IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

RANDOM FOREST MACHINE LEARNING ALGORITHM FOR CROP YIELD PREDICTION

[1]Dr. Reshma Banu, [2]Harshavardhan N, [3]Bharath S, [4]Dileepa K P, [5]Vishal V Rao

[1]Professor, [2]B.E Student, [3]B.E Student, [4]B.E Student, [5]B.E Student,
[1]Department of Computer science and engineering,
[1]Vidya Vikas Institute of Engineering and technology, Mysore, India

Abstract: Agriculture has played a crucial role in the growth of India's economy, constituting approximately 26% of its GDP and catering to employment for 61% of the population. The motivation for this project comes from the increasing suicide rates among farmers. In 2022 the country witnessed nearly 1875 suicidal cases which may be due to low harvest in crops, and unable to repay loans in banking or private sectors. Climate and other environmental changes have become a major threat in the agriculture field. Machine learning is an essential approach for achieving practical and effective solutions to this problem. The proposed work predicts the yield of the crop using historically available data such as weather, soil, rainfall parameters, and historic crop yield and achieves this by using a machine learning algorithm. Random Forest algorithm can analyze crop growth related to the current climatic conditions and biophysical change and gives better solutions for the system. The proposed work develops a web application to predict crop yield in general and for a particular crop. Additionally, it suggests the right fertilizer to be used for the preferred crop to the farmer.

Key Words - Crop_ yield prediction; logistic regression; naive bayes; random forest; weather; Agriculture; Machine Learning; Supervised Algorithms; Data Mining.

I. INTRODUCTION

Throughout human history, agriculture has remained a primary and preeminent activity in every culture and civilization since its inception. It not only contributes significantly to the growing economy but is also crucial for human survival, especially in India where it supports a vast portion of employment. However, with the increasing demand for mass production, farmers often misuse technology, resulting in soil degradation and environmental harm. To reduce losses and increase yields, accurate information on crop yield is crucial, and machine learning can help achieve this goal. By applying various machine learning classifiers, such as Logistic Regression, Naïve Bayes, and Random Forest, a pattern can be obtained to predict crop yield from past data on weather, temperature, and other factors. The developed application uses the Random Forest algorithm, which has proven to be the most accurate, to provide a list of crops suitable for the entered data, along with a predicted yield value, with a focus on precision agriculture to ensure quality and minimize environmental factors.

Currently, many modern individuals lack awareness regarding proper crop cultivation methods in terms of timing and location. Such cultivation practices are contributing to changes in seasonal climate conditions, which in turn adversely affect essential resources such as soil, water, and air. These issues create food insecurity, and despite various analyses of factors such as weather, temperature, and other conditions, there are inadequate technologies and solutions to overcome these problems. In India, there are several approaches to enhance economic growth within the agriculture sector, including multiple strategies to increase crop yield and improve crop quality. Data mining can also aid in predicting crop yield production. The main objectives of this study are to utilize machine learning techniques for crop yield prediction, provide an easy-to-use user interface, increase the accuracy of crop yield prediction, and analyze different climatic parameters such as cloud cover, rainfall, and temperature.

II. LITERATURE REVIEW

In [1]. The publication named International Journal of Engineering Science Research Technology has released a research article that discusses the employment of machine learning algorithms to forecast crop yield. The paper specifically focuses on the Random Forest algorithm and how it can be used to predict the yield of crops based on existing data. The researchers used real data from Tamil Nadu to build and test their models. The results showed that the Random Forest algorithm can accurately predict crop yield.

In [2]. The study titled "Random Forests for Global and Regional Crop Yield Prediction" published in the PLoS ONE journal concludes that Random Forest (RF) is an efficient and versatile machine-learning approach for predicting crop yields at global and regional levels. The research findings indicate that RF has a high degree of accuracy and precision, is user-friendly, and

facilitates data analysis. Additionally, the study highlights that RF outperforms multiple linear regression (MLR) and is the most effective strategy for crop yield prediction.

In [3]. The International Journal of Computer Science and Software Engineering (IJCSSE) recently published a paper on projecting crop production over a certain period through the use of Ensemble Machine Learning models. The paper proposed using AdaNaive and AdaSVM as the models for this purpose. Implementation was carried out using AdaSVM and AdaNaive, to increase the efficiency of the Naive Bayes algorithm through the use of AdaBoost.

In [4]. A research paper presented at the International Conference on Computer Communication and Informatics (ICCCI) discusses a machine-learning approach for predicting crop yield based on climate parameters. The study developed a software tool called Crop Advisor, which is a user-friendly webpage for forecasting the impact of climatic parameters on crop yields. There are 5 algorithms available that can be used to identify the most influential climatic parameter affecting crop yields for specific crops in particular districts of Madhya Pradesh.

In [5]. In October 2016, the International Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE) released an article discussing crop cultivation forecasts. The article highlighted that the current process of soil analysis and interpretation of soil test results is still paper-based, which has led to poor interpretation of soil test results. This, in turn, has resulted in inadequate recommendations for crops, soil amendments, and fertilizers to farmers. This has resulted in poor crop yields, micro-nutrient deficiencies in soil, and excessive or insufficient application of fertilizers. The article proposed formulae to match crops with soil and fertilizer recommendations.

In [6]. The International Journal of Research in Engineering and Technology has published a paper on the analysis of crop yield prediction using data mining methods. The objective of the study is to create an interface that is easy to use for farmers, providing them with an analysis of rice production based on existing data. To maximize crop productivity, different data mining techniques were employed to anticipate the yield. One of these techniques is the K-Means algorithm, which was used to predict the pollution factor in the atmosphere.

In [7]. The article published in the International Journal of Research in Engineering and Technology discusses the use of data mining techniques to predict crop yield. The main objective of the paper is to develop an easy-to-use interface for farmers that provides an analysis of rice production based on available data. To increase crop productivity, various data mining methods were utilized, including the K-Means algorithm to forecast pollution levels in the atmosphere.

In [8]. This paper provides a thorough overview of the research that delves into the use of machine learning in agricultural production systems. The advent of big data technologies, techniques, methods, and high-performance computing has given rise to machine learning, which helps to explore, measure, and analyze data-intensive processes in the agricultural sector. The paper makes use of Support Vector Machines (SVM) for its implementation.

In [9]. Research conducted by Symbiosis Institute of Geoinformatics Symbiosis International University aimed to determine crop yield for insurance purposes through precision agriculture on an aerial platform. Precision agriculture involves the use of geospatial techniques and remote sensors to identify variations in crops and address them with different strategies. Crop growth variability in agricultural fields can be attributed to various factors such as crop stress, irrigation methods, pest and disease incidence, etc. The study utilized Ensemble Learning (EL) for implementation.

In [10]. The Symbiosis Institute of Geoinformatics at Symbiosis International University conducted a study on determining crop yield for insurance using precision agriculture on an aerial platform. Precision agriculture involves using geospatial methods and remote sensors to identify differences in the field and address them with various strategies. These variations in crop growth could be caused by crop stress, irrigation practices, or the presence of pests and diseases. The study utilized Ensemble Learning (EL) to implement the paper's findings.

III. METHODOLOGY

A. Pre-processing of Data

The method of data pre-processing is utilized to convert raw data into a clean and usable dataset. Since data is collected from various sources, it is usually in an unmanageable format for analysis. By utilizing various techniques, such as replacing null and missing values, data can be transformed into a comprehensible format. The final step in data pre-processing involves dividing the data into training and testing datasets. Due to the requirement of training models with as much data as possible, data is usually unevenly divided. The training dataset is the initial dataset used to train machine learning algorithms and produce accurate predictions. In this paper, the training dataset consists of 80% of the dataset.

B. Factors Influencing Crop Yield and Production

Several factors impact the yield and production of any crop. These factors serve as features that aid in predicting crop production over a year. This paper focuses on factors such as temperature, rainfall, area, humidity, and wind speed.

C. Analysing and Choosing the Best Machine Learning Algorithm

Before selecting an algorithm, it is crucial to evaluate and compare the available options and select the one that best fits the dataset. Machine learning is a practical solution to the problem of crop yield. There are several machine learning algorithms available for predicting crop yield. This paper uses logistic regression, naive Bayes, and random forest algorithms for accurate comparison and selection. The random forest can analyze crop growth related to current climatic conditions and biophysical changes. Random forest algorithm creates decision trees on different data samples and predicts the data from each subset, and by voting, provides a better solution for the system. For this paper's data, the random forest provides the highest accuracy of 92.81%, whereas logistic regression and naive Bayes algorithms provide an accuracy of 87.8% and 91.50%, respectively.

D. Model of Crop Prediction by Random Forest

The random forest model aggregates tree predictors such that each tree relies on the values of a random subset that is independently sampled and has the same distribution for all trees in the forest. The bagging method is used by random forest to train the data, which improves the accuracy of the result. This paper uses the random forest algorithm to achieve high accuracy in crop yield prediction. The predicted accuracy of the model is analyzed to be 91.34%.

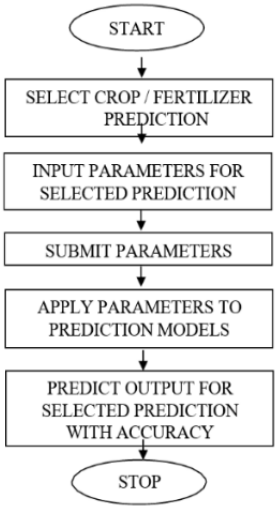

Fig 4.22 Flowchart of Crop Yield Prediction System

IV. PROPOSED SYSTEM

The proposed system is a mobile application that can predict the name of a crop and its yield. The crop's name is determined by analyzing various features such as temperature, humidity, wind speed, and rainfall, while its yield is calculated based on the area and production. The Random Forest classifier is utilized in this paper for prediction purposes, and it is expected to produce the most accurate crop predictions.

V. CONCLUSION

Accurately predicting crop yields in various districts of India can assist farmers in achieving higher profits. Agriculture relies on yield estimation models to increase production and meet demands. Model inputs are based on readings, and linear regression algorithms provide convincing estimation accuracy with high prognostic power. These algorithms can even be reformulated using alternative crop assessments for longer periods. Improvements to the system can lead to greater yields and better utilization of fields for crops.

VI. REFERENCES

[1] Rumelhart DE, Hinton GE, Williams RJ," Learning internal representations by errorpropagation". vol. 1, chapter 8. The MIT Press, Cambridge, MA (USA), pp:418-362, 1986.

[2] The article "Neural Network for Setting Target Corn Yields" by Liu J, Goering CE, and Tian L was published in the T ASAE journal in 2001. It presents a neural network model for determining target corn yields.

[3] "Statistical and Neural Methods for Site-Specific Yield Prediction" authored by Drummond ST, Sudduth KA, Joshi A, Birrel SJ, and Kitchen NR was published in T ASABE in 2003.

[4] "Statistical and neural methods for site-specific yield prediction" was authored by Drummond ST, Sudduth KA, Joshi A, Birrel SJ, and Kitchen NR in 2003 and published in T ASABE, with a volume of 46 and issue number 1, spanning from pages 5 to 14.

[5] "Integrating spatial data collection, modeling and analysis for precision agriculture" by Sudduth K, Fraisse C, Drummond S, and Kitchen N was presented at the First International Conference on Geospatial Information in Agriculture and Forestry in 1998. The paper is published in volume 2 and spans pages 166-173.

[6] "Artificial neural network model as a data analysis tool in precision farming" was authored by Irmak A, Jones JW, Batchelor WD, Irmak S, Boote KJ, and Paz JO. It was published in T ASABE in 2006

[7] Heuristic Prediction of Crop Yield using Machine Learning Technique by S. Pavani, Augusta Sophy Beulet P, International Journal of Engineering and Advanced Technology (IJEAT) Volume-9, December 2019 Smola A, Schlkopf B," A tutorial on support vector regression". Stat Comput 14(3): 199-222, 2004.

[8] A Survey on Crop Prediction using Machine Learning Approach by Sriram Rakshith. K, Dr. Deepak. G, Rajesh M, Sudharshan K S, Vasanth S & Harish Kumar, International Journal for Researchin Applied Science & Engineering Technology (IJRASET) Volume 7, Issue IV, Apr 2019

[9] Machine learning approach for forecasting crop yield based on climatic parameters by S.Veenadhari, Dr. Bharat Misra & Dr. CD Singh, International Conference on ComputerCommunication and Informatics(ICCCI -2014), Jan 2014

[10] "Support vector method for function approximation, regression estimation, and signal processing" authored by Vapnik V, Golowich S, and Smola A was released in 1997. It was published by MIT Press in Cambridge, MA, USA, and spans 281-287 pages.

[11] Predicting Yield of the Crop Using Machine Learning Algorithm by P.Priya, U.Muthaiah & M.Balamurugan, International Journal of Engineering Sciences & Research Technology (IJESRT), April 2018

[12] "Principles of Data Mining" authored by Hand D, Mannila H, and Smyth P was published by MIT Press in 2001.