

Crop Prediction using Machine Learning

M.Kalimuthu /Assistant Professor
Bannari Amman Institute of Technology
Tamil Nadu, India

P.Vaishnavi /Assistant Professor
Bannari Amman Institute of Technology, Bannari Amman Institute of Technology,
Tamil Nadu, India

M.Kishore
Tamil Nadu, India

Abstract —In general, agriculture is the backbone of India and also plays an important role in Indian economy by providing a certain percentage of domestic product to ensure the food security. But now-a-days, food production and prediction is getting depleted due to unnatural climatic changes, which will adversely affect the economy of farmers by getting a poor yield and also help the farmers to remain less familiar in forecasting the future crops. This research work helps the beginner farmer in such a way to guide them for sowing the reasonable crops by deploying machine learning, one of the advanced technologies in crop prediction. Naive Bayes, a supervised learning algorithm puts forth in the way to achieve it. The seed data of the crops are collected here, with the appropriate parameters like temperature, humidity and moisture content, which helps the crops to achieve a successful growth. In addition as the software, a mobile application for Android is being developed. The users are encouraged to enter parameters like temperature and their location will be taken automatically in this application in order to start the prediction process.

Keywords — *Precision Agriculture, Machine learning, Crop prediction, Naive Bayes, Supervised Learning, Effective farming*

I. INTRODUCTION

From ancient days, agriculture is considered as the main source of supply to satisfy the daily needs of human lives. It is also considered a primary occupation, and also one of the India's major industrial sectors. The farmers are ought to follow a traditional naked eye observation and yielded healthy crops without the involvement of chemicals for animals and also to their cultivation land in order to keep healthy diversity. But nowadays, weather conditions are being rapidly changing against the elemental assets to deplete the food and increase the security. In meantime, the GDP in agricultural sector is keep on decreasing, where in 2005 it was about 17.2%, in 2012 it was 11.1, in 2018 it was 5% and in first quarterly year of 2109-2020 it came down to 2%. Approximately 80 percent of farmers come from rural areas, and if the revenue from crop production goes down, their lifestyle would be influenced by the farms at industry level.

This makes sense to farmers in India to show some special concern towards effective and precision farming. In India there are multiple ways to rise the crop learn profit and improve the standard of the crops so as to keep up the economic growth within the field of agriculture. So, the deployment of one of the recent advancement in technology such as, Machine learning is one among the answer for

predicting the crop with relation to atmospheric & soil parameter of the agricultural land. Since, now-a-day's climatic conditions aren't predictable like decades ago. It is changing day by day due to globalization.

Hence, the farmers are facing difficulties in forecasting the weather and crops based on climate data. In recent years the advancement of Machine Learning plays a crucial role in every field including agriculture, here the crop prediction process done with consolidating the preceding data and the present data of a particular month to prove the accuracy of climatic data. Machine learning may be a methodology of analyzing information to automatize the given model and may be a branch of AI depend on the concept that systems will study from data to form selections with minimal human intervention. There may be a logical classifier, where a naive mathematician who predicts membership opportunities for each group, such as the possibility that knowledge belongs to a specific class.

The proposed system analyzes the application of supervised machine learning approaches the class with the very best chance is taken into account as the possibly class. Here the category is nothing however the crop that get foretold for the given input parameters. Once the crop is foretold, it will facilitate the farmers to predict the affordable crop for their individual land. Then, the farmers is guided with an application in mobile tend to make them to understand that what quite seeds we will tend to sow in land to induce higher yielding. Within the past preceding data, crop prediction was calculated by analyzing farmer's previous expertise on climatic condition. So, the correct data regarding history of climatic condition is a vital factor for creating selections in choosing crops. Therefore, this paper proposes a thought to predict the affordable crop for the given input parameter for the poor farmers using machine learning. Thereby this proposed work will suggest the farmers with effective solutions for more profitable cultivation,

II. RELATED WORK

Agricultural machine learning is a new technology; where a large number of researches have been done with the technology imposing in the field of agriculture using machine learning. In Machine learning, where the system is capable of learning by itself without specified any

programming so it improves machine efficiency by sensing and describing drive data consistency and pattern. Machine learning can be classified into 3 types based on the learning techniques– supervised learning, unsupervised learning, and recurrent learning.

Arun Kumar & et al., “Efficient Crop Yield Prediction Using Machine Learning Algorithms” [1], In this study, the classification of crop yields was performed to batch using Artificial neural networks based on yield productivity. And it will define the range of productivity. Regression is carried out to obtain the real crop yield and the expected cost. Nithin Singh & saurabh chaturvedi, “Weather forecasting using machine learning” [2] is consider for collecting historical weather data from various weather stations to forecast weather conditions for future. Aakash Parmar & Mithila Sompura, “Rainfall prediction using Machine Learning Techniques” [3] is taken into account to get the knowledge in predicting the weather for crop prediction. Sachee Nene & Priya, “Prediction of Crop yield using Machine Learning” [4], to know about the prediction crop with respect to atmospheric & soil parameters. Ramesh Medar & Anand M. Ambekar, “Sugarcane Crop prediction Using Supervised Machine Learning” [5], is considered in order to get to predict the unique crop by applying descriptive analytics using three datasets like as Soil dataset, Rainfall dataset, and Yield dataset as a combined dataset. Andrew Crane Droesch, “Machine learning methods for crop yield prediction and climate change impact assessment in agriculture” [6], is proposed semi parametric variant of a deep neural network model for crop prediction and evaluate the effects of climate change. Vinita Shah & Prachi Shah, “Groundnut Prediction Using Machine Learning Techniques” [7], is taken soil, environment and abiotic attributes for predicting the groundnut yield using different ML algorithms. The accuracy of the prediction was compared using RMSE. Renuka & Sujata Terdal, “Evaluation of Machine Learning Algorithms for Crop Prediction” [8], deals with estimation of crop yield from precipitation and soil input. For prediction, supervised learning algorithms were used. The algorithms were compared using MSE for finding an optimal crop prediction. P. Vinciya, Dr. A. Valarmathi, “Agriculture Analysis for Next Generation High Tech Farming in Data Mining” [9], discuss MLR method for analyzing crops and decision tree algorithm for classification of more than 350 data. It classifies real estate, organic and inorganic soil types. Shivnath Ghosh, Santanu Koley, “Machine Learning for Soil Fertility and Plant Nutrient Management using Back Propagation Neural Networks” [10], detailed the Back Propagation Network to estimate the testing data. The hidden layers of Back Propagation Network responsible for the prediction of soil properties. This method provides great accuracy than the usual method.

III. PROPOSED SYSTEM

The proposed system follows Naive Bayes classifier, the supervised learning algorithm consist of the four levels to calculated and predict the crop for the suitable climate in phenomenon such as,

A. Data Collection

Data is composed from a different source and optimized for data sets. And the data is used to evaluate descriptively. Several abstract online outlets, like Kaggle, Google weather forestation and data government, provide the data for up to 10years in series. The data sets such as soil nature, climatic conditions and seed data are used for the crop prediction and better crop yields.

B. Preprocessing Step

Preprocessing the data is considered as a significant step machine learning phase. Preprocessing involves adding the missing values, the correct set of data, and extracting the functionality. Data set form is important to the process of analysis. The data collected in this step will induced in Google Colab platform in the form of python programming in order to get the desired output.

C. Feature Extraction

Extraction of the features would reduce the data size involved to characterize a wide collection of data. The characteristics of soil, crop and weather collected from the pretreatment process establish the final training data collection. This approach selects the features based on the correlation matrix i.e. the features that has more correlation value is selected as an important predictive function for yield as shown in fig.1

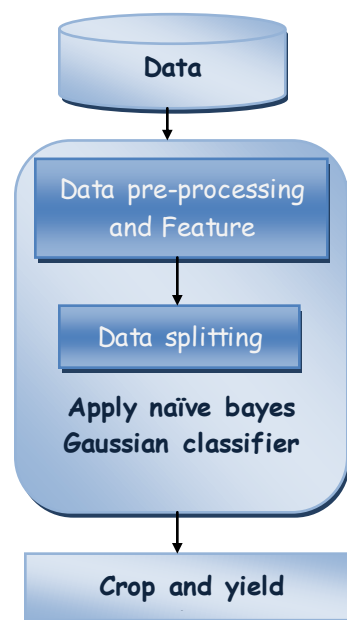


Fig. 1. Steps in proposed system

D. Data Prediction

In advance to this step there need to split the data into train dataset and test dataset. By applying the Naïve Bayes Gaussian classifier the data is trained with available input and output data. In the test phase, the data are tested if the accuracy of e model is satisfied. Then the new data is predicted by machine learning module.

IV. SYSTEM WORKING

A. Hardware and Software Implementation

Energized ATmega imposed arduino UNO microcontroller has 20 digital input/output pin used to control the given process for the connected input & output components through programming in Arduino software taken into account for data collection from DHT 11 sensor module. The DHT11 is a digital temperature and humidity sensor that has two inputs & one output pin. This sensor is used to measure the surrounding temperature & humidity of air at high accuracy with more timing to grab the data. The sensor starts dysfunction when its finds the temperature over 680°C. In meantime, the soil moisture sensor is used to measure the moisture content (water) of the soil.

In the figure 2 hardware setup, where the components used in the proposed work are mentioned as, Arduino UNO with Atmega processor is cited as 1, Dht11 temperature and moisture sensor is cited as 2 and the soil moisture is cited as 3.

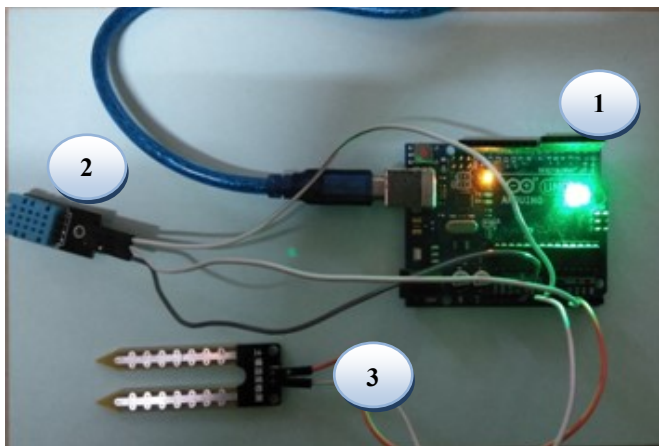


Fig. 2: Hardware setup

The collected data then send to the Google Colab, an open-source platform which is written completely in Python and is designed for data analyst. Google Colab offers built-in integration with many popular scientific packages including Numpy, Pandas.

B. Process Block Diagram of the System

The proposed system consist of four major process includes, collection of preceding data about the weather,

collecting present data, data consolidation and seed data collection and the block diagram representing the above process are shown in fig. 3

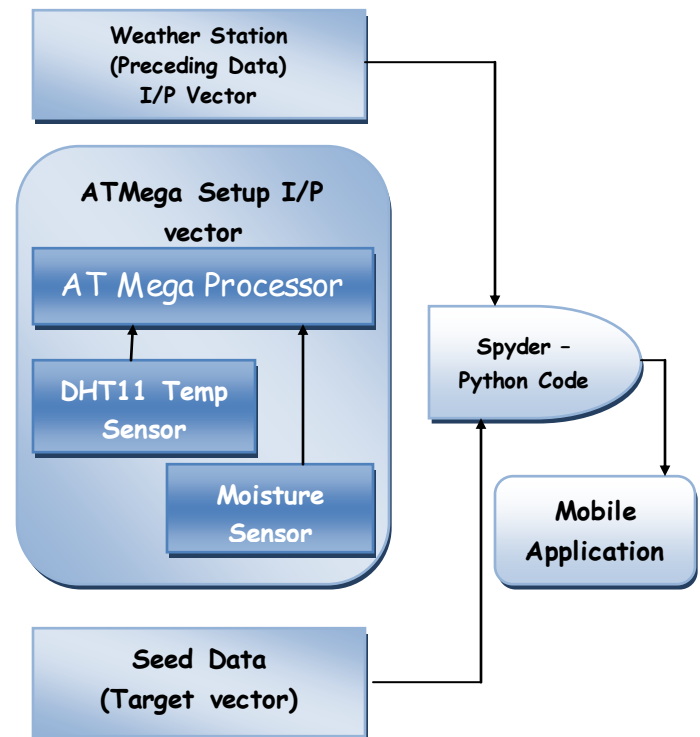


Fig. 3: Block Diagram

Data collection, pre-processing of data, selection of features, and prediction are consider as various steps involved in crop yield prediction and the process flow shown in the fig.4

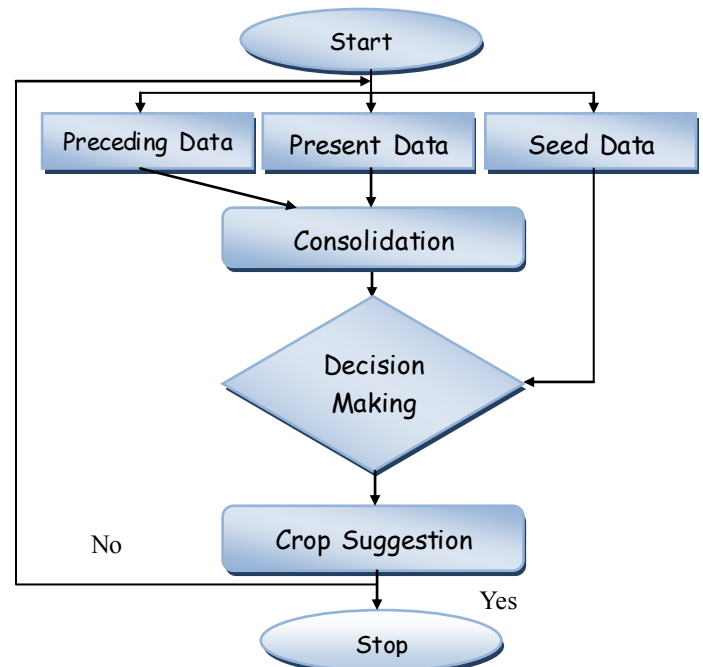
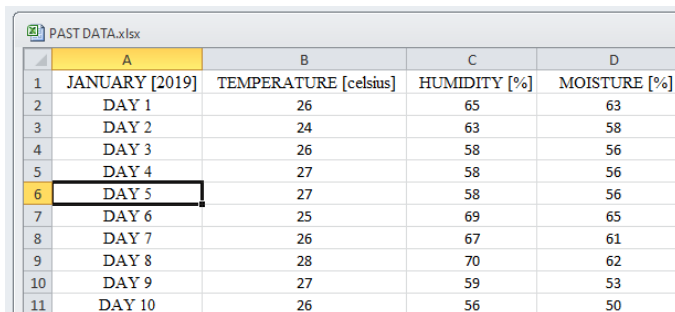


Fig. 4: System process flow

B1. Collecting Preceding Data

The preceding data collection includes the temperature, humidity of air, moisture of the soil & rainfall for a month from the platform that contains data at high accuracy for example weather station. The need for collecting the preceding data will be explained in the following section and the collected data represented in the sheet as shown in fig.5.



	A	B	C	D
1	JANUARY [2019]	TEMPERATURE [celsius]	HUMIDITY [%]	MOISTURE [%]
2	DAY 1	26	65	63
3	DAY 2	24	63	58
4	DAY 3	26	58	56
5	DAY 4	27	58	56
6	DAY 5	27	58	56
7	DAY 6	25	69	65
8	DAY 7	26	67	61
9	DAY 8	28	70	62
10	DAY 9	27	59	53
11	DAY 10	26	56	50

Fig. 5: Preceding data [Excel Sheet]

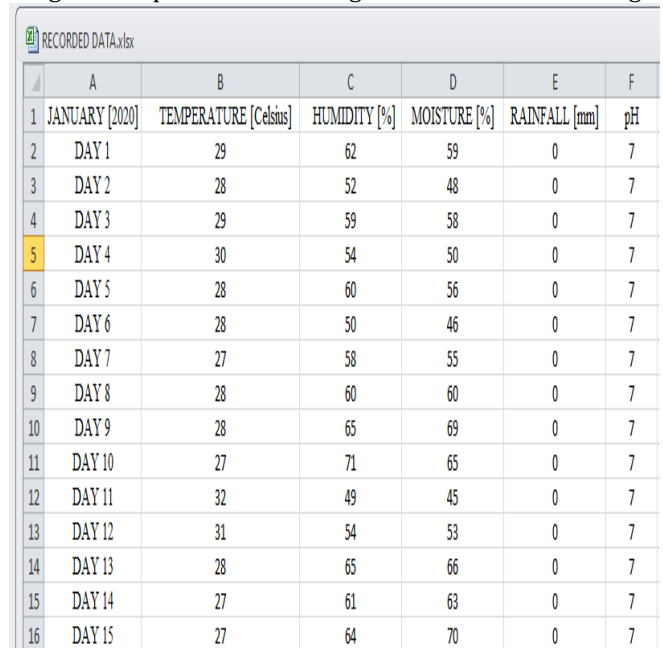
B.2 Collecting Present Data

After collecting the preceding data, the present data is collected for a month by using the designed Arduino setup as shown in fig.8 that has Arduino Uno board, DHT11 sensor & moisture sensor. Each component in this set-up performs particular operation to collect the data at present. Once the controller gets powered up the sensor connected to it will sense the data with some time delay, stored as .csv file and transfer to the controller board then to cloud. The sensor value is displayed on the screen as output value sensed by the respective sensor.



Fig. 6: Arduino Uno Set-Up

After getting the temperature, humidity, moisture value additionally added sensor to get pH & rainfall value as a present data. The need for collecting the present data will be explained in the following section and the resultant image of the present data being collected is shown in fig. 7

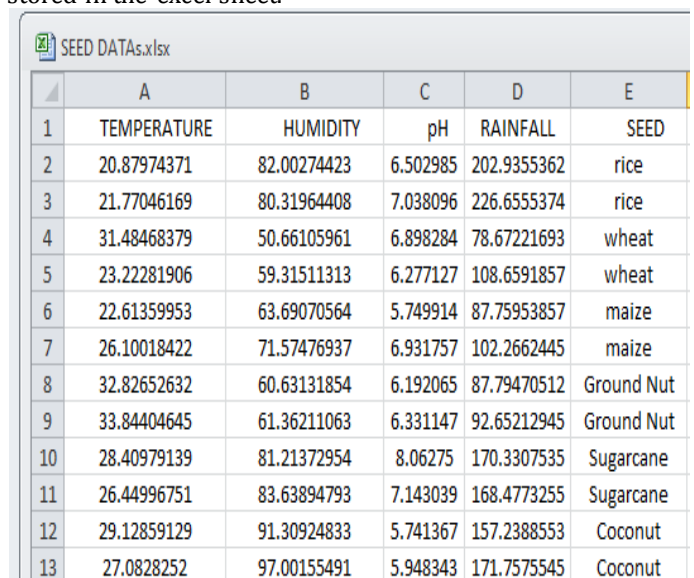


	A	B	C	D	E	F
1	JANUARY [2020]	TEMPERATURE [Celsius]	HUMIDITY [%]	MOISTURE [%]	RAINFALL [mm]	pH
2	DAY 1	29	62	59	0	7
3	DAY 2	28	52	48	0	7
4	DAY 3	29	59	58	0	7
5	DAY 4	30	54	50	0	7
6	DAY 5	28	60	56	0	7
7	DAY 6	28	50	46	0	7
8	DAY 7	27	58	55	0	7
9	DAY 8	28	60	60	0	7
10	DAY 9	28	65	69	0	7
11	DAY 10	27	71	65	0	7
12	DAY 11	32	49	45	0	7
13	DAY 12	31	54	53	0	7
14	DAY 13	28	65	66	0	7
15	DAY 14	27	61	63	0	7
16	DAY 15	27	64	70	0	7

Fig. 7: Present Data [Excel Sheet]

B.3 Collection of Seed Data

After getting graphical output from the consolidation process, expressing the average to determine the accuracy between them. The Seed data is collected for different crops considering different parameter values to grow healthy. The collected crops will be favorable to the climate of city erode for better yielding. Thus the collected data is stored in the excel sheet.



	A	B	C	D	E
1	TEMPERATURE	HUMIDITY	pH	RAINFALL	SEED
2	20.87974371	82.00274423	6.502985	202.9355362	rice
3	21.77046169	80.31964408	7.038096	226.6555374	rice
4	31.48468379	50.66105961	6.898284	78.67221693	wheat
5	23.22281906	59.31511313	6.277127	108.6591857	wheat
6	22.61359953	63.69070564	5.749914	87.75953857	maize
7	26.10018422	71.57476937	6.931757	102.2662445	maize
8	32.82652632	60.63131854	6.192065	87.79470512	Ground Nut
9	33.84404645	61.36211063	6.331147	92.65212945	Ground Nut
10	28.40979139	81.21372954	8.06275	170.3307535	Sugarcane
11	26.44996751	83.63894793	7.143039	168.4773255	Sugarcane
12	29.12859129	91.30924833	5.741367	157.2388553	Coconut
13	27.0828252	97.00155491	5.948343	171.7575545	Coconut

Fig. 8: Seed Data [Excel Sheet]

B.4 Consolidation of Data

The need for collecting the preceding and present is to perform consolidation process for checking accuracy between them. When there is a reasonable accuracy then there will be no occurrence of problem in predicting output. To perform consolidation process there need to perform the python programming in a Google Colab platform that has panda & numpy library in it to perform mathematical operation like average. Thus, both the data is fed into the program to obtain the average output value in graphical form.

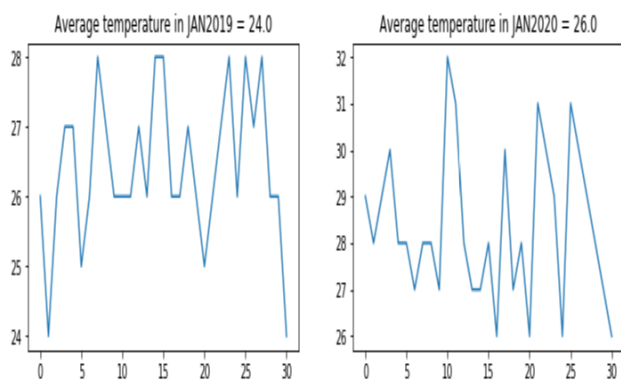


Fig. 9: Average output of data

The consolidation of data and decision making is the major step involved in this mentioned work. Wherein. Naive Bayes classifier, a Supervised learning algorithm in this precision agriculture helps to construct the most accurate and efficient model to predict crop and also the yield.

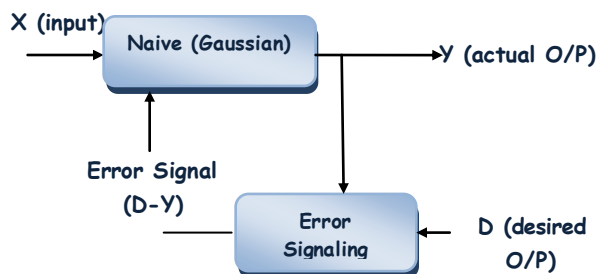


Fig. 10: Supervised Naive Bayes process flow

The Naive Bayes extension is known as the Gaussian Naive Bayes. The mathematics behind the algorithm is Bayes theorem. The Bayes Theorem predict the probability of an occurring event when the probability of already occurs event is known. Bayes' theorem equation:

$$P\left(\frac{h}{d}\right) = \frac{P\left(\frac{d}{h}\right) * P(h)}{P(d)} \quad (1)$$

Where,

$P\left(\frac{h}{d}\right)$ - Probability of hypothesis h given the seed data d.

$P\left(\frac{d}{h}\right)$ -Probability of data d given that the hypothesis h was known.

$P(h)$ -Hypothesis probability of which h being true called the prior probability of h.

$P(d)$ -Probability of the data (independent of the hypothesis).

For finding the maximum probability of hypothesis then the equation 1 can be rewritten as

$$MAP(h) = \max\left(P\left(\frac{h}{d}\right)\right) \quad (2)$$

In this type of algorithms, where the training is labeled and the goal is to find an appropriate function for mapping input with an output. It helps to get the desired output under the supervision of a teacher and also it has the ability to predict the target based on various data using high probability technique as shown in fig .10,. This algorithm also plays a main role in predicting seeds based on various data [Temp, Humidity, Moisture, pH & rainfall] using high probability technique.

V. PROPOSED METHOD

In Naïve Bayes method the accuracy of the model is 97%. In order to improve the accuracy boosting method is used. In boosting method the weak rule has to find and combine these rules to make it stronger. The weak rule has been found by iterative process. The weak rule was found by distribution of Naïve Bayes algorithm.

For right distribution

1. The distribution taken by base learner and allot same weight to all observation
2. If any error occurred by prediction of 1st Naïve Bayes base algorithm, then more attention has to pay. Then next Naïve Bayes base algorithm has to apply
3. Repeat the step 2 until it gives higher accuracy

VI. Result

After collecting the data, it is prepared to get implement in the Google Colab platform by altering it to create a ratio of 70 % for training and 30% for testing which is done by train test library. When the data is rationed correctly, the seed data is fed into the program as an input as shown in fig. 11 to get predicted output.

	A	B	C	D	E	F
1252	25.09898	92.36099	6.047044	157.7593	Coconut	
1253	27.18723	92.19907	6.137103	141.3221	Coconut	
1254	29.03065	90.79094	5.894027	205.572	Coconut	
1255	27.7543	95.94644	5.562224	131.09	Coconut	
1256	27.31156	99.96906	5.832608	201.8259	Coconut	
1257	28.84271	99.64329	6.218572	224.4017	Coconut	
1258	26.61423	96.97301	6.142011	191.0067	Coconut	
1259	29.70143	95.65754	6.078807	215.1968	Coconut	
1260	25.5476	91.64195	5.702485	212.8676	Coconut	
1261	26.37978	91.49883	5.547595	167.0471	Coconut	
1262	29.4844	63.19915	7.454532	71.89091	Black gram	
1263	26.73434	68.14	7.040056	67.15096	Black gram	
1264	26.27274	62.28815	7.418651	70.23208	Black gram	
1265	34.03679	67.21114	6.501869	73.23574	Black gram	
1266	28.03644	65.06602	6.814411	72.49508	Black gram	
1267	28.81461	65.33538	7.581443	62.26243	Black gram	
1268	34.03619	64.28791	7.741419	66.85511	Black gram	
1269	33.86429	61.57072	6.573532	68.022	Black gram	
1270	32.84213	68.68401	7.543804	73.67166	Black gram	
1271	27.10053	63.36086	6.540821	73.8495	Black gram	
1272	25.65843	61.18236	7.224059	69.28608	Black gram	
1273	32.34744	66.61453	7.551364	64.55882	Black gram	

Fig. 11: Parameter Data [Input]

When the seed data and its parameter, black gram [sample] gets fed as an input as shown in fig. to get the sampled output. Where, once the parameters such as temperature, humidity, moisture, pH of the soil entered manually the crop gets predicted by considering the preceding data and the performance of naive Bayes classifier algorithm carried as mentioned. For example the data suited for the black gram has been given in the Google Colab and it does the consolidation part and get back the black gram as an output as shown in fig.12

```
var=[(29.4844,63.19915,7.454532,71.89091)]
y_pred = gnb.predict(var)
y_pred
array(['Black gram']) dtype='<U12')
```

Fig. 12: Predicted Crop [Output]

In further an android application also developed for the farmers for their convenience where, the farmer are provided with options just to enter the temperature, humidity, soil moisture and pH. Using machine learning the consolidation of data will be getting automatically and the identified crop will be predicted as shown in fig.12 where a sample has shown in fig.13 while the machine learning based crop prediction system has predicted the

crop successfully for the given input in reference with the preceding seed data.

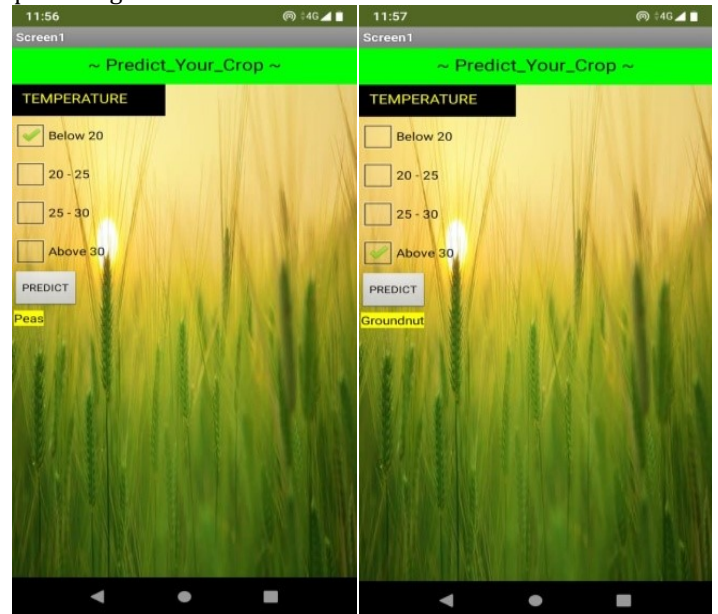


Fig-13: Mobile Application

VII. CONCLUSION

The proposed supervised machine learning using naive Bayes Gaussian classifier with boosting algorithm is developed to predict the crop at high accuracy. Thus, the seed is predicted as an output for the given input parameter. This work may greatly help the needy farmers who have less knowledge in predicting the crops for developing a sustainable future. In future it may also extend to suggest the fertilizer, suitable guidelines for cropland and crops for the given input. In addition, source of sunlight and crop health are monitored at regular intervals and it is also taken into the account for achieving a better crop yield.

References

- [1] Arun Kumar, Naveen Kumar, Vishal Vats, "Efficient Crop Yield Prediction Using Machine Learning Algorithms", International Research Journal of Engineering and Technology (IRJET)- e-ISSN: 2395-0056, p-ISSN:2395-0072, Volume: 05 Issue: 06 | June-2018
- [2] Nithin Singh & saurabh chaturvedi, "Weather Forecasting Using Machine Learning", 2019 International Conference on Signal Processing and Communication (ICSC) Volume: 05 | DEC-2019.
- [3] Aakash Parmar & Mithila Sompura, "Rainfall Prediction using Machine Learning", 2017 International Conference on (ICIIECS) at Coimbatore Volume: 3 | March 2017.
- [4] Sachee Nene & Priya, R "Prediction of Crop yield using Machine Learning", International Research Journal of Engineering and Technology (IRJET) Volume: 05 Issue: 02 | Feb-2018.
- [5] Ramesh Medar & Anand M. Ambekar, "Sugarcane Crop prediction Using Supervised Machine Learning" published in International Journal of

Intelligent Systems and Applications Volume: 3 | August 2019.

- [6] Andrew Crane Droesch, "*Machine learning methods for crop yield prediction and climate change impact assessment in agriculture*", Published by IOP Publishing Ltd Volume: 05 | OCT -2018.
- [7] Vinita Shah & Prachi Shah, "*Groundnut Prediction Using Machine Learning Techniques*" ,published in IJSRCSEIT. UGC Journal No : 64718 | March-2020.
- [8] Renuka & Sujata Terdal, "*Evaluation of Machine Learning Algorithms for Crop Prediction*" Published in International Journal of Engineering and Advanced Technology (IJEAT) Volume-8 | August 2019.
- [9] P. Vinciya, Dr. A. Valarmathi, "*Agriculture Analysis for Next Generation High Tech Farming in Data Mining*" IJARCSSE, vol. 6, Issue 5, 2016
- [10] Shivnath Ghosh, Santanu Koley, "*Machine Learning for Soil Fertility and Plant Nutrient Management using Back Propagation Neural Networks*" IJRIT CC, vol. 2, Issue 2, 292-297, 2014.