Faculty of Computer Engineering Informatics and Communications

k-NN Based Predictive Model for Optimization of Tomato Growth

BY

Rumbidzai Michelle Gwinji

Project submitted for review for
COMPUTER ENGINEERING
in the Faculty of Computer Engineering Informatics and Communications at
the ***University of Zimbabwe***
Supervisor: ***Mr. P Worsnop***

June 6, 2024

Declaration

The researcher, Rumbidzai Gwinji, declares that this project is her original work and has not been submitted to any other university . All references and citations from other sources are properly acknowledged and documented in the text, footnotes, and bibliography.

The researcher further declares that: The ideas, concepts, and methodologies employed in the project are based on her own understanding and research, supported by appropriate references and citations. The k-NN model developed to support this research, including its design, implementation, and testing has been carried out by the researcher and is the result of her efforts. Any external sources of information used in this project, including published literature, online resources, and intellectual property of others have been appropriately acknowledged through proper referencing. The researcher has complied with all the ethical guidelines and professional standards during the execution of this project ensuring the confidentiality and integrity of the information involved. The researcher takes full responsibility for any errors or inaccuracies that may be present in this project and hereby accept any consequences arising from them. The researcher understands that any form of plagiarism or dishonesty in presenting this project is a serious offense and can result in disciplinary actions by the educational institution.The researcher hereby affirms that this declaration is made in good faith, with complete honesty and integrity. The researcher understands the importance of adhering to academic principles and ethical standards, and the researcher is fully aware of the consequences of any violation.

Approval

This certifies that the project titled ***k-NN Based Predictive Model for Optimization of Tomato Growth***, submitted by ***Rumbidzai Gwinji***, has been approved as part of the requirements for the ***Bachelor of Science Honours in Computer Engineering (HCE)*** _____

Head of Department _____

Project Guide _____

Supervisor

Dedication

The significant achievements I've made are attributed to the unwavering support of my family. I consider this work a tribute to their belief in me.
And to mum and dad, this project is for you!

Acknowledgements

The researcher would like to express gratitude for the support and encouragement received from many individuals throughout the research journey. It was a great experience to discover new technologies and appreciate the art of research. The researcher would like to express gratitude to her supervisor Peter Worsnop for his kindness and patience in sharing his knowledge which in different ways has helped in the thorough work done and completion of this project.He never got tired of the researcher's numerous shouts for help, for that the researcher is grateful. The writer would also like to thank the faculty and staff at the Department of Computer Engineering for providing the necessary resources and a conducive environment for research.The researcher's profound gratitude goes to her parents (Mr and Mrs Gwinji) who in this journey supported financially and emotionally without complaints. Everyone's success is attributed to a community and the researcher would like to thank her cousins and best friends;Tendai and Fadzai who not only supported emotionally but believed in the researcher's capabilities to excel.Special thanks to the farmers and agricultural experts who provided practical insights and data for this project.Their contributions have enriched the research and ensured its relevance and applicability.Special mention goes to my aunt; Unanda who is an agronomist specialist for her guidance. Above all, the researcher would like to express gratitude to the Lord Almighty for the success of this project.

Abstract

This project aimed to enhance tomato growth prediction and optimize fertilizer use through the development of a k-NN model.The primary objective was to develop a K-Nearest Neighbors model to forecast tomato plant growth , specifically in the context of Zimbabwean agriculture. Additionally, the project sought to deploy this model in a user-friendly interface and apply optimization techniques to provide real-time fertilizer recommendations based on current soil conditions.Background research highlighted the critical role of variables such as nitrogen, phosphorus, potassium , temperature, soil moisture, and humidity in influencing tomato growth. Comprehensive data collection from reliable sources ensured these variables were well-represented in the dataset.The research was aimed at bridging the gap between assumed knowledge and data-driven insights based on experimental research.The optimized k-NN model was integrated into a user-friendly interface, allowing farmers to input current conditions and receive real-time predictions and fertilizer recommendations. hyper-parameter tuning was employed to optimize performance, ensuring accuracy and reliability. This practical application aimed to support better crop management and yield optimization. The integration of real-time NPK readings into the model provided actionable insights for farmers, enhancing the practical utility of the system.Key outcomes included the robust performance of the k-NN model for both height and fruit number after optimization using Random Search technique. The project successfully demonstrated the ability to provide optimized fertilizer recommendations, leading to improved growth rates and favorable outcomes in real-world applications.In conclusion, this project successfully developed and deployed a predictive modeling system that can significantly enhance tomato growth and optimize agricultural practices. By providing farmers with accessible and accurate tools for crop management, the project contributes to the broader goal of sustainable and efficient agriculture. Future work will focus on expanding the dataset, validating the models in diverse environments, and re-running the k-NN model with other optimization techniques for the growth metrics of the following week, and/or for the growth rate of the tomato plant.

Contents

List of Figures

List of Tables

Nomenclature

R^2	Coefficient of Determination
GSM	Global System for Mobile Communications
k-NN	K-Nearest Neighbors
MSE	Mean Squared Error
NPK	Nitrogen, Phosphorus, Potassium
pH	Potential of Hydrogen

List of Appendices

- Appendix A: Survey Questionnaire
- Appendix B: Data Collection Methods
- Appendix C: Model Performance Metrics
- Appendix D: Fertilizer Optimization Algorithm
- Appendix E: Additional Graphs and Charts

Chapter 1

Introduction

1.1 Introduction

Crop growth prediction is a crucial aspect of precision agriculture, enabling farmers to optimize resource allocation and maximize crop yields.k-nearest neighbors (k-NN), a versatile machine learning algorithm, proved to be a promising tool for prediction of crop growth due to its simplicity, effectiveness, and capability to manage diverse data types. This research aims to investigate the application of k-NN for tomato crop growth prediction taking into account the nutrient and water requirements of the tomato crop.By analysing historical data on environmental conditions, soil properties, and cultivation practices, the researcher aims to train a predictive model capable of estimating tomato growth based on key input variables . The resulting model will serve as a valuable tool for farmers to make informed decisions regarding irrigation, application of fertilizers, and other critical management practices. The proposed research involves several essential steps in the machine learning pipeline. Through exploratory data analysis (EDA) and feature engineering, the researcher will extract meaningful features and identify key factors affecting tomato crop growth.The researcher will develop a predictive model that accurately forecasts nutrient uptake and tomato growth based on historical sensor data and plant growth parameters. Using k-NN, the model will provide valuable insights to farmers, enabling them to make data-driven decisions concerning resource allocation and fertilizer application. Most modern day farmers assume the ideal conditions for growth rate of the tomato plant based on generational knowledge.Whilst this generational knowledge is valuable in many cases, in a complex multi-dimensional system, a data-based approach can provide values that are more custom to particular scenarios. The research aims to predict growth of the tomato plant using plant height and fruit number as the measurement metrics based on inputs such as nitrogen, phosphorus, potassium, temperature , humidity and soil moisture. This can then be used to find ideal values for controllable inputs based on the value of the other inputs. The research will complement the already existing

smart greenhouses which will contain multiple sensors which will be used for data collection.

1.2 Background of the study

Crop growth [8] refers to the process by which agricultural crops develop and mature over time. It could also be defined [21] as the irreversible increase in size, while development involves the ongoing changes in plant form and function, marked by distinct transitional phases. It involves various physiological and biochemical changes, such as germination, vegetative growth, flowering, and fruiting. In the context of tomatoes, height measured over a period of time can be used as a metric for growth rate. However it is imperative to factor in the fruit number and the relationship with plant height. Vegetative parameters such as leaf area, stem length, fresh weight, and stem diameter can be used to assess the growth rate of tomato plants. [2]. The tomato plant is a widely cultivated crop, known for their economic importance and nutritional value. Understanding the growth patterns and factors affecting tomato plants is crucial for optimizing growth rate, quality yields and overall crop productivity.

Crop growth is important to agriculture because it is the foundation of food production.The research will be designed carefully to be in line with UN Sustainable development goals addressing several of them some of which include SDGs 2, 12 and 13. The research's focus on increasing food production efficiency, reducing water use, and improving crop quality aligns with Zero Hunger, Responsible Consumption and Production SDGs. Additionally, the research's focus on reducing the environmental impact of agriculture aligns with Climate Action SDG. Addressing these goals will create a significant positive impact on both the environment and society.

A number of researches have been done for prediction of crop growth rate in tomatoes as well as other plants, a number of them being implemented in the hydroponics farming techniques. A number of researches have been done on tomatoes, some highlighting the important inputs for optimization of tomato growth rate [13] while some emphasized the importance of nitrogen based fertilizers as well as good irrigation schedules for improved growth [29]. Researchers have been been successful implementing ANN for prediction of tomato growth and yield [29]. [16] proposed a system that implements k-NN Algorithm to predict the absolute crop growth of leafy vegetables in a greenhouse using hydroponics farming techniques. k-NN was shown to be accurate on hydroponics farming of leafy vegetables. [6] focused on the use of different machine learning algorithms on different crops in nutrient management. k-NN was shown to be accurate for nutrient management on potatoes. k-NN was also researched for disease detection [7] and crop selection prediction [11]. k-Nearest Neighbors is a machine learning algorithm that has been successfully applied in crop-related studies. k-NN can be used as a classifier or as a regressor and this research will

make use of its regression capabilities. k-NN has been of use in crop machine learning in tasks such as crop yield prediction [27], disease prediction [23] as well as crop growth rate as well as yield of leafy vegetables in hydroponics as mentioned above.

[25] proposed the use of plant height and vegetable index maps to measure the growth rate of the tomato for use in prediction of the crop yield. Besides the evidence that k-NN is an effective algorithm in the agricultural prediction, it has also been used to predict the growth rate of other crops such as corn and soy beans. There are various reasons why k-NN is the proposed machine learning algorithm for use in the research. Some of these reasons include:

- Non-linearity: k-NN works very well with data that does not follow a linear approach. Tomato growth rate is a non-linear parameter which will not follow a linear relationship thus k-NN is best suited for its prediction.
- Proximity-based Approach: k-NN uses the concept of proximity to make predictions, making it suitable for predicting plant growth because neighbouring plants often exhibit similar growth patterns under similar conditions. This is consistent with the assumption that neighbouring plants respond together to environmental factors and management practices.

The proposed model seeks to enhance traditional greenhouse management. By implementing this research, farmers will have access to a user-friendly interface where they get actionable insights, in order to make informed decisions, optimize plant growth and reduce resource consumption.

1.3 Problem definition

Most modern day farmers assume the ideal conditions for growth rate of the tomato plant based on generational knowledge.Whilst this generational knowledge is valuable in many cases, in a complex multi-dimensional system, a data-based approach can provide values that are more custom to particular scenarios. The aim of the research is to predict growth rate of the tomato plant using the plant height(cm/week) and fruit number per week based on the following input parameters: nitrogen, phosphorus, potassium, temperature , humidity, ph and soil moisture.This can then be used to find ideal values for controllable inputs based on the value of the other inputs using the k-NN algorithm.

Technically speaking, the researcher's aim is to minimize error between the predicted values and real outcomes by finding appropriate cluster sizes (k).

$$L(f,k) = \frac{1}{P}\sum_{p=1}^{p}(\hat{y} - f(x_p,k))^2 \tag{1.1}$$

The k-NN model considers the parameters which include soil moisture, pH, humidity, temperature, and nitrogen-phosphorous-potassium levels (NPK). Based

on the below equation; i (row) and j (column) are positions of the scalar in a matrix first row first column. These will be used as the feature data for the model,

$$X = x_{11}x_{12}....x_{1p}x_{21}x_{22}x_{2p}x_{p1}x_{p2}.....x_{pp}(1.2)$$

1.4 Aim

To develop a k-nn model to predict growth of the tomato plant using fruit number and plant height as measurement metrics against controllable and uncontrollable inputs and in turn maximize yield by manipulating the controllable inputs.

1.5 Objectives

- Develop and train a model using the k-NN Algorithm to predict growth of tomato plant used in a Zimbabwean Context.
- Deploy the k-NN model to a user-friendly interface where farmers can manipulate controllable inputs.
- To apply optimization on the model for recommendation of optimal values for controllable inputs to maximize plant growth.

1.6 Scope and limitations of the project

1.6.1 Scope

Development of Predictive Model: This research will focus on the development of a k-NN predictive model specifically designed to predict the growth rate of tomatoes.

Data Inputs: The model will be trained on data including:

- Plant characteristics(outputs of model): height and fruit number
- Growth-influencing metrics(inputs of model):
- Soil chemistry: Nitrogen (N), Phosphorus (P), Potassium (K) levels
- Environmental factors: Temperature, Soil Moisture, pH, and Humidity

Model Training and Validation: The k-NN model will be trained using historic data and validated using collected data from an experimental set-up to ensure accuracy and reliability in tomato growth rate predictions.

Application : The outcomes of this research are intended to assist farmers and agricultural professionals in optimizing growth conditions and improving crop yield through data-driven insights.

1.6.2 Delimitations and Limitations of the research

The research focuses on a variety of tomatoes called Solanaceae . Different tomato varieties may respond differently to environmental factors, and the results that will be yielded from this research may not generalize to all tomato cultivars.

The inputs which are considered for the research are pH, EC , NPK, humidity, temperature. There are other relevant inputs which can be considered for tomato crop growth rate such carbon dioxide CO_2 and light intensity.

The performance of the predictive model largely relies on the quality and completeness of the data collected. Any shortcomings or discrepancies within the data may affect the precision of the model.

Training and validation of the k-NN model, particularly with large datasets, can be computationally intensive and may require significant processing power and time.

1.7 Feasibility Analysis

1.7.1 Technical Feasibility

The project uses open-source platforms to develop the model which is accessible to the researcher. Data from Total Solution greenhouse farmers will be used to validate the k-NN model.Infra-red sensor, NPK sensor and Soil moisture sensor. Humidity sensor will be put across about $20m^2$ of tomatoes running for 3 months and this is feasible as the same sensors will be used for the period of time the data is collected from the greenhouse. The soil moisture sensors will be placed on different tomato crops and data will be collected on the edge to ensure accurate collection of data.

According to [22] define that technical feasibility posing a question are the technical resources available to implement the proposed system. The technical feasibility of this research addresses the question asked in the paper above as stated below:

- The researcher posses the necessary technical expertise.
- The researcher has access to the necessary technology.
- The researcher has access to the data set which will be used for training and has access to Total Solutions farm greenhouse where data will be collected from for validation to make it generalized to the Zimbabwean context.

Item	Quantity	Cost
ESP32	1	50
GSM Module for broadcasting communication	1	70
Infra-red sensor	1	20
NPK sensor	1	180
pH sensor	1	25
Soil Moisture sensor	3	300
Humidity Sensor	1	15
Data sets and travel to greenhouse	1	30

Table 1.1: Budget

1.7.2 Economical Feasibility

a) Cost of Sensors and Equipment: The cost of purchasing and installing the required sensors and related equipment will be sourced from the researcher's funds with support from family therefore the choice of components is made economically.

b) Operational Costs: The facilities that the research will be using are already operational and self-funded, and using them does not increase any operational burden that would require compensation.

1.8 Justification and rationale

Being able to predict the crop growth based on controllable variables of the growing would lead to higher yield efficiency and the following important aspects:

- Precision Resource Management: The development of the k-NN predictive model that estimates nutrient uptake, water requirements, and growth rate provides farmers with valuable insights into crop health and resource needs. This information enables precise resource management, such as tailored watering schedules and fertilizer applications. By optimizing resource usage, farmers can minimize waste, reduce costs, and promote sustainable agricultural practices.

- Informed Decision-Making: The research's focus on providing a system that provides actionable recommendations and graphical representations of growth rate empowers farmers to make informed decisions regarding crop management. By visualizing growth patterns over time, farmers can identify trends, compare growth rates, and detect any abnormal growth behaviours. This information helps farmers adjust their cultivation strategies, identify nutrient deficiencies or excesses, and optimize the growth of tomatoes.

Overall, the research is important because it has the potential to enhance environmental monitoring, optimize resource management, enable informed decision-

making and promote sustainable and productive greenhouse farming practices. By leveraging sensor integration and predictive modelling, the project offers a comprehensive solution that empowers farmers with insights and tools to optimize their operations and achieve better agricultural outcomes. Significance to the researcher The researcher will benefit through acquiring an in-depth understanding of designing, developing, deploying and maintaining the above described system. The researcher will also boost their knowledge on best IT systems design practises and current solutions that are the epicentre of delivering excellent engineered IT products and services. This research project will also provide the researcher with practical experience through interaction with the various components that constitute the system.

1.9 Work Plan

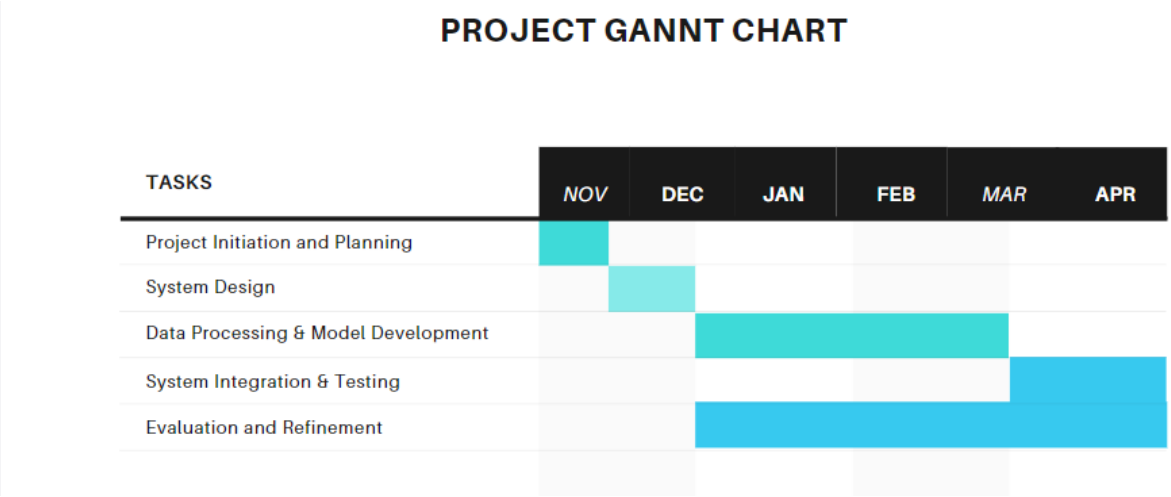

Figure 1.1: Project Timeline

1.10 Conclusion

This chapter introduces and explores the application of k-NN as a regressor for tomato crop growth prediction taking ph, soil moisture, temperature, humidity, NPK as input parameters. By leveraging the machine learning technique and analysing relevant data sources, the researcher sought to provide valuable insights for farmers to make informed decisions and enhance tomato crop productivity. A feasibility study was used to assess if the goal of the project would be attainable. It was determined that the necessary technical resources, including data sources, pre-processing techniques, and the k-NN algorithm, are available and suitable for the project‘s objectives. The feasibility study also highlighted the potential benefits, such as optimized agricultural practices and increased crop yields, which contribute to the economic viability of the project.

Considering nutrient management, soil data and nutrient levels are important input variables in the k-NN predictive model. By analysing the relationship between nutrient levels and tomato crop growth, the model can assist farmers in recommending optimal nutrient application rates, ensuring balanced nutrition for the crop, and mitigating nutrient deficiencies or excesses that can hinder growth. Water requirements are crucial for tomato crop growth, and this research recognizes the significance of considering that. By utilizing climate and weather data, variables such as rainfall, humidity, and evapo-transpiration rates will be incorporated into the predictive model. This enables farmers to optimize irrigation strategies, ensuring that water is applied efficiently, minimizing wastage, and preventing both water shortage and water logging issues that can impact tomato crop growth.The integration of nutrient management and water requirements into the predictive model will enhance practical utility. By accounting for these factors, farmers can make well-informed decisions regarding fertilization and irrigation practices, promoting sustainable resource management and improving overall crop performance.

In conclusion, the application of k-NN as a regressor for tomato crop growth prediction as well providing recommendations for best fertilizer application schedules and irrigation schedules will be significant in the agricultural sector. The research project provides valuable insights to farmers so that they make data-driven decisions. Ultimately, this contributes to sustainable agriculture by improving crop growth and yield and resource-efficient farming practices.

Chapter 2

Literature Review

2.1 Review Relevant Research

2.1.1 Machine learning based crop growth management in greenhouse environment using hydroponics farming techniques

V.Mamatha et al 2023 [16] proposed a system that implements k-NN as a classifier to predict the absolute crop growth of leafy vegetables in a greenhouse using hydroponics farming techniques. In the research, the entire greenhouse is considered for crop plantation. Environmental factors such as water, temperature, air, total dissolved solid (TDS), pH, humidity, and light conditions inside the greenhouse will have a direct impact on the growth of plants. The data values that are collected from different growth substrate shows different accuracies when it is implemented using k-NN. The use of k-NN algorithm along with Nutrient Film Technique (NFT) technique produces an accuracy of 93% in the prediction of crop growth. In this proposed system of research, the experiments are carried out only on leafy vegetables which are hydroponically grown and future work can be done on fruits.

2.1.2 Improving the CROPGRO-Tomato Model for Predicting Growth and Yield Response to Temperature

The primary goal of the study [4] is to enhance the CROPGRO-tomato model, a tool used to predict tomato growth and yield, by improving its ability to account for temperature variations. The model was re-calibrated and evaluated using 10 datasets from field experiments in Florida from 1991 to 2007. Results showed that the modified parameters significantly improved the model's accuracy in predicting crop and fruit dry matter accumulation, with reductions in root mean square error (RMSE) of 44% for leaf area index, 71% for fruit number, and 36% for both aboveground biomass and fruit dry weight. The

Willmott d index, measuring model performance, consistently exceeded 0.92. The study highlighted the significant impact of temperature on tomato growth, suggesting future research could explore how temperature interacts with other input variables and how they affect tomato growth especially in the greenhouse context.

2.1.3 A machine learning approach for prediction system and analysis of nutrients uptake for better crop growth in the Hydroponics system

This paper [28] proposes a framework for predicting the absolute crop growth rate (CGR) in hydroponic tomato crops using machine learning. Key input variables include electric conductivity (EC) limits, nutrient solution (NS), ion concentration uptake, and the dry weight of the fruits. The study examines the correlations between these variables and the absolute CGR. It highlights the impact of nutrient ion uptake (Na, K, Mg, N, and Ca) on growth. The correlation analysis identifies critical factors influencing CGR, aiding in optimizing nutrient supply for better crop growth. The proposed system offers a smart and efficient method for predicting CGR and achieving high-quality yields by estimating essential parameter values. While the current approach focuses on predicting absolute CGR, future research could explore additional metrics like dry weight matter and relative growth rate, as well as more detailed nutrient uptake analysis.

2.1.4 Growth analysis of tomato plants in controlled greenhouses

The research [3]The research analyzed various vegetative and yield parameters of tomato plants, including leaf area, stem length, fresh and dry weights of leaves and stems, and stem diameter, measured at 14-day intervals. Additionally, fruit traits such as fruit length, diameter, fresh weight, number, and total yield were recorded at each harvest. Plant growth indices such as leaf area index (LAI), leaf area duration (LAD), leaf weight ratio (LWR), stem weight ratio (SWR), fruit weight ratio (FWR), specific leaf area (SLA), leaf area ratio (LAR), net assimilation rate (NAR), relative growth rate (RGR), and crop growth rate (CGR) were calculated. Climatic data, including humidity and temperature inside greenhouses, were recorded to investigate their relationship with other controllable variables affecting tomato plant growth.The study does not employ machine learning or predictive modeling techniques to integrate the collected data and forecast growth outcomes.While climatic data such as humidity and temperature were recorded, their integration with other growth parameters for predictive analytics was not fully explored.The research did not focus on optimizing controllable inputs (e.g., nutrient levels, water supply) to enhance growth and yield based on the collected data.

2.1.5 Growth rate and yield of two tomato varieties under green manure and NPK fertilizer rate Samaru Northern Guinea Savanna

The research [10] involved a treatment with two tomato varieties (Roma VF and UC82B), four rates of NPK 15-15-15 fertilizer (0, 150, 300, and 450 kg ha1), and three rates of green manure (0, 5, and 10 t ha1), using a split-plot design with three replications. Both varieties showed linear growth responses in plant height, relative growth rate, and crop growth rate (CGR) at 5 and 7 weeks after transplanting. UC82B outperformed Roma VF in terms of CGR at 5–7 weeks, net assimilation rate (NAR) at 7–9 weeks, and total fruit yield, with a 10.6% higher yield. NPK fertilizer application significantly enhanced plant height, crop dry weight, CGR, and overall yield. Future research could explore how NPK interacts with other controllable and uncontrollable vari

2.1.6 Application of Smart Techniques, Internet of Things and Data Mining for Resource Use Efficient and Sustainable Crop Production

Technological advancements have increased the use of the Internet of Things (IoT) to enhance resource efficiency, productivity, and cost-effectiveness in agriculture, especially amid climate change. With the rising global population, climate variations, and growing food demand, this paper reviews modern IoT and smart techniques for sustainable crop production.IoT aids farmers not only in smart crop cultivation but also in post-harvesting and managing consumer deals. It contributes to precision farming through technologies like agricultural drones, remote sensing, smart greenhouses, smart livestock management, computer imaging, and climate monitoring. Data mining and simulation modeling for crops and environmental management are gaining attention, with new algorithms being developed for better decision-making. These techniques are used for fertilizer application timing and rates, disease and yield predictions, soil moisture detection, and irrigation scheduling.The study aims to summarize the latest applications of smart techniques, including yield estimation, irrigation and fertilizer management, and pest and disease monitoring and management in crop production under changing climates.[1] reported symptoms such as bud drop, abnormal flower development, poor pollen production, dehiscence, and yield losses in tomatoes due to increased temperature.The paper highlights various IoT applications but does not address the integration of these technologies into a cohesive, scalable system for broader agricultural use.The paper does not sufficiently explore how these technologies can be adapted to local conditions and specific crop needs, particularly in developing regions.

2.2 Discussion of similar projects or systems

V.Mamatha et al 2023 [16] which is the base paper of the research used the k-NN for prediction of growth rate of the leafy vegetables. The model showed an accuracy of 93% in hydroponics . However its applicability to a different context was yet to be researched. The model was tailor-made for leafy vegetables only and research was encouraged to be done on fruits. The short comings of this base paper were addressed in the current research.

This paper [28] proposes a framework for predicting the absolute crop growth rate (CGR) in hydroponic tomato crops using machine learning. Key input variables include electric conductivity (EC) limits, nutrient solution (NS), ion concentration uptake, and the dry weight of the fruits. The study examines the correlations between these variables and the absolute CGR. It highlights the impact of nutrient ion uptake (Na, K, Mg, N, and Ca) on growth. This paper paid attention to controllable as in the current research. However,the research did not include the NPK which are the nutrients needed to foster tomato growth.

2.3 Identification of gaps or areas for improvement

k-NN was shown to be accurate on hydroponics farming of leafy vegetables and the efficiency depends on the k-NN as well as hydroponic technique. There is an opportunity to test the efficiency of k-NN on fruits and other farming techniques. The researchers proposed further exploration of the CROPGRO-Tomato model's application, particularly in forecasting growth and fruit yield of indeterminate tomato varieties, especially relevant for greenhouse cultivation. This entails examining additional cultivars with indeterminate growth patterns and considering other input parameters. It is important to map the relationship between yield and height and if experimental evidence show that height can be used a metric for tomato growth.NPK was never investigated against controllable inputs. Modern day farmers relied on assumed knowledge of the optimal values of NPK needed for optimal tomato growth however the current research provides experiment-backed knowledge to provide farmers with insights to foster growth.

2.4 Conclusion

So therefore, we have seen researches in prediction of tomato growth rate using dry weight, leaf are index, fruiting and flowering and other metrics for measuring growth rate incorporating imagine classifications, training of Ml learning models using regression algorithms as well as classifiers. K- was used for prediction of growth rate in hydroponics for leafy vegetables, it is also very popular for prediction of tomato crop yield, disease detection, future crop prediction and other areas pertaining to Agriculture. However, the researcher intends to use the k-NN algorithm in predicting the growth of the tomato plant using height and fruit number recorded over a specific period of time as metrics for measurement of crop growth, against input variables that affect crop growth which include the soil moisture, NPK, soil moisture and temperature in a greenhouse setup. The research will also include a smart way of monitoring the NPK fertilizer application for optimum growth and in turn higher yields.

Chapter 3

Methodology

3.1 k-NN - existing algorithm explanation

The k-NN model is a non-parametric model i.e., no-one assumes the distribution of the data [12]. It can be used for classification or regression. This research uses k-NN as a regressor. [18] mentions that k-NN algorithm uses distance metric techniques to identify the similarities and dissimilarities between the data points.The similarity between the data points increases with a decrease in the distance between them. The k-NN algorithm stores the training data and then uses these data instances to predict the test data. From the training dataset, k nearest neighbors are determined, for every new test data point. Cross validation can be used for choosing the value of k [20]. The study will employ cross-validation methods, such as k-fold cross-validation, to evaluate how the model performs across various values of k. This methodology entails partitioning the dataset into k subsets and sequentially training and testing the model on different permutations of these subsets. After which,metrics (e.g., accuracy, mean squared error) to measure performance for each k value will be measured and select the one that provides the best overall performance. It is envisioned in this study that data from the green house will be collected over three months, subdivided into folds, and used for the cross-validation of the model, then the outputs will be averaged for improved model performance.

3.2 Tomato Plant Growing in Greenhouses

The tomato crop belongs to the Solanaceae family of crops, and is widely grown worldwide [17]. Total Farm Solutions use a 5 x 6m greenhouse to grow a variety of tomato crops including the determinate tomato which has been selected as a point of focus for this research. The determinate tomato crop grows to an affine height then produces flowers and gives off fruit, all in a small period of time [14].

Figure 3.1: Total Farm Solutions' 5 x 6m case study greenhouse for the research

3.3 Problem Formulation

To identify the k nearest neighbors, the Euclidean distance is typically utilized, which measures the Cartesian distance between two points in a plane. This widely adopted metric calculates the distance between two data points. For Euclidean distance Φ illustrated in the diagram the following can be concluded that Eq. 1 holds,

$$\varphi(u, v) = ((u_i - u_j)^2 + (v_i - v_j)^2)^{\frac{1}{2}} \tag{3.1}$$

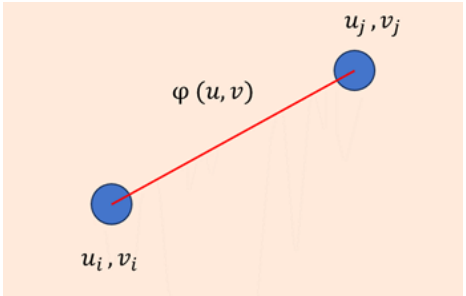

Figure 3.2: The Euclidean distance illustration between two magnified points

3.3.1 The pseudo code for the model

Start

Load training data ( ).

Load validation data ( ).

Normalize the loaded data in (1 – 2).

Compute distances ( ), between test points and training points

Sort results obtained from the above step in ascending order

Set the appropriate value of k

Predict the plant rate growth rate ( ).

End

3.3.2 The Improvement/modification to the k-NN model for plant growth rate prediction

In this study the model is used for regression which is generally represented by the regression expression in machine learning.

$$f(x,k) = \sum_{d=1}^{D} k_d f_d(x) \tag{3.2}$$

With loss function:

$$L(f,k) = \frac{1}{P}\sum_{p=1}^{p}(\hat{y} - f(x_p,k))^2 \tag{3.3}$$

It is considered to improve the performance of the model by using the Fuzzy theory approach[14]. The research [14] illustrated that in the conventional approach an element belongs or does not belong to a set, here the element belongs to a set by a discrete degree, i.e., for a set $X = X_i|_{i=1,2....P}$ The element x_1 can belong to a set by an m degree e.g., $0 \rightarrow 1$. The research will employ a fuzzy K-Nearest Neighbors (Fk-NN) regression model. Unlike the traditional k-NN algorithm, Fk-NN applies an unbiased weighting system in its decision-making process, taking into account the distances between the test sample and its closest neighbors. The classification of a new sample X in a class i, represented by its k nearest neighbors, is determined by calculating its membership degree. This method seeks to improve prediction accuracy by integrating fuzzy logic into the k-NN structure..[15]. The membership degree of a new sample X in class i, determined by the k nearest neighbors, is calculated as follows:

$$u_i(y) = \frac{\sum_{j=1}^{k} u_{ij}(1/|X - X_j|)^{2/q-1}}{\sum_{j=1}^{k}(1/|X - X_j|)^{2/q-1}} \tag{3.4}$$

The parameter q in the fuzzy strength equation regulates the Euclidean distance $(|X - X_j|)^2$ between X and X_j, determining the influence of each nearest

neighbor on the membership value. Here, u_{ij} represents the membership of the sample X_j from the training data to class i among the k-nearest neighbors. This is projected to produce a high-performance algorithm which gives more accurate predictions at a slight expense of the compute power and time. To counter this, the researcher proposes that the data will be set into batches compatible with the Zimbabwe Centre for High Performance Computing (ZCHPC)'s computing resource of 16 GB from 12 cores on a single compute node. Depending on the data sizes, multiple nodes may be applied to run the model. This way, the algorithm will remain robust to new farm data and the outputs will have high accuracy and usability to the farmer.

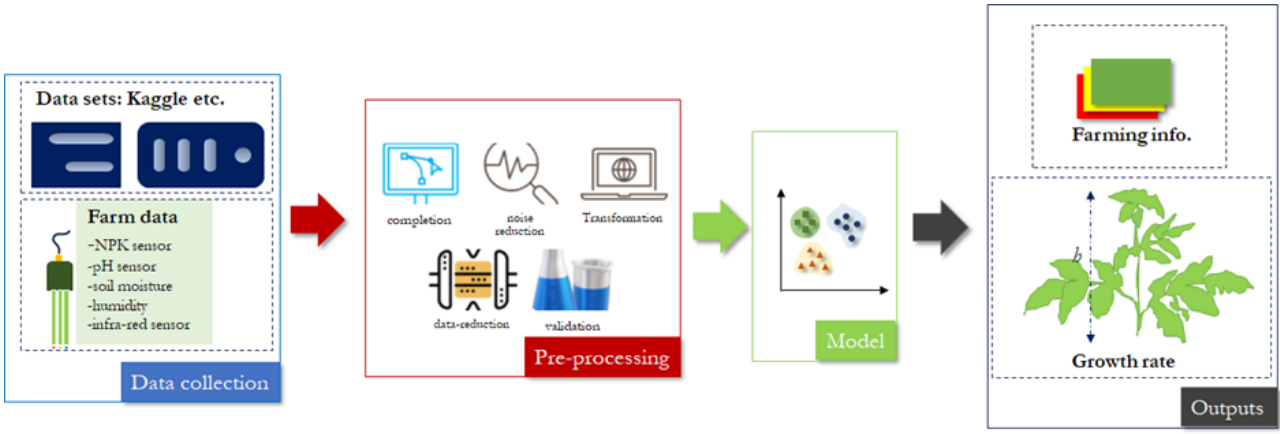

Figure 3.3: Proposed system development methodology

3.4 Dataset

This research uses an existing dataset to build the k-NN model, and collects a new data set to test its applicability to Zimbabwean tomato crops. The exisiting data set is shown below:

grid_3x3 K sort	grid_3x3 temperature sort	grid_3x3 humidity sort	grid_3x3 ph sort	grid_3x3 rainfall sort	te
5 205	8.83 43.7	14.3 100	3.5 9.94	20.2 299	
43	20.87974371	82.00274423	6.502985292000001	202.9355362	ri
41	21.77046169	80.31964408	7.038096361	226.6555374	ri
44	23.00445915	82.3207629	7.840207144	263.9642476	ri
40	26.49109635	80.15836264	6.980400905	242.8640342	ri
42	20.13017482	81.60487287	7.628472891	262.7173405	ri
42	23.05804872	83.37011772	7.073453503	251.0549998	ri
38	22.70883798	82.63941394	5.70080568	271.3248604	ri

Figure 3.4: Kaggle Data Set for Model Training

3.5 Experimental Setup

The sensors will be placed in the green house in proximity to the tomato plant. Calibration will be performed before the deployment of the sensors. Height of the tomato crop will be monitored periodically until the peak height at which fruits will be obtained. Levels of the NPK, pH, moisture, temperature, and humidity will be profiled against this measurement, moreover a control crop will be used to compare the monitored crop and the conventionally grown crop. The illustration of the experimental setup is given below:

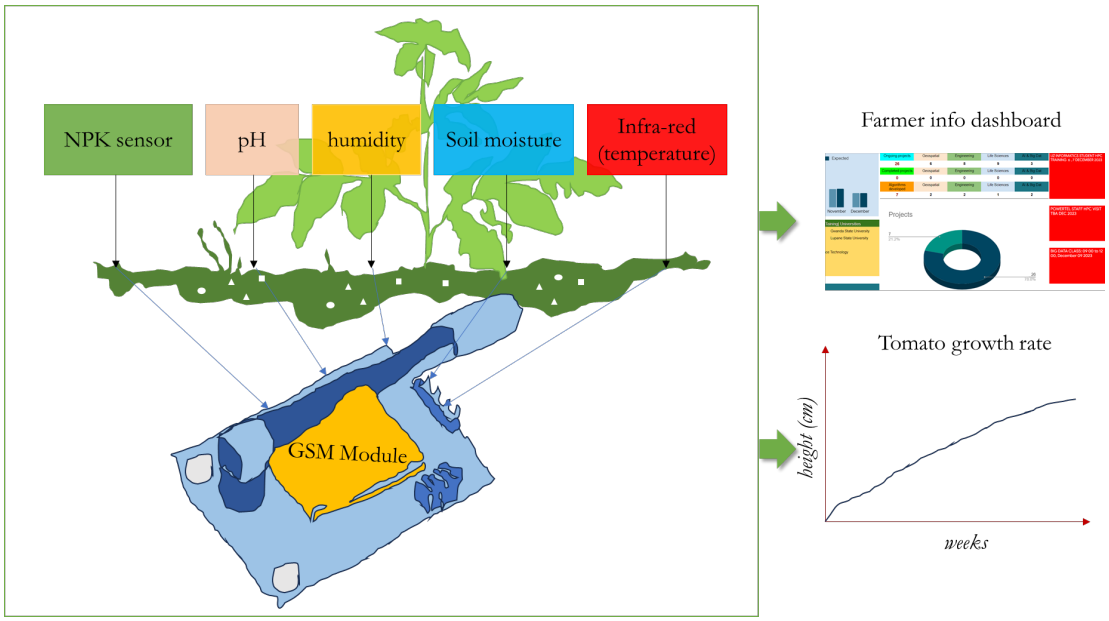

Figure 3.5: The proposed experimental setup in the green house

3.6 Evaluation

The dataset used will be split into training and test data, to establish the accuracy of the model. The efficacy of the algorithm will also be measured by the F1 score. Moreover, the algorithm will be validated using the data collected from the experimental unit Fig. 3.5 over three months from Total Farm Solutions' greenhouse, projected to be from beginning of February – end of May 2024. The predictions of the algorithm will be measured against the actual collected data from the greenhouse. The algorithm will also be compared to other published works which used other methods to perform crop growth rate predictions or condition monitoring of greenhouses/farms. Particular focus will be given to comparison of the % accuracy level to that of [24]who achieved average efficacy of 93.75% on feature recognition of crops towards precision farming.

3.6.1 Data Handling and Feature Engineering

Data pre-processing for k-NN Model

Before training the k-NN model, it was crucial to pre-process the data to ensure that it was clean, standardized, and suitable for modeling. This pre-processing involved several steps, each aimed at addressing potential issues that could impact the model's performance. Here's an overview of the pre-processing steps undertaken by the researcher :

- Data Collection and Integration

The researcher combined data from various sources, ensuring that all relevant features—such as nitrogen , phosphorus , potassium , temperature, humidity, soil moisture, number of fruits, and plant height—were included in a single dataset. The researcher made use of two datasets for training and used the Zimbabwean data set for validation.

- Handling Missing Values

Checked for missing values in the dataset and applied appropriate strategies to handle missing data.

- Data Normalization

Normalized the feature variables to ensure that they were on a similar scale. This was particularly important for the k-NN algorithm, which relies on distance calculations.Used Min-Max scaling to transform the data, bringing all feature values into the range [0, 1]. This helped prevent features with larger ranges from dominating the distance metric.

- Feature Engineering

Developed new features and adjusted existing ones to improve the model's predictive accuracy. For instance, combined temperature and humidity into a single feature to represent overall environmental conditions. Also applied polynomial feature expansion to account for non-linear relationships between the features and the target variables.

- Outlier Detection and Removal

Identified outliers in the dataset that could skew the model's performance.

- Data Splitting

Divided the pre-processed dataset into training and testing sets. This approach helped in assessing the model's performance on new data and ensures that the model does not over-fit.

- Ensuring Consistency

Verified that all pre-processing steps were applied consistently across the entire dataset, ensuring no data leakage from the training set to the test set.

By meticulously following these pre-processing steps, the researcher ensured that the data fed into the k-NN model was clean, well-scaled, and representative of the real-world scenario. This pre-processing laid the foundation for accurate and reliable model predictions, ultimately contributing to the success of our research in optimizing tomato plant growth.

3.6.2 Optimization for Fertilizer Recommendations in Tomato Growth

- ***Define the Parameter Space for Random Search***

Specify the ranges or distributions for n-neighbors, weights, and metric that you want to explore. This involves setting up the space in which the Random Search will look for the optimal hyper-parameters.

- ***Set Up and Run Random Search***

Use Randomized Search CV to perform Random Search over the defined parameter space. This step involves setting up the search algorithm with the number of iterations, cross-validation splits, and other settings. Fit the Random Search model on the data to find the best hyper-parameters.

- ***Predict Using Optimized NPK Values***

Use the optimal hyper-parameters found by Random Search to make predictions with the trained KNN model. This step involves applying these parameters to the model to get the predicted outputs.

- ***Evaluate and Compare Results***

Assess the predictions made using the optimized NPK values against the actual data.

- ***Analyze the Impact of Optimization***

Analyze the impact of the optimized NPK values on the predicted growth, comparing it to the initial predictions and actual growth data.

3.7 Conclusion

In this methodology chapter, we have detailed the experimental approach employed to develop a k-NN model for predicting the growth rate of the tomato plant . The experimental methodology provided a comprehensive framework for developing a robust predictive model. By systematically integrating various environmental and growth-related variables, we aimed to create a model that offers accurate and actionable insights for optimizing tomato plant growth. This methodological approach not only ensures the reliability of predictions but also supports the practical application of our findings in the Zimbabwean context.In summary, our experimental methodology has laid a solid foundation for the development of the k-NN based predictive model as well as the validation, contributing significantly to the field of precision agriculture and providing valuable tools for enhancing crop yield and resource management.

Chapter 4

Analysis and Design

4.1 Introduction

The tomato growth prediction model using k-NN algorithm is deployed onto sensor hardware including, pH, moisture, and NPK sensors controlled by an Arduino micro-controller for a web-based platform which retains the results. The chapter discusses the steps involved in building the model including the domain and user requirements, system components and design, and technique of implementation, feature selection, model training, and evaluation.

4.2 Detailed analysis of the problem domain and user requirements

Most modern day farmers assume the ideal conditions for growth rate of the tomato plant based on generational knowledge.Whilst this generational knowledge is valuable in many cases, in a complex multi-dimensional system, a data-based approach can provide values that are more custom to particular scenarios. There is need for a model that provides data-driven insights to foster the growth of the tomato plant.During the exploration of machine learning to produce a growth prediction model, the researcher mostly struggled with finding suitable training data from the local pool. Data on tomato plants carrying parameters such as temperature, humidity, pH, soil conditions, historical yields, and tomato crop types proved challenging to come-by especially for the Zimbabwean context. The researcher therefore resorted to collecting own data and mixing it with other data from sources available online to create usable datasets for model training. From this data, select features were then specified for the model and used in training the k-NN algorithm used in the study. The model was validated and balanced out to avoid over fitting and underfitting.

4.2.1 Functional Requirements

User Interface

1. Develop a user-friendly interface for inputting data and displaying predictions and recommendations.
2. Allow users to adjust input variables and immediately see the predicted outcomes and optimized recommendations.

Prediction and Optimization

1. Provide real-time predictions of tomato plant growth based on current input variables.
2. Offer optimization recommendations for NPK fertilizer application and irrigation schedules to maximize yield.

- ***Training pipeline requirements***

Data Collection

1. Collect data on tomato plant growth metrics, including plant height and fruit number, over specific time intervals.
2. Gather data on input variables such as soil moisture, NPK levels, electrical conductivity (EC), pH, and soil temperature within a greenhouse setup.

Data pre-processing

1. The collected data is cleaned and normalized to ensure accuracy and consistency.
2. Missing values are addressed, and relevant variables impacting tomato growth are identified through feature selection.

Model Training

1. Implement the K-Nearest Neighbors (k-NN) algorithm to train the predictive model using the pre-processed data.
2. Optimize the number of neighbors (k) and select appropriate distance metrics to enhance model accuracy.
3. Train the Fuzzy K-Nearest Neighbors (Fk-NN) model to incorporate an unbiased weighting scheme.

Model Evaluation

1. Evaluate the model's performance using metrics such as MSE and RMS values.
2. Apply cross-validation techniques to validate the model and prevent overfitting.

4.2.2 Non-functional Requirements

Performance: The system must provide predictions and recommendations within a reasonable time frame with the model should handling large datasets efficiently without significant performance degradation.
Accuracy: The predictive model should achieve MSE and RMS values within acceptable ranges, ensuring reliable predictions.
Scalability Developed system should be scalable to accommodate increasing data volumes and the addition of new variables or features in the future.
Usability: The user interface should be easy to use and interact with, requiring minimal training for users.Documentation and help features should be provided to assist users in understanding and utilizing the system.
Reliability: The system should be reliable and available with minimal downtime, ensuring continuous data collection and prediction capabilities.
Maintainability: The system should be designed for easy maintenance, with well-documented code and modular components that can be updated or replaced without affecting overall functionality.
Compatibility: The system should be compatible with various data input sources and formats to facilitate seamless data integration.The system should be designed to integrate with other agricultural management tools and platforms for comprehensive farm management.

4.3 Identification of system components and functionalities

The system includes input devices/sensors which include, pH sensor, humidity sensor, infra-red temperature sensor, and the nitrogen, phosphorus and potassium sensor which captures the input data for the system. The system then feeds this input into an Arduino Uno micro controller which would be hosting the k-NN model for processing. The output is then relayed to the server, and to a GSM for broadcasting to the farmers mobile device or remote device. The system offers suggestions and insights on the measured condition of the farm. Insights such as highlighting when it is time for addition of inputs, or boosting of the N, P, K levels.

4.3.1 Use Case Diagram

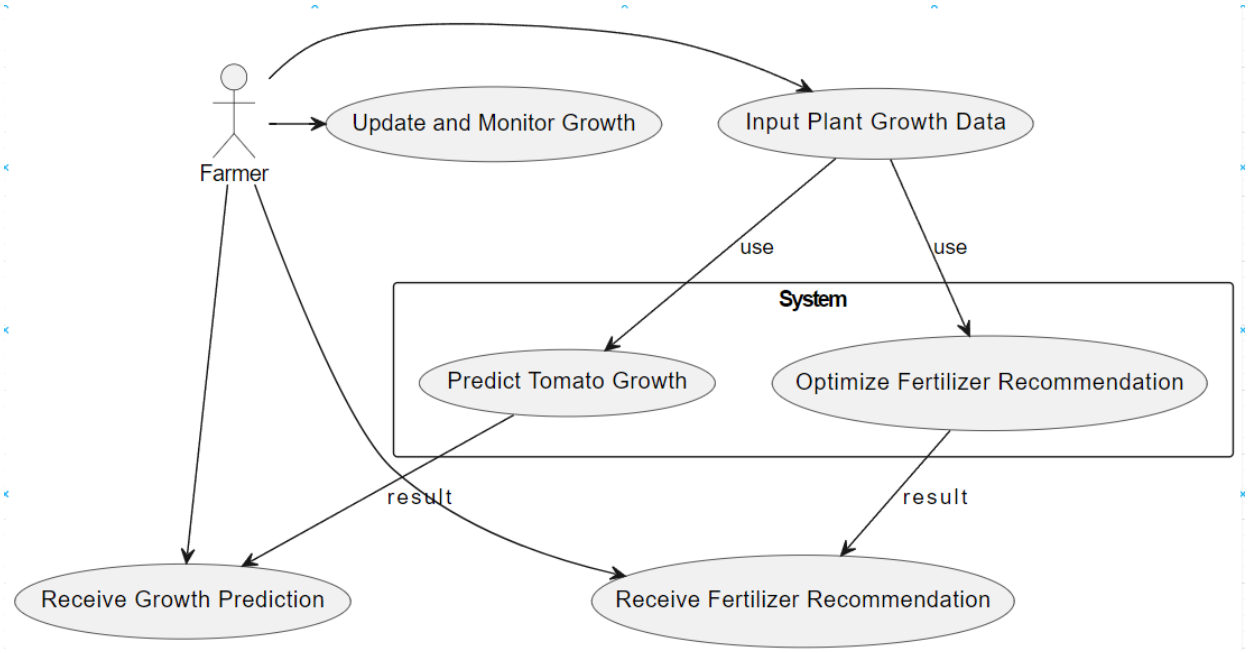

Figure 4.1: Use Case Diagram

4.3.2 Sequence Diagram

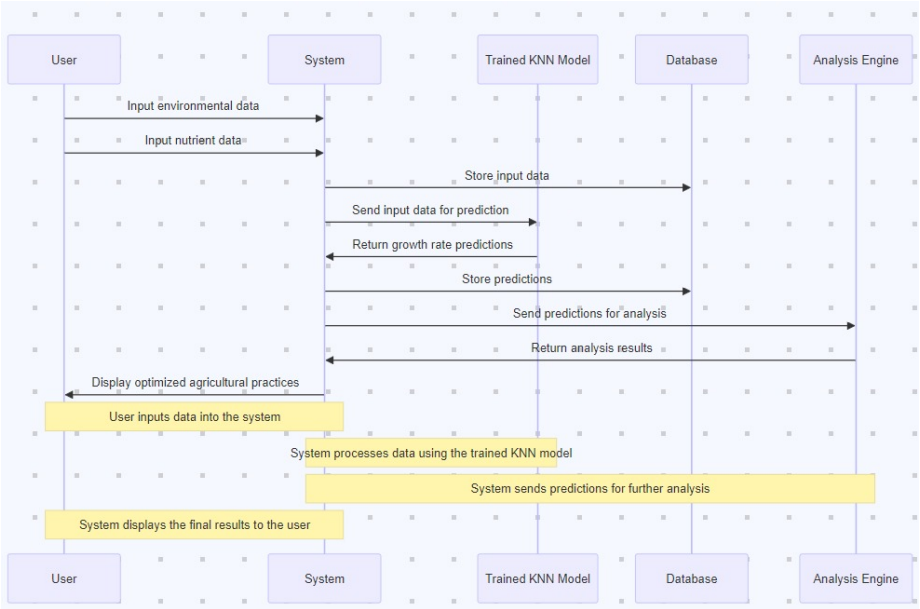

Figure 4.2: Sequence Diagram

4.4 System architecture and design considerations

The high-level system architecture for the k-NN tomato prediction system was designed using a modular and scalable approach. The system design involved data collection and pre-processing where real green-house data was collected by the researcher and prepared for use together with online based data, the interface design followed which interacted with external data sources such as the soil data, and historical tomato yield data. Data was then collected and pre-processed to make it suitable for input into the k-NN algorithm. The algorithm was then made to give prediction for the yields and growth-rate. The algorithm finds the K nearest neighbours to a new data point and assigns the majority class label to the new data point. A prediction output is received and the predicted yield from the k-NN Algorithm component is displayed to the user, and it is also stored in the database for further analysis. The researcher also considered the re-usability of the data collected from the system especially, as it slowly builds into a dataset for the Zimbabwean landscape. The system also uses a simple design with locally available components for ease of maintenance of the system.

4.4.1 Context Diagram and DFD Diagram

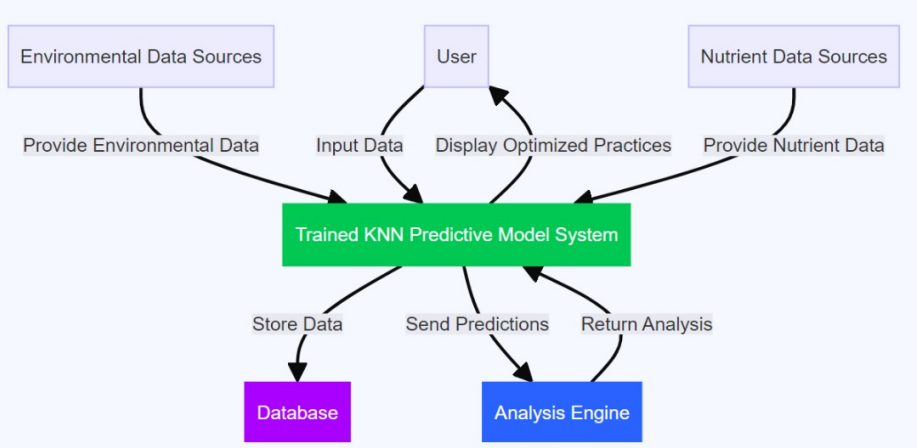

Figure 4.3: Context Diagram

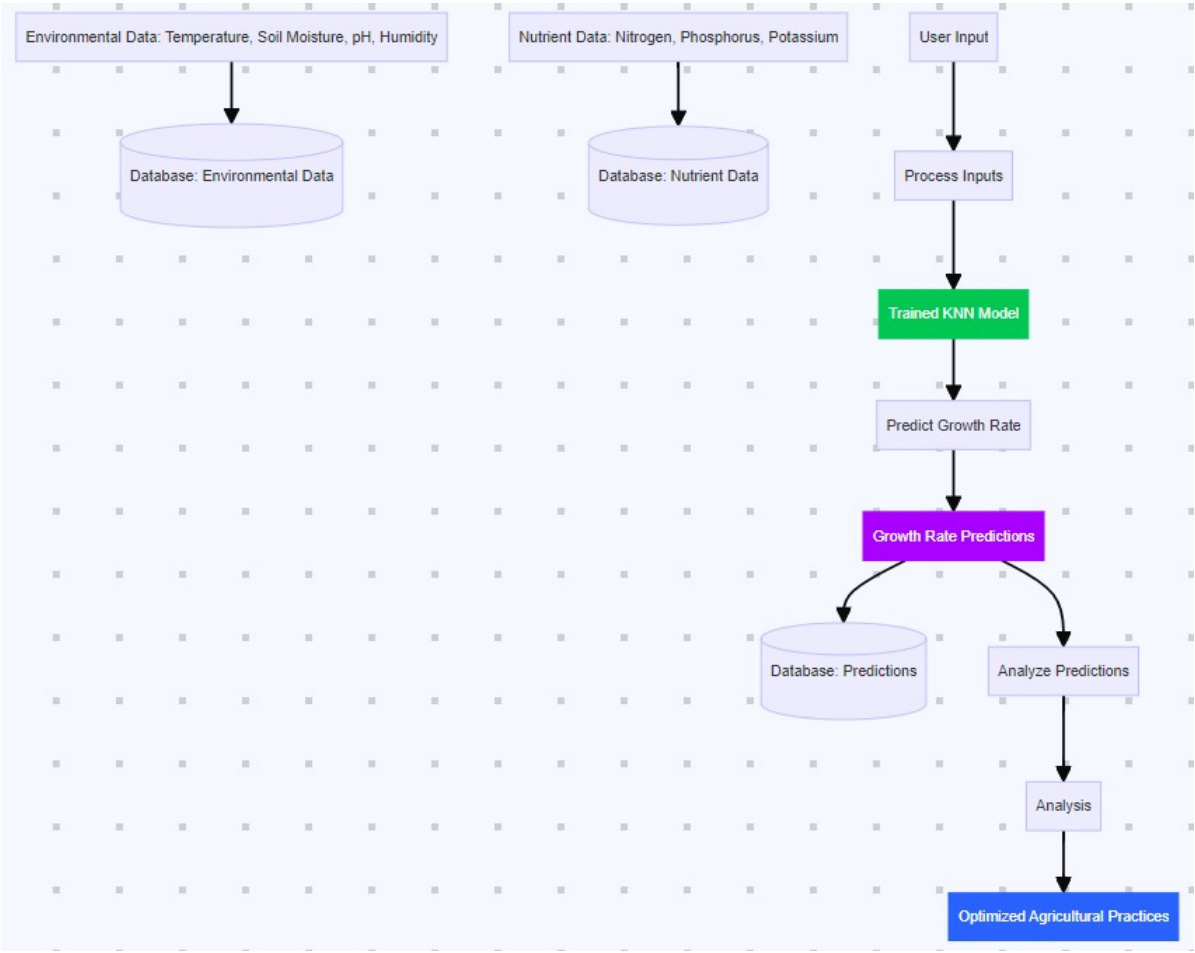

Figure 4.4: Data Flow Diagram

4.4.2 Architectural Design

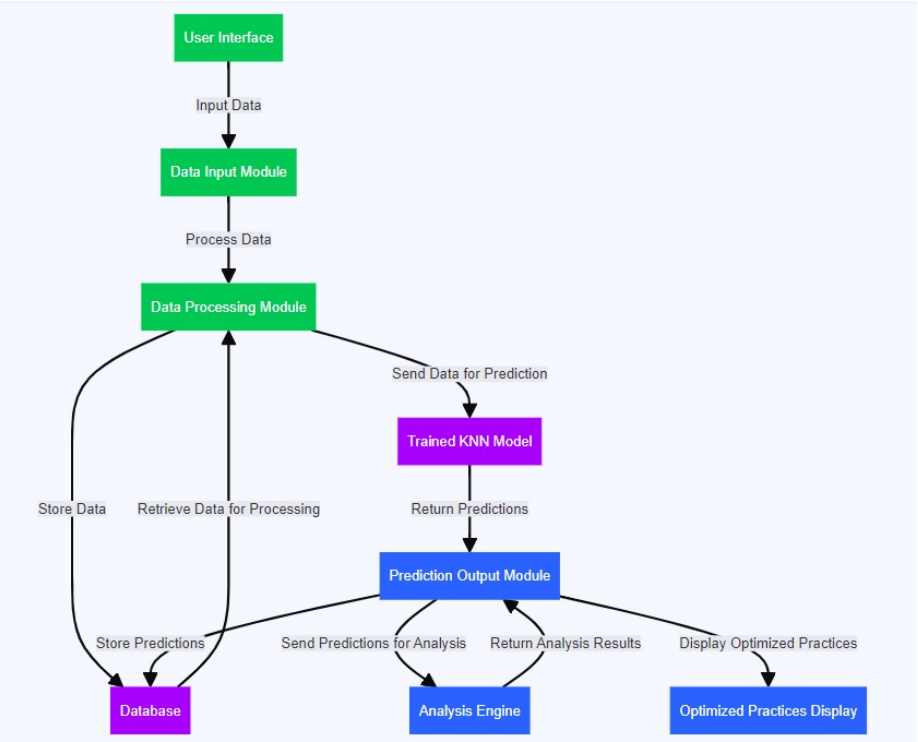

Figure 4.5: Architectural Design

4.4.3 Physical Design

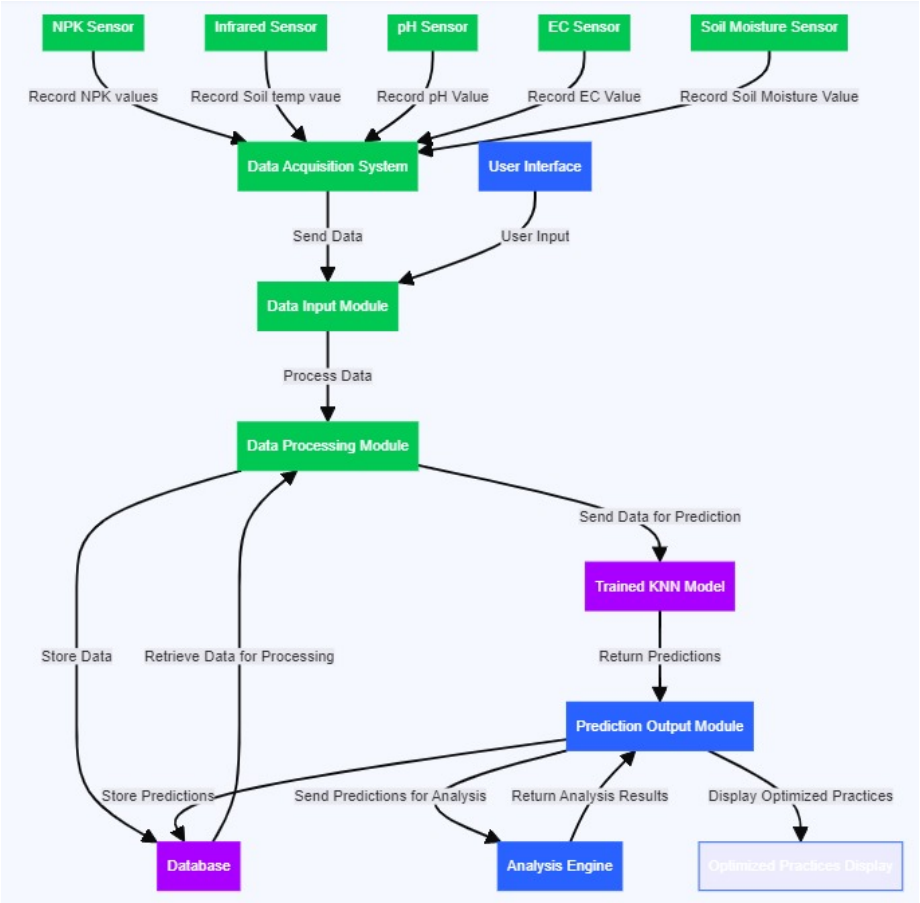

Figure 4.6: Physical Design

4.4.4 Database Design

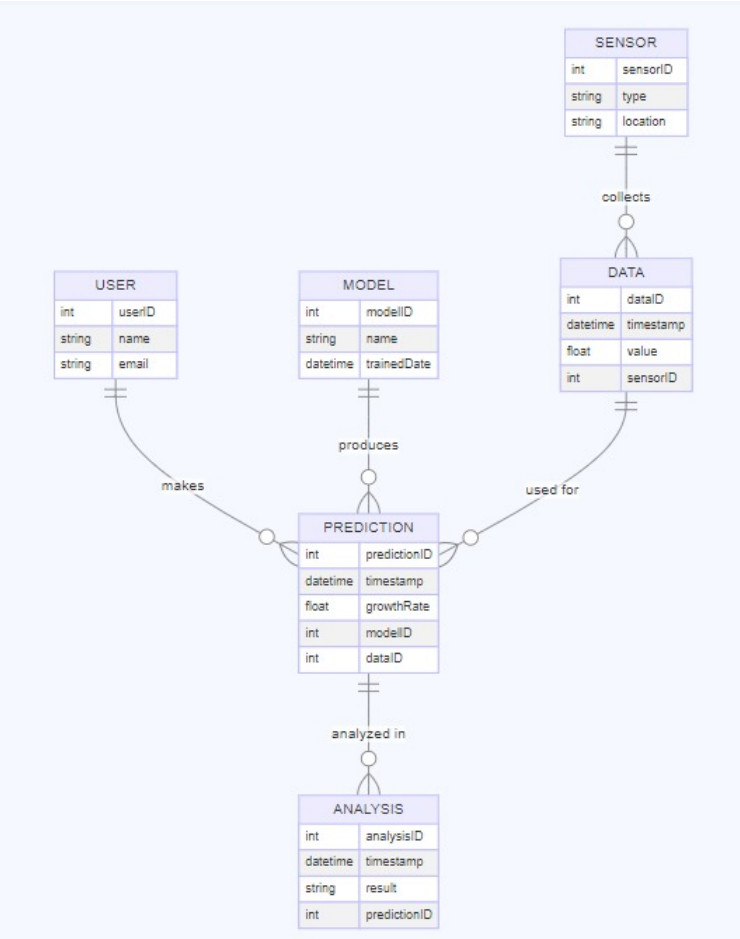

Figure 4.7: ERD

4.5 Interface design

4.5.1 Menu Design

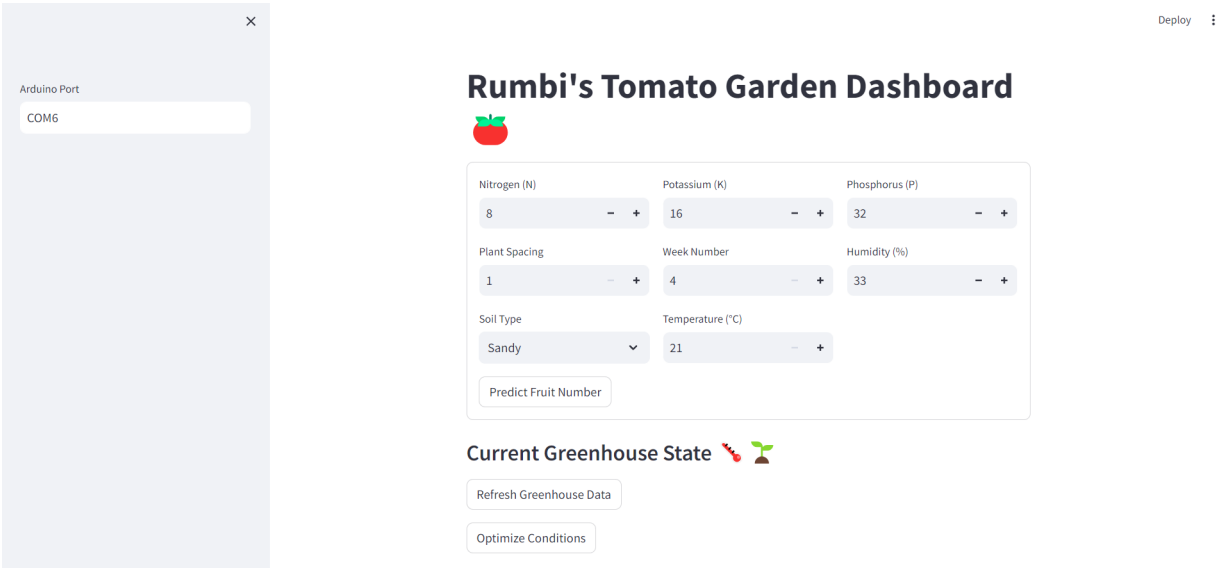

Figure 4.8: Main Menu

4.5.2 Input Design

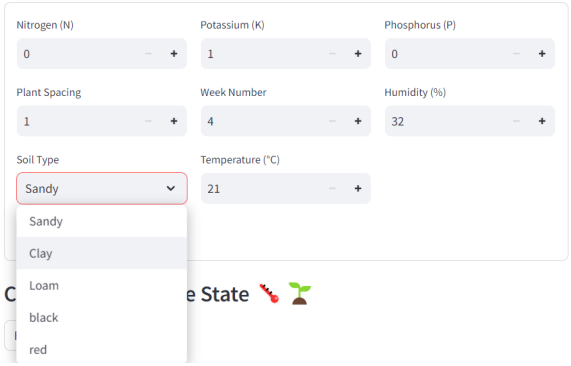

Figure 4.9: Inputs

4.5.3 Output Design

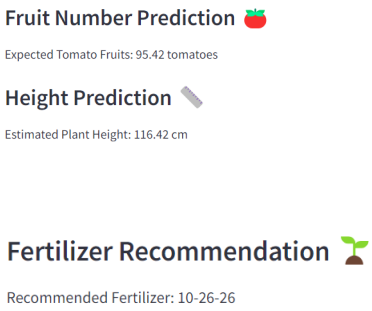

Figure 4.10: Outputs

4.5.4 User Interface Overview

- ***Loading the Web Application***:

The user starts by loading the web-based application on their device. Upon loading, they are automatically directed to the main Dashboard.

- ***Automatic Data Collection:***

The dashboard is continuously updated with data collected from various sensors placed in the greenhouse. These sensors monitor critical parameters such as soil moisture, temperature, humidity, and NPK levels. This data is automatically inputted and displayed on the dashboard in real-time.

- ***Predicting Fruit Number and Plant Height:***

There is a "Predict Fruit Number" button on the dashboard. When the user clicks this button, the application processes the real-time data from the sensors and predicts the expected number of fruits and the plant height. This prediction is based on the current growth conditions and the historical data patterns.

- ***Optimizing Growth Conditions:***

Another key feature on the dashboard is the "Optimize Conditions" button. When this button is clicked, the application analyzes the real-time sensor data and recommends the best fertilizer to use. This recommendation is based on the current percentages of NPK in the soil. The goal is to optimize these nutrient levels to foster optimal plant growth.

- ***Refreshing Greenhouse Data:***

The dashboard also includes a "Refresh Greenhouse Data" button. Clicking this button ensures that the data displayed on the dashboard is continuously reset and updated with new readings from the sensors. This feature allows the user to keep track of the latest changes and trends in the greenhouse conditions in real-time.

4.6 Conclusion

In this chapter, the researcher outlined how user requirements were translated into a comprehensive software solution for predicting tomato growth using the k-NN algorithm. The researcher demonstrated her ability to understand these requirements and create a well-structured design that serves as a clear guide for the implementation phase. By providing detailed descriptions of system components and their functionalities, along with supporting diagrams, the researcher effectively illustrated the system's structure and relationships. This groundwork ensures a smooth transition to implementation, paving the way for a robust and functional predictive model.

Chapter 5

Results

5.1 Introduction

The findings from this study are presented in this chapter, where the researcher used K-Nearest Neighbors as a regression model to predict key growth metrics of tomato plants which are height and fruit number. The analysis is based on data collected from a tomato farm in Zimbabwe used for validation, focusing on essential controllable inputs like nitrogen, phosphorus, potassium against uncontrollable inputs: temperature, soil moisture, pH, and humidity. Additionally, the model has been optimized to offer practical fertilizer recommendations tailored to real-time NPK levels sensored and sent to the user interface. In this chapter, the researcher will present how well the predictive model performed, show the relationship between height and fruit number as growth metrics.

5.2 Presentation of Findings

5.2.1 Model Score by RMS Values

Parameters	Score	R^2	Score:Z	Model
{'algorithm': 'auto', 'leaf_size': 30, 'metric': 'minkowski', 'metric_params': None, 'n_jobs': None, 'n_neighbors': 7, 'p': 2, 'weights': 'uniform'}	0.919497	0.919497	0.919497	k-NN_regression_model.pk

Table 5.1: k-NN Regression Model Results for Fruit Number Prediction

Parameters	Score	R^2	Score:Z	Model
{'algorithm': 'auto', 'leaf_size': 30, 'metric': 'minkowski', 'metric_params': None, 'n_jobs': None, 'n_neighbors': 5, 'p': 2, 'weights': 'distance'}	0.84978	0.84978	0.84978	k-NN_regression_model.pkl

Table 5.2: k-NN Regression Model Results for Height Prediction

5.2.2 Fruit Number Prediction

- **Parameters**:
  - **Algorithm**: `auto`
    - Automatically selects the appropriate algorithm for computing nearest neighbors based on the input data.
  - **Leaf Size**: 30
    - The leaf size affects the speed of the construction and query of the KDTree and BallTree data structures.
  - **Metric**: `minkowski`
    - A generalization of Euclidean and Manhattan distances. When $p = 2$, it becomes the Euclidean distance.
  - **Metric Parameters**: None
  - **Number of Jobs**: None
    - Uses the default number of parallel jobs, which is typically 1.
  - **Number of Neighbors**: 7
    - Considers the 7 nearest neighbors to make predictions.
  - **p**: 2
    - Uses Euclidean distance (Minkowski distance with $p = 2$).

- **Weights**: `uniform`
    - All neighbors are weighted equally when making predictions.

- **Performance Metrics**:
    - **Score**: 0.9194967788796279
    - **R^2 Value**: 0.9194967788796279
    - **Score:Z**: 0.9194967788796279

- **Model File**: `k-NN_regression_model.pkl`

5.2.3 Height Prediction

- **Parameters**:
    - **Algorithm**: `auto`
        - Automatically selects the appropriate algorithm for computing nearest neighbors based on the input data.
    - **Leaf Size**: 30
        - The leaf size affects the speed of the construction and query of the KDTree and BallTree data structures.
    - **Metric**: `minkowski`
        - A generalization of Euclidean and Manhattan distances. When $p = 2$, it becomes the Euclidean distance.
    - **Metric Parameters**: None
    - **Number of Jobs**: None
        - Uses the default number of parallel jobs, which is typically 1.
    - **Number of Neighbors**: 5
        - Considers the 5 nearest neighbors to make predictions.
    - **p**: 2
        - Uses Euclidean distance (Minkowski distance with $p = 2$).
    - **Weights**: `distance`
        - Neighbors closer to the query point are weighted more heavily in making predictions.

- **Performance Metrics**:
    - **Score**: 0.8497795584657768
    - **R^2 Value**: 0.8497795584657768
    - **Score:Z**: 0.8497795584657768

- **Model File**: `k-NN_regression_model.pkl`

Euclidean Distance Formula

In this research, the Euclidean distance is used to measure the distance between points in the feature space. The formula for Euclidean distance is:

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

where x and y are two points in n-dimensional space, and x_i and y_i are the coordinates of these points.

5.2.4 Graphical Comparison

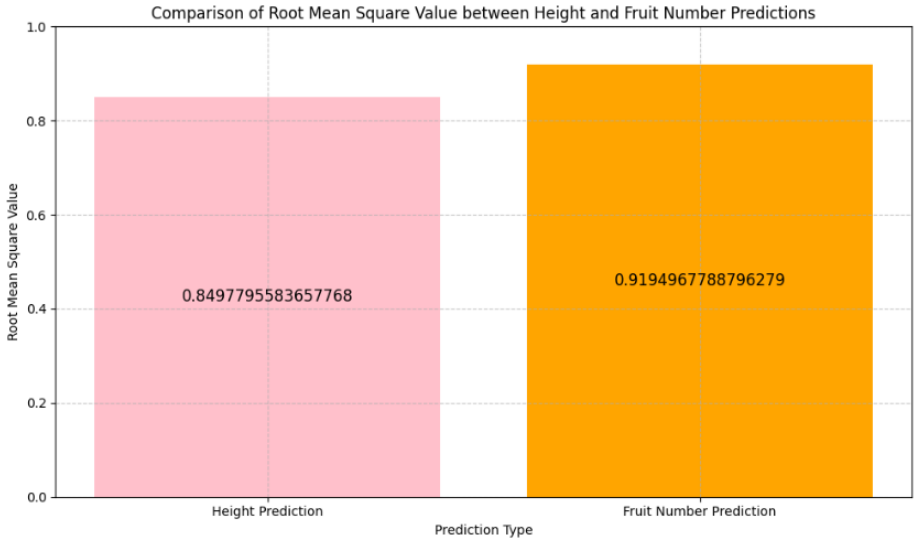

Figure 5.1: Comparison of RMS values

5.2.5 Model Score by MSE Values

Prediction Type	Parameters	Mean Squared Error	Model
Height Prediction	{'n$_n eighbors'$: 5}	0.222324	k-NN_regression_modelnumbers.pl
Fruit Number Prediction	{'n$_n eighbors'$: 5}	0.157461	k-NN_regression_modelnumbers.pl

Table 5.3: Combined Model Results (MSE)

Prediction Type: Height Prediction

- **Parameters**:
  - Number of Neighbors: 5
- **Performance Metrics**:
  - Mean Squared Error: 0.222324
- **Model File**: `k-NN_regression_modelnumbers.pkl`

Prediction Type: Fruit Number Prediction

- **Parameters**:
  - Number of Neighbors: 5
- **Performance Metrics**:
  - Mean Squared Error: 0.157461
- **Model File**: `k-NN_regression_modelnumbers.pkl`

5.2.6 Graphical Comparison

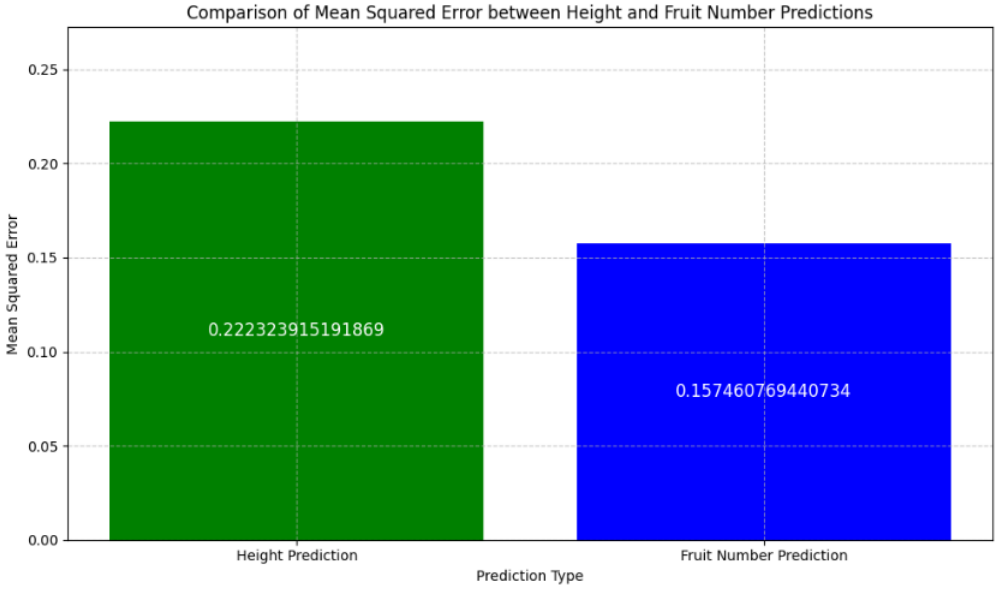

Figure 5.2: Comparison of MSE values

5.2.7 Relationship Between Fruit Number and Height

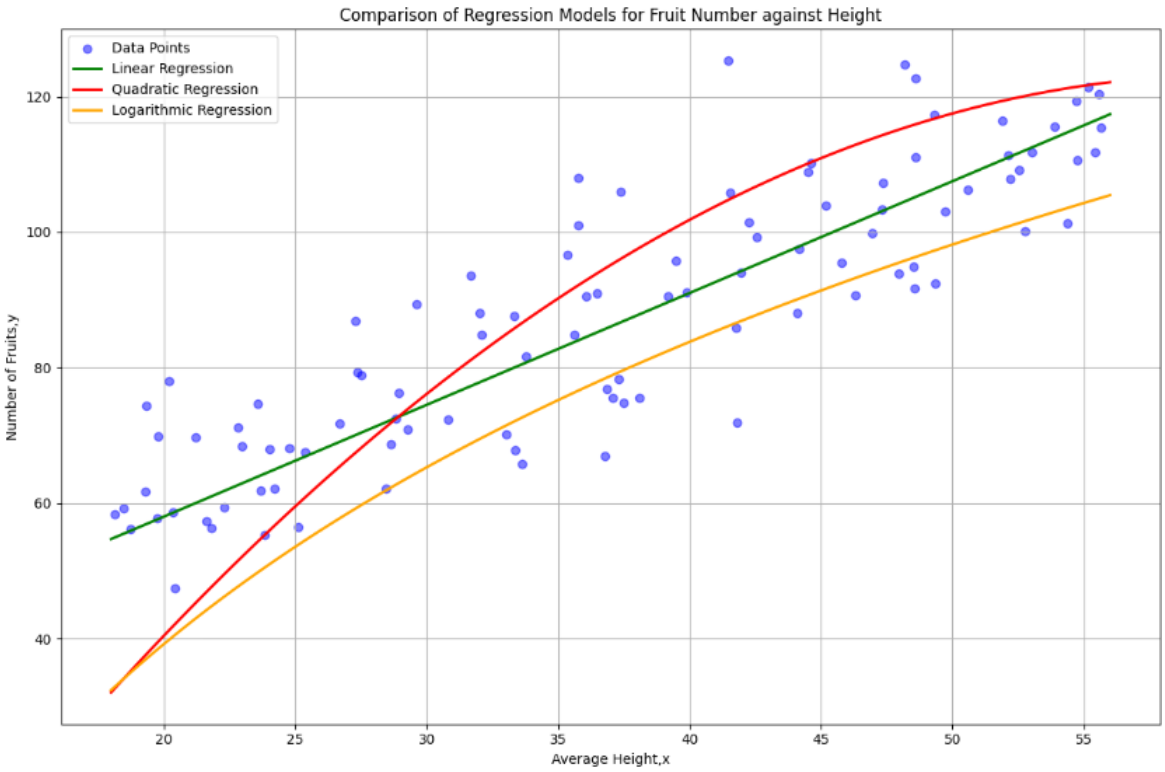

Figure 5.3: Number of Fruits and Height Relationship

The graph compares the effectiveness of three regression models—linear, quadratic, and logarithmic in predicting the number of fruits based on the average height of plants.
Linear Regression (green line): Despite the simplicity of the model, it achieved the best R^2 value, indicating that it explains the highest proportion of variance in the number of fruits relative to the height. This suggests a strong linear relationship between plant height and fruit number.
Quadratic Regression (red line): While it captures some non-linear patterns in the data, it does not perform as well as the linear regression in terms of R^2 value.
Logarithmic Regression (orange line): This model is less accurate compared to the other two, indicating that a logarithmic relationship is not as suitable for this data set.

5.2.8 Optimization Results

cccc

Metric	Before Optimization	After Optimization	Improvement (%)
Average Nitrogen	45.98	52.34	13.84
Average Phosphorus	16.23	18.56	14.36
Average Potassium	59.67	64.39	4.72
Average Fruit Number	33.78	38.12	12.86

Table 5.4: Summary of Average values of NPK and number of fruits Before and After Optimization

Optimizing NPK values using this model leads to improved tomato growth by providing tailored fertilizer recommendations. This increases tomato growth and there by increasing yields.

5.3 Conclusion

In this chapter, the researcher presented the findings from the study on predicting tomato growth using k-NN as a regressor. The results show how well these models can predict important growth metrics like plant height and fruit number based on factors such as NPK,soil moisture and humidity. The research also included practical fertilizer recommendations based on NPK readings. These findings confirm that our models are effective and can be useful in real-world farming to improve tomato growth and optimize agricultural practices.

Chapter 6

Discussion

6.1 Introduction

The results from the study on predicting tomato growth using k-NN are discussed in depth.. The above chapter looked at important growth metrics like plant height and fruit number, and also explored how optimizing fertilizer recommendations based on real-time NPK readings could improve outcomes. These findings will be interpreted , compared to what's already known in the literature, and considered for their practical implications for farming. Additionally, the researcher will discuss the limitations she faced during the study and suggest areas for future research. This discussion will help understand how effective and useful the predictive model and optimization techniques are in the context of agricultural science.

6.2 Summary of Findings

As shown in 5.3 , the linear regression model showed the best R^2 value, meaning it provides the best fit to the data among the three models tested. This suggests a strong linear relationship between the average height of tomato plants and the number of fruits produced. Given the linear model's superior fit, it can be inferred that for the range of heights studied, the number of fruits produced by tomato plants increases proportionally with plant height. This linear relationship is useful for predicting tomato yield based on plant height in practical agricultural settings because plant height can be much easier to measure than fruit number. Its important to note that while the quadratic and logarithmic models provide alternative perspectives on the relationship, their lower R^2 values indicate they are less effective at capturing the overall trend in the data compared to the linear model. The K-Nearest Neighbors (k-NN) algorithm is well-suited for agricultural data due to its flexibility, simplicity and non-parametric nature which allows it to handle complex, non-linear relationships. By leveraging optimization techniques like the random search technique implemented in this

research, the k-NN model was fine-tuned to provide precise fertilizer recommendations, optimizing the combination of NPK to maximize crop growth.This optimization process enhances the practical utility of the predictive model for farmers by improving accuracy, efficiency and relevance to local conditions. As a result, farmers receive tailored recommendations that increase yields,This approach serves as a valuable decision support tool, empowering farmers with data-driven insights to make informed agricultural decisions instead of relying on assumed knowledge to foster growth.

6.3 Model evaluation and analysis

The graph in 5.1 indicates how accurately the k-NN regression model's predictions match the observed data.

The below equation was useful in the context of this research as used in prior researches [9] which implemented k-nn as a regressor:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{6.1}$$

- **Fruit Number Prediction**:
- r^2 *Value: 0.9194967788796279*
- Approximately 91.95% of the variance in the number of fruits is explained by the model. This indicates an even stronger correlation between the model's predictions and the actual number of fruits.
- **Height Prediction**:
- r^2 *Value: 0.8497795583657768*
- Approximately 84.98% of the variance in the height of tomato plants is explained by the model. This shows a strong correlation between the model's predictions and the actual heights.

The graph in 5.2 illustrates the disparity between the predicted and actual values, providing insight into how closely the predictions of the model align with the actual data. The below equation was useful in the context of this research:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{6.2}$$

- **Height Prediction:**
- *MSE: 0.222323915191869*

- The model's predictions for height have an average squared error of approximately 0.22. This value indicates the average deviation between the predicted and actual heights of tomato plants.

- **Fruit Number Prediction:**

- *MSE: 0.157460769440734*

- The model's predictions for the number of fruits have an average squared error of approximately 0.16. This value indicates the average deviation between the predicted and actual number of fruits.

The MSE scored the model lower than the RMS due to the presence of outliers in the data. The quadratic nature of MSE amplifies the impact of these outliers, resulting in a higher error metric.

6.3.1 Analysis of Random Search Optimization

— ***Detailed Discussion***

- ***Average Nitrogen***

Before Optimization: The average nitrogen level in the dataset before optimization was 45.98.
After Optimization:After applying Random Search optimization, the average nitrogen level increased to an average 52.34.
Average Improvement (%): This represents roughly a 13.84% increase. Higher nitrogen levels are generally associated with better vegetative growth in tomatoes[18], which contributes to an increase in fruit production.

- ***Average Phosphorus***

Before Optimization: The average phosphorus level before optimization was 16.23.
After Optimization: Post-optimization, the average phosphorus level rose to 18.56.
Average Improvement (%): This constitutes a 14.36% increase. Phosphorus is critical for root development and fruit setting in tomatoes, explaining its impact on fruit yield.

- ***Average Potassium***

Before Optimization: The average potassium level was 59.67 before optimization.
After Optimization: Following optimization, the average potassium level slightly increased to 64.39.
Improvement (%): There is a 4.72% increase, potassium is essential for fruit quality and disease resistance.

- ***Average Fruit Number***

Before Optimization: Initially, the average number of fruits produced was 33.78.
After Optimization: The optimization process increased the average fruit number to 38.12.
Improvement (%): This improvement of 12.86% in fruit number highlights the efficacy of the optimization process. By fine-tuning the NPK values, the model could better predict the optimal conditions for maximizing fruit production.

6.4 Theoretical Implications

The researcher will highlight the relevance and impact of using k-NN for predicting tomato growth within the context of agricultural research giving emphasis to its adaptability, practicality and support for sustainable practices:

- ***Enhanced Predictive Accuracy with k-NN***

The findings from the current research provides evidence that the k-NN regressor accurately predicts key tomato growth metrics which are fruit number and plant height. The high values of r^s and MSE scores observed in the models indicate that k-NN can effectively capture the complex relationships between environmental variables and tomato growth outcomes.[19] supports the above notion.This enhances the predictive capabilities within the agricultural domain particularly for tomato cultivation which allows for more precise yield forecasts.

- ***Importance of Locality in Agricultural Predictions***

The emphasis on locality and proximity in k-NN aligns well with agricultural practices.[26] found that in tomato cultivation particularly, local conditions such as micro-climates and soil patches significantly impact plant development. k-NN's reliance on local data points for prediction reflects this agricultural principle, suggesting that nearby conditions provide the most relevant information for predicting growth outcomes. This theoretical implication supports the practical use of k-NN in localized agricultural settings.

- ***Sensitivity to Feature Scaling***

The research highlights the importance of proper feature scaling when using k-NN. Since k-NN calculates distances between data points, the scale of each feature can significantly affect the predictions. Normalizing or standardizing nutrient levels and environmental factors ensures that no single feature disproportionately influences the model's output. This reinforces the need for meticulous data pre-processing in agricultural research using k-NN.

- ***Practical Applications in Precision Agriculture***

The findings suggest that implementing k-NN in a user-friendly interface can provide real-time predictive insights for farmers. By inputting current environmental and soil data, farmers can receive accurate predictions on fruit yield

and plant growth along with recommendations for optimal fertilizer use. This practical application can lead to more informed decision-making, potentially increasing productivity and resource efficiency in tomato farming. The integration of k-NN into precision agriculture practices supports the movement towards data-driven farming, optimizing crop management at a granular level.

- ***Support for Sustainable Farming Practices***

The ability of k-NN to provide accurate predictions and optimize growth conditions aligns with sustainable farming objectives. By recommending precise fertilizer amounts and predicting yield outcomes, the k-NN regressor helps minimize waste and environmental impact. This theoretical implication highlights the role of machine learning in promoting sustainable agriculture, ensuring that farming practices are both productive and environmentally responsible.

6.4.1 Emerging Frameworks and Models

- Hybrid Modeling Framework:

A hybrid model which uses linear regression to identify strong linear relationships and k-NN to capture local variations and non-linearities providing a more comprehensive predictive approach.

- Integrated Nutrient and Environmental Predictive Model:

This integrated model combines predictive analytics with agronomic expertise, offering a holistic tool for optimizing crop yields based on a comprehensive set of growth factors.

6.5 Practical Implications

6.5.1 Real-World Applications of Predictive Models in Agriculture

- ***Agricultural Decision Support Systems (DSS)***

The predictive model developed in this study can be integrated into DSS used by farmers and agronomists. These tools can help make informed decisions about nutrient management against uncontrollable conditions. The model's simplicity and accuracy in predicting fruit numbers make these tools user-friendly and reliable enabling farmers to optimize yields with minimal effort. Adapting this model to various regions and crop varieties may necessitate additional data collection and fine-tuning which presents logistical challenges.

- ***Precision Agriculture Platforms***

The k-NN model can be utilized in precision agriculture platforms to provide real-time predictions and recommendations based on current environmental and soil conditions, using sensor data to continuously update and inform

decisions.This model is well-suited for handling large datasets with multiple variables enhancing the precision and accuracy of agricultural practices.It is important to note that integrating these models requires robust data infrastructure and real-time processing capabilities, which can be resource-intensive.

- ***Farm Management Software***

The model can be integrated into farm management software to help farmers plan and manage resources more effectively such as predicting growth rates and yields to optimize harvesting schedules and market supply.This integration can improve resource efficiency, reduce waste, and enhance financial planning for farmers.However, variability in farming conditions and practices may necessitate frequent updates and customization of the models, which can be time-consuming and costly.

- ***Research and Development***

Research institutions can build on these findings to further explore the relationships between growth factors and crop yield. They can develop more sophisticated hybrid models that combine various predictive techniques.Continuous improvement in predictive models can lead to more accurate and reliable agricultural practices fostering innovation in the agricultural field.Ongoing funding and collaboration between agronomists, data scientists and software engineers are necessary for sustained research and development which can pose as a challenge.

6.5.2 Potential Benefits

1. Increased Yield and Efficiency: By applying these predictive models, farmers can optimize resource use, leading to higher yields and more efficient operations.

2. Cost Reduction: Accurate predictions help minimize resource wastage, reducing the overall cost of inputs like fertilizers and water.

3. Improved Sustainability: Optimized use of inputs based on predictive models can lead to more sustainable farming practices, reducing environmental impact.

6.5.3 Limitations and Challenges

1. Data Quality and Availability: The accuracy of the models depends heavily on data quality and availability. Inconsistent or incomplete data can lead to inaccurate predictions.

2. Adaptability to Diverse Conditions: Models trained on specific datasets might not perform well in different environmental conditions or with different crop varieties. Ensuring the models' generalizability is challenging.

3. Resource Requirements: Implementing real-time predictive models requires significant computational resources and infrastructure, which can be problematic for small-scale farmers.

4. User Training and Acceptance: Farmers and agricultural stakeholders need training to use these models effectively and interpret their predictions correctly. Resistance to adopting new technologies can be a problem.

6.6 Validation and Reliability

6.6.1 Steps to Ensure Rigor and Trustworthiness

- ***Data Collection and pre-processing***

The researcher gathered comprehensive data from reliable sources, covering key variables like nitrogen, phosphorus, potassium, temperature, soil moisture, pH, and humidity. She also collected data from a greenhouse in Zimbabwe so that the model is generalized to the Zimbabwean context. Including these diverse factors enhances the study's content validity by addressing all major elements known to influence tomato growth.

- ***Model Selection and Optimization***

The researcher tested the model and optimized them through hyper-parameter tuning.She used random search optimization technique to optimize these models to ensure consistency and reliability in the findings.

- ***Cross-Validation and Testing***

The researcher split the dataset into k parts (folds) and trained the model k times, each time using $k1$ folds for training and the remaining fold for testing. The researcher used k-Fold Cross-Validation with the below equation as accordance to [26]

$$Cross - ValidationScore = \frac{1}{k}\sum_{i=1}^{k} MSE_{fold_i} \tag{6.3}$$

6.7 Limitations and Methodological Reflections

6.7.1 Limitations and Potential Weaknesses

- ***Research Design Limitations***

1. Model Selection and Scope Although the researcher tested linear, quadratic, logarithmic regression,the researcher mainly focused on these traditional methods.This approach might have overlooked more advanced models like neural networks and ensemble methods which could potentially offer better predictive performance.

2. Environmental Control The study was conducted in controlled environmental conditions.These conditions might not fully capture the variability and complexity of real-world farming environments, possibly affecting how well the findings apply to broader more varied contexts.

6.7.2 Data Collection Methods

- ***Measurement Accuracy***

There could be inaccuracies in measuring variables such as soil moisture, nutrient levels, and environmental conditions.Measurement errors can introduce noise and bias into the data, affecting the reliability and accuracy of the models and leading to less precise predictions and insights.

- ***Data Diversity and Representatives***

The data mainly represents specific regions, growing conditions and tomato cultivars.Limited diversity in the data can constrain the models' ability to generalize to different regions, climates, or cultivars, reducing the applicability of the findings in varied agricultural contexts.

- ***Sample Size Constraints***

The limited sample size affected the robustness of the statistical analysis and model training, potentially leading to over-fitting. Over-fitting occurs when models perform exceptionally well on the training data but fail to generalize to new, unseen data, reducing their reliability. Despite this challenge, using a dataset from Zimbabwe helped to mitigate the risk by ensuring the model's applicability and relevance to local conditions, thereby enhancing its generalizability and reliability in the targeted agricultural context.

6.7.3 Implications for Generalizability and Applicability

1. ***Generalizability*** Due to controlled conditions and specific datasets, the findings might not easily generalize to other crops, regions, or farming practices.While the model shows promising results for tomato growth under certain conditions, their effectiveness in different contexts remains uncertain. Further research is needed to test and adapt the models for broader applications.

2. ***Applicability in Real-World Scenarios*** Implementing these models in real-world agricultural systems may face challenges such as data integration, real-time processing capabilities, and user acceptance.Although the theoretical models are valuable, their real-world application requires additional infrastructure, user training, and adaptation to specific farming conditions and practices.

3. ***Evolving Conditions***Agricultural conditions and practices evolve over time, which might render static models less effective.Continuous updating and validation of the models are necessary to maintain their relevance and accuracy, posing an ongoing challenge for implementation.

6.8 Conclusion

This chapter analyzed how well k-NN works for predicting tomato growth. The results show that k-NN is effective at predicting important growth metrics making it a useful and flexible tool for handling complex agricultural data. This supports its role in precision farming where accurate predictions are crucial.The findings challenge the idea that only non-linear models are best for biological data[5]. Instead, k-NN, which does not rely on predefined assumptions proves to be a good fit depending on the data and conditions. This flexibility makes k-NN valuable for various agricultural applications.These results have practical implications, especially for improving tools like decision support systems, precision farming platforms, and farm management software. Using k-NN models can help increase yields, cut costs, and promote sustainability. However, issues like data quality, model adaptability, and resource needs must be addressed for successful implementation in the real world.The study was thorough, using detailed data collection, strong model optimization, and cross-validation techniques. Still, it's important to note limitations like research design, data diversity, and sample size, which impact how broadly the findings can be applied. More research and model testing in different agricultural settings are needed.

Chapter 7

Conclusion and Future work

7.1 Introduction

This final chapter provides a comprehensive summary of the study's findings and insights gained from predicting tomato growth using k-NN as a regressor. It also discusses the practical implications of these findings, acknowledges the limitations of the research and suggests areas for future work. The chapter emphasizes the importance of continuous model evaluation and adaptation in agricultural predictive modeling to ensure accuracy and relevance. Additionally, it outlines potential future research directions that could enhance the generalizability and practical applicability of the models thereby fully realizing its potential in diverse agricultural environments.

7.2 Summary of the Project

The aim of this research was to train and develop a k-NN model used for prediction of tomato growth against controllable and uncontrollable inputs such as nitrogen, phosphorus and potassium, temperature, humidity and soil moisture. The model was deployed into a user friendly interface where farmers could manipulate the controllable inputs for improved tomato growth. The researcher used Generic Algorithm Optimization technique to optimize conditions to provide farmer with fertilizer recommendations to foster tomato growth. The research aimed to bridge the gap between assumed values for tomato growth and data-driven insights for optimization of the tomato growth. The project incorporated an experimental methodology in which data was collected from a Zimbabwean farm for validation so that it is applicable and generalized to a Zimbabwean context.

7.3 Key findings and contributions

- ***Effectiveness of k-NN Model***

k-NN models demonstrated strong predictive capabilities for both plant height and fruit number. The performance was particularly enhanced by optimizing weighting strategies, validating the versatility of k-NN in handling multidimensional agricultural data.

- ***Critical Role of Growth Factors***

The study confirmed the significant impact of key growth factors such as nitrogen, phosphorus, potassium , temperature, soil moisture, pH, and humidity on tomato growth. This alignment with existing literature underscores the importance of these variables in predictive modeling.

- ***Fertilizer Optimization***

The incorporation of fertilizer recommendations based on temperature,humidity and soil moisture readings provided actionable insights for optimizing NPK application, directly addressing the practical needs of farmers to enhance crop yield and health.

7.4 Contributions to the Existing Body of Knowledge

- ***Expanded Application of k-NN***

The successful application of k-NN models in predicting growth metrics and optimizing fertilizer recommendations extends the use of this algorithm beyond traditional domains. This adds to the body of knowledge by showcasing k-NN's potential in agricultural and environmental sciences.

- ***Integrated Fertilizer Recommendation System***

The novel integration of a fertilizer recommendation system based on real-time NPK readings addresses a significant practical challenge in agriculture. This innovation not only enhances predictive accuracy but also provides farmers with direct, actionable recommendations to optimize nutrient application.

7.4.1 Addressing the Research Problem

- ***Predictive Modeling for Tomato Growth***

The primary research question focused on developing and training a model using the k-NN algorithm to predict tomato growth in a Zimbabwean context. The project successfully addressed this by creating robust models that accurately predict growth metrics based on critical environmental and nutrient factors.

- ***User-Friendly Deployment***

The deployment of the k-NN model to a user-friendly interface enables farmers to manipulate controllable inputs and receive real-time predictions and recommendations. This practical application ensures that the research outcomes are directly beneficial to end-users.

- ***Optimization for Enhanced Growth***

By applying optimization techniques to recommend optimal NPK values, the project directly contributes to maximizing plant growth by using data-driven insight and not just assumed or generational knowledge. This aligns with the objective of providing farmers with useful information to enhance crop yield and sustainability.

7.5 Evaluation of objectives

The k-NN model was successfully developed and trained using data relevant to the Zimbabwean context, incorporating key growth factors such as nitrogen, phosphorus, potassium , temperature, soil moisture and humidity. The model demonstrated strong predictive capabilities for predicting plant height and fruit number.The trained model was successfully deployed to a user-friendly interface, allowing farmers to input and adjust controllable variables.This interface provided real-time predictions and recommendations, enhancing the practical applicability of the research. Generic Algorithm Optimization technique was applied to the model to recommend optimal NPK values and in turn providing actionable insights for farmers to maximize tomato plant growth.The system was designed to offer real-time fertilizer recommendations based on current NPK readings, directly addressing the objective of enhancing crop yield.

7.5.1 Challenges and Mitigation Techniques

Challenge: The diversity and volume of data were constrained by available resources, affecting the robustness of the model.
Mitigation: Additional data collection efforts were made, but future work should focus on expanding data sources and increasing sample sizes.
Challenge:The controlled conditions under which the study was conducted may not fully represent the variability of real-world farming environments.
Mitigation: Field trials were planned to validate the models in real-world settings, ensuring practical applicability.
Challenge: Integrating the models into a seamless, user-friendly interface required significant technical development and testing.
Mitigation: Iterative design and user feedback were utilized to ensure the interface met the needs of farmers, with plans for ongoing support and improvement.

7.6 Reflection on the project process

7.6.1 Strengths of the Chosen Methodology, Tools, and Techniques

The experimental methodology allowed for controlled manipulation of variables, ensuring precise measurement of their impact on tomato growth. This control is crucial for establishing clear cause-and-effect relationships.It facilitated the systematic testing of different model parameters and configurations, enhancing the reliability of the results and ensuring that the most effective models were identified and optimized.

Utilizing Python and its powerful libraries, such as Scikit-learn for machine learning, Pandas for data manipulation, and Matplotlib for visualization, provided a robust and efficient framework for data analysis and model development.The deployment of the model to a user-friendly interface ensured practical application of the research findings, making the technology accessible to farmers and directly addressing their needs for real-time, actionable insights.

7.6.2 Weaknesses and Challenges

The controlled environment may not fully replicate the variability and complexity of real-world agricultural conditions. This can limit the generalizability of the findings to broader, more diverse contexts.The methodology's focus on precision may overlook broader ecological and environmental factors that also influence tomato growth, potentially simplifying the complexity of agricultural systems.

The reliance on specific software tools may present a barrier to adoption among users unfamiliar with these technologies. Ensuring that the tools are user-friendly and providing adequate training and support is essential. Weakness: While the chosen models were effective, exploring a wider range of machine learning techniques, such as deep learning or ensemble methods, could potentially yield even better predictive performance.

7.6.3 Lessons Learned and Insights Gained

The study highlighted the need for diverse and comprehensive datasets to improve model generalizability. Future projects should prioritize collecting data from various regions, climates, and cultivars to ensure models are robust and applicable across different contexts. The researcher learnt that investing in broader data collection efforts early in the project can significantly enhance the validity and applicability of the findings.

Deploying the model to a user-friendly interface proved crucial for practical application. Ensuring that the technology is accessible and intuitive for end-users, such as farmers, is essential for the successful implementation of research findings.It is important to Involve end-users in the design and testing phases can provide valuable feedback and improve the usability and relevance of the technology. Real-World Validation:

Incorporating real-time NPK readings for fertilizer recommendations significantly enhanced the practical value of the project. Real-time data integration can provide timely and actionable insights for farmers. Prioritizing real-time data integration and ensuring the infrastructure to support it can greatly improve the relevance and impact of predictive models in agriculture.

7.7 Future work and recommendations

- ***Mobile Application Development***

Develop a mobile application that allows farmers to easily access the predictive models and fertilizer recommendations in the field. A mobile app would make the technology more accessible and practical, especially for farmers who may not have access to desktop computers.

- ***Advanced User Interface Features***

Implement advanced features such as interactive dashboards, real-time alerts and visualization tools within the user interface.These features would enhance user experience and provide farmers with more intuitive and actionable insights.

- ***Cloud-Based Data Storage and Processing***

Use cloud-based infrastructure for data storage and processing to handle larger datasets and support scalability.This would ensure that the system can accommodate increasing amounts of data and provide faster processing times.

- ***Generalizing the optimization Problem***

This can be done by incorporating additional variables, developing a dynamic framework, addressing multiple objectives, and ensuring scalability and real-time data integration. These advancements will enhance the model's accuracy, applicability, and sustainability, thereby significantly benefiting precision agriculture and crop management practices.

- ***Validation and Adaptation for Different Crops***

Future work could involve validating and adapting the predictive models and recommendations for various crops in soil-based farming, expanding the technology's applicability and benefiting more farmers. Building on the project's findings and addressing its limitations, there are many opportunities for enhancements. Improvements could include developing a mobile application and advanced user interface features to make the technology more accessible and practical for farmers. Additionally, exploring advanced machine learning models, incorporating remote sensing data, and creating hybrid and time-series models could further improve predictive accuracy and robustness.

Further, extending the project to validate models for different crops, conducting field trials, and evaluating economic and environmental impacts will ensure the broader applicability and sustainability of the technology.It is beneficial to re-run the k-NN model to foster the growth metrics of the following week, and/or for the growth-rate. Customizing and localizing the model for various agricultural contexts will also be crucial in making the technology universally beneficial. These future directions and enhancements will build upon the current project's findings and contribute to more effective and sustainable agricultural practices.

Bibliography

[1] Awais Ali, Tajamul Hussain, Noramon Tantashutikun, Nurda Hussain, and Giacomo Cocetta. Application of smart techniques, internet of things and data mining for resource use efficient and sustainable crop production. *Agriculture*, 13(2):397, 2023.

[2] AA Alsadon, IM Al-Helal, AA Ibrahim, MR Shady, and WA Al-Selwey. Growth analysis of tomato plants in controlled greenhouses. In *XXX International Horticultural Congress IHC2018: III International Symposium on Innovation and New Technologies in Protected 1271*, pages 177–184, 2018.

[3] Abdullah Alsadon, I. Al-Helal, Abdullah Derahim, M.R. Shady, and Wadei Al-Selwey. Growth analysis of tomato plants in controlled greenhouses. *Acta Horticulturae*, pages 177–184, 02 2020.

[4] Kenneth J Boote, Maria R Rybak, Johan MS Scholberg, and James W Jones. Improving the cropgro-tomato model for predicting growth and yield response to temperature. *HortScience*, 47(8):1038–1049, 2012.

[5] Leo Breiman. Statistical modeling: The two cultures. *Statistical Science*, 16(3):199–231, 2019.

[6] Oumnia Ennaji, Leonardus Vergutz, and Achraf El Allali. Machine learning in nutrient management: A review. *Artificial Intelligence in Agriculture*, 2023.

[7] V Gurunathan, J Dhanasekar, S Suganya, et al. Plant leaf diseases detection using knn classifier. In *2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS)*, volume 1, pages 2157–2162. IEEE, 2023.

[8] E Heuvelink and M Dorais. Crop growth and yield. In *Tomatoes*, pages 85–144. Cabi Publishing Wallingford UK, 2005.

[9] K Hossny, S Magdi, Abdelfattah Y Soliman, and Ahmad Hany Hossny. Detecting explosives by pgnaa using knn regressors and decision tree classifier: A proof of concept. *Progress in Nuclear Energy*, 124:103332, 2020.

[10] AS Isah, EB Amans, EC Odion, AA Yusuf, et al. Growth rate and yield of two tomato varieties (lycopersicon esculentum mill) under green manure and npk fertilizer rate samaru northern guinea savanna. *International Journal of Agronomy*, 2014, 2014.

[11] HK Karthikeya, K Sudarshan, and Disha S Shetty. Prediction of agricultural crops using knn algorithm. *Int. J. Innov. Sci. Res. Technol*, 5(5):1422–1424, 2020.

[12] Kelvin López-Aguilar, Adalberto Benavides-Mendoza, Susana González-Morales, Antonio Juárez-Maldonado, Pamela Chiñas-Sánchez, and Alvaro Morelos-Moreno. Artificial neural network modeling of greenhouse tomato yield and aerial dry matter. *Agriculture*, 10(4):97, 2020.

[13] Martin Makgose Maboko et al. *Growth, yield and quality of tomatoes (Lycopersicon esculentum Mill.) and lettuce (Lactuca sativa L.) as affected by gel-polymer soil amendment and irrigation management*. PhD thesis, University of Pretoria, 2005.

[14] Gabriel Mascarenhas Maciel, Guilherme Repeza Marquez, Ernani Clarete da Silva, Vanessa Andaló, and Igor Forigo Belloti. Tomato genotypes with determinate growth and high acylsugar content presenting resistance to spider mite. *Crop Breeding and Applied Biotechnology*, 18:1–8, 2018.

[15] Mahinda Mailagaha Kumbure and Pasi Luukka. A generalized fuzzy k-nearest neighbor regression model based on minkowski distance. *Granular Computing*, 7(3):657–671, 2022.

[16] V Mamatha and JC Kavitha. Machine learning based crop growth management in greenhouse environment using hydroponics farming techniques. *Measurement: Sensors*, 25:100665, 2023.

[17] Nina KACJAN Maršić, Jože Osvald, and Marijana Jakše. Evaluation of ten cultivars of determinate tomato (lycopersicum esculentum mill.), grown under different climatic conditions. *Acta Agriculturae Slovenica*, 85(2):321–328, 2005.

[18] Swathi Nayak, Manisha Bhat, NV Subba Reddy, and B Ashwath Rao. Study of distance metrics on k-nearest neighbor algorithm for star categorization. In *Journal of Physics: Conference Series*, volume 2161, page 012004. IOP Publishing, 2022.

[19] José Ortiz-Bejar, Mario Graff, Eric S. Tellez, Jesús Ortiz-Bejar, and Jaime Cerda Jacobo. k-nearest neighbor regressors optimized by using random search. In *2018 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC)*, pages 1–5, 2018.

[20] Robbi Rahim, Ansari Saleh Ahmar, and Rahmat Hidayat. Cross-validation and validation set methods for choosing k in knn algorithm for healthcare case study. *JINAV: Journal of Information and Visualization*, 3(1):57–61, 2022.

[21] Victor O Sadras, Francisco J Villalobos, and Elias Fereres. Crop development and growth. *Principles of agronomy for sustainable agriculture*, pages 141–158, 2016.

[22] Gary B Shelly and Harry J Rosenblatt. Systems analysis and design nineth edition. *United States of America: Course Technology*, 2012.

[23] Jaskaran Singh and Harpreet Kaur. *Plant Disease Detection Based on Region-Based Segmentation and KNN Classifier*, pages 1667–1675. 01 2019.

[24] Hanqing Sun, Xiaohui Zhang, Zhou Yu, Gang Xi, et al. Feature recognition of crop growth information in precision farming. *Complexity*, 2018, 2018.

[25] Kenichi Tatsumi, Noa Igarashi, and Xiao Mengxue. Prediction of plant-level tomato biomass and yield using machine learning with unmanned aerial vehicle imagery. *Plant Methods*, 17(1):1–17, 2021.

[26] S. K. Tripathy and P. Sivakumar. An efficient crop yield prediction using machine learning algorithms. In *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*, pages 1–4. IEEE, 2020.

[27] Thomas Van Klompenburg, Ayalew Kassahun, and Cagatay Catal. Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, 177:105709, 2020.

[28] Ms Swapnil Verma and Sushopti D Gawade. A machine learning approach for prediction system and analysis of nutrients uptake for better crop growth in the hydroponics system. In *2021 international conference on artificial intelligence and smart systems (ICAIS)*, pages 150–156. IEEE, 2021.

[29] Xiukang Wang, Jia Yun, Peng Shi, Zhanbin Li, Peng Li, and Yingying Xing. Root growth, fruit yield and water use efficiency of greenhouse grown tomato under different irrigation regimes and nitrogen levels. *Journal of Plant Growth Regulation*, 38:400–415, 2019.

Appendices

Appendix A: Templates of data collection tools Appendix B User manual of the working system (Should be detailed) Evidence of research Appendix C: Source Code Appendix D: etc