Inform : Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi
Vol.10 No.2 July 2025, P-ISSN : 2502-3470, E-ISSN : 2581-0367

A Comparative Analysis of Explainable AI Techniques for Sentiment Classification of TikTok & Tokopedia Reviews

Muhammad Rafi Haidar Arsyad[1*], Eri Zuliarso[2], Sarah Hussein[3]

[1,2]*Information Technology Department, Universitas Stikubank, Semarang, Indonesia*

[3]*Roads and Transport Department, College Of Engineering, University Of Al-Qadisiyah, Iraq*

[1]rafihaidar@edu.unisbank.ac.id (*)

[2]eri299@edu.unisbank.ac.id, [3]sarah.hussein@qu.edu.iq

Received: 2025-07-18; Accepted: 2025-08-21; Published: 2025-08-30

Abstract—Explainable Artificial Intelligence (XAI) using the LIME (Local Interpretable Model-Agnostic Explanations) method to enhance the interpretability of sentiment classification for user reviews on TikTok and Tokopedia. Using TF-IDF for feature extraction, three machine learning classifiers —Random Forest, Decision Tree, and K-Nearest Neighbours—were evaluated through K-Fold Cross-Validation. Random Forest achieved the highest classification accuracy at 89.9%, followed by Decision Tree at 88.25%, and KNN at 81.51%. The most prominent terms in positive sentiment reviews included "*mantap*" (16,057.23) and "*bagus*" (15,310.02), while negative sentiment was associated with "*biaya*," "*sistem*," and "*ganti*." LIME provided localized, interpretable insights by highlighting the important terms that influence each prediction. In terms of positive sentiment, words such as "Tokopedia," "update," and "go" had strong weights, whereas negative classifications were triggered by terms like "offline" and "error." ROC Curve analysis further confirmed Random Forest's strong performance, showing AUC scores of 0.88 for the negative class, 0.87 for the neutral class, and 0.88 for the positive class, outperforming the other models. The network graph also identified "Tokopedia" as a central node, with frequent co-occurrence of terms like "*diskon*" and "*pengiriman*," reflecting key user expectations. These findings demonstrate that combining interpretable AI with high-performing classifiers offers a powerful approach for sentiment analysis in digital platforms. It enables stakeholders to understand user feedback better and make data-driven decisions to improve customer satisfaction and trust.

Keywords— Sentiment Analysis; Explainable AI Lime; Decision Tree; K-Nearest Neighbors; E-Commerce; Random Forest.

I. INTRODUCTION

Sentiment analysis has become a pivotal tool for extracting meaningful insights from the vast and unstructured user-generated content on digital platforms [1]. With the exponential growth of e-commerce platforms like Tokopedia and social media platforms like TikTok, analyzing user sentiments is crucial for understanding consumer perception and improving service quality [2]. Machine learning (ML) algorithms such as Random Forest have been widely adopted in this domain due to their robustness, high classification accuracy, and ability to handle high-dimensional data [3]. Despite these advantages, traditional models often lack interpretability, posing challenges for decision-making in practical business applications [4].

To address the interpretability gap in ML predictions, Explainable Artificial Intelligence (XAI) has emerged as a solution that bridges the divide between model performance and human understanding [5]. One of the most effective XAI techniques is LIME (Local Interpretable Model-Agnostic Explanations), which provides localized, model-independent explanations for each prediction by highlighting the contributions of individual features [6]. In this context, LIME is applied to sentiment classification tasks using three widely used classifiers: Random Forest (RF), Decision Tree (DT), and K-Nearest Neighbors (KNN) [7]. The integration of LIME allows stakeholders to gain insight into why a particular review is classified as positive, neutral, or negative, making the model's reasoning transparent and actionable [8].

Previous works in sentiment classification have demonstrated strong predictive capabilities using various ML models [9]. However, most of them fall short in providing explainability, limiting their practical value for decision support systems [10]. This research introduces a comparative approach that not only evaluates classification performance but also visualizes feature importance using tools such as Explicit LIME, WordCloud, ROC Curve, and K-Fold Cross-validation on datasets from both TikTok and Tokopedia [11]. The novelty lies in the dual-platform, interpretable sentiment analysis pipeline, which fills a critical gap in existing literature and provides actionable insights for digital platform managers, recommendation systems, and policy formulation [12].

II. RESEARCH METHODOLOGY

This study adopts a systematic approach to sentiment classification by integrating traditional ML techniques with Explainable AI (XAI), specifically the LIME (Local Interpretable Model-Agnostic Explanations) framework [13]. The methodological flow consists of several stages: planning, data collection, pre-processing, model training, validation, and interpretability analysis [14]. The primary goal is to evaluate and interpret sentiment polarity in user reviews from TikTok and Tokopedia, using three classification algorithms: RF, DT, and KNN [15]. Each stage is carefully designed to ensure data quality, model accuracy, and transparency in predictive outcomes. The stages of the research process are illustrated in Fig.1.

DOI : https://doi.org/10.25139/inform.v10i2.10748

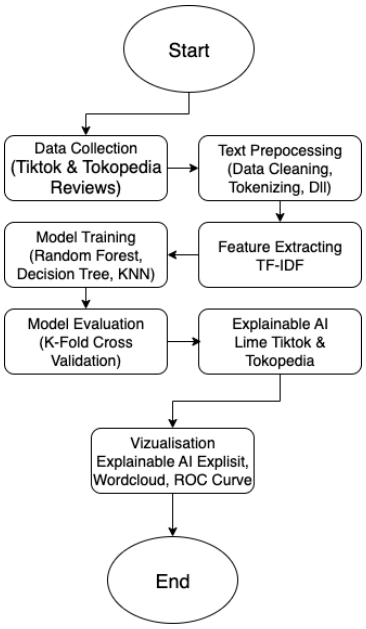

Fig.1.Workflow of Research Stage

A.Data Collection

The data used in this study were collected from user-generated reviews of the Tokopedia and TikTok applications available on the Google Play Store [16]. These reviews, which represent unstructured textual data, were extracted using an automated web scraping process developed in Python with the assistance of the BeautifulSoup and Scraper libraries [17]. The dataset encompasses five years and includes essential attributes such as review content, timestamp, and user rating [18]. The entire data collection workflow is systematically illustrated in Fig.2, which outlines the sequential stages for retrieving and preparing the review data from both platforms.

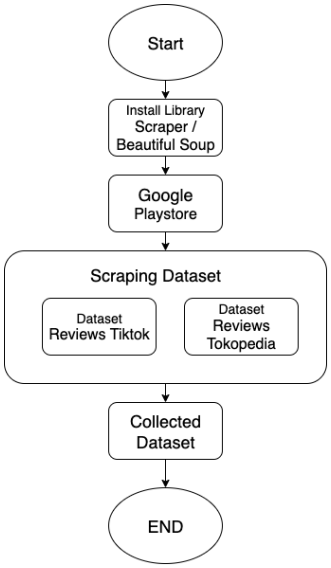

Fig.2.Workflow Of Data Collecting Review

B. Text Pre-processing

Text pre-processing plays a pivotal role in enhancing the quality and consistency of unstructured textual data, particularly in sentiment analysis involving user-generated content [19]. The pre-processing phase was designed to prepare review data from Tokopedia and TikTok for Explainable AI-based classification using the LIME method [20]. Two distinct cleaning strategies were employed: Tala Stopwords removal, which focuses on eliminating common but semantically weak Indonesian words, and manual cleaning, which targets noise such as emojis, URLs, and non-informative characters to ensure contextual clarity and data relevancy [21]. The structured overview of the pre-processing workflow is systematically presented in Fig.3, which reflects the robust pipeline developed to transform raw user reviews into analyzable inputs.

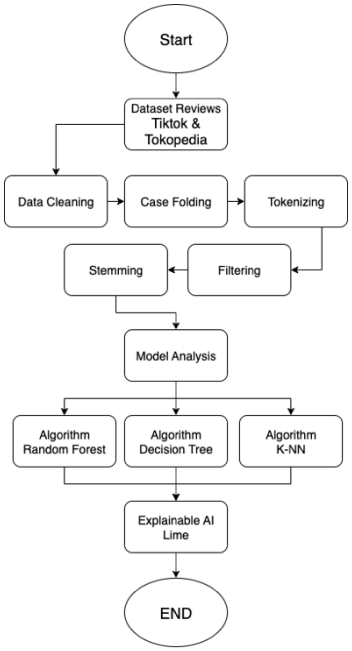

Fig.3.Workflow Of Text Pre-processing

*1) Cleaning: Data cleaning is the process of identifying and removin*g inaccurate, incomplete, or irrelevant elements within a dataset to enhance its quality and reliability [22]. In this Explainable AI study, which utilizes LIME on Tokopedia and TikTok reviews, data cleaning plays a crucial role in standardizing textual inputs and minimizing noise that may compromise the accuracy and interpretability of sentiment classification models [23].

2) Case Folding: A fundamental pre-processing step that standardizes textual data by converting all characters to lowercase, ensuring uniformity across the dataset.

3) Tokenizing: Tokenizing is the process of dividing a stream of text into individual units, commonly referred to as tokens, which represent words, phrases, or symbols. Tokenizing is an essential pre-processing step that enables the transformation of raw textual input into structured data suitable for analysis.

DOI : https://doi.org/10.25139/inform.v10i2.10748

4) Filtering: Filtering refers to the process of selecting and refining user review data to ensure its relevance and quality for sentiment analysis [24]. Two filtering techniques were applied: Tala Stopwords removal, which eliminates high-frequency but non-informative words in the Indonesian language, and manual cleaning, which targets irrelevant elements such as emojis, symbols, hyperlinks, and empty or overly short reviews [25].

5) Stemming: Stemming is an essential pre-processing step in sentiment analysis, aimed at reducing words to their root or canonical form to ensure uniformity in textual representation. Stemming was carried out using two approaches: the TALA stopword removal technique and manual cleaning, both designed to standardize each term into its correct base word in Bahasa Indonesia. This dual-method approach significantly improves the accuracy of sentiment classification by eliminating redundant morphological variations and aligning words with their formal dictionary equivalents.

C. Labelling Data

Each review from TikTok and Tokopedia included a user-generated rating score, which served as the basis for binary sentiment labelling positive for ratings of 4 and 5, and negative for ratings of 1 to 3. This labelling method enabled the conversion of unstructured textual feedback into clearly defined sentiment categories. The resulting dataset was stored in a structured format to support reproducible analysis and ensure a balanced distribution between positive and negative classes.

D. TF-IDF Weighting

The Term Frequency-Inverse Document Frequency (TF-IDF) weighting method is widely utilized to convert textual sentences into vector representations by quantifying the importance of words within a document [26]. This technique combines term frequency (TF) and inverse document frequency (IDF) to highlight words that carry significant weight in a review, filtering out commonly used but less informative terms. Relying solely on TF can be misleading, as frequent yet generic words may dominate the representation and distort the classification output [27]. Therefore, the IDF component is applied to balance word relevance, producing a more precise numerical representation, as formulated in Equation (1).

$$Tf.IDF = TF_{ij}\ X\ IDF_{ij} = TF_{ij} x\ log \frac{N}{DFj} \quad (1)$$

In Equation (1), which presents the TF-IDF calculation, N represents the total number of documents within the dataset, TF refers to the term frequency that measures how often a term appears in a document, and IDF denotes the inverse document frequency used to reduce the weight of commonly occurring terms across documents.

E.Model Analysis Data

This study's analysis compares the performance of RF, DT, and KNN classification algorithms on two datasets comprising user reviews from TikTok and Tokopedia. The objective is to identify the most accurate model for sentiment classification based on evaluation metrics. A detailed comparison of the algorithm results is presented in Fig.4.

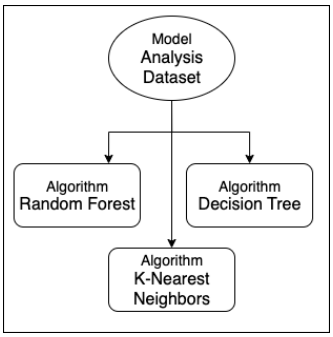

Fig.4.Workflow Of Model Analysis

1) Random Forest: Random Forest (RF) is an ensemble learning algorithm that builds multiple independent DT using bootstrap sampling and random feature selection at each node, resulting in robust classification performance. This method reduces overfitting and is particularly effective for handling large or imbalanced datasets [28]. The algorithm employs information gain calculations, as outlined in Equations 2-4, to determine the optimal attribute for data splitting. Info(D) represents the entropy of the dataset, while Info_A(D) quantifies the expected information needed to classify data points based on specific attributes. For continuous or numerical attributes, the algorithm identifies the most informative splitting point by sorting the data and evaluating each possible threshold.

$$Gain\ (A) = info\ (D) - Info\ (D) \quad (2)$$

$$Info\ (D) = \sum_{i=1}^{n} Pilog2\ (Pi) \quad (3)$$

$$Info\ A^{(D)} = \sum_{j=1}^{v} \frac{Dj}{D}\ x\ Info\ (Dj) \quad (4)$$

2) Decision Tree (DT): The DT classification algorithm employs a top-down strategy to assign data into predefined classes through a hierarchical tree-like structure. To determine the most effective data split, entropy and information gain are used as critical measures to evaluate the level of impurity or uncertainty in the dataset. These values are computed using mathematical formulas that quantify the discriminative power of each attribute. The DT theorem, as outlined in Equations (5) and (6), involves variables such as S for the set of all possible outcomes, i for individual outcomes, n for frequency or count, p_i for the probability of each outcome, and A representing the attribute under consideration.

$$Entropy\ (S) = \sum_{i=1}^{n} - pi.log_2 pi \quad (5)$$

$$Gain\ (S,A) = Entropy\ (S) - \sum_{i=1}^{n} \frac{|si|}{|s|} * Entropy\ (Si) \quad (6)$$

DOI : https://doi.org/10.25139/inform.v10i2.10748

3) K-Nearest Neighbours (KNN): KNN is a supervised classification algorithm that determines the class of a data point based on its similarity to its k nearest neighbours within the training dataset. The classification process involves calculating the distance between the test data and training instances using a metric such as Euclidean distance, as shown in Equation (7). Where *D* variable represents the computed distance, *x* denotes the training data, *y* the testing data, *n* is the number of attributes, and *i* indexes each attribute from 1 to *n*. The predicted class is assigned based on the most frequent category among the closest neighbours, providing a simple yet effective approach to pattern recognition.

$$D_{x,y} = \sqrt{\sum_{i=1}^{n} (xi - yi)^2} \quad (7)$$

F. Explainable AI With Lime

LIME (Local Interpretable Model-agnostic Explanations) is an interpretability technique that offers localized explanations for predictions generated by sentiment classification models, regardless of their underlying architecture. It achieves this by identifying key features within the input data that significantly influence the model's decision, thereby providing transparency in how a text is classified as positive, neutral, or negative. As defined in Equation (8), LIME seeks to approximate the complex model f with a simpler interpretable model g, focusing on a specific data instance x. This approximation is optimized by minimizing a loss function that measures the fidelity between f and g along with a complexity term $\Omega(g)$ to maintain interpretability. The interpretable models G typically consist of simpler structures like linear regressors or DT, which allow for clearer insight into the model's reasoning process.

$$\zeta(x) = argmin\ L(f, g, \pi_x) + \Omega(g) \quad (8)$$

III. Result and Discussion

This study conducts a comparative analysis by combining two user review datasets from TikTok and Tokopedia to evaluate the performance of sentiment classification. Three ML algorithms, RF, DT, and KNN, were applied to assess accuracy and interpretability. The analysis is enhanced using Explainable AI, specifically the LIME method, to generate transparent explanations of the classification results.

A. Data Collection

The data used in this study were derived from unstructured textual reviews posted by users on the TikTok Seller and Tokopedia Seller platforms over the past five years. Data collection was conducted using a Python-based scraping approach, resulting in a total of 28,541 reviews from TikTok Seller and 1,048,575 from Tokopedia Seller. These reviews contained free-form text content and numeric rating scores, offering a rich source of unstructured data essential for sentiment classification. All extracted information was systematically organized and saved in CSV format to support subsequent pre-processing, analysis, and model development stages.

B. Text Pre-processing

The data were obtained from user reviews on the Google Play Store, comprising unstructured textual content collected over the past five years. Before the classification process, the data underwent comprehensive pre-processing to ensure analytical relevance and structural consistency. Two filtering methods were employed: Tala Stopwords removal, which eliminates commonly used non-informative words in Bahasa Indonesia, and manual cleaning, which targets extraneous elements such as special characters, emojis, and irrelevant short reviews. These pre-processing techniques enhanced the dataset's integrity and optimized it for sentiment classification into positive, negative, and neutral categories.

TABLE I
TEXT PREPOCESSING EXPLAINABLE AI TIKTOK & TOKOPEDIA

Table Prepocessing	Before	After
Cleaning	*@!Gk bsa login dr tadi pagi*	*tidak bisa login dari tadi pagi*
Case Folding	*Aplikasi Ini Sangat Membantu*	*aplikasi ini sangat membantu*
Tokenizing	*Aplikasinya bagus banget dan mudah dipakai*	*["Aplikasinya", "bagus", "banget", "dan", "mudah", "dipakai"]*
Filtering	*fitur-fitur di dalamnya kurang lengkap dan lambat*	*fitur kurang lengkap lambat*
Stemming	*Aplikasinya membantu banget*	*aplikasi bantu banget*

Table I presents the outcomes of the text pre-processing stage. This process involves several steps, including cleaning, case folding, tokenization, filtering, and stemming, to optimize the quality of textual data. These procedures are essential to ensure the data is well-prepared before proceeding to the labelling phase.

C. Labeling Text

The Results of Table II sentiment labelling in this study are based on the star rating system provided by users. Reviews with 4 to 5 stars are categorized as positive, while those with 3 stars are labelled as neutral. Meanwhile, reviews receiving 1 to 2 stars are classified as negative to reflect user dissatisfaction.

TABLE II
LABELING TEXT EXPLAINABLE AI TIKTOK & TOKOPEDIA

Clean_Text Review	Score	Polarity
Memudahkan Penjual Online Khususnya Pemula	5	Sentiment Positive
Verifikasi Gagal Error Mulu	1	Sentiment Negative
Membantu Umkm Berkembang	3	Sentiment Neutral

D. Distribution Sentiment

The sentiment analysis of user reviews from TikTok and Tokopedia reveals a significant predominance of positive sentiment, accounting for 74.2% of the total dataset. Neutral sentiment follows with 24.1%, while negative sentiment

DOI : https://doi.org/10.25139/inform.v10i2.10748

constitutes only 1.7%. This disparity suggests a general tendency among users to express favourable impressions about the platforms' services or products. Such sentiment distribution underscores a high level of user satisfaction and affirms the positive perception of these e-commerce environments. The sentiment proportions are visually depicted through the bar and pie charts, as shown in Fig.5.

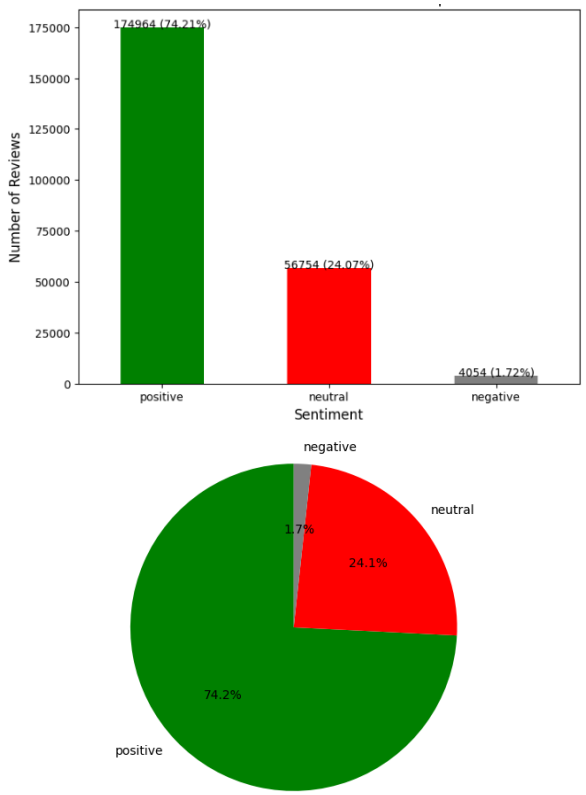

Fig.5. Sentiment Distribution of TikTok and Tokopedia Review Data

E. *Classification Model*

The classification performance on the combined TikTok and Tokopedia sentiment dataset emphasizes the superior capability of the RF model, which achieved the highest accuracy of 89.9% and a notable F1-score of 0.93 for the positive sentiment class, along with a macro average F1-score of 0.82. The DT model, utilizing the criterion min_samples_leaf and max_depth=None, achieved an accuracy of 88.25% and an F1-score of 0.92 for the positive class; however, its effectiveness declined in classifying negative sentiment, as indicated by a lower F1-score of 0.65. In comparison, the KNN classifier with n_neighbors=5 and weights='uniform' demonstrated the lowest accuracy at 81.51% and struggled significantly to detect negative reviews, reflected in its recall of only 0.19.

The classification process involved merging unstructured text data from both platforms and labelling them into three sentiment classes (positive, negative, and neutral). This was followed by vectorization using TF-IDF and model training with the specified algorithms. To enhance interpretability, the Explainable AI method LIME was applied after each classification step to reveal the influential features contributing to the predicted sentiment labels, thereby supporting transparent and explainable decision-making, as illustrated in Fig.6.

```
=== RANDOM FOREST ===
Accuracy: 0.8990
Report Classification:
              precision    recall  f1-score   support

     negatif       0.77      0.64      0.70       811
      netral       0.76      0.88      0.82     11351
     positif       0.96      0.91      0.93     34993

    accuracy                           0.90     47155
   macro avg       0.83      0.81      0.82     47155
weighted avg       0.91      0.90      0.90     47155


=== DECISION TREE ===
Accuracy: 0.8825
Report Classification:
              precision    recall  f1-score   support

     negatif       0.59      0.73      0.65       811
      netral       0.75      0.81      0.78     11351
     positif       0.94      0.91      0.92     34993

    accuracy                           0.88     47155
   macro avg       0.76      0.82      0.78     47155
weighted avg       0.89      0.88      0.88     47155


=== K-NEAREST NEIGHBORS ===
Accuracy: 0.8151
Report Classification:
              precision    recall  f1-score   support

     negatif       0.78      0.19      0.31       811
      netral       0.69      0.50      0.58     11351
     positif       0.84      0.93      0.88     34993

    accuracy                           0.82     47155
   macro avg       0.77      0.54      0.59     47155
weighted avg       0.80      0.82      0.80     47155
```

Fig.6. Model Performance

F. *Validation Model*

The evaluation results, obtained using the K-Fold Cross Validation technique on the combined review data from TikTok and Tokopedia, indicate that the RF model achieved the highest performance, with an average accuracy of 84.38% and a low standard deviation of 0.0089, reflecting strong consistency across the folds. The DT model achieved an average accuracy of 82.22%, although it exhibited greater variability with a higher standard deviation of 0.0186. In contrast, the KNN model recorded the lowest accuracy at 77.14%, despite maintaining a relatively low standard deviation of 0.0047. These findings are visually presented in Fig.7.

```
=== RANDOM FOREST ===
Average Accuracy: 0.8438
Standard Deviation:   0.0089
Score per Fold:     [0.843 0.852 0.854 0.829 0.841]

=== DECISION TREE ===
Average Accuracy: 0.8222
Standard Deviation:   0.0186
Score per Fold:     [0.818 0.807 0.848 0.799 0.839]

=== K-NEAREST NEIGHBORS ===
Average Accuracy: 0.7714
Standard Deviation:   0.0047
Score per Fold:     [0.767 0.778 0.767 0.769 0.776]
```

Fig.7. Validation Model K-Fold Cross Validation

DOI : https://doi.org/10.25139/inform.v10i2.10748

G. Explainable AI Lime

The Explainable AI (XAI) approach is implemented through the use of LIME (Local Interpretable Model-agnostic Explanations) to provide local interpretability of sentiment classification predictions. LIME generates feature-based explanations that highlight the influential variables driving each prediction, presented through informative visualizations such as LIME plots, explicit instance-level interpretations, and word clouds that emphasize dominant terms based on sentiment polarity. Additionally, scatter plots are used to visualize the distribution of predictions, while confusion matrices offer a detailed assessment of model performance by comparing predicted and actual labels.

The LIME Explicit Overview visualization for TikTok and Tokopedia review data illustrates how a complex model is interpreted locally using a simplified linear approximation. The purple dots represent local samples surrounding the instance explained, marked by a yellow '*X*', which are leveraged to construct the local LIME model, shown as a dashed green line. This linear boundary closely approximates the behaviour of the non-linear complex model within the vicinity of the target instance, as shown in Fig.8.

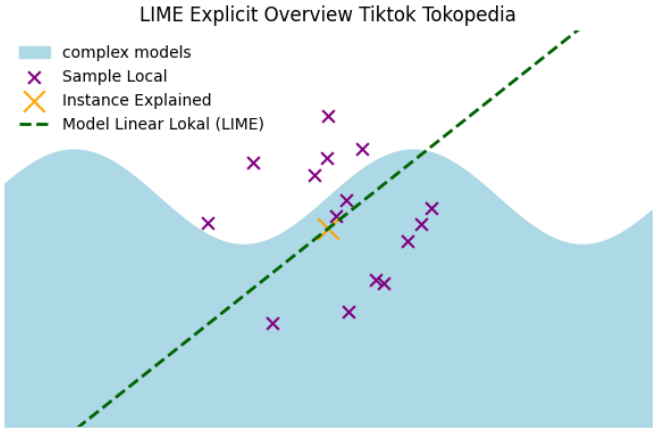

Fig.8. Visualization of Explainable AI Using LIME Explicit

The visualization of Explainable AI using LIME for positive sentiment classification on TikTok and Tokopedia reviews reveals consistent predictive behaviour across three models: RF, DT, and KNN. Each model identified influential keywords, such as "Tokopedia", "go", and "update", which contributed significantly to the classification outcome. Despite contextual ambiguity in the reviews, all models confidently predicted a positive sentiment, with highlighted terms confirming the role of brand mentions and product accessibility in shaping user satisfaction, as illustrated in Fig.9.

The LIME visualizations for negative sentiment classification of TikTok and Tokopedia reviews reveal distinct model sensitivities across RF, DT, and KNN classifiers. Keywords such as "biaya" (cost), "sistem" (system), "*ganti*" (replace), and "offline" were consistently highlighted, indicating user dissatisfaction related to service fees, technical restrictions, and account limitations. These interpretations suggest that all three models successfully identify negative sentiment drivers rooted in complaints about usability, costs, and access constraints, thereby validating the explainability and reliability of the sentiment classification pipeline, as illustrated in Fig.10.

The Explainable AI LIME visualization for neutral sentiment classification on TikTok and Tokopedia reviews demonstrates consistent predictions across the RF, DT, and KNN models. Key terms such as "*susah*" (difficult), "*masuk*" (login), and "*kesel*" (annoyed) were highlighted as influential, reflecting expressions of frustration that were still interpreted as neutral rather than overtly negative. This outcome suggests that the models prioritize the absence of explicitly polarised sentiment and consider ambiguity or mixed emotional cues as indicative of a neutral stance, as illustrated in Fig.11.

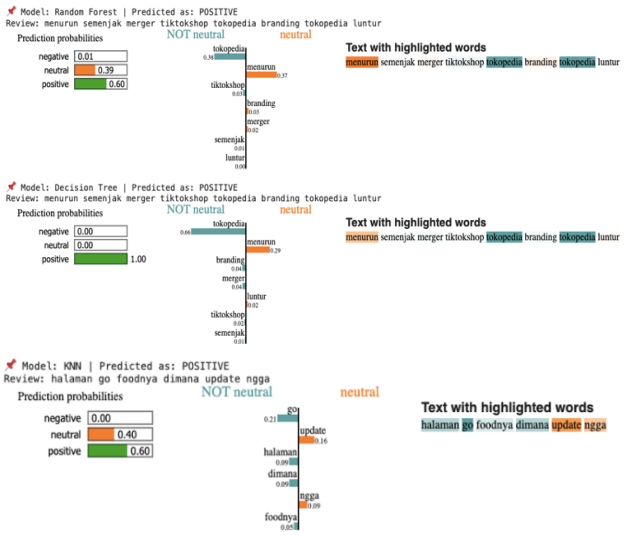

Fig.9. Explainable AI LIME Visualization of Positive Sentiment

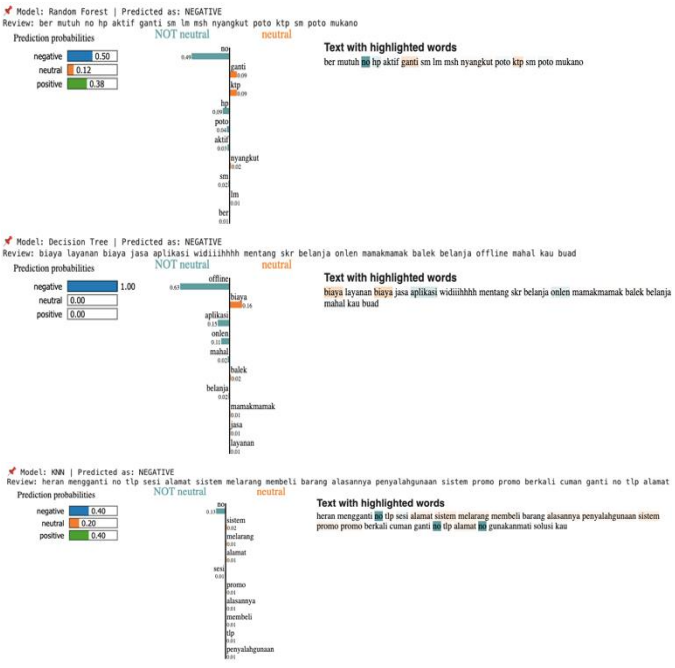

Fig.10. Explainable AI LIME Visualization of Negative Sentiment

DOI : https://doi.org/10.25139/inform.v10i2.10748

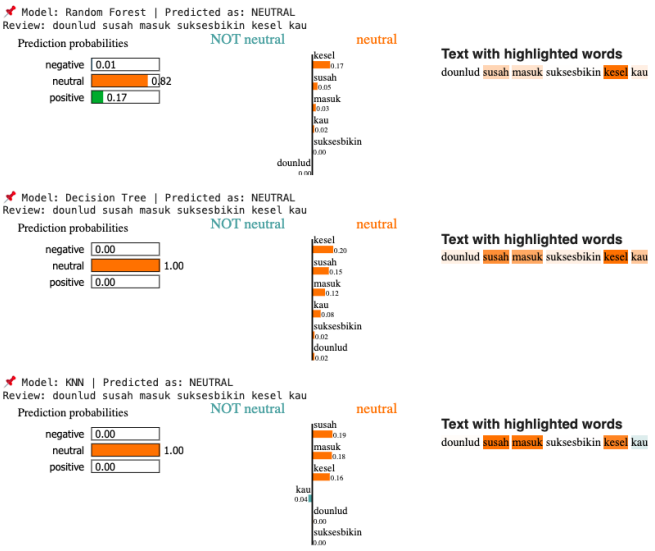

Fig.11. Visualization of Explainable AI LIME for Neutral Sentiment

H. Visualization Wordcloud

The word cloud visualization of positive sentiment in TikTok user reviews reveals the frequent appearance of terms such as “seller”, “good”, “product”, “helpful”, and “sales”, reflecting strong user appreciation for the platform’s service quality and selling experience. Words like “excellent”, “easy”, and “application” are also prominently featured, indicating favorable impressions regarding the app’s usability and system performance. Additionally, expressions such as “thank you”, “fast”, and “satisfying” reinforce the notion that TikTok provides added value for sellers through efficient transactions and customer satisfaction. This visualization clearly demonstrates a high level of satisfaction and a positive perception of TikTok Seller features, as shown in Fig.12.

The negative sentiment word cloud from TikTok user reviews highlights dominant terms such as “banned,” “product,” and “seller,” suggesting users’ concerns over account suspensions and product management issues. Keywords like “violation,” “account,” “TikTok,” and “verification” further emphasize dissatisfaction with the platform’s moderation and policy enforcement. Overall, this visualization reveals a sense of discomfort and distrust among users regarding TikTok Seller’s account review and regulatory systems, as illustrated in Fig. 13.

The word cloud visualization for neutral sentiment reviews on the TikTok application reveals frequently mentioned terms such as “*aplikasi*” (application), “*pelanggaran*” (violation), “*produk*” (product), and “seller”, indicating that users are describing operational issues or routine interactions without expressing strong positive or negative emotions. Words like “*masuk*”, “*udah*”, “*akun*”, and “susah” suggest that users are neutrally reporting experiences related to login difficulties, account access, or system functions. This pattern reflects a descriptive tone in user feedback, commonly used to convey suggestions, clarifications, or observations without explicit judgment, as shown in Fig. 14.

The word cloud visualization for positive sentiment reviews on the Tokopedia application highlights prominent terms such as “*gratis ongkir*” (free shipping), “*promo*”, “tokopedia”, and “bagus”, indicating high user appreciation for promotions and overall service quality. The frequent appearance of phrases like “*terima kasih*”, “*aplikasi*”, and “*toped*” reflects user satisfaction with the platform’s convenience and performance. Overall, these findings suggest that Tokopedia’s positive sentiment is largely driven by its cost-saving features, reliable service, and effective branding, as illustrated in Fig.15.

Fig.12. Word Cloud Visualization of Positive Sentiment in TikTok Review

Fig.13. Word Cloud Visualization of Negative Sentiment in TikTok Review

Fig.14. Word Cloud Visualization of Neutral Sentiment in TikTok Review

DOI : https://doi.org/10.25139/inform.v10i2.10748

Fig. 15. Word Cloud Visualization of Positive Sentiment in Tokopedia Review

The word cloud visualization for negative sentiment reviews on the Tokopedia application reveals dominant terms such as “error,” “cancel,” “*barang*” (product), and “*aplikasi*,” indicating user frustration with technical issues and transaction failures. Frequent mentions of “*kecewa*” (disappointed), “*transaksi*,” and “*akun*” suggest dissatisfaction with the platform’s reliability, particularly in order processing and account management. Overall, these keywords reflect recurring concerns among users regarding system errors, service interruptions, and unmet expectations, as illustrated in Fig.16.

The word cloud visualization of neutral sentiment reviews on the Tokopedia application highlights frequently used terms such as “*aplikasi*,” “tokopedia,” “*barang*,” and “*belanja*,” indicating descriptive or transactional mentions without strong emotional polarity. The presence of words like “*ribet*,” “*kecewa*,” and “*transaksi*” suggests that some users provided factual feedback about their experiences, possibly mixed with minor dissatisfaction or constructive criticism. Overall, these findings reflect a balanced expression of user opinions that neither strongly praise nor condemn the platform, but rather convey situational or procedural observations, as shown in Fig.17.

Fig.16. Word Cloud Visualization of Negative Sentiment in Tokopedia Review

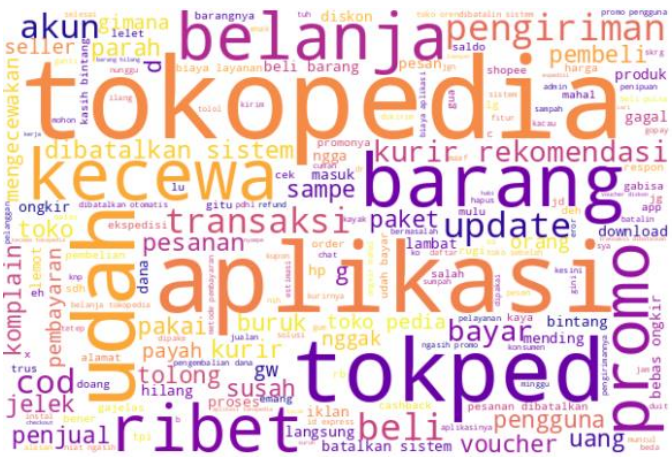

Fig.17. Word Cloud Visualization of Neutral Sentiment in Tokopedia Review

I. *Vizualization Network Graph*

The network graph visualization of the TikTok and Tokopedia review data highlights the central role of the word “Tokopedia” as a hub node, indicating its strong associative frequency with multiple relevant terms. Words such as “gratis,” “layanan,” “bagus,” “pengiriman,” “diskon,” and “cepat” appear prominently connected, reflecting key themes users frequently mention in their feedback. This structure reveals how user perception clusters around service quality, promotional offerings, and user experience, emphasizing the dominant presence of Tokopedia in the sentiment network. The graph effectively illustrates the semantic relationships and term importance in shaping user discussions, as shown in Fig.18.

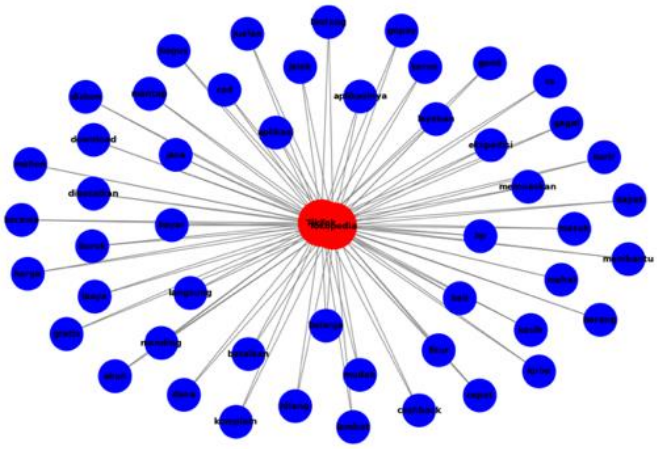

Fig.18. Network Graph Visualization of TikTok and Tokopedia Review Data

J. *Vizualisation ROC Curve*

The ROC Curve analysis for the sentiment classification task using unstructured review data from TikTok and Tokopedia confirms that the RF classifier outperforms other models, achieving excellent AUC scores of 0.88 for Negative (Class 0), 0.87 for Neutral (Class 1), and 0.88 for Positive (Class 2) sentiments. This result signifies the model's strong ability to discriminate between sentiment classes across both platforms

DOI : https://doi.org/10.25139/inform.v10i2.10748

consistently. The DT classifier demonstrated moderate accuracy, with AUC values ranging from 0.74 to 0.78. In contrast, the KNN classifier underperformed, particularly for the negative sentiment class, with a notably low AUC of 0.52, indicating near-random classification. These findings can be visually examined in Fig. 19 and are quantitatively summarized in Table III, highlighting the comparative performance of the models.

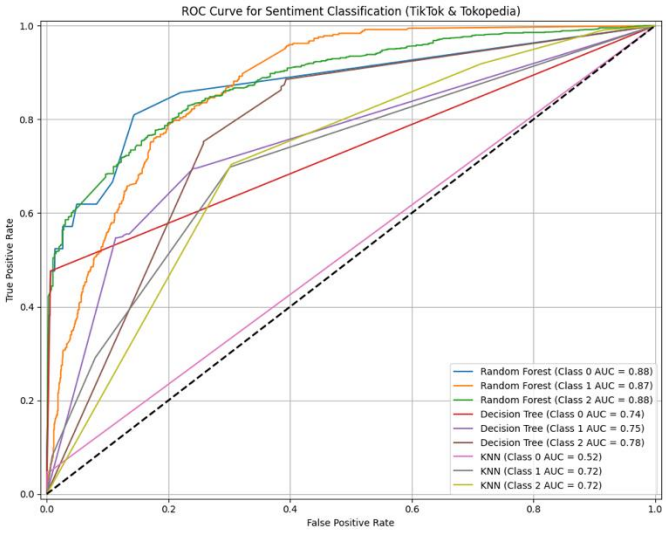

Fig.19. ROC Curve Visualization of Explainable AI LIME on TikTok and Tokopedia Reviews

TABEL III
ROC CURVE VISUALIZATION RESULTS FOR EXPLAINABLE AI USING LIME ON TIKTOK AND TOKOPEDIA DATASETS

Model Algorithm	Class 0 (Negative)	Class 1 (Neutral)	Class 2 (Positive)
RF	0,88	0,87	0,88
DT	0,74	0,75	0,78
KNN	0,52	0,72	0,72

IV. CONCLUSION

The comparative analysis of sentiment classification models using user review datasets from TikTok and Tokopedia demonstrated that the RF algorithm's output performed better than DT and KNN in terms of both predictive accuracy and interpretability. Specifically, RF achieved the highest accuracy of 89.9% and a macro-average F1-score of 0.82, supported by robust validation through K-Fold Cross-Validation with an average score of 84.38% and a low standard deviation (±0.0089), indicating high consistency. The application of LIME (Local Interpretable Model-Agnostic Explanations) provided clear, feature-based interpretations that clarified the classification rationale across all sentiment categories — positive, neutral, and negative for highlighting influential keywords such as "*tokopedia*," "*biaya*", and "*kesel*." Furthermore, visualization tools, including word clouds, ROC curves, and LIME plots, reinforced the model's transparency and its ability to reflect user perceptions accurately. These findings confirm that integrating Explainable AI with LIME significantly enhances the trustworthiness and analytical value of sentiment classification models in real-world e-commerce contexts.

REFERENCES

[1] A. Dzulkarnain, B. Rahmy Lidiawaty, S. Hidayati, R. Andi Hidayah, and A. Pramesta Setyaningtitah, "Sentiment Analysis of Student Complaint Text for Finding Context," *Inform : Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 9, no. 2, pp. 121–125, Jun. 2024, doi: 10.25139/inform.v9i2.7743.

[2] E. Yoshua and W. Maharani, "Depression Detection of Users in Social-Media Twitter Using Decision Tree with Word2Vec," *Inform : Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 9, no. 1, pp. 95–100, Feb. 2024, doi: 10.25139/inform.v9i1.7617.

[3] M. R. Adi Nugraha and Y. Sibaroni, "Classification of Depression Expressions on Twitter Using Ensemble Learning with Word2Vec," *Inform : Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 9, no. 1, pp. 67–74, Jan. 2024, doi: 10.25139/inform.v9i1.7559.

[4] M. Rusdi Rahman, A. Febri Diansyah, and H. Hanafi, "Sentiment Analysis on the Shopee Application on Playstore Using the Random Forest Classification Method," *Inform : Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 9, no. 1, pp. 20–24, Nov. 2023, doi: 10.25139/inform.v9i1.5465.

[5] A. R. Putra and D. E. Ratnawati, "Analisis Sentimen Berbasis Aspek pada Aplikasi Mobile Menggunakan Naïve Bayes berdasarkan Ulasan Pengguna Playstore (Studi Kasus : Jconnect Mobile)," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 12, no. 2, pp. 293–300, Apr. 2025, doi: 10.25126/jtiik.2025127556.

[6] F. Istighfarizky, N. A. Sanjaya ER, I. M. Widiartha, L. G. Astuti, I. G. N. A. C. Putra, and I. K. G. Suhartana, "Klasifikasi Jurnal menggunakan Metode KNN dengan Mengimplementasikan Perbandingan Seleksi Fitur," *JELIKU (Jurnal Elektronik Ilmu Komputer Udayana)*, vol. 11, no. 1, p. 167, Jul. 2022, doi: 10.24843/jlk.2022.v11.i01.p18.

[7] Yoga Religia, Agung Nugroho, and Wahyu Hadikristanto, "Klasifikasi Analisis Perbandingan Algoritma Optimasi pada Random Forest untuk Klasifikasi Data Bank Marketing," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 1, pp. 187–192, Feb. 2021, doi: 10.29207/resti.v5i1.2813.

[8] Riza Adrianti Supono and Muhammad Azis Suprayogi, "Perbandingan Metode TF-ABS dan TF-IDF Pada Klasifikasi Teks Helpdesk Menggunakan K-Nearest Neighbor," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 5, pp. 911–918, Oct. 2021, doi: 10.29207/resti.v5i5.3403.

[9] H. Nalatissifa, W. Gata, S. Diantika, and K. Nisa, "Perbandingan Kinerja Algoritma Klasifikasi Naive Bayes, Support Vector Machine (SVM), dan Random Forest untuk Prediksi Ketidakhadiran di Tempat Kerja," *Jurnal Informatika Universitas Pamulang*, vol. 5, no. 4, p. 578, Dec. 2021, doi: 10.32493/informatika.v5i4.7575.

[10] N. N. Marpid, Y. I. Kurniawan, and S. P. Rahayu, "Analysis Of The Movie Database Film Rating Prediction With Ensemble Learning Using Random Forest Regression Method," *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 1, pp. 1–10, Feb. 2025, doi: 10.52436/1.jutif.2025.6.1.1563.

[11] D. R. K. Saputra, Y. V. Via, and A. N. Sihananto, "Deteksi Anomali Menggunakan Ensemble Learning Dan Random Oversampling Pada Penipuan Transaksi Keuangan," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3, Aug. 2024, doi: 10.23960/jitet.v12i3.4910.

[12] N. Charibaldi, A. Harfiani, and O. Samuel Simanjuntak, "Comparison of the Effect of Word Normalization on Naïve Bayes Classifier and K-Nearest Neighbor Methods for Sentiment Analysis," *Inform : Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 9, no. 1, pp. 25–31, Dec. 2023, doi: 10.25139/inform.v9i1.7111.

[13] N. Rikatsih, M. Anshori, R. Siwi Pradini, and F. Faurika, "K-Nearest Neighbor Method for Early Detection of Diabetes Patients Based on Symptoms and Clinical Data," *Inform : Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 9, no. 2, pp. 187–193, Aug. 2024, doi: 10.25139/inform.v9i2.8582.

[14] C. M. Tri Yunanda, M. Hanafi, and W. M. Pradnya Dhuhita, "Sentiment Analysis on TikTok Shop Reviews Using Long Short-Term Memory Method to Find Business Opportunity," *Inform : Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 9, no. 1, pp. 1–7, Sep. 2023, doi: 10.25139/inform.v9i1.6524.

DOI : https://doi.org/10.25139/inform.v10i2.10748

[15] A. Maghfiroh, Y. Findawati, and U. Indahyanti, "Klasifikasi Penipuan pada Rekening Bank menggunakan Pendekatan Ensemble Learning," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 4, Mar. 2023, doi: 10.47065/bits.v4i4.3212.

[16] A. Nugroho and Y. Religia, "Analisis Optimasi Algoritma Klasifikasi Naive Bayes menggunakan Genetic Algorithm dan Bagging," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 3, pp. 504–510, Jun. 2021, doi: 10.29207/resti.v5i3.3067.

[17] Yoga Religia, Agung Nugroho, and Wahyu Hadikristanto, "Klasifikasi Analisis Perbandingan Algoritma Optimasi pada Random Forest untuk Klasifikasi Data Bank Marketing," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 1, pp. 187–192, Feb. 2021, doi: 10.29207/resti.v5i1.2813.

[18] I. L. Kharisma, D. A. Septiani, A. Fergina, and K. Kamdan, "Penerapan Algoritma Decision Tree untuk Ulasan Aplikasi Vidio di Google Play," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 9, no. 2, pp. 218–226, Sep. 2023, doi: 10.25077/teknosi.v9i2.2023.218-226.

[19] M. Reza, A. Dores, S. N. Ambo, and P. Meilina, "Perbandingan Metode Machine Learning Untuk Sentimen Analisis Review Penjualan Produk," 2025.

[20] R. Tsania, S. A. Putri, D. E. Ratnawati, and D. W. Brata, "Perbandingan Naïve Bayes dan K-Nearest Neighbor untuk Analisis Sentimen Aplikasi Gapura UB Berdasarkan Ulasan Pengguna pada Playstore," 2023.

[21] R. A. Khomeini Noor Bintang and N. T. Romadloni, "Perbandingan Kinerja Algoritma Klasifikasi Pada Review Pengguna Aplikasi Netflix," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 2, Apr. 2025, doi: 10.23960/jitet.v13i2.6303.

[22] M. F. Y. Herjanto and C. Carudin, "Analisis Sentimen Ulasan Pengguna Aplikasi Sirekap Pada Play Store Menggunakan Algoritma Random Forest Classifer," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 2, Apr. 2024, doi: 10.23960/jitet.v12i2.4192.

[23] N. B. Sidauruk, N. Riza, R. Nuraini, and S. Fatonah, "Penggunaan Metode Svm Dan Random Forest Untuk Analisis Sentimen Ulasan Pengguna Terhadap KAI Access Di Google Playstore," 2023.

[24] D. Saputro and D. Wahyu Utomo, "Rekomendasi Produk E-commerce Berbasis Klasifikasi Ulasan Menggunakan Ensemble Random Forest dan Teknik Boosting," vol. 15, no. 02, 2024, doi: 10.35970/infotekmesin.v15i2.2315.

[25] S. Syafrizal, M. Afdal, and R. Novita, "Analisis Sentimen Ulasan Aplikasi PLN Mobile Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 1, pp. 10–19, Dec. 2023, doi: 10.57152/malcom.v4i1.983.

[26] M. Iqbal, A. Davy Wiranata, R. Suwito, and R. Faiz Ananda, "KLIK: Kajian Ilmiah Informatika dan Komputer Perbandingan Algoritma Naïve Bayes, KNN, dan Decision Tree terhadap Ulasan Aplikasi Threads dan Twitter," *Media Online*, vol. 4, no. 3, pp. 1799–1807, 2023, doi: 10.30865/klik.v4i3.1402.

[27] D. E. Sondakh, S. W. Taju, M. G. Tene, and A. E. T. Pangaila, "Sistem Analisis Sentimen Ulasan Aplikasi Belanja Online Menggunakan Metode Ensemble Learning Sentiment Analysis System for Online Shopping Application Reviews Using Ensemble Learning Method," *Cogito Smart Journal* |, vol. 9, no. 2, 2023.

[28] C. G. Indrayanto, D. E. Ratnawati, and B. Rahayudi, "Analisis Sentimen Data Ulasan Pengguna Aplikasi MyPertamina di Indonesia pada Google Play Store menggunakan Metode Random Forest," 2023.

DOI : https://doi.org/10.25139/inform.v10i2.10748