

**KARATINA UNIVERSITY**

SCHOOL OF PURE AND APPLIED SCIENCES

DEPARTMENT OF COMPUTER SCIENCE AND INFORMATICS

INTELLIGENT CROP RECOMMENDATION MODEL USING MACHINE LEARNING ALGORITHMS

submitted in partial fulfillment for the award of the degree

BACHELOR OF SCIENCE

in

COMPUTER SCIENCE

By

Benedict Omusindobole

P101/0869G/18

DATE………………………

Under the Guidance of

Ms. Vancy Kebut

**DECLARATION**

I, Benedict Omusindobole(P101/0869G/18), student of Karatina university, hereby declare that the project entitled, ”Intelligent Crop Recommendation Model Using Machine Learning” has been carried out by myself under the guidance of Ms. Vancy K, in partial fulfillment of the requirements for the award of the degree of Bachelor of Science in Computer Science, of the Karatina University, during the academic year 2018-2022. The work done in this project report is original and it has not been submitted for any other degree in any university.

Declaration by student

BENEDICT OMUSI NDOBOLE

Signature…………………. Date…………………………

Declaration by supervisor

MRS. VANCY KEBUT

Signature…………………. Date………………………….

# DEDICATION

I dedicate my project work to my family and many friends. A special feeling of gratitude to my loving parents, George and Becky whose words of encouragement and push for tenacity ring in my ears. My sister Vivian has never left my side and is very special. I also dedicate this work to my many friends and church family who have supported me throughout the process. I will always appreciate all they have done and lastly my colleagues and my supervisor for creating an enabling environment to carry out this project.

# ACKNOWLEDGEMENT

I would like to acknowledge the Almighty God for without His continued love, support and mercy I would not have had the courage or strength to undertake this particular project proposal. I would also like to acknowledge my lecturer, Mrs. Kebut for the moral support and guidance that he offered for the success of my project proposal. Lastly, I would like to acknowledge the Karatina University fraternity, more so the ICT and library department for provision of resources necessary for success of my research.

# ABSTRACT

Choosing the right type of crop to plant in a farm is very important. This is because different areas have different soil properties which in turn keeps changing every season due to the different inputs farmers keeps adding to their farms as a way of increasing productivity. Farmers thus, find it hard to decide on which crop to plant and when, based on the changing climatic factors and soil properties. This research project aims at developing a recommendation model for discovering the type of crop a farmer can plant using research data of soil characteristics such as Ph of the soil and natural factors such as temperature.

The development of the model will use the crop natural datasets which contains the required features to help in achieving the accuracy. The model works by using the historical datasets of crops recorded. The datasets have been collected from various agricultural data hubs which provides relevant datasets which can be used in deriving patterns . The data preprocessing is done to prepare the datasets ready for training the model which is later used in prediction by recommending the type of crop suitable for a particular farm field based on the environmental factors. All these processes are to be achieved using the following machine learning techniques: Support Vector Machine, K-Nearest Neighbors and Random Forest. Random Forest techniques is later selected as the main algorithm due to its higher performance accuracy than the other two algorithms that is Support Vector Machine and K-Nearest Neighbors.

Contents

[DEDICATION ii](#_Toc106882333)

[ACKNOWLEDGEMENT iii](#_Toc106882334)

[ABSTRACT iv](#_Toc106882335)

[CHAPTER ONE: INTRODUCTION 1](#_Toc106882336)

[1.0 INTRODUCTION 1](#_Toc106882337)

[1.2 BACKGROUND OF THE STUDY 1](#_Toc106882338)

[1.3 PROBLEM STATEMENT 2](#_Toc106882339)

[1.4 PROJECT SCOPE 2](#_Toc106882340)

[1.5 JUSTIFICATION OF STUDY 2](#_Toc106882341)

[1.6 OBJECTIVES OF STUDY 2](#_Toc106882342)

[1.7 GENERAL OBJECTIVES 2](#_Toc106882343)

[1.8 SPECIFIC OBJECTIVES 2](#_Toc106882344)

[1.9 LIMITATION OF STUDY 3](#_Toc106882345)

[1.10 PROJECT RISKS AND MITIGATIONS 3](#_Toc106882346)

[CHAPTER TWO: LITERATURE REVIEW 4](#_Toc106882347)

[INTRODUCTION 4](#_Toc106882348)

[2.1 CURRENT RELATED SYSTEMS: 4](#_Toc106882349)

[2.2 CRITIQUE OF THE LITERATURE REVIEW 5](#_Toc106882350)

[2.3 RESEARCH GAP 5](#_Toc106882351)

[2.4 PROPOSED MODEL 5](#_Toc106882352)

[CHAPTER THREE: METHODLOGY 6](#_Toc106882353)

[3.0. INTRODUCTION 6](#_Toc106882354)

[3.1. ITERATIVE WATERFALL MODEL 6](#_Toc106882355)

[**3.1.1.** Feasibility Study 7](#_Toc106882356)

[3.1.2. Requirement Analysis and Specification 7](#_Toc106882357)

[3.1.3 Design 7](#_Toc106882358)

[3.1.4 Coding and Unit Testing 7](#_Toc106882359)

[3.1.5 Integration and System Testing 7](#_Toc106882360)

[3.1.6 Maintenance 7](#_Toc106882361)

[3.2 TESTING AND IMPLEMENTATION TOOLS 7](#_Toc106882362)

[3.2 .0 Software Requirements; 7](#_Toc106882363)

[3.2 .1 Hardware Requirements; 8](#_Toc106882364)

[3.2 .2 Programming Language Requirements; 8](#_Toc106882365)

[3.2 .3 Library Requirements; 8](#_Toc106882366)

[3.3. PROJECT TIMELINE 9](#_Toc106882367)

[3.4. PROJECT COSTS 9](#_Toc106882368)

[**CHAPTER FOUR:** SYSTEM ANALYSIS AND REQUIREMENT MODELLING 10](#_Toc106882369)

[4.0 INTRODUCTION 10](#_Toc106882370)

[4.1 STUDY OF THE CURRENT SYSTEM 10](#_Toc106882371)

[4.2 REQUIREMENT DEFINITIONS OF DEVELOPED SYSTEM 10](#_Toc106882372)

[4.2.1 Functional requirements 10](#_Toc106882373)

[4.2.2 Non-functional Requirements 11](#_Toc106882374)

[4.2.3 User requirements 11](#_Toc106882375)

[4.2.4 Business drivers 11](#_Toc106882376)

[4.2.5 Acceptance Criteria 12](#_Toc106882377)

[**4.5** DFD – Data Flow Diagram 14](#_Toc106882378)

[CHAPTER FIVE: SYSTEM DESIGN 17](#_Toc106882379)

[5.0. INTRODUCTION 17](#_Toc106882380)

[5.1. SYSTEM ARCHITECTURE 17](#_Toc106882381)

[5.1.1 Datastore 17](#_Toc106882382)

[5.1.2 Data preprocessing 17](#_Toc106882383)

[5.1.3 Application of Machine Learning Algorithms 17](#_Toc106882384)

[**5.2. SYSTEM MODULES AND INTERFACE** 19](#_Toc106882385)

[**5.1 Home Page** 19](#_Toc106882386)

[**5.3.2 Crop Inputs Page Without Values** 20](#_Toc106882387)

[**5.3.4 Crop Inputs Page With Values** 20](#_Toc106882388)

[**5.3.5 Crop Prediction Page** 21](#_Toc106882389)

[CHAPTER SIX: SYSTEM IMPLEMENTATION 23](#_Toc106882390)

[6.0. INTRODUCTION 23](#_Toc106882391)

[6.2. PLAN OF IMPLEMENTATION 24](#_Toc106882392)

[**6.3. MACHINE LEARNING MODELS** 25](#_Toc106882393)

[6.3.1. RANDOM FOREST 25](#_Toc106882394)

[6.3.2. SUPPORT VECTOR MACHINE 26](#_Toc106882395)

[6.3.1. K-NEAREST NEIGHBORS 26](#_Toc106882396)

[6.4. CODING AND TESTING TOOLS 26](#_Toc106882397)

[6.4.1. Visual Studio Code 26](#_Toc106882398)

[6.4.2. GitHub 27](#_Toc106882399)

[6.4.3. Chrome Browser 27](#_Toc106882400)

[6.5. SYSTEM TEST PLANS 27](#_Toc106882401)

[6.5.1. Goal of the Test Plan 27](#_Toc106882402)

[6.6 Features to be examined 27](#_Toc106882403)

[6.7 Test Environments 28](#_Toc106882404)

[6.8 Chrome Web browser. 28](#_Toc106882405)

[6.9 Pass Criteria 28](#_Toc106882406)

[6.10 TESTING APPROACHES 28](#_Toc106882407)

[6.10.1 Unit Testing 28](#_Toc106882408)

[6.11 Component Testing 29](#_Toc106882409)

[6.12 Integration Testing 29](#_Toc106882410)

[6.13 CHANGE – OVER TECHNIQUE 29](#_Toc106882411)

[6.13.1 Parallel Changeover 29](#_Toc106882412)

[6.13.2 DATASET USED 30](#_Toc106882413)

[CHAPTER SEVEN 32](#_Toc106882414)

[7.0. LIMITATIONS, CONCLUSIONS AND RECOMMENDATIONS 32](#_Toc106882415)

[7.1. INTRODUCTION 32](#_Toc106882416)

[7.2. LIMITATIONS 32](#_Toc106882417)

[7.3. CONCLUSION 33](#_Toc106882418)

[7.4. RECOMMENDATIONS 33](#_Toc106882419)

[Table 1 Project risks and mitigations 3](#_Toc106882420)

[Table 2 Project Timeline 9](#_Toc106882421)

[Table 3 Project Cost 9](#_Toc106882422)

[Figure 1 Iterative waterfall methodology 6](#_Toc106882423)

[figure 2: Activity diagram 13](#_Toc106882424)

[Figure 3 Activity diagram 14](#_Toc106882425)

[Figure 4: Data flow diagram 14](#_Toc106882426)

[Figure 5: Use case diagram 15](#_Toc106882427)

[Figure 6: Sequence diagram 16](#_Toc106882428)

[Figure 7: system architecture 18](#_Toc106882429)

[Figure 8: Home page 19](#_Toc106882430)

[Figure 9: crop input page 20](#_Toc106882431)

[Figure 10: crop input page 20](#_Toc106882432)

[Figure 11: crop prediction 21](#_Toc106882433)

[Figure 12:steps during Preprocessing 24](#_Toc106882434)

[Figure 13:Working processesof trained model 25](#_Toc106882435)

[Figure 14:crop Recommendation data set 30](#_Toc106882436)

[Figure 15:crop soil properties of N,P,K 31](#_Toc106882437)

# CHAPTER ONE: INTRODUCTION

## 1.0 INTRODUCTION

The recommendation model is a model that aims at making the right type of crop for farmers in their respective farms which ensures high crop yield production. The recommendation model by the help of machine learning techniques is meant to deal with the climatic factors and soil type properties which are the main determinant for the crop yield in farming. All farmers will be in a position to determine which crop is suitable for the farm at the time of crop cultivation. Therefore, resulting in high crop yield production.

## 1.2 BACKGROUND OF THE STUDY

Crop cultivation in Kenya is Key. As a developing country, farmers need more support to maintain the yield production in the market and to the Kenyan citizens. However, in the process of crop farming, natural factors have been found to be the most undermining problem which at large lowers the crop production yield.

The government of Kenya has tried at large to support the crop farming system even financial leaving behind the natural factors. This is the reason behind low productivity since farmers at their own find it hard to deal with these natural factors.

Understanding the ph. of the farm soil and at large the whole properties of the far, soil is essential as this helps in determine which crop best suites the soil. In addition, farmers find it hard to understand the most suitable farm input to apply in their farm. An example is the use of fertilizer in the farms. Farmers keeps on mixing fertilizer in out on every season without knowing the properties of the farm soil so that they can understand which best fertilizer best fits their soil based on factors such as soil ph.

Getting the right type of crop for the farm in a big deal. Farmers out of knowledge are advised wrongly while buying the crop seeds. Thus, the crop types end ups yield low in its production, hence a loss to the farmer.

## 1.3 PROBLEM STATEMENT

Farmers in Kenya should be honored for their great work indeed. They help to feed total of around 44million Kenyans. . However, their productivity is threatened by some natural factors that can ruin their crops and their livelihoods. The proposed classification model is a small initiative that aims to enhance agriculture, make smart decisions to consider the demographics of their field, the factors affecting their crop, for a super awesome yield.

## 1.4 PROJECT SCOPE

The basic purpose of this project is to develop a recommendation model which is a small initiative enhancing agriculture, and help farmers to make smart decisions to consider the demographics of their field, the factors affecting their crop for a super awesome yield.

## 1.5 JUSTIFICATION OF STUDY

The model will enhance the process of solving farmers problem thus, minimizing time, cost and effort.

The model will help Kenyan farmers be in a position to find out the natural factors that affect crop yield production.

## 1.6 OBJECTIVES OF STUDY

## 1.7 GENERAL OBJECTIVES

The aim of this study is to develop a recommendation model that provide recommendation of the best suitable crops in the area so that the farmer does not incur any losses.

## 1.8 SPECIFIC OBJECTIVES

1. To study and evaluate the three aspects of feasibility study which are the operational, economic and technical part of the application.

1. To analyze and specify the requirements of the overall applications and analyze them.
2. To design a functional recommendation model based on the analysis done on the system requirements.
3. To develop the proposed model and as well as perform the unit testing.
4. To integrate the tested system units and perform the system testing(testing the overall system).
5. To maintain the system developed by finding out errors and bugs and correcting them for effectiveness of the overall system.

# 1.9 LIMITATION OF STUDY

It is obvious that, a study of this nature would have some constraints such as money, rich datasets collection method and time. Lack of time is an obstacle that hindered a wider research on the project. Also, the method of collecting rich datasets was less rigid.

## 1.10 PROJECT RISKS AND MITIGATIONS

Even though this project seeks to assist farmers to minimize their own time spent in solving farm problem, it’s associated to some risks as listed below:

|  |  |
| --- | --- |
| RISK | MITIGATION |
| Project complexity such that the features are developed using quite number of technologies. Therefore, complexity in integration till the application operates | Use of Iterative method will allow more time for testing. |

Table 1 Project risks and mitigations

# CHAPTER TWO: LITERATURE REVIEW

## INTRODUCTION

This chapter includes a review of the current systems that also try to solve the farmers’ problems outlined using various methods. It tries to show how current systems have tried to solve the problem outlined and their shortcomings and it also narrows down to various interest and concerns of previous researchers at local and international arena on the subject of the study. It will be based on the project objectives.

## 2.1 CURRENT RELATED SYSTEMS:

(Rakesh K. , et. Al., 2015) in “Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique” paper proposed a method named Crop Selection Method (CSM). This method was established to solve crop selection problem, and maximize net yield rate of crop over season and subsequently achieves maximum economic growth of the country. The proposed method may improve net yield rate of crops.

However, (Haedong L. , et. Al., 2014), did a research that resulted in development of an effective agricultural yield forecasting system based on real-time monthly weather. Since it is difficult to predict the agricultural crop production because of the abnormal weather that happens every year and rapid regional climate change due to global warming, the development of agricultural yield forecasting system that leverages real-time weather information is urgently required. In this research, we cover how to process the number of weather data(monthly, daily) and how to configure the prediction system. We establish a non-parametric statistical model on the basis of 33 years of agricultural weather information. According to the implemented model, we predict final production using the monthly weather information. This paper contains the results of the simulation.

In addition, in the paper of “Analysis of Soil Behaviour and Prediction of Crop Yield using Data Mining Approach”, a system is developed. This system uses data mining techniques in order to predict the category of the analyzed soil datasets. The category, thus predicted will indicate the yielding of crops. The problem of predicting the crop yield is formalized as a classification rule, where Naive Bayes and K-means methods are used.( Hakkim, V. A., et. Al., 2020).

## 2.2 CRITIQUE OF THE LITERATURE REVIEW

From the reviewed literature, different mechanism has been established as a way of helping the farmer in crop type selection so that they can improve their crop yield during harvest time. Though, most of the mechanisms haven’t been found effective. Example the invention of Crop Selection Method was made with an aim of predicting the crop type. However, considering the processes involved to get at the point of predicting is not effective enough especially in our changing world. Therefore, I proposed this project as a recommendation model is built using the approved and realized machine learning techniques which works well in making the correct predictions based on the historical datasets of the crops.

## 2.3 RESEARCH GAP

My approach differs from all previous approaches in that it focuses on developing an intelligent crop recommendation model. Following the literature review above, the following are some of the gaps that have been identified as a result of reviewing. With the reviewed system for predicting weather;

Small and poor farmers are not able to decide which type of crop they are best suitable for their farm based on the natural environmental factors like soil properties of the farm.

Lack of knowledge among farmers with regards to maximizing crop yield and crop rotation techniques.

# 2.4 PROPOSED MODEL

The main focus of this project is to develop a recommendation model that will focus on solving farm natural problems. Unlike the current systems, farmers will be able to determine which type of crop best suites their soil.

In summary, this proposed project was to be helpful for farmers in sorting out their crop yield farm issues with easy than they have been used to. Through the proposed classification model, farmers increase their crop yield production since they can control the natural environmental factors which are the main determinant when growing crops.

# CHAPTER THREE: METHODLOGY

## 3.0. INTRODUCTION

## 3.1. ITERATIVE WATERFALL MODEL

After proper research of the project in hand, it was decided that the iterative waterfall methodology would be adopted. The methodology was best suited due to the short-term period of the project (Alshamrani & Bahattab, 2017) . It allowed the developer to go back to earlier stages to correct undetected errors after releasing the product .

According to (Shylesh, 2017), the chosen methodology will allow for effective control of the entire management process such as effective decision making and problem solving. It will also ensure success of specific processes, approaches and technologies applied. The project life cycle is laid out in successive stages as shown below:

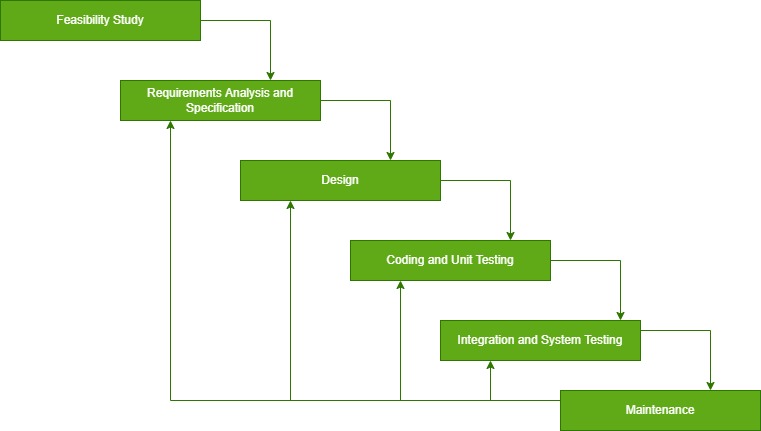


Figure 1 Iterative waterfall methodology

According to (Liu & Wei, 2018), the following are the descriptions of the iterative waterfall methodology;

### **3.1.1.** Feasibility Study

During this stage, the various operations of the system were evaluated and analyzed for new improvement to be done.

### 3.1.2. Requirement Analysis and Specification

As for my project, the features, functions, and tasks that I completed for a it to be deemed successful were both functional and non-functional. The gathering and grouping the requirements were guided by the scope of the project.

#### 3.1.2.1. Functional requirement

After the completion of my model, I tested it and found it to be able to analyze and display the crop types based on the environmental conditions inputted by the user.

#### 3.1.2.2. Non-functional

The model was able to receive the user inputs through application interface and pre-process them ready for analysis giving back the responses to the user.

### 3.1.3 Design

The structure of the system in terms of input, output, database was formulated in this phase to fulfil the gaps in the current system.

### 3.1.4 Coding and Unit Testing

During testing, the system was analyzed to check if it conforms to the requirements it was designed for.

### 3.1.5 Integration and System Testing

During this stage, the system was now put up on the user’s site for operation.

### 3.1.6 Maintenance

This is the final stage whereby bugs and errors found by the users during testing are to be fixed up.

## 3.2 TESTING AND IMPLEMENTATION TOOLS

### 3.2 .0 Software Requirements;

**Google Colab –** provides a platform for executing python code through the browser.

**Google Chrome –** This browser was handy in accessing online research whenever a consultation was needed. As well used in testing the functional requirements of the online web chatting application.

**Operating system -** windows 10 and above is needed in the development of this online web chatting system.

**Visual Studio –** this is the official IDE for developing different kind of online web applications. Has got all features required in development and testing of the developed application before release to the market.

### 3.2 .1 Hardware Requirements;

**Laptop –** This is the main device that was used to code the whole application and store the application modules for proper execution of the project during and after development.

**Core i5:** 2.4 GHz processor and above

**Hard drive:** 40 GB

**Monitor:** 15 VGA color monitor

**Ram:** 4gb and above

### 3.2 .2 Programming Language Requirements;

**Python -** is a scripting programming language to be used in development of the proposed application.

**JavaScript -** client-side language for adding interactivity to the web pages.

### 3.2 .3 Library Requirements;

**Scikit Learn –** machine learning library used in training and testing of the model developed.

**Flask -** backend framework used in developing web servers to serve the client side of an application.

**Pandas -** isa python library open source tool used for data analysis and manipulation.

**Numpy -** isa python library open source tool that adds support for large multi-dimensional arrays and matrices along with large collection of high-level math functions that operates on these arrays.

## 3.3. PROJECT TIMELINE

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Weeks | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 13 | 14 |
| Feasibility study |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Requirement analysis & specification |  |  |  |  |  |  |  |  |  |  |  |  |  |
| System design |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Coding and unit testing |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Integration and system testing |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Maintenance |  |  |  |  |  |  |  |  |  |  |  |  |  |

Table 2 Project Timeline

## 3.4. PROJECT COSTS

Having identified this project, the following resources were identified for its completion and a budget estimation laid down as below

|  |  |  |  |
| --- | --- | --- | --- |
| NUMBER | TOOL/ITEM | Approximate number | Approximate Cost |
| 1 | NOTEBOOK | 1 | 200 |
|  |  |  |  |
| 2 | INTERNET | (recommended) | 6500 |
|  |  | TOTAL COST | 6,700 |
|  |  |  |  |

Table 3 Project Cost

# **CHAPTER FOUR****: SYSTEM ANALYSIS AND REQUIREMENT MODELLING**

## 4.0 INTRODUCTION

This chapter serves the purpose of describing the working of current system using tools such as data flow diagrams, unified modelling language, flowcharts and use case diagrams. Requirement definition as well as specifications of the current and developed system are also narrated in this chapter.

## 4.1 STUDY OF THE CURRENT SYSTEM

The study of the current system was out on a model predicting the crop type using either soil properties only or climatic conditions only to perform the predication. The current system only focuses on one side but not both to perform the predication. The data are then analyzed using one of the proposed technique Crop Selection Method (CSM). They are then preprocessed before analysis are done on them. The cleaned data are uploaded on the model to be worked on.

## 4.2 REQUIREMENT DEFINITIONS OF DEVELOPED SYSTEM

According to the gathered information, a number of requirements had to be formulated to show how the design as well as development process had been undertaken through the development stages. The requirement formulated were properly analyzed to a certain level in that they contained all tasks needed for proper functioning of the proposed system. They were also analyzed to ensure that there were no conflicts among the expectation of the system and even the customer together with the stakeholders.

These requirements are categorized into various groups such as: business drivers, user requirements, functional requirements, non-functional requirements, acceptance criteria as well as system requirements.

### 4.2.1 Functional requirements

They are meant to specify the behavior of the developed system and the functions that it was designed to perform.

**The desired functionalities are**:

1. Allow users/farmers to feed in the measured/tested values of the different parameters/factors
2. The users/farmers should be able to submit the feed in values into the model system to be worked on by the recommendation model developed.
3. The model should be able to determine work on the requested values from the users .
4. The model should give back the response to the user via the system client side as the result recommended crop.

### 4.2.2 Non-functional Requirements

1. All data set used by the model should be secure from physical access or from any form of modification to get accuracy in prediction.
2. The system must be designed with user friendly interface and easy to use.
3. The system should have minimal surprise in its behavior throughout its operation time.
4. The system should be easy to maintain and allow for easier update.
5. Fast response by the system upon generating responses back to the user.

### 4.2.3 User requirements

For the effectiveness of the system to be realized, it is critical to specify how the user should interact with it for efficient operation. Users’ participation is the key to success of the system hence dictates the needs for laying out all the actions that are required from their side.

### 4.2.4 Business drivers

These are brief descriptions of why the system is being developed. This brings out the rationale for the proposed system and the motivation behind the need for its existence. These drivers guide the decision making concerning the development, analysis as well as formulation of the system architecture.

They include some of the following:

1. The system should provide text field options where the user types the values of the parameters.
2. The system should provide an interface where results are displayed for viewing. The system to provide both text and image-based response for easy understanding of the user.

### 4.2.5 Acceptance Criteria

The requirement of the system to be accepted by the user and deployed to undertake the operations include:

1. All user acceptance should be successfully passed by the system.
2. All errors should be corrected before the system is deployed.

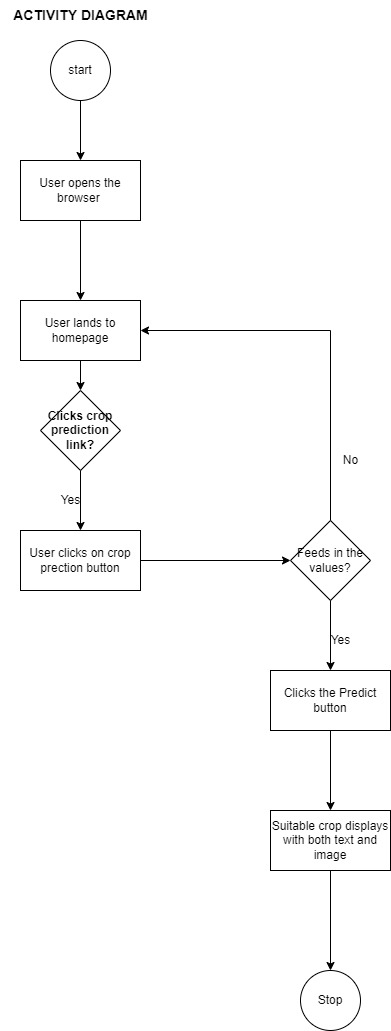


figure 2: Activity diagram

Figure 3 Activity diagram

## **4.5** DFD – Data Flow Diagram

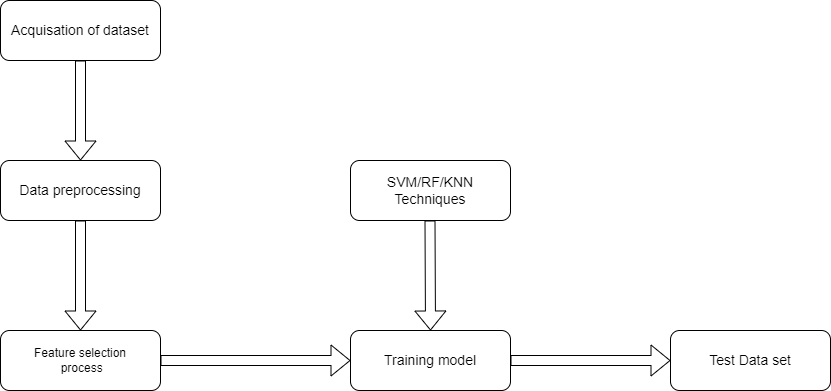


Figure 4: Data flow diagram

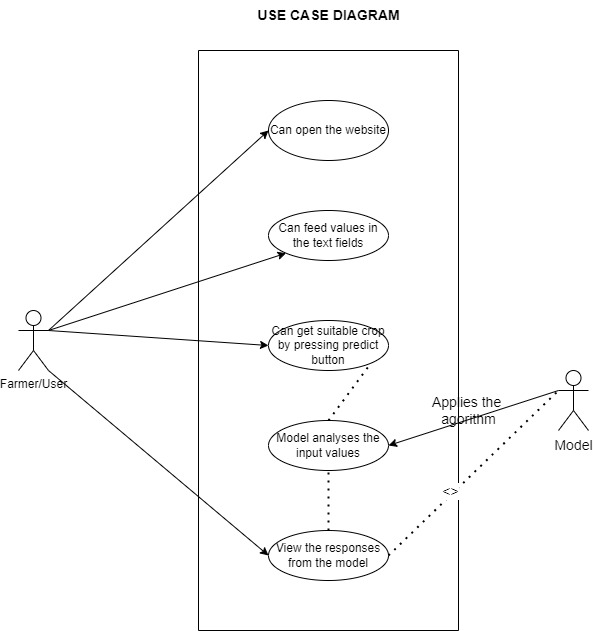


Figure 5: Use case diagram

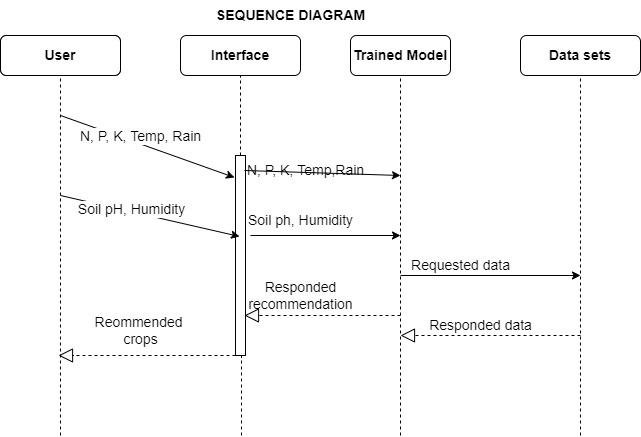


Figure 6: Sequence diagram

# CHAPTER FIVE: SYSTEM DESIGN

## 5.0. INTRODUCTION

This chapter expounds an overview of how the entire system has been designed. The system's architecture, its modules, security and integrity controls, interfaces as well as roles and responsibilities are explained and supplemented by the use of appropriate modelling tools in each scenario. The design process of the whole application was guided by the requirements specifications that were identified prior to the analysis phase. The design guidelines help with the implementation of the system as well as maintenance of the system long after it has been implemented.

## 5.1. SYSTEM ARCHITECTURE

The system is structured in terms of functional components that interact with each other by the use of pre-defined interfaces to achieve the main objective. The behavior of these components brings about general functionality as well as the reliability of the required system. These components include:

### 5.1.1 Datastore

File with relevant datasets provides data in different in the right format to be used in the training and testing of the model that is used in recommending the crop to plant.

### 5.1.2 Data preprocessing

This component is made to perform all forms of data cleaning including fixing missing values, outlier detection to make the data clean and ready for training.

### 5.1.3 Application of Machine Learning Algorithms

This component is made up of all techniques required in training the model based on the area of interestingness.

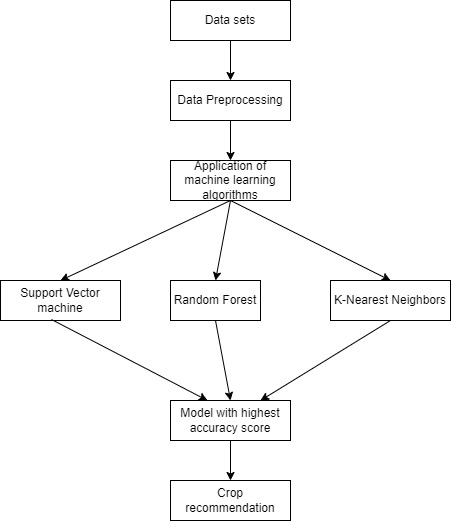


Figure 7: system architecture

## **5.2. SYSTEM MODULES AND INTERFACE**

All User Interfaces, which are designed as the various pages within the application are outlined in this section. All user interactions as well as styling have been programmed using bootstrap, CSS and html with backend built in Flask, a python light web framework and JavaScript languages.

### **5.1 Home Page**



Figure 8: Home page

### **5.3.2 Crop Inputs Page Without Values**

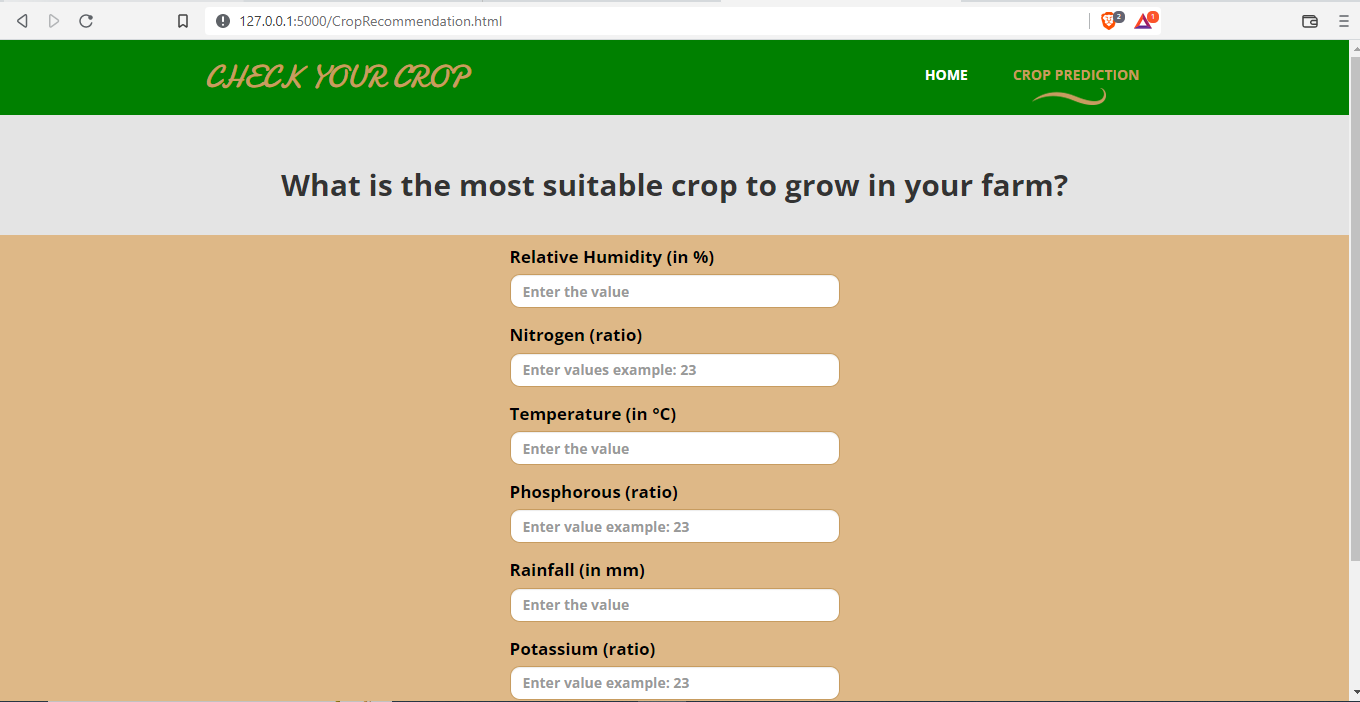


Figure 9: crop input page

### **5.3.4 Crop Inputs Page With Values**

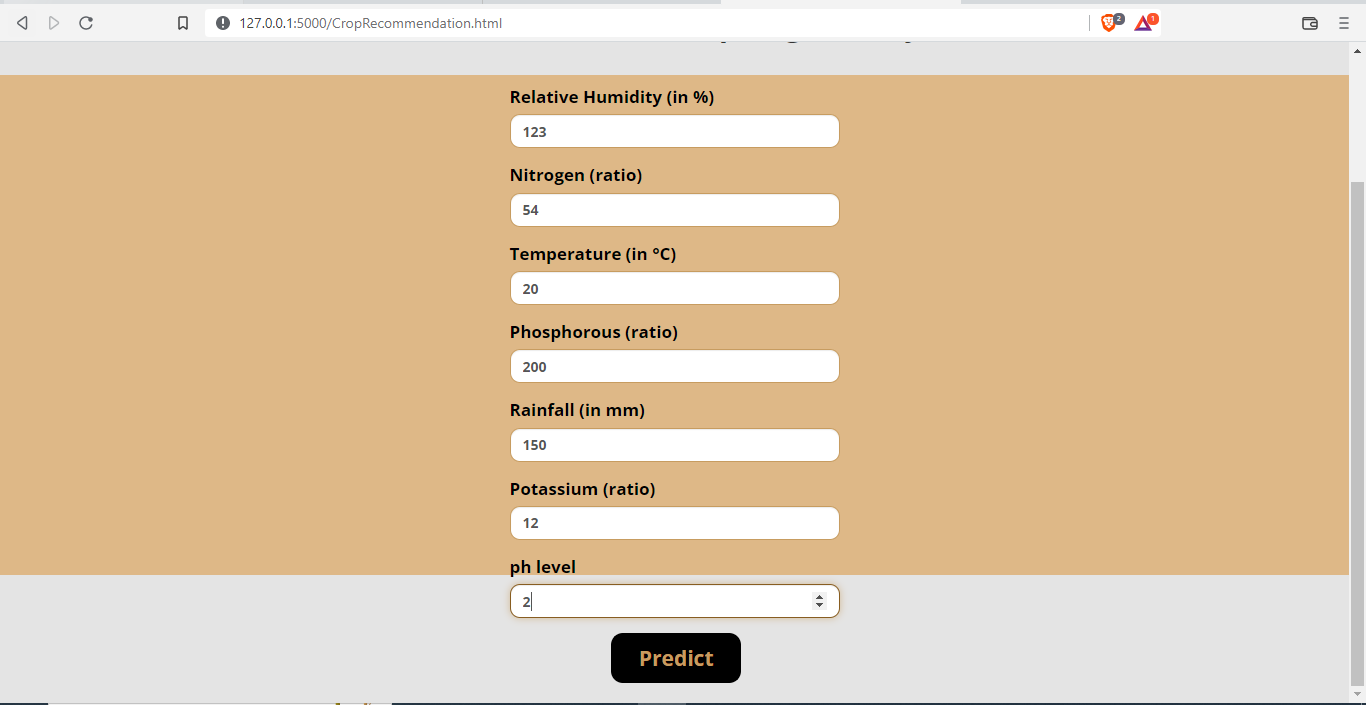


Figure 10: crop input page

### **5.3.5 Crop Prediction Page**

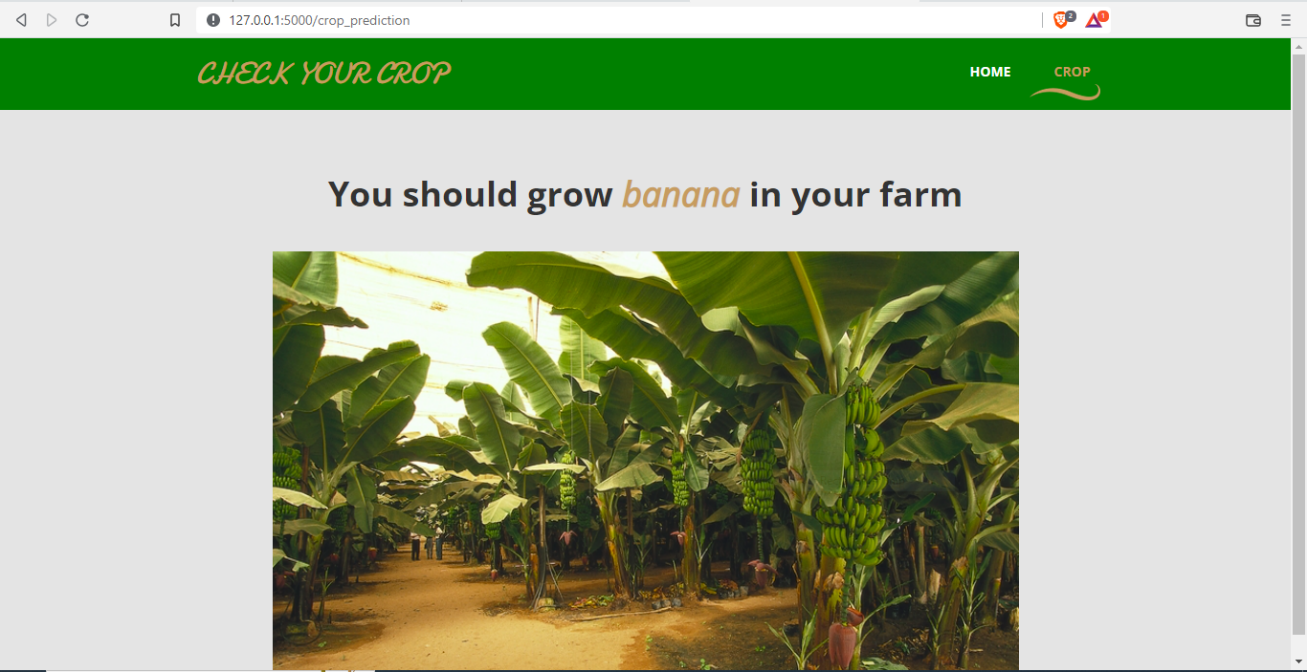


Figure 11: crop prediction

In this project I created a crop recommendation model on flask app (web based). This project hierarchy is in a tree format: images; outlook of the screen where parameter values to be feed into the system via inputs are typed, a screen when the parameter values have been typed but results not yet requested/predict button not yet pressed, screen showing the final results of the requested parameter values. The second layer is for app.py file which is the main body for the project; in this main file I imported many python libraries that includes, OS, numpy, keras, pandas, request, Flask and pickle just to name a few. These libraries helped in various functions as seen in the above previous chapter. The system uses two methods that is ‘GET’ and ‘POST’ to get data from the user through either of the request fields; the parameter values such as N, K, Humidity, Temperature, PH, P both at the same time. The system is validated to show a pop up “Please fill out this field” if a request for analysis is done with blanks on both fields.

During the analysis, if the results of the requests made by the user are met based on the values of inputs made. If the values fit a certain range which is equivalent to a particular crop, therefore such crop is the one recommended as the most suitable to be grown in that environment.

The next layer is ‘README.md’ this part describes little about the structure of the project, technologies applied. Finally, the last layer climaxes the project by indicating the libraries required to set up conducive environment to test the code, ‘requirement.txt’.

This project used methodology that consisted of four stages: data collection from various agricultural data hubs and sites, data preprocessing to ensure the data being worked on by the model is clean enough for the task, machine learning techniques, and data visualization.

Firstly, I prepared collected data in order to obtain a data set, the training set was through means of preprocessing and feature selection methods. Then, such a data set I involved in the learning step, which used ML algorithms and yields a trained classifier. Finally, I tested the classifier on various data set; the test set.

# CHAPTER SIX: SYSTEM IMPLEMENTATION

## 6.0. INTRODUCTION

In this section and with the requirements well defined, I embarked on the implementation of the findings so far. It is in this stage that we developed the programs that will help meet the expectations of the system. Moreover, we developed the interfaces that will interact with the various users of the system. This section also provides a brief summary of the tools that have been used for coding as well as testing purposes, the system test plan, testing approaches, and change-over technique to be applied.

For the successful implementation of this project, I decided to use python language for coding, visual studio as my editor IDE. I installed the current python version of 10.5 which came with an updated pip. Python as a general-purpose language, is used for web development, AI, machine learning, operating systems, mobile application development, and video games. Python is a high level, dynamically typed language developed by Guido Van Rossum in the early 1980s.

In the visual studio, I searched under extension market for various python extensions/libraries which supported the proposed project and I installed them. For the coding of the project, I created a virtual environment to hold my work and installed python libraries as indicated in the requirements.txt file on my project structure. In addition, among the installed libraries was flask , library to be used in interface development of this proposed project. Flask helped me to develop the webpage where the user of the model will be able to interact with the model by sending requests and getting/viewing the responses of the analysis displayed.

The system plan for this work involves the following highlighted steps;

step1: Acquisition of Training Dataset

step2: Data Preprocessing

step3: feature extraction and selection

step4: Training model and crop recommendation

After the above steps, a trained model is obtained, it does pre-process, feature extraction, feature selection and then applies recommendation mode. The obtained results are then compared to check the accuracy of the model.

## 6.2. PLAN OF IMPLEMENTATION

The steps involved in this system implementation are:-

Acquisition of Training Dataset - The accuracy of any machine learning algorithm depends on the number of parameters and the correctness of the training dataset. For the system, I am using various datasets all downloaded for government website and Kaggle. Many works done in this field have considered environmental parameters to predict crop type. I have tried to use all the environmental parameters like rainfall, temperature, ph., nutrients in soil to provide accurate and reliable recommendation to the farmer on which crop will be most suitable for his land.

Data Preprocessing: This step includes replacing the null and 0 values for yield by -1 so that it does not affect the overall prediction. Further I had to encode the data-set so that it could be fed into the model.

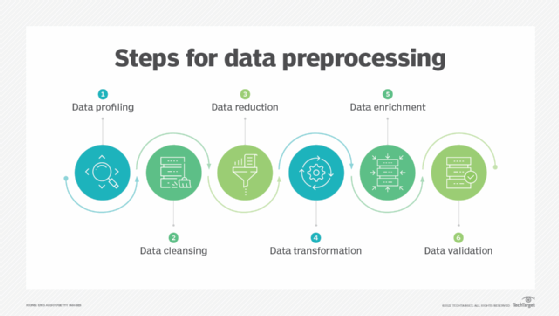


Figure 12:steps during Preprocessing

Feature Extraction - this is a technique in Deep learning and machine learning used for turning raw data into numerical features that can be handled while keeping the information in the original data set is known as feature extraction. Compared to using machine learning on the raw data directly, it produces better outcomes.

Feature Selection - By using only pertinent data and eliminating data noise, feature selection is a technique for lowering the input variable to your model. It involves automatically selecting features for your machine learning model that are pertinent to the problem you are attempting to solve. Only a small portion of the dataset's variables can be used to build a machine learning model; the others are either redundant or irrelevant. The overall performance and accuracy of the model may suffer if all these redundant and pointless information are included in the dataset. As a result, it is crucial to identify and choose the most pertinent aspects from the data, excluding any irrelevant or

Training model and crop recommendation - After the preprocessing step we used the data-set to train different machine learning models like random forest and support vector machine to attain accuracy as high as possible.

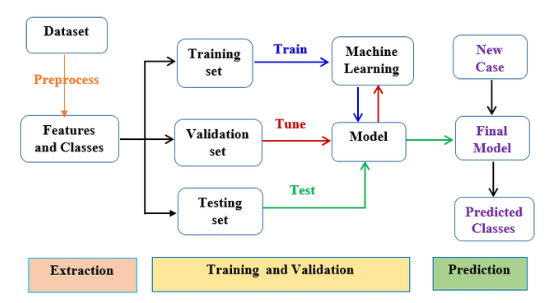


Figure 13:Working processesof trained model

## **6.3. MACHINE LEARNING MODELS**

### 6.3.1. RANDOM FOREST

A supervised learning approach called random forest is employed for both classification and regression. But it is primarily employed for classification issues. As is common knowledge, a forest is made up of trees, and a forest with more trees will be sturdier. Similar to this, Probst, P., & Boulesteix, A. L. (2017) adds that, random forest algorithm builds decision trees on data samples, obtains predictions from each one, and then uses voting to determine the optimal option. Because it averages the results, the ensemble method—which is superior to a single decision tree—reduces over-fitting.

### 6.3.2. SUPPORT VECTOR MACHINE

According to Cherkassky, V., & Ma, Y. (2019), Support vector machines (SVMs) are potent yet adaptable supervised machine learning algorithms used for both regression and classification. However, they are typically employed in classification issues. SVMs were first presented in the 1960s, however they were later improved in 1990. Compared to other machine learning algorithms, SVMs are implemented in a different method. Because they can manage numerous continuous and categorical variables, they have recently become very popular.

An SVM model is just a hyperplane in multidimensional space that represents several classes. SVM will generate the hyperplane in an iterative manner in order to reduce error. SVM seeks to discover a maximum marginal hyperplane by classifying the datasets (MMH).

### 6.3.1. K-NEAREST NEIGHBORS

The k-nearest neighbors (KNN) technique calculates the likelihood that a data point will belong to one group or another based on which group the data points closest to it do. (Guo G., et. Al., 2018) adds that, the supervised learning technique K-nearest neighbors (KNN) is used for both regression and classification. By calculating the distance between the test data and all of the training points, KNN tries to predict the proper class for the test data. Then choose the K spots that are closest to the test data. The KNN method determines which classes of the "K" training data the test data will belong to, and the class with the highest probability is chosen. The value in a regression situation is the average of the 'K' chosen training points.

## 6.4. CODING AND TESTING TOOLS

### 6.4.1. Visual Studio Code

This is the main IDE that was used for the development and debugging of the application's code. Python extensions had to be integrated into the IDE so as to allow for model development. The IDE was also linked to my GitHub account so as to allow for saving of the code into an online repository for backup purposes. I chose this IDE since it's a free, open-source, and cross-platform editor. The interface also suited my preferences.

### 6.4.2. GitHub

This is a version control system that was used to keep track of any changes that were made as the application was being developed. This is important in a situation whereby the developer might require to roll back to an earlier system and correct changes. It also allows for collaborations among developers on the same project.

### 6.4.3. Chrome Browser

The web browser that was used to run the compiled html files using flask python web framework. This is the physical device that was used for testing the system’s usability as well as functionality. It allowed for easy interaction with the graphical interface of the system.

## 6.5. SYSTEM TEST PLANS

### 6.5.1. Goal of the Test Plan

The Test Plan aimed at ensuring that all functional and design requirements are implemented as clarified in the documentation and It also acted as a check which ensured that the code that is developed is free of bugs and performs the required functions.

### 6.6 Features to be examined

Some the features and modules to be examined include;

User Interface – here all the horizontal input controls and buttons will be tested to ensure that they work properly.

Services - these include all services that are provided for by the machine learning model. They should be working properly and ensure data can be passed between the front-end as well as the back-end(model side) of the application.

## 6.7 Test Environments

Test environment helps to provide a dedicated environment for isolation of code and verify behavior of the application. Environments used for testing are as follows:

## 6.8 Chrome Web browser.

Visual Studio Debugger – this was used to navigate through the code to inspect the state of the application and show its execution.

## 6.9 Pass Criteria

The application in general should meet all the intended user requirements.

It should easily respond to specified user interactions with minimal effort applied.

The application should allow for seamless communication with the model as well as handle user requests with minimal surprise.

## 6.10 TESTING APPROACHES

### 6.10.1 Unit Testing

This was put in place to ensure that individual units of source code, sets of one or more programs modules together associated control data, usage procedure and operating procedures are tested to determine whether they are fit for use. The approach allowed for better re-usability of code because codes were more modular. The approach allowed for easier debugging and also allowed for earlier identification of errors as opposed to when the system has already been developed and the developer has a huge task to review the code and do the necessary changes.

### 6.11 Component Testing

In visual studio code, this is equivalent to plugin testing and python path set up. It involved the process of testing single plugins in the application which act as the basic building blocks. The main purpose of this test was to evaluate the design looks of the User Interface plugins as well as their required functionalities. The plugins were put through a context in which their lifecycle was put to test.

## 6.12 Integration Testing

This is the second level of software testing process that comes after unit testing. The test was used to test units or individual components of the software in a group. In addition, the test focused on exposing the defects at the time of interaction between integrated components or units. Most of these tests were done on the chrome web browser to ensure better performance of the system.

## 6.13 CHANGE – OVER TECHNIQUE

### 6.13.1 Parallel Changeover

This is where the new and old systems are run simultaneously for some time and their results are compared until the users have confidence in the new system and then discard the old system. This technique is suitable for this particular application system since it has a lower risk of failing services in case the new system does not meet some targeted objectives. In this technique, the new system will run at the same time with the Spreadsheet application after which its performance as well as the effect will be evaluated.

### 6.13.2 DATASET USED

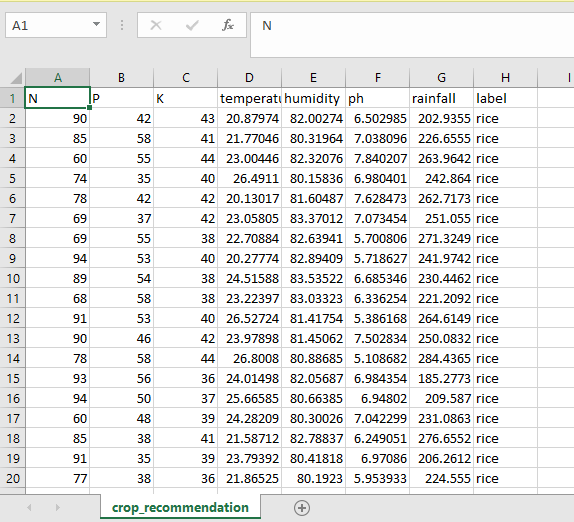
****

Figure 14:crop Recommendation data set

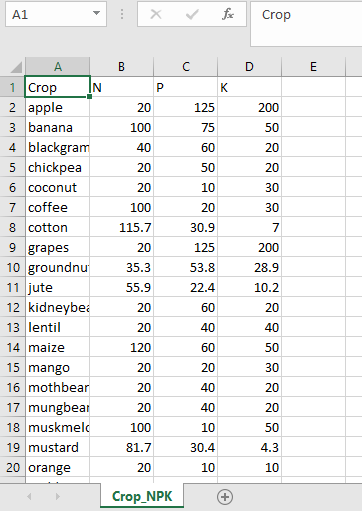
****

Figure 15:crop soil properties of N,P,K

# CHAPTER SEVEN

## 7.0. LIMITATIONS, CONCLUSIONS AND RECOMMENDATIONS

## 7.1. INTRODUCTION

Within this final chapter, the difficulties of the developed project that were encountered all through its development are briefly explained. A logical conclusion is also brought out in this chapter after which recommendations are suggested on how the system can be improved further.

## 7.2. LIMITATIONS

The development of the crop recommendation model has not been as easy as it may appear as severally, I had to face a number of challenges, most of these challenges ran throughout from start to finish. Some of these challenges include:

**Learning and implementation** - The most challenging tasks i faced for this project were machine learning model’s implementation, using pandas, NumPy and scikit learn to recommend crop using the collected dataset.

**Lack of Finance –** Due to lack of funds, I was unable to purchase the electronics components such as Node MCU model which could allow me to use real time data in crop recommendation.

**Information Source** - Information required to develop the application was not readily available on the Internet and reference sources and if it did, this was not exact despite the high internet connection. This did not only apply for the source codes only but also did apply for the Literature review still. The codes available on the internet required a bit of customization and thorough editing in order to realize the required functionality. This was not easy as for the information such as the codes to be edited they required to first be understood before procession to edit the contents from their current form.

**Internal Validity** - It is a threat that arose when the time selecting algorithms went wrong. As this is the first step of the research, this area is to be concentrated so that there wouldn’t be a chance of misplacing an algorithm with others.

**External Validity -** The extent to which the generalization of results from an experiment was achieved discusses this validity threat. For training and testing the chosen machine learning algorithms, I used real-world data to address this threat.

## 7.3. CONCLUSION

This system helps the farmer to choose the right crop by providing insights that ordinary farmers don’t keep track of thereby decreasing the chances of crop failure and increasing productivity. It also prevents them from incurring losses. We could achieve an accuracy of 99.88 percent from the Random Forest and an accuracy of 98.26 percent from the K-Nearest Neighbors model.

## 7.4. RECOMMENDATIONS

The system can be extended to the mobile and be accessed by farmers for interactivity purposes across the country. Further development is to integrate the crop recommendation system with another subsystem, yield predictor that would also provide the farmer an estimate of production if he plants the recommended crop. To remove the issue of providing statistical data in crop recommendation feature in future research, I recommend integrating this system with an IOT model where perhaps a Node MCU model will collect all data like soil pH , Rainfall and send it to the real-time database where we can access for recommendation.

8.0 REFERENCES

Alshamrani, A., & Bahattab, A. (2017). A comparison between three SDLC models waterfall model, spiral model, and Incremental/Iterative model. *International Journal of Computer Science Issues*.

Cherkassky, V., & Ma, Y. (2019). Practical selection of SVM parameters and noise estimation for SVM regression. *Neural networks*, *17*(1), 113-126.

Rakesh Kumar , M.P. Singh, Prabhat Kumar and J.P. Singh, “Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique”, International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials, 2015

Guo, G., Wang, H., Bell, D., Bi, Y., & Greer, K. (2018, November). KNN model-based approach in classification. In *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"* (pp. 986-996). Springer, Berlin, Heidelberg.

Hakkim, V. A., Joseph, E. A., Gokul, A. A., & Mufeedha, K. (2020). Precision farming: the future of Indian agriculture. *Journal of Applied Biology and Biotechnology*, 0-7.

Rakesh Kumar , M.P. Singh, Prabhat Kumar and J.P. Singh, “Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique”, International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials, 2015

Probst, P., & Boulesteix, A. L. (2017). To tune or not to tune the number of trees in random forest. *The Journal of Machine Learning Research*, *18*(1), 6673-6690.

Jr, H., O., L., Hammond, K. R., Lusk, C. M., & Mross, E. F. (2019). The application of signal detection theory to weather forecasting behavior. *Monthly Weather Review*, 863-883.

Liu, S., & Wei, J. (2018). The Satelite on the Research of the Software Life Cycle: V+ Iterative Waterfall. . *In 2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference*, 1716-1720.

Shylesh, S. (2017). A study of software development life cycle process models. . *In National Conference on Reinventing Opportunities in Management, IT, and Social Sciences*, 534-541.

Jay Gholap, Anurag Ingole, Jayesh Gohil, Shailesh Gargade and Vahida Attar, “Soil Data Analysis Using Classification Techniques and Soil Attribute Prediction”, International Journal of Computer Science Issues, Volume 9, Issue 3

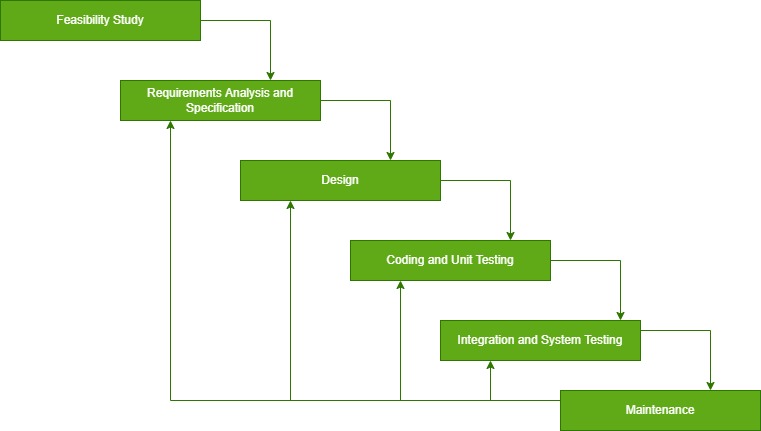
APPENDIX A:

**PROJECT RISKS AND MITIGATIONS**

|  |  |
| --- | --- |
| RISK | MITIGATION |
| Project complexity such that different features are developed using different technologies. Therefore, complexity in integration till the application operates | Use of Iterative method will allow more time for testing. |

APPENDIX B:

**ITERATIVE WATERFALL METHODOLOGY**



APPENDIX C:

**PROJECT TIMELIME**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Weeks | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 13 | 14 |
| Feasibility study |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Requirement analysis & specification |  |  |  |  |  |  |  |  |  |  |  |  |  |
| System design |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Coding and unit testing |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Integration and system testing |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Maintenance |  |  |  |  |  |  |  |  |  |  |  |  |  |

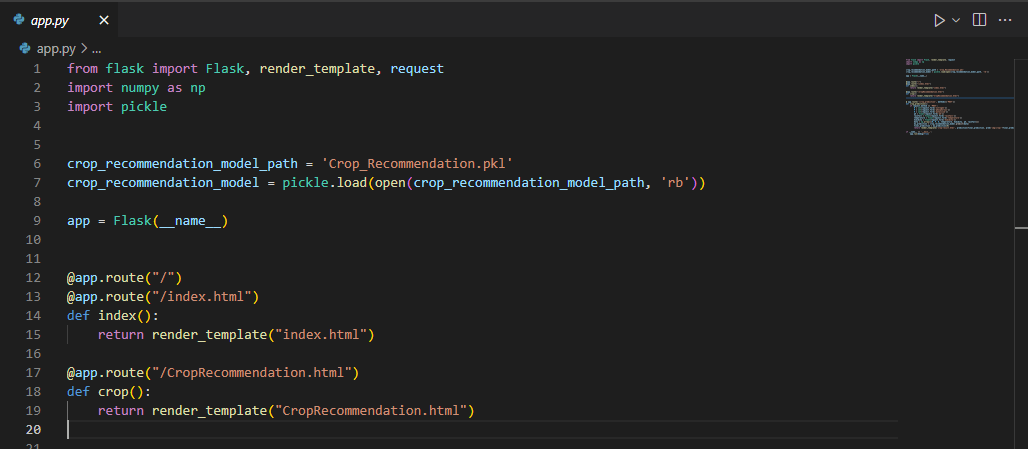
APPENDIX D:

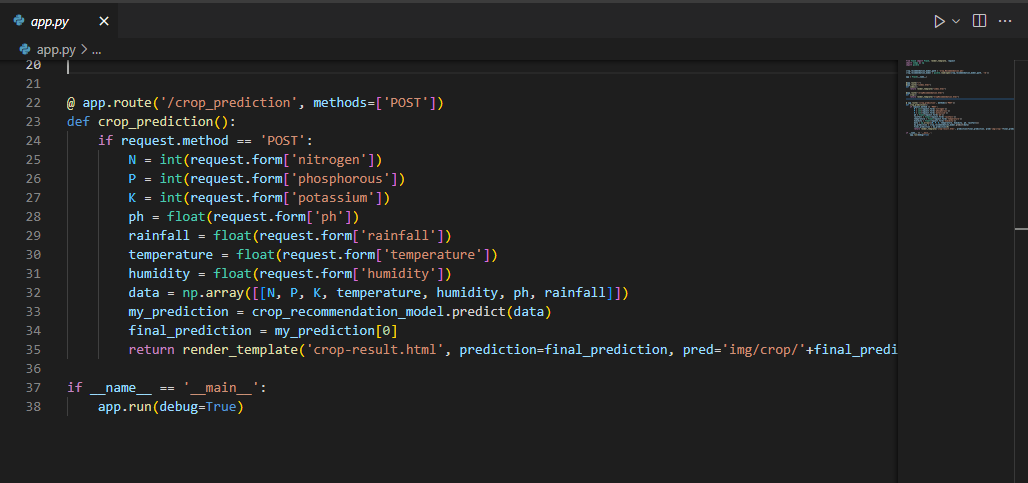
**PROJECT COST**

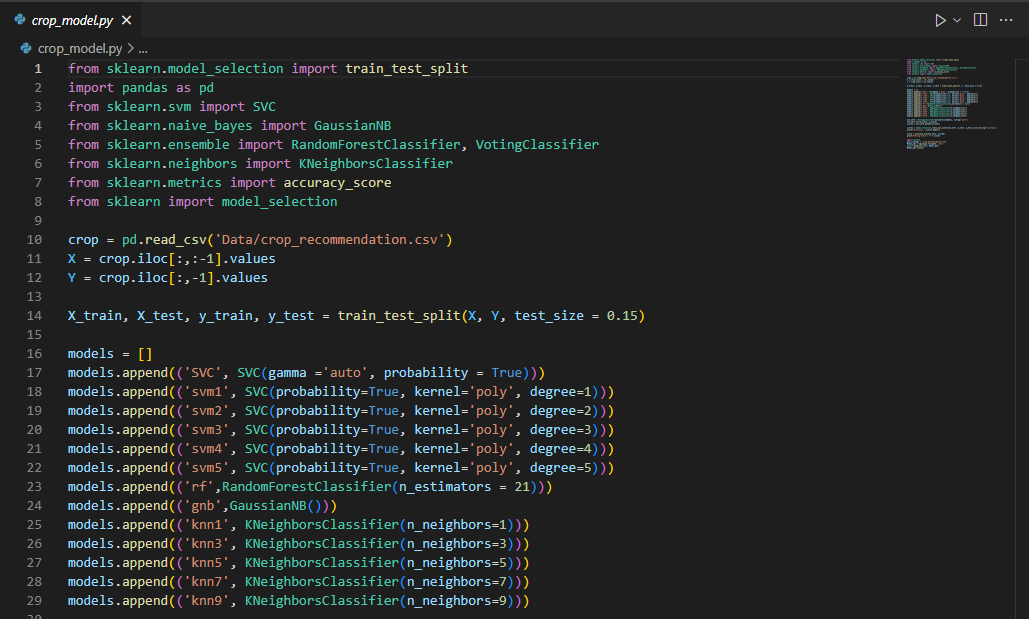
|  |  |  |  |
| --- | --- | --- | --- |
| NUMBER | TOOL/ITEM | Approximate number | Approximate Cost |
| 1 | NOTEBOOK | 1 | 200 |
| 2 | LAPTOP | 1 | 41,000 |
| 3 | INTERNET | (recommended) | 6500 |
|  |  | TOTAL COST | 65,700 |

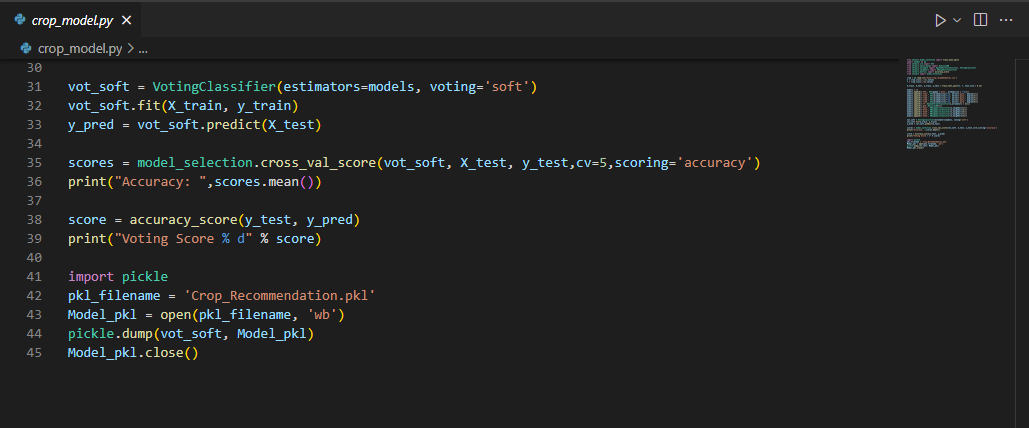
APPENDIX E:

**MAIN CODE**

****

****

****

****