Cultivation Recommendation System using Machine Learning

BACHELOR OF COMPUTER ENGINEERING

by

Avishkar Dalvi (20202002)

Riddhi Narkar (19102003)

Radha Rakshe (19102067)

Aarya Totey (19102070)

Guide

Prof. Deepak Khachane

Department of Computer Engineering

A. P. SHAH INSTITUTE OF TECHNOLOGY, THANE

2022-2023

A. P. SHAH INSTITUTE OF TECHNOLOGY, THANE

CERTIFICATE

This is to certify that the project entitled "**Cultivation Recommendation System using Machine Learning**" is a bona fide work of "**Avishkar Dalvi" (20202002), "Riddhi Narkar" (19102003), "Radha Rakshe" (19102067), "Aarya Totey" (19102070)** submitted to the University of Mumbai in partial fulfilment of the requirement for the award of the degree of **Bachelor of Engineering in Computer Engineering.**

Guide Prof. Deepak Khachane	Project Coordinator Prof. Rushikesh Nikam
Head of Department Prof. Sachin Malave	Principal Dr. Uttam Kolekar

Date:

A. P. SHAH INSTITUTE OF TECHNOLOGY, THANE

Project Report Approval for B.E.

This project report for Sem-VIII entitled ***Cultivation Recommendation System using Machine Learning by "Avishkar Dalvi" (20202002), "Riddhi Narkar" (19102003), "Radha Rakshe" (19102067), "Aarya Totey" (19102070) is approved for the degree of Bachelor of Engineering in Computer Engineering, 2022-23.***

Examiner Name Signature

1. _____

2. _____

Date:

Place:

Declaration

We declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Avishkar Dalvi (20202002)

Riddhi Narkar (19102003)

Radha Rakshe (19102067)

Aarya Totey (19102070)

Date:

Abstract

India is an agrarian society. Majority of our population is involved in agriculturally based activities and are dependent on crops related business for their livelihood. Farmers often are not able to produce a good yield due to poor farming decisions. This works aims to be an assistance tool for a farmer to make his farming decisions. It comprises of three machine learning models namely a crop recommendation system, a fertilizer recommendation system, and a crop disease identification system with cures for the diseases. The crop recommendation and the fertilizer recommendation use regression models whereas the crop disease identification uses convolutional neural network model.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Abbreviation

ICT	Information Communication Technology
ML	Machine Learning
DL	Deep Learning
CNN	Convolutional Neural Networks
LVQ	Learning Vector Quantization

CHAPTER 1
Introduction

India has a primary sector economy, that means that majority of our population is involved in agriculturally based activities and are dependent on crops related business for their livelihood. As per the census in 2011; in India, approximately 118 million people are farmers and 144 billion people are laborers working in an agricultural field. Total Indian population of India in 2011 was nearly about 121 crore and out of which nearly about 2630 lack people are farmers. This makes India, an agrarian society. With the production of agriculture activity of $375.61 billion, India is 2nd larger producer of agriculture products. India accounts for 7.39 percent of total global agricultural output. India's economy is being driven by majority of the population engaging in primary sector occupations. The agriculture sector's contribution to the Indian economy is much higher than the world's average (6.4%). The industry and services sector's contribution is lower than the world's average 30% for the Industry sector and 63% for the Services sector [10].

The Government of India, since its inception in 1947, has being running huge and many schemes that help farmers. For example, the "Parampragat Krishi Vikas Yojana (PKVY)" which is one of the important government schemes which encourages farmers for

traditional and organic farming in India. Under this scheme, the Government of India provides a financial assistance to the farmers of Rs 50,000 per hectare every three years for organic inputs, certification, labelling, packaging, transportation and marketing of organic produce. The scheme focuses on reducing the ill effects of overuse of fertilisers and agrochemicals by promoting organic manures, bio-fertilisers and bio-pesticides. It helps improve the soilfertility by improving organic carbon in the soil which results in enhancing moisture holdingcapacity in the field too.

In spite of many such schemes, the factor that hurts the farmer more is the unpredictable weather conditions and poor planning for harvest. Multiple unsustainable technologies took a huge toll on our Earth in the recent past decades giving rise to global warming and climate as worldwide crisis. Lal Bahadur Shastri once said, “Jai Jawan Jai Kisaan”, which believed in uplifting the role of a farmer. Through the course of this work, wenoticed that ICT can help farmers. ICT in agriculture is an emerging field focusing on the enhancement of agricultural and rural development in India. It involves innovative applications using ICT in the rural domain. The advancement of ICT can be utilized for providing accurate and timely relevant information and services to the farmers, therebyfacilitating an environment for remunerative agriculture [3].

Crop diseases are a major threat to food security, but their rapid identificationremains difficult in many parts of the world due to the lack of the necessary infrastructure. The combination of increasing global computer penetration and recent advances in computervision made possible by deep learning has paved the way for disease diagnosis.[4]

Crop cultivation anywhere in the world depends on the climate so called seasons and soil properties, however, the enhancing the production of crops depend on various factors like mainly on temperature. The ideal harvest prior to planting it, it would be of extraordinary assistance to the farmers and others required to settle on fitting choices on upgrading the creation of yields for neighborhood utilization needs and may prompt the capacity andexpanded fare choice for business. For that, ML procedures with the end goal that it proposes

the appropriate corps dependent on the temperature.

Most of the farmers are untutored and have little to no scientific knowledge of farming. So, they have to rely on the hit and trial method to learn from experience which leads to wastage of time and resources. We need a system to focus on building a predictive model to recommend the most suitable crops to grow in a particular farm based on various parameters. This can be helpful for the farmers to be more productive and competent without wasting any resources by farming the most competent crops.

With the vision of providing assistance and guidance to farmers during the crucial decision-making process required at various stages during farming, we propose a solution in the form of three ML models. Under many government schemes, farmers are able to get their land tested for nutrients. A farmer has all the lab results, but is not able to turn those results into actions and implementation. This work bridges that gap.

This work determines to assist and guide the farmers in 3 important and crucial areas of agricultural decisions, namely, which crop to be planted and which fertilizer to be used, both of which pertaining to the soil and weather conditions; and the last one to predict plant diseases and provide the farmer with organic solutions for the same as well.

CHAPTER 2
Literature Survey

Sunidhi N et.al crop production is affected by number of diseases worldwide. Timely detection of diseases will allow to monitor and implement control measures with greater efficiency. The main aim of this project is to design an AI-Based disease detection system that detects the type of disease present in tomato leaf by clicking the images of various leaves through camera and spray the respective pesticide to the diseased part of the plant. [2]

V Rajesh Kumar et.al in the field of agriculture, image processing is a constantly evolving field of research and progress. Currently, several plant disease identification studies are underway. Identifying plant diseases can not only help farmers increase yields, but also promote a variety of agricultural practices. This paper proposes an algorithmic program for the disease detection and categorization with the assistance of machine learning mechanisms and image recognition tools. [3]

Melike Sardogan et.al the early detection of diseases is important in agriculture for an efficient crop yield. Automatic methods for classification of plant diseases also help taking action after detecting the symptoms of leaf diseases. This paper presents a Convolutional Neural Network (CNN) model and Learning Vector Quantization (LVQ) algorithm based method for tomato leaf disease detection and classification. [6]

Sayed Mazhar Ali et.al it has been a major problem to identify what to grow, any man has adequate space in the owner's land. Not only domestic lands but also for farming lands. Why it has become a problem is that environmental factors such as temperature, water levels, and soil conditions are uncertain as they change from time to time. Due to these problems, this solution of crop recommendation system predicts the user, what crop type would be the most suitable for the selected area by collecting the environmental factors for plant growth and processing them with the trained sub-models of the main of the system.[4]

Sharada P. Mohanty et.al using Deep Learning for Image-Based Plant Disease Detection Crop diseases are a major threat to food security, but their rapid identification remains difficult in many parts of the world due to the lack of the necessary infrastructure. The combination of increasing global smartphone penetration and recent advances in computer vision made possible by deep learning has paved the way for smartphone-assisted disease diagnosis. [9]

Dhruv Piyush Parikh et.al this system focuses on building a predictive model to recommend the most suitable crops to grow in a particular farm based on various parameters. This can be helpful for the farmers to be more productive and competent without wasting any resources by farmingthe most competent crops.[5]

Pradeepa Bandara et.al recommendation system through integrated models of collecting environmental factors using Arduino microcontrollers, Machine learning techniques such as Naïve Bayes (Multinomial) and Support Vector Machine (SVM), Unsupervised machine learning algorithm such as K-Means Clustering and also Natural Language Processing (Sentiment Analysis) concerned with the Artificial Intelligence to recommend a crop for the

selected land with site-specific parameters with high accuracy and efficiency. It has been a major problem to identify what to grow, any man has adequate space in the owner's land. Not only domestic lands but also for farming lands. [7]

Title of the research paper	Year published	Authors	Abstract
AI Based Automatic Crop Disease Detection System **IEEE Xplore**	2021	Sunidhi N, Jalaja S VLSI Design and Embedded Systems Bangalore Institute of Technology, Bangalore, India; Dept. of Electronics and Communication, Bangalore Institute of Technology, Bangalore, India	Timely detection of diseases will allow to monitor and implement control measures with greater efficiency. The main aim of this project is to design an AI-Based disease detection system that detects the type of disease present in tomato leaf by clicking the images of various leaves through camera and spray the respective pesticide to the diseased part of the plant. Deep Convolutional neural networks (D-CNN) and transfer learning techniques are used to detect and classify the disease. Three transfer-learning models i.e., Alex Net, InceptionV3 and ResNet152V2 were used for training and classification and the results were compared which gave the accuracy of 95%, 97% and 99% respectively.
Identification of Plant Diseases Using Image Processing and Image Recognition **IEEE Xplore**	2021	V Rajesh Kumar, K Pradeepan, S Praveen, M Rohith, V Vasantha Kumar Department of	In the field of agriculture, image processing is a constantly evolving field of research and progress. Currently, several plant disease identification studies are underway. Identifying plant diseases can not only help farmers increase yields, but also promote a variety of agricultural practices. This paper proposes an algorithmic program for the diseases detection and categorization with

		Computer Science and Engineering, KCG College of Technology, Chennai, India	the assistance of machine learning mechanisms and image recognition tools. First detect and record the contaminated area and then perform image pre-processing.
Plant Leaf Disease Detection and Classification based on CNN with LVQ algorithm **IEEE Xplore**	2020	Melike Sardogan, Adem Tuncer, Yunus Ozen Department of Computer Engineering, Yalova University	The early detection of diseases is important in agriculture for an efficient crop yield. The bacterial spot, late blight, Septoria leaf spot and yellow curved leaf diseases affect the crop quality of tomatoes. Automatic methods for classification of plant diseases also help taking action after detecting the symptoms of leaf diseases. This paper presents a Convolutional Neural Network (CNN) model and Learning Vector Quantization (LVQ) algorithm based method for tomato leaf disease detection and classification. The dataset contains 500 images of tomato leaves with four symptoms of diseases. We have modelled a CNN for automatic feature extraction and classification.
Crop Recommendation System **International Journal of Computer Applications (0975**	2020	Pradeepa Bandara, Thilini Weerasooriya, Ruchirawya T.H., W.J.M. Nanayakkara, Dimantha M.A.C, Pabasara M.G.P.	Recommendation system through integrated models of collecting environmental factors using Arduino microcontrollers, Machine learning techniques such as Naïve Bayes (Multinomial) and Support Vector Machine (SVM), Unsupervised machine learning algorithm such as K-Means Clustering and also Natural Language Processing

– 8887) Volume 175– No. 22, October 2020		Sri Lanka Institute of Information Technology, Sri Lanka	(Sentiment Analysis) concerned with the Artificial Intelligence to recommend a crop for the selected land with site-specific parameters with high accuracy andefficiency. It has been a major problem to identify what to grow, any man has adequatespace in the owner's land.
Using Deep Learning for Image-Based Plant Disease Detection Frontiers in Plant Science Online Publication	2019	Sharada P. Mohanty, David P. Hughes, and Marcel Salathé Digital Epidemiology Lab, EPFL, Geneva, Switzerland, School of Life Sciences, EPFL, Lausanne, Switzerland.	Crop diseases are a major threat to food security, but their rapid identification remains difficult in many parts of the world due to the lack of the necessary infrastructure. Using a public dataset of 54,306 images of diseased and healthy plant leaves collected under controlled conditions, we train a deep convolutional neural network to identify 14 crop species and 26 diseases (or absence thereof). The trained model achieves an accuracy of 99.35% on a held-out test set, demonstrating the feasibility of this approach.
Machine Learning based Crop Recommendation System for Local Farmers of Pakistan	2021	Sayed Mazhar Ali, Bhagwan Das, Dileep Kumar Department of Electronic Engineering, Quaid	In Pakistan, the most part of the land is used for agriculture cultivation to meet the desires of nearby people and export want as properly. Crop cultivation anywhere in the world depends on the climate so called seasons and soil properties, however, the enhancing the production of crops depend on various factors like mainly on temperature. In

A Research Gate Preprint Online Publication		University of Engineering, Science and Technology, Nawabshah, Sindh, Pakistan.	order to address the issue of increasing crop production for Pakistan, a crop recommendation system is proposed in this work. In this work, idea of ideal harvest prior to planting it, it would be of extraordinary assistance to the farmers and others required to settle on fitting choices on upgrading the creation of yields for neighborhood utilization needs and may prompt the capacity and expanded fare choice for business. subsequently diminishes the monetary misfortunes looked by the farmers.
Machine Learning Based Crop Recommendation System International Journal of Advanced Research in Science, Communication and Technology	2021	Dhruv Piyush Parikh, Jugal Jain, Tanishq Gupta, Rishit Hemant Dabhade Vellore Institute of Technology, VIT Chennai, India	The three most basic amenities required for the survival of a human being arefood, shelter and clothing. In today's tech-savvy generation, the latter two have witnessed a huge scientific boost. Unfortunately, even today, agriculture is considered as more of a man-power oriented field. Most of the farmers are untutored and have little to no scientific knowledge of farming. So, they have to rely on the hit and trial method to learn from experience which leads to wastage of time and resources. Our system focuses on building a predictive model to recommend the most suitable crops to grow in a particular farm based on various parameters. This can be helpful for the farmers to be more productive and

			competent without wasting any resources by farming the most competent crops.

Table 2.1 Literature Survey

CHAPTER 3
Limitation of Existing system

In the works we went through for this, we identified that although all these works were pertaining for the betterment of the farmers, but they failed to implement slight nuances which would make the system really efficient for the farmer. There have been many review papers on this which summarize the work of many papers, but no competitive study of different ML models for crop recommendation system.[1]

After the study of all the aforementioned works, we found 3 key limitations:

1) Absence of a unified product/service to solve problem
2) Products/services demanding a high computer literacy rate to be used by farmers
3) Lack of solution-based approach in products/services with a classification or an identification system

CHAPTER 4

Problem Statement, Objectives and Scope

4.1 Problem Statement

To implement a system for farmers with features of crop recommendation, fertilizer recommendation, and crop leaf disease identification.

As easy as it might be to be a city dweller, we cannot ignore to the plights of our farmers. India has a primary sector economy, that means that majority of our population is involved in agriculturally based activities and are dependent on crops related business for theirlivelihood. India, no doubt is an agrarian society, and we often tend to ignore this fact.

Multiple unsustainable technologies took a huge toll on our Earth in the recent past decades giving rise to global warming and climate as worldwide crisis. This hurts the farmer the most, as they are often lacking proper guidance and resources to solve their problems. Lal Bahadur Shastri once said, “Jai Jawan Jai Kisaan”, which believed in uplifting the role of a farmer.

This project is aimed at assisting the farmer by deploying a crop recommendation system, a fertilizer recommendation system, and a crop disease identification system withcures. All these would be implemented as three different ML models and integrated into asingle webpage for easier and convenient access.

4.2 Objectives

The objectives of this work include:

1. To provide a single place to access help for crop related problems, the causes, and how to cure them, thus providing a solution-based approach.
2. To make these services available over the internet in order to maximize reach.
3. To try to make the UI as simple as possible so that a naïve user, whose computer literacy is lower than average, too, can easily navigate.

4.3 Scope

The scope of this work includes:

1. This project has the potential to assist a farming enthusiast or an occupational farmer to make smarter decisions on his harvest strategy.
2. It can provide necessary guidance related to cultivation and fertilizer ideas and curing crop diseases.

CHAPTER 5

Proposed System

5.1 Project Design

5.1.1 Architecture diagram

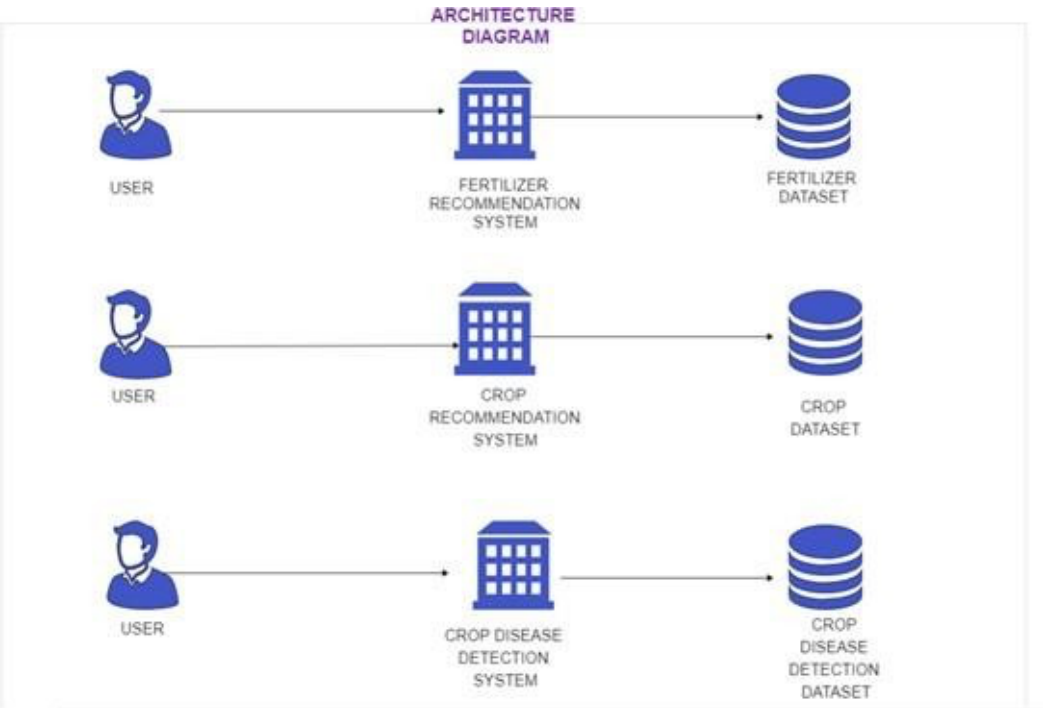


Fig 5.1 Architectural diagram

An architectural diagram is a visual representation that maps out the physical implementation for components of a software system. It shows the general structure of the software system and the associations, limitations, and boundaries between each element.

This architectural diagram depicts how the three ML models are being deployed. Each has its own dataset and the user needs to invoke a different request for each service.

5.1.2 DFD

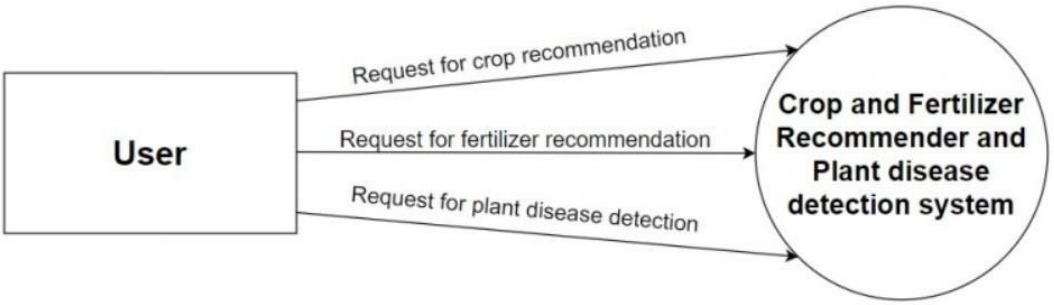


Fig 5.2 DFD level 0

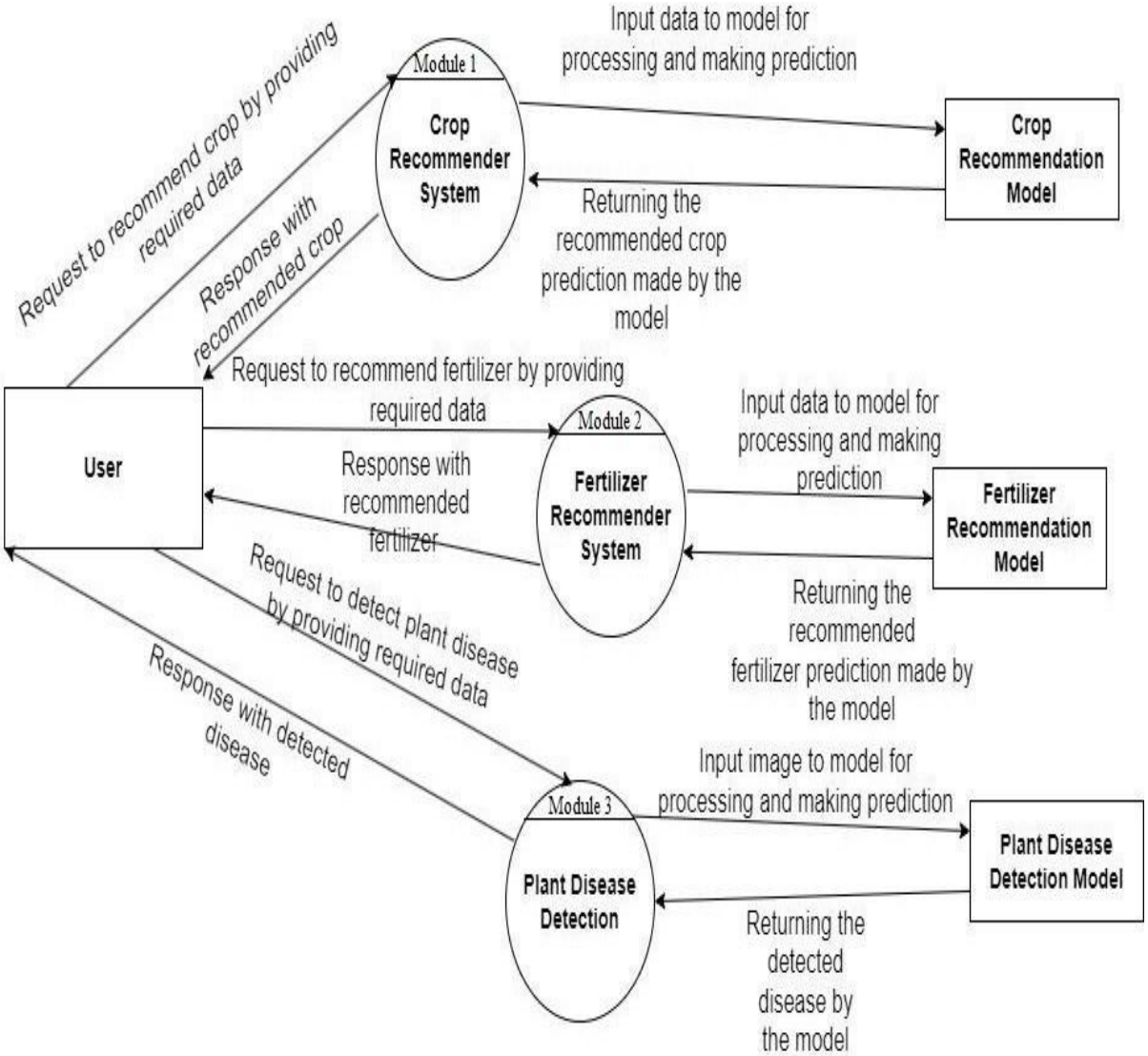


Fig 5.3 DFD level 1

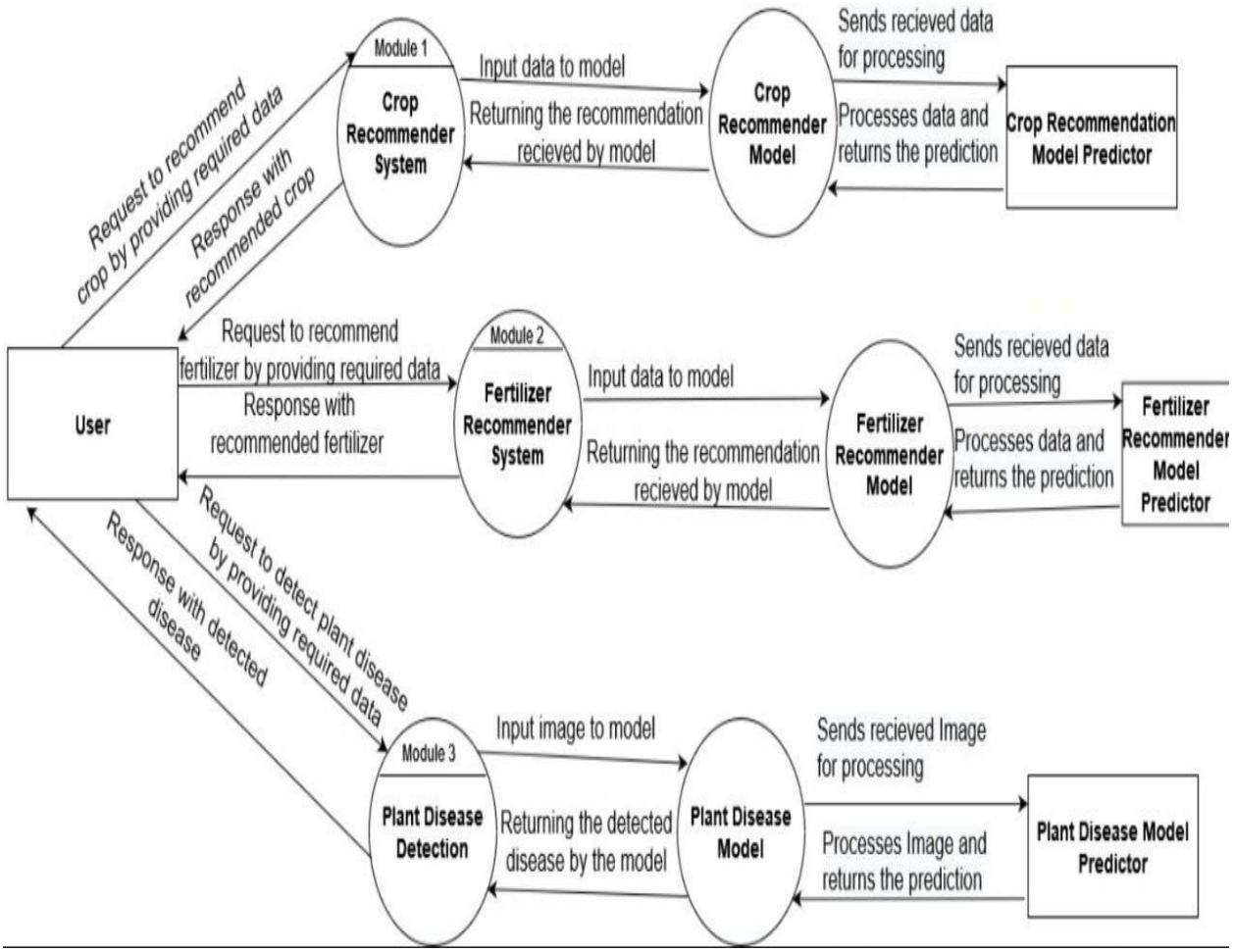


Fig 5.4 DFD level 2

The user provides soil and weather data to the model, which in turn feeds the crop recommendation system model which makes the prediction. Same happens in case of fertilizer recommendation system, just the user needs to input only the soil data. For the plant disease identification, a photo needs to be uploaded by the user which is fed into the model, and it classifies the disease. The solution is not integrated into the model itself, it would be externally integrated, as there was no dataset with inbuilt cures for the identified diseases present.

5.1.3 Use case diagram

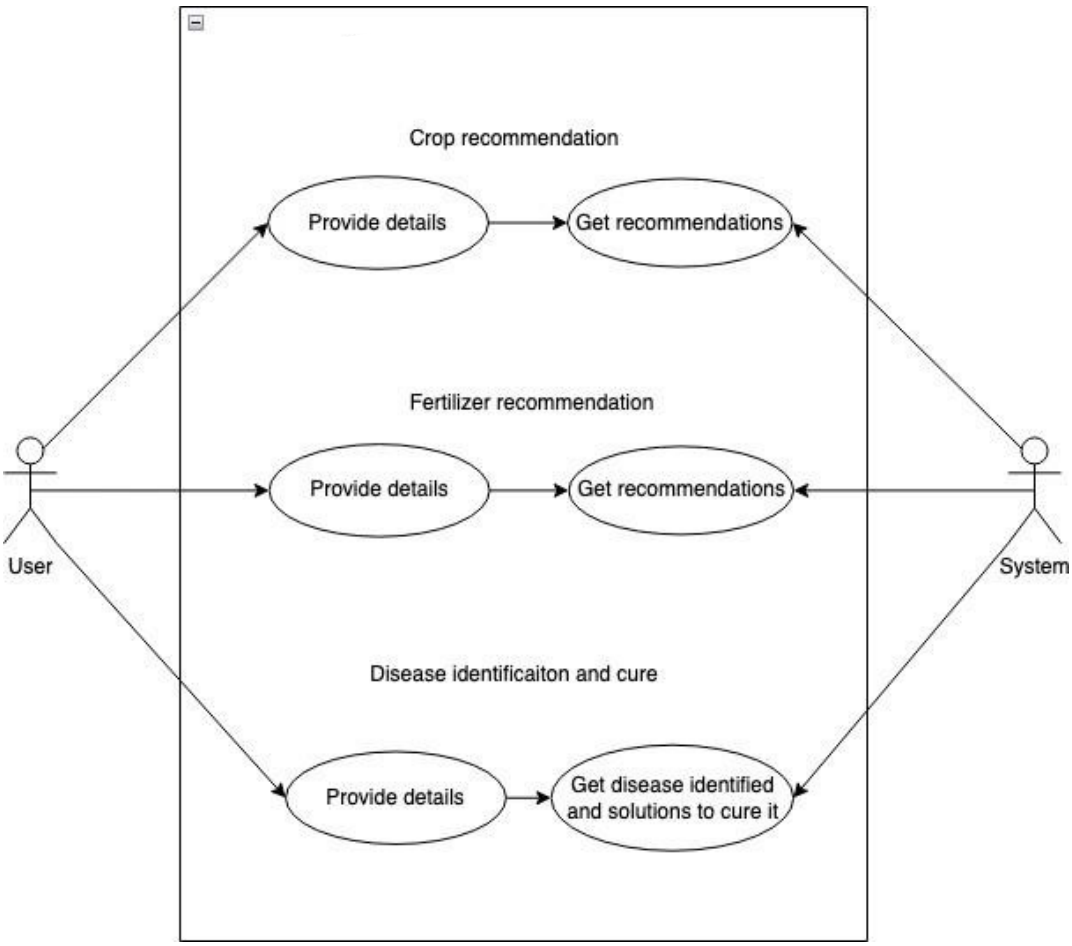


Fig 5.5 Use case diagram

Consider these following use cases:

1) User invokes Crop recommendation system:

   In this case the user would be taken to the crop recommendation page and after that he can enter the required details. After the model receives all the details, the user is then brought to the result page, where the recommendation is being done.

2) User invokes Fertilizer recommendation system

   In this case the user would be taken to the fertilizer recommendation page and after that he can enter the required details. After the model receives all the details, the user is then brought to the result page, where the recommendation is being done.

3) User invokes Crop disease identification system

   In this case the user would be taken to the crop disease identification page and after that he needs to provide an image of the crop. After the model receives it, the user is then brought to the result page, where the identification is being done and the solution (cure) for the disease is also presented on the screen.

4) User invokes a mixture of any of these above systems

   The user can use any of the above systems in any order and any number of times. He would just need to navigate to the proper page he needs. The navigation would be provided in the nav bar of the website for easier access. Once the request for any use case is taken, and the result is produces, none of that is saved anywhere. So the user can keep using the same feature repeatedly as well.

5.1.4 Sequence diagram

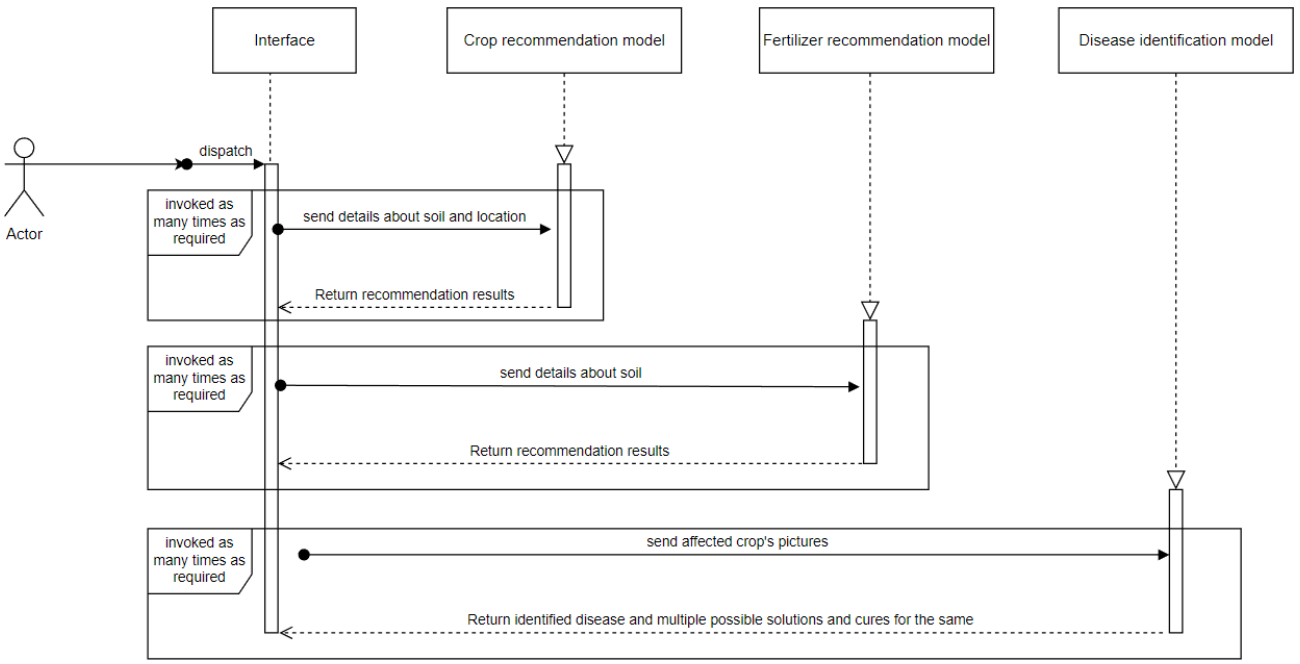


Fig 5.6 Sequence diagram

As explained in the use case diagram, a user can request for any of these services any number of times, the sequence diagram depicts three different models for the three aforementioned features. An interface keeps track of all requests and provides the result back to the user.

First, the user needs to initialize by entering the required details. After that these details are processed in the model and result is generated. This is the same sequence followed for all three features. The sequence in which the features would be use depends on the will of the user.

5.1.5 Activity diagram

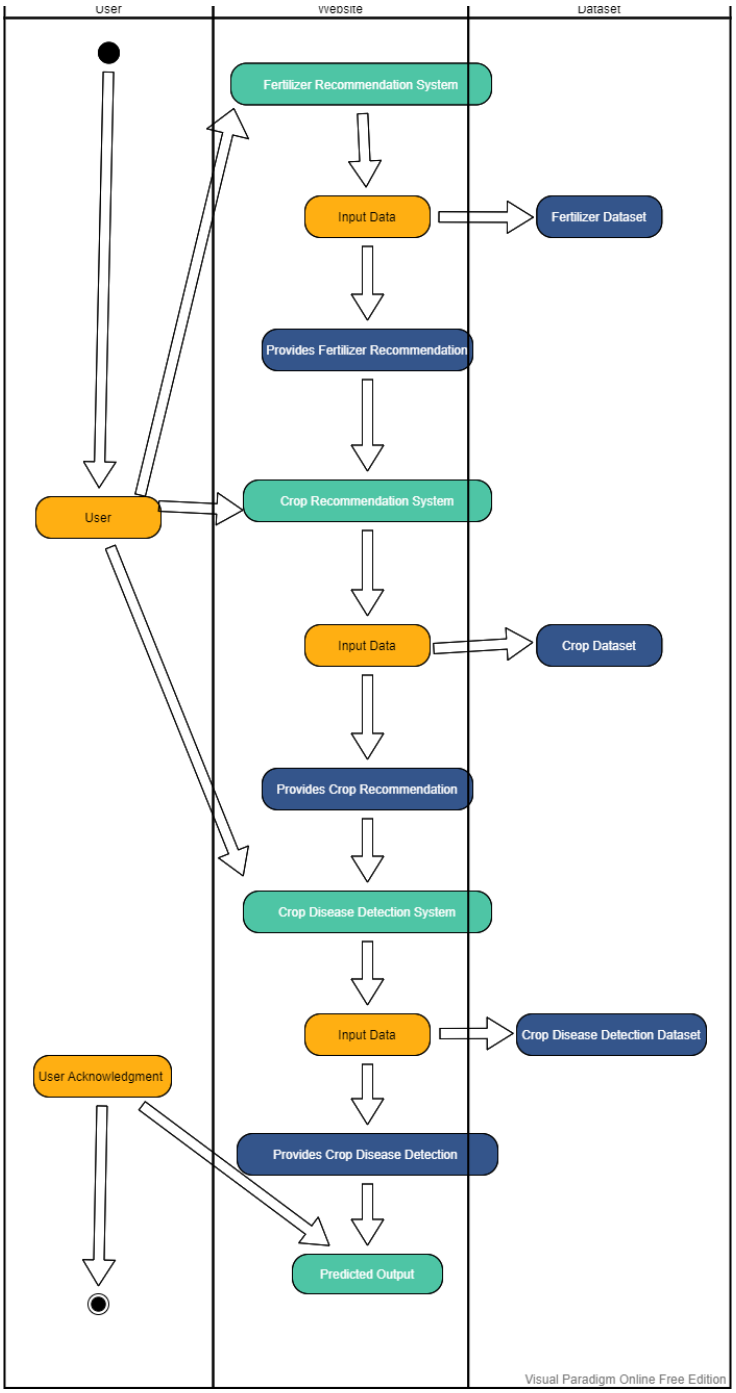


Fig 5.7 Activity diagram

CHAPTER 6

Experimental Setup

Software requirements:

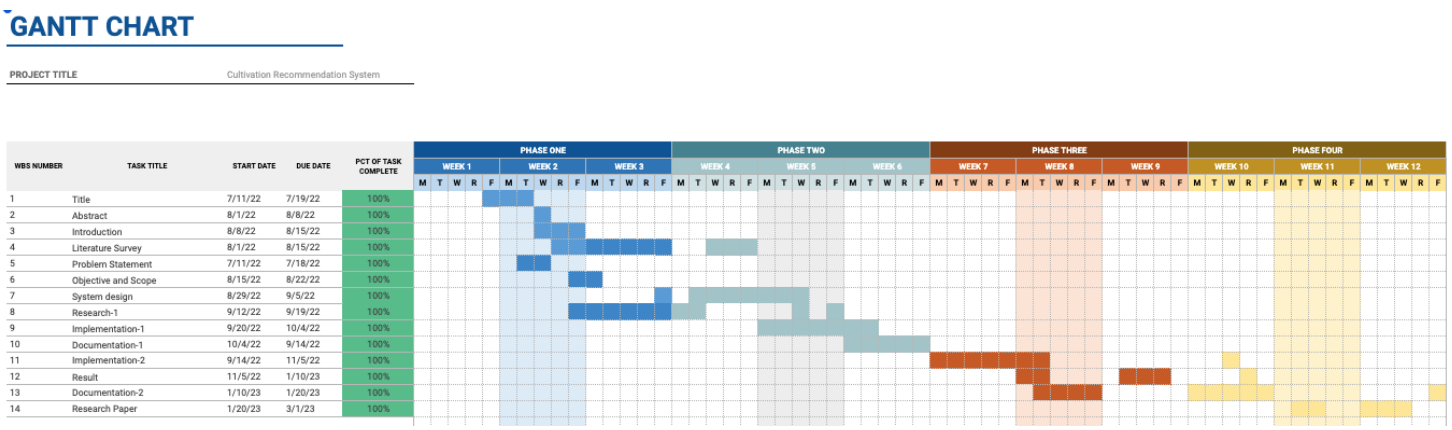
1. Python 3.8 or higher
2. Numpy
3. Pandas
4. Scikit
5. Matplotlib
6. PyTorch
7. Flask

Hardware Requirements:

1. 8 / 16 GB RAM
2. NVIDIA 1050 TI GPU

CHAPTER 7

Project Plan



GANTT CHART

PROJECT TITLE: Cultivation Recommendation System

WBS NUMBER	TASK TITLE	START DATE	DUE DATE	PCT OF TASK COMPLETE
1	Title	7/11/22	7/19/22	100%
2	Abstract	8/1/22	8/8/22	100%
3	Introduction	8/8/22	8/15/22	100%
4	Literature Survey	8/1/22	8/15/22	100%
5	Problem Statement	7/11/22	7/18/22	100%
6	Objective and Scope	8/15/22	8/22/22	100%
7	System design	8/29/22	9/5/22	100%
8	Research-1	9/12/22	9/19/22	100%
9	Implementation-1	9/20/22	10/4/22	100%
10	Documentation-1	10/4/22	9/14/22	100%
11	Implementation-2	9/14/22	11/5/22	100%
12	Result	11/5/22	1/10/23	100%
13	Documentation-2	1/10/23	1/20/23	100%
14	Research Paper	1/20/23	3/1/23	100%

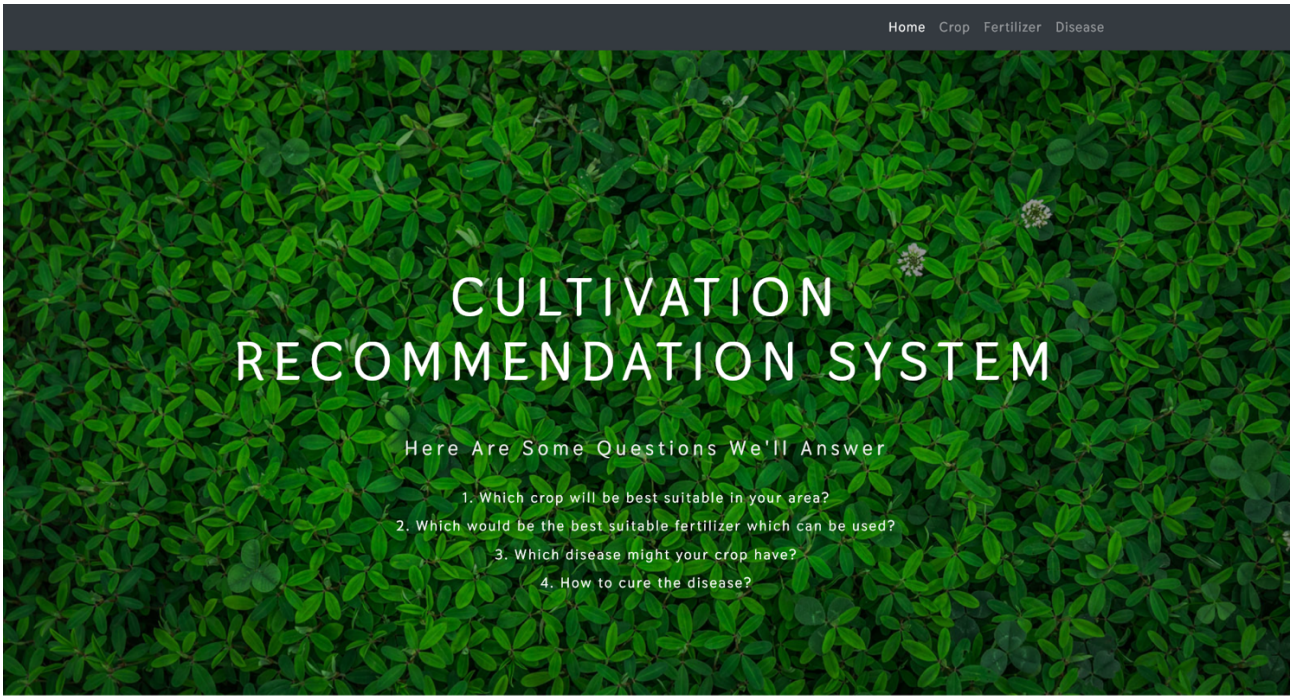
PHASE ONE: WEEK 1, WEEK 2, WEEK 3
PHASE TWO: WEEK 4, WEEK 5, WEEK 6
PHASE THREE: WEEK 7, WEEK 8, WEEK 9
PHASE FOUR: WEEK 10, WEEK 11, WEEK 12

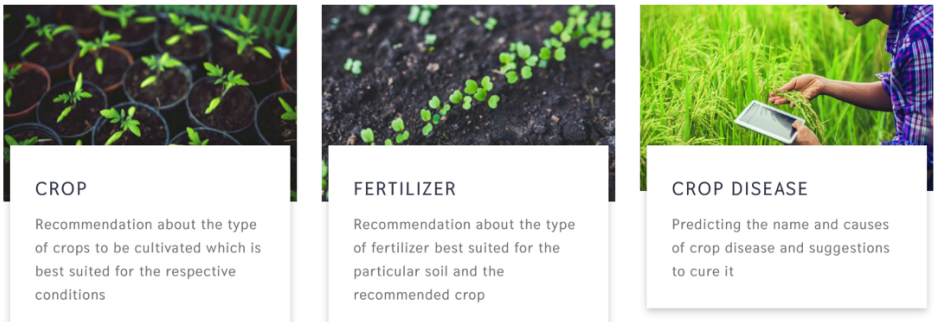
M T W R F M T W R F M T W R F M T W R F M T W R F M T W R F M T W R F M T W R F M T W R F M T W R F M T W R F M T W R F M T W R F

Fig 7.1 Gantt Chart

CHAPTER 8

Results

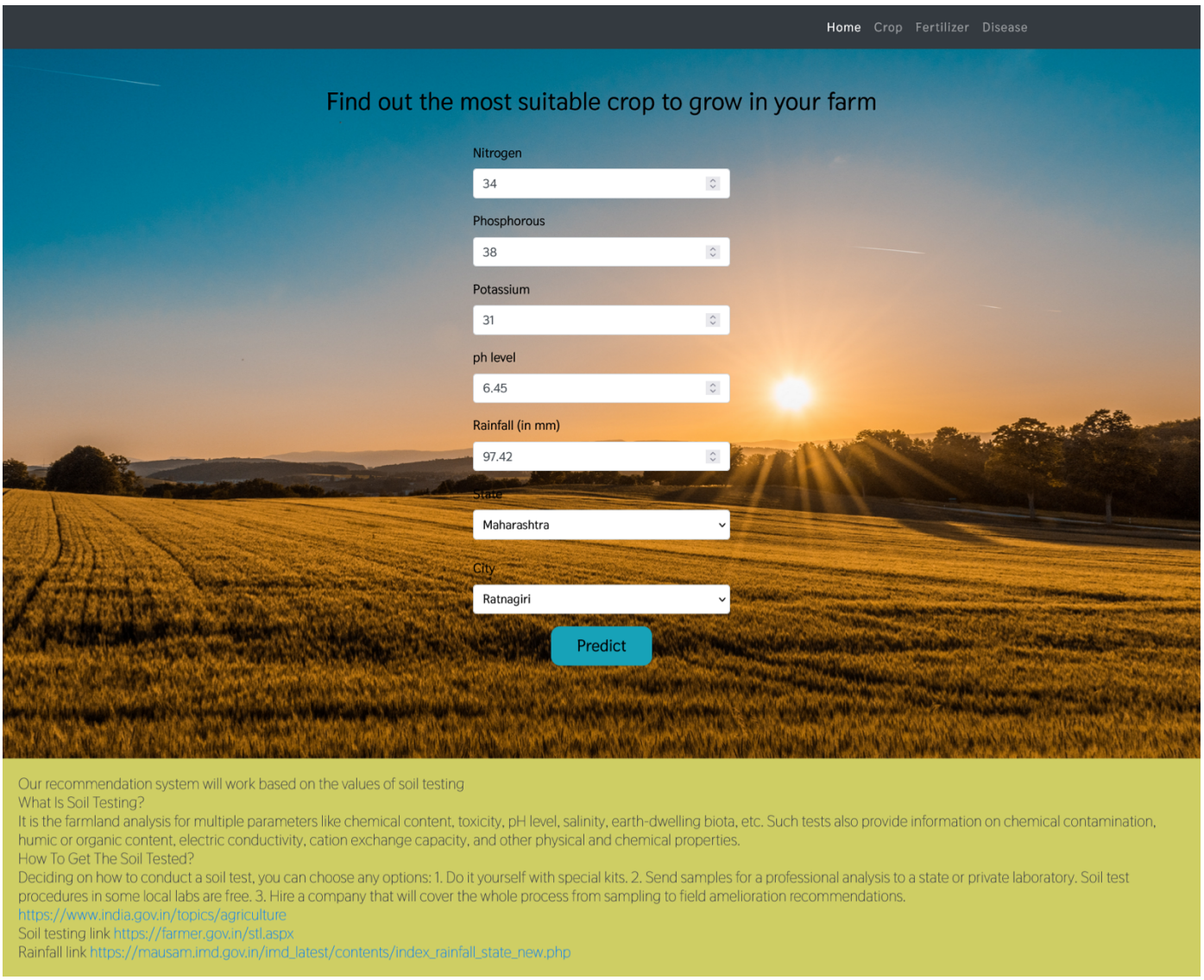




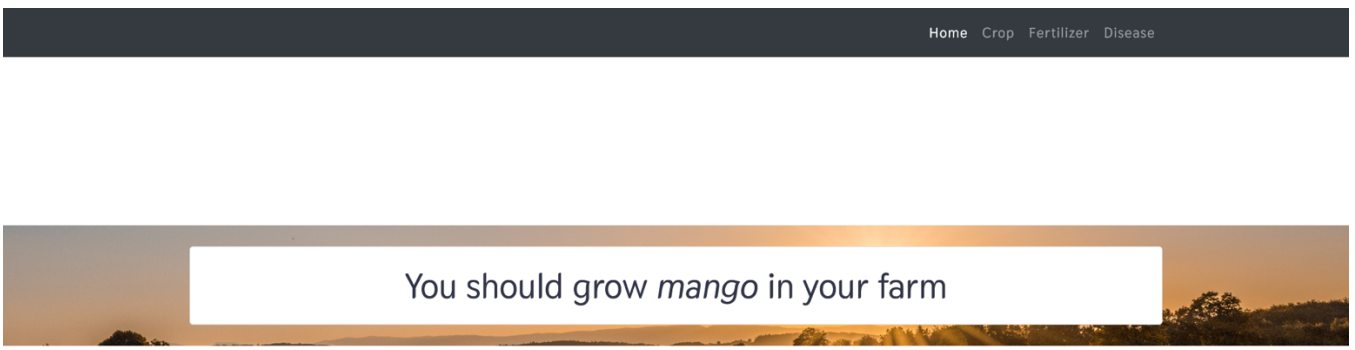
1) Crop Recommendation System:

Sample input (to be taken from the soil report):

Nitrogen: 34 ppm; Phosphorous: 38 ppm; Potassium: 31 ppm; pH: 6.45; Rainfall: 97.42 mm; State: Maharashtra; City: Ratnagiri



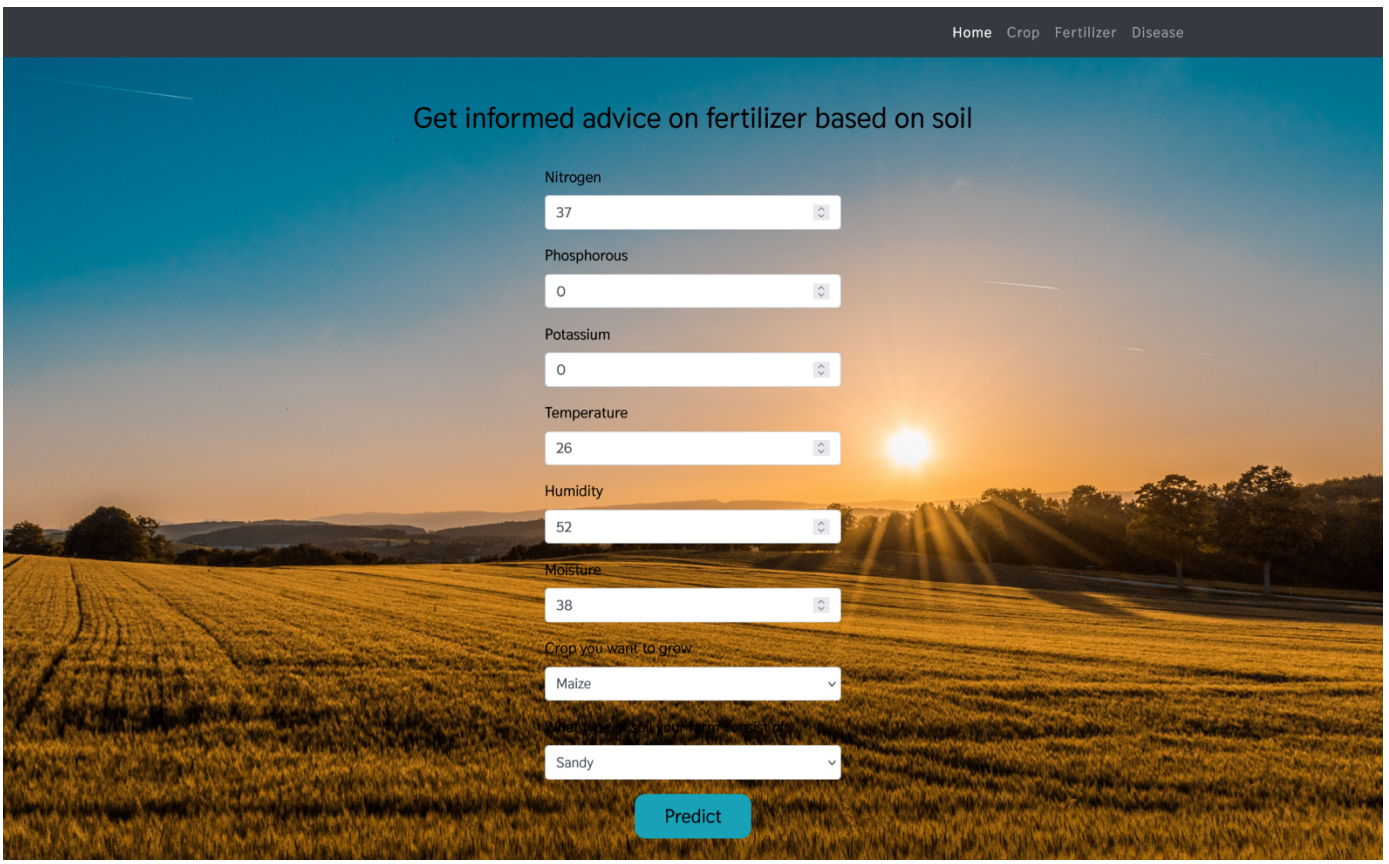
Output by the crop recommender on the basis of the above-mentioned input:



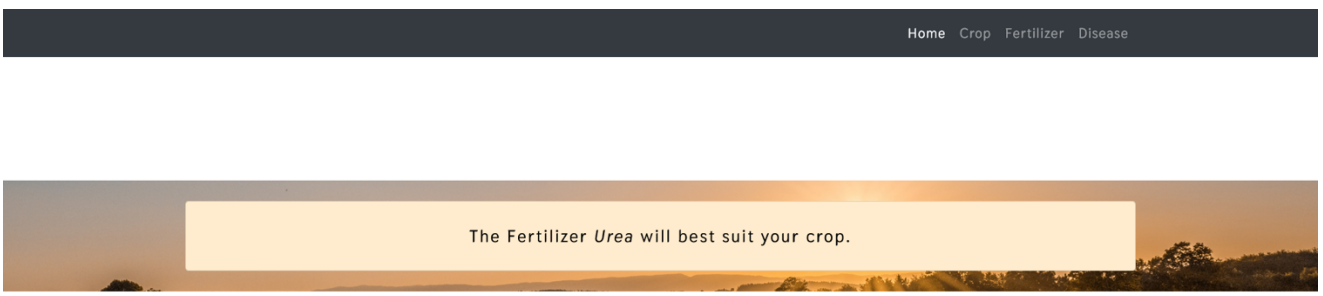
2) Fertilizer recommendation system:

Sample input (to be taken from the soil report):

Nitrogen: 37 ppm; Phosphorous: 0 ppm; Potassium: 0 ppm; Temperature: 26 ºC; Moisture: 38 gm^{-3}; Crop: Maize; Soil: Sandy

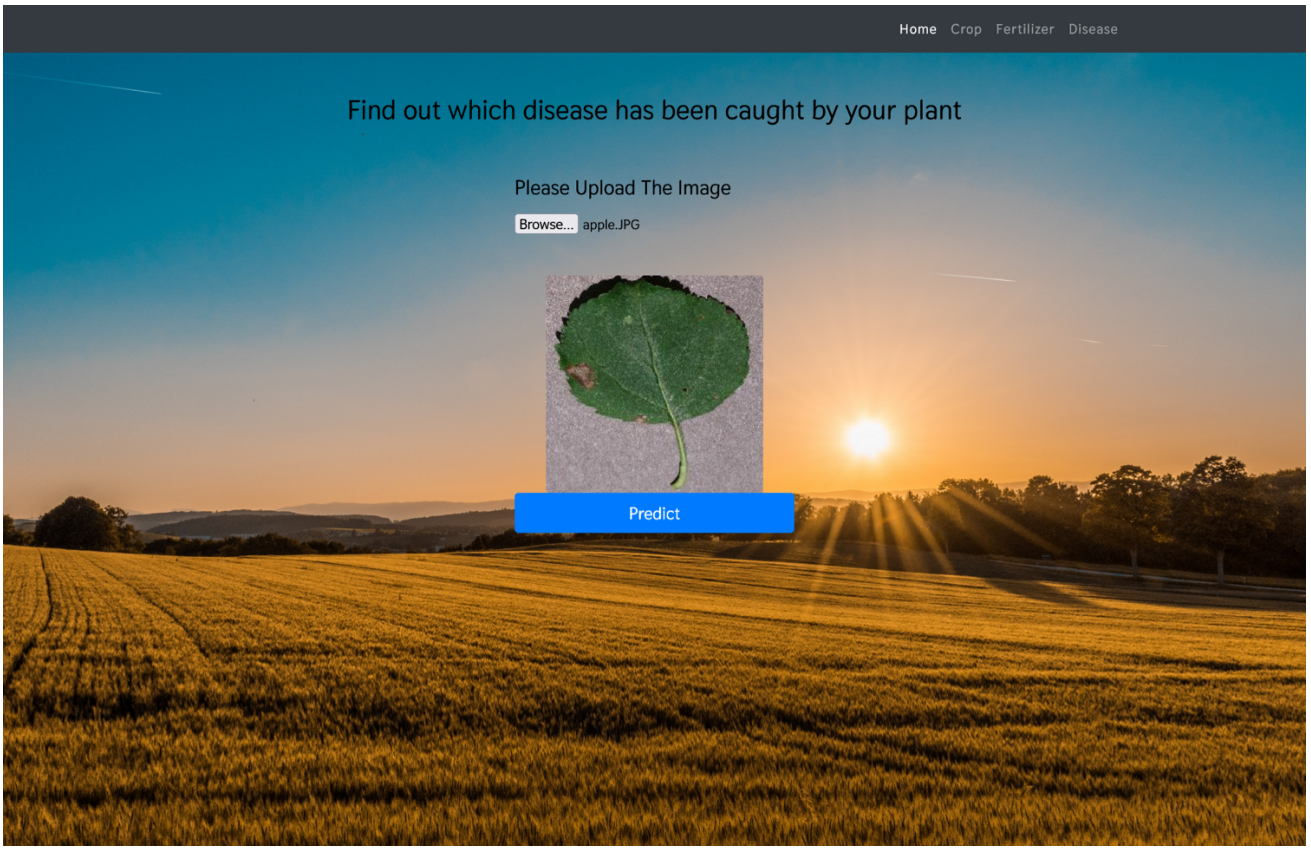


Output by the fertilizer recommender on the basis of the above-mentioned input:

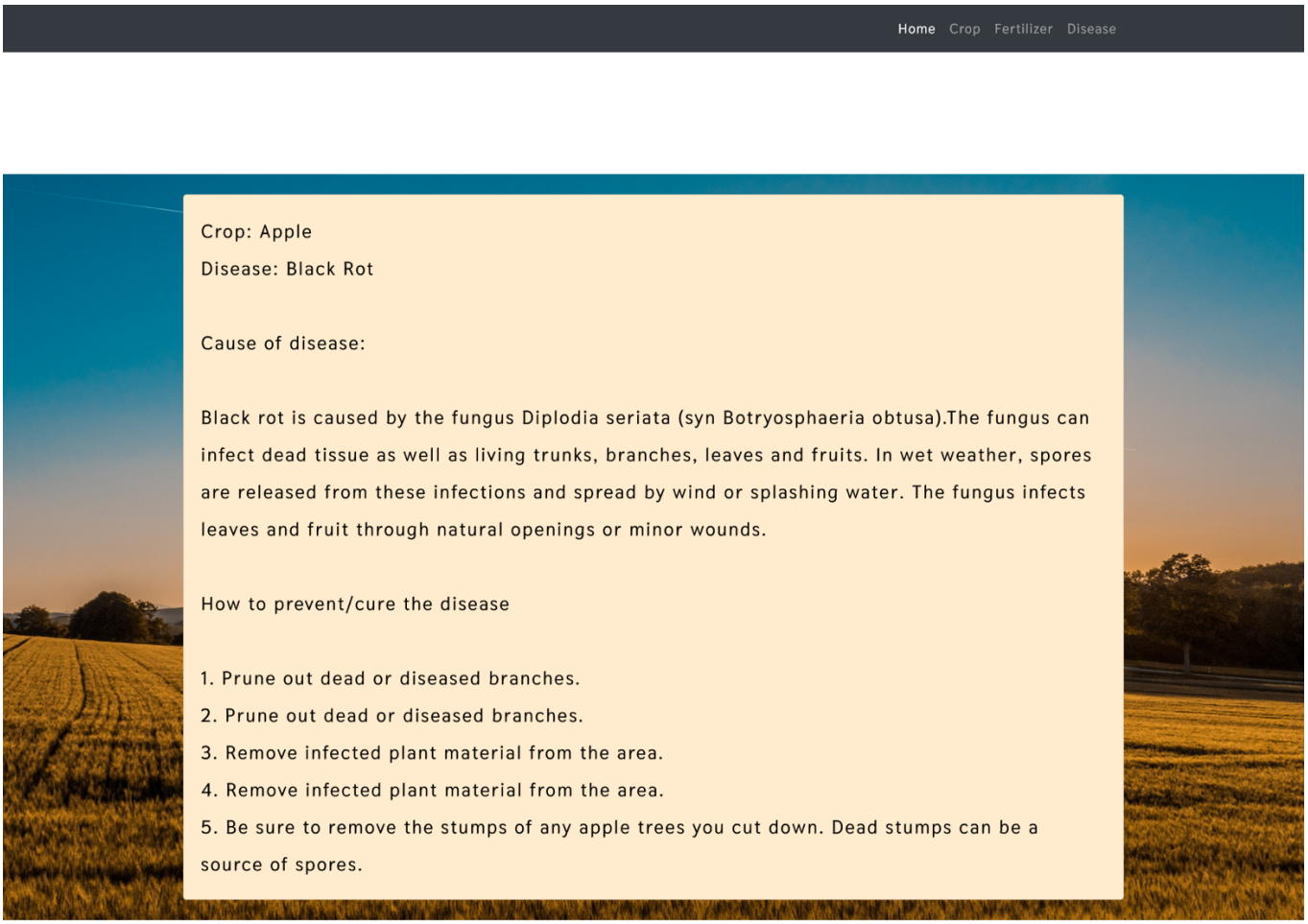


3) Crop disease prediction:

Sample Input:



Output by the disease predicter on the basis of the above-mentioned input image:



CHAPTER 9

Conclusion

1. **Crop recommendation system:** Predicts the favorable crops according to soil nutrient levels, pH, and current weather conditions within an acceptable accuracy range, given a soil test report is available. For the comparative study of different ML models, here's the accuracy comparison:

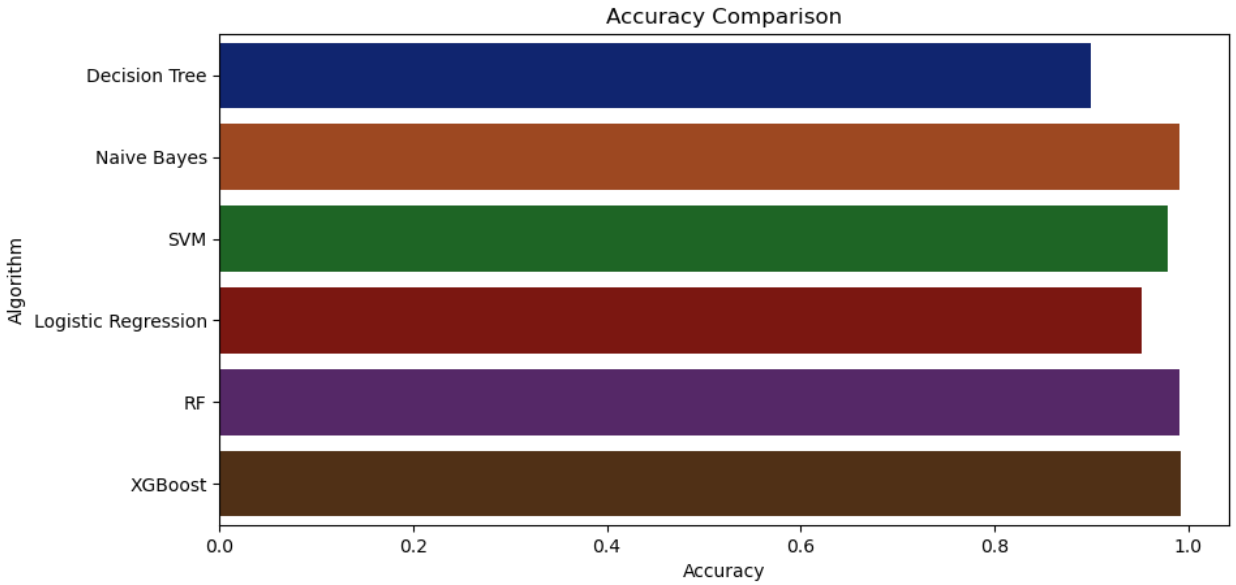


Fig 9. Accuracy Comparison

2. **Fertilizer recommendation system:** Predicts the favorable fertilizers according to soil nutrient level within an acceptable accuracy range.

3. **Crop Disease Identification system:** Classifies different crop diseases using a photo provided by the user, and provides solutions for the mitigating the disease.

CHAPTER 10

Future Scope

1) An extensive on ground research could be conducted to get better dataset with more accurate attributes to enhance the quality of recommendation.

2) This software can be coupled with technicians in service centers to eliminate the dependency of user's literacy.

3) Government's helpful schemes can be dynamically updated on the website.

References

[1] Younes OMMANE, Mohamed Amine RHANBOURI, Hicham CHOUIKH, Mourad JBENE, Ikram CHAIRI, "Machine Learning based Recommender Systems for Crop Selection: A Systematic Literature Review", A Research Gate Preprint Online Publication, 2022.

[2] Sunidhi N, Jalaja S, "AI Based Automatic Crop Disease Detection System", IEEE Xplore 2021.

[3] V Rajesh Kumar, K Pradeepan, S Praveen, M Rohith, V Vasantha Kumar, "Identification of Plant Diseases Using Image Processing and Image Recognition", IEEE Xplore 2021.

[4] Sayed Mazhar Ali, Bhagwan Das, Dileep Kumar, "Machine Learning based Crop Recommendation System for Local Farmers of Pakistan", A Research Gate Preprint Online Publication 2021, Pakistan.

[5] Dhruv Piyush Parikh, Jugal Jain, Tanishq Gupta, Rishit Hemant Dabhade, "Machine Learning Based Crop Recommendation System", International Journal of Advanced Research in Science, Communication and Technology, 2021, Chennai, India.

[6] Melike Sardogan, Adem Tuncer, Yunus Ozen, "Plant Leaf Disease Detection and Classification based on CNN with LVQ algorithm", IEEE Xplore 2020.

[7] Pradeepa Bandara, Thilini Weerasooriya, Ruchirawya T.H., W.J.M. Nanayakkara, Dimantha M.A.C, Pabasara M.G.P., "Crop Recommendation System", International Journal of Computer Applications (0975 – 8887) Volume 175– No. 22, October 2020.

[8] Manav Singhal, Kshitij Verma, Anupam Shukla, "Krishi Ville-Android based solution for Indian agriculture", 2011 Fifth IEEE International Conference on Advanced Telecommunication Systems and Networks (ANTS), 18-21 Dec. 2011, Bangalore, India.

[9] Sharada P. Mohanty, David P. Hughes, and Marcel Salathé, "Using Deep Learning for Image-Based Plant Disease Detection", Frontiers in Plant Science Online Publication 2019, Switzerland.

[10] "Sector wise GDP of India", https://statisticstimes.com/economy/country/india-gdp-sectorwise.php, sourced from Ministry of Statistics and Programme Implementation, 17th June, 2021.

Publications

Comparative Study of Machine Learning models for Crop Recommendation System

Riddhi Narkar[1], Avishkar Dalvi[2], Aarya Totey[3], Radha Rakshe[4], and Prof. D.S. Khachane

A.P. Shah Institute of Technology, Thane, India.
19102003@apsit.edu.in[1], 20202002@apsit.edu.in[2], 19102070@apsit.edu.in[3], 19102067@apsit.edu.in[4]

Abstract. Farming involves a lot of processes and steps to be done for a good harvest. Many of these processes involve decision-making, for example, the type of fertilizer, irrigation, crop, insecticide, etc. to be used. This work explores one such sector-crop recommendation. Crop recommendation. We did a comparative study of different machine learning models on a crop dataset to find out which model is the best for this task. ...

Keywords: machine learning, crop recommendation system, decision-making systems, intelligent systems, precision agriculture

1 Introduction

India has a primary sector economy, which means that the majority of our population is involved in agriculturally based activities and are dependent on crops related business for their livelihood. As per the census in 2011; in India, approximately 118 million people are farmers and 144 billion people are labourers working in agricultural fields [5].
Crop cultivation anywhere in the world depends on various factors and a good yield is directly dependent on studying and knowing these factors prior to harvesting. Precision agriculture is the science of improving crop yields and assisting management decisions using high-technology sensor and analysis tools [4]. Crop recommendation systems are an integral part of precision agriculture [2]. A crop recommendation system is a technology that suggests suitable crop options for farmers based on various factors such as soil quality, climate, soil pH, temperature, rainfall, and other relevant parameters and historical records to provide personalized recommendations for a particular farm or region. This can help farmers increase their yields, optimize resource use, and maximize profits while reducing the risk of crop failure.
In this work, we used a public crop dataset to train different Machine Learning models and found the best model for this task.

2 Dataset

The dataset we used contains a total of 2200 entries. The attributes are N, P, K, temperature, humidity, pH, rainfall, and crop. Here N, P, K are the values of macro-nutrients nitrogen, phosphorous, and potassium respectively and are measured in ppm (parts per million), temperature is measured in °C (degree celsius), humidity is measured in gm^{-3} (grams of water vapour per cubic metre of air), pH is the acidity/basicity of the soil, and rainfall is measured in mm (millimetres). The target variable is label, which has crop names. The data is collected from Indian regions and soil.
There are a total of 22 different crops in the dataset namely, muskmelon, kidneybeans, papaya, pigeonpeas, blackgram, cotton, mothbeans, mungbean, watermelon, orange, mango, banana, rice, pomegranate, chickpea, apple, jute, grapes, lentil, coffee, maize, and coconut.

Table 1. Dataset attributes' details

Attributes	Details
N	Value of Nitrogen in ppm
P	Value of Phosphorous in ppm
K	Value of Potassium in ppm
Temperature	Measured in °C
Humidity	Measured in gm^{-3}
pH	Measured on a scale of 0 to 14
Rainfall	Measured in mm
Label	Target variable (crop name)

3 Methodology

The dataset was split into 3 parts - 80% for training, 10% for testing and 10% for validation. Since this is a categorical type dataset, the first choice was naturally, logistic regression. Linear regression for classification, however, won't be a good choice, as its cost function is not optimal for classification problems [6]. Along with this, we used a decision tree, another well-known and powerful model for classification models.
The dataset we had had many attributes, and since this problem is a classification, SVM was another excellent option. This work was initiated with higher hopes from SVM due to its powerful boon of saving from the 'Curse of Dimensionality'.
The dataset had many factors and as crops need the perfect conditions of rainfall and soil nutrients, it was concluded that these attributes are related to each

other in some or the other way [3]. To test that out, Naive Bayes was also implemented, as it assumes that no attribute is dependent or related to any other attribute and all are completely independent.
Lastly, we implemented a few ensemble techniques for more effective training [1]. Of the ensemble techniques, we used random forest, which uses a forest of a myriad of decision trees; and XGBoost.

3.1 Logistic Regression

Logistic regression is a statistical machine learning model which estimates the probability of an event (in this case, classification) and needs a dataset of independent variables. The output here is a mere probability, and hence, is bounded in the interval [0, 1]. The logistic function is as follows:

$$f(x) = \frac{1}{1 + e^{-x}}$$

$$ln(\frac{f(x)}{1 - f(x)}) = \beta_0 + \beta_1 * x_1 + ... + \beta_k * x_k$$

3.2 Naive Bayes

Naive Bayes, just like Logistic Regression 'naively' assumes complete conditional independence. Naive Bayes calculates the probability of a result assuming strong conditional independence between all attributes. It uses Bayes' theorem for its calculation. Bayes' theorem states that:

$$P(A|B) = \frac{P(A) * P(B|A)}{P(B)}$$

In plain English, Bayes' theorem can be demonstrated as:

$$posterior = \frac{prior * likelihood}{evidence}$$

A Naive Bayes classifier uses this formula to calculate the probability of classification.

3.3 Decision Tree

A decision tree is a non-parametric classification tool and can be used for both regression and classification tasks. It has a tree structure, consisting of a root, internal nodes and leaves. It helps to evaluate different choices between several courses of action. Every internal node is a decision and on the basis of the decision, you travel from the root to the leaves. The leaves are any of the possible outcomes of the classification (target variable values).

3.4 SVM

SVM, an abbreviation of Support Vector Machine is an extremely powerful model for classification, outlier detection and regression. Its performance is independent of the total attributes in a dataset; it even works in cases where the total number of attributes (or dimensions) is greater than the total number of samples. SVM calculates the most optimal boundary between the classes in the dataset to classify.

3.5 Random Forest

It is an ensemble learning technique which creates a forest of 1-level decision trees. It can be used for both regression and classification tasks. A decision tree classifier becomes more vulnerable to overfitting when the tree grows. Hence, to avoid that, only trees of height one are taken in a random forest classifier. This method helps maintain the accuracy of decision trees. The individual trees need to be as uncorrelated to each other as possible to further increase accuracy.

3.6 XGBoost

XGBoost is an open source optimized gradient-boosting library. It is an ensemble technique, and it makes a stronger prediction using multiple weak models and tweaking their weights and improving their performance. XGBoost stands for 'Extreme Gradient Boosting'. It is efficient even in large dataset sizes, It can be used for both classification an regression tasks. In XGBoost, decision trees are created in sequentially. Weights, which are very integral and important in XGBoost, are assigned to all the independent variables. This is then fed into a decision tree and predictions are made. After this first iteration, the weights of wrongly predicted variables are increased and a 2nd iteration is made in a similar fashion. Such ensembling makes for a more precise model.

4 Observations

Ensemble techniques proved to be more powerful than regular techniques. Random Forest and Naive Bayes have the second highest accuracy, out of which Random Forest is an ensemble technique. The other ensemble technique implemented was XGBoost which has the highest accuracy.
Naive Bayes, even being a regular model stood alongside Random Forest. Thus, opposing our previous assumption that the attributes are fairly dependent on each other, Naive Bayes got the second-highest accuracy. Hence, the assumption thus is false, and the attributes, mathematically are quite independent.

5 Results

The resulting accuracies of all models are given below:

Fig. 1. Bar Graph comparing accuracies

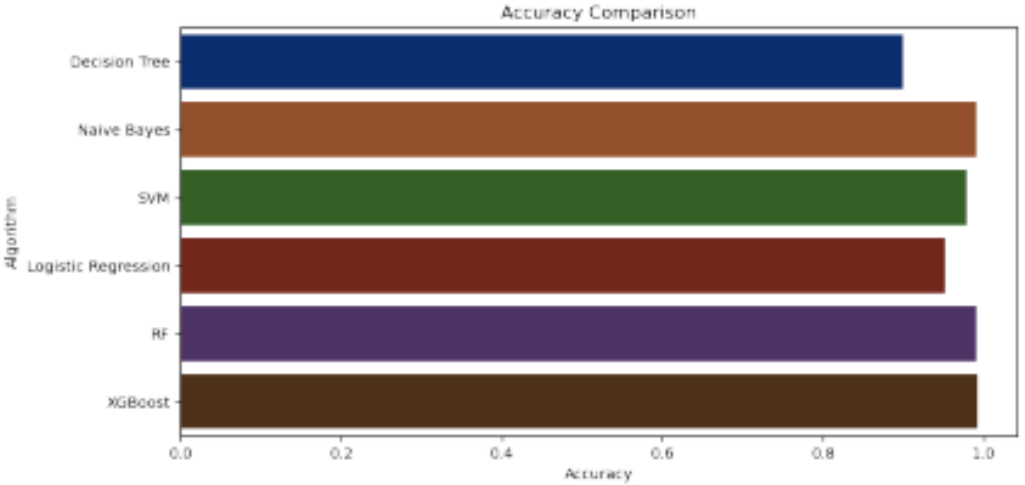


Table 2. Accuracy comparision

Models	Accuracies (in%)
Decision Tree	90.0000%
Naive Bayes	99.0909%
SVM	97.9545%
Logistic Regression	95.22727%
Random Forest	99.09091%
XGBoost	99.3182%

Hence, the best classification machine learning model is XGBoost, for crop recommendation using the dataset we used to implement this classification problem.

References

1. Nidhi H Kulkarni, Dr. G N Srinivasan, Dr. B M Sagar, Dr.N K Cauvery, "Improving Crop Productivity Through A Crop Recommendation System Using Ensembling Technique", 3rd IEEE International Conference on Computational Systems and Information Technology for Sustainable Solutions 2018.

2. Pradeepa Bandara, Thilini Weerasooriya, Ruchirawya T.H., W.J.M. Nanayakkara, Dimantha M.A.C, Pabasara M.G.P., "Crop Recommendation System", International Journal of Computer Applications (0975 – 8887) Volume 175– No. 22, October 2020.26.
3. Sayed Mazhar Ali, Bhagwan Das, Dileep Kumar, "Machine Learning based Crop Recommendation System for Local Farmers of Pakistan", A Research Gate Preprint Online Publication 2021, Pakistan.
4. Dhruv Piyush Parikh, Jugal Jain, Tanishq Gupta, Rishit Hemant Dabhade, "Machine Learning Based Crop Recommendation System", International Journal of Advanced Research in Science, Communication and Technology, 2021, Chennai, India.
5. "Sector wise GDP of India", https://statisticstimes.com/economy/country/india gdpsectorwise.php, sourced from Ministry of Statistics and Programme Implementation, 17th June, 2021.
6. Younes OMMANE, Mohamed Amine RHANBOURI, Hicham CHOUIKH, Mourad JBENE, Ikram CHAIRI, "Machine Learning based Recommender Systems for Crop Selection: A Systematic Literature Review", A Research Gate Preprint Online Publication, 2022.