

PREDICTION OF CROP YIELD USING MACHINE LEARNING

Karan Chaudhary

Student, Atria Institute of Technology,
Visvesvaraya Technological University, Bangalore

Prof. Farhana Kausar

Assistant Professor, Atria Institute of Technology
Visvesvaraya Technological University, Bangalore

Abstract: Looking at the current situation faced by farmers, there is an increase in suicide rate over the years. Main reason behind it includes poverty, unpredictable weather, financial problem, parental issue and frequent change in Indian government norms. Education also plays an important role because due to lack of education they are not aware of the soil quality and the climatic change and sometimes farmer was not aware about the feasible crops. Because of lack of knowledge they are not aware of the soil quality, soil nutrients and soil composition. This paper focused on prediction of crop based on the existing data by Artificial Neural Network. We are implementing Artificial Neural Network which is used for classification and prediction as most of intense computation takes place during training phase only, there is no requires of testing phase.

Key Words: Crop Yield, Machine Learning, Artificial Neural Network, Farmer, Prediction

I. INTRODUCTION

Agriculture is one of the important factors and it plays an important role in individual's life. Being the backbone of India, Agriculture sector has been improving by the needs of public, as the technology improves. Agriculture is the primary occupation of most of Karnataka's rural people. As per the statistics, a total of 123,100 km² of land is cultivated in Karnataka, constituting 64.60% of the total geographical area of the state. Agriculture remains the primary activity and main source of livelihood for the rural population in the state. It is characterized by wide crop diversification and remains highly dependent on the vagaries of the southwest monsoon. As per the statistics of 2018, agriculture employed 50% of Indian workforce and contribute 17 to 18% of country GDP. Suicide rate of farmer in India had ranged between 1.4 to 1.8 per 1 lakh total population over 10 years of period from 2005. However, in 2018, the farmer suicide rate is more than 10 per day. With this increasing suicide rate, our aim is to help farmer to get knowledge about the importance of prior crop prediction. Because of knowledge about the soil and its quality, they will be able to grow crop in huge amount.

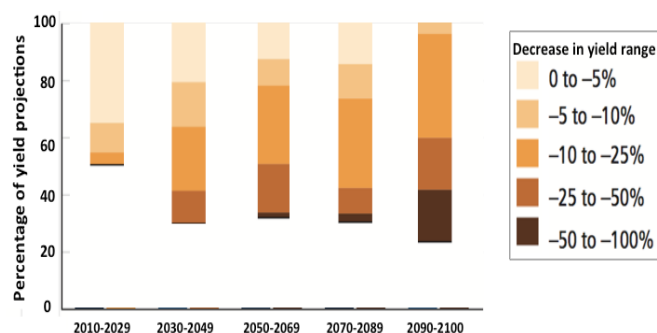


Figure 1: Decrease of Crop Yield under Global Climate Change

II. LITERATURE SURVEY

Agriculture plays a vital role in India's economy. 54.6% of the total workforce is engaged in agricultural and allied sector activities (Census 2011) and accounts for 17.1% of the country's Gross Value Added (GVA) for the year 2017-18 (at current prices). But there is continues decline in agriculture's contribution to Gross Value Added. Food is most essential element on the life which depends upon outputs of agriculture, so farmers plays an important role, Following Comparison is shown below:

The study in [1] used Kohonen's SOM (Self-Organizing Map) and Multilayer perception feed-forward neural network for analysis of crop. In the multilayer perception feed-forward neural network, the neurons of the first layer direct the output to the second layer, in a unidirectional manner so that the neurons are not received from the reverse order. Incorporating the three layers- input, output and hidden, the feed-forward neural network designates distinct roles for each. The output layer comprises a number of neurons equal to the used number of quantities computed from the input and made accessible the perceptron responses. The hidden layer takes care of approximating non-linear problems. Input layer maintains equal number of neurons corresponding to that of the variables in the problem. The input and output layer are only required for processing linear. The risk of overfitting occurs in feed-forward neural network if we want to create the database of huge data based on deep learning strategies. The main drawback of the SOM is that it requires neuron weights be necessary and sufficient to cluster inputs. When a SOM is provided too little information or too much extraneous information in the weights, the groupings found in the map may not be entirely accurate or informative.

The study in [2] used Random Forest Classifier for analysis of crop. Decision tree classifier uses greedy approach which may reduce the value of important data. The result are accurate and accuracy ranges from 0 to 100%. Class assignment was made by the number of votes from all the trees and for regression the average of result was used. Random forest Classifier is only applicable for the collection of decision tree. Due to the collection of decision tree, the complexity arises and they are much harder and time-consuming method. It's fairly good for training even small samples and can be easily parallelized in R/python/other software. However, it fails when there are rare predictors or outcomes, as algorithm is based on bootstrap sampling.

The study in [3] used Multiple Linear Regression (MLR) technique for analysis of crop. [5] Multiple linear regression (MLR), also known simply as multiple regression, is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. The goal of multiple linear regression (MLR) is to model the linear relationship between the explanatory (independent) variables and response (dependent) variable. ID3 algorithm and Classification is used to analyze 362 dataset and gives the output. Training dataset was classified according as organic, inorganic and real estate for predicting soil's type. The output computed from this system were accurate and reliable.

In [4], 2 regression supervised machine learning methods are used: Relevance vector Machine (RVM) and Support Vector Machine (SVM) to show the effectiveness in predicting soil quality. Here a smart wireless device is used for sensing moisture of the soil and other soil's factor. The wireless device used didn't give the accurate reading which results error. However, the device is not tested with real time data. And, also agricultural system which deals with so many factors affecting the crop yield needs a non-linear method to interpret the relationships between these factors and crop performances [5]. Thus, regression is insufficient to show the interactions of the factors and crop yield.

III. PROPOSED SYSTEM

Crop prediction analysis is used for predicting the suitable crop by considering various parameter of the soil such as nitrogen, fertility, pH, Phosphate, Potassium and some atmospheric parameter such as sunshine, rainfall and humidity. Above factors have direct impact on crop yield. Complexity arises as agriculture system deals with large amount of data that comes from number of factors. Lots of technique have been used for predicting the crop's yield using different algorithm. The proposed system aims to help the farmer to predict the crop's yield based on various factor such as temperature, rainfall, humidity, soil nutrients, etc. It can be achieved using Artificial Neural Network algorithm which is based on the human brain's biological neural network. The ANN is based on human brain's neural process and used to solve complex problems. Artificial Neural Network Consist of three different layer such as Input Layer, Hidden Layer, Output Layer. Input layer is very beginning layer of neural networks that takes the input and serves that input to the subsequent layer of neural network.

Input layer feeds the output to input of the hidden layer, and the output of which is connected to the output layer.

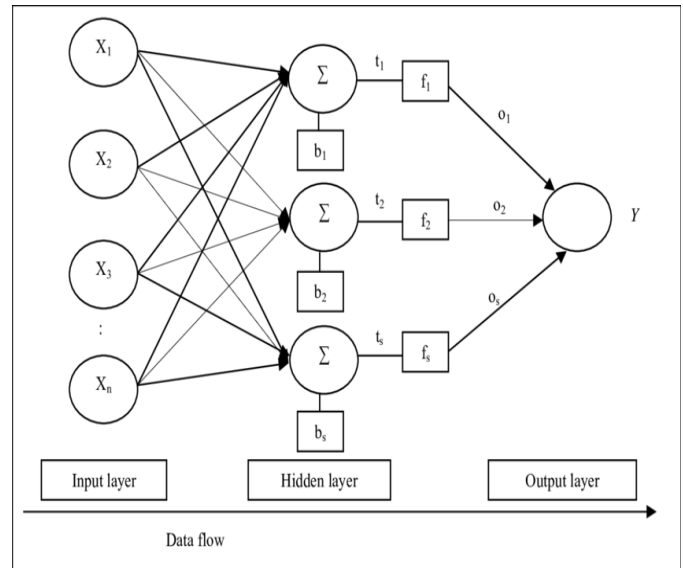


Figure 2: Layers and connection of a feed-forward back propagation ANN

ANN has become most popular technique in predicting in case of biological science. ANN model produces more accurate and precise result in crop prediction technique because ANN require no explicit mathematical equation and assumptions. In ANN most of the intense computation take place during training process. Since there are huge amount of factor which influence crop prediction, ANN technique uses the major factor with influence the prediction directly or indirectly. Some of factors are:

- Environmental Factor
- Soil Factor

3.1 ANN in Environmental Factor

The Environmental Factor such as temperature, humidity, rainfall was the most important factor that control crop yields. Crops like barley, rice, maize, etc. is highly influenced by the environmental factor. Due to global climate change, every year there is decrease in crop's yield. Using ANN one can take the specific factor and feeds to the input layer based on the types of crops. For maize, rainfall is most important factor. So, we can consider the maximum rainfall, minimum rainfall and optimal rainfall and feed that input to the ANN to show the relation between output and input factor.

3.2 ANN in Soil Factor

Based on the types of soil, one can grow the crops. Considering the soil factor (such as fertility, pests), we can predict with types of crop should be grown using the technique of back-propagation method. Based on the previous data of the soil, ANN technique can predict which crop should be grown in that land.

In ANN, Backpropagation algorithm attempts to find the minimum of the error function in the weight space using methods based on gradient descent which sends their signal forward and error are propagated backwards.

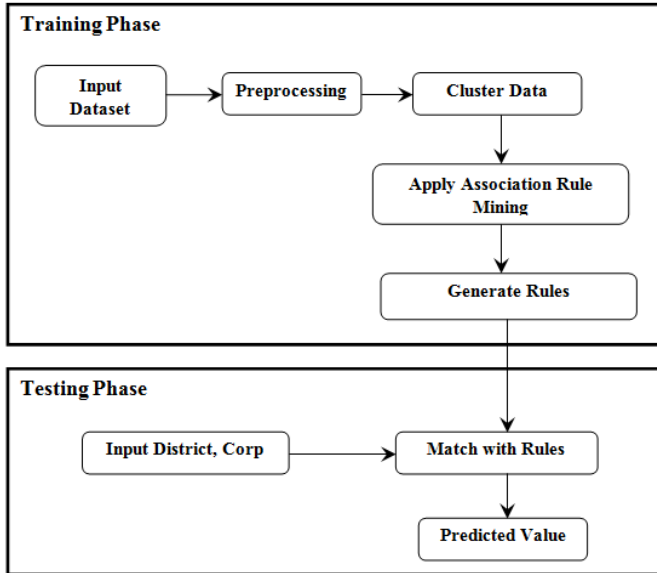


Figure 3: Workflow Model

Figure 3, show the workflow model which consist of two phases i.e. Training phase and Testing phase. In training phase, the input data is feed to the input layer and preprocessing is done using the hidden layer. Hidden layer takes the cluster of data and then it processes the data and feeds the output to the input of the output layer. Training phase comprises of Input layer, Hidden layer which Testing phase only contain output layer. Training data is always greater than testing data. All the major functionality is included in training phase (such as preprocessing of data, clustering). Preprocessing of data includes redundancy check, data filtering, outlier detected, correlation testing, feature extraction and so on. Clustering means group the data with similar characteristics. After the completion of the training phase, the data is feed to the testing phase. Training phase mainly consist of classification and prediction of data. Classification classify the data based on the related algorithm and prediction is used to predict the value present in the data.

IV. SYSTEM ARCHITECTURE

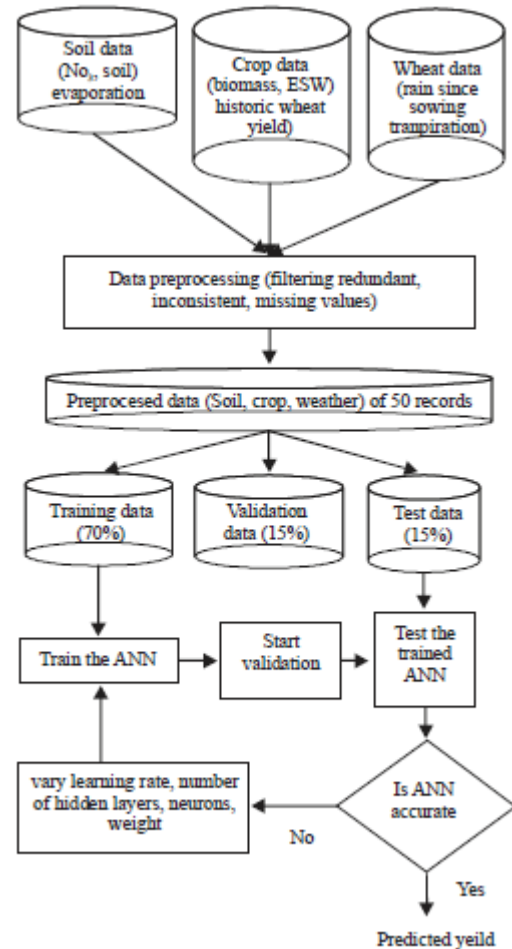


Figure 4: System Architecture

Figure 4, shows the flow diagram of Artificial Neural Network for crop prediction. In ANN technique, Input layer is feed with three datasets i.e. soil data (such as soil fertility, soil evaporation, soil moisture), crop data (such as biomass, previous year crop yield) and the specific grain data (to be planted). After preprocessing the data, the input layer sends the preprocessed data to the hidden layer. After collection the dataset by input layer, ANN performs Exploratory Data Analysis (EDA) and filter redundant for inconsistent and missing values. Data preprocessing is the major step in building model. Redundant and missing value in dataset affects the learning rates of the model and that is the accuracy of the model.

After preprocessing of dataset, ANN divide the dataset into testing data and training data with 70 percent training data and 15 percent testing data and 15 percent validation data. After that the model is built with ANN and then the model was trained using training set. When training completes, Data should be classified and tested using testing set to test the learning rate/accuracy of the model. In between validation dataset will validate the predicted ANN is accurate or not.

If the result is accurate, we can predict the sample of crop to be yield and it will give us the list of the suitable crop to be grown



in the given condition. But, if the result is not accurate, we have to reshape the ANN model with varying learning rate, number of hidden layers, neurons, data, weight and the specific crop data. We will repeat the process until we can an accurate result.

V. CONCLUSION

The proposed system takes into consideration the data related to environmental factor, soil factor, weather, fertility of soil and past year production and suggests the best profitable crops that can be cultivated in the given environment condition. ANN technique lists out all the possible crops so that the farmer has option to choose the profitable one. The combination of artificial intelligence and the agriculture will help most of the farmers in near future. In comparison with other, Neural Network is best solution for agriculture problems (such as prediction of crop yields). The application of ANN plays an important role in prediction of the agriculture based on the certain important factors and it also plays a major role in future evaluation of crop yields based on the present/past data.

VI. ACKNOWLEDGEMENT

I would like to thank my Research Guide Prof. Farhana Kausar, and Prof. Pallavi N, Professors in Computer Science Department, Atria Institute of Technology, Bangalore for their continuous support and guidance. Authors are also thankful to the reviewer for going through the manuscript and giving valuable suggestions for the renovation of manuscript. I would also like to thank the Department of Computer Science, Atria Institute of Technology, Bangalore for providing me this opportunity. Last, but not the least I would like to thank my family, who has acted as a beacon of light throughout my life. My sincere gratitude goes out to all my comrades and well-wishers who have supported me through all the ventures.

VII. REFERENCES

- [1] Ogwueleka, Francisca Nonyelum,” Crop Growth Prediction Using Self-Organizing Map and Multilayer Feed-forward Neural Network”, in American-Eurasian Journal of Sustainable Agriculture, 5(2), ISSN 1995-0748, pp. (168-176).
- [2] D.S Zingade, Omkar B, Nilesh Mehta, Chandan Mehta(2017),”Crop Prediction System using Machine Learning”, in International Journal of Advanced Engineering and Research Development”, Vol 4, Special Issue 5, pp. (1-6)
- [3] Marizel B. Villanueva, Ma. Louella M. Salenga(2018),” Bitter Melon Crop Yield Prediction using Machine Learning Algorithm, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 9, No. 3, pp. (1-6)
- [4] Zhihao Hong, Z. Kalbarczyk, R.K. Iyer,(2016) “A Data-Driven Approach to Soil Moisture Collection and Prediction” IEEE Xplore, vol. 2, Issue 2, pp. (292-297)
- [5] Siti Khairunniza-Bejo, Samihah Mustaffha and Wan Ishak Wan Ismail(2014),” Journal of Food Science and Engineering 4”, Issue-4, pp. (1-9)
- [6] E. Manjula, S. Djodiltachoumy(2017),” International Journal of Computational Intelligence and Informatics”, Vol. 6: No. 4, pp. (298-305)
- [7] Rushika Ghadge, Juilee Kulkarni, Pooja More, Sachee Nene and Priya R L(2018), Prediction of Crop Yield using Machine Learning, IRJET, Vol. 5, Issue- 02, pp. (1-7)
- [8] Mitunkumar Balar and Ronakkumar Patel(2019), Crop Prediction Estimator, International Journal for Scientific Research & Development| Vol. 6, Issue 11, 2019 | ISSN (online): 2321-0613, pp. (438-441)
- [9] T. Venkat Narayana Rao, Gaddam Rishitha Reddy(2019), Prediction Of Soil Quality Using Machine Learning Techniques, INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 8, ISSUE 11, pp. (1309-1313)
- [10] T.R. Green, J.D. Salas, A. Martinez and R.H. Erskine(2007), Relating crop yield to topographic attributes using spatial analysis neural network and regression, Geoderma 139, pp. (23-37)
- [11] <https://www.investopedia.com/terms/m/mlr.asp>
- [12] https://en.wikipedia.org/wiki/Farmers%27_suicides_in_India
- [13] http://agricoop.nic.in/sites/default/files/AR_2018-19_Final_for_Print.pdf
- [14] Aditya Shastry, H.A Sanjayand E.Bhanushree(2017), “Prediction of crop yield using Regression Technique”, International Journal of computing 12 (2), ISSN:1816-9503, pp. (96-102).