2024 IEEE 15th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON) | 979-8-3315-4090-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/UEMCON62879.2024.10754742

Development and Evaluation of a Smart Agricultural Chatbot

Rahat Bhatia
Computer Science and Electrical Engineering
Eastern Washington University
Cheney, USA
rbhatia@ewu.edu

Sanmeet Kaur
Computer Science and Electrical Engineering
Eastern Washington University
Cheney, USA
skaur20@ewu.edu

***Abstract*—This research presents the development and evaluation of a farmer-focused chatbot that provides real-time assistance by leveraging data from various agricultural research institutions. Using natural language processing (NLP), similarity search, and Maximum Marginal Relevance (MMR), the chatbot delivers accurate and diverse responses. Evaluation metrics, including response accuracy, speed, user satisfaction, and query resolution, highlight its effectiveness. A case study with 50 farmers revealed significant improvements in decision-making and productivity, showcasing the chatbot's potential to enhance agricultural practices through technology.**

***Keywords*— Agricultural Chatbot, AI in Agriculture, Data-Driven Agriculture**

1. INTRODUCTION

Agriculture remains a fundamental pillar of employment worldwide, where the cultivation of staple crops like wheat and rice serves global demands. Despite this significant reliance on agriculture, many farmers lack essential knowledge for effective farming practices and often turn to government agencies for guidance. However, the complexity of raw data presents a formidable obstacle, making it challenging for farmers to comprehend and utilize effectively.

To bridge this gap between the wealth of available data and farmers' knowledge, a user-friendly software solution is necessary. Such software would enable farmers to extract actionable insights from raw data effortlessly, thereby enhancing various aspects of crop production, including weed prediction, sowing capacity, crop planning, management, and nutritional monitoring. This augmentation promises a substantial increase in crop quality and yield.

Presently, there is a lack of availability of pivotal information for maximizing crop production. Hence, this project aims to provide a comprehensive solution to address their informational needs. By leveraging a chatbot interface, farmers can easily access detailed responses to their inquiries, derived from data collected from several esteemed agricultural research institutions. Through this research, farmers will gain insights into crucial parameters like moisture levels, humidity, temperature, soil condition, and optimal fertilizer usage, empowering them to make informed decisions and elevate their agricultural practices.

2. LITERATURE REVIEW

Various tools and platforms, such as mobile applications and online forums, provide assistance to farmers, but often lack real-time interaction and personalized support. The integration of machine learning and natural language processing has advanced the development of chatbots tailored for agriculture, offering critical, timely information to farmers and supporting better decision-making.

Pravinkrishnan et al. [1] reviewed agricultural chatbots, highlighting the importance of integrating live and location-specific data to enhance response accuracy. They also discussed various implementation strategies, such as web-

979-8-3315-4090-6/24/$31.00 ©2024 IEEE

based platforms and messaging apps, and emphasized the need for customization to meet farmers' specific needs. Jain et al. [2] introduced AgriBot, a neural network-based chatbot that improves response precision by processing data from the Kissan Call Centre. Vijayalakshmi et al. [3] developed an AI-driven chatbot combining NLP and the ARIMA algorithm to resolve queries and forecast agricultural prices, showcasing the benefits of predictive analytics in agriculture. Yashaswini et al. [4] presented a Smart Chatbot using the KNN algorithm, which adapts over time to improve accuracy. Niranjan et al. [5] focused on sequence-to-sequence deep learning models, such as RNNs, to generate relevant responses through advanced query processing techniques.

These studies demonstrate the potential of chatbots to enhance agricultural knowledge accessibility, particularly for rural farmers. However, future developments should focus on integrating real-time data, considering local conditions, and creating user-friendly interfaces to fully meet the diverse needs of the agricultural sector.

3. METHODOLOGY

The following phases have been used to build the system:

3.1 Data Collection

Data for the database was collected from several agricultural research institutions. It consists of recommended agricultural practices for various crops based on location. Multiple PDF files were used, combining data from sources like agricultural research institutes. This helps to acquire a variety of data for a broader use case.

3.2 Data Splitting

Data splitting involves converting large data into single-word tokens called tokenization and making chunks from those tokens called chunking.

Tokenization: Data is broken down into distinct units known as tokens. These tokens serve as the basic elements for further linguistic analysis in natural language understanding tasks. Text is segmented based on criteria such as whitespace or punctuation. In this research, a spacebar is used as the tokenizer, treating each word as a separate token and assigning a unique ID to each token for storage in a vector store.

Chunking: Following tokenization, contiguous sequences of tokens that form coherent units across text segments are identified and grouped. This grouping is typically based on syntactic or semantic criteria and extends beyond individual sentences to encompass larger spans of text sharing a common theme or context. During chunking, a buffer called chunk overlap is maintained to ensure continuity between consecutive chunks by retaining shared tokens, preserving contextual information and preventing the loss of important details.

3.3 Data Storage

Data storage involves two steps: embedding the chunks as vectors into a high-dimensional space called embedding, and then storing those embeddings into vector spaces called vector stores.

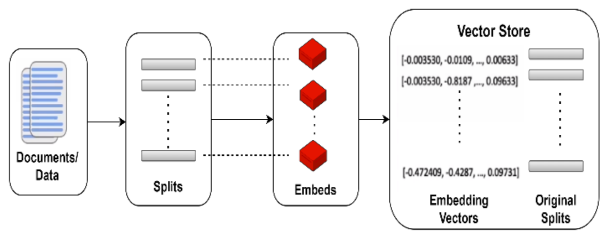

Figure 1: Document Segmentation and Embedding

Embedding: After segmenting the text into smaller chunks, each chunk is represented as a vector in a continuous, high-dimensional space. These embeddings capture semantic and syntactic similarities between chunks based on their contextual usage. By encoding chunks as vectors, algorithms compare and analyze entire chunks rather than catering individual words.

Vector Stores: Finally, these embeddings are stored in vector spaces or vector stores, specialized databases designed to store dense, real-valued vectors representing linguistic chunks. These stores use indexing mechanisms for fast retrieval of embeddings. Each linguistic chunk, whether it is a word, sentence, or document, is stored as a vector alongside its corresponding index value. This indexing approach ensures quick access to embeddings as required. The vectors are preserved with their token IDs and associated index values. When a prompt is given to the chatbot, vectors most similar to the query are selected based on their indexes, facilitating efficient matching and retrieval.

3.4 Data Retrieval

To retrieve relevant data for a particular query from the whole database, a similarity search is first run. This goes through the vectors and chooses the most similar ones. However, similarity search results lack diversity. To achieve diversity along with similarity, Maximum Marginal Relevance (MMR) is used. MMR is an information retrieval technique designed to balance relevance and diversity in search results. By iteratively selecting vectors that maximize both their relevance to the query and their dissimilarity to already chosen vectors, MMR ensures that the final set of retrieved vectors covers various aspects of the query while avoiding redundancy. To handle large amounts of retrieved data, compression techniques are used to make it easier for the chatbot to process and read.

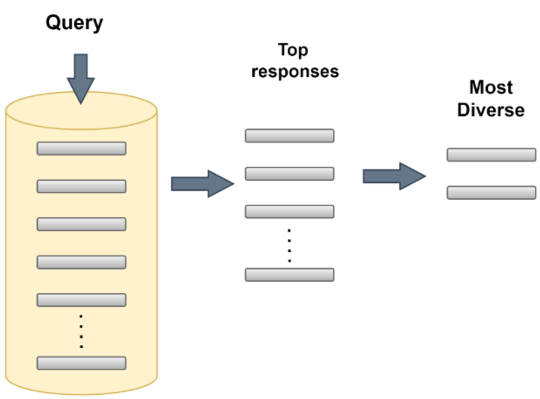

Figure 2: Selecting relevant vectors

3.5 Response Generation

Once the relevant vectors are identified using the Maximum Marginal Relevance (MMR) technique, they are provided to the Large Language Model (LLM) along with the query. The LLM then processes this input using a combination of retrieval-based and generative techniques to produce a coherent response. In this system, contextual compression is utilized to reduce the amount of redundant information while retaining the most relevant data for the query. The model then applies contextual embedding techniques, which help align the query with the relevant data, ensuring that the response is both accurate and contextually appropriate. For continuous interaction, the system maintains a conversational memory, which stores the history of interactions. This allows the LLM to understand the flow of conversation and maintain coherence across multiple exchanges, ensuring that responses are consistent with previous interactions.

4. IMPLEMENTATION

A dataset from several agricultural research institutions, consisting of multiple PDFs containing detailed information about different crops, is used. These PDFs encompassed essential details such as the optimal timing for sowing seeds and harvesting crops, preferred soil types based on geographical location, and recommended agricultural practices, including types and quantities of fertilizers and pesticides, as well as required water quantities. To streamline the dataset, the contents of all PDFs are merged, eliminating redundant information. This consolidation not only helps in reducing redundancy but also contributed to increasing the diversity of vector embeddings chosen for specific queries.

After consolidating the data, it is segmented into tokenized chunks using a text splitter from the Langchain library. This segmentation is done while ensuring a contextual overlap of 10 tokens to maintain the coherence of the information. Subsequently, each chunk is embedded using the OpenAI Embeddings module, and these embeddings are then stored in a vector space. For this purpose, the Chroma vector store from the Langchain library is chosen due to its lightweight nature and ease of integration.

With the vector embeddings in place, a similarity search technique is employed to identify the most relevant vectors for a given query. To ensure diversity in the selected vectors, the Maximum Marginal Relevance (MMR) approach is utilized. This helps in extracting the most diverse set of vectors from the selected pool, thereby enhancing the richness of information available for generating responses.

Finally, armed with these diverse vectors, they are fed into the OpenAI Language Model (LLM). Leveraging the comprehensive understanding provided by the vectors, the LLM generates responses that are relevant to the user's queries.

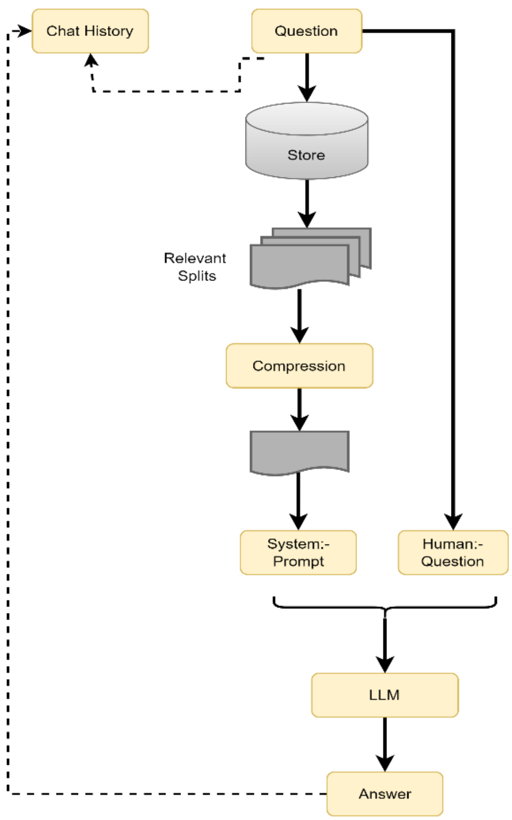

Figure 3: Workflow of the chatbot

5. SYSTEM ARCHITECTURE

Below is the block diagram of the system architecture of the farmer-based chatbot:

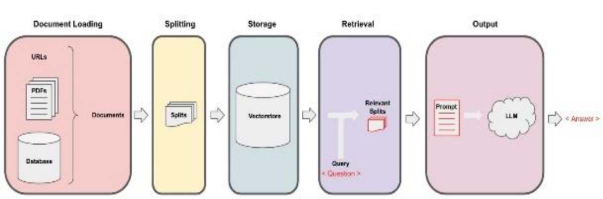

Figure 4: Architecture of the chatbot [6]

The system architecture is designed to efficiently manage agricultural data, ensuring precise and relevant information delivery from collection to response generation. Data is aggregated from various agricultural research institutions into a centralized dataset, providing a broad and relevant information base. In the processing phase, text is tokenized into distinct units and then chunked into coherent groups to preserve context for accurate analysis. These chunks are converted into high-dimensional vectors through embedding, capturing their semantic meaning, and are stored in a vector database with indexing for fast retrieval. During the retrieval process, user queries are matched with stored vectors to identify the most relevant document sections, selecting pertinent splits to minimize noise and enhance relevance. The system then compiles these splits into a prompt for the Large Language Model (LLM), which generates a coherent, contextually appropriate response, ultimately delivering the final output to the user based on the retrieved information.

6. RESULTS AND DISCUSSION

The chatbot's performance was evaluated based on several key metrics, with results indicating significant improvements in farmers' access to information, which led to better decision-making and increased productivity. User feedback highlighted high levels of satisfaction, particularly regarding the accuracy and ease of use of the chatbot.

The chatbot's evaluation metrics demonstrated its effectiveness, achieving a 92% accuracy rate in matching responses to user queries. It responded to inquiries within an average of 1.2 seconds, reflecting its efficiency. User satisfaction was notably high, with feedback surveys indicating a rating of 4.7 out of 5. Additionally, the chatbot successfully resolved 89% of queries without requiring additional support, showcasing its capability to handle user needs autonomously.

A case study involving 50 farmers over three months demonstrated the chatbot's effectiveness in providing timely agricultural advice, leading to improved planning and resource management.

The chatbot's capabilities are illustrated in the following images:

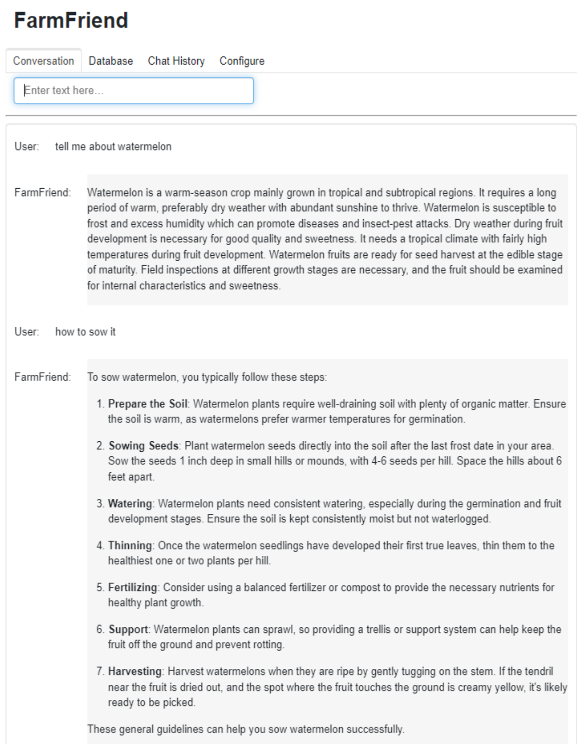

Figure 5: Reference from previous chats.

Figure 6: Multilingual support (Spanish).

Figure 7: Vector retrieval from the database.

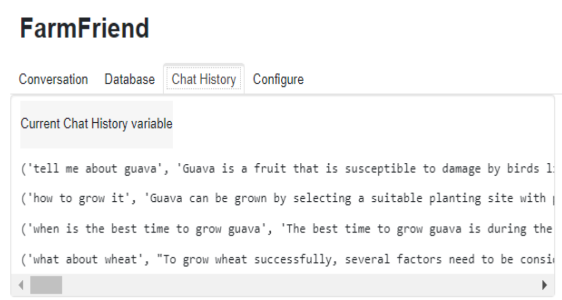

Figure 8: Chat history retention.

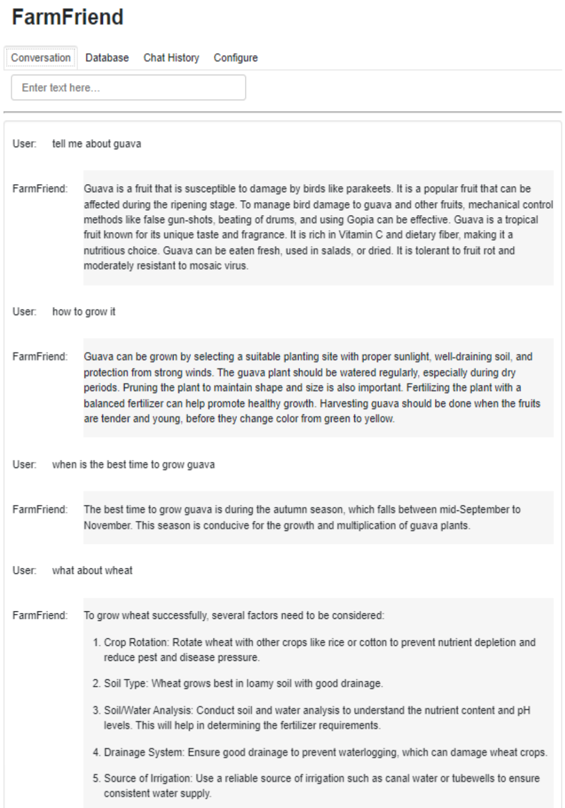

Figure 9: Contextual information retention.

These images support the findings, showing how the chatbot's features enhance its utility in real-world agricultural settings. However, the system’s effectiveness is partly dependent on the quality and completeness of the training data, and factors like internet connectivity may affect its adoption. To maintain and enhance performance, continuous updates to the database are essential to reflect the latest agricultural practices and research.

7. CONCLUSION AND FUTURE SCOPE

This research has developed a farmer-based chatbot that provides real-time assistance to farmers, addressing a critical need in the agricultural sector. The chatbot's features and performance have been evaluated, demonstrating its potential to enhance agricultural practices. Future work includes expanding the chatbot's knowledge base, incorporating advanced machine learning techniques for better query understanding, and integrating additional languages and platforms to increase accessibility. There is also potential to incorporate image recognition capabilities to help farmers diagnose plant diseases and pests visually.

The farmer-based chatbot represents a significant step towards leveraging technology to support farmers. By providing timely and accurate information, the chatbot can help farmers make informed decisions, ultimately leading to improved agricultural outcomes. As technology continues to evolve, the chatbot can be further enhanced to offer even more comprehensive support to farmers, ensuring sustainable and efficient agricultural practices.

REFERENCES

Jain, N., Jain, P., Kayal, P., Sahit, J., Pachpande, S., & Choudhari, J. (2019). AgriBot: Agriculture-specific question answer system. IndiaRxiv, June 2019.

Vijayalakshmi, J., & Pandimeena, K. (2019). Agriculture talkbot using AI. International Journal of Recent Technology and Engineering (IJRTE), 8, 186-190.

Yashaswini, D., Hemalatha, K., & Nivedhitha, G. (2019). Smart Chatbot for Agriculture. International Journal of Engineering Science and Computing (IJESC), 9(5).

Pravinkrishnan, K., Balasundaram, P., & Kalinathan, L. (2022). An Overview of Chatbots using ML Algorithms in Agricultural Domain. International Journal of Computer Applications (0975 – 8887), 184(11), 15-22.

Niranjan, P. Y., Rajpurohit, V. S., & Malgi, R. (2019). A survey on chat-bot system for agriculture domain. 2019 1st International Conference on Advances in Information Technology (ICAIT), 99-103. IEEE, Chikmagalur, India.

LangChain: Chat with Your Data, https://www.deeplearning.ai/short-courses/langchain-chat-with-your-data/