Jurnal ELTIKOM :
Jurnal Teknik Elektro, Teknologi Informasi dan Komputer

Vol. 9, No. 1, June 2025, page. 69-77
ISSN 2598-3245 (Print), ISSN 2598-3288 (Online)
DOI: http://doi.org/10.31961/eltikom.v9i1.1481
Available online at http://eltikom.poliban.ac.id

# APPLICATION OF NATURAL LANGUAGE PROCESSING AND LSTM IN A TRAVEL CHATBOT FOR MEDAN CITY

**Syarifah Atika[1*], Mahara Bengi[2], Shekainah Kim A. Sardeng[3]**
[1)] Faculty of Science and Technology, Universitas Prima Indonesia, Medan, Indonesia
[2)] Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Medan, Indonesia
[3)] College of Information Technology, Don Mariano Marcos Memorial State University, La Union, Philippines
e-mail: syarifahatika@unprimdn.ac.id, maharabengi@students.usu.ac.id, ssardeng@dmmmsu.edu.ph

Received: 7 March 2025 – Revised: 20 April 2025 – Accepted: 1 May 2025

***ABSTRACT***

*The tourism sector plays a vital role in economic growth and regional development. Medan, a major city in North Sumatra, offers rich religious, historical, and cultural attractions. However, fragmented and inconsistent information presents challenges for both tourists and destination managers, often complicating travel planning. To address this issue, this study proposes the development of an AI-based chatbot aimed at enhancing the tourism experience in Medan. By integrating Natural Language Processing (NLP) and Long Short-Term Memory (LSTM), the chatbot is designed to deliver accurate, contextual, and conversational responses tailored to users' tourism-related queries. It was trained on a comprehensive dataset compiled from various sources concerning Medan's tourism. The training ran over 100 epochs, achieving an accuracy of 84.31% and a loss of 0.7594. Validation testing yielded an accuracy of 77.14% and a loss of 2.4233, indicating good generalization to unseen data. End-to-end testing with 312 queries covering all defined intents resulted in a testing accuracy of 75.64%, confirming the model's practical effectiveness. The findings demonstrate that the chatbot can accurately interpret user input, classify information, and enhance user interaction. supports the digital transformation of Medan's tourism sector by introducing a reliable, AI-driven tool for seamless travel planning and engagement.*

***Keywords****: artificial intelligence, long short-term memory (LSTM), natural language processing (NLP), travel chatbot.*

## I. INTRODUCTION

TOURISM is a vital driver of economic growth and regional development. This sector encompasses a wide range of industries involving public, private, and community actors and includes various forms of tourism such as mass tourism, ecotourism, adventure tourism, and volunteer tourism. It also contributes to shaping a country's national image. Improving efficiency and fostering innovation in this sector require strong collaboration among stakeholders. Therefore, both the government and the private sector should work together to enhance its productivity [1].

Indonesia is widely recognized for its thriving tourism sector, which significantly contributes to both local and national development. As the capital of North Sumatera Province, Medan offers a diverse array of tourist attractions. The city presents a unique blend of natural sites, historical landmarks, rich cultural heritage, and renowned culinary experiences that attract both domestic and international visitors each year. To support these tourists, an easily accessible platform is needed to provide reliable information on Medan's attractions.

However, challenges such as fragmented content, inconsistent information quality, and information overload often hinder tourists from accessing accurate and timely details. These issues complicate travel planning and diminish the overall travel experience. To address these concerns, artificial intelligence (AI)-powered chatbots present an effective solution.

Chatbots offer a practical medium for delivering information. A chatbot is an AI-based system designed to simplify and personalize interactions between humans and machines. It can interpret user inputs and provide appropriate responses without human intervention, thereby offering round-the-clock service. Chatbots improve productivity, accessibility, and service quality across various aspects of daily life.

Advancements in natural language processing and machine learning have enabled chatbots to become more sophisticated, supporting seamless and context-aware communication. Chatbots have shown practical benefits in multiple sectors, including healthcare [2][3]. There are primary approaches to chatbot development: rule-based systems, which respond based on predefined rules, and machine learning systems, which are trained using large datasets.

This study employs a machine learning-based chatbot that uses Natural Language Processing (NLP) techniques to process user input and generate appropriate responses. To handle sequential language data effectively, the model adopts Long Short-Term Memory (LSTM), a method widely used in conversational AI systems [4]. LSTM is typically applied in training chatbot models to learn conversational patterns and relationships, enabling more accurate responses to user input. Previous studies [5], [6] have implemented LSTM models to improve chatbot accuracy and performance.

In [6], researchers developed an educational chatbot using LSTM during the training process. The chatbot achieved 100% accuracy after 200 epochs and recorded a loss rate of 1.14%, demonstrating its ability to recognize patterns in the training data. By leveraging LSTM's capacity to understand context and word relationships, the chatbot could generate accurate and relevant responses.

This study aims to develop a tourism chatbot that enhances the experience of visitors exploring Medan. NLP techniques are applied to interpret user input, while LSTM modeling is implemented to maintain context and retain essential information during interactions. This enables the chatbot to generate more relevant and contextual responses.

Unlike existing chatbots, this study introduces a contextual, AI-driven approach specifically designed for Medan's tourism sector, offering more accurate and personalized interactions. The chatbot is primarily intended for domestic tourists from outside Medan, helping them access relevant travel information, recommendations, and real-time assistance. By addressing the unique needs of non-local visitors, the chatbot supports a more seamless and informative travel experience, making it easier for users to explore the city with confidence.

The main contribution of this study lies in its domain-specific adaptation. Unlike previous chatbot implementations that often rely on generic datasets and broad applications, this study focuses on tourism within a specific region, using localized content and intent-based user interaction. The chatbot consolidates fragmented tourism information into a single, unified interface, making it easier for users to access accurate and reliable data. A comparative analysis with similar works is included to emphasize how this approach differs in contextual design, real-world applicability, and integration with local tourism content. By offering a focused and practical solution to a specific challenge faced by tourists and destination managers in Medan, the chatbot delivers both functional and experiential value, contributing modestly yet meaningfully to the development of AI applications in the tourism sector.

## II. RELATED WORKS

A chatbot is an artificial intelligence program designed to simulate conversation with human users, particularly over the internet [7]. Chatbots have proven to be valuable tools for delivering information across various domains. They can facilitate transactions, provide personalized recommendations, and deliver accurate and timely responses to user inquiries. In education, chatbots assist students with learning, exam preparation, and staying informed about current developments [8]. In healthcare, chatbots improve operational efficiency and support informed decision-making [9]. In tourism, chatbots enable users to book trips, plan vacations, explore new experiences, and reserve hotels with favorable ratings [10].

Chatbots have been implemented across sectors using different models, ranging from rule-based to generative systems [11]. Two common approaches in chatbot development are rule-based and machine learning methods. Rule-based systems rely on predefined patterns from training data and require control over responses to adapt to user input. In contrast, machine learning-based chatbots can process inputs more dynamically and respond in a more complex, context-aware manner.

NLP plays a critical role in chatbot development by enabling effective human-computer interaction. Chatbots have gained popularity due to their adaptable learning capabilities, enhanced user experiences,

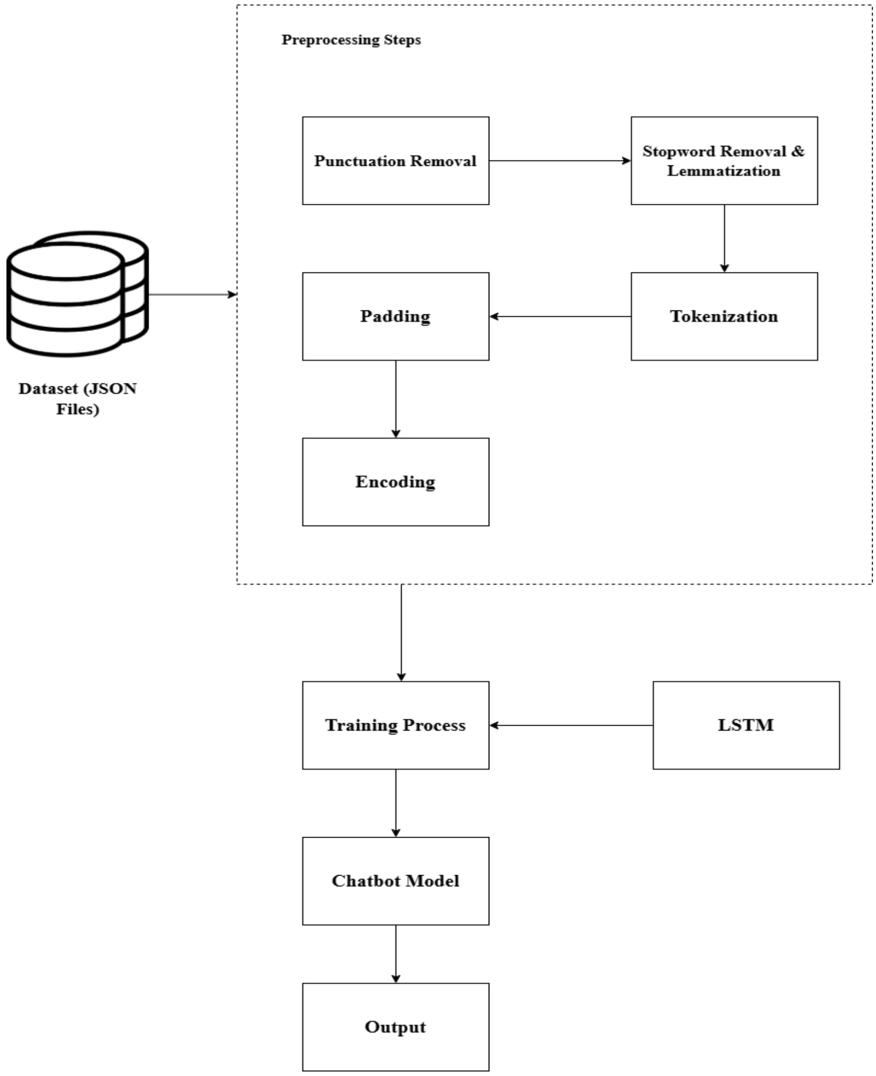


Figure 1. Training Process General Architecture

and improved interpretation through personalized communication. These advances are made possible through text analysis techniques and classification algorithms, which allow systems to process and understand large volumes of online text, including messages and social media content [8].

Artificial intelligence (AI) enhances chatbots by enabling them to handle complex queries [12]. AI-powered chatbots apply machine learning models to generate appropriate responses based on user input. Models such as neural networks and LSTM help chatbots recognize input-output relationships and maintain accuracy. LSTM, a type of Recurrent Neural Network (RNN), is particularly effective in processing sequential data, which is critical for understanding conversational context [13].

In chatbot training, LSTM is widely used to identify conversational patterns and relationships, allowing the system to respond accurately to user input. LSTM models are often employed to predict the next word in a sequence [5]. Research [14] demonstrates that, with LSTM modeling, AI-powered tourism chatbots in Jeddah, Saudi Arabia, can analyze, forecast, and predict data, ensuring quick and responsive interactions.

## III. RESEARCH METHODS

The methods that used in this study are divided into two parts: the model training phase and the prediction model phase.

### *A. Chatbot Model Training Phase*

The architecture of the chatbot modeling training process is illustrated in Figure 1, showing the functional and logical components involved in both training and testing.

{"tag": "imwhat",
"pattern": ["Apa itu istana maimun?", "Boleh ceritakan tentang istana maimun?", "Saya ingin tahu soal istana maimun"],
"respones": ["Istana Maimun adalah istana peninggalan Kesultanan Deli yang dibangun pada tahun 1888. Tempat ini memiliki arsitektur yang unik dengan dominasi warna kuning cerah dan hiasan ukiran yang rumit. Istana ini juga menjadi simbol kebesaran Kesultanan Deli."]},

{"tag": "imwhere",
"pattern": ["Istana Maimun itu di mana", "Alamat istana maimun di mana? ", "Istana maimun itu letaknya di jalan mana?" ],
"respones": ["Istana Maimun berada di Jl. Sultan Ma'moen Al Rasyid No.66, Kesawan, Medan Baru, Kota Medan, Sumatera Utara"]},

{"tag": "imwhen",
"pattern": ["Kapan istana maimun buka / tutup?", "Jam berapa istana maimun buka / tutup?", "Jam operasional istana maimun itu kapan?" ],
"respones": ["Istana Maimun dapat dikunjungi Setiap hari, dari pukul 08.00 - 17.00 WIB"]},

Figure 2. Dataset Example

['a', 'air', 'alam', 'alamat', 'barisan', 'berkunjung', 'bersejarah', 'budaya', 'buka/tutup', 'bukit', 'cambridge', 'camping', 'city', 'danau', 'delipark', 'dikunjungi', 'fie', 'gunung', 'hai', 'hairos', 'halo', 'harga', 'hiburan', 'hiking', 'hotel', 'hutan', 'informasi', 'istana', 'jam', 'jumpa', 'kasih', 'lawang', 'lokasi', 'lumbini', 'maimun', 'mall', 'masuk', 'medan', 'menarik', 'mengunjungi', 'menyenangkan', 'merci', 'murah', 'museum', 'negeri', 'opsi', 'pagi', 'park', 'penginapan', 'plaza', 'rating', 'raya', 'rekomendasi', 'rekomendasikan', 'selamat', 'siang', 'sibolangit', 'sore', 'square', 'sumatera', 'sun', 'taman', 'terima', 'terjun', 'tiket', 'timur', 'tjong', 'toba', 'transportasi', 'trekking', 'utara', 'vihara', 'wajib', 'warna', 'water', 'wisata']

Figure 3. Lemmatization Text Results

25 classes ['accommodation', 'air_terjun', 'alam', 'budaya', 'bukit_lawang', 'cambridge_city_square', 'delipark_mall', 'goodbye', 'greeting', 'hairos_water_park', 'help', 'hiburan', 'istana_maimun', 'medan_mall', 'museum_sumut', 'museum_tjongafie', 'noanswer', 'sejarah', 'sun_plaza', 'taman_alam_lumbini', 'taman_hutan_raya_bukit_barisan', 'toba', 'transportation', 'vihara_gunungtimur', 'wisata_merci']

Figure 4. Unique Classes Result

1) Data Collection Method

Data for this study were collected from various websites that provide tourist information about Medan, including Advistor.com, Google Maps, and local tourism news platforms. The goal is to compile a comprehensive dataset by integrating information from these sources to support the development of the tourism chatbot.

The collected data were compiled into a CSV file, a common format for storing tabular data. Each table contains four rows, with each row representing one data entry. The columns include the name of the place, description, location, ticket prices, operating hours, and rating.

The data in the CSV file were then processed to identify common patterns or keywords related to user queries about tourist destinations. Responses were generated using information such as location names, descriptions, ticket prices, operating hours, and ratings. These responses were organized into a JSON file named *intents.json*, which includes tags, patterns, responses, and context. A sample of this dataset is presented in Figure 2.

From this data collection process, a total of 62 intent tags, 62 responses, and 168 question patterns were gathered. The chatbot is designed to operate in English, as its primary target users are international tourists. Once the data collection is completed, the process continues with data preprocessing.

2) Preprocessing Steps

The intents.json file is part of the training dataset used to develop the chatbot model. Using NLP and the Natural Language Toolkit (NLTK), text preprocessing is carried out through the following steps. The steps are punctuational removal, stopword removal and lemmatization, tokenization, padding, and embedding.

Punctuation removal involves eliminating symbols such as '!' (exclamation mark), ',' (comma), '.' (period), '?' (question mark), and others to enhance the performance of NLP tasks. This step helps standardize the text, reduce complexity, and improve model accuracy in tokenization, stemming, lemmatization, and text classification. By removing punctuation, the text becomes more uniform, allowing NLP models to focus on semantic content without interference from extraneous characters.

During stopword removal, common words such as "and," "the," and "in" are excluded from the text, as they carry minimal semantic value and may obstruct meaningful analysis. This enhances efficiency and improves the accuracy of tasks such as text classification and information retrieval.

Lemmatization then converts inflected words into their base or canonical forms, helping the model treat different word forms as a single entity and improving overall performance [15]. This normalization step is particularly useful for tasks involving text analysis and information retrieval. As a result of this process, 76 unique lemmatized words were identified. The complete list is shown in Figure 3. Next, dataset classes were sorted alphabetically, and duplicates were removed to produce a list of 25 distinct classes, as illustrated in Figure 4. Following word preprocessing and class sorting, the document creation phase begins. This step involves pairing each pattern with its corresponding intent, resulting in a total of 415 documents. This stage is critical for preparing the data for subsequent processing and analysis.

Tokenization is a fundamental step in NLP that converts human-readable text into a sequence of discrete tokens usable by statistical models [16]. Depending on the desired granularity, tokens may consist of words, phrases, or characters. This process simplifies unstructured text into analyzable units, enabling more efficient processing by the model.

Padding is applied to numerical sequences to ensure consistent input lengths across the dataset. This is essential because most models require fixed-length inputs, and batch-based training relies on sequences within each batch having uniform dimensions.

The last is embedding, model parameters are configured to optimize performance. These include input_shape, representing the length of the input sequence; output_length, indicating the number of output classes; and vocabulary, referring to the number of unique words in the dataset. When using an embedding layer, the vocabulary size helps define the dimensionality of word representations, allowing the model to better capture semantic relationships among words.

3) Model Creation

The training process is carried out over multiple epochs, with each epoch consisting of several iterations that process subsets of the dataset in batches. The batch size determines the number of samples processed before the model updates its weights, balancing learning efficiency with computational constraints. During training, the model continuously adjusts its internal parameters based on the calculated loss, improving its ability to detect patterns and generate accurate predictions. This iterative optimization reduces errors and enhances overall model performance. Once training is complete, the LSTM model is saved, preserving the learned weights and architecture. The saved model can then be loaded during the prediction phase to process new, unseen data and produce forecasts or classifications based on prior training.

### *B. Chatbot Prediction Model Phase*

After the chatbot model is trained, it enters the prediction model testing phase. The workflow of this process is shown in Figure 5. In this phase, user input undergoes several preprocessing steps, including punctuation removal, conversion to lowercase, and elimination of irrelevant characters. The cleaned text is then tokenized and padded to meet the model’s input requirements.

The model analyzes the input and calculates the probability for each potential response label. These probabilities reflect the likelihood of each response being appropriate for the given input. The predicted probabilities are then processed using a LabelEncoder, which converts them into recognizable response labels. After identifying the most suitable label, the chatbot selects a response from a predefined list associated with that label. To increase variation and create more natural interactions, the chatbot randomly chooses one of the possible responses instead of using the same reply every time.

## IV. Results and Discussion

### *A. Results*

The model was trained over 100 epochs, resulting in a final training accuracy of 0.8431 and a training loss of 0.7594. The validation accuracy reached 0.7714, with a corresponding validation loss of 2.4233.

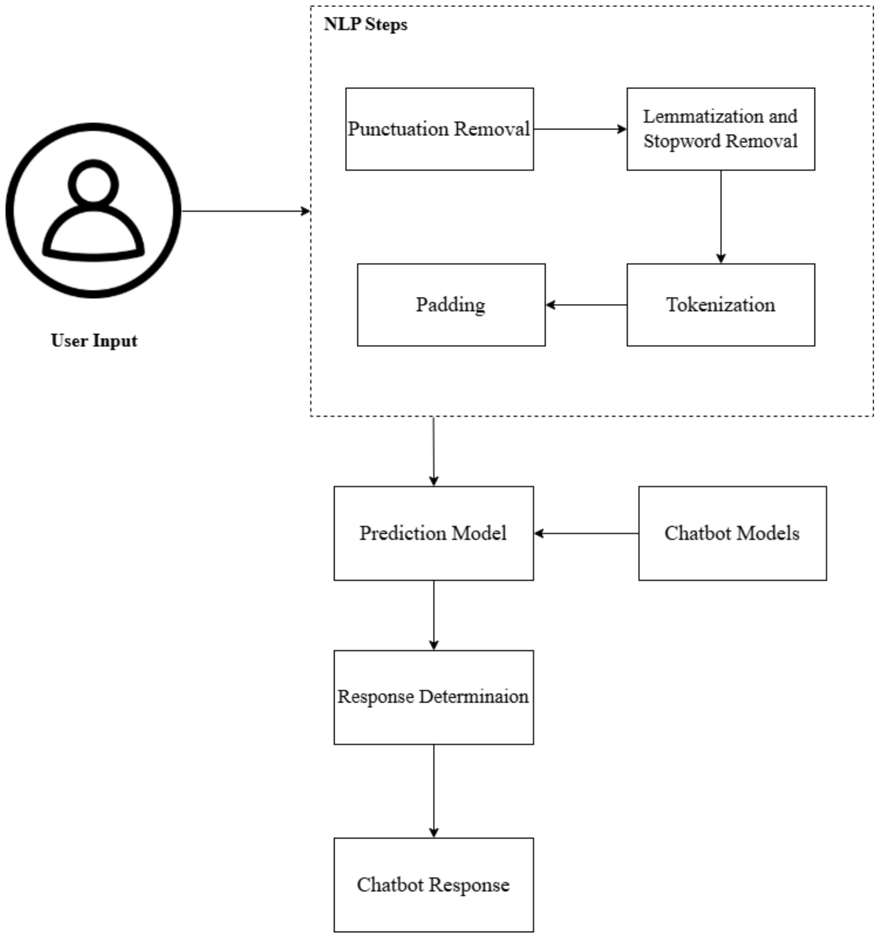


Figure 5. Prediction Model Workflow

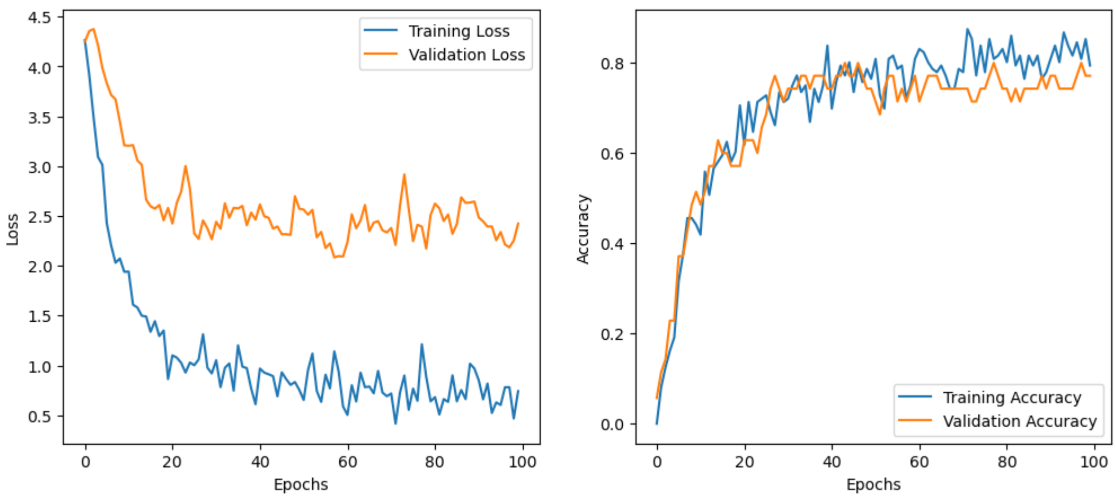


Figure 2 Training and Validation Accuracy and Loss

Figure 6 presents the accuracy and loss trends throughout the training process, illustrating the model's learning progression.

As shown in Figure 6, both training and validation losses decreased significantly, indicating effective learning, despite some fluctuations during the later epochs. The consistency between training and validation accuracy suggests that the model generalizes well to unseen data. These results demonstrate that the model effectively interprets and learns complex patterns, leading to accurate predictions and improved performance in real-world applications.

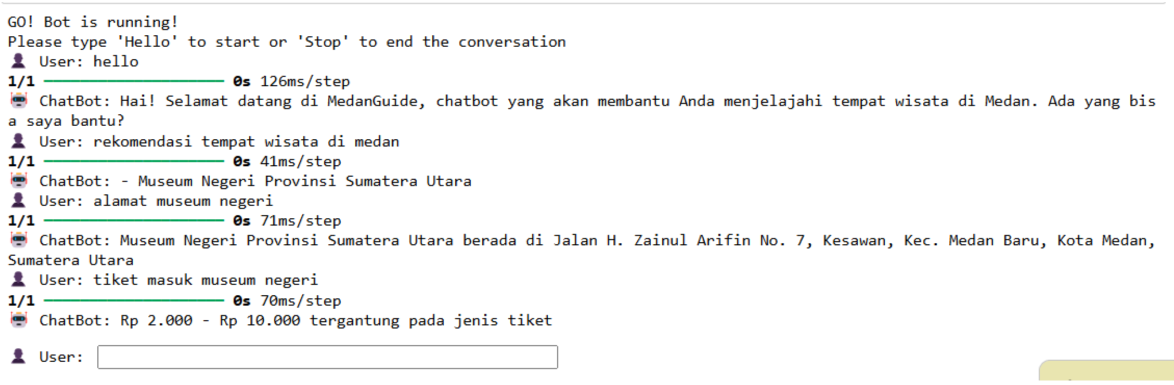


Figure 7. Model Prediction Testing Results

This study also evaluates the predictive capabilities of the trained model by testing it with a set of questions related to tourist information in the city of Medan. The model was given various queries concerning tourist attractions, accommodations, local transportation, and other relevant topics to assess its ability to generate accurate and meaningful responses. The results indicate that the model effectively interprets input and delivers well-structured, contextually appropriate answers based on the information it has learned. It also demonstrates a strong ability to process and analyze queries, ensuring that the responses align with expected outcomes. The entire prediction process, including the interaction between the model and the test questions, is illustrated in Figure 7, which visually represents how the system processes inputs and generates responses.

This study also conducted end-to-end testing to evaluate the chatbot's ability to understand user queries. The testing process involved using a labeled dataset to assess the chatbot's accuracy in identifying user intents. A total of 312 questions were tested, covering all predefined intents. Each intent was represented by five questions to evaluate how effectively the chatbot captured and classified user input. The model achieved a testing accuracy of 75.64%, indicating correct responses in most cases, with a macro average precision of 84.93%, recall of 74.65%, and an F1 score of 75.62%. These results reflect balanced performance across all intent categories and identify areas for future refinement to handle more diverse or ambiguous queries.

Based on the testing results, the chatbot was further refined by incorporating more suitable sentences and responses to enhance its accuracy and overall effectiveness. The finalized model can be saved in either *.h5* or *.pkl* (pickle) format, allowing for easy reuse and deployment. This trained model can then be integrated into AI-powered chatbot applications on various platforms, including websites and Android devices. By adopting this approach, the chatbot becomes more adaptable and accessible, offering a valuable tool to improve user interaction and communication.

### *B. Discussion*

The results of this study demonstrate that the developed chatbot model effectively learns and generalizes tourism-related information specific to the city of Medan. The training process showed a substantial reduction in both training and validation losses over 100 epochs, indicating that the model successfully captured meaningful patterns from the data. With a training accuracy of 84.31% and a validation accuracy of 77.14%, the chatbot shows strong performance in understanding user queries and delivering relevant responses. The consistency between training and validation accuracy further suggests that the model generalizes well to unseen data, supporting reliable user interactions.

The chatbot was evaluated using a variety of queries related to tourist attractions, accommodations, local transportation, and other travel-related topics. The results confirm the model's ability to interpret inputs accurately and provide well-structured, contextually appropriate answers. The integration of NLP and LSTM techniques is key to maintaining conversational context, which enhances the overall user experience. This ensures that tourists receive accurate and relevant information, streamlining the travel planning process.

In addition to training and validation metrics, the model underwent an end-to-end testing process and achieved a testing accuracy of 75.64%. These results demonstrate the model's capacity to apply its learning in real-world scenarios and maintain consistent communication across diverse topics.

Despite its strong performance, the model offers room for improvement. The slightly higher validation loss (2.4233) compared to the training loss (0.7594) suggests potential overfitting or insufficient data diversity. Future refinements could involve expanding the training dataset or optimizing hyperparameters to improve generalization further. However, the stable accuracy levels indicate that the model is both robust and effective for practical use.

Overall, the findings highlight the chatbot's efficiency in processing and analyzing tourism-related queries, making it a useful tool for enhancing visitor experiences in Medan. By organizing and centralizing information, the chatbot addresses the issue of fragmented and inconsistent tourism data, ensuring tourists have access to reliable, contextually relevant details. Future development could focus on expanding the knowledge base, integrating real-time updates, and improving conversational fluency to further enhance its impact on the tourism industry.

## V. CONCLUSION

This study proposed a chatbot to provide tourism information for Medan City. A comprehensive dataset was compiled from various sources to support chatbot development. The model was trained using LSTM over 100 epochs, achieving a training accuracy of 84.31% and a training loss of 0.7594. The validation accuracy reached 77.14%, with a loss of 2.4233. These results demonstrate the model's effectiveness in minimizing the gap between actual and predicted values, as well as its ability to learn patterns and generate accurate responses.

The chatbot's capacity to deliver well-structured, contextually appropriate answers makes it a valuable asset for Medan's tourism sector. By consolidating tourism-related data, it addresses the challenge of fragmented and inconsistent information, improving accessibility for both tourists and destination managers. The integration of NLP and LSTM allows the chatbot to maintain conversational context, resulting in a more engaging and informative user experience.

Additionally, the chatbot underwent an end-to-end testing process involving 312 queries covering all predefined intents, with each intent represented by five questions. The model achieved a testing accuracy of 75.64%, confirming its capability to interpret real user input and respond accurately. This result supports its readiness for deployment in real-world tourism applications.

Looking ahead, there are important considerations for implementing generative models and enhancing data interaction in chatbots. Future improvements should focus on expanding the knowledge base, integrating real-time updates, and refining conversational abilities to deliver more relevant and personalized information. These enhancements will strengthen the chatbot's role as a reliable, efficient tool for assisting tourists and contributing to the continued growth of Medan's tourism industry.

## REFERENCES

[1] M. E. Putri, "Peran Sektor Pariwisata Terhadap Ppendapatan Asli Daerah (Studi Kasus Kabupaten/Kota Provinsi Sulawesi Selatan)," Mar. 2020.

[2] L. Laranjo *et al.*, "Conversational agents in healthcare: a systematic review," *Journal of the American Medical Informatics Association*, vol. 25, no. 9, pp. 1248–1258, Sep. 2018, doi: 10.1093/jamia/ocy072.

[3] M.-H. Su, C.-H. Wu, K.-Y. Huang, Q.-B. Hong, and H.-M. Wang, "A chatbot using LSTM-based multi-layer embedding for elderly care," in *2017 International Conference on Orange Technologies (ICOT)*, IEEE, Dec. 2017, pp. 70–74. doi: 10.1109/ICOT.2017.8336091.

[4] D. Calvaresi, A. Ibrahim, J.-P. Calbimonte, R. Schegg, E. Fragniere, and M. Schumacher, "The Evolution of Chatbots in Tourism: A Systematic Literature Review," in *Information and Communication Technologies in Tourism 2021*, Cham: Springer International Publishing, 2021, pp. 3–16. doi: 10.1007/978-3-030-65785-7_1.

[5] P. Anki, A. Bustamam, H. S. Al-Ash, and D. Sarwinda, "Intelligent Chatbot Adapted from Question and Answer System Using RNN-LSTM Model," *J Phys Conf Ser*, vol. 1844, no. 1, p. 012001, Mar. 2021, doi: 10.1088/1742-6596/1844/1/012001.

[6] E. Kasthuri and S. Balaji, "A Chatbot for Changing Lifestyle in Education," in *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, IEEE, Feb. 2021, pp. 1317–1322. doi: 10.1109/ICICV50876.2021.9388633.

[7] E. Adamopoulou and L. Moussiades, "Chatbots: History, technology, and applications," *Machine Learning with Applications*, vol. 2, p. 100006, Dec. 2020, doi: 10.1016/j.mlwa.2020.100006.

[8] O. Zahour, "Towards a Chatbot for educational and vocational guidance in Morocco: Chatbot E-Orientation," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 2, pp. 2479–2487, Apr. 2020, doi: 10.30534/ijatcse/2020/237922020.

[9] C. Bulla, "A Review of AI Based Medical Assistant Chatbot," 2020, doi: 10.5281/zenodo.3902215.

[10] A. Alnefaie, S. Singh, B. Kocaballi, and M. Prasad, "An Overview of Conversational Agent: Applications, Challenges and Future Directions," in *Proceedings of the 17th International Conference on Web Information Systems and Technologies*, SCITEPRESS - Science and Technology Publications, 2021, pp. 388–396. doi: 10.5220/0010708600003058.

[11] T. A. Baha, M. El Hajji, Y. Es-Saady, and H. Fadili, "Towards highly adaptive Edu-Chatbot," *Procedia Comput Sci*, vol. 198, pp. 397–403, 2022, doi: 10.1016/j.procs.2021.12.260.

[12] E. Adamopoulou and L. Moussiades, "An Overview of Chatbot Technology," 2020, pp. 373–383. doi: 10.1007/978-3-030-49186-4_31.

[13] S. A. Thorat and V. Jadhav, "A Review on Implementation Issues of Rule-based Chatbot Systems," *SSRN Electronic Journal*, 2020, doi: 10.2139/ssrn.3567047.

[14] A. S. Lokman and M. A. Ameedeen, "Modern Chatbot Systems: A Technical Review," 2019, pp. 1012–1023. doi: 10.1007/978-3-030-02683-7_75.

[15] R. Alotaibi, A. Ali, H. Alharthi, and R. Almehamdi, "AI Chatbot for Tourist Recommendations: A Case Study in the City of Jeddah, Saudi Arabia," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 14, no. 19, p. 18, Nov. 2020, doi: 10.3991/ijim.v14i19.17201.

[16] O. Toporkov and R. Agerri, "On the Role of Morphological Information for Contextual Lemmatization," Feb. 2023, [Online]. Available: http://arxiv.org/abs/2302.00407

[17] H. van Halteren, Ed., *Syntactic Wordclass Tagging.* Dordrecht: Kluwer Academic Publishers, 1999.

[18] C. W. Schmidt, V. Reddy, H. Zhang, A. Alameddine, O. Uzan, Y. Pinter, and C. Tanner, "Tokenization Is More Than Compression," *Proc. 2024 Conf. Empirical Methods in Natural Language Processing*, pp. 678–702, Nov. 2024.

[19] Y. Dang, Y. Liu, E. Yang, G. Guo, L. Jiang, X. Wang, and J. Zhao, "Repeated Padding for Sequential Recommendation," *Proc. 18th ACM Conf. Recommender Syst. (RecSys '24)*, pp. 1–11, Oct. 2024, Bari, Italy. [Online]. Available: https://github.com/KingGugu/RepPad.

[20] L. Benaddi, H. Boulaknadel, and Y. A. Bassir, "Seq2Seq Model-Based Chatbot with LSTM and Attention Mechanism for Enhanced User Interaction," *arXiv preprint arXiv:2501.00049*, Jan. 2024. [Online]. Available: https://arxiv.org/abs/2501.00049

[21] S. Mustakim, "An Intelligent Chatbot for Faculty Administration Using Bidirectional LSTM and Seq2Seq Architecture," in *Proc. 2024 Int. Conf. on Artificial Intelligence and Education*, 2024. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3604281

[22] Y. Wang and X. Zhang, "A Task-oriented Chatbot Based on LSTM and Reinforcement Learning," in *Proc. 2022 ACM Int. Conf. on Intelligent Computing and Human-Computer Interaction*, 2022. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3529649

[23] R. Kumar, A. Pandey, P. Verma, and S. Srivastava, "Development of Medical Chatbots Using Aho-Corasick and Siamese LSTM Networks," in *Lecture Notes in Networks and Systems*, vol. 707, pp. 245–254, 2023. [Online]. Available: https://link.springer.com/chapter/10.1007/978-981-97-8526-1_23

[24] J. Smith and H. Lee, "MarkBot – A Language Model-Driven Chatbot for Interactive Marketing," *J. Business Analytics*, vol. 4, no. 2, pp. 67–80, 2021. [Online]. Available: https://doi.org/10.1007/s10796-021-10184-y

[25] L. Chen, M. Zhao, and F. Tan, "Leveraging RNNs and Transformers for Personalized Mental Health Chatbots," in *Lecture Notes in Computer Science*, vol. 14434, pp. 317–329, 2024. [Online]. Available: https://link.springer.com/chapter/10.1007/978-981-97-8526-1_30