JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Agro Consultant for farmers Using Machine Learning

Prof.A.B.Gadewar

Prathamesh.M.Girme[1] , Akshay.G.Joshi[2] , Tanvi.G.Satav[3] , Saurabh.V.Dubey[4]

"Information Technology Department , Savitribai Phule Pune University

PDEA's College of Engineering Manjri(Bk),Pune-412307(Maharashtra)India

Abstract— Agriculture is the single most crucial element for living. Machine learning (ML) may be a key viewpoint for finding a practical and workable solution to the crop yield problem. The results aren't particularly accurate given the current method, which includes manual counting, climate-smart pest management, and satellite photography. The major goal of this research is to use several machine learning approaches to forecast the agricultural production. Among the classifier models utilized here, Random Forest has the highest accuracy, followed by Logistic Regression and Naive Bayes. By taking into account variables like temperature, rainfall, area, and other characteristics, the predictions provided by machine learning algorithms will assist farmers in choosing which crop to cultivate to induce the greatest yieThis ties the technology and agricultural sectors together.

Keywords: naive bayes, random forest, logistic regression, crop yield prediction, and weather APIld.

I. INTRODUCTION

Since its inception, agriculture has been the main activity in every society and civilization that has existed throughout human history. It is not only a huge part of the expanding economy, but it is also necessary for our survival. It is also a vital sector for the Indian economy and the future of humanity. Additionally, it makes up a sizable amount of employment. As time goes on, the demand for production has dramatically expanded. People use technology in an utterly incorrect manner in order to produce in large quantities. Every day, new hybrid kinds are created. These kinds, however, don't offer the same critical components as a crop grown naturally. These artificial methods degrade the soil. All of this causes more environmental deterioration. Most of these unconventional methods are used to prevent losses.

However, the loss is reduced when agricultural growers have access to reliable crop production information. Machine learning is a rapidly expanding methodology that supports decision-making across all industries to provide the most useful of its applications. The majority of modern gadgets benefit from models being examined before deployment. The fundamental idea is to use machine learning models to boost the agricultural sector's throughput. The amount of knowledge imparted during the training period is another element that influences the prediction, as the number of parameters was higher in comparison. Precision agriculture, which priorities quality over unfavourable environmental variables, would be the main focus. To make an accurate prediction and take a stand against the divergent tendencies in Climate and precipitation To encourage a pattern, a variety of machine learning classifiers including Logistic Regression, Naive Bayes, Random Forest, etc. are used. Our analysis of the aforementioned machine learning classifiers led us to the conclusion that the Random Forest method offers the highest level of accuracy. The system forecasts crops based on the collection of historical data. The information is provided using historical data on the weather, temperature, and a number of other variables. Our application runs an algorithm and displays a list of crops that match the inputted data and their anticipated yield values.

II. LITERATURE SURVEY

On the Indian government dataset, Aruvansh Nigam, Saksham Garg, and Archit Agrawal[1] carried out trials and found that the Random Forest machine learning method provides the best yield forecast accuracy. Simple Recurrent Neural Network, a sequential model, is more effective at predicting rainfall than LSTM is at predicting temperature. For the purpose of yield forecast, the article combines

variables such as rainfall, temperature, season, area, etc. When all parameters are considered, the results show that Random Forest is the best classifier.

Leo Brieman [2] is an expert in the strength, accuracy, and correlation of the random forest method. The random forest algorithm builds decision trees using several data samples, predicts the data from each subset, and then determines the best solution for the system through voting. The data was trained in Random Forest using the bagging approach. The randomness must reduce correlation while retaining strength in order to increase accuracy.

Crop yield prediction has been implemented by Balamurugan [3] using simply the random forest classifier. To anticipate the agricultural output, various factors like rainfall, temperature, and season were considered. On the datasets, no further machine learning methods were used. Because alternative algorithms were lacking, comparison and quantification could not be done, making it impossible to provide the best algorithm.

Mishra [4] has provided a theoretical overview of a number of machine learning approaches that can be used in different forecasting contexts. But because they don't use any algorithms in their work, they can't really say if the suggested work is viable or not.

According to Dr. Y. Jeevan Nagendra Kumar [5], supervised learning allows machine learning algorithms to forecast an objective or outcome. This study focuses on supervised learning methods for predicting crop yields. To obtain , It must create an acceptable function from a set of variables that may map the input variable to the desired output in order to produce the given outputs. According to the paper, crop predictions may be made using the Random Forest ML method, which achieves the best accuracy value while taking into account the fewest number of models.

III. METHODOLOGY

Data pre-processing, first

A technique called data preprocessing is used to turn the raw data into a clean data set. The data are acquired from various sources, however because they are collected in raw form, analysis is not possible. We can change data into a comprehensible format by using various strategies, such as substituting missing values and null values. The division of training and testing data is the last step in the data preprocessing process. Due to the fact that training the model typically requires as many data points as possible, the data typically tend to be distributed unevenly. The training dataset, which in this case makes up 80% of the dataset, is the initial dataset used to teach machine learning algorithms how to learn and make accurate predictions.Fig.1. shows the few rows of the preprocessed data.

B. Factors Affecting Crop Production and Yield

	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	ERNAKL	1998	Whole Y	Arecanu	3604	1941	262.6	24.78	79.9	2.15
9	Kerala	ERNAKL	2003	Whole Y	Arecanu	5275	3813	199.6	24.48	80.6	1.89
10	Kerala	ERNAKL	2007	Whole Y	Arecanu	5207	6395	290.8	24.35	79.08	1.97
11	Kerala	ERNAKL	2010	Whole Y	Arecanu	4549.9	4889.9	261	24.54	80.84	1.99
12	Kerala	ERNAKL	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

The yield and productivity of any crop are impacted by a wide range of variables. These are essentially the characteristics that aid in estimating a crop's annual yield. We take into account variables like temperature, rainfall, area, humidity, and wind speed in this essay (Fig.1 shows the attributes for the crop name prediction and its yield calculation).

C. Evaluation and Algorithm Selection for Machine Learning

We must first assess and compare potential algorithms before selecting the one that best fits this particular dataset. The best method for solving the crop production problem practically is machine learning. Numerous machine learning methods are employed to forecast agricultural yield. The following machine learning techniques for selection and accuracy comparison are included in this paper:

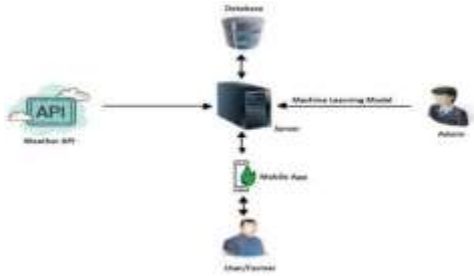
Logistic regression is a classification approach for supervised learning that is used to forecast the likelihood of a target variable. Since the dependent variable's nature is dichotomous, there are only two viable classes. The accuracy of the logistic regression approach when applied to our dataset is 87.8%.The Naive Bayes classifier makes the assumption that the existence of one feature in a class is unrelated to the existence of any other feature. Simple to construct and especially helpful for very big data sets is the naive Bayes model. In addition to simplicity,Naive Bayes is known to perform better than even the most complex classification techniques. It offers a 91.50% accuracy.

Random Forest is able to examine crop development in relation to the current meteorological conditions and biophysical change. The random forest algorithm builds decision trees using several data samples, predicts the data from each subset, and then determines which answer is best for the system through user voting. The bagging approach is used by Random Forest to train the data, increasing the accuracy of the outcome. RF offers a 92.81% accuracy for our data.

It is obvious that Random Forest provides the highest accuracy among the three algorithms when compared to other algorithms.

D. Predictive Random Forest Model for Crops

A random subset sampled individually, with the same distribution across all of the trees in the forest, is what makes up a random forest, which is an accumulation of tree predictors. The bagging method was utilized by Random Forest to train the data, increasing the accuracy of the outcome. The Random Forest approach, which provides accuracy for model-based predictions and the actual results of predictions in the dataset, was utilized to achieve high accuracy.



Analyzed is the model's anticipated accuracy, which is 91.34%. The random forest model's flowchart for predicting crop yield is shown in Fig. 2.

System Architecture (E)

The weather API is the key component of the system architecture shown in Fig. 3, from which we retrieve data on temperature, humidity, rainfall, and other variables. The server module receives the data that was retrieved from the API. The server's database is where the data is kept. The user can provide information such as location, area, etc. through the mobile application. The user can register once for an account on the mobile app And the server receives all of the entered data. The trained Random forest model installed on the server analyses all the input and fetched data to estimate crop yields and locates the yield of the named crop in the specified location.

Proposed System, Section F

A mobile application that we've suggested predicts the name of the crop and determines its related yield. Numerous factors, such temperature, humidity, wind speed, rainfall, etc., define the name of the crop, and output is influenced by area and production. The Random Forest classifier is employed in this study to make predictions. It will achieve the best accuracy levels for crop forecast.

System Analysis (G)

```
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
```

The platform for machine learning analysis is Python 3.8.5 (Jupyter Notebook), a programming language. Jupyter Notebooks provides the required output and illustrates the analysis process.

An application programming interface called Weather API (Open Weather Map) is used to obtain a location's current weather information. Current weather information required for crop prediction is displayed in the generated API key.

The official integrated development environment (IDE) for creating Android applications is called Android Studio (version 3.4.1). Java serves as the frontend design framework in this essay. The connection between the IDE and app is made using the USB debugging mechanism.

Python's Flask Framework (Version 2.0.1) is a small framework for the language. Flask is built using the Jinja2 template engine and the WSGI (Web Server Gateway Interface) tools. In this

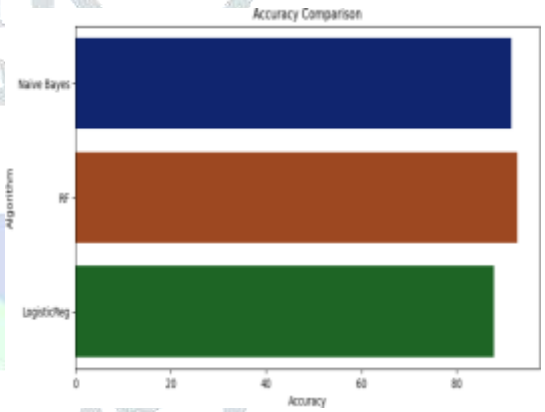


Fig. 3.Comparison Plot

study, the back-end framework for creating the application is flask. The developer can create applications without writing the low-level code necessary for protocols, thread management, etc. thanks to a collection of modules and libraries.

Heroku is a container-based cloud platform that enables programmers to create, launch, and manage apps solely in the cloud. For the server portion of this essay, Heroku is used. Once a Heroku account has been created, it can be linked to a GitHub repository before being deployed.

IV..Results and discussion are in section

This paper uses machine learning approaches to improve crop output. The method that produces high accuracy foretasted the yield of the correct crop. With input libraries like Scikit-Learn, Numpy, Keras, and Pandas, Python 3.8.5 (Jupyter Notebook) is used to implement the machine learning algorithms. An Android application that was created questioned the outcomes of the

machine learning analysis. The crop name and accompanying yield were displayed through an Android app built with Flutter.

A. Used Datasets

The datasets were retrieved from many official government websites, including:

Details about area, productivity, and crop name are available atdata.gov.in[8].

 indianwaterportal.org -Depicts specifics of rainfall[9].

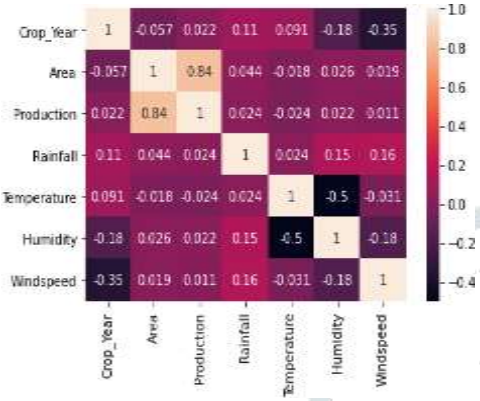


Fig. 4. Heat Map

Power.larc.nasa.in - Details on temperature, humidity, and wind speed[10].

name, area, productivity, temperature, rainfall, humidity, and wind speed. The pre-processing of the data produced the precise dataset that was required. Figure 4 illustrates a heat map used to depict the many properties present. The method that achieves high accuracy correctly forecasted

Classifiers Used

Logistic Regression, Random Forest, and Nave Bayes are examples of machine learning classifiers that were used for accuracy comparison and prediction. The ed dataset, which contains 4261 occurrences, was used to train these three classifiers.

Results and discussion are in section IV.

This study uses machine learning CombinB to improve crop output, and a performance comparison graph was created to indicate how well the models performed. The models' performance is shown in Fig. 5. The most accurate of the three classifiers was Random Forest.

Used C. Weather_ API

Utilizing the weather API allowed access to the necessary location's most recent weather information. API was utilized to retrieve weather data.

"OpenWeatherMap" was the term. Using the obtained API key, the district name was entered to access the required metro logical parameters, such as near-surface elements like temperature, wind speed, humidity, and precipitation. The current weather description for the entered location is shown in Fig. 6.Following data-set training, API data was provided as input to match the crop name with its yield.

D. Crop Name Forecast

The crop that can be cultivated in a certain district at a given period was predicted using the Random Forest Classifier, which had the highest accuracy.

ALGORITHM	ACCURACY	
RANDOM FOREST	92.81407991690006	
NAÏVE BAYES	91.49621790098573	
LOGISTIC REGRESSION	87.82982929223341	

Table I : Accuracy Table

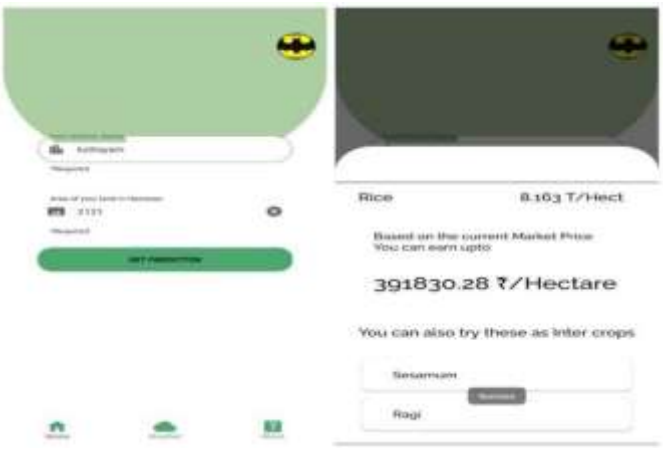
Abundantly growing crops in Kerala were chosen and their name was predicted and yield was calculated on the basis of area, production, temperature, humidity, rainfall and windspeed. The preprocessed dataset was trained using Random Forest classifier. Chosen district's instant weather data accessed from API was used for prediction. Trained model resulted in right crop prediction for the selected district.

A. Crop Yield Calculation

The crop which was predicted by the Random Forest Classifier was mapped to the production of predicted crop. Then the area entered by the user was divide from the production to get crop yield[1].

Yield = Area/Production

Farmers may choose the best time to plant the best crop for the most yield by using the estimated yield for each crop name.



F. An Android app

An Android app has been created to allow users to search the outcomes of machine learning analyses. Version 7 of the

Android OS is compatible with the app. Java was used to create the pages. The software offers an intuitive design that just needs a few taps to acquire the needed results. The Android app simply needs the field's location and size to identify the appropriate crop to grow there.

The app will query the machine learning analysis by accessing the user-inputted information. The weather information will be provided via the API using the location. In order to forecast the crop and determine the yield, a machine learning classifier gathers the received weather data. The server then fetches the output so

that the application may display the outcome.The application's primary functions were account creation, detail entry, and results retrieval. The account creation facilitates active user participation in the application interface. To go to the results action, the user fills out the field on the home page. The recovered data is fed into a machine learning model, and the crop name and calculated yield value are foretasted.

V. CONCLUSION

The use of machine learning algorithms to anticipate crops and determine their yield is the main topic of this study. The calculation of accuracy uses a variety of machine learning techniques. The crop prediction for the selected district made use of the Random Forest classifier. created a technique to anticipate crops using data gathered in the past. The suggested method aids farmers in choosing which crop to plant in the field. This work is done to learn more about the crops that can be used to harvest things in an effective and helpful way. Farmers in Kerala will benefit from accurate crop forecasting across many districts. As a result, the yield rate of crop production is maximized, which benefits our Indian economy.

This study's major focus is on the application of machine learning algorithms to forecast crops and estimate their yield. Several machine learning approaches are employed in the accuracy computation. The Random Forest classifier was used to predict the crop for the chosen district. developed a method to predict harvests using historical data. Farmers can choose which crop to plant in the field with the help of the provided approach. This research is being conducted to understand more about the crops that can be used to efficiently and effectively harvest various items. The accurate crop predictions across numerous areas will help farmers in Kerala. As a result, crop yield rates are maximized, which is good for the Indian economy.

VI. REFERENCES

[1] Aruvansh Nigam, Saksham Garg, Archit Agrawal “Crop Yield Prediction using ML Algorithms “,2019
[2] Leo Brieman, “Random Forests”, 2001
[3] Priya, P., Muthaiah, U., Balamurugan, M.”Predicting Yield of the Crop Using Machine Learning Algorithm”,2015
[4] Mishra, S., Mishra, D., Santra, G. H.,“Applications of machine learning techniques in agricultural crop production”,2016
[5] Dr.Y Jeevan Kumar,”Supervised Learning Approach for Crop Production”,2020
[6] Ramesh Medar,Vijay S, Shweta, “Crop Yield Prediction using Machine Learning Techniques”, 2019
[7] Ranjini B Guruprasad, Kumar Saurav, Sukanya Randhawa,”Machine Learning Methodologies for Paddy Yield Estimation in India: A CASE STUDY”, 2019
[8] Sangeeta, Shruthi G, “Design And Implementation Of Crop Yield Prediction Model In Agriculture”,2020
[9] https://www.data.gov.in
[10] https://power.larc.nasa.gov/data-access-viewer/
[11] https://en.wikipedia.org/wiki/Agriculture
[12] https://www.ibm.com/weather
[13] https;//flutter.dev
[14] https://openweathermap.org
[15] https;//builtin.com/data-science/random-forest-algorithm
[16] https://tutorialspoint/machine-learning/logistic-regression
http://scikit-learn.org/modules/naive-bay