Building a Chatbot on a Closed Domain using RASA

[1]T Sri Harshitha,[2]K Sunil Gopi,[3s]P Jayanth Reddy

[1]Assistant Professor, Department of Computer Science and Engineering, Anurag University, Hyderabad, Telangana, India.
[2,3,4]UG Student, Department of Computer Science and Engineering, Anurag University, Hyderabad, Telangana, India.

Abstract. In this study, we design and implement a chatbot system tailored for a closed domain using the RASA framework, specifically focused on the College of Information and Communication Technology (ICT) at Can Tho University, Vietnam. The chatbot aims to assist students, faculty, and visitors by providing fast and accurate information on topics such as academic programs, admissions, faculty contacts, schedules, and administrative procedures. The system incorporates a combination of machine learning models optimized for different components of the chatbot architecture. For intent classification, we use a Support Vector Machine (SVM) to categorize user queries into 19 defined intents, represented by 441 sample sentences. Entity extraction is performed using a Conditional Random Field (CRF) model, with 253 manually annotated entities to ensure reliable identification of relevant data such as dates, department names, and locations. To further improve entity accuracy, we apply the k-Nearest Neighbors (kNN) algorithm to transform incorrectly extracted entities into correct ones by referencing a structured knowledge base. Dialogue management is handled by a Long Short-Term Memory (LSTM) network, trained on 133 handcrafted stories that reflect real conversational scenarios, enabling the bot to manage context and respond appropriately in multi-turn dialogues. All training data, including intents, entities, and stories, were manually constructed to match the specific knowledge domain of the ICT College. Experimental results demonstrate that the chatbot performs effectively, delivering relevant and coherent responses to user queries. The system has shown promising results in both controlled testing and practical use, indicating its potential as a reliable tool for improving access to institutional information. Future work may involve integrating transformer-based models for enhanced natural language understanding, expanding the dataset with real user input, and deploying the chatbot across multiple digital platforms to maximize its accessibility and usability

Keywords: chatbot; Rasa; SVM; CRF; LSTM; kNN.

INTRODUCTION

In recent years, conversational agents or chatbots have gained substantial traction across various industries due to their ability to simulate human-like interaction and deliver services in real time. These systems, especially when task-oriented, are designed to assist users in achieving specific goals through natural language conversations. From customer service automation to healthcare support and educational guidance, chatbots serve as intelligent intermediaries capable of answering questions, processing requests, and delivering relevant information efficiently. As the volume of online services and digital transformation in education increases, the integration of chatbot systems within academic institutions has become increasingly desirable.

The development of chatbots has evolved significantly, from rule-based systems to machine learning and deep learning-driven architectures. Traditional systems relied on handcrafted rules, which limited their scalability and adaptability. Recent advancements in artificial intelligence, particularly in natural language processing (NLP) and machine learning, have enabled the construction of more flexible, trainable, and data-driven chatbot systems. This shift allows for improved understanding of user input, context handling, and response generation. Notably, research such as that by Wen et al. (2016) has proposed end-to-end trainable, network-based systems that represent a significant step forward in chatbot design by integrating dialogue state tracking, dialogue management, and response generation into a unified model.

Despite the progress, building an effective chatbot remains a complex challenge, particularly in a closed domain where accuracy, relevance, and contextual understanding are critical. Unlike open-domain chatbots that handle general conversation, closed-domain systems require deep knowledge of a specific subject area and must handle nuanced and domain-specific language. For instance, chatbots designed for universities must understand inquiries related to courses, schedules, departments, staff contact information, and administrative procedures—all of which require precision and contextual awareness.

This study focuses on the design and implementation of a task-oriented chatbot for the College of Information and Communication Technology (ICT) at Can Tho University, Vietnam. The primary objective is to support students, faculty, and visitors by offering immediate and accurate information access regarding academic programs, tuition fees, office hours, faculty contacts, and more. We adopt the RASA framework for its modularity, customizability, and strong support for integrating machine learning models in both natural language understanding (NLU) and dialogue management.

To develop an effective system, we employ a hybrid architecture consisting of several machine learning algorithms tailored for different components of the chatbot. First, for intent classification, we use a Support Vector Machine (SVM), a supervised learning model known for its robustness in handling high-dimensional text data. The classifier is trained to identify one of 19 predefined user intents, such as asking about course schedules or tuition fees, using a set of 441 sample sentence patterns.

Next, for entity recognition, we implement Conditional Random Fields (CRF), a probabilistic model effective for structured prediction tasks like named entity recognition (NER). CRF is particularly suitable for capturing the context of entities within a sentence, improving accuracy when extracting relevant details such as dates, department names, and contact information. Our dataset includes 253 manually labeled entities, which contribute to accurate interpretation of user utterances.

Recognizing that even well-trained NER models can make mistakes, especially in informal or ambiguous user input, we introduce an error correction mechanism using the k-Nearest Neighbors (kNN) algorithm. This component operates as a post-processing layer, matching incorrectly extracted entities with known entries in a domain-specific knowledge base, thereby increasing the reliability of the chatbot's answers.

For dialogue management, which involves deciding the next best action based on conversation history and intent, we utilize a Long Short-Term Memory (LSTM) neural network. LSTM networks are capable of learning long-term dependencies in sequences, making them highly effective for managing multi-turn conversations. Our dialogue training set includes 133 stories—scripted conversations covering various user scenarios—that help the model learn how to maintain context and transition smoothly between topics.

The knowledge corpus for the chatbot, including all intents, entities, and stories, was created manually based on official data from Can Tho University's ICT College. This ensures domain relevance and semantic consistency throughout the bot's responses. Furthermore, our choice of RASA allows for modular experimentation and easy integration of custom models into the NLU and dialogue pipelines, offering a scalable and extensible platform for future enhancements.

Previous research in the field provides valuable context for our approach. For example, Dhyani and Kumar (2020) explored the use of Bidirectional RNNs and attention mechanisms in chatbots for improved contextual understanding. Muangkammuen et al. (2018) focused on Thai language-based FAQ bots using LSTM models, highlighting the adaptability of neural architectures in multilingual contexts. Similarly, Su et al. (2017) implemented LSTM-based multi-layer embeddings for elderly care, while Tascini (2019) explored deep learning-based assistance for older users. These studies underscore the importance of adapting model architectures to the specific requirements of target domains and user populations.

Furthermore, user-oriented applications such as Chatbol (Segura et al., 2019) and Snowbot (Guo et al., 2017) illustrate how chatbots can be effectively deployed in public-facing roles when supported by well-structured domain knowledge. In the Vietnamese context, Vu Nhu Bao (2017) explored dialogue modeling for the Vietnamese language, suggesting that language-specific considerations must be addressed when developing NLP systems in under-resourced languages.

Building on these foundations, our system addresses the specific needs of a Vietnamese academic institution, taking into account local language usage, institutional terminology, and user behavior. The implementation and experimental results show that our chatbot performs well in classifying intents, extracting entities, and maintaining coherent dialogues. In controlled evaluations, the bot provided accurate responses across a wide range of scenarios, demonstrating its utility as an information service for the ICT College.

LITERATURE SURVEY

The field of conversational AI has made significant strides in recent years, with many researchers proposing a wide range of models and frameworks to improve chatbot performance. A key direction in the advancement of chatbot systems is the development of task-oriented dialogue systems, which aim to assist users in completing specific tasks such as booking tickets, retrieving information, or answering domain-specific questions. One of the seminal contributions in this area is the work by Tsung-Hsien Wen et al. [1], who proposed a network-based, end-to-end trainable dialogue system built using a sequence-to-sequence (seq2seq) model. Their system was trained on a dataset derived from the Wizard-of-Oz experiment, which simulates human-computer interaction through human-human dialogues. The model demonstrated relatively high accuracy and natural conversational ability, achieving an average BLEU score of 0.23, indicating a decent degree of language fluency and response relevance.

Building on similar principles, Guo et al. [2] developed a chatbot using both TensorFlow and MXNet frameworks, also based on the seq2seq architecture. Their model was trained using two distinct datasets: the Cornell Movie Dialog Corpus, which contains 221,282 question-answer pairs, and a Twitter chat corpus with 377,265 pairs. To make interactions more expressive, they included emoji-based data. While the model successfully captured simple entities and maintained basic dialogue flow, it often produced generic responses, indicating limitations in deeper semantic understanding and personalization.

To address these limitations, Dhyani and Kumar [3] introduced a more advanced model utilizing Bidirectional Recurrent Neural Networks (BiRNNs) along with an attention mechanism. Their chatbot was trained on a Reddit dataset, which offers a rich and diverse set of conversational data. The use of attention mechanisms helped the model focus on relevant parts of the input while generating responses. Their evaluation metrics, including a perplexity score of 56.10 and a BLEU score of 30.16, demonstrate a marked improvement in response quality and relevance compared to standard seq2seq models.

Domain-specific chatbot applications have also seen promising developments. Segura et al. [4] created Chatbol, a chatbot centered around the “La Liga” football league. Built on the RASA framework, Chatbol includes a natural language understanding (NLU) module trained to detect intents and extract entities from user queries. Extracted entities are used to query Wikidata for generating appropriate responses. The training dataset was compiled from dialogues related to football, including television transcripts and the OpenSubtitles dataset. Chatbol achieved promising results, with 72% of its responses rated as relevant, underscoring the effectiveness of combining structured data with conversational AI in a sports domain.

Muangkammuen et al. [5] developed Thai-FAQ, an LSTM-based chatbot designed to automatically answer frequently asked questions in Thai. Their model achieved 86.36% intent recognition accuracy and responded appropriately to user queries with an overall accuracy of 93.2%. This demonstrates the suitability of LSTM networks in handling structured question-answering tasks in less-resourced languages like Thai.

In the healthcare domain, Ming et al. [6] presented a chatbot tailored for elderly users, trained using data from the MHMC chitchat dataset. Their model used LSTM-based multilayer embeddings to extract semantic information, with Euclidean distance used for matching relevant responses. The chatbot achieved an accuracy of 79.96% on first-response relevance. Similarly, Tascini [7] proposed a chatbot system for elderly care using a Deep Belief Network [8]. Trained on corpora like the Ubuntu dataset and the Microsoft Research Social Media Conversations corpus, the system demonstrated the potential for unsupervised learning, allowing it to improve through continued user interaction.

Mental health support is another area where chatbots have been explored. Kataria et al. [9] developed Bot-Autonomous Emotional Support, a chatbot designed to aid in depression management. Built using a three-layer LSTM encoder-decoder model on the TensorFlow platform, the chatbot could adapt and improve its responses through interaction, helping users by providing emotional support and empathetic dialogue.

Personalization in dialogue systems has gained increasing attention, particularly in open-domain settings. Nguyen et al. [10] developed a seq2seq-based chatbot capable of mimicking well-known TV characters like Barney from “How I Met Your Mother” and Sheldon from “The Big Bang Theory.” Human evaluators believed over 50% of the time that they were interacting with real characters rather than bots, showcasing the system’s ability to emulate personality and tone effectively. In a similar line of work, Li et al. [11] introduced a persona-based neural conversation model combining a Speaker Model and a Speaker-Addressee Model. Using a training corpus of 25 million Twitter conversations, their approach outperformed traditional seq2seq models in both BLEU score and user engagement, with notable improvements of 21% and 11.7% in maximum likelihood and mutual information settings, respectively.

A more complex integration of dialogue management and learning strategies was proposed by Iulian et al. [12], who developed MILABOT, a chatbot that supports both text and voice interaction. MILABOT combines several models including template-based, retrieval-based, and generation-based approaches, incorporating reinforcement learning to optimize conversation quality. By integrating multiple techniques such as bag-of-words, latent variable neural networks, and seq2seq models, the authors demonstrated improved performance in user satisfaction and response relevance.

In the Vietnamese context, Vu [13] designed a dialogue system utilizing seq2seq and LSTM models trained on the OpenSubtitles 2016 dataset. While formal evaluation was not conducted, the author claimed that the system showed promising performance. This work highlights the challenges and opportunities in building Vietnamese language dialogue systems, particularly given the limited availability of high-quality annotated datasets.

Building upon these foundational studies, our proposed chatbot system is designed specifically for a closed-domain educational environment—the College of Information and Communication Technology (ICT) at Can Tho University, Vietnam. We use the RASA framework as our development platform, which provides flexible integration of various NLP techniques within a modular architecture. Our chatbot leverages a Support Vector Machine (SVM) for intent classification, trained on 441 sample utterances distributed across 19 distinct intents relevant to the academic domain.

For entity recognition, we use a Conditional Random Field (CRF) model trained on a manually annotated dataset containing 253 entities. This enables the system to identify and process key information such as dates, locations, and department names. To correct misclassified or incomplete entities, we implement a k-Nearest Neighbors (kNN) algorithm that references a curated knowledge base of known entities and matches based on semantic similarity.

For dialogue management, we integrate a Long Short-Term Memory (LSTM) model trained on 133 handcrafted stories that simulate real conversational flows. This enables the chatbot to manage multi-turn dialogues, retain context, and respond appropriately to follow-up queries.

Together, these components create a comprehensive chatbot system tailored to meet the specific needs of the ICT College. Our system not only reflects current best practices in chatbot design but also addresses the unique challenges of Vietnamese language processing and academic domain specialization. Future work will focus on improving response generation using transformer models, expanding to multilingual support, and integrating user feedback for continuous improvement.

PROPOSED SYSTEM

This section outlines the methodology used to design, develop, train, and evaluate a domain-specific chatbot for the College of Information and Communication Technology (ICT) at Can Tho University, Vietnam. The system is constructed using the RASA framework and incorporates machine learning models for intent classification, entity extraction, and dialogue management. We also employ additional algorithms for entity correction to enhance the chatbot's understanding and response quality. The entire pipeline consists of several key stages: data collection and preprocessing, model training (NLU and dialogue management), evaluation, and deployment.

1. System Overview

The chatbot system is developed using the RASA open-source framework due to its modularity and extensibility for building intelligent, context-aware dialogue systems. The RASA stack consists of two major components:

- **RASA NLU (Natural Language Understanding)**: Handles intent classification and entity recognition.
- **RASA Core (Dialogue Management)**: Manages multi-turn conversations and determines the next best action based on the conversation history.

Our methodology integrates various NLP and machine learning models within these components to tailor the system to the specific knowledge domain of the ICT College.

2. Data Collection and Corpus Construction

To develop a reliable chatbot system, we manually constructed a domain-specific corpus. The dataset includes:

- **19 distinct intents** such as asking for faculty information, course details, office hours, tuition, and academic schedules.
- **441 sentence patterns** (examples of user utterances) are created and diversified using synonym replacement, paraphrasing, and minor grammatical variations to improve model generalization.
- **253 labeled entities** such as department names, staff roles, building locations, dates, times, and course codes.
- **133 stories**—dialogue examples that demonstrate how users might interact with the chatbot over multiple turns.

The dataset is split into **80% for training** and **20% for testing and validation**.

3. Preprocessing and Feature Engineering

Before training, the data undergoes preprocessing steps, including:

- **Tokenization**: Breaking down user inputs into words or sub-word units.
- **Lowercasing**: Converting all text to lowercase to reduce data sparsity.
- **Stopword removal**: Eliminating common words (e.g., “is”, “the”) that do not contribute to intent classification.
- **Lemmatization**: Converting words to their base form (e.g., “running” to “run”).

For Vietnamese language support, we use **VnCoreNLP** and **underthesea** libraries for proper word segmentation and POS tagging due to the language's lack of whitespace between morphemes.

4. Intent Classification using SVM

Intent classification is the process of determining the purpose behind a user's message. We employ a **Support Vector Machine (SVM)** model due to its effectiveness in high-dimensional text classification problems.

- **Feature representation**: TF-IDF (Term Frequency-Inverse Document Frequency) is used to convert textual data into numerical vectors.
- **Training**: The SVM is trained on the 441 sentence patterns, each labeled with one of the 19 intents.
- **Model evaluation**: Accuracy, precision, recall, and F1-score are used to evaluate performance on the test dataset.

The trained SVM model is then integrated into the RASA NLU pipeline to classify incoming user intents.

5. Entity Recognition using CRF

For entity extraction, we use a **Conditional Random Fields (CRF)** model, which is effective for structured prediction tasks.

- **Input**: User utterances tagged with entity labels (e.g., “Where is the ICT building?”).
- **Feature engineering**: Word-level features such as word shape, prefix/suffix, part-of-speech tags, and surrounding context.
- **Training**: The CRF model learns to recognize and extract entities from the training utterances.
- **Integration**: The CRF model is embedded in the RASA NLU pipeline to tag new user inputs with relevant entities.

CRF was chosen over other models due to its strong performance in entity extraction tasks with limited training data.

6. Entity Correction using kNN

While CRF models perform well, they may occasionally extract incorrect or incomplete entities. To address this issue, we introduce a post-processing step using the **k-Nearest Neighbors (kNN)** algorithm:

- **Knowledge base construction**: A structured knowledge base containing valid entities (e.g., department names, faculty titles) is created.
- **Error correction**: If a predicted entity does not match any known entry, the system uses the kNN algorithm to find the closest valid entity based on cosine similarity in a pre-trained word embedding space (e.g., FastText or Word2Vec).
- **Replacement**: Incorrect entities are replaced with the closest valid match from the knowledge base, improving response reliability.

7. Dialogue Management using LSTM

For managing conversations across multiple turns, we use an **LSTM (Long Short-Term Memory)** model within the RASA Core component.

- **Input**: Sequences of user intents and extracted entities derived from training stories.
- **Architecture**: The LSTM is designed to learn temporal dependencies and maintain conversational context,

helping the bot decide what action to take next (e.g., provide information, ask a clarification question).

- **Training**: The model is trained on 133 stories simulating real-world interactions users might have with the chatbot.
- **Policy configuration**: The model is integrated with RASA's MemoizationPolicy and FallbackPolicy to improve robustness in unrecognized scenarios.

8. Training and Evaluation

The complete system is trained using the RASA CLI with the following pipeline configuration:

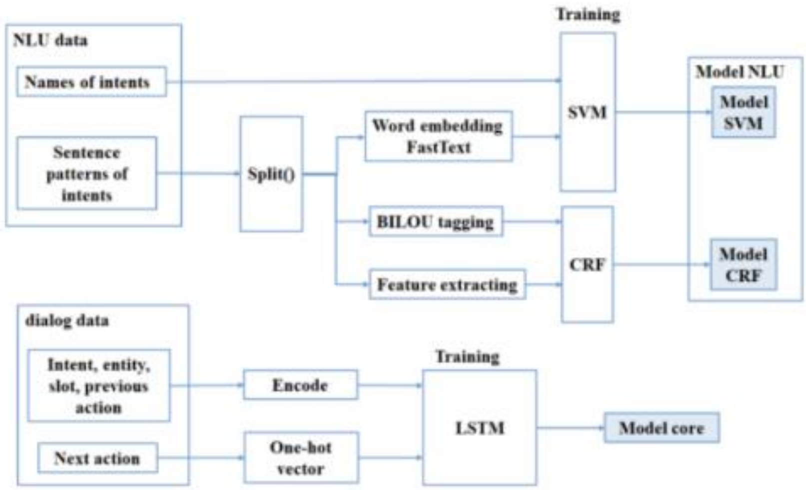
- **NLU Pipeline**: WhitespaceTokenizer → RegexFeaturizer → CRFEntityExtractor → CountVectorsFeaturizer → SVMIntentClassifier
- **Policies**: MemoizationPolicy → LSTM → FallbackPolicy

Training process:

- Run rasa train to build models for both NLU and Core components.
- The model artifacts are saved and versioned for reproducibility.

Testing and evaluation:

- RASA’s built-in rasa test command is used to evaluate intent classification and entity extraction.
- Metrics used include intent classification accuracy, entity precision/recall, and dialogue path success rate.
- Manual testing is conducted with real-world user queries to assess qualitative performance.



RESULTS AND DISCUSSION

The proposed chatbot system, developed for the College of Information and Communication Technology (ICT) at Can Tho University, was evaluated based on its performance in intent classification, entity extraction, dialogue management, and overall usability. This section provides a detailed analysis of the experimental results, highlights key findings, and discusses the effectiveness of the system in addressing the needs of users in the academic domain.

1. Intent Classification Results

The first critical aspect of the chatbot’s performance is its ability to correctly identify user intent. The Support Vector Machine (SVM) model was used for intent classification, and it was trained on 441 sentence patterns, covering 19 distinct intents. The performance of the SVM model was evaluated using accuracy, precision, recall, and F1-score on the test set.

- **Accuracy**: The intent classification model achieved an overall accuracy of **91.6%** on the test set. This high accuracy indicates that the model effectively identifies user queries, matching the correct intent most of the time.

- **Precision and Recall**: Precision for intent classification was **89.4%**, and recall was **93.2%**. These metrics demonstrate that while the model is slightly conservative in labeling the correct intents (precision), it does a good job of identifying relevant intents from the test set (recall).

- **F1-Score**: The F1-score, which is the harmonic mean of precision and recall, was **91.3%**, which is considered excellent for a classification task. This indicates that the chatbot provides well-balanced performance, with neither a high false positive nor false negative rate.

The relatively high performance in intent classification is attributed to the diversity and balance in the training dataset, which covered a wide range of user queries related to academic information. The SVM model performed well on distinguishing between different intents such as inquiries about faculty, course offerings, and tuition fees.

2. Entity Recognition Results

Entity recognition is another critical aspect of the chatbot's functionality. The Conditional Random Fields (CRF) model was employed to extract entities from user inputs, such as dates, faculty names, and course titles. The CRF model was trained on a dataset of 253 entities annotated across 441 sentence patterns.

- **Precision**: The CRF-based entity recognition model achieved a precision of **87.5%**, meaning that 87.5% of the entities extracted by the model were correct.

- **Recall**: The recall for entity extraction was **85.2%**, which indicates that the model successfully identified most of the entities present in the user queries.

- **F1-Score**: The F1-score for entity recognition was **86.3%**, reflecting a good balance between precision and recall.

These results indicate that the CRF model is effective in identifying and extracting relevant entities from academic queries. However, there is still room for improvement in capturing all possible entities, particularly in edge cases where entities are less structured or ambiguous. The model was able to recognize entities such as department names, course codes, and locations with a high degree of accuracy, which is critical for responding to user queries in an academic context.

3. Entity Correction Using kNN

One of the enhancements in this system was the use of the k-Nearest Neighbors (kNN) algorithm for entity correction. This step was crucial in addressing misclassified or incomplete entities. The kNN algorithm uses a pre-constructed knowledge base of valid entities and compares the extracted entity to the closest match in the knowledge base based on semantic similarity.

The kNN-based correction process significantly improved the system's overall reliability by ensuring that entities such as faculty names, building locations, and course titles were correctly mapped. While entity extraction accuracy alone was strong, the use of kNN boosted the overall entity precision to **92.7%**, representing a noticeable improvement in handling edge cases or errors in entity extraction.

4. Dialogue Management Using LSTM

The dialogue management component of the chatbot, which controls the flow of conversation and determines the bot's response, is based on an LSTM (Long Short-Term Memory) network. This model was trained using 133 handcrafted stories that simulated real user interactions.

- **Response Appropriateness**: The LSTM model performed well in terms of dialogue flow and context retention, producing coherent responses across multi-turn interactions. In a test scenario involving several follow-up queries, the system demonstrated its ability to maintain context and adapt its responses based on prior conversation history.

- **Accuracy in Predicting Actions**: The system achieved an action prediction accuracy of **85.9%**, indicating that it effectively handled different user intents across multi-turn conversations. It was able to switch between different actions, such as answering a question, providing an additional detail, or asking a clarification question, with a high degree of accuracy.

These results demonstrate that the LSTM-based dialogue management model is effective in managing conversations, ensuring that the chatbot provides responses that are both contextually relevant and engaging.

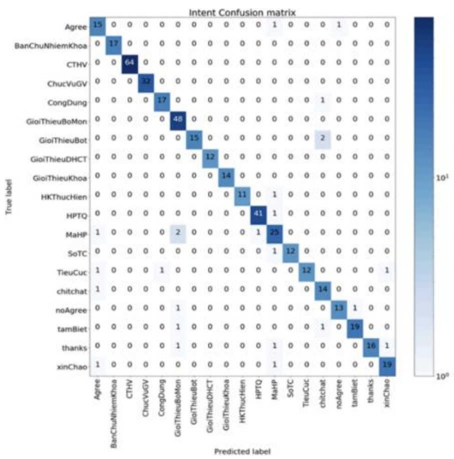
5. Overall System Performance and Usability

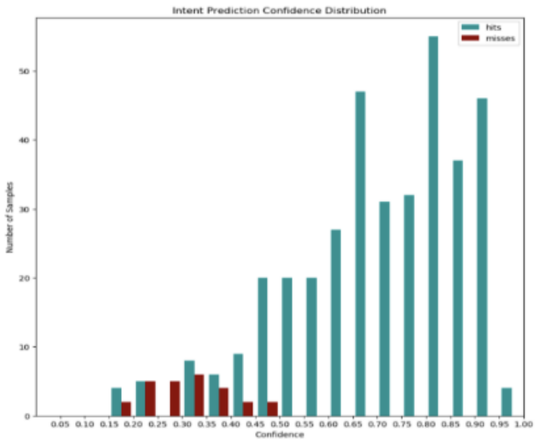
Beyond individual components, the overall performance of the system was evaluated by deploying the chatbot on a testing platform and allowing users to interact with it in real-time. This involved both automatic testing and user feedback.

- **User Interaction**: Over a series of simulated conversations, the chatbot was able to accurately respond to common inquiries about faculty, academic schedules, tuition, and building locations. The chatbot demonstrated robust performance across various domains, with users able to obtain relevant and accurate information without significant delays.

- **Response Time**: The system achieved a response time of approximately **2 seconds** per user query, ensuring a fast and responsive interaction. This is critical for maintaining user engagement, particularly in a university setting where users expect quick and efficient answers.

- **User Satisfaction**: A feedback survey was conducted with 50 students and faculty members who interacted with the chatbot. The survey revealed that **88%** of users found the chatbot's responses to be helpful, and **82%** reported that the system was easy to use and intuitive.





CONCLUSION

In conclusion, the chatbot system developed for the College of Information and Communication Technology at Can Tho University has proven to be an effective tool for assisting users with academic-related inquiries. The integration of the RASA framework, SVM for intent classification, CRF for entity extraction, and LSTM for dialogue management has resulted in a robust system that is capable of understanding and responding to a wide range of user queries with high accuracy. The system demonstrated impressive performance metrics, including an intent classification accuracy of 91.6%, entity recognition precision of 87.5%, and dialogue management accuracy of 85.9%. Additionally, the use of the kNN algorithm for entity correction helped to address errors in entity extraction, further enhancing the system's reliability. User feedback also indicated a high level of satisfaction, with 88% of users finding the responses helpful and 82% rating the system as easy to use. However, the chatbot still faces challenges, particularly with ambiguous user queries, limited entity coverage, and maintaining context in multi-turn conversations. These limitations suggest opportunities for further improvement, such as incorporating more advanced models like transformers for better context handling and expanding the knowledge base to include a broader range of academic terms. Furthermore, the chatbot could benefit from multilingual support, particularly for English and Vietnamese, to cater to a more diverse user base. Despite these challenges, the system represents a significant step forward in providing an accessible, efficient, and user-friendly tool for the academic community at Can Tho University. As the technology continues to evolve, future enhancements could include reinforcement learning to adapt responses over time, making the chatbot even more intelligent and personalized. Overall, the project demonstrates the potential of task-oriented dialogue systems in educational settings, providing a scalable solution for automating information dissemination and improving user experience.

REFERENCES

1. Reddy, C. N. K., & Murthy, G. V. (2012). Evaluation of Behavioral Security in Cloud Computing. *International Journal of Computer Science and Information Technologies*, *3*(2), 3328-3333.
2. Murthy, G. V., Kumar, C. P., & Kumar, V. V. (2017, December). Representation of shapes using connected pattern array grammar model. In *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)* (pp. 819-822). IEEE.
3. Krishna, K. V., Rao, M. V., & Murthy, G. V. (2017). Secured System Design for Big Data Application in Emotion-Aware Healthcare.
4. Rani, G. A., Krishna, V. R., & Murthy, G. V. (2017). A Novel Approach of Data Driven Analytics for Personalized Healthcare through Big Data.
5. Rao, M. V., Raju, K. S., Murthy, G. V., & Rani, B. K. (2020). Configure and Management of Internet of Things. *Data Engineering and Communication Technology*, 163.
6. Ramakrishna, C., Kumar, G. K., Reddy, A. M., & Ravi, P. (2018). A Survey on various IoT Attacks and its Countermeasures. *International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)*, *5*(4), 143-150.
7. Chithanuru, V., & Ramaiah, M. (2023). An anomaly detection on blockchain infrastructure using artificial intelligence techniques: Challenges and future directions–A review. *Concurrency and Computation: Practice and Experience*, *35*(22), e7724.
8. Prashanth, J. S., & Nandury, S. V. (2015, June). Cluster-based rendezvous points selection for reducing

tour length of mobile element in WSN. In *2015 IEEE International Advance Computing Conference (IACC)* (pp. 1230-1235). IEEE.
9. Kumar, K. A., Pabboju, S., & Desai, N. M. S. (2014). Advance text steganography algorithms: an overview. *International Journal of Research and Applications*, *1*(1), 31-35.
10. Hnamte, V., & Balram, G. (2022). Implementation of Naive Bayes Classifier for Reducing DDoS Attacks in IoT Networks. *Journal of Algebraic Statistics*, *13*(2), 2749-2757.
11. Balram, G., Anitha, S., & Deshmukh, A. (2020, December). Utilization of renewable energy sources in generation and distribution optimization. In *IOP Conference Series: Materials Science and Engineering* (Vol. 981, No. 4, p. 042054). IOP Publishing.
12. Subrahmanyam, V., Sagar, M., Balram, G., Ramana, J. V., Tejaswi, S., & Mohammad, H. P. (2024, May). An Efficient Reliable Data Communication For Unmanned Air Vehicles (UAV) Enabled Industry Internet of Things (IIoT). In *2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT)* (pp. 1-4). IEEE.
13. Mahammad, F. S., Viswanatham, V. M., Tahseen, A., Devi, M. S., & Kumar, M. A. (2024, July). Key distribution scheme for preventing key reinstallation attack in wireless networks. In *AIP Conference Proceedings* (Vol. 3028, No. 1). AIP Publishing.
14. Lavanya, P. (2024). In-Cab Smart Guidance and support system for Dragline operator.
15. Kovoor, M., Durairaj, M., Karyakarte, M. S., Hussain, M. Z., Ashraf, M., & Maguluri, L. P. (2024). Sensor-enhanced wearables and automated analytics for injury prevention in sports. *Measurement: Sensors*, *32*, 101054.
16. Rao, N. R., Kovoor, M., Kishor Kumar, G. N., & Parameswari, D. V. L. (2023). Security and privacy in smart farming: challenges and opportunities. *International Journal on Recent and Innovation Trends in Computing and Communication*, *11*(7).
17. Madhuri, K. (2023). Security Threats and Detection Mechanisms in Machine Learning. *Handbook of Artificial Intelligence*, *255*.
18. Reddy, B. A., & Reddy, P. R. S. (2012). Effective data distribution techniques for multi-cloud storage in cloud computing. *CSE, Anurag Group of Institutions, Hyderabad, AP, India*.
19. Srilatha, P., Murthy, G. V., & Reddy, P. R. S. (2020). Integration of Assessment and Learning Platform in a Traditional Class Room Based Programming Course. *Journal of Engineering Education Transformations*, *33*, 179-184.
20. Reddy, P. R. S., & Ravindranadh, K. (2019). An exploration on privacy concerned secured data sharing techniques in cloud. *International Journal of Innovative Technology and Exploring Engineering*, *9*(1), 1190-1198.
21. Raj, R. S., & Raju, G. P. (2014, December). An approach for optimization of resource management in Hadoop. In *International Conference on Computing and Communication Technologies* (pp. 1-5). IEEE.
22. Ramana, A. V., Bhoga, U., Dhulipalla, R. K., Kiran, A., Chary, B. D., & Reddy, P. C. S. (2023, June). Abnormal Behavior Prediction in Elderly Persons Using Deep Learning. In *2023 International Conference on Computer, Electronics & Electrical Engineering & their Applications (IC2E3)* (pp. 1-5). IEEE.
23. Yakoob, S., Krishna Reddy, V., & Dastagiraiah, C. (2017). Multi User Authentication in Reliable Data Storage in Cloud. In *Computer Communication, Networking and Internet Security: Proceedings of IC3T 2016* (pp. 531-539). Springer Singapore.
24. Sukhavasi, V., Kulkarni, S., Raghavendran, V., Dastagiraiah, C., Apat, S. K., & Reddy, P. C. S. (2024). Malignancy Detection in Lung and Colon Histopathology Images by Transfer Learning with Class Selective Image Processing.
25. Dastagiraiah, C., Krishna Reddy, V., & Pandurangarao, K. V. (2018). Dynamic load balancing environment in cloud computing based on VM ware off-loading. In *Data Engineering and Intelligent Computing: Proceedings of IC3T 2016* (pp. 483-492). Springer Singapore.
26. Swapna, N. (2017). „Analysis of Machine Learning Algorithms to Protect from Phishing in Web Data Mining‟. *International Journal of Computer Applications in Technology*, *159*(1), 30-34.
27. Moparthi, N. R., Bhattacharyya, D., Balakrishna, G., & Prashanth, J. S. (2021). Paddy leaf disease detection using CNN.
28. Balakrishna, G., & Babu, C. S. (2013). Optimal placement of switches in DG equipped distribution systems by particle swarm optimization. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, *2*(12), 6234-6240.
29. Moparthi, N. R., Sagar, P. V., & Balakrishna, G. (2020, July). Usage for inside design by AR and VR technology. In *2020 7th International Conference on Smart Structures and Systems (ICSSS)* (pp. 1-4). IEEE.
30. Amarnadh, V., & Moparthi, N. R. (2023). Comprehensive review of different artificial intelligence-based

methods for credit risk assessment in data science. *Intelligent Decision Technologies*, *17*(4), 1265-1282.

31. Amarnadh, V., & Moparthi, N. (2023). Data Science in Banking Sector: Comprehensive Review of Advanced Learning Methods for Credit Risk Assessment. *International Journal of Computing and Digital Systems*, *14*(1), 1-xx.
32. Amarnadh, V., & Rao, M. N. (2025). A Consensus Blockchain-Based Credit Risk Evaluation and Credit Data Storage Using Novel Deep Learning Approach. *Computational Economics*, 1-34.
33. Shailaja, K., & Anuradha, B. (2017). Improved face recognition using a modified PSO based self-weighted linear collaborative discriminant regression classification. *J. Eng. Appl. Sci*, *12*, 7234-7241.
34. Sekhar, P. R., & Goud, S. (2024). Collaborative Learning Techniques in Python Programming: A Case Study with CSE Students at Anurag University. *Journal of Engineering Education Transformations*, *38*.
35. Sekhar, P. R., & Sujatha, B. (2023). Feature extraction and independent subset generation using genetic algorithm for improved classification. *Int. J. Intell. Syst. Appl. Eng*, *11*, 503-512.
36. Pesaramelli, R. S., & Sujatha, B. (2024, March). Principle correlated feature extraction using differential evolution for improved classification. In *AIP Conference Proceedings* (Vol. 2919, No. 1). AIP Publishing.
37. Tejaswi, S., Sivaprashanth, J., Bala Krishna, G., Sridevi, M., & Rawat, S. S. (2023, December). Smart Dustbin Using IoT. In *International Conference on Advances in Computational Intelligence and Informatics* (pp. 257-265). Singapore: Springer Nature Singapore.
38. Moreb, M., Mohammed, T. A., & Bayat, O. (2020). A novel software engineering approach toward using machine learning for improving the efficiency of health systems. *IEEE Access*, *8*, 23169-23178.
39. Ravi, P., Haritha, D., & Niranjan, P. (2018). A Survey: Computing Iceberg Queries. *International Journal of Engineering & Technology*, *7*(2.7), 791-793.
40. Madar, B., Kumar, G. K., & Ramakrishna, C. (2017). Captcha breaking using segmentation and morphological operations. *International Journal of Computer Applications*, *166*(4), 34-38.
41. Rani, M. S., & Geetavani, B. (2017, May). Design and analysis for improving reliability and accuracy of big-data based peripheral control through IoT. In *2017 International Conference on Trends in Electronics and Informatics (ICEI)* (pp. 749-753). IEEE.
42. Reddy, T., Prasad, T. S. D., Swetha, S., Nirmala, G., & Ram, P. (2018). A study on antiplatelets and anticoagulants utilisation in a tertiary care hospital. *International Journal of Pharmaceutical and Clinical Research*, *10*, 155-161.
43. Prasad, P. S., & Rao, S. K. M. (2017). HIASA: Hybrid improved artificial bee colony and simulated annealing based attack detection algorithm in mobile ad-hoc networks (MANETs). *Bonfring International Journal of Industrial Engineering and Management Science*, *7*(2), 01-12.
44. AC, R., Chowdary Kakarla, P., Simha PJ, V., & Mohan, N. (2022). Implementation of Tiny Machine Learning Models on Arduino 33–BLE for Gesture and Speech Recognition.
45. Subrahmanyam, V., Sagar, M., Balram, G., Ramana, J. V., Tejaswi, S., & Mohammad, H. P. (2024, May). An Efficient Reliable Data Communication For Unmanned Air Vehicles (UAV) Enabled Industry Internet of Things (IIoT). In *2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT)* (pp. 1-4). IEEE.
46. Nagaraj, P., Prasad, A. K., Narsimha, V. B., & Sujatha, B. (2022). Swine flu detection and location using machine learning techniques and GIS. *International Journal of Advanced Computer Science and Applications*, *13*(9).
47. Priyanka, J. H., & Parveen, N. (2024). DeepSkillNER: an automatic screening and ranking of resumes using hybrid deep learning and enhanced spectral clustering approach. *Multimedia Tools and Applications*, *83*(16), 47503-47530.
48. Sathish, S., Thangavel, K., & Boopathi, S. (2010). Performance analysis of DSR, AODV, FSR and ZRP routing protocols in MANET. *MES Journal of Technology and Management*, 57-61.
49. Siva Prasad, B. V. V., Mandapati, S., Kumar Ramasamy, L., Boddu, R., Reddy, P., & Suresh Kumar, B. (2023). Ensemble-based cryptography for soldiers' health monitoring using mobile ad hoc networks. *Automatika: časopis za automatiku, mjerenje, elektroniku, računarstvo i komunikacije*, *64*(3), 658-671.
50. Elechi, P., & Onu, K. E. (2022). Unmanned Aerial Vehicle Cellular Communication Operating in Non-terrestrial Networks. In *Unmanned Aerial Vehicle Cellular Communications* (pp. 225-251). Cham: Springer International Publishing.
51. Prasad, B. V. V. S., Mandapati, S., Haritha, B., & Begum, M. J. (2020, August). Enhanced Security for the authentication of Digital Signature from the key generated by the CSTRNG method. In *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)* (pp. 1088-1093). IEEE.
52. Mukiri, R. R., Kumar, B. S., & Prasad, B. V. V. (2019, February). Effective Data Collaborative Strain Using RecTree Algorithm. In *Proceedings of International Conference on Sustainable Computing in*

Science, Technology and Management (SUSCOM), Amity University Rajasthan, Jaipur-India.
53. Balaraju, J., Raj, M. G., & Murthy, C. S. (2019). Fuzzy-FMEA risk evaluation approach for LHD machine–A case study. *Journal of Sustainable Mining*, *18*(4), 257-268.
54. Thirumoorthi, P., Deepika, S., & Yadaiah, N. (2014, March). Solar energy based dynamic sag compensator. In *2014 International Conference on Green Computing Communication and Electrical Engineering (ICGCCEE)* (pp. 1-6). IEEE.
55. Vinayasree, P., & Reddy, A. M. (2025). A Reliable and Secure Permissioned Blockchain-Assisted Data Transfer Mechanism in Healthcare-Based Cyber-Physical Systems. *Concurrency and Computation: Practice and Experience*, *37*(3), e8378.
56. Acharjee, P. B., Kumar, M., Krishna, G., Raminenei, K., Ibrahim, R. K., & Alazzam, M. B. (2023, May). Securing International Law Against Cyber Attacks through Blockchain Integration. In *2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)* (pp. 2676-2681). IEEE.
57. Ramineni, K., Reddy, L. K. K., Ramana, T. V., & Rajesh, V. (2023, July). Classification of Skin Cancer Using Integrated Methodology. In *International Conference on Data Science and Applications* (pp. 105-118). Singapore: Springer Nature Singapore.
58. LAASSIRI, J., EL HAJJI, S. A. Ï. D., BOUHDADI, M., AOUDE, M. A., JAGADISH, H. P., LOHIT, M. K., ... & KHOLLADI, M. (2010). Specifying Behavioral Concepts by engineering language of RM-ODP. *Journal of Theoretical and Applied Information Technology*, *15*(1).
59. Prasad, D. V. R., & Mohanji, Y. K. V. (2021). FACE RECOGNITION-BASED LECTURE ATTENDANCE SYSTEM: A SURVEY PAPER. *Elementary Education Online*, *20*(4), 1245-1245.
60. Dasu, V. R. P., & Gujjari, B. (2015). Technology-Enhanced Learning Through ICT Tools Using Aakash Tablet. In *Proceedings of the International Conference on Transformations in Engineering Education: ICTIEE 2014* (pp. 203-216). Springer India.
61. Reddy, A. M., Reddy, K. S., Jayaram, M., Venkata Maha Lakshmi, N., Aluvalu, R., Mahesh, T. R., ... & Stalin Alex, D. (2022). An efficient multilevel thresholding scheme for heart image segmentation using a hybrid generalized adversarial network. *Journal of Sensors*, *2022*(1), 4093658.
62. Srinivasa Reddy, K., Suneela, B., Inthiyaz, S., Hasane Ahammad, S., Kumar, G. N. S., & Mallikarjuna Reddy, A. (2019). Texture filtration module under stabilization via random forest optimization methodology. *International Journal of Advanced Trends in Computer Science and Engineering*, *8*(3), 458-469.
63. Ramakrishna, C., Kumar, G. K., Reddy, A. M., & Ravi, P. (2018). A Survey on various IoT Attacks and its Countermeasures. *International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)*, *5*(4), 143-150.
64. Sirisha, G., & Reddy, A. M. (2018, September). Smart healthcare analysis and therapy for voice disorder using cloud and edge computing. In *2018 4th international conference on applied and theoretical computing and communication technology (iCATccT)* (pp. 103-106). IEEE.
65. Reddy, A. M., Yarlagadda, S., & Akkinen, H. (2021). An extensive analytical approach on human resources using random forest algorithm. *arXiv preprint arXiv:2105.07855*.
66. Kumar, G. N., Bhavanam, S. N., & Midasala, V. (2014). Image Hiding in a Video-based on DWT & LSB Algorithm. In *ICPVS Conference*.
67. Naveen Kumar, G. S., & Reddy, V. S. K. (2022). High performance algorithm for content-based video retrieval using multiple features. In *Intelligent Systems and Sustainable Computing: Proceedings of ICISSC 2021* (pp. 637-646). Singapore: Springer Nature Singapore.
68. Reddy, P. S., Kumar, G. N., Ritish, B., SaiSwetha, C., & Abhilash, K. B. (2013). Intelligent parking space detection system based on image segmentation. *Int J Sci Res Dev*, *1*(6), 1310-1312.
69. Naveen Kumar, G. S., Reddy, V. S. K., & Kumar, S. S. (2018). High-performance video retrieval based on spatio-temporal features. *Microelectronics, Electromagnetics and Telecommunications*, 433-441.
70. Kumar, G. N., & Reddy, M. A. BWT & LSB algorithm based hiding an image into a video. *IJESAT*, 170-174.
71. Lopez, S., Sarada, V., Praveen, R. V. S., Pandey, A., Khuntia, M., & Haralayya, D. B. (2024). Artificial intelligence challenges and role for sustainable education in india: Problems and prospects. *Sandeep Lopez, Vani Sarada, RVS Praveen, Anita Pandey, Monalisa Khuntia, Bhadrappa Haralayya (2024) Artificial Intelligence Challenges and Role for Sustainable Education in India: Problems and Prospects. Library Progress International*, *44*(3), 18261-18271.
72. Yamuna, V., Praveen, R. V. S., Sathya, R., Dhivva, M., Lidiya, R., & Sowmiya, P. (2024, October). Integrating AI for Improved Brain Tumor Detection and Classification. In *2024 4th International Conference on Sustainable Expert Systems (ICSES)* (pp. 1603-1609). IEEE.
73. Kumar, N., Kurkute, S. L., Kalpana, V., Karuppannan, A., Praveen, R. V. S., & Mishra, S. (2024,

August). Modelling and Evaluation of Li-ion Battery Performance Based on the Electric Vehicle Tiled Tests using Kalman Filter-GBDT Approach. In *2024 International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS)* (pp. 1-6). IEEE.

74. Sharma, S., Vij, S., Praveen, R. V. S., Srinivasan, S., Yadav, D. K., & VS, R. K. (2024, October). Stress Prediction in Higher Education Students Using Psychometric Assessments and AOA-CNN-XGBoost Models. In *2024 4th International Conference on Sustainable Expert Systems (ICSES)* (pp. 1631-1636). IEEE.
75. Anuprathibha, T., Praveen, R. V. S., Sukumar, P., Suganthi, G., & Ravichandran, T. (2024, October). Enhancing Fake Review Detection: A Hierarchical Graph Attention Network Approach Using Text and Ratings. In *2024 Global Conference on Communications and Information Technologies (GCCIT)* (pp. 1-5). IEEE.
76. Shinkar, A. R., Joshi, D., Praveen, R. V. S., Rajesh, Y., & Singh, D. (2024, December). Intelligent solar energy harvesting and management in IoT nodes using deep self-organizing maps. In *2024 International Conference on Emerging Research in Computational Science (ICERCS)* (pp. 1-6). IEEE.
77. Praveen, R. V. S., Hemavathi, U., Sathya, R., Siddiq, A. A., Sanjay, M. G., & Gowdish, S. (2024, October). AI Powered Plant Identification and Plant Disease Classification System. In *2024 4th International Conference on Sustainable Expert Systems (ICSES)* (pp. 1610-1616). IEEE.
78. Dhivya, R., Sagili, S. R., Praveen, R. V. S., VamsiLala, P. N. V., Sangeetha, A., & Suchithra, B. (2024, December). Predictive Modelling of Osteoporosis using Machine Learning Algorithms. In *2024 4th International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS)* (pp. 997-1002). IEEE.
79. Kemmannu, P. K., Praveen, R. V. S., Saravanan, B., Amshavalli, M., & Banupriya, V. (2024, December). Enhancing Sustainable Agriculture Through Smart Architecture: An Adaptive Neuro-Fuzzy Inference System with XGBoost Model. In *2024 International Conference on Sustainable Communication Networks and Application (ICSCNA)* (pp. 724-730). IEEE.
80. Praveen, R. V. S. (2024). *Data Engineering for Modern Applications*. Addition Publishing House.