**Crop Recommendation System**

Ankit kumar (PF24), Swayam Sharma (PF30), Parth Mogadpally(PF43)

Machine Learning Research Paper – MIT World Peace University

Abstract

The majority of Indians see agriculture as their main line of work. The economy of our country depends heavily on crop cultivation. Too little or too much fertiliser is frequently the cause of poor crop quality. Crop cultivation used to be done by farmers with practical experience. Based on the characteristics and features of the soil, farmers are no longer able to select the best crop. The crop that can be grown in that specific soil type has been suggested by a recommendation system that uses machine learning algorithms. This method helps maintain crop health while reducing the likelihood of soil deterioration. This system examines the soil's pH, temperature, humidity, and NPK nutrients. These sensors include ones for pH, soil moisture, and soil temperature., plus others. This model uses a variety of algorithms, including Random Forest, Naive Bayes, KNN, etc., to suggest the best crop to grow. The most effective algorithm is chosen after comparing the accuracy of various models.

***Keywords*** — Machine Learning, Crop Recommendation, KNN, Decision Tree, Naive Bayes, Random Forest, Agriculture, Machine Learning, Farming

**1.Introduction**

We can all agree that agriculture plays a significant role in the Indian economy.

Agriculture is a crucial industry in India. More than 60% of the country's land is used for agriculture, which provides food for 1.3 billion people.

Agriculture is the cultivation of plants and animals. Agriculture aided in the development of India's civilization. We require soil in order to grow crops. A crucial component of agriculture is soil. Food production cannot be successful without a healthy soil. It provides the roots with support, water, oxygen, and essential nutrients. All food-producing plants are raised on soil, which also serves as

the core of the system. In India, there are various types of soil. Red soil (corn, ragi) and black soil (sugarcane, sunflower) are a few examples, as are alluvial soil (cotton, rice) and laterite soil (pulses, tea, coffee), among others.

**2.Background and Motivation**

For both human survival and the Indian economy, agriculture is a crucial sector. It is one of the primary occupations required to preserve human life. It significantly improves the quality of our daily lives as well. Most frequently, farmers commit suicide due to a drop in productivity as a result of their inability to pay back the bank loans they took out for their businesses.

**3.Challenges and opportunities in precision agriculture**

We have observed that the environment is steadily changing at the moment, which is detrimental to the crops and causes farmers to go bankrupt or commit suicide. These risks can be minimised by applying various mathematical or statistical techniques to the data. Using these methods, we can advise the farmer on the best crop to plant in his agricultural region to help him earn the most money. Today's agriculture in India has advanced significantly. "Site-specific" farming is the secret weapon of precision farming. Although advances have been made, precision agriculture still has some issues.

Precision agriculture has a big impact on crop recommendations.

Crop recommendations are significantly influenced by precision agriculture. Crop recommendations are based on a number of factors. In order to spot problems, precision agriculture focuses on identifying these parameters in a site-specific manner.

Precision agriculture does not always produce accurate results, but in agriculture it is important to have accurate and precise recommendations because errors can have very negative effects. financial and material loss. Numerous studies are being conducted to develop a more precise and effective model for crop prediction.

**4** **. Objectives**

In order to suggest the best crop, this paper uses soil PH, moisture content, temperature, nitrogen (N), phosphorus (P), potassium (K), and rainfall data. Based on the predicted future yields of eleven different crops, including rice, maize, jute, cotton, coconut, papaya, orange, apple, muskmelon, watermelon, grapes, mango, banana, pomegranate, lentil, black gramme, mung bean, moth beans, pigeon peas, kidney beans, chickpea, and coffee, the most suitable crop is suggested.

**5.Literature Survey**

A literature review was conducted in the context of crop recommendation using machine learning to examine the body of knowledge and research studies related to soil conditions and potential algorithms. The survey found that while there are a number of algorithms that can be used to predict crops, the majority of them are either overly complicated or fail to take into account crucial aspects for precise crop predictions. Our project's design and development benefited greatly from the information this literature review gave us about the research holes in the body of existing literature.

The team looked for and examined numerous patents, research papers, documents, newspapers, and magazine articles from various scenes for this project's literature review. The requirements and the reasons why people are moving towards precision agriculture [2], which is a result of globalisation, are stated in the paper [1]. Site-specific farming refers to precision agriculture. Although there have been improvements in precision agriculture over time, there are still some problems. As was already mentioned, site-specific methods of such systems require supervision to produce better results. Only a small number of the outcomes have a specific outcome. Despite this, farming is necessary because any lapse or error could seriously harm the plants as well as the resources in the event that it occurs.The aim of the earlier studies was also to predict the best crop type. But once the farmer or user plants the anticipated crop type, the system's job is done. The method put forth in this study also has a feedback system, though. The system is able to monitor plant growth and identify undernutrition in the farm by providing feedback even after recommending the best crop variety. so that the user can take the proper precautions in advance.

**6. Methodology**

Plants have varying requirements when it comes to their growth and development. Different plant species have adapted to different environmental conditions, which means that the optimal conditions for one plant may not be suitable for another. The five key factors that affect plant growth - temperature, humidity, soil pH, sunshine, and soil moisture - can vary greatly depending on the plant species.

Temperature is crucial for plant growth, and different plants thrive in different temperature ranges. For example, tropical plants require warm temperatures of around 20-30°C, while cool-season crops like lettuce and broccoli prefer temperatures between 10-20°C. Similarly, humidity levels can vary greatly depending on the plant species. While some plants, like ferns and tropical plants, prefer high humidity levels, others like succulents and cacti prefer dry conditions.

Soil pH is another important factor for plant growth. Some plants, like blueberries and rhododendrons, require acidic soils with pH levels below 6.0, while other plants, like beets and broccoli, prefer slightly alkaline soils with pH levels above 7.0.

Sunshine is also crucial for plant growth, and different plants have varying requirements for the amount of sunlight they need. Some plants, like tomatoes and peppers, require full sun for optimal growth, while others, like ferns and mosses, prefer shady conditions.

Finally, soil moisture levels can vary greatly depending on the plant species. Some plants, like succulents, have adapted to arid conditions and require very little water, while others, like rice and lotus, grow best in flooded conditions.

In conclusion, meeting the specific requirements of each plant is essential for achieving a satisfactory yield. Understanding the unique needs of different plant species in terms of temperature, humidity, soil pH, sunshine, and soil moisture is critical to successfully growing healthy plants.

**7. Data collection**

This component involves collecting data from various sources such as weather stations, soil sensors, satellite imagery, historical crop yield data, and other related factors. The data can be collected either in real-time or at regular intervals.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| N | P | K | temperature | humidity | ph | rainfall | label |
| 90 | 42 | 43 | 20.87974 | 82.00274 | 6.502985 | 202.9355 | rice |
| 85 | 58 | 41 | 21.77046 | 80.31964 | 7.038096 | 226.6555 | rice |
| 60 | 55 | 44 | 23.00446 | 82.32076 | 7.840207 | 263.9642 | rice |
| 74 | 35 | 40 | 26.4911 | 80.15836 | 6.980401 | 242.864 | rice |
| 78 | 42 | 42 | 20.13017 | 81.60487 | 7.628473 | 262.7173 | rice |
| 69 | 37 | 42 | 23.05805 | 83.37012 | 7.073454 | 251.055 | rice |
| 69 | 55 | 38 | 22.70884 | 82.63941 | 5.700806 | 271.3249 | rice |
| 94 | 53 | 40 | 20.27774 | 82.89409 | 5.718627 | 241.9742 | rice |
| 89 | 54 | 38 | 24.51588 | 83.53522 | 6.685346 | 230.4462 | rice |

*Fig. 1. Sample of dataset used*

* 1. **Data pre-processing**

This component involves cleaning and processing the collected data to remove any inconsistencies or errors. This step may also involve feature extraction, where relevant features are extracted from the raw data.

**8.Machine learning models**

* 1. **Decision Tree**: A decision tree is a supervised machine learning algorithm that utilizes a tree-like model of decisions and potential outcomes to generate a predictive model for both classification and regression problems. The algorithm splits the data set into subsets based on the features in the data, with each internal node in the tree representing a test on an attribute and each branch indicating the outcome of that test. The training phase of the algorithm involves selecting the best attribute to split the data at each node based on a measure of the attribute's impurity or information gain. The prediction phase involves using the trained decision tree to make predictions on new data points. Decision trees are known for their interpretability, ease of use, and ability to handle various data types. However, overfitting can be a concern, and techniques like pruning, regularization, and ensemble methods are used to enhance their performance.

**8.2 Random Forest:**

Random forest is a supervised machine learning algorithm that combines multiple decision trees to create a more robust and accurate model. It works by randomly selecting a subset of features and building decision trees on each subset. During the prediction phase, the algorithm aggregates the predictions of each decision tree to make the final prediction. Random forest is known for its ability to handle high-dimensional data, avoid overfitting, and provide feature importance rankings. However, it can be computationally expensive and less interpretable compared to single decision trees.

This component involves developing and training machine learning models such as decision trees, Naive Bayes, or Random Forest, to analyse the collected data and provide recommendations on the best crops to be cultivated.

* 1. **Support Vector Machine(SVM):**

SVM, or support vector machine, is a machine learning algorithm used for classification and regression analysis. It works by finding a hyperplane that separates the data points into different classes. The hyperplane is chosen to maximize the margin between the two classes, which increases the generalization ability of the model. SVM can handle both linear and non-linear data by mapping the input data to a higher dimensional space using a kernel function. SVM is widely used in various fields such as image classification, text classification, and bioinformatics due to its effectiveness in handling high-dimensional data with a relatively small number of training samples.

1. **Crop recommendation engine**

This component uses the trained machine learning models to provide recommendations on the best crops to be cultivated based on the input parameters.

1. **Environment Factors**

The environmental elements needed to be obtained in order to compare and forecast the initial data set. They are sensors that measure the intensity of the sun, soil moisture, soil pH, and humidity and temperature. To transmit the values to the next component of crop recommendation and to save the results in the database, the collected data are cleaned and processed using clustering and other techniques.

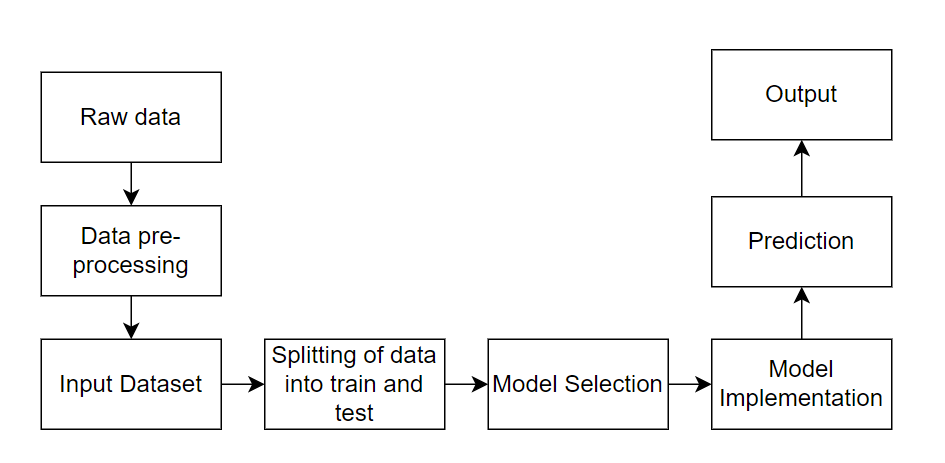
**11. Crop Forecast**

Because regional environmental variables differ, machine learning algorithms are used to estimate the ideal crop variety for the selected area. In order to train the crop-suggesting model using the information gathered from the Arduino sensors, machine learning algorithms [8] are used to determine the best crop to cultivate with the highest likelihood of growing. Naive Bayes and random forest algorithms are used to select the best crop type.

Using this model, it was determined what kinds of crops the farmer ought to grow. To do this, it is necessary to analyse variables such as humidity, temperature, soil moisture, pH level, and sunlight. utilising the Random Forest machine learning algorithm to look at the aforementioned traits.

1. **Monitoring and feedback**

This suggested system product can distinguish between twenty-two different types of crops depending on the environmental circumstances of the selected plot of land. The soil composition or any other changes in the chosen area, however, would be the reason to obtain a chance of more than 90% for the aforementioned crops. To avoid these factors having an impact on crop prediction, the approach also takes into account the farmer's feedback system.



*Fig 2. Block Diagram of Crop Recommendation System*

**13.Result**

Based on the results of the tests and the data available, it can suggest the best crop to plant for the biggest harvest.

Based on the regional environmental conditions in the preferred area, the best crop types are suggested to the farmer.

Decision trees’s Accuracy is: 90.0

support precision recall f1-score

apple 1.00 1.00 1.00

banana 1.00 1.00 1.00

blackgram 0.59 1.00 0.74

chickpea 1.00 1.00 1.00

coconut 0.91 1.00 0.95

coffee 1.00 1.00 1.00

cotton 1.00 1.00 1.00

grapes 1.00 1.00 1.00

jute 0.74 0.93 0.83

kidneybeans 0.00 0.00 0.00

lentil 0.68 1.00 0.81

maize 1.00 1.00 1.00

mango 1.00 1.00 1.00

mothbeans 0.00 0.00 0.00

mungbean 1.00 1.00 1.00

muskmelon 1.00 1.00 1.00

orange 1.00 1.00 1.00

papaya 1.00 0.84 0.91

Naive Bayes's Accuracy is: 0.990909090909091

support precision recall f1-score

apple 1.00 1.00 1.00

banana 1.00 1.00 1.00

blackgram 1.00 1.00 1.00

chickpea 1.00 1.00 1.00

coconut 1.00 1.00 1.00

coffee 1.00 1.00 1.00

cotton 1.00 1.00 1.00

grapes 1.00 1.00 1.00

jute 0.88 1.00 0.93

kidneybeans 1.00 1.00 1.00

lentil 1.00 1.00 1.00

maize 1.00 1.00 1.00

mango 1.00 1.00 1.00

mothbeans 1.00 1.00 1.00

mungbean 1.00 1.00 1.00

muskmelon 1.00 1.00 1.00

orange 1.00 1.00 1.00

papaya 1.00 1.00 1.00

pigeonpeas 1.00 1.00 1.00

pomegranate 1.00 1.00 1.00

rice 1.00 0.75 0.86

watermelon 1.00 1.00 1.00

...

accuracy 0.99

macro avg 0.99 0.99 0.99

weighted avg 0.99 0.99 0.99

SVM's Accuracy is: 0.9795454545454545

support precision recall f1-score

apple 1.00 1.00 1.00

banana 1.00 1.00 1.00

blackgram 1.00 1.00 1.00

chickpea 1.00 1.00 1.00

coconut 1.00 1.00 1.00

coffee 1.00 0.95 0.98

cotton 0.95 1.00 0.98

grapes 1.00 1.00 1.00

jute 0.83 0.89 0.86

kidneybeans 1.00 1.00 1.00

lentil 1.00 1.00 1.00

maize 1.00 0.95 0.98

mango 1.00 1.00 1.00

mothbeans 1.00 1.00 1.00

mungbean 1.00 1.00 1.00

muskmelon 1.00 1.00 1.00

orange 1.00 1.00 1.00

papaya 1.00 1.00 1.00

pigeonpeas 1.00 1.00 1.00

pomegranate 1.00 1.00 1.00

rice 0.80 0.75 0.77

watermelon 1.00 1.00 1.00

...

accuracy 0.98

macro avg 0.98 0.98 0.98

weighted avg 0.98 0.98 0.98

RF's Accuracy is: 0.990909090909091

support precision recall f1-score

apple 1.00 1.00 1.00

banana 1.00 1.00 1.00

blackgram 0.94 1.00 0.97

chickpea 1.00 1.00 1.00

coconut 1.00 1.00 1.00

coffee 1.00 1.00 1.00

cotton 1.00 1.00 1.00

grapes 1.00 1.00 1.00

jute 0.90 1.00 0.95

kidneybeans 1.00 1.00 1.00

lentil 1.00 1.00 1.00

maize 1.00 1.00 1.00

mango 1.00 1.00 1.00

mothbeans 1.00 0.95 0.97

mungbean 1.00 1.00 1.00

muskmelon 1.00 1.00 1.00

orange 1.00 1.00 1.00

papaya 1.00 1.00 1.00

pigeonpeas 1.00 1.00 1.00

pomegranate 1.00 1.00 1.00

rice 1.00 0.81 0.90

watermelon 1.00 1.00 1.00

accuracy 0.99

macro avg 0.99 0.99 0.99

weighted avg 0.99 0.99 0.99



**14.Conclusion**

In a contemporary setting with less space and less agricultural knowledge, all the factors are considered from the standpoint of the farmer and the plant, and the farmer is appropriately guided until the harvest. It's important to understand the factors that influence a plant's growth and how to maintain or manage them before deciding on one to cultivate. This system automatically evaluates these criteria before selecting the type of crop to be grown.

This system requires less upkeep and does not need professional guidance. The user won't incur any additional expenses as a result of using this system.

**15.Future Work**

It is capable of many extra system functionalities. Currently, it recommends a very suitable crop to be grown by taking relevant environmental parameters as inputs. To take things to the next level, however, an automation component and a mechanism for gathering farmer input can be implemented.

Currently, all environmental factors are fed into the system as inputs, but as an added feature, an algorithm can be used to forecast one factor using data from two other factors.

References

1. Doshi, Z. et al. (2018) “Agroconsultant: Intelligent crop recommendation system using machine learning algorithms,” 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)
2. P. A, S. Chakraborty, A. Kumar and O. R. Pooniwala, "Intelligent Crop Recommendation System using Machine Learning," 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2021, pp. 843-848, doi: 10.1109/ICCMC51019.2021.9418375.
3. G. Chauhan and A. Chaudhary, "Crop Recommendation System using Machine Learning Algorithms," 2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART), MORADABAD, India, 2021, pp. 109-112, doi: 10.1109/SMART52563.2021.9676210.
4. N. H. Kulkarni, G. N. Srinivasan, B. M. Sagar and N. K. Cauvery, "Improving Crop Productivity Through A Crop Recommendation System Using Ensembling Technique," 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS), Bengaluru, India, 2018, pp. 114-119, doi: 10.1109/CSITSS.2018.8768790.
5. 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, doi: 10.1109/ICCMC51019.2021.9418351.
6. Israni, Dishant and Masalia, Kevin and Khasgiwal, Tanvi and Tolani, Monica and Edinburgh, Mani, Crop-Yield Prediction and Crop Recommendation System (April 8, 2022). Available at SSRN: <https://ssrn.com/abstract=4111856> or [http://dx.doi.org/10.2139/ssrn.4111856](https://dx.doi.org/10.2139/ssrn.4111856)