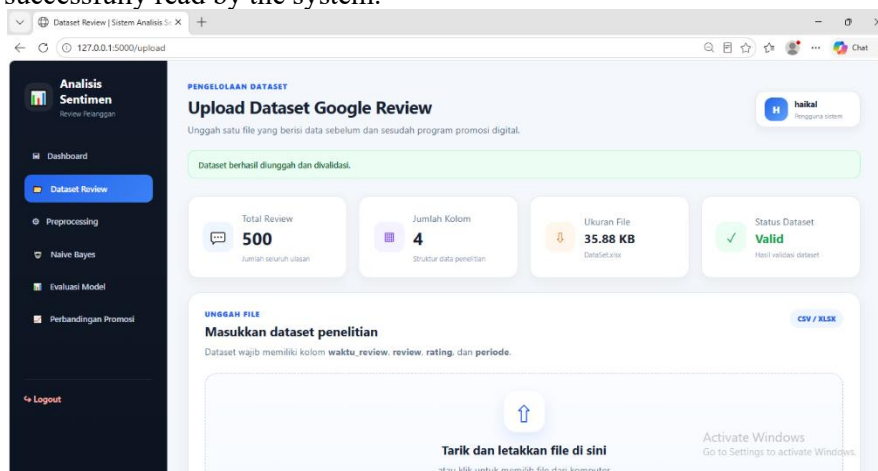


Figure 4. Dataset Review Page

5. Preprocessing Page

The Preprocessing page is used to prepare review text for use in the classification process. The steps include case folding, cleaning, tokenization, normalization, stopword removal, and stemming.

Once the process is complete, the system displays the amount of data processed, the processing time, the preprocessing file name, and the distribution of positive, neutral, and negative sentiment. This page also provides examples of text changes from the initial review to the results at various preprocessing stages.

Based on the system's results, all 500 reviews were successfully processed. Rating-based labeling yielded 279 positive reviews, 91 neutral reviews, and 130 negative reviews. The preprocessing results were then saved in an Excel file for use in the Multinomial Naive Bayes training phase.

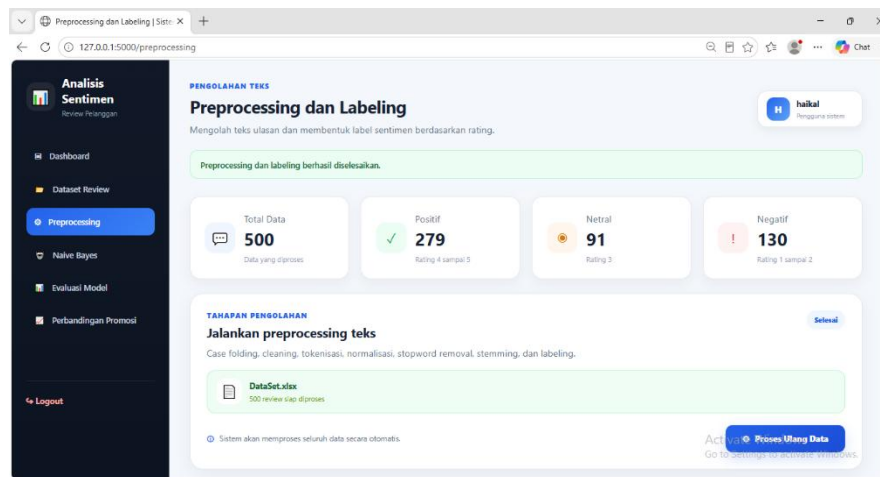


Figure 5. Preprocessing Page

6. Naïve Bayes Page

The Naive Bayes page was used to train the sentiment classification model. The preprocessed data was automatically divided into 80% training data and 20% testing data, resulting in 400 training data sets and 100 testing data sets.

The review text was then converted into a TF-IDF representation using a combination of unigrams and bigrams. To address class imbalance, the system applied oversampling to the training data, resulting in 669 data sets. A Multinomial Naive Bayes model was then constructed using the best parameters, namely a min_df of 1 and an alpha value of 0.7.

After the training process is complete, the system displays the data volume, the number of TF-IDF features, the class distribution before and after oversampling, the model parameters, the training time, and an evaluation summary. The model and TF-IDF vectorizer are also saved in Joblib format for reuse.

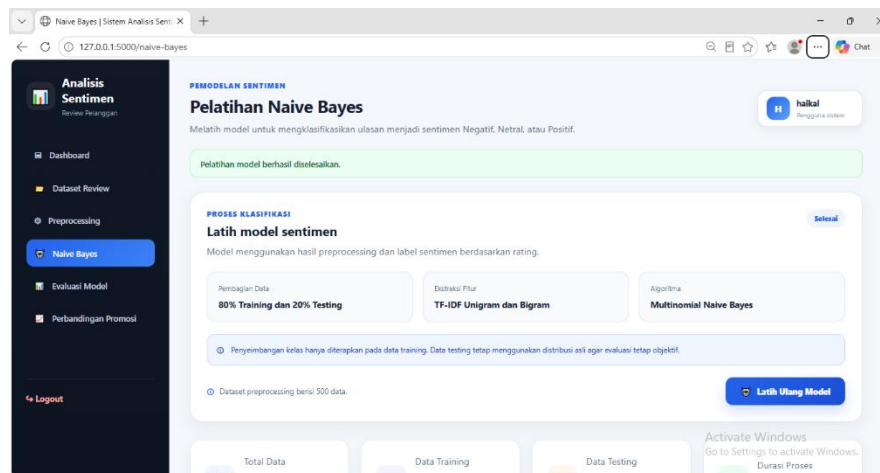


Figure 6. Naive Bayes page

7. Model Evaluation Page

The Model Evaluation page displays the results of the Multinomial Naive Bayes test using 100 test data sets. Information displayed includes accuracy, macro precision, macro recall, macro F1-score, the number of correct and incorrect predictions, and a confusion matrix.

The test results showed that the model achieved an accuracy of 71.00%, a macro precision of 61.62%, a macro recall of 61.36%, and a macro F1-score of 61.48%. Of the total test data, 71 reviews were correctly classified, while 29 were misclassified.

A confusion matrix was used to assess the model's ability to recognize each sentiment category. The positive class performed best with an F1-score of 86.73%, while the neutral class achieved the

lowest F1-score of 38.89%. This indicates that the model is quite capable of sentiment classification, but still has limitations in distinguishing neutral reviews.

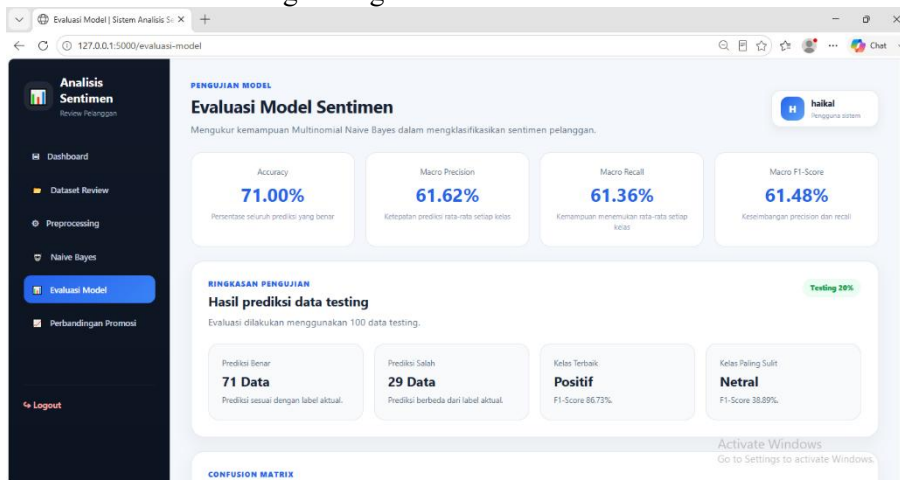


Figure 7. Model Evaluation Page

8. Promotion Period Comparison Page

The Promotion Comparison page displays changes in customer sentiment distribution between the Pre-Promotion and Post-Promotion periods. The calculations on this page are based on sentiment labels derived from Google Review ratings. Ratings 1 and 2 are categorized as negative, 3 as neutral, and 4 and 5 as positive.

The information displayed includes the number of reviews, dominant sentiment, percentage of each category, direction of change, and managerial recommendations. Of the 250 reviews, positive sentiment increased from 121 (48.40%) to 158 (63.20%). Negative sentiment decreased from 79 (31.60%) to 51 (20.40%). Meanwhile, neutral sentiment decreased from 50 (20.00%) to 41 (16.40%).

Based on these results, the Post-Promotion period showed a trend toward a more positive sentiment distribution compared to the preceding period. However, these results are descriptive in nature and cannot be used to conclude that the promotion was the direct cause of the change in customer sentiment.

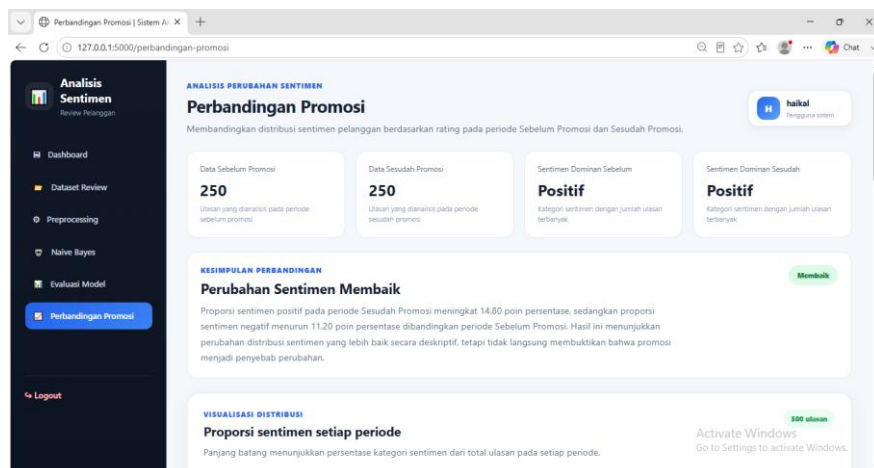


Figure 8. Period Comparison Page

Dataset Processing Results

This study used 500 Kopi Kenangan customer reviews obtained from Google Reviews. The dataset was evenly divided into 250 reviews from the Pre-Promotion period and 250 reviews from the Post-Promotion period. Each data point contained the attributes review_time, review, rating, and period.

Table 1. Distribution of Customer Ratings

Rating	Before Promotion	After Promotion	Total
--------	------------------	-----------------	-------

Ambidextrous: Journal of Innovation Efficiency and Technology in Organizations

<https://doi.org/10.61536/ambidextrous.v5i03.893>

This open-access article is distributed under a Creative Commons Attribution (CC-BY-SA) 4.0 license



1	34	20	54
2	45	31	76
3	50	41	91
4	53	64	117
5	68	94	162
Total	250	250	500

Sentiment labeling is based on ratings. Ratings 1 and 2 are grouped as negative sentiment, rating 3 as neutral sentiment, and ratings 4 and 5 as positive sentiment. The results of this grouping are shown in Table 2.

Table 2. Distribution of Sentiment Labels

Sentiment	Before Promotion	After Promotion	Total	Presentation
Negative	79	51	130	26.00%
Neutral	50	41	91	18.20%
Positive	121	158	279	55.80%
Total	250	250	500	100%

Before being used in the classification process, all reviews were processed through case folding, cleaning, tokenization, normalization, stopword removal, and stemming. The results showed that all 500 data sets were successfully processed and saved in an Excel file for use in the model training phase.

Based on the label distribution, positive sentiment was the most dominant class, with 279 reviews. Negative sentiment accounted for 130 reviews, while neutral sentiment was the least numerous, with 91 reviews. This imbalance was the reason for oversampling the training data to ensure a more proportional model development process.

Training Results and Model Evaluation

The preprocessed and labeled data were then used to build a Multinomial Naive Bayes classification model. The system automatically divided the 500 data sets into an 80:20 ratio: 400 training data sets and 100 testing data sets. This division was stratified to ensure that positive, neutral, and negative classes were equally represented in both data sets.

The preprocessed text was then converted into numeric data using the TF-IDF method. Based on the parameter selection process, the combination of unigrams and bigrams with a min_df value of 1 produced 4,976 text features. To address the difference in data size for each class, oversampling was applied to the training data, increasing the number from 400 to 669. The testing data was not included in the oversampling or model training process.

Table 3. Results of the Model Training Process

Component	Results
Total dataset	500 data
Training data	400 data
Testing data	100 data
Training data after oversampling	669 data
Number of TF-IDF features	4,976 features
Best N-gram	Unigrams and bigrams
min_df	1
alpha	0.7

The selected model used an alpha value of 0.7. Compared to the baseline model, the selected model experienced an increase in accuracy from 69.00% to 71.00%. Macro recall and macro F1-score also increased, although macro precision decreased.

Table 4. Comparison of Base Model and Selected Model

Metric	Basic Model	Selected Model
Accuracy	69.00%	71.00%
Macro precision	82.51%	61.62%
Macro recall	51.26%	61.36%
Macro F1-score	50.52%	61.48%

The selected model was then tested using 100 test data sets. The test results showed that 71 data sets were correctly classified, while 29 were misclassified.

Table 5. Model Evaluation Results

Metric	Results
Accuracy	71.00%
Macro precision	61.62%
Macro recall	61.36%
Macro F1-score	61.48%
Correct prediction	71 data
Wrong prediction	29 data

The accuracy value is calculated from the number of correct predictions compared to all testing data.

$$Accuracy = \frac{71}{100} \times 100\% = 71,00\%$$

Details of the errors and accuracy of the classification of each class can be seen through the following confusion matrix.

Table 6. Confusion Matrix

Actual \ Prediction	Negative	Neutral	Positive
Negative	15	8	3
Neutral	6	7	5
Positive	4	3	49

Based on the confusion matrix, the model successfully recognized 15 negative data points, 7 neutral data points, and 49 positive data points correctly. The performance of each class is shown in Table 7.

Table 7. Evaluation of Each Sentiment Class

Sentiment	Precision	Recall	F1-score
Negative	60.00%	57.69%	58.82%
Neutral	38.89%	38.89%	38.89%
Positive	85.96%	87.50%	86.73%

The positive class performed best with an F1-score of 86.73%. Conversely, the neutral class had the lowest F1-score of 38.89%. This indicates that the model is quite capable of recognizing positive reviews, but still has difficulty distinguishing neutral reviews from negative and positive classes. Overall, the model has moderate performance and can be used to support sentiment analysis, while still being aware of its limitations.

Comparative Sentiment Analysis

Table 8. Comparison of Sentiment Before and After Promotion

Sentiment	Before	Presentation Before	After	Presentation After	Change
Negative	79	31.60%	51	20.40%	-11.20 points
Neutral	50	20.00%	41	16.40%	-3.60 points
Positive	121	48.40%	158	63.20%	+14.80 points
Total	250	100%	250	100%	-

The percentage of positive sentiment in the Pre-Promotion period is obtained by comparing the number of positive reviews to the total number of reviews in that period. The calculation is as follows:

$$Persentase\ Positif\ Sebelum = \frac{121}{250} \times 100\% = 48,40\%$$

Next, the percentage of positive sentiment in the Post-Promotion period is calculated using the same method, namely:

$$Persentase\ Positif\ Sesudah = \frac{158}{250} \times 100\% = 63,20\%$$

The magnitude of the change in positive sentiment is obtained from the percentage difference between the Post-Promotion and Pre-Promotion periods.

$$63,20\% - 48,40\% = 14,80 \text{ poin persentase}$$

Based on the comparison, the number of positive sentiments increased from 121 reviews to 158, representing a 14.80 percentage point increase. Conversely, the number of negative sentiments decreased from 79 reviews to 51, representing a 11.20 percentage point decrease. Neutral sentiments also decreased from 50 reviews to 41, representing a 3.60 percentage point decrease.

Positive sentiment was the most dominant category in both periods. However, the percentage was higher in the Post-Promotion period, at 63.20%, compared to the Pre-Promotion period, which reached 48.40%. The decrease in the proportion of negative sentiment also indicates that fewer customers gave low ratings in the Post-Promotion period.

Descriptively, the analysis results show a shift in customer sentiment distribution toward a more positive direction in the Post-Promotion period. These findings can be used as information for management in evaluating the development of customer perceptions. However, this study only illustrates the difference in sentiment distribution between the two periods and does not prove that the promotional program was the direct cause of these changes.

System Functional Testing

Table 9. System Functional Test Results

No.	Feature	Testing Scenario	Expected results	Status
1	Login	User inputs correct username and password	The system directs the user to the dashboard page.	Succeed
2	Registration	Users fill in account registration data	The new account is saved and can be used to log in.	Succeed
3	Upload Dataset	Users upload datasets in Excel format	The dataset was successfully read and displayed by the system.	Succeed
4	Preprocessing	The user runs the preprocessing stage	The system displays the results of processing the review text.	Succeed
5	Naive Bayes	The user runs the model training process	The system generates sentiment prediction models and results.	Succeed
6	Model Evaluation	User opens evaluation page	The system displays the accuracy, precision, recall, F1-score, and confusion matrix values.	Succeed
7	Promotion Comparison	User opens comparison page	The system displays a comparison of sentiment before and after the promotion.	Succeed
8	Logout	The user selects the logout menu	The system ends the user session and redirects the user to the login page	Succeed

Based on the testing conducted, the system's main functions performed as expected. The system can perform user authentication, dataset processing, preprocessing, sentiment classification, model evaluation, and presentation of sentiment comparison results. This testing focused on the system's core functions and did not include security or application performance testing.

Discussion

This research focuses on the performance of the classification model and changes in sentiment distribution between periods. Multinomial Naive Bayes performance was measured using 100 test datasets, while pre- and post-promotion sentiment comparisons were calculated from all 500 reviews based on sentiment labels derived from ratings. This separation was done to prevent model prediction errors from affecting the inter-period comparison results.

Based on the test results, the model achieved an accuracy of 71.00%, a macro precision of 61.62%, a macro recall of 61.36%, and a macro F1-score of 61.48%. These values indicate that the model has moderate classification performance, but its ability to recognize each sentiment class is still uneven.

Positive sentiment was the class with the best results, with an F1-score of 86.73%. Negative sentiment achieved an F1-score of 58.82%, while neutral sentiment had the lowest score at 38.89%. The

low performance of the neutral class indicates that word patterns in neutral reviews are more difficult to distinguish because they can be similar to both positive and negative reviews.

The comparison results show that positive sentiment increased from 48.40% to 63.20%. Negative sentiment decreased from 31.60% to 20.40%, while neutral sentiment decreased from 20.00% to 16.40%. These findings indicate a trend toward more positive customer responses in the Post-Promotion Period.

However, changes in sentiment distribution cannot yet be attributed to the direct impact of digital promotions. This study did not identify each customer's use of promotions and did not control for other factors, such as product quality, price, service, and in-store customer experience.

The developed system can assist users in processing reviews, evaluating model performance, and observing changes in sentiment trends between periods. The results can be used to support managerial evaluations, but still require consideration in conjunction with operational data and other business conditions.

Conclusion

This study shows that Google Review sentiment analysis using Multinomial Naive Bayes can be used to process and classify Kopi Kenangan customer responses. Of the 500 reviews analyzed, 279 reviews or 55.80% included positive sentiment, 130 reviews or 26.00% included negative sentiment, and 91 reviews or 18.20% included neutral sentiment. The model was built using the division of 400 training data and 100 testing data, TF-IDF with a combination of unigrams and bigrams, and oversampling on the training data. The evaluation results showed an accuracy of 71.00%, macro precision 61.62%, macro recall 61.36%, and macro F1-score 61.48%. Descriptively, the distribution of positive sentiment increased from 48.40% in the Pre-Promotion Period to 63.20% in the Post-Promotion Period, while negative sentiment decreased from 31.60% to 20.40% and neutral sentiment from 20.00% to 16.40%.

The study's findings are limited by the fact that the sentiment comparison was based on only two observation periods, and digital promotions were not tested as a causal variable for sentiment changes. The study also failed to identify the level of promotional exposure for each customer and did not control for other factors that could influence customer responses, such as product quality, price, service, and experience during the visit. Furthermore, the model's performance in the neutral class was still low, with an F1-score of 38.89%. Therefore, the classification results should be interpreted with caution. Future research could utilize a longer observation period, a larger dataset, and consider other variables related to customer experience and promotional activities. Practically, the study's findings can be used as supporting information for Kopi Kenangan management to monitor customer sentiment trends, evaluate changes in review distribution, and determine service aspects that require further attention.

Bibliography

- Abidin, Z., Junaidi, A., & Wamiliana. (2024). Text stemming and lemmatization of regional languages in Indonesia: A systematic literature review. *Journal of Information Systems Engineering and Business Intelligence*, 10(2), 217–231. <https://doi.org/10.20473/jisebi.10.2.217-231>
- Anggina, S., Setiawan, N. Y., & Bachtiar, F. A. (2022). Analisis ulasan pelanggan menggunakan Multinomial Naïve Bayes Classifier dengan Lexicon-Based dan TF-IDF pada Formaggio Coffee and Resto. *@is The Best: Accounting Information Systems and Information Technology Business Enterprise*, 7(1), 76–90. <https://doi.org/10.34010/aisthebest.v7i1.7072>
- Anim, A., & Indiani, N. L. P. (2020). Pengaruh promosi dan kualitas layanan terhadap keputusan pembelian kembali (Studi kasus pada Start Up Coffee Renon Denpasar). *Jurnal Ilmiah Manajemen dan Bisnis*, 5(1), 99–108. <https://journal.undiknas.ac.id/index.php/manajemen/article/view/2432>
- Aprilianti, A., Pertiwi, S. H., Putra, M. R. I., Apriani, E., Latif, A., & Permana, I. (2025). Strategi digital marketing terhadap keputusan pembelian pada brand Kopi Kenangan. *Journal Economics and Strategy*, 5(2), 32–42. <https://doi.org/10.36490/jes.v5i2.1457>

- Azhar, N., Adikara, P. P., & Adinugroho, S. (2021). Analisis sentimen ulasan kedai kopi menggunakan metode Naïve Bayes dengan seleksi fitur algoritme genetika. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 8(3), 609–618. <https://doi.org/10.25126/jtiik.202184436>
- Azlan, M., & Rahmi, E. (2025). Analisis sentimen ulasan pada Google Review di sebuah penginapan menggunakan algoritma Naïve Bayes: Studi kasus Grand Jatra Hotel Pekanbaru. *Neptunus: Jurnal Ilmu Komputer dan Teknologi Informasi*, 3(3), 240–247. <https://doi.org/10.61132/neptunus.v3i3.1003>
- Holle, K. F. H., Alfianita, R., & Putri, H. M. (2024). Evaluasi teknik preprocessing terhadap kinerja Multinomial Naïve Bayes dalam klasifikasi pertanyaan insincere. *JUSTIN (Jurnal Sistem dan Teknologi Informasi)*, 12(4), 707–715. <https://doi.org/10.26418/justin.v12i4.82758>
- Kumar, S., Kar, A. K., & Ilavarasan, P. V. (2021). Applications of text mining in services management: A systematic literature review. *International Journal of Information Management Data Insights*, 1(1), 100008. <https://doi.org/10.1016/j.ijime.2021.100008>
- Munaspin, Z. D. P., Titiana, N. M., Tsabitah, L., Karima, D. A., Tania, K. D., & Meiriza, A. (2025). Knowledge discovery terhadap sentimen pelanggan Kopi Kenangan menggunakan Naïve Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(3), 5303–5308. <https://doi.org/10.36040/jati.v9i3.14080>
- Romadhony, A., Al Faraby, S., Rismala, R., Wisesty, U. N., & Arifianto, A. (2024). Sentiment analysis on a large Indonesian product review dataset. *Journal of Information Systems Engineering and Business Intelligence*, 10(1), 167–178. <https://doi.org/10.20473/jisebi.10.1.167-178>
- Sakdiyah, H. (2023). *Analisis sentimen customer review brand Kopi Kenangan menggunakan metode Naive Bayes* [Skripsi, Universitas Islam Negeri Maulana Malik Ibrahim Malang]. e-Theses UIN Malang. <http://etheses.uin-malang.ac.id/50274/>
- Sari, T. A., Sinduningrum, E., & Hasan, F. N. (2023). Analisis sentimen ulasan pelanggan pada aplikasi Fore Coffee menggunakan metode Naïve Bayes. *KLIK: Kajian Ilmiah Informatika dan Komputer*, 3(6), 773–779. <https://doi.org/10.30865/klik.v3i6.884>