International Journal of Computer Applications (0975 – 8887)
Volume 184 – No.11, May 2022

An Overview of Chatbots using ML Algorithms in Agricultural Domain

Pravinkrishnan K.
Sri Sivasubramaniya Nadar College of Engineering, Chennai

Prabavathy Balasundaram
Sri Sivasubramaniya Nadar College of Engineering, Chennai

Lekshmi Kalinathan
Sri Sivasubramaniya Nadar College of Engineering, Chennai

ABSTRACT
The agricultural sector plays a vital part in a country's economic growth. It has already made a major contribution to advanced countries' economic growth. The impact it hason less-developed countries' economic development is vitally important. The farmers involved in agricultural activities lack the resources to stay updated with the information related to the latest advancements in technologies and farming practices. Existing human-involved operations such as Kissan Call Center (KCC), even though capable of delivering expected results, has its own drawbacks. Hence there is a need for an automated chatbot system that can function as a substitute to KCC. A chatbot system is a system that delivers domain-specific knowledge to its users. Such a system in the field of agriculture is very helpful in keeping the farmers updated. In this paper, existing works on such question-answer systems focusing entirely on works involving machine learning techniqueshave been reviewed. Suggestionsto improve the overall usability of the existing systemshave also been made.

Keywords
Chatbot, Query processing, Intent identification, Similarity function, Answer extraction.

1. INTRODUCTION
Agriculture is the basis for the sustainability of human life and it plays a major part in the development of a country. The developments in bio-technology and chemical engineering has had a huge impact on making agriculture more profitable and easier. But the shortage of awareness and knowledge regarding these new developments and technologies is still persistent among conventional rural farmers. On the other end of the scale, urban farmers who take up farming lack even the basic ideas behind agriculture. On either side, the problem arises due to the shortage of latest and updated knowledge related to specific problems that they face. A chatbot is a simple system that generates answers to human queries in a particular domain, based on existing knowledge.

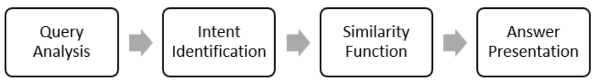

Fig 1: Workflow of a chatbot.

A significant amount of research has been done in the betterment of the ability of the chat-bot to generate accurate answers to the given query and also present it in a human-like way. Like in a factory as shown in Fig. 1, a chat-bot has an assembly line of processes that it needs to do to generate an answer for the query presented to it. To start with, the query received is analyzed and processed to convert it to a machine understandable way. Secondly, the processed query is used to identify the intent with which the query was asked. After which, a similarity function uses the intent and the processed query to try and get an answer from the existing knowledge base, that best fits the query. Finally, the extracted answer is presented back to the user as human-like response.

2. LITERATURE SURVEY
Jain et al., [1] have used a simple Neural network-based model to create an agribot, that can be accessed through any electronic device, to answer queries from farmers. Their model is based on the dataset from Kissan Call Centre which consists of previous queries and their answers recorded state wise. The input queries are processed in the neural network which extracts an entity from it and groups it with similar sentences using the Sen2Vec Model, after which the most relevant answer to the query is outputted. Their implementation has improved the accuracy of a sentence embedding model from 56% to 86%.

du Preezet al., [2] have developed a web-based voice chatbot, implementing a black-box approach. The chatbot takes a voice input which is formatted to an XML type and encapsulated as a SOAP message pack. There on, by the use of an artificial self-learning brain, the chatbot generates appropriate responses to the user's query. Each query and responseare archived and used to improve the capability of the system. The system uses a modular design for all its components. This distributed environment reduces the overall load on one system when compared to systems that run on a single module. The usage of a self-training AI model is claimed to prevent ser-vice bottlenecks and the competing of resources by modules.

Vijayalakshmi et al., [3] have presented a chatbot based solution to interact with users and use that to process and provide solutions to their queries. They have used NLP to parse the input sentences and identify the keywords from them to match them with the existing knowledge base to produce an appropriate answer to the query. The root words in the query are filtered and converted to a bag of words and then converted to a vector form so that they can be processed. A neural network-based model is constructed to classify the pre-processed data. The neural network was then optimized using gradient descent to produce the best results on the training data set. They have also incorporated an existing prediction algorithm, ARIMA, into their system to help predict the future cost of agricultural products. This system takes input through text and voice and can also output its answer through text and voice. This paper focuses mainly on improving the relevance of the output produced by the intelligent machine to make it more personal and factual.

Gawadeet al., [4] have developed an intellectual chatbot system that can respond to queries by students on college-related matters. This system has a step-wise regiment to process and answer each query. First, the chatbot gets the query as input from the user. Then, the query is pre-processed to remove stopping words and the remaining keywords are fetched. Further, the keywords are matched with the existing knowledge base to muster up an appropriate response. Finally, once the response is returned to the bot, the chatbot packages the data into a proper response to be displayed for the user.

Yashaswiniet al., [5] developed a smart chatbot based to answer questions related to agriculture practice and technology. Their model is based on data collected from various sources like government websites and repositories. They have used the K Nearest Neighbors (KNN) algorithm as the similarity measure to return the predicted class. The whole system is implemented using a Django architecture, which consists of three modules to ensure the proper functioning of the system. The system updates its knowledge base with each query that it answers which in turn, improves the accuracy of the system as a whole. This system is claimed to provide answers with an accuracy of up to 90%.

Jain et al., [6] developed a conversational agent to answer farmer queries that have a simple UI that can be used easily. The system has been developed to be used in two ways, the Audio-only FarmChat and Audio + text FarmChat. They have developed the model using the KCC dataset and information collected from formative interviews with farmers and agri-experts. The proposed system uses cloud-based scalable services to implement the conversational agent and a Sequential neural network for training the model for the chatbot. The speech-based conversational chatbot caters even to the low-literate and digitally illiterate users. The system has been built based on inputs from potato farmers and rural India and focuses mainly on the usability and viability of such smart systems among farmers.

Ong et al., [7] have created a domain-specific chatbot system that acts as a platform where farmers can get to know each other and share their experiences and easily approach the experts in agriculture to solve their problems and seek recommendations. As an added benefit to agricultural researchers, they can work together with farmers in helping their research and explore more into their research domain. The proposed system is a web-based information sharing platform involving a database that is managed with Python using MongoDB. The chatbot is a rule-based one that uses NLP to make the user convenient when interacting with the Chatbot and looking for information.

Arora et al., [8] have developed a Telegram bot that can answer crop disease and weather-related queries. The system uses a multi-layered CNN based approach which is trained using the KCC dataset which contains logs of calls at KCC by farmers. The weather prediction feature of this system is implemented by using the OpenWeather-Map API. For the disease detection feature, the queries are processed and are con-verted to a word vector which is compared with the existing knowledge base to classify the query to a particular disease. The model is said to produce answers with an accuracy of 70%.

Niranjan et al., [9] have designed a chatbot by implementing a sequence-to-sequence deep learning RNN system to answer farmer queries related to agriculture. The system works in three phases, question analysis, document processing and answer extraction. The queries are processed in the analysis phase through POS tagging, stemming and removal of keywords. The processing step uses these keywords to fetch documents that contain keywords equivalent to those present in the query. Based on the documents retrieved by the model, an answer is extracted from the vocabulary of words. The RNN system is claimed to produce answers with high efficiency.

Sawant et al., [10] have proposed an interactive web-based system to assist farmers in agricultural activities. The system takes various parameters like rainfall, temperature, area and previous yield into account to answer the queries. The model had been trained using crop data from data.gov.in and rainfall data from maharani.gov.in both of which are datasets based on farming and rainfall in Maharashtra. They have trained and compared the testing and training accuracy of algorithms like KNN, decision tree and random forest. The random forest algorithm is observed to be producing the highest testing accuracy of 78%.

Lalwaniet al., [11] have proposed an AIML based chatbot to answer college-related queries. The proposed system can be incorporated into any Institution's website as a pop-up chat feature. The system AIML files to store question-answer pairs and uses that as the knowledge base to match the query to the patterns listed in the existing knowledge base to extract a response. The query goes through a pre-processing stage where the keywords are separated using techniques like Lemmatization and POS tagging. The keywords are then matched to the knowledge base using Path similarity and Wu-Palmer similarity tests. The system also maintains a log file where the inputs to which the system could not produce a response are stored. These log files are reviewed and the unanswered queries are answered by an admin and are used to improve the knowledge base of the system.

Feineet al., [12] have covered a variety of areas in which improvements can be made to existing technologies to create chatbot systems that generate better responses to human queries. They have studied the existing usage of adjectives, adverbs and verbs in a sentence by humans and existing chatbot dialogues from the ConvAI2 challenge. They have proposed chatbot improvement responses mainly along two areas, the introduction of a chatbot developer to review and improve the quality of the generated response, and the mechanism's restrictiveness.

Han et al., [13] have presented a keyword QA system for linked data to interpret every possible user intention and report related answers in NL form. The system is said to extract every possible triple of information from linked data that are related to input keywords and report the extracted information in NL. This system uses entity disambiguation and distributed word similarity to match keyword-related entities and properties in linked data. The query is input in keyword forms which is used to generate SPARQL queries to extract the possible triples. The triples are then used to generate reports in natural language.

Singh et al., [14] have done a detailed survey of the general architecture of a chat-bot system and have discussed the improvements made to generic, mainstream chat-bot assistants. The chatbots they have discussed are ELIZA by Joseph Weizenbaum, PARRY by Kenneth Colby, JABBERWACK by Pollo Carpenter, ALICE by Richard Wallace, SMARTERCHILD by Activebuddy Inc., WATSON by IBM, SIRI by Apple Inc., MITSUKU by Steve Worsick, CORTANA by Microsoft, ALEXA by Amazon, TAY by Microsoft Research, Google Assistant by Google and Bixby by Samsung.

Liu et al., [15] have improved how chatbots respond to user queries by attempting to make those responses more personalized and human-like. The system uses user data to gather information about their interests, characteristics, language style, etc. The system takes a post and converts it into a word vector. After which 3 vectors, post, response and user, are maintained to classify the post and fill up the response based on the user. The response with the highest confidence score for the given post and user is then displayed.

Jadhav et al., [16] have developed a crop recommender system that can also give information about weather and locations of fertilizer vendors. The model has been trained using a self-made dataset using the decision tree regression algorithm to pro-duce the answers to user queries. They have used Socket programming to implement the system as a chatbot application and have used APIs to answer weather and vendor location queries. The system is claimed to have produced 90% testing accuracy.

Wang et al., [17] have proposed a method to generate answers according to the in-put question and its similar questions. The proposed method mines the mapping relationship of questions-answers, questions-questions through a constructed matrix. The answer is generated by making use of the relationship and extracted focus information. They have also shown experimental results to prove that the proposed method is promising.

Wang et al., [18] have proposed a system to substitute tutors in answering student queries in online learning platforms. The responses are also analyzed based on three metrics, correctness, professionalism and timeliness. The system takes in student queries, separates keywords and uses pattern matching techniques to find out the best possible response from the knowledge base. The response is then converted to sentence form using NLP to provide a human-like response. The intelligent teaching assistant system adopts MVC's framework mode, namely model M, view V and con-troller C, and maintains the intelligent teaching assistant system through Django.

Vijayabaskaret al., [19] have aimed to create a device implementing IoT to predict the crop which will yield the maximum profit for that particular soil in that particular harvest. They have analyzed 5 different implementations of the predictive algorithm which have considered various soil properties to predict the crop. They have observed that various soil samples taken from different places can be tested using the NPK sensors which are portable and have low time consumption and prediction based on the atmosphere which would not be accurate, as the climatic conditions may differ.

Yamada et al., [20] have proposed a system to process queries through statistical machine translation instead of entity extraction to improve the exactness of the query that is input. The sentence which has the maximum translational probability of all the possible translations of the input query is computed using the Bayes rule. Using the extracted question sentence and the original question sentence, the ideal answer sentence is generated based on the existing learning data. The system mainly concentrates on Who~?, When~? And Where~? questions and has produced 74%,67% and 80% accuracies respectively and overall accuracy of 74%.

Alves et al., [21] have done work to present the development of a chatbot capable of assisting different RPG roles in decision-making. Game agents consult trends using statistics and make predictions about pollution levels based on a model constructed with ML, applied to the data collected in the RPG pilot sessions. For the pollution predictor, they have implemented 4 algorithms, Linear regression, SVR, regression tree, and random forest regression, to compare and contrast the accuracies produced by each of them. Based on their conclusion, the Random Forest Regression algorithm produced the least MAE and MSE scores of 0.3948 and 6.5115 respectively and the highest R2 Score of 0.9907.

Patel et al., [22] have discussed the various ways in agriculture in which Machine Learning could be used to help farmers. They have also given a general insight into the various stages in Machine Learning. The applications discussed here are the usage of crop selection and crop yield prediction using ML techniques like KNN, ANN, Decision Trees, etc., water management in agriculture using regression, and chatbots that answer queries related to farming. This paper has described various machine learning models and the scope of machine learning in various applications of the agriculture field.

Mostacoet al., [23] have developed AgronomoBot which was developed focusing on the search and display of data acquired from a Wireless Sensor Network deployed on a vineyard. It has been developed as a Telegram Bot API and can access information collected by eKo field sensors, bringing it back to a user through interaction over the Telegram application. The IBM Watson cognition services platform was used for improving the user experience by enabling the use of natural language during the conversation experience, providing intention detection. AgronomoBot is a chatbot that uses NLP and AI to interact with the user and search for the desired information in a WSN, adapting to different forms or languages of dialogue to achieve the same intention.

Fernandes et al., [24] have discussed the implementation and efficiency of various mainstream and domain-based chatbots. The domains in which AI and ML-based chatbots were implemented that are discussed here are medical, psychiatric counsel-ling, weight control, agriculture and E-commerce. They have given a generic template of the factors to address and the process to follow while building such systems. Their template suggests places to collect datasets to train the ML Model, ML/DM algorithms that can be used to develop a model, usage of Dialogflow which gives users ways to implement interactive conversational systems using NLP, usage of MATLAB IDE for numerical computations like matrix manipulations, data plotting, etc., and using AIML for creating conversational agents as it is easy to understand and highly maintainable.

Chen et al., [25] have done research aimed at building an intelligence commerce platform system and apply it to the innovative management of agricultural firms and farms. This study refers to the interfirm synergy, including the highly uncertain phenomenon of technology and market, and the formation of clustering synergy. This study also uses the systematic development research methodology as the development step of this research system. The system uses sequence-to-sequence (seq2seq) architecture to implement Machine Translation, Text Summarization and Conversational Modelling. The intelligence customer service system uses a Neural Conversational Model to generate answers to queries. The research results combine ICT and intelligence on the e-commerce platform to develop intelligence business with exemplary and standard, apply GPS positioning system to the farm and develop intelligence customer service system with deep learning calculus.

Nayaket al., [26] have developed a chatbot which is designed to act as a farming assistant that clears all the doubts of the farmers in an efficient manner. The proposed system uses Machine learning algorithms to generate query responses and the system fails to respond to the farmer's queries, these queries are forwarded to experts. The knowledge base for the chatbot is obtained from the data that is collected and grouped from the most asked question and answers about farming or agriculture from the internet. The proposed chatbot uses the Levenshtein distance formula for calculating the difference in the string and returns a value which is named as the confidence value. The system they have proposed is claimed to answer queries with an accuracy of 96%.

Kale et al., [27] have done research proposing the Neural Network model to predict crop yield and success rate of crops depending on the dataset provided by the Indian government. The dataset is huge containing data for all the regions of India which were filtered to get data for Maharashtra state. The crop yield prediction model uses the backpropagation algorithm of an Artificial neural network. A multilayer perceptron technique is used. The performance of the model is evaluated using parameters like Mean Absolute Error, Mean Squared Error and Root Mean Squared Error. The ANN with linear regression with forward and backward propagation model predicted the dependent variable with 82% accuracy and very little loss.

Kavita et al., [28] have done research estimates the crop yield for India using data from 1950 to 2018. The prediction is made for five crops which are Rice, Wheat, Jowar, Bajra, Tobacco, and Maize using parameters including the area used for the crop sowing, production, Yield, and Area under irrigation. The prediction is attained using Decision tree and Random forest. The dataset used for the experiment in this research was collected from ww.mospi.gov.in and https://data.gov.in, which is made public by government authorities. The evaluation parameters used are Mean Absolute Error and Root Mean Squared Error. The paper concludes that the Decision tree algorithm produces the best result for the given dataset with an accuracy of 98.62%, MAE of 1.45 and RSME of 2.11 compared to other algorithms like Linear Regression (89.38%), Lasso Regression (86.33%) and Ridge Regression (89.53%).

Manjula et al., [29] have built an architecture for crop yield prediction besides pro-posing a new methodology that combines the usage of vegetation indices derived from remote sensing images and other attributes. The framework is designed to be flexible and dynamic so that it can be used for crop yield prediction for different crops. Their methodology makes use of vegetations indices collected through remote sensing technology, climate-related variables, agronomic related variables, and weather disturbance information. The proposed approach uses the crop yield model and data mining approach coupled with the crop yield model for accurate prediction results.

Momayaet al., [30] have proposed a system that answers queries related to weather, plant protection, animal husbandry, market price, fertilizer uses, government schemes, soil testing, etc. The data for the database is collected from the official website of Kisan Call Centre (KCC). The system takes the input query and gives it as an input to RASA NLU which then classifies it into intents, identifies entities and finds out the matching intent with the highest confidence score. Further, it is checked if the intent corresponds to the weather query or not. If the intent is not a weather query, then its corresponding intent is found from the database. The answer to the query from Rasa core is then channeled to ngROK to Twilio and finally to WhatsApp for easy access by the user. The system is claimed to have 95.67% Intent Accuracy, 94.285% Story Accuracy, F1 score of 97.90 and overall Precision and Accuracy of 98.46% and 96.1% respectively.

Table 1. Analysis of the various chatbot implementation strategies and their limitations.

S.No	References	Methodology	Reported testing accuracy	Limitations
1	[1], [2], [4], [5], [7], [10], [25], [26]	Chat bots were developed aswebsites for the end users whoare farmers	[1]-86%, [5]- 90%, [10]-78%, [26]-96%	• Needs proper internet connectivity to be accessed • Does not cater well to digitally illiterate people
2	[6], [11], [19], [21], [31], [32]	Chat bots were developed to workin standalone machines withsimple UI	[31]-75.55%	• Huge amount of storage is required asthe whole system is stored on the user's device • Updates and bug fixes require data connectivity and high amount of download able data
3	[3], [8], [23], [30]	Chat bots were built as Telegramand Whatsapp using appropriate APIs	[8]-70%, [30]-intent-98.46% story-96.1%	• Failure in API connectivity results in failure of the whole system • Does not cater well to digitally illiterate people
4	[1], [2], [4], [9], [10], [16], [19], [22], [27], [28], [29], [31]	Answers extracted from the chatbots were keywords	[1]-86%, [9]-78%, [16]-90%, [28]-98.62%, [31]-75.55%	• Keywords in their raw form are difficult to interpret • Misinterpreted keywords might misguide the users

5	[3], [5], [8], [11], [13], [18], [20], [21], [25], [26], [30], [32]	Chat bots generates answers whichthe human can easily understand	[5]- 90%, [8]-70%, [20]-74%, [21]- R2 Score of 0.9907, [26]-96%, [30]- intent 98.46% and story 96.1%	• Lack of native language support.
6	[6], [17], [23]	Chat bots generates answers whichthe human can easily understandwith the language selected by theuser	Not reported.	• Language support is limited tovery few popular languages

Medaret al., [31] have presented a comprehensive way in which machine learning algorithms can make use of agricultural factors to predict the yield of a particular crop in a particular field. They have implemented and compared the performance of Machine learning algorithms like Naïve Bayes and KNN in predicting the crop yield. They have concluded that, for the dataset they have used, the Naïve Bayes algorithm produced a better testing accuracy of 91.11% compared to the KNN algorithm, which produced 75.55%.

Gounderet al., [32] have developed a mobile application that will help farmers by answering agriculture-related queries. The dataset used to develop this system has been obtained from data.gov.in that consists of data from recorded calls from KCC Odisha. They have used the NLTK library to implement NLP algorithms for extracting the required information to be converted to a word vector. The word vector is used to compare to the existing knowledge base using a cosine similarity function, which outputs an appropriate answer to the user query.

Palasundramet al., [33] have worked to improve the sequence-to-sequence (Seq2Seq) model for natural answer generation in chatbots. This literature review has identified and reviewed the methods proposed to address the weakness such as utilizing additional embedding and encoders, using different loss functions and training approaches, as well as utilizing other mechanisms like copying source words and paying attention to a certain portion of the input. They have proposed changes such as structural modifications, augmented learning, beam search and complementary mechanisms to the Seq2Seq model to make it perform better at Natural Answer Generation. The structural modifications suggested are to have additional embeddings and encoders. Augmented learning introduces the usage of approaches like Alternative loss function learning, Multi-Task Learning, Deep reinforcement learning and Adversarial learning to further improve the answer generation quality of Seq2Seq. They have proposed the usage of the Diverse Beam Search algorithm as an improvement on the existing Beam Search method. They have concluded that these enhancements provide support for the Seq2Seq model during training and prediction to generate meaningful answers.

Mishra et al., [34] have made an extensive survey of Machine learning techniques used in crop yield prediction models by various scholars. The techniques discussed are Artificial Neural Networks, Information Fuzzy Network, Decision Tree, Regression Analysis, Clustering, Bayesian Belief Network, Principal Component Analysis, Time Series Analysis and Markov Chain Model. They have given a clear-cut conclusion on which algorithms produce the best result for which application using a table that shows the algorithm type and its application area.

Sun et al., [35] have used Remote Sensing (RS) and Geographic Information System (GIS) techniques to monitor the crop yield and its variation rate. The research presented in this paper has focused on the use of fusion techniques to solve the problem of crop yield modelling, including measurement to statistics result association, track fusion, and decision fusion in complex environments of remote sensing. A fusion technique for measurement to statistics result association has been proposed. This approach is based on the data fusion model and is suitable for TM and statistics sensors having different types of attributes. An optimum fusion method for combining the output of TM and statistics sensors in the system of crop yield modelling has also been investigated.

3. ANALYSIS OF EXISTING WORK

This section discusses the existing literature with respect to key features, namely, implementation, query analysis, intent identification, similarity function and answer presentation.

3.1 Implementation

A farm help system, even though being a simple idea to comprehend, the implementation of such a system has several factors that determines its usability. One of the major factors that needs to be considered in the development of such systems to be used in agricultural domain is accessibility to the target audience. Out of the various ways, messenger-based Chat/Talk Bots, web based conversational agents and dedicated mobile applications are observed to be the most usable and accessible methods of implementations.

Messenger-based Chat/Talk Bots:

These are systems that are developed to be incorporated with existing popularly used messaging applications like Telegram [8][23] and WhatsApp [30] using appropriate APIs. Its implementation is such that, the system is easily accessible to be used in the same way the user would message any of their contacts in the messaging application.

Web-based Chat/Talk Bots:

In this approach the system is hosted on the internet as a website [5][11][30] to be used by its beneficiaries. The aspect of not having to store any files except for cookies in the user’s smart device, to use the system is a huge advantage as these systems can require a huge amount of storage space.

Dedicated Applications:

In this method, an application having a simple UI [6] and the ability to work offline is developed to ensure easy access by everyone. The only drawback in this approach is that all the data required for the application to work has to be stored in the user’s device and the response time depends totally on the user’s hardware capabilities.

3.2 Query Analysis
Analyzing the input query and extracting the important information is the primary step that a chatbot system does. This step takes care of processing the human language input by removing stopping words, marking nouns, etc. and assigning weightage of each word in the processed input, so that the similarity function can produce an appropriate response. The various ways in which researchers have adopted to process input queries are, Sen2Vec [1] which is a convenient solution to transform sentences to vectors that contain high level of information of the original sentences. Usage of SOAP message pack [2] to present the extracted information based on the voice input which is primarily converted to an XML type format [2]. Implementation of Bag of Words approach by the use of NLP to parse sentences to extract keywords and root words from the input [3]. Further, a combination of several task-specific algorithms like tokenization, noise removal, lexicon normalization, stopping word removal [4], POS tagging [9][11], stemming [9], lemmatization [11], etc., have also been used to develop such systems.

3.3 Intent Identification
Intent identification is the process of probing into the output from the first step to identify and mark the words which form the basis of the query. Through this process, even after getting a word vector from the input, distinct words in the vector that decide the intent behind the question are given more weightage before sending the vector to the next stage in the process. The techniques using which this process is done include tagging nouns in the word vector as important words and increasing their weightage, as nouns are the intent deciders in most sentences. Using DBpedia to create entity-property-entity groups which can be queried on using SPARQL [13]. Other implementations include usage of Statistical Machine Translation [20] to identify words which decide the intent based on the frequency of the usage of the word and RASA NLU which is an open-source conversational AI that performs intent classification, entity extraction and response retrieval [30].

3.4 Similarity Function
The third and most important process in the working of a chatbot system is the answer extraction stage. This is where the pre-processed queries from the previous steps are used to extract an answer that is most appropriate based on its similarity to the existing knowledge base. The best fit for a given query can be obtained by the usage of a similarity function which has various ways in which it can be implemented. The word2vec algorithm uses a neural network model to learn word associations from a large corpus of text. It can cluster similar sentences without taking into account of order of the words. Another implementation is a black box approach [2] which implements a self-learning AI algorithm that find paths to the best answer from the given pre-processed query. Further, classification algorithms optimized using gradient decent are also being used as a function to extract answers to user queries. Other implementations include K-Nearest Neighbor algorithm [5], a multi-layered CNN approach for image-based queries [8], sequence to sequence deep learning RNN for speech-based queries [25] and other algorithms like Path similarity and Wu-Palmer similarity functions [11]. Further, an analysis was done to find the efficiency of majorly used ML algorithms. Using the R^2 score as the judging parameter, linear regression, support vector regression, decision tree and random forest regression were concluded to be the increasing order of their R^2 scores [21].

3.5 Answer Presentation
This is the final stage in the working of a chatbot, which is responsible for presenting the answer extracted from the model in a human-like way. One method of approach was to include human domain experts [26] in the process to improve the response generation process of a chatbot system. Other implementations that do not involve human work include the use of RASA X which is an open-source tool for Conversation-Driven Development, template-based Natural Language Generation, APIs provided IBM cloud [23] and by incorporating neural conversational models [30].

4. RESULTS OF EXISTING WORK
For the website-based implementation of chatbots, the testing accuracies of the models presented range from78% [10] to as high as 96% [26]. Out of the few standalone applications with simple UI, the reported accuracies were sparse and was found to be 75.55% [31]. The messenger integrated applications have reported accuracies ranging from 70% [8] to 96.1% [30] for answer generation and with the highest accuracy of 98.46% [30] for intent identification. Among the chatbots that use NLP to produce human-like answers, the accuracy ranges from 70% [8] to 96.1% [30].

5. RESEARCH CHALLENGES
Kissan Call Center (KCC) is a government organization which was formed to answer farmer’s queries on a telephone call in their own dialect. KCC is a really successful initiative that supports farmers by answering their queries related to farming and many other areas like animal husbandry, Nutrient management, Livestock products processing and packaging, etc. But it takes a lot of money to set up and run these KCC call centers. Hence, in recent times, a notable amount of research has gone into developing chatbot systems that can replace these call centers. These systems that are currently in use can be made to work better if the following ideas are incorporated into them.

5.1 Usage of live information
The queries from farmers can be classified into two major categories, queries based on recorded data and queries based on live data. For example,

General query: How long does it take to grow paddy in a field?

Live query: Is it profitable to grow paddy in the current climate?

KCC is observed to give back better responses to both kind of queries when compared to automated systems. It is because the people who work there have live access to constantly changing data and are able to make informed decisions based on it. However, chatbot systems work based on previously recorded data with which it was trained. The information that was relevant when the data was recorded may or may not be relevant when the query is being asked. To address this problem, the system should take into account live information that are relevant to the query asked and make decisions based on that.

5.2 Location based dataset
Natural features like soil fertility, type of soil, climatic trends, availability of water, etc. are major factors that influence farming in a particular location. These factors lead us to

consider that generic processing of queries without considering the location from which it was asked leads to the production of faulty or irrelevant answers. Data collected from one location might not produce appropriate answers for queries from other locations. To resolve this, it is suggested that the data set be stored as clusters based on the location it was collected from. When the system gets an input query, the location from which the query was asked should be considered first. The answer should be extracted based on data collected from the location that the query was asked. This approach will make the answer extracted to be more relevant and appropriate.

5.3 Implementation based on target user

A chatbot system is commonly developed as a messenger-based, a web-based or as an app-based. These three implementations should be adjusted to better suit the target users who are broadly split as digitally literate naïve farmers and digitally illiterate expert farmers. The messenger-based and web-based implementations can be easily used by the digitally literate naïve farmers. This is because they do not have issues related to connectivity and are assumed to be digitally literate. The app-based system updates live information at regular intervals. It can also work without any connectivity. Hence it will be useful for the digitally illiterate expert farmers.

6. CONCLUSION

Developments in technology and modern systems like ML-based chatbots have enabled farmers to be updated with the latest developments in the field of agriculture. There are a lot of such chatbot question-answering systems, but they fail to produce contemporary responses. This work was done to analyze the existing research work on chatbots in the agricultural domain and suggest how it could be made to produce better results by taking new factors like live information, location-based datasets and implementation styles into consideration. The existing systems can be made to produce more accurate and relevant answers to farmer queries by incorporating the suggested changes to the existing system. These suggested improvements will help in enhancing the overall usability of such systems. The future scope of research in this domain lies in improving and extending native language support provided by these applications so that it can be put to use by a greater number of people.

7. REFERENCES

[1] Jain, N., Jain, P., Kayal, P., Sahit, J., Pachpande, S., Choudhari, J.: AgriBot: agriculture-specific question answer system. IndiaRxiv (June 2019).

[2] du Preez, SJ., Lall, M., Sinha, S.: An intelligent web-based voice chat bot. In: IEEE EUROCON 2009, pp. 386-391. IEEE 18 May 2009.

[3] Vijayalakshmi, J., Pandimeena, K.: Agriculture talkbot using AI. International Journal of Recent Technology and Engineering (IJRTE), Vol. 8, pp 186-90 (Jul 2019).

[4] Gawade, H., Patil, V., Vishe, P., Kolpe, S.: College Enquiry Chat-Bot System. International Journal of Engineering Research & Technology (IJERT), Vol. 9, Issue 9 (Sept 2020).

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

[6] Jain, M., Kumar, P., Bhansali, I., Liao, QV., Truong, K., Patel, S.: FarmChat: a conversational agent to answer farmer queries. In: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, pp.1-22. Association for Computing Machinery, New York, NY, United States (Dec 2018).

[7] Ong, RJ., Raof, RA., Sudin, S., Choong, KY.: A Review of Chatbot development for Dy-namic Web-based Knowledge Management System (KMS) in Small Scale Agriculture. In: Journal of Physics: Conference Series, Vol. 1755, Issue 1, p. 012051. IOP Publishing (Feb 2021).

[8] Arora, B., Chaudhary, DS., Satsangi, M., Yadav, M., Singh, L., Sudhish, PS.: Agribot: a natural language generative neural networks engine for agricultural applications. In: 2020 In-ternational Conference on Contemporary Computing and Applications (IC3A), pp. 28-33. IEEE, Lucknow, India (Feb 2020).

[9] Niranjan, PY., Rajpurohit, VS., Malgi, R.: A survey on chat-bot system for agriculture do-main. In: 2019 1st International Conference on Advances in Information Technology (ICAIT), pp. 99-103. IEEE, Chikmagalur, India (Jul 2019).

[10] Sawant, D., Jaiswal, A., Singh, J., Shah, P.: AgriBot-An intelligent interactive interface to assist farmers in agricultural activities. In: 2019 IEEE Bombay Section Signature Conference (IBSSC), pp. 1-6. IEEE, Bombay, India (Jul 2019).

[11] Lalwani, T., Bhalotia, S., Pal, A., Rathod, V., Bisen, S.: Implementation of a Chatbot Sys-tem using AI and NLP. International Journal of Innovative Research in Computer Science & Technology (IJIRCST) Vol. 6, Issue 3. (May 2018).

[12] Feine, J., Morana, S., Maedche, A.: A chatbot response generation system. In: Proceedings of the Conference on Mensch und Computer, pp. 333-341. ACM, New York, United States (Sep 2020).

[13] Han, S., Shim, H., Kim, B., Park, S., Ryu, S., Lee, GG.: Keyword question answering system with report generation for linked data. In: 2015 International Conference on Big Data and Smart Computing (BIGCOMP), pp. 23-26. IEEE, Jeju, South Korea (Feb 2015).

[14] Singh, S., Thakur, HK.: Survey of Various AI Chatbots Based on Technology Used. In: 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), pp. 1074-1079. IEEE, Noida, New Delhi, India (Jun 2020).

[15] Liu, B., Xu, Z., Sun, C., Wang, B., Wang, X., Wong, DF., Zhang, M.: Content-oriented us-er modeling for personalized response ranking in chatbots. IEEE/ACM Transactions on Au-dio, Speech, and Language Processing. pp. 122-33. (Oct 2017).

[16] Jadhav, M., Kolambe, N., Jain, S., Chaudhari, S.: Farming Made Easy using Machine Learning. In: 2021 2nd International Conference for Emerging Technology (INCET), pp. 1-5. IEEE, Belgaum, India (May 2021).

[17] Wang, R., Ning, C., Lu, B., Huang, X.: Research on

answer generation method based on focus information extraction. In: 2012 IEEE International Conference on Computer Science and Automation Engineering (CSAE), Vol. 2, pp. 724-728. IEEE, Zhangjiajie, China (May 2012).

[18] Wang, Y., Sun, Y., Chen, Y.: Design and Research of Intelligent Tutor System Based on Natural Language Processing. In: 2019 IEEE International Conference on Computer Science and Educational Informatization (CSEI), pp. 33-36. IEEE, Kunming, China (Aug 2019).

[19] Vijayabaskar, PS., Sreemathi, R., Keertanaa, E.: Crop prediction using predictive analytics. In: 2017 International Conference on Computation of Power, Energy Information and Communication (ICCPEIC), pp. 370-373. IEEE, Tamil Nadu, India (Mar 2017).

[20] Yamada, T., Arakawa, T.: A study on output sentence generation method for question answering using statistical machine translation. In: 2013 13th International Conference on Control, Automation and Systems (ICCAS 2013), pp. 1199-1202. IEEE, Gwangju, South Korea (Oct 2013).

[21] Alves, BC., de Freitas, LA., de Aguiar, MS.: Chatbot as support to decision-making in the context of natural resource management. In: Anais do XII Workshop de Com-putaçãoAplicada à Gestão do MeioAmbiente e RecursosNaturais, pp. 29-38. SBC, Brazil, South America (Jul 2021).

[22] Patel, D., Suryakantbhai, N.: AGRICULTURE WITH MACHINE LEARNING. European Journal of Molecular & Clinical Medicine. Vol. 7, Issue 8, pp. 5491-5493 (Feb 2021).

[23] Mostaco, GM., De Souza, IR., Campos, LB., Cugnasca, CE.: AgronomoBot: a smart answering Chatbot applied to agricultural sensor networks. In: 14th international conference on precision agriculture, Vol. 24, pp. 1-13. Montreal, Quebec, Canada (Jun 2018).

[24] Fernandes, S., Gawas, R., Alvares, P., Femandes, M., Kale, D., Aswale, S.: Survey on var-ious conversational systems. In: 2020 International Conference on Emerging Trends in In-formation Technology and Engineering (ic-ETITE), pp. 1-8.IEEE, Vellore, Tamil Nadu, In-dia(Feb 2020).

[25] Chen, AP., Chansilp, K., Kerdprasop, K., Chuaybamroong, P., Kerdprasop, N., Shun-Fa, H., Yi-In, L., Ching-Tzu, H., Chih-Hua, C., Chen, ST., Zhang, RC.: B2B Marketing Crafts Intelligence Commerce: How a Chatbot Is Designed for the Taiwan Agriculture Service. In-ternational Journal of e-Education, e-Business, e-Management and e-Learning, Vol. 10, Is-sue 2, pp. 114-24 (2020).

[26] Nayak, V., Sowmya, NH.: Agroxpert-Farmer assistant. Global Transitions Proceedings, Vol. 2, Issue 2, pp. 506-12 (Nov 2021).

[27] Kale, SS., Patil, PS.: A Machine Learning Approach to Predict Crop Yield and Success Rate. In: 2019 IEEE Pune Section International Conference (PuneCon), pp. 1-5. IEEE, Pu-ne, India (Dec 2019).

[28] Kavita, M., Mathur, P.: Crop Yield Estimation in India Using Machine Learning. In: 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA), pp. 220-224. IEEE, Noida, New Delhi, India (Oct 2020).

[29] Manjula, A., Narsimha, G.: XCYPF: A flexible and extensible framework for agricultural Crop Yield Prediction. In: 2015 IEEE 9th International Conference on Intelligent Systems and Control (ISCO), pp. 1-5. IEEE, Coimbatore, Tamil Nadu, India (Jan 2015).

[30] Momaya, M., Khanna, A., Sadavarte, J., Sankhe, M.: Krushi–The Farmer Chatbot. In: 2021 International Conference on Communication information and Computing Technology (ICCICT), pp. 1-6. IEEE, Mumbai, India (Jun 2021).

[31] Medar, R., Rajpurohit, VS., Shweta, S.: Crop yield prediction using machine learning techniques. In: 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), pp. 1-5. IEEE, Mumbai, India (Mar 2019).

[32] Gounder, S., Patil, M., Rokade, V., More, N.: Agrobot: An Agricultural Advancement to Enable Smart Farm Services Using NLP. Journal of Emerging Technologies and Innovative Research (2021).

[33] Palasundram, K., Sharef, NM., Kasmiran, KA., Azman, A.: Enhancements to the Sequence-to-Sequence-Based Natural Answer Generation Models. IEEE Access. pp. 45738- 45752 (Mar 2020).

[34] Mishra, S., Mishra, D., Santra, GH.: Applications of machine learning techniques in agricultural crop production: a review paper. Indian Journal of Science and Technology, Vol. 9, Is-sue 38, pp. 1-4 (Oct 2016).

[35] Sun, L., Zhu, Z.: A RS/GIS-Based System for Monitoring Crop Yield. In: 2011 International Conference on Computer Distributed Control and Intelligent Environmental Monitoring, pp. 17-20. IEEE, Changsha, China (Feb 2011).