**Farmer Assistant Application**

A project report submitted in fulfillment of the

requirements for the degree of

**Bachelor of Computer Engineering (Sem -VlII)**

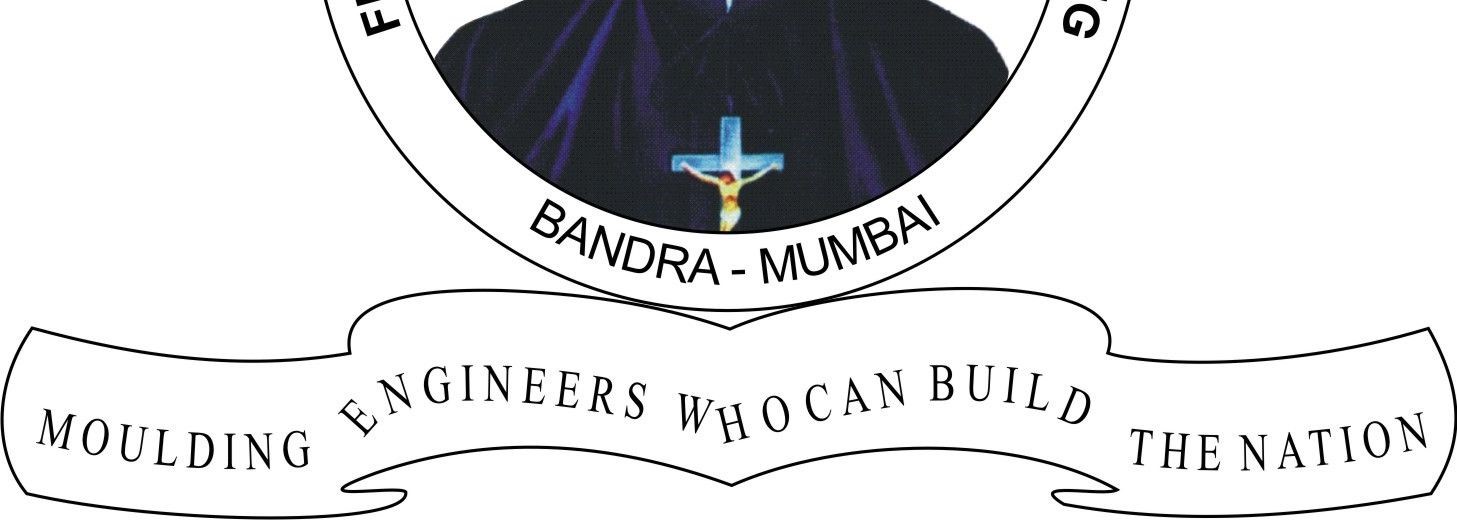
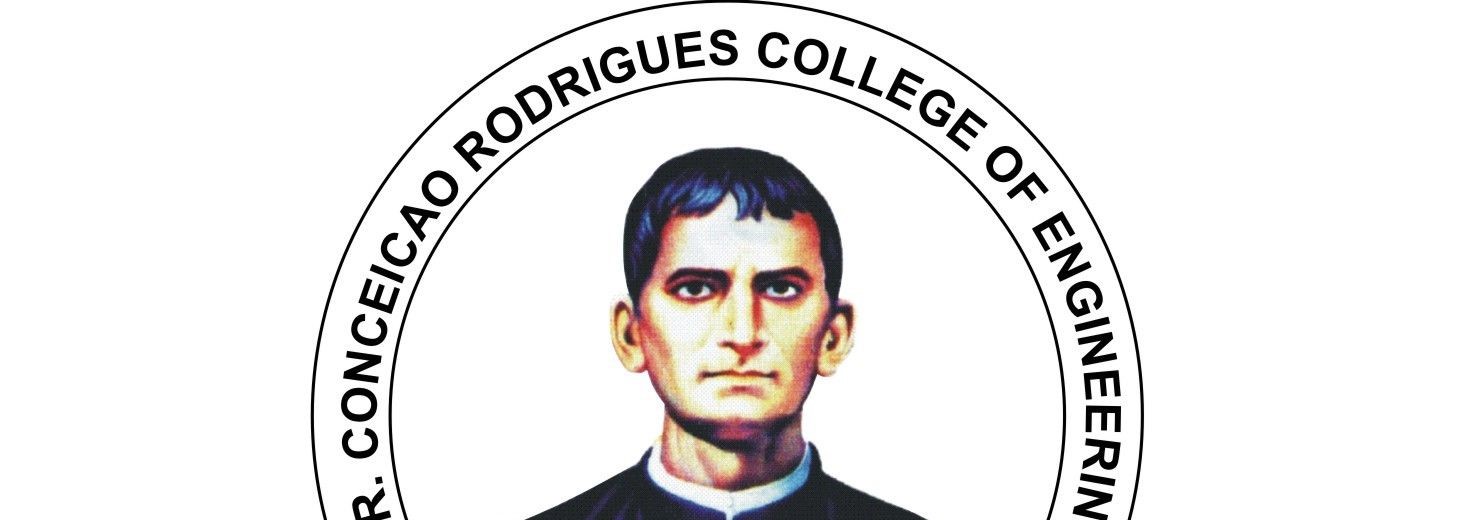
by

**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 Dr. Surendra Rathod**

Head of Department 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. Thekmost frequent issue that Indiankfarmers encounter is that they choosekthe incorrect crop forktheir soil, use the incorrect fertiliser, and are unable to identify plant diseases that are caused by their plants.As a result,kthey 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 agriculturekuses 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**

**Abstract**k **iv**

**List of Figures**k **ix**

**List of Tables**k **x**

1. **Introduction . . 1** 
   1. Motivation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 2
   2. Objectivesk. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 2
2. **Literature Review** k **3** 
   1. Smart Farming ……………………………………………………………………………………...3
   2. High ResolutionkMapping of Soil Properties Using Remote Sensing Variables in South-Western Burkina Faso……………………………………………. .. 3
   3. Internetkof things (IoT) applications to fight against COVID-19 pandemic . 3
   4. Classification of Soil and Crop Suggestion…………………………………………… ... 4
   5. Smart Agriculture Using WSN………………………………………………………………...4
   6. Intelligent Agriculture System To Assist Farmers In Smart Decision………..4
   7. Literature Review table…………………………………………………………………………5
3. **Problem Statement 7** 
   1. Drawbacks …………………………………………………………………………………………...7
   2. Solution To Above Problem…………………………………………………………………..7
4. **Project Designs 8** 
   1. Block Diagram …………………………………………………………………………………… 8
   2. Algorithms…………………………………………………………………………………………..9
   3. UML Diagram …………………………………………………………………………………13

4.3.1 Architecture ……………………………………………………………………………...14

* 1. Dataset………………………………………………………………………………………………15
  2. Hardware and Software requirements……………………………………………… ..17

1. **Implementation Details 18** 
   1. Methodology……………………………………………………………………………………….18
2. **Results**……………………………………………………………………………………………………..20
3. **Conclusion and Future Enhancments**……………………………………………………..28

**References** k **. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .**

**Appendix -1 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .**

**Appendix -2 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .**

vii

**List of Figures**

4.1. Block Diagramk

4.2. Disease Prediction Diagram

4.3. Crop Recommendation Diagram

4.4. DFD Level 0 and DFD Level 1k

4.5. UML Diagramk

4.6. Sequence Diagram

4.7 Architecture

**List of Tables**

6.1 Algorithm and Accuracy

viii

**Glossary**

SVM Support Vector Machines

RF Random Forest

DFD Data Flow Diagrams

ML Machine Learning

KNN K-Nearest Neighbor

**Chapter 1**

**Introduction**

A farmer'sddecision regarding which crop toacultivate is typically influencedaby his knowledge as well as unimportant variables like the desire toamake quickamoney, ignoranceiof market demand,iexaggeration of a soil'siability to support aiparticular crop, and soiforth. The farmer's family'safinancial situation could beaseverely strained if he makes the wrong choice.iMaybe this is one of the numerousafactors contributing to the innumerable farmerasuicide cases thatithe media reports on every day. Suchia wrong judgement wouldahave detrimentalieffects on not only the farmer's family butathe entire economyaof an area in aicountry like India,iwhere agriculture and associated sectors contributeato abouti20.4%aof its Gross ValueiAdded (GVA). Becauseiof this, we have determined that a farmer's decision on which crop to cultivate during a specific season is a very serious one. The urgentirequirement is to create aasystem that mightioffer Indian farmersipredictive information soatheyicould choose which crop to produceiwith knowledge.i a

In light of this, weisuggest a system, an intelligent system, thatiwould evaluateasoil characteristics (pH value, N, P, K, soil type, and nutrientsaconcentration) asiwell as environmental factors (rainfall, temperature, geographic location in termsaof state), before advising the useraon the crop that would grow best. Additionally, a fertiliser recommendationabased on the ideal nutrients of the produced crops is also made.

**1.1 Motivation**

Asiabouti70%iofitheipopulationidependsioniagriculture,iitiservesiasithe economicifoundationiforidevelopinginationsilikeiIndia.pIndia'sieconomy andiemploymentiareiheavilyireliantioniagriculture.

Indianifarmersifrequentlyistruggleiwithichoosingitheirighticropitoisupport theirisoil'sineedsiandiwhichifertiliseritoiutiliseiforitheiriparticularicrop. Thisiproblemiofitheifarmersihasibeeniaddressedithroughiprecision agriculture.

**1.2 Objectives**

* Toadevelop aareliableamodel thatacan accurately estimate cropasustainabilityain a givenastate under specific climatic andasoil 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 + MLadesign for smartafarmingawith severalaapplications, a system for water management systems, and an improvement to the current irrigation techniques are all included. An IoT and ML-based farming systemaconstantly keeps farmers informedaof theapotential weather patterns and offers the bestarecommendations for crops and airrigation techniques, aresulting inaincreasedaproduction.

**Smart**a**Management of**a**Crop Cultivation**a**using IoT and Machine**

**Learning [2]**

The most suited crop to grow in that climate is predicted by a smart system that takes into aaccount measuredaparameters (temperature, humidity), asawell as otheraparameters (soila type,afarm location, arainfall), into account.

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

The four modules iMODIFIEDiSUPPORTiVECTORiREGRESSION,iaiwell-known machineilearning technique, are used to determine real-time sampling of soil parameters. The modules consist of a sensor connected to an IoT device, aniagriculturalicloud,ian analysisiofireal-timeisensoridata,iandianiagriculturaliuseriinterfacei(AUI).Theifirst module consists of a NodeMCU portable IoT device with environmental sensors, including pH and soilimoistureisensors. Storage is part ofitheiagriicloudimodule. Usingia modifiedisupportivectorimachineimethod, the real-time data module is analysed to process the various types of crops and small plants. It isiaisimple online interface called Agri-user interface. Thus, using a modified support vector machine algorithm, a farmer can determine the types of cropsiandismalliplants that can be cultivated on their land using theihelpiofithe soil'siqualities.

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

InternetiofiThingsi(IOT) and machine learning are two of the proposed new technologies. Using an IOT system, real-timeidata fromitheifield may be gathered. Thei trainedimodelireceives inputifrom the field area's collectedidata. 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 theikind ofisoil and suggests a suitablei crop that can be cultivated in that soil. Several machine learningitechniques,iincluding KNN,iSVM, and logistic regression, haveibeen used to test theimodel. Compared to other models, the accuracy of the current model is highest.

**Smart Agriculture Using WSN [6]**

Proposed a smartiagricultureisystem based oniIOT and deepilearning.iWithithe aid of a wireless sensor network, thisisystemikeeps an eyeion and gathers dataionitheisoil characteristicsifromitheifield. After then, the dataiis uploadedito theicloud.iFinally,iby anticipatingitheicrop that willibe grown for theifollowing crop cycle,ithe systemsiadvise farmersion theiappropriate irrigation strategies.iThe farmers will receive this information via SMS. Temperatures of the soil, the atmosphere, and the humidity are among the characteristics.iThisitechniqueiadvisesienhancingitheiefficiencyievenifurtheriibyiii forecastingiwheniit will be appropriate to applyipesticides, fertiliser,iand 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 (weatherior soil) foripredictingithe appropriateness oficropidevelopmentiwasione 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 soilitype may be ideal for supporting a certain crop variety, the yield will suffer if the local climate doesn'tiappearitoibe favourable forithaticropikind.

**3.2 Solution to the Problem**

By proposing anieffectiveicropirecommendationisystemithat accounts for all relevant factors, such asitemperature,irainfall,ilocation,iandisoiliquality, 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 beiusediforicropsigrowniinivariousistates.

**Chapter 4**

**Project Design**

**4.1 Block Diagram**

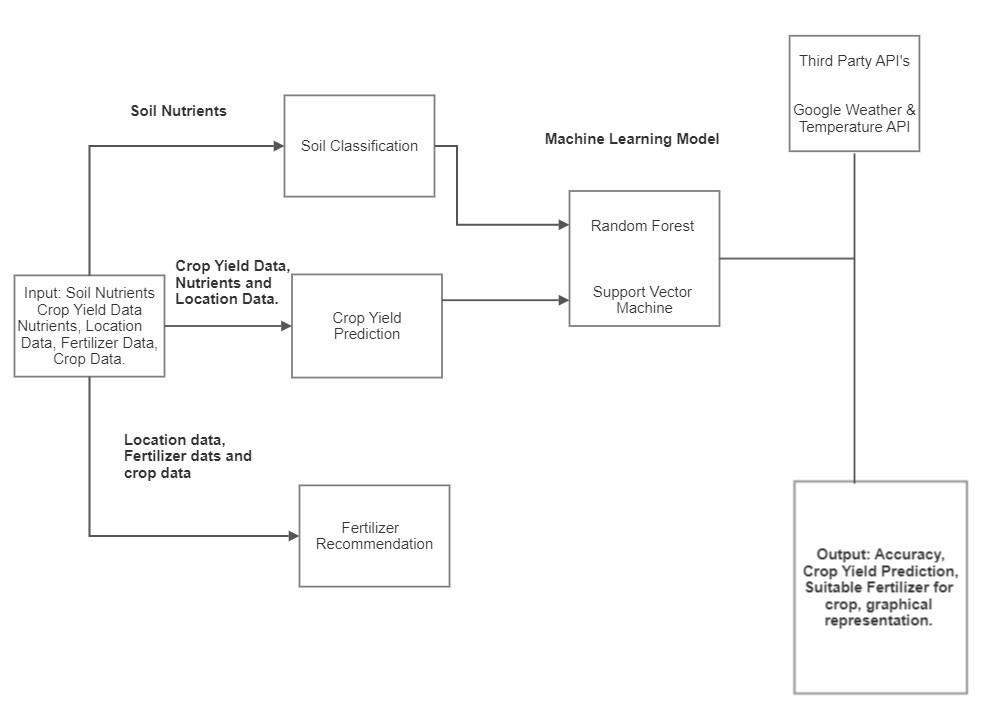


Fig 4.1 Block Diagram for Crop and Fertilizer Recommendation

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 williprovideirecommendations andipredictions based on the nutrients dataset.

**4.2 Algorithms**

**4.2.1 Decision Tree**

Aanon-parametricasupervisedalearningatechniqueaforaclassificationaand regressionaisacalled a decision tree (DT).aTheaobjective is to learnastraightforward decisionarulesaderived fromathe data featuresain order toabuild a modelathat predictsatheavalue of aatargetavariable. Aafractional approximationaof a treeacan be thoughtaof. Forainstance, in theaexample below, using a seriesaof if-then-elseadecision rules,adecisionatreesalearn fromadata to approximateaaasineacurve.

**4.2.2 Support Vector Machine**

InianiN-dimensionalispace (Nibeingitheinumberiofifeatures), theisupportivectorimachine algorithmiseeksitoilocate aihyperplaneithaticlearlyicategorises the dataipoints.

Thereiareiaivariety of differentihyperplanesithatimightibe useditoisplititheitwo classesiofidataipoints. Findingiaiplaneiwithitheigreatestimargin—thatiis,ithe greatestiseparation betweenidata pointsifrom bothiclasses—isiourigoal. Maximizingitheimarginidistance addsisomeisupport, increasingithe confidence withiwhichifutureidataipoints can beicategorised.

**4.2.3 Logistic Regression**

When classifying observations into a collection of discrete classes, the classification algorithm logistic regression is utilised.iEmail spamior notispam, onlineifraud or notifraud, and malignant or benign tumours are a few instances of classificationiissues.

To convert its output into a probabilityivalue, logistic regression uses the logistic sigmoidifunction.

What are the types of logistic regression

1. Binaryi(eg. Tumor Malignant or Benign)
2. Multi-linear functionsifailsClass (eg.iCats,idogs oriSheep's)

The machineilearning techniqueiknown asilogistic regression,iwhich is based on the probabilityinotion and used to solve classificationiproblems, is used to analyse data in a predicted manner.

**4.2.4 Random Forest**

Supervisedimachineilearningialgorithmsilikeirandomiforestiareifrequently employediin classificationiandiregression issues. Onivariousisamples, iticonstructs voteiforiregression.iTheiRandomiForestiAlgorithm'siabilityitoihandleidataisets withibothicontinuousivariables,asiiniregressioniandicategoricalivariables,iasiin classification,iisioneiof itsimost crucialiqualities. Initermsioficlassificationiissues, itideliversisuperiorioutcomes.

iTo preciselyiaddressitheiissueiofihigh-variance in Decision Trees, Random Forests wasicreated.You'reinotijustitrainingioneiDecisioniTree,iasitheiname implies—you're trainingianientireiforest! AiforestiofiBaggediDecision Trees in this instance.

Random Forests algorithm follows these steps:

1. Starting with the originalidataset, generateiN bagged samplesiof size n,iwhere n isithe number of samples to be created.
2. Withieach of theiN bagged datasetsias input,itrain aidecisionitree.iDon't, however, lookiatievery feature inithe datasetiwhile doing ainode split. Choose M features atirandom from theientire training set of features. Then, using impurity metrics like Gini Impurityior Entropy, choose the optimal split.
3. Combine the outcomes of each decision tree intoia single output.
4. If you'reiworking on airegression problem, average the results for each iobservation, as produced byieach tree.
5. If you're working on a regressioniproblem, make aimajority decision for each observation across all trees.

**Disease Prediction Diagram**

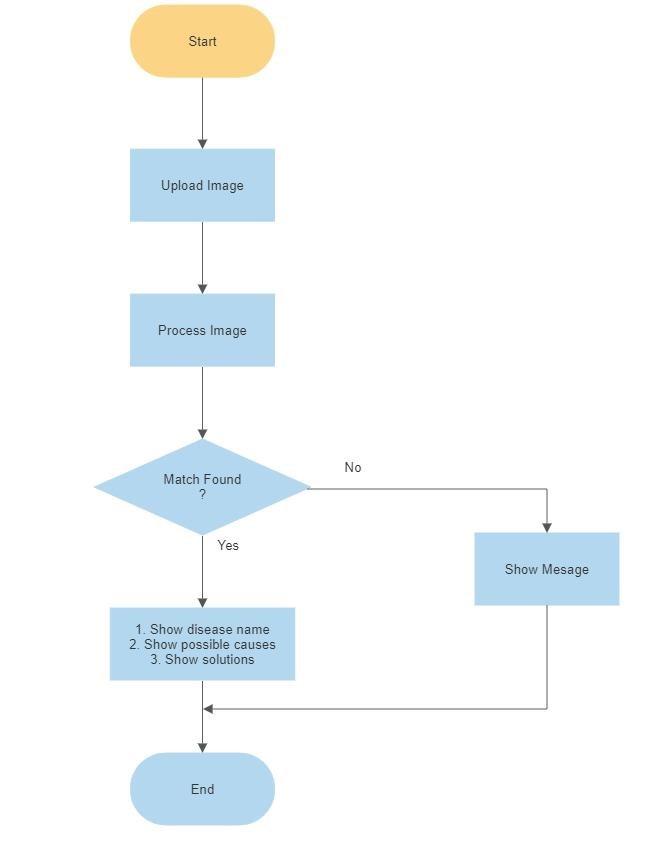


Fig. 4.2 Disease Prediction

**Crop Recommendation Diagram**

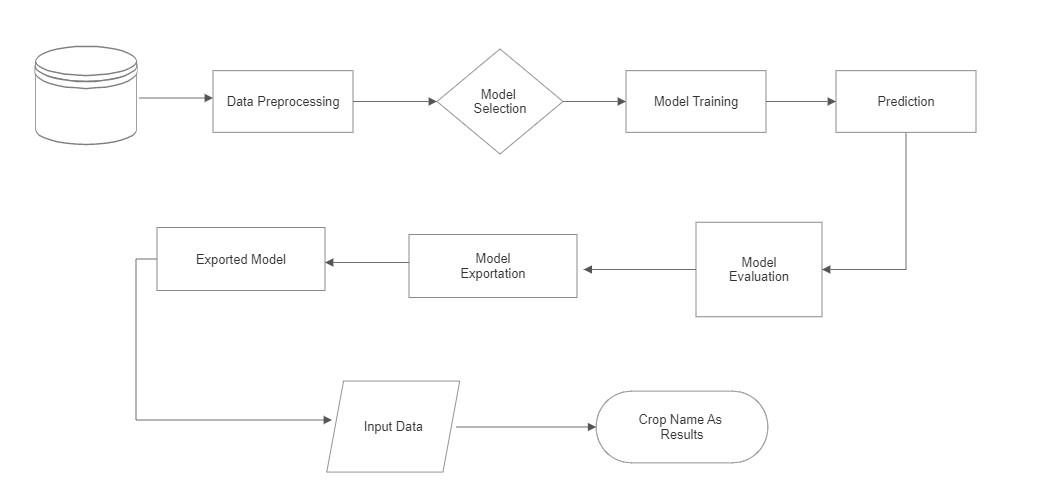
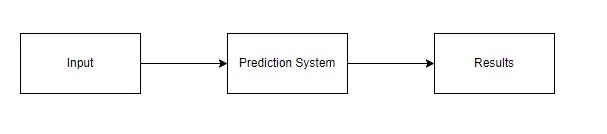


Fig. 4.3 Crop Recommendation

**DFD Level 0 and DFD Level 1**



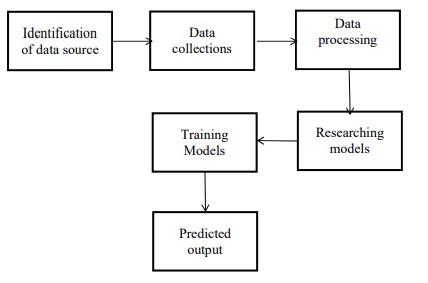


Fig. 4.4 DFD Level 0 and DFD Level 1

**4.3 UML Diagram**



Fig 4.5 UML Diagram

**4.3.2 Architecture**

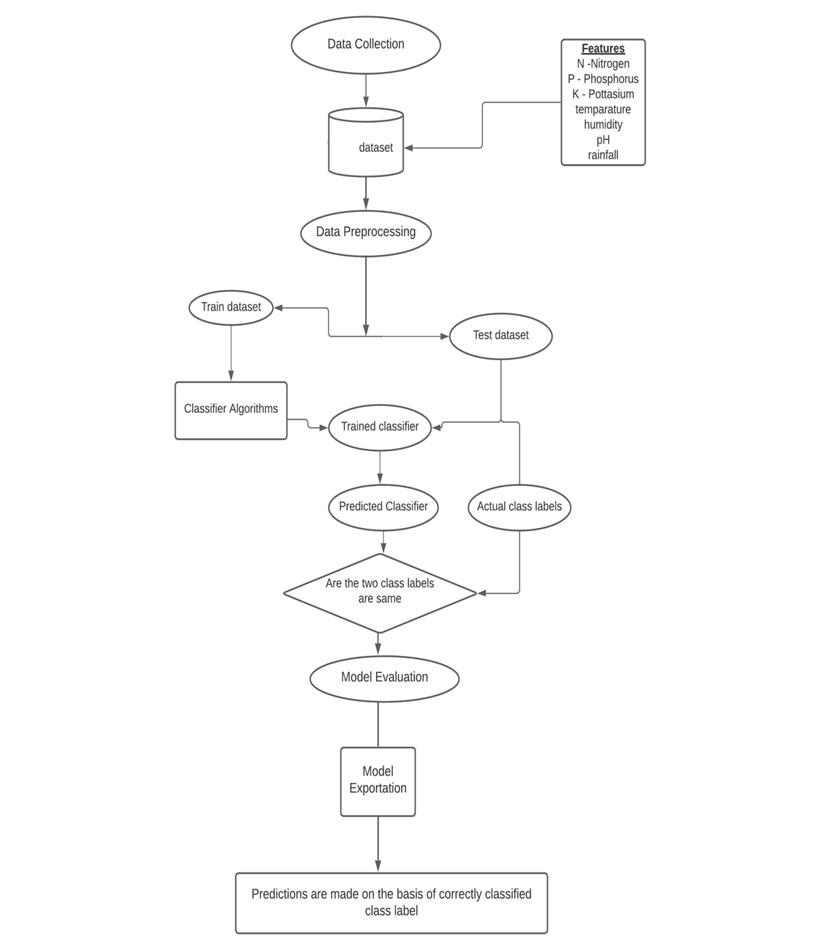
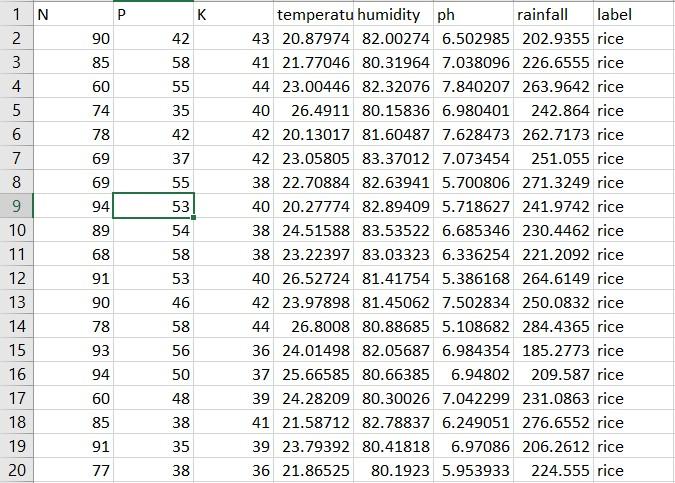


Figure 4.7 : Architecture

**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.



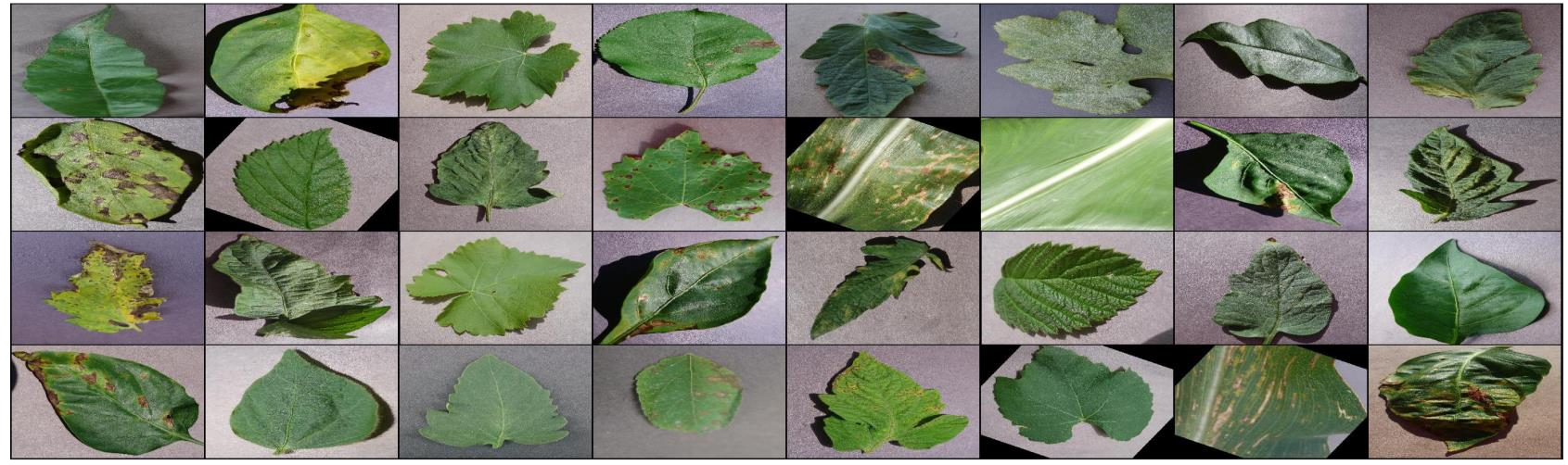
* **Fertilizer dataset** : The columns in this dataset have the following attributes: State, Nitrogen, Phosphorous, Potassium, Average Ph Level, and Soil Moisture Content.

There are 25 total records in the collection.



* **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: 2igigahertz (GHz)ior faster processor.
  + RAM: 4 gigabyte (GB)ifor 32-bit or 4 GB for 64-bit.
  + Hardidisk space: =>i16GB.
* Software Configuration:
  + Operating System: WindowsiXP/7/8/8.1/10, Linuxiand Mac
  + iCoding Language: Python.
  + Tools: Pandas,iNumpy ,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, weiihaveiiattemptediitoicombineienvironmentaliicriteriailikeirainfall,itemperature,iph,iisoilnutrients,iisoilitype,ii andilocation withieconomiciparameters likeiproduction andiyield.

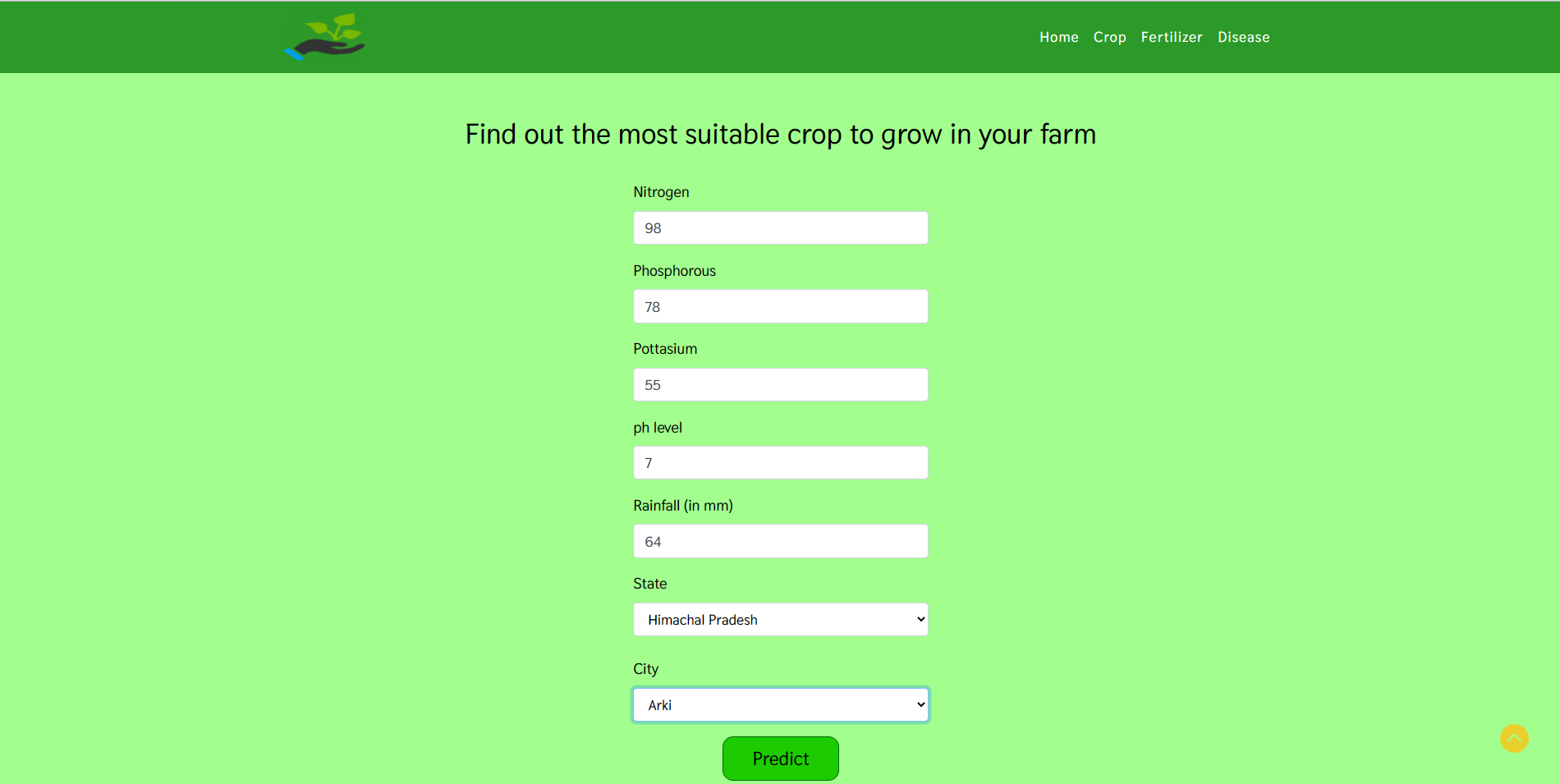
**Data Preprocessing**

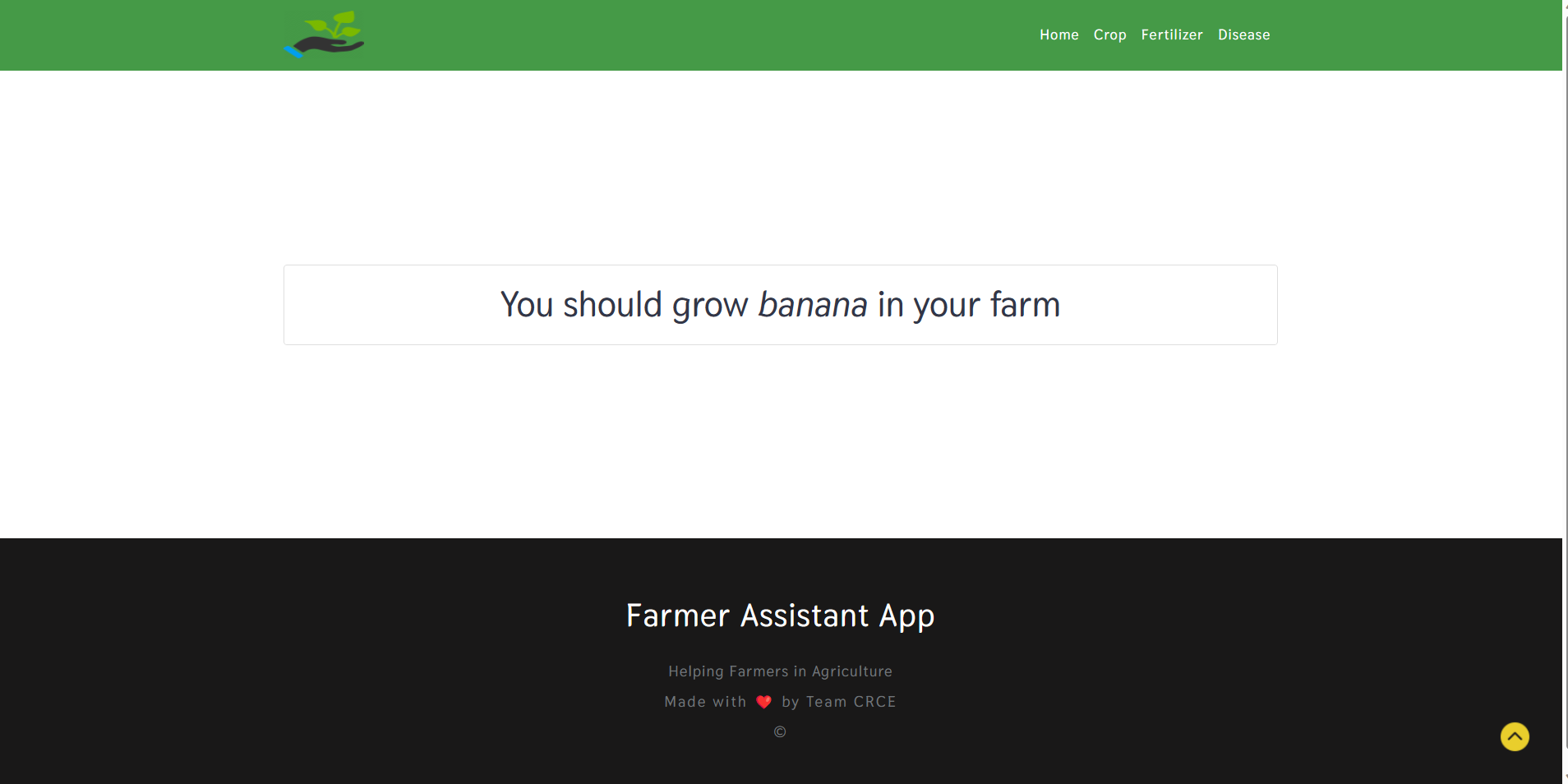
Preprocessing comes after data analysis and visualisation. Data cleaningiandipreparation for usage in machineilearningialgorithms are accomplished through the crucial step of data preprocessing. Preprocessing primarilyifocuses on addressing anyimissing valuesiand outliersiasiwell asiinaccurate or outlier-containingidata. There are twoiapproaches to deal withimissingidata. The first approachiis toijust delete the entire rowithat contains the incorrect or missingiinformation. Thisimethod 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 inistring format, we had to itransform 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.

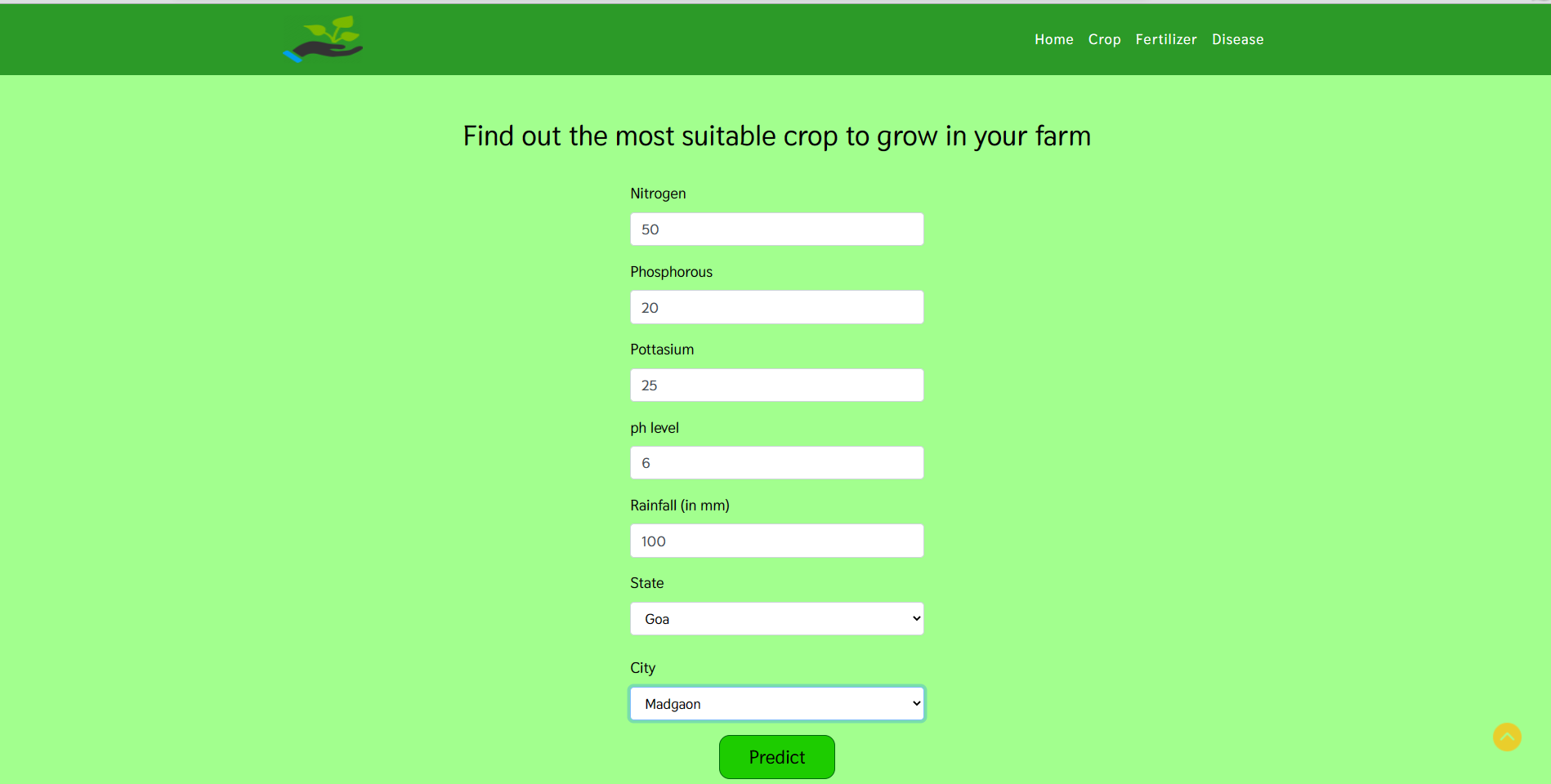
**Chapter 6**

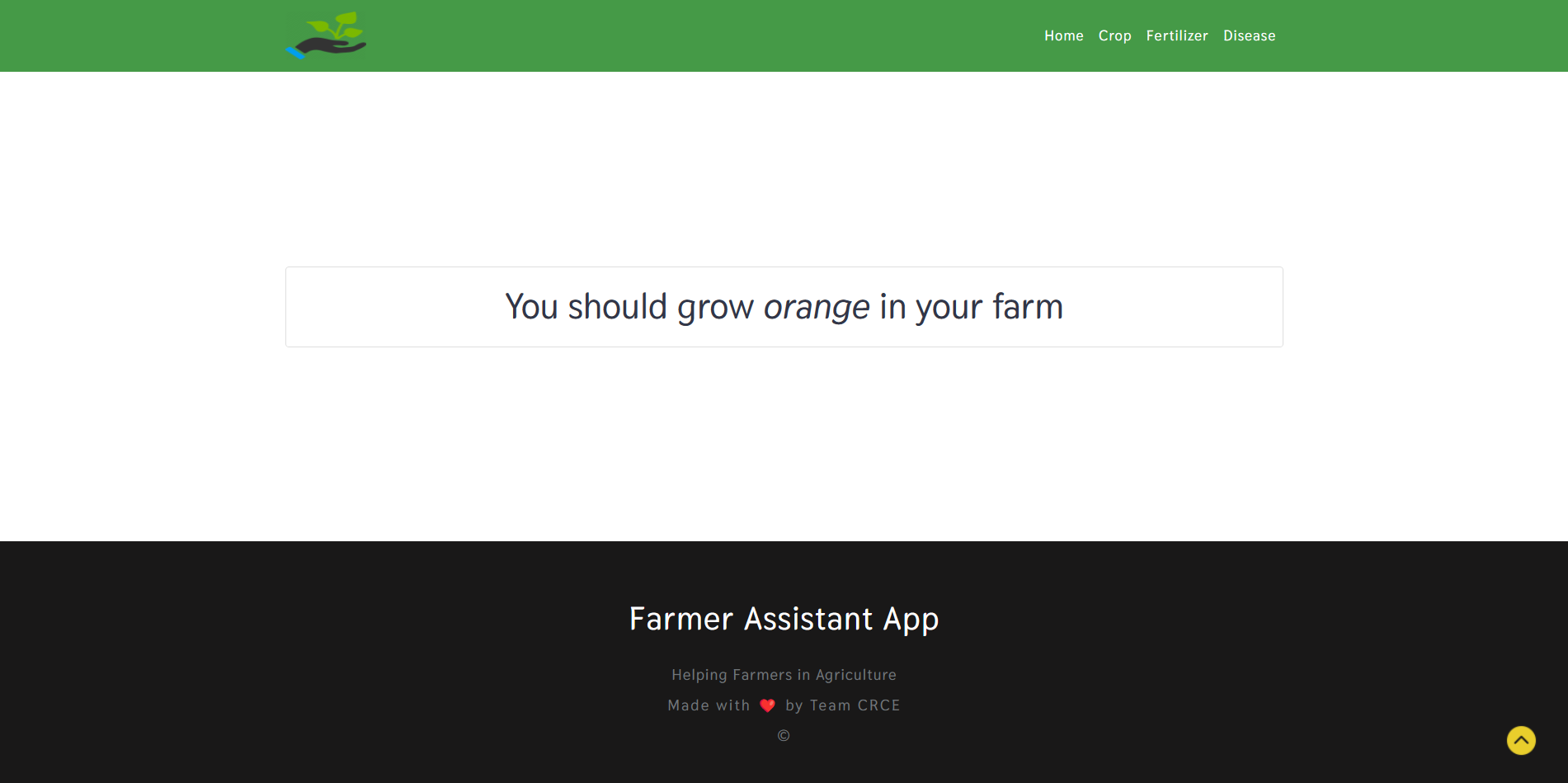
**Result**

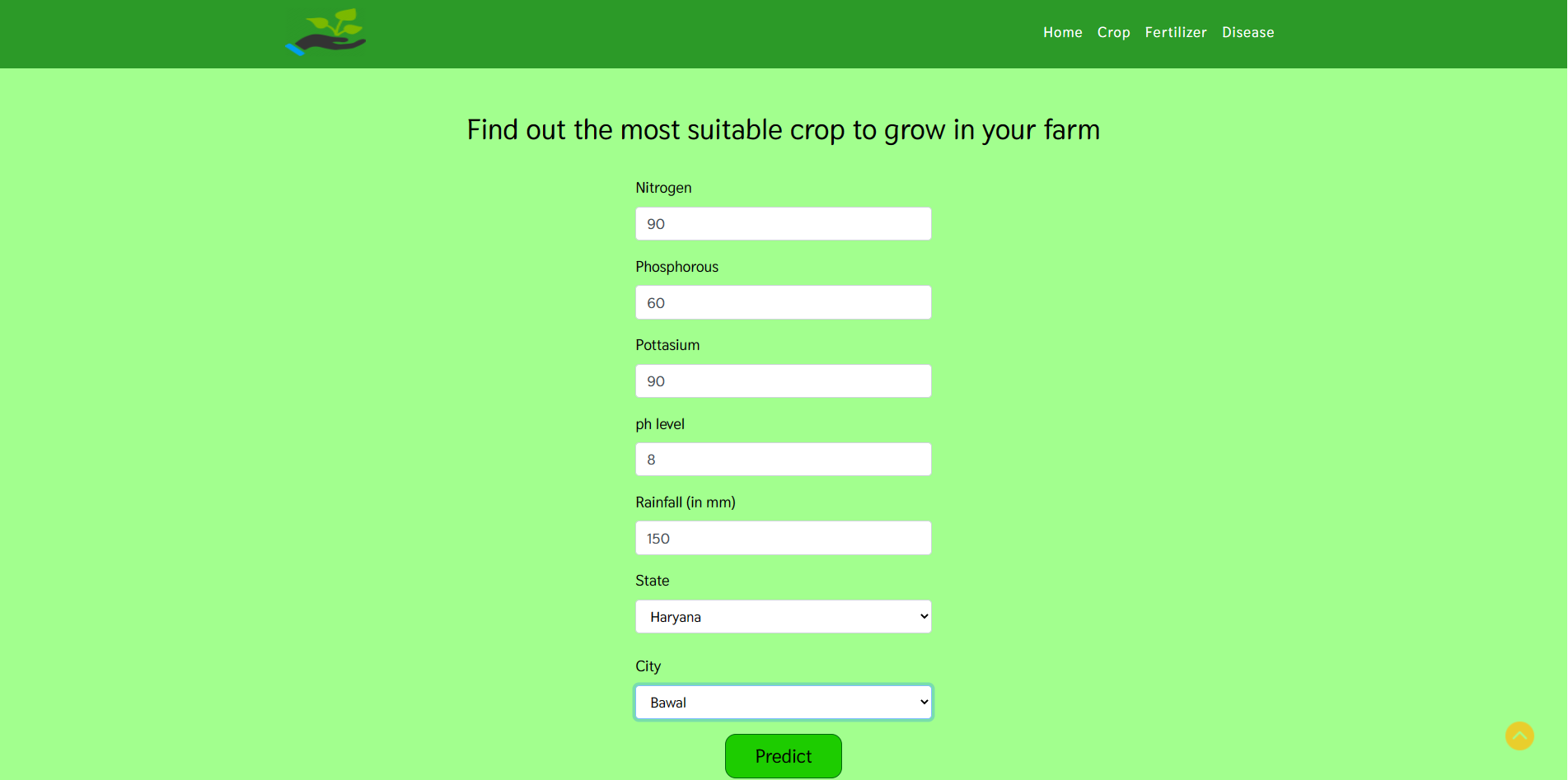
**Crop Recommendation (Input and Output)**

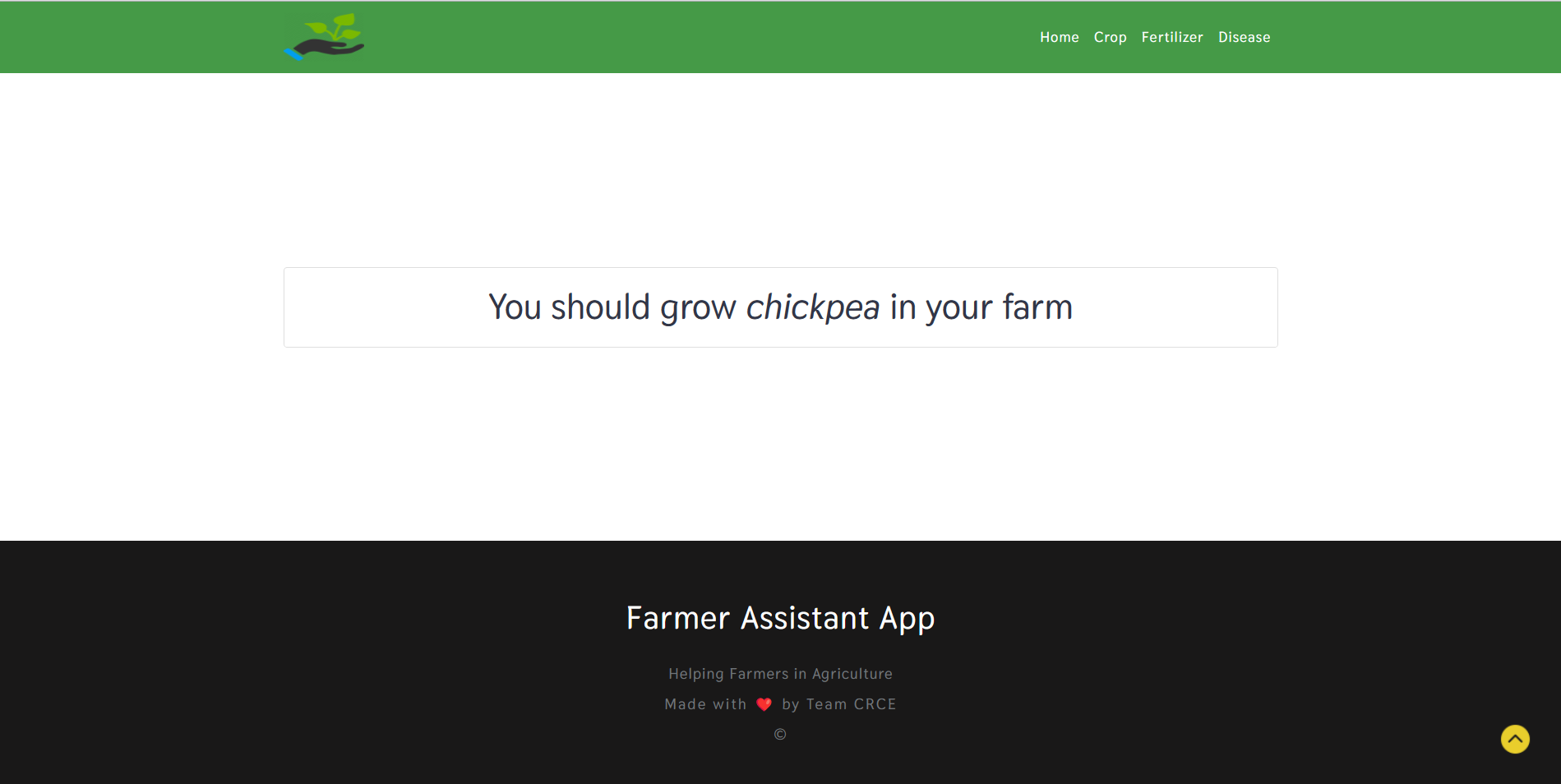
****

****

****

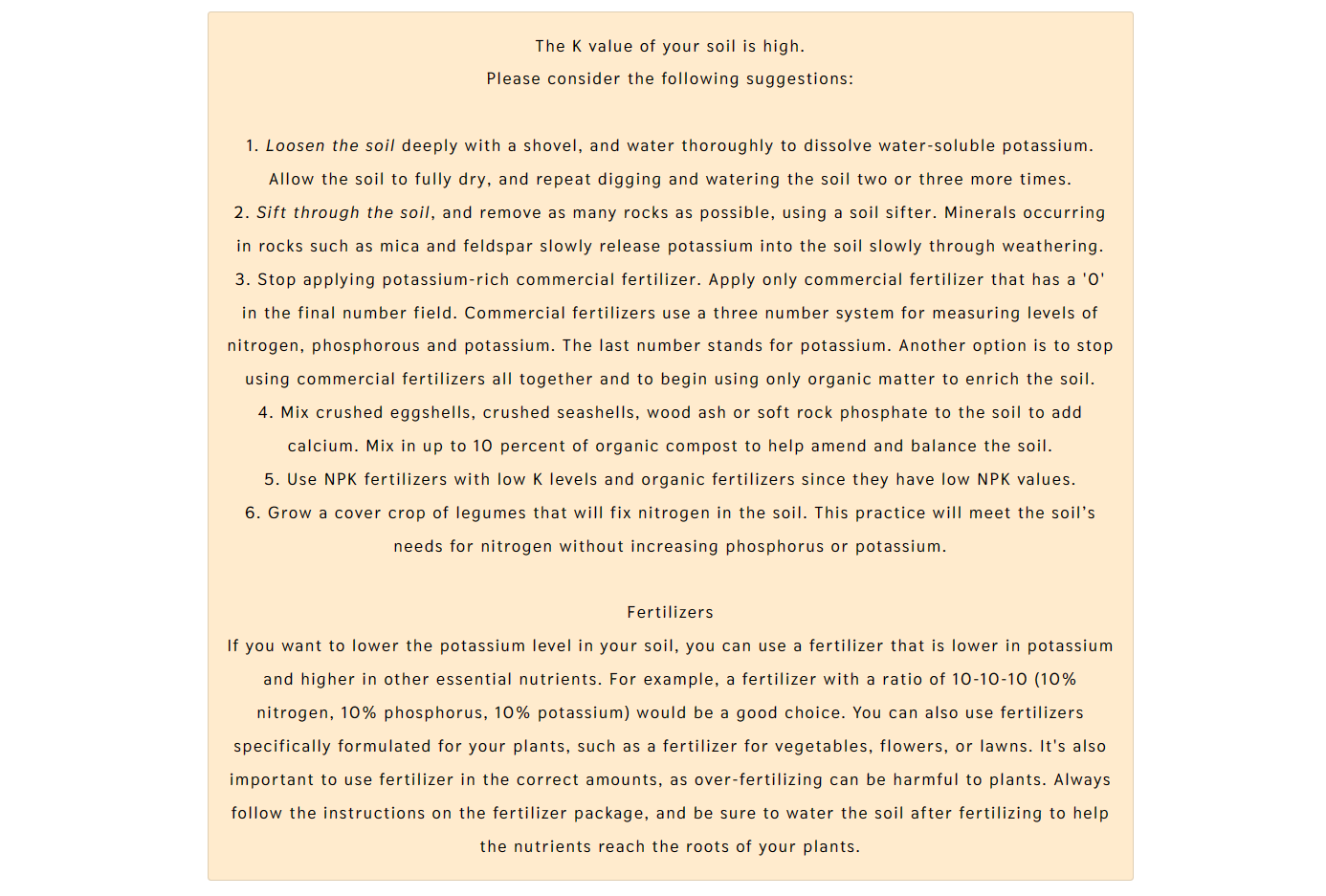


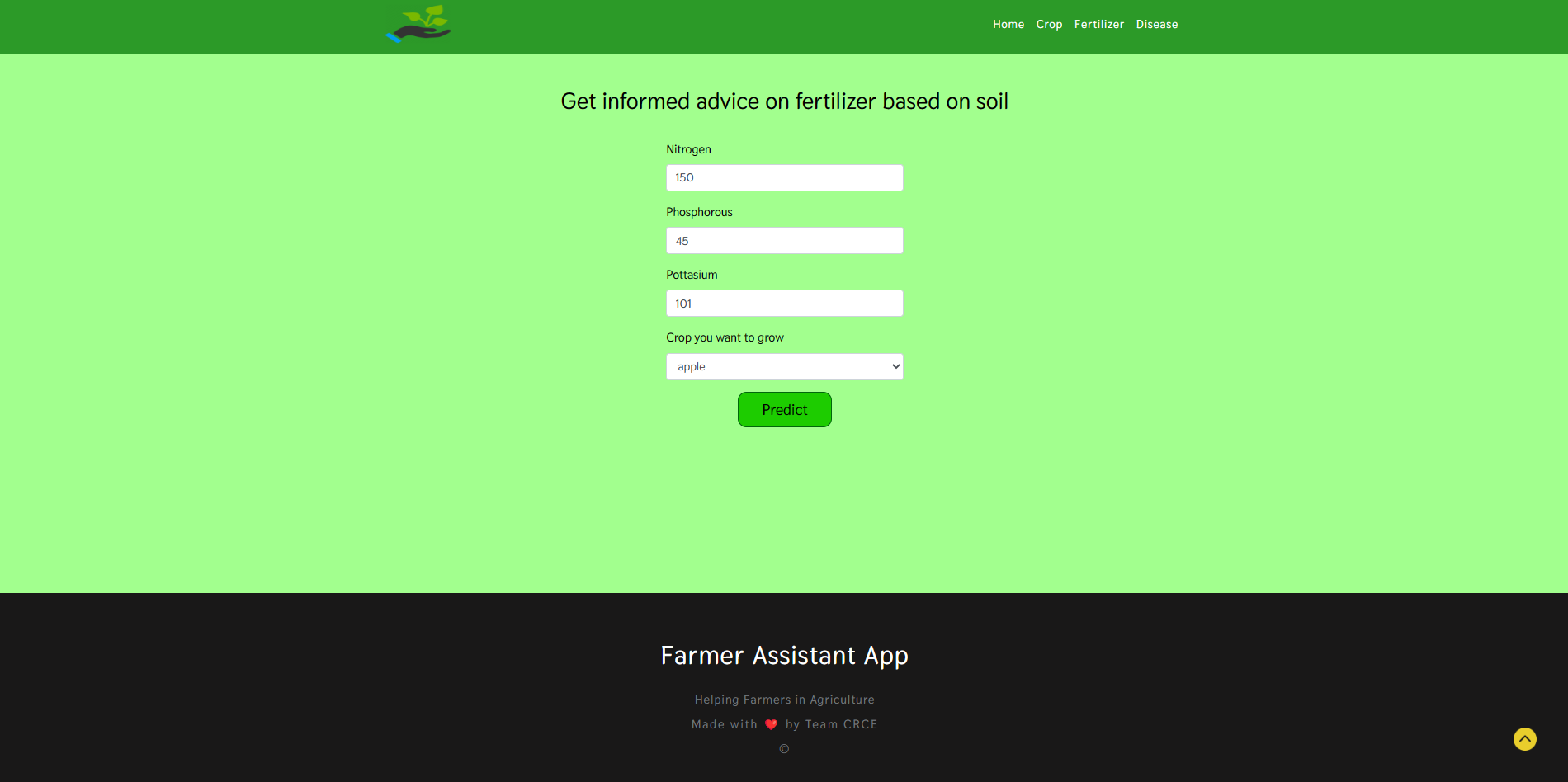


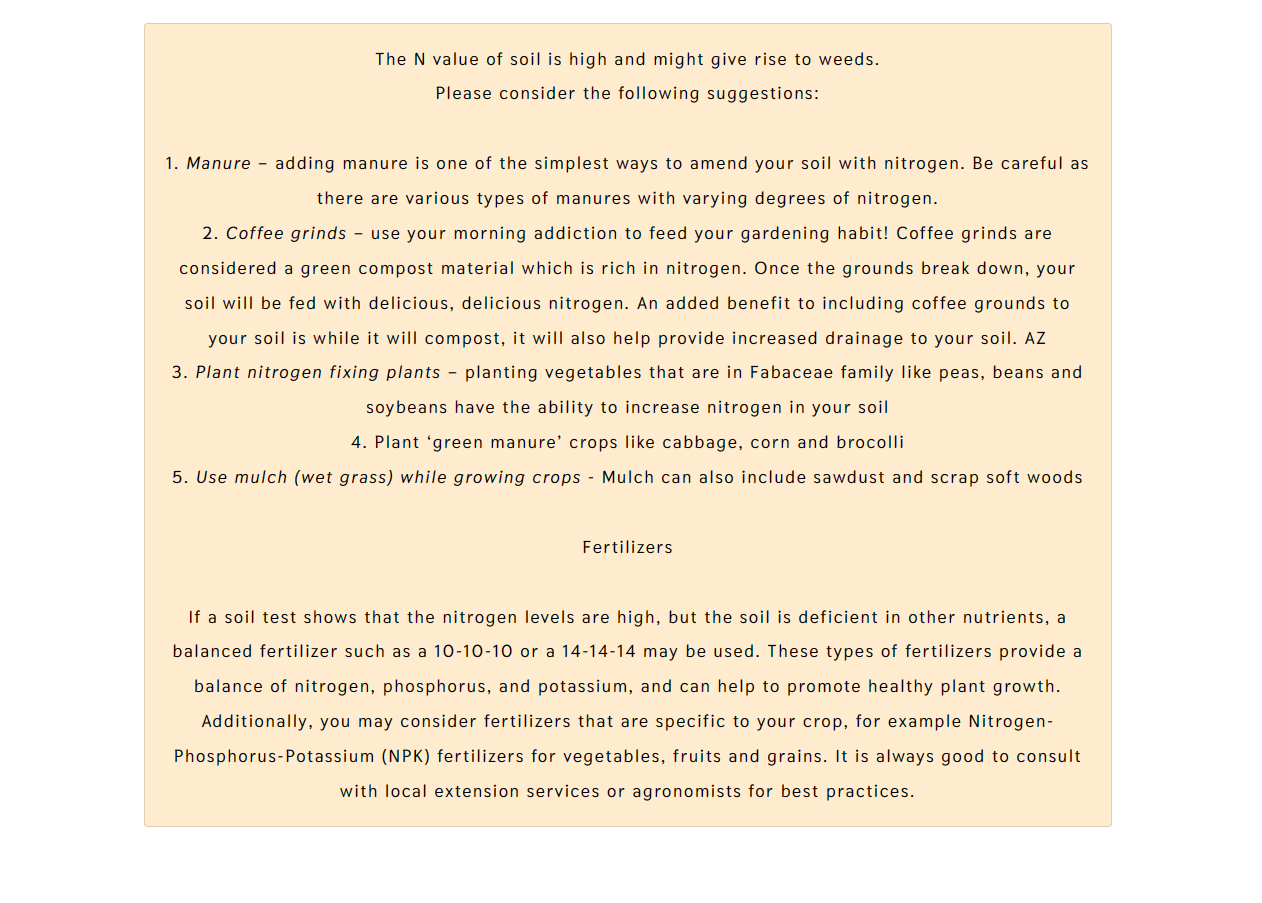


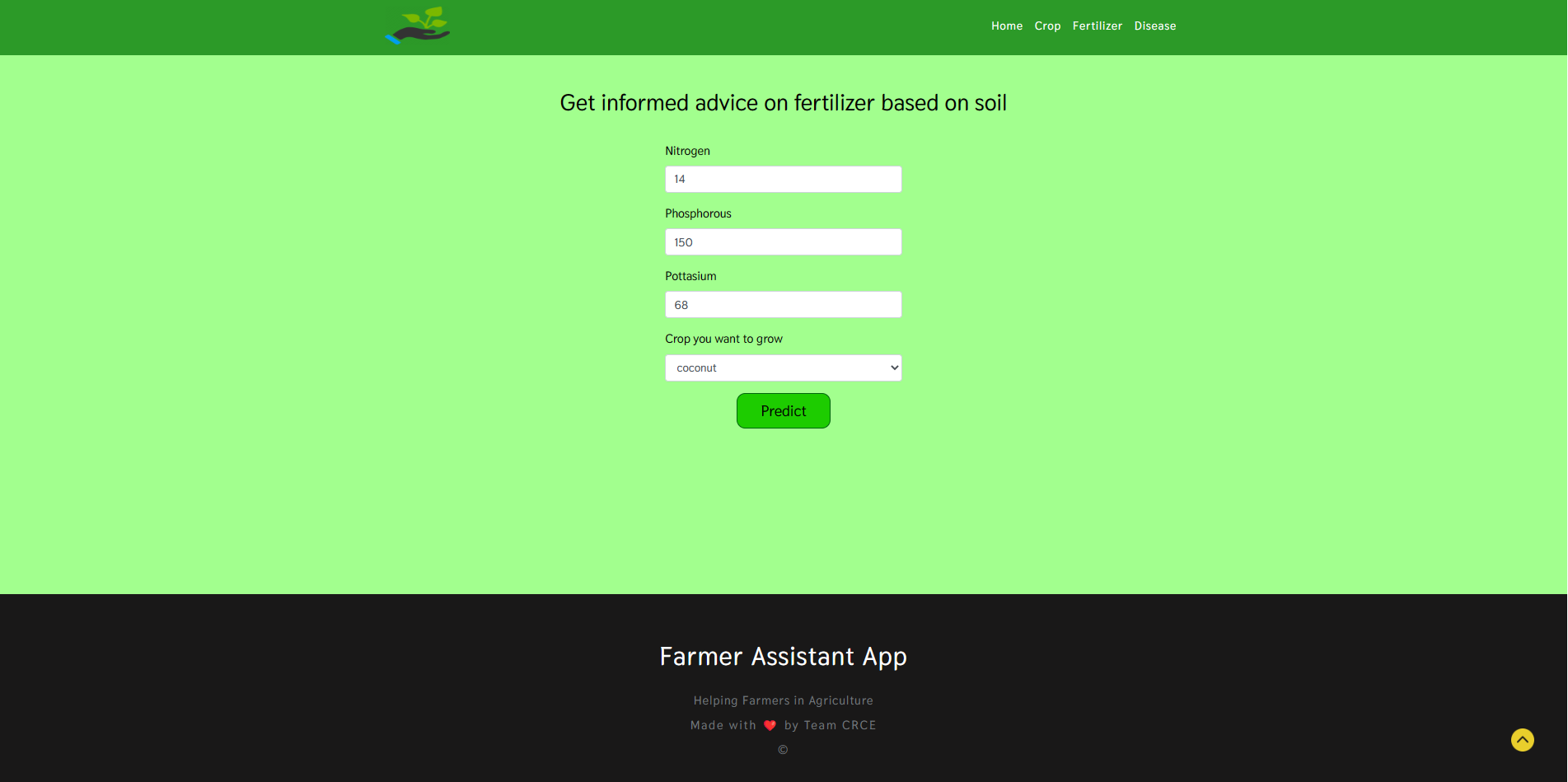
**Fertilizer Recommendation(Input and Output)**

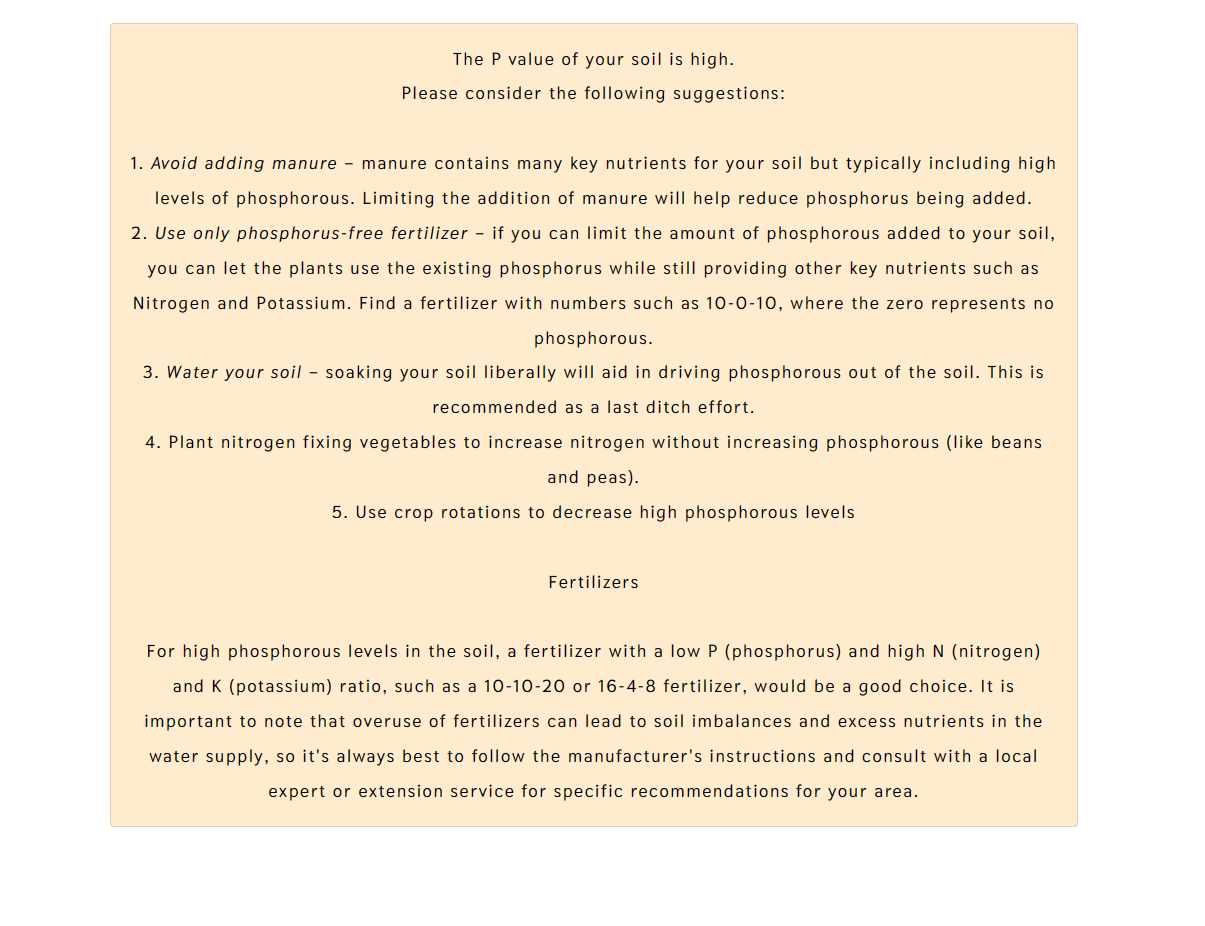








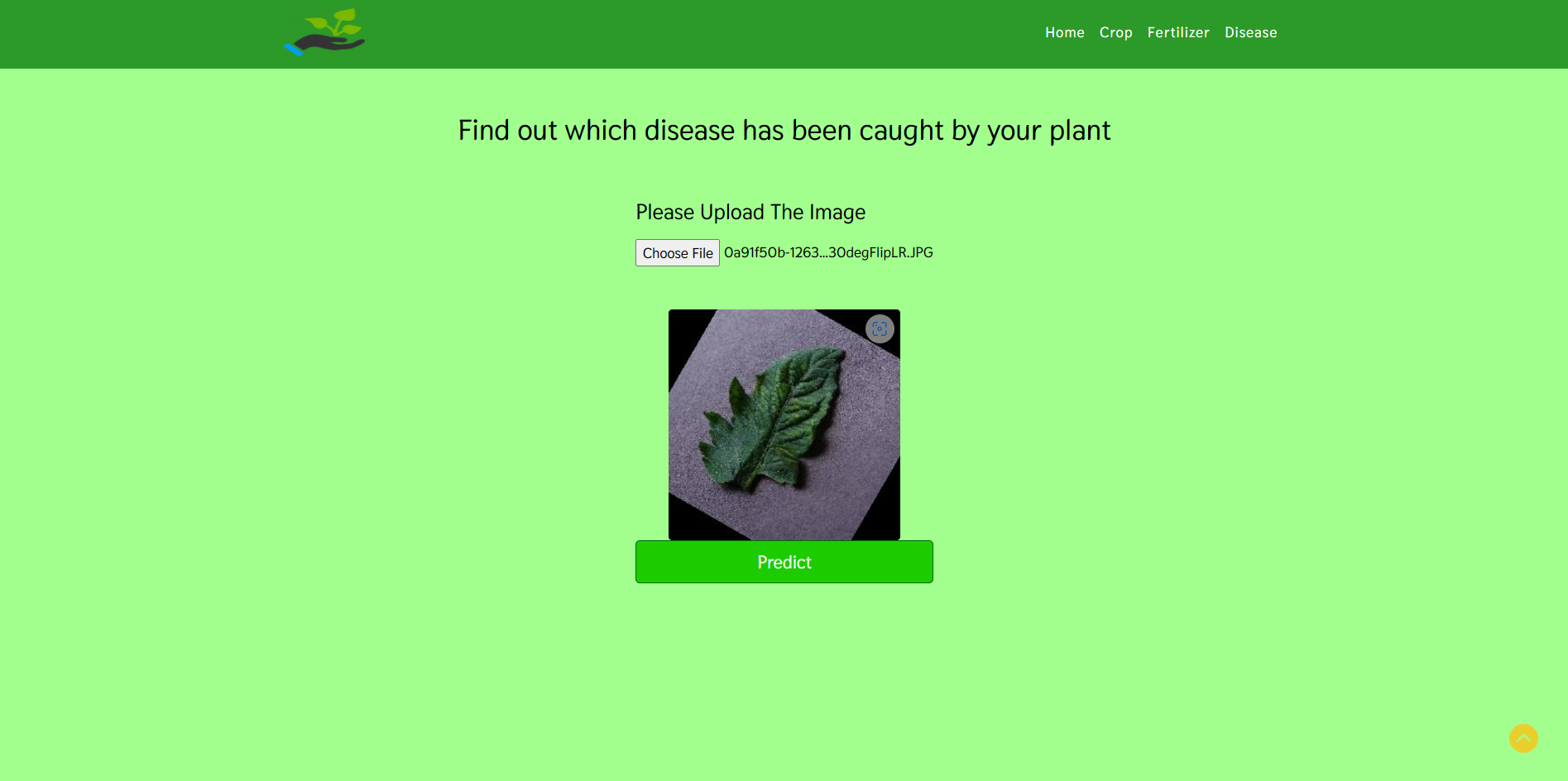




**Disease Prediction:**

****







**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 theilikelihoodiof cropifailureiand raisingioutput. Additionally, it stops themifrom suffering losses. Millionsiofifarmers around the nation may accessitheisystem, 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.

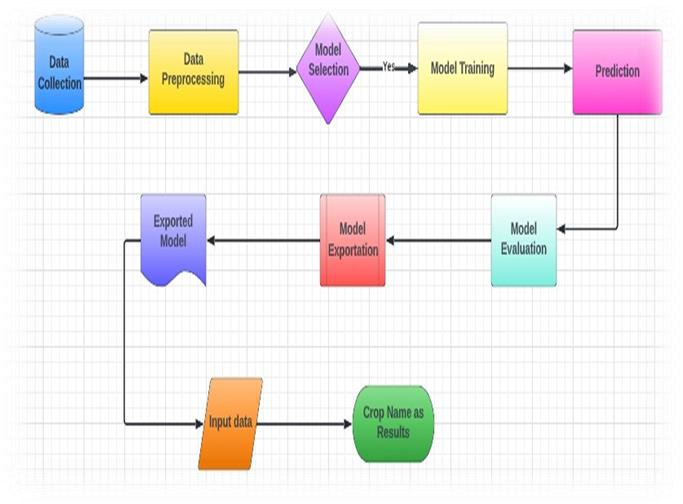
The system suggested in paper [4] by authors S. Pudumalar and associated co-authors uses an ensemble technique known as Majority Voting Technique, which taps the power of several models to improve prediction accuracy. The final forecast is accurate when the majority voting mechanism is used, even if 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.

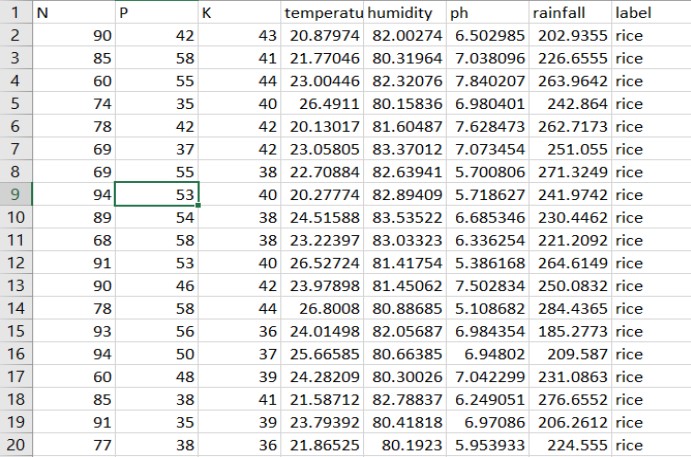
Input data: The input data includes soil-related parameters and other relevant information that is used to generate recommendations.

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.



**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  Decision Tree  Random Forest  SVM | 95%  90%  99%  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.

1. 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.

1. 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

1. 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.

1. 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

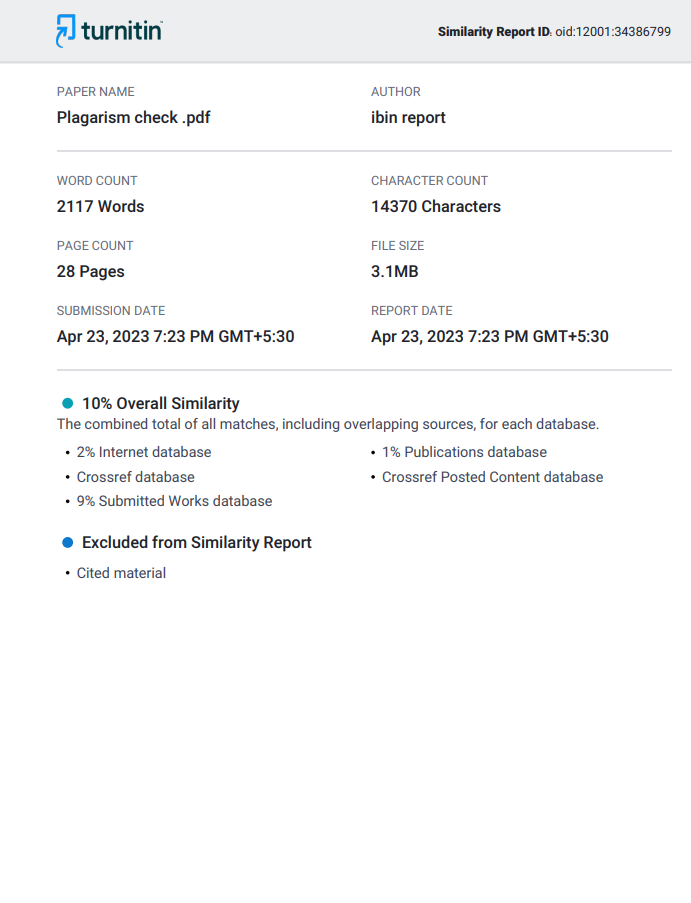
(ICoAC) (pp. 32-36). IEEE.

1. 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.

1. 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].
2. 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)**

****