FARMER-BOT: AN INTERACTIVE BOT FOR FARMERS USING ARTIFICIAL INTELLIGENCE

1. Abstract

India is an agro-based economy, and knowledge of agricultural methods is critical to achieving optimal agricultural growth and output. We have created an agricultural chatbot based on the Kisan Call Center dataset to help farmers with their questions. This system is capable of answering questions on the weather, market prices, plant protection, etc.. It may be accessed from any electronic device, and the information is presented in an understandable manner. The method is built on a Neural Networks model that has a 73% accuracy rate. Farmers can develop towards easier knowledge regarding farming techniques and thus a better agricultural output with such a system. The Call Center workforce's job would be made easier, and the hard work of numerous such individuals may be diverted toward a more worthwhile purpose.
Key Words: NLP, Neural Network, AI, Chatbot.

2. Introduction

In India, farmers and policymakers are desperately seeking a smart agriculture decision support system that will allow them to make rapid and effective decisions based on agricultural data. We are certain that digital agriculture, which uses technologies like Artificial Intelligence (AI), Machine Learning, and advanced analytics to empower and benefit small-holder farmers and decision-makers in agrarian economies like India. In 2020, 41.49 percent of the workforce in India were employed in agriculture. Chatbots are conversational virtual assistants which automate interactions with end users. Artificial intelligence powered chat bots, using machine learning techniques, understand natural language and interact with users in a personalized way.

In early days chat bots were used mostly in retail, travel, media or insurance players. Agriculture could also leverage this emerging technology by assisting farmers with answers to their questions, providing advice and recommendations on farming related problems. This conversational assistant uses Natural Language Processing

techniques to understand the user queries in their natural language. This will make the system understand even the grammatically not well defined sentences as input queries. It also answers all the basic queries of a farmers and can also provide a possible information and solution related to agriculture.

A chatbot has been created to specifically address the needs of Indian farmers. The information was gathered from farmer discussions at Kisan Call Centers (KCC). This project is based on the idea of assisting farmers by providing them with technology. This project will deliver services such as cultivating a variety of crops, protecting diseased plants with appropriate fertilisers, and weather-related irrigation information, among other things.

Features of System:
The following set of features covers the two districts, Tiruppur and Thiruvannamalai respectively.
1. Historic Rainfall data from 2017 can be visualized from various stations representing the whole district can be visualized.
2. For Quick and Easy interpretation, the User has the option to see all the stations market information of crops, weather, cultivation, nutrient management etc., representing the whole district. 3. Plant Protection: In the KCC dataset, 60.6% of the farming calls were related to remedies for protecting
4. Pests and diseases: Agri-experts recommends the dosage of pesticides for a particular crop disease to farmers. 5. Weather: Farmers eagerly sought weather information, as rains can wash away expensive sprayed pesticides and weather conditions determine the best time to harvest crops. 6. Best Practices: Information related to best practices can help increase yield in terms of the quantity or quality of potatoes with respect to irrigation, harvesting, etc. 7. Unbiased Recommendations on Products: Farmers wanted recommendations from agri-experts on the purchase of products such as seeds, fertilizer, etc.

3. Literature Survey

Sohampachpande et al. [1] used a simple Neural Network to create an Agribot to answer the questions of the farmers based on the dataset at Kisan Call Centre [KCC]

Here the queries are processed under basic text formatting, regional spell check and stemming in order to produce the query corpus. Then the corpus is processed in neural network and the model is generated. The Bot thus provides the answers to queries based on higher priority.

Yashaswini et al. [2] have proposed a technique machine learning and Natural Language Processing to analyse the parameters such as weather , season, rainfall ,and type of soil of a region . These parameters are required to provide suggestions about the crops to grow for the farmers to get good yield. These data are collected from different sources like government websites and repositories. These data is processed to remove the noise and unwanted non- related data. Then they are stored securely in a database. This database is trained thoroughly using machine learning in tensor flow architecture using KNN (K-Nearest Neighbor) algorithm. The trained data is tested for its accuracy repeatedly by using a small validation data.

Jain et al. [3] have proposed a cloud-based scalable services to provide usable solution for queries in KCC dataset. It is worked with two interface modalities: Audio-only and Audio-Text. It uses Natural Language Processing(NLP) to convert speech to text and vice versa. The process of tokenization on the dataset has created the corpus. The Sequential neural network is used for training the corpus and the model is obtained. Chatbots which use cloud services as core technologies might have repeated limited access to Internet in rural areas.

Vijayalakshmi et al. [4] have proposed the TalkBot based on NLP techniques to understand the user queries from different government websites in their natural language. This facilitates the system understand even the grammatically not well defined sentences as input queries. The user queries underwent the pre-processing stage where the query is first tokenized into words, then the stop words like a, is, the, etc., are removed. Because the stopwords are not needed to contribute to the probability of classifying the queries based on their respective classes. Subsequently, the stemming process is carried out where the words are converted into their root words. These root words are then converted to a bag of words which are further converted to a vector form. These vectors are processed and trained efficiently by the classification algorithm. A neural network of 3 layers is constructed and the error

is optimized using the gradient descent algorithm. The class with the highest probability is iterated to get the accurate results. It can be further developed by making the bot to response in their regional language.

Du Preez et al. [5] have presented the design and development of an intelligent voice recognition chat bot using a web Service technology. A black box approach is used to control the communication structure, to and from the web-service, the web-service allows all types of clients to communicate to the server from any platform. The service provided is accessible through a generated interface which allows for seamless XML processing; whereby the extensibility improves the lifespan of such a service. By introducing an artificial brain, the web-based bot has generated customized user responses and aligned to the desired character. Questions that are not well defined are further processed using a third-party expert system (an online intelligent research assistant), and the responses are archived, improving the artificial brain capabilities for future generation of responses. Due to the use of online intelligent research assistant, their system always shows a good network capality. Rural areas have only limited access to internet facility.

Gawade et al. [6] have implemented the College Enquiry chatbot using the keyword matching algorithm to help the students to stay updated with their college activities. First, this bot has read the input query from the user. The query is pre-processed with the removal of stopwords. Further the remaining keywords are fetched and performed a match with the keywords in Knowledge base and thus provided an appropriate response. Finally the Database module is used to call proper services using entity information to find proper data. The keywords will be matched with the help of keyword matching algorithm. It returns the query response to the bot. Chat-bot packages the data into proper response for display by the client.

Anand et al. [7] have implemented the neural networks s (with an input layer, single hidden layer and an output layer) to detects the e Tuberculosis found in the 3D CT-images of patients' chests. The model has achieved validation accuracy of 0.20. The accuracy can be improved by adding more meaningful layers and/or adding more neurons per layer.

In conclusion, the findings of the studies[2,4] can further be improved by using appropriate classifiers. Also from the studies[3,5,6], it is inferred that there will be only limited access to Internet in rural areas while the core technology used is Cloud/Web based Services.

4. Methodology

NLP based Farm AI-Bot model Queries of farmers are available as unstructured data is in the form of text, chats, emails, social media, survey responses are present everywhere today. Queries of farmers are available as an unstructured data in KCC government website. Text can be a rich source of information. Due to its unstructured nature it is very hard to extract insights from it. Farmers query classification in a chatbot system is one of the important task in supervised Machine Learning (ML). It is a process of assigning tags/intents to queries helping farmers to automatically and quickly structure and analyze text in a cost-effective manner. It is one of the fundamental tasks in Natural Language Processing with broad applications for Farmer's chatbot such as topic-labeling, intent-detection etc.

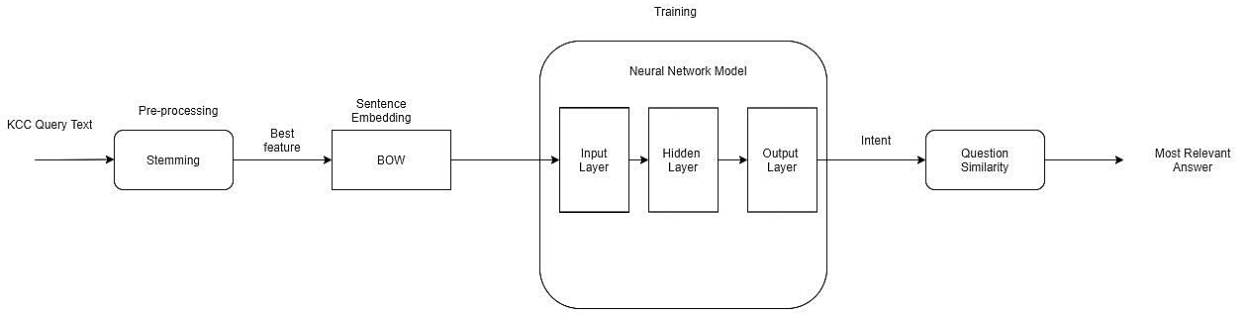

Fig 2: Farm AI-Bot Architecture

4.1 Collection of dataset

Farmers' information inquiries are obtained from the Kisan Call Center

(KCC) in the website. The Government of India has made all logs of calls to the KCC from specific month year to publicly available for the districts Tiruppur and Thiruvannamalai. In total, this corpus contains data for 3225 calls. Each call log has 11 fields, including the date and time of the call, location, crop (one of the 306 crop types), query, and the answer provided by the KCC agri-expert

For implementing Farm AI-Bot, our system is restricted to 2 districts of the state of Tamilnadu. We have relevant information mapped and dataset contain maximum information. There were 3225 calls related to Tiruppur and Thiruvannamalai in the KCC dataset.

4.2 Preprocessing Corpus Generation

We need to preprocess our dataset by removing punctuations & special characters, cleaning texts, and applying lemmatization Simple text cleaning processes: Some of the common text cleaning process involves: • Removing punctuations, special characters, URLs & hashtags • Removing leading, trailing & extra white spaces/tabs • Typos, slangs are corrected, abbreviations are written in their long forms • Stemming: Refers to the process of slicing the end or the beginning of words with the intention of removing affixes(prefix/suffix) • Lemmatization: It is the process of reducing the word to its base form

4.3 Embeddings

Bag of Words

Bag-of-Words technique we can convert variable-length texts into a fixed-length vector. If your dataset is small and context is domain specific, BoW may work better than Word Embedding. Context is very domain specific which means that you cannot find corresponding Vector from pre-trained word embedding models (GloVe, fastText etc).

5. Implementations

Neural Network model A simple shallow neural network model has been designed to generate the responses for the farmer's queries. The model has three layers in it as shown in Fig 1. The first layer accepts the preprocessed text data and are flattered before passing then to the next two layers.

5.1 NeuralNet based on farmer's queries and answers

In Farm AI-Bot, firstly we have trained the model by giving input as farmer queries and answers as labels using NeuralNet model of Deep Learning. In such case, the model couldn't predict the answers accurately for queries as there is lot of similarities in the answers of different question.

For example,

Case 1:

Query Type: Nutrient Management
Query: Asking for watermelon top dressing fertilizer management
Answer:Recommended for apply urea 44kg / ac

Case 2:

Query Type: Plant Protection
Query: Asking about cowpea aphids management
Answer: Recommended for spray imidachloprid -5 kg/area

In the above two cases, the queries are belong to different categories might have similarities in answers. When we train the model directly with query text and answers, it couldn't predict the answers accurately as the model is not able to differentiate the category of query text.

5.2 NeuralNet based on farmer's queries and its intents

Now, we have trained the model by giving input as farmer's queries and its intents as labels using NeuralNet model of Deep Learning. Here the model predicts the query type from the user's query better than the other models discussed with the intent obtained, similarity of the queries has been further analysed. Query similarity finds for the closest match of the user query with the collected query and thus provides the most relevant answer corresponding to the matched one.

5.3 Fine Tuning

In the fine tuning phase, We have concatenated the queries with area of location from where the query arises The concatenated queries undergoes pre-processing stages to get processed input (in form of vector). Then train the model with this dataset using NeuralNet algorithm of Deep Learning to increases the accuracy of Farm AI-Bot.

6. Experimental Results

6.1 Hardware used:

Google Colab notebook was used to train the model. A general purpose RAM size of 8GB was alloted with a 2.3GHz Intel Xenon CPU.

6.2 Software used:

Implementation is done using Python and the libraries used are numerous, nltk, Jason, PyTorch.

7. Results:

The two models are trained for 20 epochs, accuracy metric is used to study the performance of the model during training. We have build a simpler Neural Network model for classifying and have achieved validation accuracy of 0.73. We also infer that BOW embeddings yield the good accuracy in neural network model when compared to others. Hence in our FarmAI-Bot, we have chosen BOW embeddings to generate the response of query.

8. Conclusion:

FarmAI-Bot is based on simple and shallow neural networks (with an input layer, single hidden layer and an output layer. We have build a neural network model with accuracy of 73%. However, a rigorous assessment of the study revealed that this neural network model can be improved by adding more meaningful layers and/or adding more neurons per layer in such a way that the model doesn't become intractable.