IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 5, May 2023

Crop Yield Prediction and Recommendation in Agriculture using Machine Learning Algorithms

Prof. Ajit Karanjkar[1], Adesh Bhansali[2], Pravin Shende[3], Mayur Badgujar[4], Kunal Pawar[5]

Asst. Professor, Department of Computer Engineering[1]

UG Students, Department Computer Engineering[2,3,4,5]

Sinhgad College of Engineering, Pune, India

Abstract: The accurate prediction of crop yield is a complex and multifaceted task, requiring consideration of various factors such as climate conditions, soil properties, and geographic location. To achieve precise crop yield prediction, it is necessary to identify the relationships between these factors and crop yield using comprehensive datasets and advanced algorithms. In this paper, we propose the use of machine learning techniques, specifically Decision Tree and Random Forest models, to predict crop yield and provide crop recommendations. By analysing factors such as temperature, rainfall, and area, these models enable farmers to make informed decisions about crop selection and cultivation practices, while also mitigating the depletion of soil nutrients caused by continuous cultivation of the same crop. Our proposed crop recommendation system takes into account a range of factors, such as annual temperature, rainfall, and soil type and content, providing farmers with tailored recommendations for optimal crop selection and yield.

Keywords: Crop Yield Prediction, Agriculture, Machine Learning Algorithms, Recommendation Systems, Data Analysis.

I. INTRODUCTION

Crop yield prediction involves forecasting the amount of crop that will be harvested before it is actually harvested. This is done by analysing historical data and various factors that affect the yield of crops, such as rainfall, temperature, and soil properties. Predictive models can be built using machine learning algorithms to accurately forecast crop yields. Accurate crop yield predictions can help farmers plan their cultivation practices, optimize the use of resources, and mitigate the risks of crop failure.

As the global population continues to grow rapidly, agriculture has become more important than ever. However, despite the growth of the agriculture sector, it is still not enough to meet the needs of the population. To address food security challenges and reduce the impact of climate change, it is crucial to maximize crop yield. Accurate predictions of crop yield can help farmers make informed decisions about what and how much to grow. Governments and non-governmental organizations also rely on accurate predictions to make policies that support national food security. In short, accurate predictions of crop yield are essential for ensuring food security, and addressing the question of how much food can be produced each year is a critical step in achieving this goal.

India, agriculture is a major source of livelihood for 58% of the population. Each state has its own staple crop, which is typically grown in specific districts or areas. In the past, farmers used to make rough estimates of crop yield, but these estimates often failed due to various factors such as global warming, pollution, irrigation problems, and nutrient-deficient soil. However, with the advent of modern technology and IT companies entering the agriculture sector, crop yield prediction has become a key area of focus. Despite this, there haven't been significant improvements in recent years that can accurately predict crop yield while taking into account various factors, which could significantly reduce the burden on farmers.

DOI: 10.48175/IJARSCT-10062

II. RELATED WORKS

The paper discusses recent research on crop yield prediction and recommendation for the Indian agriculture industry. This paper employs advanced regression techniques like Kernel Ridge, Lasso, and ENet algorithms to predict crop yield and Stacking Regression to enhance prediction accuracy. The model takes into account easily available features such as State, District, Crop, Area, and Season, which allows farmers to access the application and get direct information on crop yield. However, this simplicity also presents some limitations, as the features used do not show significant variation over time, which may affect the accuracy and impactfulness of the predictions. Overall, the study provides an insightful analysis of the use of advanced regression techniques for crop yield prediction, although more research is needed to improve the model's accuracy and consider other relevant factors.[1]

In this paper, three models, namely Polynomial Regression, Decision Tree, and Random Forest were compared for crop yield prediction. The paper used features like Country, State, Humidity, Crop Name, Temperature, and Wind Speed for the prediction. Based on the comparison, it was found that the Random Forest model gave the best accuracy of 85% in crop yield prediction. However, the paper did not incorporate any recommendation system for suggesting crops to the farmers. This study's limitations include the lack of consideration of other important factors that affect crop yield, such as soil quality, irrigation, and fertilizers. Overall, the study provides valuable insights into the use of machine learning techniques for crop yield prediction, but more research is required to incorporate additional factors and develop a comprehensive recommendation system for farmers.[2]

The paper proposed using random forest instead of decision trees for crop yield prediction as decision trees tend to over-fit the data. The features used were humidity, crop name, temperature, and pH. The proposed solution was a random forest classifier to overcome the drawbacks of decision trees. However, the paper did not mention the accuracy of the model, which is an important evaluation metric.

This paper proposed a hybrid approach for crop yield prediction by combining machine learning and deep learning models such as SVM, LSTM, and RNN. The features used for prediction included area, rainfall, temperature, humidity, and pH value. The proposed pipeline approach led to higher accuracy and better predictions compared to using a single model. However, the paper lacked any interface to implement the model in technology, which limited the practical application of the idea. Therefore, the proposed approach remained theoretical without practical implementation.[3]

This paper utilizes Recurrent Neural Networks (RNN) such as LSTM and GRU, as well as their extensions BLSTM and BGRU, along with other models like Convolutional Neural Network and Multi-Layer Perceptron, to estimate tomato yield. The features used include Evapotranspiration, soil water volume, irrigation scheduling, solar radiation, temperature, and wind speed. However, the model requires a large amount of clean data for training and takes longer to train than other models. Despite the advantages of using RNNs, this approach also has some limitations, such as the need for high-quality data and significant computational resources. Overall, this study provides insights into the potential of deep learning methods for crop yield estimation, but further research is needed to overcome the challenges and limitations of this approach.[4]

III. METHODOLOGY

A system is proposed to predict crop yield and recommend crops based on weather and soil parameters using Decision Tree and Random Forest algorithms in machine learning. The system takes into account factors like rainfall and soil that influence crop growth.

A] Data

The Indian Crop Production dataset has crop production data for 124 crops in India from 1997 to 2015. The Rainfall dataset contains monthly rainfall data from 1901 to 2015 for all states in India. The Crop Recommendation dataset has soil attributes such as Nitrogen, Phosphorous, and Potassium content, soil pH, temperature, humidity, and rainfall, used for crop recommendation.

Top of Form

B] Data pre-processing and Features selection

The datasets were pre-processed to remove redundancy and null values. The data was also restructured by merging and de-merging various attributes to suit the models used. The most influential features were selected by doing correlation analysis between the output variable and different input variables from numerous features available in the dataset.

C] Models

The authors reviewed various models used in previous studies and found that the random forest and decision tree algorithms were the most suitable for their dataset and problem statement. Previous literature also supported these algorithms as the best in terms of accuracy.

D. System Architecture

The prediction models developed in this project are aimed to be used by farmers, government and non-government institutions. To provide an easy-to-use interface for users, a web interface and a mobile application were built. Users will enter input parameters through the client and send a request to the server. The pre-trained models will compute crop yield prediction and crop recommendation, and the result will be sent back to the client in response.

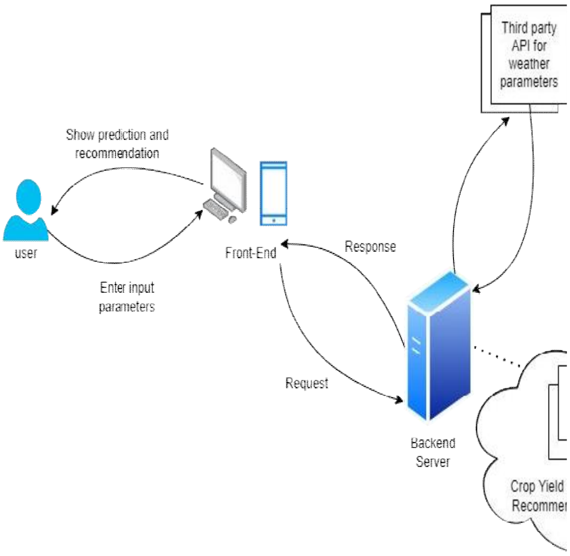

Fig. 1: System Architecture

IV. EXPERIMENTAL SETUP

A] Data Merging and Restructuring

To predict crop yields, we merged two separate datasets-crop production and rainfall. However, these datasets had different spatial parameters- crop production used states, while rainfall used subdivisions. To make the datasets compatible, we created subdivisions within each state in the crop production dataset and calculated the mean monthly rainfall for each of these subdivisions. We then merged the two datasets by adding the mean monthly rainfall for each subdivision to its corresponding state in the crop production dataset. This resulted in a single dataset containing both crop production and rainfall data, with a uniform spatial parameter of subdivisions.

B] Data Pre-processing

Null values were handled by dropping all the rows containing null values. Encoding categorical data to ensure it could be used effectively by machine learning models.

The 'state' and 'crop name' attributes in our dataset were encoded using a label encoder, as shown in Table I. Meanwhile, the 'season' attribute was One Hot encoded, as it had multiple categorical values including 'Autumn', 'Summer', 'Winter', 'Kharif', 'Rabi', and 'WholeYear', Removing Insignificant Data - Removed some data of crops due to a very less number of records available Normalizing Numerical Variables – normalization is performed to the numerical input variables in our dataset by using the Min-max scaler. This transformation rescaled the range of the variables to a standard range of 0-1, which helped to prevent any single variable from dominating our machine learning models.

V. RESULTS AND DISCUSSION

A] Data analysis and features selection -

After data gathering, merging, and pre-processing, attribute correlations were computed to choose input features.

State and Crop Production-

By examining the average yearly production per area of specific crops across different states (as shown in Figure 5.1), we observed that certain crops tend to yield better in specific states. This suggests that there is a correlation between the state and the crop yield

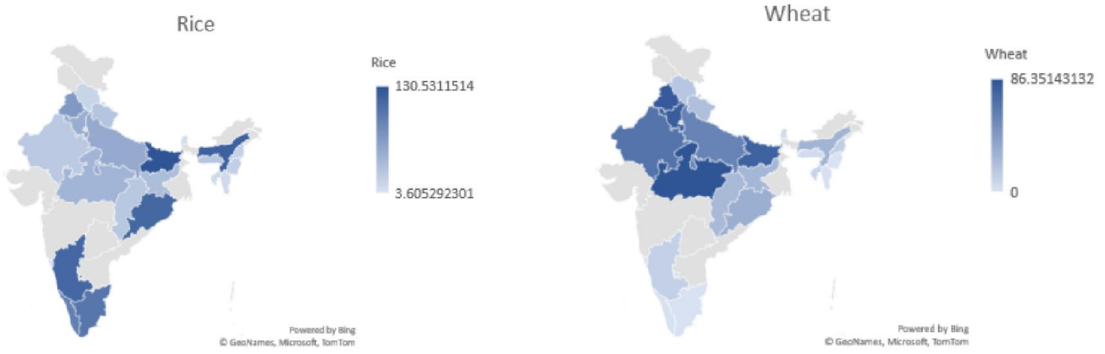

Fig. 5.1: Production per Area of crops in India

Rainfall and Crop yield - Rainfall vs crop yield plots for different crops revealed varying levels of rain sensitivity and requirements. Examples of crop yield dependency on rainfall are presented in Figure 5.2.

Season and Crop yield - The season in which a crop is grown is another crucial attribute that impacts crop yield. Figure 5.3 shows that yields of specific crops are concentrated in particular seasons, indicating that there is a correlation between the crop yield and the season in which it is cultivated.

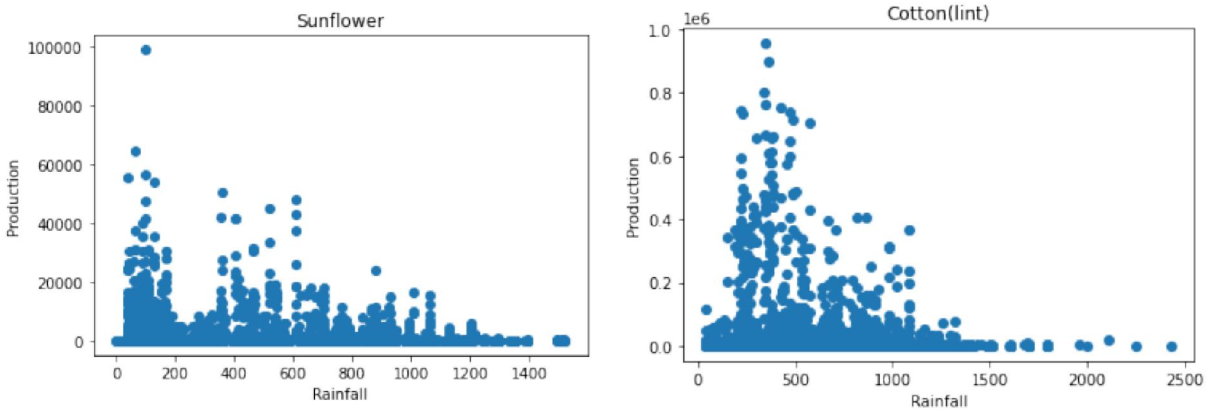

Fig. 5.2: Rainfall vs Crop Yield Scatter Plot for Sunflower and cotton

B] Crop Yield Prediction-

Decision Tree Regression and Random Forest Regression models are used For crop yield prediction.

DOI: 10.48175/IJARSCT-10062

Decision Tree Regression -

Figure 5 displays the accuracy of various parameters used, and the "absolute error" criterion was chosen as it provided the best accuracy. After experimenting with different max depth values, the highest accuracy of 0.92 was attained at a max depth of 14, as depicted in Figure 5.4.

Random Forest Regression-

Figure 7 displays the accuracy for various values of the number of estimators, indicating that an accuracy of 0.92 was obtained for 70 estimators. The accuracy for different max depth values is shown in Figure 8, and a maximum accuracy of 0.922 was achieved for a max depth of 12.

For crop yield prediction, Random Forest Regression has given the highest accuracy of 0.922

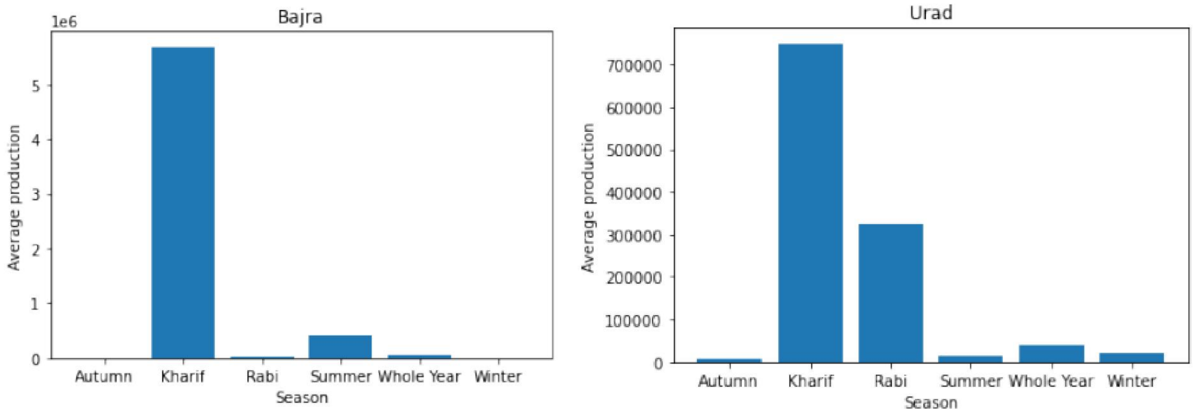

Fig. 5.3: Season wise distribution of the Crop Yield

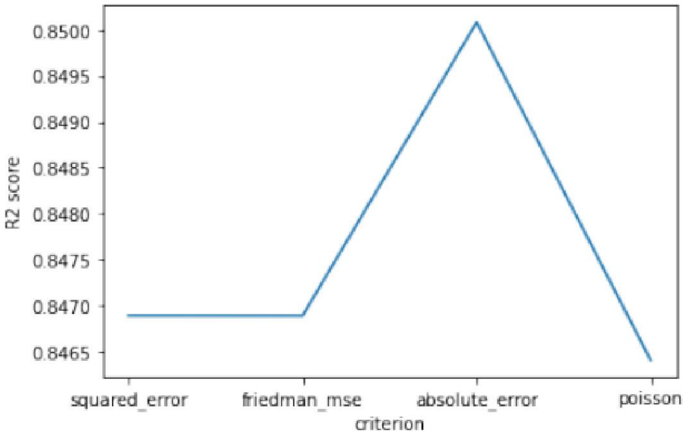

Fig. 5.4: Accuracy for different criterion in Decision Tree Regression

C] Crop Recommendation-

Decision Tree Regression and Random Forest Regression models are used For Crop Recommendation.

Decision Tree Classifier-

In Decision Tree Classifier, the accuracy for "Gini" and "Entropy" criterion parameters are 0.984 and 0.981, respectively. "Gini" was chosen for further experiments. The maximum accuracy of 0.988 was achieved at a max depth of 10 after trying different values. Refer to Fig. 5.5.

DOI: 10.48175/IJARSCT-10062

Random Forest Classifier -

The Random Forest Classifier was experimented with different values for number of estimators and max depth. The highest accuracy of 0.993 was achieved at 30 estimators, and the highest accuracy of 0.995 was achieved at max depth 10. Refer to Fig. 5.6.

For crop recommendation, Random Forest Classifier has given the highest accuracy of 0.995.

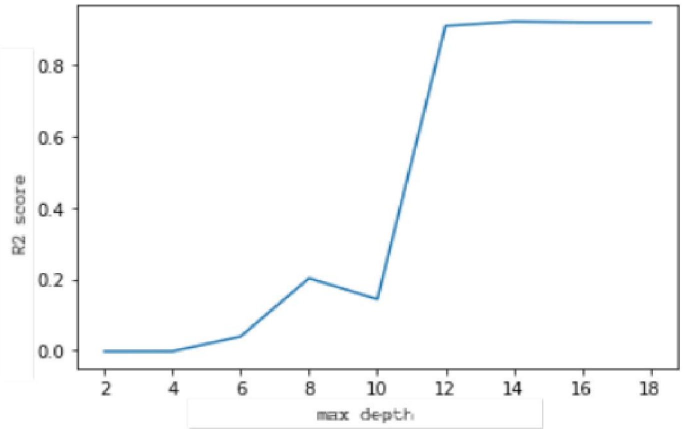

Fig. 5.5: Accuracy for different max depth in Decision Tree Regression

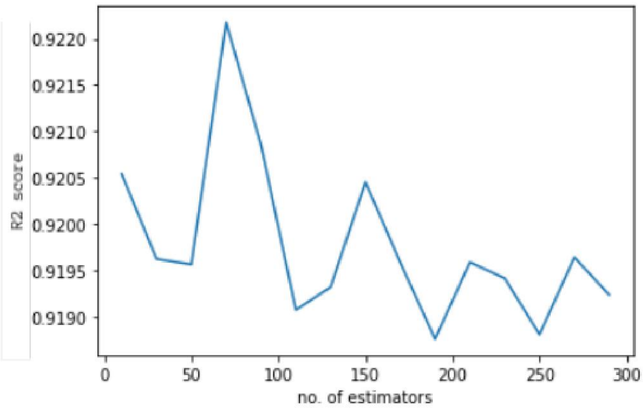

Fig. 5.6: Accuracy for different no. of estimators in Random Forest regression

VI. CONCLUSION

This paper presents a study on crop yield prediction and crop recommendation using machine learning algorithms. Three models were chosen based on a literature survey: Decision Tree and Random Forest regressor for crop yield prediction, and Decision Tree and Random Forest classifier for crop recommendation.

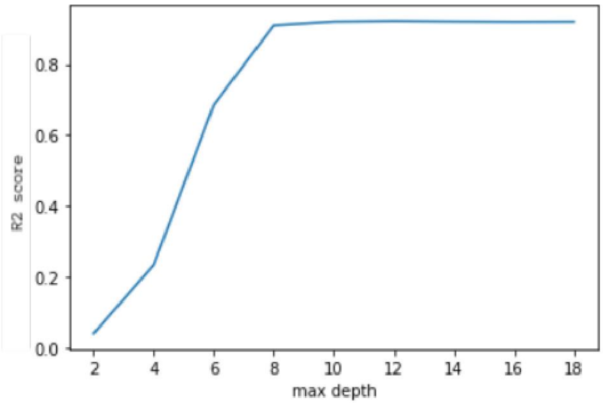

Fig. 6.1: Accuracy for different max depths in Random Forest regression

DOI: 10.48175/IJARSCT-10062

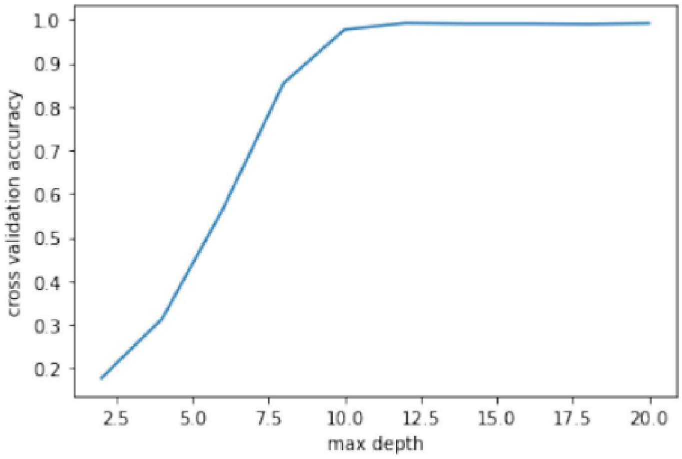

Fig. 6.2: Accuracy for different max depths in Decision Tree classifier

REFERENCES

[1]. Shilpa Mangesh Pande, Dr. Prem Kumar Ramesh, Anmol, B.R Aishwarya, Karuna Rohilla, Kumar Shaurya, "Crop Recommender System Using Machine Learning Approach", International Conference on Computing Methodologies and Communication (ICCMC), 2021, DOI: 10.1109/ICCMC51019.2021.9418351

[2]. P. S. Nishant, P. Sai Venkat, B. L. Avinash and B. Jabber, 'Crop Yield Prediction based on Indian Agriculture using Machine Learning,' International Conference for Emerging Technology (INCET), 2020, doi: 10.1109/INCET49848.2020.9154036

[3]. S. P. Raja , Barbara Sawicka , Zoran Stamenkovic, And G. Mariammal, "Crop Prediction Based on Characteristics of the Agricultural Environment Using Various Feature Selection Techniques and Classifiers", International Conference on Computing Methodologies and Communication (ICCMC), 2022, DOI 10.1109/ACCESS.2022.3154350

[4]. Agarwal, Sonal Tarar, Sandhya. (2021). A HYBRID APPROACH FOR CROP YIELD PREDICTION USING MACHINE LEARNING AND DEEP LEARNING ALGORITHMS. Journal of Physics: Conference Series. 1714. 012012. 10.1088/1742-96596/1714/1/012012.

[5]. Namgiri Suresh, :N.V.K.Ramesh, :Syed Inthiyaz, :P. Poorna Priya, :Kurra Nagasowmika, Kota.V.N.Harish Kumar, :Mashkoor Shaik and 2B. N. K. Reddy, "Crop Yield Prediction Using Random Forest Algorithm", International Conference on Advanced Computing & Communication Systems (ICACCS), 2021, DOI: 10.1109/ICACCS51430.2021.9441871.

[6]. D. J. Reddy and M. R. Kumar, 'Crop Yield Prediction using Machine Learning Algorithm,' 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 2021, doi: 10.1109/ICICCS51141.2021.9432236.

DOI: 10.48175/IJARSCT-10062