Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 2109−2117
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8811

Sentiment Analysis of the Matahari Application to Provide User Experience Insights using Support Vector Machine

Moch Arif Samsul Rizal[1], Anik Vega Vitianingsih[1,*], Hewa Majeed Zangana[2], Anastasia Lidya Maukar[3], Fitri Marisa[4]

[1] Informatics Department, Universitas Dr. Soetomo, Surabaya, Indonesia
[2] IT Department, Duhok Technical College, Duhok Polytechnic University, Duhok, Iraq
[3] Industrial Engineering Department, President University, Bekasi, Indonesia
[4] Informatics Department, Universitas Widyagama Malang, Malang, Indonesia
Email: [1]mochrizal1616@gmail.com, [2,*]vega@unitomo.ac.id, [3]hewa.zangana@dpu.edu.krd, [4]almaukar@president.ac.id, [5]fitrimarisa@widyagama.ac.id
Email Penulis Korespondensi: vega@unitomo.ac.id
Submitted: **30/11/2025**; Accepted: **31/12/2025**; Published**: 31/12/2025**

Abstract−The expansion of Indonesia's digital commerce ecosystem has pushed retail companies to strengthen the quality of their online services to remain competitive. Matahari, one of the country's leading retail brands, launched its mobile app as a platform for shopping, promotions, and customer interaction. However, user feedback on the Google Play Store indicates persistent problems with system responsiveness, ease of use, and the consistency of promotional information. This study examines sentiment patterns in 2,500 user reviews and classifies them using a Support Vector Machine (SVM) based model that incorporates three kernel types: Linear, RBF, and Polynomial. Before modelling, the text corpus underwent several pre-processing steps—such as tokenization, stopword filtering, and stemming represented numerically using TF-IDF weighting. Among all tested configurations, the Linear kernel produced the strongest results, achieving an accuracy rate of 88%. Despite a moderate distribution across categories (1030 negative, 886 neutral, and 584 positive), the model achieved consistent performance across all classes. Evaluation using Precision, Recall, and F1-Score confirmed the validity of the 88% accuracy without the need for additional sampling techniques. From a scholarly standpoint, this research adds insight into sentiment analysis for retail applications within the Indonesian context by applying a machine-learning approach. In practice, the outcomes highlight areas for improvement, particularly technical stability, the intuitiveness of user flows, and promotional clarity to support a better overall user experience.

Keywords: Sentiment Classification; SVM; TF-IDF; Matahari Mobile Application; User Perception

1. INTRODUCTION

In the era of digital transformation, the success of retail companies is not only determined by product quality, but also by their ability to deliver superior user experiences through digital channels. Modern consumers have become increasingly powerful entities, demanding that companies adapt to higher expectations for quality, transparency, and service speed, as well as the ability of brands to provide valuable, enjoyable experiences [1]. These increased expectations require businesses to make optimal use of digital technology, including data analysis and artificial intelligence, to gain a deep understanding of consumer behaviour and opinions. Understanding customer opinions is now an integral part of a sustainable customer experience-based business strategy.

In Indonesia, the e-commerce sector continues to grow rapidly in line with the increasing penetration of internet users. It changes in people's lifestyles, which are becoming increasingly dependent on digital technology [2]. Matahari Department Store, one of Indonesia's largest retailers, responded to these changes by launching a mobile application in 2016. The Matahari app serves as a vital channel for customer interaction, offering online shopping features, promotional information, and store locations. To date, the application has surpassed 1 million installations on the Google Play Store, with tens of thousands of user reviews reflecting a wide range of experiences and opinions. These reviews are a valuable source of data for evaluating the app's quality and identifying areas for improvement [3] [4]. However, the massive volume makes manual analysis inefficient, prone to bias, and slow in capturing sentiment trends. As a result, developers and management find it difficult to identify key issues and determine the right priorities for improvement. This condition underscores the need for an automated text-mining-based approach to gain large-scale insights into user sentiment.

Previous studies have explored sentiment analysis of mobile app reviews using a variety of methodological approaches. One such work [5] analyzed Shopee user feedback using a Naïve Bayes–based classifier. The authors processed a dataset of 3,000 reviews through several preparation steps, including converting text to lowercase, normalizing informal words, and removing non-essential terms. The resulting model achieved an accuracy of 87.58%. However, the study noted limitations in handling imbalanced data and standard stopword removal which potentially reduced the precision for negative sentiment detection.

In contrast, the SVM shows greater consistency. Research [6] on Zalora reviews (1,000 data points) using k-fold cross-validation achieved 85% accuracy. The study confirms that SVM is suitable for classifying Indonesian text and suggests using a larger dataset for more representative results. Similarly, study [7] on Tokopedia reviews reported an accuracy of 81.25% while also demonstrating the sensitivity of SVM performance to the quality and quantity of training data.

In addition to the general classification approach, Random Forest has also been applied. Research [8] on Alfagift application reviews employed the Random Forest algorithm on a dataset of 4,379 reviews collected from the

Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 2109−2117
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8811

Google Play Store. The study reported an excellent accuracy of 97.6% using an 80:20 data split. However, the preprocessing stage in this study was limited to standard techniques such as cleaning, normalization, and stopword removal, without explicitly addressing negation handling to preserve the context of negative phrases.

Deep learning techniques have also been investigated within comparable research settings. One study [9] analyzed Shopee user reviews using TF-IDF features integrated with a Long Short-Term Memory (LSTM) network. The optimal configuration achieved 83% classification accuracy. Despite its ability to capture sequential patterns, the study explicitly removed standard stopwords, including the negation term *'tidak'* (not), during the preprocessing stage. This approach represents a significant limitation, as removing negations can invert the sentiment polarity of a sentence.

Although various methods have been applied to e-commerce platforms and mobile applications, critical gaps remain in both the domain and the methodology. First, most previous studies [5], [6], [7], [8], [9] relied on standard preprocessing techniques (stopword removal) that explicitly eliminate negation words, causing the model to lose the context of refusal or disappointment. Second, no studies have specifically highlighted user reviews on the Matahari app. Considering these gaps, this research aims to provide empirical contributions by implementing Semantic-based Normalization with Negation Concatenation. Unlike standard approaches, this study specifically maps negation phrases into single semantic tokens (e.g., *'tidak sesuai'* becomes *'tidak_sesuai'*) to prevent information loss during the cleaning process. This technique is applied to the under-studied case of the Matahari application to improve classification validity.

Accordingly, this research investigates how effectively a Support Vector Machine–based classifier performs multi-class classification to categorize user sentiments into three distinct classes (Positive, Neutral, and Negative), while also uncovering key user-experience factors that contribute to those sentiments. From an academic standpoint, the work adds to existing studies on sentiment analysis within Indonesia's fashion-retail application ecosystem by employing a quantifiable machine-learning framework. In practice, the findings may guide the development of improvement strategies targeting application stability, the accuracy of pricing and promotional information, and the reliability of transaction workflows, enabling continuous enhancement of Matahari app's overall user experience.

2. RESEARCH METHODOLOGY

To maintain methodological transparency and facilitate reproducibility, this study was conducted in a series of systematically arranged phases, as illustrated in Figure 1. The process began by retrieving user comments from the official Matahari application page on the Google Play platform. These textual entries were then processed through a comprehensive pre-processing pipeline to remove noise and standardize linguistic structure. After the cleaning phase, sentiment categories were assigned using a lexicon-driven scoring mechanism. The refined text corpus was then converted to numerical features using TF–IDF weighting. Finally, the dataset was partitioned into separate development and evaluation sets to train and validate an SVM-based classifier, followed by an analysis of its predictive outcomes.

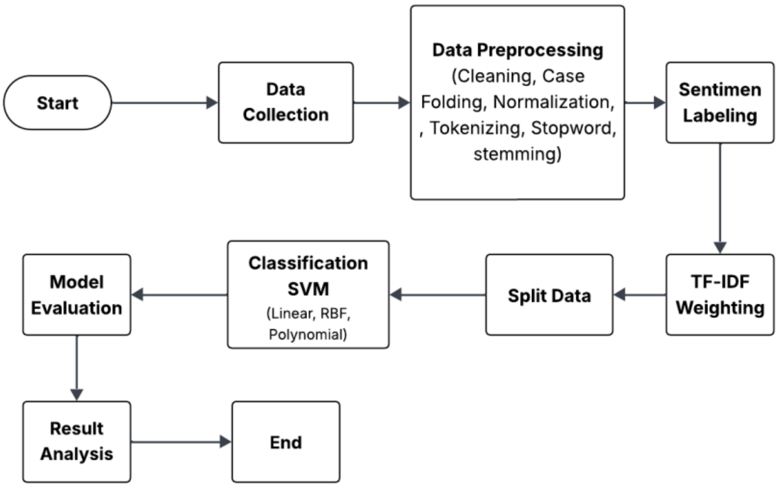

Figure 1. Research Stages

2.1 Data Collection

The dataset for this research was obtained from user feedback posted on the official Matahari application page in the Google Play environment. Review retrieval was performed using an automated Python extraction script, enabling the systematic collection of application comments. In total, 2,500 entries were retrieved and arranged chronologically, without applying any rating-based restrictions, to ensure that the collected data reflected recent user experiences in a balanced manner.

Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 2109−2117
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8811

2.2 Data Pre-processing

The raw user reviews must be cleaned and standardized. This pre-processing phase is crucial for enhancing the final model's reliability by eliminating textual noise, inconsistencies, and irrelevant data that could otherwise compromise the classification results for these steps [10] [11] [12]:

a. Cleaning*:* This stage focuses on removing elements that do not contribute to semantic meaning, such as punctuation, numerical symbols, emoticons, and hyperlinks. Eliminating these components ensures that only the linguistically relevant content of each review remains.
b. Case Folding: All characters were converted into lowercase to standardize word forms. This normalization step prevents the model from treating words with different case forms as separate tokens.
c. Normalization*:* A targeted normalization dictionary containing 55 high-frequency entries was used to convert informal expressions, abbreviations, and non-standard Indonesian terms into their formal equivalents. This list was curated based on the most frequent slang terms found in the corpus (e.g., "bgus" to "bagus"). Crucially, this stage also applied 'Negation Concatenation', where negation particles were joined with the following adjectives using underscores (e.g., 'tidak sesuai' becomes 'tidak_sesuai'). This technique prevents the loss of negative context during the stopword removal process.
d. Tokenizing**:** Each review was segmented into individual lexical units (tokens). Tokenization allows the model to process text at the word level and serves as the foundation for subsequent computational steps.
e. Stopword Removal**:** This stage excludes frequently occurring terms that contribute little to the meaning of the sentence. These high-frequency terms—often functional words—are filtered out so that the remaining content places greater emphasis on information-bearing vocabulary.
f. Stemming**:** Indonesian words were reduced to their morphological roots by removing affixes. Stemming harmonizes related word variations (such as verb inflexions and derivatives) so that the model treats them as a single conceptual unit.

2.3 Sentimen Labeling

Data labelling constitutes the process of assigning categorical tags to dataset entries [13]. High-quality annotation is fundamental for developing robust models, as it ensures predictions align with specific application requirements. In this study, sentiment categories were determined using a lexicon-driven approach. An Indonesian sentiment dictionary was used to map words to predefined polarity values [14], assigning positive terms a score of 1 and negative terms a score of –1. The overall polarity of each review was calculated by aggregating these values using Equation (1). Finally, the sentiment class was assigned based on the net score: reviews with a total value above zero were categorized as positive, those with a value below zero were identified as negative, and those with a total of exactly zero were labelled as neutral.

However, it is essential to note that a lexicon-driven approach to data labelling has certain limitations. Since sentiment classes are derived from predefined polarity values, the resulting labels may not fully capture contextual nuances, sarcasm, or implicit sentiment expressions found in user reviews. Consequently, the SVM model may tend to learn patterns inherent to the lexicon rather than deeper semantic representations of human sentiment. To mitigate this limitation, a small random sample of lexicon-labelled data was manually spot-checked by human annotators to ensure reasonable alignment with ground truth. Despite this limitation, the lexicon-based approach was adopted due to its practicality and efficiency in automating the annotation process. This limitation is acknowledged, and future research could incorporate additional manual annotation or expert validation to enhance the quality of the ground truth.

$$Sentiment\ (S^i) = \sum_{i=1}^{n} Sentiment(w_i) \tag{1}$$

2.4 Feature Weighting Using TF-IDF

After pre-processing, the refined text was converted into feature representations through the TF-IDF weighting scheme [15]. This approach assigns significance scores to terms by considering how frequently a word appears in an individual review and adjusting that value based on its frequency across the entire corpus [16].

a. Term Frequency (TF): Indicates how often a term appears within a particular document. Its value is determined using Equation (2).

$$tf(t,d) = \frac{number\ of\ words\ t\ in\ document\ d}{total\ words\ t\ in\ document\ d} \tag{2}$$

b. Inverse Document Frequency (IDF): IDF represents the degree to which a word is distributed across the entire corpus. As expressed in Equation (3), its value is calculated by taking the logarithm of the quotient between the total number of documents (N) and the number of documents containing that particular word (df(t)). A higher IDF value indicates that the word appears in fewer documents, making it more informative for distinguishing one text from another.

$$idf(t) = log\frac{N}{df(t)} \tag{3}$$

Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 2109−2117
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8811

c. TF-IDF: The combined TF-IDF value, computed through Equation (4), reflects how significant a term (t) is within a given document (d). This value is obtained by multiplying the term's TF(t,d) with its corresponding IDF(t). A larger TF-IDF score signifies that the term carries greater relevance in the context of that particular document [15, 16].

$$tfidf(t,d) = tf(t,d) * idf(t) \tag{4}$$

2.5 Classification Using Support Vector Machine (SVM)

The dataset was partitioned into two subsets: 80% for training and 20% for evaluation. For classification, this research uses an SVM-based approach [17]. The SVM framework determines a separating hyperplane that maximizes the margin between the classes. When the data distribution is not linearly separable, kernel functions transform the input features into a more discriminative, higher-dimensional representation [18]. Three kernel variants were examined to identify the configuration that produced the most effective results:

a. Linear Kernel, calculated using Equation (5).

$$K(x, x_k) = x_k^T x \tag{5}$$

b. Polynomial Kernel, calculated using Equation (6).

$$K(x, x_k) = (x_k^T x + 1)^d \tag{6}$$

c. Radial Basis Function (RBF) Kernel, calculated using Equation (7).

$$K(x, x_k) = exp\{- || x - x_k ||_2^2 /\} \sigma^2 \tag{7}$$

2.6 Model Evaluation

To measure the performance of each SVM variant, this study used a confusion matrix to assess how the model classifies instances across all sentiment categories during testing. From this matrix, several numerical evaluation indicators were computed to assess the classifier's overall predictive capability [19]. The metrics used include:

a. Accuracy: represents the percentage of predictions that the model correctly identifies for all sentiment categories [20]. Calculated using Equation (8).

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{8}$$

b. Precision: the proportion of reviews assigned a predicted label that are actually labelled correctly. This metric indicates the share of predictions labelled as positive that were actually correct according to the true class of the data [20]. The value is computed based on Equation (9).

$$Precision = \frac{TP}{TP+FP} \tag{9}$$

c. Recall: indicates the model's ability to correctly identify all data points that genuinely belong to a given class. Put simply, this metric shows how many of the actual positive instances were successfully recognized by the classifier during prediction [20]. The recall value is calculated using Equation (10).

$$Recall = \frac{TP}{TP+FN} \tag{10}$$

d. F1-Score: F1-Score is a performance metric that combines precision and recall to produce a balanced assessment of the classifier. It is particularly valuable in situations where both the accuracy of positive predictions and the model's ability to capture all relevant instances must be evaluated together [20]. The score is obtained using Equation (11).

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \tag{11}$$

3. RESULT AND DISCUSSION

3.1. Data Collection

The dataset for this research was obtained from user-generated reviews posted on the official Matahari mobile app page within the Google Play ecosystem. An automated scraping workflow developed using Python libraries was utilized to extract the data. A total of 2,500 of the most recent reviews were collected to ensure the analysis reflects up-to-date user perceptions and sentiment patterns.

3.2 Text Pre-processing Results

Pre-processing was conducted to convert raw text into a clean, structured form for analysis. The stages applied included cleansing, case folding, tokenizing, normalization, filtering, and stemming using an Indonesian dictionary. Table 1 shows the transformation of a review at each stage. The stopword removal process used a customized list that

Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 2109−2117
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8811

preserved key modifiers, such as "paling" (most), which act as sentiment intensifiers. Slang normalization standardized informal terms, e.g., "the bess" was converted to "terbaik" (best), thereby improving sentiment capture. This process removes irrelevant noise while retaining essential features for analysis.

Table 1. Text Pre-processing

Stage	Result
Cleaning	'cukup murah terjangkau harga hemat luar biasa toko matahari paling the bess' 'mending belanja offline setelah hari tidak ada kejelasan komplain via email akhirnya refund'
Case Folding	'cukup murah terjangkau harga hemat luar biasa toko matahari paling the bess' 'mending belanja offline setelah hari tidak ada kejelasan komplain via email akhirnya refund'
Normalisasi	'cukup murah terjangkau harga hemat keren toko matahari paling terbaik' 'mending belanja offline setelah hari tidak_ada kejelasan komplain via email akhirnya refund'
Tokenizing	'cukup', 'murah', 'terjangkau', 'harga', 'hemat', 'luar', 'biasa', 'toko' 'matahari' 'paling' 'terbaik' 'mending' 'belanja' 'offline' 'setelah' 'hari' 'tidak_ada' 'kejelasan' 'komplain' 'via' 'email' 'akhirnya' 'refund'
Stopword	'cukup', 'murah', 'terjangkau', 'harga' ,'hemat', 'keren', 'toko' 'matahari' 'paling', 'terbaik' 'mending' 'belanja' 'offline' 'hari' 'tidak_ada' 'kejelasan' 'komplain' 'via' 'email' 'akhirnya' 'refund'
Stemming	'cukup', 'murah', 'jangkau', 'harga' ,'hemat', 'keren', 'toko' 'matahari' 'paling', 'baik' 'mending' 'belanja' 'offline' 'telah' 'hari' 'tidak_ada' 'jelas' 'komplain' 'via' 'email' 'akhir' 'refund'

To further demonstrate the effectiveness of the proposed Semantic-based Normalization, we analyzed complex reviews containing implicit negation. For instance, one user remarked: 'mending belanja offline, setelah 7 hari tidak ada kejelasan. komplain via email, akhirnya refund' (better to shop offline, no clarity after 7 days, finally refunded). Standard methods might strip the word 'tidak' (no). However, our method successfully mapped the phrase 'tidak ada' into a single token 'tidak_ada'. By preserving this token, the system ensures that the negative polarity is retained for the subsequent labeling stage, preventing the loss of crucial sentiment information.

3.3 Sentiment Labelling

The sentiment tagging process in this study used an Indonesian lexical dictionary to generate a polarity score for each review. The score was determined by summing the sentiment values of all terms appearing in the text. Words expressing positive meaning were given a value of +1, while terms with negative connotations received −1. Reviews whose final total reached zero were categorized as neutral. An overview of the labelling outcome is presented in Table 2.

Table 2. Sentiment Labeling

Clean Text	Score	Sentiment
cukup murah jangkau harga hemat keren toko matahari paling baik	+4	Positive
guna huawei aplikasi matahari baru matahari baru aplikasi huawei	0	Neutral
mending belanja offline hari tidak_ada jelas komplain via email akhir refund	−3	Negative

3.4 TF-IDF Weighting

After the sentiment labels were assigned, the text data was converted to a numerical representation using a TF-IDF–based vectorization. The weighting produced by this method reflects the influence of each term within the corpus, as summarised in Table 3. Terms with higher weights, such as 'bagus' (0.54) and 'lemot' (0.31), indicate a stronger contribution to distinguishing sentiment classes. These weighted features serve as the input vector for the SVM classification process in the next stage.

Table 3. TF-IDF weighting

Word	Weight	Class
lemot	0.31	Negative
gagal	0.16	Negative
bagus	0.54	Positive
mantap	0.27	Positive
baik	0.21	Positive

3.5 Model Classification and Evaluation

Once the pre-processing, labelling, and feature extraction stages were completed, the next step was to build the classification model and assess its performance. To identify the most effective configuration, three SVM kernels—

Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 2109−2117
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8811

Linear, RBF, and Polynomial were tested and compared. The review dataset was split into two parts: a training set of 2,000 entries and a test set of 500 entries. Model performance was evaluated using several standard metrics, including accuracy, precision, recall, and F1-score. The comparative results obtained from these metrics are summarized in Table 4.

Table 4. Performance Comparison of SVM Kernels

Kernel SVM	Accuracy	Precision	Recall	F1-Score
Linear	88%	88%	88%	88%
RBF	83%	84%	83%	82%
Polynomial	65%	69%	65%	63%

Analysis of the performance summary in Table 4 indicates that the SVM model utilizing a linear kernel delivers the strongest results among all tested configurations, achieving an accuracy of 88%. Its performance clearly surpasses both the RBF and Polynomial variants, making the linear kernel the most appropriate choice for further interpretation. To illustrate how this model distinguishes among sentiment categories, a confusion matrix is provided in Figure 2, summarising the distribution of correct and misclassified predictions for the negative, neutral, and positive classes. As detailed in Figure 2, the confusion matrix demonstrates strong classification performance, with dominant diagonal values (True Positives): 191 for Negative, 150 for Neutral, and 101 for Positive. Nevertheless, some off-diagonal misclassifications are observed, particularly between the Neutral and Negative classes. This confusion stems from contextual ambiguity in user reviews and the inherent limitations of lexicon-based labelling, which may not fully capture implicit sentiment.

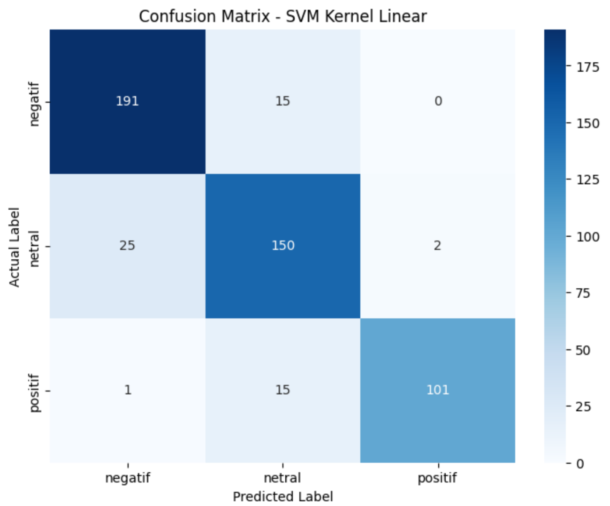

Figure 2. Confusion Matrix for Linear Kernel SVM Model

To provide a clearer overview of the model's performance, Table 5 presents a classification summary derived from the confusion matrix. The table provides evaluation scores for all sentiment categories, including metrics for predictive accuracy, retrieval ability, and the F1 score.

Table 5. *Classification Summary*

Class	Precision	Recall	F1-Score	Support
Negative	88%	93%	90%	206
Neutral	83%	85%	84%	177
Positive	98%	86%	92%	117
Accuracy			88%	500

The classification summary results show that the linear SVM achieves strong consistency in handling negative reviews. For this class, the model achieved 88% precision, indicating that most reviews predicted as negative were indeed negative. Its recall reached 93%, showing that nearly all actual negative reviews were successfully identified, while the F1-score of 90% reflects a balanced level of performance. Taken together, these findings demonstrate that the classifier is highly reliable when handling critical or unfavourable feedback.

For neutral reviews, the classifier produced 83% precision, 85% recall, and an F1-score of 84%. Although slightly lower than the metrics for the other sentiment categories, these numbers still indicate a reasonably stable ability to detect statements that clearly express neither positive nor negative sentiment. This behaviour is typical because neutral comments often contain ambiguous expressions that closely resemble both extremes.

The strongest performance is in the positive category, where the model achieved 98% precision, 86% recall, and an F1-score of 92%. These results indicate that favourable comments are identified with very high reliability and

Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 2109−2117
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8811

that the model rarely misclassifies genuine positive feedback. Overall, the linear SVM performs well across all sentiment classes, with positive reviews achieving the highest predictive accuracy.

3.6 Analysis of Key Factors Affecting User Sentiment

To understand why users give positive or negative reviews, an analysis was conducted of the keywords most influential on the model's predictions. By analyzing the coefficient weights of the linear SVM classifier, the most influential keywords for each sentiment category were identified, as illustrated in Figure 3.

(a) Negative Sentiment Keywords (b) Positive Sentiment Keywords

Figure 3. Word Cloud for Sentiment Analysis

Drawing from the keyword distribution illustrated in the word cloud (Figure 3) and further supported by the category patterns shown in the bar chart (Figure 4), user complaints can be organized into three dominant themes that strongly reflect their overall experience with the application.

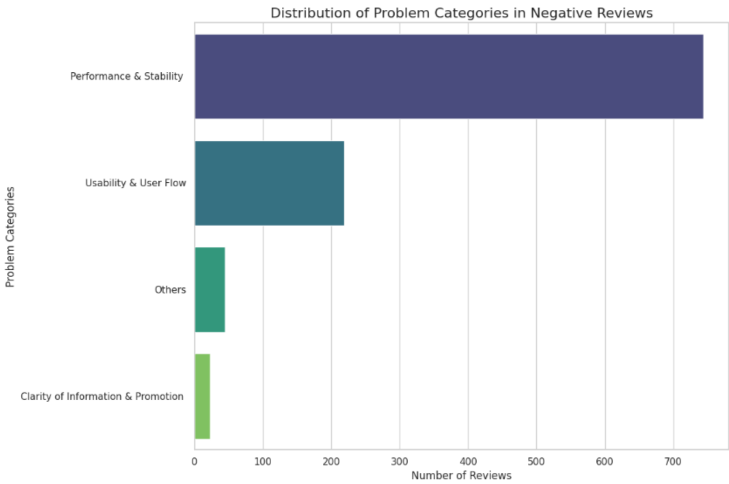

Figure 4. Distribution of Problem Categories in Negative Reviews

Based on the analysis of negative reviews, several main themes that most frequently appear from Matahari app users were found. The first theme is performance and stability, the most dominant category of complaints, with 744 reviews. Users consistently report various technical problems, such as "*lemot*" (slow), "*lama*" (long), errors, bugs, "*lelet*" (sluggish), and "*buruk*" (poor) applications. Complaints like this illustrate that many users experience performance issues with the app, especially when launching or when it suddenly closes, hindering their main activities and causing frustration.

The second prominent theme is usability and user flow, which relates to users' difficulties with the app or with completing a process. Complaints on this theme cover aspects of interface design (UI/UX) and user interaction flow. Several reviews highlight that the application feels complicated, difficult, and complex, indicating that navigation or the use of certain features is not yet intuitive. In addition, there are problems with the account login process, including keywords such as "*gagal*" (failed), "*masuk*" (login), and "*akun*" (account).

The next theme is other issues, which includes 44 negative reviews. Unlike the previous two categories, this theme contains random complaints that do not form a specific pattern. Analysis of the reviews in this category shows that there are overly general comments such as "*jelek*" (bad), "sampah" (rubbish), or "*gak jelas*" (unclear)," complaints about third-party services such as delivery couriers, and product *availability* issues such as "frequent stock shortages." Due to its diverse and unsystematic nature, this category serves as a container for minor complaints that cannot be classified into other themes.

Finally, there was the theme of information clarity and promotion, reflecting users' disappointment with the discrepancy between the app's information and their experience. Keywords that frequently appeared in this category

Building of Informatics, Technology and Science (BITS)
Volume 7, No 3, December 2025 Page: 2109−2117
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8811

included vouchers, discounts, prices, and promotions. Many users reported that vouchers could not be used, the prices displayed at checkout differed from the initial prices, or the advertised discounts did not match. This created a negative perception of the transparency and reliability of the promotion system within the app.

Overall, these four main themes show that user complaints about the Matahari app generally focus on technical performance and user experience, followed by issues related to the clarity of promotional information and several other minor issues. These findings indicate that major improvements should focus on increasing system stability and simplifying user flows to enhance user satisfaction and trust in the app.

On the other hand, positive sentiment keyword analysis provides insight into what users like. Words such as "*bagus*," "*mudah*," "*diskon*," "*mantap*," and "*murah*," are the most influential. This shows that when the app works well, users greatly appreciate the ease of transactions and, most importantly, attractive promotional offers and discounts. Praise is often brief, such as "Lots of discounts, easy to use," which suggests that, in users' eyes, the app's main selling point is the combination of convenience and financial benefits.

4. CONCLUSION

This study confirms that a text-mining approach using a linear kernel SVM can effectively analyze user sentiment toward the Matahari mobile application, achieving 88% accuracy on lexicon-labelled data. These findings validate the research objective of identifying dominant sentiment patterns and key factors influencing user experience. The analysis reveals that negative sentiment is primarily driven by issues related to application performance, usability, and information clarity, particularly regarding promotional offers and pricing discrepancies such as vouchers and discounts. While the results demonstrate the feasibility of an automated sentiment analysis system for Indonesian e-commerce platforms, future research is encouraged to employ larger, manually annotated datasets and to explore advanced transformer-based models, such as IndoBERT, to improve contextual understanding and classification reliability.

REFERENCES

[1] Rini Tresnasari, *Era Pemasaran 5.0 : Paradigma Baru Strategi Bisnis*. Padang: Takaza Innovatix Labs, 2024.
[2] Badan Pusat Statistik, *STATISTIK E-COMMERCE 2023*. Jakarta: Badan Pusat Statistik, 2025.
[3] M. Z. Siregar, A. M. Elhanafi, and D. Irwan, “Analisis Sentimen Terhadap Ulasan Aplikasi Media Sosial Di Google Play Menggunakan Algoritma Naive Bayes,” *Jurnal Mahasiswa Teknik Informatika*, vol. 9, no. 2, Apr. 2025, doi: 10.36040/jati.v9i2.12841.
[4] R. A. E. V. T. Sapanji, D. Hamdani, and P. Harahap, “Sentiment Analysis of the Top 5 E-commerce Platforms in Indonesia using Text Mining and Natural Language Processing (NLP),” *Journal of Applied Informatics and Computing (JAIC)*, vol. 7, no. 2, p. 202, 2023, doi: 10.30871/jaic.v7i2.6517.
[5] A. Nurian, M. S. Ma’arif, I. N. Amalia, and C. Rozikin, “Analisis Sentimen Pengguna Aplikasi Shopee Pada Situs Google Play Menggunakan Naive Bayes Classifier,” *Jurnal Informatika Dan Teknik Elektro Terapan*, vol. 12, no. 1, 2024, doi: 10.23960/jitet.v12i1.3631.
[6] P. A. A. Anam, D. A. Fatah, and M. A. Syakur, “Analisis Sentimen Review Aplikasi Zalora Di Google Playstore Menggunakan Metode Support Vector Machine,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 2, Mar. 2025, doi: 10.36040/jati.v9i2.13083.
[7] D. Fadila and M. Ikhsan, “Analisis Sentimen Pada Aplikasi Tokopedia Menggunakan Metode Support Vector Machine,” *Progresif: Jurnal Ilmiah Komputer*, vol. 21, no. 1, pp. 238–247, Feb. 2025, doi: 10.35889/progresif.v21i1.2593.
[8] M. Bagus Prayogi and G. Masitoh, “Analisis Sentimen Ulasan Pengguna Aplikasi Alfagift Menggunakan Random Forest,” *Jurnal Informatika Sunan Kalijaga)*, vol. 10, no. 2, pp. 158–170, 2025, doi: 10.14421/jiska.2025.10.2.158-170.
[9] A. Tholib and Z. Arifin, “Analisis Sentimen Terhadap Ulasan Aplikasi Shopee di Google Play Store Menggunakan Metode TF-IDF dan Long Short-Term Memory (LSTM),” *Journal homepage: Journal of Electrical Engineering and Computer (JEECOM)*, vol. 6, no. 2, 2024, doi: 10.33650/jeecom.v4i2.
[10] Wororomi, *Data Mining (Memahami Pola Di Balik Angka)*. Purbalingga: EUREKA MEDIA AKSARA, 2024.
[11] D. Rifaldi, Abdul Fadlil, and Herman, “Teknik Preprocessing Pada Text Mining Menggunakan Data Tweet ‘Mental Health,’” *Decode: Jurnal Pendidikan Teknologi Informasi*, vol. 3, no. 2, pp. 161–171, Apr. 2023, doi: 10.51454/decode.v3i2.131.
[12] R. R. Salam, M. F. Jamil, and Y. Ibrahim, “Analisis Sentimen Terhadap Bantuan Langsung Tunai (BLT) Bahan Bakar Minyak (BBM) Menggunakan Support Vector Machine,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, pp. 27–35, Apr. 2023, doi: 10.57152/malcom.v3i1.590.
[13] C. Sager, C. Janiesch, and P. Zschech, “A survey of image labelling for computer vision applications,” *Journal of Business Analytics*, vol. 4, no. 2, pp. 91–110, 2021, doi: 10.1080/2573234X.2021.1908861.
[14] R. Darman, “Analisis Sentimen Respons Twitter terhadap Persyaratan Badan Penyelenggara Jaminan Sosial (BPJS) di Kantor Pertanahan,” *Widya Bhumi*, vol. 3, no. 2, p. 113, Oct. 2023, doi: 10.31292/wb.v3i2.61.
[15] D. F. Surianto and D. F. Surianto, “Enhancing K-Means Clustering for Journal Articles using TF-IDF and LDA Feature Extraction,” *Brilliance: Research of Artificial Intelligence*, vol. 4, no. 2, pp. 964–972, Mar. 2025, doi: 10.47709/brilliance.v4i2.5547.
[16] V. B. Lestari and C. A. Hutagalung, “Evaluation of TF-IDF Extraction Techniques in Sentiment Analysis of Indonesian-Language Marketplaces Using SVM, Logistic Regression, and Naive Bayes,” *Journal of Computer Science and Applications*, vol. 8, no. 1, pp. 22–2025, 2025, doi: 10.21009/jkoma.v8i1.05.
[17] I. S. Aisah, B. Irawan, and T. Suprapti, “Algoritma Support Vector Machine (Svm) Untuk Analisis Sentimen Ulasan Aplikasi Al Qur’an Digital,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 6, 2023, doi: 10.36040/jati.v7i6.8263.

Volume 7, No 3, December 2025 Page: 2109−2117
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i3.8811

[18] J. A. P. Ginting, R. M. Sari, M. R. D. Siregar, and D. Kiswanto, "Analisis Support Vector Machine (SVM) Untuk Klasifikasi Jenis Kelamin Pada Ikan Cupang Dengan Bantuan Local Binary Pattern (LBP)," *JATI(Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 6, 2024, doi: 10.36040/jati.v8i6.12028.
[19] R. Nurhidayat and K. E. Dewi, "Penerapan Algoritma K-Nearest Neighbor Dan Fitur Ekstraksi N-Gram Dalam Analisis Sentimen Berbasis Aspek," *Komputa : Jurnal Ilmiah Komputer dan Informatika*, vol. 12, no. 1, May 2023, doi: 10.34010/komputa.v12i1.9458.
[20] S. Riyadi, F. D. Daffa, C. D. Damarjati, and M. S. A. M. A. Ali, "Sentiment Analysis on Social Media Uusing CNN-RNN Hybrid: A Case Study of Indonesian Presidential Candidate," *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, May 2025, doi: 10.22219/kinetik.v10i2.2125.