Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 3, June 2026, Page. 2720-2733
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5659

Development of a Hybrid Machine Learning-Based E-Commerce Chatbot Using Jaccard Similarity and K-Nearest Neighbor for Accurate Intent Classification

Andrian Sah*[1], Andi Ilham[2], Rasna[3], Siti Nurhayati[4]

[1,2,3,4]Information Systems, Faculty of Computer Science, Universitas Yapis Papua, Indonesia

Email: [1]cyberdefance23@gmail.com

Received : Jan 29, 2026; Revised : Jan 31, 2026; Accepted : Feb 1, 2026; Published : Jun 15, 2026

Abstract

The advancement of technology in the e-commerce industry requires fast and accurate information services, particularly through the use of Natural Language Processing (NLP)-based chatbots. However, many existing chatbots rely on a single method, which often limits their ability to understand user question contexts effectively. This study proposes a hybrid approach integrating Jaccard Similarity and K-Nearest Neighbor (K-NN) to improve answer retrieval accuracy and intent classification in e-commerce chatbot systems. Jaccard Similarity is employed to measure the similarity between user queries and Frequently Asked Questions (FAQ) data, while K-NN is used to determine intent based on the nearest neighbor with the highest similarity values. The dataset, consisting of FAQ questions and answers, is preprocessed through case folding, tokenization, stopword removal, and stemming. System performance is evaluated using accuracy, precision, recall, and F1-score metrics. The experimental results show that Jaccard Similarity effectively selects relevant answer candidates, achieving similarity values of up to 66%, while K-NN produces stable intent classification results. The proposed hybrid model achieved an accuracy of 87%, precision of 86%, recall of 85%, and an F1-score of 85%, outperforming single-method implementations. Furthermore, confidence score analysis indicates that most chatbot responses fall into the high confidence category (>0.70). Rule-based NLP evaluation also provides insights into unclassified inputs, which can be used as a basis for future dataset development. The implementation results demonstrate that the chatbot system can be operated effectively on both customer and admin sides and monitored through analytical features. Overall, the proposed hybrid approach enhances the reliability, relevance, and stability of chatbot responses, making it a practical and effective solution for real-time intent classification and FAQ retrieval in e-commerce customer service environments.

Keywords : *Chatbot, E-Commerce, Hybrid Machine Learning, Jaccard Similarity, K-Nearest Neighbor.*

1. INTRODUCTION

The rapid development of digital technology has encouraged the increasing adoption of Natural Language Processing (NLP) in various digital services, one of which is chatbot technology that is capable of simulating human-like conversations automatically [1]. Recent studies show that NLP-based chatbots have been widely adopted in various domains such as education and customer service due to their interactive and scalable characteristics [2]. The flexibility of chatbot systems has also supported their implementation in different application contexts, including service-based information systems [3]. Chatbot implementation has also been widely adopted in tourism, academic services, and public information systems, demonstrating its adaptability across different domains [4]. In social media and public opinion analysis, NLP-based systems are increasingly used to process short and informal text efficiently [5]. In the e-commerce context, chatbots play a crucial role in improving customer service quality by providing fast and consistent responses related to product information, transactions, shipping, and payment processes [6], [7]. Recent implementations of customer service chatbots on popular messaging platforms such as WhatsApp further demonstrate the practical effectiveness of chatbot systems in supporting service excellence and operational efficiency in business environments [8]. In addition, chatbot intelligence has been shown to influence user trust and engagement in digital platforms [9].

E-commerce systems operating on web and mobile application platforms require chatbots that are not only responsive but also capable of accurately understanding the context of user queries [10]. The increasing volume of user interactions in e-commerce environments demands chatbot systems that can efficiently handle short and repetitive queries [2]. To meet these requirements, various machine learning approaches have been applied in the development of text-based chatbots to enhance text similarity measurement and intent classification capabilities [11]. Several chatbot development studies also emphasize the importance of systematic design and maintainable architecture to ensure long-term adaptability of NLP-based chatbot systems [12]. Various machine learning algorithms have been explored for text classification and sentiment analysis tasks, particularly in Indonesian-language datasets [13]. Performance evaluation of machine learning models shows that algorithm selection must consider data size, feature representation, and computational efficiency [14]. Supervised learning algorithms such as K-Nearest Neighbor and Support Vector Machine are commonly used for text classification tasks in Indonesian-language processing [15], [16]. Several studies also highlight that classification performance is highly influenced by dataset quality and variation [17].

Several methods commonly applied in text-based chatbot systems include K-Nearest Neighbor (K-NN), Naïve Bayes, and Jaccard Similarity. K-NN is widely used for classification based on data proximity and ease of implementation [18]. Its effectiveness in text-based classification has been validated in several NLP-related studies [16]. Jaccard Similarity is employed to measure text similarity based on word token overlap and set-based comparison [17]. This similarity approach is frequently used in recommendation and information retrieval systems [19]. These methods are frequently utilized in FAQ matching and intent classification tasks [20]. Other studies have compared similarity-based methods such as Jaccard and Cosine Similarity for information retrieval tasks, indicating that Jaccard performs well on short-text matching scenarios [21]. Additional studies on short-text FAQ matching confirm that similarity-based approaches remain effective for handling sparse user queries [22], [23]. In e-commerce systems, the combination of similarity-based approaches such as Jaccard and Cosine Similarity has also been proven effective for handling sparse and short-text data in recommendation and service-related applications [24]. In e-commerce service chatbots, comparative studies between K-NN and ensemble classifiers highlight the importance of filtering irrelevant responses early in the process [25]. However, most existing studies still apply Jaccard Similarity and K-NN independently, and only limited research has specifically analyzed the performance of integrating these two methods in e-commerce chatbot systems.

While traditional chatbot systems have demonstrated acceptable performance using individual methods such as K-Nearest Neighbor or Jaccard Similarity, each method has inherent limitations when applied independently. Jaccard Similarity relies primarily on lexical overlap and may fail to capture semantic variation [21]. This limitation becomes more evident in Indonesian-language user-generated text, which often contains informal expressions and contextual ambiguity, as highlighted in recent sentiment analysis studies on social media data [19]. Related studies on Indonesian short-text processing also emphasize the challenges of semantic variation in informal user input [23].

Meanwhile, K-NN performance is highly dependent on feature relevance and candidate selection [25]. Integrating both methods into a hybrid model offers a potential solution by combining efficient similarity filtering with stable intent classification. Previous hybrid-based studies reported improved response relevance and classification stability compared to single-method approaches [18]. However, this integration has not been sufficiently explored or quantitatively evaluated in the context of e-commerce chatbot systems, particularly for FAQ-based short-text interactions.

P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5659

Recent sentiment analysis studies also emphasize that traditional machine learning models may struggle with semantic variation when dealing with dynamic user-generated text [12]. Advanced deep learning approaches such as CNN-based models have shown promising results but often require larger datasets and higher computational resources [21]. Based on this gap, this study aims to develop a Hybrid Machine Learning-based e-commerce chatbot by integrating Jaccard Similarity as a text similarity measurement method and K-Nearest Neighbor (K-NN) as an intent classification method, as well as to evaluate its performance in producing more accurate and relevant responses. The main contributions of this study are as follows:

1. proposing a hybrid machine learning mechanism that integrates Jaccard Similarity for text similarity measurement and K-Nearest Neighbor (K-NN) for intent classification in an FAQ-based e-commerce chatbot;
2. presenting a quantitative performance evaluation of the proposed hybrid approach using accuracy, precision, recall, and F1-score metrics, and comparing it with single-method implementations; and
3. analyzing confidence scores as an additional indicator to assess the stability and reliability of chatbot responses.

These contributions are expected to enrich research on machine learning-based e-commerce chatbot development and to provide a practical solution to the limitations of single-method approaches in short-text processing. This also supports scalable customer service automation in real-world e-commerce environments.

2. METHOD

This research method describes the stages involved in developing a hybrid machine learning-based e-commerce chatbot that integrates Jaccard Similarity and K-Nearest Neighbor (K-NN). The chatbot system is designed to automatically process user questions through the integration of Natural Language Processing (NLP) and machine learning techniques. In general, a chatbot is a computer program that interacts with users using natural language, and its development has evolved since the early introduction of chatbot intelligence in the 1960s [22]. Modern chatbot systems commonly combine NLP and machine learning approaches to simulate human-like conversations, enable simultaneous user interactions, ensure continuous availability, and reduce manual response requirements [23], [2].

The proposed chatbot architecture consists of several main components, including user input reception, text preprocessing, similarity calculation using Jaccard Similarity, intent classification using K-Nearest Neighbor (K-NN), and an FAQ database as the primary answer source. The research process begins with FAQ dataset collection, followed by text preprocessing to prepare the data for similarity measurement and intent classification. Jaccard Similarity is applied to calculate text similarity values between user queries and FAQ entries. The similarity results are then used to select the most relevant FAQ candidates, which serve as input features for the K-NN algorithm to perform intent classification. K-NN classifies the user intent based on the proximity of the selected candidates, and the outputs of both methods are integrated in a hybrid stage to determine the most relevant chatbot response.

System performance is evaluated using accuracy, precision, recall, and F1-score metrics to assess the effectiveness of the proposed hybrid approach compared to single-method implementations. The overall research workflow, from user input to chatbot response generation, is illustrated in Figure 1. Figure 1 shows that the process begins with text input and preprocessing, followed by similarity measurement using Jaccard Similarity to select relevant FAQ candidates. These candidates are then classified using the K-Nearest Neighbor (K-NN) algorithm to determine the most appropriate intent before generating the final chatbot response. This workflow ensures that irrelevant responses are filtered

Vol. 7, No. 3, June 2026, Page. 2720-2733
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5659

at an early stage while maintaining classification accuracy for short-text interactions in e-commerce chatbot systems.

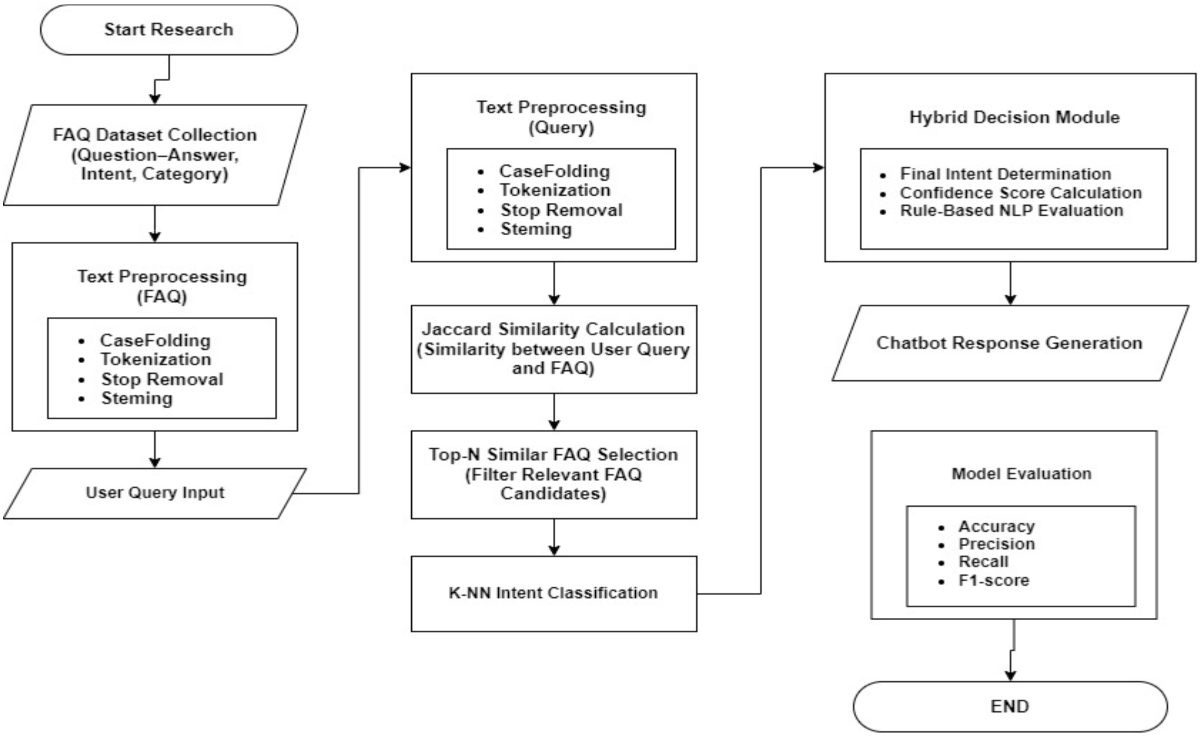

Figure 1. Research Flow Chart

2.1. Dataset Preparation and Text Preprocessing

Dataset quality plays a crucial role in determining the performance of machine learning models, as models can only learn valid patterns when the data is clean, balanced, and accurately labeled. Datasets containing noise, class imbalance, or domain-irrelevant information tend to reduce model generalization capability and lead to lower prediction accuracy. Previous studies emphasize that proper data preparation and preprocessing are essential steps before machine learning training is conducted [3]. Similarly, research on IndoBERT-based NLP models reports that data imbalance and quality variation significantly affect model performance, indicating the need for representative and well-prepared datasets [2].

The importance of dataset preparation is also highlighted in studies on RASA-based chatbots for tourism applications, which state that data understanding, text cleaning, and intent formulation are fundamental stages, particularly when datasets originate from internal FAQ sources with diverse formats [9]. In addition, research applying the K-Nearest Neighbor algorithm to chatbot data shows that classification performance is strongly influenced by text structure, label consistency, and preprocessing quality [15]. These findings confirm that the effectiveness of NLP-based chatbot systems, including those utilizing Jaccard Similarity and K-NN, highly depends on how the dataset is structured and prepared.

In this research, the dataset consists of Frequently Asked Questions (FAQs) collected from e-commerce systems, comprising question–answer pairs along with intent and category labels. This dataset is divided into training and testing data and is used as input for similarity calculation using Jaccard Similarity and intent classification using K-NN. The structure of the dataset is illustrated in Table 1. Table 1 presents sample FAQ entries including user questions, corresponding system responses, intent labels, and category classifications used to support similarity measurement and intent classification processes.

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 3, June 2026, Page. 2720-2733
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5659

Table 1. Chatbot FAQ Data Example

Question	Answer	Intent	Category
How to change address?	Please go to the profile menu and change the address	address_shipping	general
How long is delivery time?	Jabodetabek: 1–2 days, outside Jabodetabek: 3–7 days	delivery_time	shipping
How to return goods?	Contact customer service within 7 days with a photo	return_item	general
Is there a warranty for the product?	Yes, every product has a warranty. Please check	warranty	product
How to track an order?	You can track orders in the Transaction History menu	track_order	shipping
What payment methods are available?	We accept COD and Bank Transfer	payment_method	payment

To ensure data readiness for similarity measurement and classification, text preprocessing is applied to the entire dataset. Text preprocessing aims to transform unstructured text into a more standardized and machine-readable form, thereby improving feature extraction efficiency and classification accuracy [4]. In this study, preprocessing is conducted through several stages. Case folding is applied to convert all characters to lowercase in order to eliminate inconsistencies caused by capitalization differences [16]. Tokenization is then performed to split text into individual word tokens, which is particularly important for calculating word intersections and unions in Jaccard Similarity [5].

Furthermore, text cleaning is carried out to remove noise such as punctuation and irrelevant characters, resulting in cleaner and more meaningful text [5]. Stopword removal is applied to eliminate commonly occurring words that do not contribute significantly to semantic meaning, thereby enhancing the effectiveness of similarity calculation and classification processes [13]. Finally, stemming is used to reduce words to their base forms by removing prefixes and suffixes, minimizing word variation and producing a more compact and consistent text representation [24].

All preprocessing steps are applied consistently to both training and test data to ensure uniform text structure. This integrated dataset preparation and preprocessing process plays a vital role in supporting the performance of the proposed hybrid machine learning model that combines Jaccard Similarity and K-Nearest Neighbor for e-commerce chatbot development.

2.2. Jaccard Similarity

Jaccard Similarity, also known as the Jaccard Index or Jaccard Coefficient, is a metric used to measure the similarity between two sets based on shared elements [8]. In text processing, this method is commonly applied to compare document similarity using word tokens. The Jaccard Similarity is mathematically defined in Equation (1):

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \tag{1}$$

where A and B represent two word sets, $A \cap B$ denotes the number of common words (intersection), and $A \cup B$ represents the union of all words contained in both sets [17]. Previous studies indicate that Jaccard Similarity is effective for short text matching tasks, such as FAQ retrieval, due to its simplicity, computational efficiency, and direct token-based similarity measurement [8]. Therefore, in this research, Jaccard Similarity is applied to calculate similarity scores between user queries and

Vol. 7, No. 3, June 2026, Page. 2720-2733
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5659

FAQ questions, which are then utilized as input for intent classification using the K-Nearest Neighbor (K-NN) method.

2.3. K-Nearest Neighbor (K-NN)

The K-Nearest Neighbor (K-NN) algorithm is a non-parametric supervised learning method originally introduced by Fix and Hodges in 1951 [12]. K-NN classifies new data by assigning the majority class among its K nearest neighbors within the training dataset [19]. The proximity between data instances is calculated using Euclidean Distance, as defined in Equation (2):

$$d(x_{i,}x_d) = \sqrt{\sum_{k=1}^{m}(x_i(a_k) - x_d(a_k))^2} \qquad (2)$$

where x_i and x_d represent data vectors, a_k denotes the k-th feature, and m is the total number of features [19]. In this research, K-NN is applied to perform intent classification based on text feature vectors generated from the preprocessing stage. The value of K is selected to achieve a balance between classification accuracy and decision stability. K-NN is chosen because of its simplicity, effectiveness for short text data, and the absence of a complex training process, making it suitable for integration with Jaccard Similarity in a hybrid chatbot approach.

2.4. Hybrid Integration of Jaccard Similarity and K-NN

The integration of Jaccard Similarity and K-Nearest Neighbor (K-NN) in the chatbot system is designed to enhance the accuracy of matching user queries with relevant answers. Jaccard Similarity is employed as an initial similarity measurement to calculate word-based similarity scores between user questions and FAQ data based on the intersection and union of token sets. This set-based approach has been proven effective for short text retrieval tasks due to its simplicity and computational efficiency [14].

The similarity scores generated by Jaccard Similarity are subsequently utilized as input features for the K-NN algorithm to perform intent classification using a nearest neighbor-based approach. Previous studies indicate that K-NN produces stable classification results when applied to structured and representative text features [12], [25]. By filtering candidate answers through Jaccard Similarity before classification, K-NN operates on more relevant data, thereby improving decision stability and classification accuracy.

Overall, the hybrid Jaccard and K-NN approach combines efficient token-based similarity measurement with robust proximity-based classification, resulting in a more accurate FAQ retrieval process. This hybrid model is particularly suitable for e-commerce chatbot systems that handle short and concise user queries.

2.5. Evaluation and Testing

Model evaluation is conducted to assess the accuracy and consistency of the proposed hybrid Jaccard Similarity and K-Nearest Neighbor (K-NN) method in generating chatbot responses. The evaluation process employs accuracy, precision, recall, and F1-score metrics, which are commonly used in NLP and text classification research [11].

The evaluation setup considers factors that influence classification performance, including dataset structure, label consistency, and the selection of the K value, as reported in previous studies [21]. These considerations ensure that the evaluation results accurately reflect the performance of the hybrid approach.

In this research, the dataset is divided into training and testing subsets to support model generalization. The predicted results are compared with ground-truth labels using a confusion matrix, from which evaluation metric values are derived. This evaluation aims to measure the effectiveness of

Vol. 7, No. 3, June 2026, Page. 2720-2733
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5659

integrating Jaccard Similarity and K-NN compared to single-method implementations in improving chatbot response accuracy.

3. RESULT

3.1. Dataset Preprocessing

Text preprocessing was conducted to standardize text formats and reduce noise so that similarity calculations become more stable. This preprocessing stage included case folding, tokenization, stopword removal, and stemming. The preprocessing results are presented in Table 2. Table 2 illustrates examples of FAQ questions before and after preprocessing, showing how raw user queries are transformed into simplified token representations through stemming to support similarity measurement and intent classification.

Table 2. Example of FAQ Dataset Preprocessing Results

No	Original Question	Tokenization Results (Stemming)
1	How to change address?	change address
2	How long is delivery time?	delivery time
3	How to return goods?	return of goods
4	Is there a warranty for the product?	product warranty
5	How to track an order?	track orders
6	What payment methods are available?	payment method

The preprocessing process produces concise token representations without altering the original meaning of the text. By reducing lexical variation and noise, this preprocessing stage improves the reliability of Jaccard Similarity calculations and supports more stable intent classification using the K-Nearest Neighbor (K-NN) algorithm.

3.2. Jaccard Similarity Calculation

Jaccard Similarity was applied to calculate the similarity between user queries and FAQ data as an initial filtering stage. The similarity calculation involved token set generation, intersection computation, union computation, and similarity value determination. For example, the user query *"Cara melacak pesanan"* (How to track an order) and the FAQ question *"Bagaimana cara melacak pesanan?"* (How to track an order) produced token sets {how, track, order} and {track, order}. The intersection contained two tokens, while the union contained three tokens, resulting in a similarity value of 0.66. This value indicates a high level of contextual similarity between the two sentences.

The quality of text preprocessing plays a crucial role in the effectiveness of the similarity calculation and intent classification stages. Case folding, stopword removal, and stemming contribute to reducing vocabulary variation and textual noise, which leads to more consistent token representations. As a result, Jaccard Similarity is able to produce more reliable similarity scores, particularly for short and repetitive user queries commonly found in e-commerce environments. Without proper preprocessing, similarity values tend to be lower and less stable, which negatively affects the downstream K-NN classification process.

3.3. K-NN Intent Classification

After similarity calculation, intent classification was performed using the K-Nearest Neighbor (K-NN) algorithm with a k value of 3. The system selected the three FAQ entries with the highest similarity values as nearest neighbors, as shown in Table 3. Table 3 presents the candidate FAQ entries ranked by their Jaccard Similarity scores, which are used as input for the K-NN algorithm to determine the most relevant user intent.

Vol. 7, No. 3, June 2026, Page. 2720-2733
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5659

Table 3. K-NN Candidates Based on Jaccard Similarity Values

FAQ	Similarity	Label Intent
"How to track an order?"	0.66	track_order
"How long is delivery time?"	0.20	delivery_time
"How to return goods?"	0.14	return_item
"What payment methods are available?"	0.00	payment_method

Based on the classification results, the final intent was determined as track_order, since it had the highest similarity score among the selected neighbors. This result indicates that the hybrid filtering process effectively narrows down relevant candidates, allowing K-NN to produce a stable and accurate intent classification.

3.4. Hybrid Model Performance

The performance of the hybrid Jaccard Similarity and K-NN model was evaluated using accuracy, precision, recall, and F1-score. Table 4 summarizes the performance comparison between the single-method approaches (Jaccard Similarity and K-NN) and the proposed hybrid method across all evaluation metrics.

Table 4. Performance Comparison of Single Methods and the Hybrid Method

Method	Accuracy	Precision	Recall	F1-Score
Jaccard Similarity	0.78	0.76	0.75	0.75
K-Nearest Neighbor (K-NN)	0.81	0.80	0.79	0.79
Hybrid Jaccard + K-NN	0.87	0.86	0.85	0.85

The hybrid approach achieved the highest performance across all evaluation metrics compared to single-method implementations. As shown in Table 4, the hybrid model outperformed Jaccard Similarity and K-NN individually, achieving an accuracy of 0.87, precision of 0.86, recall of 0.85, and an F1-score of 0.85. The superior performance of the hybrid Jaccard Similarity and K-NN model can be attributed to the complementary roles of both methods in the classification process. Jaccard Similarity acts as an effective filtering mechanism by eliminating irrelevant FAQ entries at an early stage, thereby reducing noise in the input data. By limiting the candidate set to only the most contextually relevant FAQs, the subsequent K-NN classification process becomes more focused and accurate. This reduction in the search space allows K-NN to perform intent classification more efficiently, which directly contributes to the observed improvements in accuracy, precision, recall, and F1-score. In contrast, when Jaccard Similarity or K-NN is applied individually, each method exhibits inherent limitations. Jaccard Similarity relies solely on lexical overlap and does not consider class distribution, which may lead to incorrect intent assignment in cases of similar vocabulary across different intents. Meanwhile, K-NN without prior filtering is more susceptible to noisy or less relevant data, especially in short-text chatbot queries. The hybrid approach overcomes these limitations by combining similarity-based filtering with distance-based classification, resulting in more stable and reliable performance.

3.5. Rule-Based NLP Evaluation

A rule-based NLP mechanism was implemented to evaluate unclassified inputs. This evaluation focused on out-of-context messages, fallback responses, and short messages. Table 5 presents the evaluation indicators used in the rule-based NLP mechanism, along with their corresponding descriptions.

Vol. 7, No. 3, June 2026, Page. 2720-2733
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5659

Table 5. Rule-Based NLP Evaluation on Chatbot

Evaluation Indicator	Description
Out-of- Context	Message not relevant to e-commerce context
Fallback	Message failed to be classified and produces default response
Short Messages	Message too short (≤ 3 words) thus reducing text understanding context

The implementation of rule-based NLP evaluation provides additional insight into chatbot limitations that are not fully addressed by machine learning methods alone. As shown in Table 5, out-of-context messages refer to inputs that are unrelated to the e-commerce domain, fallback messages represent cases where the system fails to determine an intent and produces a default response, and short messages indicate inputs with very limited textual context. Out-of-context messages and very short inputs frequently result in low similarity scores, making intent classification unreliable. By explicitly identifying these cases through rule-based evaluation, the system can prevent incorrect responses and provide fallback messages instead. Furthermore, the analysis of unclassified inputs serves as valuable feedback for improving dataset coverage, indicating which types of user queries should be added to future FAQ expansions to further enhance chatbot performance.

3.6. System Chatbot Implementation

The chatbot system was implemented with two main interfaces: a customer-side interface and an admin-side interface. Figure 2 illustrates the customer-side chatbot interface, which provides an interactive chat window where users can submit questions and receive automatic responses generated by the hybrid Jaccard Similarity and K-NN model. Through this interface, users can directly interact with the chatbot to obtain information related to products, orders, payments, and shipping services.

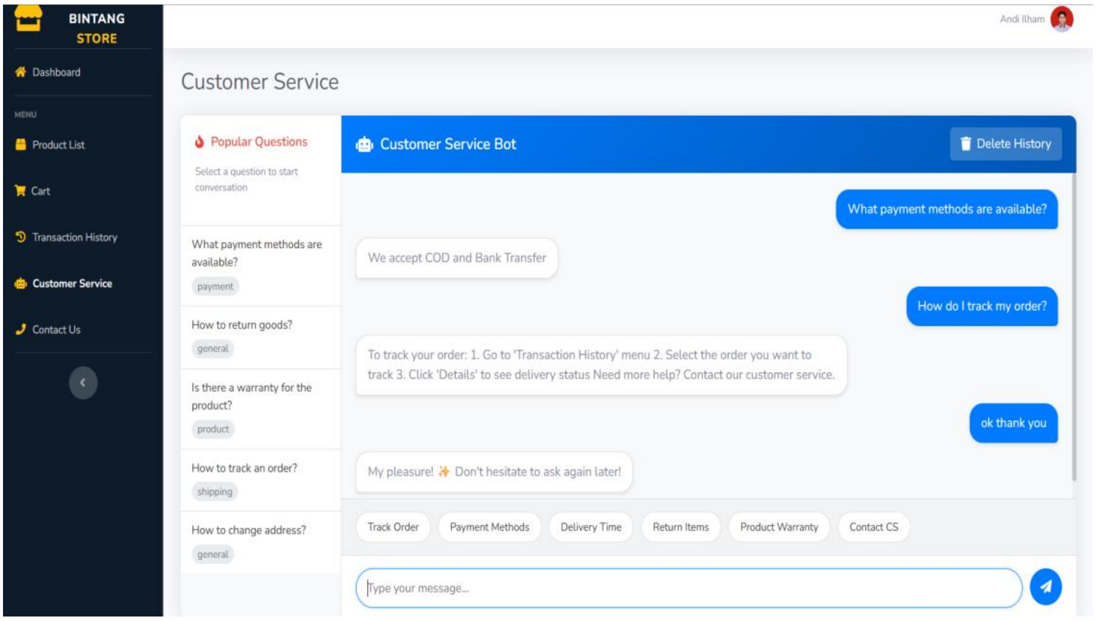

Figure 2. Customer Service Chatbot Interface on Customer Side

The admin-side interface is designed to manage the FAQ dataset. Through this interface, administrators can add, update, and delete FAQ entries to ensure response accuracy. Figure 3 presents the FAQ management interface on the admin side, showing features for dataset maintenance, including question–answer editing, intent labeling, and category management. This interface allows administrators to continuously refine the chatbot knowledge base.

Vol. 7, No. 3, June 2026, Page. 2720-2733
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5659

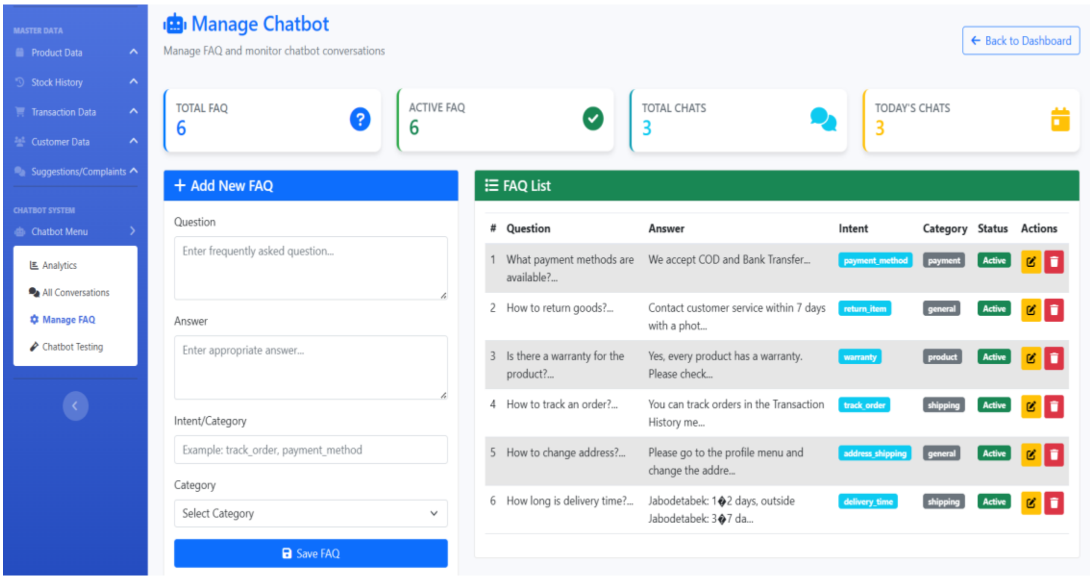

Figure 3. FAQ Management Interface on Admin Side

In addition to dataset management, the system provides analytical features to monitor chatbot performance, including confidence score distribution, fallback responses, and short-message occurrences. Figure 4 shows the chatbot performance analysis interface on the admin side, which visualizes chatbot interaction statistics and evaluation indicators to support performance monitoring and system improvement. These analytical features help administrators identify system weaknesses and guide future enhancements.

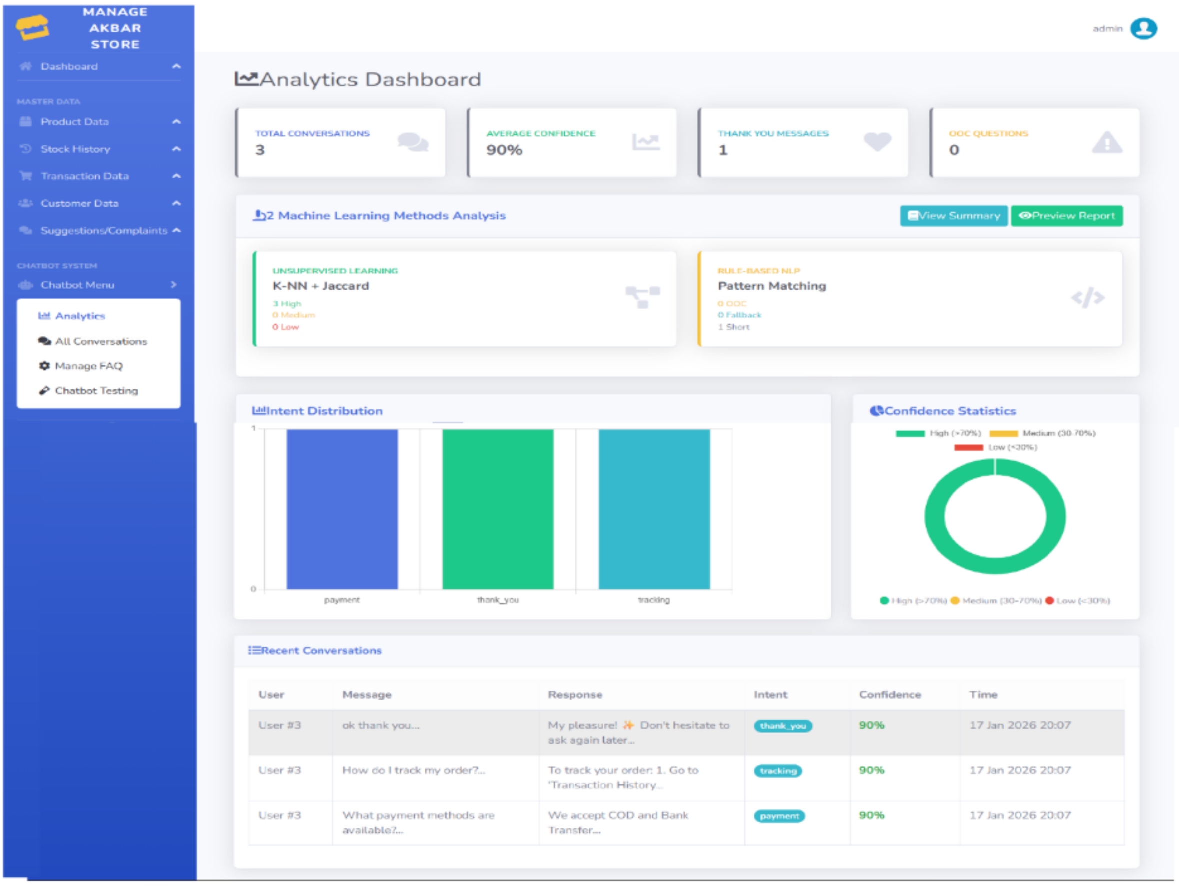

Figure 4. Chatbot Performance Analysis Display on Admin Side

Vol. 7, No. 3, June 2026, Page. 2720-2733
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5659

With the implementation of both customer and admin interfaces, the chatbot system is capable of providing automatic responses while allowing continuous monitoring and improvement of system performance.

4. DISCUSSION

The experimental results indicate that the integration of Jaccard Similarity and K-Nearest Neighbor (K-NN) provides significant improvements in chatbot response accuracy compared to single-method approaches. Jaccard Similarity effectively filters irrelevant FAQ entries by measuring lexical similarity, while K-NN strengthens intent determination by classifying only the most relevant candidates. This finding is consistent with previous studies which reported that hybrid approaches are able to combine the strengths of multiple text processing methods to improve classification performance and response relevance in FAQ-based chatbot systems [26], [27], [20].

When compared with other chatbot development techniques commonly used in e-commerce environments, such as rule-based systems or single-method machine learning models, the proposed hybrid approach demonstrates clear advantages. Rule-based chatbots rely heavily on predefined patterns and lack flexibility when handling variations in user queries, while single-method approaches such as standalone K-NN or similarity-based matching are more sensitive to noise and limited contextual representation. In contrast, the hybrid Jaccard Similarity and K-NN model reduces the search space through similarity-based filtering and improves classification stability through distance-based learning, resulting in higher intent classification accuracy and more relevant responses.

The confidence score analysis further demonstrates that the hybrid model produces stable responses for queries with strong contextual similarity. Low confidence scores were mainly associated with short messages, out-of-context inputs, or variations not sufficiently represented in the dataset. Compared to systems that rely solely on accuracy-based evaluation, the incorporation of confidence score analysis provides an additional layer of interpretability, allowing the system to identify uncertain predictions and reduce the risk of incorrect responses. This finding highlights the importance of confidence-based evaluation in improving chatbot reliability, especially in customer service applications where incorrect answers may negatively impact user trust.

Additionally, the rule-based NLP mechanism acts as a safeguard for handling ambiguous or unrecognized inputs. Beyond preventing incorrect responses, this mechanism provides valuable feedback for identifying dataset weaknesses and guiding future dataset enrichment. The combination of confidence score evaluation and rule-based fallback responses contributes to a more robust chatbot system and supports continuous improvement through dataset refinement. From a scalability perspective, the proposed hybrid model has the potential to be extended to handle more complex queries and larger datasets. Future research may explore the integration of multilingual datasets, as well as the incorporation of more advanced NLP techniques such as word embeddings or transformer-based models (e.g., BERT or GPT) to further enhance contextual understanding. Overall, the discussion confirms that the hybrid approach effectively overcomes the limitations of single-method implementations by combining similarity measurement and intent classification. This makes the proposed model suitable for e-commerce chatbot systems, which typically handle short, direct, and repetitive user queries.

5. CONCLUSION

This research successfully implemented a hybrid model based on Jaccard Similarity and K-Nearest Neighbor (K-NN) for an e-commerce chatbot system. Based on experimental results and analysis, the proposed hybrid approach was able to improve the relevance of chatbot responses by utilizing similarity values as an initial filtering stage prior to intent classification. Jaccard Similarity proved effective in selecting the most relevant answer candidates, while the K-NN algorithm provided

Vol. 7, No. 3, June 2026, Page. 2720-2733
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5659

stable intent classification results for questions with similar contextual patterns. The confidence score analysis indicated that most chatbot responses fell into the high confidence category (> 0.70), demonstrating that the system was able to generate reliable and relevant answers. In addition, the rule-based NLP evaluation revealed several question patterns that could not be optimally classified, particularly short or out-of-context inputs. These findings provide valuable insights for future dataset enrichment and system refinement.

Furthermore, the implementation of the chatbot interfaces on both customer and admin sides showed that the system is easy to operate and supports continuous performance monitoring through analytical features. This implementation enables administrators to manage datasets effectively and evaluate chatbot behavior in real time. From a practical perspective, this hybrid chatbot can be directly applied to e-commerce platforms to improve customer service efficiency, reduce response time, and support automated handling of frequently asked questions.

Overall, the results of this study demonstrate that integrating Jaccard Similarity and K-NN is an effective approach for improving the accuracy and stability of text-based chatbot responses in e-commerce environments. Future research should focus on deploying the proposed system in real-world e-commerce environments to evaluate real-time performance, scalability, and user interaction behavior. In addition, further studies may explore the use of larger and more diverse datasets, as well as the integration of embedding-based or transformer-based NLP models to better handle complex language variations and improve semantic understanding.

REFERENCES

[1] H. A. Mumtahana and R. Pamungkas, "Architecture Design Development of e-Learning for Primary School Learning in Madiun City," *INTENSIF: Jurnal Ilmiah Penelitian dan Penerapan Teknologi Sistem Informasi*, vol. 6, no. 2, pp. 285–292, Aug. 2022, doi: 10.29407/intensif.v6i2.18276.

[2] N. R. P. Syallya, A. A. Pravitasari, and A. Helen, "NLP-Based Intent Classification Model for Academic Curriculum Chatbots in Universities Study Programs," *Jurnal RESTI*, vol. 9, no. 1, pp. 111–117, Feb. 2025, doi: 10.29207/resti.v9i1.6276.

[3] A. P. Segara, M. Andryan, W. Saputra, and N. A. Ranggianto, "Perbandingan Performa Algoritma Random Tree, K-NN, dan A-NN untuk Deteksi Serangan DDoS pada Software Defined Network (SDN) Article Info ABSTRAK," *JSAI : Journal Scientific and Applied Informatics*, vol. 08, no. 2, 2025, doi: 10.36085/jsai.

[4] V. Agustina and A. Herliana, "Analisis Sentimen Publik atas Kebijakan Efisiensi Anggaran 2025 dengan Text Mining dan Natural Language Processing, Jurnal Media Informatika (JUMIN)," *JURNAL MEDIA INFORMATIKA [JUMIN]*, vol. 6, no. 3, pp. 2182–2194, 2025, doi: doi.org/10.55338/jumin.v6i3.6301.

[5] A. W. Saputra, H. Setiawan, and R. Dijaya, "Analisis Sentimen Layanan Perwalian Mahasiswa UMSIDA Menggunakan Metode Support Vector Machine (SVM)," *JSAI: Journal Scientific and Applied Informatics*, vol. 8, no. 1, 2025, doi: 10.36085/jsai.v8i1.7387.

[6] B. Wijaya, "Implementasi Chatbot Dalam Sistem E-Commerce Menggunakan Natural Language Processing Dengan Metode Extreme Programming," *Jurnal Ilmiah Komputasi*, vol. 24, no. 2, Jun. 2025, doi: 10.32409/jikstik.24.2.3806.

[7] H. Y. Panjaitan, Y. Masnita, and K. Kurniawati, "Penambahan Variabel Tingkat Kecerdasan dari Chatbot untuk Mempengaruhi Kepercayaan Pengguna dalam Aplikasi Telekonsultasi Kesehatan," *J. Sistem Info. Bisnis*, vol. 13, no. 1, pp. 78–87, Sep. 2023, doi: 10.21456/vol13iss1pp78-87.

[8] M. R. Waskito, A. D. Rahajoe, and A. L. Nurlaili, "Implementasi Metode Collaborative Filtering Menggunakan Algoritma Cosine Similarity Dan Jaccard Similarity Pada Sistem E-Commerce," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3S1, Oct. 2024, doi: 10.23960/jitet.v12i3s1.5315.

[9] Z. H. Pradana, H. Nafi'ah, and R. A. Rochmanto, "Chatbot-based Information Service using RASA Open-Source Framework in Prambanan Temple Tourism Object," *Jurnal RESTI*, vol. 6, no. 4, pp. 656–662, Aug. 2022, doi: 10.29207/resti.v6i4.3913.

[10] M. Khusnah, R. Gernowo, and B. Surarso, "Implementasi E-Commerce dengan Sistem Informasi Rekomendasi menggunakan Metode Collaborative Filtering untuk Pengembangan Penjualan pada UMKM," *Jurnal Sistem Informasi Bisnis*, vol. 15, no. 1, pp. 134–141, 2025, doi: 10.14710/vol15iss1pp135-142.

[11] R. S. Nurhalizah, R. Ardianto, and P. Purwono, "Analisis Supervised dan Unsupervised Learning pada Machine Learning: Systematic Literature Review," *Jurnal Ilmu Komputer dan Informatika*, vol. 4, no. 1, pp. 61–72, Aug. 2024, doi: 10.54082/jiki.168.

[12] J. A. Firdaus, A. Setia Budi, and E. Setiawan, "Analisis Performa Algoritma Machine Learning Pada Perangkat Embedded Atmega328p," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)* , vol. 10, no. 2, pp. 245–254, 2023, doi: 10.25126/jtiik.2023106196.

[13] A. A. Kurniawan and M. Mustikasari, "Evaluasi Kinerja Mllib Apache Spark Pada Klasifikasi Berita Palsu Dalam Bahasa Indonesia," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 9, no. 3, 2022, doi: 10.25126/jtiik.202293538.

[14] Y. Imanuel, S. J. Pinastika, H. Soepandi, and K. Lebai, "Analisis Komparatif Jaccard Dan Cosine Similarity Untuk Pencarian Ayat Alkitab," *SINTA Jurnal Sistem Informasi dan Teknologi Komputasi*, vol. 2, no. 4, pp. 171–177, 2025, doi: https://doi.org/10.61124/sinta.v2i4.102.

[15] Sujacka Retno, Rozzi Kesuma Dinata, and Novia Hasdyna, "Evaluasi model data chatbot dalam natural language processing menggunakan k-nearest neighbor," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 4, no. 1, pp. 146–153, Apr. 2023, doi: 10.37859/coscitech.v4i1.4690.

[16] D. A. Prabowo and S. Sudianto, "Analisis Sentimen Sepak Bola Indonesia pada Twitter menggunakan K-Nearest Neighbors dan Random Forest," *JSAI : Journal Scientific and Applied Informatics*, vol. 06, no. 02, 2023, doi: 10.36085/jsai.v6i2.5337.

[17] M. M. Amin, A. Firdaus, and Y. Dwitayanti, "Model Rekomendasi Jurnal dengan Algoritma Jaccard Similarity dan Protokol OAI-PMH," *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 4, no. 10, pp. 489–499, Feb. 2025, doi: 10.52436/1.jpti.637.

[18] S. Sandiwarno, T. B. A. Hartanto, and E. Pitaloka, "Penerapan Machine Learning Untuk Prediksi Bencana Banjir," *Jurnal Sistem Informasi Bisnis*, vol. 14, no. 1, pp. 62–76, Jan. 2024, doi: 10.21456/vol14iss1pp62-76.

[19] A. Tarisa Akbar, N. Yudistira, and A. Ridok, "Identifikasi Gagal Ginjal Kronis Dengan Mengimplementasikan Metode Support Vector Machine Beserta K-Nearest Neighbour (SVM-KNN)," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)* , vol. 10, no. 2, pp. 301–308, 2023, doi: 10.25126/jtiik.2023106059.

[20] C. D. Sasongko, R. Isnanto, and A. P. Widodo, "Review of Systematic Literature about Sentiment Analysis Techniques," *Jurnal Sistem Informasi Bisnis*, vol. 15, no. 2, pp. 227–236, Jun. 2025, doi: 10.14710/vol15iss2pp227-236.

[21] H. Handoko, A. Asrofiq, J. Junadhi, and A. S. Negara, "Sentiment Analysis of Sirekap Tweets Using CNN Algorithm," *INTENSIF: Jurnal Ilmiah Penelitian dan Penerapan Teknologi Sistem Informasi*, vol. 8, no. 2, pp. 312–329, Aug. 2024, doi: 10.29407/intensif.v8i2.23046.

[22] N. B. Puspitasari, A. Noventi, A. B. Ginting, A. A. Fransiskus, and D. Saribu, "Bot Innovation Realizing Service Excellence: Designing a WhatsApp Chatbot as a Customer Service Solution Using the Waterfall Method," *Jurnal Sistem Informasi Bisnis*, vol. 15, no. 2, pp. 251–259, 2025, doi: 10.14710/vol15iss2pp251-259.

[23] L. Anindyati, "Analisis Dan Perancangan Aplikasi Chatbot Menggunkan Framework Rasa Dan Sistem Informasi Pemeliharaan Aplikasi (Studi Kasus: Chatbot Penerimaan Mahasiswa Baru Politeknik Astra)," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 10, no. 2, pp. 291–300, Apr. 2022, doi: 10.25126/jtiik.2022106409.

[24] A. R. Sembiring and C. K. Dewa, "Sentiment Analysis On Indonesian Tweets about the 2024 Election," *Sinkron*, vol. 9, no. 1, pp. 413–422, Jan. 2025, doi: 10.33395/sinkron.v9i1.14481.

Vol. 7, No. 3, June 2026, Page. 2720-2733
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5659

[25] F. Zamakhsyari *et al.*, "Comparison of KNN and Random Forest Algorithms on E-Commerce Service Chatbot," *Jurnal Informatika Sunan Kalijaga (JISKA)*, vol. 10, no. 1, pp. 100–109, 2025, doi.org/10.14421/jiska.2025.10.1.100-109.

[26] V. R. Prasetyo, M. F. Naufal, and K. Wijaya, "Sentiment Analysis of ChatGPT on Indonesian Text using Hybrid CNN and Bi-LSTM," *Jurnal RESTI*, vol. 9, no. 2, pp. 327–333, Apr. 2025, doi: 10.29207/resti.v9i2.6334.

[27] B. Kommey, E. Tamakloe, D. Opoku, T. Crispin, and J. Danquah, "Disease Detection in Tropical Tomato Leaves via Machine Learning Models," *Jurnal ELTIKOM*, vol. 8, no. 2, pp. 179–191, Dec. 2024, doi: 10.31961/eltikom.v8i2.1340.