

# AN APPROACH FOR PREDICTION OF CROP YIELD USING MACHINE LEARNING AND BIG DATA TECHNIQUES

**Kodimalar Palanivel**

Department of Computer Science,  
Bharathidasan University Constituent Arts & Science College,  
Navalurkuttapattu, Tiruchirappalli, TamilNadu, India

**\*Chellammal Surianarayanan**

Department of Computer Science,  
Bharathidasan University Constituent Arts & Science College,  
Navalurkuttapattu, Tiruchirappalli, TamilNadu, India

\*Corresponding Author

## ABSTRACT

*Agriculture is the primary source of livelihood which forms the backbone of our country. Current challenges of water shortages, uncontrolled cost due to demand-supply, and weather uncertainty necessitate farmers to be equipped with smart farming. In particular, low yield of crops due to uncertain climatic changes, poor irrigation facilities, reduction in soil fertility and traditional farming techniques need to be addressed. Machine learning is one such technique employed to predict crop yield in agriculture. Various machine learning techniques such as prediction, classification, regression and clustering are utilized to forecast crop yield. Artificial neural networks, support vector machines, linear and logistic regression, decision trees, Naïve Bayes are some of the algorithms used to implement prediction. However, the selection of the appropriate algorithm from the pool of available algorithms imposes challenge to the researchers with respect to the chosen crop. In this paper, an investigation has been performed on how various machine learning algorithms are useful in prediction of crop yield. An approach has been proposed for prediction of crop yield using machine learning techniques in big data computing paradigm.*

**Key words:** ISTA, IISTA, image restoration, inverse problems,  $l_0$  norm,  $l_1$  norm,  $l_2$  data fidelity term, regularization function, total variation.

**Cite this Article:** Kodimalar Palanivel and Chellammal Surianarayanan, An Approach for Prediction of Crop Yield Using Machine Learning and Big Data Techniques, *International Journal of Computer Engineering and Technology* 10(3), 2019, pp. 110-118.

<http://www.iaeme.com/IJCET/issues.asp?JType=IJCET&VType=10&IType=3>

---

## 1. INTRODUCTION

Agriculture proves to be a major factor of Indian economy and it involves production of crops. Crops may be either food crops or commercial crops. Food crops include paddy, wheat, maize, grams, millets, etc., whereas commercial crops are sugarcane cotton, groundnut, cashew, etc. The productivity of the crops is significantly influenced by weather conditions [1]. Hence, accurate yield prediction is a major problem that ought to be addressed. Early prediction of yield would facilitate the farmers to make precautionary actions to improve productivity. Early prediction is possible through collection of previous experience of the farmers, weather conditions and other influencing factors and; store it in a large database. The common input parameters are rainfall, temperature, humidity, solar radiation, crop population density, fertilizer application, irrigation, tillage, type of soil, depth, farm capacity, and soil organic matter. By applying data mining techniques such as prediction, classification and clustering, early decisions are possible.

### 1.1. Need for prediction

Estimating agricultural yield prior to harvest is an important issue in agriculture, as the changes in crop yield from year to year influence international business, food supply, and global market prices. Also, early prediction of crop yield provides useful information to policy planners. Appropriate prediction of crop productivity is required for efficient planning of land usage and economic policy. In recent times, forecasting of crop productivity at the within-field level has increased. The most influencing factor for crop productivity is weather conditions. If the weather based prediction is made more precise, then farmers can be alerted well in advance so that the major loss can be mitigated and would be helpful for economic growth. The prediction will also aid the farmers to make decisions such as the choice of alternative crops or to discard a crop at an early stage in case of critical situations. Further, predicting crop yield can facilitate the farmers to have a better vision on cultivation of seasonal crop and its scheduling. Thus, it is necessary to simulate & predict the crop yield before cultivation for efficient crop management and expected outcome. As there exists a non-linear relationship between crop yield and the factors influencing crop, machine learning techniques might be efficient for yield predictions.

## 2. MACHINE LEARNING TECHNIQUES

Machine Learning involves problems in which the input and output relationship is not known. Learning specifies the automatic acquirement of structural descriptions. In contrast to traditional statistical methods, machine learning does not make assumptions about the exact construct of the data model, which describes the data. This feature is very helpful to describe complex non-linear behaviors such as a crop yield prediction. Machine learning is a part of artificial intelligence employed to build an intelligent system [2]. By utilizing the training samples, the test samples can be identified. The accuracy of the system can be measured using metrics such as mean square error, root mean square error, precision, recall, sensitivity specificity etc. Further, machine learning can be employed to address a variety of applications including crop yield prediction [3] through supervised, unsupervised and reinforcement learning methods. Classification, clustering, regression, prediction are some of the techniques involved to attain the intelligent system. In this study, prediction is considered and the methods used for prediction are elaborated in the following subsection.

Meteorological conditions, such as precipitation, temperature, soil conditions, topography and socio-economic factors are responsible for about 30% growth of the crops. Several works were proposed in the literature for predicting the yield of crop using expert systems,

regression, , artificial neural network, data mining, support vector machine etc. The expert system such as fuzzy logic is based on logical rules to predict the yield. However, it requires wide interaction with the experts for obtaining the rules for prediction. In addition, these rules are based on certain set of input data [4].

## 2.1. Linear Regression

Prediction based on linear regression is discussed in many works. It is a statistical method applied over linear systems. Using this, the relationship between dependent and independent variables can be measured. If independent variable is having more than one input attribute, multiple regressions can be applied. Regression based models are used in prediction [5] as this technique shows consistent results during standard tests. Even though regression based models work fine for linear data, they do not fit for complex and non-linear data. Also, these models may not be able to perform better due to its limitation with regression assumptions for multiple co-linearity among the dependent and independent variables. Regression Models were used to Predict Tea Crop Yield [6]. The results revealed that most of the observed events were correctly predicted. An equation for predicting the yield of tea from the climate variables is presented. Analysis of crop yield prediction using Multiple Linear Regression (MLR) technique and Density based clustering technique in the East Godavari district of Andhra Pradesh is presented[7].

## 2.2. Artificial neural Networks

Artificial Neural Network (ANN) is another technique attempted by many researchers to address prediction. Back Propagation Neural Network (BPNN) is widely used among available neural networks. The back propagation algorithm involves three layers namely input, output and hidden layers. The weights are adjusted depending upon error. Multiple Linear Regression (MLR) and neural networks utilizing data from various tests conducted for three years were considered as predictive models for potato yield [8].

A 3-input NN using Leaf Area Index (LAI) functions produced accurate results and forecasts of potato yields. Neural Network is used for wheat prediction. [9]. The inputs were rainfall, guaranteed price, area under cultivation, subsidy, insured area etc. Crop prediction based on ANN is proposed to reduce losses when the conditions are not favorable [10]. The crops used were cotton, groundnut, rice, soya bean, corn, wheat etc. The parameters such as soil type, PH, nitrogen, depth, temperature and rainfall were taken into account for estimation. ANN models for predicting maize in the major maize producing provinces of South Africa is presented [11]. The prediction is performed using the climate variables namely; precipitation (PRE), maximum temperature (TMX), minimum temperature (TMN), Potential Evapotranspiration (PET), soil moisture (SM) and land cultivated (Land) for maize. Among the datasets used from 1990 to 2017, 80% were used for training and 20% were utilized for testing.

ANN solves the complex relations and strong nonlinearity between crop production and various predictor parameters. They contain objective mathematical functions rather than subjective rules and can be easily automated. Also, they produce outputs with acceptable accuracy for conditions which are not mentioned in the input. Further, they do not involve pre-established association and can be obtained using available data. ANN is considered to be the efficient technique for obtaining results from inexact, vague and non-linear data. ANN turned out to be an important tool for a large number of applications including crop production prediction. A maize yield prediction system based on ANN with soil and weather data as input parameters is proposed [12]. The output can be further tuned by using Convolutional Neural Networks (CNN) and Deep Neural Network (DNN) or deep learning.

Yield prediction of wheat from using both the spatial and temporal neural network models is presented [13]. Simulation results proved that the spatial NN model is capable of predicting the wheat yield with a high accuracy for a specific field area compared with the temporal models. An artificial neural network (ANN) technique for predicting the production of wheat using extensive data collected is discussed [14]. The ANN model was able to predict wheat production for various conditions and farming methods through direct and indirect technical factors with an error margin of  $\pm 9\%$  ( $\pm 0.89 \text{ t ha}^{-1}$ ).

### 2.3. Support vector Machine

Prediction can be possible through Support Vector Machine (SVM) developed by Vapnik (1998). SVM creates a hyperplane or set of hyperplanes in a high- or infinite dimensional space, which is utilized for regression, classification or other tasks. SVM uses linear functions for learning. In case of nonlinear cases, SVM uses a kernel technique to plot the data into a higher dimensional feature space, in which linear functions can be applied. The main advantage of SVM over other data mining techniques, such as ANN is its simplicity as it requires only a few parameters to be adjusted to optimize the model.

This technique finds wide applications in signal processing, time series analysis, weather prediction, crop productivity prediction etc. SVM is less vulnerable to over fitting problem as it is capable of mapping the functions to hyper plane. A study on the impact of climate factors such as temperature, rainfall, relative humidity, sunshine hours, daily temperature range, and rainy days are considered to predict paddy yield is proposed [15]. Further SVM Classifier is utilized to distinguish the plant as either crop or weed. The method is implemented to detect maize and weed plants. SVM-based Open Crop Model (SBOCM) for rice development stage and yield prediction is attempted [16].

Soil classification to identify suitable crop for the soil using SVM is proposed in [17]. The key issue in SVM modeling is the determination of the kernel functions, hyper parameters, and penalty coefficient. In [18], the application of SVM in analyzing aerial hyperspectral observations taken over a corn field to predict crop yield, biomass, plant height, and leaf-greenness is proposed. The results were compared with a stepwise regression method, and found to be better. Supervised learning algorithms such as KNN, SVM and LS-SVM are proposed for prediction of sugarcane yield in [19].

## 3. METRICS

From the above section, it is understood that different machine learning algorithms are being used for prediction of crop yield. Though different machine learning algorithms are available for use, selection of a particular algorithm is based on the nature of application and accuracy of prediction algorithm. Prediction accuracy of classifiers is validated by different metrics as *Mean Absolute Error (MAE)*[8], *Root Mean Squared Error(RMSE)*[3],[10], [14] , *Mean Absolute Error(MAE)* as discussed below.

### Mean Squared Error

Mean Squared Error(MSE) measures the average of squared error of predictions. That is, it calculates the square of different between the predicted valued and actual value and then averages those values. Computation of MSE is shown in (1)

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y}_i)^2 \quad (0)$$

In (0),  $y_i$  denotes the expected output and  $\bar{y}_i$  denotes the predicted output for an  $i^{\text{th}}$  data and 'i' ranges from 1 to N. Low value of MSE indicates that the accuracy of the classifier is good. For a perfect classifier, the value of MSE would be zero.

**Root Mean Squared Error (RMSE)**

Root Mean Squared Error(RMSE) is the square root of MSE. RMSE is expressed as in (2)

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \bar{y}_i)^2} \tag{2}$$

It is the standard *deviation* of the prediction *errors*. Initially, the difference between an actual target value and the model output value is calculated. This difference is then squared and averaged over all data items before the root of the mean value is computed.

*Mean Absolute Error (MAE)*

Mean Absolute Error(MAE) is computed as the average of absolute differences between the target values and predicted values. This is expressed as in (3)

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \bar{y}_i| \tag{3}$$

**4. PROPOSED APPROACH**

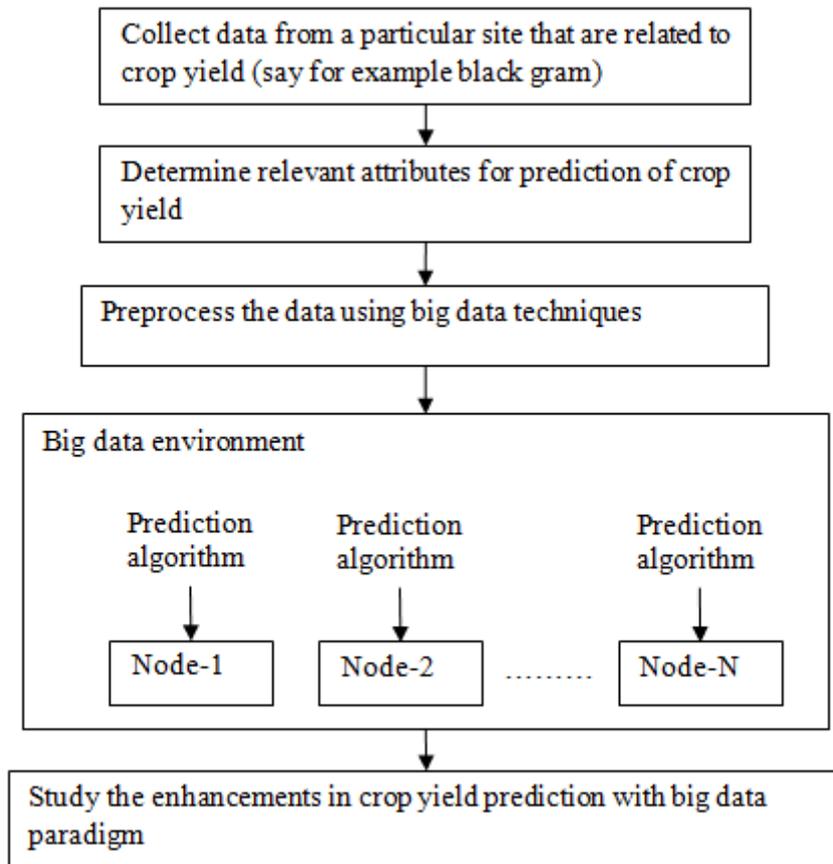
To provide an insight into how different research works have employed various machine learning algorithms, tools and data sets for predicting the yield of different crops are tabulated in Table 1.

S.No	Research reference	Crop	Factors	Method	Tool	Data sets
1	[11]	Maize	Precipitation, Maximum temperature, Minimum temperature, potential evapotranspiration, soil moisture, land cultivated	ANN	Neural net package in R software	CRU TS 3.24.01
2	[5]	Pepper, common bean, corn, potato, tomato	Irrigation water depth, Solar radiation, rainfall, temperature, relative humidity	ANN, MLR, SLR, regression tree	SPSS, WEKA	Agricultural production data from spriter-GIS system and weather information data from SMN
3	[12]	Rice	Precipitation, Minimum, average and maximum temperature, Evapotranspiration, Area,	ANN	WEKA	1998-2002 Maharashtra government records
4	[20]	Rice	precipitation, minimum temperature, average temperature, maximum temperature and reference crop evapotranspiration, area, production and yield	SVM	WEKA	publicly available Indian Government records
5	[21]	Rice	Soil code, section thickness, soil composition entropy,	SVM	v2.1 software in the	weather observation stations in China (i.e., daily published

			organic matter, pH, nitrogen, phosphorus, potassium, air pressure, average temperature, average relative humidity, precipitation, wind speed		MATLAB	meteorological data and the 1:1000000 soil database published by the Chinese Academy of Sciences [CAS]
6	[22]	Maize, Soyabean, Sugar beet	maximal (Tmax), minimal (Tmin) and average (Tavg) monthly air temperatures, precipitation and evapotranspiration	SVM	R package e1071	(from 1999 to 2008) in Serbian province of Vojvodina, from the internal database of the Department of field and vegetable crops at the Faculty of Agriculture in Novi Sad
7	[23]	Prediction of suitable crop according to the factors	Temperature, rainfall, soil parameters	Multiple Linear Regression	----	Indian Meteorological Department
8	[24]	Rice paddy, cotton, sugarcane, groundnut, black gram	Climate variables	SVM, AdaSVM, NaiveBayes, AdaNaive		Indianwaterportl.org
9	[25]	Prediction of suitable crop according to soil quality, soil nutrients and soil composition	Nutrients present in the soil	Kohonen Self Organizing Map and Back Propagation Network		Not mentioned

From the literature, it is understood that the weather inputs such as rainfall, temperature, humidity in addition to non-weather inputs namely soil moisture, pH, salts in soil (nitrogen, phosphate, potassium, organic carbon, calcium, magnesium, sulphur etc.), crop type and seed variety are fed as input to the prediction model. Initially the raw data is preprocessed to remove erroneous and unformatted noisy data. Then, the processed input is furnished to the prediction model. By comparing the performance metrics of different machine learning models, Artificial Neural Network and Support Vector Machine based prediction models are found to be more suitable for crop yield prediction. Further, from literature, it is found that with the digital advancements in the field of agriculture a large amount of data is being produced constantly. Hence, agriculture data has entered the world of big data[26]. With this idea in mind, in this paper, an approach is proposed to analyze crop yield prediction in big data environment. The block diagram of the proposed approach is shown in Fig. 1.

As in Fig. 1, it is proposed to using big data techniques at two different stages, namely, (i) preprocessing and (ii) prediction. When prediction is taking place in different data nodes with a parallel programming model, it may enhance the performance of prediction. So, it is proposed to determine how big data techniques are useful in enhancing the prediction of crop yield.



**Figure 1** Approach to predict the yield of black gram yield with big data tools and techniques

## 5. CONCLUSION

An extensive study on the crop prediction is made. The literatures reveal diverse machine learning techniques adopted for prediction of crop yield. Further, the performance metrics of the machine learning algorithms such as root mean square error are studied. Along with machine learning algorithms for prediction, it is planned to study the impact of big data techniques in the prediction of crop yield. A conceptual approach is proposed for the same. The proposed approach is being implemented.

## REFERENCES

- [1] Food and Agriculture Organization of the United Nations (FAO). World Food Situation: FAO Cereal Supply and Demand Brief. Available online: <http://www.fao.org/worldfoodsituation/csdb/en/> (accessed on 17 September 2018).
- [2] AnnaChlingaryan, SalahSukkarieh, BrettWhelan, “Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review”, Computers and Electronics in Agriculture, Volume 151, August 2018, Pages 61-69
- [3] S. Veenadhari , Bharat Misra , CD Singh,” Machine learning approach for forecasting crop yield based on climatic parameters “,International Conference on Computer Communication and Informatics, 2014, Coimbatore, India
- [4] M. A. Jayaram and Netra Marad,” Fuzzy Inference Systems for Crop Yield Prediction”, Journal of Intelligent Systems, 2012,21(4),pp.363-372.
- [5] Gonzalez-Sanchez, A.; Frausto-Solis, J., Ojeda-Bustamante, W.,” Attribute selection impact on linear and nonlinear regression models for crop yield prediction”, Sci. World Journal, 2014.

- [6] Kenya Betty J., Sitienei , Shem G. Juma, and Everline Opere, "On the Use of Regression Models to Predict Tea Crop Yield Responses to Climate Change: A Case of Nandi East, Sub-County of Nandi County", MDPI Sensors, 2017
- [7] D Ramesh and B Vishnu Vardhan, "Analysis Of Crop Yield Prediction Using Data Mining Techniques", International Journal of Research in Engineering and Technology, 2015, 4(1)
- [8] Fortin JG, Anctil F, Parent L, Bolinder MA, "Site- specific early season potato yield forecast by neural network in Eastern Canada", Precision Agriculture, Springer, 2011,12(6), 905-923
- [9] Ghodsi, R.; Yani, R.M.; Jalali, R.; Ruzbahman, M, "Predicting wheat production in Iran using an artificial neural networks approach", Int. J. Acad. Res. Bus. Soc. Sci. 2012, 2, 34–47
- [10] Snehal S.Dahikar, Dr.Sandeep V.Rode, "Agricultural Crop Yield Prediction Using Artificial Neural Network Approach", International Journal Of Innovative Research In Electrical, Electronics, Instrumentation And Control Engineering, 2(1),2014, 683-686
- [11] Omolola M. Adisa ,Joel O. Botai, Abiodun M. Adeola , Abubeker Hassen et. al "Application of Artificial Neural Network for Predicting Maize Production in South Africa", Sustainability, 2019, 11(4), 1145
- [12] Niketa Gandhi, Owaiz Petkar and Leisa J. Armstrong "Rice Crop Yield Prediction Using Artificial Neural Networks", IEEE International Conference on Technological Innovations in ICT For Agriculture and Rural Development, 2016, 105-110.
- [13] William W. Guo and Heru Xue, "Crop Yield Forecasting Using Artificial Neural Networks: A Comparison between Spatial and Temporal Models", Mathematical Problems in Engineering, Hindawi, 2014
- [13] Suhas S Athani and CH Tejeshwar, "Support Vector Machine-Based Classification Scheme of Maize Crop" , IEEE Conference, 2017,84-88.
- [14] M. Safa1, S. Samarasinghe, and M. Nejat, "Prediction of Wheat Production Using Artificial Neural Networks and Investigating Indirect Factors Affecting It: Case Study in Canterbury Province, New Zealand", J. Agr. Sci. Technology, 2015, Vol. 17, 791-803
- [15] Suhas S Athani and CH Tejeshwar, "Support Vector Machine-Based Classification Scheme of Maize Crop" , IEEE Conference, 2017,84-88.
- [16] Monali Paul, Santosh K. Vishwakarma and Ashok Verma, "Analysis of Soil Behaviour and Prediction of Crop Yield using Data Mining Approach", IEEE International Conference on Computational Intelligence and Communication Networks, 2015,766-771.
- [17] Rohit Kumar Rajak, Ankit Pawar, Mitalee Pendke , Pooja Shinde, Suresh Rathod and Avinash Devare, " Crop Recommendation System to Maximize Crop Yield using Machine Learning Technique", International Research Journal of Engineering and Technology, 2017, 4( 12), 950-953.
- [18] Y. Karimi, S.O. Prasher , A. Madani and S. Kim, " Application of support vector machine technology for the estimation of crop biophysical parameters using aerial hyperspectral observations", Canadian Biosystems Engineering, 2008, Vol.50
- [19] Arun Kumar, Naveen Kumar, Vishal Vats, Arun Kumar, Naveen Kumar, Vishal Vats, "Efficient Crop Yield Prediction Using Machine Learning Algorithms", International Research Journal of Engineering and Technology, 05(06),2018.
- [20] Niketa Gandhi, Leisa J Armstrong, RiPtakar, Owaiz Petkar, Amiya Kumar, "Rice crop yield prediction in India using support vector machines", International Joint Conference on Computer Science and Software Engineering, 2016, IEEEexplore

- [21] Su YX, Xu H, Yan LJ, “Support vector machine-based open crop model (SBOCM): Case of rice production in China. Saudi J Biol Sci. 2017;24(3):537–547. doi:10.1016/j.sjbs.2017.01.024
- [22] Sanja Brdar, Dubravko Culibrk, Branko Marinkovic, Jovan Crnobarac and Vladimir Crnojevic, “Support Vector Machines with Features Contribution Analysis for Agricultural Yield Prediction”, in the Proc. of Second International Workshop on Sensing Technologies in Agriculture, Forestry and Environment, 2011.
- [23] D.S.Zingade, Omkar Buchade, Bilesh Mehta, Shubhm Ghodekar, Chandan Mehta, “Crop prediction System using Machine Learning”, International Journal of Advance Engineering and Research Development, vol. 4, Special Issue 5, Dec 2017, pp. 1-6
- [24] Narayanan Balakrishnan, Dr. Govindarajan Muthukumarasamy, “Crop production-ensemble machine learning model for prediction, International Journal of Computer Science and Software Engineering, vol. 5, Issue 7, July 2016, pp. 148-153
- [25] Rushika Ghadge, Juilee Kulkarni, Pooja More, Sachee Nene, Priya R L, “Prediction of Crop Yield using Machine Learning”, International Research Journal of Engineering and Technology, vol. 5, Issue 2, Feb-2018, pp.2237-2239
- [26] Athmaja S., Hanumanthappa M, “Applications of Mobile Cloud Computing and Big data Analytics in Agriculture Sector: A survey”, October 2016