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 10, May 2023

'Random Forest' Machine Learning Algorithm for Crop Yield Prediction

Dr. Reshma Banu[1], Harshavardhan N[2], Bharath S[3], Dileepa K P[4], Vishal V Rao[5]
Professor, Department of Computer Science and Engineering[1]
B.E Students, Department of Computer Science and Engineering[2,3,4,5]
Vidya Vikas Institute of Engineering and Technology, Mysore, India

Abstract: *With agriculture accounting for about 26% of India's GDP and providing work for 61% of the population, agriculture has been essential to the country's economic expansion. The rising suicide rates among farmers serve as the driving force for this project. The nation saw almost 1875 suicide cases in 2022, which may have been caused by poor crop yields or an inability to repay loans from the banking or private sectors. The field of agriculture is now seriously threatened by changes in the climate and other environmental factors. For this problem to be solved effectively and practically, machine learning is a crucial strategy. The proposed work uses a machine to estimate crop output utilizing data that has already been made accessible, including weather, soil, rainfall factors, and historical crop yield*

Keywords: Crop_ yield prediction; logistic regression; naive bayes; random forest; weather; Agriculture; Machine Learning; Supervised Algorithms; Data Mining

I. INTRODUCTION

Agriculture has always been the major and most important occupation in every culture and civilization that has existed throughout the history of mankind. It is essential for human survival as well as making a substantial contribution to the expanding economy, particularly in India where it provides a sizable number of jobs. However, because of the rising demand for mass production, farmers frequently abuse technology, which damages the environment and degrades the land. Accurate agricultural yield information is essential for reducing losses and increasing yields, and machine learning can aid in this Endeavor. It is possible to identify a pattern to forecast crop output using historical data on weather, temperature, and other parameters by applying different machine learning classifiers, such as Logistic Regression, Nave Bayes, and Random Forest. The created app makes use of the Random.

II. LITERATURE REVIEW

In [1]. A study article on the use of machine learning algorithms to anticipate crop yield has been published in the journal known as International Journal of Engineering Science study Technology. The Random Forest technique and its application to agricultural production prediction using historical data are the main topics of this research. The models were developed and tested by the researchers using actual data from Tamil Nadu. The outcomes demonstrated the accuracy of the Random Forest algorithm in predicting crop yield.

In [2]. Random Forest (RF) is an effective and adaptable machine-learning approach for predicting crop yields at global and regional levels, according to the paper "Random Forests for Global and Regional Crop Yield Prediction" that was published in the PLoS ONE journal. The results of the study show that RF is very accurate and precise, user-friendly, and helps with data processing. The study also shows that RF is the best method for predicting agricultural yield and exceeds multiple linear regression (MLR).

In [3]. A recent study using ensemble machine learning models to predict crop production during a specific time period was published in the International Journal of Computer Science and Software Engineering (IJCSSE). For this reason, the article suggested employing the AdaNaive and AdaSVM models. In order to increase the effectiveness of the Naive Bayes method by utilizing AdaBoost, implementation was carried out using AdaSVM and AdaNaive.

In [4]. A method to machine learning for forecasting agricultural production based on climate parameters is covered in a research article that was given at the International Conference on Computer Communication and Informatics (ICCCI).

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 10, May 2023

The research led to the creation of a user-friendly website called Crop Advisor, a software tool for predicting how meteorological factors will affect crop yields. There are five methods that can be used to determine the climatic factor that has the greatest impact on agricultural yields for particular crops in particular areas of Madhya Pradesh.

In [5]. Crop cultivation estimates were covered in an article published in October 2016 by the International Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE). The report emphasized that the interpretation of soil test data has been subpar due to the current paper-based soil analysis and testing process. As a result, farmers have received insufficient crop, soil improvement, and fertilizer suggestions. Poor crop yields, a lack of micronutrients in the soil, and an excessive or inadequate application of fertilizers are the results of this. The article suggested formulas to match recommended soil and fertilizer applications to crops.

In [6]. An article on the investigation of crop yield prediction using data mining techniques was published in the International Journal of Research in Engineering and Technology. The goal of the study is to develop a user-friendly interface that offers farmers an analysis of rice production based on current data. A variety of data mining approaches were used to predict the yield and maximize crop output. The K-Means algorithm is one of these methods, and it was used to forecast the amount of pollution in the atmosphere.

In [7]. The use of data mining techniques to forecast crop yield is covered in the International Journal of Research in Engineering and Technology article. The major goal of the article is to provide a user-friendly interface for farmers that analyses rice production using the data currently available. Several data mining techniques were used, including the K-Means algorithm to forecast atmospheric pollution levels, to boost crop output.

In [8]. The study on the application of machine learning in agricultural production systems is thoroughly reviewed in this publication. Machine learning has emerged as a result of the development of big data technologies, methodologies, approaches, and high-performance computers, and it now aids in the exploration, measurement, and analysis of data-intensive processes in the agricultural sector. Support Vector Machines (SVM) are used in the paper's implementation.

In [9]. Precision farming on an aerial platform was used in a study by the Symbiosis Institute of Geoinformatics at Symbiosis International University to estimate crop production for insurance purposes. Precision agriculture involves the use of geospatial tools and remote sensors to spot crop variations and take appropriate action. Variability in crop growth in agricultural fields can be attributable to a number of factors, including crop stress, irrigation techniques, the prevalence of pests and diseases, etc. Ensemble Learning (EL) was used in the study's implementation.

In [10]. Precision agriculture on an aerial platform was used in a study by the Symbiosis Institute of Geoinformatics at Symbiosis International University to estimate crop production for insurance purposes. Precision agriculture entails identifying variations in the field and addressing them with various ways by employing geospatial techniques and remote sensors. Crop stress, irrigation techniques, the presence of pests and diseases, or any of these factors may be the source of these variances in crop development. Ensemble Learning (EL) was used in the study to put the findings of the paper into practice, Crop stress, irrigation methods, or the presence of pests and pathogens may be to blame for these variances in crop development. To put the study's findings into practice, it used ensemble learning (EL).

III. METHODOLOGY

A. Pre-processing of Data

To transform unclean and unusable raw data into a clean and useable dataset, data pre-processing is used. Data is typically in a format that makes it impossible to analyses because it is gathered from numerous sources. Data can be converted into a readable format by using a number of strategies, such as substituting null and missing values. Creating training and testing datasets from the data is the last step in the pre-processing process. Data is typically distributed unevenly since training models must use as much data as feasible. The initial dataset used to train machine learning algorithms and provide precise predictions is known as the training dataset. In this study, 80% of the dataset is used for training.

B. Factors Influencing Crop Yield and Production

Any crop's productivity and yield are influenced by a number of factors. These elements act as features that help forecast crop yield over the course of a year. This essay concentrates on elements including temperature, rainfall, location, humidity, and wind speed.

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 10, May 2023

C. Datasets

Machine Learning depends heavily on data. It's the most crucial aspect that makes algorithm training possible. It uses historical data and information to gain experiences. The better the collection of the dataset, the better will be the accuracy.

The first step is Data Collection. For this project, we require two datasets. One for modelling the yield prediction algorithm and other for predicting weatherise., Average Rainfall and Average Temperature. These two parameters are predicted so as to be used as inputs for predicting the crop yield. The sources of our datasets are: https://en.tutiempo.net/ for weather data and https://www.kaggle.com/srinivas1/agricuture-crops-production-in-india for crop yield data.

	Crop	SoilType
2	Maize	Sandy
3	Arhar/Tur	Loamy
4	Bajra	Black
5	Gram	Loamy
6	Jowar	Loamy
7	Moong(Green Gram)	Loamy
8	Pulses total	Loamy
9	Ragi	Sandy
10	Rice	Loamy

Table 3.1 Soil and Crop data sample

The yield prediction module dataset requires the following columns: State, District, Crop, Season, Average Temperature, Average Rainfall, Soil Type, Area and Production as these are the major factors that crops depend on. _Production_ is the dependent variable or the class variable. There are eight independent variables and 1 dependent variable. We achieved this by merging the datasets. The datasets were merged taking the location as the common attribute in both.

	Year	Season	Avg Rainfall(mm)	Avg Temperature
2	1997	Rabi	42.35	27.7
3	1998	Rabi	46.2	27.8
4	1999	Rabi	44.4	27.7
5	2000	Rabi	15.42	27.6
6	2001	Rabi	34.02	27.3
7	2002	Rabi	10.97	27.7
8	2003	Rabi	8.47	27.5
9	2004	Rabi	12.57	27
10	2005	Rabi	16.57	27.5

Table 3.2 Rain and Temperature data

D. Analysing and Choosing the Best Machine Learning Algorithm

Prior to choosing an algorithm, it is critical to assess, contrast, and choose the alternative that best suits the dataset. The issue of crop yield has a workable solution in machine learning. For estimating crop yield, there are numerous machine learning techniques available. For accuracy comparison and selection, this study used the methods of logistic regression, naive Bayes, and random forest.

Crop growth in relation to present climatic conditions and biophysical changes can be analysed using random forest. By building decision trees based on various data samples, the random forest algorithm predicts the data from each subset and, through user voting, offers a more effective solution for the system. With an accuracy of 92.81% for the data in this study, random forest outperforms logistic regression and naive Bayes algorithms.

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 10, May 2023

E. Model of Crop Prediction by Random Forest

The random forest model combines tree predictors so that each tree depends on the results of an independent sampled random subset whose values have a uniform distribution over the forest as a whole. Random forest trains the data using the bagging technique, which increases the accuracy of the outcome. The random forest technique is used in this study to estimate agricultural yields with a high degree of accuracy. Analyses show that the model's anticipated accuracy is 91.34%.

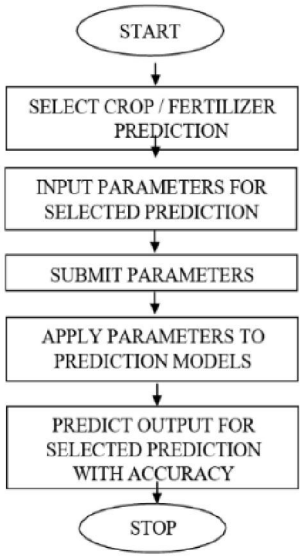

Fig 3.1 Flowchart of Crop Yield Prediction System

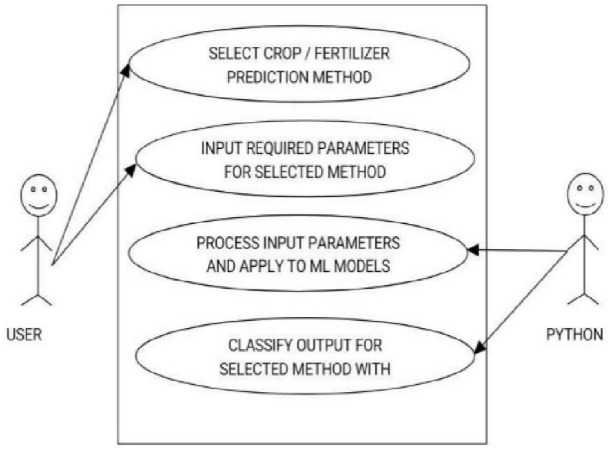

Fig 3.2 Use Case Diagram

A use case is a methodology used in system analysis to identify, clarify and organize system requirements. The use case is made up of a set of possible sequences of interactions between systems and users in a particular environment and related to a particular goal. A use case document can help the development team identify and understand where errors may occur during a transaction so they can resolve them.

The figure 3.2 represents the farmers (users) and their functional requirements provided by the system. The system involves two actors – End user (Farmer) and Admin. The functionalities provided by the system are represented in ovals. The arrows represent the dependencies and visibility of the functionalities

IV. RESULTS AND DISCUSSIONS

In the final implementation of the application the first screen the user can view is the login page.

Fig 5.1 Crop yield prediction screen

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 10, May 2023

V. CONCLUSION

In conclusion, a crop yield prediction and recommendation system utilize data analysis, machine learning algorithms, and crop growth models to provide farmers with valuable insights. By considering factors such as historical climate data, soil characteristics, and crop-specific parameters, the system can recommend suitable crops and estimate potential yields. This system assists farmers in making informed decisions about crop selection, optimizing resource allocation, and mitigating risks. Ultimately, implementing such a system can enhance agricultural productivity, resource efficiency, and profitability for farmers.

VI. FUTURE ENHANCEMENT

Future enhancements for crop yield prediction and recommendation systems may include incorporating advanced machine learning techniques such as deep learning and reinforcement learning to improve accuracy. Integration of real-time data from IoT devices, drones, and satellite imagery can enhance data collection. Integration with blockchain technology can provide transparent and secure data sharing. Additionally, leveraging emerging technologies like edge computing and 5G connectivity can enable faster data processing and analysis, enabling more timely and precise recommendations for farmers.

REFERENCES

[1]. RumelhartDE,HintonGE,WilliamsRJ," Learning internalrepre-sentationsbyerrorpropagation".vol. 1,chapter8. TheMIT Press, Cambridge,MA(USA),pp:418-362, 1986.

[2]. The article "Neural Network for Setting Target Corn Yields" by Liu J, Goering CE, 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, 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 across 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, AugustaSophy 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 ResearchinAppliedScience&EngineeringTechnology(IJRASET)Volume 7,Issue IV,Apr 2019

[9]. MachinelearningapproachforforecastingcropyieldbasedonclimaticparametersbyS.Veenadhari,Dr.BharatMisra &Dr.CDSingh,InternationalConferenceonComputerCommunicationandInformatics(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 across 281-287 pages.

[11]. Predicting Yield of the Crop Using Machine Learning Algorithm by P.Priya, U.Muthaiah &M.Balamurugan,InternationalJournalofEngineeringSciences&ResearchTechnology(IJESRT), April,2018