Multi Crop Recommendation Using Machine Learning with Internet of Things (IoT)

Md Fahad Mir
2044851032

Mohammad Mahbub Hasan
2044851043

N. M Shomirul Hayder Shourav
2044851052

MD Alamin Islam
2044851053

In partial fulfilment of the requirement for the award of Bachelor of Science (B. Sc.) in Computer Science and Engineering (CSE)

Supervised By

Al-Imtiaz

Assistant Professor and Head
Department Computer Science and Engineering

University of Information Technology and Sciences (UITS)

June 2024

Declaration

This to certify that the thesis work entitled "**Multi Crop Recommendation Using Machine Learning with Internet of Things (IoT)**" has been carried out by **Md Fahad Mir, Mohammad Mahbub Hasan, N. M Shomirul Hayder Shourav, MD Alamin Islam** in the Department of Computer Science and Engineering (CSE), University of Information Technology and Sciences (UITS), Dhaka, Bangladesh. The above thesis work or any part of this work has not been submitted anywhere for the award of any degree or diploma.

Supervisor:

……………………………………
Al-Imtiaz,
Assistant Professor and Head,
Department of CSE, UITS

Candidate:

……………………………………
Md Fahad Mir
2044851032

……………………………………
Mohammad Mahbub Hasan
2044851043

……………………………………
N. M Shomirul Hayder Shourav
2044851052

……………………………………
MD Alamin Islam
2044851053

Approval

This to certify that the thesis work submitted by **Md Fahad Mir, Mohammad Mahbub Hasan, N. M Shomirul Hayder Shourav, MD Alamin Islam** entitled **"Multi Crop Recommendation Using Machine Learning with Internet of Things (IoT)"** has been approved by the Thesis Review Committee partial fulfillment of the requirements for the degree of Bachelor of Science in Computer Science and Engineering (CSE) in the Department of Computer Science and Engineering (CSE), University of Information Technology and Sciences (UITS), Dhaka, Bangladesh in June, **2024.**

Capstone Project Review Committee

1. **……………………..** **Convener & Supervisor**
   Al Imtiaz
   Assistant Professor & Head,
   Department of Computer Science and Engineering (CSE)
   University of Information Technology and Sciences (UITS)

2. **……………………..** **Member**
   A. S. M. Shafi
   Assistant Professor
   Department of Computer Science and Engineering (CSE)
   University of Information Technology and Sciences (UITS)

3. **……………………..** **Member**
   Ms. Sultana Rokeya Nahar
   Assistant Professor
   Department of Computer Science and Engineering (CSE)
   University of Information Technology and Sciences (UITS)

4. **……………………..** **Member Secretary**
   Fatema Tuj Tarannom Esty
   Lecturer
   Department of Computer Science and Engineering (CSE)
   University of Information Technology and Sciences (UITS)

Abstract

The agricultural sector plays a pivotal role in sustaining human life and ensuring food security. However, inefficient water distribution and suboptimal crop selection remain persistent challenges, particularly in regions like Bangladesh where agricultural practices are deeply intertwined with the socio-economic fabric. This project addresses these challenges through a novel fusion of Machine Learning (ML) techniques and Internet of Things (IoT) technology.

The inspiration for this project "**Multi Crop Recommendation Using Machine Learning with Internet Of Things (IoT)**" stems from the pressing need to enhance agricultural productivity while conserving water resources. Traditional farming methods often rely on manual intervention and lack precision in water management, leading to water wastage and decreased yields. Moreover, suboptimal crop selection further exacerbates these issues, contributing to decreased profitability for farmers. Through this interdisciplinary approach, our project strives to revolutionize agricultural practices by promoting sustainability, increasing efficiency, and empowering farmers with actionable insights. By addressing the fundamental issues of water management and crop selection, we envision a future where agriculture is not only productive but also environmentally conscious and economically viable.

Keywords: Internet of Things; sensors; soil nutrients; pH value; precision agriculture; crop recommendation; machine learning; decision tree; naive bayes; SVM; logistic regression; random forest.

Preface

This B.Sc. thesis is outlined based on the results obtained from the laboratory experiment. This is carried out in the Department of Computer Science and Engineering (CSE), Faculty of Engineering, at University of Information Technology and Sciences (UITS), Dhaka, Bangladesh. This thesis includes 6 chapters which are briefed as follows:

Chapter-1
Basic introduction to motivation, objective and significance as well as scope of the study.

Chapter-2
Literature review literature review, machine learning & internet of things & its relation to agriculture.

Chapter-3
Methodology & overall view of the whole project

Chapter-4
Machine Learning execution for multi crop prediction

Chapter-5
Implementation of Internet of Things (IoT) for automatic water & fertilizer distribution system

Chapter-6
The website for the users where they would be shown the crop recommendation

Chapter-7
Result Analysis and Comparison detail description of the propose system with proper calculation.

Chapter-8
Conclusion and Future work provide conclusions, references.

Contents

List of Figure

List of Table

Chapter 1

Introduction

The agricultural sector stands as a cornerstone of human civilization, providing sustenance and livelihood to billions worldwide. However, its sustainability faces formidable challenges, including water scarcity, inefficient resource utilization, and fluctuating market demands. In Bangladesh, where agriculture serves as a linchpin of the economy and supports a vast population.

A multi-crop model simultaneously predicts yield for multiple crops and can exploit soil-climate-landscape data for crops that can grow in the same regions [1]. This project embarks on a journey to tackle key issues in agriculture through a synergistic blend of Machine Learning (ML) and Internet of Things (IoT) technologies. By harnessing the power of data-driven insights and real-time monitoring, we aim to revolutionize traditional farming practices and usher in an era of efficiency, sustainability, and resilience. Central to our approach is the utilization of various ML algorithms for crop recommendation, including Random Forest, Support Vector Machine (SVM), Decision Tree, Naive Bayes, and Logistic Regression. These algorithms analyze historical data on crop performance, soil characteristics, climate conditions, and other relevant factors to provide personalized recommendations to farmers. Notably, our experiments have demonstrated the efficacy of Random Forest, yielding the highest accuracy among the tested algorithms, thereby serving as a cornerstone of our recommendation system. Complementing our ML-based crop recommendation system is the integration of IoT devices tailored for precision water management. Soil moisture sensors deployed across the field enable real-time monitoring of soil conditions, allowing for precise irrigation scheduling and water allocation. IoT-enabled soil nutrient classification and crop recommendation (IoTSNA-CR) model are being used to recommend Crops. As a soil temperature sensor, a soil moisture sensor, a water level indicator, a pH sensor, a GPS sensor, and a color sensor, along with an Arduino UNO board [2]. This data-driven approach minimizes water wastage, optimizes resource utilization, and mitigates the risk of over- or under-watering, thereby promoting sustainable agricultural practices. Furthermore, our IoT infrastructure encompasses solenoid valves and water flow sensors, facilitating automated water distribution based on soil moisture levels. This proactive approach not only reduces manual intervention but also ensures timely and targeted irrigation, thereby enhancing crop yields and conserving precious water resources.

Smart Irrigation and Fertilizer Management System: Enhancing Crop Health with Sensor-based Monitoring and Precision," employs IoT devices and an array of sensors—water flow meter, soil moisture sensor, water pump controller, water level indicator, and Arduino UNO ESP8266 Wi-Fi module. This innovative system operates by detecting soil moisture, intelligently deciding watering periods, and managing the water pump's activation based on soil conditions. It further transmits water flow sensor data to a web server for graph generation, facilitating detailed analysis to optimize irrigation practices [3].

Mainly we are doing the upgrade of this work. In the paper [3] the irrigation and fertilizer management system are just monitored. But we are aiming to distribute water & fertilizer automatically.

From harvesting to marketing, various crops – amounting to Tk11,000 crore – are wasted every year during this process[https://www.tbsnews.net/bangladesh/crops-worth-tk11000cr-go-waste-annually-bbs]. Of the total, paddy alone faces more than 28 lakh tons in damages whose market value is about Tk7,000 crore[https://www.tbsnews.net/bangladesh/crops-worth-tk11000cr-go-waste-annually-bbs]. So, we want to decrease this wastage of resources. By implementing machine learning & IoT we are ensuring that the wastage of crops would be minimum. At a certain point there would be no crop waste.

1.1 Motivation

In a developing country like Bangladesh, at first, we have to think about efficiency. And efficiency in the food sector is one of the crucial things. Every year a vast number of crops and money are wasted just because the farmers don't have the latest technology in their hand. And the sad part is they don't even know that. So, our motivation comes from here. We have made an IoT device which is inspired from an existed device but is very costly. And we are going to make a miniature version of that device which would cost a lot less than the original device. And with the passage of time out device would cost less & less because we are going to make the device much more compact.

1.2 Objectives

Farming is the core of human civilization. And it serves to the greater good of the humanity. Our project is a small contribution into that greater good with these objectives:

- **Resource Optimization:** To precisely assess soil moisture, nutrient levels, and environmental conditions, optimizing the distribution of water and fertilizers for each crop.
- **Environmental Sustainability:** To minimize the environmental impact of agriculture by reducing water wastage and chemical runoff, promoting sustainable farming practices.
- **Crop Health Improvement:** To enhance overall crop health by continuously monitoring and responding to specific needs, ensuring optimal growth conditions and maximizing yields.
- **Cost Efficiency:** To reduce input costs for farmers by minimizing the use of resources, ultimately improving economic sustainability in agriculture.
- **Technology Integration:** To integrate cutting-edge technologies, such as IoT and sensor systems, into farming practices to automate and streamline irrigation and fertilizer management.

- **Data-driven Decision Making:** To enable farmers to make informed decisions based on real-time data, improving the efficiency and effectiveness of agricultural operations.
- **Global Food Security:** To contribute global food security by increasing agricultural productivity and efficiency, ensuring a stable and sufficient food supply for the growing world population.

1.3 Challenges

Implementing a project that combines Machine Learning (ML) and Internet of Things (IoT) for precision agriculture in a developing country like Bangladesh presents several significant challenges. Here are some key challenges and considerations:

- **Technical Literacy and Adoption:** The majority of farmers may lack familiarity with IoT devices and ML-based systems. Comprehensive training programs and continuous support are essential to ensure they can effectively use these technologies.
- **Infrastructure and Connectivity:** Reliable internet connectivity is a prerequisite for IoT devices to function correctly. Many rural areas in Bangladesh may have limited or no access to stable internet, which could hinder the deployment of IoT solutions.
- **Economic Constraints:** The cost of IoT devices and setting up the required infrastructure might be prohibitive for small-scale farmers. Financial support, subsidies, or micro-financing options could be necessary to facilitate adoption.
- **Data Management and Security:** Ensuring the privacy and security of the data collected by IoT devices is paramount. Farmers need assurance that their data is protected and used responsibly.
- **Local Adaptation and Customization**: Agricultural practices vary significantly across different regions. The ML models and IoT systems must be adaptable to local conditions, crop types, and farming practices.
- **Sustainability and Scalability:** Ensuring that the project can be sustained in the long term without continuous external support is crucial. This includes building local capacity and fostering ownership among farmers.

Addressing these challenges requires a multi-faceted approach that combines technological innovation with community engagement, policy support, and continuous learning. By proactively tackling these issues, the project can enhance its potential for success and significantly contribute to the sustainability and resilience of agriculture in Bangladesh. As we embark on this interdisciplinary journey, our goal is not merely to address the immediate challenges facing Indian agriculture but to lay the foundation for a resilient and sustainable future. By empowering farmers with actionable insights and cutting-edge technologies, we aspire to foster a paradigm shift towards precision agriculture, wherein data-driven decisions drive productivity, profitability, and environmental stewardship.

Chapter 2

Literature Review

A thorough literature review was conducted to identify efficient technologies for the design and implementation of a multi crop with the best machine learning model & water distribution system using IoT in agriculture. Several noteworthy publications have contributed to this field:

A data-driven multi-layer perceptron that simultaneously predicts the land suitability of several crops in Canada in this paper. A multi-crop model simultaneously predicts yield for multiple crops and can exploit soil-climate-landscape data for crops that can grow in the same regions [1].

This system operates by detecting soil moisture, intelligently deciding watering periods, and managing the water pump's activation based on soil conditions. It further transmits water flow sensor data to a web server for graph generation, facilitating detailed analysis to optimize irrigation practices [2].

In this paper they have proposed the design and the experiment of a smart farming system based on an intelligent platform which enables prediction capabilities using artificial intelligence (AI) Techniques [3].

A fast-learning classification methodology known as extreme learning method (ELM) is trained using the data to identify the micronutrients present in the soil. Activation functions such as hard limit, triangular basis, hyperbolic tangent, sine-squared, and Gaussian radial bases are used to optimize the methodology [4].

In this paper they have proposed a novel framework for smart framing with the federation of Machine learning prediction systems and real-time data collections by sensors, Raspberry pi3 and IOT technology [5].

In this paper machine learning techniques such as multiple linear regression, support vector regression and recurrent neural networks for prediction of soil moisture for 1 day, 2 days and 7 days ahead. The performance of the predictor is evaluated on the basis of mean squared error (MSE) and coefficient of determination (R^2) [6].

In this paper an IoT-enabled soil nutrient classification and crop recommendation (IoTSNA-CR) model are being used to recommend Crops. As a soil temperature sensor, a soil moisture sensor, a water level indicator, a pH sensor, a GPS sensor, and a color sensor, along with an Arduino UNO board [7].

The proposed IoT system is composed of pH sensors, Humidity and temperature sensors, Soil moisture sensors, soil nutrient sensors (NPK) probes, microcontroller/microprocessor equipped with Wi-Fi and Cloud storage [8].

The main objective of this case study is to end up with a model which predicts the high yield crop and precision agriculture. The proposed system modeling incorporates the trending technology, IoT, and Agriculture needy measurements [9].

The proposed model is a smart irrigation system which predicts the water requirement for a crop, using a machine learning algorithm. This system consists of a temperature, humidity and moisture sensor, deployed in an agricultural field, sends data through a microprocessor, developing an IoT device with cloud. Decision tree algorithm, an efficient machine learning algorithm is applied on the data sensed from the field in to predict results efficiently [10].

This study is aimed at making an efficient and accurate system using IoT devices and machine learning (ML) algorithms that can correctly select a crop for maximal yield. As a contribution, they have proposed an ML-based model, Smart Crop Selection (SCS), which is based on data of metrological and soil factors [11].

The paper aims to discover the best model for crop prediction, which can help farmers decide the type of crop to grow based on the climatic conditions and nutrients present in the soil [12].

Sl. No	Key Findings	Model	IoT Devices	Accuracy
1	A data-driven multi-layer perceptron that simultaneously predicts the land suitability of several crops in Canada in this paper	Semi-supervised learning with a multivariate multilayer perceptron, k fold cross validation	A data-driven multi-layer perceptron that simultaneously predicts the land suitability of several crops in Canada in this paper	
2	In this paper they have proposed the design and the experiment of a smart farming system based on an intelligent platform which enables prediction capabilities using artificial intelligence (AI) Techniques.	EDGE, LSTM-based models and GRU-based models	DHT22 Sensor, Water Flow Sensor, Soil Moisture Sensor, Mega Arduino	
3	This system operates by detecting soil moisture, intelligently deciding watering periods, and managing the water pump's activation based on soil conditions. It further transmits water flow sensor data to a web server for graph generation, facilitating detailed analysis to optimize irrigation practices		Soil Moisture Sensor, Water Flow sensor, Servo Motor	
4	A fast-learning classification methodology known as extreme learning method (ELM) is trained using the data to identify the micronutrients present in the soil	ELM	pH Sensor, Soil Moisture Sensor, Arduino UNO	
5	In this paper they have proposed a novel framework for smart framing with the federation of Machine learning prediction systems and real-time data collections by sensors, Raspberry pi3 and IOT technology		DHT11, Raspberry pi 3	

6	In this paper machine learning techniques such as multiple linear regression, support vector regression and recurrent neural networks for prediction of soil moisture for 1 day, 2 days and 7 days ahead	MLP, SVM, RNN	(ICICCT) conference	97.5%
7	In this paper an IoT-enabled soil nutrient classification and crop recommendation (IoTSNA-CR) model are being used to recommend Crops	MSVM-DAG-FFO, SVM, SVM Kernel	Soil Moisture Sensor, pH Sensor, Temperature Sensor, Arduino UNO, ESP 8266	97.3%
8	The proposed IoT system is composed of pH sensors, Humidity and temperature sensors, Soil moisture sensors, soil nutrient sensors (NPK) probes, microcontroller/microprocessor equipped with Wi-Fi and Cloud storage.	SVM, Decision Tree Algorithm	pH Sensor, Temperature Sensor, Soil Moisture Sensor	94%
9	The main objective of this case study is to end up with a model which predicts the high yield crop and precision agriculture.	MLP, JRip, Decision Table		98.2%
10	The proposed model is a smart irrigation system which predicts the water requirement for a crop, using a machine learning algorithm	Decision Tree	Soil Moisture Sensor, DHT11/22, Raspberry pi	91.4%
11	This study is aimed at making an efficient and accurate system using IoT devices and machine learning (ML) algorithms that can correctly select a crop for maximal yield.	Decision tree, SVM, KNN, Random Forest, and Gaussian Naïve Bayes. Multiple Linear Regression (rainfall prediction)	DHT11, pH Sensor, Temperature Sensor	97%
12	The paper aims to discover the best model for crop prediction, which can help farmers decide the type of crop to grow based on the climatic conditions and nutrients present in the soil	KNN, Decision Tree, Random Forest	Journal of Physics: Conference Series	97%

Table 1: Literature Review

Chapter 3

Methodology

Our system recommends multi crop through prediction. From our website the system could recommend crop based on the NPK values, pH values & rainfall with the assistance of machine learning model of the system. This system would also recommend the crop based on the NPK values of a particular land. This would be primary crop recommendation. The user could choose wither chose to proceed with the primary recommendation or the user could proceed with the secondary crop recommendation, where the user would be able to cultivate two crops in a single field at a time.

3.1 Role of Machine Learning

Based on the selection of the user the system would provide a cultivation guideline as shown in the Figure 1.

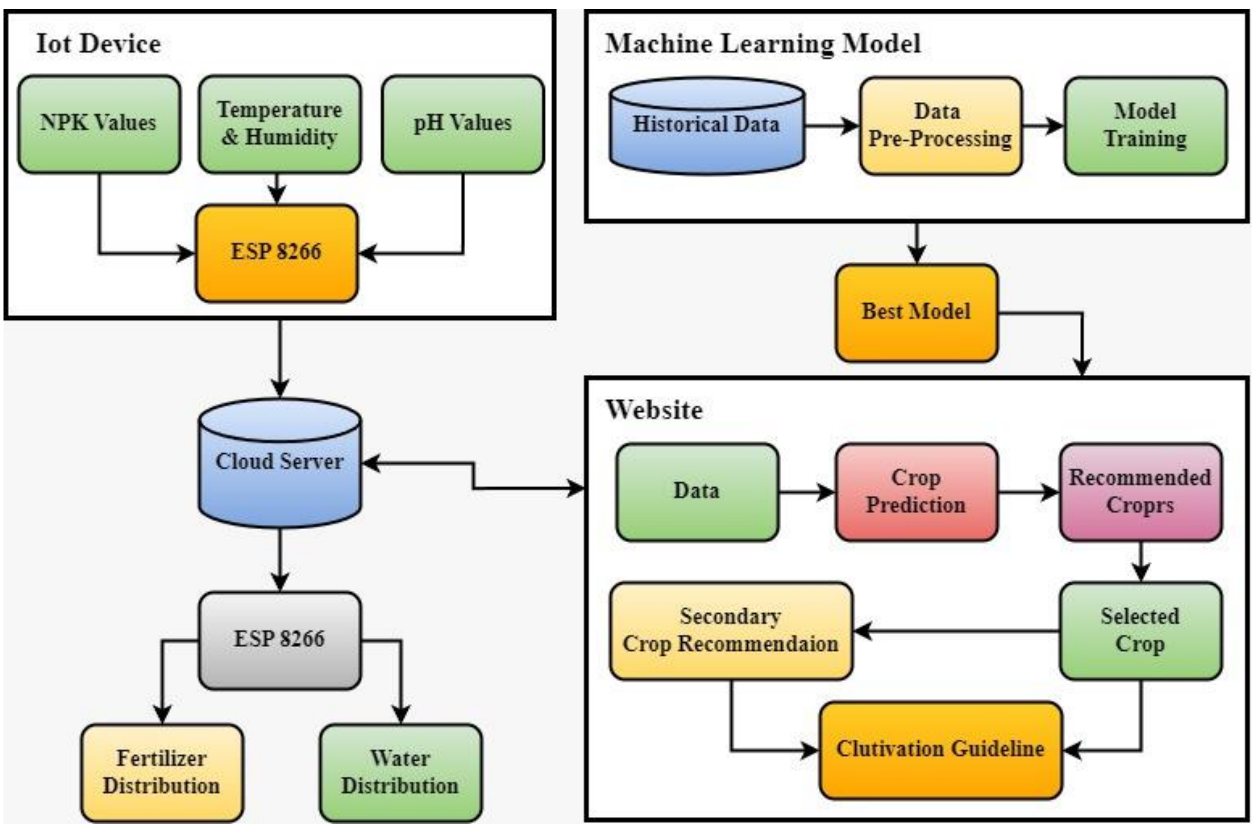


Figure 1: System Architecture

Here you can see that the IoT and Machine Learning part are divided into different parts. But both are essential. The machine learning model is being used here to recommend the crop and also it will give a guideline according to the recommended crops. Here is the detailed description about the role of Machine learning:

Data Analysis and Prediction: ML algorithms analyze historical data on NPK values, temperature, humidity, pH levels, and rainfall to identify patterns and trends. Based on historical and real-time data, ML models can predict crop yields, helping farmers make informed decisions. ML models can detect anomalies in sensor data, such as sudden drops in soil moisture or unexpected changes in pH levels, indicating potential issues that need attention.

Personalized Recommendations: ML models analyze soil and climate data to recommend the most suitable crops for a particular region. Based on soil fertility data (NPK values and pH levels), ML algorithms suggest the optimal type and amount of fertilizer needed. ML models predict the best times for irrigation by analyzing soil moisture, temperature, humidity, and rainfall data.

Optimization: ML algorithms optimize the use of water and fertilizers, reducing waste and improving efficiency. Provides real-time decision support to farmers, helping them take actions that maximize crop yield and quality.

The website is going to be the interactive part for our users. The role of the website is described here:

User Interface and Interaction: The website presents data collected from sensors and analyzed by ML models in an intuitive and user-friendly format, including graphs, charts, and dashboards. Farmers receive alerts and notifications about critical conditions, such as low soil moisture, need for fertilization, or predicted rainfall, through the website. Provides a control panel for farmers to manually override automated systems, set preferences, and monitor system performance.

Communication and Integration: The website communicates with the cloud server via API to fetch real-time data from sensors and ML models. Allows farmers to monitor and manage their fields remotely, accessing real-time data and system status from anywhere. Farmers can input additional data and feedback, which can be used to refine and improve ML models.

3.2 Role of Internet of Things

Then the cultivation guideline would be sent into the cloud server. Based on that guideline the IoT device would then distribute the water & fertilizer automatically as shown in the Figure 2.

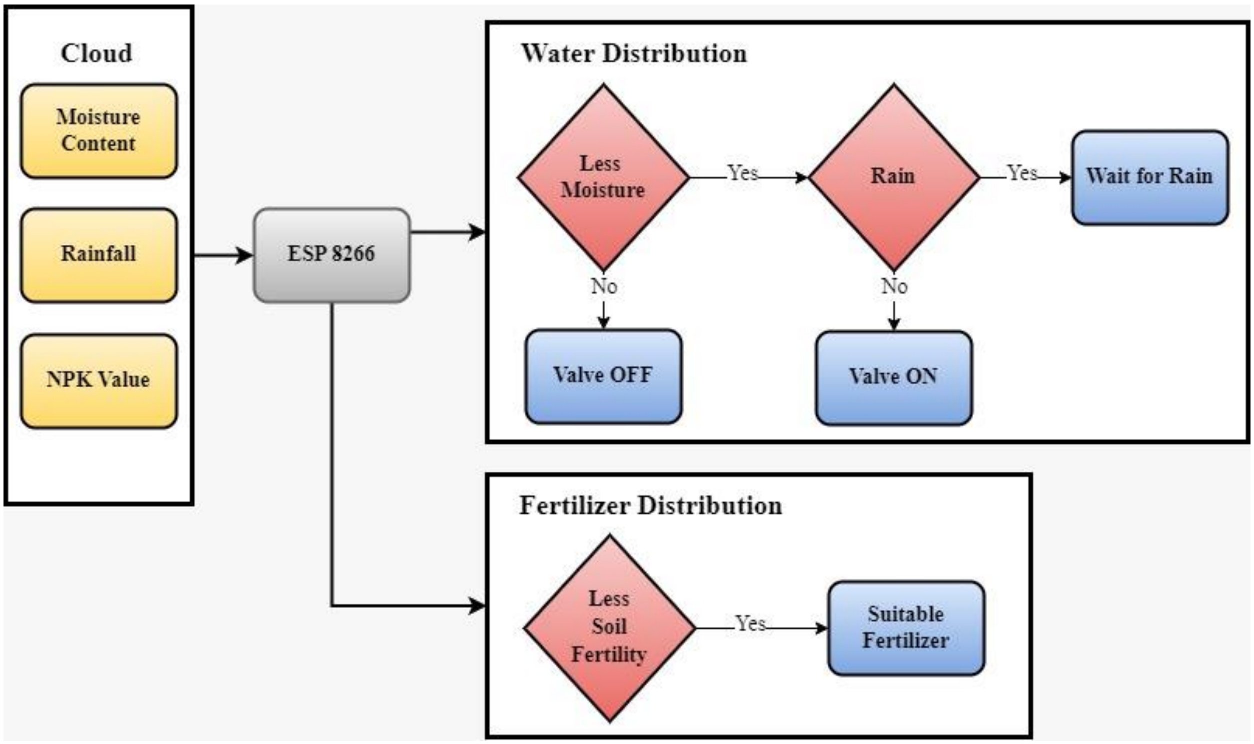


Figure 2: Automated Water & Fertilizer Distribution System

The role of IoT (Internet of Things) in the depicted system is to automate and optimize the processes of water and fertilizer distribution based on real-time data from various sensors and cloud-based information. Here's a detailed breakdown of the IoT components and their roles:

Sensors: The sensors that we are going to use here will measure the soil moisture levels & also will monitors rainfall. By taking the reading of Nitrogen (N), Phosphorus (P), and Potassium (K) in the soil. The sensors will collect data continuously and send it to the cloud for storage and processing.

Cloud: Here the data would be stored on moisture content, rainfall, and NPK values. Then the collected data would be analyzed to make decisions about water and fertilizer distribution.

ESP8266: It acts as a central hub that receives data from the cloud. It also connects to the cloud to receive sensor data and process control commands. Based on the received data, the ESP8266 makes decisions about turning on/off valves for water and determining fertilizer requirements.

Water Distribution System: The ESP8266 controls valves based on soil moisture and rainfall data. If soil moisture is low and it is not raining, the valve is turned ON to irrigate the field. If soil moisture is sufficient or it is raining, the valve is turned OFF to conserve water. Reduces manual intervention by automatically adjusting irrigation based on real-time data.

Fertilizer Distribution System: The ESP8266 decides on the need for fertilizer based on soil fertility (NPK values). If soil fertility is low, the system applies a suitable fertilizer.

Chapter 4

Multi Crop Recommendation

We are using machine learning in crop prediction to analyze vast amounts of agricultural data and identify complex patterns and relationships between environmental factors and crop performance. By leveraging machine learning algorithms, we can make accurate predictions about suitable crops for specific conditions, optimize resource allocation, and improve overall agricultural productivity while minimizing risks and resource wastage.

4.1 Model Workflow

From the collected dataset we have preprocessed all the data. Then we have split those data then implanted various machine learning models. From those models we have acquired the best model for crop recommendation from various evaluation matrices as shown in Figure 3.

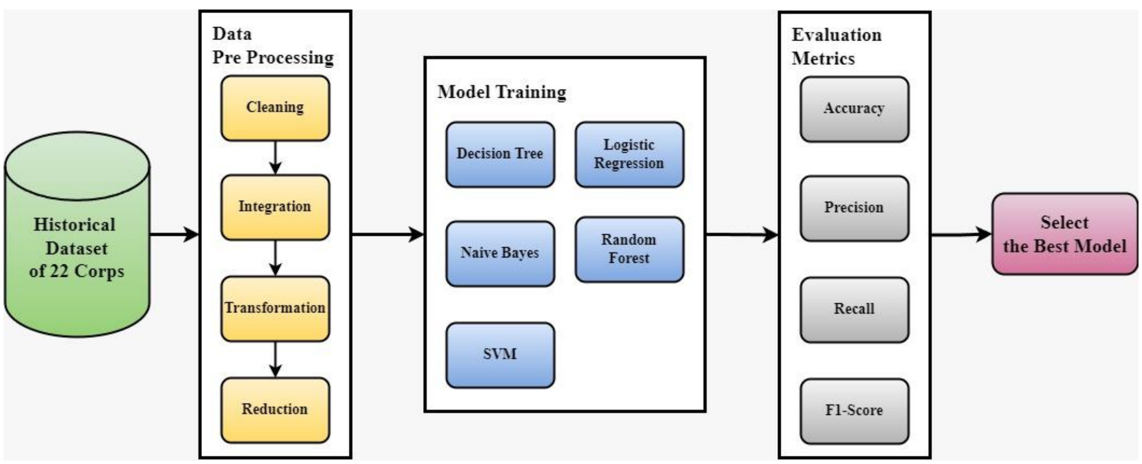


Figure 3: Workflow of crop recommendation system using machine learning

This model workflow represents a typical machine learning pipeline for crop yield prediction or crop recommendation based on historical data. Here is a detailed breakdown of each step:

Historical Dataset

The process begins with a historical dataset comprising data on 22 different crops. This dataset may include various features such as soil properties, weather conditions, crop yields, and more. Historical data is crucial for training machine learning models, as it provides the patterns and trends necessary for making predictions and recommendations.

Data Preprocessing

Preprocessing is a critical step to ensure the data is clean and suitable for model training. It includes several sub-steps:

- **Cleaning:** This involves removing or correcting erroneous data entries. Techniques such as imputation or deletion are used to deal with missing data.
- **Integration:** If the dataset is from multiple sources, it is combined into a single coherent dataset. Ensuring all data points are in a consistent format (e.g., date formats, units of measurement).
- **Transformation:** Scaling the data so that features contribute equally to the model's performance. Converting categorical data into numerical format using techniques like one-hot encoding.
- **Reduction:** Identifying and retaining the most important features that have the highest impact on the model's performance. Techniques such as Principal Component Analysis (PCA) may be used to reduce the number of features while retaining essential information.

Model Training

Different machine learning algorithms are trained on the preprocessed data to predict crop yields or recommend crops. The algorithms used include: Decision Tree, Naïve Bayes, Support Vector Machine (SVM), Logistic Regression, Random Forest

Evaluation Metrics

Once the models are trained, they are evaluated using various metrics to determine their performance:

- **Accuracy:** The proportion of correct predictions out of all predictions made.
- **Precision:** The proportion of true positive predictions out of all positive predictions made.
- **Recall:** The proportion of true positive predictions out of all actual positives.
- **F1-Score:** The harmonic means of precision and recall, providing a balance between the two.

Model Selection

Based on the evaluation metrics, the best-performing model is selected. The model with the highest accuracy, precision, recall, and F1-score, depending on the specific requirements, is chosen for deployment.

4.2 Dataset

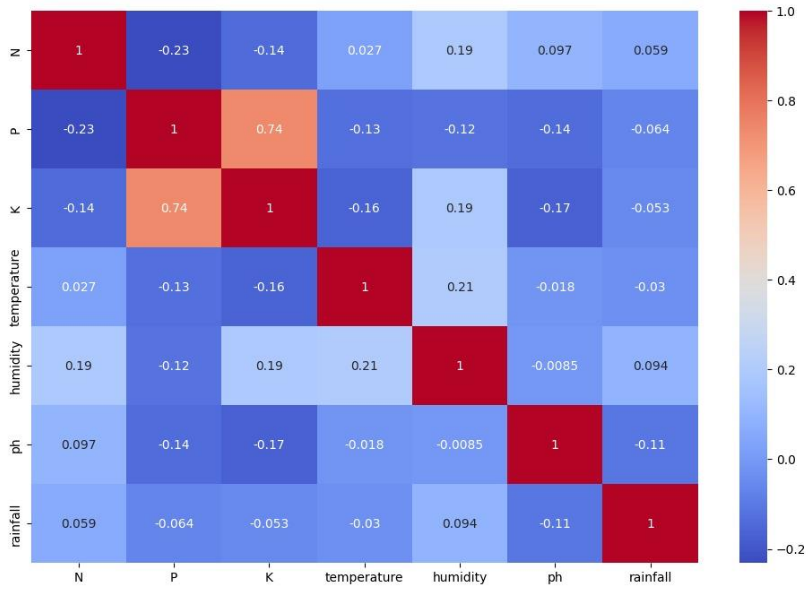
The historical crop dataset of Bangladesh is unavailable. The existing dataset are not historical, that's why we could use them to implement our model. So, we have chosen the dataset from a public repository known to be the Kaggle database. The link of the dataset is given here. [https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset].

From Table 2 we can have an overview of the whole dataset. This dataset has seven different parameters like **Nitrogen, potassium, phosphorous, pH, Rainfall, Temperature, and Humidity**. This data set contains 2200 instances for 22 different crops.

Sl. No	No. of Parameters	Instances Count	Crop Recommended
1	7	100	Rice
2	7	100	Maize
3	7	100	chickpea
4	7	100	Kidney beans
5	7	100	Pigeon peas
6	7	100	Moth beans
7	7	100	Mung bean
8	7	100	Black gram
9	7	100	lentil
10	7	100	pomegranate
11	7	100	Banana
12	7	100	Mango
13	7	100	Grapes
14	7	100	Watermelon
15	7	100	Muskmelon
16	7	100	Apple
17	7	100	Orange
18	7	100	Papaya
19	7	100	Coconut
20	7	100	Cotton
21	7	100	Jute
22	7	100	Coffee
	Total	2200	

Table 2: Overview of Crop Recommendation Dataset

Form the Figure 4 we can the correlation between the parameters of the dataset. Here we can get the idea of the relationship of the data.



	N	P	K	temperature	humidity	ph	rainfall
N	1	-0.23	-0.14	0.027	0.19	0.097	0.059
P	-0.23	1	0.74	-0.13	-0.12	-0.14	-0.064
K	-0.14	0.74	1	-0.16	0.19	-0.17	-0.053
temperature	0.027	-0.13	-0.16	1	0.21	-0.018	-0.03
humidity	0.19	-0.12	0.19	0.21	1	-0.0085	0.094
ph	0.097	-0.14	-0.17	-0.018	-0.0085	1	-0.11
rainfall	0.059	-0.064	-0.053	-0.03	0.094	-0.11	1

Figure 4: Data Correlation Heatmap

In Figure 5 we can see the percentage of the crop growth if the rain fall is greater 150 mm (rainfall > 150 mm).

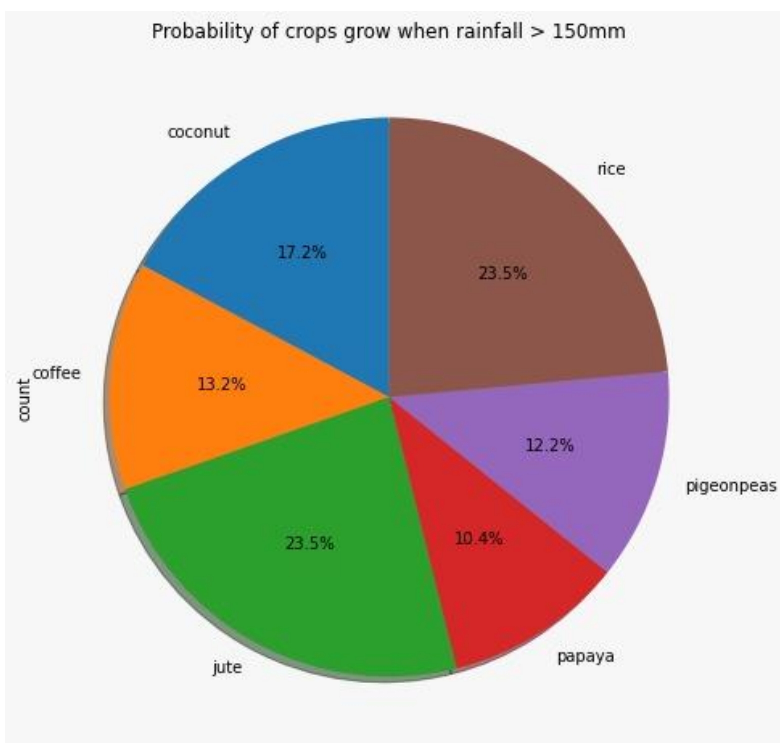


Figure 5: Probability of crops grow when Rainfall > 150mm

4.3 Feature Engineering & Data Processing

Effective feature engineering and data preprocessing were crucial steps in developing robust crop recommendation models. Prior to training the machine learning algorithms, a series of transformations were performed on the raw data to extract meaningful features and ensure data quality.

Firstly, the data was meticulously cleaned by handling missing values, either through imputation techniques or removal of incomplete instances. Outliers, which can adversely impact model performance, were identified and addressed appropriately. Furthermore, data consistency was ensured by resolving any discrepancies or inconsistencies present in the dataset.

Feature scaling was an essential preprocessing step, as machine learning algorithms can be sensitive to the scale of input features. Techniques such as min-max normalization or standardization were employed to transform the features to a common scale, preventing any individual feature from unduly influencing the learning process.

Feature selection played a vital role in identifying the most relevant and informative features that contributed significantly to the models' predictive capabilities. This process not only enhanced computational efficiency by reducing dimensionality but also mitigated the risk of overfitting and improved model interpretability. Various feature selection methods, including filter techniques (e.g., correlation analysis, mutual information), wrapper methods (e.g., recursive feature elimination), and embedded methods (e.g., regularization techniques), were utilized to select the optimal subset of features.

4.4 Selected Classifiers

The models considered in this study included Logistic Regression, Support Vector Machines (SVMs), Random Forests, Gradient Boosting, Decision Tree, and K-Nearest Neighbors (KNN). Each of these algorithms has its own strengths and weaknesses, and their performance was assessed using appropriate evaluation metrics.

Logistic Regression: It is a supervised machine learning algorithm widely used for binary classification tasks, such as identifying whether an email is spam or not and diagnosing diseases by assessing the presence or absence of specific conditions based on patient test results. The general mathematical equation for logistic regression is –

$$y = 1/(1+e\text{^}-(a+b1x1+b2x2+b3x3+...))$$

Here,

- **y** is the response variable.
- **x** is the predictor variable.
- **a** and **b** are the coefficients which are numeric constants.

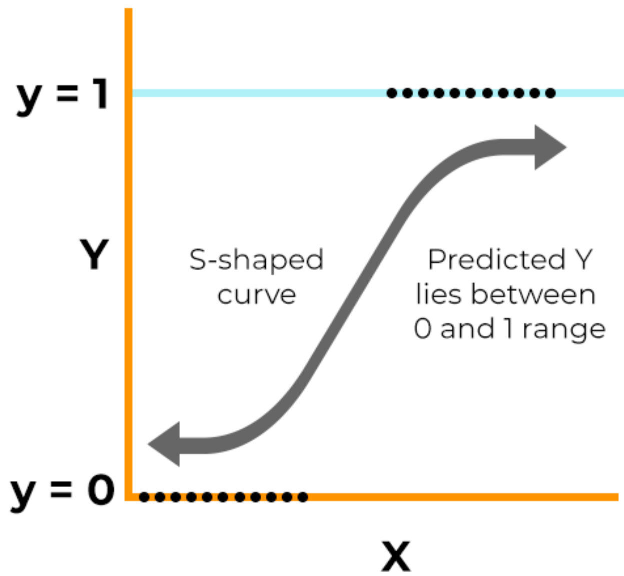


Figure 6: Linear Regression

Support Vector Machine (SVM): The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane. SVMs are particularly good at solving binary classification problems, which require classifying the elements of a data set into two groups. The data points or vectors that are the closest to the hyperplane and which affect the position of the hyperplane are termed as Support Vector. Since these vectors support the hyperplane, hence called a Support vector.

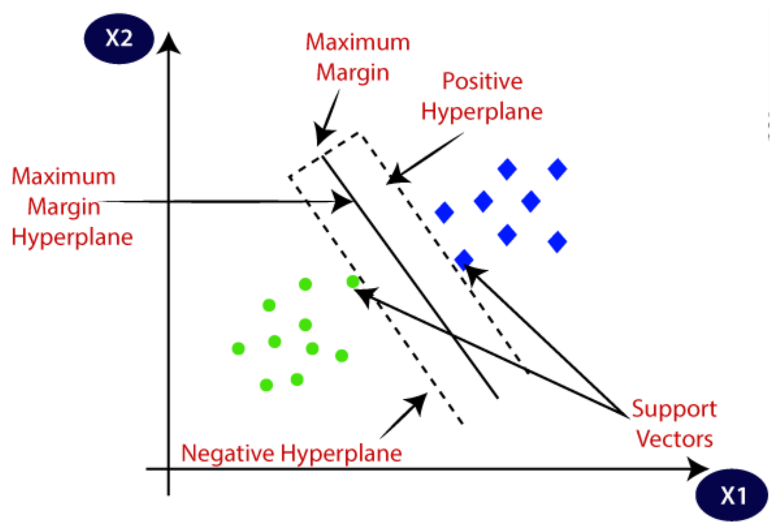


Figure 7: Support Vector Machine (SVM)

Random Forest: Random forests or random decision forests is an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees. As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

When performing Random Forests based on classification data, you should know that you are often using the Gini Index or the formula used to decide how nodes on a decision tree branch.

$$\text{Gini} = 1 - \sum_{i=1}^{c} (p_i)^2$$

You can also use entropy to determine how nodes branch in a decision tree.

$$\text{Entropy} = \sum_{i=1}^{c} -p_i * \log_2(p_i)$$

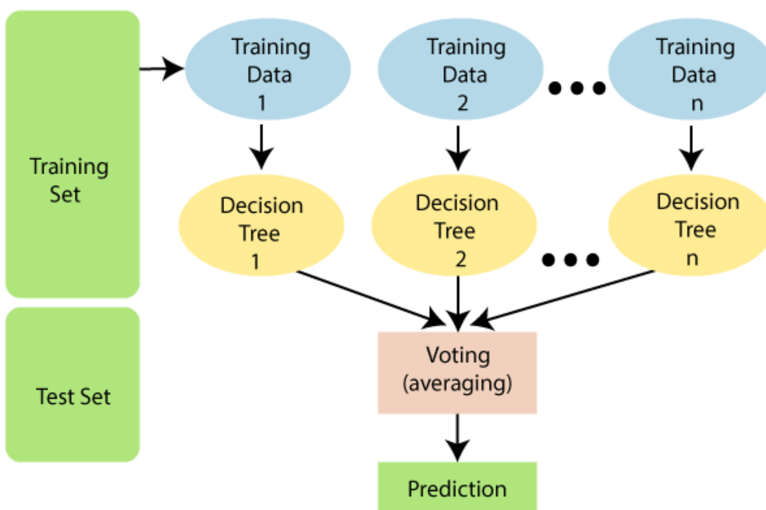


Figure 8: Random Forest

Gradient Boosting: Gradient Boosting is a functional gradient algorithm that repeatedly selects a function that leads in the direction of a weak hypothesis or negative gradient so that it can minimize a loss function.

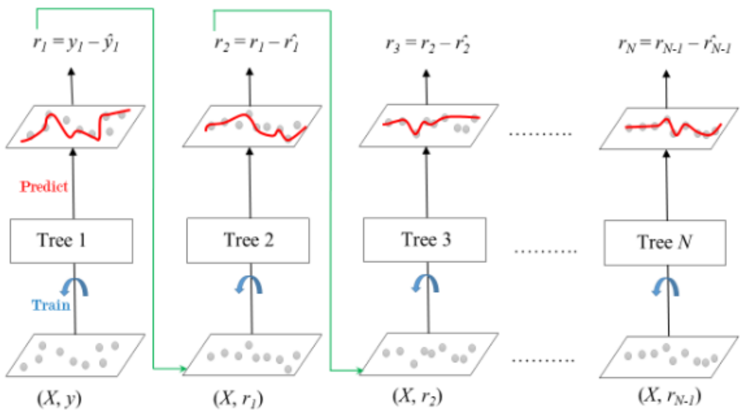


Figure 9: Gradient Boosting

The ensemble consists of M trees. Tree1 is trained using the feature matrix X and the labels y. The predictions labeled y1(hat) are used to determine the training set residual errors r1. Tree2 is then trained using the feature matrix X and the residual errors r1 of Tree1 as labels. The predicted results r1(hat) are then used to determine the residual r2. The process is repeated until all the M trees forming the ensemble are trained. There is an important parameter used in this technique known as Shrinkage. Shrinkage refers to the fact that the prediction of each tree in the ensemble is shrunk after it is multiplied by the learning rate (eta) which ranges between 0 to 1. There is a trade-off between eta and the number of estimators, decreasing learning rate needs to be compensated with increasing estimators in order to reach certain model performance. Since all trees are trained now, predictions can be made.

Let's assume X, and Y are the input and target having N samples. Our goal is to learn the function f(x) that maps the input features X to the target variables y. It is boosted trees i.e the sum of trees.

The loss function is the difference between the actual and the predicted variables.

$$L(f) = \sum_{i=1}^{N} L\big(y_i\,, f(x_i)\big)$$

We want to minimize the loss function L(f) with respect to f.

$$F_0(x) = \arg\min_{\gamma} \sum_{i=1}^{n} L(y_i, \gamma)$$

Decision Tree: Rule-based classifier which uses a simple decision table for classification. This classifier consists of the hierarchical table whose entries are broken down by the values of a pair of added features to form another table. This is analogous to dimensional stacking. In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches. It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions. A decision tree simply asks a question, and based on the answer (Yes/No), it further split the tree into subtrees.

Mathematically entropy for 1 attribute is

$$\text{Entropy (S)} = \sum_{i=1}^{c} -p_i * \log_2(p_i)$$

Mathematically entropy for multiple attributes is

$$\text{Entropy (T , X)} = \sum_{c \in x} p(c)E(c)$$

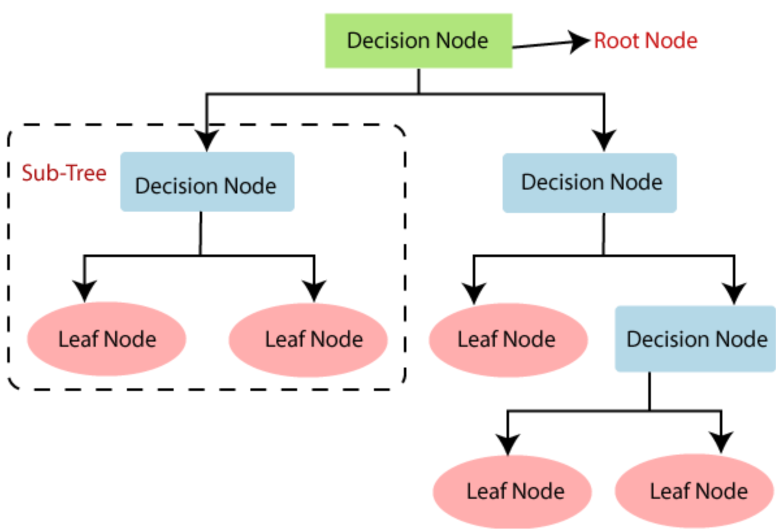


Figure 10: Decision Tree

K-Nearest Neighbors (KNN): It is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point. K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems. K-NN is a **non-parametric algorithm**, which means it does not make any assumption on underlying data. KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much similar to the new data. Suppose there are two categories, i.e., Category A and Category B, and we have a new data point x1, so this data point will lie in which of these categories. To solve this type of problem, we need a K-NN algorithm. With the help of K-NN, we can easily identify the category or class of a particular dataset. Consider the below diagram:

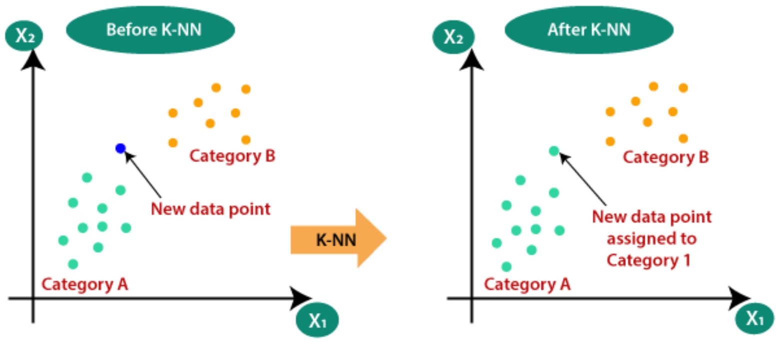


Figure 11: KNN

Suppose we have a new data point and we need to put it in the required category. Consider the below image:

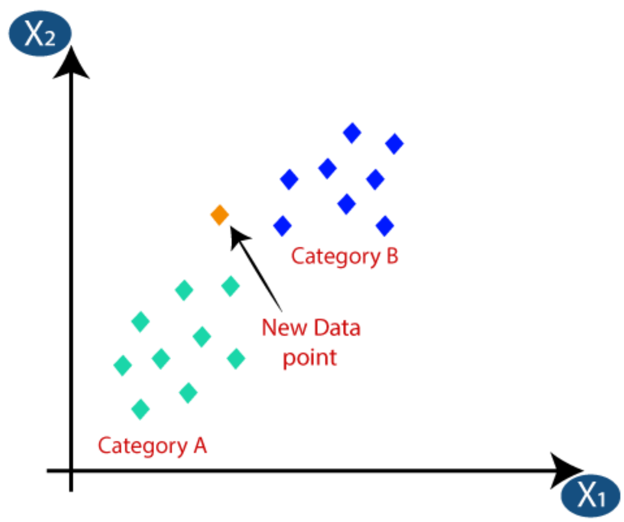


Figure 12: KNN

Firstly, we will choose the number of neighbors, so we will choose the k=5. Next, we will calculate the **Euclidean distance** between the data points. The Euclidean distance is the distance between two points, which we have already studied in geometry. It can be calculated as

$$\text{Euclidean Distance(A,B)} = \sqrt{(x_2 - x_1)^2 - (y_2 - y_1)^2}$$

By calculating the Euclidean distance we got the nearest neighbors, as three nearest neighbors in category A and two nearest neighbors in category B. Let's consider the image below:

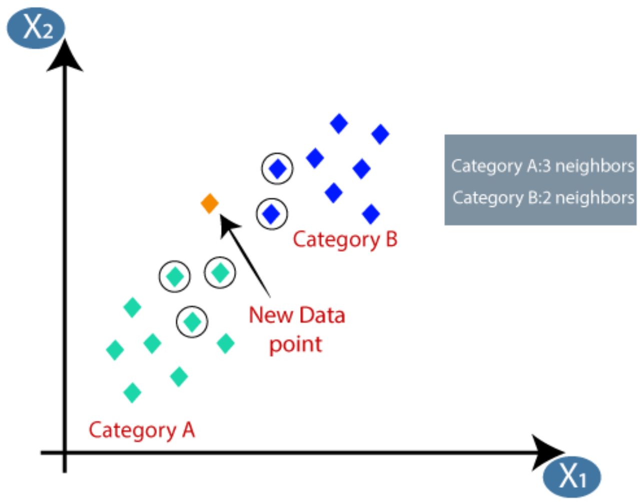


Figure 13: KNN

There is no particular way to determine the best value for "K", so we need to try some values to find the best out of them. The most preferred value for K is 5. A very low value for K such as K=1 or K=2, can be noisy and lead to the effects of outliers in the model. Large values for K are good, but it may find some difficulties. It is simple to implement. It is robust to the noisy training data. It can be more effective if the training data is large.

4.5 Motive Behind Chosen Algorithms

Logistic Regression was chosen due to its ability to model the probability of binary or multi-class outcomes, making it well-suited for crop recommendation tasks. Despite its simplicity, Logistic Regression can effectively capture linear relationships between the input features and the target variable, providing interpretable results.

Support Vector Machines (SVMs) were employed for their robustness in handling high-dimensional and non-linear data. By constructing optimal hyperplanes in a high-dimensional feature space, SVMs can effectively separate different crop classes, making them suitable for multi-cropping recommendations.

Random Forests, an ensemble learning technique, were selected for their ability to handle complex interactions between features and their resistance to overfitting. By combining multiple decision trees, Random Forests can capture intricate patterns in the data, potentially leading to improved predictive performance.

Gradient Boosting, another ensemble method, was chosen for its ability to sequentially train weak models and combine them into a strong predictive model. This algorithm can effectively capture non-linear relationships and interactions between features, making it well-suited for the multi-cropping recommendation task.

Decision Trees were included due to their interpretability and ability to handle both numerical and categorical data. By recursively partitioning the feature space, Decision Trees can capture complex decision boundaries and provide intuitive rules for crop recommendations.

K-Nearest Neighbors (KNN) was selected for its simplicity and effectiveness in capturing local patterns in the data. By considering the similarity of new instances to the existing data points, KNN can provide crop recommendations based on the most similar historical cases.

4.6 Model Evaluation & Deployment

To evaluate the performance of the crop recommendation models, a comprehensive set of evaluation metrics was employed. These metrics provided insights into the models' accuracy, precision, recall, and overall effectiveness in predicting the most suitable crops for a given set of conditions.

For classification tasks, metrics such as accuracy, precision, recall, and F1-score were utilized. Accuracy measured the overall correctness of the models' predictions, while precision quantified the proportion of true positives among the positive predictions. Recall, on the other hand, measured the fraction of actual positives that were correctly identified by the models. The F1-score provided a harmonized measure that balanced precision and recall.

In addition to these metrics, techniques like cross-validation and hold-out testing were employed to ensure the robustness and generalization capabilities of the models. Cross-validation involved partitioning the dataset into multiple subsets, training the model on a subset, and evaluating its performance on the remaining data. This process was repeated multiple times, and the results were averaged to obtain a more reliable estimate of the models' performance.

Once the best-performing model(s) were identified and thoroughly evaluated, they were prepared for deployment. This involved integrating the models into existing farm management systems or developing dedicated applications tailored to the needs of farmers and agricultural stakeholders.

Continuous monitoring and updating of the deployed models are essential to ensure their long-term effectiveness. As new data becomes available or environmental conditions change, the models may need to be retrained or fine-tuned to maintain their predictive accuracy and relevance.

Accuracy

Accuracy is the ratio of correctly predicted instances to the total instances. It measures the overall effectiveness of the model.

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

where:

- **TP (True Positive):** Number of correctly predicted positive instances.
- **TN (True Negative)**: Number of correctly predicted negative instances.
- **FP (False Positive):** Number of incorrectly predicted positive instances.
- **FN (False Negative):** Number of incorrectly predicted negative instances.

Precision
Precision, also known as Positive Predictive Value, is the ratio of correctly predicted positive instances to the total predicted positive instances. It indicates how many of the predicted positive instances are actually positive.

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

Precision is important when the cost of false positives is high.

Recall
Recall, also known as Sensitivity or True Positive Rate, is the ratio of correctly predicted positive instances to all actual positive instances. It measures the ability of the model to identify all relevant instances.

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

Recall is critical when the cost of false negatives is high.

F1-Score
The F1-Score is the harmonic mean of Precision and Recall. It provides a single metric that balances both the concerns of precision and recall.

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

The F1-Score is useful when you need a balance between precision and recall, especially in cases where you have an uneven class distribution.

4.7 Scenario of Multi Crop

Here we are going to describe a scenario after the crop recommendation system. In our proposed system the user could work with a single crop & also multiple crops based on users need. The user will have the flexibility to choose from the crops. Based on that crop our system would then recommend a cultivation guideline for the user.

Maize & Lentil could be cultivated together. The dataset we have worked on have 200 instances together on maize & lentils. Here is the intercropping guideline for maize & lentil together.

Soil Type

Well-drained loamy or sandy loam soil is good for maize & lentil. Both crops prefer soil that is rich in organic matter and has good drainage

Seed Preparation

Maize: Before planting the seed of maize, it has to be soaked in water for 12-24 hours before sowing to promote germination

Figure 14: Maize Seed

Lentil: For lentil high-quality seeds should be used. Lentils generally do not require soaking but can benefit from a seed treatment with fungicides to prevent soil-borne diseases

Figure 15: Lentil Seed

Soil Preparation

Plowing: Deep plow the field to a depth of 20-25 cm to create a fine tilth

Manure/Compost: Incorporate well-rotted manure or compost (10-15 tons/ha) during the last plowing to improve soil fertility

Planting

Row Arrangement: Plant maize in rows with lentils in between. A common practice is to alternate rows or use a specific pattern such as 2:1 or 1:2 (maize: lentil)

Spacing

For maize rows should be spaced about 60-75 cm apart with 20-30 cm between plants. And lentil seeds should be sowed in between maize rows with a spacing of about 20-30 cm between rows and 5-10 cm between plants

Temperature

For both the crops the temperature should be in a range of 20-30°C

pH

The ideal range of pH value for the both maize & lentil should be 6-7.5

Humidity

The moderate humidity should be 50-70%

Soil Moisture

Consistent soil moisture is crucial for both crops. Avoid waterlogging and ensure proper drainage

Cultivation Time

Both maize & lentil can be planted in the Rabi season (October-November) or in the Kharif season (June-July), depending on the region and climatic conditions.

Chapter 5

Water & Fertilizer Distribution

We are using Internet of Things (IoT) in our project to execute with Machine Learning to both recommend & distribute water & fertilizer automatically. In this paper a proposed prototype, design and which provides water to plant based on content of the moisture present in the soil as well as quantify the amount of water usage and send the information to a web server in real-time is shown.

5.1 System Requirements

To build the prototype we have used the following components

1. ESP 8266
2. 16*2 LED Display
3. I2C
4. DHT 11
5. Soil Moisture Sensor
6. Solenoid Valve
7. Water Flow Sensor
8. Ultrasonic Sonar Sensor
9. Battery
10. Battery Case
11. Jumper Wire

5.2 System Design & Development

- **ESP 8266** is a low-cost Wi-Fi microcontroller, with built-in TCP/IP networking software, and microcontroller capability. It enables microcontrollers to connect to 2.4 GHz Wi-Fi, using IEEE 802.11 bgn. It can be used with ESP-AT firmware to provide Wi-Fi connectivity to external host MCUs.

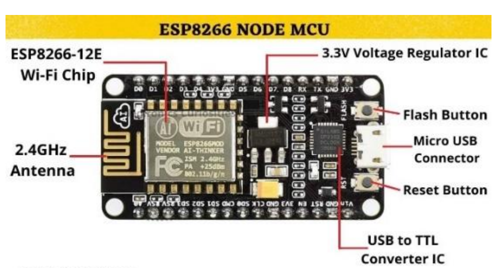


Figure 16: ESP 8266

- **16*2 LED Display** are compact displays that show 16 characters on 2 lines.

Figure 17: 16*2 LED Display

- **I2C** is a two-wire serial communication protocol using a serial data line (SDA) and a serial clock line (SCL).

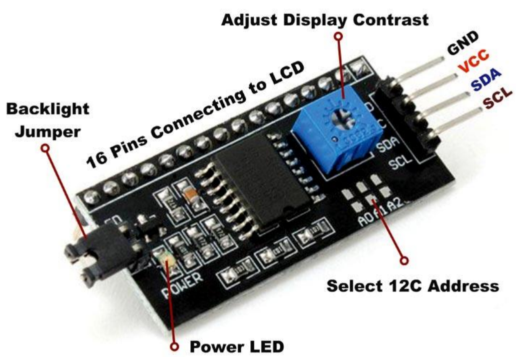


Figure 18: I2C

- **DHT 11** is a basic, ultra low-cost digital temperature and humidity sensor. It uses a capacitive humidity sensor and a thermistor to measure the surrounding air and spits out a digital signal on the data pin (no analog input pins needed)

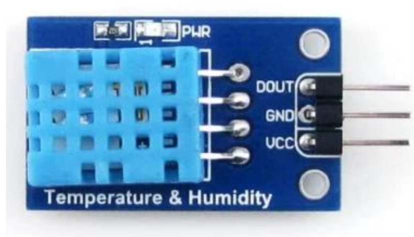


Figure 19: DHT 11

- **Soil Moisture Sensors** measure or estimate the amount of water in the soil. These sensors can be stationary or portables such as handheld probes. Stationary sensors are placed at the predetermined locations and depths in the field, whereas portable soil moisture probes can measure soil moisture at several locations.

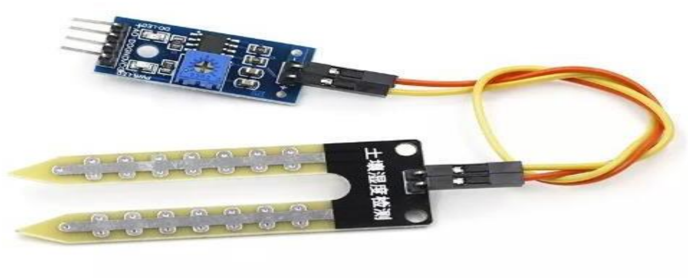


Figure 20: Soil Moisture Sensor

- **Solenoid Valve** is a control unit which, when electrically energized or de-energized, either shut off or allow fluid flow

Figure 21: Solenoid Valve

- **Water Flow Sensor** consists of a plastic valve from which water can pass. A water rotor along with a hall effect sensor is present the sense and measure the water flow. When water flows through the valve it rotates the rotor. This change is calculated as output as a pulse signal by the hall effect sensor.

Figure 22: Water Flow Sensor

- **Ultrasonic Sonar Sensors** can scan an empty room and then know when one or more people are present. In our project it would detect if the water tank if empty or not.

Figure 23: Ultrasonic Sonar Sensor

- **Battery** is being used here to power the device.

Figure 24: Battery

- **Battery Case** is being used to hold the battery


Figure 25: Battery Case

- A **Jumper Wire** is an electrical wire, or group of them in a cable, with a connector or pin at each end which is normally used to interconnect the components of a breadboard or other prototype or test

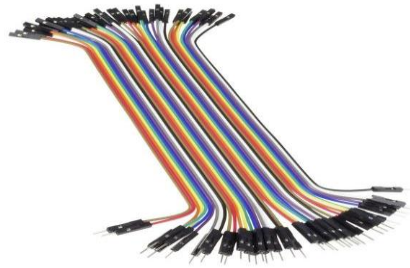


Figure 26: Jumper Wire

5.3 Simulation

Esp8266 is the brain of this system and all sensors and displays responds and are controlled by them. Here we have connected the DHT 11 sensor & Soil moisture sensor. From these 2 sensors we are getting the pH value, humidity & temperature values. The simulation is shown in Fig 5.12.

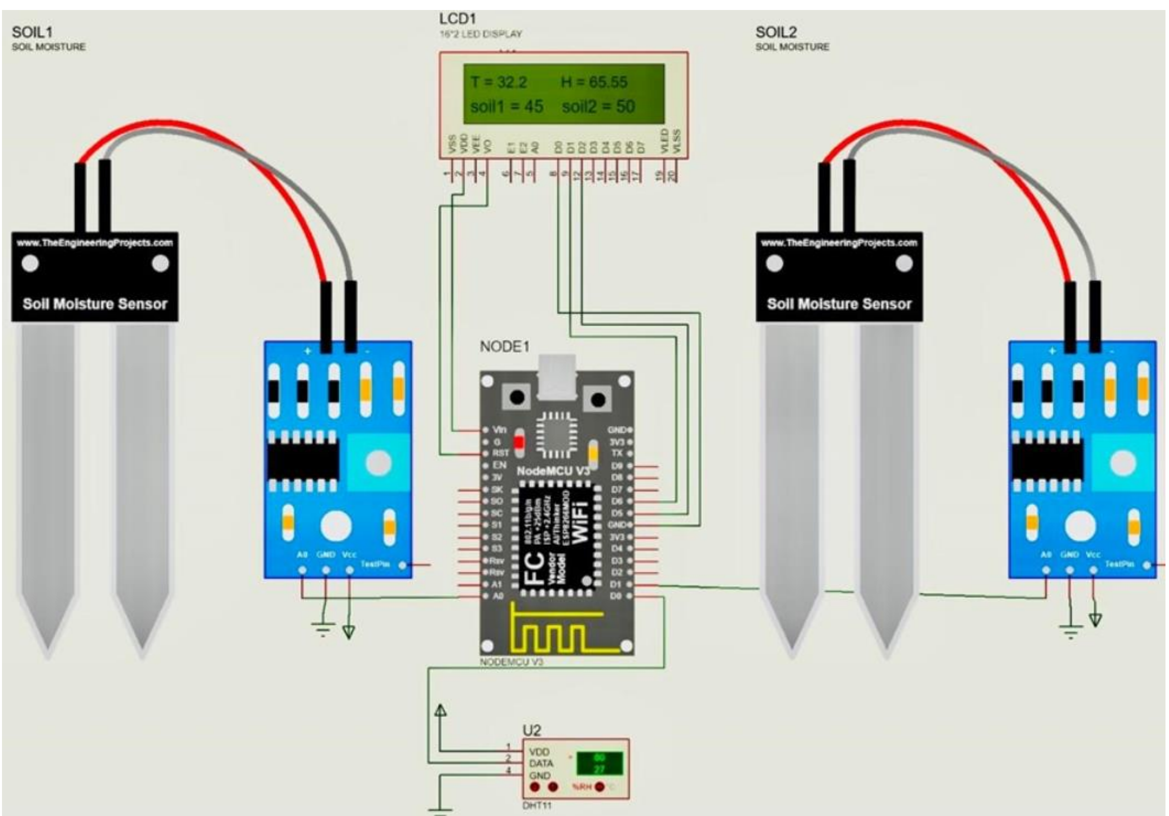


Figure 27: Simulation

And this it the actual demo of the simulation.

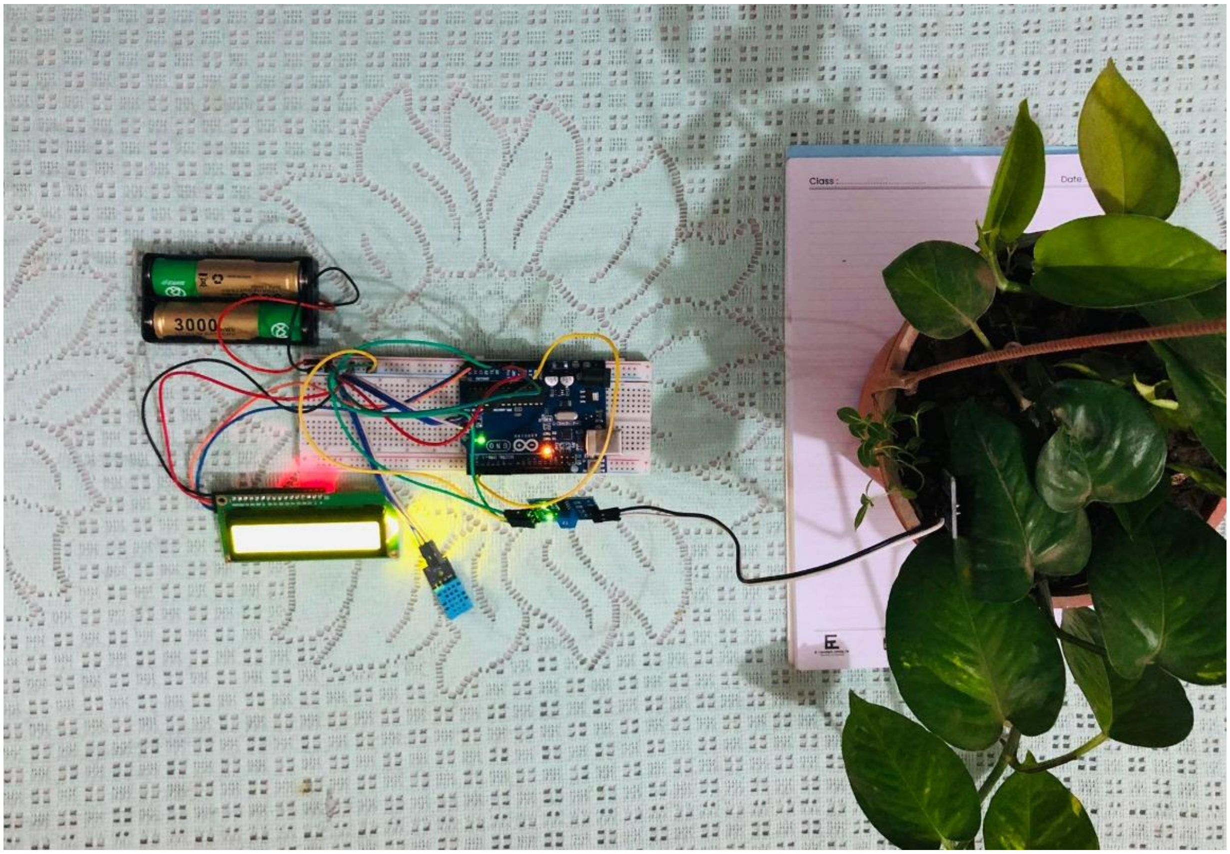


Figure 28: Actual Device

5.4 Water & Fertilizer Distribution Mechanism

The device we have made is the water & fertilizer distribution system for farming. Here in the Fig 5.14, you can see the actual mechanism of the device with different fertilizer & water.

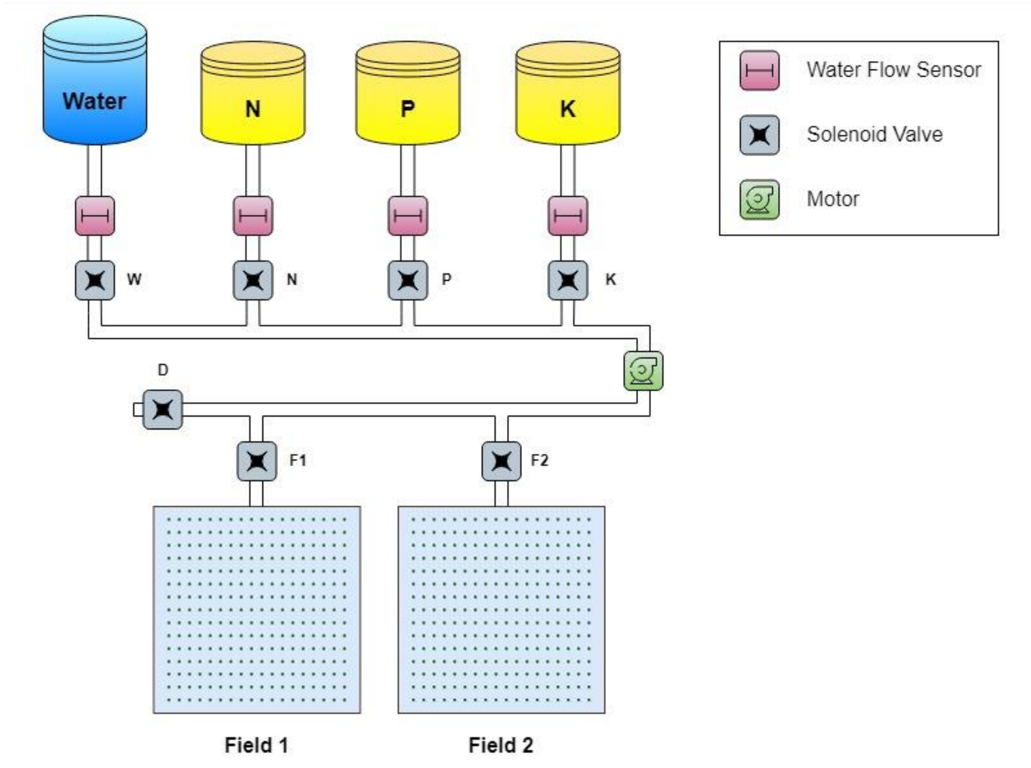


Figure 29: Water & Fertilizer Distribution Mechanism

The diagram represents an automated system for distributing water and fertilizers (Nitrogen (N), Phosphorus (P), and Potassium (K)) to different fields. Here's a detailed explanation of the mechanism depicted:

- **Tanks: Water Tank** to stores water for irrigation. **N Tank to s**tore nitrogen fertilizer. **P Tank** store phosphorus fertilizer. **K Tank** stores potassium fertilizer.
- **Sensors and Valves: Water Flow Sensors** monitors the flow of water and fertilizers. **Solenoid Valves** controls the flow of water and fertilizers. **Motor** drives the distribution mechanism.
- **Fields: Field 1 (F1)** & **Field 2 (F2)** receives water and fertilizer.

Mechanism

Water Distribution: Water from the Water Tank flows through a Water Flow Sensor (W). The flow sensor monitors the quantity of water being distributed. Water then passes through a Solenoid

Valve (D) that controls the flow direction to either Field 1 or Field 2. Based on the control signal, the solenoid valve directs water to the designated field (F1 or F2).

Fertilizer Distribution: Similar to water, each fertilizer (N, P, and K) flows through their respective Water Flow Sensors (N, P, K). These sensors monitor the flow rate of each fertilizer. Each fertilizer then passes through a dedicated Solenoid Valve that regulates its distribution. Fertilizers join the main distribution line after their respective valves.

Combined Distribution: Once the water and fertilizers are regulated by their respective solenoid valves, they are combined in the main distribution line (D). The motor assists in driving the combined mixture towards the fields. Another set of Solenoid Valves (F1 for Field 1 and F2 for Field 2) at the end of the distribution line determines which field receives the mixture.

Field Distribution: The solenoid valves at F1 and F2 open or close based on control signals, allowing the water-fertilizer mixture to be delivered to the selected field.

This mechanism ensures precise and automated distribution of water and fertilizers to each field based on the requirements.

Control & Automation

Sensors continuously monitor the flow rates and ensure the correct amounts of water and fertilizers are being distributed.

Solenoid Valves act as gatekeepers, controlling the flow and directing it to the appropriate fields.

Motor provides the necessary force to drive the mixture through the distribution system.

Control Signals (not shown in the diagram but implied) from a central controller or automated system determine when and how much of each resource is distributed to each field.

Chapter 6

Project Showcase

To showcase the recommended crops, we have implemented the model on a website. In this website the users could input the values of NPK, pH, rainfall, temperature & humidity. Based on those values our model would recommend the user a best yield crops in our website interface. And the user would be also provided cultivation guideline on that recommended crop.

6.1 ER Diagram

This ER diagram depicts the relationships between different entities in a system related to agriculture. Below is a detailed description of each entity and their relationships. The ER diagram of our website is shown in the Fig 6.1.

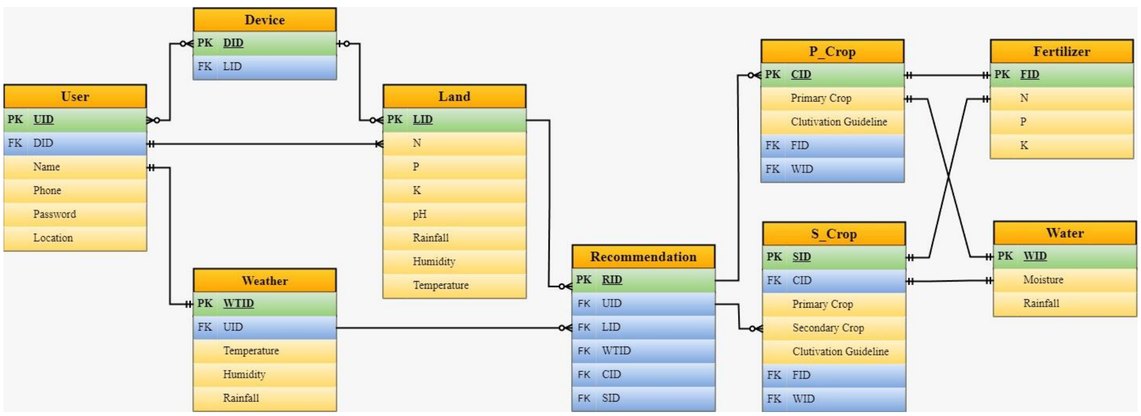


Figure 30: ER Diagram

Relationships

User – Device

The relation **relationship is** One-to-Many (One user can have multiple devices). Each user (`UID`) can be associated with multiple devices (`DID`). This allows the system to track which devices are owned or operated by which users. The `Device` table has a foreign key `DID` that references the primary key `UID` in the `User` table.

Device – Land

The relation **relationship is** Many-to-One (Multiple devices can be associated with one piece of land). Each device (`DID`) can be linked to a specific piece of land (`LID`). This is helpful in monitoring the land using multiple devices. The `Device` table has a foreign key `LID` that references the primary key `LID` in the `Land` table.

User – Weather

The relation **relationship is** One-to-Many (One user can have multiple weather data entries). Weather data is collected for each user. Each entry in the `Weather` table is linked to a user through the `UID` foreign key, which references the `User` table's primary key `UID`.

User – Recommendation

The relation **relationship is** One-to-Many (One user can have multiple recommendations). Each user can receive multiple recommendations. The `Recommendation` table has a foreign key `UID` that links to the `User` table's primary key `UID`.

Land – Recommendation

The relation **relationship is** One-to-Many (One piece of land can have multiple recommendations). Recommendations are generated for specific pieces of land based on their characteristics. The `Recommendation` table has a foreign key `LID` that links to the `Land` table's primary key `LID`.

Weather – Recommendation

The relation **relationship is** One-to-Many (One weather data entry can inform multiple recommendations). Weather data impacts the recommendations provided. The `Recommendation` table has a foreign key `WTID` that references the `Weather` table's primary key `WTID`.

P_Crop – Recommendation

The relation **relationship is** One-to-Many (One primary crop can be part of multiple recommendations). Primary crop data influences the recommendations. The `Recommendation` table has a foreign key `CID` that links to the `P_Crop` table's primary key `CID`.

S_Crop – Recommendation

The relation **relationship is** One-to-Many (One secondary crop can be part of multiple recommendations). Secondary crop data influences the recommendations. The `Recommendation` table has a foreign key `SID` that links to the `S_Crop` table's primary key `SID`.

P_Crop – Fertilizer

The relation **relationship is** Many-to-One (Multiple primary crops can use the same fertilizer). Fertilizer information is linked to primary crops. The `P_Crop` table has a foreign key `FID` that references the `Fertilizer` table's primary key `FID`.

P_Crop – Water

The relation **relationship is** Many-to-One (Multiple primary crops can use the same water data). Water data is linked to primary crops. The `P_Crop` table has a foreign key `WID` that references the `Water` table's primary key `WID`.

S_Crop - P_Crop

The relation **relationship is** Many-to-One (Multiple secondary crops are associated with one primary crop). The `S_Crop` table links secondary crops to primary crops through the foreign key `CID`, referencing the primary key `CID` in the `P_Crop` table.

S_Crop – Fertilizer

The relation **relationship is** Many-to-One (Multiple secondary crops can use the same fertilizer). Fertilizer information is linked to secondary crops. The `S_Crop` table has a foreign key `FID` that references the `Fertilizer` table's primary key `FID`.

S_Crop - Water: Many-to-One (Multiple secondary crops can use the same water data). Water data is linked to secondary crops. The `S_Crop` table has a foreign key `WID` that references the `Water` table's primary key `WID`.

Flow of Data and Purpose

User Interaction: Users register in the system and are assigned devices for monitoring their land.

Device Monitoring: Devices collect data about the land, which includes soil nutrients (N, P, K), pH, rainfall, humidity, and temperature. This data is recorded in the `Land` entity.

Weather Data Collection: Weather data is periodically collected for each user, including temperature, humidity, and rainfall, and stored in the `Weather` entity.

Crop and Resource Management: Information about primary and secondary crops, including cultivation guidelines and associated resources like fertilizers and water requirements, are maintained in `P_Crop` and `S_Crop` entities.

Recommendations: Based on user data, land characteristics, weather conditions, and crop information, the system generates recommendations. These recommendations aim to optimize agricultural practices, suggesting suitable crops, fertilizer usage, and water management strategies. Recommendations are stored in the `Recommendation` entity, linking all relevant data points.

6.2 Use Case Diagram

This use case diagram illustrates the interactions between users, administrators, and the Multi Crop Recommendation System. The use case diagram of our website is shown in the Fig 6.2

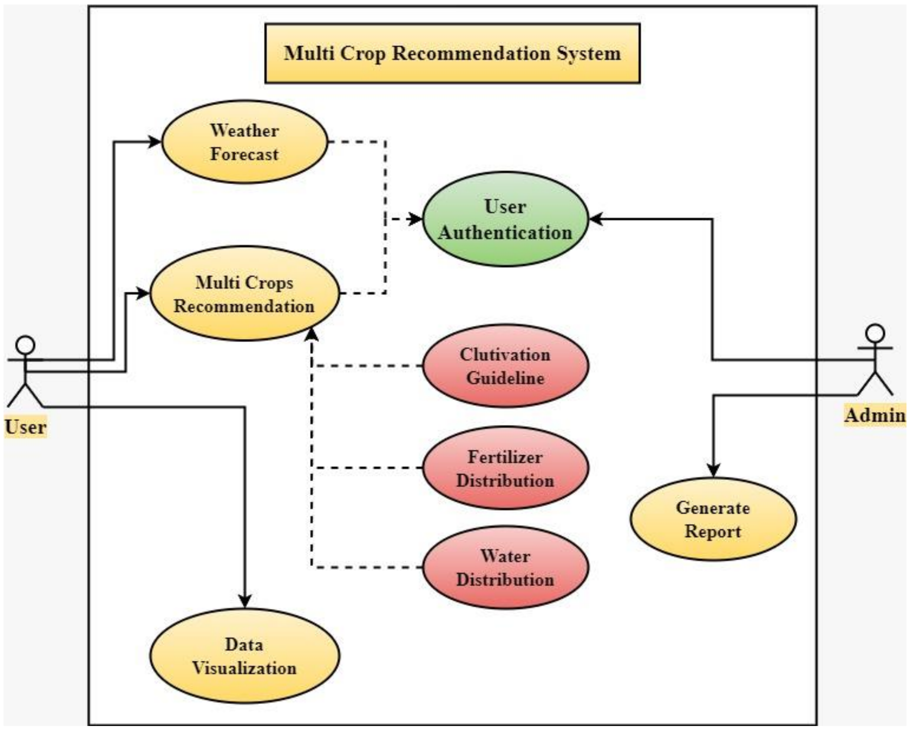


Figure 31: Use Case Diagram

Here's a detailed description of each component and their interactions:

Actors

User represents the end-user, typically a farmer or agricultural practitioner. Admin represents the system administrator who manages and oversees the system.

User Authentication

The process by which users log in to the system. Actors involved User; Admin. Users must authenticate to access the system functionalities. Admin can also authenticate to manage the system.

Weather Forecast

Provides weather forecasts to the system. Actors involved User. Users can view weather forecasts which influence the multi-crop recommendations.

Multi Crops Recommendation Generates recommendations for multiple crops based on various factors. Actors involved User. Users receive crop recommendations influenced by weather forecasts and other data. This use case is central and influences several other use cases.

Cultivation Guideline

Provides guidelines for cultivating recommended crops. Actors involved User. Users can view detailed guidelines on how to cultivate the recommended crops. This use case is connected to Multi Crops Recommendation.

Fertilizer Distribution

Provides recommendations on the distribution and use of fertilizers. Actors involved User.

Users receive advice on which fertilizers to use and how to apply them for the recommended crops. This use case is connected to Multi Crops Recommendation.

Water Distribution

Provides recommendations on water usage and distribution for the crops. Actors involved User. Users get guidance on how to effectively water the crops based on the recommendations. This use case is connected to Multi Crops Recommendation.

Data Visualization

Allows users to visualize data related to their crops, weather, and recommendations. Actors involved User. Users can see graphical representations of the data to make better-informed decisions. This use case is directly accessed by the user.

Generate Report

Allows the admin to generate reports based on system data. Actors involved Admin. Admin can create reports that may include user activity, recommendation efficacy, and system performance.

This use case is crucial for monitoring and improving the system.

Relationships

User Authentication is a prerequisite for users and admin to access the system and its features.

Weather Forecast directly influences Multi Crops Recommendation, as weather data is crucial for accurate crop recommendations. Multi Crops Recommendation is the central use case that connects to Cultivation Guideline**,** Fertilizer Distribution**,** and Water Distribution. These use cases provide specific recommendations based on the overall crop recommendation.

Data Visualization allows users to view and interpret the recommendations and other data visually. Generate Report is an administrative function that helps in system management and performance evaluation.

This is the homepage of our website.

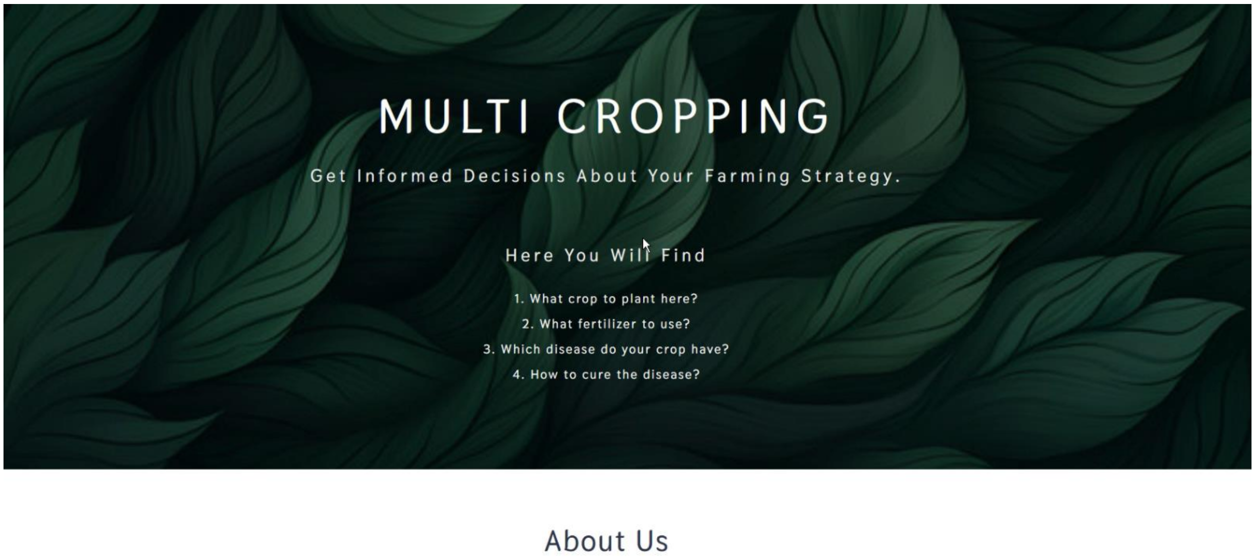


Figure 32: Homepage

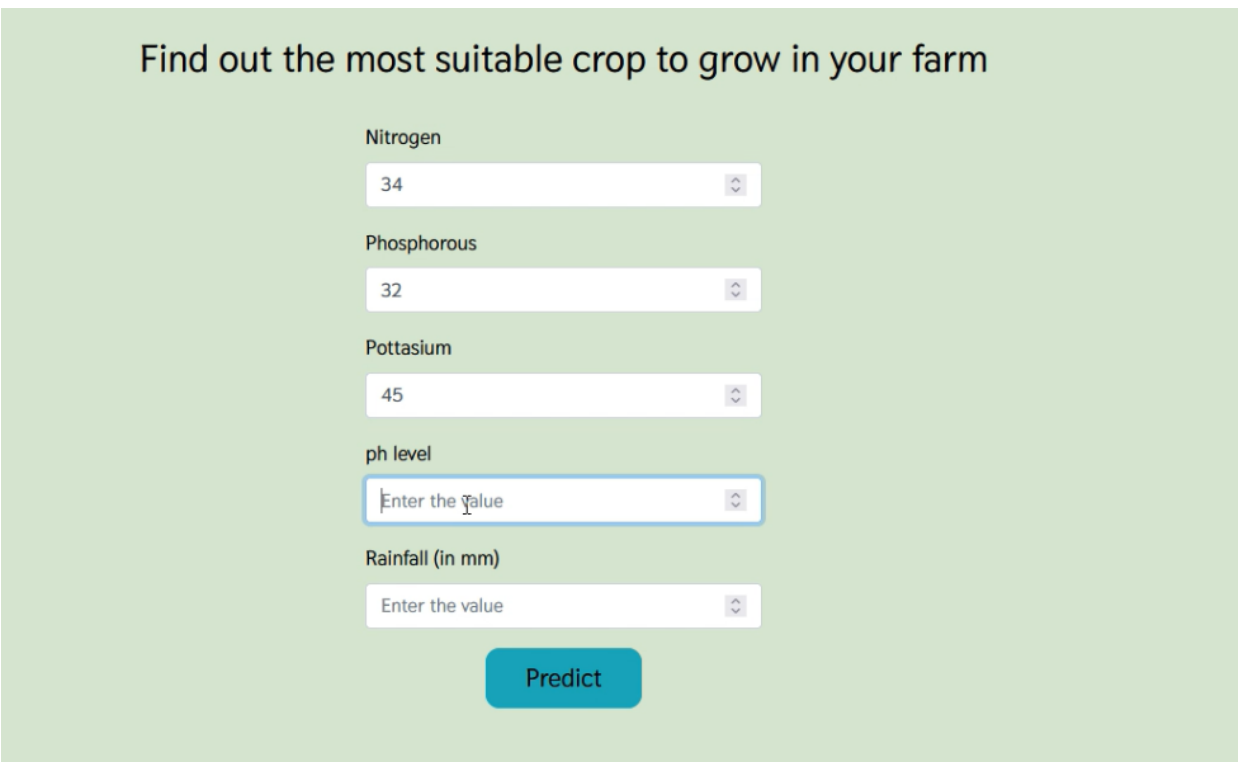


Figure 33: Recommendation Page

The N value of your soil is low.

Please consider the following suggestions:

1. *Add sawdust or fine woodchips to your soil* – the carbon in the sawdust/woodchips love nitrogen and will help absorb and soak up and excess nitrogen.
2. *Plant heavy nitrogen feeding plants* – tomatoes, corn, broccoli, cabbage and spinach are examples of plants that thrive off nitrogen and will suck the nitrogen dry.
3. *Water* – soaking your soil with water will help leach the nitrogen deeper into your soil, effectively leaving less for your plants to use.
4. *Sugar* – In limited studies, it was shown that adding sugar to your soil can help potentially reduce the amount of nitrogen is your soil. Sugar is partially composed of carbon, an element which attracts and soaks up the nitrogen in the soil. This is similar concept to adding sawdust/woodchips which are high in carbon content.
5. Add composted manure to the soil.
6. Plant Nitrogen fixing plants like peas or beans.
7. *Use NPK fertilizers with high N value.*

Figure 34: Cultivation Guideline

Chapter 7

Result & Analysis

In this study, we evaluated the performance of several machine learning models and a deep learning model for predicting crop types based on various agricultural and environmental factors. The models were trained and tested on a dataset that included features such as nitrogen (N), phosphorus (P), potassium (K) content, temperature, humidity, pH, and rainfall. The models evaluated were Logistic Regression, Support Vector Machine (SVM), Random Forest, Gradient Boosting, Decision Tree, and K-Nearest Neighbors (KNN). Additionally, a deep learning model was developed and evaluated.

7.1 Model Performance Metrics

The performance of each model was assessed using four key metrics: accuracy, precision, recall, and F1 score. The results are summarized in Table 3.

Model	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.96	0.96	0.96	0.96
Support Vector Machine	0.96	0.97	0.96	0.96
Random Forest	0.98	0.98	0.98	0.98
Gradient Boosting	0.98	0.98	0.98	0.98
Decision Tree	0.97	0.97	0.97	0.97
K-Nearest Neighbors	0.96	0.96	0.96	0.96

Table 3: Evaluation Metrics of Different Machine Learning Models

7.2 Analysis

The results indicate that all models performed exceptionally well on the crop prediction task, achieving high scores across all metrics. Notably, the Random Forest and Gradient Boosting models outperformed other models with an accuracy of 0.98, precision of 0.98, recall of 0.98, and F1 score of 0.98. These models demonstrated a superior ability to correctly classify crop types, making them the most effective models for this dataset.

The Decision Tree model also performed well, with an accuracy of 0.97, precision of 0.97, recall of 0.97, and F1 score of 0.97. This model's performance was slightly lower than that of Random Forest and Gradient Boosting but still robust, indicating its suitability for crop prediction tasks.

Logistic Regression, SVM, and KNN models each achieved an accuracy of 0.96. The SVM model demonstrated slightly higher precision (0.97) compared to Logistic Regression and KNN (both 0.96). The consistency in performance across these models suggests that they are reliable choices for this classification task, although they do not match the top performance of Random Forest and Gradient Boosting.

The deep learning model also achieved high performance with an accuracy of 0.96, precision of 0.96, recall of 0.96, and F1 score of 0.96. This indicates that the neural network was able to capture the underlying patterns in the data effectively, although it did not outperform the top-performing ensemble models.

With the dataset from Kaggle[https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset] the function-based and rules-based are taken for implementing the machine learning technique had given us an accuracy performance percentage of 88.5909% for Decision Table.

And with the same dataset [https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset] we have acquired 97% accuracy.

Chapter 8

Future work & Conclusion

8.1 Future Work

Our future work focuses on several key advancements to enhance the usability, affordability, and effectiveness of the Multi Crop Recommendation System. First, we plan to develop a multilingual application tailored for Bangladeshi farmers. This app will support local languages such as Bengali, making the services more accessible and user-friendly. We aim to incorporate features like language selection, translated content, and voice assistance to cater to farmers' diverse needs, especially those who may not be proficient in English or are less literate.

Additionally, we are working on creating a cost-effective, all-in-one IoT device. Currently, our assembled IoT device is not economically feasible, but as we move towards production, we intend to integrate multiple sensors into a single, compact device. This device will monitor soil moisture, nutrient levels, and weather conditions, offering a versatile and durable solution for multi-crop farming.

Another significant component of our future work is the development of a market analysis model. This model will analyze market conditions, including crop prices and demand trends, to recommend the most profitable crops for the current season. By utilizing machine learning algorithms, we can provide farmers with real-time, actionable insights, helping them make informed decisions and minimize financial risks.

Furthermore, we plan to enhance our system's weather forecasting capabilities. By analyzing historical weather data, we aim to predict weather conditions for up to a year in advance. This long-term forecast will help farmers plan their planting and harvesting schedules more effectively, ensuring better crop management.

Finally, we are implementing a plant disease detection feature on our website. Using machine learning and image recognition techniques, farmers will be able to upload photos of their crops and receive instant diagnostics and treatment recommendations. This feature will be supported by a comprehensive database of plant diseases, symptoms, and management practices, providing farmers with the necessary information to prevent and manage crop diseases efficiently.

In summary, these advancements will collectively enhance the Multi Crop Recommendation System, making it more accessible, cost-effective, and beneficial for farmers in Bangladesh. Our goal is to provide comprehensive support that reduces farming risks, improves crop yields, and increases profitability, thereby contributing to the overall productivity and sustainability of agriculture in the region.

8.2 Conclusion

In conclusion, integrating Machine Learning (ML) and Internet of Things (IoT) technologies can significantly transform agricultural practices, particularly by optimizing resource use and promoting sustainability. Our project's primary objectives focus on resource optimization, environmental sustainability, crop health improvement, cost efficiency, technology integration, data-driven decision making, and global food security. By employing advanced ML algorithms, such as Random Forest, to deliver precise crop recommendations, and deploying IoT devices like soil moisture sensors and solenoid valves to manage water and fertilizers efficiently, we aim to address these objectives comprehensively.

This approach empowers farmers with real-time data and insights, enabling them to make informed decisions that enhance productivity and profitability while conserving resources. By minimizing water wastage and chemical runoff, the project supports sustainable farming practices and contributes to environmental preservation. The integration of cutting-edge technologies automates and streamlines agricultural operations, reducing input costs and improving economic sustainability for farmers.

Ultimately, this project envisions a future of precision agriculture where increased efficiency and productivity meet the demands of a growing global population. By improving agricultural practices in India and beyond, we aim to ensure a stable and sufficient food supply, thereby contributing to global food security. The seamless blend of ML and IoT technologies in this project sets the stage for a sustainable, efficient, and profitable agricultural sector that can meet future challenges head-on.

References

[1] K. N. R. Amanjot Bhullar, "Simultaneous multi-crop land suitability prediction from remote sensing data using semi-supervised learning," *Scientific Reports,* vol. 13, pp. 3-5, 2023.

[2] K. S. S. ,. A. N. a. I. A.-I. Rezwan MD, Smart Soil Moisture Balancer Using Internet of Things (IoT), Dhaka: UITS, 2023.

[3] R. B. A. A.Dahane, "An IoT Based Smart Farming System Using Machine Learning," in *ISNCC*, Canada, 2020.

[4] L. D. V. A. G. M. S. S. C. B. W. K. M. S. P. P. R. K. B. T. A. T. Blesslin Sheeba, "Machine Learning Algorithm for Soil Analysis and Classification of Micronutrients in IoT-Enabled Automated Farms," *Hindawi,* pp. 3-4, 2022.

[5] J. L. L. K. A. M. M. Q. Y. Abdul Rehman, "Machine Learning Prediction Analysis using IoT for," *IJETER,* vol. 8, no. 9, pp. 6484-6485, 2020.

[6] A. S. S. S. S. Shikha Prakash, "SOIL MOISTURE PREDICTION USING MACHINE LEARNING," in *ICICCT*, India, 2018.

[7] A. R. N. P. Murali Krishna Senapaty, "IoT-Enabled Soil Nutrient Analysis and Crop Recommendation Model for Precision Agriculture," *MDPI,* vol. 12, pp. 14-19, 2023.

[8] V. S. T. S. K. S. R. Reshma, "IoT based Classification Techniques for Soil Content Analysis and Crop Yield Prediction," in *I-SMAC*, Palladam, India, 2020.

[9] B. K. V. T. S. M. D. J. K. K. V. V. Kalaiselvi Bakthavatchalam, "IoT Framework for Measurement and Precision Agriculture: Predicting the Crop Using Machine Learning Algorithms," *MDPI,* vol. 10, pp. 5-9, 2022.

[10] Y. M. R. K. R. L. N. N. S. N. Kasara Sai Pratyush Reddy, "IoT based Smart Agriculture using Machine Learning," in *ICIRCA*, Ciombatore, India, 2020.

[11] W. A. R. H. H. A. F. N. G. A. M. S. I. M. S. A. A. M. A. I. U. Amna Ikram, "Crop Yield Maximization Using an IoT-Based Smart Decision," *Hindawi,* pp. 4-6, 2022.

[12] A. S. N. S. R. D. U. A. Madhuri Shripathi Rao, "Crop prediction using machine learning," in *JPCS*, Canada, 2022.