IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, May 2022
Impact Factor: 6.252

Crop Yield Prediction Using Machine Learning

Omkar Solunkhe[1], Mayur Bhamare[2], Prashant Bothara[3], Harsh Borwal[4], Prof. Priyanka Kinage[5]
Students, Department of Computer Engineering[1,2,3,4]
Guide, Department of Computer Engineering[5]
Smt. Kashibai Navale College of Engineering Ambegaonbk, Pune, Maharashtra, India

Abstract: *Agriculture is that the pillar of the Indian economy and over 50% of India's population are keen about agriculture for his or her survival. Variations in weather, climate, and other such environmental conditions became a serious risk for the healthy existence of agriculture. Machine learning (ML) plays a big role because it has decision support tool for Crop Yield Prediction (CYP) including supporting decisions on what crops to grow and what to try to to during the season of the crops. the current research deals with a scientific review that extracts and synthesize the features used for CYP and furthermore, there are a spread of methods that were developed to research crop yield prediction using computing techniques. Many studies were recommended for agriculture development and also the goal was to form an accurate and efficient model for crop classification like crop yield estimation supported the weather, crop disease, classification of crops supported the growing phase etc., This paper explores various ML techniques utilized within the field of crop yield estimation and provided an in depth analysis in terms of accuracy using the techniques.*

Keywords: Agriculture, Artificial Neural Network, Convolution Neural Network, Crop yield prediction, Machine learning method

I. INTRODUCTION

The worldwide population was estimated at 1.8 billion in 2009 and is predicted to extend to 4.9 billion by 2030, resulting in an extreme increase in demand for agricultural products. within the future, agricultural products will have higher demand among the human population, which can require efficient development of farmlands and growth within the yield of crops. Meanwhile, because of warming, the crops were frequently spoiled by harmful climatic situations [2]. one failure because of climatic variation, floods, lack of soil fertility, lack of groundwater and other such factors destroy the crops which successively affects the farmers. In other nations, the society advises farmers to extend the assembly of specific crops in step with the locality of the world and environmental factors [3]. The population has been increasing at a significantly higher rate, that the estimation and monitoring of crop production is critical [4]. Accordingly, an appropriate method must be designed by considering the affecting features for the higher selection of crops with relevancy seasonal variation [5]. The core objective of crop yield estimation is to realize higher agricultural crop production and plenty of established models are exploited to extend the yield of crop production. Nowadays, ML is getting used worldwide because of its efficiency in various sectors like forecasting, fault detection, pattern recognition, etc. The ML algorithms also help to enhance the crop yield production rate when there's a loss in unfavorable conditions. The ML algorithms are applied for the crop selection method to scale back the losses crop yield production regardless of distracting environment.

II. LITERATURE REVIEW

Ananthara, M. G. et al. (2013, February) proposed a prediction model for datasets regarding agriculture which is termed as CRY algorithm for crop yield using beehive clustering techniques. They considered parameters namely crop type, soil type, soil pH value, humidity and crop sensitivity. Their analysis was mainly in paddy, rice and sugarcane yields in India. Their proposed algorithm was then compared with C&R tree algorithm and it outperformed well with an accuracy of 90 percent. t [2]. Awan, A. M. et al. (2006, April) built a brand new, smart framework focused on farm yield prediction clustering kernel methodology and that they considered parameters like plantation, latitude, temperature and precipitation of rainfall in this latitude. that they had experimented weighted k-means kernel method with spatial constraints for the analysis of feather palm fields [3]. Chawla, I. et al. (2019, August) used symbolic logic for crop yield prediction through statistical statistic models. They considered parameters like rainfall and temperature for prediction. Their prediction was classification with

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, May 2022
Impact Factor: 6.252

levels ‘good yield ,‘very good yield’[4]. Chaudhari, A. N. et al. (2018, August) used three algorithms namely clustering kmeans, Apriori and Bayes algorithm, then they hybridized the algorithm for better efficiency of yield prediction and that they considered parameters like Area, Rainfall, Soil type and also their system.

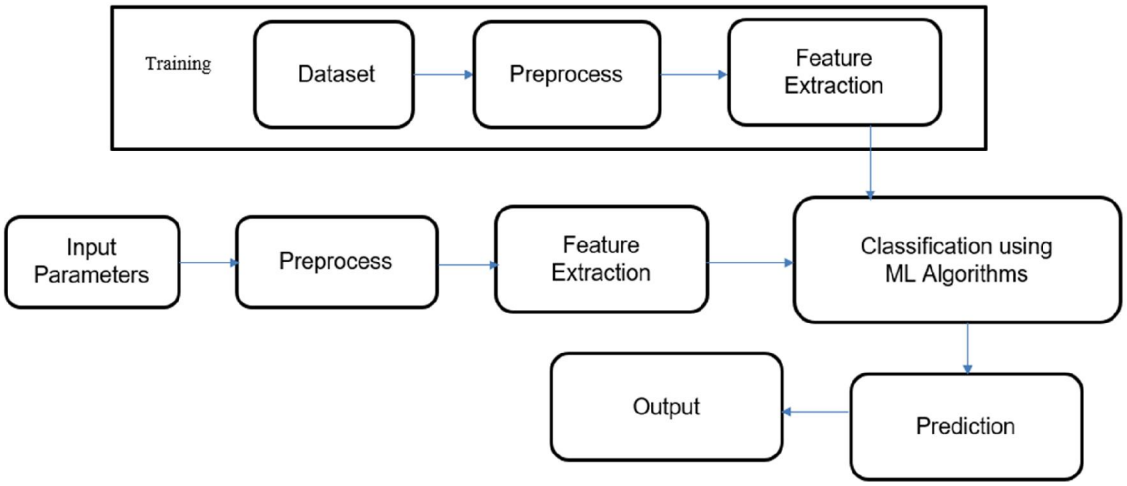

Figure: System Architecture

III. System Requirement

System Requirements Specification may be a document or set of documentation that describes the features and behavior of a system or software application. It includes a range of elements that attempts to define the intended functionality required by the users to satisfy their different users.

3.1 Functional Requirements

The Functional Requirements Definition reports and tracks the fundamental information expected to effectively portray business and handy necessities. The Functional Requirements Definition report is created within the midst of the design Phase of the endeavor. Its objective gathering is that the endeavor boss, errand gathering, wander bolster, client/customer, and any accomplice whose information/respect into the necessities definitions system is required.

3.2 Non-Functional Requirements

Non-Functional Requirement (NFR) specifies the quality attribute of a software system. They judge the software system based on Responsiveness, Usability, Security, Portability and other non-functional standards that are critical to the success of the software system. Failing to meet non-functional requirements can result in systems that fail to satisfy user needs. Nonfunctional Requirements allows you to impose constraints or restrictions on the design of the system across the various agile backlogs. Example, the site should load in 3 seconds when the number of simultaneous users are > 10000. They specify the criteria that can be used to judge the operation of a system rather than specific behaviours. They may relate to emergent system properties such as reliability, response time and store occupancy.

3.3 Hardware Requirements

Hardware specifications are technical descriptions of the computer's components and capabilities. Processor speed, model and manufacturer. Processor speed is usually indicated in gigahertz (GHz). the upper the quantity, the faster the pc. Random Access Memory (RAM). this is often typically indicated in gigabytes (GB). The more RAM in an exceedingly computer the more it can do simultaneously. fixed disk (sometimes called ROM) space. this is often typically indicated in gigabytes (GB).

- Processor Intel Pentium/Core – 1.7GHz and above
- Memory 1GB and above
- Storage 80GB minimum fee space
- Graphics 1GB and above 5.4

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, May 2022
Impact Factor: 6.252

3.4 Software Requirements

- Operating System - Windows 7,8,10
- Programming Language - Python
- Framework - Pycharm

IV. METHODOLOGY

Data may be a very important a part of any Machine Learning System. To implement the system, we decided to specialize in Maharashtra State in India. because the climate changes from place to put, it had been necessary to induce data at district level. Historical data about the crop and also the climate of a specific region was needed to implement the system. This data was gathered from different government websites. the information about the crops of each district of Maharashtra was gathered from www.data.gov.in and therefore the data about the climate was gathered from www.imd.gov.in. The climatic parameters which affect the crop the foremost are precipitation, temperature, overcast, pressure, wet day frequency. So, the data about these climatic parameters was gathered at a monthly level.

- **Dataset Collection:** during this phase, we collect data from various sources and prepare datasets. and also the provided dataset is within the use of analytics (descriptive and diagnostic). There are several online such as Data.gov.in and indiastat.org. For a minimum of 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 whole dataset is partitioned into 2 parts: as an example, say, 75% of the information set is employed for training the model and 25% of the data is ready aside to check the model. To predict future events Machine Learning Algorithms: Supervised learning: Supervised machine learning algorithms can apply what has been learned within the past to new data using labelled examples. After Sufficient training the system can provide targets for any new input. so as to vary the model accordingly the training algorithm may also differentiate its results with the right, intended output and find errors. Unsupervised learning: compared, unsupervised machine learning algorithms are used when the knowledge accustomed train is neither labelled nor classified. Unsupervised learning does analysis of how systems can infer a function to explain a hidden structure from unlabelled data. so as to explain hidden structures from unlabelled data the system doesn't work out the correct output, but it examines the data and might draw inferences from datasets.
- **Random Forest Classifier:** Random forest is that the most popular and powerful supervised machine learning algorithm capable of performing both classification and regression tasks, that operate by constructing a mess 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 very forest the more robust the prediction.

V. PROPOSED ALGORITHMS

5.1 SVM (Support Vector Machine)

Support vector machine is utilized in crop yield prediction is called support vector regression. The aim of the support vector technique is to get non-linear function using kernel function (a linear function or polynomial function). The radial basis function and therefore the polynomial function are the widely used kernel function. The merit of support vector regression is to avoid difficulties of using linear function in large input samples space and optimization of a posh problems transformed into simple linear function optimization.

5.2 RF (Random Forest)

Random Forest could be a ML algorithm. At training situation multitude decision trees are built and therefore the output are divided supported on number of classes i.e., classification, prediction of sophistication i.e., regression. The dataset comprises factors like rainfall, perception, temperature and production. These factors in dataset are used for training. The algorithm random forest consists of three parameters like: n-tree which describes the n number of trees which require to grow, m try - mentions what percentage variables must need to be taken at a node split. Node size - In terminal nodes it advise us the amount of observation have to take.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, May 2022
Impact Factor: 6.252

VI. Conclusion And Future Scope

Based on the climatic input parameters this 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 therefore the accuracy of predictions are above 75 per cent all told the crops and districts selected within the study indicating higher accuracy of prediction. By providing climatic data of that place the user-friendly online page developed for predicting crop yield is employed by any user their choice of crop.

References

[1]. P.P riya, 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]. ShwetaK 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,PatriziaBusato,Dimitrios Moshou, Simon Pearson ID,DionysisBochtis. Machine Learning in Agriculture. Lincoln Institute for Agri-food Technology (LIAT), University of Lincoln, Brayford Way, BrayfordPool,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.