

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/351481738>

# Crop Recommendation on Analyzing Soil Using Machine Learning

Article in Turkish Journal of Computer and Mathematics Education (TURCOMAT) · April 2021

---

CITATIONS  
26

READS  
9,273

---

5 authors, including:



Angu Raj

Sona College of Technology

12 PUBLICATIONS 150 CITATIONS

[SEE PROFILE](#)



Dr.Thiyaneswaran Balashanmugam

Sona College of Technology

48 PUBLICATIONS 636 CITATIONS

[SEE PROFILE](#)

## Crop Recommendation on Analyzing Soil Using Machine Learning

**Anguraj.K<sup>a</sup>, Thiyaneswaran.B<sup>b</sup>, Megashree.G<sup>c</sup>, Preetha Shri.J.G<sup>d</sup>, Navya.S<sup>e</sup>, Jayanthi.J<sup>f</sup>**

<sup>a,b,c,d,e</sup>Department of ECE, Sona College of Technology, Salem, India.

<sup>f</sup>Department of CSE, Sona College of Technology, Salem, India.

**Article History:** Received: 10 November 2020; Revised 12 January 2021 Accepted: 27 January 2021;  
Published online: 5 April 2021

---

### ABSTRACT:

Agriculture plays a significant role in increasing the economic development of our nation. Crop production has greatly affected due to changes in weather pattern. Emerging technologies can be used to improve productivity of the crops by converting traditional farming to precision farming. The new technologies that are used include data analysis and Internet of things (IOT). The major issue yet to be resolved is cultivating precise crop at precise time. This can be done with the help machine learning algorithms which is found to be an effective method for predicting the suitable crop. The soil parameters such as soil moisture, temperature, humidity and pH are collected from the sensors using IOT and given to Graphical User Interface (GUI). GUI gets the inputs and suggests the suitable crops. The system developed using IOT and ML greatly helps the farmers to take a valuable decision.

**Keywords:** Agriculture, Crop Suggestion System, Machine Learning, Precision Farming, Prediction, IOT.

---

### 1. Introduction

The main challenge faced in agriculture sector is the lack of knowledge about the changing variations in climate. Each crop has its own suitable climatic features. This can be handled with the help of precise farming techniques. The precision farming not only maintains the productivity of crops but also increases the yield rate of production. In India it is very important to maintain sustainable agriculture to meet growing needs. Though many steps have been carried out to minimize the loss of crops, traditional methods has its own disadvantages. These disadvantages can be overcome with the help of precision farming. With the use of IOT and prediction system, precision farming makes decision. The data from the field area are collected using IOT system. The prediction framework is fed with the collected data from the sensors for obtaining suggestion. Crop selection and changing climatic conditions are the two major problems faced among the farmers. With the existing prediction and monitoring methods we can address this problem. Though these methods are useful, there is no optimum solution for the crop suggestion. Some of the drawbacks found in the existing system are the improper analysis, choosing effective algorithms, and efficient selection of attributes all these parameters may affect the crop yield.

The proposed system helps in overcoming the drawbacks found in the existing system. The methods in the proposed system includes increasing the yield of crops, real-time analysis of crops using IOT, selecting efficient parameters, making smarter decisions and getting better yield. Effective algorithms need to be used for early prediction of crops [2]. By choosing correct factors most useful data can be generated using ML model by intake of various parameters and this generated data helps the farmers by suggesting a right crop to be sown. To suggest the precise crop for a particular field area is the main aim of crop suggestion model. By choosing the suitable crops for the field area we can minimize the loss of crops. Suitable algorithms with selective features has to be chosen since the accuracy level of suggesting crops may vary based on the type of algorithm chosen

has been chosen. Machine Learning is found to be the most effective technology for predicting suitable crops and its yield.

## **2. Related Works**

Literature review on nurturing technologies employed in agriculture was carried out. Here, glimpse on IOT, deep learning, sensing technologies deployed in agriculture were provided.

Aruul Mozhi Varman S et al proposed an IOT and deep learning based smart agriculture systems. This system monitors and collects the soil parameters from the field with the help of a wireless sensor network. The collected data is then uploaded in the cloud. Finally, the systems suggest best irrigation practices to the farmers by predicting the crop to be sown for next crop rotation. This information will be sent as an SMS to the farmers. The parameters include soil temperature, atmospheric temperature, and humidity [3]. This system suggests further improving the effectiveness by predicting the suitable time for applying pesticides, fertilizer, and manures [4].

Smart farming system based on sensing technology was proposed by Radhika et al [5]. Their system basically uses sensors such as pH, temperature, moisture, and sensor to measure the intensity of light. The sensors are positioned in such a way in which it effectively senses about the environmental conditions. This system simply collects the data from the environment without any processing. Hence it is suggested to apply the collected data into the algorithm for better crop production.

Suhas Athani et al [1] proposed IoT enabled soil moisture monitoring for predicting the future harvest. This system detects soil moisture level for obtaining good plant growth. Initially, soil samples are given as an input to the sensors then the respective pH content, salinity content and moisture content is measured using Wi-Fi shield[6]. From the database, the moisture reading can be fetched for Neural Network processing. In the final module, the generated correction factor is given as an input and the result produced by the neural network will get displayed in the form of mobile application. This work suggests neural network can be used for predictions based on seasonal variations.

Kumar et al [7] in their study proposed a prediction framework for farming which includes sensors that sense the agricultural characteristics with the help of IoT framework hence the farmers will be given advice. These sensors such as soil moisture temperature and PH are deployed in the agricultural field and it uses Radio Frequency communication to transfer the data [8] [9]. The transferred data are given as an input to the Decision Support System (DSS). The DSS will receive the input and give suggestion to the farmers about the detail of crop such as fertilizer usage, time for usage and time for irrigation etc. This information reaches the farmers with the help of an android application [17].

## **3. Methodology**

The invention of new technologies has made everything possible. The new technologies include the use of Internet of Things (IOT) and Machine Learning. The real time data from the field area can be collected using IOT system [10]. The collected data from the field area is fed to the trained model. The trained model then makes the predictions using the data. The result produced by the model greatly helps in sowing the suitable crops in the particular field area.

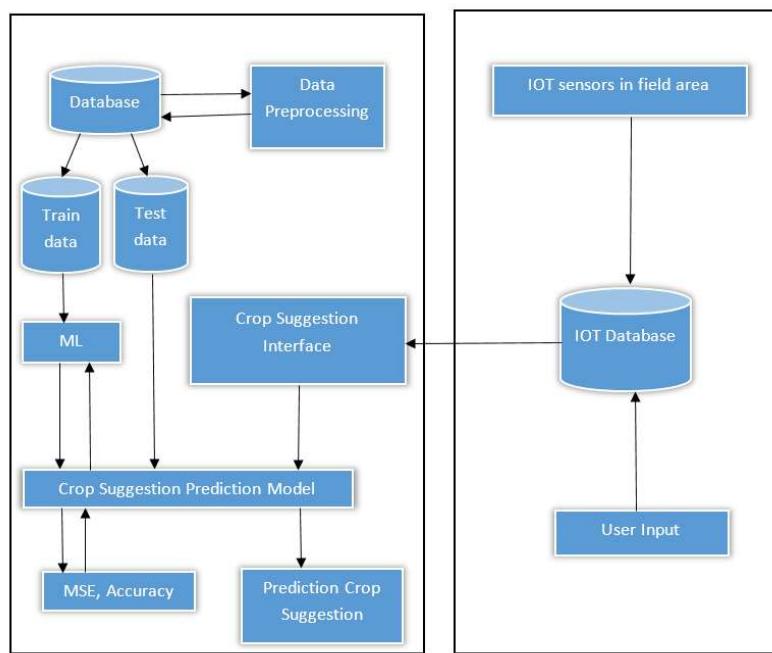
### **3.1 Implementation**

Implementation includes data collection from IOT, ML model designing, preprocessing the data and training, model testing and minimizing the error and improving the accuracy [18][19].

### **3.2 Architecture:**

The architecture contains collecting real-time data and building the prediction model and then creating a user interface for giving inputs. Initially data preprocessing is done. Once the preprocessing is done, using ML

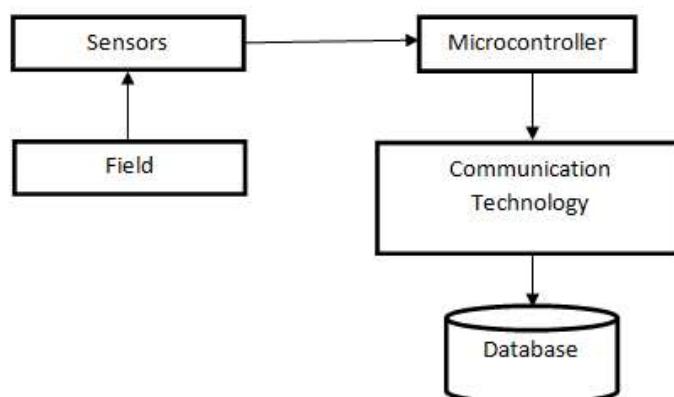
algorithm model is generated for prediction. The test data is given to the generated model for prediction [11]. The model is tested against random input values based on accuracy rate and error produced while predicting during testing. Until the error value is reduced and accuracy rate is improved this process is repeated [12]. In order to gather the input from the IOT sensors and the user, GUI is used. The inputs collected are given to the crop prediction model for predicting the precise crops.



**Figure 1:** Block diagram of Crop Suggestion System

### 3.3 Iot For Data Collection:

Firstly It Involves Collecting Data From The Field Area. The Parameters Such As Moisture, Temperature, Humidity And Ph Are Collected From The Particular Field Area. The Collected Data Is Then Stored And Given As Input To GUI. The Water Content In The Soil Is Measured Using Soil Moisture Sensor [13][14]. This Is Very Important Parameter Which Is The Basis For Providing Nutrients To The Soil. LM35 Temperature Sensor Is Used To Measure The Temperature Of The Soil. DHT22 Humidity Sensor Is Used To Measure Air Temperature And Moisture. Ph Meter Is Used To Measure The Ph Level Of The Soil And It Should Be Constantly Maintained [15][16]. To Collect The Information From The Sensors Arduino Microcontroller Is Used Which Is Responsible For Collecting Information From The Sensors. The Collected Information From The Sensor Is Stored In The Excel Sheet Using Wi-Fi.



**Figure 2:** IOT System Design

#### **4. Feature Description**

The key feature for plant growth is nutrients. The soil contains different types of nutrients. Soil helps in retaining water. This will help in proper nutrient supply to the plant growth [20]. Using moisture sensor, the feature soil moisture is measured from respective soil. This helps us in understanding the need of moisture content for the particular soil. Using temperature sensor, the temperature of the soil can be measured which indicates the optimal temperature for plant growth. Moisture level in the air is measured with humidity which is measured using humidity sensor[21]. The readings measured from the sensors are measured and used for processing. The pH of the soil is an important feature as it affects the plant growth in several ways. The optimal pH range for the soil is 5.5-7.0. The soil should be in the prescribed pH range to maintain the plant growth.

#### **5. Data Preprocessing**

The dataset collected from open source has to be cleaned. Since the machine learning algorithm cannot handle noisy, incompatible and incomplete data the data has to be cleaned [22]. The data with outliers and errors is referred here as noisy data. Any difference in features is referred as incompatible or inconsistent data. The lack features or attributes in the dataset is referred to as incomplete data. Data cleaning, data integration, data transformation and data reduction are the basic steps involved in data preprocessing.

##### **5.1 Encoding Categorical Data Using Label Encoding:**

The dataset can be divided into two categories numerical and categorical data. The data in numeric format is numerical data and the data in the form of strings is categorical data. The categorical data should be converted into numerical values since python does not deal with categorical data. LabelEncoder( ) class helps in converting categorical data into numerical data. This can be used where priority is needed in the features.

##### **5.2 Encoding Categorical Data Using One Hot Encoding:**

The encoder will assume that  $1 > 0$  and  $2 > 0$  during label encoding. This means that higher priority will be given to label 1 than label 0 and higher priority will be given to label 2 than label 1 and label 0. But we cannot apply this encoding to all types of features. Hence we can overcome this problem with the help of OneHotEncoder( ) class . It encodes the column to be encoded by splitting across multiple columns and thereby replacing it with 0 and 1.

##### **5.3 Numerical And Categorical Data Preprocessing:**

Data cleaning is done to preprocess these kind of data.. Handling missing values is the key feature of data cleaning. The result drawn from the dataset will get affected if the missing values in the dataset are not handled properly. These values can be handles using three different methods. These methods involve deleting the entire row, calculating mean, median or mode, and filling missing values by zero. Missing values can be handled using three different methods that include deleting the entire row, calculating mean, median or mode, and filling missing value by zero. Filling those missing values with zeros would the best method for handling missing values. This method does not contain any null value which in turn prevents data loss and hence it is an effective method. When compared to other methods this method will produce a better prediction.

#### **6. Feature Scaling:**

The various features in dataset are standardized using feature scaling. The range of the feature can be limited using feature scaling. This is usually done in data preprocessing and it is also known as normalization. Normalization comes under data transformation. Feature rescaling is one of the methods involved in feature scaling. This feature rescaling is also known as min-max scaling or min-max normalization. This is one of the simplest methods. This can be used to handle numerical features. The range of the features can be determined based on the nature of the dataset. The range of the features has to be minimized as it may result in inaccurate predictions. In order to avoid this problem rescaling can be done which scales all the features in the same range. In order to scale all the features in the range of 0 and 1, minmaxscaler() can be used. It is provided by scikit learn library. The formula for minmaxscaler is,

$$x = \frac{x - \min(x)}{\max(x) - \min(x)}$$

where,  $x$  is the normalized data value.

$x$  is the original value to be normalized.

$\min(x)$  is the minimal value of the feature to be normalized.

$\max(x)$  is the maximum value of the feature.

For the feature scaled value binarization can be applied to convert numerical feature value into a Boolean value.

## 7. Machine Learning For Building Prediction Model:

### 7.1 Random Forest:

Random forest is a supervised learning technique. It is a popular machine learning algorithm. It can be used for both Classification and Regression problems in Machine Learning. It is based on the process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. This concept is called ensemble learning. Random forest contains a number of decision trees. In order to improve the accuracy of the dataset predicted average is taken. Random forest takes the prediction from each tree rather than relying on one decision tree. Based on majority votes of predictions, it predicts the final output. The main advantage of Random forest is that it can handle large datasets with high dimensionality. It prevents over fitting issue by enhancing the accuracy of the model.

### 7.2 Naïve Bayes:

Naïve Bayes algorithm is a type of supervised learning algorithm. It is based on Bayes theorem . It is used for solving classification problems. Text classification involves using Naïve Bayes which includes high-dimensional training dataset. It is one of the simple and most effective algorithm for classification. It also helps in building fast machine learning models that can make quick predictions. It is used to predict on the basis of probability of an object. Hence it is known as probabilistic classifier. It is called so because it assumes the occurrence of a certain feature is independent of the occurrence of other features. For example if a fruit is identified on the bases of colour, shape and taste, then red, spherical and sweet fruit is recognised as apple. So here each feature is individually used to identify that it is an apple independently. Since it depends on the principle of Bayes theorem it is called as Bayes. Naïve Bayes can also be used for binary and Multi-class Classifications.

Bayes' Theorem:

- Bayes' theorem is used to determine the probability of hypothesis with prior knowledge and it is also known as Bayes' Rule or Bayes' law. It depends on the conditional probability.
- The formula for Bayes' theorem is,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Where,

$P(A|B)$  is the Posterior probability.

$P(B|A)$  is the Likelihood probability.

$P(A)$  is the Prior Probability.

$P(B)$  is the Marginal Probability.

## **8. Model Design By Machine Learning:**

Now-a-days for various data science projects, an open source python distribution anaconda is used. For predictive analytics spyder and jupyter can be used which are available on the navigator. To run a whole program in a partitioned manner jupyter can be used. For powerful editing, visualization, debugging and analysis option spyder can be used.

### **8.1 Package Description:**

Using command prompt the packages are installed. Packages such as matplotlib, numpy, pandas and scikit-learn should also be installed. The various types of graphs such as histogram, scatter plot, two-dimensional plots etc., are plotted using matplotlib. The different operations on multidimensional array and to store resultant value is done using numpy. In order to work with labelled data and to perform data manipulation pandas package is used. The function based on scikit-learn is sklearn-preprocessing. This function supports the splitting a dataset for training and testing. It is also used to preprocess the data using label encoder which is MinMaxScaler.

### **8.2 Training The Model:**

The dataset is read once the packages are imported. Then it is subjected to data preprocessing. Data encoding, filling the missing values and feature scaling are some of the data preprocessing methods. Once the data preprocessing is done, the dataset is divided into train and test dataset. The train data is around 80% of the data and the test data is around 20% of the dataset.

### **8.3 Learning Rate:**

To train the model, learning rate should be mentioned. To train the model, the learning rate should be given low. The model gets trained incorrectly if the learning rate is high. When the number of epochs increases, the loss decreases for a good learning rate. The loss increases for a very high learning rate. The loss decreases for low learning rate. Mean Squared Error (MSE) is used to measure loss. If the learning rate is low, then the training will be reliable. Initially we can choose a large value and then we can reduce the value with a hundred epochs.

### **8.4 Gradient Descent:**

For training the model, gradient descent technique is used. A gradient descent tries to minimize the loss function by improving the prediction accuracy. The gradient is a measure that tells about the change in error. While training it is found that there is a difference in the actual and predicted output this refers there is an increase in error rate. In such cases, with the help of gradient descent finds out the minimum error. Hence at final, the accuracy is improved to be around 97%.

## **9. Crop Suggestion User Interface:**

With third party software such as Qt Designer, spyder can also be used. Qt designer is a software which is used as a tool for building Graphical User Interface (GUI). Interactive GUI can be created using PyQt5 package which is available in anaconda. With the help of this package we can run the GUI generate by Qt designer. The GUI fetches data such as moisture, temperature, humidity and pH from IOT database using Qt designer. The data collected from the field area is stored in IOT database.

The GUI database contains the features entered in the GUI. This is given to the crop prediction model for crop prediction. The output from the trained model is given to crop suggestion interface. Then the suitable crop will be suggested by GUI. The predicted result greatly helps the farmers to choose the precise crops. Hence precision decision is made by the farmers with the help of the predicted result.

```

jupyter finial_year Last Checkpoint: 12/15/2020 (autosaved)

File Edit View Insert Cell Kernel Widgets Help
Run Cell Code

In [27]: y_p = regressor.predict(np.array([[27.35152643, 55.99375012, 7.13411409, 148.9812524999998]]))
y_list = (y_p.flatten('F'))
import pandas as pd
pd.set_option('display.max_columns', 34)
print(data[:9])
y_list = int(y_list)
print("-----")
print(y_list)
print("the suitable crop is",index2[y_list])

temperature humidity ph rainfall label label1
1207 12.141907 83.568125 5.647202 69.631220 grapes 7
587 30.888831 52.626968 8.634930 55.519324 mothbeans 13
893 27.996279 68.642859 7.327110 46.105852 lentil 10
1127 29.980805 49.486133 6.442393 91.822716 mango 12
1508 21.911913 91.687481 6.499227 117.076128 apple 0
1905 24.694571 81.735888 6.628723 78.584944 cotton 6
1132 35.787774 51.941903 5.395276 100.216061 mango 12
2183 24.592457 56.468296 7.288212 137.704405 coffee 5
278 19.462104 18.728320 7.217018 68.814051 chickpea 3
-----
5
the suitable crop is coffee

```

**Figure 3:** Output Suggesting Suitable Crop

## 10. Result and Discussion:

The dataset collected from the open source is initially divided into testing and training dataset. The ML model is provided with training dataset for generating the crop suggestion prediction model. The test data is given to the model, once the model is generated with minimum error and maximum accuracy. The inputs are fed to the generated model. The model then predicts and suggests the crops to be sown with an accuracy of about 96.89%.

## 11. References:

1. Athani, S., Tejeshwar, C. H., Patil, M. M., Patil, P., & Kulkarni, R. (2017, February 1). Soil moisture monitoring using IoT enabled arduino sensors with neural networks for improving soil management for farmers and predict seasonal rainfall for planning future harvest in North Karnataka — India
2. Biradar, H. B., & Shabadi, L. (2017). Review on IOT based multidisciplinary models for smart farming.
3. Mehta, P., Shah, H., Kori, V., Vikani, V., Shukla, S., & Shenoy, M. (2015). Survey of unsupervised machine learning algorithms on precision agricultural data. IEEE Xplore.
4. Sandhiya, D., & Thiyaneswaran, B. (2017). Extraction of dorsal palm basilic and cephalic hand vein features for human authentication system. 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET). IEEE.
5. Radhika, Y., & Shashi, M. (2009). Atmospheric Temperature Prediction using Support Vector Machines. International Journal of Computer Theory and Engineering, 55–58.
6. Ahmad, S., Kalra, A., & Stephen, H. (2010). Estimating soil moisture using remote sensing data: A machine learning approach. Advances in Water Resources, 33(1), 69–80.
7. Kumar, R., Singh, M. P., Kumar, P., & Singh, J. P. (2015). Crop Selection Method to maximize crop yield rate using machine learning technique. IEEE Xplore.
8. Thiyaneswaran, B., Anguraj, K., Kumarganesh, S., & Thangaraj, K. (2020). Early detection of melanoma images using gray level co-occurrence matrix features and machine learning techniques for effective clinical diagnosis. International Journal of Imaging Systems and Technology, (ima.22514). doi:10.1002/ima.22514
9. Shah, N. P., & Bhatt, P. (2017). Greenhouse Automation And Monitoring System Design And Implementation. International Journal of Advanced Research in Computer Science, 8(9), 468–471.
10. Jayanthi, J., and J. Selvakumar. "A novel framework to facilitate personalized web search in a dual mode." *Cluster Computing* 20.4 (2017): 3527-3535.

11. Parker, C. (2012). Unexpected challenges in large scale machine learning. Proceedings of the 1st International Workshop on Big Data, Streams and Heterogeneous Source Mining Algorithms, Systems, Programming Models and Applications - BigMine '12.
12. B Thiyaneswaran, K Anguraj, M Sindhu, N. S Yoganathan, J Jayanthi. (2020). Development of Iris Biological Features Extraction for Biometric Based Authentication to Prevent Covid Spread. International Journal of Advanced Science and Technology, 29(3), 8266–8275.
13. Pudumalar, S., Ramanujam, E., Rajashree, R. H., Kavya, C., Kiruthika, T., & Nisha, J. (2017). Crop recommendation system for precision agriculture. IEEE Xplore.
14. Anitha, P., & Chakravarthy, T. (2018). Agricultural Crop Yield Prediction using Artificial Neural Network with Feed Forward Algorithm. International Journal of Computer Sciences and Engineering, 6(11), 178–181. <https://doi.org/10.26438/ijcse/v6i11.178181>
15. B Thiyaneswaran, V Bhuvaneshwaran, M.S Dharun, K Gopu, T Gowsikan (Ed.). (2020). Breathing level monitoring and alerting by using embedded IOT (Vol. 10). Journal of Green Engineering.
16. Rekha, P., Rangan, V. P., Ramesh, M. V., & Nibi, K. V. (2017). High yield groundnut agronomy: An IoT based precision farming framework. IEEE Xplore.
17. Gupta, A., Nagda, D., Nikhare, P., & Sandbhor, A. (2021). Smart Crop Prediction using IoT and Machine Learning. International Journal of Engineering Research & Technology, 9(3).
18. Biradar, H. B., & Shabadi, L. (2017). Review on IOT based multidisciplinary models for smart farming. IEEE Xplore.
19. Jayanthi, J., and J. Selvakumar. "A novel framework to facilitate personalized web search in a dual mode." *Cluster Computing* 20.4 (2017): 3527-3535.
20. Priya, P. K., & Yuvaraj, N. (2019). An IoT Based Gradient Descent Approach for Precision Crop Suggestion using MLP. *Journal of Physics: Conference Series*, 1362, 012038.
21. Thiyaneswaran, B., Saravanakumar, A., & Kandiban, R. (2016). Extraction of mole from eye sclera using object area detection algorithm. 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET). IEEE
22. Patil, A., Beldar, M., Naik, A., & Deshpande, S. (2016). Smart farming using Arduino and data mining. IEEE Xplore. <https://ieeexplore.ieee.org/document/7724599>