Crop Yield Prediction And Efficient Use Of Fertilizers

Sneha B R[1], Srushti H N[2], Srushty[3], Sushmitha V[4], Dr. Sumithra Devi K A[5]

[1]BE Student, Department of Information Science and Engineering, Dayananda Sagar Academy of Technology and Management, Karnataka, India
[2]BE Student, Department of Information Science and Engineering, Dayananda Sagar Academy of Technology and Management, Karnataka, India
[3]BE Student, Department of Information Science and Engineering, Dayananda Sagar Academy of Technology and Management, Karnataka, India
[4]BE Student, Department of Information Science and Engineering, Dayananda Sagar Academy of Technology and Management, Karnataka, India
[5] Dean Academics & Head, Department of Information Science and Engineering, Dayananda Sagar Academy of Technology and Management, Karnataka, India

ABSTRACT

As we all know that India being an agricultural country it predominantly depends on the growth in agriculture field and also the agro-industry products. The emerging research field in analyzing the crop yield by the Data mining technique and has helped the sector growth drastically. In the recent years farmers are interested in knowing the yield that he/she is about to expect. As a result we analyze the various related attributes such as state name, district name, crop year, season, crop name, soil moisture, humidity, temperature and area, along with it, the percentage of nutrients like Nitrogen (N), Phosphorous (P), and Potassium (K) are considered. Later all the above attributes are analyzed and then trained using various machine learning algorithm such as Random forest regression algorithm for creating the model. Finally the model is to be precise and accurate in predicting the yield to be expected for the particular crop grown and also the fertilizers to be added to the crops to enhance the crop yield and increase the farmers' revenue. Here we are developing a GUI application for the sake of farmers, where this application can be used on any computers and can be used to identify the crop yield. Through this application the framers can get there results in no time and is very helpful in increasing their crop yield by growing proper crop in particular seasons and adding sufficient fertilizers required.

Keywords: Crop yield prediction, Fertilizer Prediction, Random forest algorithm, Back propagation, Linear regression, KNN regression, Decision tree regression.

I. INTRODUCTION

As we all know that Indian economy mainly depends on agricultural sector, but in recent years due to various reasons like sudden climatic change or due to lack of fertilizers there has been drastic decrease in the production rate. So here the major aim is to introduce the fundamental approach of machine learning system in the agricultural process.

Through this approach one can expect the predictions with better accuracy and efficiency, as this mainly helps the formers in getting better yield and increase their revenue. Indian agriculture is characterized by mainly soil, temperature, rainfall, and cropping system. There are various other problems in Indian agricultural system, the use of poor quality seeds, neglected crop rotation, poor in soil nutrients, cropping pattern, inadequate usage of fertilizers

these are the major problems facing by the farmers**[4]**. Since independence, India has made huge progress towards achieving food security.

The latest studies have shown a steady decline in the contribution made by agriculture to the Indian economy although it is demographically the broadest economic sector and plays an important role in the entire socio-economic fabric of India**[2]**. So, keeping all these things in mind we are planning the best for the agricultural field. Many researches and calculation are made on the agricultural planning where the goal is to get an efficient and accurate model for the crop yield prediction. Crop production rate is mainly depends on cartography of the particular region eg. (hill area, river side area), weather condition of that region eg.(temperature rainfall, cloud), type of soil eg. (sandy, clay, saline soil) soil configuration(ph, nitrogen, potassium value, sulphur, copper, iron value) and harvesting method. Some of the prediction model for crop production rate is studied through the research.

Prediction models are classified into two levels: first is traditional statistics model which formulates a single predict function which holds a whole sample space i.e. it generates a model over whole sample space which is developing technology for knowledge that relates to input output variables which is difficult to obtain statically. In evaluating the factors that affect the crop growth that soil moisture, the availability of soil nutrients are the factors that limit the crop growth and hence limit the crop yield**[6]**.

The purpose of crop yield prediction is to estimate production in agriculture sector for better crop management and make strategic decision for improving crop yield in future. Changes in climate like temperature humidity and rainfall also leads to limit the crop yield. The challenge is to build the model that gives the accurate efficient crop yield. So, try with the algorithms and compare then select the suitable algorithm and predict the yield of the particular crop. Machine learning algorithm will build a mathematical model from a group of data that contains both input and output. In order to perform efficient prediction and handle climatic condition in machine learning algorithms such as Random forest, Liner Regression etc. to be applied to get a pattern. In past many researches various machine learning algorithms has been implemented to enhance the agricultural growth of the country.

The main aim of the paper is to improve the crop yield in several ways and recommend suitable fertilizer for the particular crop grown**[5]**. We have mainly focused on active utilization of machine learning algorithms and its quantifications that helps the farmers in determining the crop yield based on land area, rainfall, temperature, district and various other attributes. Comparing the predicted result of the different algorithms and determining the suitable approach is more satisfactory

II. EXISTING METHODOLOGY

India being agricultural country is mainly dependent on summer rainfall. The relation between summer and rainfall and their results has been studied[1]. The paper represents an approximation of crop-climate relationship based on the previous crop data. Relationship status says us that the Pacific and Indian Ocean sea-side temperatures, monsoon rainfall and Darwin level of the sea pressure is directly proportional to the crop production in India. Finally output says that the production at the state level and sub divisional monsoon rainfall are constant with respect to the all over India result. However, the effect of sub divisional monsoon rainfall in connection to EI Niosouthern oscillation and the sea-side temperature of the Indian ocean has an greater impact on western to central peninsia.

Niketa et al 2016 [1] has indicated that the crop yield depends on the seasonal climate. In India climatic conditions are very unconditional and so that in the time of drought, farmers face various problems. So considering all this into the consideration, using various machine learning algorithms helped the farmers predicting the crop suitable for the better yield. To classify the results they used SMO classifier in WEKA. The main attributes that they take into consideration are minimum temperature, maximum temperature, and average temperature and previous data that are related to crop and yield information. Using SMO tool the previous data has been classified into two classes they are high yield and low yield. The obtained results for crop yield predictions using SMO classifier gives less accuracy when compared to naïve bayes, multilayer percptron and bayesian network.

Eswari et al 2018 indicates that the crop yield depends on the perception, average, minimum and maximum temperature. Evapotranspiration is the other function that they have taken into the consideration.[2] The crop Evapotranspiration is the function of both growth and weather stage of the plants. The above attributes are taken into consideration to get the better decisions on yield of the crops. They then collected the dataset based on the above

attributes and sent them as the input to the Bayesian network and these inputs had been classified into two classes namely true and false classes and the results were compared with the observed classifications in the model with the confusion matrix to bring out the accuracy. Finally, they concluded saying prediction of the crop yield with Naïve Bayes and Bayesian network gives out high accuracy when compared to SMO classifier and forecasting the crop yield prediction in different climatic conditions and cropping scenarios would be more beneficial.

Shruti Mishra et al 2018, has mentioned that on applying Data Mining techniques on past climate and crop production data, and predicting data would enhance the crop productivity.[3] For the formers to take correct decisions on the crop to be cultivated the Decision Support System has to be implemented. They had collected the dataset with various attributes such as season, area and production in hectares and these attributes were analyzed through various algorithms in WEKA. They analyzed the data through four methods and then compared the results among each other. The four used in WEKA are J48, IBK, LAD tree, LWL. They finally stated that the IBK method has the greater accuracy when compared to the other methods and it depends on the nature of the dataset and also the nature type the crop grown.

Chlingaryana et al 2017 mentioned that the nitrogen level is the major factor in predicting the crop yield. In decision making, nowadays remote sensing systems are mostly used. The data collected through this remote sensing are more useful to the farmers in predicting their crop yield.[4] The large amount remote sensing data is used to make the final decision. They stated that to improve the crop yield and soil fertility nitrogen in mainly used. Here Machine Learning algorithms are used to make the major decisions. Important attributes that were considered here are nitrogen level in the soil, type of soil, yield analysis comparing with the previous data, which all helped to predict the crops to be grown and enhance the crop yield. Later to predict the future data conventional neural network of the long term memory was used.

Dakshayini Patil et al 2017[5] has indicated that, in the Indian economy the rice crops played a major role. They used various Data Mining techniques to predict the yield of the rice crops. Rice crops are the sustainable security of India. Commonly, it will give 40% to the normal yield. And the great yield of the crop depends on the proper climatic factors According to the climatic situation making a good plan to grow the crop can increase the yield of the crop.The records will use different skills of mining based on the existing data of the yield of the crop and the various climatic regions. In this, the writer utilized the data of the 27 regions of the Maharashtra for predicting the crop yield. Weighted gifted to its yield performing. One category of the structure sees the center as “Artificial neurons”.

And that is called as neural systems. The back cause calculation (Rumelhart and McClelland, in 1986) that is used in the layered feed-forward ANNs. This indicates that the counterfeit neurons are arranged in the layers and it will send a signal as “forward”, and after that the errors will spread in backwards[7].

The system will get a present by the neurons in the info layer, and the yield of the plan is given by the neurons on the layer of the yield. There may be minimum one center of the road concealed layers. This neural will sort the engineering in the serious mainstream, since it is very good it might be associated to a broad range of venture. The primary word, "feed forward" describes how this neural sort the procedure and analysis the design. In a feed forward neural system, neurons are exactly connected with ahead. Every layer of the neural structure contains connection with the following layer until there are no connections back. The equation back spread represents how this type of neural structure is made [8].

A back spread is a kind of cope up preparing. The system should be given both example data and foreseen yields in order to use the cope up preparing plans. The foreseen yields are considered against the original yields for specified information. Using the foreseen yields, the back proliferation will start preparing calculation at that point which takes a resolute error and it will replace the loads of the various layers in backward from the yielding layer to the info layer [9].

III. PROPOSED METHODOLOGY

Here we are using various algorithms such as random forest, back propagation, linear regression, KNN, decision tree regression algorithms in predicting the crop yield and finally on comparing the accuracy obtained through these algorithms we use the algorithm that gives the greater accuracy and efficiency for final predictions. Later we are predicting the fertilizers required to enhance the crop yield using random forest algorithm.

a) Data Set Description

We have collected two different datasets through kaggle website, one for crop yield prediction and the other for fertilizer prediction. Crop dataset contain attributes such as state name, district name, crop year, season, crop name, temperature, humidity, soil moisture, area and production. Fertilizer dataset contain attributes like temperature, humidity, moisture, soil type, crop type and considers the nutrients value present in the soil that is the amount of Nitrogen, Potassium, and Phosphorous are taken into consideration.

Finally on taking all the input attributes mentioned above the model undergoes through various algorithms for training and testing process. Later the accuracy of the model is checked and lastly the model comes to be precise and accurate in predicting the crop yield and also the fertilizer name for the particular crops grown[1]. This model maily helps the framers get better yield and also in increasing their revenue.

Sample crop prediction data:

	A	B	C	D	E	F	G	H	I	J
1	State_Name	District_Name	Crop_Year	Season	Crop	Temperatu	humidity	soil moistu	area	Production
2	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Arecanut	36	35	45	1254	2000
3	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Other Kharif pulses	37	40	46	2	1
4	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Rice	36	41	50	102	321
5	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Banana	37	42	55	176	641
6	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Cashewnut	36	40	54	720	165
7	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Coconut	34	45	52	18168	65100000
8	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Dry ginger	34	55	62	36	100
9	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Sugarcane	35	50	59	1	2
10	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Sweet potato	25	55	55	5	15

Table-1: Sample crop prediction dataset.

Sample fertilizer data:

1	Temparatu	Humidity	Moisture	Soil_Type	Crop_Type	Nitrogen	Potassium	Phosphorc	Fertilizer_Name
2	26	52	38	Sandy	Maize	37	0	0	Urea
3	29	52	45	Loamy	Sugarcane	12	0	36	DAP
4	34	65	62	Black	Cotton	7	9	30	14-35-14
5	32	62	34	Red	Tobacco	22	0	20	28-28
6	28	54	46	Clayey	Paddy	35	0	0	Urea
7	26	52	35	Sandy	Barley	12	10	13	17-17-17
8	25	50	64	Red	Cotton	9	0	10	20-20
9	33	64	50	Loamy	Wheat	41	0	0	Urea
10	30	60	42	Sandy	Millets	21	0	18	28-28

Table-2: Sample fertilizer prediction dataset.

The above mentioned datasets are considered in predicting the crop yield and also the fertilizers to be added.

b) Architecture Diagram:

The Fig. 1 shown below represents the architecture diagram.

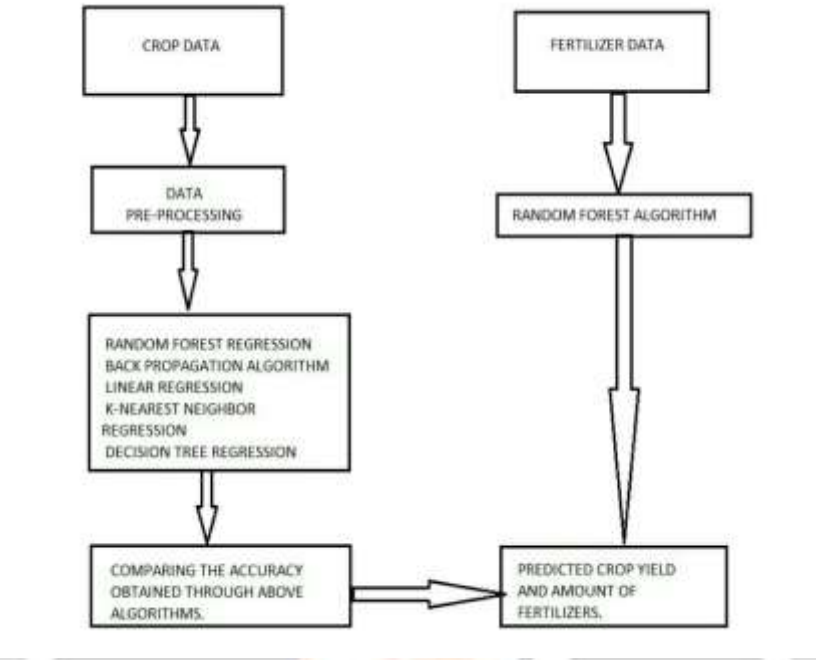

Fig-1: Architecture Diagram

This model takes inputs as state name, district name, crop year, crop name, season, soil moisture, humidity, temperature and area. Using these inputs the model undergoes through various algorithms like random forest, back propagation, linear regression, KNN regression and decision tree regression through which the prediction of the crop yield is done. Then considering the amount of nutrients like nitrogen, phosphorous and potassium the fertilizer to be added to enhance the crop yield could be predicted.

c) Data Pre-Processing:

In general the raw data may be incomplete, inconsistence or empty. Thus pre-processing includes data cleaning, integration, transformation and reduction. Data anomalies are not taken into account for future classification and prediction calculation[4]. Here all the string values get converted to the integer values. So, that the data is easily understandable by machine. In this pre-processing, we use label encoder for converting all the labels like state name, district name, season and crop name to integer values.

Then these data will be moved further and are divided into the train and the test data. For this splitting of the data into train and test we need to import train_test_split through the scikit-learn this will help the pre-processed data to split into train and test according to the given weights.

d) Random Forest Algorithm:

The Random Forest algorithm gives an efficient prediction for both smaller and larger datasets. This algorithm forms various decision trees for the given dataset and for the classification problems it considers the majority votes as the final prediction where as in regression problems it considers the mean value of all the decision trees formed and projects the final output. As explained below in fig.2 in this model the dataset has been divided into training set and testing set and this split of dataset happens in the ratio for 20:80 respectively. Greater the decision tress formed greater the accuracy of the algorithm. Based on the various conditions the decision trees which are formed by the algorithm are analyzed and trained to reach the better accuracy.

Fig-2: Working of random forest algorithm

Back Propagation algorithm:

Here the algorithm attempts to find the minimum error function by using the methods like gradient descent that sends the signals "forward", which then back propagates the errors occurred. As shown in the below figure this process undergoes three layers, firstly the input layer that is responsible for providing the inputs to the model, next comes the hidden layer, this layer gets the output from the input layer and calculate the weights present on input to hidden layer and projects the desired output. Lastly the output layer which predicts the output obtained through the network model.

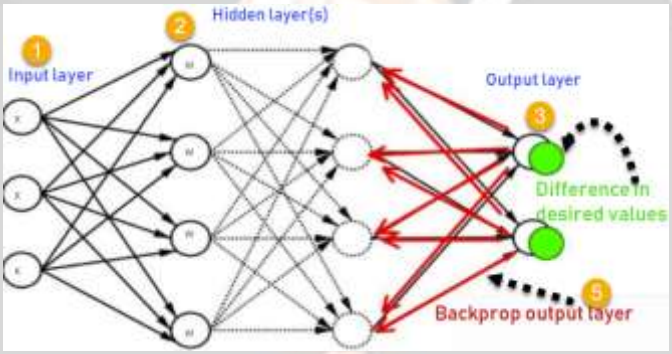

Fig-3: Working of back propagation algorithm

Linear Regression algorithm:

This algorithm performs tasks to predict the dependent variable y on the basis of the independent variable x. So, the algorithm finds the linear relationship between the input(x) and output(y). Here the regression line represents the best fit line for the model.

K-Nearest Neighbors algorithm:

KNN algorithm basically uses 'feature similarity' to predict the new data point values. This says that the new point is assigned a value based on how closely it resembles the points in the training set. Finally it calculates the average of the numerical target of the K nearest neighbors.

Decision Tree Regression algorithm:

Decision tree algorithm breaks down the dataset into smaller and smaller subsets while at the same time the associated decision tree is incrementally developed. Here the algorithm observes the features and trains the model according to the features in the structure of a tree to analyze the final predictions. As mentioned in fig.4 the final tree contains decision nodes and leaf nodes. A decision node has two branches, each representing the values for the tested attributes. Leaf node represents a decision on the numeric target.

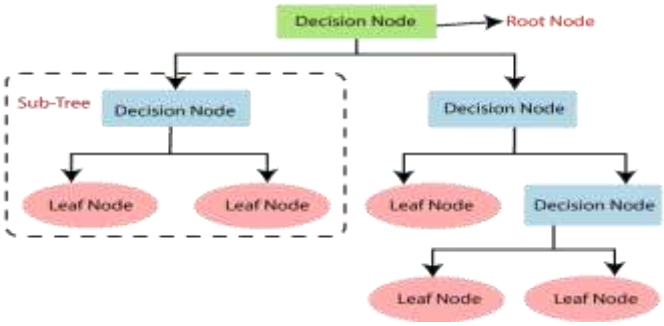

Fig-4: Working of decision tree regression algorithm

IV. RESULTS

Crop Yield Prediction

The model is trained with dataset provided in csv file. The model undergoes train_test_split with the ratio of 80 and 20 percent respectively. Here, we have presented a model that works on random forest algorithm for predicting the final crop yield for the given input values. The final accuracy obtained through this algorithm is about 97%, hence we conclude saying random forest is the efficient and best algorithm for predicting the crop yield when compared with the various other algorithms like Back Propagation, Linear Regression, K-nearest neighbor Regression, Decision Tree Regression algorithms. Below figure represents the accuracy check between various algorithms and the graph plot against them.

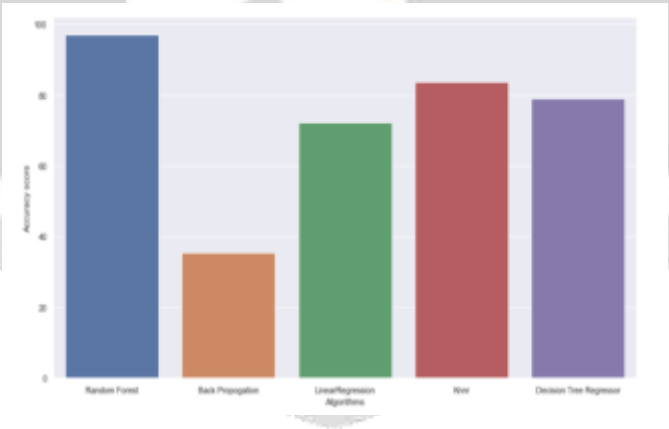

Fig-5: Accuracy check between various algorithms.

The below figure shows the results, on providing the given input attributes the model results in predicting the final expected crop yield.

Fig-6: Crop Yield prediction model based on random forest algorithm.

Basic measures derived from the algorithm:

Mean Squared Error (MSE): It is the average of the squared error that is used as the loss function for least squares regression.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y})^2$$

R2_Score: R-squared is a statistical measure that represents the goodness of fit of a regression model. The ideal value of r-square is 1.The closer the value of r-square to 1, the better is the model fitted.

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=0}^{n_{\text{samples}}-1}(y_i - \hat{y}_i)^2}{\sum_{i=0}^{n_{\text{samples}}-1}(y_i - \bar{y})^2}$$

$$\text{where } \bar{y} = \frac{1}{n_{\text{samples}}}\sum_{i=0}^{n_{\text{samples}}-1} y_i .$$

Fertilizer Prediction

The model is trained with dataset provided. Here the model undergoes train_test_split with the ratio of 80 and 20 percent respectively. Here the Random forest regression algorithm is used to predict the fertilizers to be added.

Below figure shows the application page for fertilizers prediction.

Fig-7: Fertilizer prediction model based on random forest algorithm

V. CONCLUSION

Crop yield prediction and efficient use of the fertilizers is successfully predicted and also found the best algorithm that is efficient in predicting the crop yield that is the random forest algorithm with the accuracy of 97%. In future, developing the web application based on this ideology helps the farmers to take the right decision for right crop such that the agricultural sector could be developed by innovative ideas.

VI. REFERENCES

[1]Niketa Gandhi et al, "Forecasting of rice crop yield Tropical Dry and Wet Climatic Zone of Indian Climate Zone Using Data Mining Technologies", IEEE International Conference on Advances in Computer Contribution (ICACA), 2016.

[2]K.E. Eswari. L.Vinitha. "Yeild prediction of crop Tamil Nadu Using Network bayesian", Advancements of international journal and Engineering research Computations, Volume-6, Issue-2, ISSN: 23468- 2078.

[3]Shruti Mishra, Priyanka Paygude, Snehal Chaudhary, Sonali Idate "Use of Data Mining in Crop Yield Prediction"IEEE Xplore Compliant - Part Number:CFP18J06-ART, ISBN:978-1-5386-0807-4; DVD Part Number: CFP18J06DVD , ISBN:978-1-5386-0806-7.

[4]Anna Chlingaryana, Salah Sukkarieha, Brett Whelanb Machine learning approaches for yield of the crop prediction and nitrogen position evaluation in precision agriculture: A analysis, Computers and Electronics in Agriculture 151 (2018) 61–69, Elisver, 2018.

[5]Dakshayini Patil et al, "Rice Crop Yield Prediction using Data Mining Techniques: An Overview", International Journal of Advanced Research in Computer Science and Software Engineering ,Volume 7, Issue 5, May 2017.

[6]Shakil Ahamed, Navid Tanzeem Mahmood, Nazmul Hossain, Mo Faridur Rahman, Rashedur M Rahman "Applying Data Mining Technology to Predict the Annual crop Yield of more Crops and Recommended to Planting more crops in various places in pune".

[7]Schlkopf B," A seminaronSVM(support vector regression)".Stat Computation 14(3): 198-266, 2005.

[8]K.Menaka, N.Yuvaraj,Indian Journal of Innovations and Developments Vol 5(12), October, 2016 ISSN (online) : 2277-5390 ISSN (Print): 22775382

[9]Zhang L, Zhang J, Kyei-Boahen S, Zhang M, Simulation and soybean growth prediction and enlargement of lower field position". Am-Euras J Agr Environ Sci 7(4): 374-385, 2010.

[10]Quinlan JR, "Learning with continuous classes". Proc. AI92, 5th Aust. Joint Conf. on Artificial Intelligence and its application about yield prediction (Adams & Sterling, eds.), World Scientific, Singapore, pp: 343-348, 1992.

[11]D Ramesh, B Vishnu Vardhan, "Analysis of crop yield prediction using data mining techniques", International Journal of research in engineering and technology,eISSN:2319-1163.