International Journal of Research in Engineering, Science and Management
Volume 5, Issue 6, June 2022
https://www.ijresm.com | ISSN (Online): 2581-5792

Soil Health Monitoring System using Random Forest Algorithm

Akanksha Mahangare[1*], Jatin Kumar[2], Rhea Simon[3], Shubham Mallick[4], Arvind Jagtap[5], Rishikesh Yeolekar[6]

[1,2,3,4]*Student, Department of Information Technology, MIT Art, Design & Technology University, Pune, India*
[5]*Associate Professor, Department of Information Technology, MIT Art, Design & Technology University, Pune, India*
[6]*Assistant Professor, Department of Information Technology, MIT Art, Design & Technology University, Pune, India*

***Abstract*: Agriculture is India's largest economic sector and plays a critical role in the country's overall socioeconomic development. Indian farmers produce less than farmers in other countries due to inadequate usage and implementation of modern agricultural practices and technologies and it leads to low productivity, which is impeded by neglect and high expenses. Knowing the level of soil nutrients can help farmers enhance agricultural output because insufficient nutrient levels can affect crop yield, while excess nutrient levels can either have the same effect or be wasted. A machine learning-based remote monitoring system can boost crop productivity and quality. The application of appropriate machine learning algorithms to collect data and analyze it can benefit in crop recommendation.**

***Keywords*: Crop recommendation, IoT, Machine Learning, NPK sensor, Random Forest, Soil health.**

1. Introduction

Agriculture could benefit from the adoption of IoT. Agricultural processes become more transparent as farmers gain access to sensor data analytics. Machine Learning in Agriculture, for example, with its high-precision algorithms, provides significant information into their field's performance. The goal of this project is to increase crop quantity and quality.

Soil Health Monitoring is a difficult problem in agriculture, and various models have been created and tested thus far. Soil Health Monitoring is a challenging topic in agriculture, and several models have been developed and tested to date. Because several factors influence crop productivity, including weather, fertilizer type, soil, and seed, this challenge necessitates the use of multiple datasets. As a result, measuring agricultural production is a time-consuming operation that needs a series of processes.

With this in mind, a soil health monitoring and analysis system is proposed in this project, in which the farmer can use a hardware model consisting of sensors to monitor soil moisture, soil temperature, and soil nutrient content such as Nitrogen, Phosphorus, and Potassium. The data collected from the soil sensors is transmitted to the cloud using the Arduino microcontroller. To make it easier for farmers, the information received from the sensor is accessed in the web application without requiring a login. As a result, this innovative technology assists farmers in determining correct parameters and recommending suitable crops based on soil nutrients, making the soil monitoring process easier.

2. Literature Survey

[1] Using modified support vector regression, a common machine learning technique, and four modules, this Paper calculates real-time sampling of soil parameters. Sensor interfaced to IoT device, Agri cloud, real-time sensor data analysis. The first module is a NodeMCU portable IoT device containing soil moisture and pH sensors, as well as environment sensors. Storage is included in the agri cloud module. Using a modified support vector machine method, the real-time data module is processing different types of crops and small plants. Agri-user interface is a simple web interface for farmers.

Thus, with the help of soil qualities, farmers will be able to determine the sorts of crops and tiny plants that may be cultivated in farms.

[2] By employing multiple categorization methods and analysing different characteristics, Reference Paper [2] concluded that this work aids in enhancing the success of agricultural yield rates. To predict crop yield, various machine learning techniques were examined. Artificial neural networks, support vector machines, K-Nearest Neighbours, Decision Trees, Random forests, Gradient boosted decision trees, regularized greedy forests, and the proposed CSM technique (Crop Selection Method), which aids in predicting the sequence of crops that can be considered for strategizing in the upcoming seasons, are among the algorithms included for comparative analysis.

[3] A random forest algorithm is used to calculate global and regional crop yields for potato, maize, and wheat, as well as environmental variables such soil, climate, photoperiod, fertilization data, and water. Root mean square, Nash-Sutcliffe model efficiency, index of agreement, and observed vs. predicted plots were used to compare the results with multiple linear regressions. The information is gathered from many sources throughout the United States. The results demonstrated that Random Forest is a dynamic and useful method for

*Corresponding author: aakankshac02@gmail.com

agricultural yield prediction with high accuracy, precision, and ease of use.

[4] Machine Learning Algorithm (KNN) determines the parameter to advise the best crop to produce in the area a certain field depending on data obtained in real time. A standardized dataset containing the basic data for a certain crop is kept and utilized to predict the crop. The sensors are placed in the field where the readings will be taken. Soil Moisture, DHT11, MQ2, the Light Intensity Sensor transmits values in real time data from the sensors to a cloud-based server.

[5] The main goal was to simulate and estimate crop production in order to provide optimal crop management and adequate results. This paper employed a three-layered artificial neural network (ANN) and the R programming language to predict and simulate agricultural yield.

3. Method

A. Random Forest Algorithm

Random Forest is a supervised learning technique that may be applied to both classification and regression. The Random Forest method generates decision trees on several data samples, predicts the data from each subset, and then votes on the best option for the system. The data was trained using the bagging approach using Random Forest. Essentially, the bagging approach is a combination of researching many models in order to improve the system's final outcome. To get high accuracy, we used the Random Forest technique, which predicts the real outcome in the dataset and delivers accuracy by forecasting by model. A decision tree is built from a sample of data in the random forest, and each family's prediction is given, with the best choice chosen by voting, resulting in better model accuracy. It yields the best results for the system as shown in fig. 1.

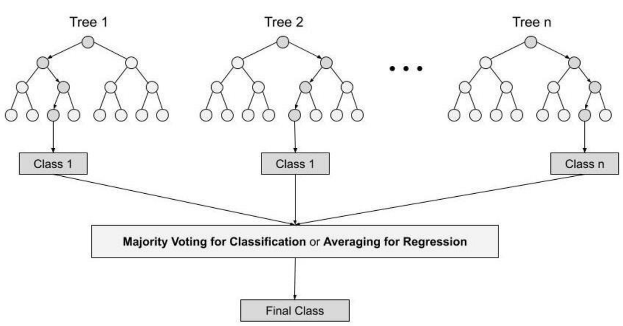

Fig.1. Random Forest algorithm

B. Proposed System

Nitrogen(N), Phosphorous(P), Potassium(K), PH value of soil, Humidity, Temperature, and Rainfall are all included in the dataset. The data was gathered through the Kaggle website. The data set contains 2203 instances. Rice, Chickpeas, Kidneybeans, Pigeonpeas, Mothbeans, Mungbean, Blackgram, Lentil, Pomegrante, Banana, Mango, Grapes, Watermelon, Muskmelon, Apple, Orange, Papaya, Coconut, Coffee are among the eighteen crops represented in this dataset.

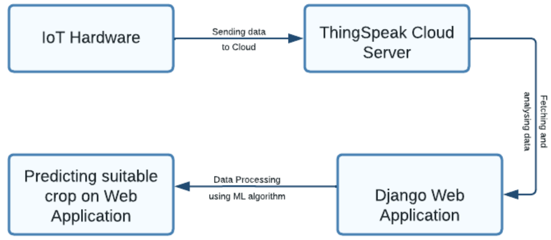

Fig. 2. Block diagram of the proposed system's overall methodology

- We are working with live data because we are using an IoT device to capture the data. The first sensor on our IoT device is NPK, which measures nitrogen, potassium, and phosphorus. The nutrients in a soil are represented by these three factors. Second, we have a temperature sensor that measures the soil's temperature. The third sensor is a humidity sensor, which is used to measure the soil's humidity. All of the sensors capture data and send it wirelessly to think speak, a statistically based cloud server. The information is presented in a graphical way.
- We are allowing live data access, retrieval, and logging of data by using ThingSpeak cloud, which provides an API to both the hardware and the Django web application.
- Using the Python API, we are retrieving data from the ThingSpeak server in a Django web application. We use ThingSpeak's secret channel ID to read and write live data that we collect from soil. The Random Forest Algorithm is then used to perform analysis on the dataset.
- The random forest algorithm, which is based on the decision tree method, is applied. To eliminate null values from the dataset, a pre-processing method is used. The scikit learn module's technique is used for training and testing. The dataset was split in half, with 80% of the data going into the training dataset and 20% going into the testing dataset. Fitting is used to increase the model's accuracy. When predicting the likelihood of a specific result, "prediction" refers to the outcome of an algorithm after it has been trained on a previous dataset and applied to new data. We can predict the model using the predict() method and the test feature dataset. As a result, it returned an array of expected values.
- A statistic is calculated by the Classification Report to analyse how effectively a classification algorithm's predictions performed. Precision, recall, and the model's f1-score are the three outcomes, where

$$\text{Precision} = \frac{True\ Positive}{Tptal\ Predicted\ Positive},$$

$$\text{Recall} = \frac{True\ Positive}{Total\ Actual\ Positive}$$

and

$$\text{F1 Score} = 2\times \frac{Precision * Recall}{Precision + Recall}$$

4. Conclusion

This project aims to anticipate the best crop for a particular area of land, allowing farmers to grow crops more effectively. Finally, we found that the Random Forest Algorithm had the accuracy with 99.09% on the crop recommendation dataset. As a result, this system will assist farmers in reducing their struggles. We are working with recommendation of Crops Suitable to the Land based on an analysis of the essential soil parameters such as soil moisture, pH, soil NPL values etc. This will serve as a simple way to provide farmers with the information they need to produce a high harvest, thereby increasing their profit and reducing their troubles.

References

[1] Radhika, Narendiran, "Kind of Crops and Small Plants Prediction using IoT with Machine Learning," International Journal of Computer & Mathematical Sciences, April 2018

[2] Rakesh Kumar, M.P. Singh, Prabhat Kumar and J. P. Singh, "2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, T.N., India. 6 - 8 May 2015, pp. 138-145.

[3] Jeong JH, Resop JP, Mueller ND, FleisherDH, Yun K, Butler EE, et al. (2016)"Random Forests for Global and Regional Crop Yield Predictions", PLoSONE 11(6): e0156571, pp. 1-15.

[4] T Raghav Kumar, Bhagavatula Aiswarya, Aashish Suresh, Drishti Jain, Natesh Balaji, Varshini Sankaran, "Smart Management of Crop Cultivation using IoT and Machine Learning," International Research Journal of Engineering and Technology, Nov. 2018.

[5] Mitra P., Mukherjee A., and Chandra A., "Machine learning for prediction of crop yield: a case study on rain-fed area rice yield from West Bengal, India," Proceedings of the DSANRM2017: Workshop on Data Science for Agriculture and Natural Resource Management; December 2017, Hyderabad, India: International Institute of Information Technology (IIIT, Hyderabad).