Crop Prediction Based On Soil And Climate

Bhagyashri Bhamare
Information Technology
Department of IT, NITK, Suratkal
Mangalore, Karnataka
bhamare.bhagyashri1999@gmail.com

C Sneha
Information Technology
Department of IT, NITK, Suratkal
Mangalore, Karnataka
snehac661@gmail.com

Gayathri Gutla
Information Technology
Department of IT, NITK, Suratkal
Mangalore, Karnataka
gayathrihogwarts@gmail.com

***Abstract*—Agriculture is considered India's backbone, despite the fact that its economic contribution to the country's GDP is gradually declining as the country's broad-based economic growth accelerates. Despite the fact that agriculture is India's largest economic sector by population and plays a significant role in the country's overall socio-economic fabric, the sector is in decline due to a lack of farmer empowerment and a lack of IT implementation in the farming sector. Farmers have less knowledge about the crops to be grown in this continuously changing climatic conditions. In this project, we intend to predict the crops which are most suitable for the farmers to grow organically and locally, based on the soil quality (ph and nutrients in soil) and climatic conditions of the area like temperature and rainfall.**

***Index Terms*—Crops, Multiple Linear Regression, temperature, rainfall, soil, nutrients**

I. Introduction

Agriculture is India's bedrock, but it has recently wilted, with almost all of the problems stemming from economics, political systems, and policymaking. Via instruments such as the "minimum support price," the government promotes the production of rice and wheat, with a few other crops thrown in for good measure. This method of subsidising wheat, rice, and a few other crops is ineffective. This is because if the crop fails, the farmer has no other option but to go into debt, on top of what they owe from the crop they cultivated. One of the most important aspects of a country's development, such as India's, is its ability to produce food for consumption worldwide.

Agriculture has been linked to the development of important food crops for centuries. In fact, however, the rapid rate of population growth has long been one of our society's most pressing concerns. As a result, the agricultural field's reach has been severely limited, especially in terms of land use and fertility. It is a fact that the amount of land under cultivation is doubtful to increase in this age of globalisation and urbanisation. The emphasis will thus have to be along the lines of capitalizing on what land and resources are accessible. Moreover, due to the unrestrained use of fertilizers comprising potassium,nitrogen and micro-nutrients, crop cultivation prediction is a challenge. Crop cultivation forecasting is an important factor in agriculture for the above reasons. Despite the fact that recent research has made agricultural statistics more available, few studies have looked into crop prediction using previously collected data for organic farming.

Crop output while prediction is influenced by parameters taken as inputs like the temperature, soil quality and texture, and rainfall during the time of the crop's growth. Agriculture input parameters differ from one region to another region, making it difficult to collect such data over vast swaths of land.The huge data sets collected from real-life scenarios can be used to forecast crop yields on a large scale. Due to the following problems,there is a need to develop technology.For farming arable land and making the most of limited land resources, new machine learning methods must be created.Agriculture researchers have been putting a variety of forecasting methodologies to the test in order to find the best crop for particular areas of land.

Monocropping on such a large scale is an offshoot of chemically intensive agriculture and assistance for a few crops that India's diverse food cultures are on the verge of extinction.The nutrients in the soil are depleted, the water is polluted, farmers' incomes are steadily declining, and food security for all of India is becoming increasingly difficult to achieve.

This situation can be managed by experimenting with new crops, but this is difficult due to farmers' lack of knowledge about new crops and the yields they would generate.If a farmer can predict the yield of a crop ahead of time, it will be easier for him to decide which crop to plant.Farmers will be able to predict crop yields using the application built in this study. A user-friendly interface is created for farmers.

Now days predicting appropriate crops for farming is a necessary part of agriculture.Recent years machine learning algorithms playing an important role in such prediction. In this particular area of technology and data science, the agricultural sector is one of the sector which benefits greatly from properly implemented techniques.

Supervised, unsupervised, and reinforcement learning are the three most popular machine learning techniques.This application employs supervised learning classification techniques to predict the best crops for a given location based on soil nutrients and weather patterns.

A. Problem Statement

Given information about crop yields in different areas of the country in different months across the years, predict the crops that are most probable to give out the best yield in the current month. Take soil composition, rainfall, and temperature of previous years into consideration and give out crops(if more than 1 crop possible, usually, it is possible as we have a very big list of options) in the possible increasing order of productivity or yield.

B. Objectives

We intend to predict the crops which are most suitable for farmers based on the soil and climatic conditions of the area like temperature and rainfall.

Agriculture's contribution to India's GDP is decreasing as the country's economy grows. Agriculture continues to play a vital role in India, as it is the country's most populous economic sector and plays a significant role in the country's overall social and economic fabric. The reason for this decline in the agriculture sector is due to the fact that farmers are not empowered and due to the lack of application of IT in the farming sector. Helping farmers choose the best crop and promote organic farming would help the economy, the farmers, and the general health and wellbeing of consumers.

II. LITERATURE SURVEY

This is such a broad and most sought-after topic and requirement. As a result, there are many papers and articles published discussing the various ways algorithms can be implemented to predict the right crop at the right time according to the conditions of a given area. Some of the works we followed are mentioned here.

Some of their references are given in the references section at the end of the paper. Chlingaryan and Sukkarieh(Chlingaryan et al., 2018) used ML to perform sturdy on nitrogen status on the soil.The paper concludes that rapid advances in sensing technology and machine learning techniques can lead to cost-effective agricultural solutions [2].Elavarasan et al. are a group of researchers who have devised a novel method for conducting a survey on machine learning techniques for crop yield prediction based on climatic variables such as temperature and rainfall. The paper recommends broadening your search to find more crop yield-related parameters[3].Liakos et al. published a paper reviewing the use of machine learning methods in agriculture. The research was conducted using publications focusing on crop livestock management, water management, and soil management. Many ML techniques and their applications in plant biology were introduced by Somvanshi and Mishra.

After that, Gandhi and Armstrong released a study paper on data mining applications in agriculture. Beulah conducted a survey of the different data mining approaches used for crop yield prediction and came to the conclusion that data mining techniques could be used to solve the problem.Then there are several review papers that go over numerous previous crop prediction papers that have been written[5].

III. METHODOLOGY

First, we curate the required datasets. Decide on which parameters will be used to predict the crops. We used the datasets from those available in http://www.data.gov.in and http://www.indianwaterportal.org. We have used rainfall, temperature, soil composition, and pH as the parameters for this.

Further, the algorithm used to predict the crops is Multiple Linear Regression after considering its simplicity and easy interpretability. We went for simplicity because there are multiple factors we are taking into consideration for prediction and don't want to make the algorithm complex and time-consuming. As we are using multiple independent variables we use multiple linear regression.

We import LinearRegression models and test_train_split from sklearn. The area where we want to predict the crop is passed as an input. The code uses month input from the user.

We are using four datasets(as already has been discussed in the introduction) for the project. The temprainfall.csv dataset stores the area, state, month, minimum and maximum temperatures recorded in that area in that month and rainfall.

Input: In order to accurately predict the future crop, different factors such as weather nutrients, soil, and previous crop production must be considered.All of these variables are position-dependent, so the device takes the location and month as inputs.

Data Acquisition: The device mines the soil properties in the respective region from the soil repository based on the user's current position.Weather parameters are derived from the weather data collection in the same way.

*Data Processing:*Only under the right circumstances can a crop be grown. These conditions include soil and weather-related variables. These constraints are weighed against each other, and the most reliable crops are chosen. The method predicts the crop using Multiple Linear Regression (MLR). The prediction is based on previous crop production data, which includes defining weather and soil parameters and comparing them to current conditions in order to predict the crop more accurately and practically.

Output: The system uses a Multiple Linear Regression algorithm to predict the most productive crop, and the consumer is given multiple crop recommendations based on the crop's length.

S = {I, S, F, Fu, O, }
I = I1,I2 .... set of Input.
I1 = Location of user
I2 = Month
Fu = { Location(), Attributes(city,State), Soil(), Weather(), FeasibleCrop(soilcondition,weather), PreviousProduction(), ProfitableCrop(FeasibleCrops,PastProduction) }.
Where,
Nutrients in soil – P ,N, K Soil components
Weather conditions – Temperature and Rainfall values
S= Success Condition{Correct forecast for increased production and profit}
F = Failure Condition{Prediction failure due to erroneous training data}
O = Set of output{Crop forecast for a specific location at a specific month}

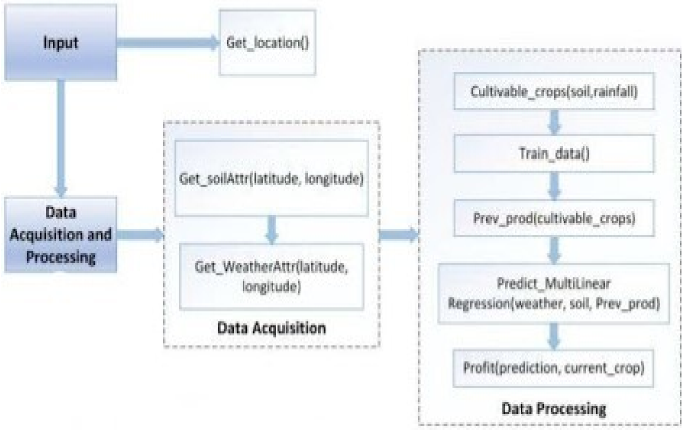


Fig. 1. Proposed system architecture

soil reports.pdf

STATE NAME	N	P	K
Andaman and Nicobar	VL	VL	L
Andhra Pradesh	L	VH	M
Arunachal Pradesh	H	M	M
Assam	M	L	VL
Bihar	VL	VH	H
Goa	M	VL	M
Gujarat	VL	VH	H
Haryana	VL	M	H
Himachal Pradesh	M	M	H
Jammu and Kashmir	M	M	M
Jharkhand	L	M	M
Kerala	VL	VH	M
Madhya Pradesh	VL	M	H
Maharashtra	VL	M	H
Manipur	VL	VH	L
Meghalya	L	M	M
Mizoram	M	M	M
Nagaland	M	VL	H
Odisha	VL	M	M
Puduchery	VL	VL	M
Punjab	M	M	H
Rajasthan	VL	M	H
Sikkim	M	M	M
Tamil Nadu	L	VH	H
Tripura	L	M	L
Uttar Pradesh	VL	M	H
Uttarkhand	L	M	M
West Bengal	M	VH	M
Chattisgarh	L	M	L
Telanagana	L	VH	H
Karnataka	L	VH	M

/ppm(parts per million)		
Nitrogen	Phosphorus	Potassium
VL: 10	VL: below 5	VL: below 100
L:10-20	L:5 -10	L: 100-150
M:20-30	M:10-20	M:150-250
H:30-40	H: 20-30	H: 250-300
VH:40+	VH: 30+	VH: 300+

Fig. 3. Soil nutrients dataset

A. Datasets

The crop production dataset stores the details regarding the crops grown in a particular region and also gives information about the production of each crop with respect to the environmental conditions like rainfall, temperatures and the pH of the soil in which the crop is grown.

State_Name	District_Name	Crop	Area	Production	prod/area	rainfall	temp	Ph
Tamil Nadu	NAMAKKAL	Arecanut	626	25	0.03994	1000.499	35	5.5
Assam	KAMRUP METR	Arecanut	675	91	0.13481	1001.685	35	5.5
Tamil Nadu	KRISHNAGIRI	Arecanut	32	5	0.15625	1001.953	34	5.5
Tamil Nadu	PERAMBALUR	Arecanut	63	11	0.17460	1002.183	33.66667	5.5
Assam	CACHAR	Arecanut	4520	858	0.18982	1002.373	33.16667	5.5
Tamil Nadu	THE NILGIRIS	Arecanut	291	57	0.19588	1002.448	32.66667	5.5
Assam	HAILAKANDI	Arecanut	2690	548	0.20372	1002.546	32.16667	5.5
Tamil Nadu	THE NILGIRIS	Arecanut	332	71	0.21386	1002.673	31.66667	5.5
Assam	CACHAR	Arecanut	4545	1037	0.22816	1002.852	31.16667	5.5
Assam	CACHAR	Arecanut	4505	1058	0.23485	1002.936	30.66667	5.5
Tamil Nadu	PERAMBALUR	Arecanut	16	4	0.25000	1003.125	30.16667	5.5
West Bengal	MEDINIPUR EAS	Arecanut	982	251	0.25560	1003.195	29.66667	5.5
West Bengal	MEDINIPUR EAS	Arecanut	1000	266	0.26600	1003.325	29.16667	5.5
Assam	CACHAR	Arecanut	4184	1133	0.27079	1003.385	28.66667	5.5
Assam	CACHAR	Arecanut	4464	1222	0.27375	1003.422	28.16667	5.5
Assam	HAILAKANDI	Arecanut	2850	783	0.27474	1003.434	27.66667	5.5
Tamil Nadu	TIRUNELVELI	Arecanut	133	37	0.27820	1003.477	27.16667	5.5
Tamil Nadu	KRISHNAGIRI	Arecanut	51	15	0.29412	1003.676	26.66667	5.5
Assam	JORHAT	Arecanut	2286	688	0.30096	1003.762	26.16667	5.5
Assam	KARIMGANJ	Arecanut	2120	641	0.30236	1003.779	25.66667	5.5

Fig. 2. Crop production dataset

Weather stats.pdf

Station Name	Month	Mean Temperature °C		Mean Rainfall in
		Maximum	Minimum	mm
Abu	January	19.3	8	5.3
Abu	February	21	10	4.4
Abu	March	25.3	14.5	6.5
Abu	April	29.4	18.7	2.6
Abu	May	31.5	21	16.4
Abu	June	29.1	19.8	101.6
Abu	July	24.5	18.7	573.2
Abu	August	22.7	17.8	600.3
Abu	September	24.5	17.6	214.2
Abu	October	26.7	16.2	19.4
Abu	November	23.8	12.1	7.9
Abu	December	20.9	9	2.4
Agartala (A)	January	25.6	10	27.5
Agartala (A)	February	28.3	13.2	21.5
Agartala (A)	March	32.5	18.7	60.7
Agartala (A)	April	33.7	22.2	199.7
Agartala (A)	May	32.8	23.5	329.9
Agartala (A)	June	31.8	24.6	393.4
Agartala (A)	July	31.4	24.8	363.1
Agartala (A)	August	31.7	24.7	298.7
Agartala (A)	September	31.7	24.3	232.4
Agartala (A)	October	31.1	22	162.5
Agartala (A)	November	29.2	16.6	46
Agartala (A)	December	26.4	11.3	10.6
Agra	January	22.3	7.7	13.2
Agra	February	25.5	10.3	17.6
Agra	March	31.9	15.5	9.3
Agra	April	37.9	21.5	6.3
Agra	May	41.7	26.5	11.3
Agra	June	40.7	28.9	55.7
Agra	July	35.3	26.8	203.3
Agra	August	33.2	25.7	243.2
Agra	September	34	24.3	129.7
Agra	October	34	19.1	24.8
Agra	November	29.2	12.5	4.3
Agra	December	23.9	8.2	6.1
Ahmedabad	January	28.7	13.1	2.1
Ahmedabad	February	31	14.8	1.2
Ahmedabad	March	35.9	19.4	1.1
Ahmedabad	April	39.5	23.5	1.9
Ahmedabad	May	41.4	26.3	9.1
Ahmedabad	June	38.5	27.2	97.4

Fig. 4. Weather statistics dataset

The soil reports dataset stores details regarding the quality of the soil of each region with respect to Nitrogen, Phosphorus and Potassium (NPK value) content in the soil. This data is helpful to compare and analyze which NPK value best suits the crop in a particular region. The weather statistics dataset stores the details regarding the temperature range(Minimum and Maximum) and the mean of the rainfall if all the regions throughout the year in a monthly fashion.

The crop dataset stores the details of different crops, which are the duration of its cultivation period(in months), minimum and maximum temperature, rainfall and pH value required for its optimal growth and also the best suited NPK value for the particular crop.

Apart from these four datasets there are two more datasets, metacrop.csv and metacrop11.csv which are created from the given datasets.

B. Algorithm for proposed system

Train data:

Y1 $=\beta_0 + \beta_1 xi_1 + ... + \beta_p xi_p \quad \epsilon$ i for i in range 1 to n inclusive here,

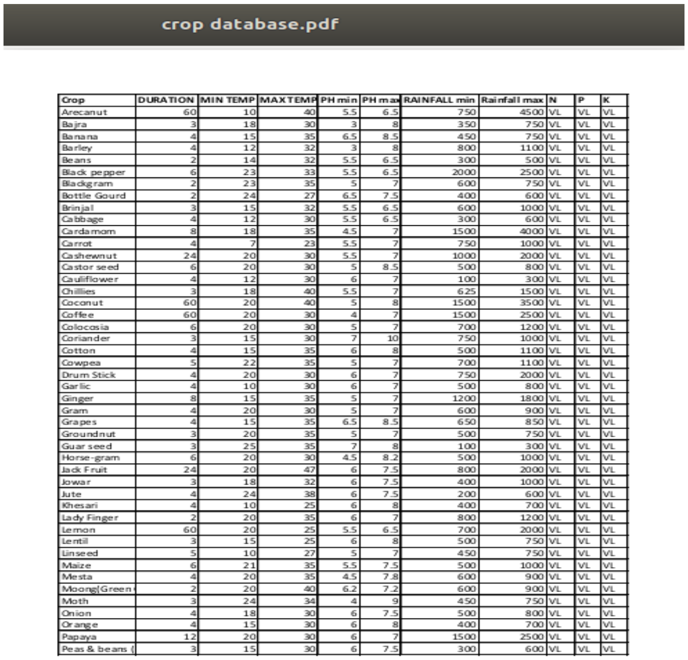
1)$\beta_0, \beta_1, \beta_2, ...$ the MLR coefficients

2)$xi_1, xi_2, ...xi_p these are the independent variables.$

3)X set of weather and soil attributes

4)Y Production

Y=Xβ + E



crop database.pdf

Crop	DURATION	MIN TEMP	MAXTEMP	PH min	PH max	RAINFALL min	Rainfall max	N	P	K
Arecanut	60	10	40	5.5	6.5	750	4500	VL	VL	VL
Bajra	3	18	30	3	8	350	750	VL	VL	VL
Banana	4	15	35	6.5	8.5	450	750	VL	VL	VL
Barley	4	12	32	3	8	800	1100	VL	VL	VL
Beans	2	14	32	5.5	6.5	300	500	VL	VL	VL
Black pepper	6	23	33	5.5	6.5	2000	2500	VL	VL	VL
Blackgram	2	23	35	5	7	600	750	VL	VL	VL
Bottle Gourd	2	24	27	6.5	7.5	400	600	VL	VL	VL
Brinjal	3	15	32	5.5	6.5	600	1000	VL	VL	VL
Cabbage	4	12	30	5.5	6.5	300	600	VL	VL	VL
Cardamom	8	18	35	4.5	7	1500	4000	VL	VL	VL
Carrot	4	7	23	5.5	7	750	1000	VL	VL	VL
Cashewnut	24	20	30	5.5	7	1000	2000	VL	VL	VL
Castor seed	6	20	30	5	8.5	500	800	VL	VL	VL
Cauliflower	4	12	30	6	7	100	300	VL	VL	VL
Chillies	3	18	40	5.5	7	625	1500	VL	VL	VL
Coconut	60	20	40	5	8	1500	3500	VL	VL	VL
Coffee	60	20	30	4	7	1500	2500	VL	VL	VL
Colocosia	6	20	30	5	7	700	1200	VL	VL	VL
Coriander	3	15	30	7	10	750	1000	VL	VL	VL
Cotton	4	15	35	6	8	500	1100	VL	VL	VL
Cowpea	5	22	35	5	7	700	1100	VL	VL	VL
Drum Stick	4	20	30	6	7	750	2000	VL	VL	VL
Garlic	4	10	30	6	7	500	800	VL	VL	VL
Ginger	8	15	35	5	7	1200	1800	VL	VL	VL
Gram	4	20	30	5	7	600	900	VL	VL	VL
Grapes	4	15	35	6.5	8.5	650	850	VL	VL	VL
Groundnut	3	20	35	5	7	500	750	VL	VL	VL
Guar seed	3	25	35	7	8	100	300	VL	VL	VL
Horse-gram	6	20	30	4.5	8.2	500	1000	VL	VL	VL
Jack Fruit	24	20	47	6	7.5	800	2000	VL	VL	VL
Jowar	3	18	32	6	7.5	400	1000	VL	VL	VL
Jute	4	24	38	6	7.5	200	600	VL	VL	VL
Khesari	4	10	25	6	8	400	700	VL	VL	VL
Lady Finger	2	20	35	6	7	800	1200	VL	VL	VL
Lemon	60	20	25	5.5	6.5	700	2000	VL	VL	VL
Lentil	3	15	25	6	8	500	750	VL	VL	VL
Linseed	5	10	27	5	7	450	750	VL	VL	VL
Maize	6	21	35	5.5	7.5	500	1000	VL	VL	VL
Mesta	4	20	35	4.5	7.8	600	900	VL	VL	VL
Moong(Green	2	20	40	6.2	7.2	600	900	VL	VL	VL
Moth	3	24	34	4	9	450	750	VL	VL	VL
Onion	4	18	30	6	7.5	500	800	VL	VL	VL
Orange	4	15	30	6	8	400	700	VL	VL	VL
Papaya	12	20	30	6	7	1500	2500	VL	VL	VL
Peas & beans (	3	15	30	6	7.5	300	600	VL	VL	VL

Fig. 5. Crop dataset

Y is the production matrix
X is the attributes matrix
β is the Partial coefficient matrix
E is error control
$\beta = ((X'X) \wedge (-1))X'Y\ldots\ldots$Least Square Estimate
X' is Transpose
X-1 is the Inverse of Matrix
Prediction:- Y = Xβ
Result:- res = Y - Y

IV. Results And Analysis

The above mentioned methodology is implemented and a User Interface(UI) is created to enable convenient use of the application by farmers. The home page has options to select the inputs i.e. location of cultivation and the month of planting crop. According to these inputs, the prediction of appropriate crops that can be planted can be made.

The location selection on the home page has an option of selecting cities out of the 108 options available. Whereas, the month selection has an option of selecting any of the 12 months according to the requirement.

The output is then displayed which shows the crops that the farmer can grow organically according to the location and month entered. Moreover, the different datasets used can be viewed from a menu on the application.

V. Conclusion

With the current efficiently working model, the front end can be further improved to include an option to take location using GPS. Also, work can be done in training the model with

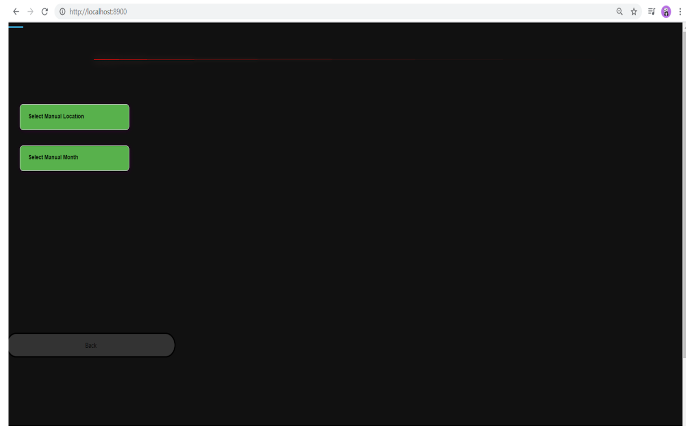


Fig. 6. Home page of application

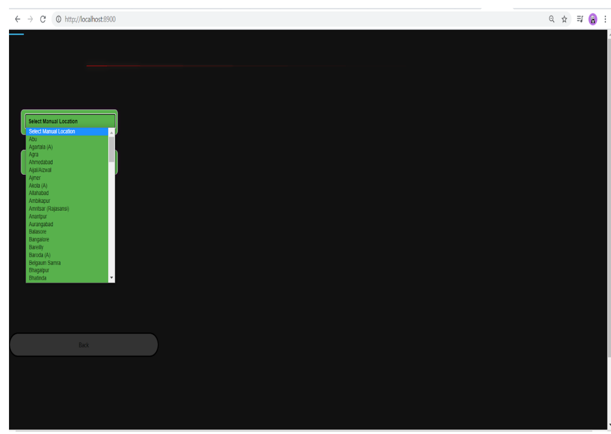


Fig. 7. Location selection on home page

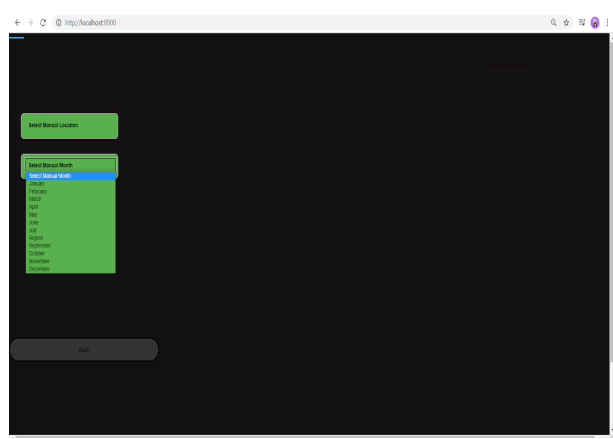


Fig. 8. Selection of month of planting crops on home page

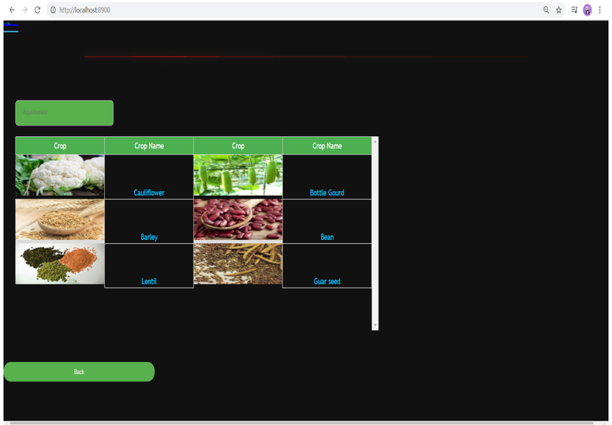


Fig. 9. Output for Aijal

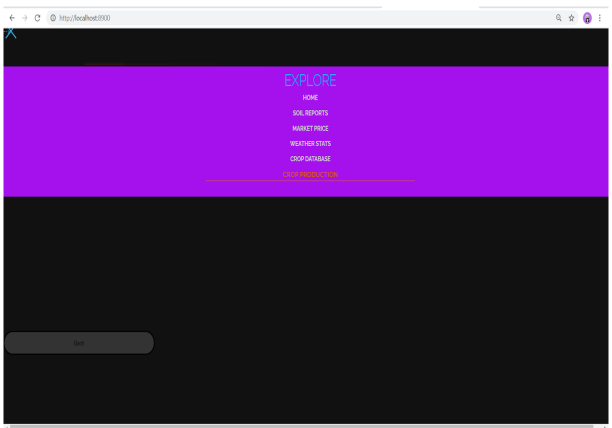


Fig. 10. Datasets on menu

larger datasets for a precise value and the smooth working of the research. With all these developments it can be used in the real world to truly benefit farmers. In the future, we will use IoT to link all of the farming devices. We can use sensors to gather data about current farm conditions, and then put devices in place to increase moisture, acidity, and other factors.The sensors and vehicles used on farms can be linked to the internet of things, allowing farmers and research centres to receive real-time data about harvesting, farm yield, diseases, and inadequate nutrient levels, among other things. All of this would aid in improving the precision of its predictions.

References

[1] A. Suruliandi, G. Mariammal, S. P. Raja, 'Crop prediction based on soil and environmental characteristics using feature selection techniques', 12 Mar 2021,

[2] Anna Chlingaryan, Salah Sukkarieh, Brett Whelan, "Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture", August 2018

[3] Dhivya Elavarasan, Durai Raj Vincent P M, "Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian Applications",May 2020

[4] Niketa Gandhi, Leisa J. Armstrong, Owaiz Petkar, Amiya Kumar, Tripathy,"Rice crop yield prediction in India using support vector machines", July 2016

[5] Thomasvan Klompenburga, Ayalew Kassahuna, Cagatay Catal"Crop yield prediction using machine learning: A systematic literature review", October 2020