Advanced International Journal for Research (AIJFR)
E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

A Study of the Application of Machine Learning to The Crop Recommendation System

Ekta Patel[1], Dr. Dand Hiran Jayantilal[2], Dr Neelam Labhade[3]

[1,2,3] Department of Computer Engineering,
Shri Jagdishprasad Jhabarmal Tibrewala University, Jhunjhunu Rajasthan, India

Abstract
Over 60% of India's workforce is employed in the agricultural sector, which contributes about 17% of the nation's GDP. With the advent of new technology like vertical farming and others, this area has witnessed some changes. Nonetheless, a lot of Indian farmers continue to use their property in accordance with customs and beliefs. For instance, instead of adapting to weather fluctuations, they wait for the weather to suit their farming methods. By forecasting which crops are compatible with the variables that affect crop growth, such as soil nutrients, soil pH, humidity, and rainfall, our research aims to assist farmers in selecting the optimum crops for their circumstances and surroundings. We employ a variety of machine learning models, including Gaussian Naïve Bayes (GNB), Logistic Regression (LR), Decision Tree (DT), and Support Vector Machine (SVM).

Keywords: Machine Learning (ML), Crop suggestion, Nitrogen-Phosphorus- Potassium (NPK), Humidity, Rainfall, pH, Decision Tree (DT), Logistic Regression (LR), and Gaussian Naïve Bayes (GNB).

1. Introduction

An important sector of the Indian economy and human life is agriculture. It is among the most important professions that are necessary for human survival. It also plays a significant role in our daily lives. Because they are unable to repay the bank loans they have taken out for farming, farmers typically commit suicide as a result of their decreased output. The environment is changing steadily these days, which is bad for agriculture and pushes farmers into debt and even to commit suicide. By applying different statistical or mathematical techniques to data, we can reduce these risks and assist farmers maximize their profits by recommending the optimal crop for their agricultural area.

India's agricultural sector has advanced significantly in recent years. "Area-specific" cultivation is the hidden weapon of precision farming. While there have been advancements, precision cultivation still has some issues. Precision agriculture has a big impact on crop recommendations. A number of factors influence crop recommendations.

The goal of precision agriculture is to pinpoint these factors in a region-specific manner in order to spot problems. While not all precision agriculture outcomes are correct, it is important to have precise and accurate recommendations in agriculture because mistakes can result in significant material and financial loss. Numerous studies are being conducted in an effort to develop a more precise and effective crop forecast model.

Algorithms such as supervised, unsupervised, and reinforcement learning are the emphasis of machine

learning, and each has pros and cons. The program uses supervised learning to build a mathematical model from a set of data that includes the inputs and the intended results. An algorithm that uses unsupervised learning builds a mathematical model from a set of data that only includes inputs and no labels for the intended outputs. Algorithms for semi-supervised learning build mathematical models from training data that is not fully supervised, meaning that some sample input is labeled.

The purpose of this study is to suggest the best crop depending on input characteristics such as soil PH, rainfall, temperature, humidity, nitrogen (N), phosphorus (P), and potassium (K). Using a variety of supervised machine learning techniques, this paper forecasts the accuracy of the future production of twenty-two distinct crops in India, including rice, maize, chickpeas, kidney beans, pigeon peas, moth beans, mungbean, black gram, lentil, pomegranate, banana, mango, grapes, watermelon, muskmelon, apple, orange, papaya, coconut, cotton, jute, and coffee crops. It then suggests the best crop. The dataset includes a number of factors, including soil PH, temperature, humidity, rainfall, phosphorus (P), potassium (K), and nitrogen (N). Several machine learning techniques, including Decision Tree (DT), Support Vector Machine (SVM), Logistic Regression (LR), and Gaussian Naïve Bayes (GNB), were used in this suggested system.

Review Of Literature

V. Spandana et. al (2020). "A Supervised Machine Learning Method for Predicting Crop Yield in the Agriculture Sector." Crop production can be predicted in this suggested approach using historical data from the past, which includes variables like temperature, humidity, ph, rainfall, and crop name. The greatest number of crop varieties will be covered in India's various districts under this strategy. We can forecast the optimal crop based on field weather conditions by implementing the suggested system. Decision trees and the random forest algorithm can be used to predict this crop. The most accurate value result was obtained by using the random forest technique. Results with greater accuracy increased agricultural output profits.

Suresh et. al (2021). "Digital Farming: An Effective Crop Yield Recommendation System Employing Machine Learning." Using certain data, this suggested system is utilized to identify a given crop. Support Vector Machines (SVM) were used to increase productivity and precision. Two datasets were the primary focus of this research paper: a sample dataset of crop data and a sample dataset of location data. This suggested system detected available nutrient values and necessary fertilizer levels for specific crops, such as rice, maize, black gram, carrot, and radish, and advised specific crops based on their nutrient (N, P, K, and PH) values.

Reddy et. al (2019). "A machine learning-based crop recommendation system to optimize crop yield in the Ramtek region." Three parameters were used by this proposed system: soil types, soil features, and crop yield data gathering. Based on these elements, the system suggested a suitable crop for the farmer to cultivate. Several machine learning methods, including random forest, CHAID, K-Nearest Neighbor, and Naïve Bayes, were tested by this suggested system. We can forecast specific crop conditions under specific weather conditions, state, and district values by implementing the suggested system. In order to boost national productivity, our suggested work would assist farmers in planting the appropriate seed depending on soil requirements.

Avinash Devare et . al (2017). "Crop recommendation system that uses machine learning to maximize crop yield." This suggested approach uses a soil database to identify a certain crop. Several crops, including groundnuts, pulses, cotton, vegetables, bananas, paddy, sorghum, sugarcane, and coriander, as

well as different characteristics including depth, texture, ph, soil color, permeability, drainage, water retention, and erosion, were all impacted by this suggested method. In order to accurately and efficiently recommend a crop for site-specific parameters, this suggested system works using a variety of machine learning classifiers, including Random Forest, ANN, support vector machine (SVM), and Naïve Bayes. In addition to preventing soil degradation on farmed land, this research would assist farmers in increasing agricultural productivity, reducing the use of chemicals in crop production, and making optimal use of water resources.

Doshi et. al (2018). "AgroConsultant: Machine Learning Algorithms for Intelligent Crop Recommendation System." AgroConsultant, an intelligent system, was created for this research article. The two subsystems that make up this suggested system are the Rainfall Predictor and the Crop Appropriate Predictor. In addition to fifteen minor crops (barley, cotton, groundnut, gram, jute, other pulses, potato, ragi, tur, rapeseed and mustard, sesame, soybean, sugarcane, sunflower, and tobacco) and five major crops (bajra, jowar, maize, rice, and wheat), this proposed system also takes into account factors like soil type, aquifer thickness, soil PH, topsoil thickness, precipitation, temperature, and location parameters. Several machine learning techniques, including Decision Tree, K Nearest Neighbor (K-NN), Random Forest, and Neural Network, were implemented in this suggested system and subjected to multi-label categorization. The accuracy of this suggested method was 71% when using a rainfall predictor model and 91.00% when utilizing a neural network algorithm on a crop-appropriate predictor system.

Shrikant et. al (2018). "Crop Recommendation System Survey." A crop recommendation system for intelligent farming was created by this suggested system. Numerous machine learning algorithms, including CHAID, KNN, K-means, Decision Tree, Neural Network, Naïve Bayes, C4.5, LAD, IBK, and SVM algorithms, were reviewed in this study. The Hadoop framework was used for this study's complex computations, which also improved the system's accuracy.

Kulkarni et. al (2018). "Improving Crop Productivity With An Ensembling Technique-Based Crop Recommendation System." This suggested approach is used to highly accurately select the appropriate crop based on the kind of soil and its properties, such as typical rainfall and surface temperature. This suggested system tested a number of machine learning methods, including Linear SVM, Random Forest, and Naive Bayes. The input soil dataset was categorized by this crop recommendation system into two recommendable crop types: Rabi and Kharif. Applying this suggested system produced an accuracy value of 99.91%.

Proposed System

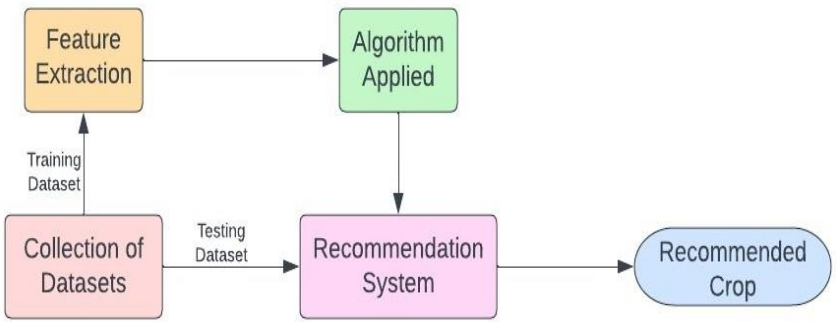

Fig 1: Block Diagram of Overall Methodology of Proposed System

In our framework, we have proposed a procedure that is separated into various stages as appeared in Figure 1.
The five phases are as per the following:

1) Collection of Datasets
2) Pre-processing (Noise Removal)
3) Feature Extraction
4) Applied Machine Learning Algorithm
5) Recommendation System
6) Recommended Crop

Flow of the Proposed System

As demonstrated in the figure, the methodology to extract the sentiment contains the several steps that are described below:

(1) Data Collection:

Nitrogen (N), phosphorus (P), potassium (K), soil PH, humidity, temperature, and rainfall are among the characteristics included in the dataset. The Kaggle website is the source of the datasets. 2200 instances or data points from historical data are included in the data set. Rice, maize, chickpeas, kidney beans, pigeonpeas, mothbeans, mungbean, blackgram, lentil, pomegranate, banana, mango, grapes, watermelon, muskmelon, apple, orange, papaya, coconut, cotton, jute, and coffee are among the twenty-two crops included in this dataset.
The dataset is divided into Train and Test sets, with 20% designated as the Test dataset and 80% designated as the Train dataset.

(2) Pre-Processing (Noise Removal)

Pre-processing is necessary for the application to be successful. Sometimes the information obtained from many sources is in its unprocessed state. It might include some conflicting, redundant, or incomplete data. Thus, it is necessary to filter such redundant data in this phase. Normalization of data is necessary. Additionally, we eliminate trash numbers, outliers, local min-max, and peak/downfall using Power BI.

(3) Feature Extraction

The goal of this stage is to find and use the dataset's most pertinent attributes. This procedure eliminates redundant and unnecessary data so that classifiers can be used.

(4) Methodology

Several machine learning techniques, including Decision Tree, Support Vector Machine (SVM), Logistic Regression (LR), and GaussianNB, were used in this suggested system.

A. Decision Tree

The greedy approach is used by decision tree classifiers. It is a supervised learning system that uses a tree to encode attributes and class labels. By learning decision rules derived from prior data (training data), decision trees are primarily used to create a training prototype that can be used to predict the class

or value of target variables. There are two different forms of decision trees: decision nodes and leaves. The outcomes or ultimate outcomes are represented by the leaves. Every node in the tree serves as a test case for a particular property, and every edge that descends from that node represents a potential response to the test case. For each sub-tree rooted at the new nodes, this recursive procedure is repeated. We have applied Decision tree approach in our model as:

(i) Importing library DecisionTreeClassifier from sklearn.tree Class
(ii) Now we create DecisionTree Classifier object
(iii) In the last we fit our data

B. Support Vector Machine (SVM)

The Support Vector Machine (SVM) is a supervised machine learning model or technique that may be applied to both regression and classification problems. But we mostly apply it to categorization problems. SVM is typically shown as training data points in space that are separated into groups by the smallest possible understandable gap. Each feature value in the SVM method is the value of a particular coordinate, and each data item is represented as a point in n-dimensional space. The categorization is then carried out by identifying the hyper-plane that effectively distinguishes the two classes.

C. Logistic Regression (LR)

In its most basic form, the logistic regression model is a widely used statistical model that models a binary dependent variable using a logistic function; there are other more intricate variations. As a type of binomial regression, logistic regression is used in regression analysis to predict a logistic model's parameters.

D. Gaussian Naïve Bayes (GNB)

A machine learning (ML) classification method called Gaussian Naive Bayes (GNB) is based on the Gaussian distribution and probabilistic methodology. According to Gaussian Naive Bayes, every parameter—also referred to as features or predictors—has the ability to independently predict the output variable. The final prediction that yields the likelihood that the dependent variable will be categorized in each group is the sum of the predictions for all parameters. The group with the highest likelihood is given the final classification.

RESULT AND ANALYSIS

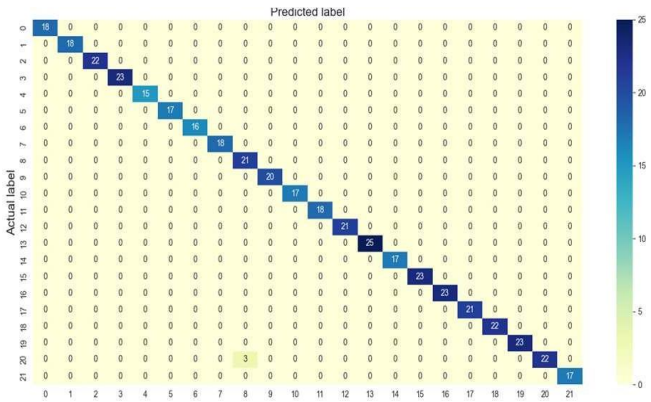

Fig 2 : Accuracy Comparison

For a given set of test data, the confusion matrix is used to assess how well the classification models perform. Only after the actual values for the test data are known can it be ascertained.
Although the matrix itself is simple to understand, some of the associated terms may be unclear.

Table 1 : Algorithm vice Accuracy Result in Percentage

Training Accuracy Score	99.5%
Validation Accuracy Score	99.3%

Based on the results provided, we can see that the model performs great with 99.3% accuracy.

Conclusion

We have created and implemented an intelligent system that can recommend appropriate crops to farmers all throughout India. Based on variables like nitrogen, phosphorus, potassium, PH value, humidity, temperature, and rainfall, this technology will assist farmers in selecting the optimal crop. We can use this research to boost national productivity and make money utilizing this method. By employing this method, the research can increase the nation's profitability and production. In this manner, farmers may cultivate the ideal crop and raise both their own and the nation's total income. After evaluating several machine learning methods, we found that Gaussian NB and Decision Tree had the highest accuracy.

Future Work

The following features could be added to the system to further improve it:

1. Improving datasets with more qualities is the primary goal of future study.
2. We must develop a model that can distinguish between crop leaves that are healthy and those that are diseased, as well as identify the type of disease present in the crop.
3. To create user-friendly mobile apps and websites.

References

1. Akshatha, K. R., and K. S. Shreedhara. "Implementation of machine learning algorithms for crop recommendation using precision agriculture." International Journal of Research in Engineering, Science and Management (IJRESM) 1, no. 6 (2018): 58-60.
2. Bhuvan, S., Alankriti Jain, and J. Sanjeetha. "DigiFarm– A Machine Learning based holistic Crop Prediction Platform for Farmers." NVEO-NATURAL VOLATILES & ESSENTIAL OILS Journal| NVEO (2021): 11112- 11129.
3. Bondre, Devdatta A., and Santosh Mahagaonkar. "Prediction of Crop Yield and Fertilizer Recommendation Using Machine Learning Algorithms." International Journal of Engineering Applied Sciences and Technology 4, no. 5 (2019): 371-376.
4. Computing Communication Control and Automation (ICCUBEA), pp. 1-6. IEEE, 2018.
5. Dighe, Deepti, Harshada Joshi, Aishwarya Katkar, Sneha Patil, and Shrikant Kokate. "Survey of Crop Recommendation Systems." (2018).
6. Doshi, Zeel, Subhash Nadkarni, Rashi Agrawal, and Neepa Shah. "AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms." In 2018 Fourth International Conference on

7. Gandge, Yogesh. "A study on various data mining techniques for crop yield prediction." In 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), pp. 420-423. IEEE, 2017.
8. Garanayak, Mamata, Goutam Sahu, Sachi Nandan Mohanty, and Alok Kumar Jagadev. "Agricultural Recommendation System for Crops Using Different Machine Learning Regression Methods." International Journal of Agricultural and Environmental Information Systems (IJAEIS) 12, no. 1 (2021): 1-20.
9. Ghadge, Rushika, Juilee Kulkarni, Pooja More, Sachee Nene, and R. L. Priya. "Prediction of crop yield using machine learning." Int. Res. J. Eng. Technol.(IRJET) 5 (2018).
10. Jain, Sonal, and Dharavath Ramesh. "Machine Learning convergence for weather based crop selection." In 2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), pp. 1-6. IEEE, 2020.
11. Jaiswal, Sapna, Tejaswi Kharade, Nikita Kotambe, and Shilpa Shinde. "Collaborative Recommendation System For Agriculture Sector." In ITM Web of Conferences, vol. 32. EDP Sciences, 2020.
12. Jejurkar Siddhi, S., S. Bhosale Meghna, and D. N. Wavhal. "Crop Predication and Diseases Detection Using Machine Learning." (2021).
13. Kamatchi, S. Bangaru, and R. Parvathi. "Improvement of Crop Production Using Recommender System by Weather Forecasts." Procedia Computer Science 165 (2019): 724-732.
14. Kulkarni, Nidhi H., G. N. Srinivasan, B. M. Sagar, and N. K. Cauvery. "Improving Crop Productivity Through A Crop Recommendation System Using Ensembling Technique." In 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS), pp. 114-119. IEEE, 2018.
15. Kumar, Avinash, Sobhangi Sarkar, and Chittaranjan Pradhan. "Recommendation system for crop identification and pest control technique in agriculture." In 2019 International Conference on Communication and Signal Processing (ICCSP), pp. 0185-0189. IEEE, 2019.
16. Kumar, Y. Jeevan Nagendra, V. Spandana, V. S. Vaishnavi, K. Neha, and V. G. R. R. Devi. "Supervised Machine learning Approach for Crop Yield Prediction in Agriculture Sector." In 2020 5th International Conference on Communication and Electronics Systems (ICCES),
17. Mahule, Ankit Arun, and A. J. Agrawal. "Hybrid Method for Improving Accuracy of Crop-Type Detection using Machine Learning." International Journal 9, no. 2 (2020).
18. Medar, Ramesh, Vijay S. Rajpurohit, and Shweta Shweta. "Crop yield prediction using machine learning techniques." In 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), pp. 1-5. IEEE, 2019.
19. Meshram, Vishal, et al. "Machine learning in agriculture domain: A state-of-art survey." Artificial Intelligence in the Life Sciences 1 (2021): 100010.
20. Nigam, Aruvansh, Saksham Garg, Archit Agrawal, and Parul Agrawal. "Crop yield prediction using machine learning algorithms." In 2019 Fifth International Conference on Image Information Processing (ICIIP), pp. 125-130. IEEE, 2019.
21. pp. 736-741. IEEE, 2020.
22. Pudumalar, S., E. Ramanujam, R. Harine Rajashree, C. Kavya, T. Kiruthika, and J. Nisha. "Crop recommendation system for precision agriculture." In 2016 Eighth International Conference on Advanced Computing (ICoAC), pp. 32-36. IEEE, 2017.

23. Qiao, Mengjia, et al. "Crop yield prediction from multi- spectral, multi-temporal remotely sensed imagery using recurrent 3D convolutional neural networks." International Journal of Applied Earth Observation and Geoinformation 102 (2021): 102436.
24. Rajak, Rohit Kumar, Ankit Pawar, Mitalee Pendke, Pooja Shinde, Suresh Rathod, and Avinash Devare. "Crop recommendation system to maximize crop yield using machine learning technique." International Research Journal of Engineering and Technology 4, no. 12 (2017):
25. Reddy, D. Anantha, Bhagyashri Dadore, and Aarti Watekar. "Crop recommendation system to maximize crop yield in ramtek region using machine learning." International Journal of Scientific Research in Science and Technology 6, no. 1 (2019): 485-489.
26. S. M. PANDE, P. K. RAMESH, A. ANMOL, B. R. AISHWARYA, K. ROHILLA and K. SHAURYA,
27. Suresh, A., P. Ganesh Kumar, and M. Ramalatha. "Prediction of major crop yields of Tamilnadu using K- means and Modified KNN." In 2018 3rd International Conference on Communication and Electronics Systems (ICCES), pp. 88-93. IEEE, 2018.
28. Suresh, G., A. Senthil Kumar, S. Lekashri, and R. Manikandan. "Efficient Crop Yield Recommendation System Using Machine Learning For Digital Farming." International Journal of Modern Agriculture 10, no. 1 (2021): 906-914.