

# Crop Prediction System Using Machine Learning

Yash Bhandari<sup>1,\*</sup>, Aruna Gawade<sup>2</sup>

<sup>1</sup>D.J.Sanghvi College Of Engineering, Mumbai, yashb21@gmail.com

<sup>2</sup>Department of Computer Engineering, D.J.Sanghvi College Of Engineering, Mumbai, [aruna.gawade@djsce.ac.in](mailto:aruna.gawade@djsce.ac.in)

**Abstract.** One of the most significant professions in India is agriculture. Agriculture is the primary industry for more than half the population. Additionally, the rising suicide rate over time is evident. The weather, family troubles, and debt are the main causes of this. We may also remark that farmers frequently lack knowledge of the crops that would be best for their soil's quality, nutrition, and rainfall potential. Early identification of the variables causing a fall in output can be aided by previous crop yield predictions. Additionally, it may help with the right application of fertiliser and pesticides, the selection of the best crop kinds, and early forecasting, which offers farmers the opportunity to make advance preparations for storage and selling. Additionally, they lack sophisticated methods for predicting crop output at the time of seeding. This approach is suggested to forecast soil fertility using a decision tree. The proposed method primarily focuses on evaluating soil fertility and rainfall amounts to forecast the most suited crop and to suggest viable alternatives for cultivation that can boost farmers' output.

## 1 INTRODUCTION

In India, one of the most significant industries is agriculture. It is crucial to the nation's overall growth. The kind of soil, the amount of rainfall, and the appropriate crop are the three key elements that affect the agriculture industry. In the past, yield production was done based on a farmer's experience with a particular field and crop. Farmers are under pressure to plant an ever-increasing variety of crops due to the rapidly changing present circumstances. Given the current scenario, many farmers are unaware of the advantages of growing new crops and lack sufficient understanding about them. By comprehending and predicting crop output under various environmental situations, we may also raise agricultural productivity. For novice farmers, selecting a crop based on soil and rainfall might be challenging. To improve our country's food output, agricultural land must be used effectively.

Understanding, interpreting, and drawing conclusions from data sets is the process of data analytics. One can utilise the conclusion to create customised systems or software. Computer science includes the topic of machine learning. The science of machine learning has advanced significantly in recent years and aids in the automation of processes, which tends to lessen the need for manual labour. Artificial intelligence known as machine learning (ML) enables computers to learn without being explicitly programmed. It is feasible to create a computer software that can adapt to new data by using machine learning (ML). Machine

Learning is classified into three categories: [1] Supervised Learning-Inputs for supervised learning come in the form of a labelled dataset. The primary

tasks carried out in supervised learning are regression and classification. The outcome of supervised learning is predicted by the kind of class. The primary objective is to produce formulas based on inputs and results. [2] Unsupervised Learning-Unsupervised learning does not come with a clear or well-labelled dataset. The primary activity in unsupervised learning is clustering. Unsupervised learning made predictions based on underlying patterns. In this, we group the input values together based on their associations. [3] Reinforcement Learning-Algorithms are taught to respond to their surroundings on their own using reinforcement learning. Exploitation is an issue that is dealt with via reinforcement learning. The outcome in reinforcement learning is defined by the agent. An agent in this situation learns through environment interaction and delayed input.

## 2 LITERATURE REVIEW

[1] Using a (MLR) crop analysis method The decision tree approach and classification are used to analyse over 362 datasets in order to produce findings. Numerous linear regression was used in the study by [P. Vinciya, et al. [13]] to predict the type of soil using a dataset that was split into organic, inorganic, and real estate categories. This method yields reliable and accurate results. [2] The study by [Shivnath Ghosh, et al [14]] transferred data to a back propagation network in order to analyse the test data set. The Back Propagation Network uses a hidden layer to enhance performance when predicting soil properties. In this instance, a back propagation network is used to develop a self-learning function to forecast soil quality using inputs. This is more accurate and performs better than the traditional methods, albeit occasionally the system may become slow and the outcomes may be

---

Corresponding author: yashb21@gmail.com

unpredictable. [3] Support Vector Machine (SVM) and Relevance Vector Machine (RVM) are two regression supervised machine learning techniques that are employed in [Zhihao Hong, et. al. [15]] to demonstrate their efficacy in predicting soil quality. a smart wireless gadget that can measure weather and soil moisture. The wireless gadget provides accuracy of 95% and a 15% error rate. It hasn't been tried with real-time data, though. [4] The article [Sabri Arik, et al. [16]] uses a back propagation method to check for soil fertility and plant nutrient levels. The correct findings allow for bettering soil characteristics. It works more effectively than conventional procedures. However, the system is unstable and slow-moving. [5] The work [Vaneesbeer Singh, et al. [17]] analyses soil and forecasts crop yield using three methods: Decision Tree, Naive Bayes Classifier, and KNN Classifier. The conclusions are not accurate, however principle induction and SVM can be utilized for more accuracy.

## 2 METHODOLOGY

The solution that we suggested accepts user location and soil composition as inputs. To improve accuracy, processing components evaluate numerous data sets. To find patterns in the data and handle it appropriately, the suggested system employs machine learning and prediction algorithms. This process will make it easier to supply the best alternative crop under the circumstances. As a result, the algorithm will use the farmer's location and soil type as inputs to recommend the most lucrative crop. Additionally, it will give more details about the crop.

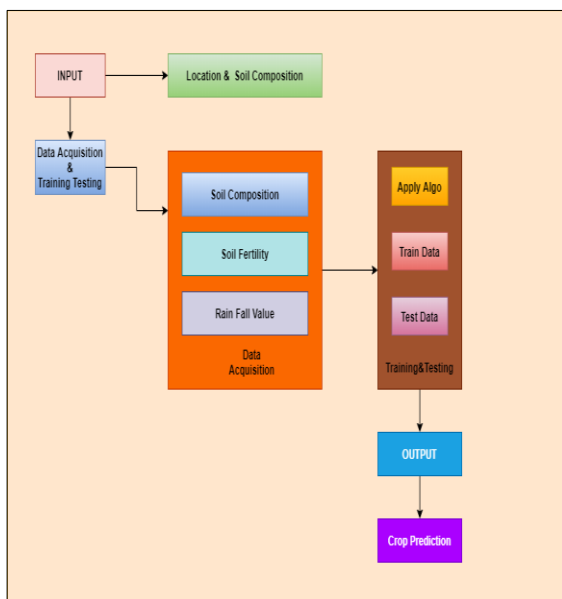


Fig1. System Architecture

Corresponding author: yashb21@gmail.com

### 3.1 Inputs

Data entry into a computer is referred to as an input. After entering the data into the computer, it can be processed and whatever command was given may be executed. The initial phase in the procedure involves gathering information such as location and soil fertility parameters (pH value of soil, phosphorus, potassium, pH value of soil, Electrical conductivity).

### 3.2 Data Acquisition

Data acquisition is the process of taking signals that measure actual physical occurrences in the real world and digitising them so that a computer and software may alter them. It is generally acknowledged that data acquisition is separate from prior methods of recording to tape recorders or paper charts. Data must first be collected, filtered, and cleaned before being stored.

### 3.3 Training and testing

#### 3.3.a Decision Tree Algorithm

The decision tree technique for supervised learning may be used to resolve classification and regression issues. By learning decision rules obtained from training data, we may construct Decision Trees to provide a training model that can be applied to anticipate a class of attribute values. Terms to Know About Decision Trees

[1] Root Node- Root nodes represent whole samples after they have been split up into several homogenous groupings. [2] Decision Node- When sub-nodes are divided into smaller sub-nodes. Such nodes are thus referred to as decision nodes. [3] Splitting- A node is split when it is divided into its child nodes. [4] Leaf node- Terminal nodes are nodes that cannot divide further. [5] Pruning- A decision tree is pruned when its sub-nodes are removed. This is splitting carried out backwards. [6]- Parent and Child Node: When a node is split into a sub-node, the resulting node is referred to as the child node. Parent nodes are non-sub nodes.

The decision tree algorithm begins by guessing the class of a given dataset from the tree's root node. This method compares the values with roots to the information or record, and then iterates based on the results. Once further comparing value with other sub-nodes for the following node, it proceeds. If we want to make the decision tree algorithm's operation as simple as possible, let's start the tree at the root node, let's call it "S," which has a complete dataset. The best attribute for the dataset will then be chosen using the Attribute Selection Measure. Later, smaller groupings of the dataset with probable values for the traits with the best fit were to be created. Following that, the decision tree node with the best attributes will be created. The nodes should repeat this process up until the point at which they are no longer capable of being classified.

#### 3.3.b C 4.5 Algorithm

C4.5 serves as a statistical classifier and is utilised for classification. It can be utilised as a data mining tool. C4.5 generates decision trees that represent categorization rules from a set of supplied examples. We can make meaningful decisions by using a sample of the data. It uses data for categorization and determines the data class. It functions with both continuous and discrete data. It also functions when the data is incomplete. Using a decision tree classifier, C4.5 makes decisions. Similar to a flowchart, a decision tree contains various factors that will affect your choice. In the C4.5 methods, we first choose an attribute from the instances. Then we choose a preliminary subset from the training examples, and based on the chosen characteristic, we construct a decision tree. The generated decision tree will then be compared against the remainder of the attribute to ensure correctness. If every case has been accurately identified, stop. If instances are not appropriately categorised, add the initial subset once again and build a new tree. Continue until the full training dataset has been used to build the tree or the tree properly classifies every occurrence. The reason for choosing is Able to manage both numerical and descriptive data, Easy to comprehend, interpret, and visualise, Decision trees do variable filtering inadvertently and C4.5 is adaptable in that it can be computed across a variety of timescales

### 3.4 Train Data

The machine learning algorithm is built up using data. The data provides input to the algorithm, which produces the desired result. The model continuously evaluates the data to understand the behaviour of the data and then adapts itself to accomplish its objective. Models for machine learning require data. Without a foundation of high-quality training data, even the most effective algorithms may become useless. Indeed, robust machine learning models can suffer greatly if they are initially trained on insufficient, inaccurate, or irrelevant data.

### 3.5 Test Data

Testing data after the model's creation reaffirms its capacity to make accurate predictions. Test data provide an immediate, useful verification of an unidentified dataset to demonstrate that the ML system was properly trained.

## 3 DETAILED DISSCUSION

There have been several studies undertaken to create an effective yield prediction system, but the emphasis has always been on statistical approaches, and nothing has been accomplished with a machine learning approach. The agricultural sector supports

more than half of the population. Farmers still lack modern methods for predicting crop output at the time of sowing, as well as fundamental information about their soils, which crops to plant in which types of soil, and how to apply fertilizers effectively. Crop yield forecasts made in advance can be useful in identifying the causes of output drop early on. In addition, it may help farmers choose the right crop kinds, apply fertilizers effectively, and save losses by giving them the opportunity to plan ahead for storage and distribution. Farmers still lack modern methods for predicting crop output at the time of sowing, as well as fundamental information about their soils, which crops to plant in which types of soil, and how to apply fertilizers effectively. Crop yield predictions made in advance can be useful in identifying the causes of output drop early on. The right use of fertilizer and insecticides might also benefit from it.

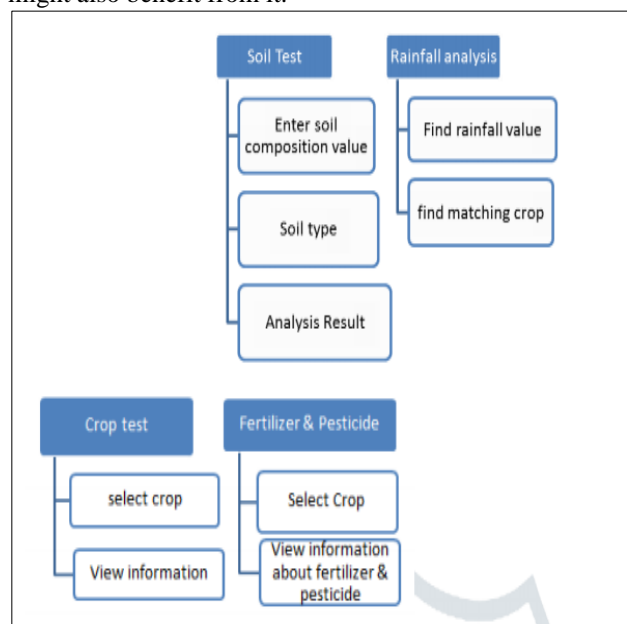


Fig2. Module Diagram

There are four modules in the system. The user enters the pH and location in the soil test analysis. The examination of the proportion of nutrients in the soil is the module's output. Through comparison with the crop database, the soil crop matching module identifies the appropriate crop that might be cultivated in that soil. The user is given recommendations for fertilizer and pesticide in the module that will result in the maximum crop output. The user may choose a crop and examine details on it in the crop test module. Using the system, it is possible [1] To give the farmer an easy-to-use GUI for handling. [2] To determine the optimal crop for the weather and soil conditions. [3] To find a potential replacement crop and boost output. [4] To provide a platform for farmers to access further crop, fertilizer, and pesticide information. The findings and specifics of the experiment are covered in Section 5. The primary goal of the study is to adopt the most cutting-edge technologies currently available to get better outcomes and determine how they work better than conventional techniques for crop prediction.

#### 4 EXPERIMENT AND RESULTS

The research paper shows implementation of Decision Tree Algorithm and C 4.5 algorithm to predict the crop. The experiment was carried out to achieve better accuracy results and predict the soil fertility and most suitable crop. After the model is trained, Graphical User Interface was created in which we entered the soil composition values. On the basis of inputs, we decided the fertility and type of soil. We consider factors like pH value of soil, phosphorous, potassium, Electrical conductivity which helped to decide the soil fertility (high, medium, low). The following visualization is created for soil prediction.

The figure shows a graphical user interface for soil fertility prediction. It contains five input fields with their respective units and a 'Predict Quality' button. The values entered are: N value (105.00 Kg/ha), P value (9.86 Kg/ha), K value (173.21 Kg/ha), pH value (6.79 Kg/ha), and EC Value (0.32 dS/m). Below the inputs, the text 'Quality Low' is displayed.

Fig3. Soil Fertility GUI

Later we consider the rainfall index of the specific area and of specific month and matched it with the soil fertility to predict the suitable crop. Following visualization is created for the rainfall index.

The figure shows a graphical user interface for rainfall index prediction. It contains three input fields: State (Maharashtra), District (Nasik), and Select month (May). Below these inputs, the text 'Rain 7.155714 mm' is displayed. A 'Rain' button is located at the bottom.

Corresponding author: yashb21@gmail.com

Fig4. Rainfall Index GUI

Considering the soil fertility and rainfall index of specific month, now can predict the most suitable crop for the specific farmer.

Crop	Period	Pesticide	Fertilizer
Grapes	180-200 Days	Azoxystrobin,Carbendazim	Potassium,Calcium,Boron,Nitrogen
Bajra(Pearl Millet)	.	.	.
Gooseberry(Amla)	.	.	.

Fig5. Final Output

#### 6 CONCLUSION

The system employs supervised machine learning techniques and provides the most accurate results. One or more alternative crops and those crops' fundamental requirements are the outputs of the machine learning algorithms. Thus, the method will aid in easing the issues the farmers confront. It will serve as a conduit for giving farmers the effective knowledge they need to obtain high yields and optimise revenues, which will lower rates of difficulty.

#### 7 REFERENCES

- [1] R Sujatha,Dr P.Issaki,"A Study of Crop Yield Prediction Using Data Mining Techniques" IEEE 2016.
- [2] Jeetendra Shenoy, Yogesh Pingle, "IOT in agriculture",IEEE 2016
- [3] N.Heemageetha, "A survey on Application of Data Mining Techniques to Analyze the soil for agricultural purpose", IEEE 2016
- [4] Essentials of Machine Learning Algorithms. [Online] Available at <https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/> (Accessed on Dec. 9, 2017).
- [5] Shweta Taneja, Rashmi Arora, Savneet Kaur, "Mining of Soil Data Using Unsupervised Learning Technique", International Journal of Applied Engineering Research, ISSN 0973-4562 Vol. 7 No.11, 2012.
- [6] D Ramesh , B Vishnu Vardhan, "Data mining technique and applications to agriculture yield data", International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 9, September 2013
- [7] Mucherino, P. Papajorgji, P.M. Pardalos, "Data Mining in Agriculture", Springer, 2009
- [8] M.R. Bendre, R.C. Thool, V.R.Thool, "Big Data in Precision agriculture",Sept,2015 NGCT
- [9] Monali Paul, Santosh K. Vishwakarma, Ashok Verma, "Analysis of Soil Behavior and Prediction of Crop Yield

using Data Mining a pproach”, 2015 International Conference on Computational Intelligence and Communication Networks.

[10] Dr.N.Suma, Sandra Rhea Samson, S.Saranya, G.Shanmugapriya, R.Subhashri, “IOT Based Smart Agriculture Monitoring System”, Feb 2017 IJRITCC.