ABSTRACT:

Farmers still favor traditional farming in today's era, but digital farming is a better option now. As technology advances, farmers need to improve their methods to produce the maximum yield with time. As digital farming becomes more popular, people are shifting towards it and maximizing their yields. DigiFarm is made with the sole purpose of helping farmers in various fields. Using DigiFarm farmers can predict which crops they can sown on their land. Climate change has become one of the most pressing concerns in modern times [5]. The world has indeed experienced some drastic changes in the recent past, from glaciers melting to erratic rainfall.  A number of challenges have been created for agriculture as a result of climate change, and they will continue to grow. In addition to affecting crop growth rate, climate change also affects rainfall variability. Nitrous oxide emissions from soils [7], fertilisers, manure, and many other sources are a major reason why agriculture is affected. They all play a role in raising global temperatures. With the introduction of green revolution in order to increase the yield farmers started using High Yielding Variety (HYV) seeds which in a long run affects soil quality and causes many other problems. But now every farmer is realizing the damage that HYV seeds have made and they are slowly choosing the path of organic farming. The website is designed to equip the farmers with digitized farming so that they can make the most out of their crops. With the aid of this website, they can receive precise information about which crops would be most suitable for their land.DigiFarm, with the aid of technology will help them with the best possible information for their farm using a single platform. DigiFarm would result in digitization of agriculture on a global scale. However, it would prove extremely beneficial to farmers in India who find it difficult to access the information as their outreach is confined which results in using obsolete methods despite the advancements.

1. INTRODUCTION

Farmers, the soul of the Indian economy, are the ones who feed the huge population of India. Without them, no one would be getting food on their plate today. DigiFarm is primarily aimed at farmers that can assist with their needs. The motive is to make the farmers aware of digital farming and make them aware about it and how is it better than traditional farming and using DigiFarm farmers can predict which crops they can sow on their land based on certain factors like their soil condition (pH of the soil, type of the soil etc.,), weather conditions to name a few. The prediction is based on the machine learning model which is trained with adequate amounts of statistical data so that the prediction is accurate and is ultimately helpful for the farmers to increase their income.

Climate change has become one of the most pressing concerns in modern times. The world has indeed experienced some drastic changes in the recent past, from glaciers melting to erratic rainfall. A number of challenges have been created for agriculture as a result of climate change, and they will continue to grow. In addition to affecting crop growth rate, climate change also affects rainfall variability. They all play a role in raising global temperatures. Traditional farming makes use of harmful chemicals and fertilisers due to which the health of the crop is compromised. In order to reduce this, DigiFarm encourages farmers to shift to organic farming as they are beneficial for the environment as it does not make use of harmful chemicals and helps in the preservation of the environment. Organic farming is also said to have only one-third the volume of water movement into the groundwater and surface when compared to traditional framing. DigiFarm also helps farmers in choosing the best crop which is suited for their land on the basis of various factors like weather conditions, soil composition, location and other factors. Through this, the maximum yield can be achieved.

The main aim of DigiFarm is to help the farmers in growing their business by growing the right crop at the right time and making them aware about Digital Farming. Due to the global accessibility of DigiFarm, even a farmer in some remote village can access it and make profit thereby. All their queries will be answered by the chatbot present in the website.

The website DigiFarm which is designed carefully and accurately to predict the most suitable crop that the farmer can produce in his region. The website is designed to equip the farmers with digitized farming so that they can make the most out of their crops. With the aid of this website, they can receive precise information about which crops would be most suitable for their land.

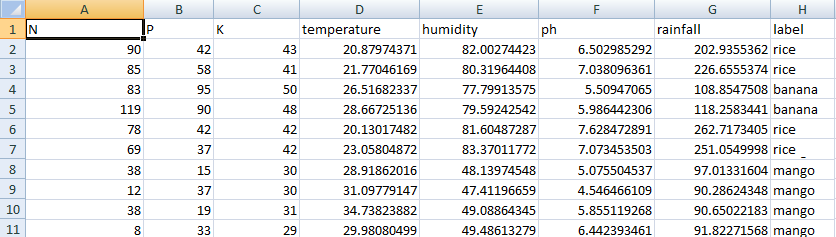
This paper is organized as follows: Section 2 presents the Methodology used to develop the website, collect the data and use different algorithms. Section 3 presents the result and description of the website. Section 4 summarizes the conclusion to our study derived from the website. Section 5 presents the future scope of the whole website and how it will benefit the mass. Finally, Section 6 presents the references used in the paper.

**2. METHODOLOGY**

Data is a very important part of a Machine Learning Model. Hence, DigiFarm is a user-friendly website designed in such a way that anyone can use it to predict the best crop that can be grown on their soil. To predict the crop we have used Machine Learning(ML) and Artificial Intelligence(AI) technologies. The prediction model is the result of testing the dataset with several algorithms and methodologies to determine which are best and highest accuracy rate.

Following are the steps that we have followed to create the crop prediction model:

1. Data collection
2. Data preparation and analysis
3. Choosing the algorithm for training the dataset
4. Testing the Machine Learning (ML) model and evaluation
5. Deploying
6. *Data collection*

To train this prediction model we are using the dataset which we have procured from Kaggle Website (https://www.kaggle.com/siddharthss/crop-recommendation-dataset) which is shown in Figure 1. This dataset was build by augmenting datasets of rainfall, climate and fertilizer data available for India. Gathered over the period by ICFA, India.

*Figure 2: Dataset used to train Machine Learning Model*

1. *Data preparation and analysis*

The Dataset which is shown in Figure 1 has 2201 samples among which we have used 90% (i.e. 1980 samples) for the purpose of training. The remaining 10% (i.e. 221 samples) are used for testing purpose. The dataset contains 8 attributes (they are Nitrogen, Phosphorous, Potassium, Temperature and humidity of the region, pH of the soil, Rainfall in mm in that region and the crop).The dataset contains 22 distinct categories (i.e., Apple, Banana, Blackgram, Chickpea, Coffee, Cotton, Grapes, Jute, Kidneybeans, Lentil, Maize, Mango, Mothbeans, Mungbean, Muskmelon, Orange, Papaya, Pigeonpeas, Pomogranate, Watermelon, Rice, Coconut).

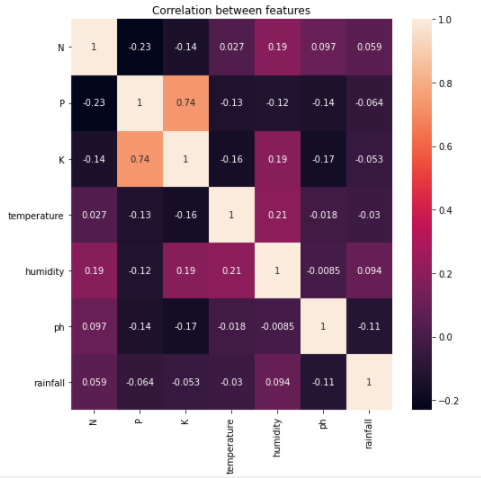
Before training our dataset we conducted following preliminary analysis of the dataset:

1. Finding out some statistical information about the data which is summarized in the Table 1

|  | *N* | *P* | *K* | *Temperature* | *Humidity* | *ph* | *Rainfall* |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *Count* | 2200.00 | 2200.00 | 2200.00 | 2200.00 | 2200.00 | 2200.00 | 2200.00 |
| *Mean* | 50.55 | 53.36 | 48.14 | 25.61 | 71.48 | 6.46 | 103.46 |
| *Standard Deviation* | 36.91 | 32.98 | 50.64 | 5.063 | 22.26 | 0.77 | 54.95 |
| *Minimum value* | 0.00 | 5.00 | 5.00 | 8.82 | 14.25 | 3.50 | 20.21 |
| *25%* | 21.00 | 28.00 | 20.00 | 22.76 | 60.26 | 5.97 | 64.55 |
| *50%* | 37.00 | 51.00 | 32.00 | 25.59 | 80.47 | 6.42 | 94.86 |
| *75%* | 84.25 | 68.00 | 49.00 | 28.56 | 89.94 | 6.92 | 124.26 |
| *Maximum Value* | 140.00 | 145.00 | 205.00 | 43.67 | 99.98 | 9.93 | 298.56 |

*Table 1: Basic statistical information about the dataset*

1. *We found out the correlation between different attributes which is summarized in the following Table 2.*



*Table 2: Correlation between features*

1. Since there were no missing cells or null values we then moved to the next stepi.e choosing the algorithm and training it.
2. *Choosing the algorithm for training the dataset*

In this research work, we have considered “crop name” as target label and used “Multiclass Classification”. We have used Light Gradient Boosting Machine (LGBM) Classifier and Extra tree classifier algorithms for training the ML model.

*Algorithms*

*Light Gradient Boosting Machine (LGBM)*

*LGBM*, short for Light Gradient Boosting Machine, is a fast, high performance gradient boosting framework based on decision tree algorithm used for machine learning tasks which was originally developed by Microsoft in 2016. LGBM does not grow a tree level-wise as other implementations do. Instead of that it grows trees leaf-wise. It chooses the leaf which it believes will yield the largest decrease in loss. Because this grows leaf wise this can lead to over fitting (Write overfitting definition here) which can be minimized by defining the depth for splitting.

*Extra tree classifier*

It is a type of ensemble learning model. They are much more faster than random forest. Let us consider a simple dataset with M features (i.e. columns) and N rows. In random forest we grow multiple trees such that each tree comprises of the square root of the total number of features that are present. In our case we have M features so each tree would have square root of M features to train on. Additionally we will make use of samples with replacement that is where we have the out of the bag error score (Out-of-bag (OOB) error, also called out-of-bag estimate, is a method of measuring the prediction error of random forests, boosted decision trees, and other machine learning models utilizing bootstrap aggregating (bagging) [2]) that we can later use for validation. In Extra tree we will be selecting samples for every decision tree without replacement so all the samples would be unique in nature. The total number of features selected still remains the same, that is the square root of the total number of features if you are using a classification task. Main difference lies in the fact that instead of computing the locally optimal split for a feature combination a random value is selected for the split for the extra tree. So rather than not spending the time in finding out the best splitting point we randomly pick up a point and split based on that this leads to more diversified trees and less splitters to evaluate when training an extremely random forest.

1. *Testing the Machine Learning (ML) model and evaluation*

Once training the dataset with LGBM Classifier and Extra tree classifier we test both the models with 10% samples of the dataset. On testing we found that the accuracy that we got from LBGM Classifier algorithm is 0.991 and that of Extra tree Classifier was 0.985. Since the accuracy of LGBM Classifier was high we chose that model to deploy and use it for predicting the crop.

1. *Deploying*

*For deploying this*  ML model we have used IBM cloud services. Since it is deployed in a cloud, we are using the ML Model for prediction through Application Program Interface (API).

**3. RESULT AND DISCUSSION**

In this section we shall see the outcome of our platform “DigiFarm” which is designed to carefully and accurately predict the most suitable crop that the farmer can produce in his region. DigiFarmis designed to equip the farmers with digitized farming so that they can make the most out of their crops. With the aid of this platform, they can receive precise information about which crops would be most suitable for their land. User can predict their crop based on two methods:

1. Making use of region's weather conditions, pH value of the soil, rainfall pattern and soil composition (i.e., nitrogen, phosphorus and potassium)
2. Making use of place/location and current season

*3.1 Software Compatibility*

Our platform DigiFarm is compatible with latest versions of browsers such as Google Chrome, Microsoft Edge, Mozilla Firefox etc. The Front-end of DigiFarm is designed using HTML 5, CSS 3 and JavaScript whereas for the Back-end we have used Django 3.2.5 (i.e., a Python back-end framework).

*3.2 Home Page*

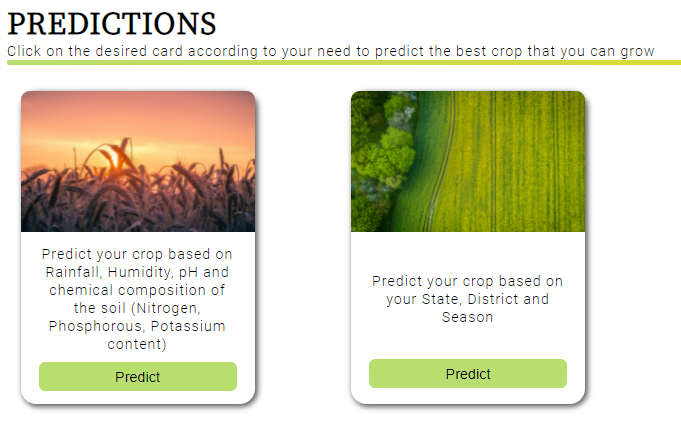
The Homepage of DigiFarm, which consists of all the major icons and the logo of DigiFarm along with its name is shown in the Figure 1. The logo displayed on the top left corner has two central elements: a hand and the water. The water is representative of rivers and oceans that forms the backbone of the irrigation system in Indian agriculture. The hand that is holding a plant represents the farmers who grow the crops. Since India is an agrarian economy, not only is the population dependent on the farmers for food, but the national economy is also dependent on the yield from the primary sector. 

This page connects the user to all the different pages on the website. The Navigation bar on the homepage as shown in Figure 1 has different buttons for various purposes such as ChatBot, Prediction, Weather and News. Additionally, there are two icons: one to get the detailed tutorial on how to use the website for farmers and another for predicting the most suited crop(s).

Now, let us see what each section of the platform does in detail:

*3.3 Prediction Section*

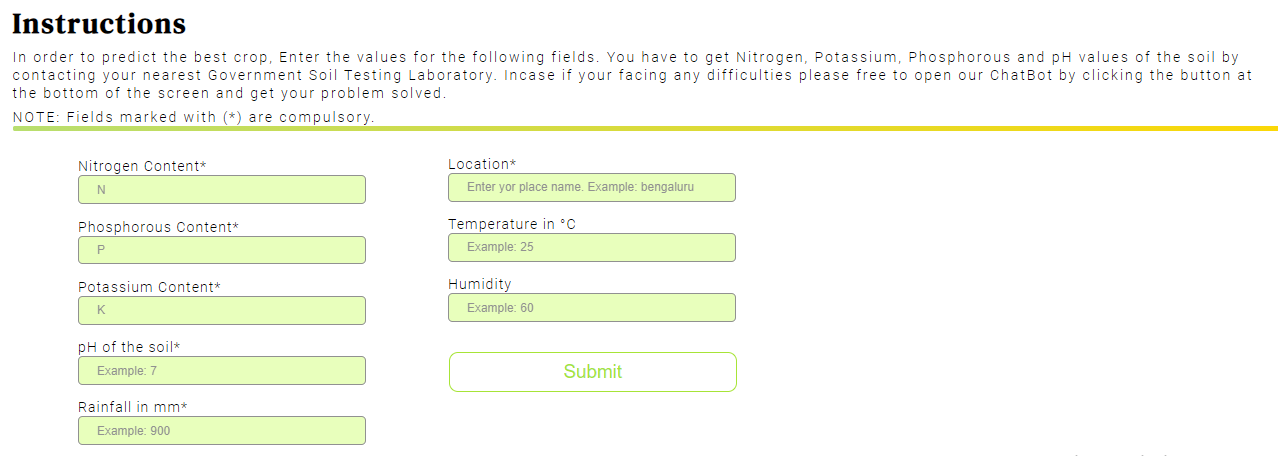
This part of DigiFarm is the main part of the platform. The “Prediction” section as shown in the Figure 2 can be used to predict the most suitable crops by using two different methods.



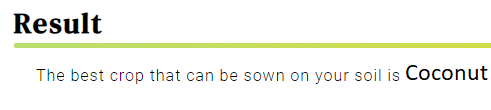
*Figure 2: Predictions Section (i.e. Different methods for predicting the crop)*

*3.3.1 Method 1*

Users can predict the crop by entering details such as rainfall, humidity, and temperature, and pH value along with the composition of the soil (i.e. nitrogen, phosphorus, and potassium content), the system will predict the crops most suitable for the given geographical conditions as shown in the Figure 3(a).

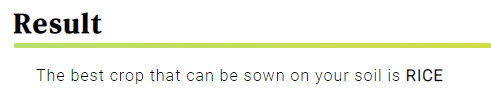
*Figure 3(a): Prediction of crops using method* 

Consider an instance, when the user enters the values of nitrogen content as 20, phosphorus content as 89, potassium content as 40, pH value as 6, rainfall as 700, location as Bengaluru (i.e. as shown in Figure 3(a)), the Machine Learning (ML) model will predict the crop based on the values entered and the result is displayed as Coconut as shown in the Figure 3(b).



*Figure 3(b): The output based on method 1*

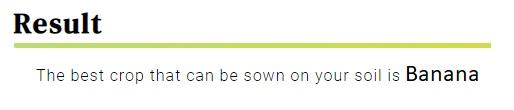
Similarly, we can get different crop names as output (as shown in the figures 3(c) to 3(f)) based on the different input combinations.



*Figure 3(c): For the input values N=89, P=58,*

*K=38, Temp=23°C, Humidity=83%, pH=6.3*

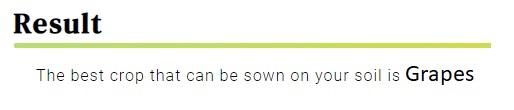
*And Rainfall=221mm.*

**

*Figure 3(d): For the input values N=86, P=76,*

*K=54, Temp=29°C, Humidity=80%, pH=5.9*

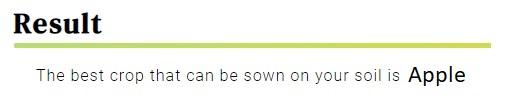
*And Rainfall=90mm.*



*Figure 3(e): For the input values N=36, P=125,*

*K=196, Temp=37°C, Humidity=80%, pH=6.1*

*And Rainfall=66mm.*



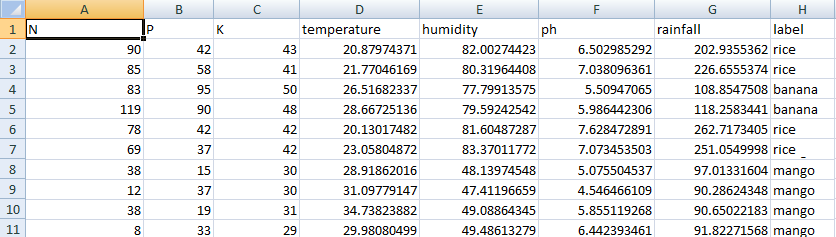
*Figure 3(f): For the input values N=34, P=140,*

*K=198, Temp=21°C, Humidity=93%, pH=5.75*

*And Rainfall=115mm.*

*3.2.1.1 Discussion on prediction of crop (method 1) dataset:*

The Dataset as shown in the Figure 4 below, has been used for training and testing our crop prediction model. In total, we have 2201 samples among which we have used 90% (i.e. 1980 samples) for the purpose of training. The remaining 10% (i.e. 221 samples) are used for testing purpose. The dataset contains 8 attributes (they are Nitrogen, Phosphorous, Potassium, Temperature and humidity of the region, pH of the soil, Rainfall in mm in that region and the crop).

*Figure 4: Crop prediction dataset for method 1*

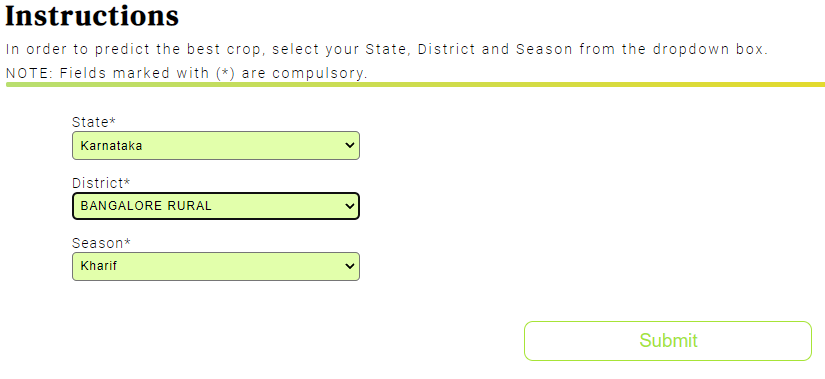
For training and deploying the ML model, the IBM Cloud Services is used which includes ML and AutoAI services. Since it is deployed in a cloud, we are using the ML Model through Application Program Interface (API).

In this research work, we have considered “crop name” as target label and used “Multiclass Classification”. The dataset contains 22 distinct categories (i.e., Apple, Banana, Blackgram, Chickpea, Coffee, Cotton, Grapes, Jute, Kidneybeans, Lentil, Maize, Mango, Mothbeans, Mungbean, Muskmelon, Orange, Papaya, Pigeonpeas, Pomogranate, Watermelon, Rice, Coconut). We have used Light Gradient Boosting Machine (LGBM) Classifier and Extra tree classifier. Accuracy is the ratio of total number of correctly predicted data points to the total number of all the data points. The accuracy that we got from LBGM Classifier algorithm on testing was found to be 0.991 and that of Extra tree Classifier was 0.985. Since the accuracy of LGBM Classifier was high we chose that model to deploy and use it for predicting the crop.

This system of crop prediction results in accuracy and efficiency which is unprecedented. Farmers generally wish to continue growing the same crops on their land either to avoid risk or due to lack of awareness. However, this system would enable them to go beyond their regular pattern by providing them with accurate scientific and data based information.

*3.3.2 Method 2*

Another option for predicting the crops is by mentioning their state, district, and season as shown in Figure 5(a).



*Figure 5(a): Input fields for approach 2*

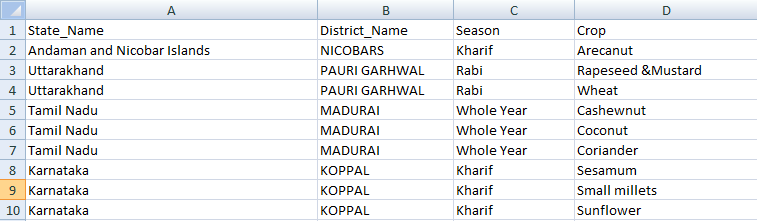
For example:- When we enter the state as Karnataka, district as Bengaluru Rural, season as kharif (ref Figure 5(a)), we get the crop prediction of ONION, DRY GINGER, RAGI, BAJRA, MAIZE, RICE, GRAM and many more as shown in Figure 5(b).



*Figure 5(b): Sample result from the second approach to Crop Prediction*

*3.3.2.1 Discussion on prediction of crop (method 2) dataset:*

For this 2nd method we have made use of a dataset which is different from the first one. This dataset contains the name of the crops which gave the best yield in each seasons (i.e. Autumn, Kharif, Rabi, Summer, Winter). This crop and season data is available for all the districts of India. This dataset has 4 columns (as shown in the Figure 6) namely the State Name, District Name, Cropping Season and crop names.

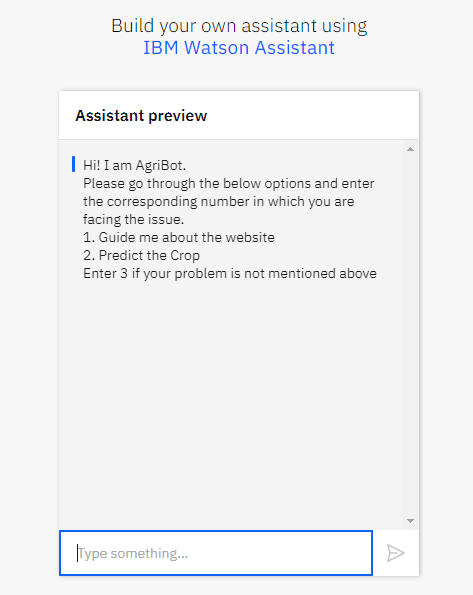
For predicting the crops in this method we are not using any ML model instead we are searching the dataset and displaying the crop names according to the information (i.e. State name, District name and the preferred season) entered by the user.

*Figure 6: Crop prediction dataset for method 2*

*3.4AgriBot*

In the current growing age of digitization, Artificial Intelligence (AI) powered chatbots are playing a leading role by exemplifying the function of a virtual assistant that could manage a conversation via speech or textual methods. It makes use of voice or textual queries to get answers, perform actions and recommendations according to user needs. They are adaptable to the user's individual language usages, searches, and preferences with continuing use. With the advent of AI, virtual assistants can be seen penetrating to the nook and corner of the world [1]. A conversational bot with a voice and/or chat interface can play a principal role in solving the user queries by giving instant service. This drastically reduces waiting time for the users (i.e. mainly farmers) to get their queries solved. The ChatBot as shown in Figure 4 is an AI based bot called the “AgriBot”. It is created using the IBM Watson Chatbot services (i.e. a part of IBM Cloud services).

The result of the AgriBotis the easy accessibility of DigiFarmplatform for the users. The AgriBot provides user three options (i.e. as shown in the Figure 7) which can be availed one at a time. This helps the user to get their queries solved.



The different options available for users are:

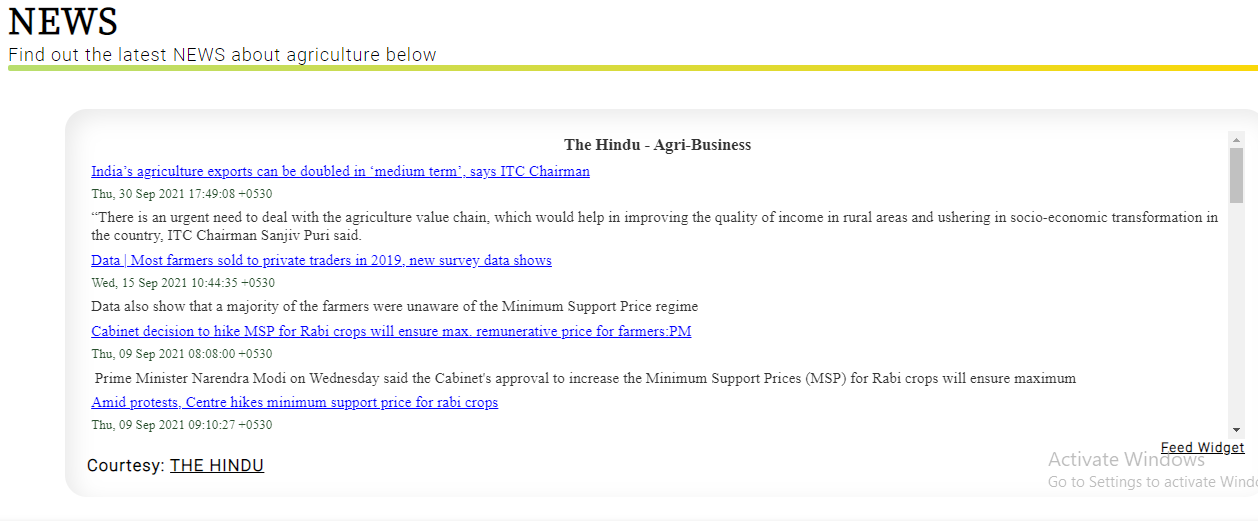
1. To guide the users about the platform and how to make use of it.
2. Various methods to predict the crop
3. Directs users to a Google Form which they can fill for any further queries

The presence of a AgriBot on this platform simplifies the user experience which is the broader aim of the Digifarm.

*Figure 7: AgriBot*

*3.6 News Segment*

The news segment results in bringing together the latest information about agriculture from across the globe as shown in Figure 7. The idea behind this is to keep our users about the technological developments happening in the agricultural field from different parts of the world. This enables them to learn from these techniques and apply the suitable ones on their land. The result would be increased productivity, developing a nature of taking risks and also equip them with the most updated advancements which they may find suitable for their farm.

In the news section of DigiFarm, we have used RSS (Really Simple Syndication) news feed from “The Hindu Agri-Business” news. 

*Figure 7: News section*

The idea is to build a platform which solves all the queries of the farmers holistically and comprehensively. The main aim of DigiFarm is to help the farmers in increasing their production and yield per square by choosing the right crop for their field at the right time. Also, it enables them to sell their crop for maximum revenue. DigiFarm would result in digitization of agriculture on a global scale. However, it would prove extremely beneficial to farmers in India who find it difficult to access the information as their outreach is confined which results in using obsolete methods despite the advancements. DigiFarm, with the aid of technology will help them with the best possible information for their farm using a single platform.

1. References

1) U. Bharti, D. Bajaj, H. Batra, S. Lalit, S. Lalit and A. Gangwani, "Medbot: Conversational Artificial Intelligence Powered Chatbot for Delivering Tele-Health after COVID-19," 2020 5th International Conference on Communication and Electronics Systems (ICCES), 2020, pp. 870-875, doi: 10.1109/ICCES48766.2020.9137944.

2) <https://en.wikipedia.org/wiki/Out-of-bag_error#:~:text=Out%2Dof%2Dbag%20(OOB,utilizing%20bootstrap%20aggregating%20(bagging>)

3. Clapp, Jennifer, and Sarah-Louise Ruder. "Precision technologies for agriculture: digital farming, gene-edited crops, and the politics of sustainability." *Global Environmental Politics* 20.3 (2020): 49-69.

4. Balakrishnan, Narayanan, and Govindarajan Muthukumarasamy. "Crop production-ensemble machine learning model for prediction." *International Journal of Computer Science and Software Engineering* 5.7 (2016): 148.

1. Van Klompenburg, Thomas, Ayalew Kassahun, and Cagatay Catal. "Crop yield prediction using machine learning: A systematic literature review." *Computers and Electronics in Agriculture* 177 (2020): 105709.
2. Palanivel, Kodimalar, and Chellammal Surianarayanan. "An approach for prediction of crop yield using machine learning and big data techniques." *International Journal of Computer Engineering and Technology* 10.3 (2019): 110-118.
3. Change, Climate. "Climate change." *Synthesis Report* (2001).
4. Wagner-Riddle, C., Thurtell, G. W., Kidd, G. K., Beauchamp, E. G., & Sweetman, R. (1997). Estimates of nitrous oxide emissions from agricultural fields over 28 months. *Canadian Journal of Soil Science*, *77*(2), 135-144.
5. Peigné, J., et al. "Is conservation tillage suitable for organic farming? A review." *Soil use and management* 23.2 (2007): 129-144.
6. Eli-Chukwu, N. C. (2019). Applications of artificial intelligence in agriculture: A review. *Engineering, Technology & Applied Science Research*, *9*(4), 4377-4383.
7. Kumar, Aayush, Omen Rajendra Pooniwala, and Swapneel Chakraborty. *Intelligent Crop Recommendation System Using ML*. Diss. CMR Institute of Technology. Bangalore, 2020.