Journal of Applied Data Sciences
Vol. 7, No. 2, May 2026, pp. 1049-1062

An Integrated Text Analytics and Ensemble Machine Learning Framework for Fake Review Detection in Online Marketplaces

Eka Praja Wiyata Mandala[1,*], Sarjon Defit[2], Gunadi Widi Nurcahyo[3]

[1,2,3]*Department of Information Technology, Faculty of Computer Science, Universitas Putra Indonesia YPTK Padang, Indonesia*

(Received: October 25, 2025; Revised: December 20, 2025; Accepted: March 12, 2026; Available online: April 4, 2026)

Abstract

The increasing prevalence of fake reviews on e-commerce platforms undermines consumer trust and affects purchasing decisions, particularly for local products by limited visibility such as those by West Sumatra, Indonesia. This study proposes a hybrid approach combining text analytics and machine learning to enhance the detection of fake reviews. Four classification models—Naive Bayes, Random Forest, Logistic Regression, and K-Nearest Neighbor—were tested on a dataset of 1,500 labeled product reviews. Among these models, Random Forest had the highest starting accuracy of 0.8533. To further enhance performance, this study introduces EKAHypeRFor (Enhanced Knowledge Augmentation of Hyperparameter Random Forest), an optimized Random Forest framework that integrates sentiment-based feature engineering with systematic hyperparameter optimization using RandomizedSearchCV. By leveraging composite sentiment polarity and sentiment strength features derived from hybrid labeling, the proposed method improves classification accuracy while maintaining model interpretability. The enhanced model reached an accuracy of 0.8778, which is 2.45% higher than the original. To support practical application, the optimized EKAHypeRFor model is integrated into a lightweight prototype system that performs real-time classification of incoming review text using a Python-based machine learning pipeline. The system enables automatic sorting of reviews into fake and genuine categories through direct model inference. Functional testing was conducted using representative review samples to verify classification accuracy and response behavior. Formal usability evaluation with end users is not included in this study and is suggested as future work. This method is simple, fast, and accurate, helping to make online product reviews more trustworthy for small and medium businesses in the area.

Keywords: Fake Review Detection, Machine Learning, Text Analytics, Random Forest, Hyperparameter Optimisation, West Sumatra Products, Ekahyperfor

1. Introduction

Online shopping in Indonesia is growing quickly. Because of this, many more customers are writing reviews. These reviews help people decide if they can trust a product or seller. But some reviews are fake. They are deliberately generated to mislead potential consumers. This practice undermines the reliability of information available on e-commerce platforms. This not only makes customers lose trust but also hurts fair competition between businesses and the online community. Online buying and selling are growing, which affects how products are bought and sold. Studies have shown that a substantial proportion of online reviews in Indonesian e-commerce platforms are deceptive, with estimates ranging from 15% to 30% in competitive product categories. The rapid expansion of user-generated reviews in Indonesia has intensified this problem, underscoring the need for scalable and automated fake review detection approaches.

Several studies have investigated the emergence and impact of fake reviews in online marketplaces. Joseph and Hemalatha [1] proposed an ensemble-based machine learning framework to detect deceptive reviews on e-commerce platforms, demonstrating the effectiveness of supervised classification methods. Han et al. [2] integrated explainable knowledge representations to improve model transparency, while Liu et al. [3] incorporated behavioral features alongside textual content to enhance detection accuracy. Luo et al. [4] introduced a mixed probability approach to identify fraudulent reviews, and Silpa et al. [5] applied traditional machine learning techniques to classify fake reviews based on textual patterns. Sentiment analysis is used to assess customer opinions about products. Reviews are distributed to enhance and disseminate customer perspectives, providing information about customer sentiment to

***Corresponding author: Eka Praja Wiyata Mandala (ekaprajawm@upiyptk.ac.id)**
DOI: https://doi.org/10.47738/jads.v7i2.1143
This is an open access article under the CC-BY license (https://creativecommons.org/licenses/by/4.0/).
© Authors retain all copyrights

distinguish between positive and negative reviews [6]. Customers use online reviews because they find third-party information more credible [7]. However, not all online reviews are true. Therefore, sellers should understand how fake online reviews affect customer purchase behaviour [8].

Fake reviews are intentionally produced by certain users to manipulate the perceptions of other consumers, and sellers write fake reviews to reduce their competitors' sales [9], [10]. The increase in fake reviews has made customers doubt more. Fake reviews are often posted by paid reviewers or competitors, which hurts the product's reputation [11]. Writing a fake review starts by collecting information, organizing it, writing the fake review, and then improving it to make it look real [12]. Deceptive fake reviews are always a difficult point to detect because they have hidden and diverse characteristics [13]. Fake reviews are created to deceive potential customers and influence their purchasing decisions [14]. Previous studies have employed machine learning techniques such as Support Vector Machine (SVM) and Naive Bayes for fake review detection [15]. These approaches are discussed to provide contextual background rather than as part of the experimental evaluation conducted in this study. Another research combined word vectors generated using TF-IDF and BERT to enhance the classification performance of Amazon product reviews. The hybrid model's performance was checked by measuring accuracy, recall, precision, and F1 score. The model reached an accuracy of 88% [16]. Another study used exploratory data analysis to classify product reviews as positive, negative, or neutral. It tested three machine learning methods—NB, LR, and RF—on Samsung reviews by Amazon on Kaggle. The results drawn that Random Forest had the highest accuracy, over 90% [17].

The growing number of customer reviews on marketplace platforms in Indonesia has created new problems by how trustworthy the information is. Many of these reviews are fake reviews that are created by manipulative intentions, either to enhance a product's image or to bring down a competitor's reputation. The main problem is that there is no accurate system that can adjust to the local language features of Indonesian. Because of this, many fake reviews are not caught by automatic checks. Also, methods that only use raw text devoid of looking at sentiment often give less accurate results. Therefore, there is a need for an approach that is able to integrate sentiment analysis by text classification algorithms more systematically. Although prior studies have explored ensemble and hybrid sentiment labeling approaches, many rely on supervised learning or complex model stacking that requires large annotated datasets and substantial computational resources. In contrast, this study adopts a lightweight hybrid labeling strategy that combines complementary lexicon-based tools (VADER and TextBlob) through transparent rule-based aggregation, enabling robust weak supervision for small-scale and domain-specific applications such as Indonesian marketplace reviews.

Unlike previous studies that primarily rely on rule-based sentiment scoring or single classifiers, this research aims to improve the accuracy and interpretability of Indonesian review analysis by introducing lightweight hybrid labeling technique called VADERBLOB and an optimized ensemble learning framework named EKAHypeRFor. The VADERBLOB method combines lexicon-based and corpus-based sentiment scoring to produce more reliable labels, while EKAHypeRFor integrates multiple classifiers with automatic hyperparameter tuning to achieve robust performance. The study evaluates the performance of classification models using metrics such as accuracy, precision, recall, and F1-score. Additionally, it contributes by developing a fake review detection system tailored to the Indonesian language and local user behavior. These innovations not only capture subtle linguistic nuances in Indonesian texts but also enhance model adaptability across different datasets. The main contribution of this study lies in presenting a practical and explainable sentiment classification model that outperforms conventional methods and can be effectively applied in demonstrates practical feasibility for deployment-oriented scenarios under offline evaluation settings.

2. Literature Review

2.1. Machine Learning

Recent advancements in Artificial Intelligence (AI) and Machine Learning (ML) have enabled researchers to address complex classification problems, including fake review detection. Prior studies have demonstrated that machine learning–based classifiers can effectively extract discriminative patterns from textual review data and improve detection accuracy [18], [19]. Ensemble approaches, such as Random Forest–based models, have shown particular robustness when dealing with noisy and high-dimensional features [20]. Combining lexicon-based and supervised machine learning approaches improves sentiment classification by leveraging their complementary strengths [21]. However, many existing methods depend on large, manually labeled datasets and complex feature representations, which limit their applicability in domain-specific or low-resource settings. These limitations motivate the need for an optimized yet lightweight ensemble framework, as proposed in this study.

2.2. Text Analytics

Text analytics plays a fundamental role in extracting meaningful patterns from unstructured textual data, particularly in applications such as sentiment analysis and opinion mining. Previous studies have shown that text analytics techniques enable the identification of latent sentiment cues and thematic structures within large volumes of user-generated content [22], [23]. These techniques have been widely applied to customer review analysis, where textual features are leveraged to infer user opinions and product perceptions [24]. Opinion strength captures sentiment intensity, where expressions with similar polarity may differ in strength, which is important in contexts such as n-star rating analysis. This intensity is commonly estimated by analyzing the modifying effects of adverbs and adjectives in opinion text [25]. Although effective, many text analytics approaches rely on complex models that require substantial computational resources and large annotated datasets, limiting interpretability and practical deployment. In contrast, lexicon-based sentiment analysis offers transparency and efficiency, motivating the use of sentiment-driven feature engineering in this study to balance simplicity, interpretability, and predictive performance.

2.3. Fake Review

The problem of fake reviews has received increasing attention due to its significant impact on consumer trust and decision-making in online marketplaces. Prior research has explored various strategies for fake review detection, including supervised machine learning models, sentiment-based analysis, and hybrid approaches that combine textual and behavioral features [26], [27]. These studies demonstrate that incorporating sentiment information can improve the detection of deceptive content, particularly when fake reviews exhibit exaggerated or inconsistent emotional expressions [28]. However, many existing approaches depend heavily on platform-specific behavioral metadata or large-scale manually labeled datasets, which are often unavailable or difficult to obtain in real-world settings. Additionally, several studies focus on complex detection frameworks that limit interpretability and scalability across different languages or domains [29], [30]. These limitations highlight the need for a lightweight and adaptable fake review detection strategy that relies primarily on textual content and weakly supervised labeling. Accordingly, this study adopts a hybrid lexicon-based labeling approach combined with an optimized ensemble classifier to address the challenges of limited labeled data and domain-specific language characteristics.

3. Methodology

Figure 1 illustrates a high-level overview of the proposed methodology, while detailed descriptions of each processing stage and parameter configuration are provided in the subsequent subsections.

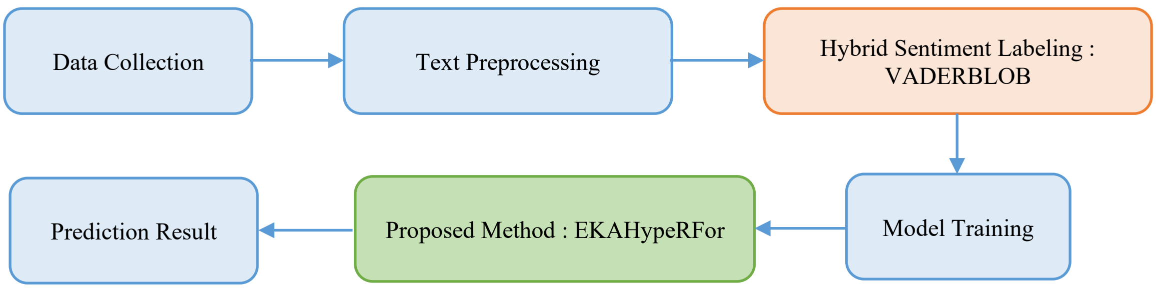

Figure 1. Overview of the proposed hybrid text analytics and ensemble learning framework for fake review detection

Figure 1 shows the full research method to find fake reviews on Indonesian marketplaces. It uses a hybrid approach by sentiment analysis and machine learning classification. The method has several main steps that work together: collecting data, cleaning the text, labeling data by the VADERBLOB method, extracting features, training the model, and making the final prediction using one optimized classification method.

3.1. Data Collection

The study was started by collecting review data by Indonesian marketplaces. The data includes reviews written in Indonesian, which are the focus of the analysis. These reviews are about typical products by West Sumatra. The dataset was compiled by publicly accessible user reviews on Shopee Indonesia, one of the most popular e-commerce platforms in the region. We focused on reviews about traditional food and snacks by West Sumatra, like keripik balado, packaged rendang and songket. Songket, a highly artistic traditional Indonesian textile crafted by interweaving cotton and silk threads across the lunsing, is widely recognized for its cultural significance and is commonly worn at weddings and other formal ceremonial events [31], [32]. Reviews were gathered by manually copying them by product pages by many sales and user comments, to ensure the data was rich and varied. Shopee was chosen as the main source because it is the top e-commerce site in Indonesia and has an easy system to access verified customer feedback. All collected reviews were made anonymous and checked to remove irrelevant or incomplete ones, leaving a final dataset of 1,500 valid reviews for analysis.

3.2.Text Preprocessing

The review corpus underwent a multi-stage preprocessing pipeline designed to enhance data quality and analytical reliability. Initially, extraneous symbols were removed and texts were standardized into modern Indonesian, which not only improved textual consistency but also supported more accurate star rating predictions [33]. The process began with the replacement of slang and foreign words to clarify meaning, followed by the removal of stopwords and punctuation to reduce textual noise. The corpus was then translated into English to ensure consistency with analytical tools and resources. To ensure compatibility with VADER and TextBlob, the Indonesian reviews were translated into English using automated machine translation, a common practice in multilingual sentiment analysis due to its practicality and consistency. The use of English-based tools was motivated by their stability and interpretability. To assess potential translation-induced sentiment drift, a validation check was conducted on a representative subset by comparing sentiment orientation before and after translation. The results showed that most reviews retained their original sentiment polarity, indicating that translation had a limited impact on the sentiment distribution used for labeling. After translation, tokenization was applied to split sentences into individual words, which were subsequently reduced to their basic forms through stemming. These systematic steps provided a linguistically robust and cleaner dataset, significantly improving the reliability of classification and prediction outcomes.

3.3. Hybrid Sentiment Labeling: VADERBLOB

Next, the pre-processed data was translated into English to work by two popular sentiment analysis tools: VADER and TextBlob. VADER (Valence Aware Dictionary and Sentiment Reasoner) is a lexicon-based sentiment analysis method widely recognized for its efficiency and capability in rapidly processing textual data [34]. In this study, the VADER lexicon was employed to annotate the dataset, thereby streamlining the labeling process and improving the accuracy of sentiment classification [35]. VADER produces a compound sentiment score ranging from –1 (strongly negative) to +1 (strongly positive). In this study, sentiment decision thresholds were used to label fake reviews based on polarity intensity, where extreme (≥ 0.85 or ≤ -0.85) or near-neutral (-0.1 to 0.1) sentiments were considered indicative of deceptive behavior. This heuristic approach, commonly adopted in weakly supervised fake review detection, balances labeling strictness and coverage while preserving interpretability. A sensitivity analysis on a representative subset showed that moderate threshold variations led to only marginal changes in label distribution and classification performance, indicating the stability of the labeling strategy.

$$compound = \frac{S}{\sqrt{S^2 + \propto}} \tag{1}$$

where S is the sum of sentiment values and α is a normalization constant (typically 15). The VADER compound score is approximately 0.899, leading to a Fake label based on the defined threshold (≥ 0.85 or ≤ -0.85). TextBlob is a sentiment analysis tool that classifies textual data by calculating its polarity score, which reflects the overall sentiment orientation. Based on this polarity value, the text is categorized as positive, negative, or neutral [36].

$$polarity\ score = \frac{1}{n}\sum_{i=1}^{n} Polarity\ (w_i) \tag{2}$$

where n represents the number of words in the sentence that carry sentiment weight, and Polarity (w_i) is the predefined polarity value of the i^{th} word based on TextBlob's sentiment lexicon. These tools give a label and sentiment score for each review. This labeling process is managed by the method VADERBLOB, which aims to combine the advantages of both approaches. If VADER and TextBlob give the same label, that label is used. But if they differ, the label by the highest absolute score is chosen to keep the labeling consistent and reliable.

When VADER and TextBlob produce conflicting sentiment labels for a given review, the final label is determined based on the sentiment score with the higher absolute magnitude, reflecting stronger sentiment intensity. Formally, let S_V denote the VADER compound score and S_{TB} denote the TextBlob polarity score. The final sentiment label L_{final} is defined as:

$$L_{final} = \begin{cases} L_V, & if\ |S_V| > |S_{TB}| \\ L_{TB}, & if\ |S_{TB}| \geq |S_V| \end{cases} \tag{3}$$

where L_V and L_{TB} represent the sentiment labels assigned by VADER and TextBlob, respectively. For example, consider a review with a VADER compound score of $S_V = 0.89$, which exceeds the predefined extreme threshold and is therefore labeled as Fake, while TextBlob produces a polarity score of $S_{TB} = 0.32$, corresponding to a Genuine label. Since $|0.89|>|0.32|$, the final label is assigned as Fake according to the above rule. This decision mechanism ensures consistent resolution of conflicting sentiment outputs. The VADERBLOB method combines VADER and TextBlob to leverage their complementary strengths in capturing sentiment intensity and lexical polarity. Through a transparent rule-based aggregation mechanism, this approach improves labeling consistency while maintaining interpretability and computational efficiency, making it suitable for weakly supervised scenarios with limited labeled data.

3.4. Model Training

In this study, four classification models were used to see how well machine learning can find fake reviews. These models are Naïve Bayes (NB), Random Forest (RF), Logistic Regression (LR), and K-Nearest Neighbors (KNN). Although Support Vector Machine (SVM) is widely used in fake review detection studies, it was not included in the experimental evaluation of this work. This decision was motivated by the study's focus on ensemble-based optimization and computational efficiency, as well as the comparable performance of tree-based ensemble methods on sentiment-driven feature representations. The selected models were chosen to provide a balanced comparison across probabilistic, ensemble, linear, and instance-based classifiers. Each model learned by and was tested on the same labeled data. The same data cleaning and testing steps were done for all models to make the comparison fair. To see how well the classification models worked, a confusion matrix was used. The data was split carefully, by 30% of the 1,500 reviews kept for testing. To mitigate class imbalance, stratified train–test splitting was applied to preserve class proportions, and class weighting was used where supported to reduce bias toward the majority class during model training.

3.5. Proposed Method: EKAHypeRFor

As explained earlier in the overall method (figure 1), this study has several steps: data collection, preprocessing, sentiment labeling, and model training. In this part, we explain the model optimization using the EKAHypeRFor (Enhanced Knowledge Augmentation of Hyperparameter Random Forest) approach. We focus on tuning hyperparameters and doing simple feature engineering to make the model more accurate and easier to understand. After feature extraction, the dataset is divided using stratified splitting, ensuring that class labels remain proportionally represented in both the training and testing sets. This step is crucial for maintaining balanced representation, particularly when certain classes have fewer examples. The subsequent step involves selecting the range of hyperparameters to evaluate. These include n_estimators (100 to 400), max_depth (up to 40), min_samples_split (by 2 to 15), class_weight (set to balanced), bootstrap (True or False), and criterion (Gini or Entropy). This chosen range helps prevent overfitting, enhances how well the model works on new data, and makes the model stronger overall. Hyperparameter optimization is performed using RandomizedSearchCV, a fast and efficient technique that tests 100 different parameter combinations. It employs 5-fold cross-validation to evaluate the effectiveness of each setting. The optimization process is performed in parallel (n_jobs=-1) to reduce computation time while preserving reproducibility. The optimal model identified through this process is then evaluated using accuracy, calculated as the ratio of correctly predicted instances (true positives and true negatives) to the total number of predictions (TP + TN / P + N).

 ISSN 2723-6471

4. Results and Discussion

4.1. Data Collection

The dataset employed in this study consists of product reviews collected from online marketplace stores specializing in traditional West Sumatran products. The selected stores include Rendang Uni Lili, Kripik Balado Christine Hakim, and Kripik Balado Shirley. In total, 1500 reviews were gathered from these three stores. For illustrative purposes, however, only five representative samples are presented in this paper, as shown in table 1.

Table 1. Reviews Datasets

Num	Original Reviews
1	The packaging is neat, thick cardboard, layered with bubble wrap. Can't wait to eat the chips, and it's proven. Often buy directly at the stand. Because there is an online shop, just buy it at Shopee.
2	Neat packaging with a box and bubble wrap. The plastic is not torn. Fast shipping. Original product. Recommended seller.
3	Favorite Balado chips 💜💜 Thank you seller,,, Packaging is ok, shipping is also fast,, definitely repeat order here 👍👍👍
4	Standard packaging, not too long, not too fast either, but the packaging is safe, the box is also tightly closed. The product is very good, crispy, the seasoning is delicious ❤😍 Shipping with JNT is very fast
5	Fast process, fast response, fast delivery, great original quality, delicious and tasty
6	My favorite chips, I've bought them several times. They're always securely packaged in cardboard, and the shipping from Padang is fast. The chips are delicious and crunchy.
7	The package landed safely.. fast packaging, fast delivery.. next order again.. Thanks seller and shopee..
8	Loved the snacks from Padang. The package arrived safely. The packaging was nice and neat 👍👍🌷 Thank you Shopee and the seller and courier.
9	It's not hard and tastes delicious. I kept asking for more. Delivery was on time. The packaging was also safe. Thank you
….	……..
1.500	Neat packaging, the chips are not broken, the taste is delicious. Thank God, the item has been received, the packaging is fast, the packaging is super safe, none of the chips are damaged, the taste is delicious as expected. Thank you...seller

4.2. Text Preprocessing

As presented in Table 1, the initial dataset consists of raw product reviews containing various informal expressions, slang, and foreign words. To ensure textual consistency and analytical reliability, the dataset was subjected to a comprehensive preprocessing pipeline. The process began with replacing slang and foreign words to clarify meaning, followed by the removal of stopwords and punctuation to reduce noise. The reviews were then translated into English, stemmed to reduce words to their base forms, and tokenized into individual words. The outcome of this preprocessing workflow is shown in Table 2, which presents the cleaned and standardized version of the original dataset, ready for further analysis.

Table 2. Text Preprocessing Result

Num	Tokenization Result
1	"The", "packaging", "is", "neat", "thick", "cardboard", "layered", "with", "bubble", "wrap", "Cant", "wait", "to", "eat", "the", "chips", "and", "its", "proven", "Often", "buy", "directly", "at", "the", "stand", "Because", "there", "is", "an", "online", "shop", "just", "buy", "it", "at", "Shopee"
2	"Neat", "packaging", "with", "a", "box", "and", "bubble", "wrap", "The", "plastic", "is", "not", "torn", "Fast", "shipping", "Original", "product", "Recommended", "seller"
3	"Favorite", "Balado", "chips", "Thank", "you", "seller", "Packaging", "is", "ok", "shipping", "is", "also", "fast", "definitely", "repeat", "order", "here"
4	"Standard", "packaging", "not", "too", "long", "not", "too", "fast", "either", "but", "the", "packaging", "is", "safe", "the", "box", "is", "also", "tightly", "closed", "The", "product", "is", "very", "good", "crispy", "the", "seasoning", "is", "delicious", "Shipping", "with", "JNT", "is", "very", "fast"
5	"Fast", "process", "fast", "response", "fast", "delivery", "great", "original", "quality", "delicious", "and", "tasty"
6	"My", "favorite", "chips", "I", "have", "bought", "them", "several", "times", "They", "are", "always", "securely", "packaged", "in", "cardboard", "and", "the", "shipping", "from", "Padang", "is", "fast", "The", "chips", "are", "delicious", "and", "crunchy"
7	"The", "package", "landed", "safely", "fast", "packaging", "fast", "delivery", "next", "order", "again", "Thanks", "seller", "and", "shopee"
8	"Loved", "the", "snacks", "from", "Padang", "The", "package", "arrived", "safely", "The", "packaging", "was", "nice", "and", "neat", "Thank", "you", "Shopee", "and", "the", "seller", "and", "courier"

ISSN 2723-6471

Num	Tokenization Result
9	"It's", "not", "hard", "and", "tastes", "delicious", "I", "kept", "asking", "for", "more", "Delivery", "was", "on", "time", "The", "packaging", "was", "also", "safe", "Thank", "you"
….	……..
1.500	"Neat", "packaging", "the", "chips", "are", "not", "broken", "the", "taste", "is", "delicious", "Thank", "God", "the", "item", "has", "been", "received", "the", "packaging", "is", "fast", "the", "packaging", "is", "super", "safe", "none", "of", "the", "chips", "are", "damaged", "the", "taste", "is", "delicious", "as", "expected", "Thank", "you", "seller"

4.3. Hybrid Sentiment Labeling : VADERBLOB

To ensure compatibility by sentiment analysis tools such as VADER and TextBlob—which are primarily designed for English-language texts—the original Indonesian reviews were translated into English prior to analysis. We used the Google Translate API because it supports Bahasa Indonesia well and is commonly used in language research for working by multiple languages. This tool was chosen because past studies drawn it translates informal and user-written content accurately. We know that automatic translation can cause errors by certain cultural or context-specific expressions. However, since the overall sentiment (positive, neutral, negative) stayed consistent, this step is accurate enough for weak supervision labeling. Also, common Indonesian slang was changed to standard words during preprocessing to reduce confusion. Next, polarity scores are created using TextBlob. These scores are then compared by those by the other method to check the sentiment results. It is labeled as a Fake Review if the score is in the neutral (-0.1 to 0.1) or extreme (≤ -0.85 or ≥ 0.85) range, and a Genuine Review for scores outside of these limits. The score of each method can be seen in Table 3.

Table 3. Initial sentiment scores and labels from VADER and TextBlob

Num	Compound Score	VADER Label	Polarity Score	TextBlob Label
1	0.8990	Fake	0.3200	Genuine
2	0.4404	Genuine	0.5167	Genuine
3	0.8578	Fake	0.7800	Genuine
4	0.8176	Genuine	0.4000	Genuine
5	0.8313	Genuine	0.5367	Genuine
6	0.8020	Genuine	0.3000	Genuine
7	0.3696	Genuine	0.5279	Genuine
8	0.5719	Genuine	0.4000	Genuine
9	0.5574	Genuine	0.2000	Genuine
….	….	….	….	….
1.500	0.7219	Genuine	0.3542	Genuine

Table 3 reveals two instances where VADER classified the reviews as Fake, whereas TextBlob labeled them as Genuine. Applying the VADERBLOB method involves determining the final label by comparing VADER's compound score by TextBlob's polarity score and selecting the label by the higher absolute sentiment value. In cases where VADER and TextBlob generate different sentiment scores for reviews, the final label is decided based on the method by the higher absolute score. Table 4 presents the final sentiment labels obtained after applying the conflict resolution rule defined in Equation (3).

Table 4. Final sentiment labels using the VADERBLOB conflict resolution rule

Num	Label VADER	Label TextBlob	Final Label
1	Fake	Genuine	Fake
2	Genuine	Genuine	Genuine
3	Fake	Genuine	Fake
4	Genuine	Genuine	Genuine
5	Genuine	Genuine	Genuine
6	Genuine	Genuine	Genuine
7	Genuine	Genuine	Genuine
8	Genuine	Genuine	Genuine

Num	Label VADER	Label TextBlob	Final Label
9	Genuine	Genuine	Genuine
….	….	….	….
1.500	Genuine	Genuine	Genuine

After using the VADERBLOB labeling method on all 1,500 reviews, we found that 1,224 (81.6%) were labeled genuine and 276 (18.4%) were labeled fake. This indicates a moderate class imbalance, by genuine reviews being significantly more prevalent. To handle this imbalance during training and testing, we used a stratified train-test split, keeping the same ratio of fake and genuine reviews in both sets. For algorithms that allow it, we also used class_weight=balanced to give more importance to the smaller fake review class. This helps the model focus on fake reviews and avoid bias toward the larger genuine review group.

4.4. Model Training

Among all the models tested, the Random Forest (RF) model had the highest accuracy, by a score of 0.8533. This means the Random Forest (RF) model was the best at correctly finding both real and fake reviews in the test data. The exact counts of true positives, true negatives, false positives, and false negatives are shown in the confusion matrix in figure 2. This also shows that the model did better than the others.

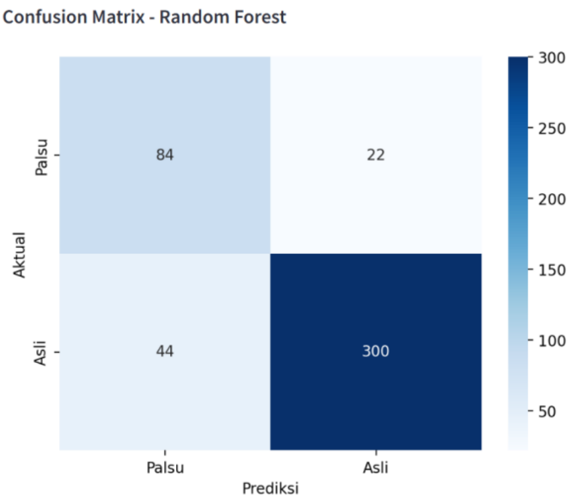

Figure 2. Confusion Matrix of Random Forest

This study compared five classification models—Naïve Bayes (NB), Random Forest (RF), Logistic Regression (LR), and K-Nearest Neighbors (KNN)—to see how well they detect fake reviews. Each model was checked using the same measurements: accuracy, precision, recall, and F1-score. The results are shown in table 5, which clearly shows that the Random Forest model did better than the others in most measurements. This comparison gives strong proof that ensemble methods like Random Forest work well for text classification, especially when the data is noisy or unbalanced.

Table 5. Model comparison results

Model	Accuracy	Precision	Recall	F1 score
NB	0.8400	0.8542	0.3868	0.5325
RF	0.8533	0.6562	0.7925	0.7179
LR	0.8489	0.9318	0.3868	0.5467
KNN	0.8500	0.7312	0.6415	0.6834

Among the tested models, the Random Forest classifier had the highest accuracy. It did better than the other algorithms at finding the complex patterns in the data. This shows that ensemble methods like Random Forest are strong and effective for handling data that has many features and may be noisy.

4.5. Proposed Method : EKAHypeRFor

The EKAHypeRFor results are evaluated using a confusion matrix and feature importance analysis, which together demonstrate effective fake review classification and model interpretability, supporting its applicability to real marketplace review data. Using the EKAHypeRFor method enhanced the classification results compared to the basic Random Forest model. The basic model, which used default or hand-picked settings, reached an accuracy of 0.8533 on the test data. In contrast, after using EKAHypeRFor, which includes smart hyperparameter tuning by RandomizedSearchCV and simple feature engineering, the accuracy rose to 0.8778, as shown in table 6. This is an enhancement of 0.0245 in accuracy, showing that the optimization method helps the model better tell fake reviews by real ones. The result shows that choosing hyperparameters carefully and using good methods can greatly enhance model performance in feasibility for deployment-oriented scenarios under offline evaluation text classification tasks.

Table 6. Accuracy Enhancement of EKAHypeRFor Compared to Baseline Random Forest Model

Model and Proposed Model	Accuracy
Random Forest baseline	0.8533
EKAHypeRFor	0.8778
Accuracy enhancement	0.0245

The performance of the EKAHypeRFor method was checked using a confusion matrix. The test data was 30% of 1,500 labeled reviews. The results show the model predicted very accurately, by an accuracy score of 0.8778. This proves that the hyperparameter tuning in EKAHypeRFor works well. The confusion matrix depicted in figure 3 shows a balanced classification between genuine and fake review categories, by a high proportion of true positives and true negatives.

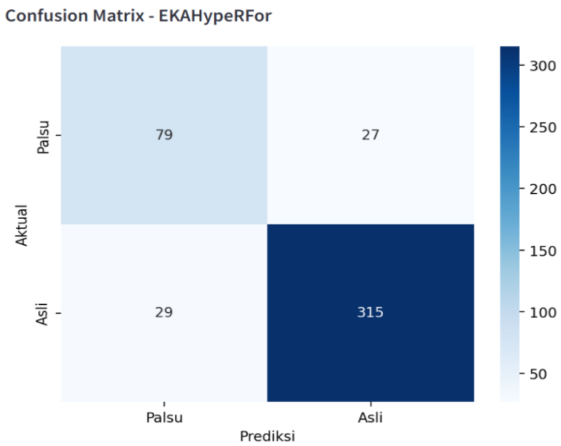

Figure 3. Confusion Matrix of EKAHypeRFor

The dataset is moderately imbalanced, with fake reviews representing approximately 18.4% of the samples, making accuracy alone insufficient for performance evaluation. Therefore, minority-class metrics such as precision, recall, and F1-score are emphasized. As shown in figure 3, the proposed EKAHypeRFor model achieves a balanced precision–recall trade-off for fake reviews, resulting in a stable F1-score that indicates effective detection of deceptive content despite class imbalance.

The EKAHypeRFor method successfully found the best hyperparameters, which greatly enhanced the model's performance. As presented in table 7, the best-performing Random Forest configuration includes a specific combination of parameters such as the number of estimators, maximum tree depth, minimum samples required for splitting and leaf nodes, bootstrap setting, class weighting strategy, and splitting criterion. These optimized settings were found using a guided RandomizedSearchCV process, helping the model learn complex patterns in the data and avoid overfitting. The results show that tuning hyperparameters is very important for improving accuracy, especially in text tasks like detecting fake reviews.

Table 7. Optimal Hyperparameter Configuration for EKAHypeRFor Using RandomizedSearchCV

Parameter	Values
n_estimators	300
max_depth	40
min_samples_split	2
min_samples_leaf	2
max_features	log2
class_weight	None
bootstrap	False
criterion	entropy

In this study, the feature engineering strategy deliberately focuses on a minimal set of sentiment-based features, namely the combined sentiment polarity score and absolute sentiment strength. This parsimonious design aims to assess how effectively hybrid sentiment signals alone can support fake review detection without relying on high-dimensional textual representations. By limiting the feature space, the proposed approach prioritizes interpretability, computational efficiency, and ease of deployment, particularly in low-resource or small-scale e-commerce environments. These were measured by the optimized Random Forest model in the EKAHypeRFor method. Both features help the model almost equally, by importance scores close to 0.5. This means that both the direction of sentiment and how strong the sentiment is have a balanced effect on the model's decisions. This result validates the effectiveness of our minimal feature engineering strategy, emphasizing that even by only two well-designed features, the model can achieve high classification accuracy. The relative contribution of the composite sentiment polarity and absolute sentiment strength to the EKAHypeRFor model is illustrated in figure 4.

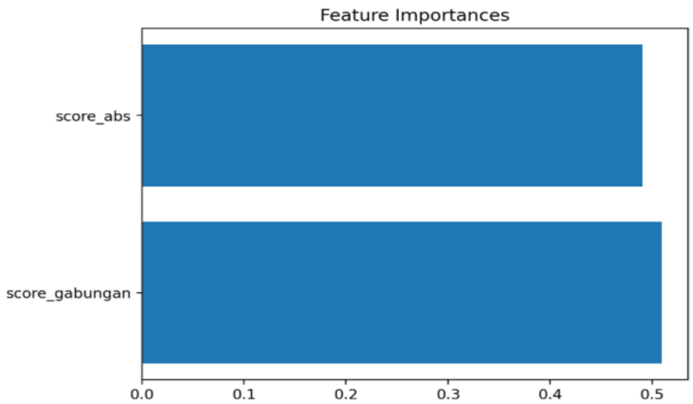

Figure 4: Feature Importance of Composite Sentiment Scores in EKAHypeRFor using Optimized Random Forest Classifier

Analyzing feature importance is essential for understanding how each input variable influences the model's decisions. As shown in figure 4, both features—score_gabungan and score_abs—play a similarly significant role in determining the classification outcome. The feature importance analysis indicates that the two sentiment-based features contribute almost equally to the classification process, with importance values close to 0.5. This observation suggests a balanced influence of sentiment polarity and sentiment strength at the model level. It should be noted that this interpretation is descriptive and based on the relative importance values produced by the optimized Random Forest ensemble, which aggregates information across multiple trees and cross-validation folds. Although formal statistical uncertainty measures such as confidence intervals or standard deviations were not computed, the ensemble-based training procedure provides a degree of stability in the estimated feature contributions.

Despite the minimal feature set, the proposed EKAHypeRFor model achieves competitive performance, indicating that sentiment intensity and polarity provide strong discriminative signals for fake review detection in the studied context.

 ISSN 2723-6471

While the results suggest that the model has not yet reached a clear performance plateau, it is plausible that incorporating richer features could further enhance classification accuracy. However, such extensions may introduce trade-offs in terms of interpretability and computational complexity, which were intentionally minimized in this study.

The results show that the proposed hybrid approach effectively overcomes the limitations of rule-based and single-classifier methods. The integration of the VADERBLOB labeling technique and the EKAHypeRFor ensemble framework yields higher accuracy, better stability, and stronger adaptability across Indonesian marketplace datasets. By combining hybrid sentiment labeling with automated hyperparameter tuning, the proposed model achieves reliable and interpretable results at the global level. Interpretability is provided through feature importance analysis of the optimized Random Forest, which reveals the relative contribution of sentiment-based features across the dataset. Although this approach does not offer instance-level explanations such as SHAP or LIME, it supports transparent and practical interpretability in resource-constrained settings.

Although not evaluated in a live or real-time environment, the proposed framework is designed with deployment considerations in mind. The EKAHypeRFor model employs a Random Forest classifier with a minimal feature set, resulting in low inference overhead and suitability for near real-time processing. Moreover, simplified feature engineering reduces preprocessing latency and supports scalability, although empirical evaluation in live environments remains an important direction for future work.

5. Conclusion

This study developed and evaluated an enhanced machine learning framework for detecting fake reviews in e-commerce, focusing on local West Sumatran products. Using text analytics and four classification models—Naive Bayes, Random Forest, Logistic Regression, and K-Nearest Neighbor—the baseline Random Forest achieved the highest accuracy of 0.8533. To improve this result, a novel optimization method called EKAHypeRFor (Enhanced Knowledge Augmentation of Hyperparameter Random Forest) was introduced, combining simple feature engineering with focused hyperparameter tuning using RandomizedSearchCV. This approach increased accuracy to 0.8778, a 2.45% improvement over the baseline, demonstrating the importance of careful parameter optimization.

Evaluation through a confusion matrix and feature importance analysis confirmed that the model is both robust and interpretable. Despite using only two sentiment-based features (score_gabungan and score_abs), it effectively distinguishes fake from genuine reviews. The integration into an interactive interface further validated its practical usability, offering a lightweight, user-friendly tool suitable for small and medium e-commerce platforms. Compared to prior approaches, the proposed framework delivers superior predictive accuracy without relying on complex deep learning models, making it more accessible and adaptable for Bahasa Indonesia content. Although EKAHypeRFor performed well, limitations remain: sentiment labeling relies on English-oriented tools (VADER and TextBlob), the dataset is relatively small (1,500 reviews) and imbalanced, and behavioral metadata is not yet included.

Future work will extend the proposed framework in several directions. First, data-driven or adaptive threshold optimization strategies may be explored to further refine sentiment-based labeling and improve label reliability under varying data distributions. Second, fold-wise analysis or bootstrapping techniques can be incorporated to quantify the variability and stability of feature importance estimates, providing stronger confidence in model interpretability. In addition, future studies may integrate domain-specific language models such as IndoBERT or Multilingual BERT to capture richer semantic and contextual representations of Indonesian reviews without relying on machine translation. Finally, deployment-oriented research will focus on evaluating runtime latency, throughput, and scalability under controlled real-world marketplace scenarios, enabling systematic assessment of real-time performance and user interaction while addressing practical challenges such as data privacy and evolving language usage.

6. Declarations

6.1. Author Contributions

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

6.2. Data Availability Statement

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

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. 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.

References

[1] S. Joseph dan S. Hemalatha, "Fake review detection using enhanced ensemble support vector machine system on e-commerce platform," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 38, no. 1, pp. 478–485, 2025, doi: 10.11591/ijeecs.v38.i1.pp478-485.

[2] S. Han, H. Wang, W. Li, H. Zhang, dan L. Zhuang, "Explainable knowledge integrated sequence model for detecting fake online reviews," *Appl. Intell.*, vol. 53, no. 6, pp. 6953–6965, 2023, doi: 10.1007/s10489-022-03822-8.

[3] J. Liu, P. Quan, dan W. Zhang, "A Study on Fake Review Detection Based on RoBERTa and Behavioral Features," *Procedia Comput. Sci.*, vol. 242, no.1, pp. 1323–1330, 2024, doi: 10.1016/j.procs.2024.08.131.

[4] J. Luo, J. Luo, G. Nan, dan D. Li, "Fake review detection system for online E-commerce platforms: A supervised general mixed probability approach," *Decis. Support Syst.*, vol. 175, no. 1, pp. 114045, 2023, doi: 10.1016/j.dss.2023.114045.

[5] C. Silpa, P. Prasanth, S. Sowmya, Y. Bhumika, C. H. S. Pavan, dan M. Naveed, "Detection of Fake Online Reviews by using Machine Learning," *2023 Int. Conf. Innov. Data Commun. Technol. Appl.*, vol. 2023, no.1, pp. 71-77, 2023, doi: 10.1109/ICIDCA56705.2023.10099776.

[6] A. A, L. N, P. D, dan Gokilavani, "Predicting the Fake Review to Amazon Product Review Dataset Using Fuzzy Optimized Convolution Neural Network," *2024 Int. Conf. Expert Clouds Appl.*, vol. 2024, no. 1, pp. 689-693, 2024, doi: 10.1109/ICOECA62351.2024.00125.

[7] P. Hajek, L. Hikkerova, dan J.-M. Sahut, "Fake review detection in e-Commerce platforms using aspect-based sentiment analysis," *J. Bus. Res.*, vol. 167, no. 1, pp. 114143, 2023, doi: 10.1016/j.jbusres.2023.114143.

[8] L. J. Harrison-Walker dan Y. Jiang, "Suspicion of online product reviews as fake: Cues and consequences," *J. Bus. Res.*, vol. 160, no.1, pp. 113780, 2023, doi: 10.1016/j.jbusres.2023.113780.

[9] M. Jaiswal dan D. Javale, "Fake Product Review Monitoring System," *2024 IEEE Int. Conf. Interdiscip. Approaches Technol. Manag. Soc. Innov.*, vol. 2024, no. 1, pp. 1-5, 2024, doi: 10.1109/IATMSI60426.2024.10503512.

[10] A. Iqbal, M. A. Rauf, M. Zubair, dan T. Younis, "An Efficient Ensemble approach for Fake Reviews Detection," *2023 3rd Int. Conf. Artif. Intell.*, vol. 2023, no. 1, pp. 70-75, 2023, doi: 10.1109/ICAI58407.2023.10136652.

[11] M. Kumar, A. Rana, A. K. Yadav, dan D. Yadav, “Leveraging Sentiment Analysis to Detect Fake Reviews Using Deep Learning,” *SN Comput. Sci.*, vol. 6, no. 3, pp. 242, 2025, doi: 10.1007/s42979-025-03792-x.

[12] S. Banerjee dan A. Y. K. Chua, “Understanding online fake review production strategies,” *J. Bus. Res.*, vol. 156, no. December 2022, pp. 113534, 2023, doi: 10.1016/j.jbusres.2022.113534.

[13] Z. Shunxiang, Z. Aoqiang, Z. Guangli, W. Zhongliang, dan L. KuanChing, “Building Fake Review Detection Model Based on Sentiment Intensity and PU Learning,” *IEEE Trans. Neural Networks Learn. Syst.*, vol. 34, no. 10, pp. 7476–7488, 2023, doi: 10.1109/TNNLS.2023.3234427.

[14] P. Kalaivani, V. D. Raj, R. Madhavan, dan A. P. N. Kumar, “Fake Review Detection using Naive Bayesian Classifier,” *2023 Int. Conf. Sustain. Comput. Smart Syst.*, vol. 2023, no.1, pp. 705-709, 2023, doi: 10.1109/ICSCSS57650.2023.10169838.

[15] A. H. Alshehri, “An Online Fake Review Detection Approach Using Famous Machine Learning Algorithms,” *Comput. Mater. Contin.*, vol. 78, no. 2, pp. 2767–2786, 2024, doi: 10.32604/cmc.2023.046838.

[16] B. Priya Kamath, M. Geetha, U. Dinesh Acharya, D. Singh, dan A. Rao, “Comprehensive Analysis of Word Embedding Models and Design of Effective Feature Vector for Classification of Amazon Product Reviews,” *IEEE Access*, vol. 13, no. February, pp. 25239–25255, 2025, doi: 10.1109/ACCESS.2025.3536631.

[17] O. Agbonifo, V. Olutayo, dan O. Oluyode, “Machine Learning-Based Sentiment Analysis of Product Reviews,” *J. Trends Challenges Artif. Intell.*, vol. 02, no. 04, pp. 257–264, 2025, doi: 10.61552/JAI.2025.04.002.

[18] A. Saeed dan E. Al Solami, “Fake News Detection Using Machine Learning and Deep Learning Methods,” *Comput. Mater. Contin.*, vol. 77, no. 2, pp. 2079–2096, 2023, doi: 10.32604/cmc.2023.030551.

[19] Z. Fadhel, H. Attia, dan Y. H. Ali, “Optimized and Comprehensive Fake Review Detection based on Harris Hawks optimization integrated with Machine Learning Techniques,” *J. Cybersecurity Inf. Manag.*, vol. 15, no. 1, pp. 11–21, 2025, doi: 10.54216/JCIM.150102.

[20] M. J. Abd dan M. H. Hussein, “Fake reviews detection in e-commerce using machine learning techniques: a comparative survey,” *BIO Web Conf.*, vol. 97, no. 1, pp. 1–12, 2024, doi: 10.1051/bioconf/20249700099.

[21] M. U. Ijairi, M. Abdullahi, dan I. H. Hassan, “Sentiment Classification of Tweets with Explicit Word Negations and Emoji Using Deep Learning,” *Int. J. Softw. Eng. Comput. Syst.*, vol. 9, no. 2, pp. 93–104, 2023, doi: 10.15282/ijsecs.9.2.2023.3.0114.

[22] S. U. Hassan, J. Ahamed, dan K. Ahmad, “Analytics of machine learning-based algorithms for text classification,” *Sustain. Oper. Comput.*, vol. 3, no. April, pp. 238–248, 2022, doi: 10.1016/j.susoc.2022.03.001.

[23] S. Stoyanov dan P. A. Kirschner, “Text analytics for uncovering untapped ideas at the intersection of learning design and learning analytics: Critical interpretative synthesis,” *J. Comput. Assist. Learn.*, vol. 39, no. 3, pp. 899–920, 2023, doi: 10.1111/jcal.12775.

[24] O. P. Suthar, A. Mishra, dan S. Singhal, “Exploring the Landscape of Natural Language Processing for Text Analytics: A comprehensive Review,” *Procedia Comput. Sci.*, vol. 259, no. 1, pp. 453–462, 2025, doi: 10.1016/j.procs.2025.03.347.

[25] M. O. D. Jr, “A Domain-Specific Evaluation of the Performance of Selected Web-based Sentiment Analysis Platforms,” *Int. J. Softw. Eng. Comput. Syst.*, vol. 9, no. 1, pp. 1–9, 2023, doi: 10.15282/ijsecs.9.1.2023.1.0105.

[26] R. Oak, “Detecting review fraud using metaheuristic graph neural networks,” *Int. J. Inf. Technol.*, vol. 16, no. 7, pp. 4019–4025, 2024, doi: 10.1007/s41870-024-01896-w.

[27] J. M. Thilini Jayasinghe dan S. Dassanayaka, “Detecting deception: employing deep neural networks for fraudulent review detection on Amazon,” *Neural Comput. Appl.*, vol. 37, no. 26, pp. 21363–21379, 2025, doi: 10.1007/s00521-025-11485-y.

[28] S. Geetha, E. Elakiya, R. S. Kanmani, dan M. K. Das, “High performance fake review detection using pretrained DeBERTa optimized with Monarch Butterfly paradigm,” *Sci. Rep.*, vol. 15, no. 1, pp. 1–26, 2025, doi: 10.1038/s41598-025-89453-8.

[29] T. Bikku, S. Thota, dan P. Shanmugasundaram, “A Novel Quantum Neural Network Approach to Combating Fake Reviews,” *Int. J. Networked Distrib. Comput.*, vol. 12, no. 2, pp. 195–205, 2024, doi: 10.1007/s44227-024-00028-x.

[30] A. M. Santos dan N. Antonio, "Improving trust in online reviews: a machine learning approach to detecting artificial intelligence-generated reviews," *Inf. Technol. Tour.*, vol. 27, no. 3, pp. 739–766, 2025, doi: 10.1007/s40558-025-00329-z.

[31] A. Ramadhanu, J. Na'am, G. W. Nurcahyo, dan Yuhandri, "Development of Affine Transformation Method in the Reconstruction of Songket Motif," *Int. J. Adv. Sci. Eng. Informatioan Technol.*, vol. 12, no. 2, pp. 600–606, 2022, doi: 10.18517/ijaseit.12.2.14069.

[32] A. Ramadhanu, J. Na'am, G. W. Nurcahyo, dan Yuhandri, "Implementation of the Affine Segmentation Point Method and Image Blending Techniques in Creating New Songket Motifs," *2022 9th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI2022)*, vol. 2022, no. 1, pp. 233–238, 2022 doi: 10.23919/EECSI56542.2022.9946616.

[33] M. Işik dan H. Dağ, "The impact of text preprocessing on the prediction of review ratings," *Turkish J. Electr. Eng. Comput. Sci.*, vol. 28, no. 3, pp. 1405–1421, 2020, doi: 10.3906/elk-1907-46.

[34] K. Barik dan S. Misra, "Analysis of customer reviews with an improved VADER lexicon classifier," *J. Big Data*, vol. 11, no. 10, pp. 1-29, 2024, doi: 10.1186/s40537-023-00861-x.

[35] V. Nurcahyawati dan Z. Mustaffa, "Vader Lexicon and Support Vector Machine Algorithm to Detect Customer Sentiment Orientation," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 9, no. 1, pp. 108–118, 2023, doi: 10.20473/jisebi.9.1.108-118.

[36] G. Kaur, A. Kaur, M. Khurana, dan R. Damasevicius, "Sentiment polarity analysis of love letters: Evaluation of TextBlob, Vader, Flair, and Hugging Face transformer," *Comput. Sci. Inf. Syst.*, vol. 21, no. 00, pp. 40–40, 2024, doi: 10.2298/csis240328040k.