RESEARCH ARTICLE OPEN ACCESS

Smart Farming an ML and DL based WebProject

Abhay Purohit, Avnish Kumar, Ritu Vaishnav, Shubham Singh
Department of Computer Science and Engineering Global Institute of Technology
(RTU Affiliated) Jaipur, Rajasthan

ABSTRACT
Agriculture is a means of survival. Machine learning (ML) and deep learning (DL) have the potential to provide a real-world operational solution to yield issues. Given the current manual counting system, climate-friendly alternatives are needed. The results acquired through pest control and satellite images are not entirely reliable. The main focus of this research is on using machine learning approaches to predict yields. Here, the classifier model comprises logistic regression, naive Bayes, and random forest, with random forest delivering the highest accuracy. Farmers can use machine learning algorithms to predict crop yield by taking into account factors such as temperature, rainfall, and area. This bridges the gap between technology and the agricultural sector.
Keywords—Crop_yield_prediction; logistic_regression; naïve bayes; random forest; weather_api

I. INTRODUCTION

Agriculture has been the primary activity of all cultures and civilizations throughout human history. It is not only essential for the economy, but it is also vital for our survival. In India, it is a crucial sector that employs a significant portion of the population. However, the demand for production has increased exponentially, leading to the misuse of technology for mass production. The use of new hybrid varieties has become common, but these do not provide the necessary nutrients found in naturally produced crops. These unnatural methods harm the soil and contribute to environmental damage. To prevent losses, many farmers resort to these techniques, but if they have accurate crop yield information, losses can be minimized.

Machine learning is a rapidly evolving approach that can assist the agricultural industry in making sound decisions to improve their applications. Most devices today are supported by models that are tested before deployment. Using machine learning models to improve agricultural throughput is the main idea. Precision agriculture is the focus, where quality is guaranteed against unfavorable environmental factors. To make accurate forecasts and deal with inconsistencies in trends, machine learning classifiers such as logistic regression, Naive Bayes, and Random Forest can be used. Among these classifiers, the Random Forest algorithm produces the most accurate results.

The system forecasts yields based on a collection of historical data on weather, temperature, and a variety of other factors. The application we created runs the algorithm and displays a list of plants that are suitable for the input data, as well as the predicted yield values.

II. LITERATURE SURVEY

A study conducted by Aruvansh Nigam, Saksham Garg, and Archit Agrawal [1] has revealed that random forest machine learning algorithms when tested on Indian government datasets, provide the highest accuracy in yield prediction. The study found that sequential models, such as Simple Recurrent Neural Networks, predict precipitation well, whereas LSTMs predict temperature well. For yield prediction, the study took into account various factors such as precipitation, temperature, season, and area. When all these parameters were combined, the results showed that Random Forest is the best classifier. Leo Brieman [2] is an expert in random forest algorithm accuracy, strength, and correlation. The Random Forest algorithm constructs a decision tree for each data sample, predicts the data from each subset, and then votes to provide a good answer to the system. Random Forest uses the bagging method to train the data. Injected randomness should minimize correlation while maintaining intensity to achieve accuracy. Balamurugan [3] used only random forest classifiers to implement yield forecasting, considering various factors such as precipitation, temperature, and season. The dataset was not subjected to any other machine learning algorithms. However, due to a lack of comparison and quantification, a suitable algorithm could not be provided in the absence of other algorithms. Mishra [4] theoretically describes various machine-learning techniques that can be used in various predictive areas. Their work, however, does not implement the algorithm and thus cannot provide clear insight into the utility of the proposed work. Y. Jeevan and M.D. Nagendra Kumar [5] concluded that machine learning algorithms can be used to generate the appropriate function through a set of variables that can map the input variables to the target output. The Random Forest ML algorithm can be used to make predictions and reduce the number of models used while still delivering the most accurate crop predictions, according to this paper.

III. METHODOLOGY

A. Data Pre-Processing

Data preprocessing is the process of transforming raw data into a clean dataset that can be used for analysis. Raw data is often collected from various sources and is in an unusable format. To make it understandable, several techniques, such as replacing missing values, can be used. The final step in data preprocessing is to split the data into training and test datasets. The training dataset is used to teach the machine learning algorithm how to make accurate predictions, and it usually contains around 80% of the data. Figure 1 illustrates a few preprocessed data rows.

Crop yields and production are influenced by numerous factors. These factors are used to predict the yield of each crop throughout the year. This paper considers the effects of temperature, precipitation, area, humidity, and wind speed on crop yields. Figure 1 displays the plant name and its attributes that are used to predict yield calculations.

	A	B	C	D	E	F	G	H	I	J	K
1	State_N	District_	Crop_Ye	Season	Crop	Area	Production	Rainfall	Temper.	Humidit	Windspeed
2	Kerala	ALAPPU	1997	Whole Y	Arecanu	2253	1518	271	24.54	79.64	1.88
3	Kerala	ALAPPU	1999	Whole Y	Arecanu	2308	1043	242.9	23.97	80.66	2.12
4	Kerala	ALAPPU	2004	Whole Y	Arecanu	2376	1006	240.5	24.28	79.87	2.05
5	Kerala	ALAPPU	2007	Whole Y	Arecanu	1696	687	290.8	24.35	79.08	1.97
6	Kerala	ALAPPU	2008	Whole Y	Arecanu	1577	955	210.4	23.98	81.34	1.87
7	Kerala	ALAPPU	2011	Whole Y	Arecanu	1615.4	659.29	252.9	24.06	80.86	1.99
8	Kerala	ERNAKU	1998	Whole Y	Arecanu	3604	1941	262.6	24.78	79.9	2.15
9	Kerala	ERNAKU	2003	Whole Y	Arecanu	5275	3813	199.6	24.48	80.6	1.89
10	Kerala	ERNAKU	2007	Whole Y	Arecanu	5207	6395	290.8	24.35	79.08	1.97
11	Kerala	ERNAKU	2010	Whole Y	Arecanu	4549.9	4889.9	261	24.54	80.84	1.99
12	Kerala	ERNAKU	2014	Whole Y	Arecanu	4133	4533	253.9	24.66	79.45	1.93
13	Kerala	IDUKKI	2005	Whole Y	Arecanu	4009	4669	252.6	24.34	82.23	2.03

Fig. 1. Preprocessed data

B. Comparison and Selection of Machine Learning Algorithm

Before selecting an algorithm, it is essential to evaluate and compare each one to determine the best fit for the dataset. Machine learning is an effective way to provide a practical solution to the yield problem. There are several machine learning algorithms available to predict yields. This white paper includes a comparison of different machine learning algorithms based on their selection and accuracy.

- . Logistic regression:-. Logistic regression is a supervised learning classification algorithm that predicts the probabilities of target variables. It is a binary algorithm, meaning that it can only classify two types of target or dependent variable. When applied to a dataset, the logistic regression algorithm provides an accuracy of 87.8%.
- *Naive Bayes:*-, The Naive Bayes classifier assumes that the presence of one attribute in a class is independent of the presence of other attributes in that class. This model is easy to construct and is especially useful for handling large datasets due to its simplicity.Even highly sophisticated classification methods have been known to outperform Naive Bayes. It has a 91.50percent accuracy rate.
- *Random Forest*:- Random forests can be used to study plant growth in relation to current climatic conditions and biophysical changes. This algorithm creates a decision tree for each data sample, predicts the data from each subset, and then votes to improve the system's solution. To train the data, the bagging method is used in random forests, which improves the accuracy of the results. According to our research, RF has an accuracy rate of 92.81 percent. Out of the three algorithms, Random Forest clearly outperforms the others in terms of accuracy.

C. Random Forest Model for Crop Prediction

A random forest is a collection of tree predictors, each of which is based on the value of an independently sampled random subset. All trees in the forest have the same distribution. The data was trained using the bagging method by Random Forest, resulting in improved accuracy. We used a random forest algorithm to achieve high accuracy. It provides the accuracy of the predicates per model as well as the actual results of the predicates in the dataset. The model's prediction accuracy is estimated to be 91.34 percent. Figure 2 depicts a flow chart of a random forest model for forecasting yields.

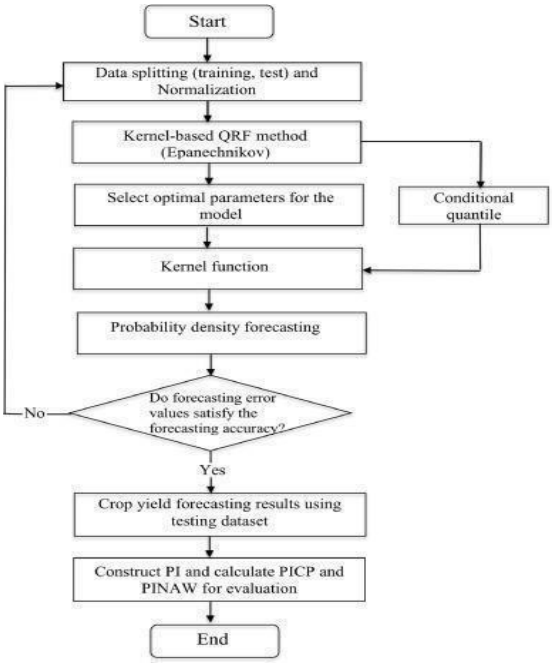

Fig. 2. Flowchart for Random Forest Model

D. System Architecture

In Figure 3, you can see the system architecture, which mainly consists of meteorological APIs. These APIs collect useful data such as temperature, humidity, and precipitation. Once the data is retrieved, it is sent to the server module where it is stored in a database. Users can create an account on the mobile app by registering once and providing basic information like their location and region.

The majority of the data that is entered is sent to the server. The server has an educated Random Forest algorithm that uses all of the fetched and entered data to predict crop yields. This algorithm can uncover the expected crop yield for a specific area..

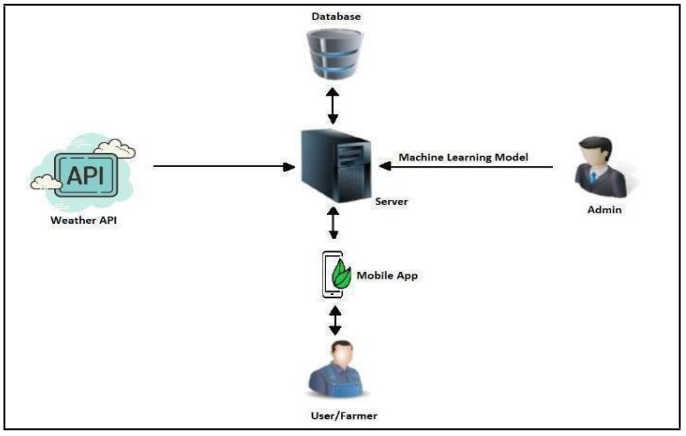

Fig. 3. System Architecture

E. Proposed System

We have developed a mobile application that predicts crop names and calculates yields based on those predictions. The yield is determined by taking into account the area and production, whereas the name of the crop is determined by considering several characteristics such as temperature, humidity, wind speed, and rainfall. We use a random forest classifier to predict the crop, and the crop forecasts with the highest accuracy values are taken into account.*G. System Analysis*

- *Python 3.8.5(Jupyter Notebook):*. Python is a programming language used for machine learning research. Jupyter Notebooks display the analysis process and outcomes.

Weather_API (Open Weather Map The Weather API is a tool that provides real-time weather information for a specific location by generating an API key. This information can be used to forecast crop yield.

- *Visual Studio* (Version 3.4.1): The official IDE for application development is Visual Studio. Flask Python is used for frontend design..

- *Python Flask Framework* (Version 2.0.1): Flask is a micro-framework written in Python.Flask is based on the

  WSGI (Web Server Gateway Interface) toolkit and the Jinja2 template engine. Use PaperFlask as the backend framework for building your application. This is a collection of modules and libraries that allow developers to write applications without writing low-level code such as protocols and thread management.

- *Heroku*: Heroku is a container-based cloud platform that enables developers to create, run, and maintain applications entirely in the cloud. For the server portion of this document, Heroku is used. You can connect your Heroku account to your GitHub repository and deploy it once it's been created..

IV. DISCUSSIONS

In this paper, crop production is boosted using machine learning techniques. This technology uses high-accuracy predictions to determine the right crop based on its yield. To implement machine learning algorithms in Python 3.8.5 (Jupyter Notebook), input libraries such as Scikit Learn, Numpy, Keras, and Pandas are used. However, the results of the machine learning analysis were questioned by the developer. The plant name and yield were then displayed in an Android app based on Flutter.

A. Datasets Used

The datasets have been obtained from different official

Government websites:

- data.gov.in-Details regarding area, production, crop name.
- indianwaterportal.org -Depicts rainfall details.
- power.larc.nasa.in –Temperature, humidity, wind speed details.

The total number of instances in the dataset is 4261. In 14 districts of Kerala, this includes information such as crop name, area, production, temperature, rainfall, humidity, and wind speed. You have the exact dataset you require during the data preprocessing phase. Figure 4 shows a heatmap with the individual attributes in the heatmap displayed.

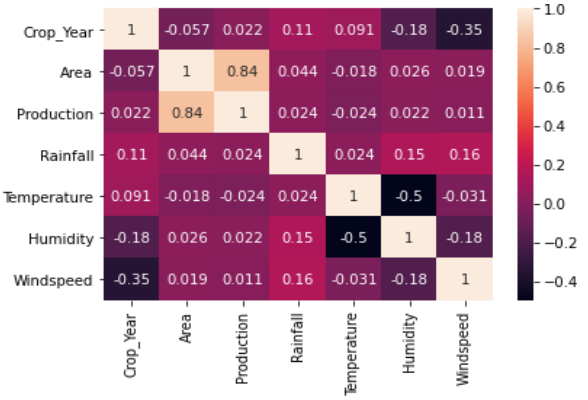

Fig. 4. Heat Map

B. Classifiers Used

Three machine learning classifiers - logistic regression, random forest, and naive bayes - were trained and compared for accuracy using the dataset.

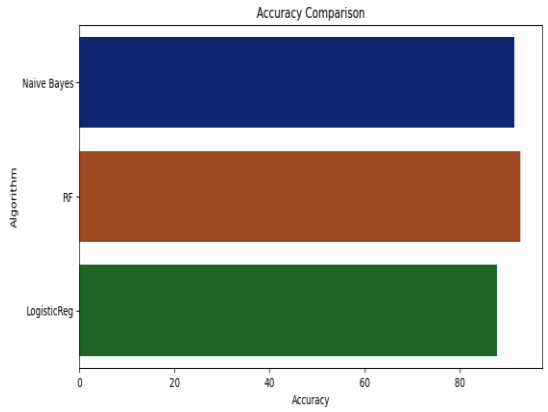

Fig. 5.Comparison Plot

To showcase the model's performance, a comparison chart has been constructed and is displayed in Figure 5. The Random Forest classifier exhibited the highest accuracy among the three classifiers utilized.

C. Weather_ API Used

You can use the Weather _API to access the current weather data where you need it. API used to get weather data

```
Enter the city name: kottayam
-------------------------------------------------------------
Weather Stats for - KOTTAYAM  || 16 May 2021 | 01:03:56 PM
-------------------------------------------------------------
Current temperature is: 26.00 deg C
Current weather desc  : moderate rain
Current Humidity      : 94 %
Current wind speed    : 1.54 kmph
Current rainfall      : 90
```

Fig. 6. API Output

The program we used was called "OpenWeatherMap". After entering the location name, we accessed the required measurement elements such as temperature, wind speed, humidity, and precipitation using the generated API key. The current weather for the entered location was displayed as a description on Graph 6.

Furthermore, we input API data into the trained dataset to indicate the crop name and yield.

D. Crop Name Prediction

As a first step towards predicting suitable crops for any given time and area, we utilized the most precise random forest classifier...

ALGORITHM	ACCURACY
RANDOM FOREST	92.81407341690006
NAÏVE BAYES	91.49626790098573
LOGISTIC REGRESSION	87.8298929223341

Table I : Comparison Table

In Kerala, we conducted a study on various plants that were readily available. We identified them and estimated the amount of yield that could be obtained based on the factors such as area, production, temperature, humidity, rainfall, and wind speed. We used a random forest classifier to train the preprocessed dataset and then used real-time meteorological data accessed via the API for the selected area to make accurate forecasts. After training the model, we were able to accurately predict the crop forecast for the selected area..

E. Crop Yield Calculation

The Random Forest classifier mapped the predicted crops to their respective production volumes and calculated yield by dividing area by volume [1].

Yield= Production/Area

The expected plant name for each yield helps farmers determine the right time to plant the right crop for maximum yield.

F. Web Application

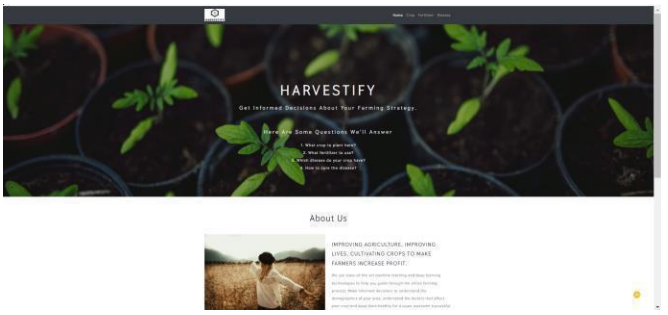

Fig. 7. Home Screen

A web application based on machine learning and deep learning has been developed to query the results of machine learning analysis. The application supports Android OS version 7 and has been created using the Java programming language. It has a user-friendly interface that requires only a few taps to get the desired results. The Android app suggests the right crop to grow based on the location and size of the field. The app accesses the user's details to query machine learning analysis and uses an API to provide weather data details based on the location. The machine learning classifiers use the retrieved weather data to predict the crop and calculate the yield. The server then fetches the output and displays it in the application.

The application has three main functions: account creation, detail entry, and results fetch. The account creation feature allows the user to interact more actively with the application interface. To proceed to the results activity, the user must fill out the field on the home page. The retrieved data is fed into a machine learning model, and the crop name is predicted based on the yield value calculated. The application is designed to be simple and efficient, making it easy for users to get the information they need quickly and easily.

V. CONCLUSION

The objective of this white research paper is to utilize machine learning techniques to predict crops and estimate crop yields. To ensure accuracy, various machine learning methods were utilized. The yield for the selected area was predicted using a random forest classifier. The proposed method involves implementing a system for predicting crops based on historical data. This system will help farmers decide which crops to plant in their fields. The information obtained from this task can be used to conduct efficient and informative harvests. Accurately predicting different specific crops in different districts will be beneficial for farmers in Kerala, and it will help boost the Indian economy by increasing crop yield rates.

VI. FUTURE SCOPE

In the upcoming years, you may consider implementing a data-independent system that works with equal accuracy regardless of the format. Incorporating soil knowledge into the system can aid in plant selection, making it more advantageous. Proper watering is also a crucial aspect of plant cultivation. Precipitation can help determine whether additional watering is required. By sharing this research work throughout India, it can reach new heights.

REFERENCES

[1]Aruvansh Nigam, Saksham Garg, Archit Agrawal "Crop Yield Prediction Using ML Algorithm", 2019

[2]Leo Brieman, "Random Forest", 2001

[3]Priya, P., Muthaiah, U., Balamurugan, M. Predicting Crop Yield Using Machine Learning Algorithms, 2015

[4]Mishra, S., Mishra, D., Santra, G.H. , "Application of Machine Learning Technology in Agricultural Production", 2016

[5]dr Y Jeevan Kumar, "Supervised Learning Approach to Crop Production",2020

[6] Ramesh Medar, Vijay S, Shweta, "Yield Forecasting Using Machine Learning Technology", 2019

[7] Ranjini B Guruprasad, Kumar Saurav, Sukanya Randhawa, Machine Learning Methodology for Paddy Yield Estimates in India: Case Study, 2019

[8] Sangeeta, Shruthi G, "Design and Implementation of Crop Yield Prediction Models in Agriculture," 2020

[9]M.Mathur, Rahul Jain, "DETECTION OF FRUIT DISEASES WITH HYBRID, DWT-GLCM APPROACH", Eur. Chem. Bull. 2023, 12(Special Issue 7), 613-624.

[10]Unmasking Embedded Text: A Deep Dive into Scene Image Analysis, Maheshwari, A., Ajmera.R., Dharamdasani D.K., 2023 International Conference on Advances in Computation, Communication and Information Technology, ICAICCIT 2023, 2023, pp. 1403–1408

[11]Internet of Things (IoT) Applications, Tools and Security Techniques, Kawatra, R., Dharamdasani, D.K., Ajmera, R,et.al. 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering, ICACITE 2022, 2022, pp. 1633–1639

[12]K. Gautam, A. K. Sharma, K. Kanhaiya and J. Dabass, "Temperature Measurement Using Fiber Bragg Grating Sensor for Industrial Applications" in International Journal of Current Research in Embedded System & VLSI Technology (Eureka Journals), vol. 7, no. 1, pp. 26-36, July 594, 2024.

2022.

[13] K. Gautam, S. K. Yadav, K. Kanhaiya and S. Sharma, "Hybrid Software Development Model Outcomes for In-House IT Team in the Manufacturing Industry" in International Journal of Information Technology Insights & Transformations (Eureka Journals), vol. 6, no. 1, pp. 1-10, May 2022.

[14]Pradeep Jha, Deepak Dembla & Widhi Dubey , "Implementation of Transfer Learning Based Ensemble Model using Image Processing for Detection of Potato and Bell Pepper Leaf Diseases", International Journal of Intelligent Systems and Applications in Engineering, 12(8s), 69–80, 2024.

[15]Pradeep Jha, Deepak Dembla & Widhi Dubey, "Deep learning models for enhancing potato leaf disease prediction: Implementation of transfer learning based stacking ensemble model", Multimedia Tools and Applications, Vol. 83, pp. 37839–37858, 2024.

[16]P. Upadhyay, K. K. Sharma, R. Dwivedi and P. Jha, "A Statistical Machine Learning Approach to Optimize Workload in Cloud Data Centre," 2023 7th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2023, pp. 276-280, doi: 10.1109/ICCMC56507.2023.10083957.

[17]Pradeep Jha, Deepak Dembla & Widhi Dubey , "Crop Disease Detection and Classification Using Deep Learning-Based Classifier Algorithm", Emerging Trends in Expert Applications and Security. ICETEAS 2023. Lecture Notes in Networks and Systems, vol 682, pp. 227-237, 2023.

[18]P. Jha, D. Dembla and W. Dubey, "Comparative Analysis of Crop Diseases Detection Using Machine Learning Algorithm," 2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS), Coimbatore, India, 2023, pp. 569-574, doi: 10.1109/ICAIS56108.2023.10073831.

[19]P. Jha, R. Baranwal, Monika and N. K. Tiwari, "Protection of User's Data in IOT," 2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS), Coimbatore, India, 2022, pp. 1292-1297, doi: 10.1109/ICAIS53314.2022.9742970.

[20]P. Jha, T. Biswas, U. Sagar and K. Ahuja, "Prediction with ML paradigm in Healthcare System," 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2021, pp. 1334-1342, doi: 10.1109/ICESC51422.2021.9532752.

[21]Gaur, P., Vashistha, S., Jha, P. (2023). Twitter Sentiment Analysis Using Naive Bayes-Based Machine Learning Technique. In: Shakya, S., Du, KL., Ntalianis, K. (eds) Sentiment Analysis and Deep Learning. Advances in Intelligent Systems and Computing, vol 1432. Springer, Singapore. https://doi.org/10.1007/978-981-19-5443-6_27

[22]P. Jha, D. Dembla and W. Dubey, "Implementation of Machine Learning Classification Algorithm Based on Ensemble Learning for Detection of Vegetable Crops Disease", International Journal of Advanced Computer Science and Applications, Vol. 15, No. 1, 2024.