CROP YIELD PREDICTION USING MACHINE LEARNING

Article *in* International Journal of Science and Research (IJSR) · April 2020

CITATIONS
0

READS
367

4 authors, including:

Mayank Champaneri
K J Somaiya Institute of Engineering and Information Technology
2 PUBLICATIONS **1** CITATION

SEE PROFILE

CROP YIELD PREDICTION USING MACHINE LEARNING

Mayank Champaneri
Department of Information technology,
K.J. Somaiya Institute of Engineering and Information Technology,
University of Mumbai, India.
m.champaneri@somaiya.edu

Chaitanya Chandvidkar
Department of Information technology,
K.J. Somaiya Institute of Engineering and Information Technology,
University of Mumbai, India
chaitanya.c@somaiya.edu

Darpan Chachpara
Department of Information technology,
K.J. Somaiya Institute of Engineering and Information Technology,
University of Mumbai, India.
darpan.c@somaiya.edu

Mansing Rathod
Department of Information technology,
K.J. Somaiya Institute of Engineering and Information Technology,
University of Mumbai, India
rathodm@somaiya.edu

***Abstract*— The impact of climate change in India, most of the agricultural crops are being badly affected in terms of their performance over a period of the last two decades. Predicting the crop yield in advance of its harvest would help the policy makers and farmers for taking appropriate measures for marketing and storage. This project will help the farmers to know the yield of their crop before cultivating onto the agricultural field and thus help them to make the appropriate decisions. It attempts to solve the issue by building a prototype of an interactive prediction system. Implementation of such a system with an easy-to-use web based graphic user interface and the machine learning algorithm will be carried out. The results of the prediction will be made available to the farmer. Thus, for such kind of data analytics in crop prediction, there are different techniques or algorithms, and with the help of those algorithms we can predict crop yield. Random forest algorithm is used. By analysing all these issues and problems like weather, temperature, humidity, rainfall, moisture, there is no proper solution and technologies to overcome the situation faced by us. In India, there are many ways to increase the economic growth in the field of agriculture. Data mining is also useful for predicting crop yield production. Generally, data mining is the process of analysing data from various viewpoint and summarizing it into important information. Random forest is the most popular and powerful supervised machine learning algorithm capable of performing both classification and regression tasks, that operate by constructing a multitude of decision trees during training time and generating output of the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.**

Keywords*— *Agriculture, Machine Learning, crop-prediction, Supervised Algorithms, Crop yield, Data Mining.

I. INTRODUCTION

Agriculture is the backbone of the Indian economy. In India, agricultural yield primarily depends on weather conditions. Rice cultivation mainly depends on rainfall. Timely advice to predict the future crop productivity and an analysis is to be made in order to help the farmers to maximize the crop production of crops. Yield prediction is an important agricultural problem. In the past farmers used to predict their yield from previous year yield experiences. Thus, for this kind of data analytics in crop prediction, there are different techniques or algorithms, and with the help of those algorithms we can predict crop yield. Random forest algorithm is used. Using all these algorithms and with the help of inter-relation between them, there are growing range of applications and the role of Big data analytics techniques in agriculture. Since the creation of new innovative technologies and techniques the agriculture field is slowly degrading. Due to these, abundant invention people are concentrated on cultivating artificial products that are hybrid products where there leads to an unhealthy life. Nowadays, modern people don't have awareness about the cultivation of the crops at the right time and at the right place. Because of these cultivating techniques the seasonal climatic conditions are also being changed against the fundamental assets like soil, water and air which lead to insecurity of food. By analysing all these issues and problems like weather, temperature and several factors, there is no proper solution and technologies to overcome the situation faced by us. In India, there are several ways to increase the economic growth in the field of agriculture. There are multiple ways to increase and improve the crop yield and the quality of the crops. Data mining is also useful for predicting crop yield production. The main objectives are

a. To use machine learning techniques to predict crop yield.
b. To provide easy to use User Interface.
c. To increase the accuracy of crop yield prediction.
d. To analyse different climatic parameters (cloud cover, rainfall, temperature)

II. LITERATURE REVIEW

In [1] Predicting yield of the crop using machine learning algorithm. International Journal of Engineering Science Research Technology. This paper focuses on predicting the yield of the crop based on the existing data by using Random Forest algorithm. Real data of Tamil Nadu were used for building the models and the models were tested

with samples. Random Forest Algorithm can be used for accurate crop yield prediction.

In [2] Random forests for global and regional crop yield prediction. PLoS ONE Journal. Our generated outputs show that RF is an effective and adaptable machine-learning method for crop yield predictions at regional and global scales for its high accuracy and precision, ease of use, and utility in data analysis. Random Forest is the most efficient strategy and it outperforms multiple linear regression (MLR).

In [3]. Crop production Ensemble Machine Learning model for prediction. International Journal of Computer Science and Software Engineering (IJCSSE). In this paper, AdaNaive and AdaSVM are the proposed ensemble model used to project the crop production over a time period. Implementation done using AdaSVM and AdaNaive. AdaBoost increases efficiency of SVM and Naive Bayes algorithm.

In [4]. Machine learning approach for forecasting crop yield based on parameters of climate. The paper provided in International Conference on Computer Communication and Informatics (ICCCI). In the current research a software tool named Crop Advisor has been developed as a user friendly web page for predicting the influence of climatic parameters on the crop yields.C4.5 algorithm is used to produce the most influencing climatic parameter on the crop yields of selected crops in selected districts of Madhya Pradesh. The paper is implemented using Decision Tree.

In[5]. Prediction On Crop Cultivation. International Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE) Volume 5, Issue 10, October 2016. Presently, soil analysis and interpretation of soil test results is paper based. This in one way or another has contributed to poor interpretation of soil test results which has resulted into poor recommendation of crops, soil amendments and fertilizers to farmers thus leading to poor crop yields, micro-nutrient deficiencies in soil and excessive or less application of fertilizers. Formulae to Match Crops with Soil, Fertilizer Recommendation.

In [6]. Analysis of Crop Yield Prediction by making Use Data Mining Methods. IJRET: The paper provided in International Journal of Research in Engineering and Technology. In this paper the main aim is to create a user-friendly interface for farmers, which gives the analysis of rice production based on the available data. For maximizing the crop productivity various Data mining techniques were used to predict the crop yield. Such as K-Means algorithm to forecast the pollution factor in the atmosphere.

In [7]. Applications of Machine Learning Techniques in Agricultural Crop Production. Indian Journal of Science and Technology, Vol 9(38), DOI:10.17485/ijst/2016/v9i38/95032, October 2016.
From GPS based colour images is provided as an intensified indistinct cluster analysis for classifying plants, soil and residue regions of interest. The paper includes various parameters which can help the crop yield for better enhancement and ratio of the yield can be increased during cultivation.

In [8] In this paper, we present a comprehensive review of research dedicated to the application of machine learning in agricultural production systems. Machine learning (ML) has emerged together with big data technologies, techniques, methods and high-performance computing to generate new opportunities to unravel, quantify, and analyse data intensive processes in agricultural operational sectors. By using Support Vector Machines (SVP) the Paper is Implemented.

In [9]. A Study to Determine Yield for Crop Insurance using Precision Agriculture on an Aerial Platform. Symbiosis Institute of Geoinformatics Symbiosis International University 5th & 6th Floor, Artur Centre, Gokhale Cross Road, Model Colony, Pune – 411016. Precision agriculture (PA) is the application of geospatial methodologies and remote sensors to identify variations in the field and to deal with them using different strategies. The causes of variability of crop growth in an agricultural field might be due to crop stress, irrigation practices, incidence of pest and disease etc. The Paper is Implemented using Ensemble Learning (EL).

In [10]. Random Forests for Global and Regional Crop Yield Predictions. institute on the Environment, University of Minnesota, St. Paul, MN 55108, United States of America. The generated outputs show that RF is an effective and different machine-learning method for crop yield predictions at regional and global scales for its high accuracy. The Paper is Implemented using k-nearest neighbour, Support Vector Regression (SVG).

III. METHODOLOGY

Data is a very important part of any Machine Learning System. To implement the system, we decided to focus on Maharashtra State in India. As the climate changes from place to place, it was necessary to get data at district level. Historical data about the crop and the climate of a particular region was needed to implement the system. This data was gathered from different government websites. The data about the crops of each district of Maharashtra was gathered from www.data.gov.in and the data about the climate was gathered from www.imd.gov.in. The climatic parameters which affect the crop the most are precipitation, temperature, cloud cover, vapour pressure, wet day frequency. So, the data about these climatic parameters was gathered at a monthly level.
Dataset Collection: In this phase, we collect data from various sources and prepare datasets. And the provided dataset is in the use of analytics (descriptive and diagnostic). There are several online abstracts sources such as Data.gov.in and indiastat.org. For at least ten years the yearly abstracts of a crop will be used. These datasets usually accept behaviour of anarchic time series. Combined the primary and necessary abstracts. Random Forests for Global and Regional Crop Yield Predictions.
Data Partitioning: The Entire dataset is partitioned into 2 parts: for example, say, 75% of the dataset is used for training the model and 25% of the data is set aside to test the model.
To predict future events Machine Learning Algorithms: Supervised learning: Supervised machine learning algorithms can apply what has been learned in the past to new data using labelled examples. After Sufficient training the system can provide targets for any new input. IN order to change the model accordingly the learning algorithm can also differentiate its results with the correct, intended output and find errors. Unsupervised learning: IN comparison, unsupervised machine learning algorithms are used when the information used to train is

neither labelled nor classified. Unsupervised learning does analysis of how systems can infer a function to describe a hidden structure from unlabelled data. In order to describe hidden structures from unlabelled data the system doesn't figure out the right output, but it examines the data and can draw inferences from datasets.

Random Forest Classifier: Random forest is the most popular and powerful supervised machine learning algorithm capable of performing both classification and regression tasks, that operate by constructing a multitude of decision trees at the time of training and generating outputs of the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. The more trees in a forest the more robust the prediction.

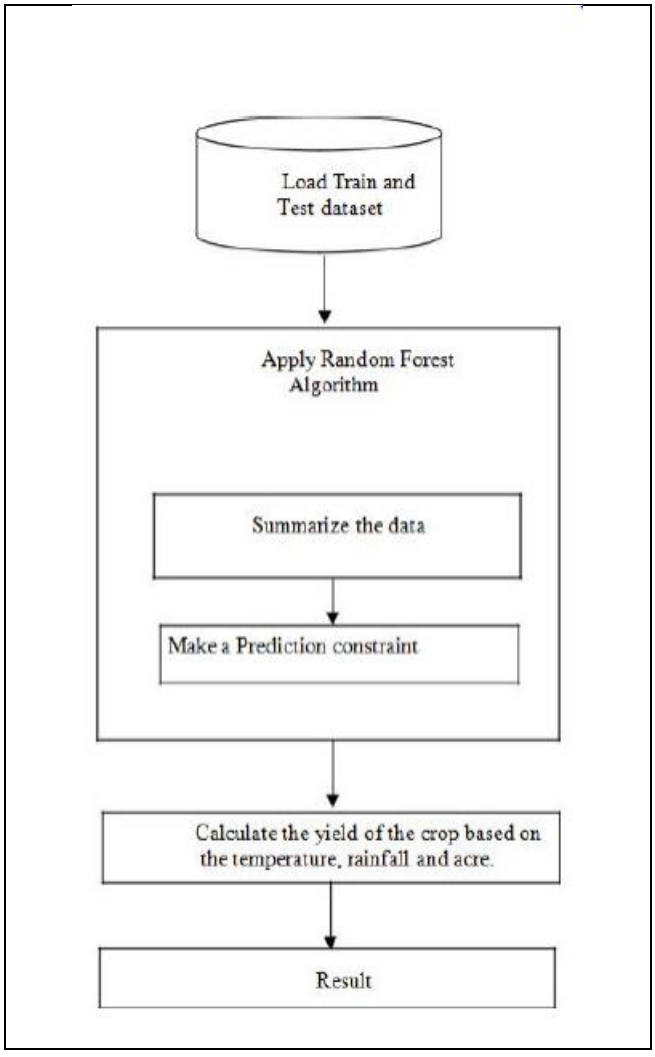

Fig. 1. Proposed Approach

Fig. 1. Shows the proposed approach and how the data is summarized, and Random Forest algorithm is applied, and the result is calculated

IV. IMPLEMENTATION

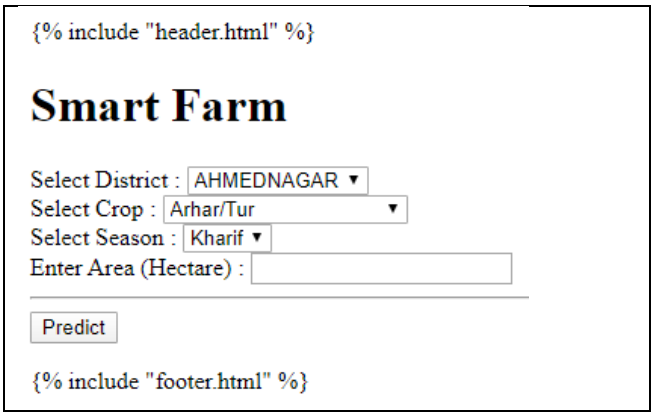

Fig. 2. Home page

Fig. 2 shows the home page of the website where the person accessing the website enters the details such as the district, crop, season and the area in Hectare and by clicking on predict the result is printed

District_Name	Season	Crop	PJan	PFeb	PMar	PApr	PMay	PJ
AHMEDNAGAR	Kharif	Arhar/Tur	3.099	0	1.671	23.129	4.646	
AHMEDNAGAR	Kharif	Bajra	3.099	0	1.671	23.129	4.646	
AHMEDNAGAR	Kharif	Gram	3.099	0	1.671	23.129	4.646	
AHMEDNAGAR	Kharif	Jowar	3.099	0	1.671	23.129	4.646	
AHMEDNAGAR	Kharif	Maize	3.099	0	1.671	23.129	4.646	
AHMEDNAGAR	Kharif	Moong(Green G	3.099	0	1.671	23.129	4.646	
AHMEDNAGAR	Kharif	Pulses total	3.099	0	1.671	23.129	4.646	
AHMEDNAGAR	Kharif	Ragi	3.099	0	1.671	23.129	4.646	
AHMEDNAGAR	Kharif	Rice	3.099	0	1.671	23.129	4.646	
AHMEDNAGAR	Kharif	Sugarcane	3.099	0	1.671	23.129	4.646	
AHMEDNAGAR	Kharif	Total foodgrain	3.099	0	1.671	23.129	4.646	
AHMEDNAGAR	Kharif	Urad	3.099	0	1.671	23.129	4.646	
AHMEDNAGAR	Rabi	Jowar	3.099	0	1.671	23.129	4.646	
AHMEDNAGAR	Rabi	Maize	3.099	0	1.671	23.129	4.646	
AHMEDNAGAR	Rabi	Other Rabi puls	3.099	0	1.671	23.129	4.646	
AHMEDNAGAR	Rabi	Wheat	3.099	0	1.671	23.129	4.646	
AHMEDNAGAR	Summer	Maize	3.099	0	1.671	23.129	4.646	

Fig. 3. Data set

Fig. 3. It is the snapshot of the final processed data set that is being used for this project

V. CONCLUSION AND FUTURE SCOPE

Based on the climatic input parameters the present study provided the demonstration of the potential use of data mining techniques in predicting the crop yield based. The developed webpage is user friendly and the accuracy of predictions are above 75 per cent in all the crops and districts selected in the study indicating higher accuracy of prediction. By providing climatic data of that place the user-friendly web page developed for predicting crop yield can be used by any user their choice of crop.

REFERENCES

[1] P.Priya, U.Muthaiah M.Balamurugan.Predicting yield of the crop using machine learning algorithm. International Journal of Engineering Science Research Technology.

[2]. J.Jeong, J.Resop, N.Mueller and team.Random forests for global and regional crop yield prediction.PLoS ONE Journal.

[3].Narayanan Balkrishnan and Dr. Govindarajan Muthukumarasamy.Crop production Ensemble Machine Learning model for prediction. International Journal of Computer Science and Software Engineering (IJCSSE).

[4]. S.Veenadhari, Dr. Bharat Misra, Dr. CD Singh.Machine learning approach for forecasting crop yield based on climatic parameters. International Conference on Computer Communication and Informatics (ICCCI).

[5]. Shweta K Shahane , Prajakta V Tawale.Prediction On Crop Cultivation. IInternational Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE) Volume 5, Issue 10, October 2016.

[6]D Ramesh ,B Vishnu Vardhan. Analysis Of Crop Yield Prediction Using Data Mining Techniques. IJRET: International Journal of Research in Engineering and Technology.

[7]Subhadra Mishra,Debahuti Mishra, Gour Hari Santra. Applications of Machine Learning Techniques in Agricultural Crop Production. Indian Journal of Science and Technology, Vol 9(38), DOI:10.17485/ijst/2016/v9i38/95032, October 2016.

[8].Konstantinos G. Liakos,Patrizia Busato,Dimitrios Moshou, Simon Pearson ID,Dionysis Bochtis. Machine Learning in Agriculture. Lincoln Institute for Agri-food Technology (LIAT), University of Lincoln, Brayford Way, Brayford Pool,Lincoln LN6 7TS, UK, spearson@lincoln.ac.uk.

[9]. Baisali Ghosh. A Study to Determine Yield for Crop Insurance using Precision Agriculture on an Aerial Platform. Symbiosis Institute of Geoinformatics Symbiosis International University 5th & 6th Floor, Atur Centre, Gokhale Cross Road, Model Colony, Pune – 411016.

[10]. Jig Han Jeong, Jonathan P. Resop, Nathaniel D. Mueller, David H. Fleisher ,Kyungdahm Yun, Ethan E. Butler,Soo-Hyung Kim. Random Forests for Global and Regional Crop Yield Predictions. Institute on the Environment, University of Minnesota, St. Paul, MN 55108, United States of America.

[11] Ecochem Online. (2009). Soil Health and Crop yields. Last modified January 28th 2009.Retrieved on March 4th 2009 from http://ecochem.com/healthy_soil.html

[12] Food and Agricultural Organization. (2006). The state of Agricultural Commodity Markets. 37-39.

[13]Aditya Shastry, H.A Sanjay And E.Bhanushree,"Prediction of crop yield using Regression Technique", International Journal of computing r12 (2):96-102 2017,ISSN:1816-914] E.14]E. Manjula , S. Djodiltachoumy,"A Model for Prediction of Crop Yield", International Journal of Computational Intelligence and Informatics, Vol. 6: No. 4, March 2017

[14]Mrs.K.R.Sri Preethaa, S.Nishanthini, D.SanthiyaK.Vani Shree ,"CropYield Prediction",International Journal On Engineering Technology and Sciences – IJETS™ISSN(P): 2349-3968, ISSN (O):2349-3976 Volume III,Issue III, March- 2016

[15]Jharna Majumdar, Sneha Naraseeyappa and Shilpa Ankalaki, "Analysis of agriculture data using data mining techniques: application of big data"Majumdar et al. J Big Data (2017) 4:20 DOI 10.1186/s40537-017-0077-4

[16]D. Ramesh and B. Vardhan, "Analysis of crop yield prediction using data mining techniques", International Journal of Research in Engineering and Technology, vol. 4, no. 1, pp. 47-473, 2015.

[17]Yethiraj N G , " Applying data mining techniques in the field of Agriculture and allied sciences", Vol 01, Issue 02, December 2012.

[18] Zelu Zia (2009). An Expert System Based on Spatial Data Mining used Decision Tree for Agriculture Land Grading. Second International Conference on Intelligent Computation Technology and Automation.Oct10-11,China