# 

**SCHOOL OF COMMUNICATION AND INFORMATION TECHNOLOGY.**

**Project Title: Smart Crop and Fertilizer Recommendation System.**

**BY**

**Name: Samuel Gicharu**

**RegNo: G126/1340/2021**

**Date: December 2024.**

**Course: Computer Science.**

# Declaration

This research project is my original work and has never been presented in any other university for academic use.

*Student’s name ………………………………………….*

Reg No. ………………………………………………………..

SIGNED ……………………………DATE ………………………………..

This research project has been submitted for examination with the approval of the university supervisor

SUPERVISOR

NAME: ………………………………………………………………………………..

SIGNED ……………………………DATE…………………………………..

# Dedication

This project is dedicated to my family, friends, and mentors, whose unwavering support and encouragement have been instrumental in the successful completion of this work. Their belief in my capabilities has been a source of inspiration throughout this journey.

# Acknowledgment

I wish to express my heartfelt gratitude to all those who have contributed to the successful completion of this project. Special thanks to my supervisor, \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, for their invaluable guidance, insightful feedback, and encouragement throughout the research and writing process.  
  
I would also like to thank my colleagues and peers for their support and collaboration. Lastly, I am deeply grateful to my family and friends for their continuous encouragement and moral support.

# Abstract

Agriculture remains a vital sector in Kenya, providing livelihoods for the majority of the population and significantly contributing to the economy. However, farmers face challenges such as low crop yields, inefficient resource utilization, and vulnerability to climate change. This study aims to address these issues by developing a machine learning-based precision agriculture system that provides data-driven recommendations for crop selection and fertilizer application.

Traditional farming methods often rely on intuition rather than data, leading to suboptimal land use and economic losses. Many farmers lack access to reliable tools that consider diverse agro-climatic conditions. This project bridges that gap by analyzing key environmental factors, including soil nutrients, weather patterns, and historical crop data, to generate tailored recommendations for specific locations.

The system employs machine learning algorithms such as Random Forest, Logistic Regression, and Support Vector Machines, trained on soil and weather datasets from government and open data platforms. The Random Forest model achieved the highest accuracy of 99%, making it the most effective for crop and fertilizer recommendations. A user-friendly web interface was developed using Python, Scikit-learn, Pandas, and Flask, ensuring accessibility for smallholder farmers.

Findings suggest that data-driven decision-making can significantly improve agricultural productivity, reduce crop failures, and optimize resource use. The study recommends adopting this system for widespread use, particularly in rural areas, and proposes future enhancements such as pest and disease prediction modules and mobile application integration. By leveraging technology and machine learning, this project contributes to a more sustainable and resilient agricultural sector in Kenya.

Keywords: Crop Recommendation, Fertilizer Optimization, Machine Learning, Precision Agriculture, Sustainable Farming, Soil Health, Climate-based Recommendations

Table of Contents

[i](#_Toc189392793)

[Declaration ii](#_Toc189392794)

[Dedication iii](#_Toc189392795)

[Acknowledgment iv](#_Toc189392796)

[Abstract v](#_Toc189392797)

[List of Tables viii](#_Toc189392798)

[List of Figures ix](#_Toc189392799)

[Definition of Terms x](#_Toc189392800)

[List of Abbreviations/Acronyms xi](#_Toc189392801)

[CHAPTER ONE 1](#_Toc189392802)

[INTRODUCTION 1](#_Toc189392803)

[1.1 Background of the Study 1](#_Toc189392804)

[1.2 Statement of the Problem 2](#_Toc189392805)

[1.3 Objectives of the Study 3](#_Toc189392806)

[1.4 Research Questions 5](#_Toc189392807)

[1.5 Purpose of the Study 7](#_Toc189392808)

[1.6 Significance of the Study 8](#_Toc189392809)

[1.7 Limitations of the Study 9](#_Toc189392810)

[1.8 Scope of the Study 11](#_Toc189392811)

[CHAPTER TWO 12](#_Toc189392812)

[LITERATURE REVIEW 12](#_Toc189392813)

[2.1 Introduction 12](#_Toc189392814)

[2.2 Review of Theoretical Literature 13](#_Toc189392815)

[2.2.1 Supervised and Unsupervised Learning 15](#_Toc189392816)

[2.3 Review of Analytical Literature 17](#_Toc189392817)

[2.4 Theoretical/Conceptual Framework 18](#_Toc189392818)

[2.5 Summary and Gaps 19](#_Toc189392819)

[CHAPTER THREE 20](#_Toc189392820)

[RESEARCH DESIGN AND METHODOLOGY 20](#_Toc189392821)

[3.1 Introduction 20](#_Toc189392822)

[3.2 Research Design 21](#_Toc189392823)

[3.3 Target Population 22](#_Toc189392824)

[3.4 Sampling Design 23](#_Toc189392825)

[3.5 Data Collection Procedure/Instruments 24](#_Toc189392826)

[3.6 Data Analysis Methods 25](#_Toc189392827)

# List of Tables

# List of Figures

# Definition of Terms

1. Precision Agriculture: A farming management concept based on observing, measuring, and responding to variability in crops and soil.
2. Machine Learning: A subset of artificial intelligence that enables systems to learn and improve from experience without being explicitly programmed.
3. Crop Recommendation: A system that suggests suitable crops for a given soil and climatic condition.

# List of Abbreviations/Acronyms

1. **AI** – Artificial Intelligence
2. **ANN** – Artificial Neural Network
3. **API** – Application Programming Interface
4. **CNN** – Convolutional Neural Network
5. **CSV** – Comma-Separated Values
6. **FAO** – Food and Agriculture Organization
7. **GIS** – Geographic Information System
8. **GPS** – Global Positioning System
9. **ML** – Machine Learning
10. **MSE** – Mean Squared Error
11. **NDVI** – Normalized Difference Vegetation Index
12. **NPK** – Nitrogen, Phosphorus, and Potassium (Key Soil Nutrients)
13. **pH** – Potential Hydrogen (Soil Acidity/Alkalinity)
14. **RF** – Random Forest
15. **RNN** – Recurrent Neural Network
16. **RMSE** – Root Mean Squared Error
17. **SVM** – Support Vector Machine
18. **WHO** – World Health Organization.

# CHAPTER ONE

## INTRODUCTION

### 1.1 Background of the Study

Agriculture is the backbone of Kenya's economy, employing approximately 75% of the population, either directly or indirectly, and contributing about 33% of the GDP. Despite its importance, the agricultural sector faces numerous challenges that hinder its productivity. Key among these challenges is the lack of scientific and data-driven approaches to farming, leading to suboptimal crop selection, inefficient use of fertilizers, and declining soil health.  
  
In recent years, the increasing effects of climate change, such as unpredictable weather patterns and prolonged droughts, have exacerbated these problems, leaving farmers vulnerable to crop failures and financial instability. Furthermore, many farmers rely on traditional farming practices and intuition, which often fail to account for the diverse climatic and soil conditions found across Kenya.  
  
Advances in machine learning provide a powerful solution to these challenges. Machine learning-based systems can analyze complex data, such as soil nutrients, weather patterns, and historical crop yields, to recommend suitable crops and fertilizers for specific locations. This study explores how such a system can help Kenyan farmers optimize their decisions, increase yields, and promote sustainable agricultural practices.

### 1.2 Statement of the Problem

Kenyan farmers face significant challenges in selecting the right crops and fertilizers for their farms due to a lack of access to reliable agricultural information. This has led to low yields, economic losses, and soil degradation. Current solutions available to farmers, such as extension services, are limited in reach and often fail to address the specific needs of individual farmers, given the country’s diverse ago-climatic zones.  
  
Moreover, there is little integration of scientific data into the decision-making process for crop selection and fertilizer application. Without tailored recommendations based on soil properties, climate, and economic conditions, farmers continue to experience low productivity and increased vulnerability to climate variability. This study addresses the need for a cost-effective, data-driven system to provide personalized agricultural recommendations to Kenyan farmers.

### 1.3 Objectives of the Study

The main objectives of this study are:

1. To develop a machine learning-based crop recommendation system that considers soil and environmental parameters unique to Kenya.
2. To provide fertilizer recommendations based on soil nutrient deficiencies and crop requirements to optimize agricultural practices.
3. To enhance agricultural productivity by reducing crop failures and increasing resource efficiency.
4. To create an easy-to-use interface that enables farmers to access recommendations without requiring technical expertise.

### 1.4 Research Questions

1. What are the key environmental and soil parameters that influence crop selection and fertilizer application in Kenya?
2. How accurately can a machine learning model recommend suitable crops and fertilizers for different regions in Kenya?
3. To what extent can the proposed system improve yields and reduce input wastage compared to traditional farming methods?

### 1.5 Purpose of the Study

The purpose of this study is to create a cost-effective and efficient precision agriculture system that empowers Kenyan farmers to make informed decisions on crop selection and fertilizer application. By leveraging machine learning models and localized agricultural data, the system aims to improve productivity, enhance food security, and contribute to the overall economic growth of the agricultural sector.

### 1.6 Significance of the Study

This study has far-reaching implications for various stakeholders:

* Farmers: The system provides reliable, data-driven insights to reduce crop failures, improve yields, and minimize input costs.
* Agricultural Extension Services: The system complements the work of extension officers by offering personalized recommendations, even in remote areas.
* Policymakers: The study provides a framework for developing agricultural policies and subsidy programs based on scientific data.
* Researchers: The system serves as a foundation for further innovations in precision agriculture, particularly in low-resource settings.
* The Kenyan Economy: By increasing agricultural productivity and ensuring better resource management, the study contributes to national food security and economic stability.

### 1.7 Limitations of the Study

The study relies on the availability and quality of agricultural data, such as soil properties, weather conditions, and crop yields. Challenges may arise from incomplete or outdated datasets, which could affect the system's accuracy. Additionally, while the machine learning model is cost-effective compared to hardware-intensive approaches, its success depends on the willingness of farmers to adopt and use the technology. Internet access and digital literacy may also pose challenges in some rural areas.

### 1.8 Scope of the Study

This study focuses on the development of a machine learning-based crop and fertilizer recommendation system tailored to Kenya's agricultural landscape. The scope includes:

* Analysis of datasets containing soil nutrient information, weather conditions, and historical crop yields.
* Development of machine learning models to predict the most suitable crops and fertilizers for specific regions in Kenya.
* Testing the system with real-world data to ensure its accuracy and reliability.

The study is intended for use by smallholder farmers, agricultural extension officers, and other stakeholders in Kenya's agricultural sector. It does not address pest and disease management or irrigation practices, although these could be considered for future system enhancements.

# CHAPTER TWO

## LITERATURE REVIEW

### 2.1 Introduction

Agriculture is a cornerstone of Kenya’s economy, supporting millions of livelihoods and contributing significantly to GDP. However, the sector faces challenges, including inconsistent yields, poor decision-making due to lack of information, and environmental degradation caused by inefficient farming practices. Machine learning (ML) has emerged as a potential solution to address these challenges by enabling data-driven decision-making.

This chapter reviews theoretical perspectives that provide the foundation for precision agriculture and ML's application in agriculture. It also delves into analytical literature to explore real-world implementations and their results, laying a basis for understanding the theoretical and conceptual framework of this study. Finally, this chapter identifies gaps in existing research and how this study seeks to address them.

The integration of machine learning into agriculture has opened new pathways for addressing age-old challenges in farming, such as low yields, resource inefficiency, and environmental degradation. With the rise of data-driven approaches, modern agriculture is transitioning from traditional practices to smarter, more adaptive systems that leverage advanced algorithms for predictive and prescriptive insights.

Machine learning models have demonstrated remarkable potential in analyzing complex data sets, predicting outcomes, and offering actionable recommendations for farmers. This chapter explores the theoretical underpinnings and analytical findings of studies related to precision agriculture, highlighting key advancements, methodologies, and gaps in current research. Special focus is placed on studies addressing crop recommendation, soil analysis, and yield optimization through machine learning.

While many existing studies rely on IoT integration to enhance real-time data acquisition, this research pivots away from IoT to focus on cost-effective machine learning solutions that are accessible to smallholder farmers, especially in resource-constrained settings like Kenya.

### 2.2 Review of Theoretical Literature

Precision agriculture operates on the principle of optimizing farming practices by managing within-field variability in soils, crops, and environmental conditions. This approach aligns with theories that advocate for scientific farming methods to reduce waste, improve yields, and promote environmental sustainability.

The **Diffusion of Innovations Theory** by Everett Rogers provides a framework for understanding how technologies like machine learning are adopted in farming. According to this theory, innovation adoption depends on factors such as perceived usefulness, complexity, and compatibility with existing practices. In Kenya, the adoption of ML-based systems is influenced by digital literacy and infrastructure availability among farmers.

Machine learning draws on the **Computational Learning Theory**, which studies algorithms designed to make predictions based on data. This study focuses on supervised learning models, including Random Forest, Support Vector Machines (SVM), and Logistic Regression. These models have theoretical underpinnings in statistical learning, which emphasizes the importance of pattern recognition and predictive accuracy for decision-making in uncertain environments.

Theoretical studies suggest that combining machine learning algorithms with precision agriculture practices can create robust systems capable of optimizing agricultural productivity. However, translating these theories into real-world systems remains a challenge, particularly in developing regions.

Machine learning in agriculture is fundamentally grounded in the principles of computational learning and precision agriculture. These theoretical perspectives provide the basis for designing models that can address challenges faced by farmers.

**Precision Agriculture Theory**  
Precision agriculture advocates for the precise management of farming resources, such as soil nutrients, water, and fertilizers, based on site-specific data. This approach seeks to minimize wastage while maximizing productivity by tailoring agricultural practices to the unique conditions of individual plots of land. The theory emphasizes the need for accurate data collection and robust analysis tools to guide decision-making, making machine learning a natural fit for such systems.

**Computational Learning Theory**  
At the heart of machine learning lies computational learning theory, which examines how algorithms learn from data to make predictions or decisions. Supervised learning models, such as Support Vector Machines (SVM), Random Forests, and Logistic Regression, play a central role in this study. These algorithms analyze historical data to identify patterns and relationships that inform future recommendations, such as crop suitability or fertilizer application.

**Applications of Supervised Learning Models**

* **Support Vector Machines (SVM)**: Widely recognized for their accuracy in classification problems, SVMs have been used in studies, where a modified SVM algorithm processes real-time soil data to predict suitable crops. By mapping input features (soil properties, climatic conditions) to higher-dimensional spaces, SVMs can effectively separate classes and make precise predictions.
* **Random Forests**: These ensemble models, comprising multiple decision trees, excel in handling diverse datasets. They offer high predictive accuracy by reducing overfitting and are particularly effective in crop yield prediction and soil classification tasks.
* **Logistic Regression**: Known for its simplicity and interpretability, logistic regression models have been employed to analyze binary outcomes, such as determining whether a specific crop is suitable for a given soil type.

**Integration with Data-Driven Decision Making**  
Theoretical frameworks also emphasize the role of machine learning in transitioning from intuition-based farming to data-driven practices. By integrating datasets that capture environmental, geographical, and historical variables, machine learning enables farmers to make informed decisions. This aligns with the broader goals of sustainable agriculture, which seek to balance productivity with ecological preservation.

#### 2.2.1 Supervised and Unsupervised Learning

Machine learning techniques can be broadly categorized into the following types: Supervised learning takes a set of feature/label pairs, called the training set. From this training set the system creates a generalized model of the relationship between the set of descriptive features and the target features in the form of a program that contains a set of rules. The objective is to use the output program produced to predict the label for a previously unseen, unlabelled input set of features, i.e. to predict the outcome for some new data. Data with known labels, which have not been included in the training set, are classified by the generated model and the results are compared to the known labels. This dataset is called the test set. The accuracy of the predictive model can then be calculated as the proportion of the correct predictions the model labeled out of the total number of instances in the test set.

Unsupervised learning takes a dataset of descriptive features without labels as a training set. In unsupervised learning, the algorithms are left to themselves to discover interesting structures in the data. The goal now is to create a model that finds some hidden structure in the dataset, such as natural clusters or associations. Unsupervised learning studies how systems can infer a function to describe a hidden structure from unlabeled data. The system does not figure out the right output, but it explores the data and can draw inferences from datasets to describe hidden structures from unlabeled data. Unsupervised learning can be used for clustering, which is used to discover any inherent grouping that are already present in the data. It can also be used for association problems, by creating rules based on the data and finding relationships or associations between them.

Semi-supervised machine learning falls somewhere in between supervised and unsupervised learning, since they use both labeled and unlabeled data for training typically a small amount of labeled data and a large amount of unlabeled data. The systems that use this method are able to considerably improve learning accuracy. Usually, semi-supervised learning is chosen when the acquired labeled data requires skilled and relevant resources in order to train it / learn from it. Otherwise, acquiring labeled data generally does not require additional resources.

Reinforcement machine learning algorithms is a learning method that interacts with its environment by producing actions and discovers errors or rewards. Machine learning algorithms are tools to automatically make decisions from data in order to achieve some overarching goal or requirement. The promise of machine learning is that it can solve complex problems automatically, faster and more accurately than a manually specified solution, and at a larger scale. Over the past few decades, many machine learning algorithms have been developed by researchers, and new ones continue to emerge and old ones modified.

### 2.3 Review of Analytical Literature

The practical application of machine learning in agriculture has gained significant attention globally, and its potential to transform farming practices in Kenya is immense. Studies have demonstrated that leveraging data on soil properties, weather patterns, and historical yields can lead to significant improvements in efficiency and productivity. For instance, integrating machine learning models with readily available agricultural data has been shown to enhance decision-making, enabling farmers to optimize crop selection and improve resource utilization.

In the Kenyan context, farming practices are shaped by unique factors such as diverse agro-climatic zones, land tenure systems, access to markets, and traditional agricultural knowledge. While these factors provide a rich foundation for innovation, they also introduce complexities that many existing studies fail to address. Current research often focuses narrowly on specific parameters, such as soil pH or rainfall, without considering the holistic nature of Kenyan agriculture, where economic, cultural, and environmental factors intersect.

Moreover, many of the models developed rely on expensive technological infrastructures that are inaccessible to the majority of smallholder farmers in Kenya. With limited access to advanced hardware or real-time IoT-enabled systems, farmers in rural areas often face barriers to adopting such technologies. This study seeks to bridge these gaps by developing a cost-effective, accessible machine learning system that leverages locally available data and addresses the diverse needs of Kenyan farmers.

By focusing on affordability and practicality, the proposed system aims to empower farmers to make data-driven decisions, improving yields and promoting sustainable agricultural practices tailored to Kenya’s unique agricultural landscape.

### 2.4 Theoretical/Conceptual Framework

This study integrates theoretical principles and practical methodologies to develop a machine learning-based system for precision agriculture. The framework is structured around three main components:

1. **Input Variables**:  
   These include soil nutrient profiles, climatic conditions (temperature, rainfall, humidity), and historical crop yield data.
2. **Processing Layer**:  
   Machine learning models analyze the data to make predictions and provide recommendations. Random Forest is the primary model, selected for its robustness and high predictive accuracy.
3. **Output Layer**:  
   Recommendations are delivered to farmers in an easy-to-understand format, detailing the most suitable crops and fertilizers for specific conditions.

This framework emphasizes accessibility, accuracy, and practical implementation, ensuring that farmers can use the recommendations to improve their yields while conserving resources.

### 2.5 Summary and Gaps

While existing studies demonstrate the potential of precision agriculture, their narrow scope and high technological requirements limit their applicability in Kenya. Most research has not addressed the need for context-specific solutions that integrate diverse variables and consider local socio-economic conditions. This study fills these gaps by proposing an affordable, user-friendly machine learning system tailored to Kenyan farmers' needs.

# CHAPTER THREE

## RESEARCH DESIGN AND METHODOLOGY

### 3.1 Introduction

This chapter outlines the research design and methodology used to develop and validate the proposed precision agriculture system. It describes the research design, target population, sampling techniques, data collection procedures, and analytical methods employed in the study.

### 3.2 Research Design

The study employs a **descriptive research design**, blending qualitative and quantitative approaches. This design facilitates an in-depth understanding of the current challenges faced by Kenyan farmers and enables the development of solutions grounded in empirical data. A descriptive design is appropriate because it supports both the exploration of patterns and the testing of hypotheses.

### 3.3 Target Population

The primary target population includes smallholder farmers across Kenya, especially in regions with varying agro-climatic conditions. These farmers are often reliant on rain-fed agriculture and face challenges due to resource constraints and climate variability. Secondary stakeholders include agricultural extension officers and policymakers, who play a crucial role in promoting and implementing precision agriculture technologies.

### 3.4 Sampling Design

**Purposive sampling** is employed to ensure that the study captures diverse agricultural practices across Kenya. Representative samples are drawn from high-potential farming regions such as the Rift Valley and semi-arid zones like Eastern Kenya. The sample includes farmers with varying levels of access to technology and agricultural resources, ensuring the model is applicable to a wide audience.

### 3.5 Data Collection Procedure/Instruments

Data collection is conducted using both primary and secondary methods:

1. **Primary Data**:
   * **Surveys and Interviews**: Structured questionnaires and interviews with farmers and extension officers to gather insights on farming challenges and practices.
   * **Focus Groups**: Discussions with small groups of farmers to understand local needs and barriers to technology adoption.
2. **Secondary Data**:
   * Data on soil properties, weather patterns, and historical crop yields sourced from government databases and open-source platforms.
   * Remote sensing data and satellite imagery for validating climatic conditions.

### 3.6 Data Analysis Methods

Data analysis involves both statistical and computational techniques:

* **Descriptive Statistics**: Summarize trends and challenges in the agricultural data.
* **Machine Learning Models**: Algorithms such as Random Forest, SVM, and Logistic Regression are trained and tested on the collected data. Model performance is evaluated using metrics like accuracy, precision, and recall.
* **Visualization**: Findings are presented using charts and graphs to enhance comprehensibility.

By integrating these methods, the study ensures robust and actionable insights tailored to the needs of Kenyan farmers.