2019 IEEE Bombay Section Signature Conference (IBSSC)

AgriBot - An intelligent interactive interface to assist farmers in agricultural activities

Divya Sawant*, Anchal Jaiswal†, Jyoti Singh‡, Payal Shah§,
Department of Electronics Engineering,Sardar Patel Institute of Technology, Andheri, Mumbai, India.
Email: *divya.sawant17497@gmail.com, †anchal.jaiswal7@gmail.com, ‡jyoti130896@gmail.com, §payal_shah@spit.ac.in

***Abstract*—In India, agriculture plays a predominant role in economy and employment. The common problem existing among the Indian farmers today is that they fail to choose the right crop based on their region specications and yield history. Hence they face a serious setback in productivity. Agricultural statistics and forecast is an important resource that the government has not explored commensurate to its impact. The paper proposes an intelligent portable system using data mining and analytics which assists farmers with various farming techniques, helps them decide most suitable crops as per current climate conditions, soil conditions and geographical characteristics of the specified region.The farmers do not have a single source which can cater to all their queries regarding seeds, fertilizers, market prices, storage facilities, government schemes,etc. To make this data analysis easily accessible to the farmers a chatbot is proposed which uses the Natural Language Processing technique. It helps to get responses of the farmer input queries regarding agricultural context in audio format, so as to make farmer interaction more user friendly. If the system fails to answer any specified query, the query is redirected to helpline centers. The system basically works as a virtual, handy assistant to assist farmers throughout the year helping them stay notified of any factor that would affect crop productivity and profit. The responses are generated based on various machine learning algorithms modelled around data set. Though the main audience under consideration are farmers any other user can also use the system to get advice regarding activities related to agriculture.**

Keywords: Machine Learning, Decision Tree, Random Forest, K- Nearest Neighbor (KNN), Chatbot, User Interface, Crop Suggestion, Web Application (app), Virtual Assistant, Kisan Call Center, Application Programming Interface (API), Dialogflo, OpenWeatherMap,Natural Language Processing Crop Statistics.

I. INTRODUCTION

Agriculture is the major provider of employment to people in many parts of the world. But, still there are millions of small and marginal farmers whose awareness levels are low as many live in remote areas. Traditionally field officers visit the fields and interact with farmers in villages and provide them training and advisory on best practices in farming and aspects of agriculture [2] [4]. In the current scenario, the government is collecting data about the rainfall, crops production but only in its raw form, and this data is of no use to the end user, that is mainly the farmers. Collecting this raw data, standardizing it, analyzing it, and feeding it to a system that will provide relational trends can be useful to the farmers. The technology in the field of agriculture is developing day-by-day. Also, a large number of software is being simultaneously developed to educate the farmers with this technological information. Most of them provide static information about farming, they require large number of searching steps to get the accurate information and they don't provide an interactive way of query and response. The proposed system overcomes the above mentioned drawbacks by providing a user interface, where farmers or any other users can interact effectively to get the desired responses with lesser number of steps [1] [2]. This system Agribot is a web application (app), which is a virtual assistant that enables farmers a very portable and handy tool which would both assist them with their farming methods as well as recommend them the best suitable crop as per given conditions so that they can decide what crops they can grow so as to get maximum yield in a better way. A web app is developed which has a dedicated User Interface for farmers of specific location. The User Interface has various features like crop recommendation, query section, weather forecasting as well as assistance to Kisan helpline centers. There can be an option of choosing local language to communicate with the chatbot in speech format would make app very feasible for use even to the farmers having very low exposure to mobile technology and chatting.

Fig. 1 depicts the entire System overview of project idea. It explains the flow of data across the system right from collection of data to compilation of data to finally recommending the best crop to the farmers by feeding data in the database and modelling it through various algorithms.

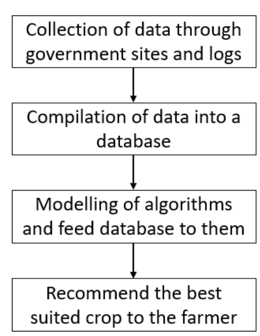

Fig. 1. System overview

Fig. 2 explains the System flow diagram of the entire project. It explains the various factors considered while training the system to suggest the best crop. Then treating the data through various machine learning algorithms for processing, the system is trained.

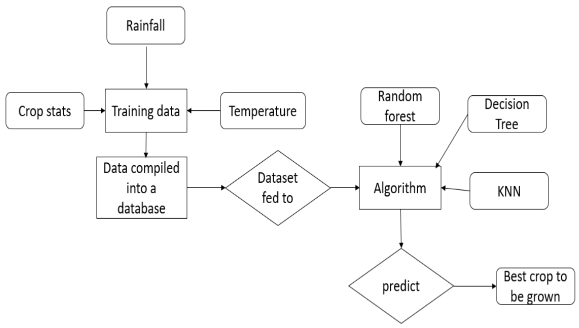

Fig. 2. System flow diagram

II. LITERATURE SURVEY

Several research studies have been undertaken in this area. They are differentiated on the basis of the method of algorithm selection or method of deployment. The following literature was referenced for our project:

- In [1], the problem among Indian farmers regarding their choice of best crop based on their soil quality is resolved. The recommendation system is developed using various classification algorithms. The system works through a GUI.
- In [2], the modelling and development of Precision farming is discussed. The main purpose of this system is to notify farmers through mails and emails directly about various advisory schemes. The model is developed so as to help farmers even with very low technological knowledge or very small scale farming.
- In [3], a comparative study of various algorithms is done so as to decide the best one for yield prediction of various crops in Precision Agriculture. All the algorithms are basically put for testing on a data set of soya crop collected across several years. The algorithms for comparison used in this paper are Random Forest, Support Vector Machine, Bayes, Bagging and Decision Tree.
- In [4], FarmBot is a chat bot, which is a virtual assistant that enable users to get their queries clarified in a user friendly manner. The input is obtained from the user. If the query is based on prediction, then the future predictions on the requested agricultural products are represented in the graphical format and displayed to the user.
- In [5], BRAC University has developed an automated system for prediction of farming. It is developed on Android Platform and suggests farmers the best crops as per their land area even before they start the cultivation process. The context of the best suggested crop totally lies over parameter of Cost effectiveness. The entire system is implemented for various regions in Bangladesh.

In [1], if the test case doesn't match with any of the keywords, by default a 'no' response output is produced.
In [2], The model is mostly designed for places like Kerala where average holding size of any farmer is much less than that of any other Indian state.
In [5], The major drawback of the entire system developed here was that it was designed on android platform and no users using windows or ios could be able to use the system.

III. DATA COLLECTION AND DATASETS

A. Parameters affecting crop yeild

1) **Rainfall** - Farming in Maharashtra is mainly rain-fed. So the crop yeild highly depends on rainfall.
2) **Temperature** - Temperature affects several aspects on crop growth. Low temperature can cause a change in the plant's ability to perform photosynthesis and water transport whereas high temperature adversely affects mineral nutrition, shoot growth and pollen development. All these factors directly or indirectly affect crop yeild.

B. Parameters determining a profitable crop

1) **Area under cultivation** - Crop yields are calculated as the harvested production per unit of harvested area. Thus the harvested area determines whether a crop will prove to be profitable to the farmer.
2) **Previous yield of crop** - Historical data is necessary to study the seasonality and trends in the crop yeilds for every season.

C. Collection of Data

1) **Crop statistics data:** Data for past year crop statistics was collected from *https://data.gov.in/catalog/district-wise-season-wise-crop-production-statistics*. It is a About Open Government Data (OGD) Platform in India. The portal is used by Government of India Ministries and their organizations to publish data sets, documents, services, tools and applications collected by them for public use. It contains data for 1997-2014 [5] [6] [10]. Fig. 3 shows the screen-shots of database of the statistics for district wise crop productivity for various regions of Maharashtra.
2) **Rainfall Data:** Data for rainfall in various regions of maharashtra has been taken from *http://maharain.gov.in/*. The web portal is a initiative of the Department of Agriculture, Maharashtra. It was launched in 1998 to record daily rainfall at circle, district and division level.It contains data for 1997-2016.

D. Preprocessing of Data

Data preprocessing is a method that is used to convert raw data into a desired and an understandable format before using it to feed the algorithm. Data which is acquired from the real world often contains missing values, errors and are

	A	B	C	D	E	F	G
1	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
2	Maharashtra	AHMEDNAGAR	1997	Autumn	Maize	1	1113
3	Maharashtra	AHMEDNAGAR	1997	Kharif	Arhar/Tur	17600	6300
4	Maharashtra	AHMEDNAGAR	1997	Kharif	Bajra	274100	152800
5	Maharashtra	AHMEDNAGAR	1997	Kharif	Gram	40800	18600
6	Maharashtra	AHMEDNAGAR	1997	Kharif	Jowar	900	1100
7	Maharashtra	AHMEDNAGAR	1997	Kharif	Maize	4400	4700
8	Maharashtra	AHMEDNAGAR	1997	Kharif	Moong(Gr	10200	900
9	Maharashtra	AHMEDNAGAR	1997	Kharif	Pulses tota	451	130
10	Maharashtra	AHMEDNAGAR	1997	Kharif	Ragi	2600	2100
11	Maharashtra	AHMEDNAGAR	1997	Kharif	Rice	5900	7200
12	Maharashtra	AHMEDNAGAR	1997	Kharif	Sugarcane	45900	38940
13	Maharashtra	AHMEDNAGAR	1997	Kharif	Total food	3384	1836
14	Maharashtra	AHMEDNAGAR	1997	Kharif	Urad	1600	800
15	Maharashtra	AHMEDNAGAR	1997	Rabi	Jowar	598400	217000
16	Maharashtra	AHMEDNAGAR	1997	Rabi	Maize	6200	9100
17	Maharashtra	AHMEDNAGAR	1997	Rabi	Other Rab	60	3
18	Maharashtra	AHMEDNAGAR	1997	Rabi	Wheat	79700	87100
19	Maharashtra	AHMEDNAGAR	1997	Summer	Maize	1100	1900
20	Maharashtra	AHMEDNAGAR	1997	Whole Yea	Cotton(lint	98	17900

Fig. 3. Past year crop statistics data set

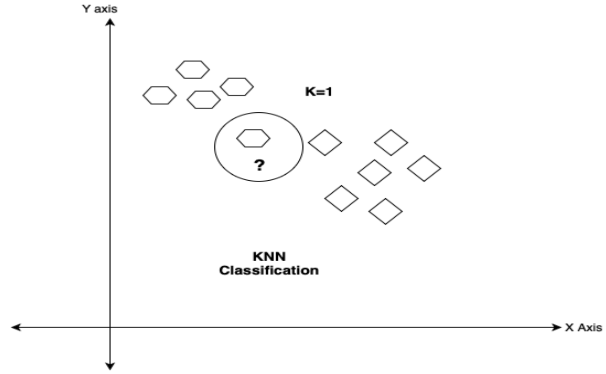

Fig. 4. KNN

inconsistent.

The steps followed in Data Preprocessing are:

1) Missing Values: The missing values in data can be processed by either replacing it with an average of the data points, frequency of the categorical values or by replacing it based on other functions.

2) Data Normalisation: When the data is comprised of attributes with varying scales, certain features having a larger range can influence the prediction more, in order to avoid this data normalisation is necessary. It is useful for algorithms like K-Nearest Neighbors in which the prediction is based on distance measures.

IV. ALGORITHM AND IMPLEMENTATION

The algorithms used for predicting the crops are as follows:

A. *K-nearest neighbor algorithm*

*1) **Working of KNN algorithm**:* K Nearest Neighbors is a most widely used algorithm in comparison to other algorithm. It stores all the available training data and uses it in the testing phase. It classifies new test data based on a similarity measure i.e. distance function.

In KNN, k means the number of nearest neighbors. As in Fig. 4, suppose the class of the new data point has to be predicted and value of k is equal to 1, then the algorithm determines the closest neighbor to the new data point and assigns it's label to that point. If the value of k is greater than 1, then k nearest neighbors are determined and then the label is assigned by the majority votes of its k neighbors.To find the closest similar points, the distance between the two points is computed by using distance measures such as Euclidean distance, Manhattan distance, Hamming distance and Minkowski distance.

*2) **Modelling**:* The accuracy of the model depends on the value of k. As the value of k decreases, the risk of overfitting increases because it captures the local points. In this case, the training accuracy will be high but the testing accuracy will be low. If the value of k is increased and becomes close to the number of data points then the model suffers from underfitting as everything will be predicted as one class which is the majority class in the dataset.

B. *Decision Tree*

*1) **Working of a Decision Tree**:* Decision tree is a supervised machine learning algorithm that is mostly used for classification problems. It can work for categorical as well as continuous input variables. In this, the samples is split into homogeneous sets based on a significant splitter. As our target variable is a categorial value i.e the crop to be predicted we use a Categorical variable Decision tree [14][15].

*2) **Splitting strategy**:* The strategy which is used for splitting is Entropy. It is the degree of disorganisation in a system. If the entropy of the sample is zero then it means it is completely homogeneous and if the sample is equally divided then it has a entropy of one. Entropy can be calculated using formula:-

$$Entropy = -plog_2p - qlog_2q$$

Here p and q is probability of success and failure of the node. The split which has lowest entropy is chosen and accordingly the decision tree is modelled. The lesser the entropy, the better is the performance of the model [15].

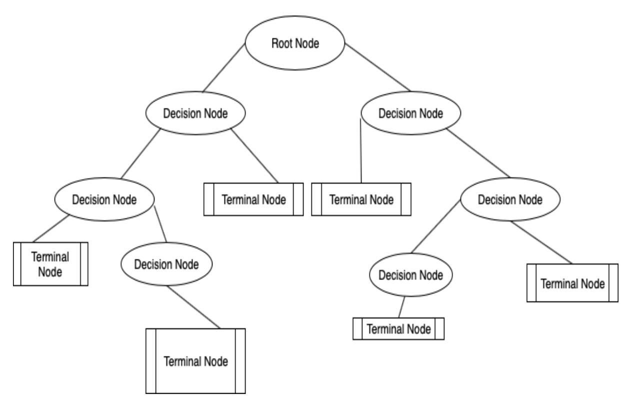

Fig. 5. Decision Tree terminologies

In Fig. 5, the root node represents the entire samples in the dataset which gets further divided into homogeneous sets. The algorithm identifies the best significant variable and splits the samples into further sub-nodes. When sub-nodes splits into further sub-nodes, then it is called decision nodes. Terminal nodes are the nodes which cannot split [14] [15].

3) Modelling: Decision tree often suffers from overfitting. If there is no limit on the decision tree modelling parameters then 100% training accuracy can be achieved. This is because it creates one leaf for every observation. But, in such modelling the testing accuarcy will be low. Thus, it is necessary to prevent overfitting by the following ways:

1) Setting constraints on tree size
2) Tree pruning

C. Random Forest Algorithm

*1) **Working of Random Forest**:* Random forest is also a supervised machine learning algorithm in which multiple trees are grown unlike the decision tree. In this case, each tree will output a single class and then the forest will select the class having the majority vote [16].

It works in the following manner. Each tree is planted and grown as follows:

1) If the training set contains N cases. Then, the sample of N cases is randomly selected with replacement. This will be training set for that tree [16].
2) Suppose there are M input variables, then m variables are selected randomly out of M such that m is less than M. The best split is considered to split the node [16].
3) Each tree is grown to the largest extent possible without pruning.
4) Predict new test data by aggregating or combining the predictions of the n tree trees [16]. In case of classification, it is the majority vote while average is taken in case of regression as shown in Fig. 6.

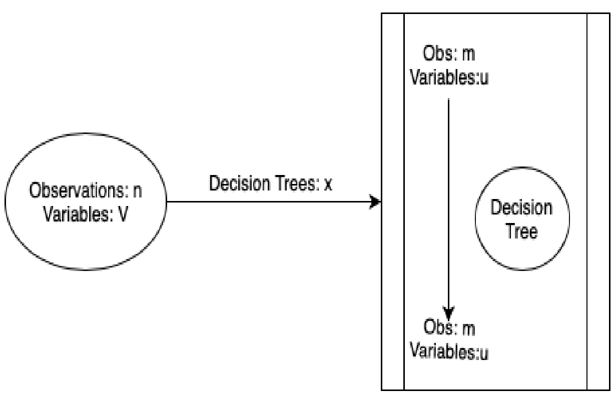

Fig. 6. Random Forest explained

V. RESULT ANALYSIS

The results of the algorithms have been compared with each other. The training and testing accuracy obtained for the algorithms are summarized in Table 1. Analysis of the algorithms and their outputs further emphasize the following points.

1) KNN -
   - Label encoding required for categorial values
   - Need to determine value of K
   - Effective if training data is large
   - High computational cost due to distance computation for every test data
2) Decision Tree -
   - Graphical representation is very intuitive and easily relatable
   - It can handle both numerical and categorical variable and less data cleaning required. It can handle missing values.
   - Can cause overfitting if no constraints are set on parameters
3) Random Forest Algorithm -
   - Can handle large data set with higher dimensionality
   - estimates missing data and maintains accuracy
   - Resembles a black box approach for statistical modelers (less control on parameters)

TABLE I
COMPARISON OF ALGORITHMS

Algorithm	Training Accuracy	Testing Accuracy
KNN	70%	72.5%
Decision Tree	99%	75%
Random Forest	99%	78%

VI. USER INTERFACE

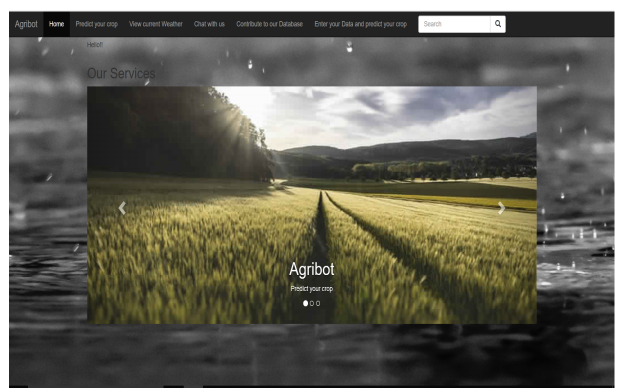

Fig. 7. Web interface home page

The user interface consists of 5 elements:

1) A web page showing the recommended crop for the particular region for given conditions.
2) A web page to give current weather details and weather forecast.
3) A web page to encase the interactive chatbot which responds to user queries regarding godowns and warehouses for storage of crops and also provides information regarding the training centers in the region. The chatbot uses DialogFlow API . The chatbot can also accept voice input. There are various keywords stored

in API which would redirect the query to intents and responses are then produced to farmers in the form of text or voice [11] [12]. If the farmer enters any query which does not contain any stored keyword, the query is then directed directly to Kisaan Call Center.

4) A webpage to allow the farmer to contribute to our database since the current dataset does not contain soil data as the Soil healthcard scheme has been introduced in 2017. The farmer can enter data about soil parameters as soon as he receives the soil healthcard. This feature will help to expand the database and use this data for future training of the model.
5) A webpage to allow farmer to enter his plot parameters and obtain the most suitable crop for his plot.

Fig. 7 depicts the basic GUI of entire system. There are multiple tabs present over the GUI to facilitate the user to choose available options as per his requirements.
Fig. 8 shows the output of the system to suggest the best voted crop. Here the best voted crop by the model is Groundnut.
Fig. 9 depicts the climatic information aspect of web app, wherein the user can also get information about current as well as future climatic aspects. This is done using the OpenweatherMap API [13].
Fig. 10 depicts the working of DialogFlo wherein an interactive Chatbox is opened for user which both takes input via speech or text and thus responds as per the intent keywords in the queries. The queries can be of anything related to farming or agricultural assistance.

Fig. 8. Displaying recommended crop

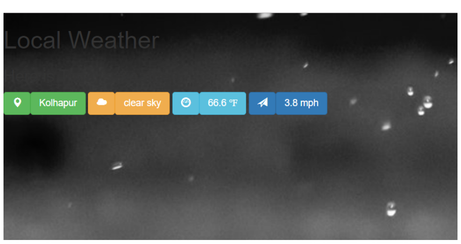

Fig. 9. Displaying climatic conditions

Fig. 11 shows how the multipage form is displayed which allows the user/farmer to contribute to the database. The current database lacks information about the soil parameters and hence this form of data would prove useful for further training of the model. The form has the following data fields :

Fig. 10. Chatbot interface for query response

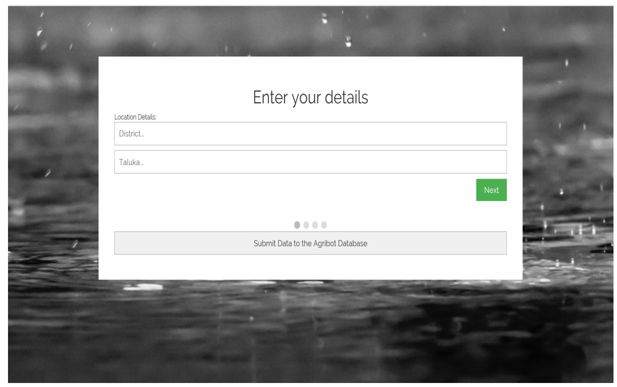

Fig. 11. Multi page form for contributing to current database

1) Page 1 -Location details comprising of the name of the district and taluka
2) Page 2 -Term details comprising of the year and cropping season (Kharif, rabi or summer)
3) Page 3 -Soil details comprising of Soil type (red, alluvial, black, laterite,etc.), Soil pH and the Nitrogen, Phosphorous and Potassium content in the soil (in %)
4) Page 4 -Climate details comprising of rainfall (in mm) and temperature (in degrees) fields
5) Page 5 - Farm plot and yeild details comprising of feilds for area of the plot (in sq.m), name of the crop planted and the yeild obtained (in tonnes).

The data entered into the form is saved in a database as shown in Fig. 12.
Fig. 13 shows the form fields which allow the user/farmer to get the most suitable crop for his plot of land. After entering the details the most suitable crop for his land is displayed as shown in Fig. 14.

VII. CHALLENGES AND LIMITATIONS

1) The data for soil and temperature is very limited and not available freely so these parameters have been dropped from the database used for training.
2) APIs which are used for interaction with the farmer do not provide services in the local languages for free.

District	Taluka	Year	Season	Soil Type	Soil pH	Nitrogen	Phosphorus	Potassium	Rainfall	Temperature	Area	Crop	Yield
aurangabad	satara	2019	kharif	red	2	12	13	14	234	23	1234	rice	1234
ratnagiri	dapoli	2019	kharif	red	4	23	43	23	140	31	400	rice	200
pune	kharadi	2019	kharif	black	3	32	12	45	300	24	400	cotton	211
nashik	igatpuri	2019	Summer	alluvial	6	12	12	20	230	30	2135	bajra	789
pune	hinjewadi	2019	kharif	red	12	23	23	42	200	32	300	sugarcane	2000
pune	bhed	2018	Rabi	red	8	12	15	20	340	29	5641	Jowar	345
Kolhapur	karvir	2018	Summer	black	9	13	14	17	234	34	2135	rice	679
raigad	murud	2019	monsoon	black	5	23	21	40	200	30	300	rice	200

Fig. 12. Information entered in the form stored in a database

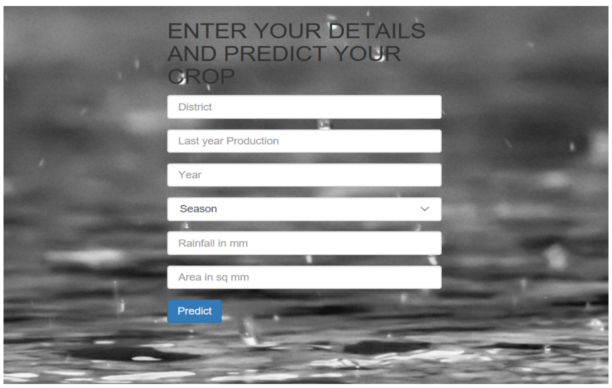

Fig. 13. Form for obtaining the most suitable crop for specific plot

3) There was no single database available which enlisted all geographical, market value and climatic parameters of an agricultural region. So the entire database had to be designed accordingly. Thus, data cleaning and organizing was one of major challenge related to this project.

VIII. CONCLUSIONS

The entire work was primarily based on objective of providing farmers with a handy and portable virtual interactive farming assistant that can communicate with farmers. The app was designed to be extremely farmer specific. Agribot, not only helps farmers get best crop recommendations but it also helps them better nurture their crops, store their crops so as to increase shelf life of crops, as well as help them find best rates as which they can trade their crops in nearby markets.

IX. FUTURE SCOPE

The system can be scaled to be implemented and deployed for usage of farmers of any region in India. The Chatbot API can be used with the local language feature which is a paid service and hence not implemented in the chatbot currently. The app can also be made more personalized for farmers by considering various features like monetary budget of the concerned farmer or the agricultural resources available to him for farming. The concept of Crop Rotation if considered while designing would give better interpretation of suggestion of crops to farmers.

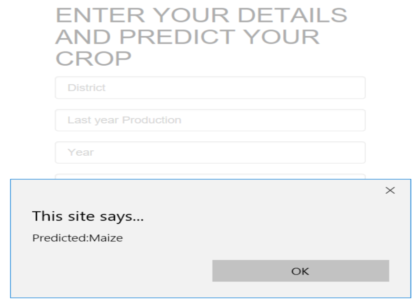

Fig. 14. Displaying the most suitable crop for specific plot

REFERENCES

[1] S.Pudumalar, E.Ramanujam, R.Harine Rajashreen, C.Kavyan, T.Kiruthikan, J.Nishan Thiagarajar College of Engineering, "Crop Recommendation System for Precision Agriculture", *IEEE Eighth International Conference on Advanced Computing, March 2016*

[2] Satish Babu, "A Software Model for Precision Agriculture for Small and Marginal Farmers" , *The International Centre for Free and Open Source Software (ICFOSS) Trivandrum, India , May 2013*

[3] Anshal Savla, Parul Dhawan, Himtanaya Bhadada, Nivedita Israni, Alisha Mandholia , Sanya Bhardwaj ,"Survey of classification algorithms for formulating yield prediction accuracy in precision agriculture" , *Innovations in Information,Embedded and Communication systems (ICIIECS), May 2015*

[4] Mrs. L. Kannagi, Ramya .C, Shreya .R, Sowmiya R Valliammai Engineering College, "Virtual Conversational Assistant The FARMBOT" *,IJETSR, ISSN 2394, Volume 5, Issue 3, March 2018.*

[5] Talha Siddique, Dipro Barua, Zannatul Ferdous, Amitabha Chakrabarty, BRAC University , "Automated Farming Prediction", *, Intelligent Systems Conference 2017 ,7-8 September 2017 London, UK*

[6] Rakesh Kumar, M.P. Singh, Prabhat Kumar and J.P. Singh, "Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique" *, International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM),2015*

[7] Monali Paul, Santosh K. Vishwakarma, Ashok Verma, "Analysis of Soil Behaviour and Prediction of Crop Yield using Data Mining Approach" *, International Conference on Computational Intelligence and Communication Networks, India , May 2015*

[8] V. Kumar, V. Dave, R. Bhadauriya, S. Chaudhary, "Krishimantra:"agricultural recommendation system" *, Proc. of the 3rd ACM Symposium on computing for development, no. 45, 2013.*

[9] J. Ranjan, "Data Mining Techniques for better decisions in Human Resource Management Systems" *, International Journal of Business Information Systems IJBIS, vol. 3, pp. 464-481, 2008.*

[10] Paul Monali, Santhosh K. Vishwakarma, Ashok Verma, "Prediction of crop yield using Data Mining approach" , *Computational Intelligence and Communication Networks (CICN) International Conference, 12-14 Dec. 2015.*

X. WEB REFERENCES

[11] https://codeburst.io/2-how-assistant-work-introduction-to-dialogflow

[12] https://dialogflow.com/docs/tutorial-build-an-agent

[13] https://openweathermap.org/

[14] https://www.analyticsvidhya.com/blog/2016/04/complete-tutorial-tree-based-modeling-scratch-in-python/

[15] https://acadgild.com/blog/decision-tree-python-code

[16] https://www.analyticsvidhya.com/blog/2015/06/tuning-random-forest-model/

[17] https://machinelearningmastery.com/k-nearest-neighbors-for-machine-learning/

[18] https://data.gov.in/catalog/district-wise-season-wise-crop-production-statistics

[19] http://maharain.gov.in/