Agrobot : An agricultural advancement to enable smart farm services using NLP

[1]Susheel Gounder, [2]Mahaleppa Patil, [3]Vishal Rokade, [4]Nikita More
[1234]Student
Department of Information Technology,
International Institute of information Technology, Pune, India

Abstract : Presently we receive around 5 lakh calls on our krushi call center (KCC) every month. This shows that our farmers want relevant information today and are willing to ask for it. However, expanding call centers is very expensive and requires huge capacity building. That is why we must develop intelligent chatbots where farmers can ask in their local language any query and our bot will give a meaningful reply. The project will involve a lot of natural language processing because we can't expect farmers to type commands to our bot (and anyways no one likes to talk to a machine!). So we can't restrict the farmer's interaction with our bot as press 1 for this, press 2 for this and so on. We have to make it as human-like as possible. Our chatbot will benefit from the localized context information about the farmer. Example, there would be a dedicated chatbot for say paddy growers in Seemandhra or orange growers of Jhalawar which will benefit from the specific knowledge there. For example, it would know what kind of varieties farmers typically grow, what are the diseases they are prone to, what are the solutions, what are the market prices there, what is the weather information and so on. This would also make our task a lot simpler and our bot more intelligent. Initially we are just looking to develop a bot for a specific context to give a proof of concept.

I. INTRODUCTION

Agriculture is the backbone of our nation. India being second in the worldwide ranking of farm outputs. As agriculture is so vastly spread in our nation farming becomes the primary occupation for most of the people living in India. Farming is mostly carried out by uneducated villagers who aren't aware about the solutions on various crop related issues. As India is also moving fast towards technology, the idea is to bridge the gap between the farmer and the knowledge they are unaware of. The basic idea is to use technology in the form of a simple mobile app that is very farmer friendly and very informative. Agriculture extension is vital to disseminate information on new agriculture techniques and agronometric practices among the farmers. Even if we develop the best varieties of seeds but if we are not able to disseminate its information amongst the farmers, it can hardly have any impact.

Once the farmers come to know about the advanced seed, only then will they demand it and use it. Also just using the seed may not be sufficient. It may need a different way of applying fertilizers or different watering time. The farmer needs to know about all this, and hence the value of extension. The main feature of the Agro BOT system includes information retrieval to its user from anywhere in the form of statistical information about fertilizer, diseases, soil concentration for particular crops, an appropriate solution to any diseases of crops. System aims to use machine learning and natural language processing for automatic response to its user at any time by recognizing any of the languages which are familiar to its user.

II. OBJECTIVES

The main objective of this project is to provide farmers a platform that is farmer friendly wherein the farmers can easily get real time services with proper accuracy.

The aims of the project are:

1. To allow farmers to get 24/7 service on their fingertips.
2. To be able to translate and give proper information in regional languages.
3. To replace the KISAN CALL CENTERS (KCC) issues.
4. To make valid and reliable agricultural knowledge available.

III. OVERVIEW

In recent years, Information Communication Technology (ICT) has begun to be actively used in agriculture extension. Some of the ICT based channels are:

1. Farmer's Portal: which is a central repository of all the information required by farmers at all stages at one location. Kisan Call Centres: where a toll free number 1800-180-1551 is available for farmers to query in regional languages. The system maintains a huge amount of information and receives over 5 lakh calls per month across India.
2. mKisan: which is used to send sms advisories by Agriculture Officers and Experts to farmers and other registered users in hindi or English or regional languages. KisanSuvidha Mobile App (available on Google Play).

There are about 14 crore farmers in India on the Internet. Most of us believe farmers don't use the internet. So why bother with this? But the situation might be changing silently right under our nose (and may already have changed!). As per the latest report by The Internet and Mobile Association of India (IAMAI), there are already about 11 crore people in rural India who are accessing the internet via their mobiles (not just mobile, but mobile internet!). However, reality means more than this.

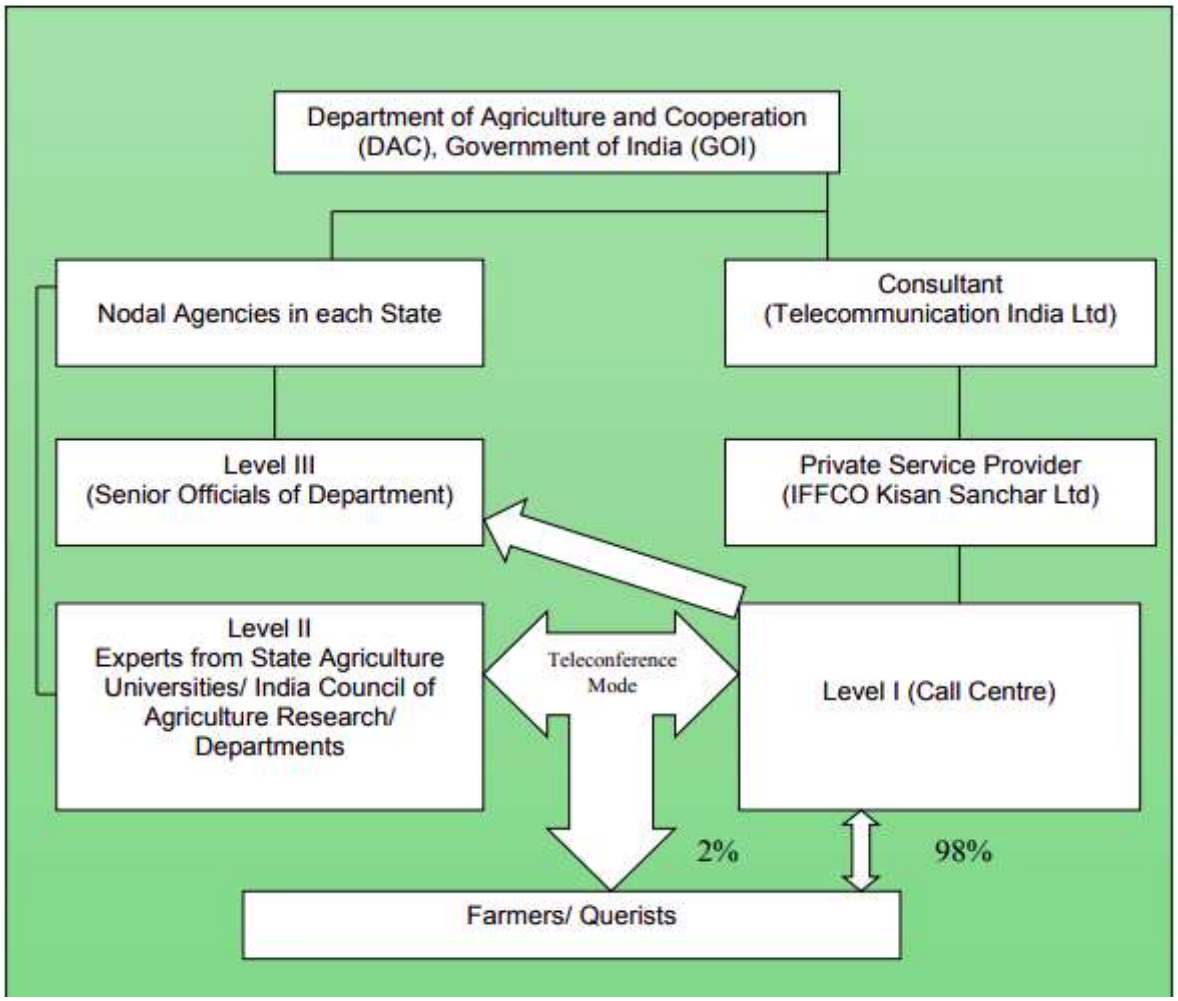
There would be many more rural-to-urban migrants who are working in cities in the unorganized sector but have a family farm back in the village cultivated by their family members. Also since an average rural household holds less number of mobiles than an average urban household, we are actually looking at a good percentage of rural households being on the internet. Hence, the use case of this application would be

IV. BACKGROUND AND LITERATURE SURVEY

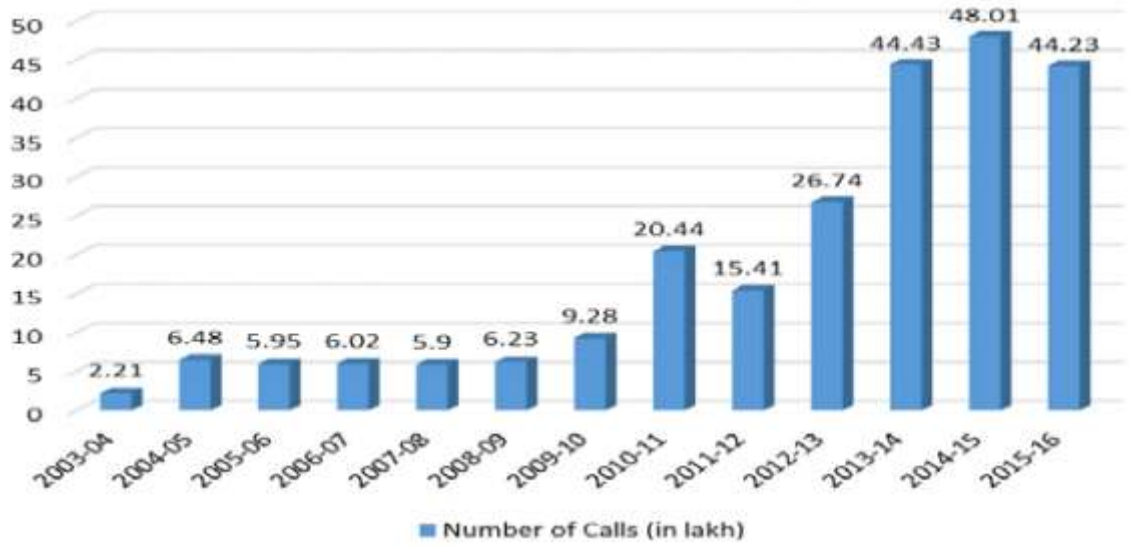
1. Kunal Verma, Dinesh Pabbi and Avnish Singh Jat ——AGRICULTURE ADVANCEMENT USING ARTIFICIAL INTELLIGENCE —— To solve agriculture issue, the Indian government has a helpline number (1800-180-1551), which aims to solve the issues faced by farmers as well as provide knowledge of modern farming techniques. But as we all know, the service could not provide 24x7 support due to improper implementation as well as the huge crowd of farmers needing help. We think we have found a solution to this issue in AI via Bots. Our idea is to design a bot that can be accessed through a website, a standalone app and also in apps such as Facebook messenger, Google allow. Bots are nothing but command based assistants that respond to predefined commands.

2. Mr. Anilkumar T.G. and Mr. T.D. Ganesh, Enable Smart Farm in Digital India with Artificial Intelligence and IoT The complete Digital ecosystem to Enable Smart Farm as a Service can be built on a Service Oriented Architecture meta model. This model not only can help the farmer but also provides a business opportunity to service providers like entrepreneurs, government authorized NGOs to develop, operate and maintain the system. The service consumers are the smart farmers who are going to benefit out of this end-to-end service in the agriculture farming activities.

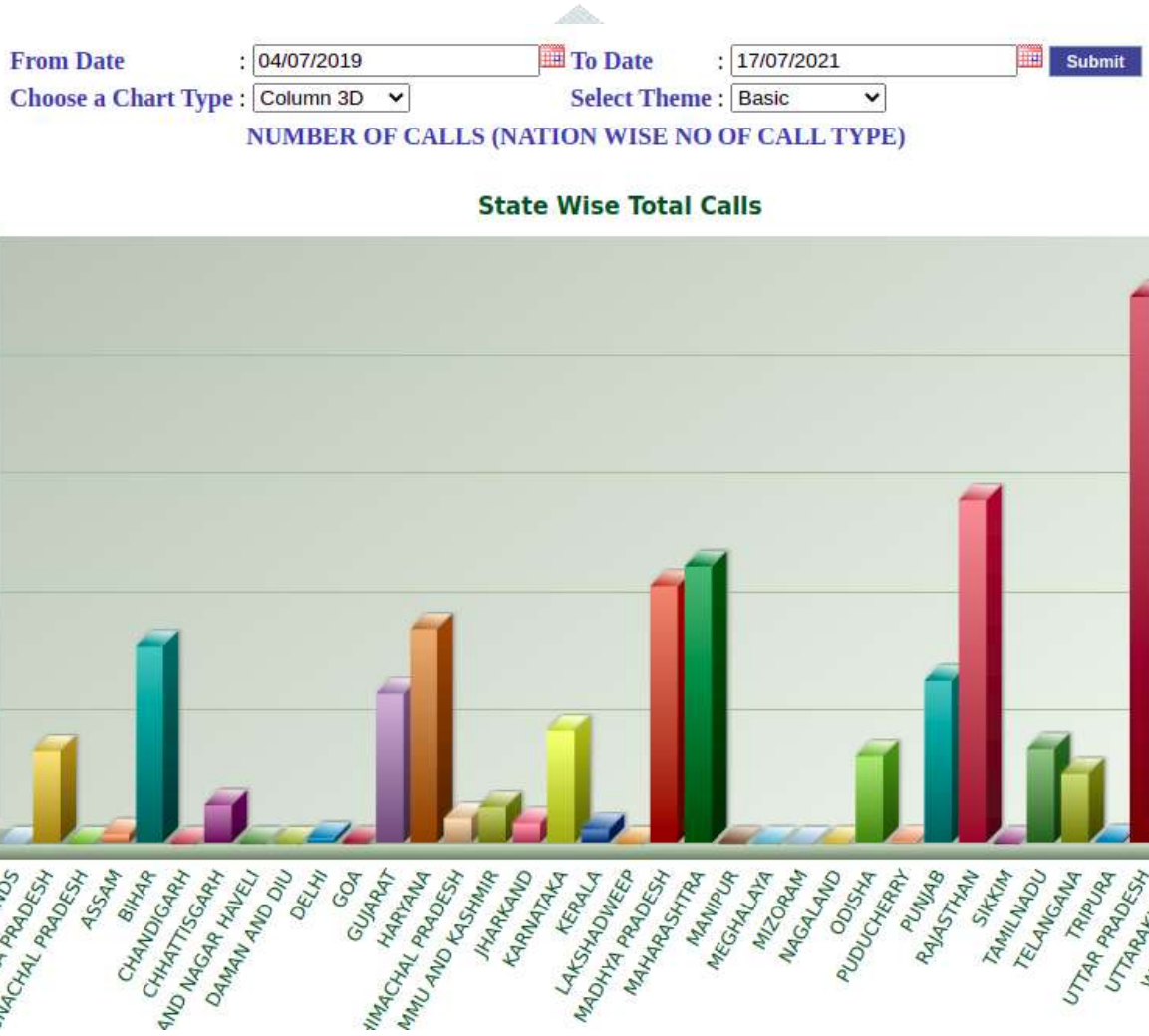
3. Other aspects that we surveyed are called the changing landscape. The present change is currently driven by 3 factors: cheaper smartphones, cheaper and faster data transfer with wide coverage and localization of the internet i.e development of audio/ visual content in local languages. With these 3 factors taken into consideration farmers will soon be on the internet and social media.
4. NLP (Natural Language Processing) is the field of study that not only helps in generating human understandable language but at the same time has the capabilities to understand this language. Natural Language generation and Natural Language understanding has been a widely grown field of study and research.



Source : Kisan Call Center Overall flow chart [Online]



Source: mkisan Year Wise number of Calls



Source: mkisan State wise number of phone calls received

From the above presented data it's easy to understand the number of calls received from farmers is enormous and expanding such call centers would be a tedious task and also would cost too much. Hence, to overcome such issues we could negate the usage of KCC and replace it with a specialized chat bot that would not only be available 24/7 but also provide valuable and reliable insights that would not hinder farmer's occupation.

V. PROPOSED THEORY

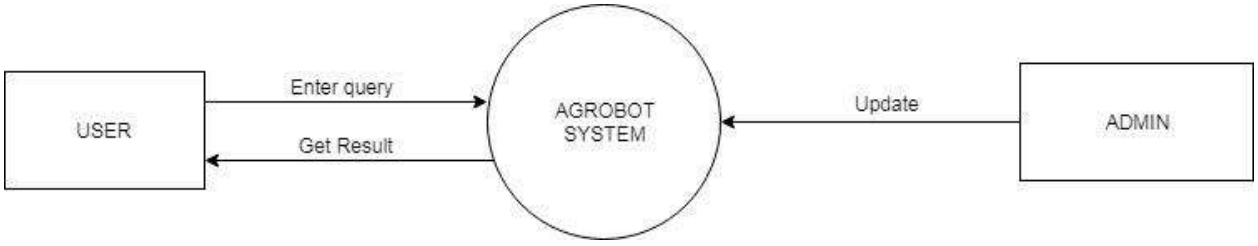
The proposed implementation will make use of NLP tools and algorithms in order to perform data extraction, data exploration, data cleaning and finally data representation. We have made use of the data source provided by the data gov to get the KCC recorded calls in the form of CSV or JSON format. data gov is an openly available data source provided by the indian government. This source provides data in two formats as stated previously (JSON and CSV). The data for KCC can be found via searching keywords. The data for each state is available separately, for study purposes we have made use of the data set of ODISHA farmers with their queries.

A Sample JSON format would look something like this,

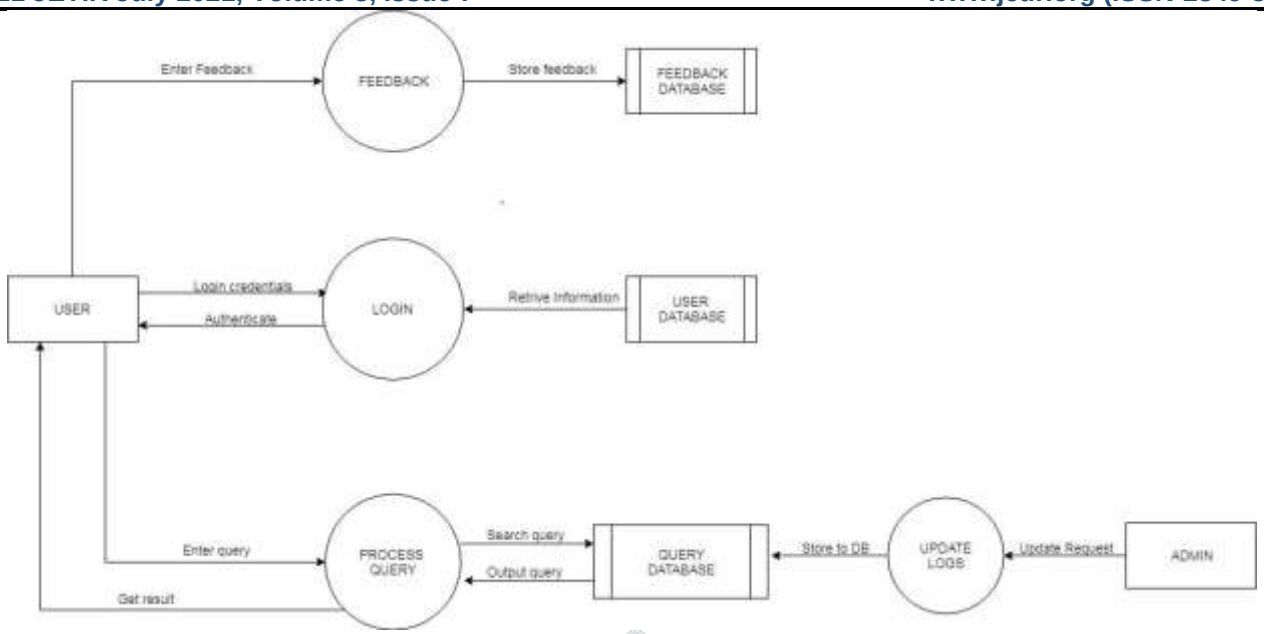

```
"data": [
    {
      "Season": "NA",
      "Sector": "HORTICULTURE",
      "Category": "Vegetables",
      "Crop": "Chillies",
      "QueryType": "\tPlant Protection\t",
      "QueryText": "Green peach aphid in Chilli",
      "KccAns": "Recommended to spray Acetamiprid 20SP @ 80gm with 200 liter of water per acre for control of Green peach aphid in Chilli.",
      "StateName": "ODISHA",
      "DistrictName": "BARGARH",
      "BlockName": "BHATLI",
      "CreatedOn": "2020-02-01T08:04:25.773"
    },
    {
      "Season": "NA",
      "Sector": "HORTICULTURE",
      "Category": "Vegetables",
      "Crop": "Brinjal",
      "QueryType": "\tPlant Protection\t",
      "QueryText": "Seedling treatment of Brinjal    ",
      "KccAns": "--Recommended for seedling treatment of Brinjal with (Thiophanate Methyl 45% + Pyraclostrobin 5% FS ) @ 3 ml per 1 Litre water as preventive measure for soil borne fungal infestation",
      "StateName": "ODISHA",
      "DistrictName": "BARGARH",
      "BlockName": "AMBABHONA",
      "CreatedOn": "2020-02-01T08:18:10.693"
    },
    {
      "Season": "NA",
      "Sector": "AGRICULTURE",
      "Category": "Others",
      "Crop": "Others",
      "QueryType": "Weather",
      "QueryText": "Farmer asked query on Weather",
      "KccAns": "--Recommended for seedling treatment of Brinjal with (Thiophanate Methyl 45% + Pyraclostrobin 5% FS ) @ 3 ml per 1 Litre water as preventive measure for soil borne fungal infestation",
      "StateName": "ODISHA",
      "DistrictName": "BARGARH",
      "BlockName": "AMBABHONA",
      "CreatedOn": "2020-02-01T08:21:11.127"
    },
```


The data shows features like season, sector, category, crop, querytype, queryText, KCCAns, statename, districtName, blockName and createdOn. Based on these features the further processes are carried out.

5.1 Data flow diagram (DFD)

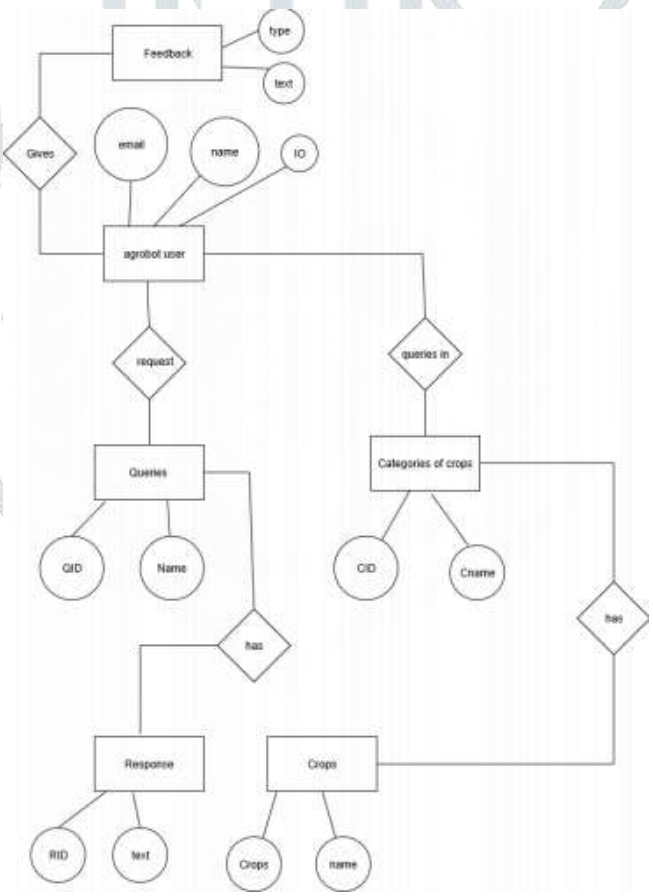


DFD LEVEL 0



DFD LEVEL 1

5.2 Entity Relationship Diagram



Entity Relationship Diagram

5.3 Implementation

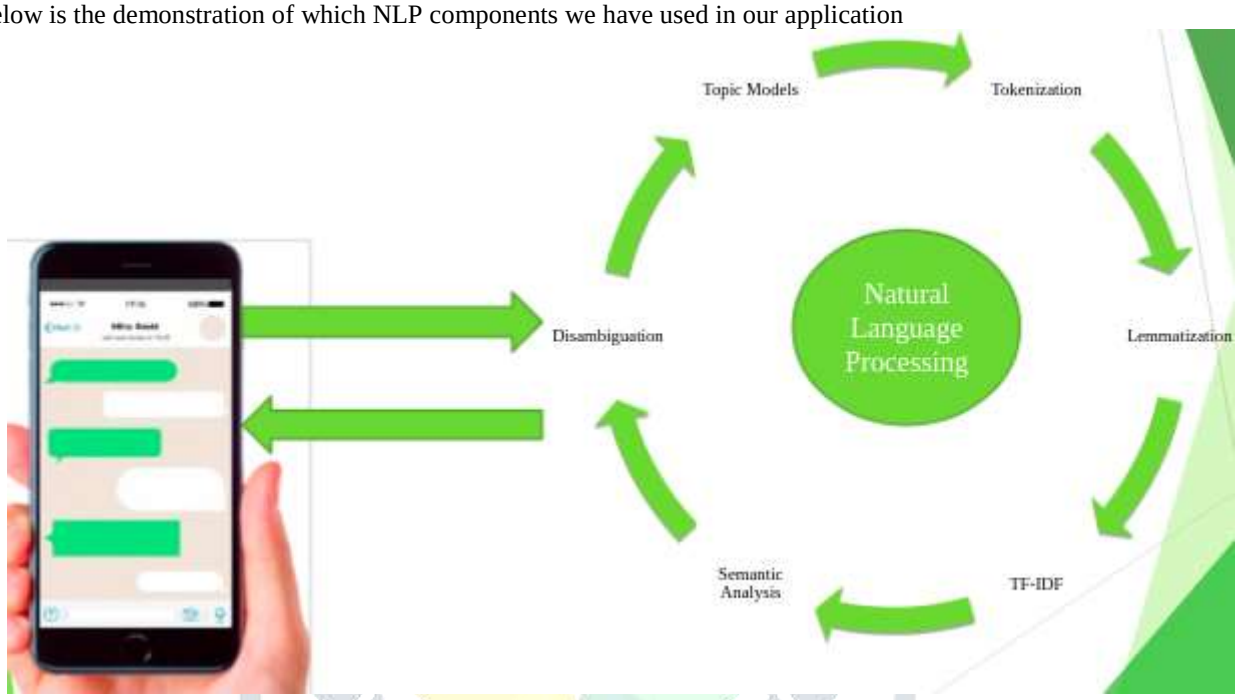
As stated previously, we have used the data source having features like season, sector, category, crop, querytype, queryText, KCCAns, statename, districtName, blockName and createdOn. Now, we would be shortlisting only the valuable features that would help in providing maximum accuracy in the results. This stage is called "feature extraction", in this stage we have analyzed the data and found out which features would provide a greater impact in the result. After analyzing the features and its relevance we found out all the features had significant value to the result. Even createdOn is useful as we could find out the peak time where farmers have a lot of queries. Out of all the features we come to a conclusion that queryText and KCCAns is the most important feature for text generation and understanding. Hence we used these two features for further NLP processing.

The implementation of the application can be divided into two parts, client and server.

A. AgroBot Server:

The server had the actual business logic and NLP algorithms that extracted human understandable text. For the server we have made use of a python flask framework that serves API endpoints for the client to interact with. For actual text mining and extraction we have made use of the NLTK library that performs various NLP algorithms on the features that we had enlisted previously.

Below is the demonstration of which NLP components we have used in our application



NLP Components

The following are the functions of these components

1. **Tokenization :** It's the process of breaking a document into sentences first and then breaking each sentence into words. This process is heavily used at the beginning of the process and is therefore called "text pre-processing". The query text and KCC answers both are sent through the tokenization function.
2. **Lemmatization :** After tokenizing the words it is sent to a lemmatization instead of stemming. The reason we have used lemmatization instead of stemming is because stemming would just cut short the word without thinking of the presence of the word, whereas lemmatization makes use of the word net dictionary that stems the word to the root which is present in a dictionary. Hence having a meaningful root will be useful in identifying the meaning. This also comes under "text pre-processing".
3. **TF-IDF :** It's an algorithm basically used for identifying the uniqueness in the document. The reason for using TF-IDF weighting system instead of other weighing systems is the popularity and accuracy it has.We checked various recommender systems and found that most of the recommender systems used TF-IDF instead of PCA (principal component analysis). The TF-IDF makes use of the following formula,

Recommended tf–idf weighting schemes

weighting scheme	document term weight	query term weight
1	$f_{t,d} \cdot \log \frac{N}{n_t}$	$\left(0.5 + 0.5\frac{f_{t,q}}{\max_t f_{t,q}}\right) \cdot \log \frac{N}{n_t}$
2	$1 + \log f_{t,d}$	$\log\left(1 + \frac{N}{n_t}\right)$
3	$(1 + \log f_{t,d}) \cdot \log \frac{N}{n_t}$	$(1 + \log f_{t,q}) \cdot \log \frac{N}{n_t}$

4. **Semantic Analysis :** Once we have done the preprocessing part there is a need to identify the meaning of the word. The semantic analysis of natural language content starts by reading all of the words in content to capture the real meaning of any text. It identifies the text elements and assigns them to their logical and grammatical role. It analyzes context in the surrounding text and it analyzes the text structure to accurately disambiguate the proper meaning of words that have more than one definition.

Once the preprocessing part is completed, we then receive a vector with weighted values. These weighted values are then fed to cosine similarity function.

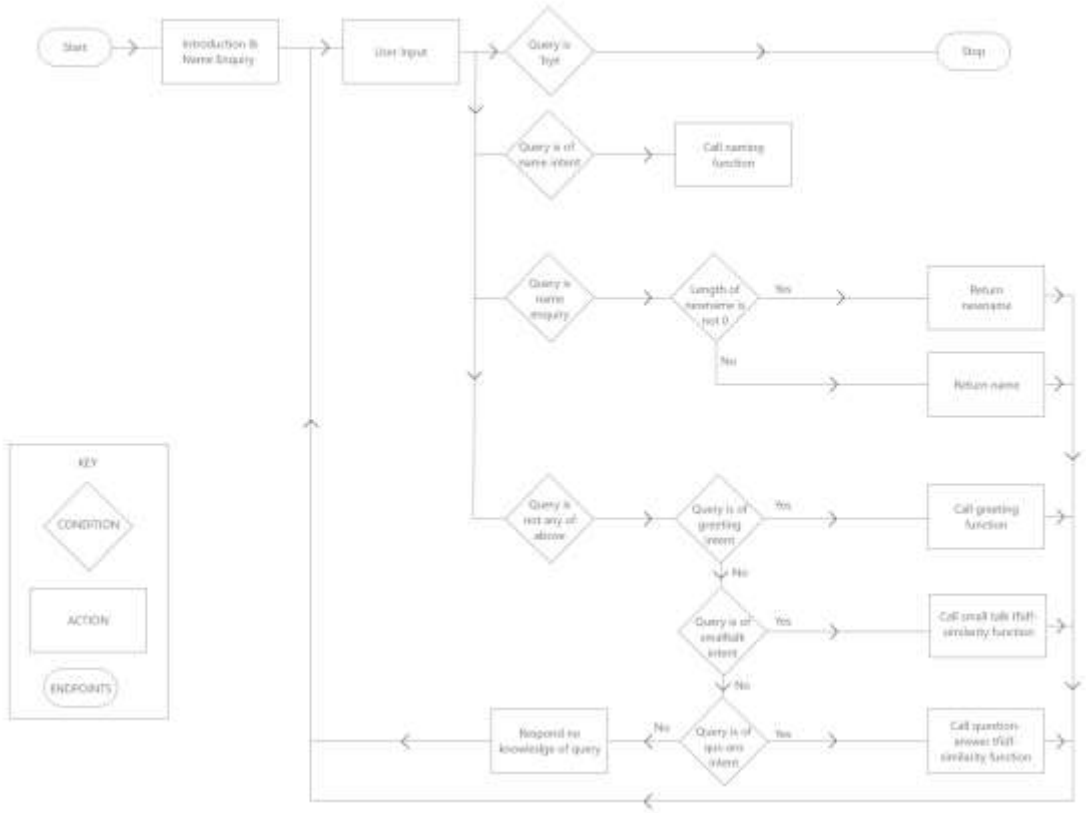
Here's a sample vector,

TD-IDF Vectorizer

	45	asked	borne	...	treatment	water	weather
Doc1	0.000000	0.5	0.000000	...	0.000000	0.000000	0.5
Doc2	0.235702	0.0	0.235702	...	0.235702	0.235702	0.0

The vector is then passed to the cosine similarity function which receives a vector again showing the similar documents to the entered document.

Following is the flow diagram for Agrobot (chatbot),



Agrobot flowchart

As per the above diagram, the chatbot is divided into 2 parts, 1st is the small talk and second is the actual data set related message. Whenever a user tries to enter any data which is very irrelevant to the context, then the small talk is used. Small talk is basically a daily routine language. For now, we have considered a data set which includes day to day verbal communication. For eg: A farmer might ask certain questions which are not present in the dataset such as what is your name? Who are you? etc. Such question answers are pre defined in the small talk function. We have specified the outcome for certain questions instead of throwing a singleton error.

Final output in the server,

```
Hello, my name is Agrobot. What is your name? : Susheel

Hi susheel , I am Agrobot. How can I help you? If you want to exit, type Bye. :
```

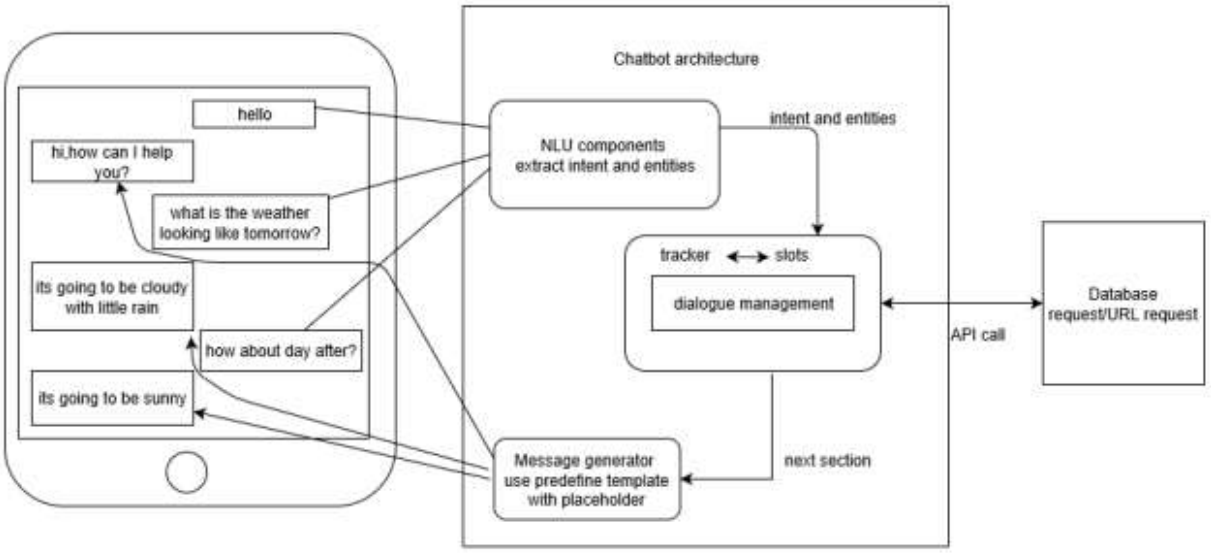
Note: This output is just for demonstration purpose, the response will be bind to JSON and handled by a particular endpoint. For this we have used python's flask framework

Next step was the deployment of the server. To deploy the server we made use of Heroku , an online platform that allows you to deploy your applications. We deployed the python flask server in heroku.

B. Agrobot Client :

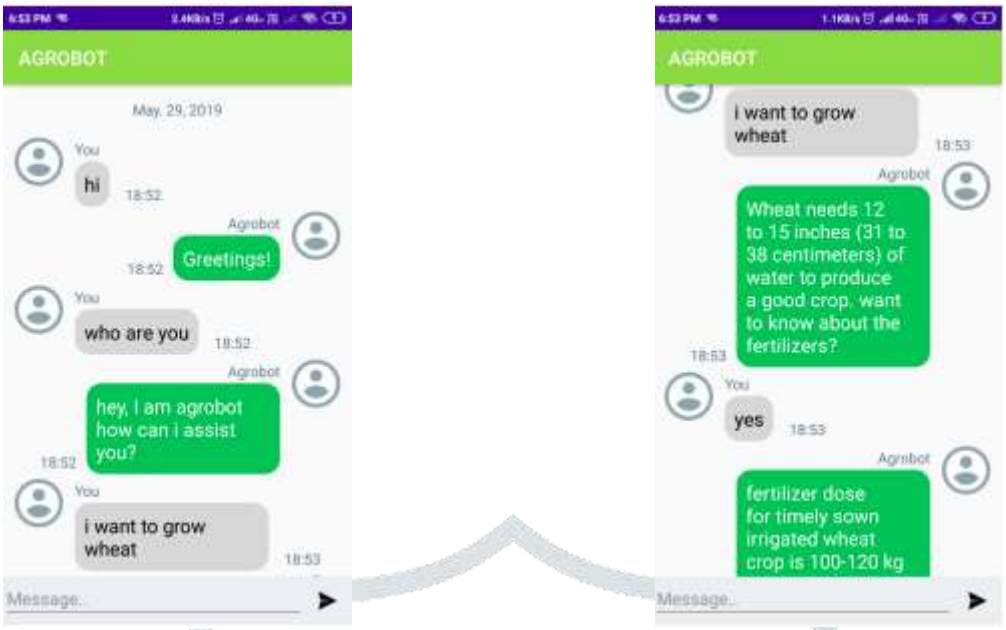
The Agrobot Client is developed in Android using Kotlin. The database used here is to store the user information like email address, mobile number, address, basic details, etc.

Following is the complete integration diagram,



Server and Client integration

VI. FINAL OUTCOME



Final Output Screen of Agrobot

Regional Language Conversion Using corpus data

VII. CONCLUSION

From our perspective, Chatbots or smart assistants with artificial intelligence are dramatically changing businesses. There is a wide range of Chatbot building platforms that are available for various enterprises, such as e-commerce, retail, banking, leisure, travel, healthcare, and so on. Chatbots can reach out to a large audience on messaging apps and be more effective than humans. They may develop into a capable information-gathering tool in the near future. As there are a wide range of mobile users in our country, this application can easily spread over the locality. It is farmer friendly application so they can use it efficiently for retrieving their precision information thus can help them to join with technology and also the growth of their business.

REFERENCES

[1] Mr. Anilkumar T. G. Mr. T. D. Ganesh. Enable smart farms in digital india with artificial intelligence and iot. Imperial Journal of Interdisciplinary Research (IJIR), 2016.

[2] Avnish Singh Jat Kunal Verma, Dinesh Pabbi. Agriculture advancement using artificial intelligence. 2nd International Conference on Recent Innovations in Science,Technology, Management and Environment, 2016.