

Classification of Customer Opinions on the Quality of Cooperative Minimarket Services Using the Lexicon Approach

Muhammad Sholeh, Uning Lestari, Dina Andayati

Universitas AKPRIND Indonesia, Yogyakarta, Indonesia

Article Info

Article history:

Received February 4, 2026

Revised February 16, 2026

Accepted March 6, 2026

Keywords:

Comments

Lexicon

Logistic Regression

Minimarket

Sentiment Analysis

ABSTRACT

Almost every strategic location today has a store that sells daily necessities. These stores compete with each other by offering prices and services that they hope will satisfy their customers. This competition must be anticipated in the management of cooperatives' minimarkets. Minimarkets run by cooperatives need to maintain the loyalty of their members and general customers by improving service quality. Customer reviews, suggestions, and criticisms from members or customers are valuable sources of data for evaluating service performance. These customer reviews are unstructured data that is difficult to process manually. This study aims to classify customer opinions on the service quality of cooperative minimarkets into positive, negative, and neutral sentiments using a Lexicon-Based approach. The research methods used are text data preprocessing, sentiment weighting using a lexicon dictionary, classification into positive, negative, or neutral classes, and system performance testing using a confusion matrix. The data labeling stage is automated using the Lexicon InSet dictionary to determine the sentiment class (positive or negative). The labeled data was then processed using TF-IDF feature extraction and used to train the logistic regression model. Model performance evaluation was carried out using a Confusion Matrix with a training data and test data ratio of 80:20. The results of this study show that the logistic regression algorithm is capable of classifying cooperative service sentiment with an accuracy rate of 81%, precision of 83%, recall of 81%, and an F1 score of 79%. These results indicate that the method is quite effective at identifying customer opinions and can serve as a decision-support system for cooperative managers to continuously improve service quality by analyzing customer sentiment data.

Copyright ©2026 The Authors.

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



Corresponding Author:

Muhammad Sholeh

Universitas AKPRIND Indonesia, Yogyakarta, Indonesia,

Email: muhash@akprind.ac.id

How to Cite: M. Sholeh, U. Lestari, & D. Andayani, "Classification of Customer Opinions on the Quality of Cooperative Minimarket Services Using the Lexicon Approach," *International Journal of Engineering and Computer Science Applications (IJECSA)*, vol. 5, no. 1, pp. 23–32, Mar. 2026. doi: [10.30812/ijecs.v5i1.6172](https://doi.org/10.30812/ijecs.v5i1.6172).

1. INTRODUCTION

Competition in the retail business is currently becoming increasingly fierce. Almost every strategic location has a kiosk or a retail store selling a variety of daily necessities. In fact, retail stores are not only located in strategic areas, but have also entered residential neighborhoods. This situation requires every retail business to be able to compete, not only in terms of price, but also in terms of the quality of service provided to consumers [1]. One important effort to increase competitiveness is to provide excellent service and pay attention to customer feedback. This feedback is usually conveyed through comments, suggestions, or criticism regarding staff service, product availability, prices, or shopping convenience. However, customer comments are generally conveyed in the form of unstructured free text, making them difficult to analyze manually when there are a large number of them [2, 3]. Analysis of customer comments is very important because it can help measure customer satisfaction levels. Positive comments indicate that customers are satisfied with the service provided, while negative comments indicate dissatisfaction that needs to be addressed immediately. This information can serve as a basis for managerial decision-making to determine strategies for improving service quality and developing the retail business in the future.

Minimarkets managed by cooperatives also face similar challenges. In addition to serving general consumers, cooperative minimarkets also have members who are expected to show high loyalty. However, member loyalty cannot be used as the sole measure of service success. Service quality evaluations must continue to be conducted continuously, both for cooperative members and general customers, to maintain the sustainability of the cooperative business. Sentiment analysis is an approach in the field of text mining that can be used to identify a person's opinion or feelings towards an object based on text [4, 5]. Using sentiment analysis, customer comments can be automatically classified into positive, negative, or neutral sentiments [6]. One commonly used approach is the Lexicon-Based method [7]. The Lexicon-Based method is carried out by utilizing a dictionary of words that have been given sentiment weights to determine the tendency of opinions in a text [8, 9].

Research on sentiment analysis has been conducted across various fields. A literature study shows the application of sentiment analysis to evaluate user reviews of online shopping platforms such as Tokopedia [10], Shopee [11], and Lazada [12]. Analysis of e-commerce reviews generally uses classification algorithms such as Naïve Bayes [12], Random Forest [13], and Decision Tree [14], which are used to classify positive, negative, and neutral sentiments. The data collection process for sentiment analysis was obtained from various sources, including the Google Play Store [15], Twitter, Instagram [16], applications [17]. In banking services, research also makes extensive use of sentiment analysis to understand customer perceptions of the quality of bank services and digital banking applications [18]. Research on sentiment analysis in the digital banking sector was conducted using social media data. The study aims to analyze public sentiment towards Bank DKI services using data obtained from Twitter [18].

In the digital age, online shopping has become part of people's daily lives. Product reviews provided by customers play an important role in helping potential buyers make purchasing decisions, while also serving as a valuable source of information for e-commerce platform managers [19]. Research was conducted by analyzing product review sentiment to understand customer perceptions and satisfaction with the products and services provided. The study applied the Naive Bayes algorithm to analyze the sentiment of product reviews on the Esrocte e-commerce platform. Another study [20] applied the K-Nearest Neighbor (KNN) algorithm to analyze the sentiment of cooperative customer comments. The research stages included filtering words in each comment, calculating the proximity of values across comment data, and determining sentiment classes based on similarity to previous data. Data processing and modeling were performed using RapidMiner software.

Against the background and introduction, the research focused on classifying customer opinions on the service quality of the "Cahaya Manunggal" cooperative minimarket in Yogyakarta using a Lexicon-Based Sentiment Analysis approach. The data used in this study were unstructured customer comments or reviews. The research began with text preprocessing, including case folding, tokenization, stopword removal, and stemming. Next, automatic sentiment labeling was performed using the Lexicon InSet dictionary to determine the positive or negative sentiment tendency in each comment. The data with sentiment labels were then feature extracted using the Term Frequency–Inverse Document Frequency (TF-IDF) method. The feature extraction results were used as input to the machine learning process, which used the logistic regression algorithm. The data were divided into training and test sets at an 80:20 ratio. Model performance was evaluated using a Confusion Matrix to measure accuracy. The 80:20 data split ratio was chosen to balance model training requirements and evaluation validity. 80% of the data was used for training, allowing the model to learn sentiment patterns optimally, while 20% was used for testing to assess the model's generalization to new data. This ratio is a common practice in machine learning research and is considered appropriate for the dataset used in this study.

Based on a literature review, there are gaps that previous studies have not addressed, including a lack of research specifically examining sentiment analysis in the context of cooperative minimarkets, which differ from modern retail and e-commerce. Most previous studies have focused on e-commerce platforms, digital applications, or banking services, using classification algorithms such as Naïve Bayes, Random Forests, Decision Trees, and K-Nearest Neighbors. Previous studies generally used data from social media or digital application platforms, while studies on customer opinions on community-based and membership-based cooperative

minimarkets are still relatively limited. Another thing is that the integration of the Lexicon-Based approach for automatic labeling, combined with TF-IDF feature extraction and the logistic regression algorithm, in the context of cooperative minimarket services, has not been explored comprehensively.

The difference between this study and previous studies is that this study specifically examines customer opinions on the quality of cooperative minimarket services by utilizing a Lexicon-Based approach for automatic sentiment labeling using the Lexicon InSet dictionary, then combining it with the TF-IDF feature extraction method and the logistic regression algorithm as a classification model. This study not only focuses on sentiment classification, but also analyzes the dominant word patterns in each sentiment category to provide an overview of the service aspects that need to be maintained or improved. Thus, this study presents the integration of lexicon and machine learning approaches in the context of retail cooperatives, an area that has rarely been studied.

The objective of this study is to develop a sentiment classification model capable of automatically, accurately, and measurably identifying customer opinions on the quality of cooperative minimarket services. The contribution of this study to the development of science lies in the application and evaluation of a combination of Lexicon-Based, TF-IDF, and logistic regression approaches in the cooperative retail domain, thereby enriching the study of sentiment analysis in the people’s economy sector. In practice, the results of this study provide benefits for cooperative managers as a data-based decision-support system for evaluating service quality, increasing customer satisfaction, and strengthening the loyalty of members and general customers on an ongoing basis.

2. RESEARCH METHOD

The research used quantitative methods, with an experimental design, to analyze customer sentiment from reviews and comments about minimarket services. The research process was carried out in several systematically arranged stages. The initial stage of the research began with the collection of unstructured data in the form of datasheets containing reviews or comments from minimarket consumers, including suggestions, criticisms, and customer impressions during transactions. The next stage was preprocessing the text data to remove unnecessary words, symbols, and other elements, making the data ready for analysis. After that, feature extraction was carried out to convert the text data into numerical form using the TF-IDF method [21]. The modeling stage involves applying the logistic regression classification algorithm to categorize customer review sentiments as positive, negative, or neutral. Next, the model is evaluated using accuracy, precision, recall, and F1-score metrics to assess its performance. Overall, the stages of this research are presented in a flowchart shown in Figure 1.

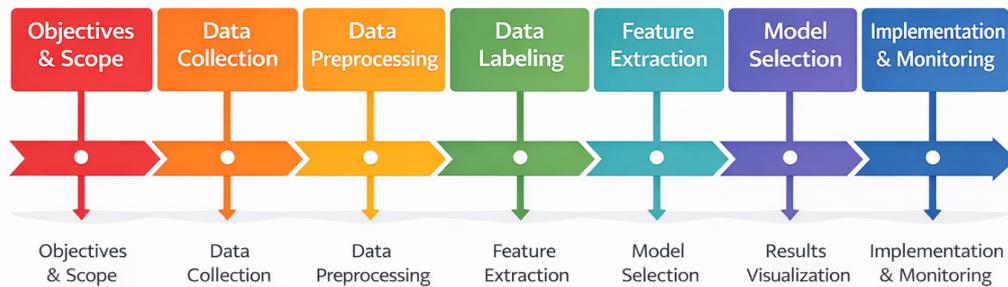


Figure 1. Effects of selecting different switching under dynamic condition

The customer reviews are sentences containing suggestions, criticism, and impressions of transactions at the minimarket. The dataset used in this study consists of 581 rows with two columns: the review date and the visitor review or comment. An example of the content of visitor reviews or comments stored in the datasheet is shown in Table 1.

Table 1. Example of The Content of Visitor Reviews or Comments Stored in the Dataset

Date	Comment
19/12/2025	Parkir cukup memadai untuk pelanggan datang
21/12/2025	Fasilitas memadai untuk ukuran minimarket kecil
22/12/2025	Pelayanan sangat baik menyenangkan
23/12/2025	Tempat kumuh tidak layak untuk berbelanja
24/12/2025	Lokasi cukup mudah ditemukan oleh orang
24/12/2025	Barang expired dijual diskon tanpa informasi jelas

Date	Comment
26/12/2025	<i>Pelayanan buruk pada saat malam hari</i>
27/12/2025	<i>Operasional normal tanpa masalah</i>
27/12/2025	<i>Kebersihan diabaikan oleh manajemen minimarket</i>
29/12/2025	<i>Barang tersedia dengan kualitas cukup memuaskan</i>
30/12/2025	<i>Harga tidak konsisten sering berubah-ubah</i>
30/12/2025	<i>Operasional lancar dari pagi hingga malam hari</i>
31/12/2025	<i>Antrian sampai keluar membuat tidak nyaman</i>

2.1. Sentiment Analysis

Sentiment analysis is an approach in text mining that aims to identify, extract, and classify opinions or expressions of feelings about a particular object based on text data. In sentiment analysis, opinions are generally grouped into three main categories: positive, negative, and neutral. This approach is widely used to evaluate customer satisfaction levels with a product or service through reviews or comments submitted in the form of free and unstructured text, thereby enabling the analysis process to be carried out systematically and automatically [22].

2.2. Text Pre-processing

Text preprocessing is the initial stage in text data processing that aims to improve data quality before further analysis. This stage includes cleaning the text of irrelevant characters or symbols, converting all letters to lowercase (case folding), removing common words that do not contribute significantly to the meaning of sentiment (stopword removal), and normalizing words. The preprocessing process is carried out to reduce noise in the data and produce a more consistent representation of the text, thereby improving the performance and accuracy of the classification model used [14].

2.3. Term Frequency–Inverse Document Frequency (TF-IDF)

In the preprocessing stage, the text data undergoes several steps: case folding, tokenization, stopwords removal, and stemming. These stages aim to reduce noise and ensure the consistency and relevance of the words used in the analysis process. In the feature extraction stage, the TF-IDF method is used to represent text documents as numerical vectors based on the importance of each term. The Term Frequency (TF) value is used to measure the frequency of occurrence of a term in a document [23]. Term Frequency (TF) Formula is shown in Equation (1) [23].

$$TF(t, d) = \frac{f(t, d)}{\sum_k f(k, d)} \quad (1)$$

In this equation, $f(t, d)$ represents the frequency of occurrence of term t in document d , while $\sum_k f(k, d)$ indicates the total number of terms in the document. Furthermore, Inverse Document Frequency (IDF) is used to measure the level of importance and uniqueness of a term based on its distribution across all documents in the corpus, which is formulated of Equation (2).

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

In this equation, N represents the total number of documents in the corpus, and $df(t)$ is the number of documents containing term t . In the final stage of feature extraction, the TF-IDF weight is obtained by multiplying the TF and IDF values, as shown in Equation (3). The numerical representation resulting from this process is then used as input in the classification modeling stage to determine the sentiment category.

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

2.4. Logistic Regression

In the classification modeling stage, the logistic regression algorithm is used to predict the probability that a data point belongs to a particular class. This algorithm is a regression-based classification method that utilizes a logistic (sigmoid) function to map a

linear combination of input features to a range of values between 0 and 1, so that it can be interpreted as class membership probability [24]. Mathematically, the sigmoid function is formulated in Equation (4).

$$\sigma(z) = \frac{1}{1+e^{-z}} \quad (4)$$

with $z = w^T x + b$

In this equation, x is the TF-IDF feature extraction vector, w represents the model weight vector, and b is the bias. The sigmoid function's output is then used to determine the data class based on a threshold.

2.5. Model Evaluation

The performance of the classification algorithm is measured to assess its ability to accurately predict sentiment classes. The evaluation is carried out using several metrics commonly used in classification, namely accuracy, precision, recall, and F1-score. These metrics assess the model's accuracy, consistency, and reliability in classifying test data. Accuracy is used to measure the proportion of correct predictions against the entire test data and is formulated in Equation (5) for accuracy and Equation (6) for precision [25, 26].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

Precision measures the accuracy of the model in predicting positive classes, which is formulated as:

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

Recall is used to measure the model's ability to recognize all data belonging to the positive class, with the Equation (7):

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

F1-score is the harmonic value of precision and recall used to provide a balance between the two metrics, formulated in Equation (8).

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (8)$$

3. RESULT AND ANALYSIS

3.1. Text Data Preprocessing

The preprocessing stage is a crucial initial step that determines the quality of the data before feature extraction and classification modeling. Based on the data processing results, all unstructured customer comments were successfully normalized through the stages of case folding, tokenizing, stopword removal, and stemming [27]. The case-folding process produced consistency in word writing by converting all text to lowercase letters, so that variations in uppercase and lowercase letters did not affect feature formation. The tokenizing stage breaks each comment into more structured word units (terms) that are ready for further processing. Furthermore, the stopword removal process can eliminate common words that do not significantly contribute to sentiment, such as conjunctions and prepositions. The stemming stage unifies various word variations into their basic forms, so that word frequency becomes more representative. Overall, this preprocessing stage successfully reduces noise in the data and produces cleaner text that is ready for feature extraction [28]. The results of the preprocessing stage are shown in Table 2.

Table 2. Results of the Preprocessing Process

Comment	Text Case Folding	Text Tokenizing	Text Stemming	Text Filtering
<i>Harga mahal tidak wajar sama sekali untuk anggota</i>	<i>harga mahal tidak wajar sama sekali untuk anggota</i>	<i>harga, mahal, tidak, wajar, sama, sekali, untuk, anggota</i>	<i>harga, mahal, tidak, wajar, sama, sekali, untuk, anggota</i>	<i>harga, mahal, wajar, anggota</i>
<i>Harga murah meriah untuk anggota koperasi setia</i>	<i>harga murah meriah untuk anggota koperasi setia</i>	<i>harga, murah, meriah, untuk, anggota, koperasi, setia</i>	<i>harga, murah, riah, untuk, anggota, koperasi, setia</i>	<i>harga, murah, riah, anggota, koperasi, setia</i>

Comment	Text Case Folding	Text Tokenizing	Text Stemming	Text Filtering
Harga sangat terjangkau untuk semua anggota koperasi	harga sangat terjangkau untuk semua anggota koperasi	harga, sangat, terjangkau, untuk, semua, anggota, koperasi	harga, sangat, jangkau, untuk, semua, anggota, kop-erasi	harga, jangkau, anggota, koperasi
Harga wajar sesuai dengan kualitas barang dijual	harga wajar sesuai dengan kualitas barang dijual	harga, wajar, sesuai, dengan, kualitas, barang, dijual	harga, wajar, sesuai, dengan, kualitas, barang, jual	harga, wajar, sesuai, kualitas, barang, jual
Karyawan lumayan membantu para pelanggan yang datang	karyawan lumayan membantu para pelanggan yang datang	karyawan, lumayan, membantu, para, pelanggan, yang, datang	karyawan, lumayan, bantu, para, langgan, yang, datang	karyawan, lumayan, bantu, langgan
Parkir nyaman aman untuk semua kendaraan	parkir nyaman aman untuk semua kendaraan	parkir, nyaman, aman, untuk, semua, kendaraan	parkir, nyaman, aman, untuk, semua, kendara	parkir, nyaman, aman, kendara
Pelayanan sangat ramah dan cepat sekali	pelayanan sangat ramah dan cepat sekali	pelayanan, sangat, ramah, dan, cepat, sekali	layan, sangat, ramah, dan, cepat, sekali	layan, ramah, cepat
Pelayanan sangat ramah dan cepat sekali memuaskan	pelayanan sangat ramah dan cepat sekali memuaskan	pelayanan, sangat, ramah, dan, cepat, sekali, memuaskan	layan, sangat, ramah, dan, cepat, sekali, puas	layan, ramah, cepat, puas
Pelayanan sangat ramah membuat nyaman berbelanja	pelayanan sangat ramah membuat nyaman berbelanja	pelayanan, sangat, ramah, membuat, nyaman, berbelanja	layan, sangat, ramah, buat, nyaman, belanja	layan, ramah, nyaman, belanja

3.2. Data Labeling Results

After going through the preprocessing stage, customer comment data was labeled with sentiment using a Lexicon-Based approach utilizing the Lexicon InSet dictionary [27, 29]. The labeling process was automated by calculating the sentiment weight of each word in the comments. This labeling resulted in three sentiment classes, namely positive, negative, and neutral, which reflect customer perceptions of the quality of cooperative minimarket services. Lexicon-based labeling enables efficient data annotation without requiring extensive manual labeling. The labeling results still depend on the completeness and accuracy of the sentiment dictionary used, which can affect the quality of the training data used in the modeling stage. The labeling results are presented in Table 3. The labeling results yielded 301 positive, 215 negative, and 65 neutral sentiments. The labeling results are presented in Figure 2.

Table 3. Labeling Results From Comments

Comment	Text Filtering	Polarity Score	Indonesia Sentiment
Pelayanan sangat ramah membuat nyaman berbelanja	['layan', 'ramah', 'nyaman', 'belanja']	5	Positive
Pelayanan sangat ramah dan cepat sekali memuaskan	['layan', 'ramah', 'cepat', 'muas']	7	Positive
Pelayanan lambat sekali saat ramai pengunjung	['layan', 'lambat', 'ramai', 'ujung']	-1	Negative
Stok selalu lengkap untuk kebutuhan sehari-hari	['stok', 'lengkap', 'butuh', 'seharihari']	-9	Negative
Pelayanan sangat ramah dan cepat sekali	['layan', 'ramah', 'cepat']	2	Positive
Parkir berbahaya untuk anak-anak dan manula	['parkir', 'bahaya', 'anakanak', 'manula']	-4	Negative
Barang tersedia dengan kualitas cukup memuaskan	['barang', 'sedia', 'kualitas', 'muas']	11	Positive
Harga tidak konsisten sering berubah-ubah	['harga', 'konsisten', 'berubahubah']	3	Positive
Operasional lancar dari pagi hingga malam hari	['operasional', 'lancar', 'pagi', 'malam']	-1	Negative
Antrian sampai keluar membuat tidak nyaman	['antri', 'nyaman']	4	Positive

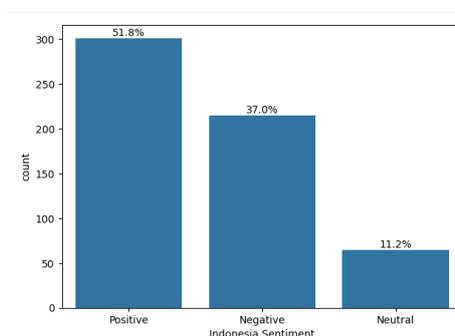


Figure 2. Negative, Positive, And Neutral Sentiment Labeling Results

A word cloud visualization representing the distribution of dominant words in each sentiment category is presented in Figure 3. The visualization shows differences in word patterns between positive, negative, and neutral sentiments, reflecting customers' focus on certain aspects of the cooperative minimarket's service quality. The word cloud visualization identifies the most frequently occurring words in each sentiment category and provides an initial overview of customer opinion trends.



Figure 3. Frequently Appearing Keywords

The clustering results presented in Figure 3 provide an overview of the linguistic patterns that distinguish each sentiment category. In positive sentiment, the dominance of the words goods, system, and stock indicates that customer satisfaction is mainly related to the operational aspects of the minimarket, particularly product availability and stock management systems, which are considered adequate. These findings indicate that technical factors and service efficiency are the main contributors to positive customer perceptions. In negative sentiment, the dominant use of the words “customer,” “employee,” and “service” reflects customer dissatisfaction with service quality and human resource interactions. In the neutral sentiment category, the words ‘clean’ and “service” are the most frequently used, indicating that customers tend to convey descriptive assessments of cleanliness and service without strong emotional expressions.

3.3. TF-IDF Feature Extraction Results

In the feature formation stage, labeled text data was extracted using the Term Frequency–Inverse Document Frequency (TF-IDF) method [21]. This method successfully converted text data into a high-dimensional vector representation that reflects the importance of each term in the document. Words that appear frequently in one comment but rarely appear in other comments obtain a higher TF-IDF weight, thus contributing more to the classification process. The feature extraction results show that TF-IDF effectively represents text data characteristics and is suitable as input to the logistic regression algorithm. This representation is also sparse, consistent with the characteristics of linear classification algorithms. The results of the TF-IDF feature extraction process are presented in Table 4.

Table 4. Results of The TF-IDF Feature Extraction Process

abai	ac	akhirakhir	akses	aktif	alami	aman	anakanak	anggota	antri	...	ukur	ventilasi	wajar	wangi	warga
0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	...	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
...
0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0

3.4. Model Evaluation

Model performance evaluation was conducted using a Confusion Matrix with accuracy, precision, recall, and F1-score metrics [30]. The evaluation results showed an accuracy of 81%, precision of 83%, and recall of 81%. These accuracy results indicate that the model can classify customer sentiment with a high level of accuracy, reflecting its ability to make general predictions. The precision of 83% indicates that most comments predicted to a sentiment class match the actual class, so prediction errors can be minimized. The recall value of 81% indicates that the model is quite capable of recognizing comments that truly belong to each

sentiment class. The F1-score value of 79% shows a good balance between precision and recall. This indicates that the model is not only accurate in making predictions but also consistent in recognizing sentiment patterns in the test data. Overall, the evaluation results prove that the combination of the Lexicon-Based approach for data labeling, the TF-IDF method for feature extraction, and the logistic regression algorithm for classification is capable of providing effective performance in analyzing the sentiment of cooperative minimarket customer opinions.

The findings of this study are that the combination of a Lexicon-Based approach for automatic labeling, feature extraction using TF-IDF, and the logistic regression algorithm is capable of classifying the sentiment of cooperative minimarket customer opinions with a good level of performance, as indicated by an accuracy value of 81%, precision of 83%, recall of 81%, and an F1-score of 79%. The model is not only effective at distinguishing positive, negative, and neutral sentiments but also at identifying dominant word patterns that represent specific service aspects. Positive sentiments are dominated by words related to product availability and inventory management systems, while negative sentiments are more related to service aspects and employee interactions. This shows that operational factors and human resource quality are the main determinants of customer perception of cooperative minimarket services.

The results of this study align with and support previous studies, which show that the TF-IDF method is effective for representing text features for sentiment classification, and that the logistic regression algorithm performs stably and can handle high-dimensional data in text analysis. Previous research also shows that the Lexicon-Based approach can serve as an efficient initial labeling method for Indonesian sentiment analysis. In addition, previous studies in the e-commerce and digital services domains report that combining TF-IDF-based feature extraction with linear classification algorithms can achieve competitive accuracy. Thus, the results of this study reinforce the state of the art that integrating lexicon and machine learning approaches is an effective strategy for sentiment analysis, while expanding its application to the context of cooperative minimarkets, which has not been widely explored in previous studies.

4. CONCLUSION

The application of the Lexicon-Based Sentiment Analysis approach in classifying customer opinions on the quality of cooperative minimarket services yielded fairly good evaluation results, with an accuracy of 81%, precision of 83%, recall of 81%, and an F1-score of 79%. The analysis was conducted on unstructured customer comments through the following stages: text pre-processing, lexicon-based sentiment labeling, feature extraction using the Term Frequency–Inverse Document Frequency (TF-IDF) method, and classification modeling using the logistic regression algorithm. These results show that the combination of methods used is capable of producing effective text data representation and supports the sentiment classification process optimally.

The sentiment analysis results show that positive customer sentiment is generally associated with the availability of goods, systems, and stock management, reflecting customer satisfaction with the operational aspects of minimarkets. Conversely, negative sentiment is dominated by words related to service and employee interaction, indicating customer dissatisfaction with certain service qualities. These findings confirm that human resources and service quality remain crucial aspects that require more attention from cooperative minimarket managers. The results of this study not only demonstrate the effectiveness of the Lexicon-Based approach combined with TF-IDF and logistic regression in classifying customer sentiment, but also provide an overview of the service aspects that need to be maintained or improved. The information obtained from the analysis of negative sentiment, in particular, can be used as a basis for strategic evaluation in efforts to improve service quality and strengthen customer loyalty to cooperative minimarkets in a sustainable manner.

ACKNOWLEDGEMENTS

We would like to express our gratitude to the Directorate of Research and Community Service, Directorate General of Research and Development, Ministry of Higher Education, Science, and Technology of the Republic of Indonesia for the funding support and trust provided in the implementation of the 2025 research. We would also like to express our appreciation to the Directorate of Research and Community Service of AKPRIND University Indonesia for the guidance, facilities, and continuous assistance during the activities.

REFERENCES

- [1] A. Sundari and A. Y. Syaikhudin, *Manajemen Ritel (Teori dan Strategi dalam Bisnis Ritel)*. Lamongan: Academia Publication, Jun. 27, 2021, 190 pp. Google Books: [jLk1EAAAQBAJ](#).

- [2] A. Wibowo, *Manajemen Bisnis Retail*. Semarang: Penerbit Yayasan Prima Agus Teknik, Jul. 12, 2021.
- [3] C. W. Utami, *Manajemen Ritel: Strategi Dan Implementasi Operasional Bisnis Ritel Modern Di Indonesia*. Jakarta: Salemba Empat, 2017.
- [4] D. Purnamasari et al., *Pengantar Metode Analisis Sentimen*. Depok: Penerbit Gunadarma, 2023.
- [5] A. Asrumi et al., *Analisis Sentimen dan Penggalian Opini*. Purbalingga: Eureka Media Aksara, Nov. 27, 2023.
- [6] Y. Yunitasari, *Teori Dan Implementasi Analisis Sentimen Menggunakan Python*. Madiun: UNIPMA Press, 2023.
- [7] S. A. S. Mola, R. V. K. I. O. Roma, and T. Widiastuti, *Text Mining: Analisis Sentimen dengan Lexicon*. Kaizen Media Publishing, Jun. 27, 2025.
- [8] F. Rachmawati, U. Azmi, and R. Azwarini, "Comparison of Lexicon-Based Methods and Bidirectional Encoder Representations for Transformers Models in Sentiment Analysis of Government Debt Market Movements," *International Journal of Engineering and Computer Science Applications (IJECSA)*, vol. 4, no. 1, pp. 13–28, Mar. 11, 2025. DOI: [10.30812/ijeCSA.v4i1.4832](https://doi.org/10.30812/ijeCSA.v4i1.4832)
- [9] M. Rodríguez-Ibáñez et al., "A review on sentiment analysis from social media platforms," *Expert Systems with Applications*, vol. 223, p. 119 862, Aug. 2023. DOI: [10.1016/j.eswa.2023.119862](https://doi.org/10.1016/j.eswa.2023.119862)
- [10] A. Faesal et al., "Sentimen Analisis pada Data Tweet Pengguna Twitter Terhadap Produk Penjualan Toko Online Menggunakan Metode K-Means," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 19, no. 2, pp. 207–213, May 30, 2020. DOI: [10.30812/matrik.v19i2.640](https://doi.org/10.30812/matrik.v19i2.640)
- [11] T. P. R. Sanjaya, A. Fauzi, and A. F. N. Masruriyah, "Analisis Sentimen Ulasan Pada E-Commerce Shopee Menggunakan Algoritma Naive Bayes Dan Support Vector Machine," *INFOTECH : Jurnal Informatika & Teknologi*, vol. 4, no. 1, pp. 16–26, Jun. 30, 2023. DOI: [10.37373/infotech.v4i1.422](https://doi.org/10.37373/infotech.v4i1.422)
- [12] T. K. A. Lutfani et al., "Penerapan Naive Bayes untuk Analisis Sentimen pada Ulasan Pelanggan di Lazada: Studi Kasus Toko Mawar Collection," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 2, Apr. 10, 2025. DOI: [10.23960/jitet.v13i2.6391](https://doi.org/10.23960/jitet.v13i2.6391)
- [13] A. P. Adhi, K. Umuri, and G. Triyono, "Sentiment Analysis and Entity Detection on News Headlines to Support Investment Decisions," *Jurnal Teknik Informatika (Jutif)*, vol. 5, no. 6, pp. 1801–1810, Dec. 29, 2024. DOI: [10.52436/1.jutif.2024.5.6.3434](https://doi.org/10.52436/1.jutif.2024.5.6.3434)
- [14] K. Kumalasari et al., "Klasifikasi Sentimen Umpan Balik Pelanggan dalam Bahasa Indonesia: Studi Kasus Global Bangunan Menggunakan SVM dan Naive Bayes," *Seminar Nasional Sekolah Tinggi Teknologi Dumai 2025*, vol. 1, no. 2, pp. 416–431, Aug. 1, 2025.
- [15] D. A. Andriati, R. L. S. Putra, and G. D. Saputra, "Analisis Sentimen Komentar Pengguna Aplikasi Klik Indomaret Di Google Playstore Menggunakan Metode Support Vector Machine (SVM)," *Jurnal Multimedia dan Teknologi Informasi (Jatilima)*, vol. 7, no. 03, pp. 663–673, Aug. 30, 2025. DOI: [10.54209/jatilima.v7i03.1653](https://doi.org/10.54209/jatilima.v7i03.1653)
- [16] A. Karim, S. F. C, and M. Mustafa, "Analisis Sentimen pada Komentar Sosial Media Instagram Layanan Kesehatan BPJS Menggunakan Naive Bayes Classifier," *Prosiding Konstelasi Ilmiah Mahasiswa Unissula (KIMU) Klaster Engineering*, no. 0, Jan. 19, 2022.
- [17] H. L. Nisa and M. H. S. Kurniawan, "Analisis Sentimen Terhadap Komentar Aplikasi Allstats BPS dengan Klasifikasi Naive Bayes: Analisis Sentimen Terhadap Komentar Aplikasi Allstats BPS dengan Klasifikasi Naive Bayes," *Emerging Statistics and Data Science Journal*, vol. 2, no. 3, pp. 315–329, Oct. 31, 2024. DOI: [10.20885/esds.vol2.iss.3.art24](https://doi.org/10.20885/esds.vol2.iss.3.art24)
- [18] F. Faturohman, M. Julaeha, and I. Kartini, "Analisis Bauran Pemasaran Pada Toko Sembako (Studi Kasus Pada Toko Sembako di Desa Sukasirna)," *Innovative: Journal Of Social Science Research*, vol. 5, no. 3, pp. 1104–1114, May 17, 2025. DOI: [10.31004/innovative.v5i3.19121](https://doi.org/10.31004/innovative.v5i3.19121)
- [19] S. Nafisyah and R. Sulistiyowati, "Analisis Sentimen Ulasan Produk Toko Online Esrocte untuk Peningkatan Pelayanan Menggunakan Algoritma Naive Bayer," *Blantika: Multidisciplinary Journal*, vol. 2, no. 8, pp. 846–852, Jun. 25, 2024. DOI: [10.57096/blantika.v2i8.189](https://doi.org/10.57096/blantika.v2i8.189)
- [20] N. K. Lingga, "Analisis Sentimen Nasabah Pada Kotak Komentar Terhadap Pelayanan Koperasi Simpan Pinjam Bangun Mandiri Menggunakan Metode K-Nearest Neighbor," *JIKTEKS : Jurnal Ilmu Komputer dan Teknologi Informasi*, vol. 2, no. 02, pp. 47–59, Apr. 30, 2024. DOI: [10.70404/jikteks.v2i02.74](https://doi.org/10.70404/jikteks.v2i02.74)

- [21] S. Winardi et al., “Utilizing TF-IDF Content-based Filtering for Job Recommendation Systems,” *Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)*, vol. 8, no. 5, pp. 2838–2845, Oct. 29, 2025. DOI: [10.32672/jnkti.v8i5.9686](https://doi.org/10.32672/jnkti.v8i5.9686)
- [22] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, “Analisis sentimen terhadap penggunaan aplikasi shopee menggunakan algoritma support vector machine (svm),” *Jambura Journal of Electrical and Electronics Engineering*, vol. 5, no. 1, pp. 32–35, Jan. 9, 2023. DOI: [10.37905/jjee.v5i1.16830](https://doi.org/10.37905/jjee.v5i1.16830)
- [23] A. R. Putra and D. E. Ratnawati, “Analisis sentimen berbasis aspek pada aplikasi mobile menggunakan naïve bayes berdasarkan ulasan pengguna playstore (studi kasus : Jconnect mobile),” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 12, no. 2, pp. 293–300, Apr. 24, 2025. DOI: [10.25126/jtiik.2025127556](https://doi.org/10.25126/jtiik.2025127556)
- [24] J. Wong, I. Sanu, and H. Irsyad, “Implementasi TF-IDF, Cosine Similarity, dan Logistic Regression Pada Rekomendasi Buku Berdasarkan Mood Pembaca Dengan Data Oversampling,” *DEVICE: Journal of Information System, Computer Science and Information Technology*, vol. 6, no. 1, Jun. 2, 2025. DOI: [10.46576/device.v6i1.6499](https://doi.org/10.46576/device.v6i1.6499)
- [25] H. Putra and R. Rumini, “Comparative study of logistic regression, random forest, and xgboost for bank loan approval classification,” *Journal of Applied Informatics and Computing*, vol. 9, no. 5, pp. 2822–2835, Oct. 19, 2025. DOI: [10.30871/jaic.v9i5.10862](https://doi.org/10.30871/jaic.v9i5.10862)
- [26] F. Madani and A. H. Lubis, “CatBoost Algorithm Implementation for Classifying Women’s Fashion Products,” *Journal of Informatics and Telecommunication Engineering*, vol. 9, no. 1, pp. 249–260, Jul. 28, 2025. DOI: [10.31289/jite.v9i1.15604](https://doi.org/10.31289/jite.v9i1.15604)
- [27] F. Chahyadi et al., “Evaluating Lexicon Weighting and Machine Learning Models for Sentiment Classification of Indonesian Mangrove Ecotourism Reviews,” *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 6, pp. 5679–5698, Dec. 23, 2025. DOI: [10.52436/1.jutif.2025.6.6.5563](https://doi.org/10.52436/1.jutif.2025.6.6.5563)
- [28] N. Agustina et al., “Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store: The Implementation of Naïve Bayes Algorithm for Sentiment Analysis of Shopee Reviews On Google Play Store,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 2, no. 1, pp. 47–54, Apr. 17, 2022. DOI: [10.57152/malcom.v2i1.195](https://doi.org/10.57152/malcom.v2i1.195)
- [29] A. O. K. Adi, A. Sanjaya, and A. B. Setiawan, “Penerapan Inset Lexicon untuk Analisis Sentimen Penonton Video JKT48 di YouTube,” *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, vol. 9, no. 2, Jul. 10, 2025. DOI: [10.29407/ynsa1795](https://doi.org/10.29407/ynsa1795)
- [30] S. Anggina, N. Y. Setiawan, and F. A. Bachtar, “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*, vol. 7, no. 1, pp. 76–90, Sep. 30, 2022. DOI: [10.34010/aisthebest.v7i1.7072](https://doi.org/10.34010/aisthebest.v7i1.7072)