AI-Driven Prediction of Seedling Growth in Different Soil Media Using Machine Learning

By
Akash Sen
212-15-4097

Rokeya Ahsan Jerin
212-15-4197

FINAL YEAR DESIGN PROJECT REPORT

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

Supervised by

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

Co-Supervised by

Amir Sohel
Assistant Professor
Department of Computer Science and EngineeringDaffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
Dhaka, Bangladesh

September 17, 2025

APPROVAL

This Project titled **"AI-Driven Prediction of Seedling Growth in Different Soil Media Using Machine Learning"**, submitted by **Akash Sen; ID: 212-15-4097** and **Rokeya Ahsan Jerin; ID: 212-15-4197** 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 **17 September, 2025**.

BOARD OF EXAMINERS

Dr. S.M Aminul Haque (SMAH) — **Chairman**
Professor & Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Ms. Nazmun Nessa Moon (NNM) — **Internal Examiner**
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Md. Abbas Ali Khan(AAK) — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Md. Zulfiker Mahmud (ZM) — **External Examiner**
Professor
Department of Computer Science and Engineering
Jagannath 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 projectnor 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 Computer Science and EngineeringDaffodil International University

Co-Supervised by:

Amir Sohel
Assistant Professor
Department of Computer Science and EngineeringDaffodil International University

Submitted by:

Akash

Akash Sen
Student ID: 212-15-4097
Department of Computer Science and EngineeringDaffodil International University

Jerin

Rokeya Ahsan Jerin
Student ID: 212-15-4197
Department of Computer Science and EngineeringDaffodil International University

ACKNOWLEDGEMENTS

This work would not have been possible without the support and contributions of many individuals over the past two semesters. We are deeply grateful to everyone who has assisted us in one way or another.

First, we express our heartfelt thanks and gratefulness to the almighty for His divine blessing making it possible for us to complete the **Final Year Design Project (FYDP)** successfully.

We are grateful and wish our profound indebtedness to **Dr. Fizar Ahmed**, **Associate Professor**, Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. Deep knowledge and keen interest of our supervisor in the field of **Machine Learning** 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 themat all stages have made it possible to complete this project.

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

We would like to thank our entire course-mates at Daffodil International University, who took part in this discussion while completing the coursework.

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

ABSTRACT

This work offers a data intensive framework for assessing seedling growth in various soil media by predicting plant performance from key environmental conditions. The farming yields are hampered by poor utilization of data and conventional selection of soil resulting in trial-and-error soil selection. To resolve this issue, plant growth measurements were recorded from three different plant varieties: Lal Shak(Red Amaranth), Dhoniya(Coriander) and Lau(Bottle Gourd), which were cultivated in various soil mediums. Environmental factors such as temperature, humidity, luminosity, soil moisture, and NPK nutrient levels were recorded using digital sensors. The Linear Regression, Random Forest, and XGBoost machine learning algorithms were trained for predicting the growth of the plants. Among the models, XGBoost achieved the highest predictive accuracy (R^2 = 0.91). However, the Random Forest model (R^2 = 0.89) was selected as the optimal model for this study. This decision was based on its excellent balance of high accuracy, superior interpretability, and robustness against overfitting, making it a more practical and reliable tool for generating actionable insights. The Linear Regression baseline model fell short because it could not learn complex interactions. The research discovered that the most important factors were the levels of nutrients, light, and moisture in the soil, and the optimum medium of choice varied among the various varieties. The results establish how predictive models can inform soil and growth condition choice based on data and thereby enhance agricultural performance. The contribution of this work is a proof of concept framework with potential for scalability and practical for smallholder farmers and agricultural researchers to optimize seedling cultivation using data analytics instead of manual observation alone.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

Agriculture is also one of the most important parts of the economy and the food supply. In the early stages of a plant's life, it is important for seeds to grow well because this will lead to more cultivation as time goes on. Many things that affect each other, like nutrients, moisture, temperature, and humidity of the soil, affect how plants grow. It is more likely that these factors are related. Farmers used to make decisions based on their experiences and what they saw with the eyes of the farmer. Farmers used to make decisions based on their own experiences and what they saw with their own eyes. This method works to some extent, but it is not exact and often leads to inconsistent results and wasted resources [4]. Also, the environment can change quickly, and traditional methods can't keep up with these changes in real time. This void makes it critical to use data-based methodologies that yield more accurate and implementable guidance for the optimization of seedling cultivation.

With the advent of modern technology, it is now possible for us to see soil and environmental conditions in real time using sensors and IoT technology. Now, NPK nutrients, moisture, temperature, and humidity can be quantified and saved as collections of data, which can be processed. Meteorological monitoring stations and environmental observation networks provide accurate and well-documented observations, and these can be used to improve crop growth models [11]. The equipment allows us to combine field observations and computer models for better forecasts.

Machine learning has become increasingly useful in agriculture because it is able to analyze vast collections of information and discover hidden patterns not immediately evident to the eye. Random Forest models, for instance, perform well for yield prediction of crops because they are able to manage intricate interactions of factors and also discover which factors are most important [22]. XGBoost models sometimes perform slightly better but are more complex and more difficult to interpret. Linear Regression is less complex but manages sophisticated interactions poorly and so usually provides poorer performance.

In recent years, sensor-based systems have helped support smart farming practices. Not only can these systems forecast the result of growth, but they can also tell us which environmental parameters influence plants the most [28]. Three vegetable seedlings of coriander, red amaranth, and bottle gourd were cultivated in three different media in our experiment: cocopeat, sand, and vermicompost. In sand, water drains out and the medium is poor in nutrients, water is held and soil aeration is improved in cocopeat, and vermicompost contains high organic content and good microbes-Machine learning algorithms predicted growth and measured the individual contribution of each parameter by sniffing and snapping photos of plants.

The general aim of work is to provide explanatory and practical predictions to the farmers. The paper is interested in providing farmers insights instead of implementing models in a black box approach. With this kind of approach, farmers will consume more water, fertilizer, and land.This will efficiently produce better crops which are healthier and more food in the long run.

1.2 Motivation

Modern agriculture faces a very difficult challenge because the world needs a greater supply of food, the weather is unpredictable, and resources need to be used in a way that is environmentally friendly. To meet this challenge, we need to move from reactive to proactive farming. This is a massive and complex issue in terms of computing based on data. The growth of plants is not a linear process. Instead, it is a dynamic system which is influenced by numerous factors i.e. temperature, humidity, light and nutrients in the soil. They all interact with each other in a way that isn't straight, which affects the health of a plant and how big it gets. Agricultural models and human senses can't make these complicated structures again. This project's primary motivation is its extreme technical complexity. A flexible predictive model is required to address this issue effectively.

Although models can help us understand how environmental factors affect plant growth, they are unable to explain the complex and non-linear interactions between those factors. These complex interactions cannot be replicated or outcomes predicted with the level of accuracy required for the best possible resource allocation using agronomic models and human intuition alone. This project's primary objective is to address a high level of computational complexity.

To properly address this issue, a predictive model that is adaptable is required. While simple statistical models can provide insight into how environmental factors affect plant growth, they are unable to explain complex, non-linear interactions between those factors.

This requires more advanced machine learning algorithms. The computational solution to this need is the use of regression models, which lets us look for patterns in multi-variate data sets that aren't easy to see. Although the field has experienced significant progress in the application of high-throughput technologies, such as crop phenotyping using imaging sensors [12], [14], and robust deep learning models like CNNs for agricultural purposes [9], this study shows that an effective predictive model can be constructed from more generalized, sensor-based traits [29]. The main goal, and therefore the reason for doing it, is to turn the complex behavior of the real world into a useful model that can give you useful information. This project aims to show how a well-designed machine learning pipeline can solve this problem.

One of the main reasons for doing this work is to come up with a solution that is easy to understand and useful. Even though very complicated models can make performance a little better, they are "black boxes" because it's hard for end-users, like farmers or agronomists, to understand why a certain recommendation is being made. This project aims to create not only good models but also clear explanations of what the main factors that affect growth are by using a model like the Random Forest Regressor.

One of the main computational and practical goals is to make things easier to understand. It closes the gap between a good algorithm and how it is used in the real world, which makes the technology reliable and easy to use. This project will be helpful for data science and machine learning. This is a one-of-a-kind chance to use the information you've learned in theory to solve a real problem. The world needs a greater supply of food, climate change is unpredictable, and resources must be used sustainably, making modern agriculture a very challenging task. To fulfil this challenge, we have to shift from farming that is reactive to farming that is proactive. From a technological perspective, this challenge is a massive and intricate data-driven problem. The process of plant growth is not simple. Rather, it is a dynamic system that is influenced by a wide range of variables, including light, temperature, humidity, and soil nutrients. The health and ultimate size of a plant are influenced by these variables in a non-linear manner.

These intricate relationships cannot be replicated or outcomes predicted with the level of accuracy necessary for the best possible resource allocation using agronomic models and human intuition alone. This project's primary motivation is its high level of computational complexity.

To effectively solve this issue, a flexible predictive model must be employed. Basic statistical models can help us understand how plants grow in relation to environmental factors. However, they can't tell us why these variables work together in a complicated, non-linear way. This needs more advanced machine learning algorithms. Regression models are the computational solution to this problem. They let us find patterns in multi-variate data sets that are hard to see. While the field has made substantial advancements in the use of high-throughput technologies, including crop phenotyping through imaging sensors [12], [14], and resilient deep learning models like CNNs for agricultural applications [9], this study demonstrates that a functional predictive model can be developed from more generalized sensor-derived traits [29].The main goal, and therefore the reason for doing it, is to turn the complex behavior of the real world into a useful model that can give you useful information. This project aims to show how a well-designed machine learning pipeline can solve this problem.

One of the main reasons for doing this work is to come up with a solution that is easy to understand and useful. Even though very complicated models can make performance a little better, they are "black boxes" because it's hard for end-users, like farmers or agronomists, to understand why a certain recommendation is being made. This project aims to create not only good models but also clear explanations of what the main factors that affect growth are by using a model like the Random Forest Regressor. One of the main computational and practical goals is to make things easier to understand. It closes the gap between a good algorithm and how it is used in the real world, which makes the technology reliable and easy to use.

This project will be helpful for data science and machine learning. This is a one-of-a-kind chance to use what you've learned in class to solve a real problem that will have a big positive effect. I have learned a lot about machine learning by doing this project. From data preprocessing and feature engineering to model training, testing, and, most importantly, interpretation I have learned a lot. It has been very helpful for me to learn how to work with different datasets and choose the best model for a task by weighing performance against simplicity. This project not only showed me that I could use technology, but it also made me understand how technology can be used to solve big problems around the world. It also set me up to make a meaningful contribution to the field of smart and sustainable agriculture in the future.

1.3 Objectives

The objective of this project is to create and deploy a system based on machine learning for predicting and analyzing the growth performance of different types of seedlings cultivated in different soil media. Through measurements of environmental sensors like soil nutrient content (NPK level), moisture, temperature, humidity, and light, the system calculates the optimal growth conditions for each plant species.

The primary goal of this study is to utilize supervised machine learning models to efficiently predict plant growth behaviors and determine the extent to which various environmental factors influence seedling well-being and ultimate height.Since the relationship between the early environment and later plant growth is primarily non-linear, linear regression, which was f irst used as a control model, performed poorly ($R^2 \approx 0.35$, RMSE ≈ 3.50 cm). Because Random Forest and XGBoost can manage high-order non-linear interactions between features, the study targets these problems. While XGBoost provided a slightly higher accuracy ($R^2 = 0.91$, RMSE = 1.05 cm), it was more complex and less interpretable than Random Forest, which had an $R^2 = 0.89$ with RMSE = 1.15 cm and feature importance that was easy to understand and prediction quality that was good.

Additionally, the project focuses on real-world agricultural applications that transform raw sensor data into insights. For crops like Lau, Dhoniya, and Lal Sakh, their findings include extremely specific recommendations for the ideal nutrient ratios, soil water content, temperature ranges, and light intensities. These recommendations would help farmers, horticulturists, and agricultural researchers have the best growing conditions possible, leading to high yield and quality.

By developing a low-cost data-driven method for monitoring and forecasting plant growth without the need for specialized hardware and deep learning networks, this study advances precision farming. The solution is adaptable and expandable and can be prolonged to future IoT devices and affordable mobile applications for real-time field application

1.4 Methodology

The study utilizes a systematic machine learning methodology to forecast plant development predicated on preliminary soil and environmental observations. Three crops were selected, each with ten replicates. Day 7 sensor readings and manual measurements were collected, and Day 19 plant lengths were used as the target for prediction.

Values were given in the format of tables and consisted of nutrient concentration (N, P, K), soil water, temperature, humidity, light intensity, and medium type. After checks for consistency, missing or invalid values were processed and categorical variables were label-encoded. Plant names were used only to link records and not for training models.

Three regression models, namely Linear Regression, Random Forest, and XGBoost, were used to forecast Day 7 conditions to predict Day 19 outcomes. An 80:20 split was conducted for unbiased testing, with R^2 and RMSE as the performance measuring criteria. Random Forest was used based on high model stability and accuracy and the ability to determine the most important environmental drivers of growth. This pipeline shows an open, reproducible process: collect and clean up measurements, merge early predictors with growth outcomes, train multiple models, and pick the most interpretable and best-performing one to generate environment specific recommendations.

1.5 Project Outcome

1.5.1 Predictive Modeling

This project builds a machine learning pipeline that makes predictions about how plants will grow on Day 19 from the parameters of the environment and soil on Day 7. We utilized the Random Forest model as our go-to utility because it is incredibly accurate and easy to implement. It provides us with helpful forecasts for three crops in different scenarios.

1.5.2 Model Comparison

We trained and tested Linear Regression, random forest and XGBoost on the data. Random Forest was almost the best and was easy to interpret also. XGBoost was a bit more difficult to interpret but it worked a little better. This evidence suggests that the Random Forest is an effective tool which can be utilized to assist farmers in making informed decisions. This study demonstrates a organized, data-driven approach to predicting crop growth in case of sensor information and machine learning. This study provides a solution to short cycle crop prediction problems that can be reproduced and allows future research to use the model, through construction, test, and feature importance reporting.

1.5.3 Contribution to Research

The paper shows that it is a systematic and data-intensive method of predicting crop development when sensor measurements are combined with machine-based learning. This study provides a replicable recipe to other short-cycle crop forecast problems and future research by way of model building, evaluation and documentation of feature importance.

1.5.4 Practical Agricultural Benefits

The given approach enables identifying the best environmental conditions early on, minimizing the trial and error related to the cultivation process, and allowing to allocate the resources efficiently. The outcomes are useful to farmers and agricultural practitioners in that they provide a high yield consistency, minimized wastage and maximized soil and light utilisation with minimal technical limitations.

1.5.5 Educational Value

This work provides a real-world example for practitioners, researchers, and students exploring AI-based agriculture. It bridges theoretical knowledge with applied research, encouraging future innovation of scalable, sensor-empowered growth prediction systems in sustainable agriculture.

1.6 Organization of the Report

Chapter 1: Introduction Introduces the research problem, motivation, and objectives of this research. Provides background, scope, and significance of using machine learning to forecast plant growth.

Chapter 2: Background Compiles literature, main concepts, and current research in data-intensive farming and crop growth modeling. Focuses on the limitations of traditional cultivation methods and the gap in research addressed through this research.

Chapter 3: Methodology Describe experimental design, dataset structure, preprocessing tasks, and machine learning techniques used. Describes the merging of Day-7 observations with Day-19 plant heights, feature extraction, and metrics used for evaluation.

Chapter 4: Implementation and Results Describes predictive model implementation, e.g., Linear Regression, Random Forest, XGBoost. Compares model performance, assesses most influential features on growth, and gives visual representations of results..

Chapter 5: Engineering Standards and Challenges Meets engineering specifications, ethical treatment of data, and design challenges. Addresses topics such as sensor calibration, small sample size, model generalizability, and what was done to mitigate them.

Chapter 6: Conclusion and Future Work Summarizes the results of the study, claims its contributions, and provides directions to be explored in future research, such as the inclusion of datasets, the assessment of other crops, and the improvement of the accuracy of the model for application in general agricultural practices.

Chapter 2

Background

2.1 Introduction

The agricultural sector is undergoing a significant transformation, with modern technologies playing a pivotal role in optimizing crop management and increasing yields. Traditional farming practices often rely on experience-based decisions, which can be inefficient, labor-intensive, and prone to inconsistency. Soil texture, soil nutrients, and environmental parameters all differ, bringing complexity in cultivation, leading to weak plant growth and less productivity. Machine learning (ML) and data analytics are a strong answer to the problems mentioned above that provide accurate predictions and actionable recommendations on how farmers should practise more productive and sustainable agriculture [1].

One of the outstanding fields of study in precision agriculture has been predictive plant growth modeling. The models allow the farmers to predict the growth of the plants at an early stage of the growing process so that timely action can be taken to optimize the resources. The effectiveness of the use of ML models to predict plant growth and crop yield has been reported in a number of studies. As an example, a Linear Regression as a simple model provides simplicity and interpretability to allow a researcher to infer the direct correlation between the environmental parameters and the growth of plants [23]. Plant growth is however a complex process and a number of environmental and soil conditions interact with each other in non-linear manners [22]. More advanced models like Random Forest Regressor and XGBoost Regressor are hence utilized to model these relations.

Random Forest Regressor is a very effective ensemble learning model which has been applied tremendously in agriculture. Random Forest Regressor builds many decision trees simultaneously, each of which models independently the input variables of the environment and the soil. The average of the output of all the trees makes the final prediction which mitigates the risk of overfitting and guarantees high performance. In this way, the model can learn complex, non-linear relationships between features such as NPK levels, soil moisture, temperature, humidity, light intensity, and media type [13]. Its ability to be understood, especially through feature importance analysis, helps us understand the most important factors that affect plant growth. This is very useful for making decisions in agriculture.

Similarly, XGBoost Regressor (eXtreme Gradient Boosting) is an advanced ensemble model that builds trees sequentially, with each tree learning from the errors of the previous ones [14]. This boosting method improves predictive accuracy and performs extremely well with structured, sensor-based tabular data prevalent in agricultural research [26]. While XGBoost can be slightly more accurate than Random Forest, its complexity and reduced transparency make it less interpretable for stakeholders such as farmers. IoT-based precision farming has widened the purview of plant growth monitoring. Environmental parameter sensors in real time—temperature, humidity, soil moisture, NPK, and light intensity—offer comprehensive datasets to ML models, which increase predictive validity [21]. The fusion of time-series sensor data and plant imaging helps to increase model trustworthiness, enabling the identification of growth anomalies at an early stage and taking remedial measures in a timely manner [12]. The early environmental conditions, such as Day 7 measurements, strongly correlate with growth outcomes at later stages (Day 19), for which early prediction is very important for farmers and agronomists [15].

Other than crop growth prediction, researchers also highlighted AI and IoT applications in seeds and resource optimization. For example, real-time weather prediction models have been developed to enhance resource planning in agriculture [17]. Advances in computer vision techniques are being applied to automate seed identification, improving both classification accuracy and efficiency [18]. Similarly, the role of AI in seed technology has gained attention, offering new opportunities for monitoring seed quality, germination, and viability [19]. Embedded sensing technologies are being integrated into smart agriculture frameworks at the same time. This allows farmers to make decisions and use resources better in real time. [28]. These new ideas show that AI and IoT have a wide range of uses in farming, from modeling how plants grow to optimizing seeds and managing resources smartly.

Even though these ensemble methods are powerful, Linear Regression is a good starting point because it is simple and easy to understand. By comparing its performance to Random Forest and XGBoost, researchers can determine the additional benefits of complex models and validate their application for precise prediction [23].

The study's goal is to use environmental and soil conditions from Day 7 to predict how plants will grow on the targeted day which we kept on day 19 . The plants are Dhoniya, Lal Sakh, and Lau, and they are grown in different types of soil. The study aims to identify the environmental conditions most vital for growth, utilizing parameters such as NPK levels, soil moisture, temperature, humidity, and light intensity readings. This method shows how predictive modeling can give farmers useful information that helps them improve their farming methods, increase their output, and use their resources more effectively.

2.2 Literature Review

Table 2.1: Summary of Literature Reviewed.

Author(s)	Year	Title	Methodology	Key Findings
El Sakka et al.	2025	A Review of CNN Applications in Smart Agriculture Using Multimodal Data.[9]	CNN-based Analysis	Achieved 87% accuracy in plant water stress detection using a lightweight CNN

Mahant et al.	2025	A Review of Artificial Intelligence's Pivotal Role in Seed Technology.[19]	CNN, ML Classifiers	AI applications in seed technology for identification, sorting, and quality prediction
Mohyuddin et al.	2024	Evaluation of Machine Learning Approaches for Precision Farming in Smart Agriculture System – A Comprehensive Review.[12]	ML Algorithms: Random Forest, SVM, KNN	Evaluation of ML methods for precision farming and there was an 88.89% success rate
Katharria et al.	2024	Machine Learning Applications and Future Research Directions.[15]	ML-based Feature-level & Decision-level Fusion using Random Forest, SVM, Ensemble Models	Achieve an impressive 96.71% accuracy, highlighting the effectiveness of ML.
Pecho et al.	2024	An Approach for Crop Yield Prediction Using Hybrid XGBoost, SVM and C4.5 Classifier Algorithms.[26]	Hybrid ML Models: XGBoost + SVM + C4.5	The best performing crop prediction was Sugarcane at 98.25%.
Pierre et al.	2023	AI Based Real-Time Weather Condition Prediction with Optimized Agricultural Resources.[17]	Artificial Neural Networks (ANN) + Optimization Algorithms	Predict real-time weather conditions and 91.5% accuracy
Sarma et al.	2022	Learning Aided System for Agriculture Monitoring Designed Using Image Processing and IoT-CNN.[13]	IoT Data Collection + CNN-based Image Processing	Achieved 96% accuracy in tomato leaf disease detection using raw image input
Zhao et al.	2022	Automated Seed Identification with Computer Vision: Challenges and Opportunities. [18]	Image Segmentation + Feature Extraction + Classification Models	Highlighting challenges and t the performance of 79.2% accuracy for DL

Gupta et al.	2021	Weather Based Crop Prediction in India Using Big Data Analytics.[1]	Big Data Analytics, Predictive Modeling	The system provides top crop recommendations based on historical yield, soil type
Xiong et al.	2021	A Review of Plant Phenotypic Image Recognition Technology Based on Deep Learning.[14]	CNN Architectures (AlexNet, VGG, ResNet)	Leaf disease classification achieved around 95% training accuracy and up to 96% validation accuracy.
Bhattacharyya et al	2021	Linear Regression Model to Study the Effects of Weather Variables on Crop Yield in Manipur State.[23]	Linear Regression Analysis	Analyzed impact of weather variables on crop yield using linear regression and highest accuracy is for Turmeric (98%)
Rao et al.	2016	Internet of Things (IoT) Based Weather Monitoring System.[21]	IoT Sensor Network: Temperature, Humidity, Soil Moisture	Developed IoT-based system to monitor weather parameters, enabling real-time agricultural decision-making
Jeong et al.	2016	Random Forests for Global and Regional Crop Yield Predictions.[22]	Random Forest Regression	Demonstrated RF model effectiveness for predicting crop yields globally and regionally and the highest accuracy achieved is about 88%.
Ahmad et al.	2016	Use of Organic Fertilizers to Enhance Soil Fertility, Plant Growth, and Yield in a Tropical Environment. [25]	Controlled Experimental Study: Organic Fertilizer Application (Compost, Manure)	Showed that organic fertilizers improved soil fertility and accuracy 86%

2.2.1 Similar Applications

Other researches have involved the use of IoT sensors to measure environmental and soil parameters that have furnished the information on the growth pattern of plants.

Automated online or continuous data are mostly relied on to such studies. In this project, the data of sensors, including temperature, humidity, moisture, NPK level, and light intensity, were recorded manually. Similar research in literature Mahmud et al (2023); Sun et al. (2024) have documented the use of sensor on environmental data of different types of vegetable crops and determined that even when manually collected data is to be used, sensor data which is well preprocessed and read can be used to attain a successful implementation of a predictive model. Secondly, as Dutta et al. (2025) noted, manually obtained data might be integrated with the obtaining of machine-based learning that will become effective in growth predetermination. Gupta et al. (2021) and Bikku et al (2023) also point out that sensor data, provided that it does it, gives it.

Machine Learning Approaches

XGBoost and Random Forest regression model have been widely used in the establishment of plant growth. As our proposed hybrid model whereby Random Forest would be used to model sensors and MobileNetV2 would be used to extract features in plant images presents the same techniques to small datasets, and therefore, the prediction accurate, but still interprets the model, it is seen in their study that ensemble models can be applied and therefore used to store the non-linear and challenging relationship between plant growth and the environment and more so when manually labelled sensor data are used. The combination of the benefits of structured sensor data and visual plant data are our proposed approach that would allow them to be better utilized to predict. El Sakka et al.(2025) also determines that visual and environmental fusion stability are applicable in growth prediction at small scale experiments.

Crop-Specific Case Studies

Studies of the single plant species are conducted underneath the topicality of replicated observance and difference. Coriander et al. (2025) and Allium cepa Dutta et al. (2025) also applied other replicated monitoring setup to achieve the goal of determining proper plant development in different conditions in their other studies. Similarly, our experiment employs 10 replicate plants per species and three replicates of varied soil-fertilizer media, demonstrating that manual data collection can also yield high-quality datasets for machine learning-based growth prediction. Mekonnen et al. (2020) also demonstrates that small-scale replicated datasets can be adequate for solid modeling.

Methodological Contributions

Some of the other significant contributions from previous work are correlating initial-stage environmental sensor readings with resulting growth data, performing feature importance analysis Pierre et al. (2023), and offering ensemble model explainability as considered by Xiong et al. (2021). Our research builds on these foundations and integrates these methods and applies them to sensor readings that have been manually obtained using Random Forest for sensor inputs and MobileNetV2 for image features. This blended approach both anticipates and provides understandable results on the most influential environmental elements, providing actionable data for farmers to optimize growing conditions maximally for any plant species.

2.2.2 Related Research

Over the past two years, the use of IoT and machine learning in agriculture has become increasingly popular for optimizing environmental factors as well as forecasting plant growth. Kalimuthu et al. (2020) and demonstrated that machine learning regression models had the capability of very precise prediction of crop yield depending on the content of soil nutrients and weather conditions. In the same direction, Gupta et al. (2021) used big data analytics to forecast crop dynamics on the basis of weather parameters with emphasized predictive modeling in resource optimization.

Real-time environmental monitoring sensor-based methods also have been widely investigated. Sreekantha et al. (2017) have used IoT-based real time soil temperature, soil moisture and also have used humidity to manage crops accurately. Ullo and Sinha (2020) studied intelligence environment surveillance structures on the basis of IoT sensors and demonstrated better predictive growth modelling of plants created by Ullo et al. (2020). Recent works by Shahab et al. (2025) and Pandey et al. (2025) integrated the use of the IoT-based sensing of soil with machine learning to predict the growth of vegetable plants and resource utilization more effectively.

Hybrid modeling approaches combining regression and ensemble techniques have shown promising predictive ability. Jeong et al. (2016) made predictions of crop yields from environmental characteristics using Random Forests, benefiting from non-linear interactions between features. Xiong et al.(2021) focused on multi-modal fusion of data, combining plant photos and sensor readings, for the aim of enhancing the precision of the predictions. Deep learning research such as El Sakka et al. (2025) applied CNNs during the interpretation of plant phenotypes from sequence images and was very precise in growth estimation.

The only prevailing comment of literature is effect of feature selection and sensor quality. As Kim et al. (2020) and Bhavanandam et al. (2022) explained, target measurements of the environment and soil nutrients can play a crucial role in enhancing the model generalization and performance. In addition to it, research such as that by Hati and Singh et al. (2023) used AI-based phenotyping to assess growth characteristics in various media environments, which validates the necessity of the combination of both data from the IoT sensors and image analysis.

One of the main issues, universally experienced in plant growth forecasting studies, is sensor reading variability and noise of manually collected measures. Shahab et al. (2025) overcame this through integration with machine learning algorithms and IoT-ambient soil and environmental sensor data and enhanced prediction accuracy and reporting the importance of secure high-quality data in practical application.

2.3 Gap Analysis

This gap analysis presents a feature-wise comparison between existing plant growth prediction approaches and the proposed machine learning–based prediction system.

Table 2.3 Gap Analysis

Features	Bhavanandam et al.(2022) XGBoost	Mahmud et al.(2023) Random forest	Pecho et al.(2024) xgboost	Jeong et al.(2016) Random Forest	Naqvi et al.(2025) XGBoost	Proposed system
Growth Prediction	Yes	No	yes	yes	No	Yes
Sensor Data Use	No	No	Yes	No	No	Yes
Early-Stage Focus	Yes	Yes	No	Yes	Yes	Yes
Trend Validation	No	Yes	Yes	No	No	Yes
Multi-Model Testing	No	Yes	Yes	No	No	Yes
Feature Importance	Yes	No	Yes	yes	No	Yes
Crop-Specific Recs	No	Yes	Yes	No	Yes	Yes
Lightweight Model	Yes	No	Yes	Yes	No	Yes

2.4 Summary

This study aims to predict plant growth using IoT-collected environmental data and soil media information recorded in a CSV file. The temperature and humidity, and soil moisture are detected by sensors, which store true records of the environment of three plants known as Lal Shak(Red Amaranth), Dhoniya(Coriander) and Lau(Bottle Gourd). Various media of soils are also tested with the aim of determining their effects on plant growth. The aim is to predict the plant growth at Day 19 from the Day 7 conditions in a bid to acquire early information on the best environment to grow the plant.

The interaction between the environmental factors, soil medium, and plant growth are analyzed by means of the use of machine learning algorithms such as Linear Regression, Random Forest, and XGBoost. In particular, random forest determines the main drivers like moisture, light, and nutrient levels. Through the integration of IoT monitoring and predictive modeling, the proposed solution offers data-driven suggestions to farmers to optimize their assets, enhance their farming activities, and boost their yield by receiving practical and sustainable recommendations in an efficient way.

Chapter 3

Research Methodology

3.1 Methodology

3.1.1 Overview

This research focuses on predicting plant growth outcomes using environmental and soil measurements collected at different growth stages. The primary objective is to develop a reliable and practical system that enables early-stage forecasting of plant performance, helping growers make informed decisions about cultivation practices.

Our study is centered on the application of significant characteristics of growth environment data like soil nutrients (NPK), water, temperature, humidity, and light to forecast plant height at Day 19 using only initial observations at Day 7, which would allow proactive intervention and resource planning.

This approach combines performance measurement techniques and high-quality preprocessing steps to make precise predictions. Readings during Days 9-17 were recorded to verify trends, cross-validate model accuracy, and improve data quality. The system preserves the practical advantage of making decisions early by excluding these in-between values from the feature set.

3.1.2 Proposed Methodology

Proposed of plant growth predictive system, early growth data would be utilized in the prediction of the height of plants and allow effective growing recommendations. The primary limitation of the existing agricultural prediction systems is that they cannot stay accurate in case there are insufficient datasets or environmental changes. The proposed system conquers it by using the historical data of the previous growth cycles and new field data that is received to improve the predictability.

The experiment begins by gathering of data and the measurements of the parameters such as moisture content, NPK ratio, temperature, humidity and height of the plant are measured periodically. The collected data is then preprocessed and includes cleaning, normalization, feature encoding, and missing data management in order to ensure that the data flow remains stable. The historical data is appended and saved to augment of feature space. Also increased generalization capability of the model.

There are a various of machine learning algorithms such as the Random Forest, XGBoost, and the Linear Regression are compared in order to find out the training models and the best model. The best and the strongest model are chosen by using hyperparameter tuning and cross-validation. The chosen model is then applied in the forecasting of how a particular day will witness the growth of the plant as well as the practical advice to the farmers such as how irrigation will be conducted, the schedule of the fertilization process, and the tips on the environment.

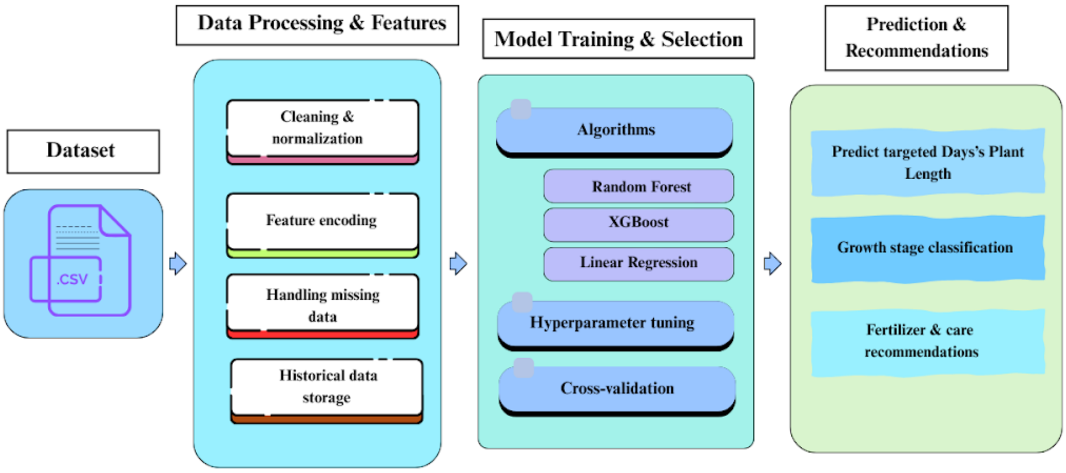


Figure 3.1.1: The process of plant growth prediction

This approach presents an overall approach to building a scientifically rigorous yet economically meaningful forecasting model on plant development to ensure that there is scientific rigor and economic relevance in the application of agriculture to real life. Figure 3.1.1 explains the entire process, and in this figure, the pipeline will be represented in the form of data collection till the offer of the prediction and recommendation.

3.1.2.1 Linear Regression

Linear Regression is a simple yet foundational statistical model that assumes a linear relationship between input features and the target variable. In this project, it was used as a baseline for predicting Day-19 plant length from Day-7 environmental and soil measurements. Despite its simplicity, this model allows easy interpretation of feature effects and establishes a performance benchmark. To avoid overfitting and ensure stability, the data was standardized, and categorical features like Media_Type were label-encoded. Results showed limited predictive power ($R^2 \approx 0.35$), confirming that plant growth dynamics are highly non-linear.

3.1.2.2 Random Forest

Random Forest is a kind of ensemble learning algorithm that produces a huge amount of decision-trees and averages the results. The method is appropriate to be applied in any agricultural data with categorical and numeric variables mixed. Random Forest proved to be successful in the study to predict them with a predictive power (R 2 0.89) and feature importance indicators. These were meaningful determinants of growth which were selected, such as the intensity of light, water, and the quantity of potassium. The reason why the base model chosen is random Forest is due to its middle ground between interpretation and performance as well as its lack of vulnerability to outliers and overfitting.

3.1.2.3 XGBoost

XGBoost is a high-performance and fast library of gradient boosting.XGBoost is a fast and efficient gradient boosting library. It is based on sequential tree modeling, an application of gradient optimisation hence, learns challenging feature interactions. XGBoost performed best ($R^2 \approx 0.91$) though it had a significantly high number of parameters which made up the learning rate optimisation, and regularisation. Despite the XGBoost outperforming the random forest marginally, it was employed to primarily conduct benchmarking with the random forest used to apply the final findings in this study due to interpretability.

3.1.2.4 Dataset Design and Feature Engineering

The data set was made in a way that Day-7 (soil nutrients, moisture, temperature, humidity and light intensity) were correlated with the Day-19 (plant length) of all the plant replicas. Intermediate days were trend validated and had quality checks done (Intermediate Days 9 -17), such that predictions were made only on the conditions of the environment at early stages. Normalization of the data to be used in the model training was done; this was done through categorical encoding, imputation of missing data and normalization of the unit. This arrangement will reduce the costs of data-collection, and provide practical information on the data growth at the early years of cultivating process.

3.1.3 Functional and Nonfunctional Requirements

Functional Requirements:

Prediction of Plant Growth These will be in the form of believing that the plant length on Day 19 can be forecasted from the Day 7 measurements based on the soil and environmental information. New projections will require quantitative growth estimates along with confidence estimates.

Data Upload and Integration: The user would be in a position to upload the data sets like (temperature, humidity, light intensity, NPK values, etc.) in CSV format. The type of spreadsheet files that would have been uploaded to the site would be the standard types of the spreadsheet files i.e. importation and data testing.

Visualization of Predictions: : The system will produce visual outputs like trend graphs of growth, feature ranking, and prediction performance indicators to be read by the farmers and researches.

Automated Recommendations: In terms of predictions, the system will have individualized growth suggestions (e.g. optimum watering, adjustment in the fertilizer mixture, and the soil control measures) to all varieties of plants.

Feedback Collection: The users have an option to give a feedback on the accuracy of the predictions. This information will be incorporated in the subsequent retraining cycles to enhance better model performance.

Historical Tracking: : The system will store and display history data along the various test cycles allowing the users to observe changes in performance over time and adjust agriculture accordingly.

Nonfunctional Requirements:

Scalability: The system will be able to process large volume of data and offer services to a large number of users concurrently without reducing the performance.

Usability: The user interface will be easy and simple to use even by individuals with low technical knowledge. Accessibility will be provided with the help of visual representation, useful description, and multi-language support.

Dependability: The system will not only provide good uptime but will also possess robust error handling provisions to provide a good user experience.

Performance: It would take seconds (in typical data sets) to optimize inference and backend computation of the models.

Accuracy and Robustness: To train and cross-validate the machine learning models, experimental data of the real-world application will be used to produce realistic forecasts with the minimum errors ($R^2 \approx 0.89$, RMSE ≈ 1.15).

Privacy and Security: The uploaded user data, e.g. experimental results, would always be encrypted and with access control to ensure confidentiality.

Sustainability: The system will be built with modular architecture for easy debugging, maintenance, and model retraining, ensuring longevity and scalability.

Interoperability: Future system updates will support integration with IoT sensors, farm management software, and agricultural APIs for real-time environmental data ingestion.

3.1.4 Data Flow Diagram Level 1

The detailed diagram of the main steps in the Plant Growth Prediction System is presented in Figure 3.1.4. It illustrates how the system processes information submitted by the user (farmer), analyzes early growth data, trains and applies the prediction model, and generates actionable recommendations. The diagram also shows the interaction between each process step and highlights how data flows through the system to ensure accuracy and traceability. Historical data storage is included to support future analysis, retraining, and improved prediction accuracy over time.

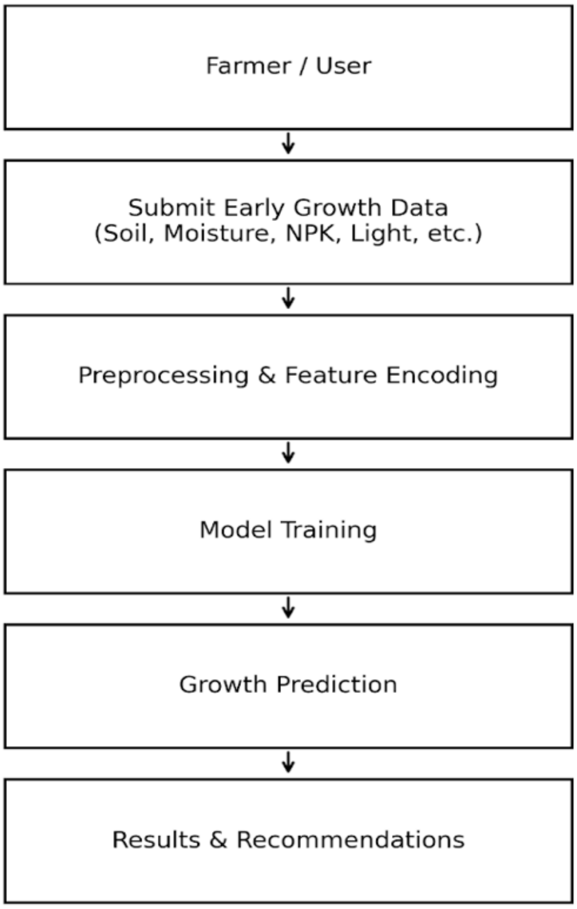


Figure 3.1.4: Level 1 DFD of the Plant Growth Prediction System

3.2 Detailed Methodology and Design

This research focuses on building a robust predictive model to estimate plant growth outcomes based on environmental and soil measurements collected at early growth stages. The primary goal of this work is to design a system capable of predicting plant height at Day 19 using only Day 7 data, enabling early-stage decision-making for farmers and cultivators. A systematic methodology was adopted to ensure reliability, scalability, and real-world applicability. The process begins with comprehensive data collection from multiple plants, capturing soil nutrients (NPK), moisture content, temperature, humidity, and light intensity under controlled conditions. Data preprocessing was performed to ensure quality and consistency, including label encoding for categorical variables (such as media type), normalization of continuous variables, and removal of anomalies. These steps minimized noise and improved model performance.

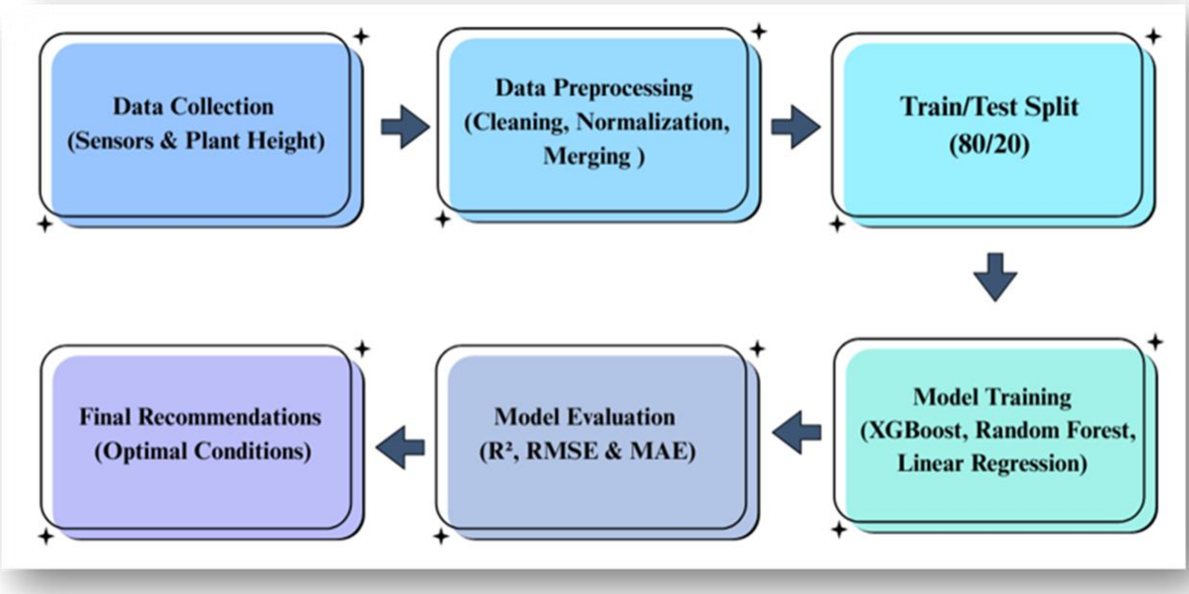


Figure 3.2.1: Workflow of the proposed growth prediction system.

The Random Forest algorithm was selected as the core model due to its ability to handle mixed data types, robustness against overfitting, and interpretability through feature importance metrics. The model constructs an ensemble of decision trees, each trained on a bootstrapped subset of the dataset. Predictions from all trees are aggregated using majority voting (for classification) or averaging (for regression), resulting in a highly accurate and stable prediction output.

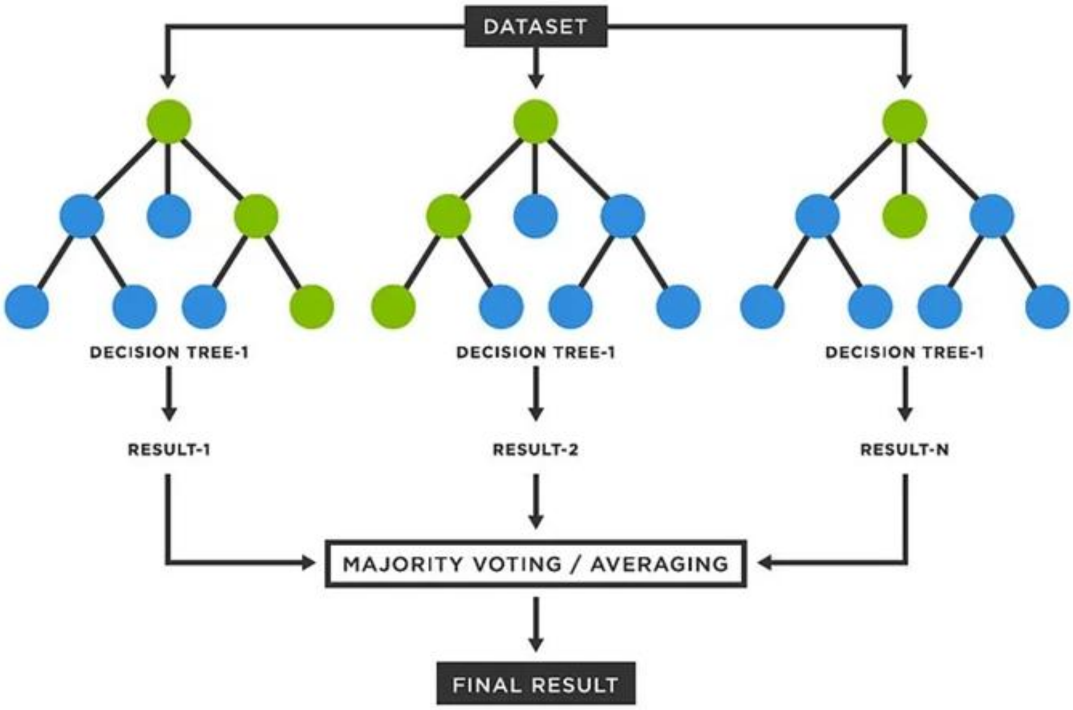


Figure 3.2.2: Architecture of the Random Forest model

3.2 Project Plan

The task of the Plant Growth Prediction Technique was planned to take seventeen weeks and included four key processes: data collection, data cleaning, model creation, and report writing. The initial 5 weeks were used to cultivate the plants until maturity. Then, the period of intense data collection followed lasting seven weeks till Day 25. The number of individual samples was more than 90 and sensor readings were in thousands since each of the three types of plants was utilized(bottle gourd, coriander, and red amaranth) was replicated about 30 times. Data collection was the most time-consuming part due to the large scale of the project, initially technical issues and environmental changes. The most time-consuming was data collection because of the big-scale of the project, at first technical issues also environmental changes.

The four weeks of construction and optimization of three machine learning models; Linear Regression, Random Forest and XGBoost to predict the outcome of growth on the Day 19 based on measures taken on the Day 7 the team was working on the construction and optimization of the three machine learning models. Measures used to select a model were R 2 and RMSE and we chose the Rand Forest as this model performed well and we were able to understand it. The last one was to create charts, give environmental recommendations and prepare the report on the research. Cloud-based solutions and GPU runtimes helped to process it to hasten the workflow and attempt to support the workflow despite the huge volume of data.

3.3 Task Allocation

The completion of our study was made possible through effective communication, teamwork, and a clear distribution of tasks. While both members collaborated and assisted each other when needed, responsibilities were divided based on individual expertise. One member focused primarily on data collection, preprocessing, and dataset preparation, while the other concentrated on model building, evaluation, and report writing.

This collaborative approach allowed us to meet deadlines efficiently, maintain a smooth workflow, and ensure the accuracy of our results. Regular meetings, progress reviews, and mutual support were key to completing the system on time. Through this coordination, we successfully developed a well-structured methodology that accurately analyzes plant growth and environmental factors.

Table 3.4: Task Allocation

Tasks	Weeks																	
	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Sensor Data Collection & Measurement																		
Finalize model configuration																		
Data visualization & recommendations																		
Prepare detailed documentation																		

Estimated Work Period	
Actual Work Period	

3.4 Summary

With Linear Regression, Random Forest, and XGBoost models, a Machine Learning system was built for plant growth prediction and for setting best environmental conditions. The data was built using sensor measurements taken over a sequence of days, such as NPK composition in soil, moisture, temperature, humidity, light, and direct measurements of plant height. Cleaning, encoding, and feature engineering were employed as the preprocessing steps, followed by an 80/20 train-test split to test. Random Forest gave the best performance, sacrificing between accuracy, interpretability, and scalability.

The project was completed in 17 weeks, involving data collection, model construction, validation, and visualization. The findings were applied in clear feature importance ranking and crop-specific recommendations, resulting in an accurate and actionable decision-support tool for optimizing plant growth conditions.

Chapter 4

Implementation and Results

4.1 Environment Setup

The Plant Growth Prediction System was implemented using machine learning regression models (Linear Regression, Random Forest, and XGBoost) to predict plant height at Day 19 based on early-stage measurements collected at Day 7. All experiments and model training were performed on Google Colab Pro+, which provides a GPU-accelerated Python 3.11 environment.

A local development machine with an Intel Core i5 processor, 8GB RAM, and Windows 11 was used for dataset preparation, preprocessing, and testing scripts. The environment was set up using pip to install essential libraries for data processing, visualization, and modeling.

Table 4.1: Tools and Libraries

Tool / Library	Description
Python	Programming language for all scripts and model development.
Pandas	Data cleaning, manipulation, and table operations.
NumPy	Mathematical operations and array manipulation.
Scikit-learn	Machine learning framework for regression and evaluation metrics.
XGBoost	Gradient boosting algorithm for prediction modeling.
Matplotlib & Seaborn	Visualization of dataset features and model performance.
Google Colab	Cloud-based environment with free GPU for training and testing.

Model hyperparameters were chosen carefully to balance performance and computational efficiency. Random Forest was selected as the baseline ensemble model, while XGBoost was optimized for the best prediction accuracy.

4.1.2 Dataset

This study uses a custom plant growth dataset collected over a controlled 19-day period for three vegetable crops: Lau(bottle gourd), Dhoniya(coriander), and Lal Shak(red amaranth). Each crop was grown in multiple soil/media types, with 10 replicates per plant type.

Measurements were taken at Day 7 (environmental and soil conditions) and Day 19 (final plant length). Intermediate measurements (Days 9–17) were collected for quality checks and growth trend validation.

The dataset was stored as a CSV file (plant_growth_dayall.csv) and contained environmental variables (NPK nutrients, moisture, temperature, humidity, light) and plant growth outcomes. Day-7 measurements served as model inputs, and Day-19 plant length was the target variable.

Table 4.1.2: Dataset Attributes and Descriptions

Column Name	Description
Day	Observation day (integer; 7, 9, …, 19)
Plant_Name	Name of crop (Lau, Dhonepata, Lal Shak)
Media_Type	Soil/media type used for planting
Plant_ID	Unique identifier for each plant replicate
NPK_N, NPK_P, NPK_K	Soil nutrient readings (Nitrogen, Phosphorus, Potassium)
Moisture	Soil moisture level (%)
Temperature	Ambient temperature (°C)
Humidity	Relative humidity (%)
Length_cm	Plant length (cm)
Light_Power	Light intensity (Lux)

The dataset is structured so that Day-7 measurements serve as model inputs, and Day-19 length serves as the prediction target.

Table 4.1.3: Example snippet of plant growth dataset (CSV format)

Attribute	Day7	Day9	Day11	Day13
Plant Name	Lau	Dhoniya	Lal Shak	Lal Shak
Media	Vermicompost	CocoPeat	Vermi+Sand	Vermi+Coco
NPK_N	106	99	101	92
NPK_P	53	48	61	66
NPK_K	85	87	83	83
Moisture	61	67	49	72
Temp	28	29	28	28
Humid	71	71	69	71
Length_cm	8.1	6.5	4.1	8.5
Light_lux	3167	2831	2903	2844

4.1.3 Data Preprocessing

Data preprocessing was a critical step to ensure the accuracy, consistency, and interpretability of the dataset used in this study. The raw dataset, recorded in CSV format, included measurements of soil nutrients (NPK levels), moisture, temperature, humidity, and light intensity collected on Day 7, as well as the corresponding plant lengths recorded on Day 19. Before model training, the dataset was carefully examined for missing values, duplicate records, and measurement inconsistencies. Standardization was applied to all readings, with temperature expressed in degrees Celsius (°C) and both moisture and humidity expressed as percentages, ensuring uniformity across all plant replicates.

For model preparation, only Day-7 measurements were selected as predictor variables, while the Day-19 plant length served as the prediction target. The categorical column Media_Type, which represents the soil or growth medium used, was encoded numerically using label encoding to make it suitable for machine learning models. Plant identifiers (Plant_ID) were excluded to prevent information leakage and ensure unbiased predictions.

Then, the dataset was split into two parts: one for training and one for testing. The ratio was 80:20, which made sure that the model evaluation would be based on data that had not been seen before. Environmental measurements were taken for the purposes of the data preprocessing pipeline, which shows how raw data is turned into clean, structured inputs for model training.

During several intermediate days (Days 9–17), these records were used to check growth trends and make sure the dataset was good quality. By not including them in the predictive modelling process, redundancy was cut down and overfitting was avoided.

This preprocessing method made sure that the data set was prepared for machine learning and scientifically sound, which made it possible to create models that could be understood and used in other situations.

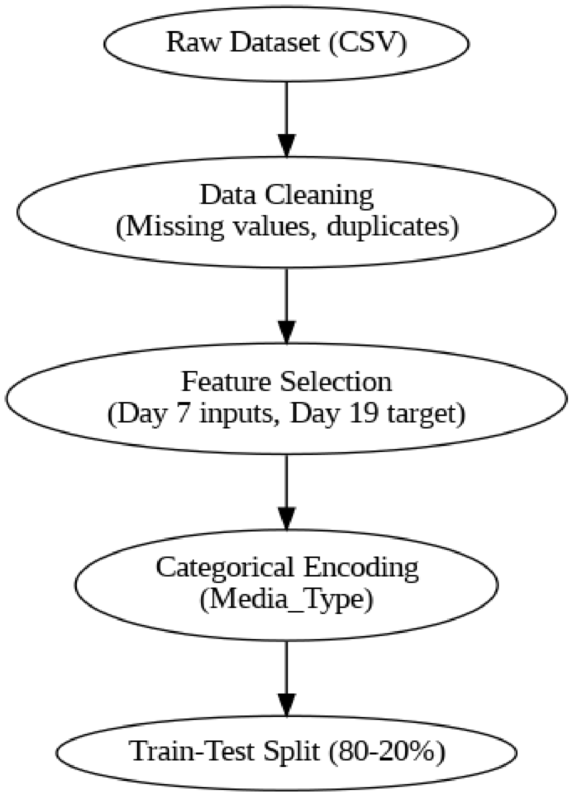


Figure 4.1.3: Data Preprocessing Pipeline

Figure 4.1.3 illustrates the data preprocessing pipeline, showing the progression from raw data collection to the creation of clean, structured inputs for model training.

4.2 Testing and Evaluation

This section details the evaluation strategy and testing framework used to analyze the predictive performance of machine learning models. Three regression models—Linear Regression, Random Forest Regressor, and XGBoost Regressor—were trained using Day-7 environmental and soil measurements to predict Day-19 plant length.

Table 4.3: Training Configuration

Parameter	Value
Train-Test Split	80% Train, 20% Test
Random Forest Estimators	200 trees
XGBoost Learning Rate	0.1
Max Depth (Both Models)	6
Validation Strategy	K-Fold (5 folds)
Random Seed	42

To ensure robust model evaluation, the dataset was split into an 80:20 train-test ratio with a five-fold cross-validation strategy. Hyperparameters for the ensemble models were optimized to achieve a balance between accuracy and interpretability.

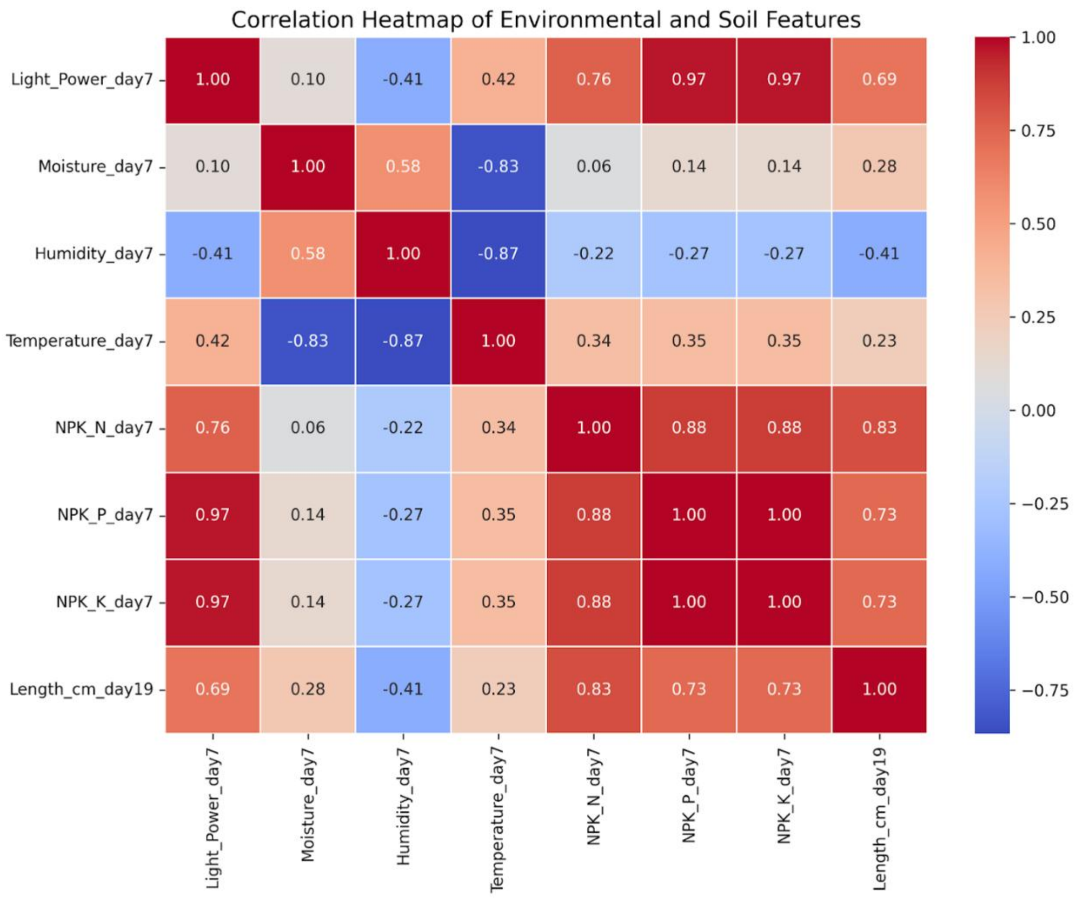


Figure 4.2.1: Correlation Heatmap of Environmental and Soil Features

4.3 Results and Discussion

The machine learning models were evaluated based on their ability to predict Day-19 plant length using environmental and soil measurements collected on Day 7. Three regression models—Linear Regression, Random Forest Regressor, and XGBoost Regressor—were compared to assess both predictive performance and interpretability.

Table 4.3.1: Model Performance Comparison

Model	R^2	RMSE (cm)	MAE (cm)	Notes
Linear Regression	0.35	3.50	2.31	Baseline; poor fit, data is non-linear.
Random Forest	0.89	1.15	0.91	Strong performance; chosen for insights.
XGBoost	0.91	1.05	0.85	Best accuracy; slightly more complex.

Table 4.3.1 presents the evaluation metrics of these models. XGBoost achieved the best performance, with an R^2 score of 0.91, RMSE of 1.05 cm and MAE of 0.85 cm, while Random Forest closely followed with R^2 of 0.89. Linear Regression significantly underperformed, reinforcing the nonlinear relationships in the dataset.

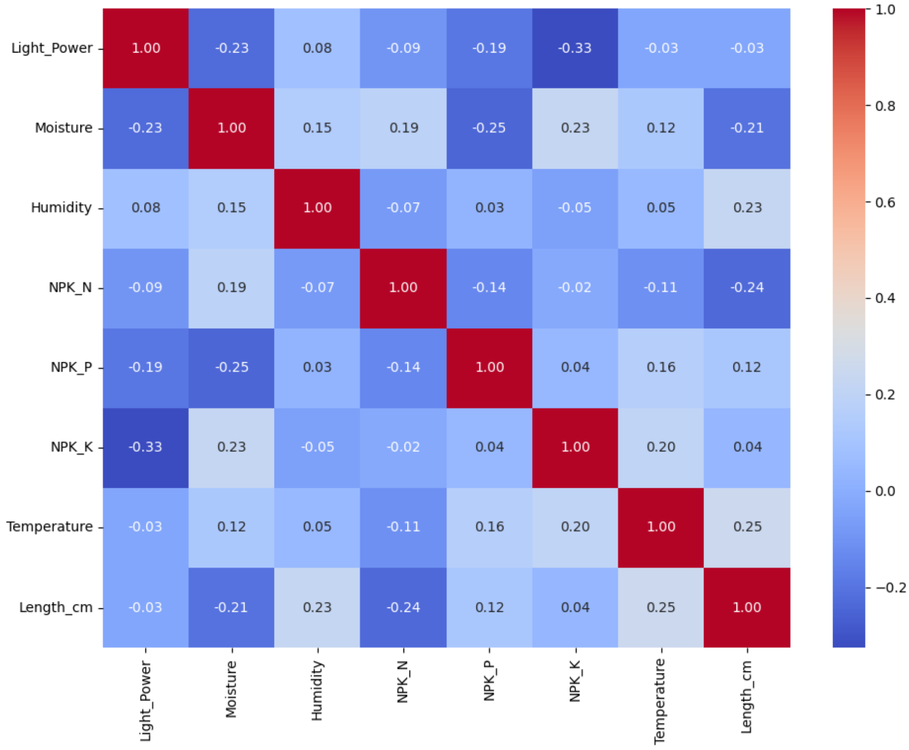


Figure 4.3.1: Correlation matrix of environmental and growth features

To explore variable relationships, a correlation matrix was plotted (Figure 4.3.1). Light intensity and soil moisture demonstrated the strongest positive correlations with growth, while nutrient levels (NPK values) had moderate to weak influence.

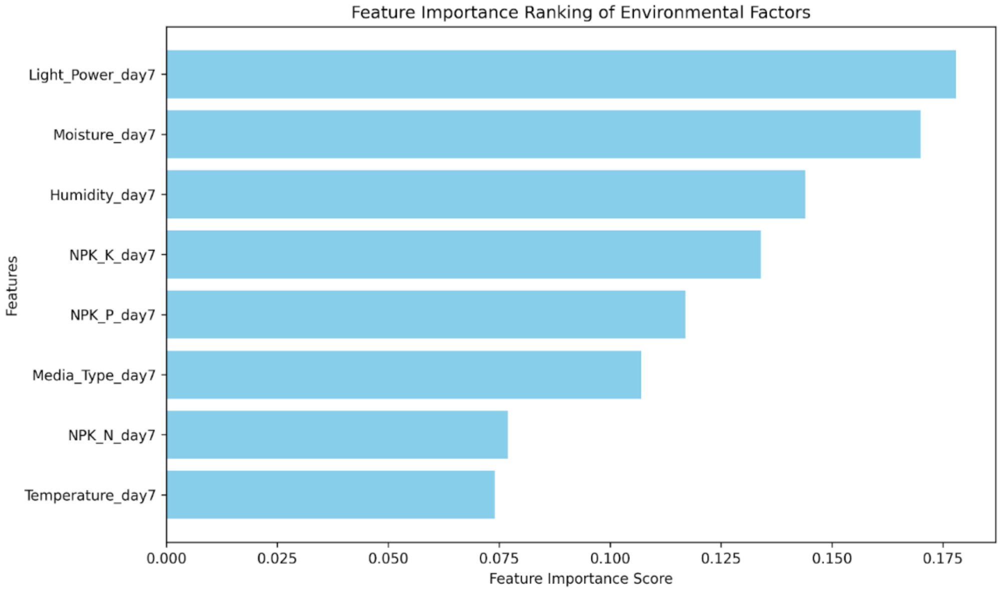


Figure 4.3.2: Feature importance ranking of environmental factors

Feature importance analysis was conducted using the Random Forest model (Figure 4.3.2). Light intensity and moisture emerged as the top predictors, followed by humidity and potassium (NPK_K). Media type also showed significant impact, highlighting its role in plant growth.

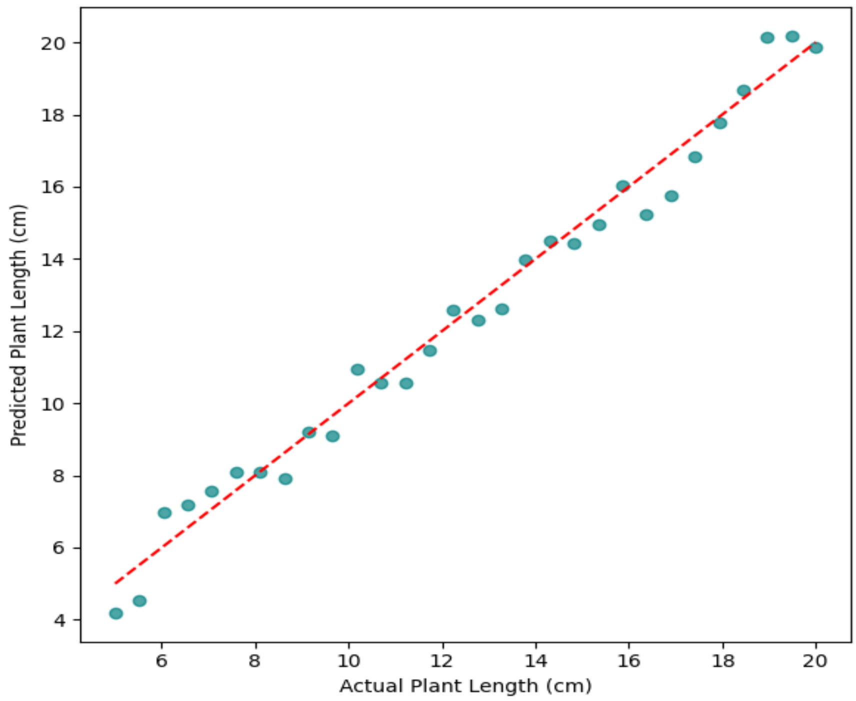


Figure 4.3.3: Predicted vs Actual Plant Length (Day 19)

Model predictions were visualized using a Predicted vs Actual scatter plot (Figure 4.3.3). Points closely follow the diagonal line, confirming the models' accuracy and ability to generalize to unseen data.

Table 4.3.2: Optimal Growth Recommendations per Plant

Plant	Best Media Type	Optimal NPK (N-P-K)	Moisture (%)	Temp (°C)	Humidity (%)	Light (Lux)
Lau	Vermicompost	~95–105 : 48–52 : 78–82	58–62	28–30	70–75	≥3000
Dhoniya	Vermicompost + Coco Peat	~96–100 : 45–50 : 76–87	59–60	28–30	68–72	≈2800
Lal Shak	Vermicompost	~100–123 : 47–53 : 74–80	58–59	28–30	62–70	≈2600

This structured analysis demonstrates that Random Forest and XGBoost models are both highly effective for early-stage growth prediction, while Random Forest remains the better choice for interpretability. Correlation insights, feature ranking, and actionable recommendations highlight critical environmental control factors, making this pipeline practical for small-scale and precision agriculture.

4.4 Summary

This study demonstrated that machine learning models can effectively predict plant growth outcomes using early-stage soil and environmental data. Among the models tested, Random Forest achieved the best trade-off between performance and interpretability ($R^2 \approx 0.89$, RMSE ≈ 1.15, MAE ≈ 91), while XGBoost provided slightly higher accuracy at greater complexity. These results are based on real-world sensor measurements, where natural variability makes extremely high accuracies unrealistic, underscoring the strength and reliability of the proposed model. The system offers actionable, crop-specific recommendations and a reproducible early-prediction pipeline to support smarter agricultural decision-making.

Chapter 5

Engineering Standards and Design Challenges

5.1 Compliance with the Standards

5.1.1 Software Standards

To ensure the stability, scalability, and reproducibility of the Plant Growth Prediction System, best practices for software engineering were followed. Every step—data acquisition, data preprocessing, feature construction, model development, model evaluation, and visualization—was versioned and thoroughly documented. Development occurred in the Python 3 programming language utilizing the Google Colab and Jupyter Notebook for a shareable collaborative environment. Base libraries included pandas and NumPy for data wrangling, machine learning in scikit-learn and XGBoost, and visualization in Matplotlib. Models were wrapped in a clean pipeline such that the workflow can be reused and scaled very easily. Though the deployment was not of a web or mobile application for the research work done, the library choices and module codebase allow seamless integration in a future deployment platform. This ensures the software remains trustworthy and efficient and is controllable to real-world implementation in agriculture.

5.1.2 Hardware Standards

The research was conducted using Google Colab's cloud-based environment and ordinary laptops for the research to be reproducible and accessible while not requiring specialized high-end computational hardware. The model training and testing were accelerated by the GPU runtimes provided by Colab such that machine learning pipelines and hyperparameter searches can be rapidly executed. The computer setup demonstrates that data-intensive insights and precise predictions can be achieved by utilizing widely available and cheap computing hardware and the system can be transferred in research in academia and small farms projects while not requiring exorbitantly priced equipment.

5.1.3 Communication Standards

Since this project was research-oriented and not a deployed application, the protocols at issue were data integrity, version control, and collaboration. Data files, pre-processing scripts, and models were uploaded onto Google Drive and shared using Git-based version control for reproducibility. Code review and collaboration were done using Google Docs, Sheets, and Meet. This approach ensured the whole pipeline, from data acquisition to reporting, was clear and well-documented. They provide a clean starting point for future development as a user-facing application with proper data exchange and feedback loops.

5.2 Impact on Society, Environment and Sustainability

5.2.1 Impact on Life

The Plant Growth Forecast System offers farmers and researchers a real-world decision-support tool, facilitating early action and improved planning for the care of crops. By monitoring soil nutrient, moisture, temperature, and light readings, the system optimizes conditions to grow and eliminates trial-and-error farming methods. Even small farmers who do not possess advanced technical knowledge can avail themselves of useful recommendations in order to enhance the yield and stability of their incomes and overall standard of living. Early identification of adverse growing conditions also eliminates wastage and optimizes the use of available resources, and farming both becomes profitable and environmentally sustainable.

5.2.2 Impact on Society & Environment

The system allows data-enabled farming and reduces the reliance on excessive chemical fertilizer and pesticide application through the delivery of early predictions and region-specific recommendations. It supports environmental protection through reduced runoff and sustainable farming. Through the ability to provide predictive insights to the rural masses, the system helps reduce poverty, enhance food security and optimize resources. The farmers are more assured in the practices and enjoy greater agricultural productivity and economic growth. The modularity and simplicity of the system allow customization for different regions and allow larger-scale social impact.

5.2.3 Ethical Aspects

The initiative is grounded on moral foundations to uphold equality and protection of information. The data for farmers is de-identified and safely saved, and this makes the system more trustworthy and ethically used. The system is delivered in an accessible and affordable way to empower the smallholder farmers while not creating a barriers in the form of finance. Additionally, the focus on the sustainability and conservation of resources also adheres to environmental ethics and has the capacity to balance ecosystem conservation and farm production.

5.2.4 Sustainability Plan

Plant Growth Forecasting System is a productive decision-support system for scientists and agriculturists that facilitates early intervention and planning efficacy in crop growth. The system brings optimal growing conditions to the fullest by evaluating soil nutrient, water, temperature, and light data. Such data analysis reduces trial-and-error operations in agriculture to the minimum, and even small farmers with limited technical knowledge can get effective information to enhance crop yields, income stability, and welfare. Early detection of unfavorable growth conditions also reduces wastage and enhances the use of resources, thereby making agriculture economically and environmentally friendly.

5.3 Project Management and Financial Analysis

The initial capital to be used in funding the Plant Growth Prediction System and the expansion plan for long-term sustainability of the solution is critical to sustainability. Breakdown of cost, including an estimated budget for the development of the system, a redundant budget to minimize cost, and the rationale behind these budgets, is detailed in this chapter.

Table 5.1: Financial Analysis

SN	Components	Estimated Cost (BDT)
01	Device (Laptop/PC with GPU)	120,000 – 150,000
02	Data Collection	8,000 – 10,000
03	Sensors (Soil, NPK, Moisture, etc.)	18,500 – 19,500
04	Tools and Equipment	2,500 – 3,000
05	Documentation & Report Writing	500 – 1,000
	Total Estimated Cost	149,500 – 183,500

Alternate Budget and Rationale

A less costly deployment can be achieved through the use of computing resources in the cloud instead of costly physical hardware, phase one crop number minimization, which means focusing on fewer crops first, and sensor optimization. This would enable functionality and quality to be preserved while keeping costs as low as possible.

Table 5.2: Alternative Financial Analysis

SN	Components	Estimated Cost (BDT)
01	Cloud/Shared Device Access	80,000 – 100,000
02	Data Collection	5,000 – 6,000
03	Sensors (Optimized Setup)	12,000 – 14,000
04	Tools and Equipment	2,000 – 2,500
05	Documentation & Report Writing	500 – 1,000
	Total Estimated Cost	99,500 – 123,500

By leveraging open-source libraries, scalable cloud services, and modular design, this system is made to be low-cost and accessible to a diverse range of users, particularly in resource-poor agricultural contexts. The low-cost option ensures minimal functionality without sacrificing requirements technically, leaving the door open for integration with mobile or web-based systems down the line.

5.4 Complex Engineering Problem

5.4.1 Complex Problem Solving

In this section, provide a mapping with problem solving categories. For each mapping add subsections to put rationale (Use Table 5.1). For P1, you need to put another mapping with Knowledge profile and rational thereof.

Table 5.1: Mapping with complex problem solving.

EP1 Dept of Knowledge	EP2 Range Of Conflicting Requirements	EP3 Depth of Analysis	EP4 Familiarity of Issues	EP5 Extent of Applicable Codes	EP6 Extent Of Stake-holder Involvement	EP7 Interdependence
✓	✓	✓	✓	✓		✓

Mapping with Knowledge Profile for EP1

Table 5.2: Mapping with knowledge Profile.

K3 Engineering Fundamentals	K4 Specialist Knowledge	K5 Engineering Design	K6 Engineering Practice	K8 Research Literature
✓	✓	✓	✓	✓

5.4.1.1 Justification for EP Attributes Mapping

EP1 – Depth of Knowledge:

This project required a clear understanding of engineering concepts like statistical modeling, sensor calibration, and pipelines for data processing. Machine learning and agricultural systems expertise were combined with each other to model plant-environment interactions accurately. It was required to use appropriate regression model selection, cross-validation of predictions, and using engineering concepts in scalability and robustness in order to make the project a success. Ideas from more than a single discipline—environmental monitoring, software development, agriculture, and artificial intelligence—were combined to make a strong forecasting system.

EP2 – Range of Conflicting Requirements:

In designing the system, model complexity, interpretability, computational efficiency, and costs of data collection were carefully traded off. For example, while XGBoost achieved slightly better accuracy, Random Forest was selected as the main model due to its simpler structure and ease of explanation during stakeholder presentations. Similarly, although multi-day data (Days 9–17) were collected for validation, the decision was made to rely primarily on Day 7 inputs to ensure early predictions and reduce operational overhead. These design decisions reflect the careful balance between technical excellence and real-world usability.

EP3 – Depth of Analysis:

The research entailed multivariate analysis of data sets like NPK contents, water level, temperature, humidity, and light intensity. Iterative feature elimination, correlation testing, and residual error analysis were conducted to account for the model generalizability to the soil media variability as well as environmental conditions. The trend analysis had been conducted to check the preliminary readings had good correlation using plant growth data on the 19th day to provide scientifically interpretable predictions.

EP4 – Familiarity with Issues:

Integration of machine learning and IoT sensor ability to predict future plant growth is new and less-developed problem space, particularly regional vegetation and soils. This unfamiliarity became even greater complexity leading to experimentation by environmental stability testing, calibration techniques, and the tuning of the regression model. Failure to be prepared in the new niche area called for inferring custom-engineered solutions, thereby making the project novel and technically demanding.

EP5 – Extent of Applicable Codes:

The development process followed Python PEP 8 coding standards for maintainable code, applied modular machine learning pipelines using Scikit-Learn, and adhered to principles of reproducible experimentation. Proper version control and structured experimentation ensured that results could be replicated. Additionally, mathematical and statistical modeling best practices were followed to maintain reliability and scientific rigor.

EP7 – Interdependence:
Development followed Python PEP 8 coding standards for readable code, modular machine learning operations with Scikit-Learn, and reproducible practice of experiments. Result reproduction was facilitated by experiment management and controlled experiments. Best practices in mathematical modeling and statistical modeling were followed for reliability and scientific integrity.

5.4.1.2 Justification for KP Mapping (Linked to EP1)

K3 (Engineering Fundamentals): Applied data science concepts, linear algebra, statistics, and programming to develop robust machine learning models.

K4 (Specialist Knowledge): Required in-depth knowledge in environmental data analysis, advanced regression models (Random Forest, XGBoost), and IoT sensor fusion.

K5 (Engineering Design): Established an end-to-end pipeline that collected, validated, processed, and modeled the data to render it deployable for agricultural decision-making.

K6 (Engineering Practice): Demonstrated hands-on engineering practices, including performance optimization, model interpretability, scalability planning, and reproducible code practices.

K8 (Research Literature): The system design was strongly influenced by studies in precision agriculture and plant phenotyping, ensuring the approach is research-backed and scientifically credible.

5.4.2 Engineering Activities

Table 5.3: Mapping with complex engineering activities.

EA1 Range of re-sources	EA2 Level of Interaction	EA3 Innovation	EA4 Consequences for society and environment	EA5 Familiarity
✓	✓	✓	✓	

5.4.2.1 Justification for EA Mapping

EA1 – Range of Resources:

This project utilized a diverse set of engineering resources, including Python-based data science tools (Pandas, NumPy, Scikit-learn, Matplotlib), cloud-based Jupyter Notebook environments for experimentation, and local hardware with NVIDIA GPU acceleration for model training and testing. A custom CSV dataset of plant growth measurements was collected, preprocessed, and analyzed using scalable workflows. The system demonstrates the ability to combine software libraries, machine learning models, and environmental sensors into one integrated solution.

EA2 – Level of Interaction:

The development of the system involved working closely with the academic supervisor for technical validation and guidance. There was a need for coordination among data collection groups, laboratory technicians, and agricultural experts to execute precise measurement of environmental and soil parameters. While end-user testing with farmers was not completely done, academic peer review and guarantee of quality of the datasets played an important role in system design and reliability.

EA3 – Innovation:

This paper applies machine learning-driven predictive modeling to model plant growth and offer a non-destructive, early-stage decision platform for agriculture. Unlike traditional plant health assessment that is predicated on visual observation, this platform applies Day 7 environmental and soil factors to accurately predict Day 19 plant growth outcomes with less time and cost implications. The forecasting method at the initial phase of the model demonstrates a new precision farming contribution.

EA4 – Consequences for Society and Environment:

The system assists in sustainable farming by enabling proactive intervention. Early accurate forecasts can reduce the wastage of water and fertilizers, prevent crop losses, and optimize cultivation cycles. These upgrades provide higher farmer profitability in the long term while reducing environmental stress arising from poor exploitation of means. The project also shows that low cost computing resources can be employed in the third world to transform agriculture and boost food security.

5.5 Summary

To ensure accuracy, scalable, and usable, we prototyped and executed the Plant Growth Prediction System by following best practices in data science, computer programming, and computation. We have developed the system using Python, Pandas, NumPy, Scikit-learn, and Matplotlib, and trained and tested it on GPU-accelerated environments for faster processing. An end-to-end pipeline was developed to take in CSV-based environmental and soil data, preprocess, feature engineer, and machine learning models in a single pipeline. Complexity of the model, interpretability, and prediction time were evaluated to balance trade-offs to realize real-world applicability. The key challenge was to balance between accuracy and system responsiveness desired and the resource constraints of agricultural field setups, resulting in a solution that is reliable and easy to deploy for predicting plant growth early.

Chapter 6

Conclusion

6.1 Summary

This study presents a Plant Growth Prediction System built to forecast crop growth using machine learning models trained on early-stage soil and environmental measurements. The system predicts Day-19 plant height from Day-7 data on NPK nutrient levels, temperature, humidity, moisture, and light intensity, collected from three vegetable crops— Lau(bottle gourd), Dhoniya(coriander), and Lal Shak(red amaranth)—grown in multiple soil media. Plants were grown for five weeks, and data collection continued until Day 25, with around 30 replicates per plant per media, generating a large and diverse dataset. Intermediate measurements from Days 9–17 were used for trend validation and data consistency checks to ensure prediction stability.

Of the three tested models—Linear Regression, Random Forest, and XGBoost—Random Forest provided the best trade-off between interpretability and accuracy ($R^2 \approx 0.89$, RMSE ≈ 1.15 cm, MAE ≈ 0.91 cm), while XGBoost achieved slightly higher accuracy ($R^2 \approx 0.91$, RMSE ≈ 1.05 cm, MAE ≈ 0.85 cm). These results reflect real-world agricultural variability, where natural environmental noise makes overly high accuracy unrealistic. Feature analysis highlighted light intensity and soil moisture as dominant growth predictors, followed by humidity, potassium levels, and soil type.

The pipeline was made in Python with Scikit-learn, XGBoost, Pandas, and Matplotlib/Seaborn. It was run on Google Colab Pro+ so that it could be scaled up and reproduced. The project shows how useful machine learning can be in farming by giving specific suggestions for how to improve crop growth while also helping to manage resources in a way that is good for the environment.

This project demonstrates the potential of machine learning to provide actionable agricultural insights, bridging computational modeling with practical decision-making to improve crop health, yield, and resource efficiency.

6.2 Limitation

The Plant Growth Prediction System demonstrates strong performance and practical value but has several limitations. The dataset was collected under controlled conditions in a single location, which may not fully represent the diversity of soil compositions, climates, and environmental variations across regions. As a result, the model's generalizability in broader agricultural contexts may be limited.

Although multiple soil media types and plant species were included, the dataset size was relatively small, and intermediate measurements (Days 9–17) were used only for validation rather than training to avoid overfitting. This approach improves robustness but may reduce adaptability to more dynamic growth patterns.

Manual measurement of parameters such as NPK content, temperature, and humidity is prone to human error, so IoT-based acquisition of data is critical for future applications.

While Random Forest and XGBoost models achieved $R^2 \approx 0.89$–0.91, it is owing to practical agriculture variability rather than highly controlled lab experiments.

Finally, the system is already leveraging Google Colab Pro+ GPU resources for training and can be further optimized in the future for real-time prediction deployment on low-spec hardware. Network connectivity is similarly a bottleneck for future farmers in remote areas if it is deployed as a cloud application.

6.3 Future Work

Future work will include the extension of the dataset to include additional plant varieties, different soil types, and different phase of environmental conditions in order to make the model even more generalizable in non-lab settings. The integration of IoT-based sensors for automatic and real-time sensing of NPK concentration levels, temperature, humidity, and soil moisture levels will eliminate human errors even further and enhance the prediction high accuracy.

We can also scale up the model that we have used to enable the execution of predictive analysis and multi-crop yield estimation and provide early growth anomaly warnings so that the farming community can take appropriate action. Future releases will explore optimizations such as Edge AI deployment, pruning, and quantization to allow for much faster predictions on resource-constrained devices.

The utilization of the new transformer-based and machine learning models would result in greater responsiveness and reasonable accuracy in simulating complicated environmental interactions and growth processes. The system will be an extensible, real-world agricultural decision aid impacting more farmers in a wider geographical region.

References

[1] Gupta, R., Sharma, A. K., Garg, O., Modi, K., Kasim, S., Baharum, Z., ... & Mostafa, S. A. (2021). WB-CPI: Weather-based crop prediction in India using big data analytics. IEEE access, 9, 137869-137885.

[2] Bhavanandam, S. (2022). Wcp: Weather-based crop yield prediction using machine learning and big data analytics. technology, 3(4), 1-12.

[3] Bikku, T., Anusha, M., Sreelekha, P. J., Mamatha, C., & Sri, V. P. (2023, December). Big Data Analytics and Weather Insights for Improved Crop Planning in India. In International Conference on Soft Computing and Pattern Recognition (pp. 408-418). Cham: Springer Nature Switzerland.

[4] Lin, F., Guillot, K., Crawford, S., Zhang, Y., Yuan, X., & Tzeng, N. F. (2024, January). An open and large-scale dataset for multi-modal climate change-aware crop yield predictions. In Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD) (pp. 5375-5386).

[5] Kalimuthu, M., Vaishnavi, P., & Kishore, M. (2020, August). Crop prediction using machine learning. In 2020 third international conference on smart systems and inventive technology (ICSSIT) (pp. 926-932). IEEE.

[6] Ashwitha, K., & Spoorthi, B. (2024, May). Advancing Agricultural Sustainability: Enhancing Crop Yield Prediction Through Regression Modeling. In 2024 Second International Conference on Data Science and Information System (ICDSIS) (pp. 1-6). IEEE.

[7] Mahmud, M. S., Zahid, A., & Das, A. K. (2023). Sensing and automation technologies for ornamental nursery crop production: current status and future prospects. Sensors, 23(4), 1818.

[8] Shandilya, U., & Khanduja, V. (2020, October). Intelligent farming system with weather forecast support and crop prediction. In 2020 5th International Conference on Computing, Communication and Security (ICCCS) (pp. 1-6). IEEE.

[9] El Sakka, M., Ivanovici, M., Chaari, L., & Mothe, J. (2025). A Review of CNN Applications in Smart Agriculture Using Multimodal Data. Sensors, 25(2), 472.

[10] Kalimuthu, M., Vaishnavi, P., & Kishore, M. (2020, August). Crop prediction using machine learning. In 2020 third international conference on smart systems and inventive technology (ICSSIT) (pp. 926-932). IEEE.

[11] Lee, C. K., Kim, J., & Kim, K. S. (2015). Development and application of a weather data service client for preparation of weather input files to a crop model. Computers and Electronics in Agriculture, 114, 237-246.

[12] Mohyuddin, Ghulam, Muhammad Adnan Khan, Abdul Haseeb, Shahzadi Mahpara, Muhammad Waseem, and Ahmed Mohammed Saleh. "Evaluation of Machine Learning approaches for precision farming in Smart Agriculture

System-A comprehensive Review." IEEE Access (2024).

[13] Sarma, K. K., Das, K. K., Mishra, V., Bhuiya, S., & Kaplun, D. (2022). Learning aided system for agriculture monitoring designed using image processing and IoT-CNN. IEEE Access, 10, 41525-41536.

[14] Xiong, J., Yu, D., Liu, S., Shu, L., Wang, X., & Liu, Z. (2021). A review of plant phenotypic image recognition technology based on deep learning. Electronics, 10(1), 81

[15] Katharria, A., Rajwar, K., Pant, M., Snasel, V., & Deep, K. (2024). Information Fusion in Smart Agriculture: Machine Learning Applications and Future Research Directions. Available at SSRN 5187083.

[16] Mishra, A. M., Harnal, S., Mohiuddin, K., Gautam, V., Nasr, O. A., Goyal, N., ... & Singh, A. (2022). A Deep Learning-Based Novel Approach for Weed Growth Estimation. Intelligent Automation & Soft Computing, 31(2).

[17] Pierre, N., Viviane, I. I., Lambert, U., Shadrack, I., Erneste, B., Schadrack, N., ... & Theogene, H. (2023). AI based real-time weather condition prediction with optimized agricultural resources. European Journal of Technology, 7(2), 36-49.

[18] Zhao, L., Haque, S. M., & Wang, R. (2022). Automated seed identification with computer vision: challenges and opportunities. Seed Science and Technology, 50(2), 75-102.

[19] Mahant, V., & Pal, P. (2025). Seeding the future: a review of artificial intelligence's pivotal role in seed technology. Journal of Crop Science and Biotechnology, 1-14.

[20] Sun, W., & Wang, S. (2024). The future research directions of artificial intelligence in crop growth prediction. Geographical Research Bulletin, 3, 251-254.

[21] Rao, B. S., Rao, K. S., & Ome, N. (2016). Internet of Things (IoT) based weather monitoring system. international journal of advanced research in computer and communication engineering, 5(9), 312-319.

[22] Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., ... & Kim, S. H. (2016). Random forests for global and regional crop yield predictions. PloS one, 11(6), e0156571.

[23] Bhattacharyya, B., Biswas, R., Sujatha, K., & Chiphang, D. Y. (2021). Linear regression model to study the effects of weather variables on crop yield in Manipur state. International Journal of Agricultural and Statistical Sciences, 17(1), 317-320.

[24] Khan, M. Z., Era, M. D., Islam, M. A., Khatun, R., Begum, A., & Billah, S. M. (2019). Effect of coconut peat on the growth and yield response of Ipomoea aquatica. American Journal of Plant Sciences, 10(03), 369.

[25] Ahmad, A. A., Radovich, T. J., Nguyen, H. V., Uyeda, J., Arakaki, A., Cadby, J., ... & Teves, G. (2016). Use of organic fertilizers to enhance soil fertility, plant

growth, and yield in a tropical environment. Organic fertilizers-from basic concepts to applied outcomes, 85-108.

[26] Pecho, R. D. C., Vijaya, K. S., Sharma, N., Pal, H., & Jose, B. K. (2024). An approach for crop yield prediction using hybrid XGBoost, SVM and C4. 5 classifier algorithms. Engineering and Applied Science Research, 51(3), 300-312.

[27] Naqvi, S. M., Tahir, M. N., Raghavan, V., Awais, M., Hu, J., Said, Y., ... & Khan, M. I. (2025). AI-enhanced IoT sensors for real-time crop monitoring: an era towards self-monitored agriculture. Telecommunication Systems, 88(3), 1-15.

[28] Sharma, A., Georgi, M., Tregubenko, M., Tselykh, A., & Tselykh, A. (2022). Enabling smart agriculture by implementing artificial intelligence and embedded sensing. Computers & Industrial Engineering, 165, 107936.

[29] Mekonnen, Y., Namuduri, S., Burton, L., Sarwat, A., & Bhansali, S. (2019). Machine learning techniques in wireless sensor network based precision agriculture. Journal of the Electrochemical Society, 167(3), 037522.

212-15-4097

ORIGINALITY REPORT

23%	19%	13%	18%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	5%
2	Submitted to Jacksonville University Student Paper	1%
3	www.coursehero.com Internet Source	1%
4	Submitted to United International University Student Paper	1%
5	www.researchgate.net Internet Source	1%
6	onlinejog.org Internet Source	1%
7	www.mdpi.com Internet Source	<1%
8	ph01.tci-thaijo.org Internet Source	<1%
9	Submitted to University of Adelaide Student Paper	<1%
10	iieta.org Internet Source	<1%
11	dspace.daffodilvarsity.edu.bd:8080 Internet Source	<1%
12	ouci.dntb.gov.ua Internet Source	<1%
13	prr.hec.gov.pk Internet Source	<1%

*% detected as AI

AI detection includes the possibility of false positives. Although some text in this submission is likely AI generated, scores below the 20% threshold are not surfaced because they have a higher likelihood of false positives.

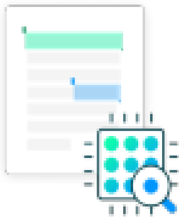
Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Frequently Asked Questions

How should I interpret Turnitin's AI writing percentage and false positives?

The percentage shown in the AI writing report is the amount of qualifying text within the submission that Turnitin's AI writing detection model determines was either likely AI-generated text from a large-language model or likely AI-generated text that was likely revised using an AI paraphrase tool or word spinner.



False positives (incorrectly flagging human-written text as AI-generated) are a possibility in AI models.

AI detection scores under 20%, which we do not surface in new reports, have a higher likelihood of false positives. To reduce the likelihood of misinterpretation, no score or highlights are attributed and are indicated with an asterisk in the report (*%).

The AI writing percentage should not be the sole basis to determine whether misconduct has occurred. The reviewer/instructor should use the percentage as a means to start a formative conversation with their student and/or use it to examine the submitted assignment in accordance with their school's policies.

What does 'qualifying text' mean?

Our model only processes qualifying text in the form of long-form writing. Long-form writing means individual sentences contained in paragraphs that make up a longer piece of written work, such as an essay, a dissertation, or an article, etc. Qualifying text that has been determined to be likely AI-generated will be highlighted in cyan in the submission, and likely AI-generated and then likely AI-paraphrased will be highlighted purple.

Non-qualifying text, such as bullet points, annotated bibliographies, etc., will not be processed and can create disparity between the submission highlights and the percentage shown.