May - June 2020
ISSN: 0193-4120 Page No. 4566-4570

Crop Yield Prediction

Aishwarya More[1], Amrutha H N[2], Chaitra B[3], Amilineni Sai Vinutha[4], Anusha J[5]

[1,2,3,4,5]School of Computing and IT, REVA University Bangalore, India
[1]aaishwaryamore@gmail.com, [2]amruthahn75@gmail.com, [3]chaitra.b@reva.edu.in,
[4]saivinutha123@gmail.com, [5]anushajha15@gmail.com

Article Info
Volume 83
Page Number: 4566-4570
Publication Issue:
May - June 2020

Abstract
The agriculture sector continues to be the backbone of Indian economy. Agricultural productivity has become a problem. This makes crop yield prediction a critical task. Data Mining and Machine Learning are emerging fields of research in crop yield analysis. It depends on varied parameters such as temperature, weather, nutrients percentages, soil type, PH value, regional rainfall, etc. Incorporating these factors into crop yield predictions is important. By analyzing these data attribute sand training the data with different machine learning algorithmic rule would boost crop yield prediction under different climatic circumstances. In this work, three supervised algorithms such as Random forest algorithm, Support Vector Machine and Artificial Neural Network are used. The performance of the three is compared and then the best model is used for prediction. The main aim of this project is to achieve optimum yield and to help framers sustain the crop. It also provides the information about maintaining the crop. This work gives a better prediction for the framers to have prior knowledge about crop production.

Article History
Article Received:*19 November 2019*
Revised: *27 January 2020*
Accepted: *24 February 2020*
Publication: *12 May 2020*

Keywords: *Data Mining, Machine Learning, Random Forest, Support Vector Machine, Artificial Neural Network.*

1. Introductions

Agriculture is the backbone of every country. In India there is a lot of population growth so indirectly there is demand for methods of food advance farming to meet the needs. Also today, agriculture is still a primary source of income in many broad areas of India. Agriculture is one of the reasons for the increase in Indian economic growth. In this paper we will provide a solution to yield predictions using different attributes such as PH value, location, percentage of nutrients, weather and temperature, form of soil, composition of soil, rainfall and etc.

Ancient people used to cultivate crops to their own needs. People nowadays have little knowledge of cultivation. This cultivation is also difficult due to climatic changes in soil, water and air. The use of artificial and hybrid products is also poor. Since it only provides 26 percent of the gdp output, we need to increase it by educating all farmers who use conventional methods to use smart methods to produce more yield and cost-effective crops in a very short time. This country is where 80% of the land is agricultural property, as farmers are lagging behind, so we need to teach them about maintaining the crop for best yield. Specific knowledge about the crop's history is important for decision-making about risk management of agriculture.

Machine learning and data mining literally help to solve all of the problems. Data mining is a process of finding the correlations and patterns within it among many data sets. Data mining also eliminates all unnecessary information from massive datasets. This method is carried out using preprocessing of the data. In agriculture machine learning is commonly used mainly to increase crop productivity and quality. This paper we're using machine learning algorithms to predict the most productive yield production in land relevant to agriculture. This method implements a model not only for crop yield prediction but also for crop maintenance. The ultimate aim of this is to provide a simple and understandable interface that shows framers to evaluate crop production, and to hold the crop for the future.

TEST Engineering & Management

May - June 2020
ISSN: 0193-4120 Page No. 4566-4570

2. Related Work

Shivam Bang et al.[1].proposed a method "Fuzzy logic based Crop Yield Prediction using Temperature and Rainfall parameters predicted through ARMA, SARIMA, and ARMAX models". In this work, three methods are used to predict weather- ARMA (Auto Regressive Moving Average),SARIMA(Seasonal Auto Regressive Moving Average) and ARMAX(ARMA with exogenous variable).The performance of the three compared and the best model for temperature prediction is the SARIMA model and ARMA model for rainfall. By using fuzzy logic technique crop yield is being predicted. The advantage of this is that it operates on a fixed range rather than discrete values. The disadvantage of this is that due to certain existing variables, climatic conditions change can affect the estimation of rainfall.

Rakesh Kumar et al.[2]. proposed a new method "Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique". The objective is to pick the best crop for agricultural planning. This paper proposed a method called Crop Selection Method (CSM) to provide a solution as to which crop can be selected in order to boost the crop's net yield rate over season and thereby achieve the country's maximum economic growth. The advantage of this model is that it improves the crop's net yield rate and also increases the land re-usability. The disadvantage of this is it uses maximum resource for cultivation.

Shriya Sahu et al.[3]. proposed a new method namely "An Efficient Analysis Of Crop Yield Prediction Using Hadoop Framework Based On Random Forest Approach". The objective of this research is to give a better prediction of which kind of crops to be grown in order to increase their productivity. In this proposed work the random forest algorithm is integrated with MapReduce programming model in Hadoop framework. The adavantage of using this method is the prediction allows the framers to increase the productivity of the plants. The main disadvantage is the processing of vast quantities of information requires additional equipment.

Neha Rale et al.[4]. proposed a method namely "Prediction of Crop Cultivation". This method propose to use machine learning techniques to build a crop yield prediction model. By comparing the performance of different linear and non-linear regressor models using 5-fold cross validation. e result showed that random forest regressor performed the best. The advantage of this algorithm is it has fixed number of hyper-parameters and by some adjustments of the parameters actually bought the high level in performance. The disadvantage is the model shows that over fitting still poses some problems.

R.Sujatha et al.[5]. proposed a new method "A Study on Crop Yield Forecasting Using Classification Techniques". The objective is to estimate the crop yield and select the most suitable crop for a given area. This proposed method uses previous crop data and apply data mining techniques and classification algorithms to predict the yield. The main advantage is it allows the farmer to select the best crop according to the climatic conditions. The disadvantage is that only few factors are considered which will affect the crop yield.

YogeshGandge et al.[6]. new method "A Study on Various Data Mining Techniques for Crop Yield Prediction" was proposed. The goal is to research the various techniques of data mining used to predict crop yields. This proposed method uses various algorithms applied on different crop and also mentioned its accuracy. The main advantage is various algorithms are used. The disadvantage is that a smaller dataset is considered.

Niketa Gandhi et al.[7]. new method "Rice Crop Yield Prediction in India using Support Vector Machines" was proposed. The objective is to predict crop yields for Indian rice cropping areas under different climatic scenarios using support vector machine algorithm. This proposed method uses the dataset of 27 districts of Maharashtra state, India and the results are produced using WEKA tool to apply SMO classifier on. The key benefit is the use of WEKA platform which is freely available and open source data mining tool. The drawback is that SMO had the lowest precision and worst efficiency.

Niketa Gandhi et al.[8]. proposed a new method "Rice Crop Yield Prediction Using Artificial Neural Networks ". The goal is to predict crop yields for Indian rice cropping areas using artificial neural network under different climatic scenarios. This proposed method uses dataset of 27 districts of the state Maharashtra, India and results are obtained by processing dataset using WEKA tool and developing Multilayer Perceptron Neural Network. The main advantage is that the result provided 97.5 percent accuracy with 96.3 percent sensitivity and 98.1 percent specificity. The disadvantage is that a linear approach is used.

S Pudumalar et al.[9]. proposed a "Crop Recommendation System for Precision Agriculture". Here we use precision farming, which uses certain data such as soil characteristics, soil type, and yield of the specific crop. In this paper this problem of selecting the best crop is solved via a recommendation method using Random tree, K-Nearest Neighbor and Naive Bayes technique. Advantage is a crop with high precision and productivity for the site specific parameters. Disadvantage is estimation during heavy rainfall and enormous changes in the atmosphere.

Tanhim Islam et al.[10]. proposed a "Deep Neural Network Approach for Crop Selection and Yield Prediction". Because we use artificial neural network it has several robust predictive and modeling methods. They have some performance prediction algorithms with others such as vector support system, logistic regression. Advantage is that we use random forest algorithms to compare the accuracy and rate of error. They also gathered 46 parameters in this paper about soil temperature and climate. Disadvantage is that they were

May - June 2020
ISSN: 0193-4120 Page No. 4566-4570

better at predicting temperature-change responses than changes in precipitation.

S Kanaga Subha Raja et al.[11]. proposed a system "Demand based crop recommender system for farmers". This paper is a tool used to predict crop yield and price by analyzing patters in the last data that the farmer gets for his land. Here process sliding window nonlinear regression techniques are based on variables such as rainfall, temperature, market prices, land area and past crop yield. The benefit is that it indicates the best crop that adapts to the demands of prevailing social crises. The disadvantage is confused in selecting the best crop during continuous fluctuation in time of high and low demand.

Mr A Suresh et al.[12]. proposed a "Prediction of major crop yields of Tamilnadu using K-means and Modified KNN". Since prediction plays an important role during demand here in this paper we use KNN and k-means for major crops of tamilnadu. We use tools like matlab and WEKA for clustering and classification. Advantage is use of data mining in preprocessing, data cleaning hence converting into understandable format. Disadvantage is having a lot of complicated data mapping problem using fuzzy algorithm that has vast rules.

3. Proposed system

Methodology

India being an agricultural country depends on agriculture for its economy. Because of the various atmospheric conditions, agricultural productivity is the biggest challenge. Framers ought to analyze these factors and take measures needed to boost the productivity of the crops. This work can facilitate framers use an efficient algorithm to predict the proper yield for the crop.

A. Data Description

The dataset used in this research is taken from the Indian Government's available records. The dataset are incorporated in to Microsoft Office Excel for further processing. The dataset is stored in the .csv format. Table 1 data is the sample data set that is used in the project. The data used for forecasting is focused on factors like state, district, year, production and etc. This data is given to different model of machine learning and it predicts performance by training the model.

Table 1: Sample Dataset of Crop data

State	District	Year	Production (Kg/Ha.)
West Bengal	Burdwan	2018-19	2842
West Bengal	Birbhum	2018-19	2743
West Bengal	Nadia	2018-19	2705
West Bengal	Hooghly	2018-19	2514

B. Architecture

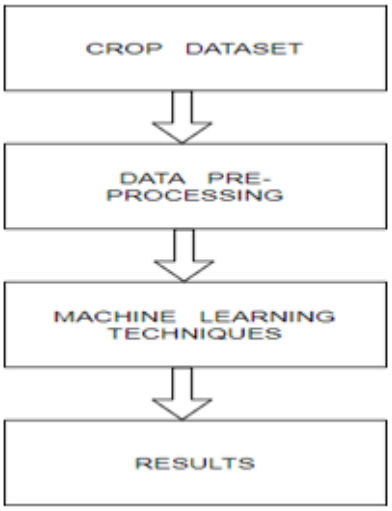

Figure 1: Block diagram

C. DataPre-Processing

The collected data is raw data in both structured and unstructured format. Instead of directly using this raw data cleaning is performed so that the data is made easy to train. In this unwanted data and fields are removed, detect and eliminate the noisy data, correct the inconsistent data and replace the missing data with mean, median or mode value in the dataset. The data are then divided into the train and the test data. We use scikit-learn and import_train_test_split to split the pre-processed data and it is done according to the weight given in the code.

D. Machine Learning Techniques

For prediction machine learning algorithm used are namely Random Forest algorithm, Support Vector Machine and Artificial Neural Network. We compare the performance of various models to find out which is the most accurate model for prediction.

a) Random Forest Algorithm

Random forest is an ensemble learning approach that is u sed for classification as well as regression. This algorithm is suitable for both large and small data, and provides an accurate prediction. The data gathered is fed into the algorithm as input. Each time the algorithm selects some random data (bootstrapping) with replacement (bagging) and then builds decision trees and tests the same prediction for several trees. This tests for which tress give the same output after that maximum trees giving the same output is considered as final output as shown in Fig.2.Every tree shall be categorized under different conditions on the basis of which it shall train the model and count the number of trees having the same output and having the highest number to be calculated as the output.

TEST Engineering & Management

May - June 2020
ISSN: 0193-4120 Page No. 4566-4570

The training dataset (t) is t= (cj,dk)
where j=1,2,3,……n and k=1,2,3, m

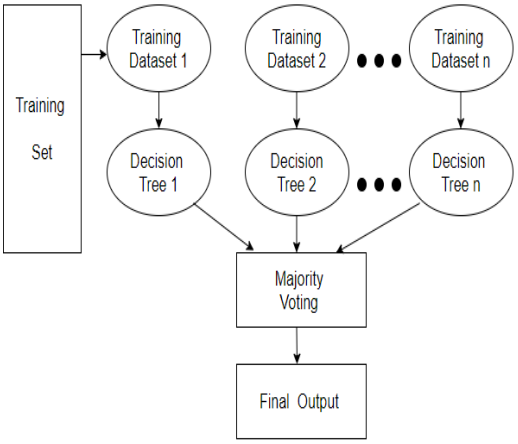

Figure 2: Working of Random Forest

b) Support Vector Machine

A support vector machine is a supervised type of machine learning. It is used to construct hyper plane in an N-dimensional space (N-the number of features). The data collected is used to map points in n-dimensional space. In a 2D plane hyper plane which is the line that divides a plane into 2 sections is constructed. On either side, the hyper plane contains classes. There are several possible ways to select the hyper plane to differentiate the two sets of data points; however we consider the plane with maximum margin which is the maximum distance between the data points in both classes. Data points may be assigned to different classes whose points are falling on either side of the hyper plane. Data points that are closer to hyper plane are Support Vectors.

w.u+b>=1 → + class w.u+b<=-1 → - class Yi(w.u+b)-1>0

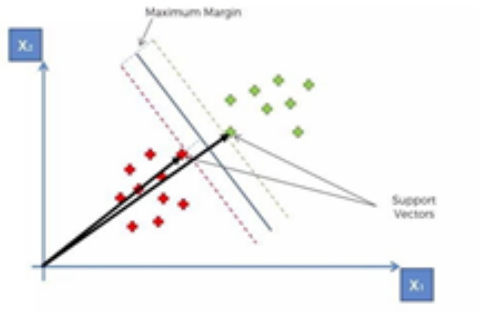

Figure 3: Support Vector Machine

c) Artificial Neural Network

Artificial Neural Networks (ANN) is a biological nervous system based data operative model. ANN is an interconnected collection of nodes, based on the simplification of neurons in the brain. The primary processing unit in this is the neurons, these neurons are bound in layers together. A feed forward method in a neural network performs in 3 layers primarily input layer, hidden layer and output layer. For every input function that is the node has one weight within the input layer. The weighted sum of its input is taken from nodes and therefore the network is equipped with special equations called activation functions. The result of the activation function is basically the anticipated output.

Dot product between inputs and weights Y.W=y1.w1+y2.w2+ yn.wn+b
Where b is the bias term which makes prediction if we don't have a non- zero term
Summation through activation function
f(x)=1/1+e^-(x)

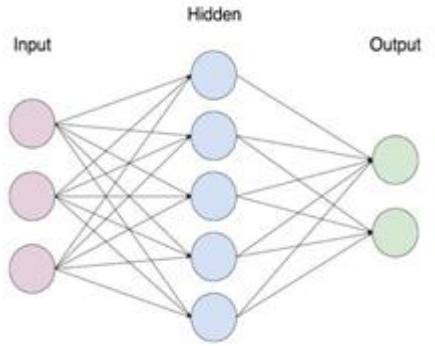

Figure 4: Artificial Neural Network

4. Results

In this paper an attempt is made to know the output of crop yield rate and also to sustain it for better yield. This is processed by using three algorithms i.e., Artificial Neural Network, Support vector machine and Random Forest. To predict the performance, these algorithms were experimented with on a crop from various regions across India. Also the algorithms are trained in such a way that it can brief on maintaining the crop for better yield. For predicting the performance all the models were compared. We compared the error rate during estimation of the output. It is obtained that ANN model has lower error rate than algorithms Random forest and SVM. The model relation by their error rate is plotted in the graph Fig. 5.The performance is achieved in such a way that the yield rate is estimated and it also defines the maintenance of the crop.

May - June 2020
ISSN: 0193-4120 Page No. 4566-4570

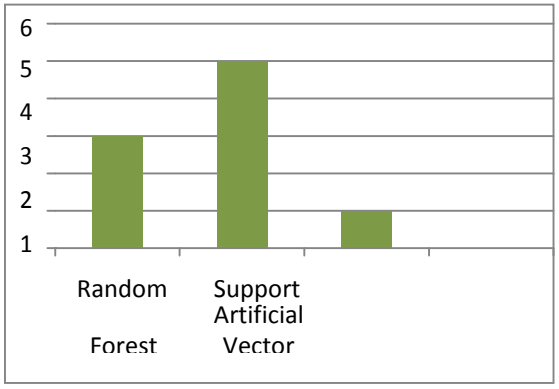

Figure 5: Comparison between all the three models.

5. Conclusion and Future Work

Crop yield prediction is successfully predicted and the effective algorithm is also found among the three algorithms and obtained the most effective performance with information of crop maintenance. It is possible to carry out a web application or smart phone application on this ideology in the future, which makes it easier for users to use this and can better understand it.

6. Acknowledgment

The result obtained will be assistance to farmers in understanding the crop yield and also helping to maintain the crop. He can use this prior knowledge of the crop to achieve better crop production which can lead to higher incomes.

References

[1] Shivam Bang, RajatBishnoi, Ankit Singh Chauhan, Akshay Kumar Dixit and Indu Chawla "Fuzzy Logic based Crop Yield Prediction using Temperature and Rainfall parameters predicted through ARMA, SARIMA, and ARMAX models" 2019IEEE.

[2] Rakesh Kumar, M.P. Singh, Prabhat Kumar and J.P. Singh "Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique" 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), Vel Tech Rangarajan Dr. Sagunthala R &D Institute of Science and Technology, Chennai, T.N., India.6- 8 May 2015.pp.138-145.

[3] Shriya Sahu, Meenu Chawla and Nilay Khare "An Efficient Analysis Of Crop Yield Prediction Using Hadoop Framework Based On Random Forest Approach" International Conference on Computing, Communication and Automation (ICCCA2017).

[4] Neha Rale, Raxitkumar Solanki, Doina Bein, James Andro-Vasko and Wolfgang Bein "Prediction of Crop Cultivation" 2019IEEE.

[5] R. Sujatha and Dr.P. Isakki Devi "A Study on Crop Yield Forecasting Using Classification Techniques" 2016IEEE.

[6] Yogesh Gandge and Sandhya "A Study on Various Data Mining Techniques for Crop Yield Prediction" 2017 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT).

[7] Niketa Gandhi, LeisaJ. Armstrong, Owaiz Petkar and Amiya Kumar Tripathy "Rice Crop Yield Prediction in India using Support Vector Machines" 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE).

[8] Niketa Gandhi, Owaiz Petkar and LeisaJ. Armstrong "Rice Crop Yield Prediction Using Artificial Neural Networks" 2016 IEEE International Conference on Technological Innovations in ICT ForAgriculture and Rural Development (TIAR2016).

[9] S. Pudumalar, E. Ramanujam, R. Harine Rajashreen, C. Kavyan, T. Kiruthikan and J. Nishan "Crop Recommendation System for Precision Agriculture" 2016 IEEE Eighth International Conference on Advanced Computing(ICoAC).

[10] Tanhim Islam, Tanjir Alam Chisty and Amitabha Chakrabarty "A Deep Neural Network Approach for Crop Selection and Yield Prediction in Bangladesh" 2018IEEE.

[11] S. Kanaga Suba Raju, Rishi. R, Sundaresan. E and Srijit. V "Demand Based Crop Recommender System for Farmers" 2017 IEEE International Conference on Technological Innovations in ICT ForAgriculture and Rural Development (TIAR2017).

[12] MrA Suresh, Dr. P. Ganesh Kumar and Dr.M. Ramalatha "Prediction of major crop yields of Tamilnadu using K-means and Modified KNN" Proceedings of the International Conference on Communication and Electronics Systems (ICCES 2018) IEEE Xplore Part Number:CFP18AWO-ART;ISBN:978-1-5386-4765-3.