International Journal of Research in Engineering, Science and Management
Volume 6, Issue 10, October 2023
https://www.ijresm.com | ISSN (Online): 2581-5792

Smart Crop – Precision Agriculture using Machine Learning

Ankit R. Sawant[1*], Yash Sivramkrishnan[2], Hemish Ganatra[3], Ameya A. Kadam[4]
[1,2,3]Student, Department of Electronics and Telecommunication Engineering, SVKM's Dwarkadas J. Sanghvi College of Engineering, Mumbai, India
[4]Assistant Professor, Department of Electronics and Telecommunication Engineering, SVKM's Dwarkadas J. Sanghvi College of Engineering, Mumbai, India

***Abstract*: The agriculture industry in India makes a significant contribution of around 17% to Gross Domestic Product (GDP) of the country and serves as the primary employer for over 60% of the country's labor force. The sector has undergone significant changes because of the introduction of innovative technology, like vertical farming, among other examples. However, it is noteworthy that a considerable proportion of Indian farmers persist in employing traditional methods and adhering to cultural beliefs when managing their agricultural land. To exemplify, farmers adhere to their established agricultural protocols by patiently expecting favorable weather conditions, instead of modifying their techniques in response to evolving weather patterns. The primary aim of this research is to help farmers in making reasonable decisions regarding crop selection, considering their unique circumstances and the prevailing environmental conditions. The achievement of this objective will be aid through the advancement of prediction models that consider many factors that affect the growth of crops, such as soil nutrients, soil pH levels, humidity levels, and rainfall patterns. A range of machine learning models are utilized, encompassing Decision Tree (DT), Support Vector Machine (SVM), Logistic Regression (LR), and Gaussian Naïve Bayes (GNB).**

***Keywords*: Crop-suggestion, Decision Tree (DT), Gaussian Naïve Bayes (GNB), Logistic Regression (LR), Support Vector Machine (SVM).**

1. Introduction

The agricultural sector has a vital role in both the Indian economy and the maintenance of human livelihood. This occupation carries substantial significance in the preservation of human existence. Moreover, it assumes a significant function in our everyday living [1]. In a significant number of cases, farmers exhibit suicidal inclinations due to reduced crop yields, leading to their inability to repay debts obtained from financial institutions for agricultural activities [2]. In recent years, there has been an observed pattern of frequent changes in the environment, which has had adverse impacts on agricultural yields and has further intensified the economic hardships faced by farmers. Consequently, this has led to occurrences of suicide [3]. The mitigation of these hazards can be accomplished through the application of mathematical or statistical methodologies to the available data. By implementing these methodologies, it becomes feasible to furnish farmers with suggestions for the most advantageous crop choice for their specific agricultural region, thereby allowing the achievement of maximum profitability [2].

In the present-day context of India, the agriculture industry has experienced significant advancements. The incorporation of "area-specific" horticulture is widely recognized as a significant benefit in the field of precision farming. Despite the introduction of several improvements, precision agriculture still faces some problems. The discipline of precision agriculture significantly influences the development of crop recommendations. The determination of crop recommendations is affected by a variety of factors.

Precision agriculture is an academic discipline that places emphasis on the detection and analysis of specific characteristics within a designated geographical area, with the purpose of detecting and effectively addressing challenges and issues. While it is acknowledged that not all outcomes offered by precision agriculture may be considered completely reliable, it remains crucial to acquire exact and accurate advice within the agricultural domain. The significance of these guidelines lies in the potential for significant losses in terms of resources and capital that may arise from any inconsistencies or disparities. There are now multiple ongoing research initiatives aimed at developing an exact and advanced model for crop prediction [4].

Machine Learning includes a multitude of algorithms, including unsupervised, supervised, and Reinforcement learning. Unique strengths and weaknesses are exhibited by these algorithms. Supervised learning involves the utilization of an algorithm to create a mathematical model by leveraging a dataset including input values and their corresponding anticipated output values. Unsupervised learning is a computational methodology in which a mathematical model is developed using a dataset comprising only input variables, without any associated desirable output labels. Semi-supervised learning is a technique that demands the use of algorithms to improve mathematical models by leveraging partial training data, where a certain amount of the input samples do not have corresponding labels [5].

The overriding purpose of this research is to offer the most appropriate crop selection by taking into account many input

*Corresponding author: ankitsawant26@gmail.com

characteristics, including Nitrogen (N), Phosphorous (P), Potassium (K), soil pH, humidity, temperature, and rainfall. The objective of this study is to predict the accuracy of future crop yields for a total of twenty-two distinct crops, encompassing rice, maize, chickpea, kidney beans, pigeon peas, moth beans, mung-bean, black gram, lentil, pomegranate, banana, mango, grapes, watermelon, muskmelon, apple, orange, papaya, coconut, cotton, jute, and coffee crops, within the geographic confines of India. In order to accomplish this objective, a variety of supervised machine learning approaches are utilized. Based on the empirical evidence presented, the paper posits the optimal crop for agricultural cultivation. The dataset comprises various parameters, namely Nitrogen (N), Phosphorous (P), Potassium (K), soil pH value, humidity, temperature, and rainfall.

2. Related Work

The proposed methodology seeks to estimate agricultural productivity by utilizing historical data, which incorporates many attributes such as humidity, pH levels, rainfall patterns, temperature, and the specific type of crop being cultivated [1]. The objective of this technique is to incorporate a diverse array of agricultural types spanning multiple areas in India. The utilization of the prospected method enables the prediction of the ideal crop choice by taking into account the current meteorological circumstances present in the field. The estimation of agricultural crop yields can be achieved by the utilization of the random forest algorithm in conjunction with decision tree approach. The utilization of the random forest methodology produced the most precise outcome. The implementation of more precise methods for evaluating crop production led to an increase in profitability.

The present study has developed a crop recommendation system that is specifically tailored for the implementation of smart farming practices [3]. This research study investigates many machine learning algorithms, such as CHAID, KNN, K-means, Decision Tree, Neural Network, Naïve Bayes, C4.5, LAD, IBK, and SVM algorithms. This study utilized the Hadoop architecture to allow computationally intricate calculations, hence augmenting the precision of the system.

The proposed methodology is employed to ascertain the identification of particular crops using the given dataset. The utilization of Support Vector Machine (SVM) yielded enhanced precision and productivity [9]. The main motive of this report was to do an analysis of two datasets. The first dataset consisted of location data, while the second dataset contained crop data. The proposed methodology employs nutrient values, including Nitrogen (N), Phosphorus (P), Potassium (K), and pH levels, to provide recommendations for suitable crops and ascertain the available nutrient values and necessary quantities of fertilizers for crops such as Rice, Maize, Black gram, Carrot, and Radish.

The system under consideration functions based on three fundamental components, namely soil characteristics, soil classifications, and the collection of crop productivity data [10]. The algorithm generated recommendations for farmers regarding the optimal crops for cultivation by employing these parameters. The proposed system employed a range of machine learning methodologies, such as Random Forest, CHAID, K-Nearest Neighbour, and Naïve Bayes. By employing the suggested methodology, it becomes feasible to predict the expansion of particular agricultural produce contingent upon distinct climatic circumstances, alongside the corresponding metrics linked to various regions and localities. Hence, the primary motive of this proposed study is to provide support to farmers in the process of seed selection based on soil characteristics, with the overarching purpose of improving agricultural production at a national level.

The current technology is utilized to determine the precise crop type by using a soil database [14]. The success of the proposed approach was proved in several crops, such as peanuts, legumes, cotton, vegetables, banana, paddy, sorghum, sugarcane, and coriander. In addition, the assessment also took into account several characteristics including depth, texture, pH, soil color, permeability, drainage, water retention, and erosion. The suggested methodology was executed by employing various machine learning classifiers, such as Support Vector Machine (SVM), Artificial Neural Network (ANN), random forest, and naïve Bayes. The classifiers were employed to determine an appropriate crop choice by considering parameters relevant to the site, so guaranteeing both precision and effectiveness. The main motive of this study endeavor is to offer support to farmers in improving agricultural output, addressing soil degradation in cultivated regions, lessening dependence on chemical inputs in crop cultivation, and optimizing the exploitation of water resources.

An intelligent system called Agro Consultant was developed by this research paper [15]. This research paper outlines the creation and advancement of a sophisticated intelligence system referred to as Agro Consultant. The proposed system can be classified into two discrete subsystems: i) a crop suitability predictor and ii) a rainfall predictor. The proposed system places emphasis on the development of five primary crops, namely bajra, jowar, maize, rice, and wheat, along with fifteen secondary crops, including barley, cotton, groundnut, gram, jute, other pulses, potato, ragi, tur, rapeseed and mustard, sesame, soybean, sugarcane, sunflower, and tobacco. Additionally, the model takes into account a range of factors, including soil type, aquifer thickness, soil pH, topsoil thickness, precipitation, temperature, and location parameters. The proposed system integrates many machine learning approaches, such as Decision Tree, K Nearest Neighbor (K-NN), Random Forest, and Neural Network. The aforementioned algorithms were employed for the purpose of performing multi-label classification. The accuracy rate of the proposed system was achieved at 71% through the application of a rainfall forecast model. Furthermore, the use of a neural network approach on the crop appropriate prediction system resulted in a notable increase in accuracy, reaching a rate of 91%.

The proposed methodology aims to offer precise suggestions for crop choice by taking into account distinct soil characteristics and factors, including average precipitation and surface temperature [21]. The constructed system incorporated many machine learning techniques, such as Random Forest, Naive Bayes, and Linear Support Vector Machines (SVM). The

crop recommendation system successfully classified the given soil dataset into the appropriate crop categories, specifically Kharif and Rabi. By employing the aforementioned methodology, an impressive level of precision up to 99.91% was achieved.

3. Proposed System

In this framework, this procedure is divided into numerous phases as shown in Fig. 1.

The five phases are as per the following:

1) Dataset Collection
2) Data Pre-processing (Noise Removal)
3) Feature Extraction
4) Applying Machine Learning Algorithm
5) Recommendation System
6) Recommend Crop

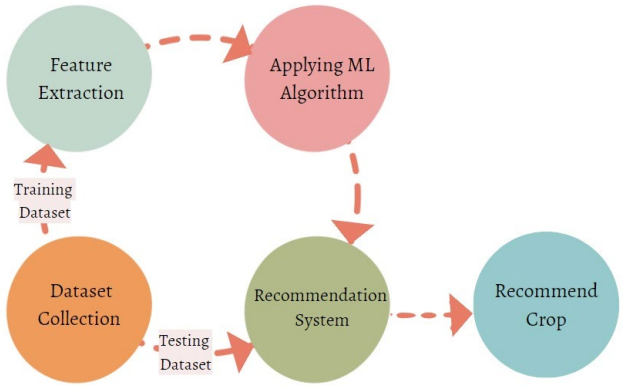

Fig. 1. Block diagram of overall methodology of proposed system

A. Flow of the Proposed System

As demonstrated in the Fig. 1, the methodology to extract the sentiment contains the several phases that are described below

1) Dataset Collection

The dataset [27] includes many characteristics, such as soil pH, humidity, temperature, rainfall, and the quantities of nitrogen (N), phosphorus (P), and potassium (K). The datasets have been made available on the Kaggle website. The dataset comprises a total of 2200 cases of data, which were derived from literal records. The dataset has a total of twenty-two distinct crops, including rice, maize, chickpeas, kidney beans, pigeonpeas, mothbeans, mungbeans, blackgram, lentil, pomegranates, bananas, mangoes, grapes, watermelon, muskmelon, apples, oranges, papayas, coconuts, cotton, jute, and coffee [27].

The dataset is split into Train set and Test set. Out of which 80% of the entire dataset is taken as Train set and remaining 20% as Test set.

2) Data Pre-Processing (Noise Removal)

Pre-processing is an essential need for the effective execution of the application. The data acquired from various sources is sometimes in its unprocessed state. The dataset may potentially contain conflicting, redundant, or missing information. Hence, it is imperative to eliminate superfluous data during this stage. It is recommended to normalize data [5].

3) Feature Extraction

This phase involves the identification and utilization of the dataset's relevant attribute. In this process, the elimination of duplicate and unnecessary data is carried out as a first step in order to prepare the data for the utilization of classifiers [5].

4. Methodology

The suggested system utilizes various Machine Learning techniques, including Decision Tree, Support Vector Machine (SVM), Logistic Regression (LR), and Gaussian Naïve Bayes (GNB).

A. Decision Tree

Decision tree classifiers utilize a greedy approach in their methodology. The solution being contemplated is a variant of supervised learning that use a tree structure to show both features and class labels [15]. The main purpose of utilizing this is to build a training model that facilitates the estimation of the class or target variable's values by utilizing decision rules drawn from historical data, also known as training data. It can be classified into two primary groups, decision nodes and leaves. The foliage symbolizes the resultant outcomes or ultimate conclusions. Within the framework of the tree structure, each individual node functions as a distinct test case for a certain property, with each subsequent edge stemming from the node symbolizing a potential answer to the aforementioned test case. The approach described above demonstrates a recursive nature and is executed recursively for every sub-tree that originates from the recently generated nodes [22].

The Decision tree was implemented in this model.

i. Importing the DecisionTree Classifier library from the sklearn.tree class
ii. Building a DecisionTree Classifier object.
iii. Finally, fit the data.

B. Support Vector Machine (SVM)

The Support Vector Machine (SVM) is a supervised machine learning methodology that may be utilized for both classification and regression applications. However, it is predominantly employed within the framework of categorization tasks. The Support Vector Machine (SVM) is commonly represented as a collection of training data points located in a geographic domain that is divided into separate groups by a noticeable gap, with the aim of maximizing the distance between them [22]. The Support Vector Machine (SVM) algorithm is employed to show individual data points as coordinates in an n-dimensional space, with each attribute value being associated with a particular position. The procedure of categorizing entails the identification of a hyperplane that efficiently differentiates between the two classifications [19].

C. Logistic Regression (LR)

The Logistic Regression model is a commonly used statistical model that employs a logistic function to express a binary dependent variable in its simplest form. Several other complex variants of this model have been developed. Within

the realm of Regression Analysis, it is a statistical methodology employed to ascertain the parameter estimates of a logistic model. This particular model can be classified as a form of Binomial regression [22].

D. Gaussian Naïve Bayes (GNB)

The Gaussian Naive Bayes (GNB) algorithm is a widely used classification approach. The methodology employed in this study is based on a probabilistic framework, which assumes that the features conform to a Gaussian distribution. The Gaussian Naive Bayes algorithm is based on the concept that each parameter, also known as predictors, exhibits independent predictive power for the output variable. The ultimate prediction is derived by amalgamating the forecasts for all parameters, leading to a probability estimation for the dependent variable to be categorized within each group. The highest categorization is assigned to the group that demonstrates a higher likelihood.

5. Results and Analysis

The performance of the classification model, for a set of test data is evaluated using a confusion matrix. Understanding the matrix itself is relatively straightforward. The accompanying terminologies can often be perplexing.

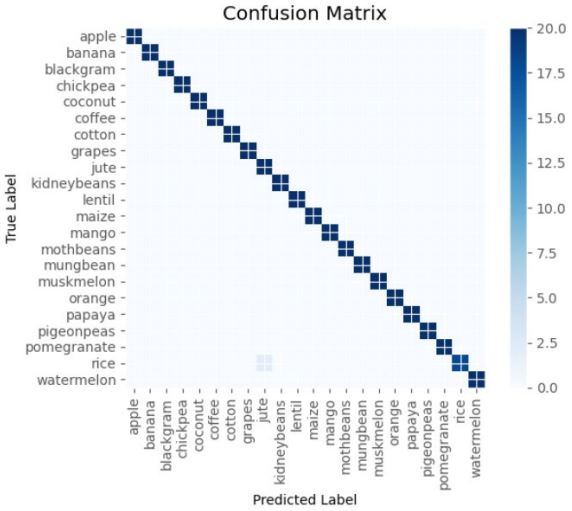

Fig. 2. Accuracy comparison

The model achieves a training accuracy score of 99.5% and a validation score of 99.3%. These outstanding outcomes demonstrate that the model has been trained well and can accurately classify data and make predictions. The small difference, between the training accuracy and validation score indicates that the model does not overfit and can effectively handle data. This high level of accuracy highlights the effectiveness of the training process.

6. Conclusion

A sophisticated system has been devised and executed to offer agricultural advice to farmers across India. This methodology would aid agricultural practitioners in determining the most preferred crop by considering many factors like Nitrogen, Phosphorous, Potassium, pH Value, Humidity, Temperature, and Rainfall. By employing this research, it is feasible to augment the efficiency of the country and generate economic benefits through the adoption of such an approach. The utilization of this methodology possesses the capacity to considerably enhance the efficiency and financial viability of the nation. In this way, agricultural practitioners possess the ability to develop suitable crop kinds, so enhancing their financial profits and making a positive contribution to the overall lucrative growth of the nation. The machine learning algorithms were evaluated, and it was concluded that Decision Tree and Gaussian NB had superior levels of accuracy in comparison to the other techniques.

References

[1] 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), IEEE, 2020, pp. 736-741.

[2] 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), pp. 1-20.

[3] Dighe, Deepti, Harshada Joshi, Aishwarya Katkar, Sneha Patil, and Shrikant Kokate. "Survey of Crop Recommendation Systems." 2018.

[4] 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), pp. 906-914.

[5] 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), IEEE, 2017, pp. 420-423.

[6] 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), IEEE, 2019, pp. 125-130.

[7] Medar, Ramesh, Vijay S. Rajpurohit, and Shweta, "Crop yield prediction using machine learning techniques." In 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), IEEE, 2019, pp. 1-5.

[8] 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), IEEE, 2020, pp. 1-6.

[9] 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), IEEE, 2018, pp. 88-93.

[10] Kamatchi, S. Bangaru, and R. Parvathi. "Improvement of Crop Production Using Recommender System by Weather Forecasts." Procedia Computer Science 165 (2019), pp. 724-732.

[11] 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), pp. 371-376.

[12] 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), pp. 485-489.

[13] 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), IEEE, 2017, pp. 32-36.

[14] Jejurkar Siddhi, S., S. Bhosale Meghna, and D. N. Wavhal. "Crop Predication and Diseases Detection Using Machine Learning." 2021.

[15] 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), pp. 950-953.

[16] Doshi, Zeel, Subhash Nadkarni, Rashi Agrawal, and Neepa Shah. "Agro Consultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms." In 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), IEEE, 2018, pp. 1-6

[17] Mahule, Ankit Arun, and A. J. Agrawal. "Hybrid Method for Improving Accuracy of Crop-Type Detection using Machine Learning." International Journal, vol. 9, no. 2, 2020.

[18] 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, vol. 1, no. 6, pp. 58-60, 2018.

[19] 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.

[20] Ghadge, Rushika, Juilee Kulkarni, Pooja More, Sachee Nene, and R. L. Priya. "Prediction of crop yield using machine learning." Int. Res. J. Eng. Technol., vol. 5, 2018.

[21] 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), IEEE, 2018, pp. 114-119.

[22] 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), IEEE, 2019, pp. 185-189.

[23] S. M. Pande, P. K. Ramesh, A. Anmol, B. R. Aishwarya, K. Rohilla and K. Shaurya, "Crop Recommender System Using Machine Learning Approach," 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2021, pp. 1066-1071.

[24] Bhuvan, S., Alankriti Jain, and J. Sanjeetha. "DigiFarm– A Machine Learning based holistic Crop Prediction Platform for Farmers." Natural Volatiles & Essential Oils Journal, pp. 11112- 11129, 2021.

[25] 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.

[26] Meshram, Vishal, et al. "Machine learning in agriculture domain: A state-of-art survey." Artificial Intelligence in the Life Sciences 1 (2021): 100010.

[27] Dataset - Link: https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset