Crop Recommendation and Prediction System

[1]CH. Rakesh D. [2]Vishnu Vardhan, [3]Babu Bhavani Vasantha, [4]G. Sai Krishna
Department of Computer science and Engineering,
Koneru Lakshmaiah EducationFoundation,
Vaddeswaram, Andhra Pradesh, I ndia.
[1]190032064@kluniversity.in, [2]190030364@kluniversity.in,
[3]vasanthabhavani@kluniversity.in,[4]190030581@kluniversity.in

2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS) | 979-8-3503-9737-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICACCS57279.2023.10113081

***Abstract*—Agriculture is a major driver of India's economic development and jobs. Worldwide, agricultural research has bolstered profit optimization and is a significant and large field to acquire more advantages. There are innumerable people who own land today, but they have no idea how to produce food. Everyone can only succeed in the future through agriculture. Most people practice unproductive agriculture because they grow crops on subpar soil. India's farmers frequently select crops that are not suitable for their soils, which reduces production. By providing the proper crops, precision agriculture has addressed this issue for farmers. Based on the state of the soil. In India, which is regarded as an agricultural nation, there are various ways to offer advice. The advice given today is based on communications amongst farmers and draws on knowledge from a number of sources. Precision farming is a cutting-edge farming method that offers the best crop to farmers based on site-specific criteria after collecting and analyzing data on soil properties, soil types, and crop production. By doing this, the likelihood of picking the incorrect crop is decreased, and productivity is raised. This structure is said to be able to forecast the ideal crop for the agronomist's region. Additionally, recommendations are offered for a variety of farming techniques and tactics, such as crop diversity, crop spacing, irrigation, sow processing, etc. as well as advice on pesticides and fertilizers. They are founded on previous soil indices. cost projections for crops and the weather. You will be able to determine the sorts of soil, the water source on that property, and what crop you should grow with this application. Additionally, you will be able to recommend which crops will thrive in that soil. Therefore, the application offers details about those who work in agriculture. For instance, we can identify the crop that will grow best given the soil, the climate, the temperature, etc. As a result, machine learning is used to determine the crop for each soil. Farmers can receive advice on what crop to plant and what fertilizer to use based on historical data from agricultural operations. Based on your previous agricultural actions, this tool can also be used to boost crop yields and suggest that you plant particular crops.**

Keywords—Agriculture, Machine Learning, Farming, Crop suggestion, k-Nearest Neighbor (KNN), Crop Recommendation.

I. INTRODUCTION

With more than 1.6 million square kilometers under cultivation, it is the second-largest nation in the world. India, one of the oldest nations in the world, continues to practice agriculture, albeit it has seen significant change in recent years due to globalization. The majority of Indians work in agriculture, making them reliant on the sector's economy. India is positioned to emerge as a global leader in agriculture. While only a small portion of the US population is reliant on farming, as much as 55% of the Indian population does. In India, where only around 45% of the land is irrigated, rainfed agriculture is the most popular form of food production.

Several causes have had a negative impact on Indian agriculture. Lack of attention to India's agricultural sector causes a decrease in the number of farmers living on farms and an increase in farmer suicide rates. In agricultural India, there is no such system of universal assistance for farmers. Farmers need to become more productive in order to work less and produce more from the same amount of land in order to make more money. Numerous innovative technologies have been created in attempt to revive agriculture. So, utilizing the extensive collection of agricultural data we have previously gathered, we can offer recommendations. through agriculture, contributes to eradicating poverty and fostering rural development. Professional farmers are now engaged in the production of commodities that are in high demand on the market. However, by blindly adhering to such an approach, they fail to take into account factors like the fact that their land is incompatible with such crops or that it lacks the resources the crop requires, which leads to a lower yield. In this case, a precision farm can be useful.

A lot of research is being done to create a crop forecasting model that is more precise and effective. This research area covers a range of machine learning methods, including assembling. These include several machine learning methods like enchantment. Precision farming, which focuses on applying precisely the right amount of soil, fertilizer, and other materials, has become increasingly popular in recent years as a result of globalization. The benefit of precision agriculture is that it has allowed us to increase input and output efficiency and improve agricultural decisions. The recommendation of crops is a key area of precision agriculture. This depends on a variety of variables. Different crops, fertilizers, and farming methods can be suggested by systems. In order to address problems with crop selection, precision agriculture aims to determine these factors in a site-specific manner.

Multiple factors have an impact on Indian agriculture. To restore health, many cutting-edge technologies are being created. There are many different kinds of precision agriculture, including ones that boost productivity and yields by giving the crop just the correct amount of pressure at the right time. Not all precision farming techniques work. A strategy used in precision farming is site specific.

II. LITERATURE SURVEY

In this literature review of this project, the team sought out and studied various patents, research papers, documents, and newspapers and magazine articles from various scenes.

Unfit crop selection is a significant problem for Indian farmers. India's main industry for generating income and jobs is agriculture. They will have to drastically cut back on their production as a result. Precision agriculture has helped farmers by enabling them to overcome their challenges. It is a method of agricultural production that uses research on soil properties, soil types, and crop production statistics to advise farmers on the best crop for their particular site. As a result, productivity rises and fewer crops are chosen improperly. [1]

In order for farmers, government organizations, and academics to make educated judgments about crop storage, selling, setting minimum support prices, and importing and exporting, Yogesh Gandge and Sandha [2] created a system of Crop yield projections. This prediction method is appropriate for data mining since crop prediction necessitates a thorough examination of numerous variables, including soil quality and pH, EC, N, P, and K. Since a lot of data is required, this prediction method is best used in data mining. Large amounts of data can be mined for knowledge using a data mining technique. This research explores various data mining techniques for estimating agricultural yields. Any crop production prediction system's performance depends on how accurately traits are retrieved and how well the classifier performs. This paper presents an overview of agricultural output forecast algorithms, their precision, and recommendations. [8]

The agricultural decision support systems that Ji-chun Zhao and Jian xin Guo [3] developed can provide both scientific underpinnings for agricultural research and direction for farm output. Implementing big data analytics will improve intelligent decision support systems. The creation of intelligent decision-making systems for agriculture is being studied by researchers and developers. The category for agricultural decision systems is introduced for the first time after examining the frame designation and design process for intelligent decision systems.

For India, agricultural production is crucial. The nation as a whole and farmers both flourish. Our work has helped farmers cultivate the right seed based on soil characteristics, increasing the output of the nation. We aim to further enhance the data collection by include more attributes and yield estimates. [4]. Precision farming is outlined in depth, along with the prerequisites and planning necessary for its development, by S. Babu et al. [5]. The paper gives a general review of precision farming's fundamentals. The author creates a framework by starting with the fundamentals of precision farming and working toward creating a model that can support it. This study demonstrates the use of Precision Agriculture (PA) concepts to enhance the rate of variability management over small, open farms.[6] The strategy intends to reach even the smallest farmer at the level of his or her smallest plot of the crop in addition to offering direct advising services via SMS and email. Because it was developed for Kerala State, where the average holding size is far smaller than that of the rest of India, this model can therefore be applied everywhere in India with just slight modifications.

III. Problem Formulation

Few platforms exist that aid farmers in developing their farming strategies. Decisions based on intuition might not turn out to be advantageous in the long run. Farmers frequently overestimate or underestimate the soil fertility in their fields. They frequently have trouble identifying plant illnesses that have an immediate impact on the rate of output. It is feasible to provide precise crop forecast results by using the right criteria, such as rain patterns, temperature patterns,[7] soil structures, and other things like crop diseases. Furthermore, it is also feasible to determine in advance what disease a crop has. Many of the current systems are exceedingly difficult to use or have numerous problems that make them unintuitive.

IV. Data Processing And Analysis

Preprocessing and Analysis of Data Pre-processing data is essential for converting a sizable dataset into a format that can be used. This includes removing Nan values (Missing Values) and incorrect noisy data from the dataset in order to narrow down the model's training to the proper level. We eliminated records with Nan values to make our dataset suitable for training the ensemble learning model. [9]. For Crop Recommendation Dataset is shown in Fig. 1. For Crop Yield Prediction Dataset is shown in Fig. 2. Distribution of Agriculture is shown in Fig. 3.

A. For Crop Recommendation Dataset

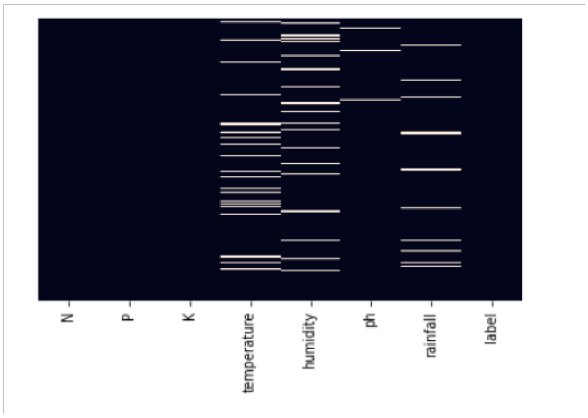

Fig. 1(a). Before

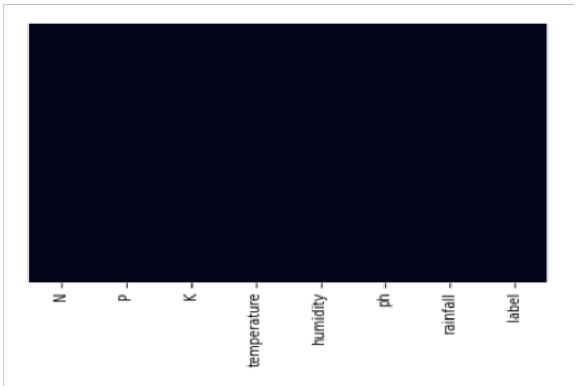

Fig. 1(b). After

Fig. 1. (a) before & (b) after, For Crop Recommendation Dataset

B. For Crop Yield Prediction Dataset

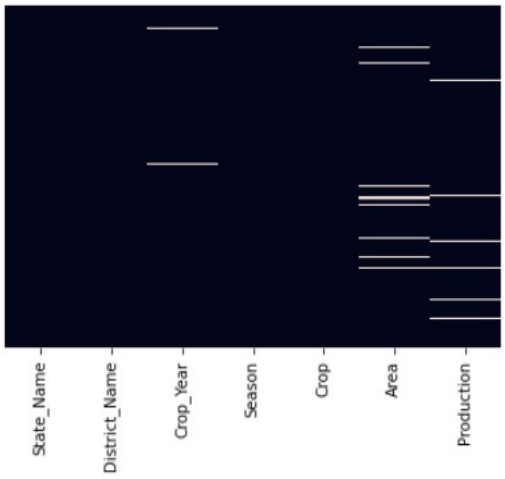

Fig. 2(a). Before

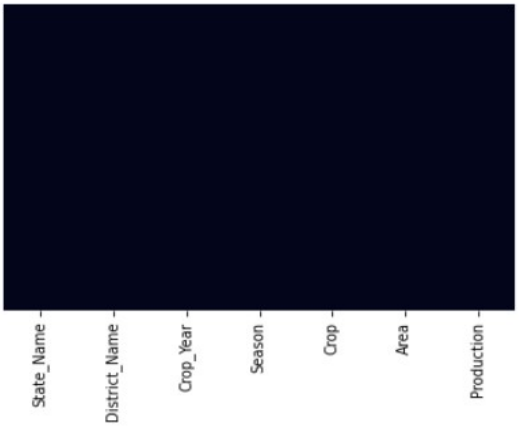

Fig. 2(b). After

Fig.21 (a) before & (b) after, For Crop Yield Prediction Dataset

C. Distribution of Agriculture

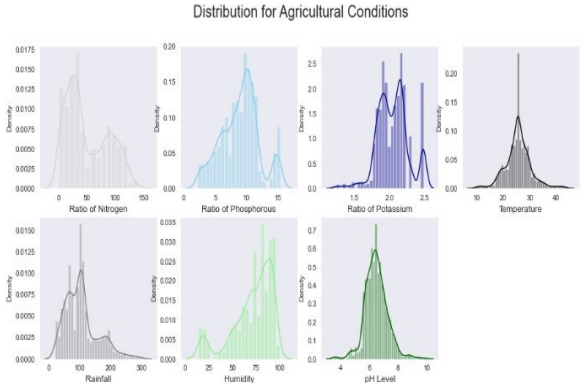

Fig. 3. Distribution of Agriculture

V. Methodology

A. Linear Regression

A linear regression model is, generally speaking, a statistical analysis technique that uses mathematical regression analysis to ascertain the quantitative relationship between two or more variables. A least-squares function known as a linear regression equation is used to depict the relationship between one or more independent variables and dependent variables. This function is a linear combination of one or more regression coefficients, which are model parameters. Simple regression is used when there is only one independent variable;[10] multiple regression is used when there are numerous independent variables. There are many uses for linear regression. By fitting the mapping between data input and output, the linear regression model predicts the proportionate connection between a dependent variable and a predictor. The traditional linear regression equation is as follows: $y = \alpha + \beta 1x1 + \beta 2x2 + ... + \beta nxn + \varepsilon$.

The dependent variables are represented by the y variable, the independent variables are represented by the beta, and the calculation error term is represented by the if. The goal of the linear regression model is to reduce the total squared errors while estimating the parameters.

B. Decision Tree

A well-known machine learning technique called the decision tree operates under the premise that the same (or similar) input leads to the same (or similar) output. By evaluating the opinions of the samples' various features and placing them in the next leaf node, decisions based on tree outcomes aim to categories or regress samples with the same characteristics. Data can be categorized using a decision tree and a set of criteria. In order to decide which values will be given out under specific conditions, it employs a rule-based technique. There are classification trees for discrete data and regression trees for continuous variables.

C. KNN

K-Nearest Neighbor (KNN) is a machine learning technique for regression and classifying predictive analysis. It is also known as a lazy learner algorithm since it categories fresh data it is tested on based on similarities rather than studying the data it was trained on. Utilizing feature similarity, KNN based on how closely incoming data corresponds to the points in the training set, in order to predict house price values.

$$\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \qquad \sum_{i=1}^{k}|x_i - y_i| \tag{1}$$

where X stands for the new argument, Y for the current point, and K for the K-Factor (number of clusters evaluated by the algorithm before allocating a value). It is a simple, adaptable, and simpleto-implement supervised learning algorithm. It is predicated on the notion that forging a bond requires close closeness. It determines the separation between two points on a graph in order to communicate the idea of similarity. The algorithm's numerical parameter 'k' indicates how many data points should be considered while casting a vote. In order to categories a new point, we encircle it with K data points and give it to the group that has the greatest number of points inside the circle. The ideal method for determining the value of K is to experiment with a range of values before choosing one that reduces error while retaining prediction accuracy.

D. Random Forest

The decision tree ensemble method known as random forests is challenging to master and is prone to overfitting. It uses a technique called "tree bagging." The issue of decision tree instability and excessive variation can be solved by creating the decision tree. These decision trees are known as random forests because they were created using random sampling techniques. Weak learners are trained independently using a subset of explanatory factors that is randomly chosen. By simultaneously training weak prediction models, like decision trees, it builds a prediction model the forecasts from each tree are integrated using a model averaging technique. The random forests approach creates a decision tree on its own after randomly choosing a set of traits. Random forests' decision trees are separate from one another. These trees can be voted on using the majority rule. Ensemble model is shown in Fig. 4.

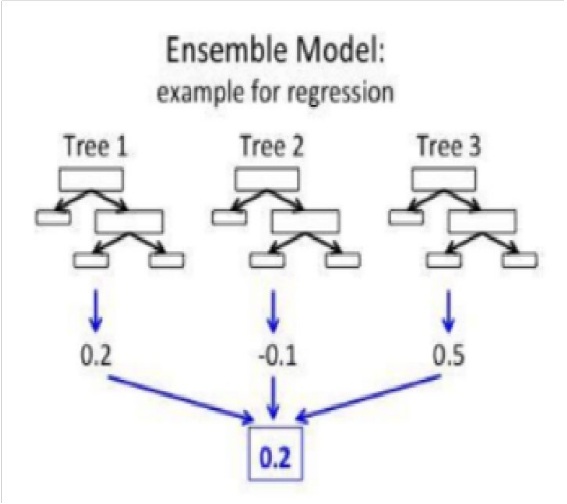

Fig. 4. Ensemble model

E. Support Vector Machine

A powerful supervised learning tool is the Support Vector Machine (SVM). The Support Vector Machine (SVM) method converts the original data into a high-dimensional space in order to discover a hyperplane for data segmentation.

The hyperplane is defined by support vectors, often known as "essential training tuples". Due to its acceptance of nonlinear boundaries, SVM is more accurate than other methods.

To implement SVM models, two sets of input variables are used. The first is based on the five essential elements of systematic subset selection. There are six PCA transformation components in the second input set.

F. Naive Bayes

The Bayes theorem is used to analyze the data, and the naive assumption of conditional independence between every pair of attributes is made given the value of the class variable.

G. Root Mean Square Error

The Root Mean Square Error will be used to test the model created in this study (RMSE). The prediction error of each data is taken into account when calculating predicted performance using the RMSE.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(d_i - p_i)^2} \tag{2}$$

H. Mean Square Error

The Mean Squared Error (MSE), which is regularly taught in Machine Learning courses, is the most fundamental and widely applied loss function. Divide the discrepancy between your model's forecasts and the actual data, square it, and average it over the entire dataset to obtain the MSE. The MSE will never be bad since we are always fixing our mistakes.Crop Recommendation Data is shown in table 1. Crop Yield prediction dataset is shown in table 2.

The following is a mathematical definition of the MSE:

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{3}$$

I. Mean Absolute Error

The Mean Absolute Error (MAE) shares almost all of the characteristics of Mean Standard Error (MSE), although conceptually being different from it. By dividing the discrepancy between your model's predictions and the actual data by the absolute value, you may determine the MAE by averaging the result across the entire dataset. For Crop Recommendation Dataset is shown in Fig. 6.

$$\mathrm{MAE} = \frac{1}{n}\sum_{j=1}^{n}|y_j - \hat{y}_j| \tag{4}$$

The following Fig. 5 is shows the proposed method.

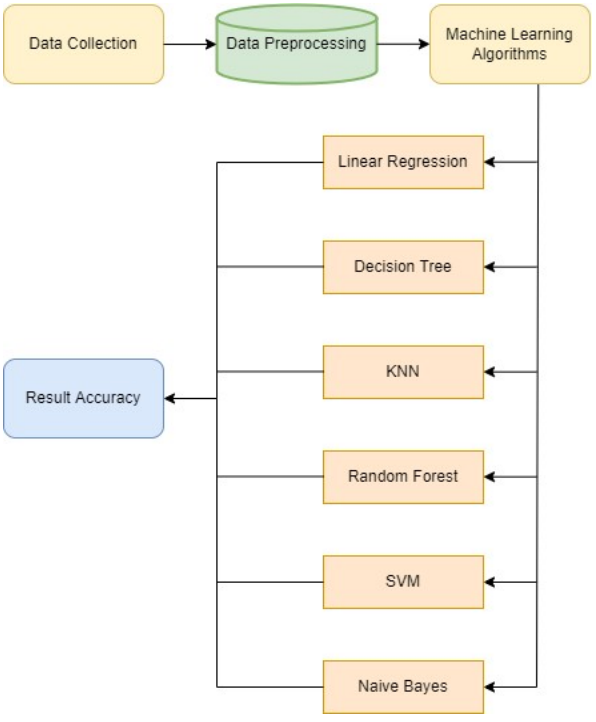

Fig. 5. proposed method

VI. Results And Experimentation

A. For Crop Recommendation Dataset

TABLE I

Algorithm	Prediction Accuracy
Decision Tree	95.45
KNN	95.00
Naive Bayes	97.28
SVM	93.18

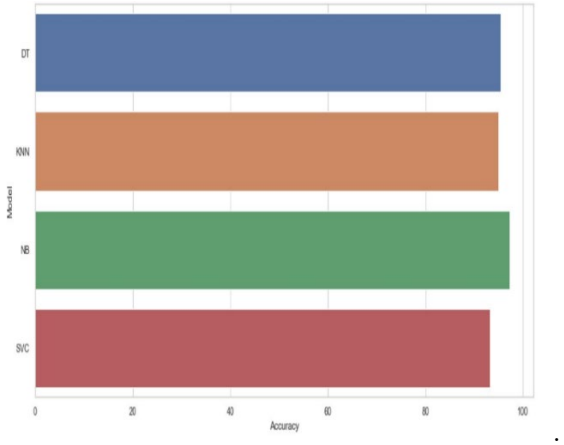

Fig.6 For Crop Recommendation Dataset

B. Crop Yield Prediction Dataset

TABLE II

Algorithm	MSE	MAE	RMS E	R2
Linear Regression	0.107	0.199	0.328	0.89
Random Forest	0.026	0.083	0.163	0.97
Decision Tree	0.0517	0.105	0.227	0.94

VII. Conclusion

Based on the findings of experiments using various models. When compared to other methods, using machine learning techniques like naive Bayes, decision trees, and random forests greatly increased the model's accuracy.

On two datasets in this study, we applied machine learning algorithms and conducted analysis. The first dataset contains information about the crop's classification. The Naive Bayes method outperformed the other three of the four machine learning algorithms used. The second set of information consists of information on yield projection. The Random Forest algorithm outperformed the other two machine learning algorithms. These algorithms can be utilized to build a web application where we can employ these models for in-depth analysis of real-world data.

References

[1] Pudumalar, S. & Elangovan, Ramanujam & Rajashree, R. & Kavya, C. & Kiruthika, T. & Nisha, J. (2017). Crop recommendation system for precision agriculture. 32-36. 10.1109/ICoAC.2017.7951740.

[2] Gandge, Yogesh & Sandhya. (2017). A study on various data mining techniques for crop yield prediction. 420-423.10.1109/ICEECCOT.2017.8284541.

[3] Zhao, Ji-chun & Guo, Jian. (2018). Big data analysis technology application in agricultural intelligence decision system.209-212.10.1109/ICCCBDA.2018.8386513.

[4] Dasari, Anantha Reddy. (2019). Crop recommendation system to maximize crop yield in ramtek region using machine learning. International Journal of Scientific Research in Science and Technology. 6. 485-489.

[5] Babu, Satish. "A software model for precision agriculture for small and marginal farmers." 2013 IEEE Global Humanitarian Technology Conference: South Asia Satellite (GHTC-SAS) (2013): 352- 35.

[6] Simon Tavasoli (202), '10 Machine Learning Algorithms You Need to Know', Simplilearn Solutions

[7] Sunil Ray (2015), '6 Easy Steps to Learn Naive Bayes Algorithm with codes in Python and R', Analytics Vidhya.

[8] Rohith Gandhi (2018), 'Support Vector Machine — Introduction to Machine Learning Algorithms', Toward Data Science Web Site.

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

[10] 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.