

# An Intelligent IoT-Based Soil Nutrient Analysis and Recommendation System

Aryan Gaonkar

Finolex Academy of Management and  
Technology, Ratnagiri  
University of Mumbai

Dewang Korgaonkar

Finolex Academy of Management and  
Technology, Ratnagiri  
University of Mumbai

Suyash Morye

Finolex Academy of Management and  
Technology, Ratnagiri  
University of Mumbai

Sarthak Parab

Finolex Academy of Management and  
Technology, Ratnagiri  
University of Mumbai

**Abstract:** This work “An Intelligent IoT-Based Soil Nutrient Analysis and Recommendation System” focuses on improving agricultural productivity and sustainability through intelligent soil analysis using machine learning techniques. In modern agriculture, understanding the nutrient composition of soil plays a vital role in determining suitable crops and fertilizer requirements. However, traditional soil testing methods are often time-consuming, expensive, and require specialized laboratory facilities, making them less accessible to small-scale farmers. To address this challenge, our project implements a data-driven approach where soil nutrient datasets were collected from Kaggle, consisting of essential parameters such as Nitrogen (N), Phosphorus (P), Potassium (K), pH, and moisture content. After preprocessing the data, various machine learning algorithms such as Random Forest and XGBoost were applied to analyze the soil properties and predict suitable crop or fertilizer recommendations. The models were trained and evaluated using key performance metrics including accuracy, precision, recall, and F1-score to identify the most effective predictive model. The system aims to provide an automated and accurate solution for soil nutrient assessment. In the next phase, the trained model will be integrated with IoT-based soil sensors to enable real-time soil data collection and analysis. A web or mobile application will also be developed to present soil health information and recommendations in a simple, user-friendly manner, making precision agriculture accessible to all farmers.

**Keywords:** IoT, ESP32, NPK sensor, XGBoost, Precision Agriculture, Soil Health Card, Crop Recommendation, Fertilizer Recommendation

## I. INTRODUCTION

### 1.1 Background and Motivation:

Agriculture continues to be the backbone of the Indian economy; however, production can be negatively affected by traditional farming practices that rely on empirical estimation based on prior experience rather than scientific

evidence. In an attempt to solve this problem, the Government of India initiated the Soil Health Card (SHC) program in 2015, which offers farmers soil nutrient test results at regular intervals. Although the SHC has set the stage for soil consciousness, the program is restricted by some logistical issues, including a 2-year test period and mandatory physical laboratory facilities. The need to develop an application that complements the SHC program through instant online assessment using Artificial Intelligence and Machine Learning is the driving factor behind this study.

### 1.2 Problem Statement:

While soil health data exists, small farmers may not have the technical knowledge required to interpret nutrient values in terms of crop selection and fertilization practices. Soil testing procedures are known for being expensive, time-consuming, and prone to human error. Moreover, existing systems lack real-time integration or accessibility to help determine the ideal crop for a particular setting, and also the right amount of fertilization required to achieve maximum crop yield without degrading the soil quality. Over-fertilization arising from the information vacuum causes economic strain on farmers and poses considerable environmental risks such as ground water pollution and soil toxicity.

**1.3 Objective:** The primary aim of this research is to design an advanced two-level automation system using machine learning techniques to increase the efficiency of digital soil health analysis. This system will analyze a range of classification algorithms, which include Random Forest, XGBoost, SVM, and KNN, to determine the best model that can be used for accurate crop prediction. This method makes use of a two-phase approach, whereby the first phase

identifies the best crop according to the soil and environmental variables. After determining the most suitable crop, the second phase will identify the correct quantity of fertilizers needed by the identified crop. Environmental variables like temperature, humidity, pH levels, and rainfall (gathered from a weather API) have been incorporated in this system to improve prediction accuracy. Moreover, this system will integrate IoT-based hardware devices, namely a 7-in-1 NPK sensor and an ESP32 microcontroller, to acquire and transmit data seamlessly. The main objective of this system is to provide a timely, efficient, and economical service that eliminates reliance on old methods of testing soil health, ensures there are no lags in the process, and encourages sustainable farming by avoiding over-application of fertilizers.

## II. LITERATURE SURVEY

**[Paper 1]** The paper titled “Revolutionizing Somali Agriculture: Applying Machine Learning and IoT for Optimal Crop Recommendations,” authored by R. Pallavi Reddy et al., explores the potential to increase agricultural productivity in Somalia by incorporating technologies such as machine learning and IoT. The study applies decision tree, random forest, and K-Nearest Neighbor models to suggest the appropriate crops using Kaggle crop and soil datasets. Among the various challenges identified in implementing the technology include weather instability, lack of sufficient resources, and inadequate infrastructure. Nevertheless, the suggested system showcases the effectiveness of IoT and ML in optimizing crop selection.

**[Paper 2]** In the paper titled “Soil Mapping for Farming Productivity: Internet of Things (IoT) Based Sustainable Agriculture,” the authors, Swapna Babu et al., (2025), present their findings on increasing farming efficiency by using data analytics to select suitable crops and soil types. The researchers use Kaggle datasets for crops and soils to conduct predictive analysis and data mining. They further explore the problems associated with security, privacy, and high cost of implementing IoT-based systems. The authors assert that using sensors with Arduino controllers is an effective and affordable means of detecting soil nutrient deficiencies.

**[Paper 3]** Dutta et al. (2023) developed a machine learning-based system for recommending suitable fruit crops by analyzing soil nutrient characteristics and climatic conditions. The study implements a Light Gradient Boosting Machine (LightGBM) model and evaluates its performance against several other algorithms, including Random Forest, Support Vector Machine, K-Nearest Neighbors, Artificial Neural Networks, and Convolutional

Neural Networks. For experimentation, datasets were collected from publicly available sources such as the UCI Machine Learning Repository along with domain-specific agricultural data. The findings indicate that boosting-based approaches are highly effective in capturing complex relationships within agricultural datasets. The authors also emphasize that accurate crop recommendation systems require continuous monitoring of changing environmental conditions and consideration of external factors that influence crop productivity.

**[Paper 4]** Folorunso et al. (2023) present a comprehensive review of machine learning techniques applied to soil nutrient prediction and digital soil mapping (DSM). The study examines a wide range of algorithms, including traditional models such as Random Forest, K-Nearest Neighbors, Support Vector Machine, Artificial Neural Networks, and Deep Neural Networks, along with advanced approaches like ANFIS, Gradient Boosting methods, Extremely Randomized Trees, Support Vector Regression, Multivariate Adaptive Regression Splines, and Genetic Programming. The analysis is based on datasets sourced from platforms such as Kaggle, focusing on major agricultural crops. One of the key challenges identified in the study is the limited availability of high-quality soil data in developing regions, which can affect model performance and generalization. Overall, the review concludes that machine learning methods play a crucial role in enhancing the accuracy and efficiency of soil property prediction and classification.

**[Paper 5]** “Kollu et al. (2023) propose an IoT-enabled framework for fertilizer recommendation aimed at improving agricultural productivity and addressing food security challenges. The study integrates machine learning techniques with real-time data acquisition to optimize fertilizer usage. For experimentation, datasets were obtained from open-source platforms and analyzed using multiple algorithms, including Multilinear Regression (MLR), Random Forest, C4.5 decision tree, Naïve Bayes, and Sequential Forward Floating Selection (SFFS) for feature selection. The results demonstrate that the MLR-based IoT approach can effectively generate accurate fertilizer recommendations by utilizing real-time soil and crop-related data, thereby supporting efficient nutrient management and sustainable farming practices.

**[Paper 6]** Musanase et al. (2023) present a data-driven Crop and Fertilizer Recommendation System (CFRS) designed to improve agricultural productivity, particularly in the context of Rwanda. The study utilizes datasets sourced from Kaggle in combination with region-specific NPK data to develop predictive models. Multiple machine learning techniques, including Random Forest, Support Vector

Machine (SVM), Decision Tree, K-Nearest Neighbor (KNN), and K-Means Clustering, are employed to generate crop and fertilizer recommendations. The system primarily relies on key soil parameters such as Nitrogen, Phosphorus, Potassium, and pH levels for decision-making. While the model demonstrates promising results, the authors highlight that variations in regional soil and climatic conditions may influence its performance when applied to different geographical areas.

**[Paper 7]** The article entitled “Rapid In-Field Soil Analysis of Plant-Available Nutrients and pH for Precision Agriculture” by Elena Najdenko et al., (2024), highlights the various approaches used in soil analysis within precision agriculture, including the comparative analysis between sensor data and reference laboratory tests on soils with different nutrient compositions. The analysis methods discussed in the article include Partial Least Squares Regression, Neural Networks, Data Mining, and AI-assisted automated soil property analysis techniques. According to the findings of the study, invasively used soil sensors may generate very localized data, which necessitates repetitive sampling to account for the variation in the field.

**[Paper 8]** “Survey on Soil Micro-Nutrients Analysis and Crop Recommendation System in Smart Agriculture”, authored by P. Abinaya et al., (2024), is an extensive study that covers the role

of data analytics and machine learning techniques in soil micro-nutrient analysis and crop recommendation in agriculture. This study uses data collected by government bodies like NBSSLUP, NRSC/ISRO, OGD Platform India, NSDI India, and State Agricultural Universities. Machine learning methods such as Random Forest, SVM, KNN, and K-Means Clustering have been used in this study to interpret soil data. Some of the challenges discussed in this study include sensor accuracy issues and economic limitations for small scale farmers.

### III PROPOSED SYSTEM

#### 3.1 System Overview:

The suggested system is a combined one that incorporates machine learning and the IoT technology for analyzing the nutrients in the soil and recommending the crops and fertilizers automatically. This system employs soil properties like nitrogen (N), phosphorus (P), potassium (K), pH, temperature, humidity, and rainfall for determining the crop that can be cultivated efficiently in the soil. In accordance with the selected crop and nutrient levels, the system recommends the right type of fertilizer.

#### 3.2 System Architecture:

The proposed system follows a layered architecture comprising data acquisition, data processing, and result

generation stages. Soil parameters are captured using a 7-in-1 NPK sensor interfaced with an ESP32 microcontroller. This sensor is capable of measuring key soil attributes, including Nitrogen (N), Phosphorus (P), Potassium (K), temperature, humidity, pH, and electrical conductivity (EC), enabling comprehensive soil analysis. In addition to sensor data, environmental information such as rainfall is obtained through a weather API, which provides location-specific real-time or historical data. The ESP32 microcontroller gathers the sensor readings and transmits them wirelessly via Wi-Fi to a backend system. At the backend, the collected data undergoes preprocessing steps such as noise filtering and normalization to improve data quality and consistency. The processed data, along with rainfall information retrieved from the RESTful API, is then analyzed using the XGBoost machine learning model to generate crop recommendations. Based on the selected crop and the corresponding nutrient levels (N, P, K), the system further determines appropriate fertilizer suggestions to correct nutrient deficiencies.

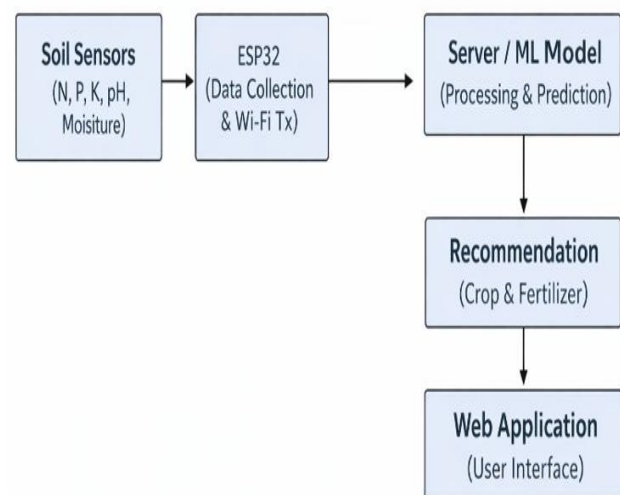


Fig.1 Block diagram of the proposed system

### IV METHODOLOGY

#### 4.1 Data Source:

This study is based on a secondary data set obtained from Kaggle that was particularly created for precision farming and recommending crops. The said data set has been acknowledged widely by the academic community for use in testing classification algorithms in soils studies.

Data set 1: Crop Recommendation

There are 2,200 samples, providing sufficient data to develop high-precision classification models. The data set shows good diversity among classes with 22 different types of crops, including food crops like rice, maize, chickpea, kidney bean, pigeon pea, moth bean, mung bean, black

gram, lentil, pomegranate, banana, mango, grape, watermelon, muskmelon, apple, orange, papaya, coconut, cotton, jute, and coffee. Each sample is described based on seven significant environmental and chemical factors affecting plant growth, namely soil nutrients (nitrogen, phosphorus, and potassium ratios), air temperature (Celsius), humidity percentage, soil pH, and average rainfall (mm).

#### Data Set 2: Data Core

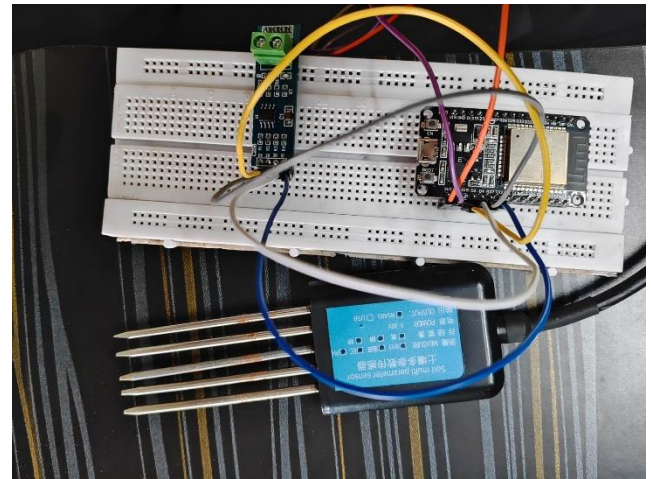
dataset that contains 8,000 samples, especially designed for automated fertilizers recommendation. The dataset comprises eight different attributes, which include environment-related, Phosphorous, Potassium, temperature, moisture along with categorical attributes like soil type (which can be sandy, loamy, black, red, and clayey) and crop type (consisting of eleven different types such as maize, sugarcane, cotton, and paddy). Most importantly, the dataset also includes soil nutrients (nitrogen, phosphorus, and potassium) to determine the required fertilizer among seven different classes: Urea, DAP, MOP, NPK fertilizers, etc.,

#### 4.2 Data Acquisition:

The data collection process consists of collecting soil-based parameters through secondary data collections and direct data acquisition systems. In this case, the data used was collected from two separate Kaggle datasets. These include the crop recommendation dataset with 2200 samples and the fertilizer recommendation dataset with 8000 samples. These contain important parameters such as nitrogen, phosphorus, potassium, temperature, humidity, pH levels, and rainfall. Apart from being trained on data collected from datasets, the proposed system will also provide support to the data acquisition process through Internet of Things (IoT) based sensors. The soil parameters including moisture content, pH levels, and NPK levels can be determined using the sensors.

#### 4.3 Sensor Interfacing:

The hardware part of the system uses ESP32 as the main controller that processes data from various sensors. Such soil sensors as the NPK sensor, pH sensor, and moisture level sensor are connected to the controller through input pins. These sensors generate electrical impulses related to specific soil parameters that are processed and digitized by the ESP32 controller. Adequate power source (either 5V or 12V, according to the sensor's requirements) and appropriate grounding are provided. As for the choice of ESP32, it is due to its integrated Wi-Fi module, affordable price, and efficient computing capabilities.



**Fig.2 Experimental setup using ESP32**

The figure illustrates the hardware prototype where the ESP32 interfaces with soil sensors to collect real-time parameters such as NPK, temperature, and humidity.

#### 4.4 Data Preprocessing:

Preprocessing is an important step that takes place prior to feeding the data into various machine learning algorithms. Some of the key steps involved in preprocessing are:

- Data Cleaning: Elimination of unnecessary columns and fixing any inconsistencies in the dataset.
- Dealing with Missing Data: Elimination of all the missing and null values present in the dataset or replacing them via suitable means.
- Label Encoding: Encoding categorical variables like type of soil and crops using the technique of label encoding.
- Feature Normalization: As the values across all different features are not consistent (example – rainfall value compared to nitrogen value), feature normalization is required.
- Data Splitting: Dividing the dataset into training set (80%) and testing set (20%).

#### 4.5 Model Deployment:

The recommendation system suggested in this paper implements machine learning based on a two-level framework.

##### Level 1: Crop Recommendation

For the first level, environmental parameters and nutrient values (N, P, K, temperature, humidity, pH, rainfall) will be used as input, and the output will include the best suitable crop for this soil and environmental combination. It is represented as a multi-class classification problem where the output will consist of one of the 22 crops.

##### Level 2: Fertilizer Recommendation

The fertilizer recommendation is based on nutrient deficiency analysis, where the measured N, P, and K values are compared with optimal crop-specific requirements. Based on the deficiency, suitable fertilizers such as Urea (Nitrogen source), DAP (Phosphorus source), and MOP (Potassium source) are recommended

#### 4.6 Machine learning Algorithms

**1.SVM:** Support vector machine is a type of supervised learning algorithm that is used for classification as well as regression problems. The idea behind SVM is to find the hyperplane that best divides the distinct classes with the largest possible margin between them.

**2.KNN:** KNN is an easy-to-understand supervised machine learning algorithm that is employed in both classification and regression tasks. It involves calculating distances between the input and other data and determining the majority among the \*k\* closest neighbors. This means the number of \*k\*, like 3, 5, or 7, influences the precision rate. It is easy to execute but takes time in large databases

**3.XGBoost:** XGBoost is a highly sophisticated machine learning model that uses boosting. This method constructs its models sequentially, meaning that each successive model rectifies the mistakes made by the preceding model. XGBoost is renowned for its high level of precision, efficiency, and effectiveness.

**4.Random Forest:** The Random Forest is a type of machine learning algorithm where many decision trees contribute to the generation of one output or prediction. In this algorithm, each tree contributes its result, which helps generate the final result.

#### 4.7 Performance Matrices:

##### 1.Accuracy:

Definition: Measures how often the model correctly predicts the output.

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+TN+FP+FN)}$$

##### 2.Precision:

Definition: Shows how many of the predicted positives are actually correct.

$$\text{Precision} = \frac{TP}{(TP+FP)}$$

##### 3. Recall:

Definition: Measures how many actual positives were correctly identified.

$$\text{Recall} = \frac{TP}{(TP+FN)}$$

#### 4.F1 Score:

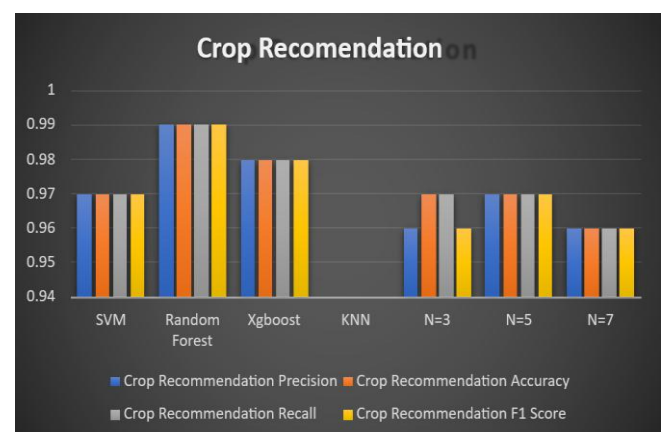
Definition: Balance between precision and recall.

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

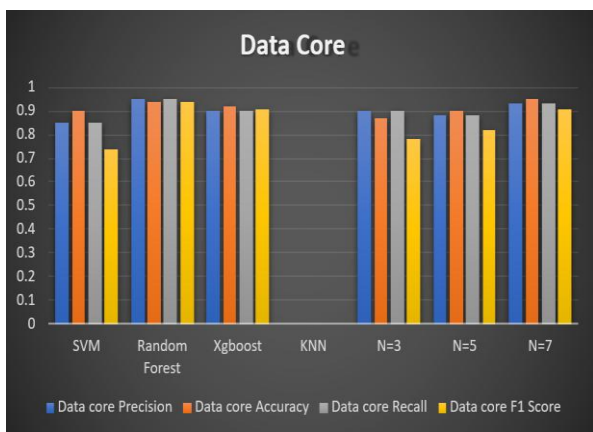
### V RESULTS AND DISCUSSION

#### 5.1 Performance evaluation of Dataset.

Crop Recommendation				
Machine Learning Algorithm	Precision	Accuracy	Recall	F1-Score
SVM	0.97	0.97	0.97	0.97
Random Forest	0.93	0.94	0.95	0.92
XGBoost	0.98	0.98	0.98	0.98
KNN				
N=3	0.96	0.97	0.97	0.96
N=5	0.97	0.97	0.97	0.97
N=7	0.96	0.96	0.96	0.96

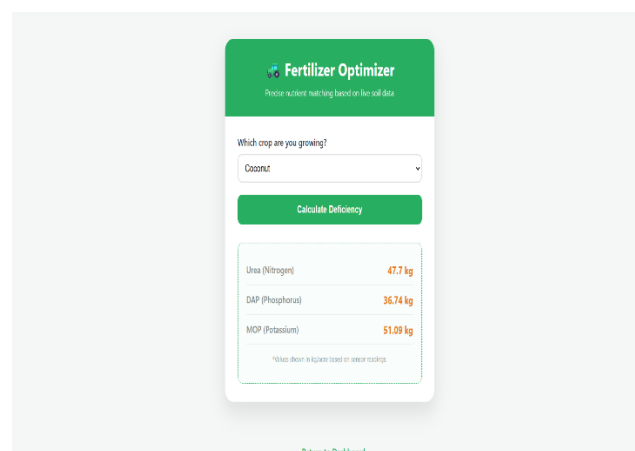
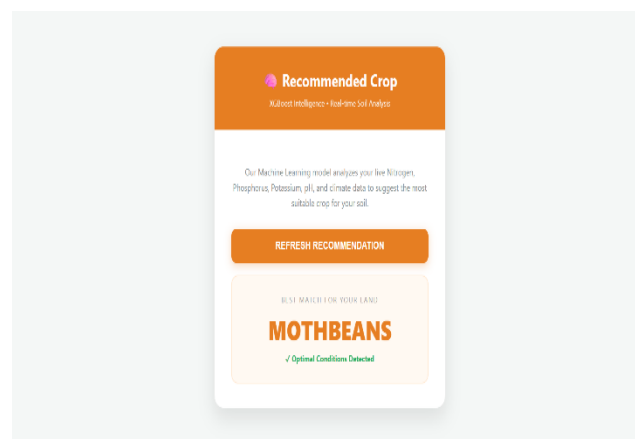
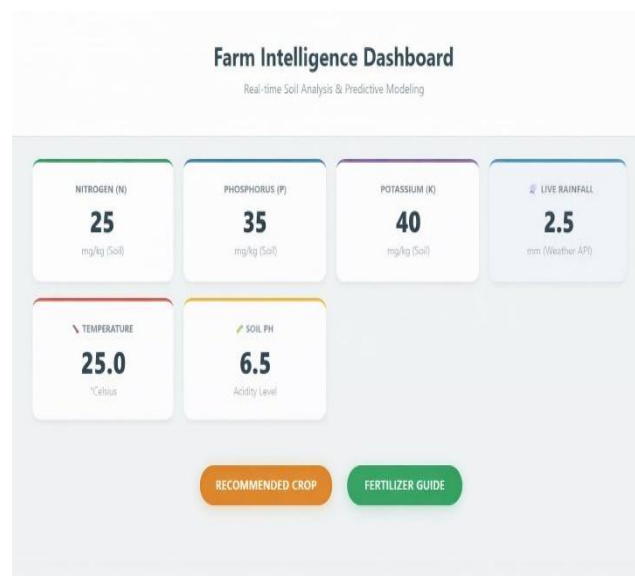


Data Core				
Machine Learning Algorithm	Precision	Accuracy	Recall	F1-Score
SVM	0.85	0.9	0.85	0.74
Random Forest	0.93	0.94	0.95	0.94
XGBoost	0.9	0.92	0.9	0.91
KNN				
N=3	0.9	0.87	0.9	0.78
N=5	0.88	0.9	0.88	0.82
N=7	0.93	0.95	0.93	0.91



For this project, we tested four machine learning algorithms—SVM, Random Forest, XGBoost and KNN—on the datasets. The models were evaluated using accuracy, precision, recall and F1 score. Random Forest and XGBoost consistently outperformed SVM and KNN across all datasets, showing reliability and robustness. Random Forest achieved 0.94 accuracy and 0.95 precision and recall, while XGBoost achieved 0.92 accuracy and 0.90 precision and recall. Random Forest reduces overfitting, handles noisy sensor data efficiently, and works well with mixed data types such as soil pH, moisture, and NPK levels. XGBoost is optimized for speed and accuracy and can manage complex relationships between soil parameters. Based on their consistent high performance and balanced F1 scores, Random Forest and XGBoost are the best algorithms for this project.

## 5.2 Snapshot of Result



## VI.CONCLUSION

In this paper, an analysis and recommendation system based on soil nutrients using machine learning is proposed. Using two different data sets for the purpose of crop recommendation and fertilizer recommendation, respectively, the study establishes the correlation between soil nutrients and their ideal use in agricultural activities. The use of four different types of machine learning algorithms, such as Random Forest, XGBoost, Support Vector Machine, and K-Nearest Neighbors helped in comparing them in terms of accuracy, precision, recall, and F1-score, whereby Random Forest and XGBoost performed better. XGBoost achieved 98% accuracy. The two-phase approach used in this system, which is crop recommendation first and then fertilizer recommendation, makes the approach efficient and effective. It is clear from the results obtained in the study that data science and machine learning have the potential to replace the current soil testing process by saving both time and money, and reducing human error. Finally, it is worth mentioning that the proposed system can easily be scaled up by integrating IoT sensors, like NPK sensor using ESP32 technology. The integration of real-time IoT data with machine learning and external weather APIs makes the system highly adaptable and suitable for precision agriculture applications.”

## REFERENCES

- [1] Abdullahi, M. O., Jimale, A. D., Ahmed, Y. A., & Nageye, A. Y. (2024). Revolutionizing Somali agriculture: harnessing machine learning and IoT for optimal crop recommendations. *Discover Applied Sciences*, 6(3).
- [2] Babu S., Madhusudanan, S., Sathiyarayanan, M., Mortka, M. Z., Szymański, J., & Rahul, R. (2025). Soil mapping for farming productivity: internet of things (IoT) based sustainable agriculture. *Microsystem Technologies: Sensors, Actuators, Systems Integration*, 31(3), 679–694.
- [3] Dutta, M., Gupta, D., Juneja, S., Shah, A., Shaikh, A., Shukla, V., & Kumar, M. (2023). Boosting of fruit choices using machine learning-based pomological recommendation system. *SN Applied Sciences*, 5(9).
- [4] Folorunso, O., Ojo, O., Busari, M., Adebayo, M., Joshua, A., Folorunso, D., Ugwunna, C. O., Olabanjo, O., & Olabanjo, O. (2023). Exploring machine learning models for soil nutrient properties prediction: A systematic review. *Big Data and Cognitive Computing*, 7(2), 113.
- [5] Kollu, P. K., Bangare, M. L., Hari Prasad, P. V., Bangare, P. M., Rane, K. P., AriasGonzález, J. L., Lalar, S., & Shabaz, M. (2023). Internet of things driven multilinear regression technique for fertilizer recommendation for precision agriculture. *SN Applied Sciences*, 5(10).
- [6] Musanase, C., Vodacek, A., Hanyurwimfura, D., Uwitonze, A., & Kabandana, I. (2023). Data-driven analysis and machine learning-based crop and fertilizer recommendation system for revolutionizing farming practices. *Agriculture*, 13(11), 2141.
- [7] Najdenko, E., Lorenz, F., Dittert, K., & Olf, H.-W. (2024). Rapid in-field soil analysis of plant-available nutrients and pH for precision agriculture—a review. *Precision Agriculture*, 25(6), 3189–3218.
- [8] P. Abinaya, A. Faaiza, S. Tharshni, G. Vishali, & M. Aravindan. (2024). Survey on soil micro-nutrients analysis and crop

recommendation system in smart agriculture. *International Research Journal on Advanced Engineering Hub (IRJAEH)*, 2(03), 457–463.

- [9] <https://www.kaggle.com/datasets/aksahaha/crop-recommendation>.
- [10] <https://www.kaggle.com/datasets/shankarpriya2913/crop-and-soil-dataset>