Farmer Assistant Application

A project report submitted in fulfillment of the
requirements for the degree of

Bachelor of Computer Engineering (Sem -VlII)

by

Pratik Harde (Roll No. 8871)
Ibin Babu (Roll No. 8872)
Ananya Sharma (Roll No. 8908)

Under the guidance of **Prof.**
Prachi Patil

DEPARTMENT OF COMPUTER ENGINEERING

Fr. Conceicao Rodrigues College of Engineering, Bandra (W), Mumbai – 400050

University of Mumbai
(2022-2023)

This work is dedicated to my family.
I am very thankful for their motivation and support.

Internal Approval Sheet

CERTIFICATE

This is to certify that the project entitled **" Farmer Assistant App "** is a bonafide work of **Pratik Harde (8871), Ibin Babu (8872), Ananya Sharma (8908)** submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of **Bachelor of Engineering** in **Computer Engineering (Sem VII).**

Prof. Prachi Patil
Supervisor/Guide

Dr. Sujata Deshmukh
Head of Department

Dr. Surendra Rathod
Principal

Approval Sheet

Project Report Approval

This project report entitled by **Farmer Assistant App** by **Pratik Harde (8871), Ibin Babu (8872), Ananya Sharma (8908)** is approved for the degree of Bachelor of Computer Engineering.

Examiners

1.———————————— 2.————————

Date: 5 Nov 2022

Place:

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we 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. We 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.

Pratik Harde (Roll No. 8871) **(sign)** _____

Ibin Babu (Roll No. 8872) **(sign)** _____

Ananya Sharma (Roll No. 8908) **(sign)** _____

Date: 29 March, 2023

Abstract

In India, agriculture is a significant source of both income and employment. The most frequent issue that Indian farmers encounter is that they choose the incorrect crop for their soil, use the incorrect fertiliser, and are unable to identify plant diseases that are caused by their plants. As a result, they will see a major decline in productivity. Farmers' problems will be resolved via the Farmer Assistant App. In order to suggest the optimum crop to farmers, as well as fertiliser suggestions based on site-specific characteristics, precision agriculture uses research data on soil qualities, soil types, and crop production statistics. By doing this, crop selection errors become less frequent and productivity rises. The proposed recommendation and prediction system in this study uses ML models as learners to highly accurately and effectively select a crop for the site-specific factors. The mechanism for recommending fertiliser is also entirely based on Python logic. Here, we compare the data (the crop's ideal nutrients) with the user-entered information. The most variable nutrient is then classified as HIGH or LOW, and recommendations are then retrieved in accordance with those results.

Contents

List of Figures

List of Tables

Glossary

SVM	Support Vector Machines
RF	Random Forest
DFD	Data Flow Diagrams
ML	Machine Learning
KNN	K-Nearest Neighbor

Chapter 1

Introduction

A farmer's decision regarding which crop to cultivate is typically influenced by his knowledge as well as unimportant variables like the desire to make quick money, ignorance of market demand, exaggeration of a soil's ability to support a particular crop, and so forth. The farmer's family's financial situation could be severely strained if he makes the wrong choice. Maybe this is one of the numerous factors contributing to the innumerable farmer suicide cases that the media reports on every day. Such a wrong judgement would have detrimental effects on not only the farmer's family but the entire economy of an area in a country like India, where agriculture and associated sectors contribute to about 20.4% of its Gross Value Added (GVA). Because of this, we have determined that a farmer's decision on which crop to cultivate during a specific season is a very serious one. The urgent requirement is to create a system that might offer Indian farmers predictive information so they could choose which crop to produce with knowledge.

In light of this, we suggest a system, an intelligent system, that would evaluate soil characteristics (pH value, N, P, K, soil type, and nutrients concentration) as well as environmental factors (rainfall, temperature, geographic location in terms of state), before advising the user on the crop that would grow best. Additionally, a fertiliser recommendation based on the ideal nutrients of the produced crops is also made.

1.1 Motivation

As about 70% of the population depends on agriculture, it serves as th e

economic foundation for developing nations like India. India's econo my and employment are heavily reliant on agriculture.

Indian farmers frequently struggle with choosing the right crop to supp ort
their soil's needs and which fertiliser to utilise for their particular crop. This problem of the farmers has been addressed through precision agriculture.

1.2 Objectives

- To develop a reliable model that can accurately estimate crop sustainability in a given state under specific climatic and soil conditions
- Give advice on the best crops to grow in the area so that the farmer doesn't suffer any losses.
- Suggest fertilizers for crops with supported chemical properties.

Chapter 2

Literature Review

Smart Farming [1]

An affordable IOT + ML design for smart farming with several applications, a system for water management systems, and an improvement to the current irrigation techniques are all included. An IoT and ML-based farming system constantly keeps farmers informed of the potential weather patterns and offers the best recommendations for crops and irrigation techniques, resulting in increased production.

Smart Management of Crop Cultivation using IoT and Machine Learning [2]

The most suited crop to grow in that climate is predicted by a smart system that takes into account measured parameters (temperature, humidity), as well as other parameters (soil type, farm location, rainfall), into account.

High Resolution Mapping of Soil Properties Using Remote Sensing Variable in South-Western Burkina Faso [3]

The four modules MODIFIED SUPPORT VECTOR REGRESSION, a well-known machine learning technique, are used to determine real-time sampling of soil parameters. The modules consist of a sensor connected to an IoT device, an agricultural cloud, an analysis of real-time sensor data, and an agricultural user interface (AUI).The first module consists of a NodeMCU portable IoT device with environmental sensors, including pH and soil moisture sensors. Storage is part of the agri cloud module. Using a modified support vector machine method, the real-time data module is analysed to process the various types of crops and small plants. It is a simple online interface called Agri-user interface. Thus, using a modified support vector machine

algorithm, a farmer can determine the types of crops and small plants that can be cultivated on their land using the help of the soil's qualities.

Internet of things (IoT) applications to fight against COVID-19 pandemic [4]

Internet of Things (IOT) and machine learning are two of the proposed new technologies. Using an IOT system, real-time data from the field may be gathered. The trained model receives input from the field area's collected data. Then, using the data, the trained model makes the predictions. The model's output significantly aids in planting the appropriate crops in the targeted field area.

Classification of Soil and Crop Suggestion [5]

Determines a model is proposed for predicting the kind of soil and suggests a suitable crop that can be cultivated in that soil. Several machine learning techniques, including KNN, SVM, and logistic regression, have been used to test the model. Compared to other models, the accuracy of the current model is highest.

Smart Agriculture Using WSN [6]

Proposed a smart agriculture system based on IOT and deep learning. With the aid of a wireless sensor network, this system keeps an eye on and gathers data on the soil characteristics from the field. After then, the data is uploaded to the cloud. Finally, by anticipating the crop that will be grown for the following crop cycle, the systems advise farmers on the appropriate irrigation strategies. The farmers will receive this information via SMS. Temperatures of the soil, the atmosphere, and the humidity are among the characteristics. This technique advises enhancing the efficiency even further by forecasting when it will be appropriate to apply pesticides, fertiliser, and manures.

Intelligent Agriculture System To Assist Farmers In Smart Decision [7]

The suggested approach will help farmers choose the best crop to cultivate based on a range of geographical and environmental parameters. By substituting intuition and inherited knowledge with more trustworthy data-driven ML models, the ML and IoT based recommendations would considerably enlighten the farmer and assist

them in minimising expenses and making strategic decisions. This enables a scalable, trustworthy solution to a significant issue affecting hundreds of millions of people.

Chapter 3

Problem Statement

3.1 Drawbacks of Current Solutions

The fact that each author of each publication concentrated on one factor (weather or soil) for predicting the appropriateness of crop development was one weakness we found in all of these important published papers. But in our opinion, the best and most accurate prediction should be made by concurrently taking into account both of these aspects. This happens frequently because even if a particular soil type may be ideal for supporting a certain crop variety, the yield will suffer if the local climate doesn't appear to be favourable for that crop kind.

3.2 Solution to the Problem

By proposing an effective crop recommendation system that accounts for all relevant factors, such as temperature, rainfall, location, and soil quality, we hope to eliminate the above shortcomings. Giving farmers crop suggestions is the main task that this system is primarily concerned with carrying out. In order to give the user a simple and trustworthy knowledge of the insight to determine and plant the crops, we also supply the fertilisers to be used for crops grown in various states.

Chapter 4

Project Design

4.1 Block Diagram

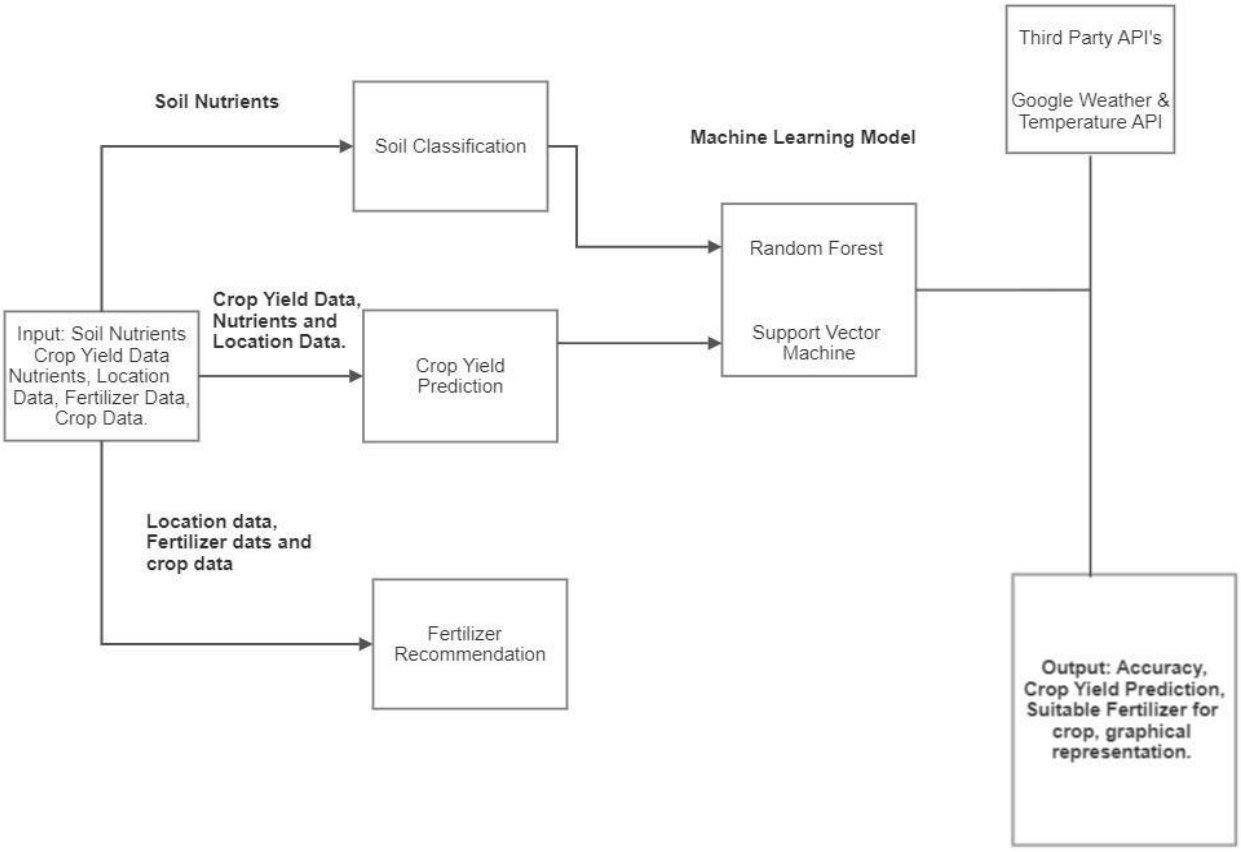


Fig 4.1 Block Diagram for Crop and Fertilizer Recommendation

Crop recommendation, fertiliser recommendation, and disease prediction are the three goals of our method. We are utilising an API key to retrieve the data for the fertiliser recommendation. The system will provide recommendations and predictions based on the nutrients dataset.

4.2 Algorithms

4.2.1 Decision Tree

A non-parametric supervised learning technique for classification and regression is called a decision tree (DT). The objective is to learn straightforward decision rules derived from the data features in order to build a model that predicts the value of a target variable. A fractional approximation of a tree can be thought of. For instance, in the example below, using a series of if-then-else decision rules, decision trees learn from data to approximate a sine curve.

4.2.2 Support Vector Machine

In an N-dimensional space (N being the number of features), the support vector machine algorithm seeks to locate a hyperplane that clearly categorises the data points.

There are a variety of different hyperplanes that might be used to split the two classes of data points. Finding a plane with the greatest margin—that is, the greatest separation between data points from both classes—is our goal. Maximizing the margin distance adds some support, increasing the confidence with which future data points can be categorised.

4.2.3 Logistic Regression

When classifying observations into a collection of discrete classes, the classification algorithm logistic regression is utilised. Email spam or not spam, online fraud or not fraud, and malignant or benign tumours are a few instances of classification issues.

To convert its output into a probability value, logistic regression uses the logistic sigmoid function.

What are the types of logistic regression

1. Binary (eg. Tumor Malignant or Benign)
2. Multi-linear functions failsClass (eg. Cats, dogs or Sheep's)

The machine learning technique known as logistic regression, which is based on the probability notion and used to solve classification problems, is used to analyse data in a predicted manner.

4.2.4 Random Forest

Supervised machine learning algorithms like random forest are frequentl y employed in classification and regression issues. On various samples, it constructs

vote for regression. The Random Forest Algorithm's ability to handle data s ets

with both continuous variables,as in regression and categorical variables, as in classification, is one of its most crucial qualities. In terms of classification issues, it delivers superior outcomes.

To precisely address the issue of high-variance in Decision Trees, Random Forests was created.You're not just training one Decision Tree, as the name implies—you're training an entire forest! A forest of Bagged Decision Trees in this instance.

Random Forests algorithm follows these steps:

1. Starting with the original dataset, generate N bagged samples of size n, where n is the number of samples to be created.
2. With each of the N bagged datasets as input, train a decision tree. Don't, however, look at every feature in the dataset while doing a node split. Choose M features at random from the entire training set of features. Then, using impurity metrics like Gini Impurity or Entropy, choose the optimal split.
3. Combine the outcomes of each decision tree into a single output.
4. If you're working on a regression problem, average the results for each observation, as produced by each tree.
5. If you're working on a regression problem, make a majority decision for each observation across all trees.

Disease Prediction Diagram

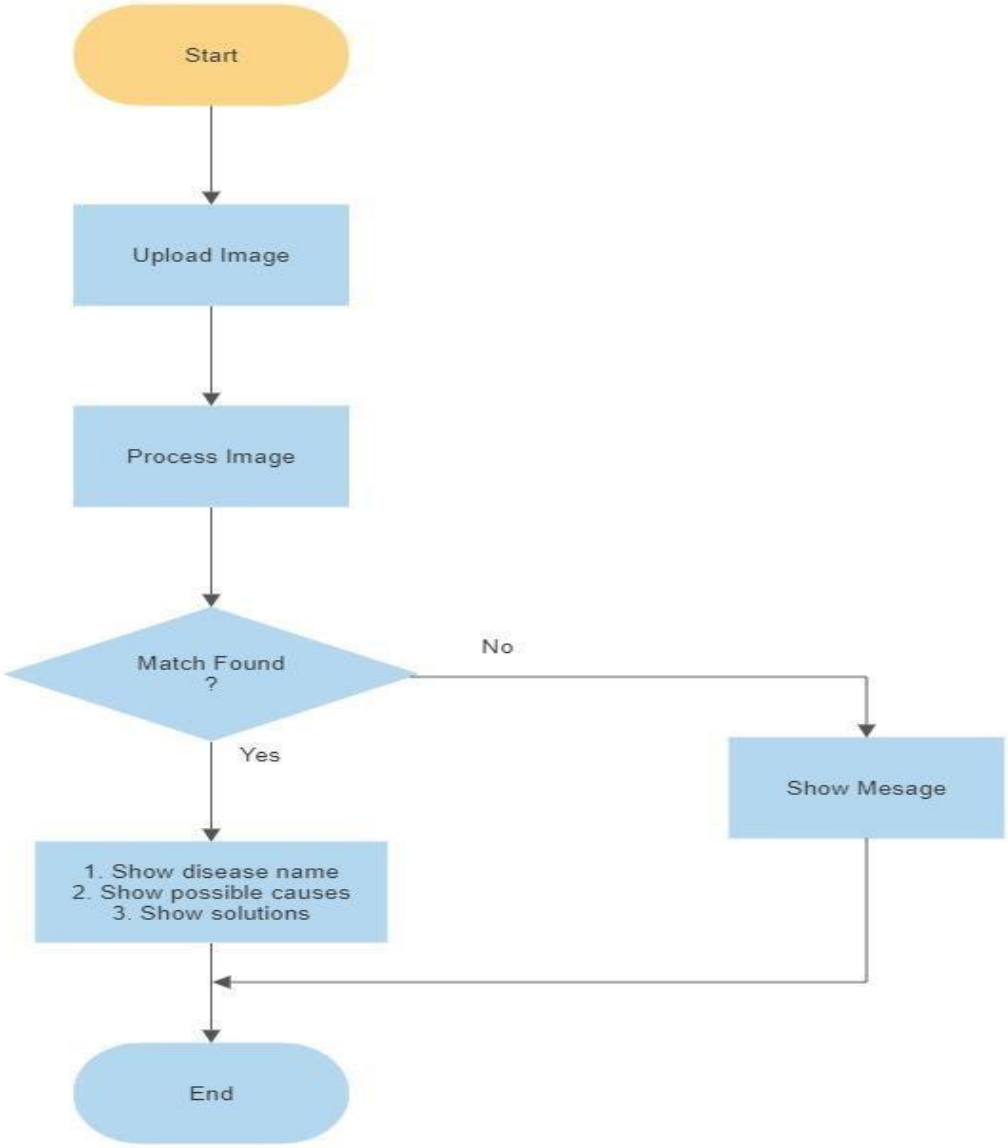


Fig. 4.2 Disease Prediction

Crop Recommendation Diagram

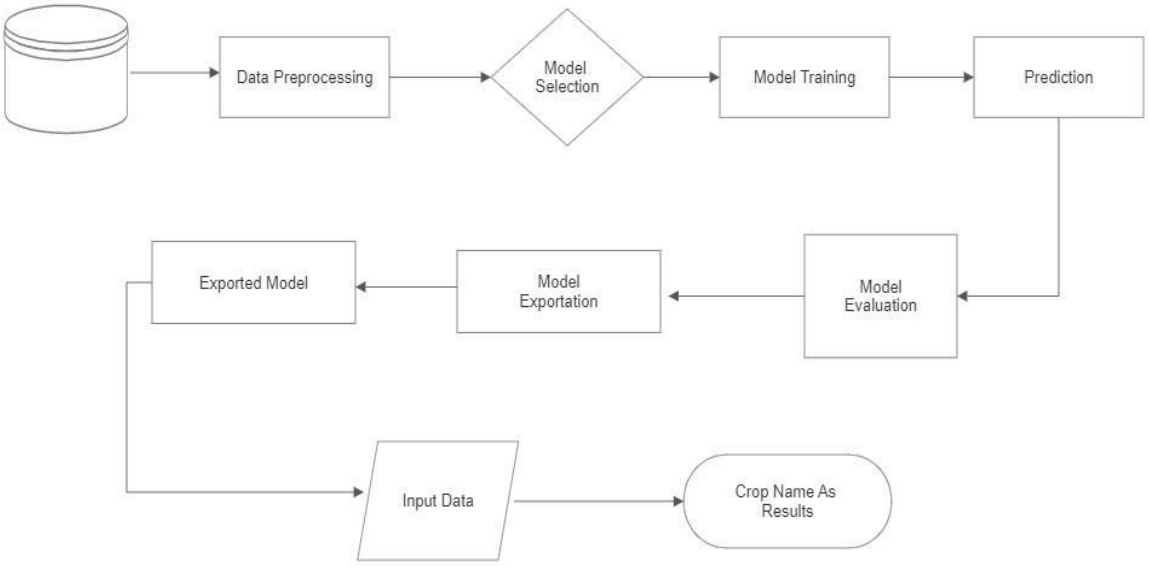


Fig. 4.3 Crop Recommendation

DFD Level 0 and DFD Level 1

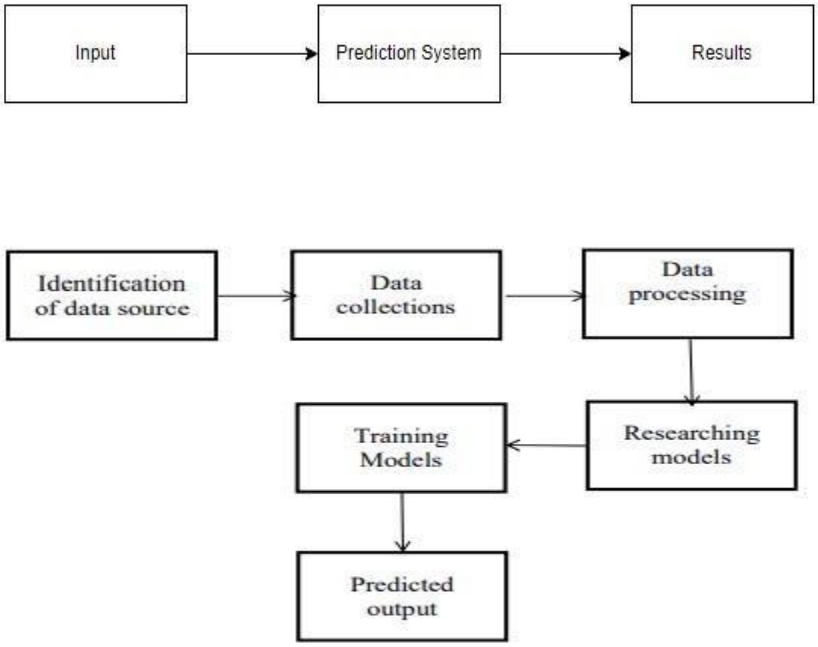


Fig. 4.4 DFD Level 0 and DFD Level 1

4.3 UML Diagram

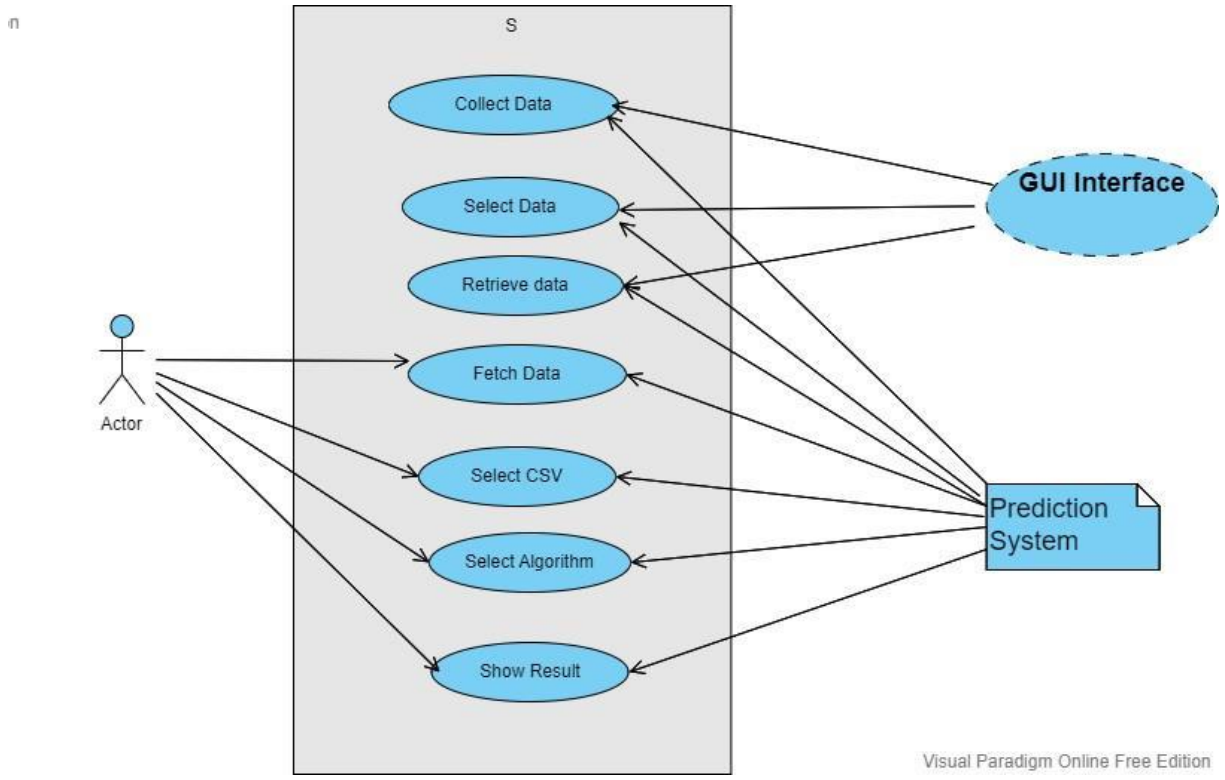


Fig 4.5 UML Diagram

4.3.2 Architecture

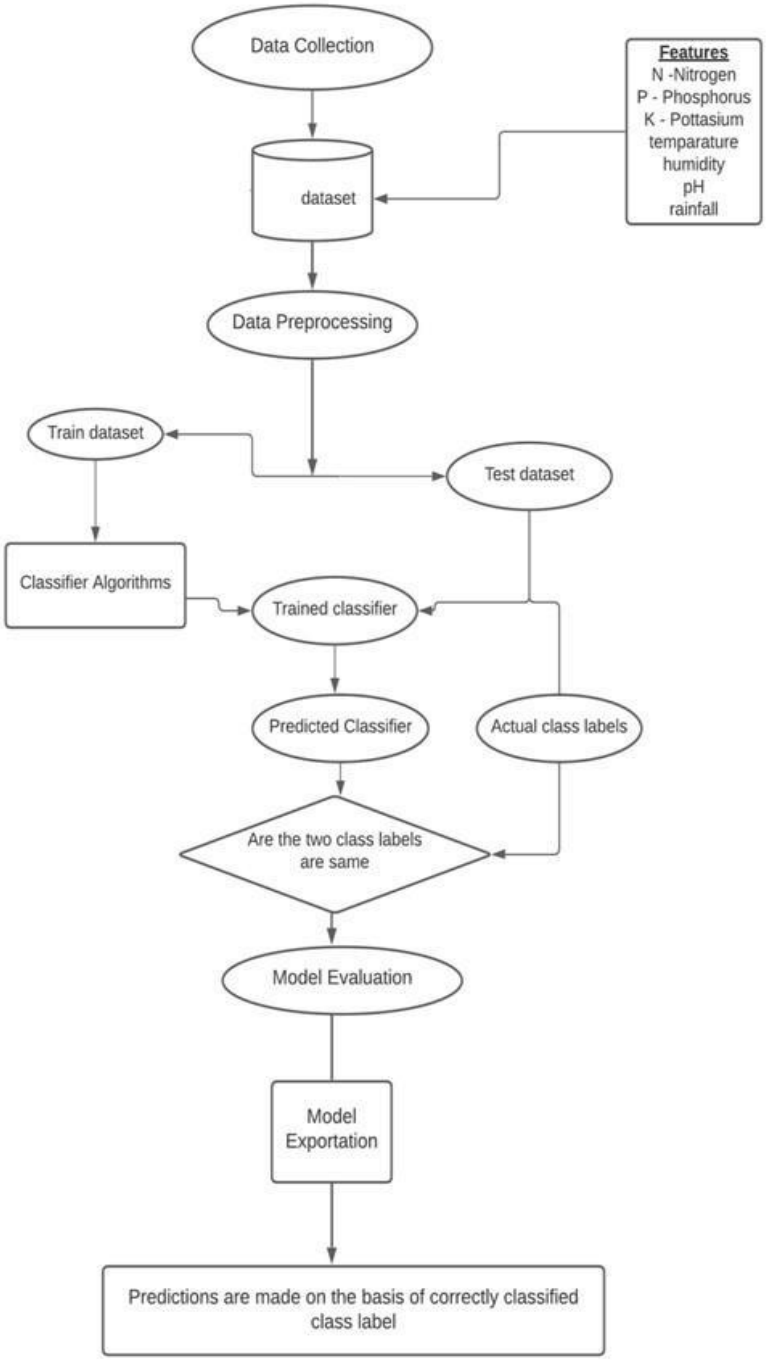


Figure 4.7 : Architecture

4.3 Dataset

Datasets include:-

- Crop Recommendation dataset
- Fertilizer dataset
- Plant Disease dataset
- State -District Wise Crop dataset A brief description of the datasets:
- **Crop Recommendation dataset :** In kg per hectare, this information shows the yield for 18 important crops farmed in every state. When the yield is 0, it means that the crop is not grown in that stage. containing a total of 2201 items in the dataset.

1	N	P	K	temperatu	humidity	ph	rainfall	label
2	90	42	43	20.87974	82.00274	6.502985	202.9355	rice
3	85	58	41	21.77046	80.31964	7.038096	226.6555	rice
4	60	55	44	23.00446	82.32076	7.840207	263.9642	rice
5	74	35	40	26.4911	80.15836	6.980401	242.864	rice
6	78	42	42	20.13017	81.60487	7.628473	262.7173	rice
7	69	37	42	23.05805	83.37012	7.073454	251.055	rice
8	69	55	38	22.70884	82.63941	5.700806	271.3249	rice
9	94	53	40	20.27774	82.89409	5.718627	241.9742	rice
10	89	54	38	24.51588	83.53522	6.685346	230.4462	rice
11	68	58	38	23.22397	83.03323	6.336254	221.2092	rice
12	91	53	40	26.52724	81.41754	5.386168	264.6149	rice
13	90	46	42	23.97898	81.45062	7.502834	250.0832	rice
14	78	58	44	26.8008	80.88685	5.108682	284.4365	rice
15	93	56	36	24.01498	82.05687	6.984354	185.2773	rice
16	94	50	37	25.66585	80.66385	6.94802	209.587	rice
17	60	48	39	24.28209	80.30026	7.042299	231.0863	rice
18	85	38	41	21.58712	82.78837	6.249051	276.6552	rice
19	91	35	39	23.79392	80.41818	6.97086	206.2612	rice
20	77	38	36	21.86525	80.1923	5.953933	224.555	rice

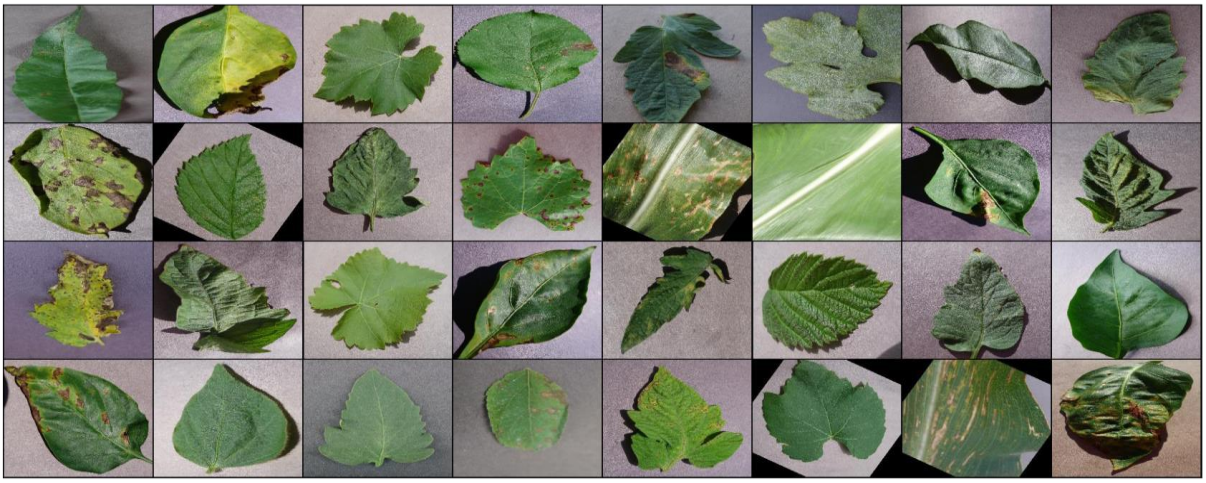
- **Fertilizer dataset** : The columns in this dataset have the following attributes: State, Nitrogen, Phosphorous, Potassium, Average Ph Level, and Soil Moisture Content.

There are 25 total records in the collection.

	Crop	N	P	K	pH	soil_moisture
0	rice	80	40	40	5.5	30
3	maize	80	40	20	5.5	50
5	chickpea	40	60	80	5.5	60
12	kidneybea	20	60	20	5.5	45
13	pigeonpea	20	60	20	5.5	45
14	mothbean	20	40	20	5.5	30
15	mungbean	20	40	20	5.5	80
18	blackgram	40	60	20	5	60
24	lentil	20	60	20	5.5	90
60	pomegran	20	10	40	5.5	30
61	banana	100	75	50	6.5	40
62	mango	20	20	30	5	15
63	grapes	20	125	200	4	60
66	watermelc	100	10	50	5.5	70
67	muskmelo	100	10	50	5.5	30
69	apple	20	125	200	6.5	50
74	orange	20	10	10	4	60
75	papaya	50	50	50	6	20
88	coconut	20	10	30	5	45

- **Plant Disease dataset:** 76100 items in this dataset include photos of illness leaves and 38 distinct classes.

  For instance, Apple->Apple scab, Apple black rot, and Apple healthy



4.4 Hardware and Software

- Hardware System Configuration:
    - ➢ Processor: 2 gigahertz (GHz) or faster processor.
    - ➢ RAM: 4 gigabyte (GB) for 32-bit or 4 GB for 64-bit.
    - ➢ Hard disk space: => 16GB.
- Software Configuration:
    - ➢ Operating System: Windows XP/7/8/8.1/10, Linux and Mac
    - ➢ Coding Language: Python.
    - ➢ Tools: Pandas, Numpy ,Seaborn , Pickle,Scikit-learn, Pytorch & ResNet.
    - ➢ Framework: Flask.
    - ➢ Other Tools : HTML, CSS, Bootstrap.
- Tools and Library used
    - ➢ Library-torch, Pandas, Numpy, Matplotlib, Seaborn, Pickle, Scikitlearn, Pytorch and resnet.
    - ➢ Tool- Flask

Chapter 5

Implementation Details

5.1. Methodology Data Analysis

An analysis of the data is one of the first tasks the system carries out during implementation. We did this in an effort to determine whether there were any connections between the different attributes that were available in the dataset. Acquisition of Training Dataset: Any machine learning algorithm's accuracy is based on the training dataset's precision and the amount of parameters it uses. In this project, we carefully chose the parameters that would produce the best results after analysing different datasets obtained from the government website and Kaggle. Environmental indicators have been taken into account in numerous studies in this sector to predict agricultural sustainability, while yield has also been a prominent component in some and just economic variables in others. To give the farmer the most accurate and trustworthy advice on which crop will be best for his land, we have attempted to combine environmental criteria like rainfall, temperature, ph , soilnutrients, soil type, and location with economic parameters like production and yield.

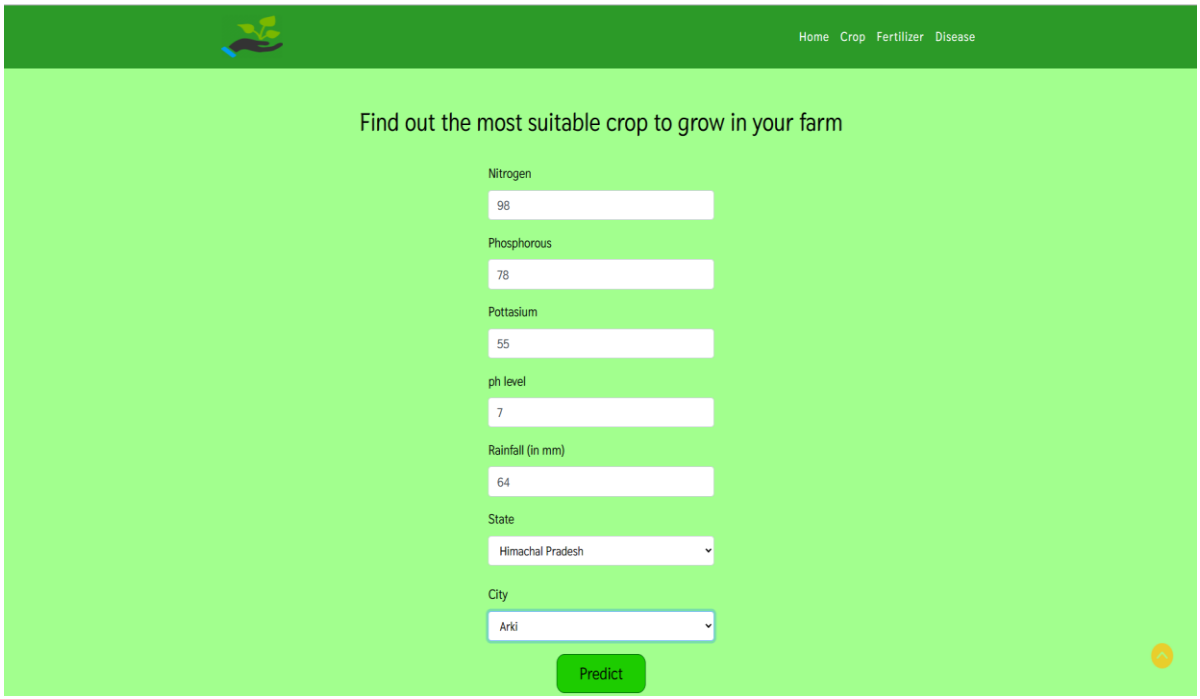
Data Preprocessing

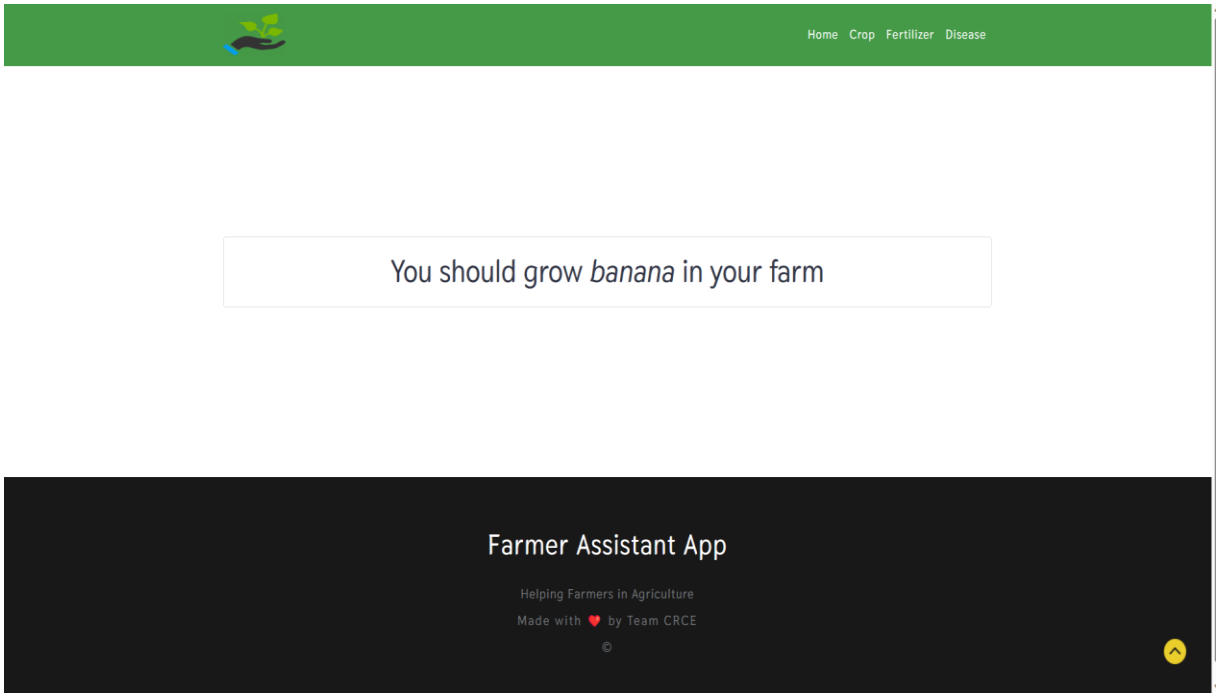
Preprocessing comes after data analysis and visualisation. Data cleaning and preparation for usage in machine learning algorithms are accomplished through the crucial step of data preprocessing. Preprocessing primarily focuses on addressing any missing values and outliers as well as inaccurate or outlier-containing data. There are two approaches to deal with missing data. The first approach is to just delete the entire row that contains the incorrect or missing information. This method is simple to apply, however it should only be applied to huge datasets. If there are a lot of missing values, using this strategy to tiny datasets may result in an excessive reduction in dataset size. The accuracy of the outcome may be seriously impacted by this. We won't be employing this strategy because our dataset is not that large. Since the values in the dataset we utilised were in string format, we had to transform and encode them into integer values before feeding them to the neural network. We first transformed the data into pandas categorical data, made codes for the various states and crops, attached these, and built distinct datasets.

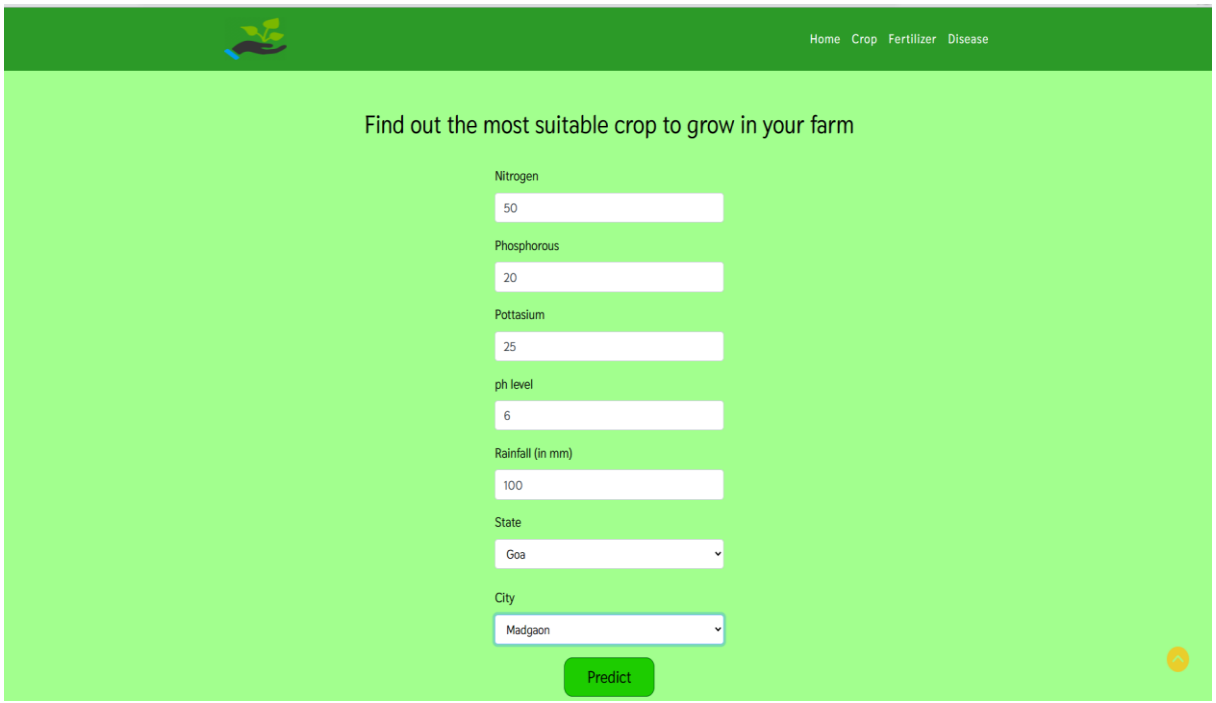
Chapter 6

Result

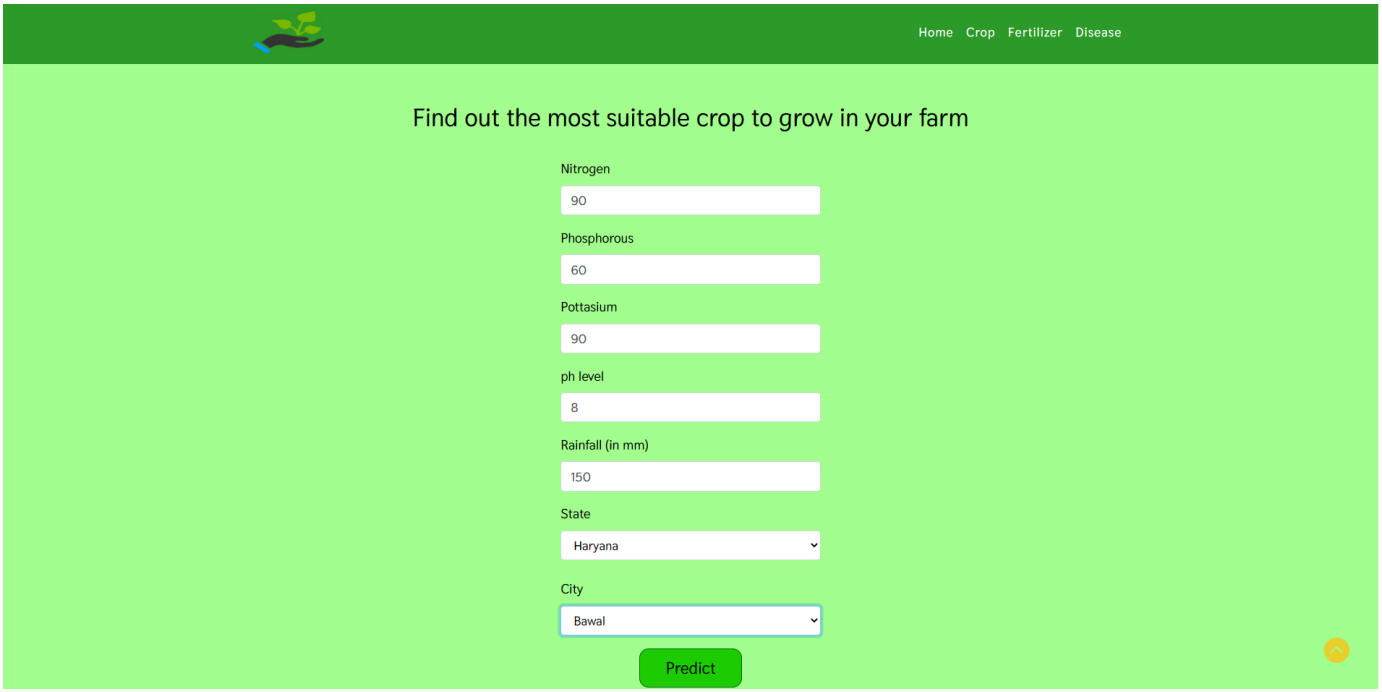
Crop Recommendation (Input and Output)

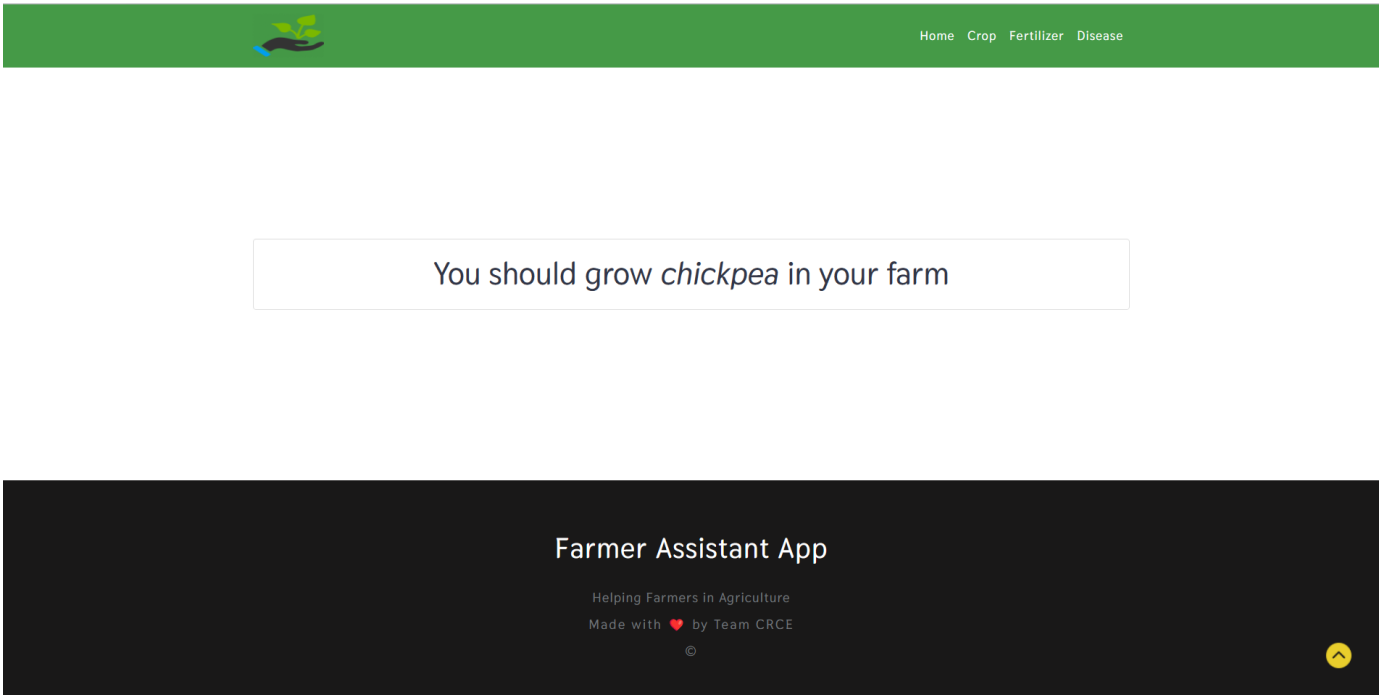



Home Crop Fertilizer Disease
You should grow *banana* in your farm
Farmer Assistant App
Helping Farmers in Agriculture
Made with by Team CRCE
©

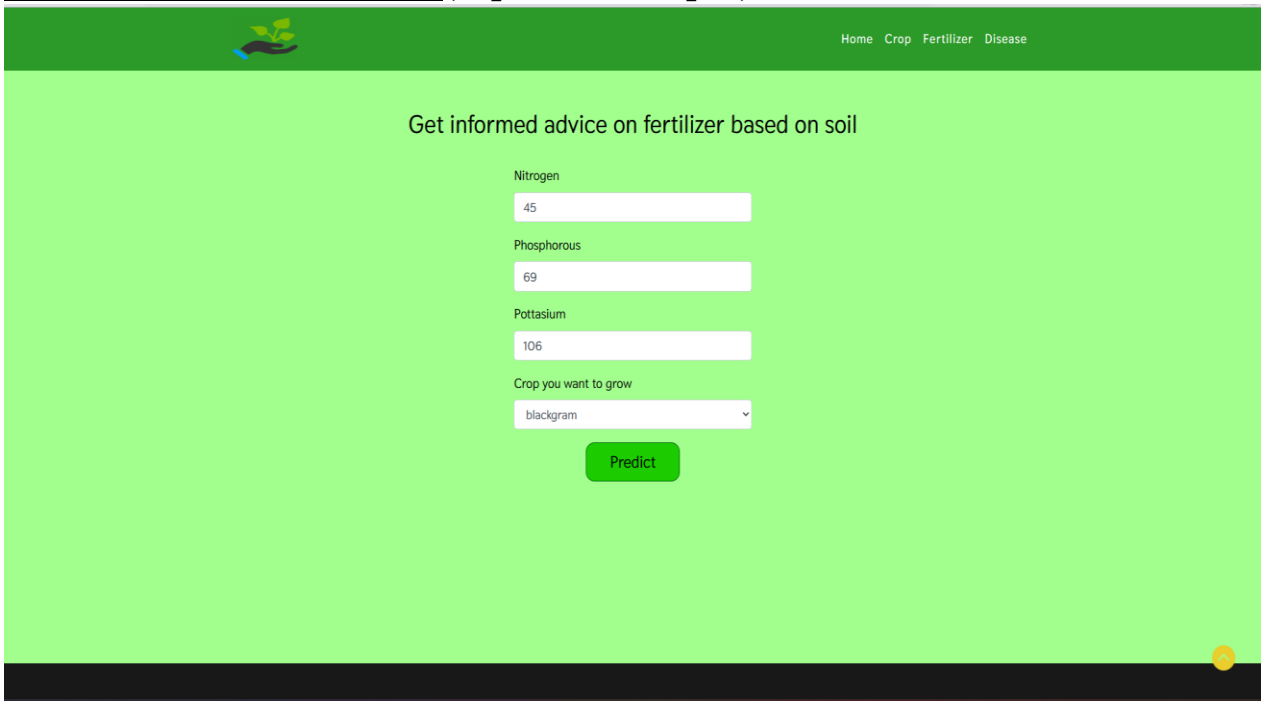

Home Crop Fertilizer Disease
Find out the most suitable crop to grow in your farm
Nitrogen
50
Phosphorous
20
Pottasium
25
ph level
6
Rainfall (in mm)
100
State
Goa
City
Madgaon
Predict

You should grow *orange* in your farm




Home Crop Fertilizer Disease
You should grow *chickpea* in your farm
Farmer Assistant App
Helping Farmers in Agriculture
Made with ❤ by Team CRCE
©

Fertilizer Recommendation(Input and Output)



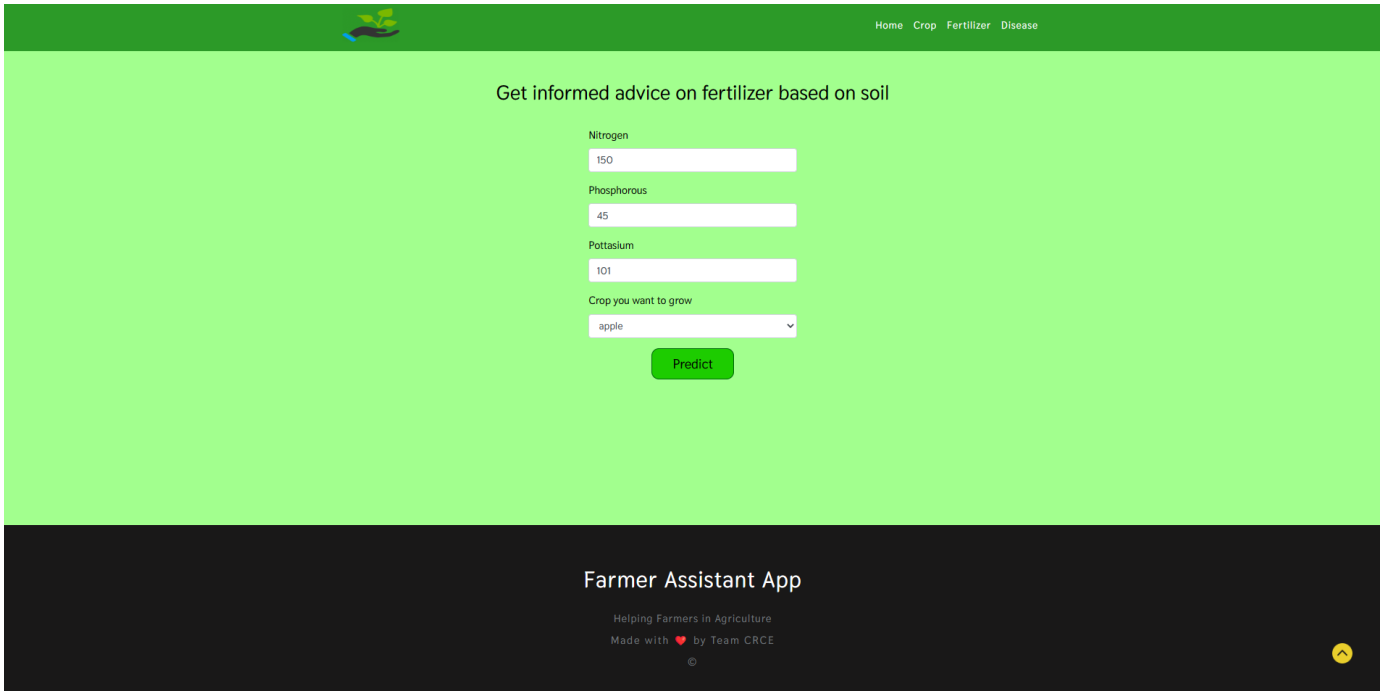
The K value of your soil is high.

Please consider the following suggestions:

1. *Loosen the soil* deeply with a shovel, and water thoroughly to dissolve water-soluble potassium. Allow the soil to fully dry, and repeat digging and watering the soil two or three more times.
2. *Sift through the soil*, and remove as many rocks as possible, using a soil sifter. Minerals occurring in rocks such as mica and feldspar slowly release potassium into the soil slowly through weathering.
3. Stop applying potassium-rich commercial fertilizer. Apply only commercial fertilizer that has a '0' in the final number field. Commercial fertilizers use a three number system for measuring levels of nitrogen, phosphorous and potassium. The last number stands for potassium. Another option is to stop using commercial fertilizers all together and to begin using only organic matter to enrich the soil.
4. Mix crushed eggshells, crushed seashells, wood ash or soft rock phosphate to the soil to add calcium. Mix in up to 10 percent of organic compost to help amend and balance the soil.
5. Use NPK fertilizers with low K levels and organic fertilizers since they have low NPK values.
6. Grow a cover crop of legumes that will fix nitrogen in the soil. This practice will meet the soil's needs for nitrogen without increasing phosphorus or potassium.

Fertilizers

If you want to lower the potassium level in your soil, you can use a fertilizer that is lower in potassium and higher in other essential nutrients. For example, a fertilizer with a ratio of 10-10-10 (10% nitrogen, 10% phosphorus, 10% potassium) would be a good choice. You can also use fertilizers specifically formulated for your plants, such as a fertilizer for vegetables, flowers, or lawns. It's also important to use fertilizer in the correct amounts, as over-fertilizing can be harmful to plants. Always follow the instructions on the fertilizer package, and be sure to water the soil after fertilizing to help the nutrients reach the roots of your plants.



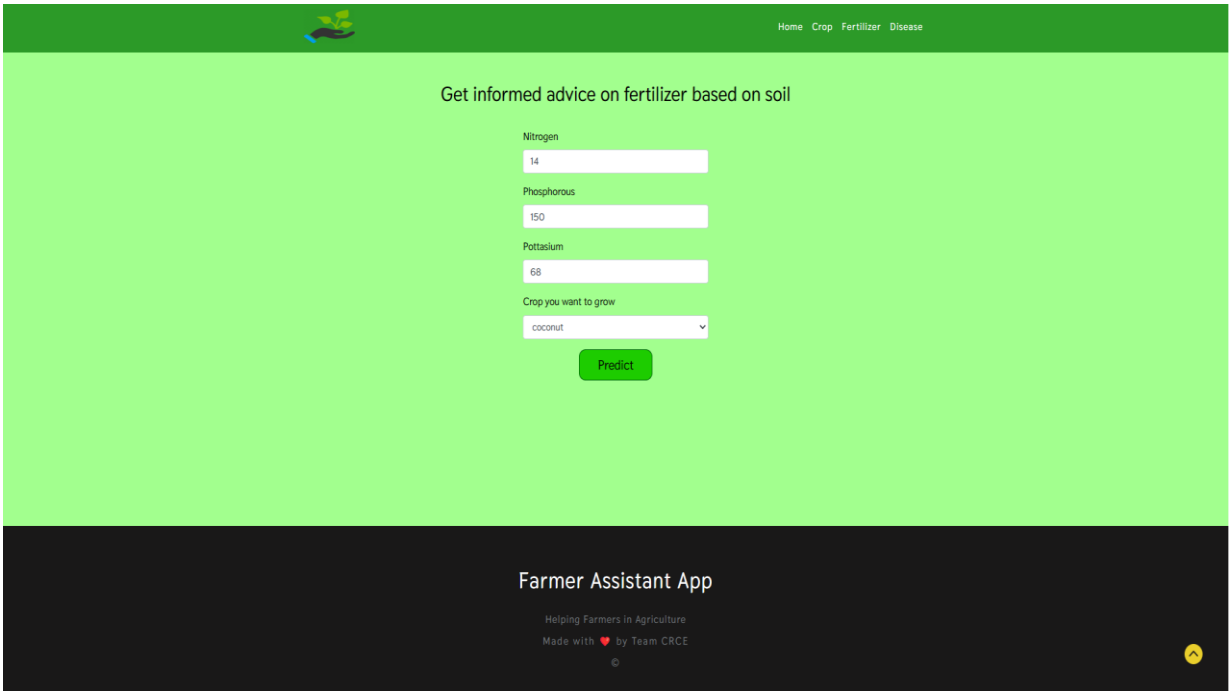
The N value of soil is high and might give rise to weeds.

Please consider the following suggestions:

1. *Manure* – adding manure is one of the simplest ways to amend your soil with nitrogen. Be careful as there are various types of manures with varying degrees of nitrogen.
2. *Coffee grinds* – use your morning addiction to feed your gardening habit! Coffee grinds are considered a green compost material which is rich in nitrogen. Once the grounds break down, your soil will be fed with delicious, delicious nitrogen. An added benefit to including coffee grounds to your soil is while it will compost, it will also help provide increased drainage to your soil. AZ
3. *Plant nitrogen fixing plants* – planting vegetables that are in Fabaceae family like peas, beans and soybeans have the ability to increase nitrogen in your soil
4. Plant 'green manure' crops like cabbage, corn and brocolli
5. *Use mulch (wet grass) while growing crops* - Mulch can also include sawdust and scrap soft woods

Fertilizers

If a soil test shows that the nitrogen levels are high, but the soil is deficient in other nutrients, a balanced fertilizer such as a 10-10-10 or a 14-14-14 may be used. These types of fertilizers provide a balance of nitrogen, phosphorus, and potassium, and can help to promote healthy plant growth. Additionally, you may consider fertilizers that are specific to your crop, for example Nitrogen-Phosphorus-Potassium (NPK) fertilizers for vegetables, fruits and grains. It is always good to consult with local extension services or agronomists for best practices.



The P value of your soil is high.

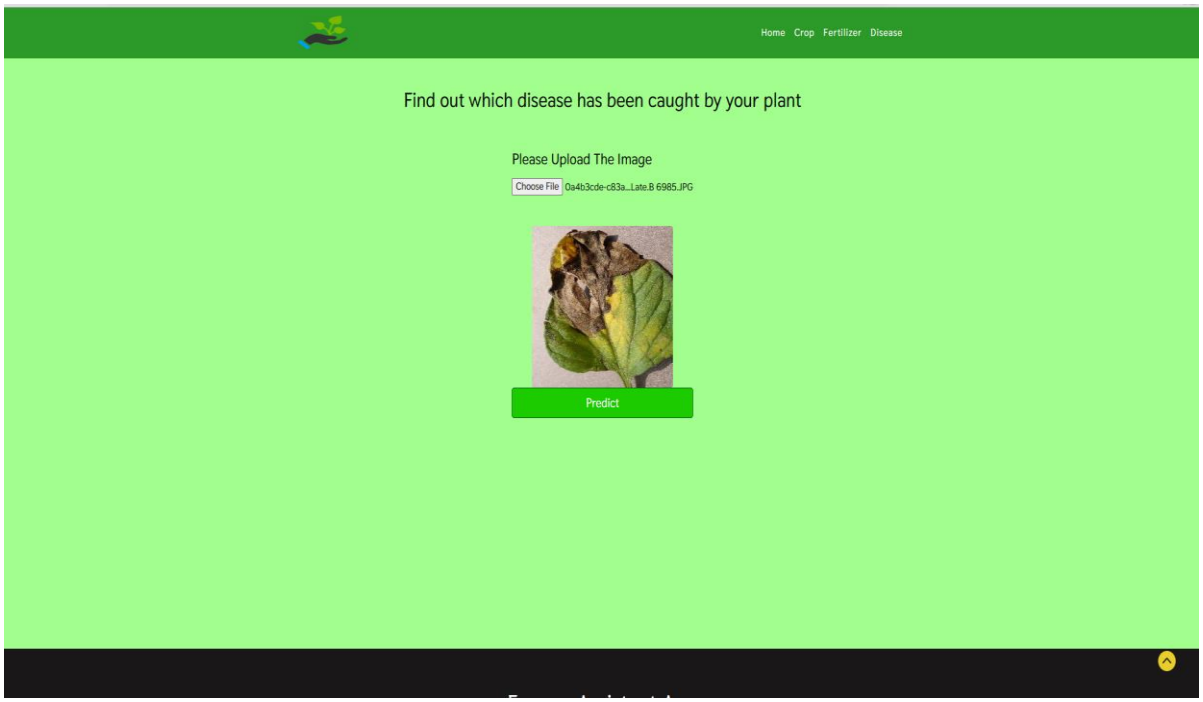
Please consider the following suggestions:

1. *Avoid adding manure* – manure contains many key nutrients for your soil but typically including high levels of phosphorous. Limiting the addition of manure will help reduce phosphorus being added.
2. *Use only phosphorus-free fertilizer* – if you can limit the amount of phosphorous added to your soil, you can let the plants use the existing phosphorus while still providing other key nutrients such as Nitrogen and Potassium. Find a fertilizer with numbers such as 10-0-10, where the zero represents no phosphorous.
3. *Water your soil* – soaking your soil liberally will aid in driving phosphorous out of the soil. This is recommended as a last ditch effort.
4. Plant nitrogen fixing vegetables to increase nitrogen without increasing phosphorous (like beans and peas).
5. Use crop rotations to decrease high phosphorous levels

Fertilizers

For high phosphorous levels in the soil, a fertilizer with a low P (phosphorus) and high N (nitrogen) and K (potassium) ratio, such as a 10-10-20 or 16-4-8 fertilizer, would be a good choice. It is important to note that overuse of fertilizers can lead to soil imbalances and excess nutrients in the water supply, so it's always best to follow the manufacturer's instructions and consult with a local expert or extension service for specific recommendations for your area.

Disease Prediction:



Crop: Tomato

Disease: Late Blight

Late blight is a potentially devastating disease of tomato, infecting leaves, stems and fruits of plants. The disease spreads quickly in fields and can result in total crop failure if untreated.

Cause of disease:

1. Late blight is caused by the oomycete Phytophthora infestans. Oomycetes are fungus-like organisms also called water molds, but they are not true fungi.
2. There are many different strains of P. infestans. These are called clonal lineages and designated by a number code (i.e. US-23). Many clonal lineages affect both tomato and potato, but some lineages are specific to one host or the other.
3. The host range is typically limited to potato and tomato, but hairy nightshade (Solanum physalifolium) is a closely related weed that can readily become infected and may contribute to disease spread. Under ideal conditions, such as a greenhouse, petunia also may become infected.

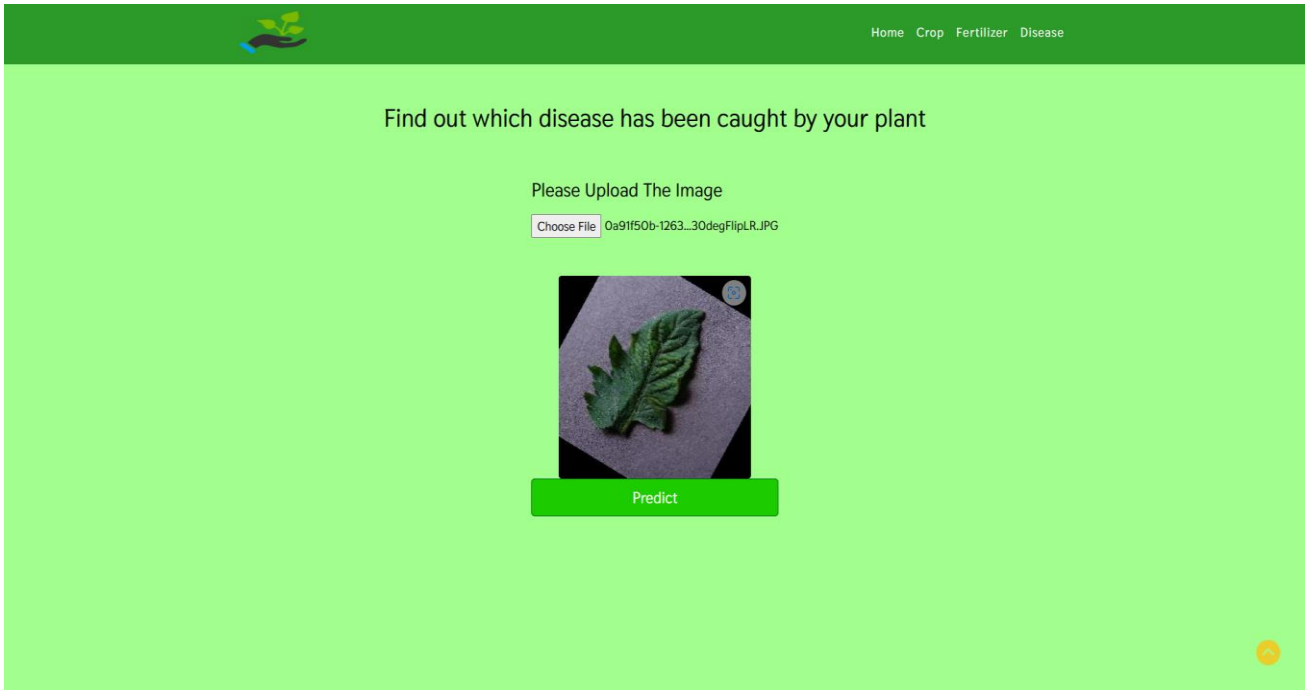
Pesticides

There are several fungicides that are effective against late blight, a fungal disease that affects tomatoes and other members of the Solanaceae family. Some examples include:

Chlorothalonil: This fungicide is a broad-spectrum protectant that is effective against a wide range of fungal diseases, including late blight. It is typically applied as a spray every 7-10 days during the growing season.

Mancozeb: This fungicide is a protectant and eradicant that is effective against late blight and other fungal diseases. It is typically applied as a spray every 7-10 days during the growing season.

Copper: This fungicide is a protectant that is effective against late blight and other fungal diseases. It is typically applied as a spray or a soil drench every 7-10 days during the growing season.


Home Crop Fertilizer Disease
Find out which disease has been caught by your plant
Please Upload The Image
Choose File
0a91f50b-1263...30degFlipLR.JPG
Predict

Crop: Tomato

Disease: Mosaic Virus

Cause of disease:

1. Tomato mosaic virus and tobacco mosaic virus can exist for two years in dry soil or leaf debris, but will only persist one month if soil is moist. The viruses can also survive in infected root debris in the soil for up to two years.
2. Seed can be infected and pass the virus to the plant but the disease is usually introduced and spread primarily through human activity. The virus can easily spread between plants on workers' hands, tools, and clothes with normal activities such as plant tying, removing of suckers, and harvest
3. The virus can even survive the tobacco curing process, and can spread from cigarettes and other tobacco products to plant material handled by workers after a cigarette

How to prevent/cure the disease

1. Purchase transplants only from reputable sources. Ask about the sanitation procedures they use to prevent disease.
2. Inspect transplants prior to purchase. Choose only transplants showing no clear symptoms.
3. Avoid planting in fields where tomato root debris is present, as the virus can survive long-term in roots.
4. Wash hands with soap and water before and during the handling of plants to reduce potential spread between plants.

Pesticides

1. Use of pesticides: There are no specific pesticides that target the ToMV virus. However, using pesticides to control the insect vectors that spread the virus, such as aphids and thrips, can help to reduce the spread of the disease.
2. Cultural control methods: Some cultural control methods that can help to prevent the spread of ToMV include crop rotation, removal of infected plant debris, and maintaining good hygiene in the greenhouse or field.
3. Resistant varieties: Planting tomato varieties that have been bred to be resistant to ToMV can also help to prevent the spread of the disease.

Accuracy

The four models we employed for our crop recommendation system, as shown in table no. 2.1, which have the highest accuracy.

Table No. 2.1 Crop Recommendation

Algorithm	**Accuracy (%)**
Logistic Regression	95
Decision Tree	90
Random Forest	99
SVM	97

In order to predict plant diseases, we now only used one model i.e the ResNet model which able to predict diseases with an accuracy of 98.03 percent.

Table no. 2.2 Disease Prediction

Algorithm	**Accuracy (%)**
ResNets	98.03

Chapter 7

Conclusion And Future Enhancements

7.1 Conclusion

The entire system aids in crop selection by offering information that most farmers would not be aware of, lowering the likelihood of crop failure and raising output. Additionally, it stops them from suffering losses. Millions of farmers around the nation may access the system, which can be expanded to the online. The crop recommendation system will be further developed to connect with a yield predictor, another subsystem that would also give the farmer an estimate of production if he plants the recommended crop. The outcomes are beneficial because they allow farmers to operate more effectively.

7.2 Future Enhancements

- We'll include new machine learning models.
- We will implement IOT-based tools to real time soil detection.

Appendix – 1

Optimizing Crop Production: An Agronomic Advisor Application Based on Soil Nutrients

Pratik Harde, Ibin Babu, Ananya Sharma, Prof. Prachi Patil,

Prof. Monali Shetty

Computer Engineering, Fr. CRCE, Mumbai, 400050, Maharashtra, India.

edu.pratik21@gmail.com, ibin.p12@gmail.com, ananya.sharma2401@gmail.com, prachi@fragnel.edu.in, shettymonalin@gmail.com

ABSTRACT

Agriculture sector works as the source of raw material for non-Agricultural sectors. As much as 60% of the land is used for farming in India. It feeds around total of 1.2 billion in population. The population has also been increasing day to day and also the agriculture is not able to meet the demanded requirements for the increasing population. Crop yields, meanwhile, have already begun to suffer as a result of climate change. Unnatural climate changes can have a negative impact on food production and forecasting, which in turn can affect farmers' economics by resulting in low yields. Droughts, floods, heat waves, storms, and various more extreme weather events can destroy crops and reduce harvests, and climate change can bring about these conditions. Crop output may also be impacted by variations in temperature and rainfall patterns, which may change when crops are planted and harvested. These changes can make it difficult for farmers to predict their yields accurately, leading to economic losses. It is possible to produce precise crop prediction results by using the right parameters, such as soil nutrients properties (Nitrogen, Phosphorous, and Potassium, nutrients concentration, soil type and pH value), rainfall patterns, temperature patterns, soil structures, and other factors, such as crop diseases. A crucial aspect of agriculture is determining the best crop to grow, and in recent years, machine learning algorithms have become increasingly important in this process. This unique research is use of ML algorithms to better precisely recommend the crops based on the location.

Supervised learning classification was used for the recomme ndations in this study. This study's main goal is to identify the most effective feature selection and classification techniques to predict the best harvest that will thrive in a particular environment, including temperature, rainfall, and geographic location in a given state, soil properties, including phosphorus (P), potassium (K), nitrogen (N), and pH value, as well as soil type. To provide recommendations for crops that are likely to thrive in a particular environment based on the available soil nutrients, an agronomic advisor application can be developed. The application can use a suitable classification algorithm to identify the most relevant features of the soil and environment. This can help farmers optimize their crop production and maximize their yields.

Keywords: Machine Learning, Agriculture, Soil nutrients, temperature patterns, Crop Recommendation, Random Forest, SVM, Decision Tree, Logistic Regression.

1 INTRODUCTION

Choosing the appropriate crop is a critical decision for farmers since it has a significant impact on the final yield and is influenced by factors such as the environment and soil type. Selecting the right crop for a particular farm is a challenging choice that affects the yield. Expert advice on crop selection or recommendations can be time-consuming and expensive, making it difficult for many farms to afford it. Traditional methods of crop selection, such as expert consultations or field trials, can be costly and require significant investments of time and resources. As a result, many farmers may not have access to expert advice, which can limit their ability to optimize their crop production and maximize their yields. The use of machine learning algorithms and agronomic advisor applications can provide a more cost-effective and efficient alternative to traditional methods of crop selection. By analyzing relevant data and identifying the most important features, these applications can provide farmers with recommendations for the crops that are most likely to thrive in their specific environment, based on the available soil nutrients and other relevant factors.

The management of system crops to maximize agricultural productivity is one of the key areas of precision agriculture.

Suggesting suitable crops based on data analysis can help increase crop production while minimizing resource usage, by identifying the most appropriate options from a dataset. These programs are crucial for decision-making because they assist users in maximizing gains or reducing risks.

It is vital to develop a system that might provide Indian farmers with predicted information so they could make informed crop decisions. In light of this, we propose a system, an intelligent system, that, before advising the user on the crop that would grow the best, would evaluate soil characteristics (N, P, K, soil type, pH value, and nutrients concentration), as well as environmental variables (rainfall, temperature, and geographic location in relation to state.

2 LITERATURE SURVEY

In paper [1], Professor Rakesh Shirsath and a number of coauthors suggested a system that helps users select the crop to be planted. The method is an online system that any registered farmer can access through a subscription and receive personalized information. The system has a module that compiles data from many sources on crops that have already been cultivated and recommends a crop that would be a match for planting. To make the procedure easier overall, artificial neural networks are used. In case the farmer has any issues while using the system, a feedback method is offered at the end to allow the developer to make any necessary improvements.

In their research [2], Ji-chun Zhao and Jianxin Guo consider knowledge databases to be big data and make conclusions from the data. The various modules considered include users, knowledge engineers, domain experts, humanmachine interfaces, inference engines, and knowledge banks. The knowledge acquisition system collects data for the decision system and creates a usable knowledge base to address the problem. The essay makes use of several Hadoop modules to extract features. It utilizes unstructured data, processes it with Hive, Mahout and NoSQL then stores the outcomes in HDFS. Only the data for the wheat crop was reported; other crops were not taken into account.

As mentioned in the paper [3], the RSF is a farming recommendation system that takes into account a data analysis, location detection module and crop growth database, storage module, and physiographic database. The related location discovery module identifies areas nearby the user's current location and researches the crops that are grown there. As a result, recommendations are offered to the user using a similarity matrix. The location detection module uses the Google API services to determine the user's current location and identify similar sites that are close by. Nevertheless, the system does not get user feedback to improve the process.

The system suggested in paper [4] by authors S. Pudumalar and associated co-authors uses an ensemble technique known as Majority Voting Technique, which taps the power of several models to improve prediction accuracy. The final forecast is accurate when the majority voting mechanism is used, even if one of the methods predicts incorrectly. KNN, Random Trees, CHAID, and Nave Bayes for ensemble are the techniques used. The key elements used in the prediction process are if-then rules. The ensemble model provided 88% accuracy.

The research by Yogesh Gandge and Sandhya [5] is a review paper that looks at several algorithms and how effective they are for use in agriculture. It was discovered that multiple linear regression offered a rice yield accuracy of 90–95%. The ID3 algorithm was used to study the decision tree and generate suggestions for the soybean crop. The third method, SVM, was applied to all the crops and had good accuracy while utilizing little computer power. A neural network was used to corn-related data to obtain 95% accuracy. Also used were LAD Tree, K-means, KNN, C4.5, J48 and Naive Bayes. The investigation came to the conclusion that the algorithms still needed to be enhanced for greater accuracy. A dataset from Kaggle.com was analyzed for a study titled Agricultural Yield Prediction using Data Mining [6]. The LAD Tree, J48, LWL, and IBK algorithms were utilized by the author to analyses the data using the WEKA tool. The accuracy was evaluated using specificity, accuracy, RMSE, mean absolute error and sensitivity. Confusion matrices were used to find the situations that each classifier correctly identified. The results suggested that pruning could lead to improved accuracy.

In their study [7] recommended employing ANN, KNN, SVM, GBDT, Random Forest, Decision Tree and Regularized Gradient Forest as seven machine-learning approaches for crop selection. The system is designed to recover every crop that was sown as well as the timing of its growth at a particular time of the year. The crops providing the best yields are selected once the yield rate of each crop has been determined. In order to have the best yields, the approach also recommends which crops should be planted in what order.

3 METHODOLOGY

The most appropriate crop will be suggested using the proposed approach, which will make use of many soilrelated characteristics. The suggested system's technique consists of multiple steps, as indicated in Fig.1.

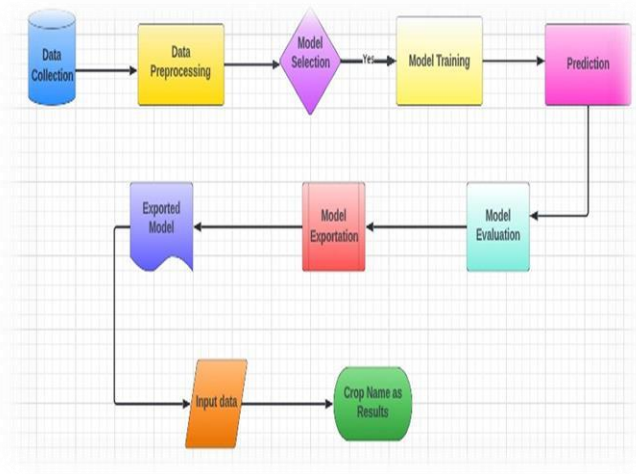


Fig.1 Proposed Architecture Diagram The process includes:

Data collection: This involves gathering data on soil-related parameters, such as soil pH, nutrient content, and other relevant information.

Data preprocessing: The collected data is then processed and cleaned to ensure that it is accurate and suitable for analysis.

Model selection: Based on the characteristics of the problem and the data at hand, the best machine learning model is chosen.

Model training: The selected model is then trained using the preprocessed data to create a predictive model.

Prediction: After being trained, the model might be used to forecast the optimal crop based on the input data.

Model evaluation: The performance of the model is then evaluated to make sure that it is accurate and reliable.

Input data: The input data includes soil-related parameters and other relevant information that is used to generate recommendations.

Crop name as an outcome: Based on the input data, the system's final output is the suggested crop name.

Overall, this process utilizes machine learning algorithms to analyze soil-related parameters and generate personalized recommendations for suitable crops. Farmers can use the agronomic advisor application to determine what crops to grow in a particular environment.

3.1 Dataset Collection

The process of building a machine learning model begins with collecting data. It is important to collect data before constructing a machine learning model. It is crucial to obtain a sizable amount of reliable information that is pertinent to the issue at hand. Data collection allows us to keep track of past events, which allows us to use data analysis to find repeated patterns. This dataset covers characteristics specific to the soil. This primarily consists of the soil and agricultural dataset for India over the preceding 20 years. This information collection consists of over 2000 observations from over 20 distinct crops. In essence, crops rest on the ground. The soil's fertility is influenced by its texture and the presence of nutrients like nitrogen, phosphorus, and potassium. This information also includes historical weather patterns for India.

	N	P	K	temperatu	humidity	ph	rainfall	label
2	90	42	43	20.87974	82.00274	6.502985	202.9355	rice
3	85	58	41	21.77046	80.31964	7.038096	226.6555	rice
4	60	55	44	23.00446	82.32076	7.840207	263.9642	rice
5	74	35	40	26.4911	80.15836	6.980401	242.864	rice
6	78	42	42	20.13017	81.60487	7.628473	262.7173	rice
7	69	37	42	23.05805	83.37012	7.073454	251.055	rice
8	69	55	38	22.70884	82.63941	5.700806	271.3249	rice
9	94	53	40	20.27774	82.89409	5.718627	241.9742	rice
10	89	54	38	24.51588	83.53522	6.685346	230.4462	rice
11	68	58	38	23.22397	83.03323	6.336254	221.2092	rice
12	91	53	40	26.52724	81.41754	5.386168	264.6149	rice
13	90	46	42	23.97898	81.45062	7.502834	250.0832	rice
14	78	58	44	26.8008	80.88685	5.108682	284.4365	rice
15	93	56	36	24.01498	82.05687	6.984354	185.2773	rice
16	94	50	37	25.66585	80.66385	6.94802	209.587	rice
17	60	48	39	24.28209	80.30026	7.042299	231.0863	rice
18	85	38	41	21.58712	82.78837	6.249051	276.6552	rice
19	91	35	39	23.79392	80.41818	6.97086	206.2612	rice
20	77	38	36	21.86525	80.1923	5.953933	224.555	rice

3.2 Pre-processing

Pre-processing entails taking away anomalies and disturbances from the CSV dataset. Data loss frequently occurs and has a direct impact on the final machine learning model's effectiveness and accuracy. This needs to be addressed using a variety of strategies, including median and overall column mean. We can quickly clean the dataset using skLearn. It offers an imputer class that addresses and replaces values when they are absent. The imputer class accepts arguments such as missing values and techniques the imputer employs internally. Using the mean on-axis, missing data are renewed using this method.

3.3 Feature Selection

To provide recommendations for suitable crops based on soil related parameters, the following features can be considered:

Nitrogen (N): Nitrogen is a essential nutrient that is vital for plant increase, and it performs a critical role in the improvement of plant life. The availability of nitrogen in the soil can influence the growth and yield of crops. Therefore, the nitrogen level in the soil can be a crucial feature for crop recommendation.

Phosphorus (P): Phosphorus is another crucial nutrient for plant growth and it is involved in various plant processes, such as photosynthesis, energy transfer, and root development. The availability of phosphorus in the soil can significantly impact crop growth and yield.

Potassium (K): Potassium is a vital nutrient that is involved in several plant processes, including water regulation, photosynthesis, and disease resistance. Therefore, the availability of potassium in the soil can be an important feature for crop recommendation.

Temperature: Temperature is critical environmental aspect which could impact crop growth and improvement. Different crops have different temperature requirements for optimal growth and yield. Therefore, temperature may be a critical feature in figuring out the most appropriate crop for a particular region.

Rainfall: Rainfall is some other critical environmental element which could impact crop growth and improvement. Different crops have different water requirements, and the availability of rainfall can significantly influence the growth and yield of crops. Therefore, rainfall can be an essential feature for crop recommendation.

Soil pH: Soil pH can have a significant impact on plant growth and development. Different crops have different pH requirements, and the availability of soil nutrients can be influenced by the soil pH. Therefore, soil pH can be a crucial feature for crop recommendation.

Humidity: Humidity is every other crucial environmental aspect which could affect crop growth and development. . Different crops have different humidity requirements for optimal growth and yield. Therefore, humidity can be a critical feature in determining the most suitable crop for a specific region.

To provide farmers with informed recommendations about what crops to grow based on their region's soil type and environmental conditions, the crop recommendation system takes into account various soil-related parameters as features. This allows farmers to make informed selections about the quality crops to develop.

3.4 Choosing Machine Learning Model

When choosing a machine learning algorithm, Random Forest is one of the most liked and widely accepted supervised learning techniques. It constitutes a number of decision trees for differential subsets of data, other than using the whole data as a single unit. This helps to enhance the accuracy of the prediction of each variable. It makes sense to use Random Forest on a large dataset as it can provide results with high accuracy in a minimum span of time. Random Forest, Decision Tree, Logistic Regression and SVM are used in this model.

3.4.1 Decision Tree

Decision trees are a type of tool that predicts outcomes based on characteristics. They use simple rules to make predictions Decision trees can handle both numerical and nominal data.

The final prediction is made by gathering and combining all the results.

The Random Tree is a special type of decision tree that selects a random subset of characteristics to make predictions. The Random Tree builds multiple decision trees and combines their results for the best prediction

3.4.2 Random Forest

Random Forest helps to minimise overfitting in decision trees by introducing randomness in the tree construction process. The technique can handle missing data and still produce accurate results. The computational complexity of Random Forest is relatively low compared to other ensemble learning methods. Random Forest can be used for feature selection to identify the most important variables in a dataset. The technique is insensitive to outliers and noisy data. Random Forest can be easily parallelized, Given that it is capable of handling datasets having continuous as well as categorical variables, the Random Forest is a flexible approach that may be utilized for a variety of applications. Random Forest often outperforms other machine learning algorithms when dealing with classification challenges.

The steps followed by the Random Forest algorithm are:

1. The original dataset is divided into n smaller bagged samples of size n.
2. A decision tree is constructed using input from all N bagged datasets. To calculate the ideal split, impurity measures such as Gini Impurity or Entropy are used, and M features are selected at random from the overall number of features in the training set to avoid looking at every feature in the dataset during a node split.
3. The unique outputs of each decision tree are combined into a single result.
4. For each observation, the outcomes produced by each tree are calculated, if working on a regression problem.
5. The majority vote is used to make a decision depending on the votes of the majority, if working on a classification problem.

3.4.3 Support Vector Machine (SVM)

SVM purpose is to identify a hyperplane in which the data points can be effectively separated (where N is the number of characteristics). In situations where there are two classes of data points, several hyperplanes can be utilized for classification. Data points are grouped into classes, and a hyperplane with the highest margin of difference is selected. By maximizing the margin distance, the accuracy of future data point classifications is enhanced.

3.4.4 Logistic regression

Observations are categorised into distinct classes using a method known as logistic regression. Examples of categorization problems include whether something is spam or not, whether an online transaction is fraudulent or not, and if a tumour is malignant or benign. The output of logistic

regression is converted into a probability value via the logistic sigmoid function.

Which logistic regression models are there?

1. Binary
2. Failure of the class of multilinear functions, such as sheep, dogs, or cats.

An approach for predictive analysis is logistic regression, a machine learning technique that is used for classification issues and is based on the probability notion.

3.5 Crop Prediction

To determine the best crop variety for a particular region, the crop recommendation system employs a machine learning algorithm that takes into account the unique environmental factors of the area. The system utilizes user input data to train the model and identify the crop with the highest probability of success. To determine the ideal crop type, machine learning techniques including SVM, RF, logistic regression, and decision trees are used. The technology evaluates variables like humidity, soil moisture, temperature, and pH levels to suggest which crops farmers should produce.

4 RESULT ANALYSIS

The proposed crop recommendation model relies on a crop database and soil factors to suggest the best crop for a given soil type. The best crop variety is identified using machine learning algorithms, and the system found that the Random Forest technique generated the most accurate results. Table 1 shows the accuracy rates for each algorithm evaluated by the system.

Table 1 Proposed methodology performance analysis

Algorithm	Accuracy
Logistic Regression	95%
Decision Tree	90%
Random Forest	99%
SVM	97%

In this field, a number of prototypes have been put out that are helping to solve agricultural problems. Indian agriculture has enormous untapped potential. The technology that will help farmers by giving them the necessary advice on crops, their growth, and other fundamental information still has to be improved to be more compact, accurate, and affordable. The majority of the approaches now in use involve manually determining the soil type. The approach had a number of drawbacks. The system might not offer the necessary support in other circumstances. Therefore, this work suggested a novel strategy that is based on the location, to recommend crops and other strategy to address the drawbacks of existing papers.

Table 2 Existing research papers performance analysis

Algorithm Used	**Accuracy**	**Drawback**	**Citation**
Subscription based system, ANN	94.5%	Earlier-planted Crops unknown to the system.	[1]
Inference engine ANN	90%	Can Hadoop and Artificial Neural Networks work together to produce better results.	[2]
Recommendation generation module.	92.4%	Crop growing season, crop output rate, and physiographic factors Database of seasonal crops.	[3]
Tools like CHAID, random tree, naive bayes KNN and WEKA	88%	Missing and out of range value.	[4]
Multiple Linear Regression, SVM, Decision tree utilising ID3, K-means, C4.5, Neural Networks and KNN	MLR= 90% NN=95%	For the algorithms to become more accurate, improvement is required.	[5]
J48, LAD tree, LWL, IBK algorithm	IBK gives the highest accuracy	The LAD tree displayed the least accuracy. The tree can be pruned to reduce errors.	[6]

5. CONCLUSION

The crop recommendation system has been effective in creating a model that can foretell which crops would do best given various environmental characteristics including topography, soil type, and climate. The system analyzes vast amounts of data to identify optimal crop varieties for different regions and soil types, considering factors such as temperature, rainfall, and soil fertility. After extensive research and analysis, the system has found that specific crops are better suited for particular regions and soil types. For example, crops that require high levels of rainfall are suitable for areas with high precipitation, while crops that can tolerate drought conditions are better suited for regions with low rainfall.

The predictive model developed by the system uses machine learning algorithms to analyze the data and

provide customized recommendations to farmers. To recommend the best crops for a given region, the model consider a number of variables, including soil pH, nutrient content, and other environmental circumstances. By providing personalized recommendations, the system can help to increase crop yields, reduce costs, and improve efficiency in the agricultural sector. This, in turn, can contribute to sustainable agricultural practices and food security.

6. REFERENCES

1. Suruliandi, A., Mariammal, G., & Raja, S. P. (2021). Crop prediction based on soil and environmental characteristics using feature selection techniques. Mathematical and Computer Modelling of Dynamical Systems, 27(1), 117-140.

2. Doshi, Z., Nadkarni, S., Agrawal, R., & Shah, N. (2018, August). AgroConsultant: intelligent crop recommendation system using machine learning algorithms. In 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA) (pp. 1-6). IEEE.

3. S. P. Raja, B. Sawicka, Z. Stamenkovic and G. Mariammal, "Crop Prediction Based on Characteristics of the Agricultural Environment Using Various Feature Selection Techniques and Classifiers," in IEEE Access, vol. 10, pp. 23625-23641,
2022

4. Kulkarni, N. H., Srinivasan, G. N., Sagar, B. M., & Cauvery, N. K. (2018, December). Improving crop productivity through a crop recommendation system using ensembling technique. In 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS) (pp. 114-119). IEEE.

5. Pudumalar, S., Ramanujam, E., Rajashree, R. H., Kavya, C., Kiruthika, T., & Nisha, J. (2017, January). Crop recommendation system for precision agriculture. In 2016 Eighth International Conference on Advanced Computing
(ICoAC) (pp. 32-36). IEEE.

6. Pande, S. M., Ramesh, P. K., Anmol, A., Aishwarya, B. R.,
Rohilla, K., & Shaurya, K. (2021, April). Crop Recommender System Using Machine Learning Approach. (Pp. 10661071),IEEE Xplore.

7. Liying Yang (2011), 'Classifiers selection for ensemble learning based on accuracy and diversity' Published by Elsevier Ltd. Selection and/or peer-review under responsibility of [CEIS].

8. Eswari, K. E., & Vinitha, L. (2018). Crop yield prediction in Tamil Nadu using Baysian network. International Journal of Intellectual Advancements and Research in Engineering Computations, 6(2), 1571-1576.

Appendix – 2 (Plagiarism Report)

turnitin

Similarity Report ID: oid:12001:34386799

PAPER NAME
Plagarism check .pdf

AUTHOR
ibin report

WORD COUNT
2117 Words

CHARACTER COUNT
14370 Characters

PAGE COUNT
28 Pages

FILE SIZE
3.1MB

SUBMISSION DATE
Apr 23, 2023 7:23 PM GMT+5:30

REPORT DATE
Apr 23, 2023 7:23 PM GMT+5:30

10% Overall Similarity
The combined total of all matches, including overlapping sources, for each database.

- 2% Internet database
- Crossref database
- 9% Submitted Works database
- 1% Publications database
- Crossref Posted Content database

Excluded from Similarity Report

- Cited material