International Journal of Modern Agriculture
ISSN: 2305-7246
Volume 10 Issue 2, 2021

Prediction and Estimation of Crop Yield using Machine Learning

Ediga Vishal[1], Ediga Hari Prasad[2],Gagan V[3], Injeti Samara Simha Reddy[4], Madan H T[5]

[1]Electronics and Communication Engineering, REVA University, (India)
[2]Electronics and Communication Engineering, REVA University, (India)
[3]Electronics and Communication Engineering, REVA University, (India)
[4]Electronics and Communication Engineering, REVA University, (India)
[5]Electronics and Communication Engineering, REVA University, (India)

ABSTRACT
In India, agriculture is the field which gave birth to civilization and also least paid occupation. Agriculture plays on important role in improving economy of our country since it occupies 59% of work force. Hence it is backbone of all business in our country. Selection of crop is important for agriculture productivity selecting suitable crop depends on various parameters such as soil, rainfall, temperature and humidity. To improve yielding and to decrease basics, Machine Learning has come up with different algorithms to predict the crop yielding and it is changing the income scenario by selecting suitable crop. This paper focus on prediction of crop yield through machine learning. This helps farmers to choose suitable crop for maximum yield

Keywords: Crop yield Prediction,machine learning.

I. INTRODUCTION

Forecasting yield of crops will help for farmer in choosing the crop. The main aim of agriculture planning is to make profits by using limited land resources and investment. Many machine learning algorithms have been introduced for crop yielding. There are some factors which influence the crop yielding such as crop seeds, soil, area, temperature, rainfall, humidity etc.., there are two types of predictions.

First type of prediction is done by traditional method which describes the collection of single samples ofsoil, seeds and other parameters to predict the crop yielding. The second type of prediction is by using machine learning techniques.

Machine learning has brought the advancement in crop agriculture methods. Many machine learning algorithms techniques goal is to get maximum yield of crops. The production of crops depends on geographical conditions like temperature, rainfall, cloud and humidity. Many parameters are used to make different predictions.

As the new technologies are introduced and over utilization of non-renewable resources, ecological balance has been disturbed and became inconsistent. This lead to global warming effecting rainfall, temperature and different parameters making farmers ambiguous to predict climate and selecting the crop.

Thus, affecting their crop yield productivity. Unnatural techniques like hybrid varieties are used to avoid losses which lead to environmental harm. Farmers need to know accurate information of crop which they are going to choose. To provide such information many researchers has introduced different approaches in machine learning.

Machine learning consists of different models which can take those parameters as input and gives output which is used for crop yielding to maximum prediction. Prediction of the crop can be performed by different models using artificial neural networks, deep learning and algorithms like mathematical method, statistical methods, exploratory data analysis, regression analysis. In these papers, we use datasets as experimental basis to predict the crop yielding.

II. LITERATURE SURVEY

This paper focus on agriculture parameter neural network technology and parameters which are used as sensor parameters such as PHvalue, N4P1Kvalue.[1]. This method utilizes time sources vegetation to predict the crop yield in advance [2]. The author describes the components which impact yield such as annual rainfall, Area under Cultivation. This harvest yielding was done by using ecological elements and regression analysis [3] Crop yield prediction is done by taking aerial pictures for decision related harvesting [4]. Instead of traditional methods, Artificial Neural Networks gave better outcomes for prediction of soybean lust [5] with the help of neural networks, rice yield prediction produced testing some of 17.3% [6].

In [7], E. Manjulla, S. Djudiltachoumy, have proposed rule-based system to predict the crop yield by using previousdata. They used K means algorithm and clustering methods Ashwani Kumar Kushwaha used one method which helps in predicting the suitable crop for lands. The lands the algorithm was Agra algorithm [8].

Balamurugan implemented crop yield prediction using random forest classifies. Since due to absence of other machine learning algorithms the quality of prediction was mixing [9].

[10] regression analysis method was used to predict rainfall and to investigate reasons for low yield by Raorane A.A, Dr. Kulkarni R.V.

[11] multiple linear regression was implemented to analyze the data by using existing data.

[12] Siti Khairunniza-Bejo, Samihah Mustaffha, Wan Ishak Wan Isma used artificial neural networks to predict crop yield but disadvantage is it consumes more time.

[13] To forecast yield IOT is utilized and Hadoop framework. Nuisances which are generated during manufacturing were eliminated with investigation and forecast classification.

[14] This paper described how Random forest shows effective forecast of yield with high accuracy and to implement they have used k-nearest neighbor and support vector regression (SVG).

[15] This paper explains how to enhance productivity with k-means algorithm, and it is used to forecast the crop yield and also data mining techniques.

International Journal of Modern Agriculture
ISSN: 2305-7246
Volume 10 Issue 2, 2021

III. SYSTEM ARCHITECTURE

To represent all the features gathered through this SLR study, we drew a feature map that shows the significant feature and sub-feature.

- Crop Information: - It refers to information about the crop itself, such as weight, growth during the growth-process, variety of plants, crop density and leaf area index.
- Soil Information: - It consists of the following variables: Soil maps, soil type, pH value, cation exchange capacity (CEC) and area of production.
  In the soil maps, general information about the nutrients in the soil, type of the soil, and location can be found.

- Humidity: - It stands for the water in the field. The features that fall under the humidity group include rainfall, humidity, forecasted rainfall and precipitation.

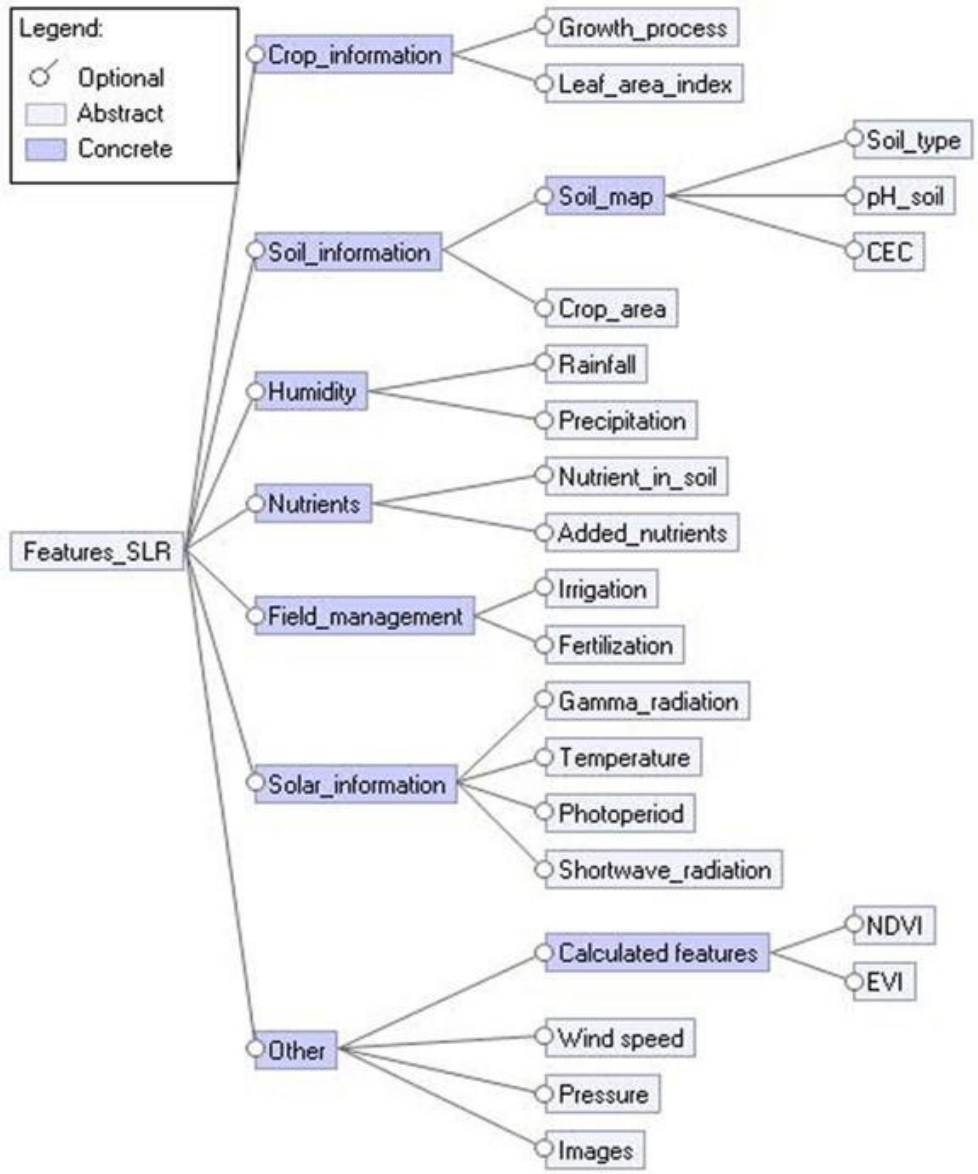

Fig.1. Feature diagram

- Nutrients: - these features measure the level of saturation. The measured nutrients are nitrogen, magnesium, potassium, sulphate, zinc, boron, calcium. These features consist of nutrients-in-soil and added nutrients.

- Field management: - With the field management, It repeats to the management of nutrients and it consists of irrigation, fertilization.

- Solar information: - it contains features related to radiation or temperature. These are gamma radiometric, temperature, Photoperiod, software radiation, degree days solar radiation.

- Others: - It consists of calculated features, wind speed, pressure, images. In this the most feature used as calculated features (Measuring vegetation (Normalized Vegetation Index & Enhanced Vegetation Index)2000). Less features are wind speed, pressure, image.

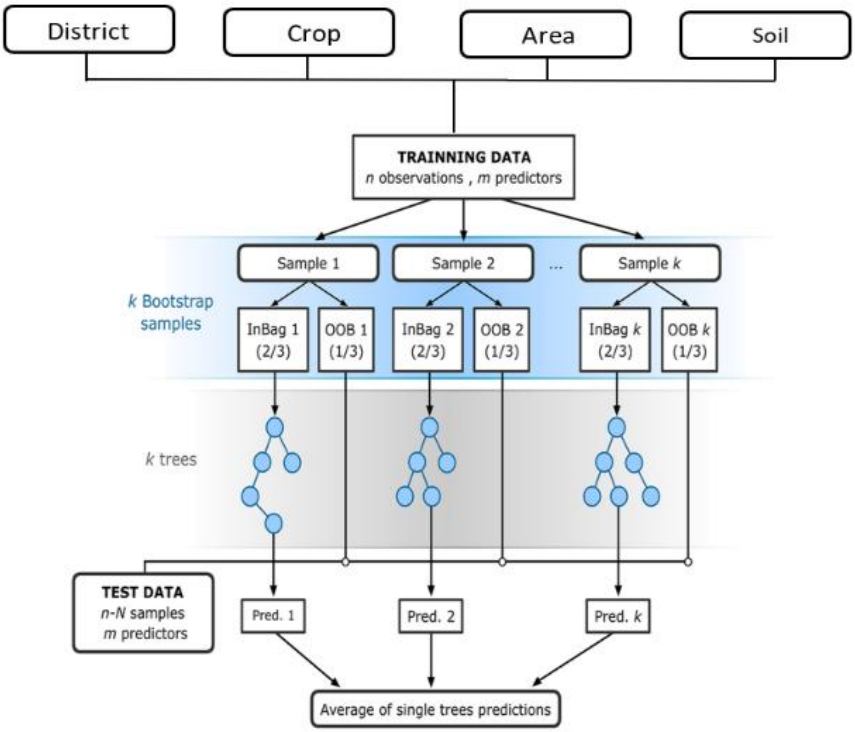

Fig.2. Working of Random Forest Algorithm

A. Random Forest Algorithm

Random Forest is a machine learning algorithm. During the training phase, a large number of decision trees are generated, and the performance is divided into two categories: classification and class prediction (regression). The number of trees is equal to estimation precision. Rainfall, vision, temperature, and production are among the variables in the dataset. This variables in the dataset are included in the training process.

About two-thirds of the dataset is taken into account. The remaining dataset is used as a test bed.

The random forest algorithm has three parameters: n tree, which specifies the number of trees

that must be grown, and m attempt, which specifies how many variables must be taken at each node break. Node size - In terminal nodes, it tellsus how many observations we'll need to make.

B. Decision

Since decision tree classifiers implement greedy methodology, a function selected from the start move cannot be consumed any longer, resulting in the better classification when used in subsequent steps. Similarly, it over fits theplanning data, which may result in bad results with non-obvious data. To get around the confinement gathering paradigm, this method is used.

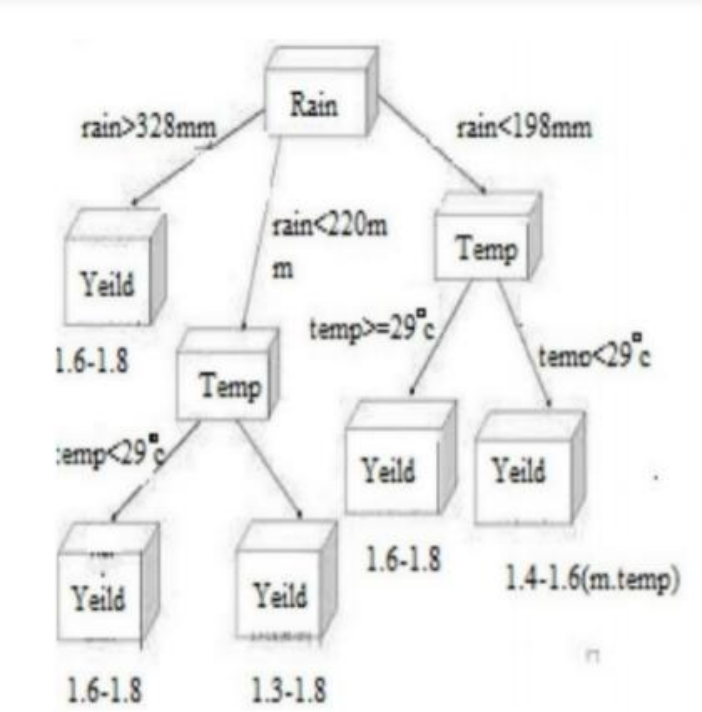

Fig.3. Working of Decision tree

C. Procedure

The dataset contains a few main attributes such as temperature, rainfall, humidity, and pH, and then used to perform classification and regression tasks using the random forest algorithm. Initially, decision trees were used to train the model, but it was discovered that using the random forest algorithm minimized overfitting and improved accuracy, and that using SVR, Random forest, and Random forest improved accuracy.

The dataset that was used was downloaded from Kaggle. The archive 80 percent of the data from the dataset was used to train the algorithm, and 20% was used to validate itresults in order to improve the model's output Using the random forest algorithm Then they were compared. With the initial data set, the expected outcome Later on, using tests, we estimated the model's accuracy.

Similarly, various algorithms were used to estimate the model's accuracy. Finally, we came to the conclusion that the random forest algorithm provides us with greater precision. As a result, the random forest algorithm was used to train the model. Finally, average of the decision trees output can be taken as result

IV. IMPLEMENTATION

Crop yield prediction is a required task in agriculture. It may depend on different area size and different parameters like temperature etc., based on that, we can predict the crop. It is very helpful for farmers in order to estimate the crop without assumptions.

Machine Learning is the subset of artificial intelligence, in this ourselves can get computers to act without being programmed. It is also the study of computer algorithms that can improve their performance through experience and by reading the data.

For this, one and all need to import python libraries like scikit-learn, NumPy, pandas, scikit-image, SciPy, virtualenv, and requests.

All of us are going to implement this using python because it is known for many features.

A. Random Forest Algorithm

Random forest is a supervised learning Algorithm. It is a bunch of decision trees, where majority outputs of the decision trees can be taken as algorithm output. Decision tree is used for classification and prediction it is a decision support tool.

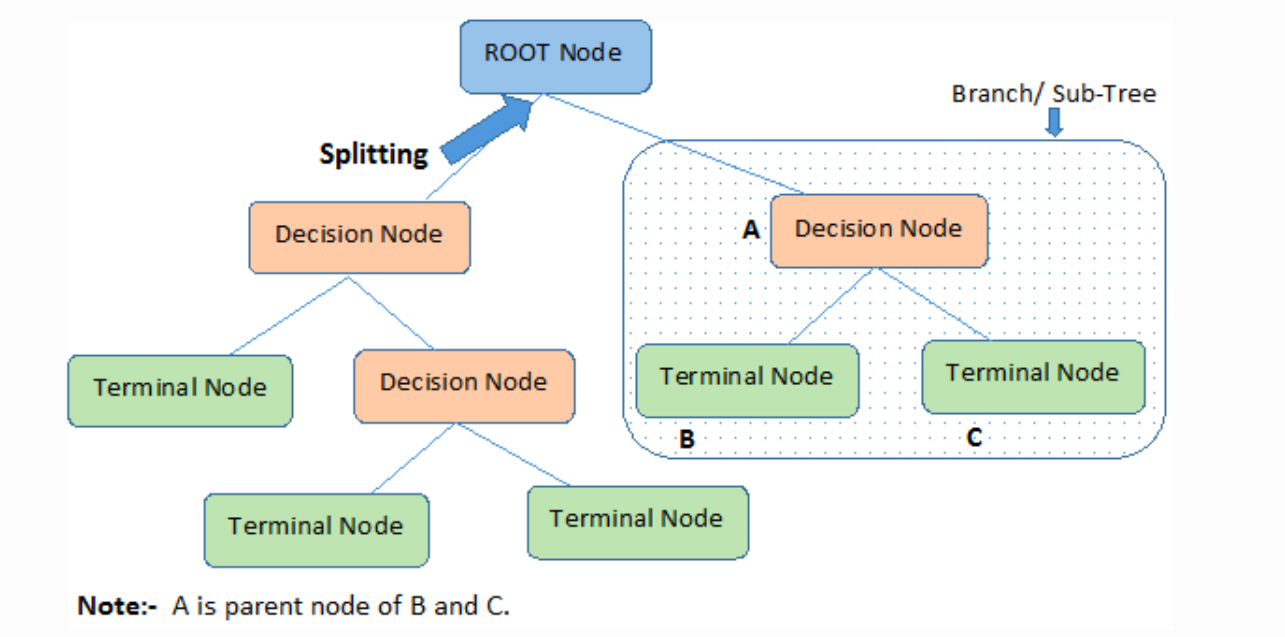

Fig.5. Represents working of decision tree

As each one said already, Random forest algorithm is a set of decision trees. Here is the algorithm of Random Forest

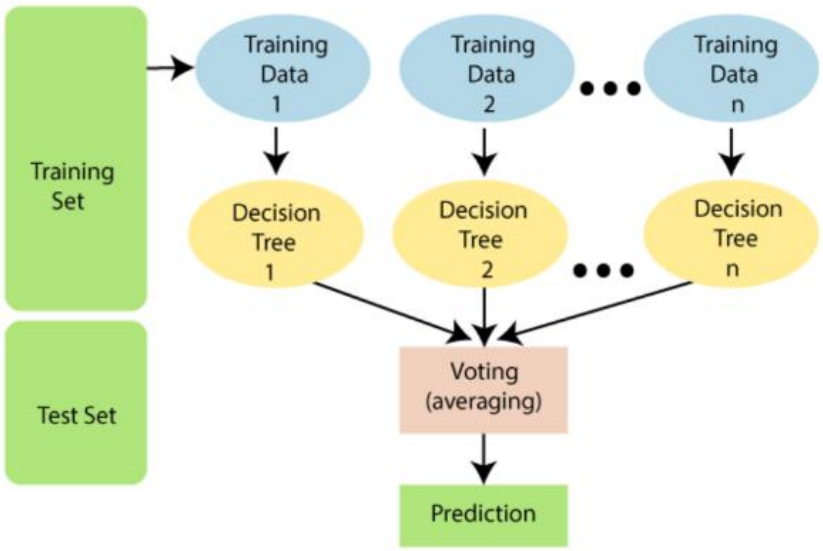

Fig.6.Represents the working of Random Forest Algorithm

As everyone all know that data plays an important role in analyzing the crucial things. Based on the given past data of parameters each one can predict the outcome which we require. So, for this one and all should give data to train the model. Now, we are taking the dataset of Maharashtra state which has 36 districts of diverse land with different parameters like area, crop, precipitation, humidity, and soil type.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Ã¯Â»Â¿Sta	Crop	Year	Area	Yield	Temperat	Precipitatic	Humidity	Soil_type:	Soil_type:	Soil_type:	Soil_type:	Soil_type:	Soil_type:	Soil_type:	District: A	District: A	District: A	District: A	District: A	District: B
2	Maharash		1997	598400	217000	24.243	42.3484	84	0	1	0	0	0	0	0	1	0	0	0	0
3	Maharash		1998	453200	227900	23.1	8.7838	70	0	0	0	0	1	0	0	1	0	0	0	0
4	Maharash		1999	498300	190900	22.8118	39.5078	35	0	0	0	0	1	0	0	1	0	0	0	0
5	Maharash		2000	518700	160800	23.2746	9.2558	33	0	1	0	0	0	0	0	1	0	0	0	0
6	Maharash		2001	528900	175000	23.5674	4.0668	29	0	1	0	0	0	0	0	1	0	0	0	0
7	Maharash		2002	494900	117300	23.5598	28.5188	85	0	0	0	0	1	0	0	1	0	0	0	0
8	Maharash		2003	478300	55100	23.7654	3.5496	87	0	1	0	0	0	0	0	1	0	0	0	0
9	Maharash		2004	555000	263300	23.456	8.3676	20	0	0	0	0	1	0	0	1	0	0	0	0
10	Maharash		2006	528400	274000	22.8787	46.7888	36	0	0	0	0	1	0	0	1	0	0	0	0
11	Maharash		2007	5034	2840	22.7811	20.8014	43	0	0	0	0	1	0	0	1	0	0	0	0
12	Maharash		2008	507200	327800	23.111	14.398	54	0	1	0	0	0	0	0	1	0	0	0	0
13	Maharash		2009	520500	350600	22.3211	35.4694	33	0	0	0	0	1	0	0	1	0	0	0	0
14	Maharash		2010	540000	318000	22.3564	28.1252	52	0	0	0	0	1	0	0	1	0	0	0	0
15	Maharash		2011	434600	123200	23.1456	26.9422	80	0	0	0	0	1	0	0	1	0	0	0	0
16	Maharash		2012	444800	67000	23.6523	32.6042	79	0	0	0	0	1	0	0	1	0	0	0	0
17	Maharash		2013	503900	354500	23.1234	9.5166	40	0	0	0	0	1	0	0	1	0	0	0	0
18	Maharash		2014	443600	279800	22.751	24.0944	79	0	0	0	0	1	0	0	1	0	0	0	0
19	Maharash		1997	3400	900	23.1122	5.213	54	0	0	0	0	1	0	0	0	1	0	0	0
20	Maharash		1998	1100	400	22.991	27.6824	34	0	0	0	0	0	0	1	0	1	0	0	0
21	Maharash		2000	200	100	23.2296	9.5258	36	0	0	0	0	0	0	1	0	1	0	0	0
22	Maharash		2001	100	100	23.8574	9.259	50	0	0	0	0	0	0	1	0	1	0	0	0
23	Maharash		2002	100	100	24.0966	5.223	47	0	0	0	0	0	0	1	0	1	0	0	0
24	Maharash		2003	100	100	23.9084	28.5188	30	0	0	1	0	0	0	0	0	1	0	0	0
25	Maharash		2004	100	100	23.8456	3.4646	48	0	0	1	0	0	0	0	0	1	0	0	0
26	Maharash		2006	100	100	22.989	8.7726	24	0	0	0	0	0	0	1	0	1	0	0	0

Fig.7. dataset of Maharashtra Districts

A Random Forest ML model trained on data from the previous 20 years predicts the expected crop yield.

Random Forest Algorithm can accommodate massive data sets of greater dimensionalities. It Predicts incomplete data and preserves precision and reminds mathematical moderators of a black box solution (less parameter control).

B. Web Application

We are developed a web application where each one can give inputs to the parameters like districts, area, crop, and soil.

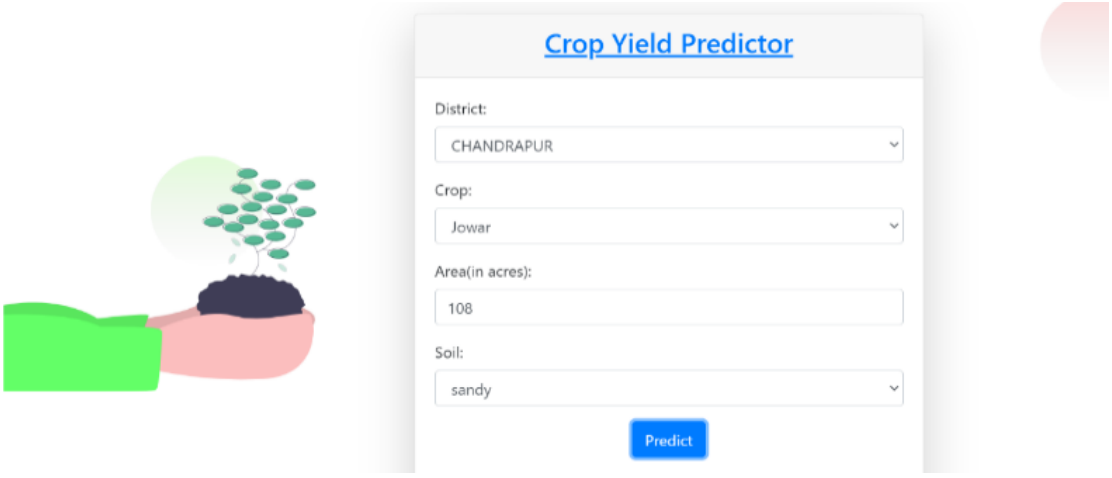

Fig.8.web application for crop prediction

V. RESULT AND DISCUSSION

Hence, after all giving input parameters like districts, crop, area required for crop and soil type we can predict how much quantity of yield can get finally based on the parameters.

This focuses on yield estimation by district based on the crop sown in the district. Yield is forecast for given crops district-by-district, and the crops with the best yield are recommended to expand in the future so that farmers can get the most out of the study.

Production/Area = Yield

Current weather information, such as temperature, humidity, and precipitation, is automatically retrieved from the internet and used for forecasting.

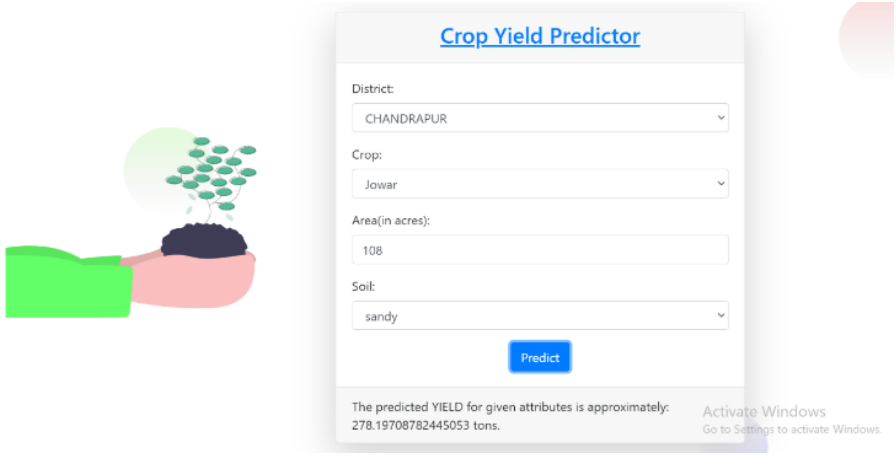

Fig.9. Shows the predicted yield.

A simple Web application created to offer farmers/users an estimate of how much crop yield will be achieved based on the given input.

For prediction, the software uses a Random Forest Machine Learning model that was educated on over 20 years of data from 30 Maharashtra districts, as well as automated live weather fetching. The model has an accuracy of about 86% and can be improved even further with more details.

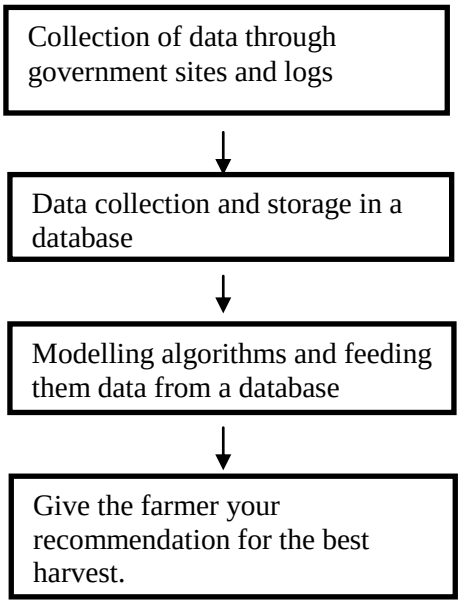

Fig.10. Crop Prediction Flow chart

It illustrates how data flows in the scheme, from data collection to data compilation to eventually suggesting the best crop to farmers, by feeding data into a database and modelling it using different algorithms.

With this result all of us can predict the accurate result of the crop and also, we can predict the loss in the market.By this, the farmers can get to know about their respective crop yields

VI. CONCLUSION

Because of temperature, rainfall, and location, the paper proposed a machine learning algorithm for predicting crop yield. Experiments were conducted on an Indian government dataset to predict yield for a specific district by combining rainfall, temperature, and other parameters such as season and region.

This will not only assist farmers in selecting the best crop to grow in the coming season, but it will also help to close the divide between technology and agriculture.

REFERENCES

[1] Y. A Pachepsky, D.Timlin and G.Varallyay, "Artificial neural networks to estimate soil water retention from easily measurable data". Soil Science Society of America Journal, vol 60(3).pp,727.733, 1996.

[2] Su,Tao Shao Huan Feng and Xing Yuan cui, a regional yield Estimation for Spring Maize with Multi-Temporal Remotely Sensed Data in Junchuan, China. "Advanced Materials Research vol

610. Trans Tech Publications, 2013.
[3] Monali Paul, Santhosh K.Vishwakarma, Ashol Varma, "Analysis of soil Behaviour and Prediction of Crop Yield using Data Mining Approach" 2015.
[4] CC.Yang, S.O.Prastier, S.Sreekanth, N.K Patni and L.Masse "An artificial neural network model for simulating pesticide concentration, in soil, Transactions of the ASAE, vol 37 pp 981-988, 1994.
[5] W.D.Batchelor, X.B.Yang and A.T.Tshanz, "Development of a neural network for soyabean rust epidemics". Transactions of the ASAE, vol 40 pp 247-252, 1997.
[6] W.Ji and Jicui, "Application of Geotableical Information System (GIS) in Agricultural Land Classification and Grading". Soil International Conference on Agricultural and Natural Resources and Engineering Advance in Biomedical Engineering, 3(5),
[7] S.Djodiltachoumy, A model for prediction of crop yield, International Journal of Computational Intelligence and Informatics, March 2017.
[8] Ashwani Kumar Kushwaha, Sweta Bhattchrya, Crop yield prediction using Ago Algorithm in Hadoop, April 2015.
[9] Balamurugan, M.Priya, P.Muthaiah, International Journal of Engineering Sciences Research Technology Prdicting Yield of the Crop Using Machine Learning Algorithm.
[10]. Raorane A.A, Dr. Kulkarni R.V, Application Of Datamining Tool To
Crop Management System, January 2015.
[11] .D Ramesh , B Vishnu Vardhan, Analysis Of Crop Yield Prediction
Using Data Mining Techniques, International Journal of Research in
Engineering and Technology, Jan-2015.
[12]. Siti Khairunniza-Bejo, Samihah Mustaffha and Wan Ishak Wan Ismail, Application of Artificial Neural Network in Predicting, journal of Food Science and Engineering, January 20, 2014.
[13].https://en.wikipedia.org/wiki/Data_analysis.
[14] ANTIBIOTIC RESISTANCE PATTERN OF UROPATHOGENIC ESCHERICHIA COLI ISOLATES OF SUSPECTED UTI PATIENTS, Yahya Ali, Arpita Sharma, International Journal Of Advance Research In Science And Engineering http://www.ijarse.com IJARSE, Volume No. 10, Issue No. 02, February 2021 ISSN-2319-8354(E).
[15]. Jig Han Jeong, Jonathan P. Resop, Nathaniel D. Mueller, David H. Fleisher, Kyungdahm Yun, Ethan E.Butler,Soo-Hyung Kim. Random Forests for Global and Regional Crop Yield Predictions. Institute on the Environment, University of Minnesota, St. Paul, MN 55108, United States of America.
[16]D Ramesh, B Vishnu Vardhan. Analysis Of Crop Yield Prediction Using Data Mining Techniques. IJRET:
International Journal of Research in Engineering and Technology.