

Efficient Crop Yield Prediction in India using Machine Learning Techniques

Payal Gulati¹

¹Department of Computer Engineering,
J.C. Bose University of Science & Technology,
YMCA, Faridabad

Suman Kumar Jha²

²Department of Computer Engineering,
Mangalmay Institute of Engineering and Technology,
MIET, Gr.Noida

Abstract:-Today Agriculture Sector is a major contributor to Indian Economy. In a country like India, which has ever increasing demand of food due to rising population, advances in agriculture sector are required to meet the needs. Therefore Crop Yield Prediction remains a challenging task in this domain. There are various parameters that affect the yield of crop like rainfall, temperature, fertilizers, pesticides, pH level, and other atmospheric conditions and parameters. Accurate yield prediction is required to be done after understanding the functional relationship between yield and these parameters. For this many researchers have applied machine learning algorithms on comprehensive datasets for predicting crop yield. This paper discusses various machine learning approaches towards crop yield prediction in India. Further in this work, Machine learning approaches have been executed on the agricultural data to evaluate the best performing technique.

1. INTRODUCTION

Agriculture is the main occupation for the people of India, covering 60% of the nation land and catering the basic needs of 1.2 billion people [1]. For the benefit of the farmers, modernization of agriculture procedures is carried out today. The crop yield or production majorly depends on the weather conditions, environmental changes, rainfall (which at times is uncertain), water management, and the utilization of pesticides. Therefore farmers are not able to accomplish expected yield of crop. Now a days data mining, machine learning as well as deep learning approaches are used by various researchers to enhance and improve the yield of crop and their quality [11,12].

Machine Learning can gain proficiency with the machine without characterized computer programming, so it improves machine execution by distinguishing and portraying the consistency and pattern of drive information. In this work various machine learning approaches such as Linear Regression, Gradient Boosting Regressor, Random Forest Regressor, Decision Tree Regressor, Polynomial Regression, Ridge Regression have been used for yield prediction on crop yield dataset of different states and considering varied crops.

The paper is organized as follows: Section II covers the related work in the area of data mining and Machine learning. Section III covers the detailed framework of crop yield prediction. The discussion and results are included in section IV. Finally, Section V concludes the research work.

2 RELATED WORK

In agriculture, Machine Learning is considered as a novel field, as variety of work has been done with the help of machine learning in the field of agriculture. There are different philosophies made and evaluated by the researchers all through the world in the field of agriculture and related sciences.

CH. Vishnu VardhanChowdary, Dr.K.Venkataramana [2], developed id3 algorithm for getting improved and great quality of crop yield of Tomato and is executed in Php platform and datasets are used as csv. Temperature, area, humidity and the production of tomato crop are the different parameters used in this study. R. Sujatha and P. Isakki [3], utilizes data mining techniques for prediction. This model worked on different parameters such as crop name, land area, soil type, pH value, seed type, water and also foreseen the boom and diseases of plants and in this way empowered to choose the descent crop dependent on climatic data and required parameters. N. Gandhi, L. J. Armstrong, O. Petkar and A. K. Tripathy [4], proposed the SVM for crop yield prediction of rice. In this method, dataset used consists of different parameters such as place, temperature, precipitation and manufacturing. On this dataset, the implemented classifier is sequential minimal optimization. They prepared the dataset through Weka tool to manufacture the set of rules on current dataset. In python, by using SVM algorithm outcomes were produced. S. Veenadhari, B. Misra and C. Singh [5], have built up an interactive site for finding the influence of climate and production of crop by utilizing c4.5 algorithm called Crop Advisor. Dependent on c4.5 algorithm, decision tree and ruled have been developed. It gives the idea how crop growth is affected by different climatic parameters. The data with respect to the related years environmental parameters like rainfall, temperature were gathered. The choices were dependent on the zone under the picked crop. Jun Wu, Anastasiya Olesnikova, Chi-Hwa Song, Won Don Lee [6], proposed selection tree which is fit for grouping all styles of farming records. A decision tree classifier turned into proposed for information of agriculture. It utilizes new facts and can address each and in whole record. 10-fold cross validation method is utilised to check dataset, horse-colic and soyabean dataset. Kiran Mai, C., Murali Krishna, I.V, A.VenugopalReddy [7], explained in their study that how data mining is incorporated with the other farming data such as meteorological data, usage of pesticides are useful for soothing out of use of pesticides. Topical information related to the business of agriculture which has contiguous

properties was represented. Verheyen, K., Adrianens, M. Hermly and S. Deckers [8], explained statistical mining techniques in their study as they are regularly used to view the characteristics of soil. As kmeans is utilized for sectioning soils in blend with GPS based innovation.

2. FRAMEWORK

In this work, different states data consisting of varied crops are taken into consideration. Supervised learning is utilized for modelling, which gives the predicted yield and their order of production. The various steps of the proposed framework are discussed in following sub sections.

A. Dataset Collection: Data is gathered from various sources [9,10] and then analyzed and prepared. This data is utilized for descriptive analysis. The dataset used in this paper consists of various states (Maharashtra, UP, West Bengal, Gujarat etc), different types of crops (sugarcane, coconut, wheat, gram etc), different seasons (Kharif, Rabi, Whole, Summer etc), different crop years and other parameters such as Rainfall, Temperature, pH, Humidity.

B. Preprocessing the data: In this module, dataset is preprocessed so as to fill the missing values, the fitting information run and separating the usefulness.

C. Feature Extraction: Feature extraction ought to streamline the amount of data required to represent a huge dataset. Its goal is to extract useful characteristics from data. The characteristics include high, low and mean temperature, air humidity, soil pH, rainfall.

D. Split dataset into Train and Test set: This step includes training and testing of the input data. The stacked information is isolated into two sets, such as preparing and testing the data. Training set is mapped with the training set and during the training phase data is to be testing after learning from previous observations. The final data is formed and is processed by machine learning module.

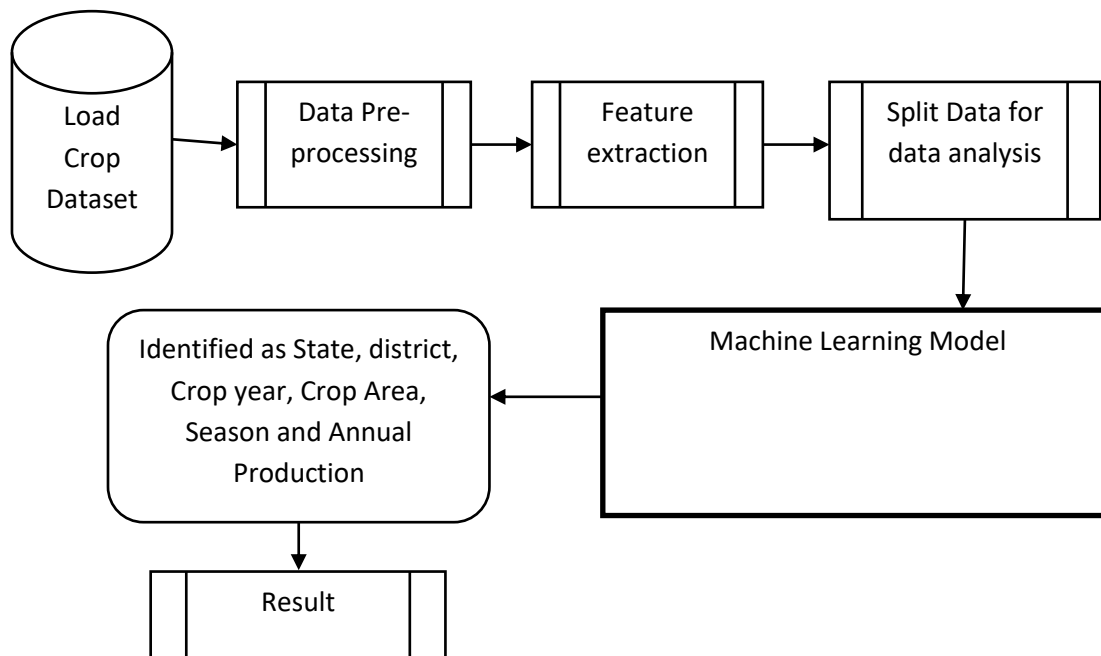
E. Apply Machine Learning Techniques: In our project, different supervised machine learning techniques for prediction of crop yield are used which is given as follows in Figure 3.1

3.1 Framework for Crop Yield Prediction

Results and Discussion

This section describes the outputs obtained after implementation of ML algorithms on the dataset obtained. Different machine algorithms such as Linear Regression, Gradient Boosting Regressor, Random Forest Regressor, Decision Tree Regressor, Polynomial Regression, Ridge Regression are applied on dataset using python programming. The different parameter set for these techniques were mean absolute error, mean squared error, root mean square error, R-square and cross validation which are used to estimate their efficiency of methods.

The formulae for calculating parameters are:



$$MAE = \frac{\sum_{i=1}^n |y_i - x_i|}{n} = \frac{\sum_{i=1}^n |e_i|}{n}$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \tilde{y}_i)^2$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \tilde{y}_i)^2}$$

$$r^2 = 1 - \frac{\sum (y_i - \tilde{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

After the analysis it is observed that Gradient Boosting Regressor is giving more accuracy with cross validation runs as 87.9% as shown in Table 1

	Test MAE	Test MSE	Test RMSE	Test R-Square	CrossValidation
Gradient Boosting Regressor	23.744808	4.676468e+05	683.847059	0.796290	0.879121
Ridge Regression	83.371910	2.289880e+06	1513.235102	0.002513	0.001220
Random Forest Regressor	19.437881	7.771160e+05	881.541815	0.661483	0.767536
Decision Tree Regressor	22.049008	1.137754e+06	1066.655599	0.504387	0.730118
Linear Regression	83.371914	2.289880e+06	1513.235103	0.002513	0.001220
Polynomial Regression	146.157727	2.236869e+06	1495.616746	0.025605	0.001220

different errors and cross validation is to be done for obtaining accuracy. Here, Gradient Boosting Regressor is giving more accuracy with cross validation runs as 87.9% when target variable is "Yield" but when target variable is "Production", the Random Forest Regressor is providing more cross validation accuracy of 98.9%. This framework will assist to reduce the issues faced by farmers and will serve as delegate to provide farmers with the information they need to gain high and maximize the profits.

REFERENCES

- [1] <https://www.hilarispublisher.com/open-access/agriculture-role-on-indian-economy-2151-6219-1000176.pdf>.
- [2] CH. Vishnu Vardhanchowdary, Dr.K.Venkataramana, "Tomato Crop Yield Prediction using ID3", March 2018,IJIRT Volume 4 Issue 10 pp,663-62.
- [3] R. Sujatha and P. Isakki, "A study on crop yield forecasting using classification techniques" 2016 International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16), Kovilpatti, 2016, pp. 1-4.
- [4] N. Gandhi, L. J. Armstrong, O. Petkar and A. K. Tripathy, "Rice crop yield prediction in India using support vector machines" 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE), KhonKaen, 2016, pp. 1-5.
- [5] S. Veenadhari, B. Misra and C. Singh, "Machine learning approach for forecasting crop yield based on climatic parameters" 2014 International Conference on Computer Communication and Informatics, Coimbatore, 2014, pp. 1-5.

- [6] Jun Wu, Anastasiya Olesnikova, Chi-Hwa Song, Won Don Lee (2009), "The Development and Application of Decision Tree for Agriculture Data" IITSI, pp 16-20.
- [7] KiranMai,C., Murali Krishna, I.V, an A.VenugopalReddy, "Data Mining o f Geospatial Database for Agriculture Related Application", Proceedings of Map India,New Delhi, 2006,pp 83-96.
- [8] Verheyen, K., Adrianens, M. Hermy and S.Deckers(2001),"High resolution continuous soil classification using morphological soil profile descriptions"Geoderma, 101:31-48.
- [9] <https://www.kaggle.com/>
- [10] <https://data.gov.in/>
- [11] P.Priya, U.Muthaiah, M.Balamurugan, "Predicting yield of the crop using machine learning Algorithm", IJESRT et al., 7(4): April-2018, pp 2277-2284
- [12] Georg Ruß, Rudolf Kruse, Martin Schneider, and Peter Wagner. "Estimation of neural network parameters for wheat yield prediction" In Max Bramer, editor, Artificial Intelligence in Theory and Practice II, volume 276 of IFIP International Federation for Information Processing, pages 109–118. Springer, July 2008.