Machine Learning-Based Crop Prediction: Boosting Accuracy and Supporting Sustainable Farming

Dr.Dasaradha Ramaiah[1], Harika Kairamkonda[2]
[1]*Professor and Head, Department of Information Technology, B V Raju Institute of Technology, Narsapur, TS, India*
[2]*Department of Information Technology, B V Raju Institute of Technology, Narsapur, TS, India*

Abstract— Agriculture employs more than half of India's population, making it a critical component of the country's economy. The unpredictability of weather and environmental circumstances, on the other hand, poses a substantial danger to agricultural productivity. Machine learning (ML) can be used as a decision-making tool for crop prediction and deciding which crops to produce to address this. Despite its utility, neural networks have drawbacks such as lower prediction efficiency and higher relative error. Supervised learning techniques also have difficulty capturing nonlinear correlations between input and output variables. Numerous studies have been conducted in order to develop reliable and effective crop categorization models, such as crop yield estimation based on meteorological conditions, disease diagnosis, and crop growth phases. This paper explores the use of the Random Forest algorithm, a type of ML technique, for crop prediction and provides a detailed analysis of its accuracy.This system takes inputs such as environmental parameters and soil characteristics before recommending the most suitable crop to grow.

Keywords—Agriculture, Neural Network, Neural Network, Crop prediction, Machine learning, Random Forest.

I. INTRODUCTION

Machine learning has proven to be an effective decision-making tool for predicting agricultural yields and deciding which crops to plant during the growing season. In crop prediction studies, a range of machine learning algorithms have been used to assist in addressing this complex topic. Machine learning is not only used in agriculture; it is also used in a variety of industries, from retail to telecommunications, to monitor client behaviour and estimate consumption trends. Yet, predicting crop yields in agriculture is a difficult process that necessitates the use of many datasets, given that several elements such as soil, seed, fertiliser, and meteorological conditions can all have an impact on crop productivity.

Farmers are currently facing difficulties in predicting weather patterns and determining the best crops to sow based on climate data. To address this challenge, machine learning has emerged as a crucial tool in various industries, including agriculture. In the context of agriculture, machine learning can be utilized to predict crop yields by combining past and present data to increase the accuracy of climate data. Machine learning is a methodology for analysing data to create automated models, and it is a branch of artificial intelligence that relies on the concept of systems learning from data to make decisions with minimal human intervention. One example of machine learning is a logical classifier, where a naive mathematician can predict the likelihood of data belonging to a specific group or class. This allows for more accurate predictions of crop growth and weather patterns, helping farmers make informed decisions about what to plant and when. Notwithstanding the difficulties, researchers have created a number of models that can estimate crop yields with reasonable accuracy. Still, they hope to enhance yield prediction accuracy. To do this, a study was created to assess various supervised learning algorithms, such as Random Forest, using a dataset that included 22 different types of crops. Entropy and the Gini Index were used to evaluate the performance of the Decision Tree and Random Forest classifiers. The results demonstrated that the proposed machine learning algorithms were highly effective, achieving the highest accuracy with precision, recall, and F1 Score. As a result, machine learning has enormous potential for use in crop prediction and might be beneficial in optimising crop yields and boosting agricultural production efficiency.

II. BLOCK DIAGRAM

The steps that are involved in crop prediction using machine learning methodology are stated as follows. Firstly, the agriculture Data is utilized for the crop prediction, Next, the data is undergone for pre-processing to remove the noisy data. The pre-processed data is undergone for feature extraction process that includes features such as soil information, nutrients, field management etc. which are used to perform the classification using ML algorithms. The results obtained by the existing models using ML algorithms are effectively described in the following section. Figure 1 shows the flow diagram of the crop prediction using ML algorithms.

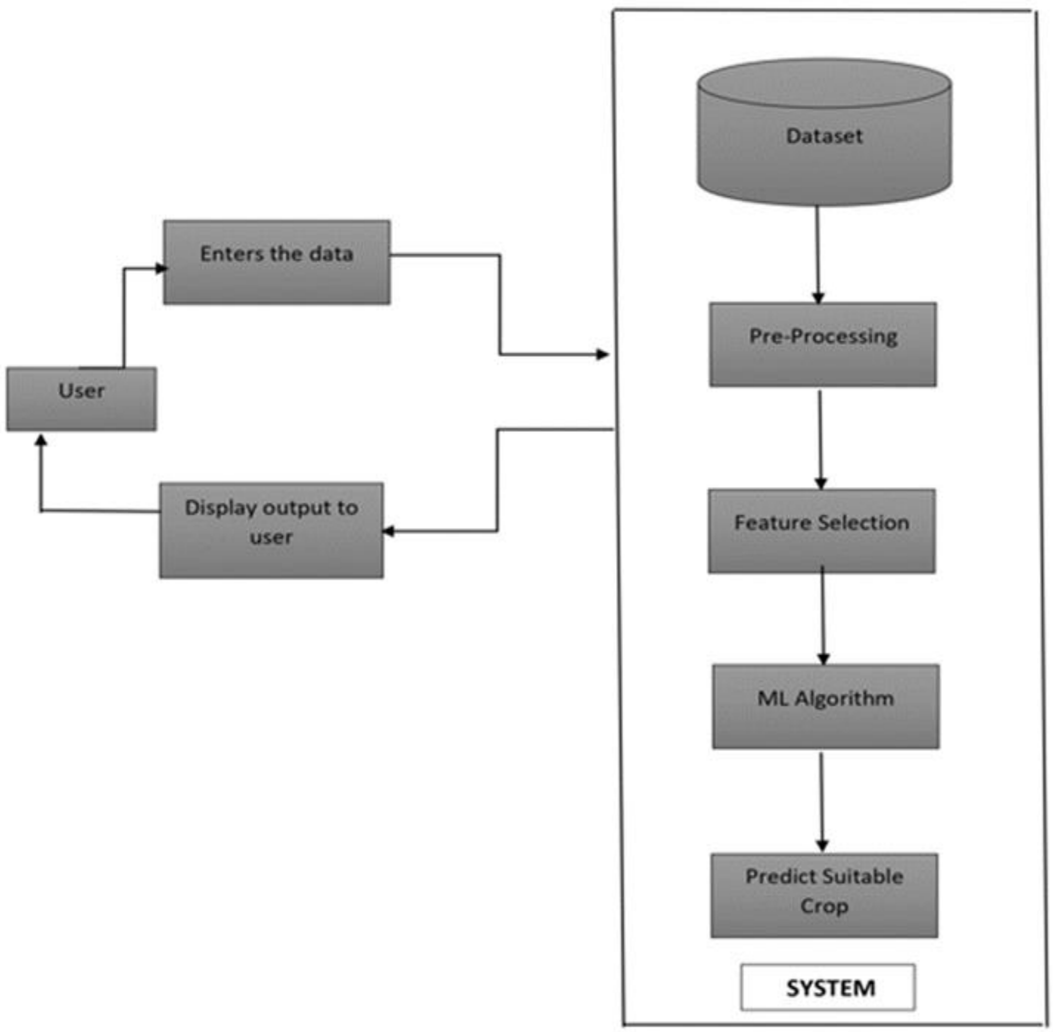

Fig. 1. Flow diagram of the crop yield prediction using ML algorithms

III. LITERATURE SURVEY

Crop prediction is an important task, and various methods have been proposed to improve its accuracy. This paper utilizes a feed-forward back propagation Artificial Neural Network to model and forecast crop yields in rural areas based on soil parameters (such as PH, nitrogen, potassium) and atmospheric parameters (such as rainfall, humidity), as well as input from farmers. However, the model displayed an unusual distribution when subjected to potential risks associated with air humidity, soil moisture content, and temperature. [1]

Tiwari and Shukla [2] new model was created to predict crop yields using CNN and Geographical Index. The existing model encountered issues with inconsistent agricultural practices due to unfavorable environmental conditions such as temperature, weather, and soil quality. The newly developed CNN model used spatial features as input and was trained with BPNN to predict errors. A benefit of the new model was that it used a real-time dataset obtained from reliable geospatial resources. However, while the new model reduced relative errors, it also decreased the efficiency of crop yield prediction.

Medar R et al.[3]It is feasible to create an application for farmers to aid in the mitigation of numerous agricultural concerns. Farmers would be able to run single or numerous experiments using input such as crop type, season, and location. After providing input, the user can choose a technique and generate outputs, which will display the expected crop yield rate. Data from past years is included in the datasets, which have been translated into a comparable format. Nave Bayes and KNN are the machine learning models used.

Fuentes et al. [4] to detect insect infestations and tomato plant illnesses in crops, a Robust Deep-Learning approach was used. Due to the prevalence of pests and illnesses, the old model had difficulty predicting crop production, resulting in large economic losses. To solve this issue, a deep metaarchitecture for predicting pests in plants was created. This model includes three main feature indicators: SDD (Single Shot Multibox Detector), Faster Region-Based CNN, and Region-Based Fully CNN. The deep meta-architecture and feature extractors implementation also provided a technique for global and local period explanation. With additional data expansion, the model's accuracy improved, and false positives in training reduced.The developed model was able to successfully identify different types of pests and diseases in complex situations from nearby areas. However, due to the use of complex pre-processing techniques, the robust deep learning method requires more time and higher computational costs.

A dataset was created using crop information from various sources, including government websites, gathered over the past ten years. An IoT device equipped with soil sensors, a Dht11 sensor for humidity and temperature, and an Arduino Uno with Atmega as a processor was set up to collect atmospheric data. Naive Bayes, a supervised learning algorithm, achieved 97% accuracy and was further improved by utilizing a boosting algorithm that uses weak rules through an iterative process to increase accuracy. To predict crop yields, advanced regression techniques such as ENet, Kernel Ridge, and Lasso algorithms were employed. These three regression techniques were improved using Stacking Regression for even better prediction.[5][6]

A comparative analysis was conducted between the current system and the proposed system that uses Naive Bayes and Random Forest, respectively. The proposed system was found to be superior. Although the Naive Bayes classifier's accuracy level is lower as it is a probability-based algorithm, the Random Forest algorithm has a high accuracy level since it is a bagging method.[7]

Andrew Crane Droesch, “Machine learning methods for crop yield prediction and climate change impact assessment in agriculture”, A semi-parametric variant of a deep neural network model has been proposed for predicting crop yields, and the impact of climate change on these predictions has been evaluated. The model combines the strengths of both parametric and non-parametric models, allowing for greater flexibility and accuracy in predicting crop yields. The effects of climate change on crop yields were also examined, and it was found that the model was able to accurately predict changes in yields under different climate scenarios. This approach could potentially help farmers and policymakers make more informed decisions regarding crop management and resource allocation in the face of changing climatic conditions.[8]

Aakash Parmar & Mithila Sompura, “Rainfall prediction usingMachine Learning Techniques” Data related to weather forecasting is considered to gain insights into predicting the crop yield.[9]

IV.METHODOLOGY

1. Collection of raw data

Data collection involves gathering and analyzing information from various sources. This is done to identify recurring patterns by analyzing past events. The 'Crop Recommendation' dataset, which includes 22 different crops classified as labels and 7 features, such as nitrogen, phosphorus, and potassium content ratios, temperature, humidity, pH value, and rainfall, was collected from the Kaggle website.

	N	P	K	temperature	humidity	ph	rainfall	label
0	90	42	43	20.879744	82.002744	6.502985	202.935536	rice
1	85	58	41	21.770462	80.319644	7.038096	226.655537	rice
2	60	55	44	23.004459	82.320763	7.840207	263.964248	rice
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice
4	78	42	42	20.130175	81.604873	7.628473	262.717340	rice
...	...	...	...	...	...	...	...	...
2195	107	34	32	26.774637	66.413269	6.780064	177.774507	coffee
2196	99	15	27	27.417112	56.636362	6.086922	127.924610	coffee

Fig. 2. Sample of Dataset

2. Data Pre-processing

Data preprocessing refers to the manipulation of raw data to make it usable for machine learning algorithms and to draw insights or predict outcomes. The primary objective of data preprocessing in this project is to detect missing values. It can be challenging to obtain every data point for all records in a dataset, as some values may be missing, denoted by null or a particular character like a question mark. However, in the dataset utilized in this project, no missing values were found.

3. Splitting data

It is a process of splitting the dataset into a training dataset and testing dataset using train_test_split() method of scikit learn module. 2200 data in the dataset has been divided as 80% of a dataset into training dataset-1760 and 20% of a dataset into testing dataset-440 data.

4. Fitting the model

"Model fitting" is the term used to describe the process of adjusting a model's parameters to improve its accuracy. To create a machine learning model, an algorithm is applied to a dataset in which the target variable is already known. The accuracy of the model is determined by comparing the model's outputs with the actual observed values of the target variable. If a machine learning model has good model fitting, it means that the model can accurately generalize data that is similar to the data on which it was trained. This is demonstrated when the model can accurately approximate the output for unknown inputs.

5. Prediction

When forecasting the likelihood of a specific result, “prediction” refers to the outcome of an algorithm after it has been trained on a previous dataset and applied to new data. Predicting the model using predict() method using test feature dataset. It has given the output as an array of predicted values.

V. CHALLENGES IN CROP PREDICTION USING MACHINE LEARNING

Below are the issues identified in previous research regarding the use of machine learning for predicting crop yields:

1. One of the problems encountered in using machine learning for crop yield prediction is the high cost involved in creating, repairing, and maintaining the algorithms due to their complex nature.
2. Another issue with using machine learning for crop yield prediction is that the technique used to combine input and output data did not produce statistically significant improvements in results.
3. The challenge encountered in using machine learning for crop yield prediction is that regression models with linear connections between parameters were unable to accurately predict outcomes in complex situations involving extreme value data and non-linear data.
4. The limitation of using machine learning for crop yield prediction is that existing K-NN models were used for classification, but their performance was reduced by the nonlinear and highly adaptable issues present in KNN. They were also operated in a locality model that increased the dimensionality of the input vector and made classification more confusing.

5. The classification decision was not accurate due to insufficient data available for estimating crop to be grown.

V. FUTURE STRATEGIES

Objectives to be followed in the future are given below:

1. To achieve accurate prediction, the modulating factor values of ML algorithms vary depending on the different crop feature categories.
2. Artificial Neural Networks (ANN) are preferred when the number of input variables is limited. Optimal features are selected through empirical analysis to improve crop yield estimation accuracy.
3. One of the benefits of using the ML method regression is that it overcomes the limitations of linear functions when dealing with a large output sample space and simplifies complex problems to linear function optimization.
4. The ML algorithm can handle large soil datasets to accurately estimate crop yields.
5. The application of ML techniques has significantly helped farmers increase crop production by providing valuable insights gained from observing agricultural fields.

VII. CONCLUSION

A machine learning model that uses a Random Forest classifier with a has been proposed to predict crops with high accuracy. The output of the model is the predicted seed based on the input parameters. This could be beneficial for farmers who lack the knowledge to predict crops for sustainable development. In the future, the model could be expanded to suggest appropriate fertilizers and guidelines for cropland and crops based on the given input. Additionally, the model monitors the source of sunlight and the health of the crops at regular intervals to achieve a better yield. The model with random forest classifier yielded an accuracy of 99.86%.

REFERENCES

[1] Dahikar S and Rode S V 2014 Agricultural crop yield prediction using artificial neural network approach International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering vol 2 Issue 1 pp 683-6.

[2] P. Tiwari, and P. Shukla, “Crop yield prediction by modified convolutional neural network and geographical indexes,” International Journal of Computer Sciences and Engineering, vol. 6, no. 8, pp. 503-513, 2018.

[3] Medar R, Rajpurohit V S and Shweta S 2019 Crop yield prediction using machine learning techniques IEEE 5th International Conference for Convergence in Technology (I2CT) pp 1-5 doi: 10.1109/I2CT45611.2019.9033611.

[4] A. Fuentes, S. Yoon, S. C. Kim, and D. S. Park, “A robust deeplearning-based detector for real-time tomato plant diseases and pests’ recognition,” Sensors, vol. 17, no. 9, pp. 2022, 2017.

[5] Nishant P S, Venkat P S, Avinash B L and Jabber B 2020 Crop yield prediction based on Indian agriculture using machine learning 2020 International Conference for Emerging Technology (INCET) pp 1-4 doi: 10.1109/INCET49848.2020.9154036.

[6] Kalimuthu M, Vaishnavi P and Kishore M 2020 Crop prediction using machine learning 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT) pp 926-32 doi: 10.1109/ICSSIT48917.2020.9214190.

[7] Geetha V, Punitha A, Abarna M, Akshaya M, Illakiya S and Janani A P 2020 An effective crop prediction using random forest algorithm 2020 International Conference on System, Computation, Automation and Networking (ICSCAN) pp 1-5 doi: 10.1109/ICSCAN49426.2020.9262311.

[8] Andrew Crane Droesch, “Machine learning methods for crop yield prediction and climate change impact assessment in agriculture”, Published by IOP Publishing Ltd Volume: 05 | OCT -2018.

[9] Aakash Parmar & Mithila Sompura, *"Rainfall Prediction using Machine Learning"*, 2017 International Conference on (ICIIECS) at Coimbatore Volume: 3 | March 2017.