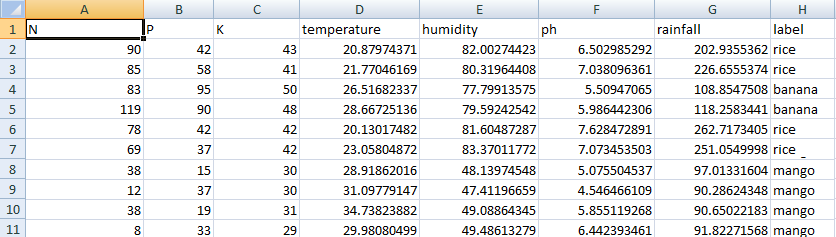
**2. METHODOLOGY**

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.

*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).

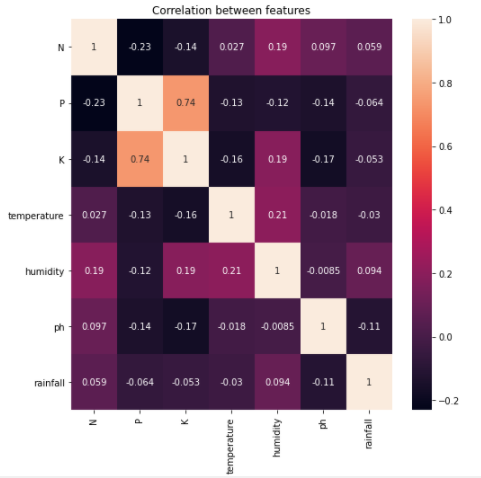
Before training our dataset we conducted following preliminary analysis of the dataset:

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 step i.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)*

*Extra tree classifier*

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.

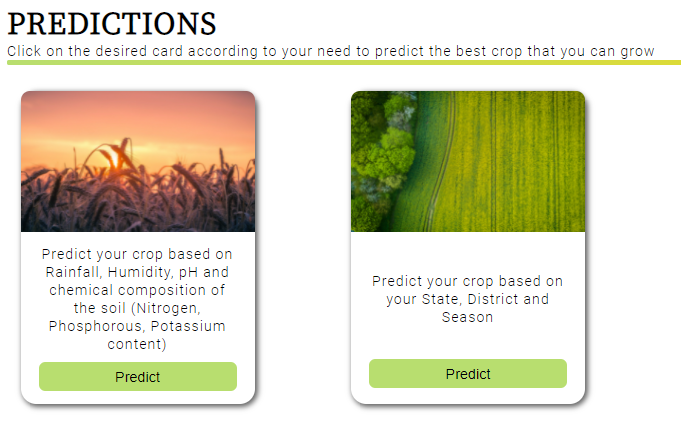
*Figure 1: Home Page of the DigiFarm*

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*

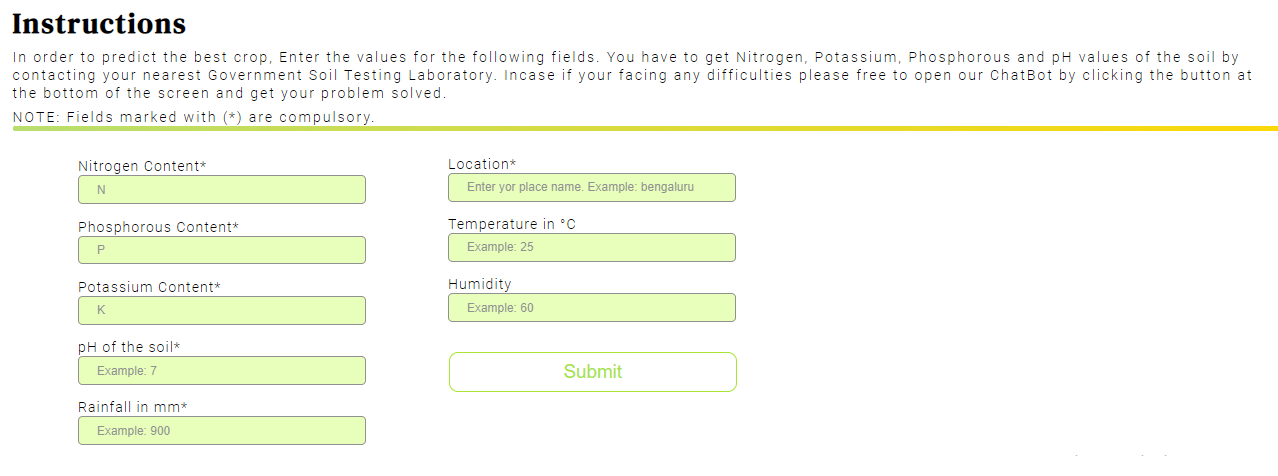
This part of DigiFarm is the main part of the platform. The “Prediction” section as shown in the Figure 2 can be used to predict the most suitable crops by using two different methods.

**

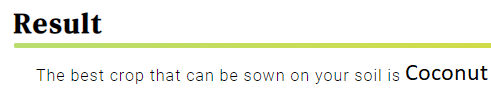
*Figure 2: Predictions Section (i.e. Different methods for predicting the crop)*

*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).

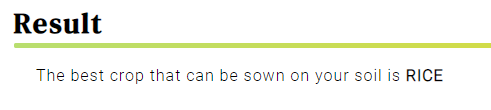
 *Figure 3(a): Prediction of crops using method*

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*

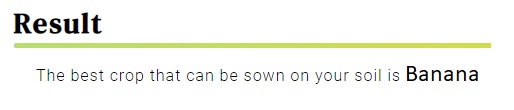
Similarly, we can get different crop names as output (as shown in the figures 3(c) to 3(f)) based on the different input combinations.



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

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

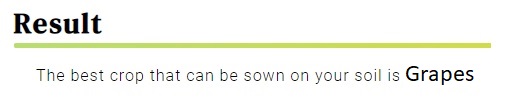
*And Rainfall=221mm.*

**

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

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

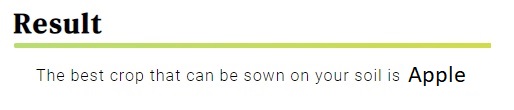
*And Rainfall=90mm.*



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

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

*And Rainfall=66mm.*

**

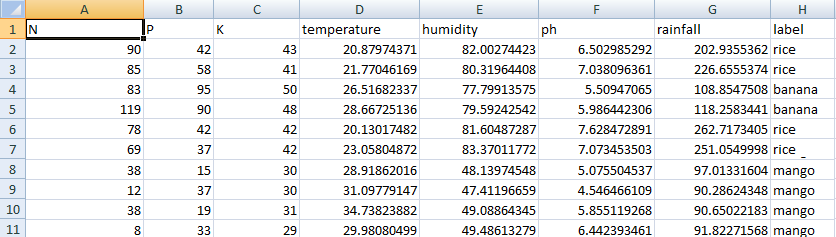
*Figure 3(f): For the input values N=34, P=140,*

*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*

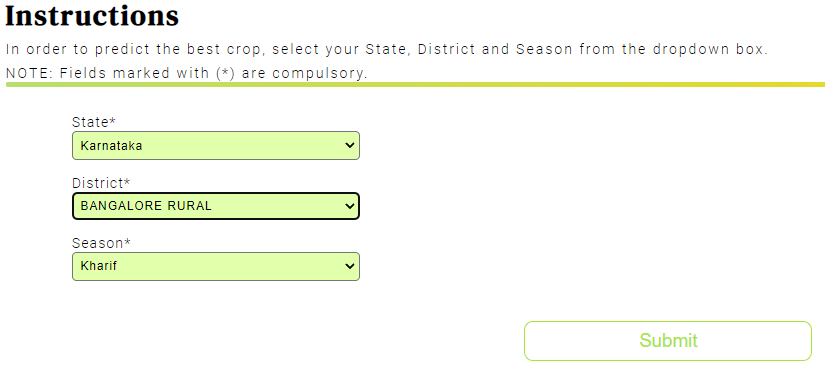
For training and deploying the ML model, the IBM Cloud Services is used which includes ML and AutoAI services. Since it is deployed in a cloud, we are using the ML Model through Application Program Interface (API).

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. The accuracy that we got from LBGM Classifier algorithm on testing was found to be 0.991 and that of Extratree 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*

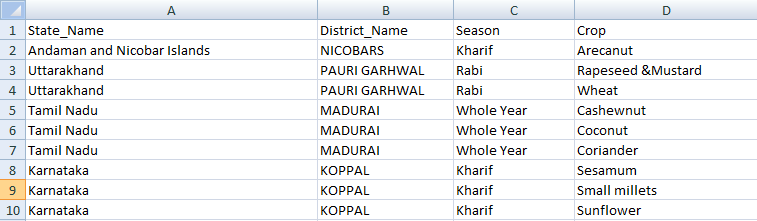
For example:- When we enter the state as Karnataka, district as Bengaluru Rural, season as kharif (ref Figure 5(a)), we get the crop prediction of ONION, DRY GINGER, RAGI, BAJRA, MAIZE, RICE, GRAM and many more as shown in Figure 5(b).



*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.

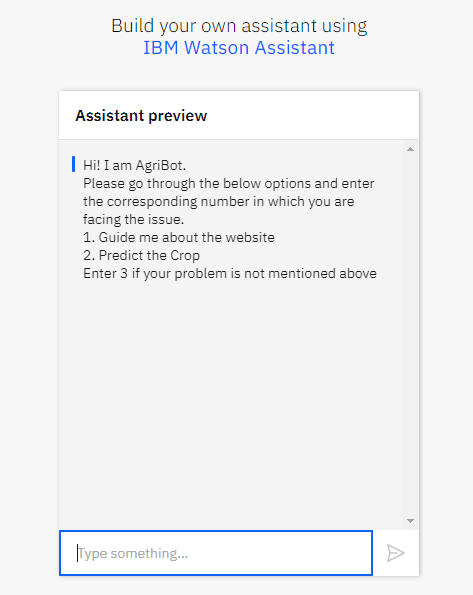
For predicting the crops in this method we are not using any ML model instead we are searching the dataset and displaying the crop names according to the information (i.e. State name, District name and the preferred season) entered by the user.

*Figure 6: Crop prediction dataset for method 2*

*3.4 AgriBot*

After observation and analysis from the different websites, we came to the conclusion that one of the most effective and modern ways of communication is the Artificial Intelligence (AI) based ChatBot. It offers easy and quick solution to the most basic queries. This drastically reduces the time consumption both at the end of the users (i.e. mainly farmers). 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 AgriBot is the easy accessibility of DigiFarm platform 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

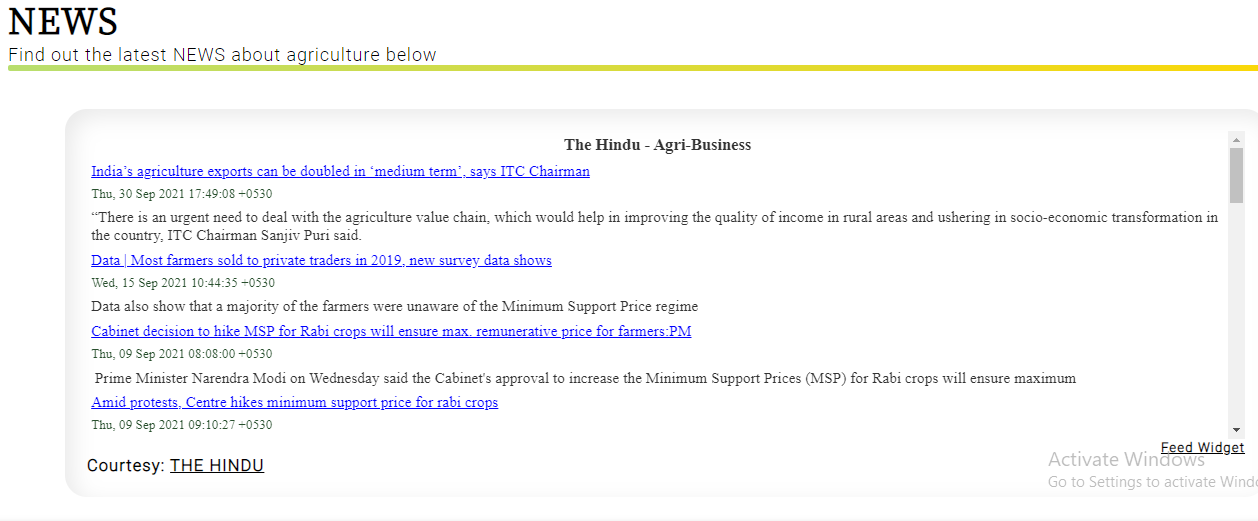
The presence of a AgriBot on this platform simplifies the user experience which is the broader aim of the Digifarm.

*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.