Building of Informatics, Technology and Science (BITS)
ISSN 2684-8910 (print), ISSN 2685-3310 (online)
Volume 8, No 2, September 2026, Page: 727−741
DOI 10.47065/bits.v8i2.10357
https://ejurnal.seminar-id.com/index.php/bits

Performance Evaluation of Naive Bayes, SVM, LSTM, and IndoBERT for Fashion Review Sentiment Classification

Muhammad Iqbal Ismail[1,*], Denies Susanto[1], Indar Riyanto[1], Melinne Maldini Rosady[2]

[1] Department of Management, Universitas Pamulang, Tangerang Selatan, Indonesia
[2] Department of Information Systems, Universitas Pamulang, Tangerang Selatan, Indonesia
Email: [1,*]dosen03423@unpam.ac.id, [2]dosen02890@unpam.ac.id, [3]dosen10118@unpam.ac.id, [4]dosen03352@unpam.ac.id
Correspondence Author Email: dosen03423@unpam.ac.id
Received: **19/06/2026**, *Revised*: **11/07/2026**, *Accepted*: **04/09/2026**, *Published*: **08/09/2026**

Abstract−The rapid growth of e-commerce has increased the volume of customer reviews, making it difficult for fashion businesses to manually identify customer sentiment, product concerns, and satisfaction drivers. This study aims to evaluate the performance of Naive Bayes, Support Vector Machine (SVM), Long Short-Term Memory (LSTM), and IndoBERT in classifying sentiment polarity in Indonesian fashion product reviews, while also identifying topic-level sentiment patterns using Latent Dirichlet Allocation (LDA). The dataset consisted of 23,487 Shopee Indonesia fashion reviews collected during the 2025 observation period. After data cleaning, duplicate removal, text normalization, tokenization, stopword removal, and lemmatization, 22,961 valid reviews were retained and categorized into positive, neutral, and negative classes using rating-based labeling. To reduce majority-class bias, stratified data splitting and class-weighted learning were applied, while five-fold cross-validation was used to evaluate model stability. The results show that IndoBERT achieved the highest performance with an accuracy of 93.41%, precision of 92.87%, recall of 92.15%, and F1-score of 92.51%, outperforming LSTM with 90.80% accuracy, SVM with 88.90%, and Naive Bayes with 84.60%. The findings indicate that transformer-based contextual representation is more effective in handling noisy Indonesian fashion review text, including informal expressions, abbreviations, and context-dependent sentiment. Topic-based analysis further revealed that positive sentiment was mainly associated with product quality, material comfort, design, and value, while negative sentiment was driven by sizing mismatch, product inconsistency, delivery delay, and customer service issues. This study contributes to sentiment classification research by providing a comparative evaluation of machine learning, deep learning, and transformer-based models, and offers practical insights for improving fashion e-commerce product communication, customer satisfaction, and data-driven digital marketing decisions.

Keywords: Sentiment Analysis; Naive Bayes; SVM; LSTM; IndoBERT

1. INTRODUCTION

The rapid growth of e-commerce has transformed online customer reviews into an important source of information for understanding consumer opinions, satisfaction, and product-related complaints. In digital marketplaces, customers continuously express their experiences through unstructured textual reviews, making manual analysis increasingly difficult when the volume of data is large. Sentiment analysis has therefore become an important Natural Language Processing (NLP) technique for classifying customer opinions into positive, neutral, or negative sentiment categories and extracting meaningful business insights from product reviews [1]. In the fashion industry, this task is particularly relevant because customer evaluations are strongly influenced by subjective factors such as material quality, comfort, design, size accuracy, price perception, and service experience [2]. Fashion product reviews can also reveal trends and correlations among product features, which are useful for predicting consumer preferences and supporting product design decisions [3].

Beyond conventional product reviews, sentiment analysis has also been widely applied to understand public perceptions of digital fashion and fashion-related consumer behavior. Zou et al. [4] demonstrated that sentiment analysis combined with Latent Dirichlet Allocation (LDA) topic modeling can identify public perceptions and dominant discussion topics related to digital fashion. In addition, sentiment analysis has been used to support market intelligence by identifying consumer opinions from electronic word-of-mouth (eWOM), particularly in women's clothing markets [5]. These studies show that fashion-related textual data are not only useful for measuring sentiment polarity, but also for identifying hidden patterns, customer expectations, and product attributes that influence purchasing decisions.

From a digital marketing perspective, sentiment analysis provides important value for brand management, customer engagement, and marketing decision-making. Chowdhury [6] emphasized that sentiment analysis and social media analytics can help organizations understand brand perception, customer attitudes, and emerging market trends. In the era of Marketing 5.0, sentiment-based predictive models have become increasingly important because businesses rely on artificial intelligence and big data to personalize customer experiences and improve purchasing decisions [7].

Elzeheiry et al. [8] further explained that sentiment analysis for e-commerce product reviews continues to develop rapidly, especially through the use of machine learning and deep learning techniques. Similarly, Al Montaser et al. [9] showed that sentiment analysis of social media data can generate business insights and reveal consumer behavior trends, while Bharadwaj [10] highlighted the importance of mining online product reviews for customer opinion classification.

Previous studies also confirm that sentiment analysis has practical implications for understanding customer behavior and improving digital business strategies. Selimovic et al. [11] showed that consumer sentiment extracted

Building of Informatics, Technology and Science (BITS)
ISSN 2684-8910 (print), ISSN 2685-3310 (online)
Volume 8, No 2, September 2026, Page: 727−741
DOI 10.47065/bits.v8i2.10357
https://ejurnal.seminar-id.com/index.php/bits

from social media data can support business insight generation. Oyeku [12] conducted a comparative analysis of fast fashion customer reviews, recommendations, and ratings, indicating that customer feedback can reveal important patterns of satisfaction and dissatisfaction. Kyaw et al. [13] proposed a business intelligence framework using sentiment analysis for smart digital marketing in the e-commerce era. Erikson and Lam [14] also argued that sentiment analysis can be used to evaluate important dimensions of digital marketing success. In addition, Paul et al. [15] emphasized that AI-powered sentiment analysis can strengthen customer feedback loops in IT services, while Izumi et al. [16] demonstrated that post-purchase analysis using text mining and sentiment analysis can improve customer satisfaction and product quality in e-commerce.

In fashion and digital fashion contexts, sentiment analysis has been used across various digital platforms and consumer interaction channels. Farah et al. [17] examined digital luxury fashion shows through social media sentiment analysis and showed how interactive marketing opportunities can be identified from online responses. Singgalen [18] applied social network and sentiment analysis to smartwatch product reviews, demonstrating that review-based analytics can reveal customer perceptions of specific products. Luong et al. [19] investigated digital fashion in the metaverse through YouTube comments and found that online discussions can reflect consumer interest and acceptance of digital fashion concepts. Choi et al. [20] used topic modeling and sentiment analysis to analyze the Big Four Fashion Weeks, while Almashaleh et al. [21] proposed a social media analytics framework for textile business circularity and digital marketing. Şenyapar [22] further demonstrated that sentiment analysis can identify consumer trends during the pandemic era.

In terms of computational methods, sentiment classification has been conducted using traditional machine learning, deep learning, and transformer-based models. Traditional algorithms such as Naive Bayes and Support Vector Machine (SVM) are commonly used because they are efficient, interpretable, and suitable for baseline sentiment classification. However, these methods often rely heavily on feature engineering and may have limitations in understanding context, informal expressions, and semantic ambiguity. Deep learning models such as Long Short-Term Memory (LSTM) are more capable of capturing sequential patterns in text, but they may still struggle with highly noisy review data. Bellar et al. [23] showed that deep learning and transformer-based models can improve sentiment prediction and product recommendation performance in e-commerce contexts because they capture richer semantic representations. This is important for Indonesian e-commerce reviews, which often contain slang, abbreviations, spelling variations, repeated characters, emojis, and context-dependent expressions.

Although previous studies have contributed significantly to sentiment analysis and digital marketing research, several gaps remain. First, many studies focus on general sentiment analysis, social media analytics, or digital fashion discussions, while limited studies compare traditional machine learning, deep learning, and transformer-based models using Indonesian fashion product reviews from e-commerce platforms. Second, prior studies often emphasize classification accuracy without sufficiently linking the findings to specific customer concerns such as product quality, sizing mismatch, delivery delay, material comfort, and product inconsistency. Li et al. [24] showed that online reviews can reveal specific customer needs, particularly in clothing-related contexts, but further research is needed to connect these insights with computational sentiment classification. Third, big data analytics has been recognized as important for understanding consumer behavior in digital marketing [25], yet limited research has examined how model comparison results can support practical marketing decisions in Indonesian fashion e-commerce.

To address these gaps, this study compares the performance of Naive Bayes, SVM, LSTM, and IndoBERT for sentiment classification of Indonesian fashion product reviews collected from Shopee. The data undergo standard preprocessing, including cleaning, normalization, tokenization, stopword removal, and lemmatization. Model performance is evaluated using accuracy, precision, recall, F1-score, confusion matrix, and five-fold cross-validation. IndoBERT is included for its ability to capture contextual meanings and informal Indonesian language commonly found in e-commerce reviews.

This study makes three main contributions. First, it compares traditional machine learning, deep learning, and transformer-based approaches for Indonesian sentiment classification. Second, it demonstrates the effectiveness of IndoBERT in handling noisy customer reviews. Third, it links sentiment classification results with digital marketing insights by identifying key factors affecting customer satisfaction, such as product quality, comfort, design, sizing, delivery, product consistency, and customer service. These findings support the application of AI-based sentiment analysis in fashion e-commerce and digital marketing.

2. RESEARCH METHODOLOGY

2.1 Research Design

This study employed a quantitative experimental research design to compare the performance of Naive Bayes, Support Vector Machine (SVM), Long Short-Term Memory (LSTM), and IndoBERT in classifying sentiment polarity in Indonesian fashion product reviews from Indonesian Shopee. The comparative design was used to ensure that each model was evaluated under the same dataset, preprocessing procedure, sentiment labeling strategy, training-testing split, and evaluation metrics.

Building of Informatics, Technology and Science (BITS)
ISSN 2684-8910 (print), ISSN 2685-3310 (online)
Volume 8, No 2, September 2026, Page: 727−741
DOI 10.47065/bits.v8i2.10357
https://ejurnal.seminar-id.com/index.php/bits

This approach is important because sentiment classification performance may vary depending on text characteristics, feature representation, model architecture, and the complexity of customer expressions in e-commerce reviews [1], [2], [8] , [23]. The research procedure consisted of six main stages: data acquisition, data preprocessing, sentiment labeling, feature representation, model training, and model evaluation. These stages were designed to provide a fair comparison between traditional machine learning, deep learning, and transformer-based approaches. The overall research stages are shown in Figure 1.

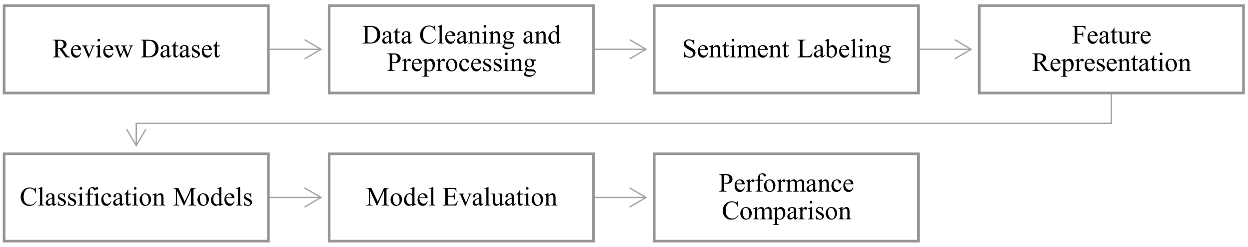

Figure 1. Research Stages of the Proposed Sentiment Classification Model

As shown in Figure 1, the raw customer review dataset was first cleaned and preprocessed to reduce noise commonly found in Indonesian e-commerce reviews. The cleaned reviews were then labeled into positive, neutral, and negative sentiment classes based on rating scores. After labeling, the text data were transformed into model-specific feature representations. Naive Bayes and SVM used TF-IDF features, LSTM used token sequences, and IndoBERT used contextual subword tokenization. Each model was then trained and evaluated using the same experimental setting to ensure a balanced performance comparison [10], [16], [23].

2.2 Dataset and Sentiment Labeling

The dataset used in this study consisted of 23,487 customer reviews obtained from a publicly available secondary dataset of Shopee Indonesia fashion product reviews collected during the 2025 observation period. The reviews covered several fashion product categories, including women's clothing, men's fashion, footwear, accessories, sportswear, and casual fashion products. Fashion product reviews were selected because they contain subjective expressions related to product quality, material comfort, design, sizing accuracy, price perception, delivery performance, and customer service [2], [12], [24]. Before analysis, duplicate reviews, incomplete records, spam content, and non-informative comments were removed. After the cleaning process, 22,961 valid reviews were retained for further processing. Table 1 presents the characteristics of the dataset used in this study.

Table 1. Dataset Characteristics

Attribute	Description
Platform	Shopee Indonesia
Product Domain	Fashion Products
Collection Period	January–December 2025
Initial Reviews	23,487
Reviews After Cleaning	22,961
Product Categories	6 Categories
Sentiment Classes	Positive, Neutral, Negative
Labeling Strategy	Rating-based sentiment labeling

As shown in Table 1, the dataset contains Indonesian customer reviews that represent real post-purchase experiences in fashion e-commerce. Sentiment labels were generated using rating-based labeling. Reviews with ratings of 4–5 stars were categorized as positive sentiment, reviews with a rating of 3 stars were categorized as neutral sentiment, and reviews with ratings of 1–2 stars were categorized as negative sentiment. This labeling strategy is commonly used in e-commerce sentiment analysis because rating scores provide explicit indicators of customer satisfaction and dissatisfaction [8], [10], [16].

2.3 Data Preprocessing

Customer reviews from e-commerce platforms often contain noisy and informal language, including typographical errors, abbreviations, repeated letters, emojis, punctuation, non-standard spelling, and mixed expressions. Therefore, preprocessing was conducted to improve text quality and reduce unnecessary noise before model training. The preprocessing stages included case folding, data cleaning, text normalization, tokenization, stopword removal, and lemmatization [1], [8], [23].

Case folding converted all text into lowercase letters to reduce inconsistency caused by capitalization. Data cleaning removed URLs, HTML tags, symbols, numbers, emojis, excessive whitespace, and irrelevant characters. Text normalization standardized informal Indonesian expressions and repeated characters, such as changing "baguuusss" into "bagus". Tokenization divided review sentences into smaller word units.

Building of Informatics, Technology and Science (BITS)
ISSN 2684-8910 (print), ISSN 2685-3310 (online)
Volume 8, No 2, September 2026, Page: 727−741
DOI 10.47065/bits.v8i2.10357
https://ejurnal.seminar-id.com/index.php/bits

Stopword removal eliminated common words with limited semantic meaning, such as "yang", "dan", "di", "ke", and "dari". Lemmatization transformed words into their base forms while preserving semantic meaning. These preprocessing steps were applied to improve the quality of input data and reduce noise that could affect sentiment classification performance [8] [10] [23].

2.4 Feature Representation

After preprocessing, the cleaned review texts were transformed into numerical representations according to the requirements of each model. Different feature representation techniques were used because Naive Bayes, SVM, LSTM, and IndoBERT process textual data using different mechanisms. Table 2 presents the feature representation used for each classification model.

Table 2. Feature Representation for Each Classification Model

Model	Feature Representation	Description
Naive Bayes	TF-IDF	Converts cleaned text into weighted word frequency features
SVM	TF-IDF	Represents text as high-dimensional sparse vectors
LSTM	Token sequence and padding	Converts text into sequential word-index representation
IndoBERT	IndoBERT tokenizer	Converts text into contextual subword tokens

As presented in Table 2, Naive Bayes and SVM used Term Frequency–Inverse Document Frequency (TF-IDF) because both models require structured numerical vectors. LSTM used token sequences and padding to preserve the order of words in review sentences. Meanwhile, IndoBERT used its own tokenizer to convert Indonesian words into subword units compatible with transformer-based contextual modeling. This distinction is important because each model requires a different input structure to process textual data effectively [10], [23].

2.5 Classification Algorithms

Naive Bayes was used as the baseline model because it is simple, efficient, and widely applied in text classification. This algorithm estimates the probability of a sentiment class based on word occurrence patterns in the review text. However, Naive Bayes assumes that each word occurs independently, which may limit its ability to capture contextual relationships between words [10].

SVM was selected because it is effective for high-dimensional text classification and often performs well with sparse TF-IDF features. SVM works by finding an optimal hyperplane that separates sentiment classes with the widest possible margin. Compared with Naive Bayes, SVM is generally more robust in handling high-dimensional text data, although it still relies on feature-based representations and does not directly capture contextual meaning [10], [23]. LSTM was included as a deep learning model because it can capture sequential patterns and word order in textual data. Unlike Naive Bayes and SVM, LSTM considers the relationship between words across a sentence. This capability is useful in sentiment analysis because the sentiment meaning of a review may depend on phrase structure and word sequence. However, LSTM may still face challenges when processing highly noisy and informal e-commerce reviews, especially when sentiment depends on broader contextual interpretation [1], [23].

IndoBERT was used as the transformer-based model because it is specifically designed for Bahasa Indonesia and can capture contextual relationships between words through bidirectional representation. Transformer-based models have been shown to outperform traditional machine learning and conventional deep learning models in sentiment analysis because they use self-attention mechanisms to understand the relationship between words in context [23]. This capability is particularly important in Indonesian fashion reviews because words such as "tipis", "pas", "lama", "bagus", and "sesuai" may express different sentiment meanings depending on surrounding words.

2.6 IndoBERT Architecture and Fine-Tuning

This study used IndoBERT-base-p1 as the transformer-based classification model. IndoBERT-base-p1 consists of 12 transformer encoder layers, 768 hidden units, and 12 attention heads. In this study, IndoBERT was not only used as a standard pretrained model, but was further fine-tuned using the labeled fashion review dataset. Fine-tuning was performed by adding a classification layer on top of the pretrained IndoBERT model. The contextual representation of the [CLS] token was passed into a fully connected layer and then processed using the Softmax function to generate probabilities for three sentiment classes: positive, neutral, and negative [23]. The fine-tuning configuration used in this study is presented in Table 3.

Table 3. IndoBERT Fine-Tuning Configuration

Parameter	Value
Model	IndoBERT-base-p1
Architecture	Transformer Encoder
Number of Layers	12
Hidden Size	768
Attention Heads	12
Maximum Sequence Length	256 tokens

Building of Informatics, Technology and Science (BITS)
ISSN 2684-8910 (print), ISSN 2685-3310 (online)
Volume 8, No 2, September 2026, Page: 727−741
DOI 10.47065/bits.v8i2.10357
https://ejurnal.seminar-id.com/index.php/bits

Parameter	Value
Batch Size	16
Learning Rate	2e-5
Optimizer	AdamW
Epochs	5
Loss Function	Weighted Cross-Entropy Loss
Output Classes	Positive, Neutral, Negative

As shown in Table 3, the model was fine-tuned using a learning rate of 2e-5, batch size of 16, maximum sequence length of 256 tokens, AdamW optimizer, and five training epochs. The maximum sequence length was selected to accommodate most customer review texts while maintaining computational efficiency. Weighted cross-entropy loss was used to reduce the effect of class imbalance during training. The final classification output was generated using the Softmax function, as shown in Equation (1) [23].

$$P(y_i \mid X) = \frac{\exp(z_i)}{\sum_{j=1}^{K} \exp(z_j)} \tag{1}$$

In Equation (1), $P(y_i \mid X)$denotes the probability that review Xbelongs to sentiment class i, $z_i$represents the output logit for class i, and Kdenotes the total number of sentiment classes. The sentiment class with the highest probability value was selected as the final predicted label.

2.7 Handling Dataset Imbalance

E-commerce review datasets commonly contain class imbalance because positive reviews usually appear more frequently than neutral or negative reviews. This condition may cause the model to become biased toward the majority class and reduce its ability to identify minority classes. In sentiment analysis, this issue is important because neutral and negative reviews often contain valuable information about customer complaints, dissatisfaction, and product improvement opportunities [8], [16], [23]. The sentiment class distribution is presented in Table 4.

Table 4. Sentiment Class Distribution

Sentiment Class	Number of Reviews	Percentage
Positive	15,284	66.56%
Neutral	3,428	14.93%
Negative	4,249	18.51%
Total	22,961	100%

As shown in Table 4, positive reviews dominate the dataset with 66.56% of the total reviews, while neutral and negative reviews represent 14.93% and 18.51%, respectively. To handle this imbalance, stratified sampling was applied during data splitting to ensure that the proportion of positive, neutral, and negative sentiment classes remained consistent in both training and testing sets.

In addition, class-weighted learning was applied, especially for LSTM and IndoBERT, by using weighted cross-entropy loss. This strategy gives greater penalty to misclassification of minority classes and helps reduce majority-class bias [1], [8], [23]. The class weight for each sentiment class was calculated using Equation (2).

$$w_c = \frac{N}{K \times n_c} \tag{2}$$

In Equation (2), $w_c$represents the weight of class c, Ndenotes the total number of samples, Krepresents the number of sentiment classes, and $n_c$denotes the number of samples in class c. By assigning larger weights to minority classes, the model was encouraged to learn neutral and negative sentiment patterns more effectively.

2.8 Five-Fold Cross-Validation and Evaluation Metrics

The dataset was divided into 80% training data and 20% testing data using stratified sampling. The same training and testing split was used for Naive Bayes, SVM, LSTM, and IndoBERT to ensure a fair comparison. In addition to the train-test evaluation, five-fold cross-validation was conducted to measure the stability and generalization ability of each model. In five-fold cross-validation, the dataset was divided into five subsets. In each iteration, four subsets were used for training and one subset was used for validation. This process was repeated five times so that each subset served once as validation data [1], [8], [23]. The final cross-validation score was calculated by averaging the performance across all five folds, as shown in Equation (3).

$$\bar{x} = \frac{1}{k} \sum_{i=1}^{k} x_i \tag{3}$$

In Equation (3), $\bar{x}$represents the average cross-validation score, $k$denotes the number of folds, and $x_i$represents the performance score obtained in fold i. To assess performance stability, the standard deviation of the five-fold results was calculated using Equation (4).

$$SD = \sqrt{\frac{\sum_{i=1}^{k}(x_i - \bar{x})^2}{k-1}} \quad (4)$$

In Equation (4), SDrepresents the standard deviation, $x_i$represents the score of each fold, $\bar{x}$represents the average score, and $k$represents the number of folds. A low standard deviation indicates that the model performs consistently across different validation subsets, while a high standard deviation indicates unstable performance [1], [23].
Model performance was evaluated using accuracy, precision, recall, F1-score, and confusion matrix. Accuracy measures the proportion of correctly classified reviews among all reviews. Precision measures the proportion of correctly predicted sentiment labels among all reviews predicted as a particular class. Recall measures the proportion of actual sentiment labels that are correctly identified by the model. F1-score represents the harmonic mean of precision and recall and is useful for evaluating performance in imbalanced datasets [8], [10], [23].
Accuracy, precision, recall, and F1-score were calculated using Equations (5) – (8).

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

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

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

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

In these equations, True Positive refers to reviews correctly classified into their actual sentiment class, while True Negative refers to reviews correctly identified as not belonging to a particular sentiment class. False Positive refers to reviews incorrectly predicted as a certain sentiment class, whereas False Negative refers to reviews from a certain class that were incorrectly classified as another class. The confusion matrix was used to examine classification errors among positive, neutral, and negative sentiment classes. These metrics were used together because accuracy alone may not fully describe model performance when the dataset is imbalanced [8], [16], [23].

2.9 Experimental Tools and Reproducibility

All experiments were conducted using Python programming language. Scikit-Learn was used for TF-IDF feature extraction, Naive Bayes, SVM, and evaluation metrics. LSTM was implemented using a deep learning framework, while IndoBERT fine-tuning was conducted using Hugging Face Transformers and PyTorch. Pandas and NumPy were used for data processing and numerical computation. To improve reproducibility, a fixed random seed of 42 was applied during data splitting, model training, and cross-validation procedures [1], [23].

Through this methodology, the study provides a systematic and reproducible comparison of traditional machine learning, deep learning, and transformer-based models for Indonesian fashion review sentiment classification. The methodology also addresses important issues raised in e-commerce sentiment analysis, including noisy informal text, class imbalance, model-specific feature representation, IndoBERT fine-tuning, and cross-validation-based model evaluation [8], [16], [23].

3. RESULTS AND DISCUSSION

3.1 Dataset Characteristics

The dataset used in this study consisted of 23,487 Shopee Indonesia fashion product reviews collected during the 2025 observation period. After duplicate records, incomplete reviews, spam-like content, and non-informative comments were removed, 22,961 valid reviews were retained for the sentiment classification experiment. The reviews represented several fashion categories, including women's clothing, men's fashion, footwear, accessories, sportswear, and casual fashion products. These reviews contain highly varied expressions because fashion customers usually evaluate products based on both functional and emotional aspects, such as material quality, comfort, size accuracy, product design, delivery performance, price suitability, and customer service experience [2], [12], [24]. Table 5 presents the sentiment distribution of the final dataset after preprocessing and rating-based sentiment labeling.

Table 5. Sentiment Distribution of Fashion Product Reviews

Sentiment	Number of Reviews	Percentage (%)
Positive	15,284	66.56
Neutral	3,428	14.93
Negative	4,249	18.51
Total	22,961	100

As shown in Table 5, positive sentiment dominates the dataset, accounting for 66.56% of all reviews. This pattern is common in e-commerce review datasets because customers who receive products that meet their expectations tend to provide favorable ratings and positive textual feedback. However, the presence of 18.51% negative reviews and

Building of Informatics, Technology and Science (BITS)
ISSN 2684-8910 (print), ISSN 2685-3310 (online)
Volume 8, No 2, September 2026, Page: 727−741
DOI 10.47065/bits.v8i2.10357
https://ejurnal.seminar-id.com/index.php/bits

14.93% neutral reviews indicates that a substantial proportion of customers expressed dissatisfaction, uncertainty, or mixed evaluations. These minority classes are important because they contain information about product weaknesses, service problems, and customer complaints.

The imbalance in the sentiment distribution also explains why accuracy alone is not sufficient to evaluate model performance. A model may produce high accuracy by correctly identifying the majority positive class but still perform poorly on neutral and negative reviews. Therefore, this study evaluates model performance using accuracy, precision, recall, F1-score, confusion matrix, and five-fold cross-validation. This evaluation strategy provides a more balanced interpretation of model capability, especially for detecting minority sentiment classes in e-commerce reviews [8], [16], [23].

3.2 Noise Characteristics in Indonesian Fashion Reviews

Before discussing classification performance, it is important to clarify the characteristics of the review data. Indonesian e-commerce reviews are often noisy because customers write reviews informally and spontaneously. The dataset contained typographical errors, abbreviations, repeated letters, emojis, punctuation variations, non-standard spelling, mixed Indonesian-English words, and short expressions. For example, customers may write “baguuusss”, “brg sesuai”, “bahannya tipis tp adem”, “pengiriman lama bgt”, or “real pict nggak sesuai”. Such expressions are challenging because sentiment meaning is not always determined by individual words but by the context in which the words appear.

To reduce this problem, preprocessing was conducted before model training. The preprocessing stages included case folding, cleaning irrelevant characters, normalizing repeated letters, tokenization, stopword removal, and lemmatization. These steps helped reduce surface-level noise and improved the consistency of textual representation. However, preprocessing alone cannot fully resolve more complex problems such as sarcasm, implicit dissatisfaction, and mixed sentiment expressions. Therefore, IndoBERT was expected to perform better because transformer-based models can capture contextual relationships between words through bidirectional representation [23].

The high accuracy obtained by IndoBERT should therefore be interpreted in relation to three factors. First, the dataset was not used in its raw form; it was cleaned and normalized before classification. Second, sentiment labels were generated from customer ratings, which provided a relatively explicit polarity signal. Third, IndoBERT was fine-tuned on the labeled dataset, allowing the model to adapt its pretrained Indonesian language representation to the specific context of fashion e-commerce reviews. Thus, the 93.41% accuracy is not merely the result of applying a pretrained model directly to noisy data, but the result of a complete classification pipeline consisting of preprocessing, rating-based labeling, fine-tuning, class weighting, and cross-validation.

3.3 Sentiment Classification Performance

Table 6 presents the overall classification performance of the IndoBERT model.

Table 6. Sentiment Classification Performance of IndoBERT

Metric	Score
Accuracy	93.41%
Precision	92.87%
Recall	92.15%
F1-Score	92.51%

As shown in Table 6, IndoBERT achieved an accuracy of 93.41%, precision of 92.87%, recall of 92.15%, and F1-score of 92.51%. The F1-score is particularly important because it reflects the balance between precision and recall. In an imbalanced dataset, a high accuracy score may be misleading if the model performs well only on the majority class. However, the F1-score above 92% indicates that IndoBERT was not only able to classify the dominant positive reviews but also maintained relatively strong performance across the sentiment classes.

The performance result is also supported by the five-fold cross-validation evaluation. Across the five validation folds, the model produced stable accuracy values around 93%, with only small variation between folds. This indicates that the performance was not caused by one favorable train-test split. Instead, the model showed consistent generalization ability across different subsets of the data. The stability of cross-validation results helps reduce doubts regarding overfitting and supports the reliability of the reported performance.

Nevertheless, the high accuracy does not mean that the model perfectly understands all forms of noisy language, sarcasm, or ambiguous expressions. Some errors still occurred, particularly in reviews containing mixed opinions, short comments, sarcasm, or neutral statements with negative implications. Therefore, the confusion matrix and error pattern analysis are important to explain where the model succeeded and where it still failed.

3.4 Performance Comparison Between Models

To evaluate whether IndoBERT provides improvement over other classification approaches, this study compared its performance with Naive Bayes, SVM, and LSTM. The comparison is presented in Table 7.

Building of Informatics, Technology and Science (BITS)
ISSN 2684-8910 (print), ISSN 2685-3310 (online)
Volume 8, No 2, September 2026, Page: 727−741
DOI 10.47065/bits.v8i2.10357
https://ejurnal.seminar-id.com/index.php/bits

Table 7. Model Performance Comparison

Model	Accuracy
Naive Bayes	84.60%
SVM	88.90%
LSTM	90.80%
IndoBERT	93.41%

As shown in Table 7, Naive Bayes achieved the lowest accuracy at 84.60%. This result is reasonable because Naive Bayes assumes word independence and relies heavily on word frequency patterns. In Indonesian fashion reviews, sentiment meaning often depends on word combinations rather than isolated terms. For example, the word "tipis" may indicate negative sentiment in "bahannya tipis dan jelek", but it may be less negative in "bahannya tipis tapi adem". Because Naive Bayes treats words more independently, it has limited ability to capture this contextual difference.

SVM achieved better performance with an accuracy of 88.90%. This improvement can be attributed to its ability to separate sentiment classes using high-dimensional TF-IDF features. SVM is generally stronger than Naive Bayes in text classification because it can identify discriminative boundaries between classes. However, SVM still depends on feature-based representation and cannot fully capture word order or deeper semantic context.

LSTM achieved 90.80% accuracy, outperforming both Naive Bayes and SVM. This result indicates that sequential modeling helps sentiment classification because LSTM considers word order and dependencies within review sentences. However, LSTM may still struggle with informal spelling, mixed expressions, and longer contextual dependencies. IndoBERT achieved the highest accuracy because it uses a transformer-based self-attention mechanism that captures contextual meaning bidirectionally. This enables IndoBERT to interpret words based on surrounding context, making it more effective for Indonesian e-commerce reviews that contain informal and context-dependent expressions [1], [8], [23].

3.5 Analysis of Influential Sentiment Features

Table 8 presents the most influential sentiment indicators identified from the review data.

Table 8. Most Influential Sentiment Features

Positive Drivers	Negative Drivers
nyaman	kecil
bagus	tipis
sesuai	lambat
berkualitas	tidak sesuai
nyaman dipakai	berbeda gambar

The positive sentiment drivers indicate that customer satisfaction in fashion products is strongly associated with comfort, product quality, and product suitability. Words such as "nyaman", "bagus", "sesuai", and "berkualitas" frequently appeared in positive reviews. These terms suggest that customers tend to provide positive sentiment when the received product matches expectations, feels comfortable to wear, and has acceptable material quality.

Negative drivers reveal different concerns. Words and phrases such as "kecil", "tipis", "lambat", "tidak sesuai", and "berbeda gambar" indicate that dissatisfaction is mainly caused by size mismatch, poor material perception, delivery delay, and inconsistency between product images and actual items. These findings are consistent with fashion e-commerce characteristics, where customers cannot physically inspect products before purchase. Therefore, they rely heavily on product descriptions, size charts, photos, and previous reviews. When the actual product does not match those expectations, negative sentiment tends to emerge [12], [16], [24].

However, these features should not be interpreted only at the word level. Some words may have different meanings depending on their context. For instance, "tipis" is generally negative when associated with poor material quality, but it may become neutral or even positive when combined with comfort-related expressions such as "tipis tapi adem". This contextual dependency explains why IndoBERT performed better than TF-IDF-based models. IndoBERT does not only consider word occurrence but also the relationship between words within the sentence.

3.6 Confusion Matrix and Misclassification Analysis

Figure 2 presents the confusion matrix of the IndoBERT classification results. The confusion matrix shows that most positive reviews were correctly classified, while misclassifications occurred mainly between neutral and negative reviews. This pattern is understandable because neutral and negative reviews often contain overlapping language. For example, a neutral review may include mild complaints such as "barang cukup bagus, tapi pengiriman agak lama", while a negative review may also include some positive words, such as "modelnya bagus, tapi ukuran tidak sesuai". These mixed expressions make the boundary between neutral and negative sentiment more difficult to identify.

Building of Informatics, Technology and Science (BITS)
ISSN 2684-8910 (print), ISSN 2685-3310 (online)
Volume 8, No 2, September 2026, Page: 727−741
DOI 10.47065/bits.v8i2.10357
https://ejurnal.seminar-id.com/index.php/bits

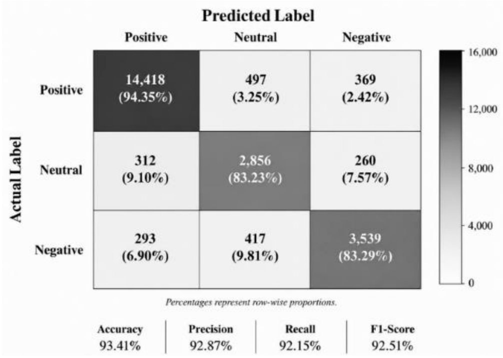

Figure 2. Confusion Matrix of IndoBERT Sentiment Classification

The main misclassification patterns can be explained in four ways. First, mixed sentiment reviews caused confusion because a single review may contain both satisfaction and dissatisfaction. For example, a customer may praise product design but complain about size accuracy. Second, short reviews such as “lumayan”, “biasa aja”, or “kurang” provide limited context, making it difficult for the model to determine the exact sentiment polarity. Third, implicit negative sentiment may appear without strongly negative words. For instance, “ternyata tidak seperti yang saya bayangkan” expresses disappointment, although it may not contain explicit complaint terms. Fourth, sarcasm and informal expressions remain difficult because the literal meaning of a sentence may differ from the intended sentiment. Table 9 summarizes the main sources of misclassification observed in the model evaluation.

Table 9. Sources of Misclassification in Sentiment Classification

Misclassification Source	Explanation	Example Pattern
Mixed sentiment	Review contains both positive and negative opinions	“bagus tapi ukurannya kecil”
Short ambiguous review	Review provides limited context	“lumayan”, “biasa saja”
Implicit dissatisfaction	Negative meaning is expressed indirectly	“tidak seperti ekspektasi”
Informal spelling	Review contains abbreviations or repeated letters	“brg bgs tp lama bgt”
Sarcasm	Literal meaning differs from intended sentiment	“bagus sekali, baru dipakai langsung rusak”

As shown in Table 9, the model’s errors were not random but were mainly associated with linguistic ambiguity. This finding strengthens the interpretation that the main challenge in Indonesian fashion review sentiment classification is not only noise at the character level but also ambiguity at the semantic and pragmatic levels. Although preprocessing can reduce spelling and formatting noise, sarcasm and implicit meaning remain difficult for supervised classification models.

3.7 Topic Modeling and Sentiment Pattern Analysis

In addition to sentiment classification, topic modeling was used to identify the dominant themes discussed in customer reviews. The LDA analysis identified seven major topics: product quality, material comfort, product design, sizing accuracy, delivery performance, customer service, and price/value. The coherence evaluation indicated that the seven-topic configuration produced the highest coherence score, suggesting that it provided the most interpretable topic structure.

Table 10. Topic Coherence Evaluation

Number of Topics	Coherence Score
5	0.49
6	0.53
7	0.58
8	0.55
9	0.52

As shown in Table 10, the highest coherence score was obtained when the number of topics was set to seven. This result indicates that seven topics were sufficient to represent the main discussion patterns in the customer reviews without creating excessive topic fragmentation.

Table 11. Extracted Customer Review Topics

Topic Number	Topic Description
T1	Product Quality
T2	Material Comfort
T3	Product Design

Building of Informatics, Technology and Science (BITS)
ISSN 2684-8910 (print), ISSN 2685-3310 (online)
Volume 8, No 2, September 2026, Page: 727−741
DOI 10.47065/bits.v8i2.10357
https://ejurnal.seminar-id.com/index.php/bits

Topic Number	Topic Description
T4	Sizing Accuracy
T5	Delivery Performance
T6	Customer Service
T7	Price and Value

The extracted topics show that customer reviews in fashion e-commerce do not only express general satisfaction or dissatisfaction. Instead, they reveal specific product and service dimensions that influence sentiment. Product quality and material comfort were strongly associated with positive sentiment, while sizing accuracy and delivery performance were more frequently associated with negative sentiment. This confirms that fashion review sentiment is attribute-dependent. Customers may evaluate one product positively because the material is comfortable, but they may still express dissatisfaction when the size does not match or delivery is delayed.

3.8 Sentiment-Topic Heatmap Interpretation

Figure 3 presents the sentiment-topic heatmap. The heatmap shows the relationship between sentiment polarity and the seven extracted topics. Product Quality and Material Comfort showed the highest concentration of positive sentiment, while Sizing Accuracy and Delivery Performance showed the strongest concentration of negative sentiment. The heatmap provides deeper insight into why certain reviews were classified as positive or negative. Positive sentiment was concentrated in topics related to product quality, comfort, design, and value for money. This indicates that customers are more likely to express satisfaction when the product meets expectations in terms of material, appearance, and usability.

On the other hand, negative sentiment was concentrated in topics related to sizing mismatch, product inconsistency, delivery delay, and customer service. These topics represent common sources of dissatisfaction in online fashion transactions.

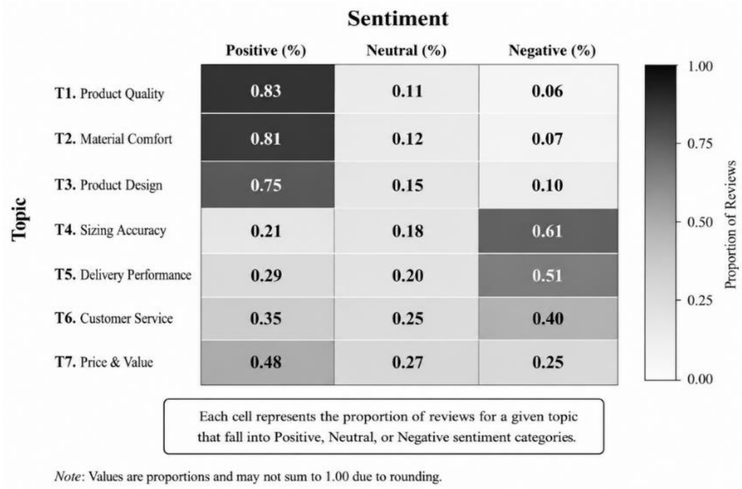

Figure 3. Sentiment-Topic Heatmap

The heatmap also helps explain some confusion matrix errors. For example, reviews related to sizing accuracy may contain mixed sentiment because customers may like the product design but complain about the size. Similarly, delivery-related reviews may be difficult to classify when the product itself is satisfactory but the shipping experience is negative. Therefore, the heatmap supports the confusion matrix analysis by showing that misclassification often occurs in reviews containing overlapping product and service evaluations.

3.9 Digital Marketing Implications

The findings provide several implications for fashion e-commerce businesses. First, positive sentiment related to quality and comfort indicates that these attributes should be emphasized in product positioning and promotional messages. Businesses can use customer language such as "nyaman", "sesuai", and "berkualitas" in product descriptions, advertisements, and customer testimonials to strengthen perceived value.

Second, negative sentiment related to size mismatch suggests the need for more accurate size charts, clearer measurement guides, and customer education regarding product fit. Since sizing issues were among the strongest negative drivers, improving size information may reduce dissatisfaction and product returns.

Third, product inconsistency between promotional images and actual items should be addressed through more transparent visual communication. Sellers should provide real product photos, detailed material descriptions, and customer-generated images to reduce expectation gaps. Fourth, delivery delay and customer service issues indicate that sentiment analysis can be used not only for marketing communication but also for operational improvement.
These findings support the argument that sentiment analysis can function as a business intelligence tool rather than merely a classification technique. By identifying what customers praise and complain about, fashion e-commerce

Building of Informatics, Technology and Science (BITS)
ISSN 2684-8910 (print), ISSN 2685-3310 (online)
Volume 8, No 2, September 2026, Page: 727−741
DOI 10.47065/bits.v8i2.10357
https://ejurnal.seminar-id.com/index.php/bits

businesses can improve product quality, communication strategy, logistics coordination, and customer relationship management [6], [7], [13], [25].

3.10 Comparison with Previous Studies

The results of this study are consistent with previous research showing that transformer-based models generally outperform traditional machine learning approaches in sentiment analysis. Kuang et al. [1] found that advanced learning models improve product review sentiment classification, while Bellar et al. [23] showed that deep learning and transformer-based approaches are effective for e-commerce recommendation and sentiment prediction. The current study extends these findings by comparing Naive Bayes, SVM, LSTM, and IndoBERT in the specific context of Indonesian fashion product reviews.

Compared with Shobha Rani et al. [2], who evaluated conventional machine-learning and neural-network models for sentiment analysis of women's fashion e-commerce reviews, this study provides a broader comparison by incorporating traditional machine-learning, sequential deep-learning, and transformer-based models. Compared with Bharadwaj [10], who focused on mining customer opinions for sentiment classification, this study further explains how classification results relate to customer satisfaction drivers. In addition, this study supports Li et al. [24], who showed that online clothing reviews can reveal specific customer needs that conventional analyses may overlook.

Overall, the results show that IndoBERT produced the strongest performance because it is better suited to contextual Indonesian text representation. However, the remaining misclassifications indicate that sentiment classification in noisy e-commerce data still faces challenges, especially in detecting sarcasm, implicit sentiment, mixed opinions, and very short reviews. Therefore, future research may improve performance by incorporating sarcasm detection, aspect-based sentiment analysis, larger manually annotated datasets, and multimodal analysis using both review text and product images.

3.11 Summary of Findings

The results demonstrate that IndoBERT achieved the highest classification performance among the four evaluated models, with an accuracy of 93.41% and F1-score of 92.51%. This performance is supported by preprocessing, contextual tokenization, supervised fine-tuning, imbalance handling, and cross-validation. The analysis also shows that positive sentiment is mainly driven by product quality, material comfort, design, and value for money, while negative sentiment is mainly associated with sizing mismatch, delivery delay, product inconsistency, material quality concerns, and customer service problems.

The deeper analysis of the confusion matrix and heatmap indicates that misclassifications mainly occurred because of mixed sentiment, ambiguous short reviews, informal expressions, implicit dissatisfaction, and sarcasm. Therefore, although the model achieved high performance, the results should not be interpreted as perfect understanding of all customer expressions. Instead, the findings show that transformer-based contextual representation can substantially improve sentiment classification performance while still leaving room for further improvement in handling complex linguistic phenomena in Indonesian e-commerce reviews.

3.12 Discussion

The findings of this study show that IndoBERT achieved the highest classification performance compared with Naive Bayes, SVM, and LSTM in classifying Indonesian fashion product reviews. The accuracy of 93.41% should not be interpreted as the result of applying a model directly to raw and highly noisy e-commerce data. Instead, this result was obtained through a systematic classification process consisting of data cleaning, text normalization, tokenization, rating-based labeling, class imbalance handling, supervised fine-tuning, and five-fold cross-validation. These stages contributed to reducing textual noise and improving the consistency of sentiment patterns before model training. The result is consistent with previous studies showing that advanced learning models and transformer-based approaches can improve sentiment classification performance in product review analysis [1], [23].

The findings also contribute to fashion informatics and digital fashion analytics. Previous studies have shown that sentiment analysis and topic modeling can reveal fashion-related trends, customer perceptions, and evolving consumer preferences. Kim et al. [26] demonstrated that topic modeling and sentiment analysis can be used to analyze genderless fashion trends from social media data. This study complements that work by focusing on post-purchase fashion product reviews rather than general social media discussions, because customer reviews provide more direct evidence of product satisfaction, dissatisfaction, and practical consumption experience.

Customer reviews should also be viewed as a source of innovation analytics. Mariani and Nambisan [27] argued that online review platforms can function as research-driven ecosystems that generate innovation insights. In this study, recurring negative topics such as size mismatch, product inconsistency, material complaints, and delivery problems indicate areas where sellers can improve product design, product information, quality control, and customer service. Positive topics such as product quality, comfort, and design also indicate attributes that should be strengthened in product positioning and marketing communication.

The relationship between sizing problems and dissatisfaction is also consistent with Sharma and Manocha [28], who showed that customer sentiment analysis can reveal important differences in consumer perceptions of online fashion retailers. In the present study, sizing accuracy appeared as one of the strongest negative topics. This finding

Building of Informatics, Technology and Science (BITS)
ISSN 2684-8910 (print), ISSN 2685-3310 (online)
Volume 8, No 2, September 2026, Page: 727−741
DOI 10.47065/bits.v8i2.10357
https://ejurnal.seminar-id.com/index.php/bits

is important because online fashion customers cannot physically inspect or try products before purchase. As a result, they depend heavily on product photos, size charts, descriptions, and previous customer reviews. When the received product does not match the expected size or appearance, negative sentiment tends to emerge.

From a methodological perspective, the relatively high performance of IndoBERT can be explained by its transformer-based architecture. Unlike Naive Bayes and SVM, which rely on TF-IDF features and treat words primarily as independent or frequency-based units, IndoBERT captures contextual relationships between words through bidirectional representation. This capability is important in Indonesian e-commerce reviews because sentiment meaning often depends on context. For example, the word "tipis" may indicate negative sentiment when used in the phrase "bahannya tipis dan jelek," but it may express a more acceptable or neutral meaning in "bahannya tipis tapi adem." This contextual sensitivity explains why IndoBERT outperformed traditional machine learning and sequential deep learning models. This finding is supported by Huang et al. [29], who identified transformer-based approaches as highly effective techniques for sentiment analysis in e-commerce platforms.

Another source of misclassification was short and ambiguous reviews. Reviews such as "lumayan," "biasa saja," "kurang," or "oke lah" provide limited contextual information. Such expressions may be interpreted differently depending on customer expectations and rating behavior. In some cases, "lumayan" may indicate neutral satisfaction, while in other cases it may imply disappointment.

This finding supports Akter et al. [30], who emphasized that consumer feedback analysis requires deeper interpretation because customer sentiment can directly influence business strategy and decision-making. The sentiment-topic heatmap provides additional insight into the classification results. Positive sentiment was strongly associated with product quality, material comfort, design, and value for money. These findings indicate that customers tend to express satisfaction when the product matches expectations in terms of comfort, appearance, and perceived quality. This is consistent with Nurcahyanie et al. [31], who emphasized the importance of online product reviews and fashion innovation in strengthening business resilience and customer-oriented improvement. In contrast, negative sentiment was concentrated in sizing accuracy, product inconsistency, delivery performance, and customer service.
The broader implication of this study is also relevant to AI-driven digital marketing and e-commerce competitiveness. Sentiment analytics can help businesses transform customer-generated reviews into actionable knowledge for product improvement, brand positioning, customer engagement, and market trend identification. Hasan [32] emphasized that AI-powered digital strategies can enhance competitiveness in e-commerce. In this study, the use of IndoBERT-based sentiment classification and LDA-based topic modeling shows how AI can support more systematic interpretation of customer feedback.

From a practical perspective, the findings show that sentiment analysis can support digital marketing and customer experience management. Positive drivers such as "nyaman," "bagus," "sesuai," and "berkualitas" indicate that fashion sellers should emphasize material comfort, product quality, and expectation matching in product descriptions and promotional content. Meanwhile, negative drivers such as "kecil," "tipis," "lambat," "tidak sesuai," and "berbeda gambar" indicate areas requiring improvement. These findings align with Sundareswaran et al. [33], who showed that sentiment analysis can support customer segmentation, product review interpretation, and recommendation systems.

The findings further support the growing role of artificial intelligence in digital marketing. Ziakis and Vlachopoulou [34], explained that artificial intelligence has become an important component of modern digital marketing because it enables businesses to analyze customer behavior, personalize communication, and improve decision-making. In this study, sentiment classification and topic modeling provide a practical example of how AI-based analytics can transform unstructured customer reviews into structured marketing intelligence.
Finally, the results are consistent with Raghav et al. [35], who argued that the future of digital marketing depends on organizations' ability to leverage artificial intelligence for competitive strategies and tactics. The present study shows that sentiment analytics should not be viewed only as a technical classification task. Instead, it can function as a strategic capability that helps fashion e-commerce sellers identify customer satisfaction drivers, detect recurring complaints, improve product communication, and support data-driven marketing decisions.

Overall, the findings demonstrate that IndoBERT is effective for Indonesian fashion review sentiment classification, but the results should be interpreted with caution. The high accuracy is supported by preprocessing, fine-tuning, imbalance handling, and cross-validation, not by direct classification of raw noisy text. The remaining classification errors show that challenges still exist, especially in mixed sentiment, short ambiguous reviews, sarcasm, implicit dissatisfaction, and aspect-level sentiment variation. Therefore, while IndoBERT provides strong performance for general sentiment classification, future research should incorporate aspect-based sentiment analysis, sarcasm detection, larger manually annotated datasets, and multimodal analysis involving review text and product images to improve interpretability and robustness in noisy e-commerce environments.

4. CONCLUSION

This study concludes that IndoBERT provides the most effective performance for Indonesian fashion review sentiment classification compared with Naive Bayes, SVM, and LSTM, achieving an accuracy of 93.41%, precision of 92.87%, recall of 92.15%, and F1-score of 92.51%. The strong performance was obtained not from raw noisy review text, but through a complete process involving text cleaning, normalization, rating-based labeling, class imbalance handling, supervised fine-tuning, and five-fold cross-validation. The findings show that positive customer sentiment is mainly associated with product quality, comfort, design, and product suitability, while negative sentiment is strongly related to sizing mismatch, material thinness, delivery delay, product inconsistency, and customer service issues. Practically, these results indicate that fashion e-commerce sellers can use sentiment analytics to improve size chart accuracy, provide more transparent product photos and descriptions, strengthen quality communication, and detect recurring customer complaints more systematically. Technically, the study is limited by the use of rating-based sentiment labels, text-only review data, and the remaining difficulty of classifying mixed sentiment, very short ambiguous comments, informal expressions, and sarcasm. Therefore, future research should develop aspect-based sentiment analysis, manually annotated datasets, sarcasm detection, and multimodal models that combine review text with product images to improve interpretability and robustness in noisy e-commerce environments.

5. DECLARATIONS

5.1 Author Contributions

Conceptualization: M.I.I., D.S., I.R.;
Methodology: M.I.I., D.S., I.R., M.M.R.;
Software: M.I.I., I.R., M.M.R.;
Validation: M.I.I., D.S., I.R., M.M.R.;
Formal Analysis: M.I.I., D.S., I.R.;
Investigation: M.I.I., D.S., I.R., M.M.R.;
Resources: M.I.I., D.S., M.M.R.;
Data Curation: M.I.I., I.R., M.M.R.;
Writing Original Draft Preparation: M.I.I., D.S., I.R.;
Writing Review and Editing: M.I.I., D.S., I.R., M.M.R.;
Visualization: M.I.I., I.R., M.M.R.;
All authors have read and agreed to the published version of the manuscript.

5.2 Data Availability

The data presented in this study are available on request from the corresponding author.

5.3 Funding

This research and the publication of this manuscript were fully funded by the authors' personal funds. No external funding was received for the research, preparation, or publication of this manuscript.

5.4 Institutional Review Board Statement

Not applicable.

5.5 Informed Consent Statement

Not applicable.

5.6 Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

5.7 Generative AI and AI-assisted Technologies in The Writing Process

The authors used Generative AI to improve the writing clarity of this paper. They reviewed and edited the AI-assisted content and took full responsibility for the final publication.

REFERENCES

[1] M. Kuang, R. Safa, S. A. Edalatpanah, and R. S. Keyser, “A hybrid deep learning approach for sentiment analysis in product reviews,” *Facta Universitatis, Series: Mechanical Engineering*, vol. 21, no. 3, pp. 479–500, 2023, doi: https://doi.org/10.22190/FUME230901038K.

[2] V. Shobha Rani, K. Deepthi, V. Ramu, and G. Shirisha, “Sentiment Forecasting in Women’s Fashion E-Commerce: A Machine Learning Perspective,” in *Multi-Strategy Learning Environment (ICMSLE 2024)*, Singapore: Springer, 2024, pp. 597–610, doi: 10.1007/978-981-97-1488-9_44.

[3] M. Sharma, N. Sharma, and P. Verma, “Visualizing trends and correlation between fashion features for product design prediction using classification and sentiment analysis,” *Computer Sciences & Mathematics Forum*, vol. 12, no. 1, Art. no. 16, 2025, doi: https://doi.org/10.3390/cmsf2025012016.

[4] Y. Zou, D. B. Luh, and S. Lu, “Public perceptions of digital fashion: An analysis of sentiment and Latent Dirichlet Allocation topic modeling,” *Frontiers in Psychology*, vol. 13, Art. no. 986838, 2022, doi: https://doi.org/10.3389/fpsyg.2022.986838.

Building of Informatics, Technology and Science (BITS)
ISSN 2684-8910 (print), ISSN 2685-3310 (online)
Volume 8, No 2, September 2026, Page: 727−741
DOI 10.47065/bits.v8i2.10357
https://ejurnal.seminar-id.com/index.php/bits

[5] Z. Nawaz, C. Zhao, F. Nawaz, A. A. Safeer, and W. Irshad, "Role of artificial neural networks techniques in development of market intelligence: A study of sentiment analysis of eWOM of a women's clothing company," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 16, no. 5, pp. 1862–1876, 2021, doi: https://doi.org/10.3390/jtaer16050104.
[6] R. H. Chowdhury, "Sentiment analysis and social media analytics in brand management: Techniques, trends, and implications," *World Journal of Advanced Research and Reviews*, vol. 23, no. 2, pp. 287–296, 2024, doi: https://doi.org/10.30574/wjarr.2024.23.2.2369.
[7] V. Gooljar, T. Issa, S. Hardin-Ramanan, and B. Abu-Salih, "Sentiment-based predictive models for online purchases in the era of Marketing 5.0: A systematic review," *Journal of Big Data*, vol. 11, no. 1, Art. no. 107, 2024, doi: https://doi.org/10.1186/s40537-024-00947-0.
[8] S. Elzeheiry, W. A. Gab-Allah, N. Mekky, and M. Elmogy, "Sentiment analysis for e-commerce product reviews: Current trends and future directions," *Preprints*, 2023, doi: https://doi.org/10.20944/preprints202305.1649.v1.
[9] M. A. Al Montaser, B. P. Ghosh, A. Barua, F. Karim, B. C. Das, R. E. R. Shawon, and M. S. R. Chowdhury, "Sentiment analysis of social media data: Business insights and consumer behavior trends in the USA," *Edelweiss Applied Science and Technology*, vol. 9, no. 1, pp. 545–565, 2025, doi: https://doi.org/10.55214/25768484.v9i1.4164.
[10] L. Bharadwaj, "Sentiment analysis in online product reviews: Mining customer opinions for sentiment classification," *International Journal of Multidisciplinary Research*, vol. 5, no. 5, 2023, doi: https://doi.org/10.36948/ijfmr.2023.v05i05.6090.
[11] M. Selimovic, A. A. Almisreb, and S. Amanzholova, "The Role of Sentiment Analysis in Brand Management and Marketing: A Comparative Study," *Procedia Computer Science*, vol. 251, pp. 579–584, 2024, doi: 10.1016/j.procs.2024.11.152.
[12] T. Oyeku, "Unveiling customer sentiments in fast fashion: A comparative analysis of customer reviews, recommendations, and ratings," Ph.D. dissertation, Dublin Business School, Dublin, Ireland, 2024. https://hdl.handle.net/10788/4583
[13] K. S. Kyaw, P. Tepsongkroh, C. Thongkamkaew, and F. Sasha, "Business intelligent framework using sentiment analysis for smart digital marketing in the e-commerce era," *Asia Social Issues*, vol. 16, no. 3, Art. no. e252965, 2023, doi: https://doi.org/10.48048/asi.2023.252965.
[14] S. Erikson and W. Lam, "Sentiment analysis in digital marketing: Evaluating success dimensions of sentiment analysis and its role in digital marketing," 2024. https://lup.lub.lu.se/student-papers/record/9162621
[15] R. Paul, M. H. Imam, and A. J. Mou, "AI-powered sentiment analysis in digital marketing: A review of customer feedback loops in IT services," *American Journal of Scholarly Research and Innovation*, vol. 2, no. 2, pp. 166–192, 2023, doi: https://doi.org/10.63125/61pqqq54.
[16] C. Izumi, S. A. Ghaffar, and W. C. Setiawan, "Enhancing customer satisfaction and product quality in e-commerce through post-purchase analysis using text mining and sentiment analysis techniques in digital marketing," *Journal of Digital Market and Digital Currency*, vol. 2, no. 1, pp. 1–25, 2025. https://doi.org/10.47738/jdmdc.v2i1.26
[17] M. F. Farah, Z. Ramadan, W. Sammouri, and P. Tawk, "Digital luxury fashion shows: Leveraging interactive marketing opportunities through social media sentiment analysis," in *Digital Marketing & eCommerce Conference*, Cham, Switzerland: Springer Nature Switzerland, 2024, pp. 23–30, doi: https://doi.org/10.1007/978-3-031-62135-2_4.
[18] Y. A. Singgalen, "Social network and sentiment analysis of product reviews: Case of smartwatch product content," *International Journal on Social Science, Economics and Art*, vol. 13, no. 4, pp. 255–267, 2024. doi: https://doi.org/10.35335/ijosea.v15i3
[19] V. H. Luong, A. Tarquini, Y. Anadol, P. Klaus, and A. Manthiou, "Is digital fashion the future of the metaverse? Insights from YouTube comments," *Journal of Retailing and Consumer Services*, vol. 79, Art. no. 103780, 2024, doi: https://doi.org/10.1016/j.jretconser.2024.103780.
[20] Y. H. Choi, S. Yoon, B. Xuan, S. Y. T. Lee, and K. H. Lee, "Fashion informatics of the Big 4 Fashion Weeks using topic modeling and sentiment analysis," *Fashion and Textiles*, vol. 8, no. 1, Art. no. 33, 2021, doi: https://doi.org/10.1186/s40691-021-00265-6.
[21] O. Almashaleh, H. Wicaksono, and O. F. Valilai, "A framework for social media analytics in textile business circularity for effective digital marketing," *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 11, no. 2, Art. no. 100544, 2025, doi: https://doi.org/10.1016/j.joitmc.2025.100544.
[22] H. N. D. Şenyapar, "Identifying pandemic era consumer trends: Sentiment analysis of social media posts," *Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, no. 61, pp. 191–213, 2024, doi: https://doi.org/10.30794/pausbed.1399760.
[23] O. Bellar, A. Baina, and M. Ballafkih, "Sentiment analysis: Predicting product reviews for e-commerce recommendations using deep learning and transformers," *Mathematics*, vol. 12, no. 15, Art. no. 2403, 2024, doi: https://doi.org/10.3390/math12152403.
[24] M. Li, L. Zhao, and S. Srinivas, "It is about inclusion! Mining online reviews to understand the needs of adaptive clothing customers," *International Journal of Consumer Studies*, vol. 47, no. 3, pp. 1157–1172, 2023, doi: https://doi.org/10.1111/ijcs.12895.
[25] L. Theodorakopoulos and A. Theodoropoulou, "Leveraging big data analytics for understanding consumer behavior in digital marketing: A systematic review," *Human Behavior and Emerging Technologies*, vol. 2024, no. 1, Art. no. 3641502, 2024, doi: https://doi.org/10.1155/2024/3641502.
[26] H. Kim, I. Cho, and M. Park, "Analyzing genderless fashion trends of consumers' perceptions on social media: Using unstructured big data analysis through Latent Dirichlet Allocation-based topic modeling," *Fashion and Textiles*, vol. 9, no. 1, Art. no. 6, 2022, doi: https://doi.org/10.1186/s40691-021-00281-6.
[27] M. M. Mariani and S. Nambisan, "Innovation analytics and digital innovation experimentation: The rise of research-driven online review platforms," *Technological Forecasting and Social Change*, vol. 172, Art. no. 121009, 2021, doi: https://doi.org/10.1016/j.techfore.2021.121009.
[28] V. Sharma and T. Manocha, "Comparative analysis of online fashion retailers using customer sentiment analysis on Twitter," in *Proceedings of the International Conference on Innovative Computing & Communication (ICICC)*, 2022. http://dx.doi.org/10.2139/ssrn.4361107
[29] H. Huang, A. A. Zavareh, and M. B. Mustafa, "Sentiment analysis in e-commerce platforms: A review of current techniques and future directions," *IEEE Access*, vol. 11, pp. 90367–90382, 2023, doi: https://doi.org/10.1109/ACCESS.2023.3307308.

Building of Informatics, Technology and Science (BITS)
ISSN 2684-8910 (print), ISSN 2685-3310 (online)
Volume 8, No 2, September 2026, Page: 727−741
DOI 10.47065/bits.v8i2.10357
https://ejurnal.seminar-id.com/index.php/bits

[30] P. Akter, S. Hossain, M. T. Siddique, M. I. Ayub, A. Nath, P. C. Nath, and M. M. Hassan, "Sentiment analysis of consumer feedback and its impact on business strategies by machine learning," *The American Journal of Applied Sciences*, vol. 7, no. 1, pp. 6–16, 2025, doi: https://doi.org/10.37547/tajas/volume07issue01-02.

[31] Y. D. Nurcahyanie, M. L. Singgih, and D. S. Dewi, "Implementing online product reviews and Muslim fashion innovation for resilience during the new normal in Indonesia," *Sustainability*, vol. 14, no. 4, Art. no. 2073, 2022, doi: https://doi.org/10.3390/su14042073.

[32] R. Hasan, "Enhancing market competitiveness through AI-powered SEO and digital marketing strategies in e-commerce," *ASRC Procedia: Global Perspectives in Science and Scholarship*, vol. 1, no. 1, pp. 465–500, 2025, doi: https://doi.org/10.63125/31tpjc54.

[33] G. Sundareswaran, H. Kamaraj, S. Sanjay, A. Devi, and P. Elangovan, "Consumer behavior analysis: Analyze customer segmentation, sentiment on product review, and a product recommendation system," *International Journal of Research and Applied Technology (INJURATECH)*, vol. 2, no. 1, pp. 82–90, 2022, doi: https://doi.org/10.34010/injuratech.v2i1.6536.

[34] C. Ziakis and M. Vlachopoulou, "Artificial intelligence in digital marketing: Insights from a comprehensive review," *Information*, vol. 14, no. 12, Art. no. 664, 2023, doi: https://doi.org/10.3390/info14120664.

[35] Y. Y. Raghav, R. K. Tipu, R. Bhakhar, T. Gupta, and K. Sharma, "The future of digital marketing: Leveraging artificial intelligence for competitive strategies and tactics," in *The Use of Artificial Intelligence in Digital Marketing: Competitive Strategies and Tactics*, IGI Global Scientific Publishing, 2024, pp. 249–274, doi: https://doi.org/10.4018/978-1-6684-9324-3.ch011.

