IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Special Issue 2, March 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

| International Conference on Recent Innovations in Engineering and Technology (ICRIET'23)|

| Sharadchandra Pawar College of Engineering, Pune, India |

|| Volume 11, Special Issue 2, March 2023 ||

Crop Yield Prediction and Recommending the Best Crop using Machine Learning

Abhishek Ghanekar, Shubham Raut, Pooja Kesarkar, Prof: Neeta.N.Bankhele

UG Student, Dept of E&TC, Sharadchandra Pawar College of Engineering Otur, Pune, India

UG Student, Dept of E&TC, Sharadchandra Pawar College of Engineering Otur, Pune, India

UG Student, Dept of E&TC, Sharadchandra Pawar College of Engineering Otur, Pune, India

UG Student, Dept of E&TC, Sharadchandra Pawar College of Engineering Otur, Pune, India

Assistant Professor, Dept of E&TC, Sharadchandra Pawar College of Engineering Otur, Pune, India

ABSTRACT: As a result of the effects of climate change in India, most crops have been hit hard in terms of productivity over the past 20 years. Pre-harvest yield forecasts will help policy makers and farmers take appropriate marketing and storage actions. This project will help farmers make appropriate decisions by knowing the yield of their crops before planting them on their fields. He wants to solve the problem by prototyping an interactive prediction system. These systems are implemented with easy-to-use graphical web user interfaces and machine learning algorithms. Forecast results are available to farmers. Therefore, there are various methods or algorithms for data analysis in crop forecasting, and crop yields can be predicted through these algorithms. A random forest algorithm is used. Weather, temperature, humidity, precipitation, humidity, etc. In analysing all these problems and problems, there is no suitable solution and technology that can overcome the situation we face. In India, there are many ways to boost economic growth in the agricultural sector. 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. In second part of our work will be building a recommendation system for crop in which we will be analysing the data which contains temperature ,humidity, pressure and soil parameters like N,P,K .We will predict which crop will be best for certain parameters. Again, a machine learning model is used, but since this is a classification problem, we use a classification model. Among them, we will focus mainly on SVMs, again on random forest classifiers, and then on KNNs. We will do a comparative study of all these models and eventually apply the best model to our application and deploy the application.

KEYWORDS: Harvest, prototyping, algorithms, weather, temperature, humidity, mean prediction, SVM,Linear Regression.

I. INTRODUCTION

Agriculture is the backbone of India's economy. In India, crop yields are largely dependent on weather conditions. Rice cultivation is mainly dependent on rainfall. It is necessary to make timely recommendations for forecasting and analysing future yields so that farmers can maximize crop yields. Yield forecasting is an important agricultural problem. In the past, farmers predicted yields based on previous year's yield experience. Therefore, there are various methods or algorithms for this kind of data analysis in crop forecasting, and with the help of these algorithms, crop yields can be predicted.The use of all these algorithms and the relationship between them is expanding the application scope and role of big data analysis methods in agriculture. Agriculture is slowly deteriorating with the advent of new innovations and technologies. 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. In India, there are several ways to increase the economic growth in the field of agriculture. There are many

ways to increase and improve crop yield and quality. Data mining is also useful for predicting crop yields. Primary Objectives

a. Predict yield using machine learning methods.
b. It provides an easy-to-use user interface.
c. Increases the accuracy of yield predictions, and Analyse various climate parameters (cloudiness, precipitation, temperature).

II. METHODOLOGY

Data is a very important part of any machine learning system. To implement the system, we decided to focus on the Indian state of Maharashtra. Because different regions have different climates, it was necessary to obtain data at the regional level. To implement the system, historical data on crops and climate in a specific area were required. This data was collected from various government websites.Crop data for each region of Maharashtra was collected from www.data.gov.in and climate data from www.imd, and using Kaggle and api. The climatic parameters that have the greatest impact on crops are precipitation, temperature, cloudiness, water vapor pressure, and frequency of rainy days. Therefore, data on these climate parameters were collected on a monthly basis. Ingest data set. In this step, data is collected from various sources and the data set is prepared.And the data set provided is in use for analysis (descriptive and diagnostic). There are several online sources for abstracts such as data.gov.in and indiastat.org.Annual crop abstracts for at least 10 years are used. These data sets typically behave as anarchic time series. Combined basic and necessary abstracts. Random forests for global and regional yield forecasting.Split Data: The entire data set is split into two parts. For example, 75% of the dataset is used to train the model and 25% of the data is used 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.To describe hidden structures in unlabelled data, the system does not determine the correct output, but can inspect the data and draw inferences from the data set.Random Forest Classifier: The Random Forest is the most popular and powerful supervised machine learning algorithm capable of both classification and regression tasks that works by building many decision trees during training and generating class outputs that are modes. Classes of individual trees (classification) or average predictions (regression). The more trees there are in the forest, the more reliable the predictions are.

III. PROBLEM STATEMENT

Need to build a high-accuracy, low-loss machine learning model to predict yield based on properties such as condition, area, soil type, N, P, K, temperature, humidity, and pressure. Compare multiple machine learning models: We need to create a recommendation system that can recommend the best crop for a specific field. Then you can get maximum yield.

IV. LITERATURE SURVEY

[1] Yield prediction using machine learning algorithms. International Journal of Engineering Research and Technology. This article focuses on using the Random Forest algorithm to predict crop yield based on existing data. We built the model using real data from the state of Tamil Nadu and tested the model on a sample. The random forest algorithm can be used to accurately predict yield.

[2] Random forests for global and regional yield prediction. PLoS ONE magazine. The results we generate demonstrate that RF is an efficient and adaptable machine learning method for regional and global yield prediction due to its high accuracy, ease of use and usefulness in data analysis. Random Forest is the most effective strategy to beat Multiple Linear Regression (MLR).

[3].A machine learning model for crop ensemble prediction. International Journal of Computer Science and Software Engineering (IJCSSE). In this article, AdaNaive and AdaSVM are proposed ensemble models used to predict crop production over a given period. Implementation is done using AdaSVM and AdaNaive. AdaBoost improves the efficiency of SVM and Naive Bayes algorithms.

[4]. A machine learning approach for predicting crop yield based on climate parameters. This paper was presented at the International Conference on Computer Communications and Informatics (ICCCI). In the current study, we developed a software tool called Crop Advisor, a user-friendly web page for predicting the effects of climate parameters on crop yields.Algorithm 5 is used to obtain the climatic parameters that have the greatest impact on the yield of individual crops in selected regions of Madhya Pradesh. Tasks are implemented using decision trees.

[5]. Crop forecasting. International Journal for Advanced Studies in Computer Science and Electronics (IJARCSEE), Volume 5, Issue 10, October 2016Currently, soil analysis and interpretation of soil test results are done on paper. This has somehow contributed to misinterpretation of soil test results, leading to erroneous recommendations for crops, soil fertilizers and fertilizers for farmers, resulting in low yields, micronutrient deficiencies in the soil and over- or under-fertilization. . Formulas to match crops to soil, fertilizer recommendations.

[6]. Yield prediction analysis using data mining methods.Article contributed to IJRET: International Journal of Research in Engineering and Technology. The main goal of this article is to create a user-friendly interface for farmers that provides rice production analysis based on available data. Various data mining techniques have been used to predict crop yield to maximize crop yield. For example, there is a K-Means algorithm for predicting air pollution factors.

[7].Application of machine learning methods to crop production. Indian Journal of Science and Technology, Volume 9 (38), DOI: 10.17485/ijst/2016/v9i38/95032, October 2016. GPS-based color images provided by focused fuzzy cluster analysis to classify plants, soils and vegetation residues. It's possible. interest. This article contains a number of parameters that can help improve yield and yield coefficients can increase during cultivation.

[8] This article presents a comprehensive review of research applying machine learning in agricultural production systems. Machine learning (ML) has emerged alongside big data techniques, methods, techniques and high-performance computing to create new opportunities to discover, quantify and analyze data-intensive processes in agricultural operations. This paper is implemented using a Support Vector Machine (SVP).

[9]. A Study on Yield Determination of Crop Insurance Using Precision Farming in Aerial Work Platform.
Symbiosis Geoinformatics Institute Symbiosis International University 5th and 6th floor, Artur Centre, Gokhale Cross Road, Model Colony, Pune - 411016. Precision Agriculture (PA) applies geospatial methodologies and remote sensors to detect and respond to changes in the field. no see. They use different strategies. In agriculture, the reasons for crop growth variability can be related to crop stress, irrigation practices, pest and disease spread, etc. The thesis is implemented using Ensemble Learning (EL).

[10].Random forests for global and regional yield forecasting. University of Minnesota Environmental Institute, St. Paul, MN 55108, United States of America. The results produced demonstrate that RF is an efficient and superior machine learning method for regional and global yield prediction due to its high accuracy. This article is implemented using k-Nearest Neighbour, Support Vector Regression (SVG).

V. RESULT

The main objective of this research work is to predict the yield ofcrops under specific climatic conditions. First, we collectdata sets and apply fouralgorithms to these data sets to test the algorithm's accuracy. We use

Linear Regression, Decision Tree Regression,Random Forest Algorithm and XG Boost Algorithm. We collected datafrom the government website www.data.gov.in, e.g. maximum temperature, minimum temperature, seasonal precipitation, arable land area and rice production. We sorted 2004–2014 data as well asby district for four selected states (MAHARASHTRA)

The more accurate this algorithm is. Therefore, in the above result, XG Boost algorithm has the best value, so will be the best algorithm for yield prediction.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

| International Conference on Recent Innovations in Engineering and Technology (ICRIET'23)|

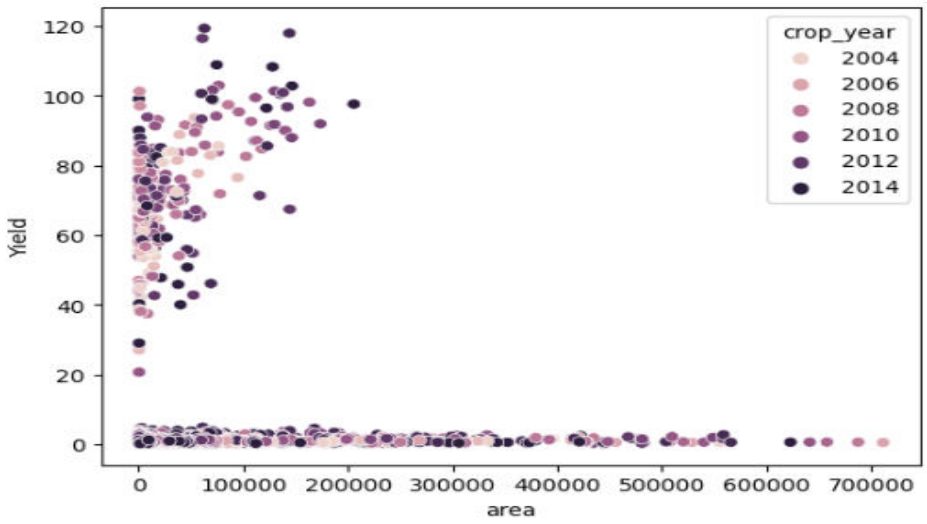
| Sharadchandra Pawar College of Engineering, Pune, India |

|| Volume 11, Special Issue 2, March 2023 ||

Model	R^2	MSE	MAE
Linear Regression	0.8885	2972466269.2335	27883.6268
SVR	0.8991	NA	27402.1696
Decision Tree Algorithm	0.8922	2847222194.9642	24187.3076
Random Forest Algorithm	0.9314	2178239719.7533	21699.0175
XG Boost Algorithm	0.9391	1999378847.4874	20613.2361

Fig1. Compare of Models

GRAPHICAL REPRESENT OF DATA:



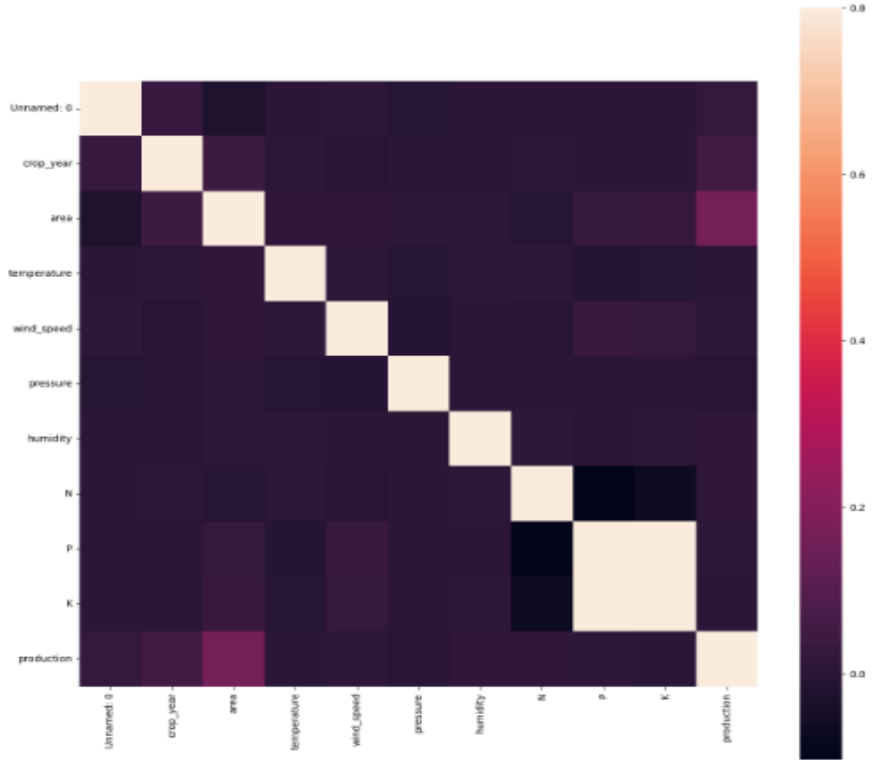


FIG2. CORELATION HEATMAP

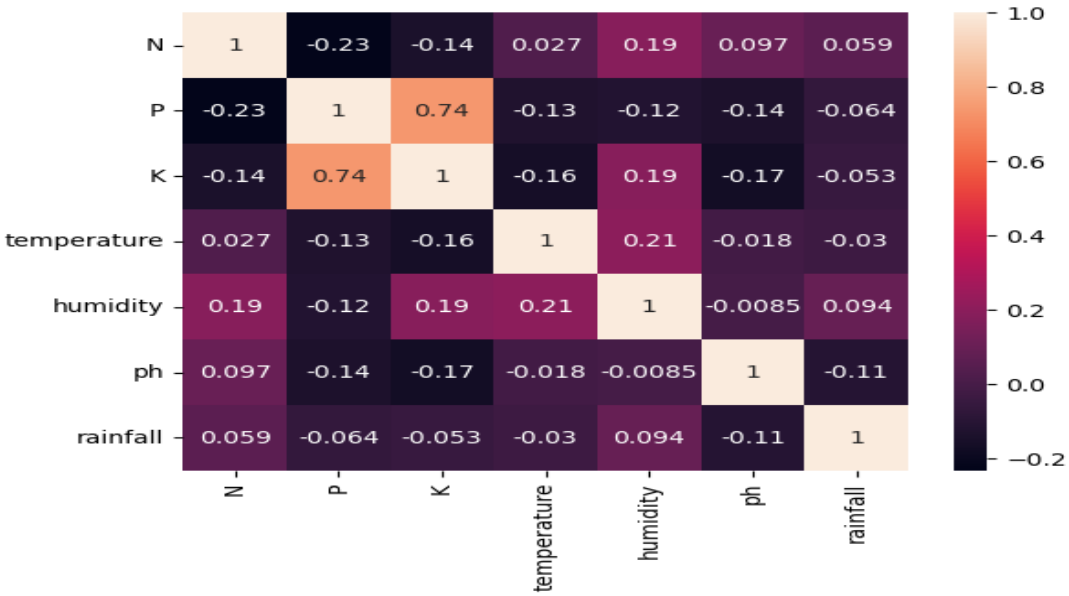


FIG3. HEATMAP OF RECOMMENDATION

VI. CONCLUSION

Farmers in India face many problems and end up losing money. It is a societal problem that technology can be used to solve or reduce losses. We're trying to build machine learning models that advise farmers to grow the best crops in specific soils and tell them what to harvest from their fields.From the above results, we can easily see that the accuracy level ofXG Boost algorithm is higher in all cases be itR2, MSE or MAE. The model was built using datacollected from the data.gov.in form from 2000 to 2014.The proposed XG Boostalgorithm is compared with linear regression, SVR,

DecisionTree, and Random Forest. The XG Boost model delivers better performance on thethan any other model. In the future, it is planned to optimize the hyperparameters of each existing algorithm and test it with a new data set.

REFERENCES

[1] P. Priya, W. Mutaya, M. Balamurugan. Yield prediction using machine learning algorithms. International Journal of Engineering Research and Technology.
[2]. J. John, J. Resop, N.Mueller and team. Random forests for global and regional yield forecasting. PLoS ONE magazine.
[3].Narayanan Balkrishnanand others.Govindarajan Muthukumarasamy. A plant ensemble machine learning model for prediction. International Journal of Computer Science and Software Engineering (IJCSSE).
[4]. S.Vinadhari, Dr. Bharat Mishra, and Dr. KD Singh. A machine learning approach to predict crop yield based on climate parameters. International Conference on Computer Communications and Informatics (ICCCI).
[5]. Svetlana K. Shahane, Project Leader V. Tavale. International Journal for Advanced Studies in Computer Science and Electronics (IJARCSEE), Volume 5, Issue 10, October 2016
[6] D. Ramesh, B. Vishnu Vardhan.Yield prediction analysis using data mining methods. IJRET: International Journal of Engineering and Technology Research.
[7] Ying Ding, Liang Wang, Yongwei Li, Daolian Li, "Predictive Control Models and Their Applications in Agriculture: A Review" Agricultural Computers and Electronics, June 2018.
[8] Anuja Chandgude, Nikita Harpale, Deeksha Jadhav, Poonam Pawar, Suhas M. Patil, "Overview of Machine Learning Algorithms Used in Crop Monitoring Systems", IRJET, April 2018
[9] Subhadra Mishra, Debahuti Mishra and Gour Hari Santra, "Application of Machine Learning Methods to Crop Production: A Review Article", Indian Journal of Science and Technology, October 2016
[10] Noran S. Uf, "Review of related applications Machine Learning in Agriculture", IJIREEICE, Aug. 2018
[11] Fabrizio Balducci, Donato Impedovo, and Giuseppe Pirlo, "Machine Learning Applications in Agricultural Datasets for Smart Farm Improvement," Machinery, Sep. 1, 2018.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.379

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details