International Journal of Innovative Research in Computer Science & Technology (IJIRCST)
ISSN: 2347-5552, Volume-10, Issue-3, May 2022
https://doi.org/10.55524/ijircst.2022.10.3.14
Article ID IRPV1039, Pages 72-75
www.ijircst.org

Design and Analysis of Prediction Model Using Machine Learning In Agriculture

Diksha Gupta[1], Dr. Yojna Arora[2], and Dr. Aarti Chugh[3]

[1] Student, Amity School of Engineering and Technology Gurugram, India
[2,3] Associate Professor, Amity School of Engineering and Technology, Gurugram, India

Correspondence should be addressed to Diksha Gupta; diksshagupta9@gmail.com

ABSTRACT- The reality of worldwide population growth and climate change demand that agriculture production can be increased. Traditional study findings which are difficult to extend to all conceivable fields since these are dependent on certain soil types, climatic circumstances, and background management combinations that aren't appropriate or transferable to all farms. There is no way for evaluating the efficacy of endless cropping system interactions (including many management practises) to crop production across the World. We demonstrate that dynamic interactions, that cannot be examined in repetitive trials, which are linked with considerable crop output variability and therefore the possibility for big yield gains, using massive databases and artificial intelligence. Our method can help to speed up agricultural research, discover sustainable methods, and meet future food demands. This is a paper attempted that at crop yield prediction using machine learning techniques with historic crop production data. For this, data has been collected from data.gov.in and data.world.

KEYWORDS- About four Machine learning, Big Data Analysis, Forecasting, Artificial Intelligence, Algorithms, Prediction and Analysis.

I. INTRODUCTION

Modern Agricultural market has a huge potential in a nation like India. Farmers in India play a major role in feeding the growing population of the country. This makes crop analysis and prediction as important as crop production. Farmers can use crop yield prediction data to make their decisions about crops. Agriculture yields' prediction is one of the major challenges in machine learning. There are many factors affecting crop yield such as crop genotype, environmental factors like soil conditions, farmer's efforts such as proper irrigation, timely plantation, etc. This forecasting is specifically relied on climate features to predict crop yield. Based on the considered datasets, we have taken into consideration the climate features specific to the agricultural seasons in India such as Rabi, Kharif and whole year and then have tried to make the crop yield predictions for the upcoming year.

The meteorological department's data sets of temperature, humidity, rainfall, and soil are analysed using big data analytics techniques. This type of study is carried out with the help of certain software tools, many of which are free source. The system will have information thanks to these tools and processes, and it will be able to make better decisions because to this processed information. As a result, greater results are guaranteed.

Farmers can usually predict the eventual yield-crop based on their previous experience with a specific crop. Farmers' yield predictions are inaccurate and ineffective. It is critical to adopt contemporary farming methods employing technology rather than traditional farming methods in order to meet the food needs of the entire population of the country and to export some agricultural goods to other countries. Modern farming practises enable farmers to plant crops in tiny areas with minimal water, fertilisers, and pesticides, resulting in a high yield and profit for the farmers.

II. LITERATURE REVIEW

Ashwani kumar Kushwaha [2] outlines crop yield prediction methods and suggests a suitable crop to boost the farmer's profit and the agriculture sector's quality. This study uses Hadoop platform and agro algorithm to acquire huge volume data, also known as big data (soil and meteorological data), for crop yield prediction. As a result of the repository data, crop suitability for certain conditions may be predicted, and crop quality can be improved.

Random forest for global and regional crop yield prediction are discussed in [4]. Journal PLoS ONE. Because of its highest accuracy and precision, ease of use, and value in data analysis, our generated outputs suggest that RF is a viable and flexible machine-learning method for agricultural production projections at regional and global scales. The most efficient technique is Random Forest, which outperforms multiple linear regression (MLR).

According to Rahul Katarya and Ashutosh Raturi,[5] they have provided various methodologies for which crop prediction for the states of Uttar Pradesh and Karnataka. These employ models such as Naive Bayes, Random Forest, KNN, and others. Cross validation, accuracy, RMSE, precision, and recall are some of the strategies used to evaluate performance on data. The model with the best performance is chosen and then use to classify and recommend crops. Wheat crop production investigates the use of machine learning in the production of wheat crops. The method entails employing digital image processing techniques to extract features for crop maturity and classifying the stage of growth using the supervised Machine

Learning (ML) technique. Data mining techniques [5] such as K-Means Clustering, KNN, SVM, and Bayesian network algorithms were used to estimate agricultural yield with great accuracy.

Crop yield forecasts based on climate parameters using supervised Machine Learning (ML). International Conference on Computer Communication and Informatics paper (ICCCI). Crop Advisor, a user-friendly online portal for estimating the impact of climatic conditions on crop yield, has been developed as part of the current project [6]. The C4.5 method is used to determine the most influential climatic parameter on agricultural production in Madhya Pradesh's selected districts. Decision Tree is used to implement the article [1].

III. PROPOSED SYSTEM

By studying the historical data of the farming area, the proposed method tries to predict or forecast crop yield. The system uses machine learning techniques to develop a predictive model by taking into consideration many elements such as soil conditions, rainfall, temperature, yield, and other things. We use a number of different of machine learning techniques here, including Random Forest, Linear Regression, and Decision Tree. The predicted accuracy is used to assess performance.

Using various data input, create, design, and implement a learning model. Using machine learning techniques, the system would learn the characteristics and predict the crop production from the data.

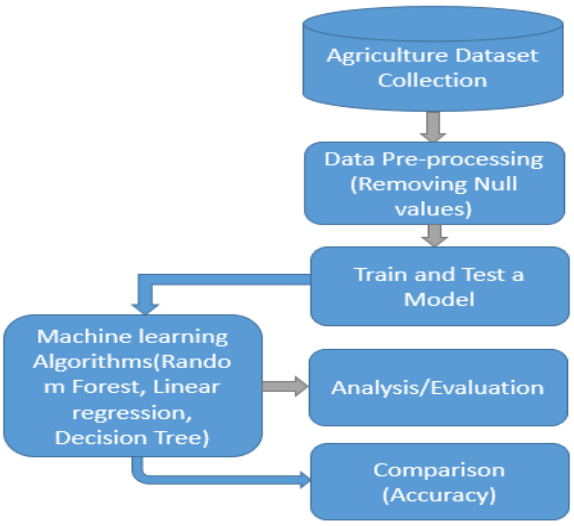

Figure 1: Block Diagram of a proposed model

IV. METHODOLOGY AND IMPLEMENTATION

A. *Dataset*

To perform our predictions, datasets that were basically required i.e., Historic data related to crop yield and data related to the climatic conditions. For crop yield production data, we used a dataset that had crop data for all districts of all states of India for all crops grown and their seasons from the year 1997 to 2015. The dataset was gathered from data.gov.in [7]. This dataset can be viewed using the dashboard that we created for demonstration purposes. Now, our main concern was to make use of this available data so as to make the predictions. The following attributes must be included in this dataset. These factors will be used for crop prediction: i) State Name ii) District Name iii) Crop Year iv) Season v) Crop vi) Area vii) Production. We have state wise data with district names and by this data we can analyse the production of crops depending on its area and season.

	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
0	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Arecanut	1254.0	2000.0
1	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Other Kharif pulses	2.0	1.0
2	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Rice	102.0	321.0
3	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Banana	176.0	641.0
4	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Cashewnut	720.0	165.0
...	...	...	...	...	...	...	...
246086	West Bengal	PURULIA	2014	Summer	Rice	306.0	801.0
246087	West Bengal	PURULIA	2014	Summer	Sesamum	627.0	463.0
246088	West Bengal	PURULIA	2014	Whole Year	Sugarcane	324.0	16250.0
246089	West Bengal	PURULIA	2014	Winter	Rice	279151.0	597899.0
246090	West Bengal	PURULIA	2014	Winter	Sesamum	175.0	88.0

Figure 2: Screenshot of Dataset

B. *Data Pre-processing*

After gathering data from a variety of sources. Before training the model, the dataset must be pre-processed. The data pre-processing process can be divided into several steps, starting with reading the acquired dataset and progressing to data cleaning. The datasets contain certain redundant attributes that are not considered for crop prediction during data cleaning. So, in order to improve accuracy, we must eliminate unnecessary attributes and datasets having some missing values, or fill them with unwanted nan values. Then decide on a model's goal. Using the sklearn library, the dataset will be separated into training and test sets after data cleaning.

```
uttar_pradesh.isnull().sum() # whether null is there, pre-processing
State_Name          0
District_Name       0
Crop_Year           0
Season              0
Crop                0
Area                0
Production        117
dtype: int64
```

```
uttar_pradesh = uttar_pradesh.dropna()  # deleting null
```

```
uttar_pradesh.isnull().sum()  #checking null
State_Name        0
District_Name     0
Crop_Year         0
Season            0
Crop              0
Area              0
Production        0
dtype: int64
```

Figure 3: Screenshot of Data Pre-processing of UP Model

For the implementation, we have basically made use of the following packages and the entire code is in python: numpy, pandas, pickle.

C. Prediction Algorithm Using Machine Learning

Machine learning predictive algorithms require highly efficient estimation based on previously taught data. Data, statistical algorithms, and machine learning techniques are used in predictive analytics to determine the likelihood of future outcomes based on historical data. The purpose is to provide the best judgement of what will happen in the future, rather than only knowing what has happened. We employed a supervised machine learning technique with classification and regression as subcategories in our system. Our system will benefit from a classification method.

```
from sklearn.metrics import r2_score
```

```
r2_score(y_test,y_pred)
0.9725662074203605
```

```
accuracy = rfregressor.score(X_test,y_test)
print(accuracy*100,'%')
97.25662074203605 %
```

```
accuracy = linear_regression.score(X_test,y_test)
print(accuracy*100,'%')
79.71654834047148 %
```

Figure 4: Screenshot of Accuracy of UP Model

V. ALGORITHMS

A. Linear Regression

Linear Regression is supervised Machine Learning technique in Python that observes continuous features and predicts a result. We can call it simple linear regression or multiple linear regression depending on whether it operates on a single variable or many features.

This is one of the most common Python Machine Learning(ML) methods, however it is often overlooked. It creates a line ax+b to anticipate the output by assigning optimal weights to variables[3]. We frequently utilise linear regression to estimate actual values based on continuous variables, such as the number of calls and housing costs. The best line that fits Y=a*X+b to denote a link between independent and dependent variables is the regression line.

B. Decision Tree

The greedy strategy is employed by decision trees, the attribute chosen in the first phase cannot be used subsequently to improve data classification. If Decision Tree is employed in the following phases, it may over fit the training dataset, resulting in unsatisfactory outcomes. To solve this flaw, an ensemble model is used, and ensemble models produce promising outcomes.

C. Random Forest

An ensemble of decision trees is known as a random forest. Trees vote for class, and each tree gives classification, in order to categorise the every new object based on its attributes. In the forest, the classification with the most votes wins. Random forest is also known as random decision forests, are an ensemble learning method for classification, regression, and other tasks that works by training the large number of decision trees and then its output of the class that

is the mode of the classes (classification) or mean prediction (regression) of the individual trees.

VI. RESULT AND DISCUSSION

As per the prediction, to evaluate the performance of the algorithm that we have implemented we checked the accuracy of our predictions and it gave the accuracy for various states is shown in the table below. This says that our implementation would always give the current prediction in terms of positive or negative directions i.e. for the chosen parameters, would the yield increase or decrease. Hence, if we compare both the algorithms which is Random Forest and Linear Regression, then as per evaluation Random Forest gave more accuracy.

Table 1: Accuracy Using Random Forest and Linear Regression

S.No.	State Name	Random Forest (Accuracy)	Linear Regression (Accuracy)
1	Uttar Pradesh	97.38	77.46
2	Rajasthan	88.54	65.34
3	Punjab	98.08	97.21
4	Maharashtra	81.20	69.22
5	Madhya Pradesh	87.23	77.81
6	West Bengal	92.63	92.26
7	Tamil Nadu	47.74	85.97

VII. CONCLUSION

We conclude that accurate yield, rainfall, and soil nutrient prediction systems are getting closer. We can forecast with excellent accuracy using ensemble learning algorithms. For greater performance, we can apply the big data analysis and mining techniques for large-scale forecasts. The data is one of the most important components here; we must analyse it by crop, season, and productivity. It is also necessary to educate farmers about such procedures. Other factors to consider for future study are fertiliser consumption on farm and the terrain of the area. A smartphone application that will notify farmers via text message when it is time to seed and harvest. It is necessary to make the technology accessible to everyone.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest

REFERENCES

[1] B M Sagar, NK Cauvery, P Abbi, N Vismita, B Pranava, Pranav A Bhat. "Chapter 105 Analysis and Prediction of Cotton Yield with Fertilizer Recommendation Using Gradient Algorithm", Springer Science and Business Media, 2022

[2] Ashwani kumar Kushwaha, Swetabhattachrya, "Crop Prediction using Machine Learning", International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181, 08 August-2020.

[3] Jeevan Kumar, Rajesh Kumar Tiwari, Vijay Pandey. "Diabetes prediction using machine learning tools", 2021 4th International Conference on Recent Trends in Computer Science and Technology (ICRTCST), 2022.

[4] Jig Han Jeong, Jonathan P. Resop, Nathaniel.D. Mueller, David H. Fleisher et al. "Random Forests for Global and Regional Crop Yield Predictions", PLOS ONE, 2016.

[5] Rahul Katarya, Ashutosh Raturi, Abhinav Mehndiratta, Abhinav Thapper, “Impact of Machine Learning Techniques in Precision Agricul- ture”,3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE- 2020), 07-08 February 2020.

[6] Pragathi Tummala, M Sobhana, Sruthi Kakumani. "Predicting crop yield with NDVI and Backscatter Networks", 2022 International Mobile and Embedded Technology Conference (MECON), 2022.

[7] Data.gov.in, https://data.gov.in.