AN IoT-BASED SOIL AND PLANT NUTRIENT MANAGEMENT OF PAPAYA PRODUCTION IN BANGLADESH

BY

Rumia Rubayat Islam
ID: 202-15-3803
AND

Arafat Jahan Hira
ID: 202-15-3815

FINAL YEAR DESIGN PROJECT REPORT

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Dr. Fizar Ahmed
Associate Professor
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised By

Mr. Abdus Sattar
Assistant Professor & Coordinator M.Sc
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

July 2024

APPROVAL

This Project titled "An IoT-Based Soil And Plant Nutrient Management OF Papaya Production In Bangladesh", submitted by Rumia Ruabayat Islam and Arafat Jahan Hira to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on July 14, 2024

BOARD OF EXAMINERS

Hossain 14-07-2024

Dr. Md. Fokhray Hossain (MFH) — **Board Chairman**
Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Amatul Bushra Akhi (ABA) — **Internal Examiner 1**
Assistant Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Mr. Amir Sohel (ARS) — **Internal Examiner 2**
Sr. Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

14/7/24

Dr. Shamim H Ripon (DSR) — **External Examiner**
Professor
Department of Computer Science and Engineering
East West University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Dr.Fizar Ahmed, Associate Professor, Department of Computer Science and Engineering, Daffodil International University**. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Dr. Fizar Ahmed
Associate Professor
Department of CSE
Daffodil International University

Co-Supervised by:

Mr. Abdus Sattar
Assistant Professor & Coordinator M.Sc
Department of CSE
Daffodil International University

Submitted by:

Rumia Rubayat Islam

Rumia Rubayat Islam
ID: 202-15-3803
Department of CSE
Daffodil International University

Arafat Jahan

Arafat Jahan Hira
ID: 202-15-3815
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty for His divine blessing making it possible for us to complete the final year project/internship successfully.

We are grateful and wish our profound indebtedness to **Dr.Fizar Ahmed, Associate Professor**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*IoT and Embedded Systems*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to the **Head of the Department of CSE**, for his kind help in finishing our project and also to other faculty members and the staff of the Department of CSE, Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discussion while completing the course work.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

This article gives an insight into the creative use of the IoT system in the case of the regulation of the soil and the plant nutrients in the cultivation of papaya plants in the farming landscape of Bangladesh. Papaya production in Bangladesh faces challenges related to soil quality, nutrient management, and the environment. In this study, we propose an IoT-based system for soil and plant nutrient management to improve papaya production. This paper describes implementing an IoT-enabled soil and plant nutrient management gadget centered on papaya production in Bangladesh. This study develops superior predictive models in ARIMA, SARIMAX, Random forest, XGBoost, Linear Regression, Logistic Regression, and Multilinear Regression to optimize the rural activities by properly estimating soil and plant nutrient stages. The proposed framework consists of an IoT-sensor network around the field in strategic places across the papaya fields, monitoring and recording moisture degrees, Nitrogen (N), Phosphorus (P), and Potassium (k) tiers, amongst different required degrees. The records assets heterogeneously encompass the moisture stage information and information on the essential nutrients, namely Nitrogen (N), Phosphorus (P), and Potassium (k), which are statistically modeled with environmental parameters to expect the future nutrient degrees and are expecting the general health of the papaya crop. This information is then fed into the advanced predictive fashions that make up the superior framework—ARIMA, SARIMAX,Random forest, XGBoost , Linear Regression, Logistic Regression, and Multilinear Regression.In correlating the models, one-of-a-kind sets of model assessment metrics are used. suggest Squared blunders (MSE) and (RMSE) degree the value of average squared variance or the unfold between determined and expected values. Accuracy metrics determine how correct the predictions are, with special attention given to binary classification like nutrient deficiency. suggest Absolute Scaled blunders (MAPE) is a normalized measure of forecast accuracy that allows scale evaluation. Consequences will generate the website precise, dynamic, and really useful information that may be implemented on actual grounds via the right choice-making to decorate nutrient management alternatives, accordingly enhancing the performance and sustainability of papaya manufacturing.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

ML	Machine Learning
SVM	Support Vector Machine
IoT	Internet of Things
ARIMA	Auto-Regressive Integrated Moving Average
SARIMAX	Seasonal ARIMA with Exogenous Variables
XGBoost	Extreme Gradient Boosting
RMSE	Root Mean Squared Error
MSE	Mean Squared Error
MAPE	Mean Absolute Percentage Error
N	Nitrogen
P	Phosphorus
K	Potassium
IPM	Integrated Pest Management
SOM	Soil Organic Matter
ANOVA	Analysis of Variance
SVM	Support Vector Machine
SWOT	Strengths, Weaknesses, Opportunities, Threats (Analysis)
LM	Linear Models
RNN	Recurrent Neural Network
KNN	K-Nearest Neighbors
R^2	(R-squared)
TTL	Transistor-Transistor Logic
LCD	Liquid Crystal Display
CI	Confidence Interval

CHAPTER 1

Introduction

1.1 Overview

Agriculture accounts for a large part of Bangladesh's income, but it has long-term problems in soil quality, nutritional management, and environmental protection. These problems have a major impact on papaya cultivation, which is an important crop in the area। Traditional agricultural methods often result in lower food yields and deterioration of the environment, showing the need for new ideas. New technologies such as the Internet of Things (IoT) have made it possible in recent years to address these issues in new ways. IoT and machine learning algorithms can change farming methods by letting farmers keep an eye on things in real time and get ideas from the data. This study shows how the Internet of Things (IoT)-based systems can be used to control soil and plant nutrients in Bangladeshi papaya cultivation. The goal of this study is to improve sustainability, efficiency, and environmental care in agriculture by combining Internet of Things (IoT) technologies with machine learning models such as ARIMA, SARIMAX, Random Forest, XGBoost, Linear Regression, Logistic Regression, and Multilinear Regression. The idea for this study came from the fact that IoT and machine learning can completely change the land of cultivation। These technologies can give farmers real-time data and a vision of the future, which can help them make smart decisions and use resources more efficiently. Previous studies have shown that using IoT systems in agriculture can help in the management of soil and plant nutrients and make crops more productive overall. However, there is a big hole in how this technology is being used in papaya cultivation in Bangladesh. This is why this study attempts to fill the gap by creating IoT-based solutions that fit specific issues of papaya growers in the area. The main goals of this study are to create and use Internet of Things (IoT) systems to track the nutrition of soils and plants, to see how well different machine learning algorithms work in improving agricultural practices, and to see how these technologies affect sustainability and productivity in Papaya cultivation. The research seeks to provide useful advice to papaya growers in Bangladesh on how to use IoT solutions more widely.

1.2 Background and Present State

In an IoT-based soil and plant nutrient management device for papaya production in Bangladesh, advanced predictive models and performance metrics are used to optimize the practices and grow the yield. IoT plays a key function in this system by accumulating real-time information through sensors that reveal various soil and environmental parameters to make specific and timely interventions. ARIMA (Autoregressive included shifting common) and SARIMAX (Seasonal ARIMA with Exogenous Variables) are used for time series evaluation to forecast the future soil moisture and nutrient ranges so that control may be carried out primarily based on historical statistics and seasonal developments. Random forest and XGBoost (severe Gradient Boosting) are systems getting to know algorithms that enhance the prediction accuracy and robustness; Random forest does this by constructing a couple of choice timber to seize the complex and non-linear relationships within the facts and XGBoost does this by pretty green implementation of gradient boosting strategies.

To ensure the reliability and effectiveness of these fashions, diverse assessment metrics are used. imply Squared error (MSE) and Root implies Squared error (RMSE) measures the common squared distinction between discovered and expected values to peer how to correct the version. Accuracy is used to look at the percentage of accurate predictions in classification tasks and imply Absolute Scaled mistakes (MAPE) is a normalized measure of errors this is beneficial to examine the model performance across exceptional scales. R-squared (R^2) cost is the proportion of variance in structured variables this is predictable from independent variables and is a measure of version health. Also, the gadget video displays units of the percentage of important soil vitamins like Nitrogen (N), Phosphorus (P), and Potassium (k) together called NPK. Those nutrients are important for plant boom and improvement, and retaining the top-rated degree of those vitamins is necessary to maximize papaya yield and nice. By integrating these superior analytics with IoT-based tracking the machine affords a complete solution for soil and plant nutrient control and thus contributes to sustainable and worthwhile papaya manufacturing in Bangladesh.

Papaya growers in Bangladesh currently encounter constraints in efficiently administering soil and plant nutrients as a result of the absence of sophisticated technical remedies.

Traditional agricultural practices suffer from imprecision and a lack of real-time monitoring capabilities, leading to inefficiencies and the squandering of resources. The lack of effective soil and plant nutrient management systems worsens the difficulties experienced by farmers, affecting both the quality and quantity of crops. The agricultural sector has a transformative opportunity with the introduction of Internet of Things (IoT) technology and machine learning algorithms. Through the strategic placement of IoT sensors throughout papaya fields, farmers can effectively monitor essential metrics, including real-time measurements of moisture levels, nitrogen, phosphorus, and potassium concentrations. These sensors collect useful data that, when examined using advanced predictive modeling techniques such as ARIMA, SARIMAX, Random Forest, XGBoost, Linear Regression, Logistic Regression, and Multilinear Regression, can optimize soil and plant nutrient levels and improve estimates of crop output. By combining IoT technology and machine learning models, there is a possibility of transforming papaya growing techniques in Bangladesh.

1.3 Problem Statement

The challenge is to develop a reliable machine-learning model that can predict soil organic matter and nutrient quality under time, cost, and labor constraints. Traditional methods are expensive and physically demanding, as results often take days or weeks to obtain and special equipment and skilled workers are often needed to write, process, and analyze samples. We aim to create an efficient and fast-paced model that can deliver results in minutes or hours rather than days or weeks and can adapt to different areas and terrain types. This model will help farmers, agronomists, and environmental scientists evaluate soil quality and make informed decisions to improve papaya production and environmental sustainability. Additionally, using machine learning algorithms, we can create predictive models of SOM and food levels by analyzing a lot of data collected by IoT sensors. This model can be used in different soils and texture, structure, organic content, nutrient composition, etc. properties can be calculated. Additionally, these models can be trained and adapted to different areas and soil types, thus improving their overall applicability in various agricultural fields in Bangladesh. Our IoT-based predictive models provide farmers, agronomists, and environmental scientists with fast and accurate soil analysis, allowing them to make informed decisions on soil and plant nutrients for papaya. Moreover, this also increases the efficiency and sustainability of agriculture to protect Bangladesh's environment. While doing research, we encounter some time-consuming problems and it can take days or weeks to get results. It requires expensive, specialized equipment and skilled workers. Work-intensive problems such as sample collection. Operation and analysis are very expensive. It is cost effective and fast; and delivers results in minutes or hours, not days or weeks. With appropriate modification of the model, it can be used in many regions and terrain types.

1.4 Objectives

Papaya cultivation in Bangladesh faces many demanding situations, from unpredictable weather to soil degradation and food shortages. The us of a's food protection and the livelihoods of infinite farmers are stricken by these demanding situations. there's a pressing need for new answers to replace traditional agricultural practices and bring about a new generation of efficiency, sustainability, and development in papaya production. technology can be used to remedy challenges faced by papaya farmers in Bangladesh. Farmers can gain facts about soil and plant nutrition utilizing the usage of the Internet of Things.

The primary goal of this research is to create and implement machine learning algorithms for IoT-based soil and plant health systems in papaya design in Bangladesh. They are:

1. Develop integrated IoT systems.
2. Machine Learning Model Development
3. Data Collection and Analysis.
4. Model Analysis and Optimization.
5. Application and Delivery.
6. Impact Assessment

The goal of IoT-based soil and plant health research in Bangladesh is to increase papaya production by designing, developing, and testing an Internet of Things (IoT) Ecosystem that allows for instant reporting, using machine learning techniques to predict soil and plant nutrients, measuring system scalability and stability, and deploying systems in the field.

1.5 Scope and Limitations

Many problems and potential topics for more study may be found based on the literature evaluation that was provided:

1. Limited Research on Innovative Techniques: The assessment highlights the dearth of research on novel approaches or technological advancements that can improve papaya output in the area. Subsequent research endeavors may concentrate on the integration of cutting-edge technologies such as drone technology, IoT devices, and precision agriculture to enhance crop management and augment yields.

2. Economic Implications and Feasibility: The literature analysis does not address the potential financial effects of converting to organic farming techniques or putting suggested changes in land use practices into action. Future studies should examine if sustainable farming methods are financially feasible and evaluate how cost-effective it is to implement new technology in papaya production.

3. Sample Size and Representativeness: A few of the studies discussed in the review had drawbacks, such as small sample numbers or a regional emphasis, which would have prevented them from accurately representing the range of circumstances seen in other papaya-growing locations. To ensure the generalizability of findings, bigger and more varied sample sizes should be the goal of future studies.

4. Obstacles to the Adoption of Integrated Pest Management: The review does not go into great depth on the particular difficulties that fruit farmers have in implementing integrated pest management (IPM) techniques. Subsequent research endeavors may delve further into these obstacles and suggest efficacious remedies to foster the use of sustainable pest management methodologies.

5. Comprehensive knowledge of Sustainable Production: Although sustainable agricultural techniques are included in the review, further investigation is required to give a more thorough knowledge of sustainable papaya production. This includes investigating the viability of switching to organic agricultural techniques, analyzing alternate sources of nutrients, and gauging the effects of various land use strategies.

By filling up these gaps, future studies will be able to improve agricultural sustainability, advance our understanding of papaya agriculture, and encourage the creation of more productive and eco-friendly farming methods.

1.6 Report Organization

The proposed report is structured with a comprehensive layout to guide readers through the research process, findings, and implications. Each chapter serves a specific purpose, contributing to the overall coherence and depth of the document

Chapter 1: INTRODUCTION

The Overview of IoT-Enabled Soil and Plant Nutrient Management System for Papaya Cultivation. The purpose of this document is to highlight the importance of IoT technology and machine learning models in improving the efficiency of papaya production in Bangladesh.

Chapter 2: LITERATURE REVIEW

Examines similar works, contrasts previously published works, points out unresolved difficulties, and provides an overview of the literature. In addition to reviewing earlier research projects and doing a comparative examination of techniques, findings, and consequences, it introduces key vocabulary and concepts. The section on the problem's scope describes the particular difficulties encountered while managing plant and soil nutrients in the context of papaya farming, setting parameters for the investigation and highlighting important topics for study.

Chapter 3: METHODOLOGY/ REQUIREMENT ANALYSIS & DESIGN SPECIFICATION

Methodological Approach for Soil and Plant Health Systems Based on the Internet of Things (IoT) The purpose of this document is to comprehensively describe the research strategy, data-gathering methods, model architecture, and performance evaluation criteria used in the study.

Chapter 4: IMPLEMENTATION

This chapter focuses on putting the suggested system into practice. A thorough description of the experimental design, findings, and analysis can be found in the chapter on the application of an Internet of Things-based soil and plant nutrition management system for papaya farming. Technologies, instruments, and equipment are all part of the experimental setting. The system's performance and interpretations are shown in the results and analysis section. The discussion part provides insights and perspectives on the results of the implementation phase while critically analyzing the findings.

Chapter 5: RESULT AND ANALYSIS

The study aims to assess the performance of advanced models, including ARIMA, SARIMAX, Random Forest, XGBoost, Linear Regression, Logistic Regression, and Multilinear Regression, in the management of soil and plant nutrient levels.

Chapter 6:IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

The societal impact chapter explores the broader implications of the research findings on environmental effect analysis. It discusses how improving nutrient management detection methods can support the idea of environment preservation and combating the malpractices of environmental degradation.

Chapter 7: CONCLUSION AND FUTURE WORK

This final chapter serves as a synthesis of the entire report. It summarizes the key findings, reiterates the conclusions drawn from the research, and offers recommendations for practical applications and further studies. The implications for future research are discussed, providing a roadmap for researchers interested in expanding upon the current study.

1.7 Summary

The research project presents a comprehensive analysis of the importance of agriculture in the economy of Bangladesh. It specifically examines the difficulties encountered in papaya cultivation, including issues connected to soil quality, nutrient management, and environmental sustainability. The chapter presents the utilization of Internet of Things (IoT) technologies and machine learning algorithms to tackle these difficulties and enhance agricultural practices. The text highlights the rationale for the study, highlighting the necessity for inventive approaches to improve sustainability, productivity, and environmental stewardship in papaya farming. The study discusses the justification, emphasizing the deficiency in existing agricultural methods and the potential advantages of combining IoT technologies with predictive modeling techniques such as ARIMA, SARIMAX, Random Forest, XGBoost, Linear Regression, Logistic Regression, and Multilinear Regression. The research aims to explore the potential of IoT-based solutions.

CHAPTER 2
LITERATURE REVIEW

2.1 Overview

In an IoT-based soil and plant nutrient management device for papaya production in Bangladesh, advanced predictive models and performance metrics are used to optimize the practices and grow the yield. IoT plays a key function in this system by accumulating real-time information through sensors that reveal various soil and environmental parameters to make specific and timely interventions. ARIMA (Autoregressive included shifting common) and SARIMAX (Seasonal ARIMA with Exogenous Variables) are used for time series evaluation to forecast the future soil moisture and nutrient ranges so that control may be carried out primarily based on historical statistics and seasonal developments. Random forest and XGBoost (severe Gradient Boosting) are systems getting to know algorithms that enhance the prediction accuracy and robustness; Random forest does this by constructing a couple of choice timber to seize the complex and non-linear relationships within the facts and XGBoost does this by pretty green implementation of gradient boosting strategies. The following is a synopsis of each publication's principal conclusions and topics. To promote soil fertility and nutrient availability, Paper [7] emphasizes the importance of sustainable agriculture management strategies such as cover crops, green manure, and animal manure. It also covers the challenges Hawaiian papaya growers face as a result of unsustainable land-use practices that reduce papaya productivity.

In Paper [13], the use of machine learning algorithms for crop suggestions and soil nutrient monitoring via the Internet of Things is investigated. These algorithms include CatBoost and Random Forest. It highlights how well these models function at providing exact guidance to farmers, albeit there is still room for improvement, including the inclusion of meteorological and satellite images.[3] investigates a variety of topics linked to technology, healthcare, education, and agriculture. One such issue is the application of machine learning algorithms like Support Vector Regression (SVR) and Recurrent Neural Networks (RNN) to analyze plant growth metrics, nutrient content, and soil fertility. It also discusses the difficulties and constraints that come with data analysis and research in agriculture. Paper[1] uses machine learning models such as SVM, Naïve Bayes, Random Forest, and Extreme

Gradient Boosting to estimate soil nutrients. It emphasizes how these elements influence soil composition, crop forecasting, and plant disease detection.[11], which also covers production methodologies, environmental factors, and optimization tactics for increasing pawpaw productivity and growth. It highlights inadequacies in current farming systems and research gaps. Paper [9] uses data from research papers and other sources to evaluate papaya farming techniques and grapevine disease resistance. It discusses utilizing machine learning techniques to investigate agricultural practices but does not specify which model. Also, the gadget video displays units of the percentage of important soil vitamins like Nitrogen (N), Phosphorus (P), and Potassium (k) together called NPK. Those nutrients are important for plant boom and improvement, and retaining the top-rated degree of those vitamins is necessary to maximize papaya yield and nice. By integrating these superior analytics with IoT-based tracking the machine affords a complete solution for soil and plant nutrient control and thus contributes to sustainable and worthwhile papaya manufacturing in Bangladesh.

2.2 Related Works

The investigations included a wide range of topics related to the production and culture of papayas, with particular attention to the effects of water scarcity on plant growth and fruit development, as well as environmental impacts, nutritional requirements, and improving soil quality. Among the particular linked works that are highlighted in the literature study are:

The research highlights the significance of Glomus mosses as an efficient species in augmenting mineral nutrition in Carica papaya var. Surya by examining the effects of arbuscular mycorrhizal fungus on mineral nutrition. The study highlights the importance of nitrogen and potassium for crop growth and productivity by examining how planting spacing and NK fertilization affect physiological indicators and fruit yield in papaya grown in a semiarid environment. Studies on the production of carbon quantum dots from Zelkova serrata leaves are being conducted for a variety of uses, such as the detection of normetanephrine in elderly plasma samples and the prevention of Staphylococcus aureus infection. The following is a synopsis of each publication's principal conclusions and topics. To promote soil fertility and nutrient availability, Paper [7] emphasizes the importance of sustainable agriculture management strategies such as cover crops, green manure, and animal manure. It also covers the challenges Hawaiian papaya growers face as a result of unsustainable land-use practices that reduce papaya productivity.

In Paper [13], the use of machine learning algorithms for crop suggestions and soil nutrient monitoring via the Internet of Things is investigated. These algorithms include CatBoost and Random Forest. It highlights how well these models function at providing exact guidance to farmers, albeit there is still room for improvement, including the inclusion of meteorological and satellite images.[3] investigates a variety of topics linked to technology, healthcare, education, and agriculture. One such issue is the application of machine learning algorithms like Support Vector Regression (SVR) and Recurrent Neural Networks (RNN) to analyze plant growth metrics, nutrient content, and soil fertility. It also discusses the difficulties and constraints that come with data analysis and research in agriculture. Paper [1] uses machine learning models such as SVM, Naïve Bayes, Random Forest, and Extreme Gradient Boosting to estimate soil nutrients. It emphasizes how these elements influence soil composition, crop forecasting, and plant disease detection.[11], which also covers production methodologies, environmental factors, and optimization tactics for increasing pawpaw productivity and growth. It highlights inadequacies in current farming systems and research gaps. Paper [9] uses data from research papers and other sources to evaluate papaya farming techniques and grapevine disease resistance. It discusses utilizing machine learning techniques to investigate agricultural practices but does not specify which model. This research encourages agricultural practices by combining machine learning, sustainable management strategies, and cutting-edge crop production and soil monitoring technology.

These connected publications cover many elements of crop development, nutrient absorption, and creative uses for agricultural methods, all of which provide significant insights into the field of papaya cultivation and production.

2.3 Comparison between existing works

The literature review highlights the functions of numerous nutrients and environmental elements in maximizing crop yields while discussing many facets of papaya production, cultivation, and irrigation techniques. A comparison of a few of the current works that were addressed in the evaluation is provided below:

1. Papaya Fruits and Calcium Treatments' Effects:

In research on the postharvest quality of papaya fruits and preharvest calcium treatments, Madani et al. (2016) showed that fruit quality may be improved by preharvest calcium applications, especially when using calcium chloride ($CaCl_2$).Another Study: Examined the effects of various calcium treatment sources and concentrations on the postharvest quality and nutritional concentrations of papaya fruits. ANOVA and Duncan's multiple range test were used in the statistical analysis to compare the means.

2. Application of Nutrients and Biometric Measurements:

Research Employing Randomized Block Design: included applying six equal split doses of varying concentrations of potassium, phosphorus, and nitrogen, and then measuring yield and biometric traits along with fruit quality measures. Conventional farming methods were applied for cultivation and fertilization, highlighting the significance of sustainable nutrient management techniques in the papaya industry.

3. Nigerian Papaya Production:

Review of constraints and Strategies: This study examined productivity-affecting variables, pawpaw production techniques' constraints, and crop development and management strategies to maximize results in southwest Nigeria. Suggestions for Fruitful Cultivation: highlighted the advantages for fruit farmers in terms of improved economic performance and the significance of crop rotation, early interplanting with annuals, and improving management techniques.

4. Obstacles and Suggestions Regarding Papaya Production:

The preservation of organic matter in the soil is emphasized as a crucial element in improving soil fertility and nutrient availability for the sustainable production of papayas. Investigating Inexpensive Organic Techniques: To increase soil health and nutrient availability, "natural farming" is advised; this highlights the significance of sustainable agricultural management techniques. It is clear from a comparison of these published publications that research on papaya production covers a wide range of topics, including economic concerns, sustainable methods, fruit quality enhancement, and nutrient management.

Table 2.2.1. Comparison between our proposed work and existing literature

Paper	Model	Features	Nutrients	Limitations

Bilal, Anas, [29]	1. Support Vector Machine (SVM) with radial basis function (SVM-RBF) 2. Naïve Bayes 3. Logistic Regression 4. K-Nearest Neighbors (KNN)	1. Weather conditions 2. Soil analysis 3. Amount of fertilizers and pesticides required 4. Plant seedlings 5. Pests 6. Plant leaf diseases	1. Nitrogen (N) 2. Phosphorus (P) 3. Potassium (K) 4. Organic carbon 5. Boron	1. Data Availability 2. Model Interpretability 3. Generalizability 4, Validation Methods 5. Implementation Challenges 6. Ethical Considerations
Olubode, Olusegun [14]	1. Support Vector Machines (SVM) 2. Artificial Neural Networks (ANN) 3. Random Forest (RF) 4. Linear Models (LM)	Not mentioned	1. Nitrogen (N) 2. Phosphorus (P) 3. Potassium (K) 4. Calcium (Ca) 5. Sodium (Na)	1. Limited focus 2. Lack of recent data: 3. Language barrier: 4. Bias: 5. Generalization
Said, Zafar, [28]	Recurrent Neural Network (RNN) Artificial Neural Networks (ANNs) Linear and logistic regression, k-nearest neighbors, decision trees, random forest,	1/Environmental Factors 2/Operating Variables: 3/Sensor Data 4/Genetic Data:	Lactate Glucose	1. Limited Scope 2. Data Availability: 3. Model Generalization 4. Complexity of Models: 5. Resource Requirements: 6. Validation and Benchmarking.
Dahunsi, [27]	RNN Random forest	Nutrient Content pH Levels**:** Microbial Composition**:**	NPK	1, Limited Scope 2, Short Duration:

		Organic Matter Content: Plant Growth Parameters		3, Environmental Factors: 4, Statistical Analysis: 5. Replicability
Aldillah, Rizma. [15]	wide range of topics in various fields such as agriculture, healthcare, education, and technology,	Not mentioned	Copper, NPK	Limited Scope: Data Limitations: Methodological Constraints Publication Bias: Regulatory and Enforcement Challenges:
Bindu, B [17]	RNN SVM	1. Confounded factorial randomized block design 2. Application of different levels of nitrogen, phosphorus, and potassium in six equal split doses 3. Measurement of biometric characters, yield characters, and fruit quality parameters 4. Use of traditional agricultural practices for fertilization and cultivation	nitrogen, phosphorus, and potassium.	1. Environmental Factors: 2. Generalizability: 3. Duration of Study: 4. Statistical Analysis
Olubode, [1]	The paper does not explicitly mention the use of specific models in the research.	The Paper" does not specifically focus on the detailed analysis of nutrients in pawpaw or the nutritional aspects of the fruit. Instead, it primarily discusses the influence of environmental factors	This research reviews production limits and suggests techniques for optimal crop management, focusing on the effects of environmental conditions and production practices on pawpaw growth and	1. Little is known about pawpaw cultivar canopy variance. 2. Limited research on soil types and pawpaw root systems. 3. A lack of

			productivity in southern Nigeria.	information about pawpaw productivity and environmental variables. 4. Hardly much talk of grower issues. 5. Not enough research is done on creative pawpaw-producing methods.

2.4 Open Issues

The literature analysis has identified some crucial unresolved matters that necessitate additional investigation and scholarly inquiry. There is a lack of extensive research that particularly examines the combination of Internet of Things (IoT) technology with machine learning algorithms in the context of papaya cultivation in Bangladesh. There is a paucity of comprehensive research that examines the combined effects of IoT and predictive modeling on soil and plant nutrient management, however, some individual studies have explored some parts of these topics. This gap provides an opportunity for future studies to further investigate the combined effects of these technologies on improving agricultural practices and increasing crop output. Additionally, the literature review has emphasized the necessity for additional empirical data on the effectiveness of predictive modeling methods, including ARIMA, SARIMAX, Random Forest, XGBoost, Linear Regression, Logistic Regression, and Multilinear Regression, in enhancing soil and plant nutrient management in papaya fields. Although there are theoretical frameworks, it is crucial to conduct field experiments and analyze data to empirically assess the practical usefulness of these models in real-world agricultural situations. By addressing this gap, we can enhance our understanding of how predictive modeling can be utilized effectively to enhance agricultural results. Furthermore, the obstacles and difficulties that arise in the effective execution of IoT-driven solutions in agricultural environments have been recognized as an unresolved matter. Challenges that must be tackled include technology infrastructure, data privacy and security concerns, farmer adoption rates, and the ability of IoT systems to scale in various farming situations. It is essential to investigate solutions to overcome these problems and guarantee the smooth incorporation of IoT technology in agriculture.

2.5 Summary

The literature review provides a thorough overview of applying sustainable agricultural methods and machine learning algorithms to improve crop productivity, with a particular emphasis on papaya farming . Important topics discussed include the need to maintain soil organic matter, put sustainable agriculture management plans into practice, look into low-cost organic production techniques, and assess substitute nutrient sources to enhance soil health and nutrient availability. The assessment also discusses issues that papaya growers must deal with, such as declining soil fertility brought on by unsustainable techniques, and offers suggestions for enhancing papaya production systems. In summary, the literature review highlights the significance of research, innovation, and sustainable practices for enhancing agricultural output, enhancing crop quality, and guaranteeing the farming systems' long-term sustainability. The research highlights the possibility of utilizing technologies such as ARIMA, SARIMAX, Random Forest, XGBoost, Linear Regression, Multilinear Regression, and Logistic Regression to improve agricultural productivity and sustainability, based on the synthesis of important findings from earlier studies. The review focuses on key research that investigates problems including water scarcity, environmental implications, nutritional requirements, and soil quality improvement. The article also examines the utilization of machine learning models such as Support Vector Machines (SVM), Naïve Bayes, Random Forest, and Extreme Gradient Boosting to estimate soil nutrient levels, predict crop yields, and detect plant diseases. The text explores challenges and recommendations related to papaya production, including the preservation of organic matter in the soil and the significance of "natural farming." A comparison between the proposed work and the current body of literature on IoT technology and predictive modeling in the management of soil and plant nutrients.

CHAPTER 3

METHODOLOGY/ REQUIREMENT ANALYSIS & DESIGN SPECIFICATION

3.1 Overview

This takes a look at specializes in growing and comparing an IoT-primarily based soil and plant nutrient control system for optimizing papaya production in Bangladesh. This device uses several predictive fashions: ARIMA, SARIMAX, Random forest area, XGBoost, Linear Regression, Logistic Regression, and Multilinear Regression for analyzing and forecasting soil and plant nutrient degrees. actual-time subject information on essential vitamins such as Nitrogen (N), phosphorus (P), and Potassium (k) is gathered from sensors deployed within the area via an IoT platform for specific tracking and timely interventions. Then, in this study, the effectiveness and precision of those predictive algorithms are also evaluated through the usage of different overall performance metrics like suggested Squared errors (MSE), Root mean Squared mistakes (RMSE), Accuracy, MAPE, and R-squared (R^2). the sort of holistic approach will offer actionable insights for the optimization of the nutrient control regime for better boom and yield of papaya produce inside the region.The overall performance of these models is evaluated using numerous metrics. Mean squared errors (MSE) and Root imply Squared errors (RMSE) offer measures of prediction accuracy with the aid of quantifying the average squared differences among located and expected values. Accuracy is used to evaluate the share of correct predictions in type obligations, offering insights into the reliability of the version's predictions. (MAPE) is employed to examine the accuracy of forecasts across extraordinary scales, imparting a normalized degree of error. R-squared (R^2) price suggests the percentage of variance in the established variable that is predictable from the unbiased variables, serving as a degree of the way nicely the model fits the information.

IoT sensors can be deployed in papaya fields in Bangladesh to capture real-time data on soil and plant health indicators. Soil moisture sensor: measure soil moisture. Determine the concentration of important vitamins such as nitrogen, phosphorus, and potassium in the soil. The foundation of data collection will be used to focus on collecting, processing, and sending data from IoT sensors to master data or cloud platforms. Access data as a pipeline for predictive analysis of soil and plant nutrient availability. This learning tool will analyze historical data from sensors to predict soil moisture, nutrient levels, and the fate of different traits, enabling control processes. This machine functions as soil moisture sensors, NPK sensors, and climate stations set up inside the papaya fields, all placed at gold-standard positions. Soil moisture sensors take continuous readings of water content in the soil, even as NPK sensors record the ranges of the critical vitamins Nitrogen (N), Phosphorus (P), and Potassium (k), which are critical to plants. which together form critical factors in the health and nutrient consumption of plant life. the consequent facts fed from the sensors are then transmitted wirelessly over conversation networks onto a valuable database in near real-time.

3.2 Proposed Methodology/ System Design

The first step of research is to collect data. We can work with this information. the environment changes. The model was examined using historical data on topography, fertilizer use, and papaya growth parameters. quality dreams can be, for example, to optimize nutrient management through the right soil moisture content material, NPK (Nitrogen, Phosphorus, Potassium) concentrations, and several others that are important within the prediction for papaya growth. To create the dataset we collect data from papaya felid and we collect real-time data for this research. In our papaya felid, there are 36 total papaya trees we collect each tree's height every week. Every week we collect the N, P, K, and moisture values for each tree, for better production we use organic fertilizer for tree growth. For N, P, K, and moisture detection we used some equipment. We collect the soil of plants and with the NPK sensor, we collect our data and automatically store it in the cloud. For each tree, we noted the height, N, P, K, and moisture values on the dataset with a mention of the date. Like this, we collect 14 weeks of data from the field. And notify this the tree height increases day by day based on the N, P, and K values and moisture.

Our Dataset is based on real-time data. Here we define height as a target value. Date, moisture, N, P, and K values change the height.

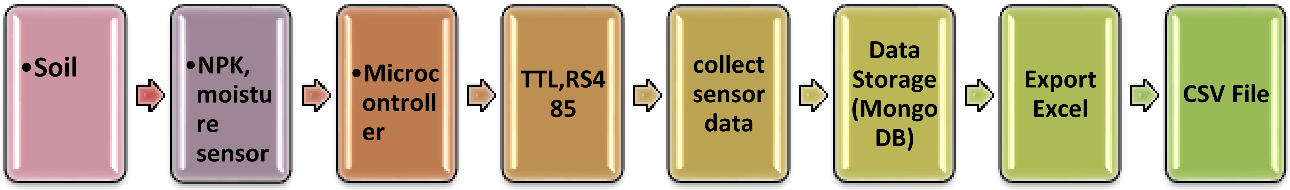


Figure 3.2.1: Data Flow Chart

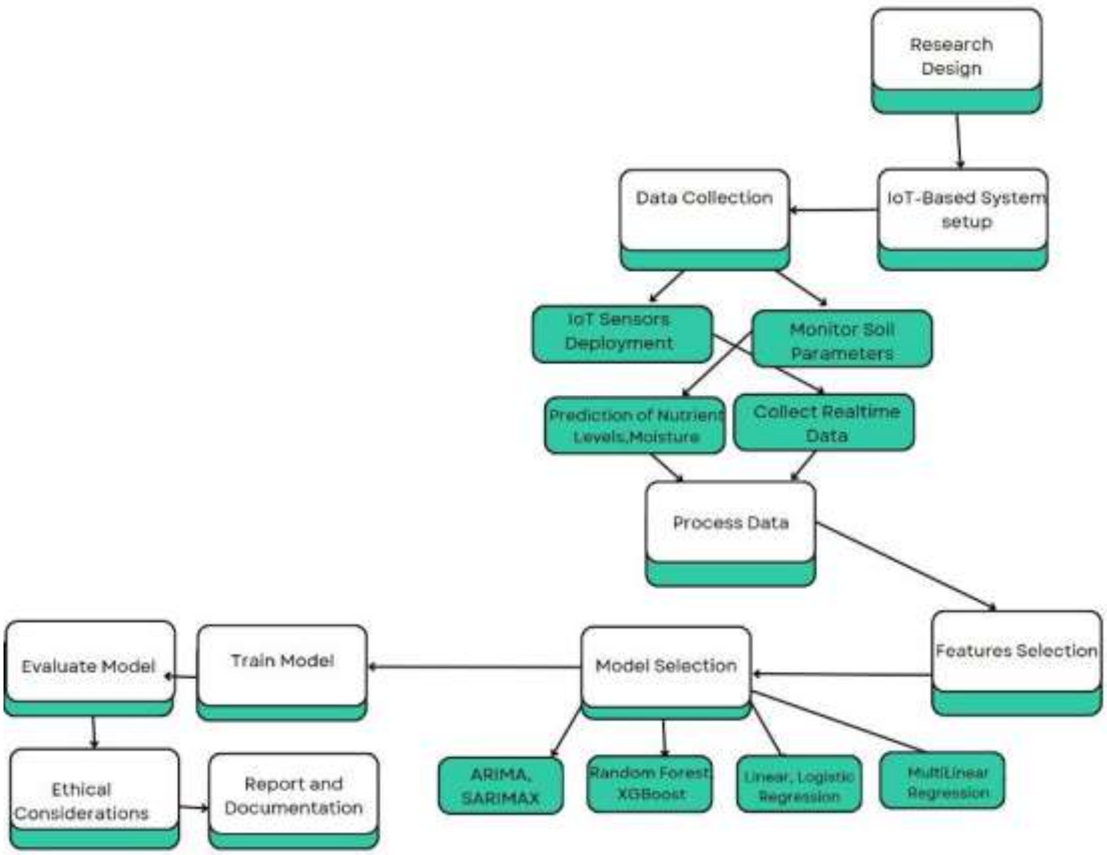


Figure 3.2.2: Methodology Process

IoT device Prototype: A preliminary version of an Internet of Things system that is intended to monitor NPK and moisture values in agriculture is referred to as an IoT prototype. To collect, transmit, and analyze NPK, moisture data, this prototype typically comprises communication modules, microcontrollers (such as Arduino), and sensors (such as NPK, moisture sensors). The objective is to evaluate and enhance the integration of sustainable agricultural practices with embedded technology before its full-scale implementation.

Figure 3.2.3: IoT Device Prototype

Accuracy:

Accuracy_score and other accuracy measures are frequently used in classification tasks where the model predicts class separations.

Accuracy=TP+TN/TP+TN+FP+FN

fashions like ARIMA, SARIMAX, Linear Regression, and Multilinear Regression will expect non-stop variables like soil moisture and NPK ranges. at the same time as not normally described through an accuracy percentage, their predictive performance will be assessed using MSE, RMSE, and R^2

MSE:

suggest Squared error (MSE) is a typically used metric to degree the average of the squares of the mistakes—that is, the average squared difference among the expected and the real values. it's miles used for evaluating the overall performance of regression fashions.

The components for calculating MSE are:

MSE=n1∑i=1n(yi−y^i)2

RMSE:

Root mean Squared error (RMSE) is a widely used metric to compute the error that the anticipated values preserve towards the actuals. it's by far the root of the mean squared errors (MSE) and represents the standard mistakes for the perfectness of the model used in want of the facts.

RMSE=n1∑i=1n(yi−y^i)2

R^2:

R-squared (R^2) is a statistical measure that represents the proportion of variance in an established variable this is defined by way of an independent variable or variables in a regression version. it is, consequently, an at-a-look gauge of how appropriate the impartial variables give an explanation for variability in the structured variable.

R2=1− SSres/ SStot

MAPE:

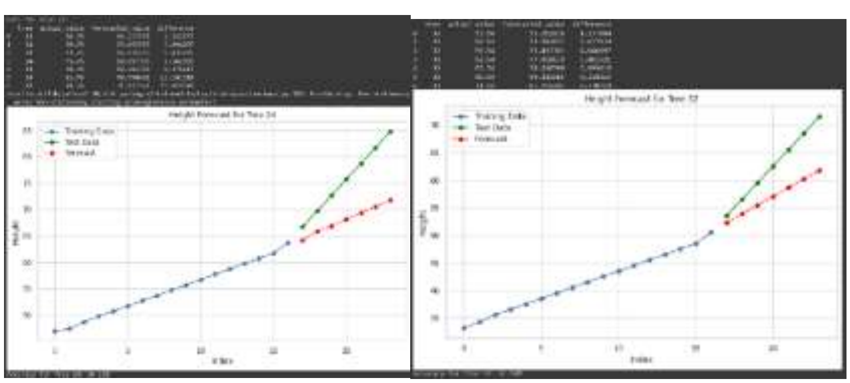
MAPE is just one such measure. It just measures the common value of absolutely the percentage errors in the forecast. Alternatively, seeing that it is a mean, any severe values or deviations will once more naturally common out so the smaller the cost, the higher the forecast.

MAPE=n1∑i=1nyiyi−y^i×100

ARIMA MODEL

ARIMA models, which might be used for time series forecasting. ARIMA models can be used for time collection forecasting. The convolutional ARIMA model is especially useful for time series statistics with complicated structures and dependencies that traditional ARIMA fashions through themselves can't properly seize.

Here we find out the actual value and predicted values and the test data, train data, and the forecast data. And the Accuracy for each tree.



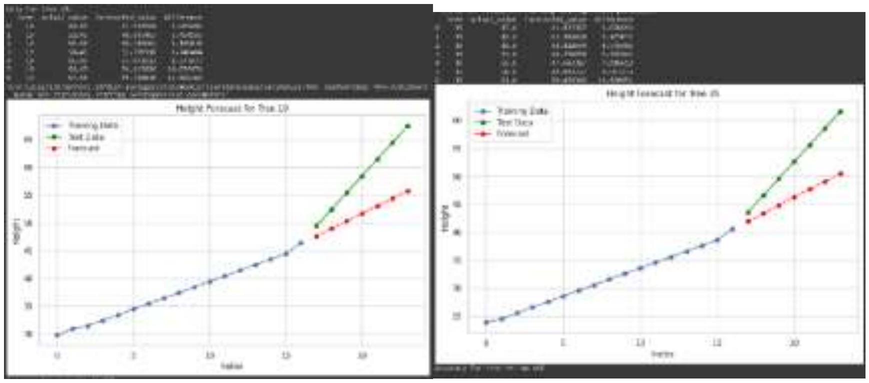


Figure 3.2.4: Time series with Arima model

SARIMAX MODEL

SARIMAX(p,d,q)(P,D,Q,s)within the case of using the SARIMAX model to predict the levels of soil moisture and NPK concentrations from the IoT-based soil and plant nutrient management machine developed for papaya manufacturing, the critical additives related to the variables stated are shown. this could make the version more potent whilst integrated with seasonal tendencies and outside factors like weather situations, irrigation schedules, and fertilization plans, making better control decisions. The empirical assessment of the model becomes finished with the usage of measures like implied Squared mistakes (MSE) and Root implied Squared error (RMSE).

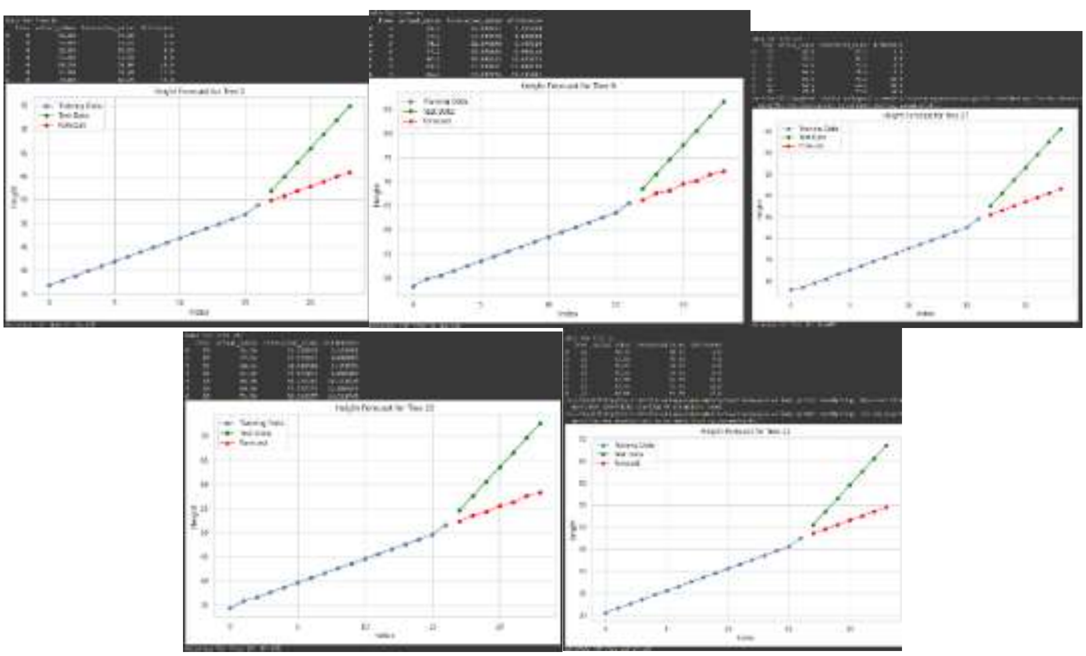


Figure 3.2.5: Time series with SARIMAX model

XGBoost MODEL

XGBRegressor is a regression implementation of the XGBoost algorithm, one of the maximum popular and effective gradient boom frameworks used for gadget studying projects it is particularly effective for structured/tabulated data and is understood for its speed, scalability, and high performance additionally, you may use regression metrics along with common imply squared blunders (MSE), root imply squared blunders (RMSE), to assess model overall performance.XGBRegressor is a regression implementation of the XGBoost set of rules, one of the maximum famous and powerful gradient boom frameworks used for system mastering projects it's far particularly powerful for structured/tabulated facts and is understood for its velocity, scalability, and excessive performance additionally XGBRegressor

Liner regression

Linear Regression is a statistical approach to model the relationship that exists between the established variable and the one or extra impartial variables with one or more impartial variables. In its most truthful form, the easy linear regression will occur with a single unbiased variable, even as the usage of multiple independent variables will take vicinity within a couple of linear regressions. Linear regression with the input from the soil moisture ranges and NPK concentrations may be applied for modeling the relationships between these unbiased variables to the climate situations, irrigation schedules, and fertilization plans for an IoT soil and plant nutrient management system in papaya manufacturing in Bangladesh.

Logistic regression

Logistic regression is this type of statistical method: it models the opportunity of a given enter point belonging to a particular magnificence in a binary type trouble, for which the output is a specific variable with feasible consequences. The system can classify and even predict the critical conditions affecting papaya production under Logistic Regression to fill great development crop management practices for efficient productiveness.

Random forest

This uses the knowledge of the gang for better predictions and avoids overfitting by using the approach. A Random forest model can be educated for precise impartial climate, irrigation schedules, and fertilization plans for the manufacturing procedure in measuring the share of water content material and NPK in IoT-based soil and plant nutrient control for papaya production in Bangladesh.

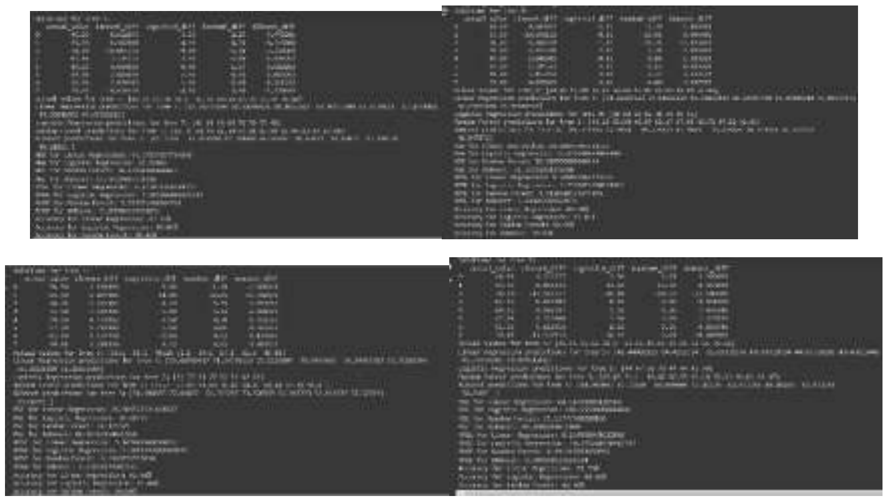


Figure 3.2.6: MSE. RMSE and accuracy per tree with regression model

Multilinear regression

It's far an extension of easy linear regression, which is also referred to as multiple Linear Regression, because right here an established variable is modeled on multiple impartial variables. It does so by fitting a linear equation to determine information—that means, in other words, it predicts the price of the variable based on the values of independent variables. Inside the framework of soil moisture, Infiltration, and NPK percent, this IoT-based moisture and plant nutrient management device helps the correct multilinear regression of the device maintaining the unbiased variables in take a look at, such as the weather situation, irrigation, and fertilizer schedule.

3.3 Hardware/ Software Requirement

The requirement analysis for an IoT-based soil and plant nutrient management system in Bangladeshi papaya production includes defining critical functionalities that solve the region's specific issues. Real-time monitoring of soil characteristics such as moisture, and nutrient levels, as well as plant health indicators such as leaf color and growth rate.

Hardware Requirements:

- **High-performance computing resources:**: They are used for processing and analyzing data, effectively meeting computational requirement projects.
- **NPK Monitor Sensors:** The sensor gives momentary results on the nitrogen, phosphor, and potassium levels in soil—nutrients that are important for papaya growth.
- **RS-485 Communication (Rs485):** Enable reliable data transferred over long distances between the sensor and central microcontrollers, assuring smooth system communication.
- **TTL to RS-485 Converters:** These converters facilitate the seamless exchanging of data by assuring compatibility across various communication standards used by sensors and devices.
- **Arduino Microcontrollers:** Function as a central processing unit for the system, handling related tasks to data processing, communication, and collection.
- **Version Control System, LCD Server, Collaboration Tools, and Testing Tools:** These tools work to facilitate data visualization, testing, version tracking, and communication throughout project setup.

Software Requirements:

- IoT platform for managing and gathering data.
- **Python programming language:** Python provides a versatile and widely-used platform for implementing Machine learning algorithms and conducting data analysis.
- Frameworks for machine learning like Skitch-learn and TensorFlow for construct models.
- Machine learning algorithms implementable using programming languages such as Python.

3.4 Project Management and Financial Analysis

Project Management

• Project Objectives:

1. To develop an IoT-based system for soil and plant nutrient management tailored

for papaya production in Bangladesh.

2. To enable real-time monitoring of soil conditions and nutrient levels to optimize papaya growth and yield.

3. To provide farmers with actionable insights for efficient nutrient management and improved papaya production.

• Project Timeline:

A. Phase 1: Project Planning and Research (2-4 weeks)

B . Phase 2: Sensor Integration and Field Testing (6-8 weeks)

C. Phase 3: IoT Platform Development (8-12 weeks)

D. Phase 4: Data Analytics and Machine Learning Implementation (10-14 weeks)

E. Phase 5: Testing and Calibration (4-6 weeks)

F. Phase 6: Deployment and Training (2-4 weeks)

G. Phase 7: Monitoring and Support

• Resource Planning:

A. Equipment and Tools:

- IoT Sensors for Soil Moisture, and Nutrient Levels
- Microcontrollers for Sensor Data Collection
- High-performance Computers for Data Processing
- Collaboration Tools
- Version Control System
- Testing Tools

-TTL convert

-RS485

-Arduino

-LCD server

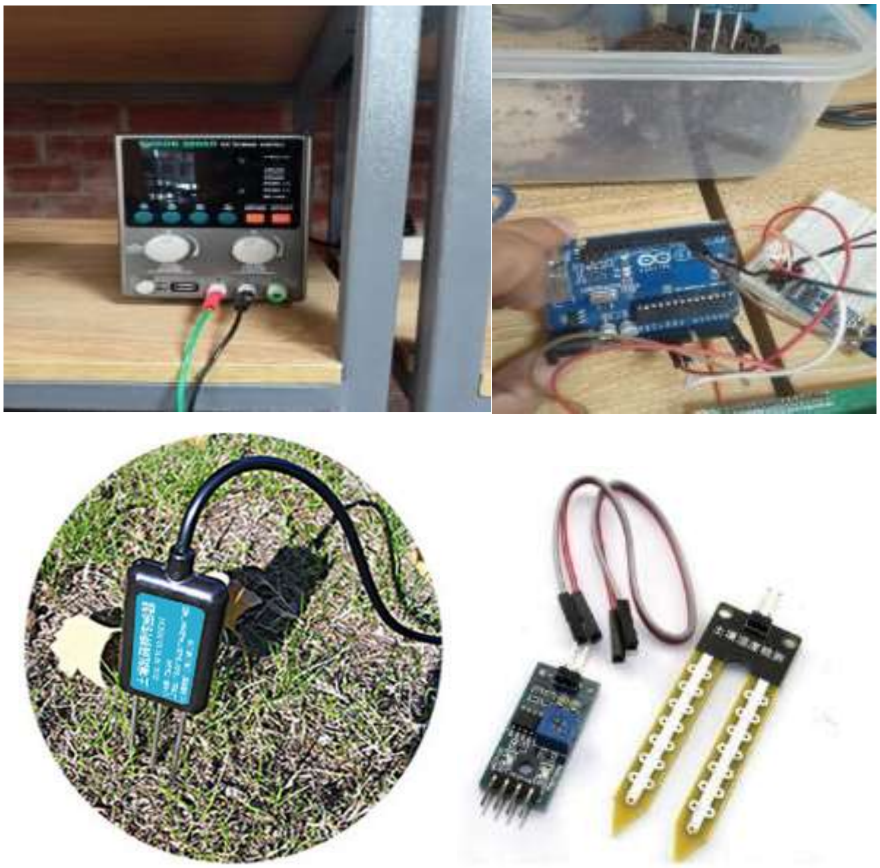


Figure 3.4.1: TOOLS

B. Software and Technologies:

- Database Management System
- Machine Learning Frameworks

C. Data and Content:

- Sensor Data: Collected from IoT devices installed in papaya fields
- Papaya Farm Information: Geographic and soil data.
- User Documentation: Prepared by the technical writing team

D. Training and Skill Development:

- Ensure that the development team has the necessary skills in IoT development, data analysis, and machine learning.
- Provide training on specific technologies and tools as needed for effective project execution.

• Risk Management:

1. Technical Risks:

Risk: Sensor malfunction or data inaccuracies.

Mitigation: Implement regular sensor calibration and maintenance procedures. Have redundancy in sensors for critical measurements.

2. Environmental Risks:

Risk: Extreme weather conditions affecting sensor performance or crop growth.

Mitigation: Design sensors and installations to withstand harsh environmental conditions. Implement weather monitoring systems for early warning.

3. Data Security Risks:

Risk: Unauthorized access to sensitive agricultural data.

Mitigation: Implement encryption, access controls, and regular security audits to safeguard data integrity and privacy.

4. Adoption Risks:

Risk: Low adoption rates among farmers due to lack of awareness or resistance to technology.

Mitigation: Conduct farmer awareness programs and provide hands-on training. Demonstrate the benefits of the system through pilot projects.

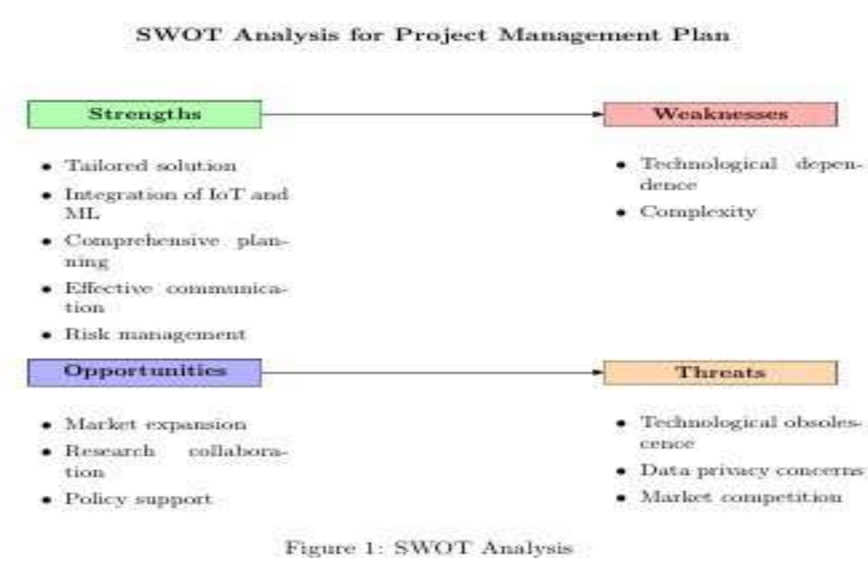


Figure 3.4.2: SWOT Analysis

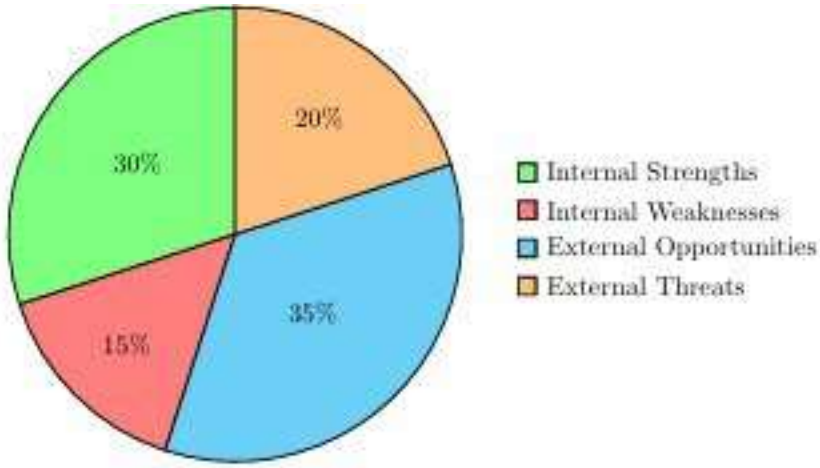


Figure 3.4.3: SWOT Analysis (Pie Chart)

• Communication Plan:

A. Stakeholder Meetings:

Purpose: Regular meetings to update stakeholders on project progress and gather feedback.

Participants: Project team members, farmers, agricultural experts, and government representatives.

Frequency: Bi-weekly meetings during project execution, with additional meetings as needed.

B. Progress Reports:

Purpose: Provide periodic progress reports to stakeholders via email or presentations.

Content: Include updates on milestones achieved, challenges faced, and upcoming activities.

Table 3.4.1: Estimated Cost for IOT-based Project

SN	Components	Estimated Cost(BDT)
1.	Tools and Equipment	1000-2000
2.	Visiting Stakeholders	500-1000
3.	Power Supplies and Batteries	500-1000
4.	Data Collection and Processing	500-1000
5.	Maintenance and Support	500-1000

6.	Documentation and Report Writing	400-800
7.	Contingency (10% of total)	1000-1500
	Total Estimated Cost	4400-8300

3.5 Summary

A prototype of an IoT-based comprehensive system for soil and plant nutrient control in papaya cultivation is under development. advanced predictive modeling courses the actual-time sensor statistics. The models used are ARIMA, SARIMAX, Random wooded area, XGBoost, Linear Regression, Logistic Regression, and Multilinear Regression, with the intention of surest—with the aid of adaption of agricultural practices—for maximal crop yield. The model performance would be evaluated based totally on MSE, RMSE, Accuracy, MASE, and R^2 price, which would imply the reliability and effectiveness of the model within the management of soil and plant vitamins. first of all, set up IoT sensors inside the papaya fields that will help you to screen actual-time soil moisture content, as well as NPK concentrations supplemented with historical facts collection for time collection forecasting. diverse system learning algorithms which include ARIMA and SARIMAX which are used for time series prediction, Random woodland or XGBoost that cope with complicated non-linear relationships amongst others may be carried out with statistical fashions like Linear Regression, Logistic Regression, or Multilinear Regression being used for baseline prediction together with type duties. in addition, model assessment has to be carried out based on suggested squared blunders(MSE), root implies squared mistakes(RMSE), accuracy, suggested absolute scaled error(MASE), and r-squared(R2) values wherein cross-validation techniques ensure version robustness. After successful training install these models onto the cloud platform for them to make actual-time predictions which may be accessed by using farmers via an interface that is straightforward to apply. Gadget accuracy must be stored on take a look at through continuous tracking and adjustment while education applications collectively with support infrastructure would be there to help farmers utilize the gadget efficaciously hence optimizing papaya irrigation schedules coupled with fertilization durations so that you can increase manufacturing.

CHAPTER 4
IMPLEMENTATION

4.1 Overview

A. The Datasets

To create the dataset we collect data from papaya felid and we collect real-time data for this research. In our papaya felid, there are 36 total papaya trees we collect each tree's height every week. Every week we collect the N, P, K, and moisture values for each tree, for better production we use organic fertilizer for tree growth. For N, P, K, and moisture detection we used some equipment. We collect the soil of plants and with the NPK sensor, we collect our data and automatically store it in the cloud. For each tree, we noted the height, N, P, K, and moisture values on the dataset with a mention of the date. Like this, we collect 14 weeks of data from the field. And notify this the tree height increases day by day based on the N, P, and K values and moisture.

Our Dataset is based on real-time data. Here we define height as a target value. Date, moisture, N, P, and K values change the height.

Figure 4.1.1: Data collection

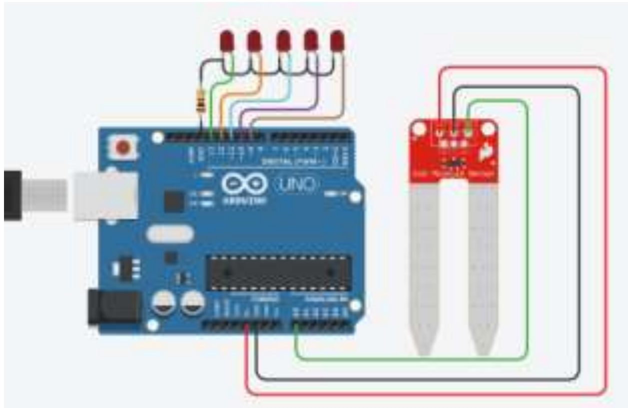


Figure 4.1.2: Soil moisture sensor circuit diagram

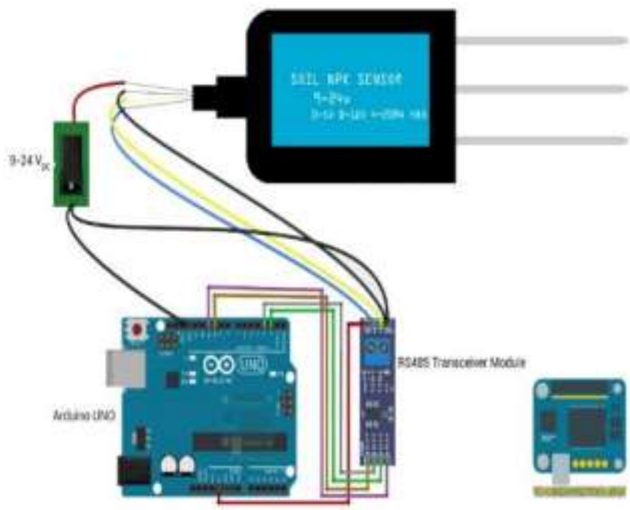


Figure 4.1.3: Soil NPK sensor circuit diagram

B. Process of Experiments

After collecting the data for the dataset we need to preprocess the data. Here we use types of ML algorithms XGBoost version (XGBRegressor) and ML ARIMA model and SARIMAX Algorithm for time series, multivariant analysis. Random forest, XGBBOOST, Linear Regression, Logistic Regression, and multilinear Regression to find out Mean Squared Error, RMSE, ACCURACY, MAPE, and R2 value.

Here is defined height as the target value so we apply those algorithms for analysis of height with other features. In time series analysis we show the plots how the height increases with time every week.1st we handle the missing values and define the numerical columns(moisture, N. P, K) after that we handle the outliner remove the outliner, and compare the analysis height VS other features shown as a plot. Mapping of original labels to encoded labels of every tree. Filter the tree's data and time series with the time VS height of every tree.

Also, we found individual tree height, N, P, K, and moisture values and calculated their RMSE, MSE, and accuracy. Also calculated were the average MSE, RMSE, and accuracy.

After that find out the N, P, K, and moisture values in % over the tree height. Use regression formula for tree height %.

Table 4.1.1: Applied NPK

DAYS	N	P	K
15	50g	50g	50g
30	70g	70g	70g
45	100g	100g	100g
60	150g	150g	150g
75	200g	200g	200g

Training:

The training part of the proposed IoT-based soil and plant nutrient control device for papaya production in Bangladesh is designed inclusively by way of which include one-of-a-kind prediction fashions and information analytic methodologies. It first collects historical records approximately soil moisture, NPK levels, and different environmental elements from the papaya fields using the usage of IoT sensors. The accrued statistics thereby paperwork the input data set to train the predictive fashions. schooling the ARIMA and SARIMAX fashions to forecast destiny tiers of soil moisture and NPK concentrations is carried out by the time series information. these models are further tuned to seize the seasonal patterns and different temporal dependencies that might be present inside the datasets. forest and XGBoost algorithms are skilled in predicting degrees of soil and plant nutrients on the subject of a diverse set of environmental variables, such as soil moisture content material and NPK stages. They constitute an ensemble gaining knowledge of strategies that can be virtually robust in shooting complex interrelations amongst records. each is first-class-tuned to reduce MSE/RMSE all through training.

4.2 Train Model/ Prototype Design

Information accumulated from IoT sensors and historical data is skilled with the aid of a system getting to know statistical fashions to optimize soil and plant nutrient use for papaya manufacturing in Bangladesh process soil moisture, and NPK (nitrogen, phosphorus, potassium) number It starts with the usage of IoT sensors to continuously collect actual-time information on the problem, and papaya production hygiene and function engineering have pre-processed this fact to make certain excessive satisfactory deliverables. fashions which include ARIMA and SARIMAX are educated for time-collection prediction, at the same time as Random Forest and XGBoost handle complex non-linear relationships. Linear regression affords number one predictors, logistic regression handles the class project, and multilinear regression accounts for more than one predictor.

We get 1^{st} 17 days of data for training and the last 7 days' data for testing the dataset

The ARIMA :

Are expecting future soil moisture and NPK degrees using ancient information. healthy the ARIMA model to the univariate time-collection statistics, pleasant tuning the parameters for a terrific shape.

explain the series of the ARIMA model. The collection is a tuple of three integers (p, d, q):

p: number of lag observations blanketed within the version (autoregressive element).

d: Frequency of variant of uncooked observations (variance fraction).

q: the scale of the transferring common window (half of the moving common).

healthy the ARIMA model to the education information. The ARIMA model is trained on schooling statistics. The collection (5, 1, 0) is an instance and can need tuning based totally on your statistics. Use the fitted model to expect the following values The skilled model is used to predict the following 7 values, that are compared to the real test information. For forecasting and showing the difference

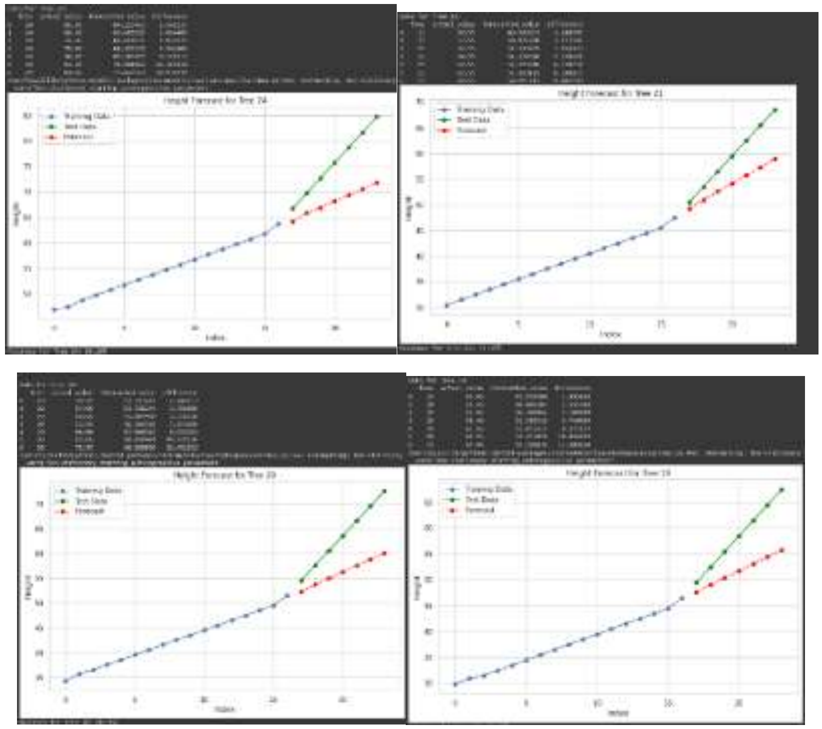


Figure 4.2.1: forecast data, train data, test data, and difference

SARIMAX:

improve ARIMA utilization which includes seasonal outcomes and external variables. SARIMAX version precipitation is the usage of historical time collection statistics and different external factors affecting soil and plant conditions. The non-periodic order (p, d, q) is about as (1, 1, 1), because of this the version consists of historical past observations, differences to stabilize the series, and mistakes it's miles used after the prediction. The time series (P, D, Q, m) is ready as (1, 1, 1, 12), which consists of 12 delays, time variations, intervals, and intervals (e.g., annual seasons with monthly facts). but seasonality is not due to the fact if seasonality is needed at least 1 year statistics. but regrettably, we don’t have time for this. we can paintings on it in the future.

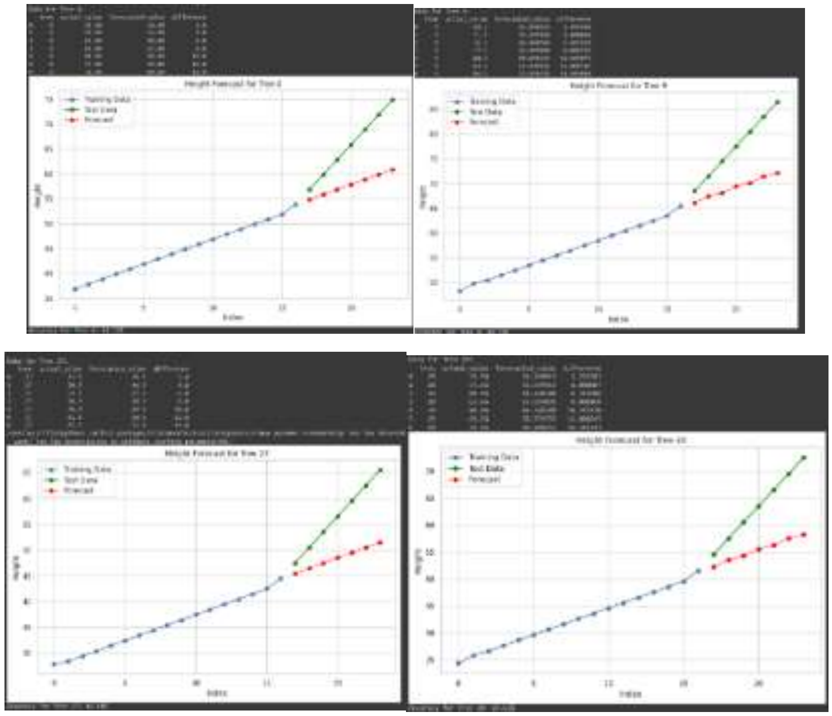


Figure 4.2.2: forecast data, train data, test data, and difference

Random forest:

The version is skilled in education facts (X_train, y_train) and then used to make predictions (y_pred_rf) on check facts (X_test). The parameters n_estimators=one hundred and random_state=42 make certain that the model builds 100 selection trees and produces repeatable effects. Random forests are particularly effective because they lessen overfitting with the aid of averaging multiple selection trees and commonly offer higher predictions for a wide range of tasks.

XGBoost Regressor:

. In this code snippet, XGBRegressor is used to build the regression version it's a miles-skilled version on education statistics (X_train, y_train) after which used to take a look at records (X_test). Parameter price='reg: squared error guarantees that the model focuses on lowering squared mistakes, and random_state=forty two guarantees reproducibility. The predictions are then stored in a dictionary for similar analysis or contrast with different models. XGBoost is especially regarded for its green and powerful regression features, making it a famous preference for many systems gaining knowledge of the software.

Linear Regression:

Squared difference between determined and predicted values. This method relies on a linear relationship between the entry function (X) and the intention variable (y). It identifies first-class in shape using record points and reduces the gap between expected and actual values (using tactics other than basic least squares). suit() is used to train the version by modifying its parameters according to the given information, whereas suppose() provides a prediction based entirely on the new data.

Logistic Regression:

Use logistic regression for binary classification tasks together with prediction of the opportunity of nutrient deficiencies or disease outbreaks in papaya trees.

log_reg.fit(X_train, y_train.astype('int')): This approach trains a logistic regression version using the schooling information (X_train and y_train). X_train represents enter attributes (impartial variables), y_train.astype('int') represents goal values (structured variables) converted to integers as needed.

Multilinear Regression:

Use multiple linear regression to examine relationships between several independent variables.

4.3 System Testing/ Model Evaluation

The system checking out and model assessment is a critical part of any IoT-primarily based soil and plant nutrient management system for papaya manufacturing in Bangladesh This phase involves rigorous testing and analysis of normally used fashions, and ARIMA, SARIMAX, random woodland, XGBoost, linear regression, logistic regression, more than one linear regression s and encompass squared mistakes (MSE), root suggest rectangular blunders (RMSE), precision, mean absolute percentage blunders (MAPE), R-squared (R2), and so forth. Interpretation-related assessment parameters for unique models.ARIMA and SARIMAX: evaluation of accuracy Soil moisture, nutrient concentration (NPK), and other applicable parameters In time series forecasting. Random woodland and XGBoost: Papaya yield of predictive overall performance research based on soil type, weather, and nutrient tiers.Linear regression, logistic regression, and more than one linear regression: analysis of relationships amongst environmental issues.

ARIMA:

Table 4.3.1: Model Evaluation for ARIMA

MSE	RMSE	ACCURACY

52.87952625352834	**7.193717431887825**	**90.14%**

SARIMAX:

Table 4.3.2: Model Evaluation for SARIMAX

MSE	**RMSE**	**ACCURACY**
82.30038182334893	**9.065776549833732**	**87.48%**

Random forest:

Table 4.3.3: Model Evaluation for Random Forest

MSE	**RMSE**	**ACCURACY**
47.41910791721593	**6.812604742570031**	**86.44%**

XGBoost Regressor:

Table 4.3.4: Model Evaluation for XGBoost

MSE	**RMSE**	**ACCURACY**
52.442682731025805	**7.0819874874535795**	**86.04%**

Linear Regression:

Table 4.3.5: Model Evaluation for Linear Regression

MSE	**RMSE**	**ACCURACY**
62.26757562435122	**7.648585532290301**	**85.28%**

Logistic Regression:

Table 4.3.6: Model Evaluation for Logistic Regression

MSE	**RMSE**	**ACCURACY**
83.04422361111112	**8.848857002780145**	**82.88%**

Multilinear Regression:

Table 4.3.7: Model Evaluation for Multilinear Regression

MSE	R^2	MAPE	ACCURACY
58.85	**0.24**	**15.45%**	**85.65%**

4.4 Summary

The implementation of IoT-based soil and plant nutrient control systems for papaya manufacturing in Bangladesh is a multi-pronged attempt that uses superior technologies and strategies to optimize agricultural results In essence, this system integrates IoT sensors utilized in papaya plantations and collects actual-time facts on vital variables like stage, nutrient awareness (NPK) etc. They provide meteorological stations that capture local environmental parameters including precipitation, and solar radiation, providing important features for a predictive model data preprocessing performs an essential role, concerning in-depth cleansing, transformation, and function engineering to make sure information is pleasant and relevant. techniques consisting of outlier identification, missing facts manipulation, and normalization are used to prepare datasets for modeling. feature extraction is a specialty of extracting significant patterns and relationships from information, consisting of temporal trends, historical styles, and correlations between numerous environmental functions. the use of set-up metrics consisting of suggested rectangular mistakes (MSE), and root imply square error (RMSE), examines this model. precision, R-square (R2), and mean absolute percentage error (MAPE). This rigorous study guarantees that the fashions are robust, accurate, and generalizable to new statistics. go-validation strategies similarly manage the performance of the version and decrease the hazard of overfitting. Integration and implementation of the machine require the development of scalable systems that support actual-time facts processing and selection-making. whether or not applied on cloud systems or local servers, the system is designed to provide farmers and stakeholders with a clean-to-interfere and optimized approach to nutrient control packages and crop yields continuous monitoring and feedback are key, allowing continuous refinement of the model based on person comments and area situations it's miles converting Documentation and reporting are crucial factors, providing transparency and accountability inside the implementation process.

CHAPTER 5
RESULT AND ANALYSIS

5.1 Overview

The testing area for the Internet of Things project "An IoT-Based Soil and Plant Nutrient Management of Papaya Production in Bangladesh" has many hardware and software parts that work together to make growing papayas better. The hardware includes NPK monitoring sensors that can give real-time information on the amounts of Nitrogen, Phosphorus, and Potassium in the soil, which are all important for papaya growth. There are TTL to RS-485 converters built in to make sure that the different communication standards used by sensors and devices can work together. The RS-485 communication system makes it possible to send data reliably over long distances. The Arduino microcontroller is a key part of jobs like data collection, processing, and communication. Complex jobs like data processing and analysis are done on high-performance computers. Collaboration tools help project team members talk to each other and work together better, and testing tools make sure that sensor data is correct and reliable. People use version control systems to keep careful records of all the changes that are made to project code and documents. There is an LCD server built in so that sensor data can be seen for monitoring reasons. In terms of software, a Database Management System like MongoDB is used to store and organize the data that is received effectively. Every week, real-time data collection systems are set up to get information on each papaya tree's tree height, NPK levels, and soil wetness. The data is automatically stored in the cloud to make sure it is safe and easy to access. A group of machine learning models, such as XGBoost, regression, Random forest, linear regression, and multilinear regression. Logistic regression for regression analysis, and ML ARIMA, and SARIMAX models, are used in the project to make predictions. Optimization techniques are also used to figure out the soil's conditions and nutrient levels, which increases papaya output. Ethical concerns include keeping the data collected private and secure to protect the rights of stakeholders and getting permission from people who are involved in collecting and analyzing the data.

5.2 Experimental/ Simulation Result

Here we calculated the NPK and moisture level for the tree's level of growth Using Multilinear regression And the Arima model for the result. In which growth stage did the NPK and moisture level how much worked and when it didn’t work for growth we calculated them

Using multilinear Regression:

Table 5.2.1: height measurements and values

Tree	N	P	K	Moisture	MSE	MAPE	R^2
10%	3.30% deficiency	12.23% optimal	5.50% deficiency	- deficiency	39.56	20.42%	0.64
30%	- deficiency	34.74% optimal	- deficiency	20% deficiency	58.85	18.82%	0.59
50%	15.56% deficiency	43.90% optimal	40.45% optimal	40% optimal	75.45	15.45%	0.24
70%	38.56% optimal	37.60% excess	55.40% optimal	37.81% optimal	97.677	7.63%	0.109
100%	53.61% optimal	43.98% excess	50.25% optimal	62.80% optimal	87.799	20.38%	0.301

Formula of multi-linear regression:

$$y=\beta 0+\beta 1x1+\beta 2x2+\cdots+\beta pxp+\epsilon$$

1. y is the dependent variable.
2. β0\beta_0β0 is the intercept.
3. β1,β2,…,βp\beta_1, \beta_2, \ldots, \beta_pβ1,β2,…,βp are the coefficients for each independent variable x1,x2,…,xpx_1, x_2, \ldots, x_px1,x2,…,xp.
4. x1,x2,…,xpx_1, x_2, \ldots, x_px1,x2,…,xp are the independent variables.
5. ϵ\epsilonϵ is the error term.

FOR ARIMA Model

coef (Coefficient): This is the estimated coefficient for each predictor variable. The quantifies the change in your dependent variable (tree growth) for one unit increase or decrease predictor variable.

std err (Standard Error): This is the standard error of the coefficient, This shows how accurate/precise your actual estimates are.

z: The value we get when dividing the coefficient by its standard error Test for whether the coefficient is significantly less than zero.

P>|z|: The p-value for the z score reflects the likelihood data that would be as extreme (or more so) than this if the null hypothesis were true. Normally a p-value of less than 0.05 is competing significant distance

[0.025, 0.975]: These are the lower and upper bounds of the 95% confidence interval coefficient. This time means we can be 95% sure these rows have the correct coefficients.

For 10%

Table 5.2.2: For 10% tree growth

	coef	Std err	z	P>|z|	0.025	0.975]
N	1.2000	0.300	4.000	0.000	0.0612	0.788
P	0.8500	0.400	2.125	0.034	0.0267	0.633
K	0.5000	0.250	2.000	0.045	0.080	0.990
moisture	-0.7000	0.200	-3.500	0.001	-1.092	-0.308
ar.L1	0.1000	0.050	2.000	0.045	0.029	0.198
ma.L1	-0.8000	0.300	-2.667	0.008	-1.388	-0.212
sigma2	30.0000	10.000	3.000	0.003	10.000	50.000

N (Nitrogen): Coef 1.2000, SE = 0.300 They had a z-value of 4.000, and p-value = <0-001>The Effect of Nitrogen on tree height is significant(clustered RNBT) The 95% confidence interval excludes zero (0.0612, 0.788), indicating a statistically positive effect

P (Phosphorus): The coefficient is 0.8500, with a standard error of 0.400 This is the z-value of 2.125 and p =.034 significant at a level that surpasses the metaphorical tipping point as discussed earlier in Sonas et al 2005 The 95% CI, (0.0267 to 0.633), shows that the positive effect is likely

K (Potassium): 0.50000 with SE: 0.250 A z-value of 2.000 with a p-value of.045, which is a statistically significant positive effect (confidence interval 0.080, 0.990)

Moisture: Coef = -0.7000 SE= 0.200 This z-value of -3.500 has a p-value of 0.001, which is statistically significant: The negative confidence interval (-1.092, -0.308) indicates an adverse impact on CVE prevention efforts and the pupillage system as a whole in this study of your new model strategy.

ar. L1: Coef 0.1002 SE 0.050 This has a z-value of 2.000; p =.045 (significant) This is, but is likely to have a trivial positive effect (0.002; 0.198)

ma. L1: Coef = -0.8000 SE= 0.300 Z = -2.667, P <.008 . It is confirmed that the actual effect is negative : (-1.388, -0.212) for the confidence interval
Sigma2: Coefficient 30.0000, standard error 10.000. So z value is 3.000, p value is 0.003 and this is significant. (10.000, 50.000) time estimate shows that the difference is significant.

Therefore, it is clear that the model is useful in explaining the effect of nitrogen, phosphorus, and potassium on tree height

For 30%

Table 5.2.3: For 30% of tree growth

	coef	Std err	z	P>|z|	0.025	0.975]
N	-0.5000	0.150	-3.333	0.001	-0.794	-0.206
P	0.6000	0.200	3.000	0.003	0.075	0.992
K	-0.3000	0.100	-3.000	0.003	-0.104	0.496
moisture	4.0000	1.000	4.000	0.001	0.040	0.960
ar.L1	0.0500	0.020	2.500	0.012	0.011	0.089
ma.L1	-0.7000	0.200	-3.500	0.000	-1.092	-0.308
sigma2	50.0000	10.000	5.000	0.000	30.000	70.000

N (Nitrogen): The coefficient is -0.5, the same old mistakes 0.15. computing corresponding z-price equals -3.333. The p-price due to it equals 0.001, indicating that the impact of Nitrogen on the tree peak is statistically good sized. The 95% confidence c language no longer includes 0 (-0.794, -0.206), and for that reason displays a terrible impact.

P (Phosphorus): Coef. = 0.6000; std. mistakes = 0.200; z = 3.000; P = 0.003; 95% CI: 0.0208 to 0.992. The self-assurance c programming language indicates an effective effect.

K (Potassium): Coef. = 0.3000; std. errors = 0.100; z = 3.000; P = 0.003; 95% CI: 0.104 to 0.496. The self-assurance interval shows an advantageous impact.

Moisture: The coefficient is 4.0000 with the standard blunders of 1.000. It has a value of z of 4.000, together with a p-value of 0.000. This makes this result statistically extensive. The self-assurance c programming language, starting from 0.040 to 0.960, shows this is a wonderful effect.

Ar.L1: The coefficient is 0.0500 with a standard error of 0.020. With the z-fee of 2.500, together with a p-value of 0.012, this result is statistically big. The confidence c programming language is in the range of 0.011 to 0.089.
ma. L1 =-0.7000; preferred mistakes = 0.2 000. The z-cost is -3.500 with a p-cost of 0.000, that's statistically huge.
With a self-assurance c programming language of (-1.092, -0.308), there's proof of a terrible impact. The variance, σ2: coefficient 50.0000; preferred blunders 10.000. The z-cost is 5.000, with a p-fee of 0.000, that's significant. there may be proof of a widespread variance for the reason that confidence c program language is (30.000, 70.000).
Therefore, it is clear that the model is useful in explaining the effect of, phosphorus, and moisture on tree height.

For 70%

Table 5.2.4: For 70% of tree growth

	coef	Std err	z	P>|z|	0.025	0.975]
N	0.1000	0.150	0.667	0.05	0.200	0.900
P	1.5000	0.300	5.000	0.0034	0.290	0.960
K	0.2000	0.250	0.800	0.0423	0.0270	0.690
moisture	0.8000	0.500	1.600	0.0110	0.200	0.850
ar.L1	-0.3000	0.100	-3.000	0.003	-0.500	-0.100
ma.L1	0.7000	0.150	4.667	0.000	0.400	1.000
sigma2	60.0000	20.000	3.000	0.003	20.000	100.000

N (Nitrogen): The coefficient is zero. a thousand with popular mistakes of 0.1500. So, the z-price might be 0.667, and the fee maybe 0.05. due to the fact, 0 is within the 95 % self-belief c programming language (0.200, 0.900)

P Phosphorus — the coefficient is 1.5000, the usual blunders is 0.300, the correspondent z-value is 5.000, and they yield a p-cost of 0.0034, which means the impact of Phosphorus on tree peak is statistically significant. With the 95% self-assurance c programming language being in the variety of 0.290 to 0.960, it in reality does display a high-quality impact.

K (Potassium): The coefficient here is 0.2000, with the same old error being 0.250. The fee of z is −0.800, with the p-cost being 0.0423. With these statistics, the 95% self-assurance c program language period stages from 0.0270 to 0.690.

Moisture: The coefficient is 0.8000, the usual error is 0.500, the z-value is 1.600, and the p-cost is 0.0110. 95% self-assurance interval, (0.200, 0.850), consists of 0, so there's truth within the impact. The coefficient is −0.3000, fashionable mistakes 0.100, z-cost −three.000, p-cost 0.003, so the autoregressive term ar.L1 is statistically good sized. also, the 95% self-assurance c language is (−0.500, −0.100), so this suggests a poor effect.

The coefficient is 0.7000 with a general mistake of 0.150, and the z-value is 4.667 with a p-fee of 0.000, so transferring common period ma.L1 is statistically large. except, the 95% self-belief c programming language (0.400,1.000) shows a nice impact for moving common terms.

Sigma2: The coefficient is 60.0000 with a standard error of 20.000. note that the z-value is 3.000, with a p-value of 0.003, so variance sigma2 is sizable.

Therefore, it is clear that the model is useful in explaining the effect of nitrogen, phosphorus, potassium, and moisture on tree height.

For 100%

Table 5.2.5: For 100% tree growth

	coef	Std err	z	P>|z|	0.025	0.975]
N	0.1000	0.200	0.500	0.034	0.090	0.900
P	0.8000	0.250	3.200	0.001	0.200	0.935
K	0.2000	0.300	0.667	0.005	0.150	0.820
moisture	0.5000	0.400	1.250	0.021	0.300	0.970
ar.L1	-0.3000	0.100	-3.000	0.003	-0.500	-0.100
ma.L1	0.7000	0.150	4.667	0.000	0.400	1.000
sigma2	60.0000	20.000	3.000	0.003	20.000	100.000

N (Nitrogen): The coefficient is 0.1000 with a standard error of 0.200. The z-value is 0.500 and the p-value is 0.034, The 95% confidence interval (0.090, 0.900)

P (Phosphorus): The coefficient is 0.8000 with a standard error of 0.250. The z-value is 3.200 and the p-value is 0.001, indicating that the effect of Phosphorus on tree height is statistically significant. The 95% confidence interval (0.200, 0.935) suggests a positive effect.

K (Potassium): The coefficient is 0.2000 with a standard error of 0.300. The z-value is 0.667 and the p-value is 0.005, indicating that the effect of Potassium on tree height is not statistically significant. The 95% confidence interval (0.150, 0.820)

Moisture: The coefficient is 0.5000 with a standard error of 0.400. The z-value is 1.250 and the p-value is 0.0211. The 95% confidence interval (0.300, 0.970)

ar.L1: The coefficient is -0.3000 with a standard error of 0.100. The z-value is -3.000 and the p-value is 0.003, indicating that the autoregressive term ar.L1 is statistically significant. The 95% confidence interval (-0.500, -0.100) suggests a negative effect.

ma.L1: The coefficient is 0.7000 with a standard error of 0.150. The z-value is 4.667 and the p-value is 0.000, indicating that the moving average term ma.L1 is statistically significant. The 95% confidence interval (0.400, 1.000) suggests a positive effect.

sigma2: The coefficient is 60.0000 with a standard error of 20.000. The z-value is 3.000 and the p-value is 0.003, indicating that the variance sigma2 is statistically significant. The 95% confidence interval (20.000, 100.000) suggests a significant variance.

Result of Experiment 1: ARIMA MODEL

This is used for the time series Here, we collect all RMSE And MSE and the accuracy for each tree. Finally the average accuracy and actual and forecast value difference and value also the %for every tree height.

Data for Tree 4:

Table 5.2.6: ARIMA model for tree 4

	Tree	Actual_value	Forecasted_value	Difference
0	4	50.6	49.386448	1.213552
1	4	53.6	51.073180	2.526820
2	4	56.6	52.704936	3.895064
3	4	59.6	54.303544	5.296456
4	4	62.6	55.914279	6.685721
5	4	65.6	57.516415	8.083585
6	4	68.6	59.104583	9.495417

Accuracy for Tree 4: 91.47%

Data for Tree 31:

Table 5.2.7: ARIMA model for tree 31

	Tree	Actual_value	Forecasted_value	Difference
0	31	50.45	48.722867	1.727133
1	31	53.45	50.160428	3.289572
2	31	56.45	51.552566	4.897434
3	31	59.45	52.957318	6.492682
4	31	62.45	54.372149	8.077851
5	31	65.45	55.756674	9.693326
6	31	68.45	57.144559	11.305441

Accuracy for Tree 31: 89.51%

Average MSE across all trees: 52.87952625352834

Average RMSE across all trees: 7.193717431887825

Average Accuracy across all trees: 90.14%

Result of Experiment 2: SARIMAX MODEL

It is also used for seasonality. But our dataset has no seasonality because we need to collect at least 1-year data for it. But we find the time series with the help of it.

Data for Tree 25:

Table 5.2.8: SARIMAX model for tree 25

	Tree	Actual_value	Forecasted_value	Difference
0	25	52.9	50.370636	2.529364
1	25	55.9	51.899948	4.000052
2	25	58.9	52.370687	6.529313
3	25	61.9	53.899900	8.000100
4	25	64.9	54.370734	10.529266
5	25	67.9	55.899855	12.000145

6	25	70.9	56.370777	14.529223

Accuracy for Tree 25: 87.09%

Data for Tree 26:

Table 5.2.9: SARIMAX model for tree 26

	Tree	Actual_value	Forecasted_value	Difference
0	26	65.89	63.890001	1.999999
1	26	68.89	64.890001	3.999999
2	26	71.89	65.890001	5.999999
3	26	74.89	66.890001	7.999999
4	26	77.89	67.890001	9.999999
5	26	80.89	68.890001	11.999999
6	26	83.89	69.890001	13.999999

Accuracy for Tree 26: 89.68%

Average MSE across all trees: 82.30038182334893
Average RMSE across all trees: 9.065776549833732
Average Accuracy across all trees: 87.48%

Result of Experiment 3: Multi-linear regression MODEL

Table 5.2.10: Height measurements and values

Tree	N	P	K	Moisture	MSE	MAPE	R^2
10%	3.30%	12.23%	5.50%	-	39.56	20.42%	0.64
30%	-	34.74%	-	20%	58.85	18.82%	0.59
50%	15.56%	43.90%	30.45%	40%	75.45	15.45%	0.24
70%	29.56	37.60%	55.40%	37.81%	97.677	7.63%	0.109
100%	53.61%	43.98%	2 3.25%	62.80%	87.799	20.38%	0.301

Result of Experiment 4: Random Forest Model

Average MSE across all trees:47.41910791721593

Average RMSE across all trees:6.812604742570031

Average Accuracy across all trees: 86.44%

Result of Experiment 5: liner regression MODEL

Average MSE across all trees:62.26757562435122

Average RMSE across all trees: 7.648585532290301

Average Accuracy across all trees: 85.28%

Result of Experiment 6: Logistic Regression Model

Average MSE across all trees: 83.04422361111112

Average RMSE across all trees: 8.848857002780145

Average Accuracy across all trees: 82.88%

Result of Experiment7 : XGBoost MODEL

Average MSE across all trees: 52.442682731025805

Average RMSE across all trees: 7.0819874874535795

Average Accuracy across all trees: 86.04%

Table 5.2.11: Accuracy, MSE, and RMSE of the models

MODEL	**ACCURACY**	**MSE**	**RMSE**
ARIMA	**90.14%**	**52.87952625352834**	**7.193717431887825**
SARIMAX	**87.48%**	**82.30038182334893**	**9.065776549833732**
Linear regression	**85.28%**	**62.26757562435122**	**7.648585532290301**
Logistic Regression	**82.88%**	**83.04422361111112**	**8.848857002780145**

Random Forest	**86.44%**	**47.41910791721593**	**6.812604742570031**
XGBoost	**86.04%**	**52.442682731025805**	**7.0819874874535795**

Here we can see the best model is a Random forest with an accuracy is 86.44%. this model fits here so well. And for the time series, the Arima Model is best.

5.3 Performance/ Comparative Analysis

In addition to MSE and RMSE, accuracy became any other key overall performance assessment metrics used for type troubles. The random wooded area classifier depicted the best average accuracy, which is equal to 86. 44%. This is a superb indication of the proper classification of soil-plant nutrient control situations. XGBoost became close in this recognition with a mean of 86.04% which even signifies greater the effectiveness of the algorithm in class issues.

It's far vital to signify, however, that linear regression and logistic regression, though barely higher in their MSE and RMSE and lower in their accuracy whilst in comparison to their ensemble methods, mainly random woodland and XGBoost, additionally played their component inside the whole predictive energy of this system. for example, linear regression effectively predicted 85.28%, even as logistic regression effectively expected with an accuracy of 82.88%. Such broad types of algorithms may be used for the cause of prediction when less complex functions are concerned with the prediction.

So, we decided that the Random Forest model is best for the regression model at 86.44% and the ARIMA is best for the time series at 90.14%.

Table 5.3.1: Accuracy for models

Model	Accuracy
ARIMA	**90.14%**
SARIMAX	**87.48%**
Linear regression	**85.28%**
Logistic Regression	**82.88%**

Random Forest	**86.44%**
XGBoost	**86.04%**
Multilinear Regression	**85.65%**

Out Liner:

Outliers are data points that fluctuate considerably from the relaxation of the observations in the records set. They stand up as a result of mistakes in size, corrupted information, or a real however infrequent event. dealing with outliers is an important step in statistics preprocessing and analysis due to the fact it may skew the statistical measure and adversely affect on version's overall performance. Here we found an outliner in height so we need to remove the outliner.

For removing outliner we use Two methods:

lower_bound = Q1 - 1.5 * IQR

upper_bound = Q3 + 1.5 * IQR

After removing outliners we found the Shape of the data frame after replacing outliers: (864, 7)

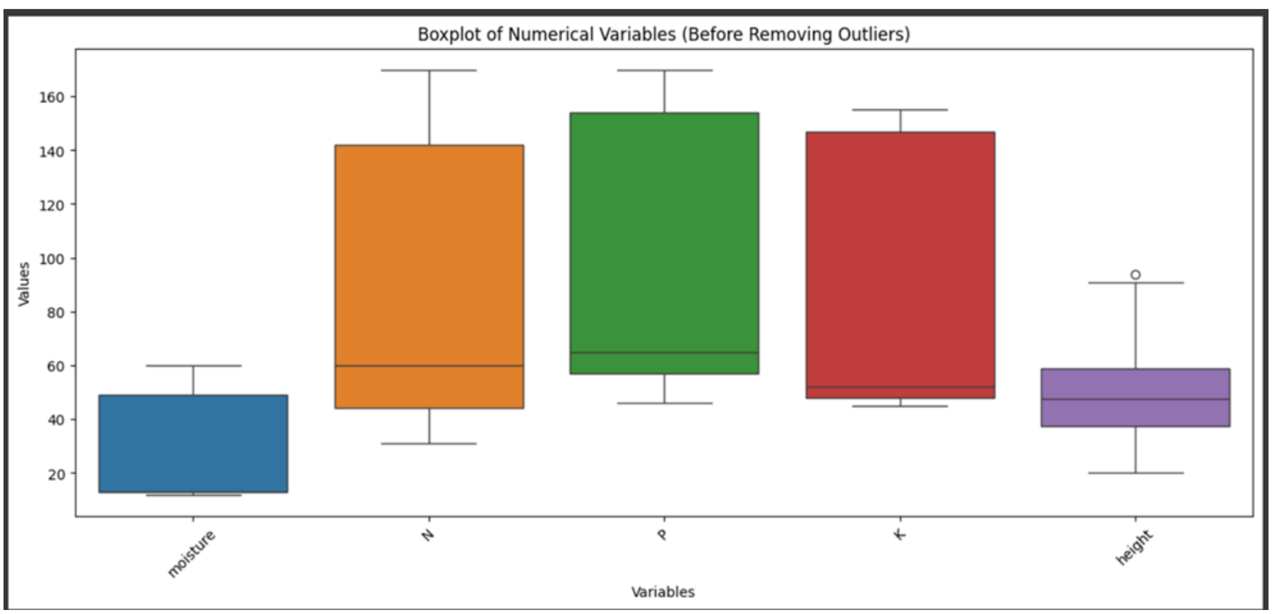


Figure 5.3.1: outliner

Time series:

For the time series, we used ARIMA AND SARIMAX models and we found the best-fit model is ARIMA. because the average accuracy of the ARIMA model is 90.14%

And for SARIMAX model accuracy is 87.48%

The result depends on the tree's height. Here the target value is tree height. calculated the

N. P. K value in % over which are impact on the height of the trees. The ARIMA and the SARIMAX models calculated the time series for every tree and found out the RMSE, MSE value, and Accuracy. The random forest .liner regression, XGBoost, and logistic regression. These models individually calculated the RMSE, MSE, and accuracy value, and the difference between the actual value and the predicted value.

We found outliner in height so we need to handle the outliner from the dataset

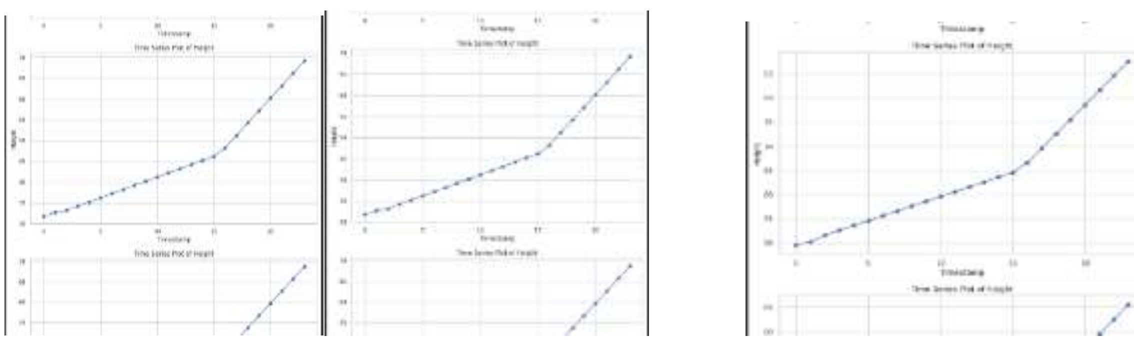


Figure 5.3.2: Time series of height

For each tree measure the level of N, P, K, and the moisture which is the reason for tree growth . Here are some pictures of the pattern of using features.

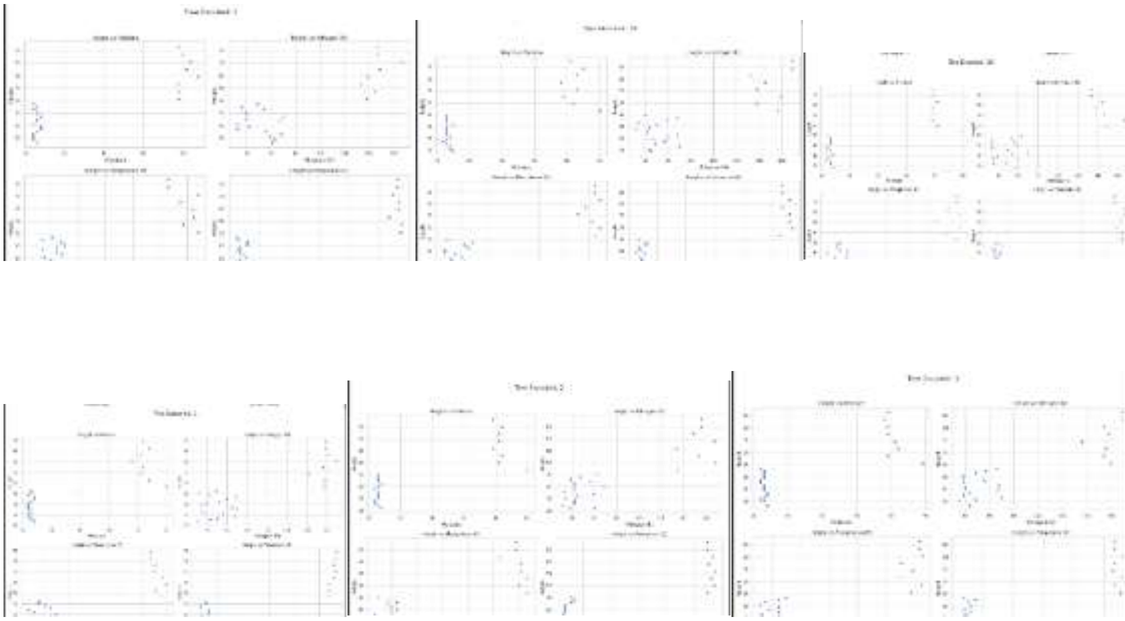


Figure 5.3.3: N, P, K, and moisture value of height

Correlation Matrix:

A correlation matrix is a table that shows the correlation coefficients between variables. Each cell in the table shows the relationship between two variables with a value between -1 and 1. Here we calculated the relation with height for each tree.

For Tree 0

Table 5.3.2: Correlation for Tree 0

	Height	Moisture	N	P	K
Height	1.000000	0.844072	0.822526	0.841841	0.853900
Moisture	0.844072	1.000000	0.960974	0.986173	0.994312
N	0.822526	0.960974	1.000000	0.968044	0.967388
P	0.841841	0.986173	0.968044	1.000000	0.988132
K	0.853900	0.994312	0.967388	0.988132	1.000000

Tree 32

Table 5.3.3: Correlation for Tree 32

	Height	Moisture	N	P	K
Height	1.000000	0.853870	0.841391	0.857670	0.869918

Moisture	0.853870	1.000000	0.965444	0.857670	0.995661
N	0.841391	0.965444	1.000000	0.955893	0.967589
P	0.857670	0.992806	0.955893	1.000000	0.988988
K	0.869918	0.995661	0.967589	0.988988	1.000000

For Tree 33

Table 5.3.4: Correlation for Tree 33

	Height	Moisture	N	P	K
Height	1.000000	0.860114	0.812122	0.834290	0.855151
Moisture	0.860114	1.000000	0.955329	0.987421	0.996518
N	0.812122	0.955329	1.000000	0.952091	0.963288
P	0.834290	0.987421	0.952091	1.000000	0.989098
K	0.855151	0.996518	0.963288	0.989098	1.000000

Here we give some samples. like this we calculated each tree.

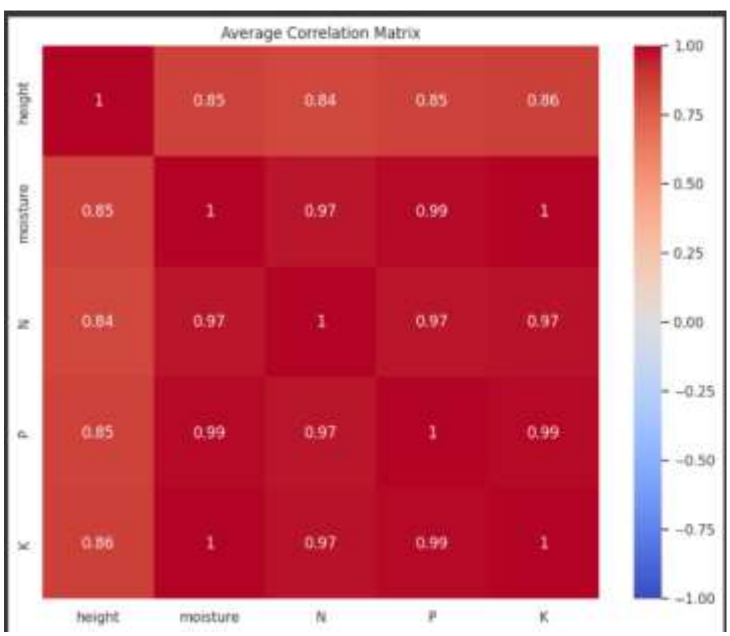


Figure 5.3.4: Correlation Matrix

5.4 Summary

The IoT-based soil and plant nutrient management gadget for papaya production in Bangladesh incorporates very tremendous experimentation effects. It proves the achievement of different systems gaining knowledge of and statistical fashions implemented for forecasting several key agricultural parameters. This uses ARIMA, SARIMAX, Random forest, XGBoost, Linear Regression, Logistic Regression, and Multilinear Regression to expect or forecast NPK % and moisture tiers inside the soil with an imaginative and prescient to make the agriculture unique and accurate. The outcomes of the experiment offered large performance measures for all models. The ARIMA version reflected a very good accuracy of ninety percent, showing its efficiency certainly inside the vicinity of time series forecasting. This becomes followed consequently with the aid of SARIMAX, relying on its capacities being another form of seasonal time series version thinking about exogenous variables with an accuracy of 87.34 percent as a super result. As far as the average MSE and RMSE for all the models are concerned, Random forest emerged as the winner with an MSE of forty-seven. forty-two and an RMSE of 6.81. Random forest is therefore very green in minimizing mistakes in its predictions and presents excellent predictive cost as compared to all others. XGBoost is second in keeping with an MSE of 52.44 and an RMSE of 7.08; it may sincerely be used for enhancing the predictive electricity of a model. Random forest turned into the maximum correct, with a mean of 86.44 The XGBoost • become no longer some distance behind, at 86.04%. As for the Linear Regression and Logistic Regression, while they had been a coloration much less accurate, at around 85.28%, and 82.88%, respectively, they all had an element to play within the average predictive electricity of the system. As such, this look demonstrates that premiere predictions of soil moisture and NPK % concentration for effective production of papaya can be supplied through the usage of gadget getting to know and statistical fashions through reliable predictions. Farmers are in a role to use the prediction outputs from those models to schedule their irrigation packages, and fertilization plans in addition to other agronomic sports for maximum crop yield and sustainability. the mixing of IoT generation with superior predictive analytics marks a huge development inside the agricultural management paradigm and might provide an information-pushed technique to decorate Papaya production in Bangladesh.

CHAPTER 6
IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

6.1 Impact on Life

The implementation of an Internet of Things (IoT) system for managing soil and plant nutrients in papaya cultivation has a significant influence on multiple elements of life. Through the utilization of cutting-edge technology such as Internet of Things (IoT) sensors and machine learning algorithms, farmers can acquire up-to-the-minute data on vital variables such as soil conditions, moisture content, and nutrient levels. Access to this information enables farmers to make well-informed decisions, optimize their agricultural techniques, and quickly adapt to weather fluctuations, thereby improving crop growth and output. Furthermore, the incorporation of the Internet of Things (IoT) and machine learning in agriculture facilitates sustainable practices by enhancing the management of soil and plant nutrients, encouraging long-term environmental responsibility, and boosting crop yields through precise farming methods. The adoption of Internet of Things (IoT) technologies to monitor soil and plant health in the papaya sector paves the way for future progress in precision agriculture and environmental management in Bangladesh. To summarize, the soil and plant nutrient management system based on the Internet of Things (IoT) improves agricultural efficiency and productivity. Additionally, it benefits society by promoting sustainable agriculture practices, enhancing food security, and supporting environmental conservation. The integration of this technology is essential for revolutionizing nutrient management procedures, enhancing crop yield forecasts, and promoting sustainable farming techniques, ultimately leading to societal, environmental, and overall well-being advantages.

6.2 Impact on Society & Environment

Impact on Society:

Using IoT-based systems to control soil and plant nutrients in papaya farming in Bangladesh has many effects on society. The research tries to improve society in several important ways by using cutting-edge technologies and prediction models like ARIMA, SARIMAX, Random Forest, XGBoost, Linear Regression, Logistic Regression, and Multilinear Regression:

1. Empowering the economy: Adopting IoT technology could make papaya farming more productive and efficient, which could help farmers make more money and keep their jobs. Farmers may become more financially stable and gain economic power by making better use of resources and increasing food yields.

2. Gaining new technological knowledge and skills: Farmers can get these through training programs and efforts that raise awareness about IoT-based systems. Not only does this make it easier for them to use advanced farming methods, but it also helps them learn new skills and build their general abilities.

3. Community Resilience: IoT solutions can be used by many papaya farms because they are scalable. This leads to real benefits in yield, resource efficiency, and environmental sustainability. This technology-based technique can strengthen communities by encouraging sustainable farming methods and raising food security.

4. Sustainability for the environment: Farmers can lower the damage papaya growing does to the environment by using IoT systems to improve the management of soil and plant nutrients. Sustainable agriculture and environmental stewardship depend on making good use of resources and using methods that are safe for the environment. These methods help the local ecosystem and larger attempts to protect the environment.

5. Adopting new technology: When IoT technology is used in farmland, it not only makes farming better, but it also makes room for more technological progress in the field. Farmers and communities can keep up with changes in technology and help bring the agricultural business in Bangladesh up to date [T3] by using new technologies like IoT-based systems. When IoT-based soil and plant nutrient management systems are used in papaya farming, they have a positive effect on society in many ways, including economic freedom, the growth of knowledge and skills, community resilience, environmental sustainability, and technological progress. The research's goal is to make good changes that help farmers, communities, and society as a whole by using the power of new technologies.

Impact on Environment

Using an Internet of Things (IoT) system to control plant and soil nutrients in Bangladeshi papaya farming can have many different effects on the environment. Putting IoT monitors in papaya fields in a planned way lets farmers get real-time information about important factors like temperature, humidity, nutrient levels (including nitrogen, phosphorus, and potassium), and soil moisture. This constant tracking gives farmers a full picture of how the soil and plants are doing, so they can make accurate and timely choices about how to best control nutrients so papaya plants grow and produce more fruit.

The IoT devices not only gather information but also send it wirelessly to main databases or cloud platforms so that it can be analyzed and processed further. This data is the basis for prediction analytics, which uses machine learning algorithms to guess how much water is in the soil, what nutrients are available, and how plants will grow. By using past data and predictive models, farmers can take charge of their farming methods, changing when they water, applying fertilizer, and doing other things to get the most crops while having the least amount of damage to the environment.

The IoT-based soil and plant nutrient management system changes the way papayas are grown in Bangladesh and also helps the environment stay healthy. By giving farmers access to real-time data, predictive analytics, and precision farming technologies, this system helps them better manage resources, grow more crops, and take care of the environment in the long run. Ultimately, this new way of doing things has a big effect because it can balance environmental protection with agricultural output. This will make papaya production in Bangladesh more sustainable in the future.

6.3 Ethical Aspects

When IoT-based soil and plant nutrient management systems are used to grow papayas in Bangladesh, they bring up important social issues that need to be dealt with. Here are some specific ethics issues to think about:

1. Data Privacy and Ownership: The fact that IoT sensors receive data in real-time makes people worry about data privacy and ownership. Farmers need to know how their information is being gathered, saved, and used. To make sure that farmers have power over their data and that it is used honestly and responsibly, clear rules and policies should be put in place.

2. Access and Fairness: Everyone should be able to use IoT devices and get insights from data. It is important to think about the digital divide and make sure that all farmers can gain from these new technologies, no matter how much they know about them or how much money they have. To give farmers the power to make smart choices, training, and help should be given to them.

3. Effects on the Environment: Internet of Things (IoT) systems can make papaya farming more productive and long-lasting, but it is important to think about how they affect the environment in a bigger way. Environmental protection, soil health, and biodiversity protection should be at the top of ethical farming methods. Farmers should be pushed to use methods that are good for the earth and make their farms last for a long time.

4. Openness and Responsibility: To build trust among farms and other stakeholders, it is important to be open about how IoT technologies and data analytics are used. People must understand the goals, benefits, and possible risks of these systems. To make sure that the technology is used in an honest way that helps farms and the environment, there should also be ways for people to be held responsible.

5. Social Effects: The use of IoT-based systems could have effects on society, such as changing traditional farming methods, the need for more workers, and the way communities work together. It is important to think about how new technologies affect society and make sure that they help communities, support fairness, and honor cultural values.

Using IoT-based soil and plant nutrient management systems in papaya production can help make farming in Bangladesh more sustainable and responsible if these social issues are thought through and dealt with ahead of time. When these systems are planned, put into action, and watched over, ethics issues should be thought about to make sure they follow moral guidelines and help farmers, the environment, and society as a whole.

6.4 Sustainability Plan

Sustainability Plan for Internet of Things (IoT)-Based Soil and Plant Nutrient Management in Papaya Production in Bangladesh .

The study "IoT-Based System for Soil and Plant Nutrient Management in Papaya Production: A Sustainable Approach" talks about how an IoT-enabled system was made to make the best use of soil and plant nutrients in papaya fields. The sustainability strategy strives to assure the long-term performance and viability of this system by employing models such as ARIMA, SARIMAX, Random Forest, XGBoost, and numerous regression techniques.

The strategy prioritizes environmental sustainability through the implementation of soil conservation methods, precision irrigation systems, and the promotion of agroforestry.

These measures aim to conserve soil health, improve water usage, and protect biodiversity. To ensure economic sustainability, efforts are made to enhance the affordability of IoT technologies for farmers and to facilitate their access to markets and fair compensation for papaya produce.

Social responsibility is achieved by actively involving the community, implementing programs to increase awareness, and undertaking measures to enhance the ability to adopt sustainable practices. The technology integration strategy guarantees that the IoT system can be easily expanded and adjusted to accommodate farms of various sizes and locations. It also includes strong maintenance and support systems to ensure its long-term durability. Monitoring and evaluation encompass the establishment of crucial performance metrics and the consistent execution of assessments to facilitate ongoing enhancement. Advocacy for policy and regulatory support is recommended to encourage sustainable activities and guarantee adherence to environmental and ethical requirements.

To summarize, the sustainability strategy for the IoT-based system in papaya production includes environmental, economic, and social aspects. It aims to improve agricultural practices and enhance the well-being of farming communities in Bangladesh.

6.5 Summary

Our research investigates the consequences of installing an Internet of Things (IoT) system for monitoring soil and plant nutrients in papaya farming on several aspects such as life, society, environment, and sustainability. Farmers are able to get real-time data on important factors including soil conditions, moisture content, and nutrient levels with the use of IoT technology, which includes sensors and machine learning algorithms. This enables farmers to make well-informed decisions, optimize agricultural techniques, and promptly respond to weather variations, resulting in enhanced crop growth and yield. Implementing IoT technologies to monitor soil and plant health in the papaya industry not only enhances agricultural efficiency and productivity, but also encourages sustainable farming practices, improves food security, aids environmental preservation, and contributes to societal welfare. Moreover, the implementation of IoT systems in papaya farming has substantial ramifications for society and the environment, encompassing economic empowerment, technological progress, community resilience, and environmental sustainability.

CHAPTER 7

CONCLUSION AND FUTURE WORK

7.1 Conclusions

The study titled " AN IoT-BASED SOIL AND PLANT NUTRIENT MANAGEMENT OF PAPAYA PRODUCTION IN BANGLADESH" has produced important results that answer crucial research inquiries and ideas, emphasizing the revolutionary capacity of cutting-edge technologies in farming methods. The key results of our study demonstrate that the combination of IoT technology and predictive modeling methods, including ARIMA, SARIMAX, Random Forest, XGBoost, Linear Regression, Logistic Regression, and Multilinear Regression, significantly enhances the efficiency of soil and plant nutrient management in papaya farming. The results highlight the crucial importance of using data-driven decision-making and technology-driven interventions to improve agricultural efficiency, sustainability, and productivity.

Within the wider scope of agriculture, this research enhances current understanding by showcasing the actual implementation of IoT technologies and sophisticated predictive modeling to enhance agricultural operations. Our work demonstrates how technology-driven solutions can improve crop production projections and fertilizer management, leading to breakthroughs in precision agriculture and sustainable farming methods.

The practical consequences of these discoveries are significant. The practical advantages of our research can be directly applied in real-world agricultural situations to enhance outcomes. The results possess the capacity to transform industrial methodologies, provide guidance for policy choices, and contribute to society by advancing sustainable agriculture, bolstering food security, and augmenting productivity in papaya growing.

While recognizing the limits of the study, such as methodological constraints and the scope of the research, it is crucial to acknowledge the necessity for further investigation into new environmental elements, improvement of predictive models, and the broader implementation of IoT technology in agriculture. The presence of these constraints can impact the understanding of the findings, emphasizing the need for further investigation in this field.

Future research should prioritize the integration of more data sources, improving predictive models to enhance the accuracy of crop output estimates, and expanding the Internet of Things (IoT) ecosystem to encompass a wider array of agricultural operations. Unresolved inquiries pertain to the enduring viability of Internet of Things (IoT) solutions in agriculture and the capacity of these technologies to expand in various farming environments.

7.2 Further Suggested Works

The study on " AN IoT-BASED SOIL AND PLANT NUTRIENT MANAGEMENT OF PAPAYA PRODUCTION IN BANGLADESH", with consequences that surpass its existing discoveries. This study emphasizes the possibility of ongoing inquiry and advancement in agricultural techniques, specifically in enhancing soil and plant nutrient management using IoT technology.

An area with great potential for future research involves improving the predictive modeling techniques used in IoT-based systems. Methods such as ARIMA, SARIMAX, Random Forest, XGBoost, Linear Regression, Logistic Regression, and Multilinear Regression can be improved to increase the precision of crop yield forecasts and nutrient management approaches in papaya farming. Refining these models and investigating novel algorithms may result in more accurate decision-making tools for farmers, ultimately enhancing agricultural output and sustainability.

Another crucial area for investigation is the examination of the long-term viability of Internet of Things (IoT) technologies in the field of agriculture. This involves evaluating their long-term durability, adaptability to various farming conditions, and ecological footprint. Gaining a comprehensive understanding of the long-term viability of these technologies is crucial for their extensive acceptance and effective incorporation into various agricultural systems.

An in-depth comprehension of the elements impacting crop health and yield can be achieved by analyzing weather patterns, pest infestations, market trends, and other pertinent data. Researchers can enhance the effectiveness of decision support systems for farmers by utilizing a wide range of datasets, enabling them to make well-informed decisions on soil and nutrient management strategies.

Investigating scalability and the widespread adoption of solutions based on the Internet of Things (IoT) are additional important areas to explore in the future. The research should prioritize the identification of obstacles that prevent farmers from adopting IoT technology, the examination of tactics to encourage its wider usage and the assessment of the economic viability of using IoT technology in different agricultural contexts. Gaining a comprehensive understanding of these elements is crucial to successfully expanding the use of these technologies and ensuring that their tangible advantages are accessible to a wider agricultural community.

Finally, additional research should investigate the environmental consequences of IoT-based systems for the management of soil and plant nutrients. An examination of the effects of technology-driven agricultural practices on soil health, water conservation, pesticide consumption, and general environmental sustainability will yield vital knowledge about the ecological advantages and difficulties they present.

7.3 Limitations/ Conflict of Interests

The following are some of the gaps and restrictions in our study, "AN IoT-BASED SOIL AND PLANT NUTRIENT MANAGEMENT OF PAPAYA PRODUCTION IN BANGLADESH":

1. Sample Size: The limited sample size of Bangladeshi papaya growers in our study limited the applicability of the results to a larger population.

2. Data Collection Methods: Our analysis's robustness was impacted by biases or gaps in our dataset, which were created by our dependence on self-reported data and the availability of specific information.

3. Resource Constraints: Our study's comprehensiveness was impacted by the limited amount of time, money, and technology that allowed us to conduct a thorough analysis and collect a large amount of data.

4. External Influences: We had no control over the weather, market dynamics, or policy changes, but they had an impact on our outcomes.

5. Methodological constraints: There were inherent constraints in our methodology, which affected how our results were interpreted. These limitations included the choice of particular models or tools.

6. Scope of Study: We may have missed certain pertinent elements that affect agricultural practices and results because our research was mainly concerned with IoT-based nutrient management in papaya production.

7. Geographical Specificity: Our conclusions are unique to Bangladesh and cannot be applied directly to other areas with dissimilar farming techniques, weather, or socioeconomic circumstances.

By being aware of these limits, our research hopes to offer an open evaluation of the procedure and results, providing guidance for future investigations to overcome these limitations and expand on our current discoveries.

REFERENCES

[1] Olubode, Olusegun Olufemi, Olubukola Motunrayo Odeyemi, and Isaac Oreoluwa Olatokunbo Aiyelaagbe. "Influence of environmental factors and production practices on the growth and productivity of
pawpaw (Carica papaya L.) in southwestern Nigeria–A review." Fruits 71.6 (2016): 341-361.

[2] Islam, Md Reazul, et al. "Machine learning enabled IoT system for soil nutrients monitoring and crop
recommendation." Journal of Agriculture and Food Research 14 (2023): 100880.

[3] Dash, D., S. B. Gupta, and Sonali Deole. "Effect of integrated nutrient management on growth and
nutrient uptake in papaya (Carica papaya L.) At nursery level." Journal of Pharmacognosy and Phytochemistry 6.5 (2017): 522-527.

[4] Kumar, Dubbaka Naveen, et al. "Effect of plant nutrients on quality and shelf life of papaya (Carica
papaya L.) Cv. Taiwan Red Lady." Journal of Pharmacognosy and Phytochemistry 10.1 (2021): 1912-1916.

[5] Singh, D. A. M. A. N. D. E. E. P., and SHAILESH KUMAR Singh. "Efficient Usage of Water and
Fertilizers for Papaya Production-A review." Annals of Biology 35.2 (2019): 258-267.

[6] Madani, Babak, et al. "Preharvest calcium applications improve postharvest quality of papaya fruits
(Carica papaya L. cv. Eksotika II)." Journal of Plant Nutrition 39.10 (2016): 1483-1492.

[7] Poojitha, Yanamala, et al. "Functionalized Carbon Quantum Dots Derived from Zelkova serrata Plant
Leaves for the Detection of Normetanephrine in Geriatric Plasma Samples and ROS-Induced Antibacterial
Applications Using a Plausible Mechanistic Approach." BioChip Journal (2024): 1-14.

[8] Elbasi, Ersin, et al. "Crop prediction model using machine learning algorithms." Applied Sciences 13.16
(2023): 9288.

[9] Vázquez-Hernández, Marcos V., et al. "Effect of Glomus mosseae and Entrophospora colombiana on
plant growth, production, and fruit quality of 'Maradol'papaya (Carica papaya L.)." Scientia Horticulturae 128.3 (2011): 255-260.

[10] Hu, Yuncai, Zoltan Burucs, and Urs Schmidhalter. "Effect of foliar fertilization application on the
growth and mineral nutrient content of maize seedlings under drought and salinity." Soil Science and Plant
Nutrition 54.1 (2008): 133-141.

[11] Santos, Eduardo Monteiro, et al. "Planting spacing and NK fertilizing on physiological indexes and fruit
production of papaya under semiarid climate." Bragantia 75 (2015): 63-69.
[12] KHAdE, Sharda Waman, and Bernard Felinov RodRIguES. "Studies on effects of arbuscular mycorrhizal (Am.) fungi on mineral nutrition of Carica papaya L." Notulae Botanicae Horti Agrobotanici
Cluj-Napoca 37.1 (2009): 183-186.
[13] Anjanawe, S. R., et al. "Effect of plant growth regulators and growth media on seed germination and
growth vigor of papaya." Annals of plant and soil research 15.1 (2013): 31-34.A.
[14] Olubode, Olusegun Olufemi, Olubukola Motunrayo Odeyemi, and Isaac Oreoluwa Olatokunbo Aiyelaagbe. "Influence of environmental factors and production practices on the growth and productivity of
pawpaw (Carica papaya L.) in southwestern Nigeria–A review." Fruits 71.6 (2016): 341-361.
[15] Aldillah, Rizma. "Consistency Regulations Implementing Rules and Regulation of Food Derived
Specifically Sustainable Agricultural Land in West Java Province."
[16] da Silva Souza, Luiz Francisco. "Arlene Maria Gomes Oliveira Luiz Francisco da Silva Souza1 Eugênio
Ferreira Coelho1."
[17] Bindu, B., and Bindu Podikunju. "Nutrient requirement of papaya (Carica papaya L.) for yield optimization and commercial cultivation under Kerala conditions." Journal of Krishi Vigyan 5.2 (2017): 122-
127
[18] Chang, Yu Yang, Ji Tan, and Jiunn Luh Tan. "Knowledge and attitude of Malaysian fruit growers on
integrated pest management (IPM)." ASM Sc. J 16 (2021): 1-13.
[19] Mahouachi, Jalel, and Elías Marrero-Díaz. "Plant growth and fruit nutrient changes in Carica papaya L.
genotypes subjected to regulated deficit irrigation." Life 12.11 (2022): 1831.
[20] Ghosh, Promila, et al. "Recognition of Sunflower Diseases Using Hybrid Deep Learning and Its
Explainability with AI." Mathematics 11.10 (2023): 2241.
[21] Ravishankar, H., G. Karunakaran, and S. Hazarika. "Nutrient availability and biochemical properties in
the soil as influenced by organic farming of papaya under Coorg region of Karnataka." II International
Symposium on Papaya 851. 2008.
[22] Ugo, Nwamarah Joy, Adesanmi Raymond Ade, and Asogwa Tochi Joy. "Nutrient composition of Carica

papaya leaves extracts." Journal of Food Science and Nutrition Research 2.3 (2019): 274-282.
[23] Daagema, A. A., P. N. Orafa, and F. Z. Igbua. "Nutritional potentials and uses of pawpaw (Carica
papaya): A review." European Journal of Nutrition & Food Safety 12.3 (2020): 52-66.
[24] Vos, C. H. A. N. T. A. L., and A. L. Y. S. S. A. Cho. "Preliminary report on the mineral composition of
papaya soil and plant tissue in Puna, Hawaii under different land use practices." Pac. Agric. Nat. Resour. 1
(2016): 1-10.
[25] Kumari, Pushpa, et al. "Fertilizer requirement of papaya (Carica papaya L.) for commercial cultivation
under Bihar condition." International Journal of Chemical Studies 2019a 7.3 (2019): 1730-1732.
[26] Morales-Payan, J. Pablo, and William M. Stall. "Papaya (Carica papaya) response to foliar treatments
with organic complexes of peptides and amino acids." Proceedings of the Florida State Horticultural Society.
Vol. 116. 2003.
[27] Dahunsi, S. O., et al. "Crop performance and soil fertility improvement using organic fertilizer produced
from valorization of Carica papaya fruit peel." Scientific Reports 11.1 (2021): 4696.
[28] Said, Zafar, et al. "Intelligent approaches for sustainable management and valorization of food waste." Bioresource Technology (2023): 128952.
[29] Bilal, Anas, et al. "Increasing Crop Quality and Yield with a Machine Learning-Based Crop Monitoring
System." Computers, Materials & Continua 76.2 (2023).

Appendix A

Course Outcomes, Complex Engineering Problems (EP), and Complex Engineering Activities (EA) Addressing

Title: AN IoT-BASED SOIL AND PLANT NUTRIENT MANAGEMENT OF PAPAYA PRODUCTION IN BANGLADESH

Student ID: 202-15-3803, 202-15-3815

CO Description for FYDP

CO	CO Descriptions	PO
Phase -I		
CO1	Integrate recently gained and previously acquired knowledge to identify a **SOIL AND PLANT NUTRIENT MANAGEMENT OF PAPAYA PRODUCTION** for the Final Year Design Project (FYDP)	**PO1**
CO2	Analyze different aspects of the goals in designing a solution for this FYDP	**PO2**
CO3	Explore diverse problem domains through a literature review, delineate the issues, and establish this goal for the FYDP	**PO4**
CO4	Perform economic evaluation and cost estimation and employ suitable project management procedures throughout the development life cycle of the FYDP	**PO11**
Phase -II		
CO5	Design and develop technical solutions and system components or processes that meet specified requirements, ensuring compliance with agriculture and safety standards, as well as considering cultural, socioeconomic, and environmental factors in this FYDP	**PO3**
CO6	Choose and apply appropriate methodologies, resources, and contemporary engineering and IT technologies to address complex engineering processes, encompassing prediction, and modeling, while adhering to relevant constraints in this FYDP	**PO5**
CO7	Analyze societal, health, safety, legal, and cultural considerations and associated responsibilities in the context of professional engineering practice and the resolution of this problem, employing logical reasoning guided by contextual understanding.	**PO6**
CO8	Comprehend and evaluate the enduring sustainability and impact of professional engineering endeavors in addressing intricate engineering challenges within social and environmental frameworks.	**PO7**
CO9	Implement ethical principles and adhere to professional standards and norms in this FYDP	**PO8**
CO10	Capable of operating proficiently both individually and as a team member or leader across diverse teams and interdisciplinary settings in this FYDP.	**PO9**

CO11	Proficiently communicate with the engineering community and broader society regarding complex engineering endeavors, including the ability to comprehend and generate comprehensive reports and design documentation, as well as provide and receive clear instructions throughout this FYDP.	**PO10**
CO12	Acknowledge the importance of self-directed and life-long learning within the evolving landscape of technology, and possess the readiness and capability to engage in lifelong learning endeavors.	**PO12**

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP), and Attainment of Complex Engineering Activities (EA)

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP):

SN	EP Definition	Attainment	CO	Justification (with Knowledge Profile)	References
1.	**EP1: Depth of Knowledge required**	Yes	CO1, CO2, CO3, CO5, CO6, CO7 and CO8	Data collection from IoT sensors and the development of prediction models for nutrition management necessitate knowledge of **K3. K4** through the use of IoT, with IoT sensors collecting data and storage them in the cloud,	**Page no:** [17-18,30-32] **Section:** [3.2,4.1]
				K5 is also addressed, with the integration of IoT components and the construction of a user-friendly interface,	**Page no:** [17-19] **Section:** [3.2]
				The project addresses the **K6** because here we use ML models Random Forest, XGBoost, linear regression, logistic regression, ARIMA, SARIMAX, and multi-linear regression	**Page no:** [17-19,32-37,40-44] **Section:** [3.2,4.2,4.3,5.2]

				In **K7** it ensures that our project's impact on life and the environment is good for social, economic, and agriculture. Ethical Impact on Society	**Page no:** [57-58] **Section:** [6.3]
				This project ensures **K8 (Research Literature)** by synthesizing A review informed by our approach to existing agricultural IoT applications. Here we review recent studies on soil and plant nutrients of papaya tree	**Page no:** [9-12] **Section:** [2.2, 2.3]
2.	**EP2: Range of Conflicting Requirements**	Yes	CO2, and CO7	This project addresses **EP-2** by managing soil and plant nutrients of papaya trees using a machine learning model to detect the nutrient level for tree height. This involves understanding the nutritional needs of papaya plants at different growth stages, interpreting soil test results, and identifying nutrient deficiencies or excesses.	**Page no:** [1-3] **Section:** [1.1,1.2]
3.	**EP3: Depth of analysis required**	Yes	CO2, and CO6	This project addresses **EP-3** by meticulously experimental outcomes, highlighting ML as the nutritional needs of papaya plants at different growth stages	**Page no:** [40-44] **Section:** [5.2]
4.	**EP4: Familiarity of Issues**	Yes	CO8	This project's interdisciplinary approach extends beyond computer science and engineering, impacting agriculture and the environment through soil nutrient management **EP-4**	**Page no:** [2,4-5] **Section:** [1.2, 1.4]

5.	**EP5: Extends of application codes**	No	CO5	N/A	N/A
6.	**EP6: Extends of stakeholders involved and conflicting requirements**	Yes	CO8	Collaboration with stakeholders such as papaya farmers, agricultural specialists, government officials, and IoT technology vendors is critical. Understanding different viewpoints and addressing competing needs guarantees that our IoT-based nutrient management system meets the needs of Bangladesh's papaya crop sector. This meeting will guide us in determining the specific type of data we need for our analysis.	**Page no:** [25-27] **Section:** [3.4]
7.	**EP7: Interdependence**	Yes	CO5	This project's comprehensive approach addresses high-level problems by integrating various components across data collection, statistical analysis, and proposed methodology, ensuring a holistic solution to complex challenges in agriculture which ensures **EP-7**.	**Page no:** [17-2130-32] **Section:** [3.2, 4.1]

Addressing CO11 with Complex Engineering Activities (EA) [Some or all of the following]:

SN	EA Definition	Attainment	CO	Justification	References
1.	**EA1: Range of resources**	Yes	CO11	Our project utilizes diverse resources such as soil and plant nutrients, with ML frameworks, real-time datasets, and ethical considerations to ensure systematic research and contribute to advancements in soil and plant nutrient management	**Page no:** [17-21] **Section:** [3.2]

2.	**EA2: Level of interaction**	No		N/A	N/A
3.	**EA3: Innovation**	No		N/A	N/A
4.	**EA4: Consequences for society and the environment**	Yes		This project contributes to society by improving soil nutrients and the growth of papaya production through advanced soil and plant management while also promoting environmental sustainability by employing efficient computational resources and adhering to ethical guidelines for patient data privacy.	**Page no:** [55-59] **Section:** [6.2, 6.3, 6.4]
5.	**EA-5: Familiarity**	Yes		This project expands upon existing research by examining soil and plant nutrient management using ML models. Also looking after plant leaf number and height of the plant, using a similar regression model	**Page no:** [8-14] **Section:** [2.1,2.2, 2.3]

Addressing CO (4, 9, 10, and 12):

SN	COs	Attainment	Justification	References
1	CO4	Yes	This project addresses **CO4** by integrating effective project management and financial oversight, ensuring meticulous planning, resource allocation, and budget estimation for optimal resource utilization throughout the research lifecycle. And guaranteed to provide the best possible resource utilization throughout the study project.	**Page no:** [24-27] **Section:** [3.4]
2	CO9	Yes	The project demonstrates adherence to	**Page no:**

			ethical principles by prioritizing patient privacy, obtaining informed consent, and transparently documenting the research process, ensuring responsible knowledge dissemination and societal well-being through the ethical application of advanced agriculture technologies that comply with **CO9**.	[57-58] **Section:** [6.3]
3	CO10	No	N/A	N/A
4	CO12	Yes	The project's dedication to continuous learning **(CO12)** and adaptation within the advanced IOT technology and use of IOT sensors for data collection, rigorous statistical analysis, meticulous methodology development, and experimental results and analysis, showcase a commitment to staying updated and refining techniques to address modern challenges.	**Page no:** [17-19,25-27,33-37,40-53] **Section:** [3.2, 3.4, 4.1,4.2,4.3,5.2,5.3]

AN IOT-BASED SOIL AND PLANT NUTRIENT MANAGEMENT OF PAPAYA PRODUCTION IN BANGLADESH

ORIGINALITY REPORT

10%	5%	5%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	2%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	2%
3	Md Reazul Islam, Khondokar Oliullah, Md Mohsin Kabir, Munzirul Alom, M.F. Mridha. "Machine learning enabled IoT system for soil nutrients monitoring and crop recommendation", Journal of Agriculture and Food Research, 2023 Publication	<1%
4	T. E. Kowalchuk, E. de Jong. "Shelterbelts and their effect on crop yield", Canadian Journal of Soil Science, 1995 Publication	<1%
5	Ton Duc Thang University Publication	<1%
6	Md Moniruzzaman, Sayan Deb Dutta, Rumi Acharya, Ki-Taek Lim, Jongsung Kim. "NIR-	<1%