



Crop Recommendation Using Machine Learning

Akshita Waldia¹, Pragati Garg², Priyanka Garg³, Rachna Tewani⁴, A. Kumar Dubey⁵, Anurag Agrawal⁶

^{1,2,3,5,6} Bharati Vidyapeeth's College of Engineering, New Delhi, India

⁴Data Scientist, Great Learning, India

Emails: akshitawaldia7@gmail.com; gargpragati9@gmail.com; priyanka.garg1803@gmail.com
; rachnatewani09@gmail.com; arudubey@gmail.com; ind.anurag@gmail.com

Abstract

The population of India is over one billion. Nearly 65 percent of the population of India lives in villages with the main occupation being agriculture. The diverse climatic conditions in the country result in the production of a large number of agricultural items. Many surveys have proved that the suicide rate of farmers is proliferating over years due to the selection of the wrong crop resulting in less yield. In some areas, farmers lack information about the composition of soil and weather conditions and may choose the wrong crop to sow which results in lesser yield. Production of crops depends on geographical parameters like humidity, rainfall, and properties of soil such as pH, and NPK content. Integration of technology with agriculture helps the farmer to improve his production. The main goal of agricultural planning is to achieve the maximum yield rate of crops by using a limited number of land resources. This paper mainly focuses on recommending the appropriate crop using ML Algorithms (Decision Tree, Naive Bayes, Random Forest) based on soil composition and weather conditions to maximize the yield of the farm and increase the economic condition of India's farmers.

Keywords: Machine Learning; Crop prediction; Decision tree; Naive Bayes; Random Forest; crop recommendation

1. Introduction

Agriculture is the main source of livelihood and the broadest economic sector in India. It plays an irreplaceable role in the overall development of the country. Therefore, there is a need to constantly adopt new, improving agriculture technologies to maximize the profit of the farmers of our country.

- Farmers generally select the crop to be planted based on their own prior experience. However, different factors such as soil, climate, weather, seed variety, and use of fertilizer determine the yield of crops and these are not taken into account by farmers.
- Due to this lack of knowledge, they do not carry out proper crop rotation and end up adding inadequate amounts of nutrients to the soil. This reduces crop yield and damages the soil further.
- Machine learning can be used to improve the agriculture sector by helping with crop recommendations. Using machine learning, we can determine correlations and patterns in the data and discover knowledge from datasets.

The dataset used for this purpose includes various features like Rainfall, temperature, humidity, and pH. The model is trained using this historical data. The part of the historical data not used for the training phase is used for testing. If the soil contents and weather conditions of a particular land are available, the model can predict the most suitable crop for that area. In this paper, we are recommending the crop using Decision Tree, Random Forest, and Naive Bayes machine learning algorithms.

2. Literature review

[1] analyses 50 papers that passed inclusion criteria for algorithms and features used. Most of the papers used random forest, neural networks, linear regression, and gradient boosting trees as their algorithms. But the most commonly applied algorithm was found to be the Artificial Neural Network. Talking about features, temperature, rainfall, and soil type were most commonly used. In [2], the system is fed with the soil characteristics, weather conditions, and history of the yield of crops in specific regions in Maharashtra state. The system then recommends the most suitable crop to be sown in that region using Random Forest Algorithm. K-Means clustering algorithm is also applied to find a suitable fertilizer for the region. [3] uses data of soil characteristics, soil types, and crop yield data collection to suggest to the farmers the appropriate crop based on the agricultural field. This helps in reducing the wrong selection of a crop which in turn leads to increased productivity. The system uses K-Nearest Neighbor, CHAID, Random tree, and Naive Bayes algorithms. The prediction accuracy is 88% based on soil requirements. [4] uses Naive Bayes Gaussian Classifier to predict the seed type taking into consideration the temperature, humidity, pH, and rainfall. Along with it, boosting algorithms are used for higher accuracy. [5] performs prior analysis of the data related to weather conditions and climate to map a correlation between them. This alerts the farmer of the potential risks to the crop due to bad weather and the quality of the crop and acts as an early warning so that preventions can be taken. [6] analyses humidity, yield, temperature, and rainfall, and the model is developed using random forest, multivariate polynomial regression, and support vector machine regression techniques to predict crop yield per acre. The MAE, RMSE, median absolute error and maximum R-squared values are calculated as 3.57, 5.48, 1.58, and 0.968 respectively. [7] analyses different types of supervised, unsupervised, and reinforcement machine learning algorithms in order to discuss their usage in the field of agriculture. In [8], the focus is on using district-wise weather prediction in Tamil Nadu to create a recommendation model. Thorough data preprocessing is applied to improve the quality of input. On this data, predictive modeling with Artificial Neural Networks is done to predict minimum and maximum temperature and relative humidity, and the best form of crop production is determined. A hybrid recommender system with Case-Based Reasoning and collaborative filtering technique is proposed which also takes into consideration the agriculture pattern of the district. The accuracy of the hybrid model is 96%. In most research papers, different models were used in order to get the best prediction. According to the result, some machine learning models are used more frequently although no model can be conclusively shown to be the most effective. Random forest, naive Bayes, neural networks, linear regression, and gradient boosting trees are most common. Overall, the most used features for the prediction of crops are temperature, rainfall, and soil type.

3. Machine Learning for a recommendation of crop

The proposed system is divided into five phases as shown in figure 1:

- I. Preparation of Dataset for the training of the model.
- II. Separating complete dataset into two parts: input parameters (features) & crop name (target).
- III. Splitting data into train, testing, and applying different algorithms.
- IV. Results analysis technique to evaluate the performance of the proposed approach on the test set.
- V. Comparing different Algorithms and their accuracy.

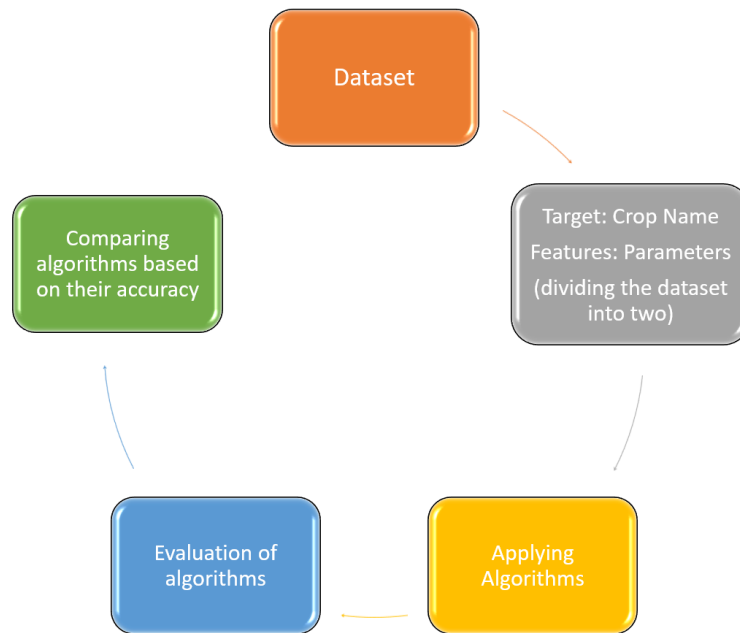


Figure 1: Proposed system

The database used for the training and testing of the classification model is publically available. The overall size of the dataset is 18600 with 3100 rows and 6 columns.

The database contains 5 different parameters for crop classification:

soil type, temperature, humidity, pH, rainfall.

And contains 31 different crops.

We have used three algorithms for the classification of crops based on the input parameters:

- I. Decision Tree
- II. Naive Bayes
- III. Random Forest

DECISION TREE

A Decision Tree is used to create a training model that can predict the value or class of the target variable (crop name) by simple decision rules inferred from prior data (training data).

Based on a comparison between the root attribute with the record's attribute, we follow the branch corresponding to that value and jump to the next node. For solving this attribute selection problem, We have used Entropy as the criteria. Entropy is the degree of randomness in the information being processed. The objective is to reduce Entropy at every subsequent step to get minimum Entropy at the leaf node and thereby classifying data for every crop.

NAÏVE BAYES

The naive Bayes Algorithm is a machine learning classification technique that is used for predictive modeling and is based on the Naive Bayes classifier. The naive Bayes classifier makes use of the assumption that the presence or absence of any feature of a class is not related to the presence or absence of any other feature. Naive Bayes classifiers are a collection of classification algorithms based on Bayes' Theorem. The theorem finds the probability of an event occurring given the probability of another event that has already occurred.

It is given by the formula:

$$P(C|X) = \frac{P(X|C) * P(C)}{P(X)}$$

Bayes theorem allows calculating posterior probability $P(C|X)$ from the given $P(X|C)$, $P(X)$, and $P(C)$ where,

$P(C|X)$ = the probability of event C occurring given that X is true.

It is also called the posterior probability of X given C.

$P(X|C)$ = the probability of event X occurring given that C is true.

It is also called the posterior probability of C given X.

$P(C)$ = prior probability of C.

$P(X)$ = probability of X

RANDOM FOREST

Random forest is an ensemble method in which the "forest" is built from a set of decision trees. The trees are used together to get a more accurate and stable prediction. Each individual decision tree is fitted from a randomly selected subset of the training set. The trees individually predict a class and the majority vote is used to decide the final class of the test object. In order to have a forest whose collective prediction is more accurate than that of individual trees, it is important to have trees with low correlation. Random forest uses Bagging and Features Randomness to create a forest of uncorrelated trees.

4. Proposed Method

The outline of the proposed method is shown in figure 2. First of all, we extract the crop name from the complete database and assign it to the 'target' vector. The remaining parameters viz. soil type, temperature, humidity, pH, and rainfall are the inputs to the model. Now, we have two subsets: input parameters (soil type, temperature, humidity, pH, rainfall) and target (crop name). We then preprocess the data to prepare a structured dataset that the machine can understand. The output to be determined is the name of the suitable crop. Further, we split the dataset into 2 parts: the training dataset, and the test dataset. We consider about 80% of the data as train data and 20% of the data as test data. Using train data, we train the model by using the above-mentioned algorithms. To check the efficacy of our model, we use test data, which are our predicted values. By comparing the predicted values and the actual name of the crop present in the test data, we determine the accuracy of our model.

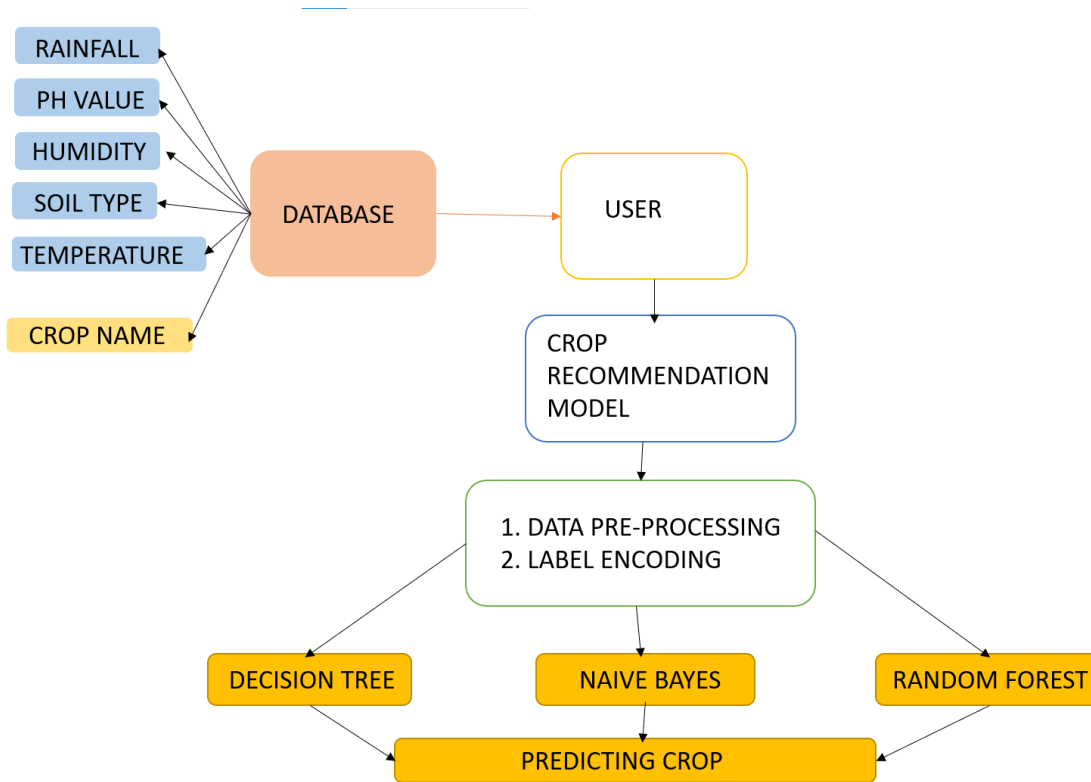


Figure 2: Working of model

5. Result

In this paper, we have proposed a system that uses machine learning algorithms to recommend the most suitable crop. We have used three algorithms- Naive Bayes, Decision Trees, and Random Forest. The accuracy achieved by each algorithm is given in Table 1.1. Using this analysis, the crop to be sown will be chosen based on the soil quality and weather conditions. This efficient information will lessen the farmer's difficulties as he will be able to achieve a high yield and maximize profit. This will lead to an increase in the economy of the agriculture sector. A bar chart comparing the accuracy of the three algorithms is shown in figure 3.

Table 1: Accuracy comparison of different papers along with their features and algorithms used

<u>Paper</u>	<u>Algorithms used</u>	<u>Parameters used</u>	<u>Accuracy</u>
	Random Forest, Naive Bayes, Decision Tree	Temperature, Rainfall, Humidity, pH, Soil Type	Decision Tree: 67.74% Random Forest: 96.61% Naive Bayes: 94.83%
Crop Recommendation System for Precision Agriculture [3]	K-Nearest Neighbor, CHAID, Random tree, and Naive Bayes	Soil types, soil characteristics, crop yield	88%

Smart Farming System: Crop Yield Prediction Using Regression Techniques [6]	Support vector machine regression	humidity, yield, temperature, and rainfall	R-squared value: 0.968
Improvement of Crop Production Using Recommender System by Weather Forecasts [8]	Weather prediction: temperature, rainfall, solar radiation, wind speed, evaporation, relative humidity, and evapotranspiration and district-wise agricultural data	Artificial Neural Networks	96%

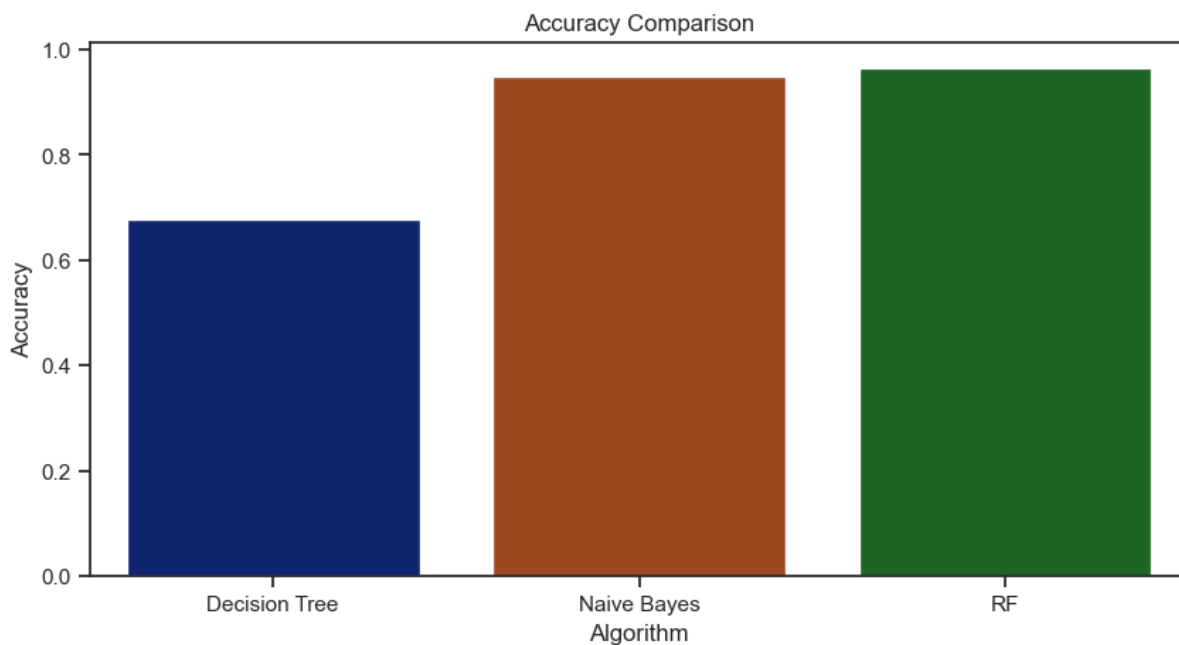


Figure 3: Accuracy comparison of algorithms

6. Conclusion

Agriculture is crucial for the economic growth of our country. But this field lacks behind in using new technologies of machine learning. Before selecting any plant to grow, it is important to have the knowledge and an understanding of the factors that affect the cultivation and how to maintain or control them. The proposed work introduces an efficient crop recommendation system using classifier models namely Decision Tree, Random Forest, and Naive Bayes. In the proposed model, the crop is recommended based on soil type, temperature, humidity, pH, and rainfall.

Successfully integrating machine learning with agriculture can lead to advancements in agriculture by maximizing yield and optimizing the use of resources that are involved. However, these recommendations must be precise and accurate. Future work can be extended to add features to the system. Currently, it takes necessary environmental factors and soil properties as inputs and suggests a very suitable crop to be cultivated. But at the next level, proper fertilizers based on the given properties of soil can be determined to help the farmer choose the accurate quantity and type of fertilizer for the crop, which will prevent excessive use of fertilizers thereby enriching the soil quality.

References

- [1] Van Klompenburg, Thomas, Ayalew Kassahun, and Cagatay Catal. "Crop yield prediction using machine learning: A systematic literature review." *Computers and Electronics in Agriculture* 177 (2020): 105709.
- [2] Chougule, Archana, Vijay Kumar Jha, and Debajyoti Mukhopadhyay. "Crop suitability and fertilizers recommendation using data mining techniques." In *Progress in Advanced Computing and Intelligent Engineering*, pp. 205-213. Springer, Singapore, 2019.
- [3] 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.
- [4] Kalimuthu, M., P. Vaishnavi, and M. Kishore. "Crop prediction using machine learning." In *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, pp. 926-932. IEEE, 2020.
- [5] Osman, Tousif, Shahreen Shahjahan Psyche, MD Rafik Kamal, Fouzia Tamanna, Farzana Haque, and Rashedur M. Rahman. "Predicting Early Crop Production by Analysing Prior Environment Factors." In *International Conference on Advances in Information and Communication Technology*, pp. 470-479. Springer, Cham, 2016.
- [6] Shah, Ayush, Akash Dubey, Vishesh Hemnani, Divye Gala, and D. R. Kalbande. "Smart farming system: Crop yield prediction using regression techniques." In *Proceedings of International Conference on Wireless Communication*, pp. 49-56. Springer, Singapore, 2018.
- [7] A. Sharma, A. Jain, P. Gupta and V. Chowdary, "Machine Learning Applications for Precision Agriculture: A Comprehensive Review," in *IEEE Access*, vol. 9, pp. 4843-4873, 2021, doi: 10.1109/ACCESS.2020.3048415.
- [8] Kamatchi, S. Bangaru, and R. Parvathi. "Improvement of Crop Production Using Recommender System by Weather Forecasts." *Procedia Computer Science* 165 (2019): 724-732.