Comparative Analysis of Deep Learning Models for Retrieval-Based Tourism Information Chatbots

Dwi Intan Af'idah, Dairoh, Sharfina Febbi Handayani
Polytechnic of Harapan Bersama, Mataram Street No.9, Tegal and 52147, Indonesia

ARTICLE INFO

Article history:

Received November 30, 2024
Revised January 31, 2025
Accepted March 13, 2025

Keywords:

Chatbot;
Deep Learning Models;
Tourism Information;
Retrieval-Based System

ABSTRACT

Despite significant advancements in deep learning models for chatbots, comprehensive analyses tailored to the tourism sector remain limited. This study addresses the gap by comparing the performance of six prominent models—MLP, RNN, GRU, LSTM, BiLSTM, and CNN—in creating chatbots designed to address traveler needs such as information about facilities, ticket prices, activity suggestions, and operational details. The methodology includes key stages such as data collection, preparation, model training, and evaluation using accuracy, precision, recall, F1-score, and qualitative assessments. The dataset, derived from interviews with managers of 11 tourism destinations, captures critical details to replicate real-world user interactions. The results indicate that the CNN model performed the best, achieving the highest accuracy (0.98), precision (0.99), recall (0.98), and F1-score (0.98), showcasing its ability to effectively handle user queries by identifying relevant patterns in data. While MLP achieved strong accuracy (0.94), its simpler design limited its capacity to manage complex questions. The RNN model had the lowest accuracy (0.82), highlighting its challenges in understanding structured information. These findings confirm CNN as the most effective model for retrieval-based chatbots in tourism, balancing accuracy and practicality. This research offers valuable insights for improving AI-driven tourism tools, providing guidelines for selecting optimal models and enhancing chatbot performance to enrich the traveler experience.

Corresponding Author:

Dairoh, Polytechnic of Harapan Bersama, Mataram Street No.9, Tegal and 52147, Indonesia
Email: dairoh@poltektegal.ac.id

1. INTRODUCTION

The advancement of digital technology has brought significant transformations across various sectors, including the tourism industry [1], [2]. In this era, travelers demand fast, accurate, and relevant access to information to support their trip planning [3]. Artificial intelligence (AI)-powered chatbots have emerged as a crucial innovation addressing these needs [4], [5]. Acting as digital guides, chatbots provide information on destinations, transportation schedules, ticket prices, activity recommendations, and accommodation suggestions—all without requiring human interaction [6], [7], [8]. The research background is that chatbots have become essential tools for enhancing user experiences by delivering efficient and responsive information [9], [10], [11].

Retrieval-based chatbots, one of the most widely implemented AI technologies, are designed to select the best response from a predefined dataset based on user input [12]. In the context of tourism, this approach is particularly relevant as travelers often require specific and structured information in a short time [13], [14]. The primary advantages of retrieval-based chatbots are their consistency and ability to deliver answers based on recognizable question patterns [15], [16], [17]. For example, retrieval-based chatbots can provide detailed information about transportation routes or activity schedules without requiring complex text processing. The development of deep learning technology has further elevated retrieval-based chatbots by enhancing their data analysis and natural language processing (NLP) capabilities [18], [19]. Models such as Multilayer Perceptron (MLP), Recurrent Neural Network (RNN), Gated Recurrent Unit (GRU), Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), and Convolutional Neural Network (CNN) have been implemented to maximize chatbot performance in understanding and responding to user inputs [20].

Multilayer Perceptron (MLP), a widely used artificial neural network model, is effective for processing structured data but less suitable for sequential data. MLP faces limitations in handling long data sequences, particularly in text processing or tasks requiring temporal dependencies [21], [22]. To address this limitation, Recurrent Neural Networks (RNNs) were introduced as a more suitable alternative for processing sequential data. However, RNNs encounter challenges such as vanishing gradients when handling very long data sequences, which can reduce performance in practical applications [23], [24].

To overcome the shortcomings of RNNs, Gated Recurrent Units (GRUs) have emerged as a solution with a more efficient architecture, simplifying the network structure while maintaining the ability to capture long-term dependencies. Recent studies have shown that GRUs perform exceptionally well in applications such as chatbots, where processing efficiency is critical [25], [26]. Meanwhile, Long Short-Term Memory (LSTM) networks have further advanced the field by enabling long-term information retention through unique gating mechanisms [27], [28], [29]. Additionally, Bidirectional LSTM (BiLSTM), which processes data in both forward and backward directions, enhances LSTM's ability to capture broader contexts, making it an ideal choice for retrieval-based chatbots that require deeper understanding of question patterns [30]. On the other hand, Convolutional Neural Networks (CNNs), though traditionally used in image processing, have demonstrated remarkable performance in short-text classification. By leveraging convolutional filters to extract critical features from text, CNNs have proven their utility in applications like short-text-based chatbots [31], [32].

The existing studies often focus on individual model performance without providing a comprehensive comparative analysis of deep learning models within the specific context of retrieval-based chatbots for tourism. This gap in the literature arises from the highly domain-specific nature of chatbot applications, where diverse user needs and varying data structures pose challenges for model generalization. Moreover, prior research rarely evaluates these models in terms of their ability to handle the diverse and dynamic queries typical in tourism information services. Addressing this gap is essential for identifying the most effective deep learning model tailored to the unique demands of the tourism sector.

This study aims to fill this gap by conducting a comparative analysis of six deep learning models - MLP, RNN, GRU, LSTM, BiLSTM, and CNN – in the development of retrieval-based chatbots for tourism information services. The methodology involves dataset preparation, preprocessing, model training, and evaluation using both quantitative metrics (e.g., accuracy, precision, recall, and F1-score) and qualitative assessments, providing a comprehensive understanding of their performance. By employing a systematic and multi-faceted evaluation approach, this research offers several key strengths that contribute to its scientific merit.

One of the primary strengths of this study is its comprehensive comparison of diverse deep learning models. By analyzing multiple architectures, including traditional (MLP), sequential (RNN, GRU, LSTM, BiLSTM), and feature-extraction-based (CNN) models, this research provides a nuanced understanding of their performance within the specific context of tourism chatbots. Additionally, the integration of qualitative assessments alongside quantitative metrics offers a more holistic evaluation of chatbot effectiveness, capturing not only accuracy and precision but also user-centric dimensions like response relevance and contextual appropriateness.

Despite these contributions, the study acknowledges certain limitations that could influence the findings and their broader applicability. First, the dataset used for model training and evaluation is relatively small, comprising data collected from in-depth interviews with managers of 11 tourism destinations. While this dataset provides rich domain-specific insights, its size may limit the generalizability of the results to broader tourism contexts or larger datasets that encompass more diverse linguistic and cultural nuances. Second, the qualitative assessments, though essential for evaluating user satisfaction and response quality, are inherently subjective. The reliance on evaluator judgments introduces potential biases, as perceptions of chatbot performance may vary across individuals. These subjective evaluations could impact the consistency and reliability of the findings, particularly when scaling the analysis to larger systems or different user groups.

To balance these limitations, the study emphasizes transparency in its methodology and results. By explicitly discussing the challenges associated with dataset size and subjective evaluations, this research seeks to provide a balanced perspective that enhances its scientific rigor. Future research is encouraged to build on these findings by employing larger and more diverse datasets, as well as incorporating standardized and objective evaluation frameworks, to further validate and expand the applicability of the results. Ultimately, this study combines its strengths-comprehensive model comparisons and a multi-dimensional evaluation framework-with a clear acknowledgment of its limitations, providing actionable insights for improving chatbot accuracy, responsiveness, and user satisfaction. In doing so, it bridges critical gaps in the literature while contributing to the development of AI-driven technologies for the tourism sector.

Theoretical contributions of this research lie in its systematic comparison of deep learning models within a domain-specific application. While prior studies have explored individual models, they often lack a direct comparative analysis tailored to the complex and dynamic requirements of tourism chatbots. This study bridges that gap by providing actionable insights into the strengths and limitations of different architectures, offering researchers a clearer understanding of which models excel under specific conditions. By doing so, it contributes to the broader body of knowledge on artificial intelligence and natural language processing in domain-specific contexts.

In terms of practical contributions, the study provides critical insights for the tourism sector by demonstrating how optimal chatbot models can enhance the delivery of information and improve user satisfaction. Specifically, the findings offer guidelines for selecting deep learning architectures that can effectively handle diverse and dynamic user queries. These insights are expected to help tourism stakeholders, such as destination managers and technology developers, implement AI-driven solutions that address traveler needs more effectively, thereby improving the competitiveness of tourism destinations. Furthermore, the study highlights the potential for chatbots to act as digital assistants, streamlining trip planning and enriching traveler experiences by providing accurate and contextually relevant information.

Overall, this study seeks to fill critical gaps in the literature by conducting a domain-specific analysis that balances theoretical rigor with practical relevance. Its findings contribute to advancing both the academic understanding of retrieval-based chatbots and the practical implementation of AI technologies in the tourism sector, ultimately supporting the development of more effective and user-centered intelligent systems.

2. METHODS

This research aims to analyze the performance of six deep learning models in the context of retrieval-based chatbots for the tourism sector. The methodology involves several stages: dataset preparation, preprocessing, model training, testing, and evaluation. Each stage is elaborated as Fig. 1.

Fig. 1. Research Stage

2.1. Dataset Preparation

Dataset preparation is a fundamental step in developing retrieval-based chatbots for the tourism sector. The dataset for this study was collected through in-depth interviews with managers of 11 priority tourism destinations in Tegal, Indonesia. This decision was driven by the study's focus on exploring localized chatbot applications in the tourism sector. Tegal was chosen due to its diverse range of tourism offerings, including natural attractions, cultural sites, and culinary experiences, providing a microcosm for examining chatbot effectiveness. This initial scale was deemed sufficient to pilot the comparative analysis of deep learning models. This approach allowed for an in-depth exploration of the methodology within resource constraints.

These sites were selected as they represent key attractions, including natural, cultural, and recreational tourism, providing a comprehensive view of the region's tourism landscape. This localized focus aligns with the study's aim of exploring chatbot applications tailored to specific regional contexts. Standardized questionnaires ensured consistent data collection, covering aspects such as destination descriptions, facilities, ticket prices, and activity recommendations. Efforts were made to minimize interviewer bias through training and adherence to structured interview protocols. Ethical considerations were thoroughly addressed to ensure adherence to research standards. Informed consent was obtained from all participants involved in data collection through in-depth interviews. Participants were briefed on the research objectives, the voluntary nature of their participation, and the measures in place to protect confidentiality. All data were anonymized and securely stored to prevent unauthorized access, ensuring compliance with ethical guidelines and data privacy regulations in research involving human subjects.

The success of retrieval-based chatbots heavily depends on the dataset's quality and structure. Previous research has emphasized the importance of well-organized datasets in creating responsive and accurate chatbots, especially in tourism applications. For instance, studies have shown that categorizing information into tags and developing specific question patterns can improve chatbot response relevance by up to 40% compared to conventional approaches [33]. The collected data was organized in JSON format with tags, patterns, and responses to enhance the chatbot's ability to provide accurate and contextually relevant answers. Tags categorize data into specific groups, such as descriptions, facilities, ticket prices, transportation, and activities. Patterns consist of sample questions relevant to each category, designed using a multiple QA approach to enhance the chatbot's response variety and accuracy. This structure supports efficient training and improves response accuracy.

2.2. Dataset Preprocessing

This is how to start another subsection. Preprocessing is a crucial stage in preparing the dataset for use in deep learning models, particularly for developing retrieval-based chatbots in the tourism sector. The goal is to transform raw data into a format that can be efficiently processed by the model. Key preprocessing steps in this study include lowercasing, tokenization, and lemmatization. Lowercasing converts all text in the dataset to lowercase [34], [35]. This process is vital for reducing word variability, which can influence model performance, especially in text-based models such as chatbots. For example, the words "Tiket" and "tiket" would be considered distinct entities without lowercasing. By standardizing all words to lowercase, consistency is improved, making it easier for the model to identify and process similar words. According to research [36], lowercasing is a fundamental step in natural language processing (NLP) that simplifies data complexity without losing essential information. It minimizes variations in word representation, making the model more efficient in learning relationships between words with similar meanings, regardless of capitalization.

Tokenization splits text into smaller units called tokens [37]. In retrieval-based chatbots, tokenization breaks down long sentences or paragraphs into manageable chunks, such as words or phrases, enabling the model to identify key elements in the text relevant to user queries. Tokenization is typically performed using methods based on spaces or punctuation. For example, the sentence "What is the ticket price for the National Park?" would be tokenized into ["What", "is", "the", "ticket", "price", "for", "the", "National", "Park"]. This process is crucial for simplifying textual data while retaining meaningful information (Automated Chatbot Implemented Using Natural Language Processing). Tokenization can be performed at various levels, including word-level and character-level tokenization. In chatbot applications, word-level tokenization is often preferred as it better handles conversational context.

Lemmatization reduces words to their base form (lemma) [38]. Unlike stemming, which only removes suffixes, lemmatization considers the meaning of words to return them to their lexical root. For example, the word "menerbangkan" (to fly) is reduced to "terbang" (fly), and "lebih baik" (better) becomes "baik" (good). Lemmatization is critical in retrieval-based chatbots as it reduces unnecessary word variations, enabling the model to recognize different forms of words with the same meaning. Research [39] has shown that lemmatization can enhance NLP model accuracy by ensuring all variations of words with identical meanings are treated as the same entity. The lemmatization process uses lemmatizers that rely on linguistic rules and dictionaries to identify a word's base form. Tools like NLTK (Natural Language Toolkit) in Python automate lemmatization across datasets, simplifying analysis and model training

2.3. Model Training

The training phase is a crucial stage in developing retrieval-based chatbots for the tourism sector using deep learning. In this study, six selected deep learning models MLP (Multilayer Perceptron), RNN (Recurrent Neural Network), GRU (Gated Recurrent Unit), LSTM (Long Short-Term Memory), BiLSTM (Bidirectional LSTM), and CNN (Convolutional Neural Network) were trained using the preprocessed dataset. The selection of models in this study was based on their established effectiveness in handling Natural Language Processing (NLP) tasks, particularly for retrieval-based chatbot systems. The CNN model was chosen for its ability to capture local patterns in text, while RNN, GRU, LSTM, and BiLSTM were selected for their strengths in modeling sequential data and long-term dependencies. MLP was included as a baseline for comparison due to its simplicity and computational efficiency.

MLP, as one of the simplest models, is used for text-based classification in retrieval chatbots by leveraging feature representations obtained from tokenization and lemmatization. It operates by connecting inputs through multiple hidden layers using non-linear activation functions, ultimately producing outputs that correspond to prediction labels based on a loss function minimized during training. Research findings [40] indicate that MLP is effective in handling static data without complex temporal dependencies. However, MLP struggles with long sequence data or complex conversational contexts, a limitation addressed by more advanced models.

In contrast, RNN is better suited for sequential data, such as conversational inputs in chatbots, due to its ability to retain information from previous steps through its hidden state mechanism, which is updated with each new token. RNN training employs the backpropagation through time (BPTT) algorithm, enabling the model to update network weights based on information from the entire data sequence. Studies (RNN Language

Processing Model-Driven Spoken Dialogue System Modeling Method) show that RNN effectively handles conversational contexts with clear temporal dependencies. However, RNN often encounters issues with training on long sequences due to vanishing gradients, a challenge that GRU seeks to address. The previous research proposes a deep learning approach to build an intelligent chatbot that utilizes a combination of Bidirectional Recurrent Neural Networks (BRNN) and an attention mechanism to handle longer conversational inputs and maintain context [41].

GRU, a variant of RNN, employs two primary gates, reset gate and update gate to control which information should be retained or forgotten in the hidden state, making it more efficient for training on longer data sequences. Research [42] asserts that GRU outperforms RNN in handling long sequences without losing critical information from previous steps. Furthermore, GRU trains faster than LSTM due to fewer parameters requiring updates during training. LSTM, a type of RNN enhanced with three gates (input gate, forget gate, and output gate), effectively addresses the vanishing gradient problem commonly encountered in traditional RNNs. LSTM is designed to retain relevant long-term information in data sequences, making it highly effective for tasks involving extended conversational contexts, such as chatbots in the tourism sector. Research [43], [44], [45] demonstrates that LSTM enhances chatbot accuracy by maintaining longer conversational contexts, making it particularly suitable for applications involving more complex user interactions. While LSTM excels at handling long sequences, BiLSTM (Bidirectional LSTM) further enhances performance by processing data in both forward and backward directions. This bidirectional processing allows BiLSTM to utilize information from both directions within a sentence or conversation, providing a more comprehensive understanding of linguistic structures. A study [46], [47] confirms that BiLSTM achieves superior results in chatbot tasks by offering a more holistic and dynamic comprehension of context. The previous study proposes a chatbot system that uses a deep learning model (BiLSTM) to classify messages and improve service quality [48].

In addition to RNN-based models, CNN is also applied in text processing despite being more widely recognized for image processing. CNN is adapted to identify local patterns in text, such as n-grams, which frequently appear in conversational data. CNN training begins with vector representations of text, which are then processed through convolutional layers to capture key features. Pooling layers are employed to reduce dimensionality and extract the most relevant features. Research [37] suggests that CNN is highly effective in recognizing patterns in text data, enabling chatbots to handle diverse user queries more efficiently than RNN-based models. The advantage of CNN lies in its ability to process text data at higher speeds while maintaining good accuracy in text pattern recognition tasks.

Another study [49] introduces a system that combines chatbot technology with a Convolutional Neural Network (CNN) for improved agricultural decision-making. The chatbot, created with React and Natural Language Processing (NLP), enhances user interaction by answering queries and providing information. The system uses a CNN-based approach for plant disease classification. After collecting and preprocessing image data of various plant diseases, a CNN model is trained and tested to predict diseases. The chatbot is implemented using React to design the interface and NLP models for dynamic, contextual responses. The integration of the CNN model and chatbot involves using TensorFlow and a Node.js application to facilitate user interaction and predict plant diseases based on user input.

This study trained and evaluated six deep learning models to develop a retrieval-based chatbot for the tourism sector. Recognizing the pivotal role of hyperparameter optimization in ensuring optimal model performance, a systematic grid search approach was employed. Common hyperparameters, including learning rate, batch size, and the number of epochs, were methodically tuned to achieve an equilibrium between model convergence and overfitting. Learning rates within the range of 0.001 to 0.01, batch sizes of 16, 32, and 64, and training epochs varying from 50 to 100 were explored. Additionally, early stopping was applied to monitor validation loss and terminate training when no further improvement was observed, thereby mitigating overfitting.

Furthermore, model-specific hyperparameters were meticulously adjusted to accommodate the distinct characteristics of each architecture. For MLP, the number of hidden layers and neurons per layer were optimized, while activation functions such as ReLU and tanh were evaluated. Sequential models, including RNN, GRU, LSTM, and BiLSTM, were fine-tuned for hidden units, dropout rates, and the number of layers to enhance their ability to model long-term dependencies while preventing overfitting. CNN models underwent optimization of filter sizes, the number of filters, and pooling strategies to improve the efficiency of feature extraction. This thorough hyperparameter tuning process enhanced the reliability and robustness of the models, addressing critical challenges in chatbot development for tourism applications.

Overall, each deep learning model has its strengths and weaknesses in handling retrieval-based chatbots for the tourism sector. MLP performs well with static data, while RNN, GRU, and LSTM excel in managing sequential data and conversational contexts. BiLSTM further enhances performance by processing data bidirectionally, while CNN provides advantages in recognizing local patterns in text. The selection of the appropriate model depends heavily on the specific requirements of the chatbot application, such as conversation complexity, interaction length, and the type of data used. The computational resources for this study were provided by Google Colaboratory, utilizing its integrated GPU environment to efficiently train and test the six deep learning models. The GPU specifications included an NVIDIA Tesla K80 or T4 with 12GB VRAM, depending on the availability, and the platform's built-in support for Python-based libraries such as TensorFlow and Keras facilitated seamless implementation. The use of Google Colaboratory ensured accessibility and scalability, enabling the reproducibility of the study for future research.

2.4. Testing and Evaluation

Testing and evaluation represent critical phases in the development of retrieval-based chatbots for the tourism sector, aimed at assessing the effectiveness of deep learning models in responding to user queries with high accuracy and relevance. The process begins with testing the trained models using a dedicated testing dataset, followed by a thorough evaluation of model performance based on metrics specifically designed for text-based chatbot applications. This approach ensures the models are capable of handling diverse queries and providing accurate and contextually appropriate responses.

The initial step in testing involves validating each deep learning model with a testing dataset prepared during the dataset development phase. This dataset comprises queries aligned with predefined categories, such as ticket prices, facilities, or operating hours, and the responses generated by the models are compared against reference answers in the dataset to determine prediction accuracy. This stage measures the model's ability to provide the correct answers based on user inputs and assess its potential effectiveness in real-world scenarios. After testing, the evaluation phase utilizes standard metrics in natural language processing (NLP), including accuracy, precision, recall, and F1-score, to measure the model's performance comprehensively. Accuracy quantifies the percentage of correct answers relative to the total number of responses, while precision evaluates the proportion of correct answers among all responses generated by the model. Recall, on the other hand, assesses the model's ability to retrieve correct answers from all relevant answers available in the dataset, and the F1-score harmonizes precision and recall to offer a balanced performance indicator.

In addition to these objective metrics, the evaluation process incorporates a qualitative assessment of response quality from the user's perspective, which includes judging the relevance, completeness, and fluency of the chatbot's answers. This subjective evaluation is often carried out through trials involving real users or simulated conversations designed to test the chatbot in diverse scenarios. Prior research indicates that the quality of chatbot responses in the tourism context is heavily influenced by the model's capacity to understand user queries' context and deliver answers that align with user preferences and needs. Such evaluations are crucial in determining the chatbot's ability to enhance user satisfaction in practical applications.

To further refine the evaluation, confusion matrix analysis is employed to investigate the model's ability to differentiate between various response categories accurately. This step ensures that the chatbot not only delivers correct answers but also avoids misclassifying queries, a critical aspect for maintaining reliability in user interactions. Moreover, the evaluation process includes measuring response time, which is a vital efficiency metric in tourism chatbot applications where users demand rapid and accurate answers. Latency, defined as the time taken by the model to produce a response after receiving user input, is assessed to ensure the chatbot meets the speed requirements of the domain.

The insights derived from testing and evaluation inform recommendations on the most suitable deep learning model for retrieval-based chatbot applications in tourism. These findings also guide further improvements, such as parameter fine-tuning or leveraging transfer learning to enhance the model's ability to process more diverse and complex datasets. By integrating these strategies, the chatbot's performance can be continuously optimized to adapt to evolving user needs.

Overall, testing and evaluation provide a holistic view of how different deep learning models perform in the context of retrieval-based chatbots for the tourism sector. Each model demonstrates unique strengths and limitations, which must be weighed against the specific goals and requirements of the application. By conducting a comprehensive evaluation using appropriate metrics and qualitative assessments, chatbots can be refined to deliver a superior user experience, particularly in a sector where quick, accurate, and contextually relevant responses are essential for user satisfaction.

3. RESULTS AND DISCUSSION

3.1. Dataset Preparation Results

The initial phase of this research involved constructing a dataset, a critical step in developing AI-powered chatbots. The study focuses on two chatbot types: retrieval-based and generative chatbots. Separate datasets were designed for each type to meet their specific requirements in responding to tourism-related inquiries in Indonesia, particularly in the Tegal region.

The dataset for the retrieval-based chatbot was structured in JSON format, comprising three primary elements: tags, patterns, and responses. This dataset enables the chatbot to match user query patterns with predefined responses categorized under specific tags. A total of 71 tags were included, representing main categories such as tourist site descriptions, facilities, activities, ticket prices, and other relevant tourism information. Each tag includes various query patterns a user might present to the chatbot as Table 1. This dataset equips the retrieval-based chatbot to deliver precise answers by identifying the patterns in user queries.

Table 1. Retrieval-Based Chatbot Dataset

Tag	Pattern	Responses
desa_cempaka_facilities *(village_cempaka_facilities)*	"Apakah ada tempat makan di Desa Cempaka?" "Fasilitas bermain di Desa Cempaka ada?" "Apakah ada penginapan di Desa Cempaka?" "Ada fasilitas apa saja di Desa Cempaka?" *("Are there any places to eat in Cempaka Village?"* *"Are there play facilities in Cempaka Village?"* *"Is there accommodation in Cempaka Village?"* *"What facilities are there in Cempaka Village?")*	"Desa Wisata Cempaka menawarkan berbagai fasilitas yang membuat kunjungan Anda semakin nyaman dan menyenangkan. Fasilitas yang tersedia termasuk Jungle Tracking untuk petualangan alam, kios souvenir untuk membeli cendera mata, serta area kuliner untuk mencicipi berbagai makanan khas." *("Cempaka Tourism Village offers various facilities that make your visit more comfortable and enjoyable. Available facilities include Jungle Tracking for natural adventures, souvenir kiosks for buying souvenirs, as well as a culinary area for tasting various typical foods.")*

3.2. Preprocessing Results

Before training the chatbot models, the prepared dataset underwent preprocessing to transform the text into a cleaner and more structured format. Preprocessing aimed to reduce data complexity and ensure consistency in word representation, which is crucial for efficient model training. The preprocessing steps involve converting all text to lowercase to eliminate discrepancies caused by different capitalizations, ensuring that terms like "Desa Cempaka" and "desa cempaka" are treated as the same. Text is then tokenized, breaking it down into smaller units such as words or phrases, which makes it easier for the model to process. Additionally, lemmatization is applied to reduce words to their base or root form, such as transforming "walking" and "walked" into "walk," in order to simplify the text and reduce variations. After preprocessing, the dataset became cleaner and more structured, enabling deep learning models to train more effectively and deliver better performance. Table 2 explain about result of dataset preprocessing.

Table 2. Preprocessing Results

Before Preprocessing	After Preprocessing
Apakah ada tempat makan di Desa Cempaka?	tempat makan desa cempaka
Fasilitas bermain di Desa Cempaka ada?	fasilitas bermain desa cempaka
Apakah ada penginapan di Desa Cempaka?	penginapan desa cempaka
Ada fasilitas apa saja di Desa Cempaka?	fasilitas desa cempaka
("Are there any places to eat in Cempaka Village?"	*Cempaka village eating place*
"Are there play facilities in Cempaka Village?"	*Cempaka village play facilities*
"Is there accommodation in Cempaka Village?"	*Cempaka village accommodation*
"What facilities are there in Cempaka Village?")	*Cempaka village facilities*

3.3. Testing and Evaluation Results

After preprocessing, six deep learning models MLP (Multilayer Perceptron), RNN (Recurrent Neural Network), GRU (Gated Recurrent Units), LSTM (Long Short-Term Memory), BiLSTM (Bidirectional LSTM), and CNN (Convolutional Neural Network) were tested. This bar chart displays Fig. 2 the confusion matrix results for the MLP model, showing its performance across various classes. The predominance of green bars (True Positives and True Negatives) highlights that the MLP model performs well overall, with high accuracy for most classes. However, occasional red (False Positives) and blue (False Negatives) segments indicate misclassifications in specific classes, suggesting limitations in handling more complex or nuanced patterns in the data.

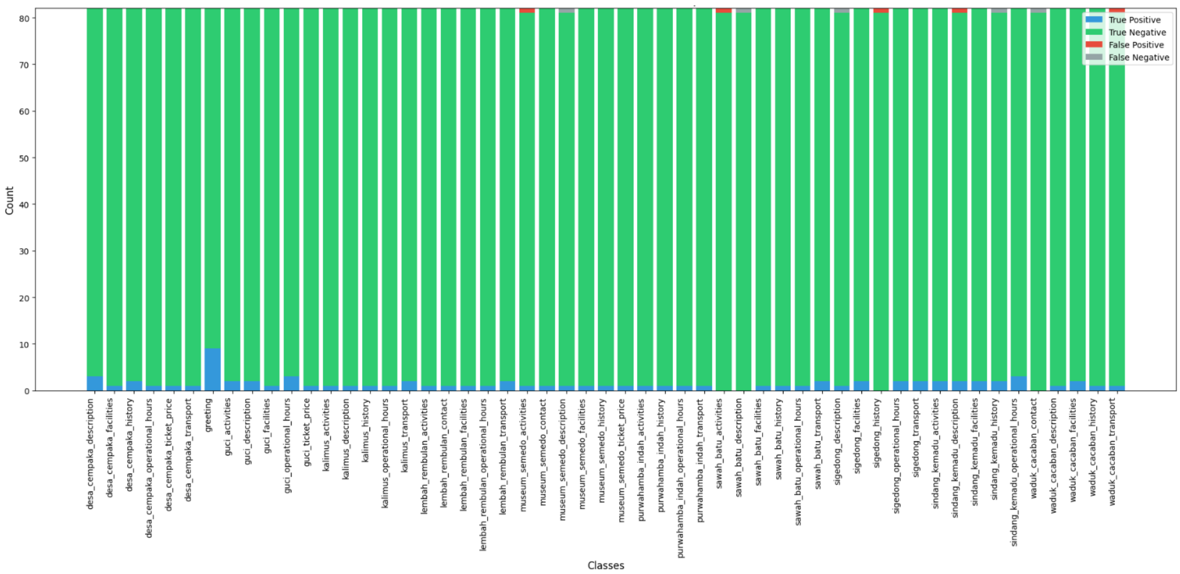

Fig. 2. Confusion Matrix of MLP

This bar in Fig. 3 chart shows the confusion matrix results for the RNN model, illustrating its classification performance across multiple classes. The abundance of green bars (True Positives and True Negatives) indicates that the RNN model achieves moderate success in identifying correct classifications. However, the presence of blue (False Negatives) and red (False Positives) segments for several classes suggests that the RNN struggles with certain patterns, indicating limited contextual understanding and misclassification tendencies in structured datasets.

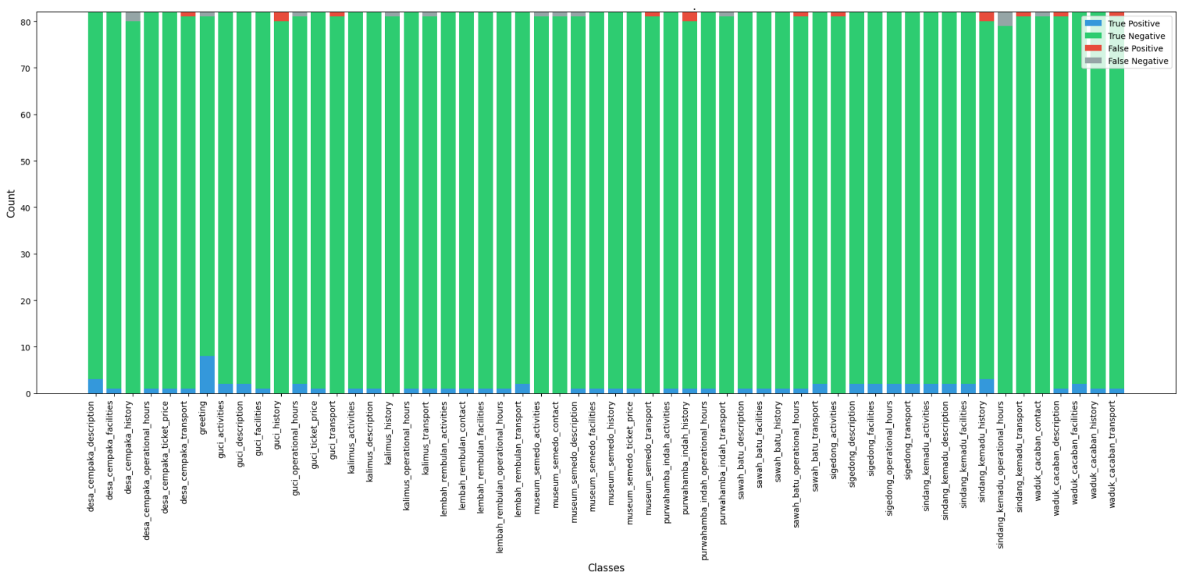

Fig. 3. Confusion Matrix of RNN

The bar chart as Fig. 4 represents the confusion matrix results for the GRU model, demonstrating its classification accuracy across different classes. The dominance of green bars (True Positives and True Negatives) suggests that the GRU performs well overall in correctly identifying most cases. However, some blue (False Negatives) and red (False Positives) bars for specific classes indicate that the GRU encounters occasional misclassifications, reflecting challenges in capturing intricate contextual patterns in certain instances.

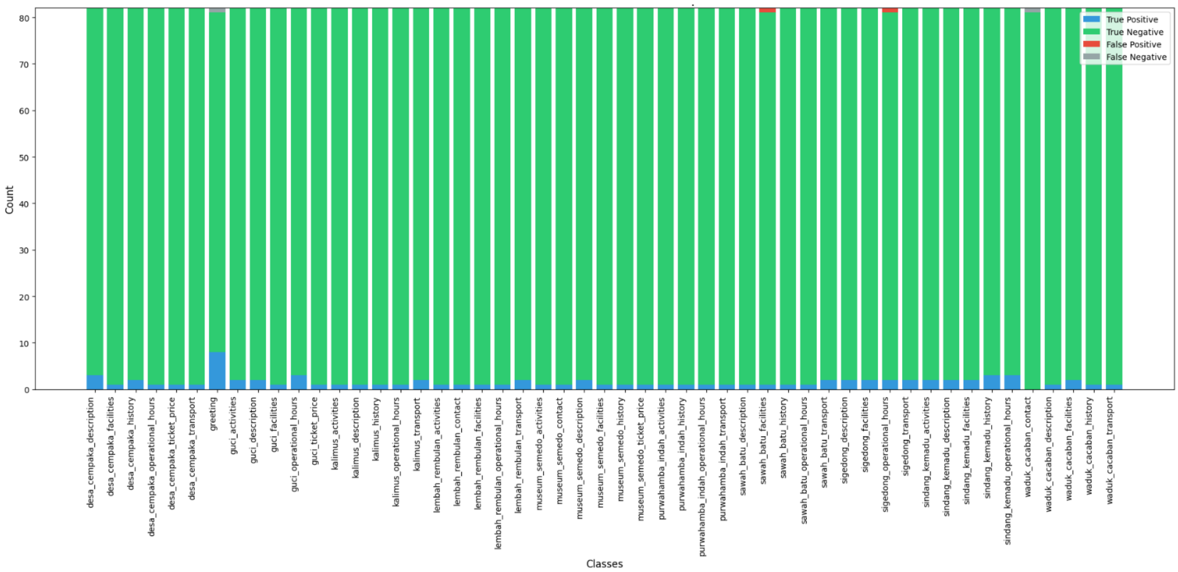

Fig. 4. Confusion Matrix of GRU

Fig. 5 represents the confusion matrix results for the LSTM model, showcasing its classification performance across various classes. The predominance of green bars (True Positives and True Negatives) indicates that the LSTM model achieves high accuracy overall, correctly classifying most instances. However, the scattered presence of blue (False Negatives) and red (False Positives) bars for certain classes suggests that the model struggles with occasional misclassifications, particularly in handling nuanced or overlapping features within the dataset.

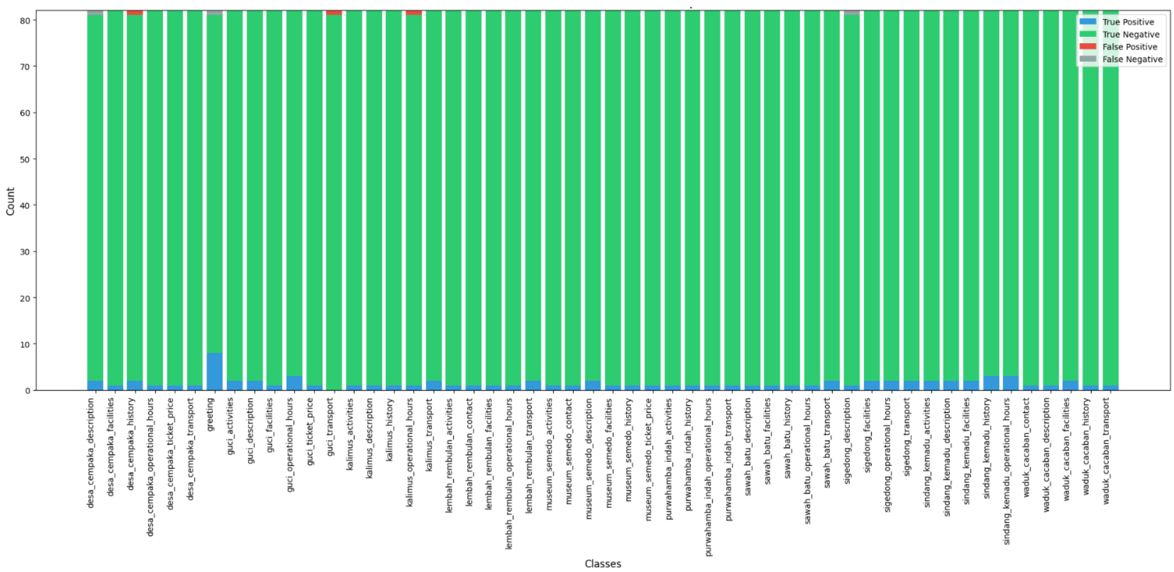

Fig. 5. Confusion Matrix of LSTM

Fig. 6 is a visual representation of the confusion matrix results for a BiLSTM model, depicting classification performance across various classes. The green bars, dominating the chart, indicate a high number of true positive and true negative predictions, suggesting the model's strong overall accuracy. However, the presence of small blue and red bars for specific classes highlights instances of false negatives and false positives, suggesting occasional misclassifications in certain categories.

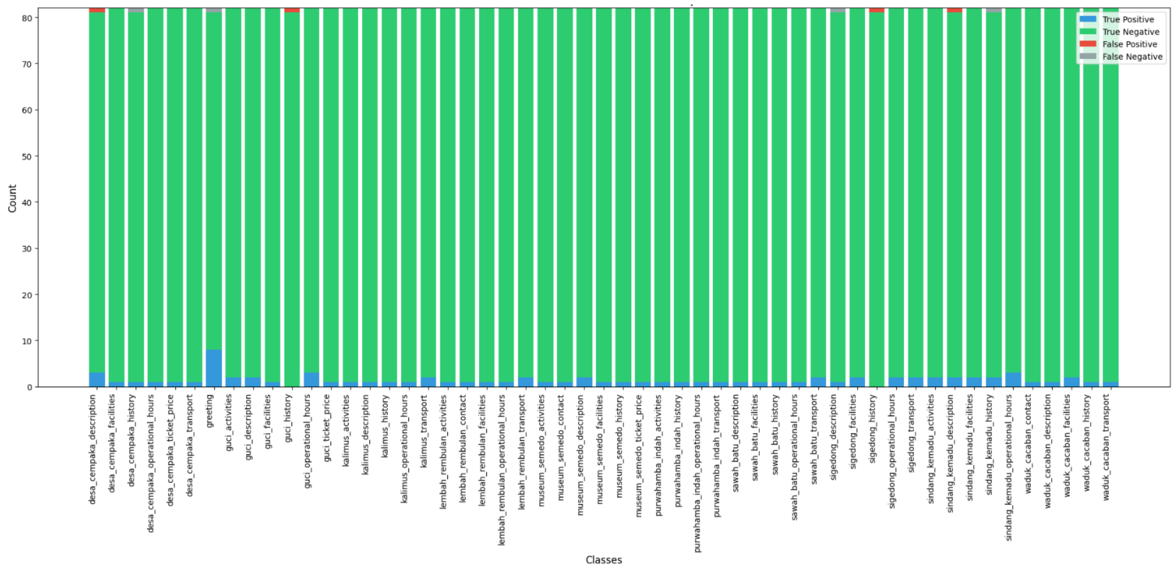

Fig. 6. Confusion Matrix of BiLSTM

Fig. 7 represents the confusion matrix results for a CNN model, with true positive and true negative predictions (green bars) dominating most classes, indicating high overall accuracy. However, small red and blue bars in specific classes point to instances of false positives and false negatives, suggesting occasional misclassification. These errors, while minimal, may require further investigation to enhance the model's performance for the affected classes.

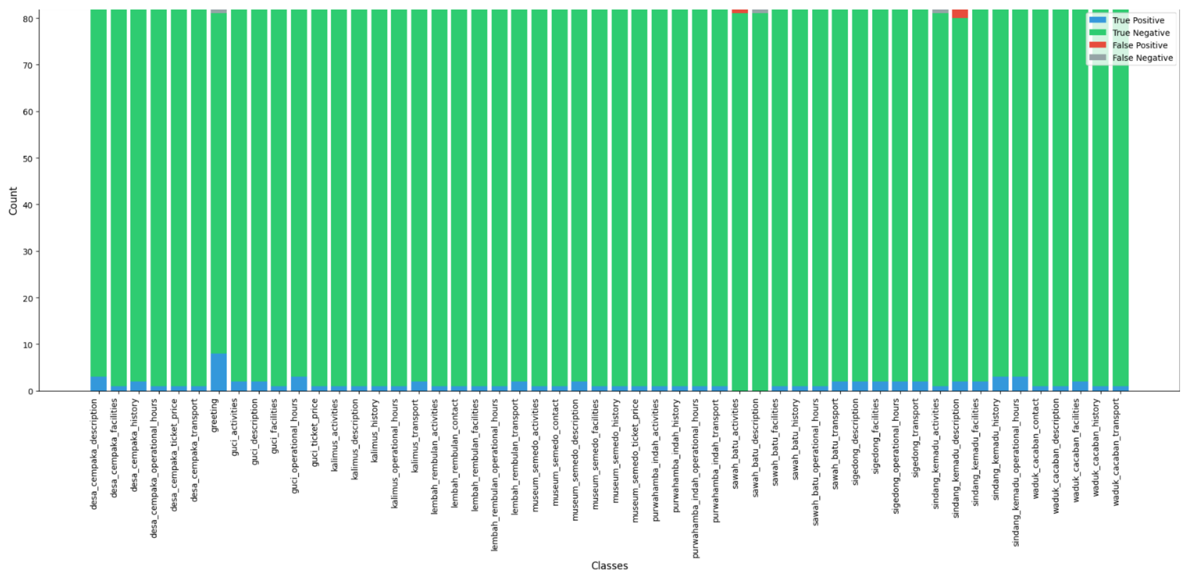

Fig. 7. Confusion Matrix of CNN

Their performance was evaluated using standard metrics: accuracy, precision, recall, and F1 score as illustrated by Table 3. The CNN model demonstrated the best performance among all tested models, achieving the highest accuracy (0.98), recall (0.98), precision (0.99), and F1 score (0.98). This result underscores CNN's ability to identify and filter information effectively, making it highly suitable for handling user query patterns in retrieval-based chatbots. CNN excels at extracting local features in text, which gives it a significant advantage in this application.

The MLP model also performed well, with an accuracy of 0.94. However, its effectiveness is somewhat limited in handling complex queries due to its simpler architecture. Conversely, the RNN model exhibited the lowest performance, with an accuracy of 0.82. While RNNs are generally effective for sequential data, their performance in this context suggests limitations in processing structured datasets requiring a deeper contextual understanding. Overall, the CNN model's superior performance highlights its suitability for retrieval-based chatbots, particularly in applications requiring high accuracy and speed in responding to user queries.

Table 3. Assessment of Deep Learning Models

Model	Accuracy	Recall	Precision	F1 Score
MLP	0.94	0.94	0.96	0.94
RNN	0.82	0.82	0.86	0.82
GRU	0.91	0.91	0.99	0.94
LSTM	0.90	0.90	0.98	0.92
BiLSTM	0.91	0.91	0.96	0.92
CNN	0.98	0.98	0.99	0.98

3.4. Discussion

Based on the experimental results, this study concludes that the Convolutional Neural Network (CNN) model is the most effective and efficient model for developing retrieval-based chatbot systems in the tourism sector. CNN demonstrated a significant advantage in handling highly complex queries and delivering accurate results in filtering and extracting relevant information from textual datasets. This model effectively captures dominant local features within the text, enabling the chatbot to provide faster and more precise responses. However, slight misclassifications observed in the confusion matrix could be mitigated by further fine-tuning hyperparameters or enhancing the dataset with more diverse queries. The lower performance of the RNN model highlights its limitations in handling structured datasets, which could be addressed by utilizing more advanced architectures like GRU or LSTM for better sequential dependency modeling. Additionally, improvements in data preprocessing, such as enhanced tokenization or normalization techniques, could reduce errors across all models and enhance their practical performance in real-world applications.

The results of this study are consistent with prior research emphasizing the effectiveness of Convolutional Neural Networks (CNNs) in natural language processing tasks. For example, previous studies have shown that CNN-based chatbots outperform Recurrent Neural Networks (RNNs) in terms of accuracy and relevance of responses, showcasing CNNs' superior capability in analyzing structured textual data [37]. Additionally, research on hybrid algorithms, combining KNN, RNN, and CNN, has demonstrated that the inclusion of CNNs significantly enhances the accuracy of sentiment analysis models [50]. These findings emphasize the adaptability and effectiveness of CNNs in managing complex textual patterns, further supporting the outcomes of this study. Collectively, the evidence confirms CNNs' suitability for retrieval-based chatbot applications, especially in scenarios requiring high precision and robust natural language processing.

The superior performance of the CNN model can be attributed to its proficiency in identifying and leveraging local patterns within textual data, which is crucial for understanding and responding to complex user queries in the tourism domain. In contrast, the RNN model's lower performance may stem from its inherent limitations in capturing long-range dependencies and handling structured data, which are better managed by advanced architectures like GRU and LSTM. The results highlight trade-offs between the computational efficiency of simpler models like MLP and the superior accuracy of more complex models like CNN. While MLP requires significantly fewer computational resources and training time, its performance is limited in handling complex patterns within structured datasets, making it less suitable for tasks requiring high contextual understanding. Conversely, the CNN model, despite its higher computational demands, delivers superior accuracy by effectively capturing local patterns and dependencies in text, which are critical for retrieval-based chatbots. Practitioners in the tourism sector must balance these trade-offs based on their specific application requirements, such as prioritizing computational efficiency for resource-constrained environments or leveraging the higher accuracy of CNN for tasks requiring robust and precise user interactions.

Nevertheless, the study's findings are constrained by the dataset's limited size, comprising only 11 tourism destinations in Tegal, Indonesia, which may impact its external validity and generalizability to broader tourism contexts. Although these destinations were selected for their priority status and diversity, the sample may not comprehensively represent the wider tourism sector. Furthermore, the reliance on data collected through in-depth interviews introduces potential biases, despite efforts to mitigate variability through standardized questionnaires. To enhance the robustness and reliability of future studies, it is recommended to expand the dataset to encompass a more diverse range of destinations and adopt alternative data collection methods, such as automated web scraping or crowdsourcing, to reduce subjective bias and improve generalizability. Despite these limitations, the study demonstrates that the CNN model consistently outperforms other deep learning models in terms of accuracy, precision, recall, and F1-score. This research, therefore, provides a significant contribution to the field by offering valuable insights into the application of deep learning models in the tourism sector, highlighting the CNN model's superior performance in developing retrieval-based chatbots, and establishing a foundation for future advancements.

4. CONCLUSION

This study demonstrates that Convolutional Neural Networks (CNNs) are the most effective deep learning models for developing retrieval-based chatbot systems in the tourism sector. CNN outperformed other tested models, including MLP, RNN, GRU, LSTM, and BiLSTM, achieving superior accuracy, precision, recall, and F1-score. Its ability to capture local patterns in structured datasets makes it particularly effective for handling complex and diverse user queries, delivering reliable and contextually relevant responses. These findings reinforce CNN's suitability for applications requiring robust natural language processing capabilities, particularly in the context of retrieval-based chatbot systems.

Despite these promising results, this study is subject to certain limitations. The dataset, which included information from only 11 priority tourism destinations in Tegal, Indonesia, is relatively small and may not fully represent the complexities of broader tourism contexts. Additionally, the reliance on data collected through in-depth interviews introduces potential biases despite efforts to mitigate them using standardized questionnaires. Furthermore, the computational requirements for CNN training and deployment, which rely on GPU-accelerated environments, could pose challenges for smaller organizations with limited technical infrastructure. These limitations should be considered when interpreting the results and applying the models in broader contexts.

The research contributes to new knowledge by offering a comparative evaluation of deep learning models in the tourism sector and highlighting the trade-offs between computational efficiency and model accuracy. The study emphasizes CNN's strength in effectively processing structured data, providing actionable insights for practitioners aiming to develop AI-driven solutions in tourism. Moreover, it establishes a foundational framework for leveraging deep learning in retrieval-based chatbots, thus advancing the field of AI in tourism applications.

Future research could focus on expanding the dataset to encompass a broader range of tourism destinations and more diverse user queries, improving generalizability and external validity. Investigating hybrid architectures that integrate CNNs with other advanced models, such as LSTMs or attention mechanisms, could further enhance chatbot performance by leveraging their combined strengths. Exploring resource-efficient deployment strategies, such as model compression or edge computing, would facilitate broader adoption of high-performing chatbot systems in resource-constrained environments. Finally, integrating generative models with retrieval-based chatbots could improve their ability to handle dynamic, context-dependent queries, increasing versatility and engagement.

Acknowledgments

This research was entirely supported by the "Penelitian Produk Vokasi" grant provided by the Ministry of Research, Technology, and Higher Education of the Republic of Indonesia, under contract 103/SPK/D.D4/PPK.01.APTV/III/2024, 028/LL6/PPM-V/AL.04/2024, 098.016/P3M.PHB/IV/2024.

REFERENCES

[1] T. Pencarelli, "The digital revolution in the travel and tourism industry," *Inf. Technol. Tour.*, vol. 22, Sep. 2020, https://doi.org/10.1007/s40558-019-00160-3.

[2] K. Muley and K. Zainab Irfan DrKirti Muley, "Chatbot Adoption in Travel and Tourism Services," *International Journal of All Research Education and Scientific Methods (IJARESM),* p. 25, 2023, https://bvimit.co.in/bvimit/pdf/Technical%20Magazine%202022%20(MCA%20SEM%20IV).pdf#page=2 .

[3] A. C. M. Nafrees and F. H. A. Shibly, "Smart Technologies in Tourism: A Study Using Systematic Review and Grounded Theory," *International Research Conference on Smart Computing and Systems Engineering (SCSE),* vol. 4, pp. 8-13, 2021, https://doi.org/10.1109/SCSE53661.2021.9568338'.

[4] O. M. Alyasiri, K. Selvaraj, H. A. Younis, T. Mueen Sahib, M. F. Almasoodi, and I. M. Hayder, "A Survey on the Potential of Artificial Intelligence Tools in Tourism Information Services Keywords Artificial Intelligence ChatGPT Hospitality and Tourism Industry Tourism Education and Research Tourism and Management," *Babylonian J. Artif. Intell.*, pp. 1–8, 2024, [Online]. Available: https://mesopotamian.press/journals/index.php/BJAI.

[5] Z. H. Syed, A. Trabelsi, E. Helbert, V. Bailleau, and C. Muths, "Question answering chatbot for troubleshooting queries based on transfer learning," *Procedia Comput. Sci.*, vol. 192, no. 2019, pp. 941–950, 2021, https://doi.org/10.1016/j.procs.2021.08.097.

[6] M. Xu, "Research on Smart Tourism System Based on Artificial Intelligence," in *2023 IEEE 3rd International*

Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA), pp. 201–205, 2023, https://doi.org/10.1109/ICIBA56860.2023.10165293.

[7] L. Benaddi, C. Ouaddi, A. Souha, A. Jakimi, and B. Ouchao, "Chatbots in Tourism Sector: Classification, Evolution, and Functionalities," *International Symposium on Signal, Image, Video and Communications (ISIVC),* pp. 1–6, 2024, https://doi.org/10.1109/ISIVC61350.2024.10577889.

[8] B. El Bakkouri, S. Raki, and T. Belgnaoui, "The Role of Chatbots in Enhancing Customer Experience: Literature Review," *Procedia Comput. Sci.*, vol. 203, pp. 432–437, 2022, https://doi.org/10.1016/j.procs.2022.07.057.

[9] M. Sehgal, A. Arya, S. Gaba, and S. Nagpal, "AI for Tourism: Pioneering New Frontiers in the Industry," in *2024 International Conference on Computational Intelligence and Computing Applications (ICCICA)*, pp. 280–285, 2024, https://doi.org/10.1109/ICCICA60014.2024.10585210.

[10] P. J. Antony, R. Kannan, and A. Professor, "Revolutionizing the Tourism Industry through Artificial Intelligence: A Comprehensive Review of AI Integration, Impact on Customer Experience, Operational Efficiency, and Future Trends," *www.chandigarhphilosophers.com International Journal for Multidimensional Research Perspectives (IJMRP*, vol. ISSN, no. 1. pp. 2584–2613, 2024. [Online]. Available: www.chandigarhphilosophers.com.

[11] H.-C. Kang, K.-B. Kang, D.-H. Kim, M.-C. Jwa, T.-S. Ko, and J.-W. Jwa, "Smart tourism chatbot system using Multi-domain Tourism Information DST," in *2023 Fourteenth International Conference on Ubiquitous and Future Networks (ICUFN)*, pp. 608–612, 2023, https://doi.org/10.1109/ICUFN57995.2023.10200288.

[12] Y. Wu, W. Wu, C. Xing, Z. Li, and M. Zhou, "Sequential matching network: A new architecture for multi-turn response selection in retrieval-based chatbots," *ACL 2017 - 55th Annu. Meet. Assoc. Comput. Linguist. Proc. Conf. (Long Pap.*, vol. 1, pp. 496–505, 2017, https://doi.org/10.18653/v1/P17-1046.

[13] K. Moore *et al.*, "A comprehensive solution to retrieval-based chatbot construction," *Comput. Speech Lang.*, vol. 83, p. 101522, 2024, https://doi.org/https://doi.org/10.1016/j.csl.2023.101522.

[14] N. Laosen, K. Laosen, and P. Sriprasert, "Development of AI Chatbot for Tourism Promotion: A Case Study in Ranong and Chumphon, Thailand," in *2024 21st International Joint Conference on Computer Science and Software Engineering (JCSSE)*, pp. 224–231, 2024, https://doi.org/10.1109/JCSSE61278.2024.10613647.

[15] M. A. Nyakonu, "Improving and Automating Knowledge Retrieval Chatbots," *Doctoral dissertation, Metropolitan State University*, 2021.

[16] V. Dutt, S. M. Sasubilli, and A. E. Yerrapati, "Dynamic Information Retrieval With Chatbots: A Review of Artificial Intelligence Methodology," in *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, pp. 1299–1303, 2020, https://doi.org/10.1109/ICECA49313.2020.9297533.

[17] M. Rukhiran and P. Netinant, "Automated information retrieval and services of graduate school using chatbot system," *Int. J. Electr. Comput. Eng.*, vol. 12, no. 5, pp. 5330–5338, 2022, https://doi.org/10.11591/ijece.v12i5.pp5330-5338.

[18] A. Vakili Tahami, K. Ghajar, and A. Shakery, "Distilling Knowledge for Fast Retrieval-based Chat-bots," in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 2081–2084, 2020¸ https://doi.org/10.1145/3397271.3401296.

[19] I. T. Aksu, N. F. Chen, L. F. D'Haro, and R. E. Banchs, "Reranking of Responses Using Transfer Learning for a Retrieval-Based Chatbot," *Increasing Naturalness and Flexibility in Spoken Dialogue Interaction: 10th International Workshop on Spoken Dialogue Systems,* pp. 239–250, 2021, https://doi.org/10.1007/978-981-15-9323-9_20.

[20] S. Pandey and S. Sharma, "Healthcare Analytics A comparative study of retrieval-based and generative-based chatbots using Deep Learning and Machine Learning," *Healthc. Anal.*, vol. 3, p. 100198, 2023, https://doi.org/10.1016/j.health.2023.100198.

[21] F. M. Shiri, R. Mohamed, T. Perumal, and N. Mustapha, "A Comprehensive Overview and Comparative Analysis on Deep Learning Models: CNN, RNN, LSTM, GRU," a*rXiv preprint arXiv:2305.17473*., 2023, https://doi.org/10.32604/jai.2024.054314%5d.

[22] R. Kruse, S. Mostaghim, C. Borgelt, C. Braune, and M. Steinbrecher, "Multi-layer Perceptrons," *Computational intelligence: a methodological introduction*, pp. 53–124, 2022, https://doi.org/10.1007/978-3-030-42227-1_5.

[23] X. Wu, B. Xiang, H. Lu, C. Li, and X. Huang, "Optimizing Recurrent Neural Networks : A Study on Gradient Normalization of Weights for Enhanced Training Efficiency," *Applied Sciences*, vol. 14, no. 15, 2024, https://doi.org/10.3390/app14156578s.

[24] D. R. Darmawan and R. Arifudin, "Enhancing Durrotalk Chatbot Accuracy Utilizing a Hybrid Model Based on Recurrent Neural Network (RNN) Algorithm and Decision Tree," *JUITA J. Inform.*, vol. 12, no. 1, p. 81, 2024, https://doi.org/10.30595/juita.v12i1.20868.

[25] M. Pham and Q. C. Ta, "Stacked Long Short-Term Memory for Vietnamese Stock Market Returns Prediction," *International Journal of Applied Sciences: Current and Future Search Trends*, 2022, https://ijascfrtjournal.isrra.org/index.php/Applied_Sciences_Journal/article/view/1237.

[26] W. Xu, J. Chen, Z. Ding, and J. Wang, "Text Sentiment Analysis and Classification Based on Bidirectional Gated Recurrent Units (GRUs) Model," *arXiv preprint arXiv:2404.17123*., 2024, https://doi.org/10.48550/arXiv.2404.17123.

[27] Y. Xie, Y. Ueda, and M. Sugiyama, "A Two-Stage Short-Term Load Forecasting Method Using Long Short-Term Memory and Multilayer Perceptron," *Energies*, vol. 14, no. 18, p. 5873, 2021, https://doi.org/10.3390/en14185873.

[28] A. Sherstinsky, "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network," *Phys. D Nonlinear Phenom.*, vol. 404, p. 132306, 2020, https://doi.org/10.1016/j.physd.2019.132306.

[29] P. F. Muhammad, R. Kusumaningrum, and A. Wibowo, "Sentiment Analysis Using Word2vec And Long Short-

Term Memory (LSTM) For Indonesian Hotel Reviews," *Procedia Comput. Sci.*, vol. 179, pp. 728–735, 2021, https://doi.org/10.1016/j.procs.2021.01.061.
[30] J. Shuford, N. S. Business, and T. Columnist, "Unveiling the Power of Deep Learning : Insights into Advanced Neural n Network Architectures," *Journal of Artificial Intelligence General science (JAIGS),* vol. 3, no. 1, pp. 1–14, 2024, http://newjaigs.com/index.php/JAIGS/article/view/60.
[31] M. Desai and M. Shah, "An anatomization on Breast Cancer Detection and Diagnosis employing Multi-layer Perceptron Neural Network (MLP) and Convolutional Neural Network (CNN)," *Clinical eHealth,* vol. 4, pp. 1-11, 2020, https://doi.org/10.1016/j.ceh.2020.11.002.
[32] A. M. A. Sai, O. Balamurali, M. Karthikeya, and S. Anand, "A Web-Based Chatbot for Indian Cities: A Comparison of CNN, ANN, and LSTM Models," in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pp. 1–6, 2023, https://doi.org/10.1109/ICCCNT56998.2023.10307912.
[33] T. R, "Chatbot Application for Tourism Using Deep Learning," *Journal for Research in Applied Science & Engineering Technology (IJRASET),* vol/ 10, no. 6, 2022, https://doi.org/10.22214/ijraset.2022.44516.
[34] Y. Windiatmoko, R. Rahmadi, and A. F. Hidayatullah, "Developing Facebook Chatbot Based on Deep Learning Using RASA Framework for University Enquiries," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1077, no. 1, p. 012060, 2021, https://doi.org/10.1088/1757-899X/1077/1/012060.
[35] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural language processing: state of the art, current trends and challenges," *Multimed. Tools Appl.*, vol. 82, no. 3, pp. 3713–3744, 2023, https://doi.org/10.1007/s11042-022-13428-4.
[36] D. W. Otter, J. R. Medina, and J. K. Kalita, "A Survey of the Usages of Deep Learning for Natural Language Processing," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 32, no. 2, pp. 604–624, 2021, https://doi.org/10.1109/TNNLS.2020.2979670.
[37] G. Tsakiris, C. Papadopoulos, G. Patrikalos, K. Kollias, and G. F. Fragulis, "The development of a chatbot using Convolutional Neural Networks," *SHS Web of Conferences*, vol. 139, p. 03009, 2022, https://doi.org/10.1051/shsconf/202213903009.
[38] D. Khyani, S. B S, N. Niveditha, and D. Y M, "An Interpretation of Lemmatization and Stemming in Natural Language Processing," *J. Univ. Shanghai Sci. Technol.*, vol. 22, 2021, https://jusst.org/an-interpretation-of-lemmatization-and-stemming-in-natural-language-processing/.
[39] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad Khasmakhi, M. Asgari-Chenaghlu, and J. Gao, "Deep Learning Based Text Classification: A Comprehensive Review," *ACM computing surveys (CSUR),* vol. 54, no. 3, pp. 1-40, 2021, https://doi.org/10.1145/3439726.
[40] A. Ciayandi, V. Mawardi, and J. Hendryli, "Retrieval Based Chatbot on Tarumanagara University With Multilayer Perceptron," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1007, no. 1, p. 012146, 2020, https://iopscience.iop.org/article/10.1088/1757-899X/1007/1/012146/meta.
[41] M. Dhyani and R. Kumar, "An intelligent Chatbot using deep learning with Bidirectional RNN and attention model," *Mater. Today Proc.*, vol. 34, pp. 817–824, 2019, https://doi.org/10.1016/j.matpr.2020.05.450.
[42] Z. Niu *et al.*, "Recurrent attention unit: A new gated recurrent unit for long-term memory of important parts in sequential data," *Neurocomputing*, vol. 517, pp. 1–9, 2023, https://doi.org/10.1016/j.neucom.2022.10.050.
[43] P. Anki, A. Bustamam, H. Al-Ash, and D. Sarwinda, "Intelligent Chatbot Adapted from Question and Answer System Using RNN-LSTM Model," *Journal of Physics: Conference Series*, vol. 1844, no. 1, p. 012001, 2021, 10.1088/1742-6596/1844/1/012001.
[44] G. Sperlí, "A cultural heritage framework using a Deep Learning based Chatbot for supporting tourist journey," *Expert Syst. Appl.*, vol. 183, p. 115277, 2021, https://doi.org/10.1016/j.eswa.2021.115277.
[45] E. Kasthuri and S. Balaji, "Natural language processing and deep learning chatbot using long short term memory algorithm," *Mater. Today Proc.*, vol. 81, pp. 690–693, 2023, https://doi.org/10.1016/j.matpr.2021.04.154.
[46] P. Anki, A. Bustamam, H. S. Al-Ash, and D. Sarwinda, "High Accuracy Conversational AI Chatbot Using Deep Recurrent Neural Networks Based on BiLSTM Model," *Int. Conf. Inf. Commun. Technol.*, pp. 382-387, 2020, https://doi.org/10.1109/ICOIACT50329.2020.9332074.
[47] N. Lhasiw, N. Sanglerdsinlapachai, and T. Tanantong, "A Bidirectional LSTM Model for Classifying Chatbot Messages," in *2021 16th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP)*, pp. 1–6. 2021, https://doi.org/10.1109/iSAI-NLP54397.2021.9678173.
[48] C. Dewi, C. S. A. Kristiantoro, H. J. Christanto, D. Riantama, " Improving service quality through classifying chatbot messages based on natural language processing: A bidirectional long short-term memory network model," *International Journal of Applied Science and Engineering,* 2024, https://ir.lib.cyut.edu.tw/bitstream/310901800/43209/1/ijase-202406-21-2-006.pdf.
[49] R. Patil, Y. Sinkar, A. Ruke, H. Kulkarni, and O. Kadam, "Smart Agri-Advisor: Integrating Chatbot Technology with CNN-Based Crop Disease Classification for Enhanced Agricultural Decision-Making," *Int. J. Eng. Trends Technol.*, vol. 72, no. 7, pp. 375–380, 2024, https://doi.org/10.14445/22315381/IJETT-V72I7P141.
[50] T. Hariguna, A. Ruangkanjanases, "Exploring the Flexibility and Accuracy of Sentiment Scoring Models through a Hybrid KNN-RNN-CNN Algorithm and ChatGPT," *HighTech and Innovation Journal,* vol. 4, no. 2, pp. 315-326, 2023, https://doi.org/10.28991/HIJ-2023-04-02-06.

BIOGRAPHY OF AUTHORS

Dwi Intan Af'idah received a bachelor of computer engineering and a master of information systems from Diponegoro University, Indonesia in 2014 and 2020. After graduating with a bachelor's degree and before continuing her master's studies, she worked as a systems analyst at an IT consulting company, which made her understand the need for technology in various industrial fields. Currently, she is a lecturer and a secretary of the undergraduate program in the informatics engineering department at Harapan Bersama Polytechnic, Indonesia. Her research interests include artificial intelligence applied, artificial intelligence-based application, machine learning, and deep learning, and she also focuses on the study of natural language processing (NLP) as well as the development of tools based on NLP like sentiment analysis systems, machine translation, and chatbots for specific domains. Some of her research has been funded by the Ministry of Higher Education and Research, Indonesia. She can be contacted at dwiintanafidah@poltektegal.ac.id.

Dairoh currently she is a doctoral candidate in the Department of Physics at Gadjah Mada University, Indonesia. She received a bachelor of physics from Jenderal Soedirman University, Indonesia, in 2010 and a master of physics from Gadjah Mada University, Indonesia, in 2014. She is a lecturer at the Department of Informatics Engineering, Harapan Bersama Polytechnic. Her research interests include machine learning, deep learning, and experiencing natural phenomena using artificial intelligence-based applications. Some of her research has been funded by the Ministry of Higher Education and Research, Indonesia. She can be contacted at dairoh@poltektegal.ac.id.

Sharfina Febbi Handayani received a bachelor of informatics engineering from Jenderal Soedirman University, Indonesia in 2015 and a master of information systems from Sepuluh Nopember Institute of Technology, Indonesia in 2019. She is a lecturer at the Department of Informatics Engineering, Harapan Bersama Polytechnic. She is also a member of the Informatics and Computer Higher Education Association in Indonesia. Additionally, she became an editorial section in the Information and Technology Developer Journal in 2022. Her research interests include business process management, software quality assurance, artificial intelligence-based information systems, and artificial intelligence-based applications. The Ministry of Higher Education and Research, Indonesia, has funded her research. She can be contacted at sharfina.handayani@poltektegal.ac.id.