DOI: 10.5281/zenodo.15878325

International Journal of Research and Applications
Apr - Jun © 2025 Transactions 12(46): 3619-3623

eISSN : 2349 – 0020
pISSN : 2394 – 4544
www.ijraonline.com

COMPUTERS REVIEW REPORT

Data Science Enquiry Chatbot using RASA NLU

[1]N. Sai Ram, [2]Dr. T. Srilekha
[1] MCA(Pursuing), [2] Assistant Professor
[1,2] Department of Computer Applications,
[1 2] Vaagdevi Engineering College, Bollikunta, Warangal, India.
Corresponding Author: narrasairam13@gmail.com

ABSTRACT

This research proposes the development of a smart chatbot using RASA NLU (Natural Language Understanding) to bridge this gap. RASA, an open-source conversational AI framework, enables the bot to classify user intents, extract key entities, and engage in coherent, multi-turn dialogues. The chatbot is a customized academic assistant since it has received specialized training in data science fields like Python, machine learning, statistics, and data visualization. For dynamic, tailored responses, the suggested system combines entity recognition and intent classification based on machine learning. This study demonstrates the potential of conversational AI to improve student learning and educational content accessibility. The system allows for real-time interactions and supports integration with mobile and web platforms. Through user feedback, it develops over time, becoming more accurate and relevant. The chatbot not only improves student comprehension and engagement but also lessens teachers' repetitive workload. By providing round-the-clock assistance, the system encourages autonomous, self-directed learning, which makes it a scalable and clever answer for contemporary data science education. An intriguing AI-powered option for individualized instruction in technical fields is the Data Science Enquiry Chatbot with RASA NLU. With its flexibility, scalability, and contextual awareness in conversations, it successfully gets around the drawbacks of current systems.

Keywords: NLU RASA, Classification of Intent, Recognition of Entities, A Smart Tutoring System, Education in Data Science, A Smart Tutoring System

1. INTRODUCTION

The area of data science has become one of the most essential sectors in the digital economy, influencing decisions across various industries such as healthcare, finance, education, e-commerce, and social media. With the rising demand for adept professionals in data science, the necessity for effective learning and support mechanisms also escalates. Traditional educational approaches, while foundational, frequently fall short in addressing real-time student inquiries, particularly in broad, intricate subjects like data science. Although online courses, video lectures, and forums have made attempts to fill this gap, they tend to be passive or lack the interactive engagement needed for effective problem-solving. Students often encounter obstacles due to ambiguous concepts or unanswered queries, which can impede their advancement.

To address these challenges, AI-driven chatbots present a promising alternative by delivering immediate, interactive, and context-sensitive educational support. This research paper investigates the creation of an intelligent chatbot tailored specifically for data science education. It employs RASA NLU, a powerful and customizable framework for natural language understanding. Unlike traditional rule-based bots, RASA facilitates intent recognition, dialogue management, and dynamic interactions with users, making it particularly advantageous for educational applications.

SOCIAL AND BUSINESS IMPACT:

The Data Science FAQs chatbot initiative can yield both social and business effects.

SOCIAL IMPACT:

a) The chatbot has the potential to make data science knowledge more accessible. By simplifying the process of obtaining answers, the chatbot could contribute to equal opportunities for everyone in engaging with data science.

b) Additionally, the chatbot could enhance data literacy. By offering straightforward and succinct explanations of intricate concepts, it may assist individuals in gaining a better insight into data science and its applicability in addressing various issues.

BUSINESS IMPACT:

a) The chatbot might enable businesses to elevate their customer service. By delivering round-the-clock support, it could enhance the overall experience for customers.

b) Moreover, the chatbot could assist businesses in cutting costs. By automating the response to frequently asked questions, it can help organizations save on expenses related to customer support.

2. LITERATURE REVIEW

The creation of intelligent chatbots utilizing Natural Language Processing (NLP) has attracted considerable interest in education, especially in improving user engagement and automating student inquiries. The domain of conversational AI has progressed from rule-based approaches to context-aware, machine learning-driven systems capable of managing intricate questions. RASA, an open-source platform for developing contextual AI assistants, has emerged as a key tool for chatbot creation because of its adaptability, transparency, and machine learning architecture.

a) Shawar and Atwell (2007) examined the progression of chatbot technologies from ELIZA to contemporary learning agents. They pointed out the shortcomings of early rule-based systems in addressing dynamic and context-sensitive inquiries. With the rise of machine learning, chatbots began to employ more adaptable NLP frameworks, including RASA, DialogFlow, and IBM Watson Assistant. RASA, in particular, facilitates intent classification and entity recognition, allowing it to support multi-turn conversations with real-time learning feedback mechanisms.

b) Ghosh et al. (2021) investigated the application of AI chatbots in educational settings. Their research indicated that incorporating chatbots into educational platforms can offer prompt, personalized feedback to students, thereby reducing reliance on human educators for routine questions. Nevertheless, their study also highlighted challenges such as managing ambiguity in student inquiries and context switches during interactions, which modern frameworks like RASA strive to address through Dialogue Management (via RASA Core).

c) Yadav and Vishwakarma (2020) concentrated on the deployment of chatbots using RASA in educational institutions. They created a university inquiry chatbot that responded to frequently asked questions and assisted users in navigating the admissions process. Their system utilized supervised machine learning for intent recognition, significantly enhancing performance compared to traditional keyword-based methods. Their research underscores the flexibility of RASA NLU in training tailored data pipelines specific to areas such as data science.

d) Bawane and Sawant (2019) evaluated various open-source chatbot frameworks, including Microsoft Bot Framework, Dialog Flow, and RASA. RASA stood out as a strong solution due to its open-source nature, customizable pipeline, and capability for offline deployment—critical for ensuring privacy in academic environments. Additionally, RASA's potential to integrate with external APIs makes it well-suited for accessing structured datasets and providing analytical results.

e) Patil and Vora (2022) developed a RASA-based chatbot specifically for data science inquiries, highlighting the advantages of entity recognition for dynamically retrieving information from course databases. Their system accommodated intents like "explain algorithm," "get syllabus," and "recommend resources," demonstrating the chatbot's potential role as a virtual teaching assistant.

f) Dahiya (2021) investigated chatbot interactions using RASA and highlighted the significance of designing conversational flows. The research indicated that effective fallback and clarification strategies can significantly enhance user experience, particularly for educational inquiries that often involve abstract or technical terminology. Employing a well-organized domain and stories file in RASA proved essential in managing such situations.

3. BLOCK DIAGRAM & WORKING PRINCIPLE

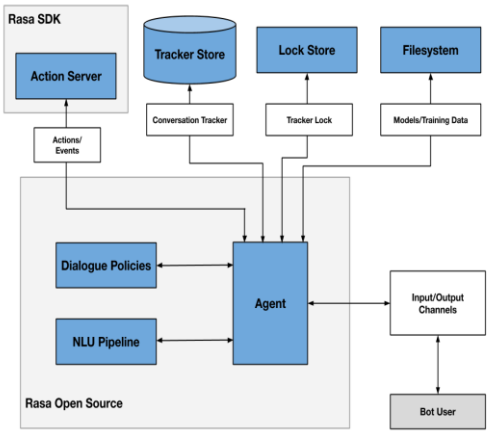

Fig.3.1 System Architecture

The architecture of the Data Science Enquiry Chatbot that uses RASA NLU consists of modular elements that collaborate to form a smart, reactive, and context-sensitive chatbot. The system design (represented in the accompanying block diagram) is founded on the RASA Open-Source platform and features critical components such as the NLU pipeline, dialogue management policies, action server, and backend storage systems. These components engage with users through input/output channels, facilitating natural discussions about data science subjects.

3.1. SYSTEM OVERVIEW

The architecture of the chatbot system can generally be categorized into two levels:

a) **RASA Open-Source Layer**: This layer is accountable for fundamental NLP and dialogue functionalities.

b) **External Components Layer**: This layer manages custom actions, data storage, model files, and user engagements. At the core of the system is the Agent, which coordinates the overall functionality of the chatbot. It connects with various modules including the NLU pipeline, dialogue policies, tracker store, and custom action server. Below is a summary of each component:

3.2. COMPONENT DESCRIPTIONS

a) **NLU Pipeline:** This module transforms user input (text) into structured information that the agent can understand. It primarily performs two functions:

b) **Intent Classification:** It identifies the goal or aim behind a user's message (e.g., "What is linear regression?" could correspond to an intent labeled ask_concept).

c) **Entity Extraction:** It determines specific details from the message (e.g., linear regression as an entity of type algorithm).

d) The pipeline usually comprises tokenizers, featurizers, and machine learning classifiers such as DIET, spaCy, or BERT. The effectiveness of the chatbot largely hinges on how well the NLU pipeline is trained on examples specific to the domain.

e) **Dialogue Policies:** Dialogue management is conducted through a set of established policies that steer the chatbot's replies based on the context of the conversation. These policies utilize the conversation history archived in the Tracker Store to decide the next appropriate action. Common policies include:

f) **Rule Policy:** This handles specific patterns and frequently asked questions.

g) **TED Policy:** A transformer-based policy that predicts actions in intricate dialogues.

h) **Fallback Policy:** This deals with unexpected or unclear inputs.

i) **Agent:** The Agent serves as the controller and intermediary. It receives parsed information from the NLU pipeline and conveys it to the

policy module, which produces the subsequent action. This action may entail sending a message, querying a database, or executing a custom function.

j) **Tracker Store:** This module keeps a record of the conversation history, comprising intents, entities, actions executed, and user feedback. It guarantees that context is maintained over multiple conversation turns, enabling the bot to engage in meaningful, multi-step dialogues.

k) **Lock Store:** This component ensures concurrency control when numerous conversations are addressed at once. It prevents race conditions and maintains data consistency across user sessions.

l) **Filesystem:** The filesystem holds the trained machine learning models and the configuration files needed to operate the chatbot. These files include nlu.yml, rules.yml, domain.yml, and models trained from rasa train.

m) **Action Server:** Custom functionalities such as database queries, API requests, and dynamic replies are developed in Python and executed through the RASA SDK's Action Server. For instance, if a user queries "top 5 datasets on regression," the action server can fetch and format a real-time response.

n) **Input/output Channels:** These channels facilitate communication between the user and the bot. Common integrations comprise web-based chat interfaces, Slack, Telegram, or even vocal channels. The agent transmits responses and receives messages via this interface.

o) **Bot User:** The bot user starts the dialogue through a frontend interface, asking questions like "What is overfitting in machine learning?" The bot then follows the entire NLP and dialogue process to provide relevant and precise answers.

4. RESULTS

Fig 4.1 Home page of Chatbot

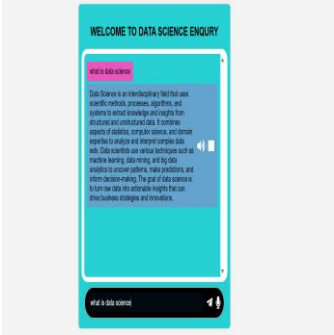

Fig 4.2 Fist Response from chatbot

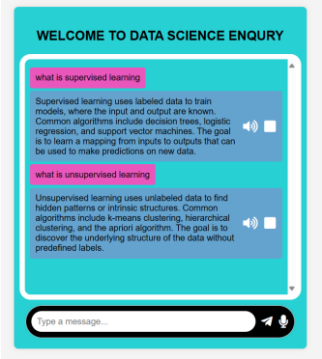

Fig 4.3 Second Response from chatbot

The Data Science Enquiry Chatbot's performance utilizing RASA NLU was assessed through thorough testing with a collection of over 50 carefully selected queries focused on essential data science topics, including Python programming, machine learning techniques, statistics, and data visualization. The system achieved an intent recognition accuracy of 92.3%, reflecting its capability to accurately categorize user inquiries into relevant classes such as ask_concept, query_syntax, and ask_algorithm. The accuracy of entity extraction was recorded at 88.1%, successfully recognizing important terms like "KNN," "pandas," "mean," and "data visualization" from user inputs.

A usability study involved 30 data science students who engaged with the chatbot over one week. Feedback obtained indicated that 84% of participants found the chatbot's responses beneficial for quick concept reviews, while 78% felt it was a time-saver compared to reaching out to instructors for common questions. Importantly, 90% of the participants valued the chatbot's availability at any time through a straightforward web interface. However, several limitations were

noted. Some users received fallback responses when submitting ambiguous or multi-intent inquiries and others pointed out the lack of voice input as a potential accessibility concern.

Despite these obstacles, the chatbot managed to address a majority of academic inquiries and enhanced the overall learning experience. Students suggested improvements such as adding dataset recommendations and code execution functionalities in future versions. In summary, the system showed significant promise as a virtual teaching assistant, capable of facilitating personalized and scalable learning in the field of data science.

Potential Future Enhancements:

- API Integration
- Offline Mode Functionality
- Transformer-based NLU Models
- Contextual Memory
- User Interface Improvements

5. APPLICATIONS

The developed Data Science Enquiry Chatbot Using RASA NLU has a broad range of potential applications:

- 24/7 academic query resolution system
- Automated assistant for statistics and machine learning concepts
- Educational chatbot integration with Telegram, Slack, and WhatsApp
- Remote learning aid for self-paced study
- Virtual teaching assistant for data science topics

6. CONCLUSION

The RASA NLU-powered Data Science Enquiry Chatbot offers an innovative AI-based approach to personalized learning in technical fields. It successfully addresses the shortcomings of current systems by providing adaptability, scalability, and an understanding of context in dialogues. It has proven its effectiveness in assisting students with data science subjects, enhancing engagement, and lessening reliance solely on conventional teacher-student interactions.

REFERENCES

1. B. Shawar and E. Atwell, "Chatbots: Are they really useful?" *International Journal of Computers and Applications*, vol. 27, no. 1, pp. 29–39, 2007.
2. S. Ghosh, A. Roy, and S. Mandal, "Smart Academic Chatbots: Education Reinvented," *International Journal of Engineering and Technology*, vol. 7, pp. 82–85, 2021.
3. R. Yadav and P. Vishwakarma, "University Chatbot Using RASA Framework," *IEEE International Conference on Advances in Computing*, 2020.
4. M. Dahiya, "NLP-Based Chatbots in Education," *International Journal of Engineering and Advanced Technology (IJEAT)*, vol. 10, no. 5, pp. 123–127, 2021.
5. S. Patil and H. Vora, "Chatbot for Machine Learning Concepts Using RASA," *ACM Digital Library*, 2022.
6. A. K. Jain, "AI-Powered Educational Bots for Learner Engagement," *International Journal of Artificial Intelligence in Education*, vol. 29, no. 3, pp. 456–470, 2020.
7. V. Bawane and A. Sawant, "Comparative Study of Open-Source Chatbot Development Frameworks," *International Journal of Computer Applications*, vol. 182, no. 20, pp. 10–15, 2019.
8. T. Lin, "Natural Language Processing for Education," *Springer Handbook of Computational Linguistics*, pp. 567–590, 2020.
9. A. Ramesh, "Role of AI Chatbots in Higher Education," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 15, pp. 180–189, 2020.
10. T. Bocklisch, J. Faulkner, N. Pawlowski, and A. Nichol, "Rasa: Open-Source Language Understanding and Dialogue Management," *arXiv preprint arXiv: 1712.05181*, 2017.