Crop Yield Prediction and Efficient use of Fertilizers

S.Bhanumathi, M.Vineeth and N.Rohit

***Abstract*—India being an agriculture country, its economy predominantly depends on agriculture yield growth and agro-industry products. Data Mining is an emerging research field in crop yield analysis. Yield prediction is a very important issue in agricultural. Any farmer is interested in knowing how much yield he is about to expect. Analyze the various related attributes like location, pH value from which alkalinity of the soil is determined. Along with it, percentage of nutrients like Nitrogen (N), Phosphorous (P), and Potassium (K) Location is used along with the use of third-party applications like APIs for weather and temperature, type of soil, nutrient value of the soil in that region, amount of rainfall in the region, soil composition can be determined. All these attributes of data will be analyzed, train the data with various suitable machine learning algorithms for creating a model. The system comes with a model to be precise and accurate in predicting crop yield and deliver the end user with proper recommendations about required fertilizer ratio based on atmospheric and soil parameters of the land which enhance to increase the crop yield and increase farmer revenue.**

***Index Terms*—Artificial neural network, Random forest algorithm, Backpropagation algorithm, Prediction.**

I. Introduction

INDIA is a highly populated country and randomly change in the climatic conditions need to secure the world food resources. Framers face serious problems in drought conditions. Type of soil plays a major role in the crop yield. Suggesting the use of fertilizers may help the farmers to make the best decision for their cropping situation [1]. The number of studies Information and Communication Technology (ICT) can be applied for prediction of crop yield [2]. By the use of Data Mining, we can also predict the crop yield. By fully analyze the previous data we can suggest the farmer for a better crop for the better yield [3].

Dr.S.Bhanumathi is an Assistant Professor with School of Computing, Sathyabama Institute of Science and Technology, Chennai. (email: banujun8@gmail.com)

M.Vineeth and N.Rohit are U.G Student with the Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai vineethmothukuri123@gmail.com and rohitnarra2@gmail.com

For the better yield we need to consider soil type and soil fertility things and also one of the major factors rainfall and groundwater availability if it is dry land it is better to go for cash crops and if is wetland it is better to go for wheat and sugarcane[4]. There are15 agro-climatic regions in India these regions are divided on the bases of a type of the land. Each agro-climatic regions can grow some specific crops. Based on that we need to suggest the farmer that which crop is best among those crops which belong to those climatic regions. Achieving the maximum crop at minimum yield is the ultimate Aim of the project. Early detection of problems and management of that problems can help the farmers for better crop yield[5]. Crop yield prediction is the important research which helps to secure food[6,7]. For the better understanding of the crop yield, we need to study of the huge data with the help of machine learning algorithm so it will give the accurate yield for that crop and suggest the farmer for a better crop.

Improving the quantity of the crop is the key goal of precision agriculture means obtaining a better understanding of the crop using the information technology methods. The main goal of precision agriculture is profitability and sustainability [8].

From ancient times agriculture has become the backbone of our country. Nowadays climatic conditions vary very often. So, it is hard to grow crops by understanding weather conditions[8]. We need to use some technology to find or understand the crop details and guide the farmers to grow crops accordingly and moreover fertilizer also one of the major factors to grow crops accordingly[9]. If fertilizer is used more or less in the field the soil may lose it fertility and crop may not give the expected yield. so, fertilizer also becomes the major factor in it [10]. mostly understanding the temperature conditions is much necessary for India because we can improve the Indian economy with the help of the crop prediction because it plays a major role in the Indian economy.

Generally, machine learning algorithms will predict the most efficient output of the yield [11]. Previously yield is predicted on the bases of the farmers prior experience but now weather conditions may change drastically so they cannot guess the yield [12]. so, technology can help them to predict the yield of the crop weather to go for that crop or no. machine learning model will understand the pattern of the crop and yield based on the several conditions and predicts the yield of the area in which he is going to crop [13].

978-1-5386-7595-3/19/$31.00 ©2019 IEEE

The challenge in it is to build the efficient model to predict the most efficient model to predict the output of the crop so try with the different algorithms and compare all the algorithms and which one has the less error and loss chose that model and predict the yield of that particular crop. From this paper, u can see the comparison of the two algorithms and predicting the output from the best model in those two[14-17].

The rest of the paper is organized as below. The related work, proposed system are described in Section II and III respectively. Section IV discuss about the results. Section V concludes the paper with conclusion and future work.

II. Related Work

Niketa et al 2016 [1] have indicated that the yield of the crop depends on the seasonal climate. In India, climate conditions vary unconditionally. In the time of drought, farmers face serious problems. So this taken into consideration they used some machine learning algorithms to help the farmers to suggest the crop for the better yield. They take various data from the previous years to estimate future data. They used SMO classifiers in WEKA to classify the results. The main factors that take into consideration are minimum temperature, maximum temperature, average temperature, and previous year's crop information and yield information. Using SMO tool they classified the previous data into two classes that are high yield and low yield. The obtained result for the crop yield prediction using SMO classifier gives less accuracy when compared to naïve Bayes, multilayer perceptron and Bayesian network.

Eswari et al 2018 [2] have indicated that yield of the crop depends on the perception, average, minimum and maximum temperature. Apart, from that, they have taken one more attribute named crop evapotranspiration. The crop evapotranspiration is a function of both the weather and growth stage of the plant. This attribute is taken into consideration to get a good decision on the yield of the groups. They all collected the dataset with these attributes and send as input to the Bayesian network and classify into the two classes named true and false classes and compared with the observed classifications in the model with a confusion matrix and bring the accuracy. Finally, they concluded that crop yield prediction with Naïve Bayes and Bayesian network give high accuracy when compared to SMO classifier and forecasting the crop yield prediction in different climate and cropping scenarios will be beneficial.

Shruti Mishra et al 2018 [3] have indicated that applying the data mining techniques on historical climate and crop production data several predictions are made which increase the crop productivity. The decision support system has to be implemented for the farmers to take proper decisions about soil and crop to be cultivated. They have collected the dataset with attributes of the crop season, Area and production in hectares and analyzed with various algorithms in WEKA. They analyzed data with four methods and found their accuracy and compared with each other. The four methods used are J48, IBK, LAD tree, LWL in WEKA. They concluded that the IBK had got more accuracy when compared to all other and that depends upon the nature type and the nature of the dataset.

Chlingaryana et al 2017 [4] indicated the major factor in the crop yield prediction is the nitrogen level in the soil. Nowadays remote sensing systems are mostly used in decision making. These remote sensing data is used to help the farmers to improve the crop yield. Huge remote sensing data is used to make a decision. Nitrogen is used to improve the crop yield and make the soil fertile. Machine Learning algorithms are used to make the decision. major factors we are going to take it into consideration is nitrogen, type of soil and yield analysis of previous data of these factors are helpful to make the accurate decision and predict the yield and helps the farmer. Now a day's precision agriculture is used to improve the yield and giving suggestion to farmers. It uses information technology to ensure the crop and soil.it says how they need to optimize the production and health of the soil. The obtained results are back-propagation neural network is used to get different vegetarian incidents. The conventional neural network of long term memory to predict feature data.

Dakshayini Patil at all 2017 [5] indicated that rice crop plays a major role in the economy. They used various data mining techniques to predict the yield of the rice crop. Rice crop is the sustainable security of India. In general, it contributes 40% to the general yield. High yield of the crop is based on the appropriate climatic conditions. Learning a better strategy to grow the crop according to the climatic conditions can improve the crop yield. The reports utilize various mining techniques based on the previous data of the crop yield and different climatic regions. In this, the authors used data of 27 regions of Maharashtra to predict the yield of the crop.

Weighted contribution to its yield enactment. One sort of system sees the hubs as “artificial neurons”. These are called neural systems. The back engendering calculation (Rumelhart and McClelland, 1986) is utilized in layered feed-forward ANNs. This implies the counterfeit neurons are sorted out in layers and send their signs "forward", and after that, the blunders are spread in reverse [7]. The system gets contributions by neurons in the info layer, and the yield of the arrange is given by the neurons on a yield layer. There might be at least one middle of the road concealed layers. This neural arrange engineering is extremely mainstream, since it very well may be connected to a wide range of undertakings. The principal term, "feed forward" depicts how this neural arrange procedures and reviews designs. In a feed-forward neural system, neurons are just associated with forward. Each layer of the neural system contains associations with the following layer (for instance, from the contribution to the covered up layer), yet there are no associations back. The expression back spread portrays how this kind of neural system is prepared [8]. A back spread is a type of managed preparing. When utilizing a managed preparing strategy, the system must be given both example inputs and foreseen yields. The foreseen yields are looked at against the real yields for given information. Utilizing the foreseen yields, the back proliferation preparing calculation at that point takes a determined mistake and changes loads of the different layers in reverse from the yielding layer to the info layer [9].

III. Proposed System Methodology

Prediction of the crop yield using the efficient algorithm and suggest how much quantity of fertilizer should be used to get the proper yield for the crop.

A. Data Set Description

This is the sample data set used in this project. The data in Table I is data used to predict crop yield based on 7 factors. These 7 factors are state, district, crop, area, season, production by this data we can create a machine learning model and train the model and we can predict the production and from Table II we can predict the amount of fertilizer should be used to get the proper yield the input parameters are the quantity of nitrogen, phosphorus, and the output is the amount of the respective fertilizer should be used. Hear in the input parameters 1, 2, 3, 4, 5, 6 represents the very high, high, above average, below average, low and very low quantity present in the soil respectively.

Table I
Sample Data Set Of Crop Data

State_Na	District	Crop_Ye	Season	Crop	Area
Bihar	GAYA	2013	Rabi	Rapesee	1086
Bihar	GAYA	2013	Rabi	Sunflow	1394
Bihar	GAYA	2013	Rabi	Wheat	58783
Bihar	GAYA	2013	Summe	Maize	1165
Bihar	GAYA	2013	Summe	Moong(	3856
Bihar	GAYA	2013	Whole Y	Sugarca	459

Table II
Sample Fertilizer Data

n	p	k	amt of n	amt of p	amt of k
3	5	6	64	50	60
1	4	2	40	46	30
5	1	5	93	16	32
3	1	3	63	20	39
4	6	6	87	37	39
2	1	1	65	19	32

B. Necessary Packages

- Numpy
- Pandas
- Matplotlib.pyplot
- Scikit-learn
- Tensorflow
- Jupyter

Store the data related to the yield and fertilizer in the csv which consists of the state_name, district_name, crop_year, season, crop, area, production, and another data set consists of level of the phosphorous, level of potassium, level of nitrogen in the soil, how much amount of phosphorous, potassium, nitrogen should be used to increase soil fertility.

C. Architecture

The Fig. 1 shown below represents the architecture diagram.

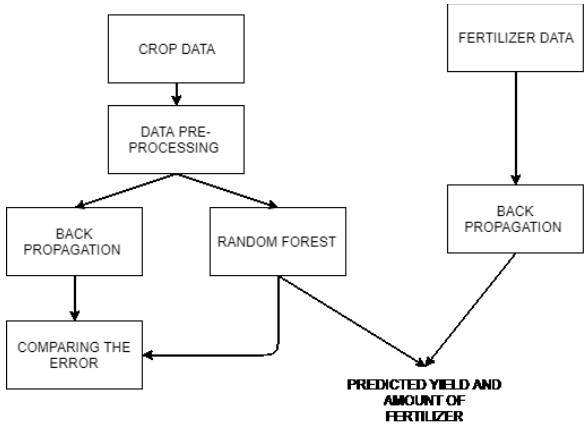

Fig. 1. Architecture diagram

D. Metadata

All the main data used in the data set are initialized with the number to use in the algorithm it is like initializing all the details. In this metadata, we are going to initialize all the crop names with the numbers.

This data makes us use the data easily in the algorithm. Hear the metadata of all the crops is given with a particular number. This number is not duplicated that is one number is given to one crop, the same number is not given to the other crop. This metadata consists of more than a hundred crops that grown all over India.

E. Data Pre-processing

Hear the raw data in the crop data is cleaned and the meta-data is appending to it by removing the things which are converted to the integer. So, the data is easy to train. Hear all the data. In this pre-processing, we first load the metadata into this and then this metadata will be attached to the data and replace the converted data with metadata. Then this data will be moved further and remove the unwanted data in the list and it will divide the data into the train and the test data

For this splitting of the data into train and test we need to import train_test_split which in the scikit-learn this will help the pre-processed data to split the data into train and test according to the given weight given in the code. The division of the test and train is done in 0.2 and 0.8 that is 20 and 80 percent respectively

F. Random Forest Algorithm [13]

This algorithm suits for both huge and small data to give an efficient prediction. Based on the given data to the algorithm it forms various decision trees and checks for how many trees give the same prediction. It is based on the votes it will count and which trees give the same output after that the output given by the maximum trees it will show as output as explained in Fig. 2. The given data in the project go to the random forest algorithm and hear it will build ten trees and pass data to it. Every tree is classified based on the various conditions and it will train the model according to it and will

count the number of trees will give the same output and which has more count it will be decided as the output.

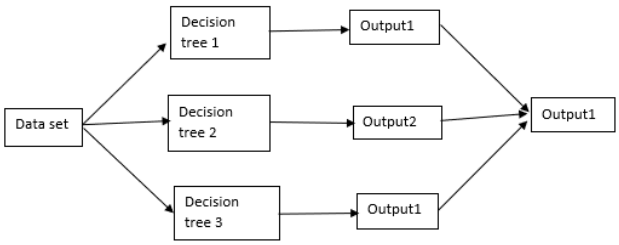

Fig. 2. Working of a random-forest algorithm

G. *Fertilizer utilization using Back-propagation [14]*

The Fertilizer dataset which is present in the CSV format is pre-processed and made ready to train the model with that dataset. First, the data set is divided into 80% for training data and 20% for the test data. The algorithm used to train the dataset into a model is the Backpropagation algorithm. The backpropagation algorithm is the concept from the multiple layer perceptron in the artificial neural network. The backpropagation algorithm is used for large datasets which have no proper relationships between the attributes of the dataset to form a network model by training the dataset and predict the output. This algorithm mainly consists of the three layers in the network model, They are the input layer, hidden layer, and output layer. Input layer in the model is mainly responsible for giving the inputs to the model, then the hidden layer which is in between the input layer and output layer and this mainly responsible to get the output from the input layer as input and calculate according to the weights present on the input to hidden layer and gives the desired output result and the last layer is output layer which gives the output predicted from the network model. Backpropagation algorithm is a supervised learning algorithm. To train a dataset in a back propagation algorithm it should have the desired output attribute in the dataset. Backpropagation algorithm is trained in a way we fix the output value or attribute to the dataset that is to be trained, At first iteration is output is calculated and observes the difference between the expected output and the obtained the output based on that observation it back-propagates the error and update the weights between the nodes in the layers and bias. The network is thus trained with many iterations until it gets the desired output. After training the network model then it's validated and generalized by the test dataset whether the prediction is accurate or not. After validation of the network model, we can predict by giving the unknown data and predict the output to the unknown data given to the model.

Method:

1. At first, to train the network we initialize the weights and bias.
2. Considering all attributes x from the dataset d compute the output for every unit in the network.
3. Then it back-propagates the error in the network

   For each output from the network as k, Calculate it's error term ∂k

$$\partial k = O_k(1-O_k)(t-O_k)[14]$$

For each hidden unit h, calculate its error term ∂h

$$\partial h = O_h(1-O_h)\sum_{K=outputs} W_{kh}\partial k$$

Update each network weight

$$W_{ji} = W_{ji}+\Delta W_{ji}$$

Where

$$\Delta W_{ji} = \eta\, \partial_j\, x_{ji}$$

IV. RESULTS

In this paper, an effort is made in order to know the crop production analysis and is processed by implementing both the Random Forest algorithm and Backpropagation algorithm. These models were experimented with different types of crops in various regions across India to predict the output. Even fertilizer data was trained using the back propagation algorithm and evaluated to get the result of how much nitrogen, phosphorus is required for the area of land. Both the models for the crop production were compared in predicting the output and by various parameters with respect to the error rate. We compared the error rate obtained while comparing the random forest algorithm and backpropagation where we got the error rate lesser to the random forest than back propagation while predicting the output for both of the models and the compassion is plotted in the graph Fig. 3. For prediction of the yield, the user will enter the data as shown in Fig. 4. The user should enter the details one after another. The output of the yield prediction is shown in Fig. 5. The input of the fertilizer data is entered as shown in Fig. 6 and the out of the fertilizer data is shown in Fig. 7.

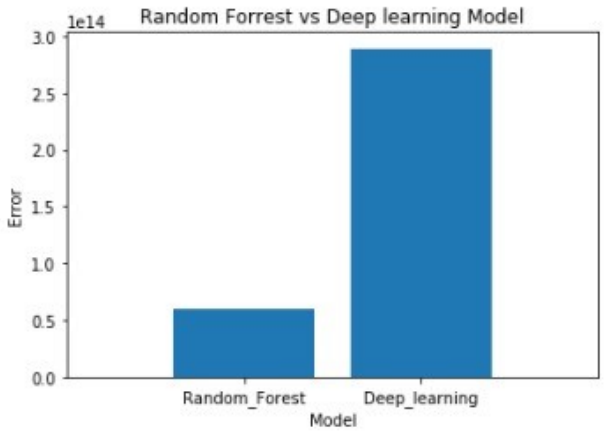

Fig. 3. Comparison between two models respective to their error rate

Fig. 4. User input-1

Fig. 5. output for the given user input-1

Fig. 6. user input-2

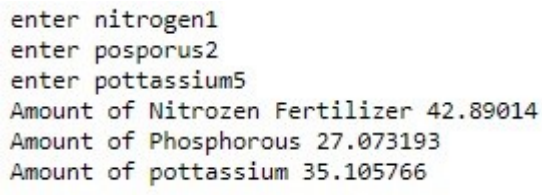

Fig. 7. Output for the above user input-2

V. Conclusion and Future work

Crop yield prediction and efficient use of the fertilizer is successfully predicted and also found the efficient algorithm from both the algorithm and obtained the most efficient output of the yield. In future developing the web application based on this ideology and make the user use this easily and help the user to understand the yield of the crop, he is going to crop in that season.

Acknowledgment

The obtained result will be helpful for the farmers to know the Yield of the crop so, he can go for the better crop which gives high yield and also say them the efficient use of fertilizer so that he can use only the required amount of fertilizers for that field. this way we can help the farmers to grow the crop which gives them better yield.

References

[1] Niketa Gandhi et al ," Rice Crop Yield Forecasting of Tropical Wet and Dry Climatic Zone of India Using Data Mining Techniques",IEEE International Conference on Advances in Computer Applications (ICACA) , 2016.

[2] K.E. Eswari. L.Vinitha. " Crop Yield Prediction in Tamil Nadu Using Baysian Network ", International Journal of Intellectual Advancements and Research in Engineering Computations, Volume-6,Issue-2,ISSN: 2348-2079.

[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 PartNumber: CFP18J06DVD , ISBN:978-1-5386-0806-7.

[4] Anna Chlingaryana, Salah Sukkarieha, Brett Whelanb — Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review , 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] A.T.M Shakil Ahamed, Navid Tanzeem Mahmood, Nazmul Hossain, Mohammad Tanzir Kabir, Kallal Das, Faridur Rahman, Rashedur M Rahman "Applying Data Mining Techniques to Predict Annual Yield of Major Crops and Recommend Planting Different Crops in Different Districts in Bangladesh" 978-1-4799-8676-7/15/$31.00 copyright 2015 IEEE SNPD 2015, June 1-3 2015.

[7] K. Menaka, N.Yuvaraj,Indian Journal of Innovations and Developments Vol 5(12), October, 2016 ISSN (online) : 2277-5390 ISSN (Print): 2277-5382

[8] Machine Learning by Tom M. Mitchell, ISBN: 0070428077

[9] Snehal S.Dahikar, Dr.Sandeep V.Rode ,"Agricultural Crop Yield Prediction Using Artificial Neural Network Approach ", International journal of innovative and research in electrical, instrumentation and control engineering, volume 2,Issue 2,2014.

[10] 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.

[11] Ramesh A. Medar, "A survey on data mining techniques for crop yield prediction", International Journal of advance in computer science and management studies, ISSN:2231-7782, volume 2, Issue 9, 2014.

[12] Pooja M C et al, —Implementation of Crop Yield Forecasting Using Data Mining , International Research Journal of Engineering and Technology (IRJET),2018,p-ISSN: 2395-0072

[13] Jig Han Jeong et al , "Random Forests for Global and Regional Crop Yield Predictions ",PLOS-ONE DOI:10.1371/journal.pone.0156571 June 3, 2016.

[14] Jharna Majumdar, Sneha Naraseeyappa and Shilpa Ankalaki, Analysis of agriculture data using data mining techniques: application of big data, Springer journal,2017.

[15] S.kanaga Subba Raju et al, Demand based crop recommender system for farmers, International Conference on Technological Innovations in ICT For Agriculture and Rural Development, 2017.

[16] Shriya Sahu et al," An Efficient Analysis of Crop Yield Prediction Using Hadoop framework based on random Forest approach, International Conference on Computing, Communication and Automation , 2017.

[17] E Manjula and S. Djodkitachoumy International journal of computation Intelligence and informatics,Vol 6:No 4 March 2017.