J. Mountain Res. P-ISSN: 0974-3030, E-ISSN: 2582-5011 DOI: https://doi.org/10.51220/jmr.v18i2.23
Vol. 18(2), (2023), 217-222

Predictive Model for Smart Agriculture Using Machine Learning

Ravisha[1*] • Neelam Sinha[2]

[1]*Computer Science Engineering, Banasthali Vidyapith, Rajasthan*
[2]*S.D. (P.G) College, Muzaffarnagar, Uttar Pradesh*

*Corresponding Author Email id: ravisha.hacs10@gmail.com

Received: 15.11.2023; Revised: 16.12.2023; Accepted: 17.12.2023

Abstract: Agriculture holds a significant position in India's economic and employment landscape. A common challenge faced by Indian farmers is the lack of adherence to appropriate crop selection based on soil requirements, resulting in the cultivation of crops without a well-structured Crop Recommendation System. This adversely affects the productivity of crop yields. Precision agriculture emerges as a solution to this issue, characterized by the utilization of a soil database derived from the farm, expert-provided crop recommendations, and the incorporation of parameters such as soil quality from soil testing lab datasets. The proposed system takes input data on soil quality, including Nitrogen, Phosphorous, Potassium, and pH values, as well as weather-related information such as rainfall, temperature, and humidity. This information is utilized to predict the optimal crop for cultivation and enhance crop productivity. The research employs datasets obtained from the Kaggle website, utilizing machine learning algorithms to analyze the data. The study focuses on key parameters to determine the most suitable crops for cultivation in specific regions, aiming for heightened productivity. Among the various machine learning algorithms applied, both the Random Forest and Naive Bayes Algorithms demonstrated comparable results, achieving a remarkable accuracy score of 99.09% although XG Boost exhibited the highest accuracy, with a score of 99.31%.

Key Words: Crop recommendation • Precision agriculture • Random Forest Algorithm • Naive Bayes Algorithm • XG Boost

Introduction

Agriculture is the basic source and occupation of livelihood of people in India. As the largest economic sector in the country, agriculture plays a pivotal role in its comprehensive development. With over 60% of the land dedicated to meeting the needs of 1.4 billion people, the imperative shift toward modern agricultural technologies becomes evident (Kulkarni, et al 2018). Embracing these advancements is crucial, promising economic gains for our nation's farmers. In the past decade, it has been observed that there is not much crop development in the agriculture sector because the farmers choose crops based on market demand rather than the productivity of their land. They follow crop and yield predictions made in the past based on farmers' experience in a specific place or they prefer the previous or nearby or more popular crop in the surrounding region solely for their land, and they lack sufficient knowledge of soil nutrients such as nitrogen, phosphorus, and potassium in the soil. In the current condition, without crop rotation and insufficient nutrients applied to the soil, yields are reduced, soil pollution (soil acidity) occurs, and the top layer is harmed. Crop recommendation (Patel and Rane 2023) strategies can benefit farmers as they help determine the most productive crops per hectare. These strategies involve predicting crop yields well in advance to optimize productivity. Such prediction systems (Tapas Kumar, et al 2022), also known as recommendation systems, involve processing large amounts of data, including soil, fertilizer, and geographical and meteorological data (Gosai, et al 2021). To handle such large datasets, machine learning approaches are often employed. Machine learning, which is a subset of artificial intelligence, has arisen with big data technologies and high-performance computers to open new avenues for data-intensive study in the multidisciplinary Agri-technology area. In agriculture, the application of machine learning enables significantly enhanced precision, empowering farmers to address the needs of plants and animals at an individual level, thereby improving the efficacy of their decisions. In the agricultural sector, machine learning is not an enigmatic or mystical technique; rather, it comprises well-defined models that gather specific data and employ precise methodologies to achieve intended outcomes. The techniques employed in machine learning for agriculture stem from the learning process. Machine learning (ML) based systems offer complete and economical ways to improve agricultural yield through proper advising and decision-making (Kumar, et al 2015). By using various ML models to clarify the complex relationships that exist between plants, their surroundings, and soil characteristics, these programs offer accurate advice on crop management. The first step in the procedure is gathering field soil samples for careful scientific analysis. This allows the chemical composition to accurately represent the nutritional status as it is impacted by variables like temperature, humidity, pH, and rainfall.

By ensuring that the soil sample precisely reflects the distinct characteristics of a specific area, a methodical sampling methodology empowers farmers thereby maximizing productivity. This is the fundamental idea behind the crop recommendation and soil testing processes used in precision agriculture (Doshi, et al 2018). By putting effort into comprehending these procedures and ideals, researchers may build a user-friendly crop prediction system that provides recommendations with an error margin as small as possible while taking agricultural season changes and other pertinent factors into consideration.

In developing nations like India, where conventional or ancient agricultural methods are still widely used, the need for precision agriculture is especially apparent (Islam, et al 2018). Site-specific agriculture, also known as precision agriculture, helps farmers manage their land by increasing production per unit of land and reducing waste from fertilizer and pesticide use. Precision farming promotes yield prediction based on data.

Objective

- Create a crop recommendation model for accurate crop selection considering pH, rainfall, humidity, and temperature attributes.
- To enhance agricultural yield by offering high-accuracy and efficient predictions via the Ensemble learning approach.
- To decrease crop selection errors using precision agricultural techniques.
- To acquire a better diversity of crops that may be cultivated across the season. The suggested approach would assist farmers in minimizing crop selection challenges and increasing productivity.

Proposed System: Machine learning, imparts machines with the capability to learn and evolve from their experiences autonomously, eliminating the need for explicit programming. The advent of advancements in computing has broadened the application spectrum of machine learning (Patil, et al 2020). Its effectiveness in a plethora of fields such as forecasting, defect identification, pattern recognition, etc., has led to its increased global adoption. A significant challenge in agriculture is the prediction of crop yield (Gangolla, et al 2022). The insights from this study will empower farmers to forecast their crop yield before cultivation, thereby facilitating informed decision-making, as shown in Fig 1. It is crucial to provide farmers with timely guidance for predicting future crop production and the necessary analysis to optimize agricultural yield.

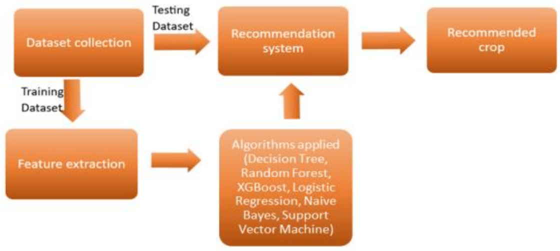

Figure 1: Block Diagram of Crop Prediction System

Dataset: The dataset, sourced from Kaggle, encompasses 2201 rows and 8 columns dedicated to crop prediction, as depicted in Table 1. A diverse set of techniques has been employed to refine this dataset, ultimately narrowing it down to four pivotal characteristics. Moreover, the dataset boasts an extensive 232705 rows and 8 columns (Chlingaryan, et al 2018). This dataset comprises eleven distinct crops, including watermelon, mango, lentil, moth beans, kidney beans, black gram, chickpea, papaya, pomegranate, maize, cotton, orange, rice, banana, muskmelon, grapes, coconut, mung bean, pigeon peas, jute, and coffee (K Ravi Kumar, et al 2023). This comprehensive dataset empowers users to suggest the optimal crops for cultivation on a given farm, leveraging a wide array of parameters. These parameters encompass data fields such as:

N: The ratio of Nitrogen content in the soil.
P: The ratio of Phosphorus content in the soil.
K: The ratio of potassium content in the soil.
Temperature: The temperature in degrees Celsius.
Humidity: The relative humidity in percentage.
pH: The pH value of the soil.
Rainfall: The amount of rainfall in mm.

Methodology: In this research paper, we utilized six different algorithms to predict the best accuracy score, subsequently incorporating it into our Crop Recommendation System (Singh, et al 2017). The procedural steps for each algorithm are outlined as follows:

(i) Importing the necessary library for the chosen machine learning algorithm.
(ii) Creating the machine learning classifier object.
(iii) Fitting the model to the provided data.

Table 1: Dataset for Crop Prediction System

	Nitrogen	phosphorus	potassium	temperature	humidity	ph	rainfall	label
2								
3	90	42	43	20.87974371	82.00274423	6.502985292	202.9355362	rice
4	85	58	41	21.77046169	80.31964408	7.038096361	226.6555374	rice
5	60	55	44	23.00445915	82.3207629	7.840207144	263.9642476	rice
6	74	35	40	26.49109635	80.15836264	6.980400905	242.8640342	rice
7	78	42	42	20.13017482	81.60487287	7.628472891	262.7173405	rice
8	69	37	42	23.05804872	83.37011772	7.073453503	251.0549998	rice
9	69	55	38	22.70883798	82.63941394	5.70080568	271.3248604	rice
10	94	53	40	20.27774362	82.89408619	5.718627178	241.9741949	rice
11	89	54	38	24.51588066	83.5352163	6.685346424	230.4462359	rice
12	68	58	38	23.22397386	83.03322691	6.336253525	221.2091958	rice
13	91	53	40	26.52723513	81.41753846	5.386167788	264.6148697	rice
14	90	46	42	23.97898217	81.45061596	7.50283396	250.0832336	rice
15	78	58	44	26.80079604	80.88684822	5.108681786	284.4364567	rice
16	93	56	36	24.01497622	82.05687182	6.98435366	185.2773389	rice
17	94	50	37	25.66585205	80.66385045	6.94801983	209.5869708	rice
18	60	48	39	24.28209415	80.30025587	7.042299069	231.0863347	rice
19	85	38	41	21.58711777	82.7883708	6.249050656	276.6552459	rice
20	91	35	39	23.79391957	80.41817957	6.970859754	206.2611855	rice
21	77	38	36	21.8652524	80.1923008	5.953933276	224.5550169	rice
22	88	35	40	23.57943626	83.58760316	5.85393208	291.2986618	rice
23	89	45	36	21.32504158	80.47476396	6.442475375	185.4974732	rice
24	76	40	43	25.15745531	83.11713476	5.070175667	231.3843163	rice

Crop_recommendation

Decision Tree: A Decision Tree is a tree-like model where internal nodes represent tests on attributes, branches represent outcomes, and leaves represent class labels. In crop prediction, it may evaluate factors like soil type, precipitation, and temperature. Its syntax often involves using the `DecisionTreeClassifier` in Python's scikit-learn library. Decision Trees are interpretable, allowing farmers to understand the decision-making process behind crop recommendations (Priya, et al 2018). However, they may overfit the training data, and parameter tuning is essential for optimal performance.

```
from sklearn.tree import
DecisionTreeClassifier
DecisionTree =
DecisionTreeClassifier(criterion="entropy",ra
ndom_state=2,max_depth=5)
DecisionTree.fit(Xtrain,Ytrain)
```

Naive Bayes : Naive Bayes is a probabilistic algorithm based on Bayes' theorem. In crop prediction, it could estimate the probability of a particular crop given observed conditions. The syntax often involves using `GaussianNB` in scikit-learn. Despite its assumption of feature independence, Naive Bayes can perform well with limited data and is computationally efficient. It is particularly effective for text-based data, making it suitable for certain types of agricultural information.

Support Vector Machine: Support Vector Machines classify data by finding the hyperplane that maximizes the margin between classes. In crop prediction, SVM may identify optimal decision boundaries based on features like temperature, humidity, and historical yields. Syntax includes using `SVC` in scikit-learn. SVM is powerful in handling high-dimensional data and is effective in scenarios with complex decision boundaries.

Logistic Regression: Logistic Regression models the probability of a binary outcome. In crop prediction, it could assess the likelihood of success for a particular crop. The syntax often involves using `LogisticRegression` in scikit-learn. Logistic Regression is simple, interpretable, and suitable for scenarios where the relationship between features and the outcome is approximately linear.

Random Forest: Random Forest is an ensemble learning method that constructs multiple decision trees and combines their outputs. In crop prediction, RF could handle diverse features like soil quality, weather conditions, and historical data. Syntax involves using `RandomForestClassifier` in scikit-learn. Random Forest is robust, handles non-linearity well, and reduces overfitting through aggregation. It is particularly useful when dealing with large, complex datasets in agriculture.

XGBoost: XGBoost is a scalable and accurate implementation of gradient boosting. In crop prediction, XGBoost could enhance decision trees' predictive power by boosting weak learners. Its syntax often involves using `XGBClassifier` in the XGBoost library. XGBoost excels in predictive accuracy, handles missing data gracefully, and is efficient in processing large datasets. It is suitable for tasks where precision is crucial, making it valuable in optimizing crop recommendations for farmers.

Experimental Results:

Our study demonstrates the effectiveness of our crop recommendation system in India's agriculture. XGBoost emerges as the optimal choice with the highest accuracy as illustrated in Fig 2, showcasing the system's potential to optimize crop selection for improved productivity.

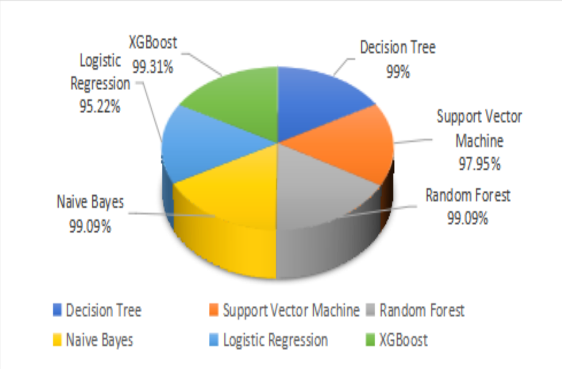

Fig 2: Accuracy Score in Percentage of Various Algorithms

Conclusion

This study presents a robust and effective crop recommendation system, designed to be user-friendly for farmers across India. The system aids farmers in making data-driven decisions on the type of crop to cultivate based on several factors such as Humidity, Nitrogen, Phosphorus, pH Value, Potassium, Rainfall, and Temperature. Utilizing this research could potentially enhance the country's productivity and profitability. It empowers farmers to select the most suitable crop, thereby maximizing their yield and contributing to the nation's overall economic gain. The study explores the application of various machine learning algorithms, including Decision Tree, Naïve Bayes, Support Vector Machine, Logistic Regression, Random Forest, and XGBoost, for crop recommendation. An analysis of these six algorithms revealed that XGBoost delivered the most accurate results, with an accuracy score of 99.31%.

References

Chlingaryan A, Sukkarieh S, Whelan B (2018) "Machine Learning Approaches for Crop Yield Prediction and Nitrogen Status Estimation in Precision Agriculture: A Review," *Computer Electronics in Agriculture*, 151, 61–69.

Doshi Z, Nadkarni S, Agrawal R, Shah N (2018) "Agroconsultant: Intelligent Crop Recommendation System using Machine Learning Algorithms," in 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA). IEEE, New York, pp. 1–6.

Gangolla P., Ashwina B., Akhila B., Rani N.D., Anuhya S., Kumar K.P. (2022) "Intelligent Crop Recommendation System using Machine Learning," *International Journal of Engineering Technology and Management Sciences*, 6, 81–87.

Gosai D, Raval C, Nayak R, Jayswal H and Patel A (2021) "Crop Recommendation System using Machine Learning," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, Volume 7, Issue 3, pp. 558–569.

Islam T, Chisty TA, Chakrabarty A (2018) "A Deep Neural Network Approach for Crop Selection and Yield Prediction in Bangladesh," in 2018 IEEE Region 10 Humanitarian Technology Conference (R10-HTC). IEEE, New York, pp. 1–6.

Kumar R, Singh M, Kumar P, Singh J (2015) "Crop Selection Method to Maximize Crop Yield Rate Using Machine Learning Technique," in 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM). IEEE, New York, pp. 138–145.

Kulkarni, N.H., Srinivasan G.N., Sagar B.M., Cauvery N.K. (2018) "Improving Crop Productivity Through A Crop Recommendation System Using Ensembling Technique," *IEEE Xplore*.

Mishra T.K., Mishra S.K., Jeevan Sai K., Peddi S., Surusomayajula M. (2022) "Crop Recommendation System Using Support Vector Machine Considering Indian Dataset," *Advances in Distributed Computing and Machine Learning*, volume 427 (First Online: 28 July 2022), pp. 501–510.

Patel, M. and Rane, A. (2023) "Crop Recommendation System," *Engineering Research eXchange*. Available at: https://engrxiv.org/preprint/view/2959/5442.

Patil A, Kokate S, Patil P, Panpatil V, Sapkal R (2020) "Crop Prediction using Machine Learning Algorithms," *International Journal of Advanced Engineering Technology*, 1(1), 1–8.

Priya P, Muthaiah U, Balamurugan M (2018) "Predicting Yield of the Crop using Machine Learning Algorithm," *The International Journal of Engineering Sciences & Research Technology (IJESTR)*, 7(1), 1–7.

Ravi Kumar K., Lahza H., Sreenivasa B.R., Shawly T., Alsheikhy A.A., Arunkumar H., Nirmala C.R. (2023) "A Novel Cluster Analysis-Based Crop Dataset Recommendation Method in Precision Farming," *Computer Systems Science and Engineering*, 46, 3239–3260.

Singh V, Sarwar A, Sharma V (2017) "Analysis of Soil and Prediction of Crop Yield (Rice) using Machine Learning Approach," *International Journal of Advanced Research in Computer Science*, 8(5).

******