JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

MACHINE LEARNING BASED CROP PREDICTION SYSTEM FOR FARMERS

Atharva Kadam[1],S.S. Bhosale[2] , Aniket Alaspurkar[3], Sumit Kamble[4], S.N.Kamble[5] ,C.G.Patil[6]
Department of Electronics and TeleCommunication Engineering,,
Smt. Kashibai Navale COE,Savitribai Phule Pune University,Pune
1 atharvakadam2310@gmail.com, 2 sonali.bhosale_skncoe@sinhgad.edu,
3 aniket2297@gmail.com, 4 s07kamble@gmail.com,
5 sarika.kamble_skncoe@sinhgad.edu, 6 chandrashekhar.patil_skncoe@sinhgad.edu

Abstract: ***India is an agricultural country around 41% of Indian population works in farm and agro industries, contributing 20% of India's total GDP. Due to unfavorable climate agricultural industry has been impacted largely and also cultivating same crop again and again in same field can lead to de-fertilization of land due to which farmers are unable to get better yield. To obtain maximum yield of best suitable crop, factors such as humidity, rainfall, temperature and soil along with its pH value can be considered to predict the best suitable crop. The main aim is to design a system using machine learning Algorithm implementing Random Forest to help the farmers in predicting the best suitable crop for their land. Random forest is ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time, for implementation of this System the database is available on Kaggle which will be used to implement this Algorithm, with accuracy of 92.84% in predicting the best crop for farmers. 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, rop-prediction, Random Forest Algorithms, Crop yield, Kaggle

I INTRODUCTION

Agriculture is the backbone of the Indian economy. In India, agricultural productivity depends largely on the climate. Cultivation of rice is highly dependent on rainfall. Timely advice on future crop production and analysis should be made to assist farmers to increase crop production. Yield forecasting is an important agricultural problem. In the past, farmers were even predicting the outcome of last year's crop. Therefore, in this type of data analysis in crop prediction, there are different strategies or algorithms, and with the help of those strategies we can predict crop yields. Using a random forest algorithm. Using all of these algorithms and with the help of the links between them, there is a growing list of applications and the role of major agricultural data analysis strategies. Since the introduction of new technologies and techniques the agricultural sector has been slowly declining. As a result, many established people are focused on growing synthetic products that are hybrid products that lead to unhealthy living. Today, most people do not have the knowledge to plant crops at the right time and in the right place. As a result of these methods of planting the climatic conditions of the year also change against important assets such as soil, water and air leading to food insecurity. By analyzing all these problems and problems such as climate, temperature and several factors, there is no suitable solution and technology to overcome the situation we are facing. In India, there are several ways to increase economic growth in the agricultural sector. There are many ways to increase and improve crop yields and crop quality. Data mining also helps predict crop yields. The objectives are as follows =

a) To use machine learning techniques to predict cropyield.
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 crop yields using a machine learning algorithm. International Journal of Scientific Research Technology. This paper focuses on predicting crop yields based on available data using the Random Forest algorithm. The actual Tamil Nadu data was used to create the models and the models were tested with samples. The Random Forest Algorithm can be used to accurately predict crop yields.

In [2] Predictable random harvest of global and regional harvest. PLoS ONE Journal. Our produced results show that RF is an effective and flexible machine learning method for predicting crop yields at regional and global scale with its high accuracy and precision, ease of use, and usefulness in data analysis. A random forest is a very efficient strategy and transcends multi-line retreat (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.

In [4]. Mechanical learning method for predicting crop yields based on weather parameters. The paper is presented at the International Conference on Computer Communication and Informatics (ICCCI). In the present study a software tool called Crop Advisor was developed as an easy-to-use webpage to predict the impact of climate on crop yields. plants in selected districts of Madhya Pradesh. Paper is used using Decision Tree.

In [5]. Predicting Crop Crops. International Journal of Advanced Research in Computer Science and Electronics Engineering (JARCSEE) Volume 5, Issue 10, October 2016. Currently, soil analysis and interpretation of soil test results is based on paper. This has in some way contributed to the negative interpretation of soil test results which has resulted in poor crop yields, soil reform and fertilizer for farmers thus leading to poor crop yields, nutrient deficiencies in the soil and excessive or minimal fertilizer use. Landscape Matching Formulas, Fertilizer Recommendations.

In [6]. Yield Prediction Analysis by Using Data Mining Methods. IJRET: Paper published in the International Journal of Research in Engineering and Technology. In this paper the main objective is to create a user-friendly interface for farmers, which provides analysis of rice production based on available data. In order to increase crop production various methods of data mining were used to predict crop yields. Like the K-Means algorithm for predicting atmospheric pollution.

In [7]. Use of Mechanical Learning Methods in Agricultural Crop Production. Indian Journal of Science and Technology, From GPS-based color photography is provided as a reinforced collection analysis of plant, soil and fossils of interested regions. This paper covers various parameters that can help crop yields to be better developed and the yield rate can be increased during planting.

In [8]. In this paper, we present a comprehensive review of the research dedicated to the use of machine learning in agricultural production systems. Computer literacy (ML) has come up with big data technologies, techniques, methods and a highly efficient computer to generate new opportunities for analyzing, evaluating, and analyzing processes that affect data in the agricultural sector. By using Support Vector Machines (SVP) paper is used.

In [9]. Research on Determining Crop Crop Crops using Precision Agriculture in the Aerial Platform. Symbiosis Institute of Geoinformatics Symbiosis International University 5th & 6th Floor, Artur Center, Gokhale Cross Road, Model Colony, Pune - 411016. . using a variety of techniques. The causes of crop diversity in the agricultural field may be due to crop stress, irrigation systems, pest and disease events etc. Paper Made using Ensemble Learning (EL).

In [10]. Global Informal Forests and Regional Plants Yield Predictions. Environmental Center, University of Minnesota, St. Paul, MN 55108, United States of America. The results obtained show that RF is an effective and unique machine learning method for predicting crop yields on a regional and global scale with its high accuracy. The paper is used using a neighbor close to k, Support Vector Regression (SVG).

In [11]. A method combining the support vector machine (SVM) the K-Nearest Neighbors (KNN), labelled the SVM-KNN method, is used to construct a solar flare forecasting model. Based on a proven relationship between SVM and KNN, the SVM-KNN method improves the SVM algorithm of classification by taking advantage of the KNN algorithm according to the distribution of test samples in a feature space.

In [12]. To help the farmers the proposed model is designed. The model is portable and used by farmers directly in the fields to measure the various basic parameters (temperature, humidity, rain, soil moisture content). The collected data from the sensors is transferred to the Thing Speak cloud. Hybrid machine learning algorithm is applied on that data. After that decision are made that which crop is suitable for that environment. Farmers can also see the results of the parameters with the help of Thing View applicatio

III THE PROPOSED SYSTEM

Data is the most important part of any machine learning program. In order to implement the plan, we have decided to focus on Maharashtra Province in India. As the climate changed from place to place, data had to be obtained at regional level. Historical data on the crop and climate of a particular region were required to operate the system. This data is collected from various government websites. Information about the plants of each Maharashtra region was collected at www.data.gov.in and weather data was collected at

www.imd.gov.in. The weather conditions that affect the crop the most are rainfall, temperature, cloud cover, humidity, and frequent wet days. Therefore, data on these weather conditions were collected at the monthly level. The proposed system is shown in Figure 1.

A) Data Collection: In this section, we collect data from a variety of sources and prepare data sets. Also the provided data is used for analysis (descriptive and diagnostic). There are several sources of online summaries such as Data.gov.in and Indiastat.org. For at least ten years the annual abbreviations of the plant will be used. These databases generally adopt the behavior of the anarchic time series. Basic and required summaries are included. Random Forests for International and Regional Harvest Prediction.

B) 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 testthe model.

Predicting future events Machine Learning Algorithms:

A) Supervised Learning: Supervised machine learning algorithms can apply past lessons to new data using labeled examples. After Enough training the system can provide the objectives of any new inputs. IN ORDER to modify the learning appropriately the learning algorithm can also differentiate its results by correct, targeted and error detection. Uncontrolled learning: By comparison, the unsupervised machine learning algorithms are used when the information used for the training is not clear and undivided. Unchecked reading analyzes how systems can perform the task of defining a hidden structure from non-labeled data. To define hidden properties from non-labeled data the system does not detect the correct output, but scans the data and may draw predictions from the data sets.

.

Fig.1 Proposed System

B) Random Forest Classifier: Random Forest is a popular and powerful machine learning algorithm capable of performing both subdivision and decontamination activities, which works by building a number of deciduous trees during training and producing class results which is class mode (separation) or mean prediction (descent) of individual trees. The more trees in the forest the more predictable the weather will be. The default way to calculate the value of diversity is to reduce the rate of contamination: In each subdivision of each tree, the development of the distinction criterion is calculated from the subdivision, and accumulates over all the trees in the forest separately. each variable. Note that this rate is the same as the reversal in the training set.

IV IMPLEMENTATION

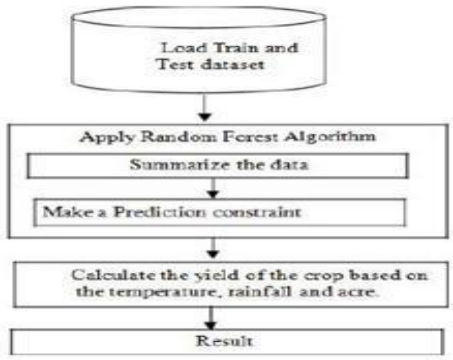

Figure 2: Proposed Approach

Fig. 2. Shows the proposed approach and how the data is summarized, and Random Forest algorithm is applied, and the result is calculated.

Table 1: Proposed Dataset

Sr. No	Nitrogen	Phosphorus	Potassium	Temperature	Humidity	pH	Rainfall	Label
1.	90	42	43	20.87	82.0	6.50	202.93	rice
2.	61	44	17	26.10	71.5	6.93	102.26	Maize
3.	40	72	77	17.02	16.9	7.48	88.52	Chickpea
4.	13	60	25	17.13	20.5	5.68	128.25	Beans
5.	40	59	23	36.89	62.73	5.26	163.72	peas
6.	8	54	20	28.33	80.77	7.03	38.79	mung
7.	42	61	22	26.27	62.28	7.41	70.23	gram
8.	32	76	15	28.05	63.49	7.60	43.35	lentil
9.	2	24	38	24.55	91.63	5.92	111.96	pomegranate
10.	91	94	46	29.36	76.24	6.14	92.82	banana
11.	2	40	27	29.73	47.54	5.95	90.09	Mango
12.	24	131	196	22.03	83.74	5.73	65.34	grapes
13.	119	25	51	26.47	80.92	6.28	53.65	watermelon
14.	117	24	53	29.17	92.21	6.29	21.30	muskmelon
15.	24	128	196	22.75	90.69	5.52	110.43	Apple
16.	22	30	12	15.78	92.51	6.35	119.03	orange
17.	61	68	50	35.21	91.49	6.79	243.07	papaya
18.	18	30	29	26.76	92.86	6.42	224.59	coconut
19.	133	47	24	24.40	79.19	7.23	90.80	Cotton
20.	89	47	38	25.52	72.24	6.00	151.88	jute

Table 1. shows the proposed Dataset with Multiple Attributes and with Multiple Crops with Different Parameters. This is final processed data set that is being used for this project. Combined dataset has 2200 instances. It includes features like crop name, Nitrogen, Phosphorus, Potassium, pH, Temperature, rainfall and humidity.

Random Forest has the ability to analyze crop growth related to the current climatic conditions and biophysical change. Random forest algorithm creates decision trees on different data samples and then predicts the data from each subset and then by voting gives better solution for the system. Random Forest uses the bagging method to train the data which increases the accuracy of the result. For our data, RF provides an accuracy of 92.81%. It is clear that among all the three algorithms, Random Forest Gives the better accuracy as compared to other algorithms

The result of the different algorithms are compared here as shown in Table 2 below.

Table 2 Comparison of the Performance of Algorithms

Algorithm	Accuracy (%)
Random Forest	92.814
Naïve Bayes	91.496
Logistic Regression	87.821

V CONCLUSION

Due to large variety of crops and varying temperature conditions present in data set, we would recommend and conclude that Random Forest Classifier is better at handling such data and can retain maximum accuracy. Random forest can be implemented in a Machine Learning system which makes prediction based on current inputs as well as data set used for train and test. Also, this project guides the farmers to predict suitable crops for farming based on different factors such as previous crop, soil content and weather conditions. It also provides weather information of the user's location. So, the overall design is simple and user friendly. Considering the present situation, project is deployed as a web application so that users from different devices such as android, iPhone, desktop can access this application using any browser.

References

[1] Dhruvi Gosai1, Chintal Raval, Rikin Nayak, Hardik Jayswal, Axat Patel (2021): "Crop Recommendation System using Machine Learning." - IJSRCSEIT, 2021.

[2] Sonal Jain, Dharavath Ramesh: "Machine Learning Convergence for Weather Based Crop Selection." - IEEE, 2020.

[3] Dr. Y. Jeevan Nagendra Kumar, V. Spandana, V.S. Vaishnavi, K. Neha: "Supervised Machine learning Approach for Crop Yield Prediction in Agriculture Sector." - IEEE, 2020.

[4] Bhawana Sharma, Jay Kant Pratap Singh Yadav, Sunita Yadav: "Predict Crop Production in India Using Machine Learning Technique: A Survey." - ICRITO, 2020.

[5] M. Kalimuthu, P. Vaishnavi, M. Kishore (2020): "Crop Prediction using Machine" Learning." - IEEE, 2020.

[6] Potnuru Sai Nishant, Pinapa Sai Venkat, Bollu Lakshmi Avinash, B. Jabber: "Crop Yield Prediction Based on Indian Agriculture Using Machine Learning." - INCET, 2020

[7] A. Nigam, S. Garg, A. Agrawal, P. Agrawal: "Crop Yield prediction using Machine Learning algorithms." - IEEE, 2019.

[8] Ramesh Medar, Vijay S. Rajpurohit, Shweta: "Crop Yield Prediction using Machine Learning Techniques."- ICCT, 2019.

[9] Rishika Anand, Dimple Sethi, Kavita Sharma, Pooja Gambhir.: "Soil Moisture and Atmosphere Components Detection System Using Iot And Machine Learning."- Ieee, 2019.

[10] Aruvansh Nigam, Saksham Garg, Archit Agrawal, Parul Agrawal: "Crop Yield Prediction Using Machine Learning Algorithms." - ICIIP, 2019

[11] Anakha Venugopal, Aparna S..: "Crop Yield Prediction using Machine Learning Algorithms", IEEE, 2019

[12] Sk Al Zaminur Rahman, Kaushik Chandra Mitra, S.M. Mohidul Islam: "Soil Classification Using Machine Learning Methods and Crop Suggestion Based on Soil Series." - IEEE, 2018