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

Abstract

As u all Know that India being an agriculture country we need to concentrate on formers problem and also India being highly populated country we need to protect the food resources. In agriculture, yield prediction is very important thing. Formers are facing so many problems on prediction of yield, and they are loosing their interest on agriculture field. So to predict the yield of the crop analyze the attributes location, ph value, nutrients of the soil, nitrogen(N), Phosphorous(P),Potassium(K)soil type by suing the help of APIS for weather and temperature, nutrients merit of soil in that region ,amount of rainfall in the particular region. Using all these attributes train the data with various machine learning algorithm. deliver the end users with the appropriate information The system comes with the accurate in yield prediction and about fertilizer based on the temperature and atmosphere of the land in which it enhances the crop yield and increase formers revenue.

Keywords: *Artificial neural network, Random forest algorithm, Back Propagation algorithm, Prediction, Agriculture.*

I. Introduction:

Agriculture is the most important field in Indian wealth. Achieving the maximum crop at raising agricultural productivity per unit of the land will be the main reason of agricultural growth. Major thing is that strengthening India's agricultural research system is one of the quantitative demand for agricultural growth. The actual problem facing by an Indian agriculture is that combination of technologies to bring the desired output Agriculture has been back bone of Indian economy it will continue for a long time. 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. 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 a important role in the entire socio-economic fabric of India. So, keeping all these things in mind we are doing best for the agricultural field. Many researches and calculation done to 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 are 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 difficulty 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.

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, Back propagation 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 recommends suitable fertilizer for every particular crop. This paper is mainly focused on active utilization of machine learning algorithms and its quantifications. This 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. The rest of the paper is organized as follows- the section two gives the problem statement, then the section three gives the related work.

II. PROBLEM DEFINITION:

Till now the work done is only concerned on crop yield and its predictions using properties of soil and Data Mining Technologies. Fertilizers Recommendation is not taken into deliberation. So it is necessary to develop prediction of crop yield and recommendation of fertilizer to get the proper yield which predicts yield of the crop based on soil nutrients crop yield data and recommend fertilizers for the particular crop based on different set of data like data of fertilizer, region data and data of the crop yield and also to compare the predicted result of the different algorithms and determine the more suitable approach.

III. RELATED WORK:

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, average temperature and previous related to crop information 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 [2] indicates that the crop yield depends on the perception, average, minimum and maximum temperature. Evapotranspiration is the other crop that they have taken into the consideration. 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 [3] has mentioned that on applying Data Mining techniques on past climate and crop production data, and predicting data would enhance the crop productivity. 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 areJ48, 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 [4] 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. 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 the27 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 (for example, from the contribution to the exposedtop layer), until there are no connection back. The equation back spread represent 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].

IV. PROPOSED SYSTEM METHODOLOGY:

To compare the predicted result of different algorithms and determine which approach is more suitable. By using efficient algorithm we predict the crop yield and also we guide how much ratio of fertilizer that have to use in order to increase the soil fertility and to get best crop yield.

a) Data Set Description:

This is the illustrative data set that we used in this project. The data which is in Table I that is for predicting the yield of the crop which is depend on 7 components. So, those 7 factors are district, state, type of crop, , season, production, area, using these as the inputs we can make a machine learning system and train that system, so that we can predict the production .Table II which is associated with the illustrative fertilizer dataset, from this data we can predict in which ratio we have to use fertilizer to get the good yield, the input values are the amount of nitrogen and the amount of phosphorus, and it will give output as the quantity of the appropriate fertilizers have to use .These are the input values i.e 1, 2, 3, 4, 5, 6 which specifies the very high, high, above average, below average, low and very low quantity present in the soil.

state name	district name	crop year	season	crop	temperature	humidity	soil moisture	area	production
AP	kadapa	2008	rabi	rice	37	42	55	14883	41271
AP	kadapa	2009	kharif	jowar	34	55	62	6749	13532
Assam	barpeta	1999	rabi	gram	36	40	54	229	92
Assam	barpeta	2000	Whole year	banana	37	42	55	5131	77222
Bihar	araria	1999	summer	maize	36	40	54	8272	9814

b) Sample fertilizer data:

state name	district name	crop year	season	crop	temperature	humidity	soil moisture	area	production
AP	kadapa	2008	rabi	rice	37	42	55	14883	41271
AP	kadapa	2009	kharif	jowar	34	55	62	6749	13532
Assam	barpeta	1999	rabi	gram	36	40	54	229	92
Assam	barpeta	2000	Whole year	banana	37	42	55	5131	77222
Bihar	araria	1999	summer	maize	36	40	54	8272	9814

Keep these data which are connected to the yields and the fertilizers. These data contains of the state_name, district_name, crop_year, season, type of crop, which area, how much is the production, and the second data set which contains the information about of level of the phosphorous, level of potassium and level of nitrogen in the soil, and in which quantity they have to use the phosphorous, potassium and nitrogen for increasing the fertility of soil.

c) Architecture diagram:

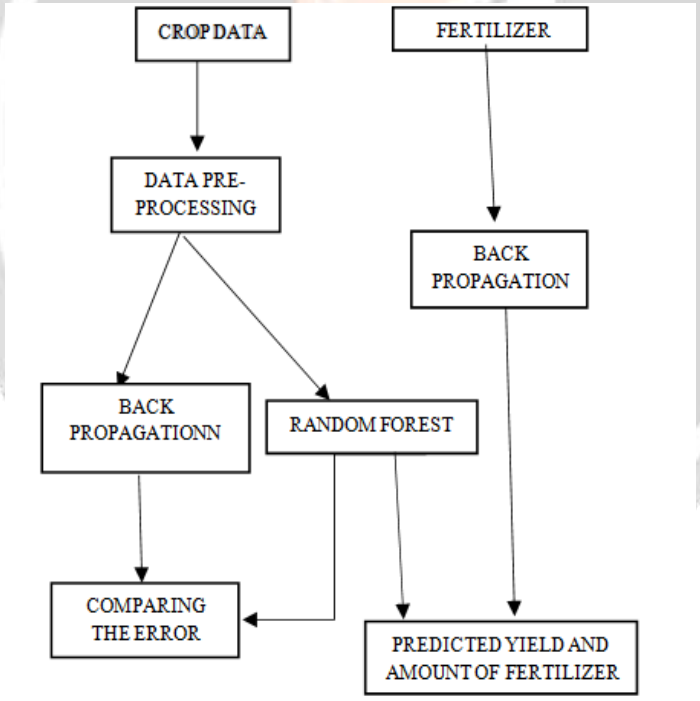

Fig.1 Architecture Diagram

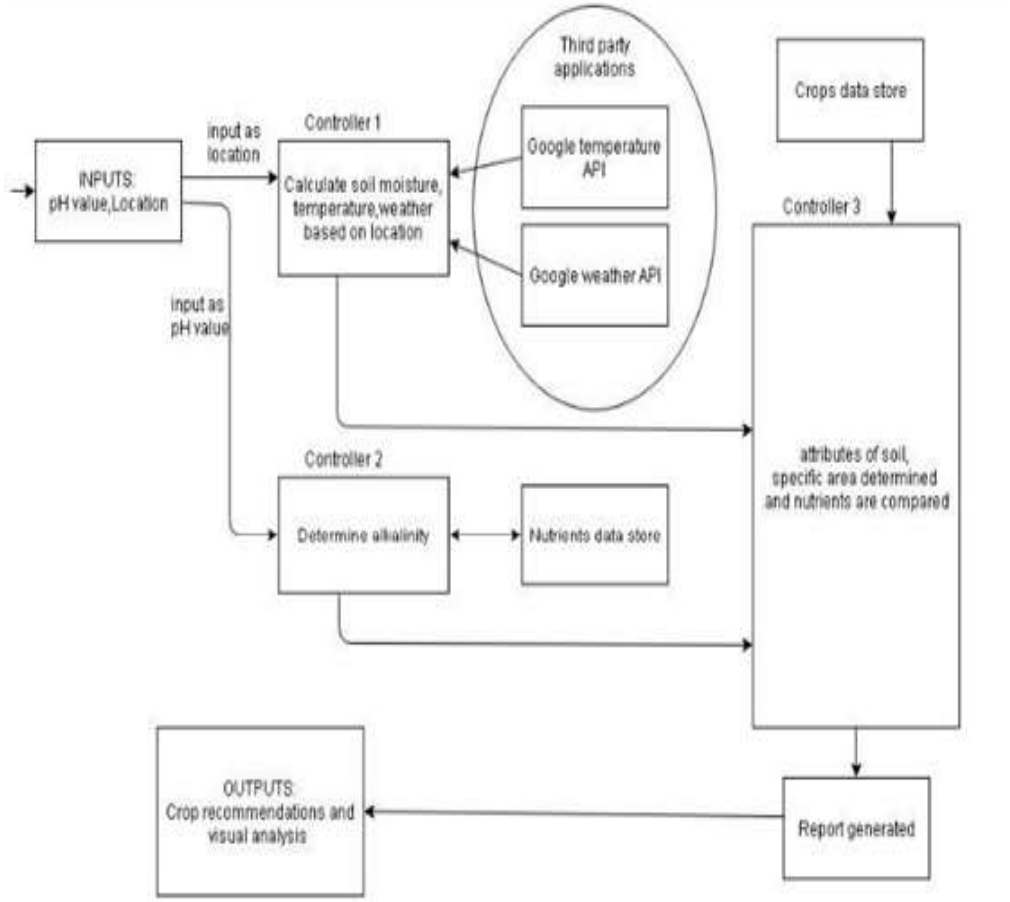

Fig.2: Prediction frame work

This framework takes inputs as pH value and area from the client. Result preparing is finished by two controllers. Area is utilized as a contribution to controller 1, alongside the utilization of APIs for climate and temperature, quality of soil, supplement estimation of dirt in that locale, measure of rainfall in a district, soil arrangement could be resolved. pH value is a contribution to controller 2, from which the alkalinity of dirt is resolved. Level of supplements like Nitrogen(N), Phosphorous(P), Potassium(K), issue can be gotten. The aftereffect of controller 1 and controller 2 are contrasted and a predefined "nutrients" information is stored. These contrasted outcomes are provided with controller 3 where in the mix of the above outcomes and the predefined informational collection present in the harvest information store is analyzed.

d) META DATA:

In the algorithm, the data set which have main data are used to initialize with the number which is like initializing the details. Here in Meta data, all the crop names are initialized with the number. This kind of initialized data

can be used easily in the algorithm. The initialized data in the meta data of all the crops are given with particular number. The number given to the particular crop must be unique, that only one number is given to the one crop, then the same number should not be given to the other crop. This metadata consists about crops which more than hundred grown all over India.

e) DATA PRE-PROCESSING:

Here in data pre-processing, crop data consists the raw data of the crop which is cleaned and metadata is appending to it by removing the things which are converted to integer. Now the data can be easily trained. In this data pre-processing we will load the metadata and this metadata will get attached to the data and then it will replace the converted data with the metadata. Then this data is going to remove the unwanted data in the list, then this data is going to split the data into test data and train data.

According to the weight given in the code, we have to divide the pre-processed data into train and test

data, to do this splitting we need to import the train_test_split which is in scikit-learn. The splitting of train and test data is done in80 and 20 percent respectively.

f) RANDOM FOREST ALGORITHM:

The Random Forest algorithm gives an efficient prediction for both huge and small data.This algorithm forms various decision trees for the given data and checks for same prediction given by other trees to know how many trees given the same prediction. The algorithm is based on the voting count and which trees gives the same output. After this the same output given by the maximum trees will be shown as the output as explained in Fig.2. The data given in this project will come under the random forest algorithm and in this algorithm it will build ten more trees and data is passed to it. Based on the various conditions every tree which is formed by the algorithm is classified and model will be trained according to it. Also it will count the number of trees which will give the same output and output will be decided on the basis of tree which has more count.

Fig.3: Working of random forest algorithm

g) FERTILIZER UTILIZATION USING BACK PROPAGATION:

The dataset of the Fertilizer is in CSV format, it has to be pre-processed and made it ready to train model with the fertilizer dataset. First step is to divide the data set into training data and test data, in this 20% for test data and remaining 80% for training data. The dataset has to be trained into model by using Back propagation algorithm. In the artificial neural network the back propagation algorithm concept is there in multiple layer perceptron. The large datasets which have no proper relationships between the attributes of the dataset uses the back propagation algorithm to form the network model by training the dataset and the output will be predicted.

In the network model the back propagation algorithm consists three layers they are input layer, hidden layer and outer layer. In the model, input layer job is to give inputs to the model, then the hidden layer which is present between the input layer and output layer, its main job is to get the input layer output as input and calculations are done according to the weights present on the input to hidden layer and this will gives the desired output and the last layer is output layer it will give the predicted output from the network model. Backpropagation algorithm is a good learning technique. In a back propagation algorithm the dataset should have the desired output attribute to train the dataset. This algorithm is trained in such a way that we can fix the attribute or output value to the dataset which has to be trained, In the first iteration the output is calculated and the difference between the expected output and the output is observed and the output obtained is based on that observation it backpropagates the error and also update the weights between the nodes in the layers and bias.In this way with many iterations the network is trained until it gets the desired output.

After trainingthe 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.

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

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

Calculate error term

$$\partial h = \square h(1-\square h)\sum Wkh\partial k \qquad K=outputs$$

Update each network weight

$$Wji = Wji+\Delta Wji$$

Where

$$\Delta Wji = \eta\, \partial j\, xji$$

V. CONCLUSION:

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, that he is going to crop in that season.This makes the farmers to take the right decision for right crop such that the agricultural sector will be developed by innovative ideas.

VI. REFERENCES:

[1]Niketa Gandhi et al,"Forecasting of rice crop yiled 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 Networkbayesian ", Advancements of international journal and Engineering research Computations, Volume-6, Issue-2, ISSN: 23468- 2078.

[3]Shruti Mishra, PriyankaPaygude, SnehalChaudhary, SonaliIdate "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]AnnaChlingaryana, Salah Sukkarieha, Brett WhelanbMachine learning approaches for yield of the croppredictionand nitrogen position evaluation in precision agriculture: A analysis, Computers and Electronics in Agriculture 151 (2018) 61–69, Elisver, 2018.

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

[6]ShakilAhamed, NavidTanzeemMahmood, NazmulHossain, MoFaridurRahman, Rashedur M Rahman"Applying Data Mining Technology to PredicttheAnnual crop Yield of more Crops and Recommended to Planting more crops in various placesin pune".

[7]K.Menaka, N.Yuvaraj,Indian Journal of Innovations and Developments Vol 5(12)

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

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

[10]DakshayiniPatil 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.

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

[12] 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.

[13]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.

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

[15] ltay HG, "An over allview of regression technology for knowl- edge discovery". Knowbang