A

Mini project report on

CROP PREDICTION USING MACHINE LEARNING

Submitted in partial fulfillment of the requirements for the award of Degree of

BACHELOR OF TECHNOLOGY

In

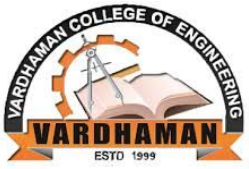
INFORMATION TECHNOLOGY

By

J Karthik reddy (17881A1213)
R Pradeepthi (17881A1224)
G Anush Reddy (17881A1204)

Under the Guidance of

Mr B.Ravinder Goud
Assistant Professor
Department of Information Technology



DEPARTMENT OF INFORMATION TECHNOLOGY
VARDHAMAN COLLEGE OF ENGINEERING
(AUTONOMOUS)
(Affiliated to JNTUH , Approved by AICTE and Accredited by NBA)
Shamshabad - 501 218, Hyderabad

DECLARATION

I hereby declare that the work described in this thesis entitled **"CROP PREDICTION USING MACHINE LEARNING"** which is being submitted by us in partial fulfillment for the award of **BACHELOR OF TECHNOLOGY** in the Department of Information Technology, Vardhaman College of Engineering to the Jawaharlal Nehru Technological University Hyderabad.

The work is original and has not been submitted for any Degree or Diploma of this or any other university.

Signature of the Student

1. J karthik reddy (17881A1213)
2. R Pradeepthi (17881A1224)
3. G Anush Reddy (17881A1204)

VARDHAMAN COLLEGE OF ENGINEERING, HYDERABAD

An autonomous institute, affiliated to JNTUH

DEPARTMENT OF INFORMATION TECHNOLOGY

CERTIFICATE

This is to certify that the project report entitled, **"CROP PREDICTION USING MACHINE LEARNING"**, done by **J Karthik Reddy (17881A1213), R Pradeepthi (17881A1224), G Anush Reddy (17881A1204),** Submitted to the Department of Information Technology, **VARDHAMAN COLLEGE OF ENGINEERING**, in partial fulfillment of the requirements for the Degree of **BACHELOR OF TECHNOLOGY** in **Information Technology**, during the year 2020-21. It is certified that he/she has completed the project satisfactorily.

Signature of Supervisor
Mr B.Ravinder Goud
Assistant Professor

Signature of Head of the Department
Dr. Muni SekharVelpuru
Associate Professor & Head

ACKNOWLEDGEMENT

The satisfaction that accompanies the successful completion of the task would be put incomplete without the mention of the people who made it possible, whose constant guidance and encouragement crown all the efforts with success.

We express our heartfelt thanks to **Mr.B Ravinder Goud**, Associate Professor & Major Project Supervisor, for his suggestions in the selection and carrying out an in-depth study of the topic. His valuable guidance and encouragement really helped us to shape this report to perfection.

We express our heartfelt thanks to **Dr. K Ramesh**, Associate Professor & Project Coordinator, for his suggestions invaluable inputs and assessment really helped us to shape this report to perfection.

We wish to express our deep sense of gratitude to **Dr. Muni SekharVelpuru**, Associate Professor & Head, Department of Information Technology, Vardhaman College of Engineering, for his intense support and encouragement, which helped us to mold our project into a successful one.

We also owe our special thanks to our honorable Principal **Dr. J. V. R Ravindra ,** of Vardhaman College of Engineering, for providing all the infrastructural facilities and congenial atmosphere to complete the project successfully.

We avail this opportunity to express our deep sense of gratitude and heartful thanks to **Dr. T. Vijender Reddy**, Chairman and **Sri T. Upender Reddy**, Secretary, of Vardhaman College of Engineering, for providing the infrastructural facilities and congenial atmosphere to complete the project successfully.

We also thank all the staff members of the Information Technology department for their valuable support and generous advice.

Finally, thanks to all my friends and family members for their continuous support and enthusiastic help.

1. **J karthik Reddy (17881A1213)**
2. **R Pradeepthi (17881A1224)**
3. **G Anush Reddy (17881A1204)**

ABSTRACT

Agricultural monitoring, in particular in developing countries,can help prevent famine and support humanitarian efforts. A central challenge is yield estimation, which is to predict crop yields before harvesting.

We introduce a scalable, accurate, and inexpensive method to predict crop yields using publicly available remote sensing data. This solution if implemented at the soil health centers which have been set up by the government could help all the farmers to use minimum fertilizers, so as to maintain the soil health and also would provide them an opportunity to gain at most revenue from the same piece of land.

Predictive analysis to suggest the top three more suitable crop based on the nutrition levels of the soil, temperature and also the expected revenue that this particular crop could generate.

Table of CONTENTS

LIST OF FIGURES

1.INTRODUCTION

1.1 INTRODUCTION TO THE PROJECT:

It is evaluated that 795 million individuals despite everything live without a satisfactory nourishment supply and thereby 2050 there were two billion individuals who take care of Dodds and Bartram. Completion hunger and improves nourishment security are essential objectives in the 2030 Agenda for Development of the United States. A focal test that addresses the nourishment security issue is yield estimation, in particular having the option to foresee crop yields a long time before reaping. Rural checking, specifically in creating nations, can improve nourishment creation and bolster helpful endeavors considering environmental change and dry spells (Dodds and Bartram). Existing methodologies depend on review information and different factors identified with crop development, (for example, climate and soil properties) to show crop yield.

This methodology is exceptionally fruitful in the United States, where the information is abundant and generally of high caliber. Far-reaching overviews of climate parameters, for example, the Daymet (Thornton) and land spread sorts, for example, the Cropland Data Layer (Bryan) are openly accessible and significantly encourage the harvest yield forecast task. Be that as it may, data about climate, soil properties, and exact land spread information are ordinarily not accessible in creating nations which have the best requirement for dependable harvest yield expectation. Remote detecting, then again, is a comprehensively accessible and practical information source that has as of late accumulated a lot of intrigues. It is much of the time utilized in computational manageability applications, for example, species circulation demonstrating, destitution mapping, atmosphere displaying, and forestalling catastrophic events.

In the project, we have proposed a methodology depends upon present-day include the learning thoughts, which as of late prompted huge enhancements in the scope of PC vision errands (Krizhevsky, Sutskever, and Hinton). We defeat the shortage of preparing information by utilizing another dimensionality decrease procedure. In particular, we treat crude pictures as the histograms of pixel tallies, rough through the high-dimensional histogram with the help of a mean-field supposition. Profound the learning designs are prepared on those histograms to foresee the crop yield. While this methodology performs

pretty well, it doesn't unequivocally represent spatial-worldly conditions between information points. We beat this constraint by fusing Gaussian Process on the profound model. We assess our methodology on an undertaking of the foreseeing province-level soybeans creation in the United States. Test results show that the model we have chosen outflanks contending methods by a huge edge while staying interpretable regarding highlight significance.

1.2 PROBLEM SCOPE:

In our country large amount of population are depending on agriculture though government is taking financial steps to help farmers still they are facing problems due to lack of data anlaysis and prediction on crops

1.3 PURPOSE/OBJECTIVE OF PROJECT:

Our Objective is to develop an application using Machine Learning for Predicting which Crop to be used based on Soil Condition using "K nearest neighbour neighbour classification".

1.4 MOTIVATION:

A harvest expectation is a boundless issue that happens. During the rising season, a rancher had an interest in knowing how much yield he is going to anticipate. In the prior period, this yield forecast becomes a self-evident truth depended on Farmer's drawn-out understanding for explicit yield, crops and climatic conditions. Rancher legitimately goes for yield forecast instead of worried on crop expectation with the current framework. Except if the right harvest is anticipated how the yield will be better and also with existing frameworks pesticides, natural and meteorological parameter identified with the crop isn't thought of. Advancing and alleviating the rural creation at an all the more quickly pace is one of the fundamental circumstances for farming improvement. Any harvest's creation shows the route either by the enthusiasm of area or improvement in yield or both. In India, the possibility of augmenting the locale under any yield doesn't exist with the exception of by restoring to increment trimming quality or harvest substitution. Along these lines, varieties in a difficult situation the region and create thorough trouble. In this way, there is have to endeavour great procedure for crop expectation so as to conquer the existing issue.

2.LITERATURE SURVEY(Existing System)

2.1 EXPLAINATION ABOUT EXISTING SYSTEM:

Remote detecting information has been generally utilized for foreseeing the crop yield in remote detecting network like (Bolton and Friedl). Be that as it may, every single existing methodology we know about depending close by made highlights, on the presumption that they can catch the vast majority of the data identified with vegetation development contained in high dimensional pictures. Some broadly utilized highlights incorporate Normalized the Difference between Vegetation Index (NDVI) (Quarmby et al. 1993; Johnson 2014), two-band Enhanced Vegetation Index (EVI2) (Bolton and Friedl 2013) and Normalized the Differences of Water Index (NDWI) (Satir and Berberoglu 2016).

2.2 LIMITATIONS OF EXISTING SYSTEM:

While the noteworthy exertion had been committed including building, Existing highlights are like genuinely unrefined lists that rely upon the modest number (typically two or more), from the accessible picture groups.

2.3 PROPOSED SYSTEM:

Our undertaking is to build up an application utilizing AI for foreseeing which yield to be utilized dependent on soil condition utilizing k closest neighbour characterization. Our trial results recommend that our scholarly highlights are substantially more compelling and that groups that are regularly overlooked could assume a significant job.

Favourable circumstances:

- Farmers can realize which harvest is possibly dependent on their dirt kind.

- Chances if expanding pay for ranchers dependent on the investigation.

3. ANALYSIS

3.1 Introduction:

Necessity Specifications gives a highly secure stockpiling to a web server effectively. Programming prerequisites manage programming and equipment assets that could be introduced on to a service which gives absolute working to the application that was developed. The product and equipment essential should be introduced before the bundles are introduced. These are the most widely acknowledge arrangement of necessities characterized by any activity structure. These product and equipment prerequisites offer a good help to the activity framework in building up an application.

3.2 Software Requirement Specification:

3.2.1 Software Requirement:

The software requirements describe the use of required shareware products like database management system. The required shareware product describes the number and version. Each interface describes the purpose of the interfacing shareware as related to this shareware product.

Operating system	:	Windows 7/10
Coding Language	:	Html, JavaScript
Development Kit	:	Flask Framework
Programming language	:	Python
IDE	:	Anaconda prompt

3.2.2 Hardware Requirement:

The equipment necessity determines every interface of the product components and the equipment components of the framework. These equipment prerequisites incorporate arrangement qualities.

System	:	Pentium IV 2.4 GHz.
Hard Disk	:	100 GB.
Monitor	:	15 VGA Colour.
Mouse	:	Logitech.
RAM	:	1 GB.

3.3 K-NN ALGORITHM:

The KNN algorithm assumes that similar things exist in close proximity. In other words, similar things are near to each other.The k-nearest neighbors (KNN) algorithm is a simple, easy-to-implement supervised machine learning algorithm that can be used to solve both classification and regression problems.

Steps:

1.Load Data

2.Initialize K to your chosen number of neighbors.

3.For each example in the data.

3.1 Calculate the distance between the query example and the current example from the data.

3.2 Add the distance and the index of the example to an ordered collection.

4. Sort the ordered collection of distances and indices from smallest to largest (in ascending order) by the distances.

5. Pick the first K entries from the sorted collection.

6. Get the labels of the selected K entries.

7. If regression, return the mean of the K labels.

8. If classification ,return the mode of the K labels.

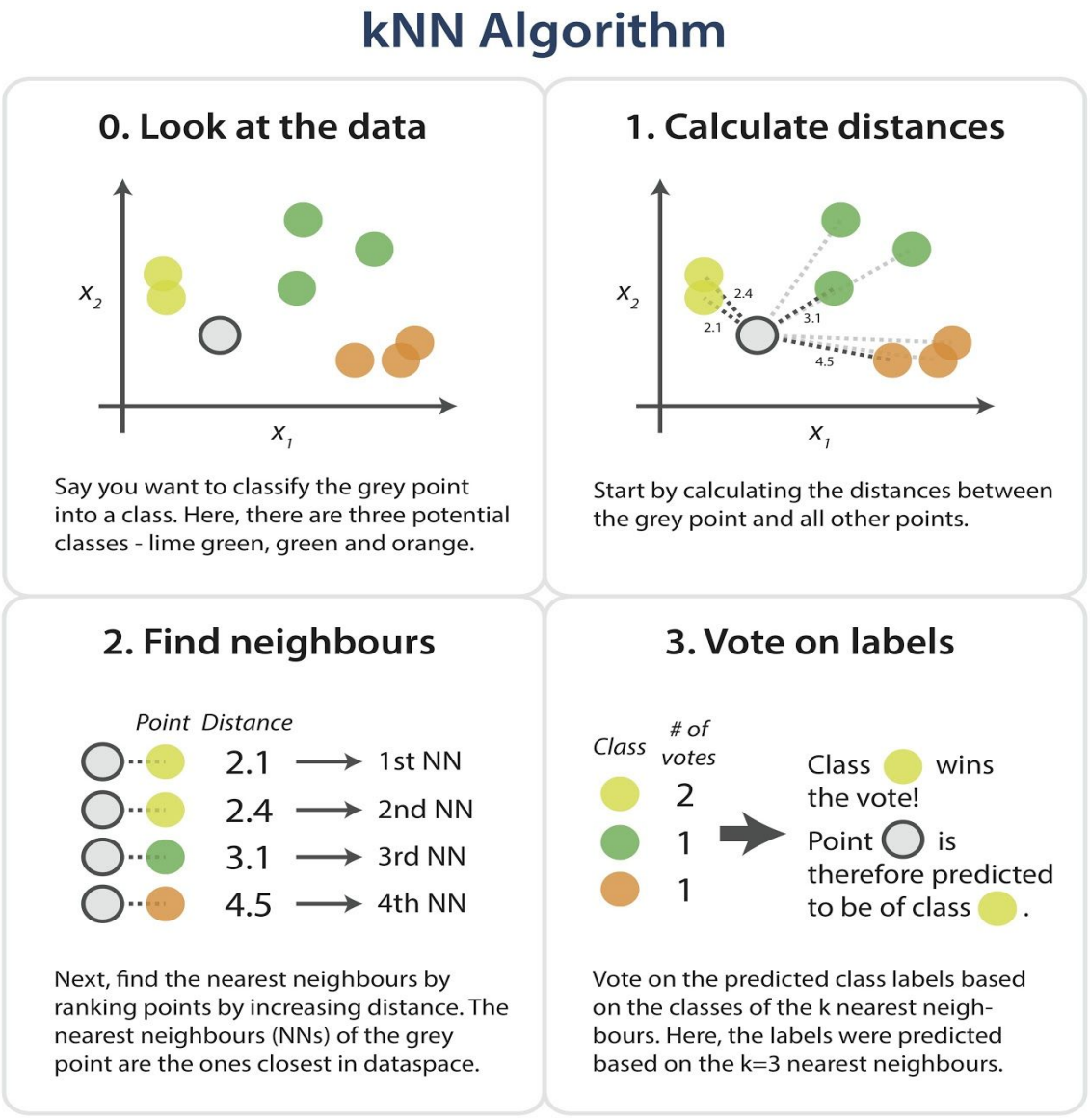


Figure 1 KNN Algorithm

4.DESIGN

4.1 INTRODUCTION TO SYSTEM ARCHITECTURE:

The motivation behind the structure stage is to orchestrate an answer to the issue, for example, by need report. This part has its initial move in the moving issue area to the appropriate response space. The planning stage fulfills all the necessities of the skeleton. The structure of the skeleton is likely the preeminent urgent issue of caring nature the standard of a product bundle. It's a genuine effect on the later part, strikingly to testing and supporting.
The Capitulate of this part is that of the style of the archive. This record closely resembles as an outline to the answer and is utilized later throughout the execution, testing and support. The structured movement is generally outlay into 2 separate stages known as System Design and Detailed Design.

Framework Design conjointly alluded to as top-positioning style expects to detect the modules that should be inside the skeleton, the determinations of those modules, and the manner in which they move with each other to provide the predefined results.
At the highest point of the framework style all the fundamental information structures, record groups, yield positions, and furthermore the significant modules inside the framework and their particulars square measure set. Framework configuration is that the strategy or speciality of the procedure the structure, segments, modules, interfaces, and information for a framework to fulfil, for example, needs. Clients will peruse it in light of the fact that the utilization of frameworks hypothesis to improvement.

Definite Design, the inward rationale of everything about modules spread out in framework configuration is resolved. All through this part, the important part contains information of a module square which estimates the measure here and there spread out in a significant level style portrayal language that is independent of the objective language inside which the product bundle can inevitably be implemented.

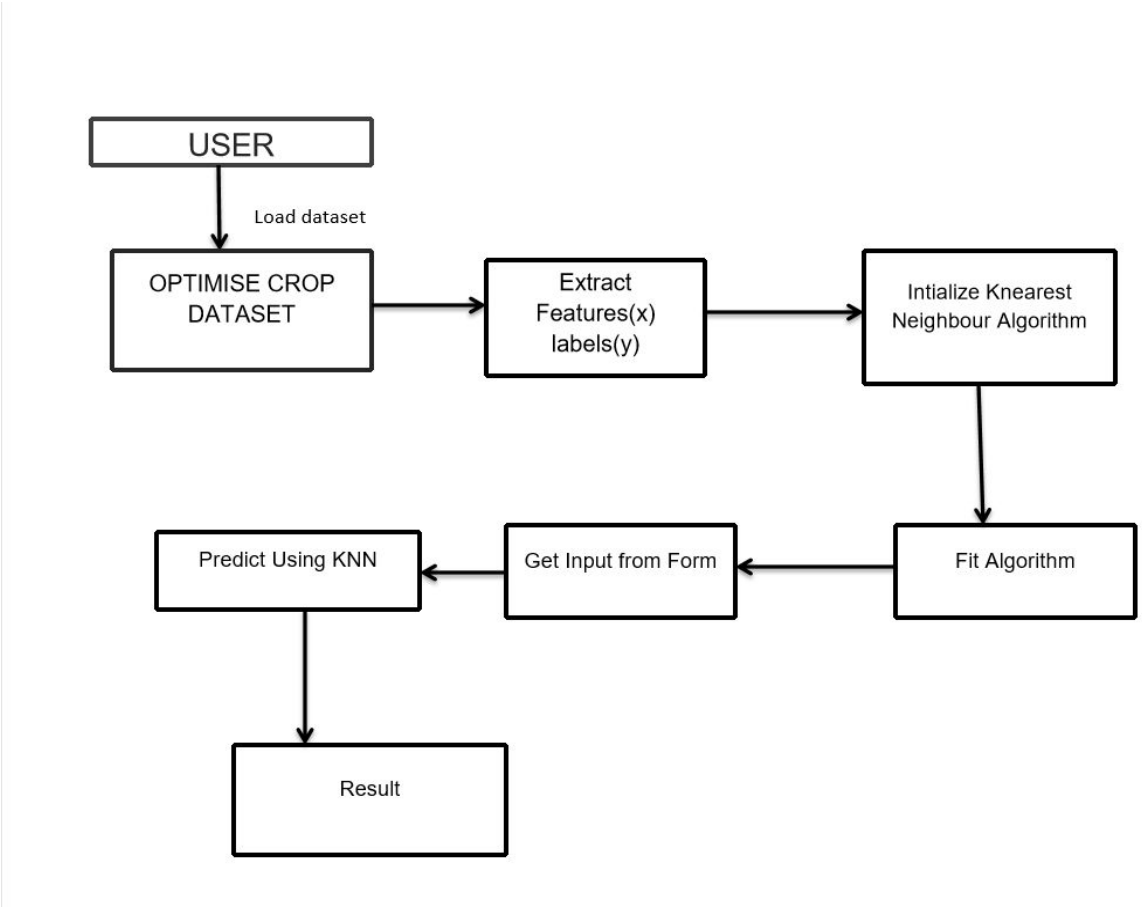


Figure 2 Architecture

USER INPUT:

The input is the input given by the user by filling all the 17 parameters.

TRAIN DATA:

The train data is the data which is taken from the dataset by taking 70% of the dataset randomly.

TEST DATA:

The test data is the data which is taken from the dataset by taking remaining 30% of the dataset**.**

PREDICTION:

It will predict, which crop is suitable for the farmer based on the predicted class.

4.2 UML Diagrams:

The UML allows the product planner to impart an examination model using the exhibiting documentation that is directed by a great deal of syntactic-semantic and down to business rules. A UML structure is addressed using five one of a kind viewpoints that portray the system from obviously exchange perspective. Each view is portrayed by a ton of diagram, which is according to the accompanying.

USE CASE DIAGRAM :

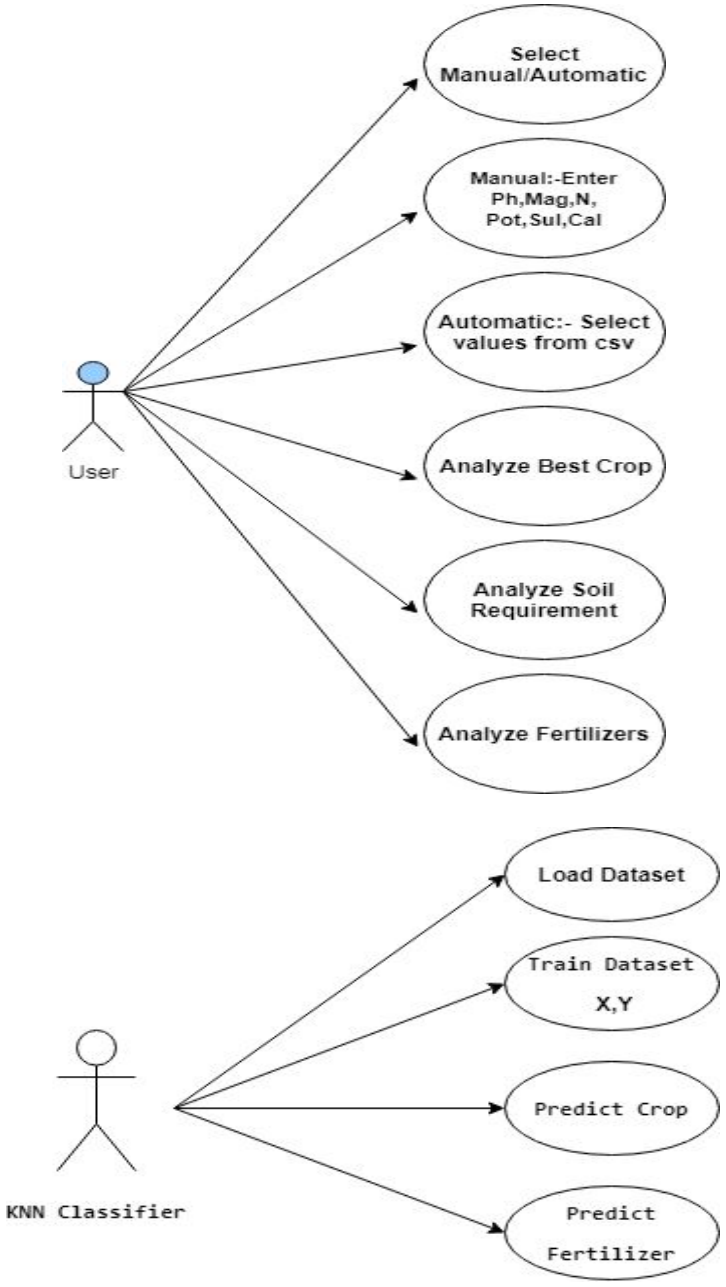


Figure 3 Use Case Diagram

CLASS DIAGRAM:

The class graph is the rule assembling square of article masterminded illustrating. It is used both for the generally determined showing of the deliberate of the application and for point by point exhibiting translating the models into programming code. Class diagrams can in like manner be used for data illustrating. The classes in a class diagram address both the essential things, associations in the application and the classes to be altered. A class with three portions, in the blueprint, classes are addressed with boxes which contain three segments:

- The upper part holds the name of the class
- The focus part contains the characteristics of the class
- The base part gives the strategies or exercises the class can take or grasp

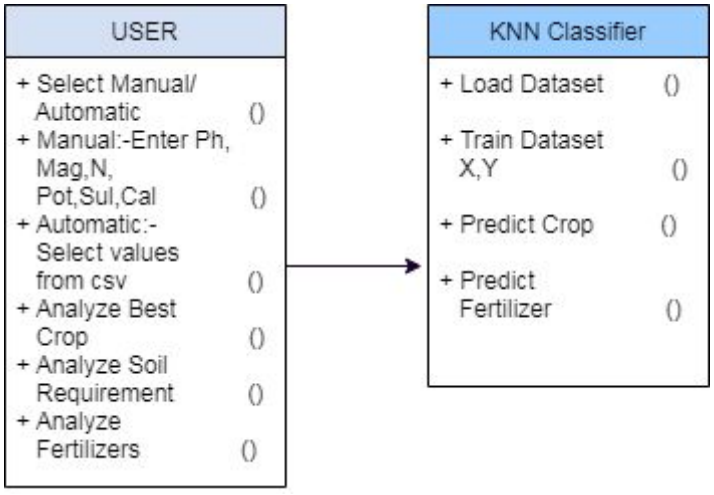


Figure 4 Class Diagram

ACTIVITY DIAGRAM:

Activity diagrams are graphical portrayals of work processes of stepwise exercises and activities with help for decision, cycle and simultaneousness. In the Unified Modeling Language, movement charts can be utilized to depict the business and operational bit by bit work processes of parts in a framework. An action chart shows the general progression of control.

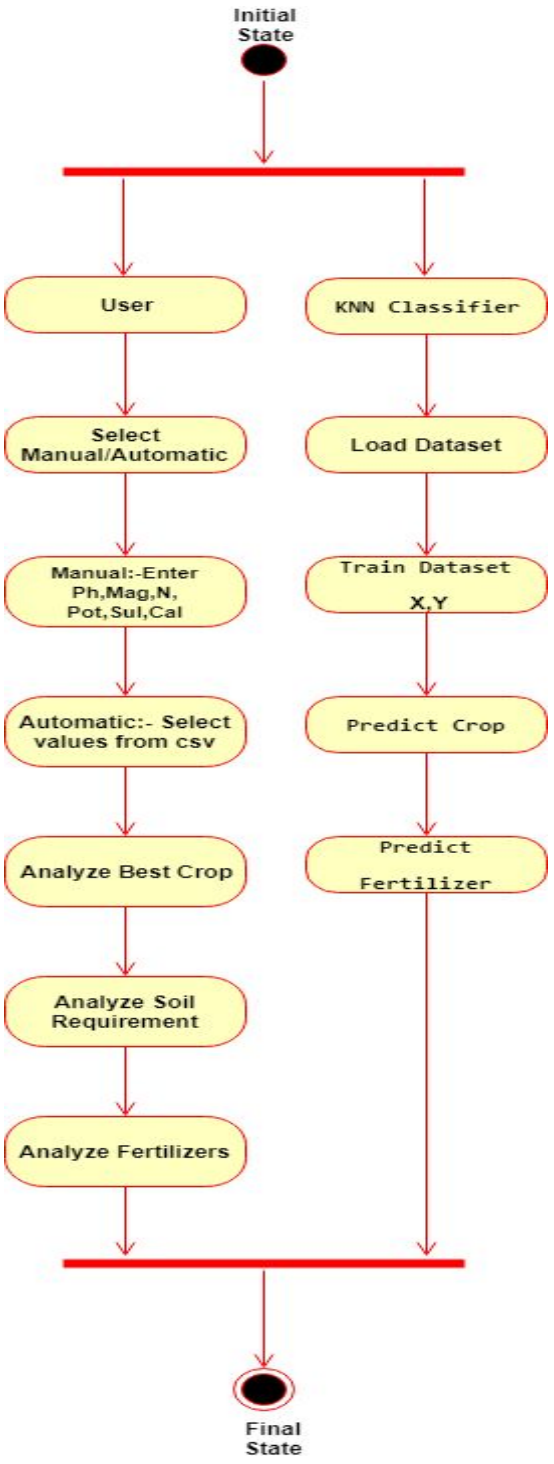


Figure 5: Activity Diagram

5. IMPLEMENTATION

5.1 INTRODUCTION TO MODULES :

- Get Soil values from User.
- Load the Dataset values.
- Match the Crop.
- Analyze the Fertilizer.
- Display the predicted Crop Information.

5.1.1 Get Soil values from User :

Take the inputs from the form.

5.1.2 Load the Dataset values:

Now, load the values from the Dataset.

```
import pandas as pd
import numpy as np
# import os
optimum = pd.read_excel("optimum2.xlsx", 'newData')
price = pd.read_excel("optimum2.xlsx", 'pricePerhr')
```

Figure 6 Load the Dataset values

```
from sklearn.neighbors import KNeighborsClassifier
clf = KNeighborsClassifier(n_neighbors=3)
clf.fit(X,y)
```

Figure 7 Fit into the K-NN

5.1.3 Match the Crop:

Now, predict the crop based on the nutrient values.

```
X = optimum.drop("CLASS",axis=1)
y = optimum.CLASS
clf = KNeighborsClassifier(n_neighbors=3)
clf.fit(X,y)
prediction1 = clf.predict(pred)
print(prediction1)
```

Figure 8 Match the Crop

If the predicted value is out of range, then it will predict using nearest value by applying K-NN.

Based on the predicted value, it will recommend the best suitable crop. It also predicts another two crops.

5.1.4 Analyse the fertilizer:

It will predict the fertilizer based on a suitable crop. It will recommend the best fertilizer to get a good outcome.

5.1.5 Display the predicted Crop Information:

Now, it will display the predicted crop along with the price, the yield and another two crops.

5.2 METHODS OF IMPLEMENTATION:

5.2.1 Output screens:

Start the server:

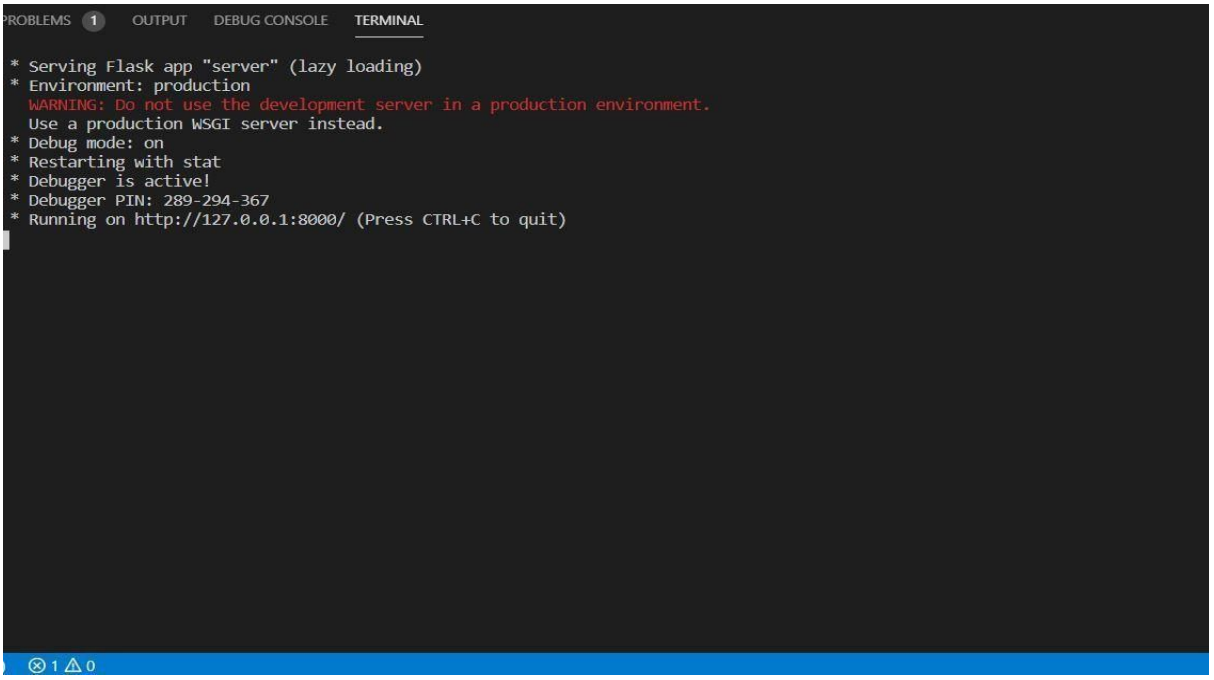


Figure 9 Start the server

MAIN PAGE:

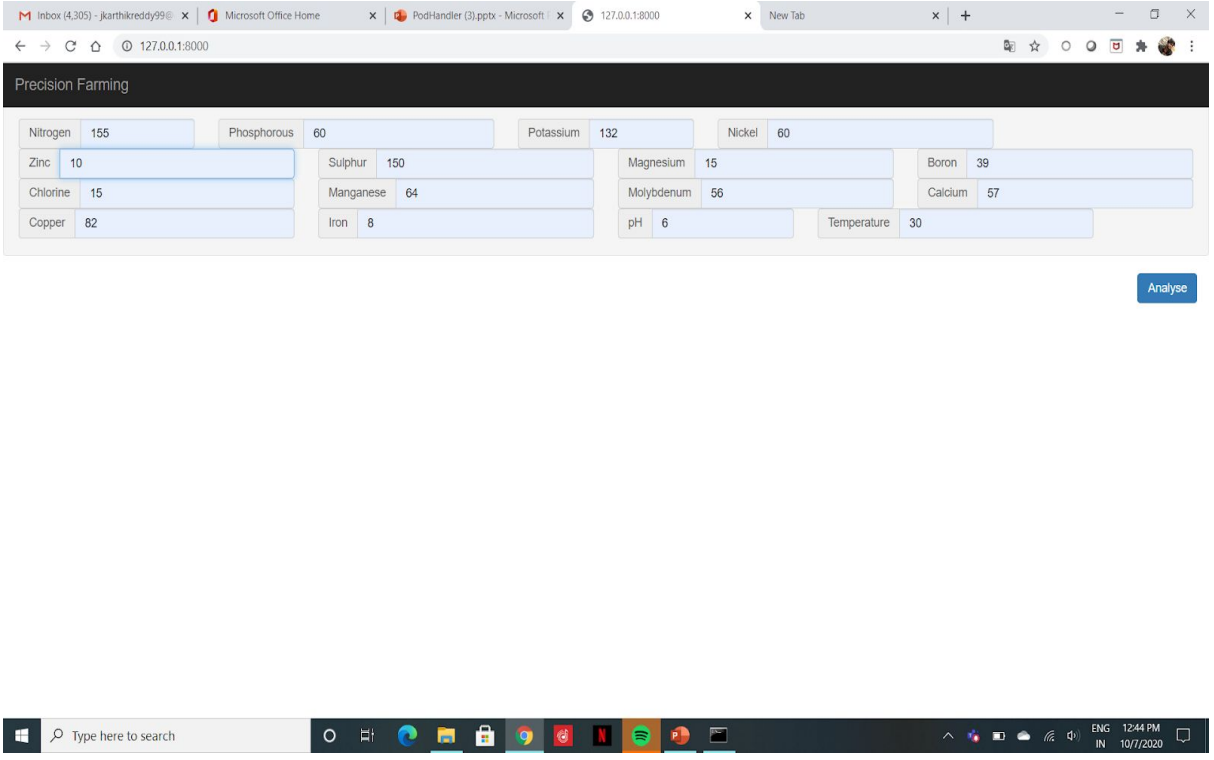


Figure 10 Main Page(Enter Required values for manual analysis)

5.2.2 Result Analylis:

View Predicted values:

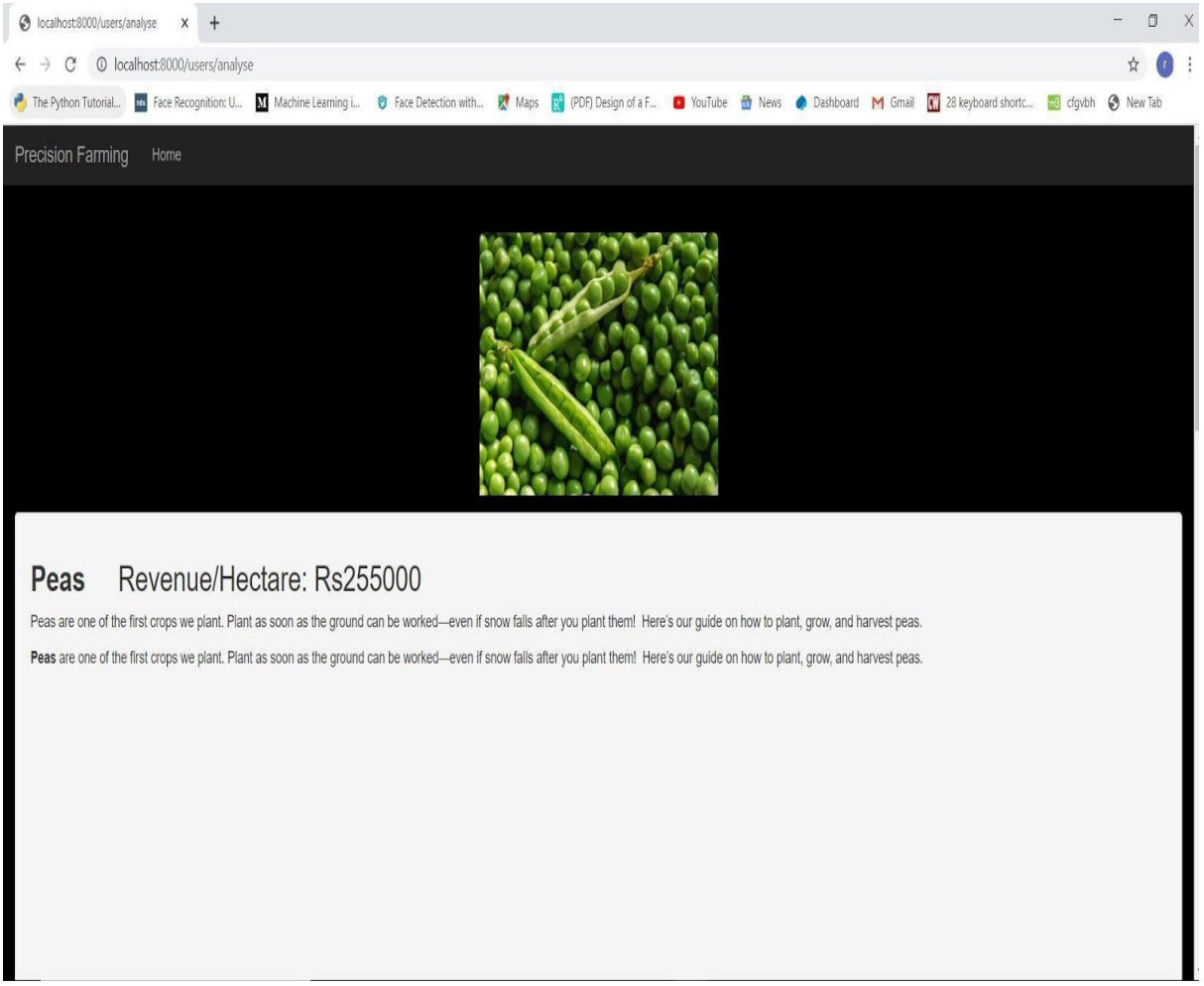


Figure 11 View Predicted values

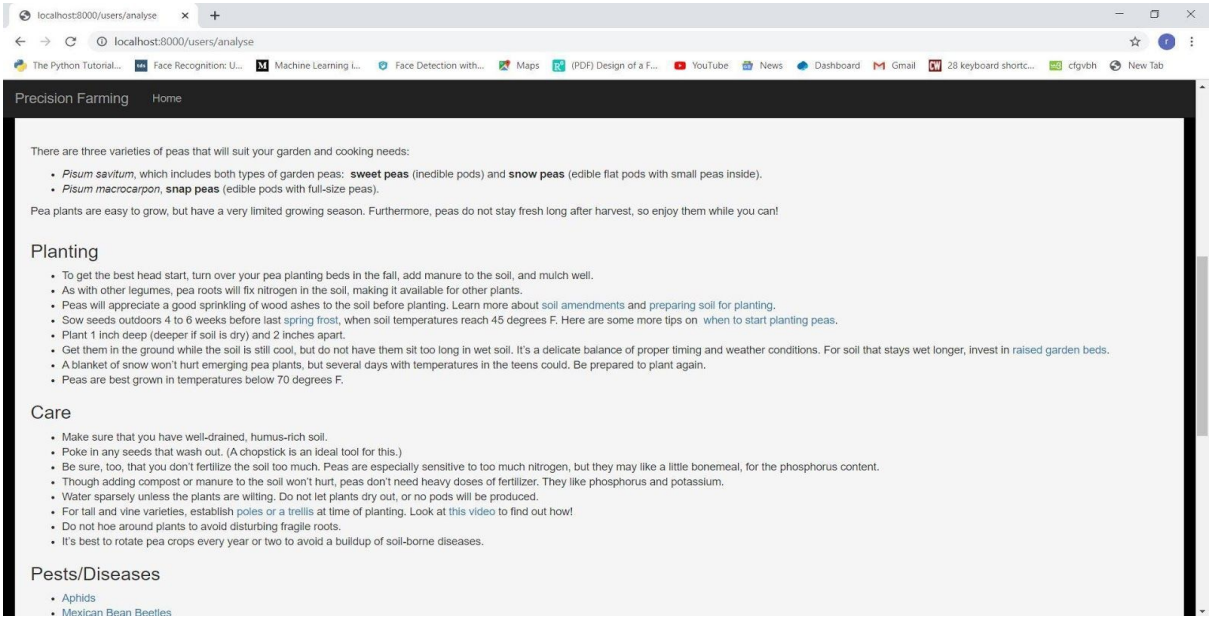


Figure 12 View Predicted values

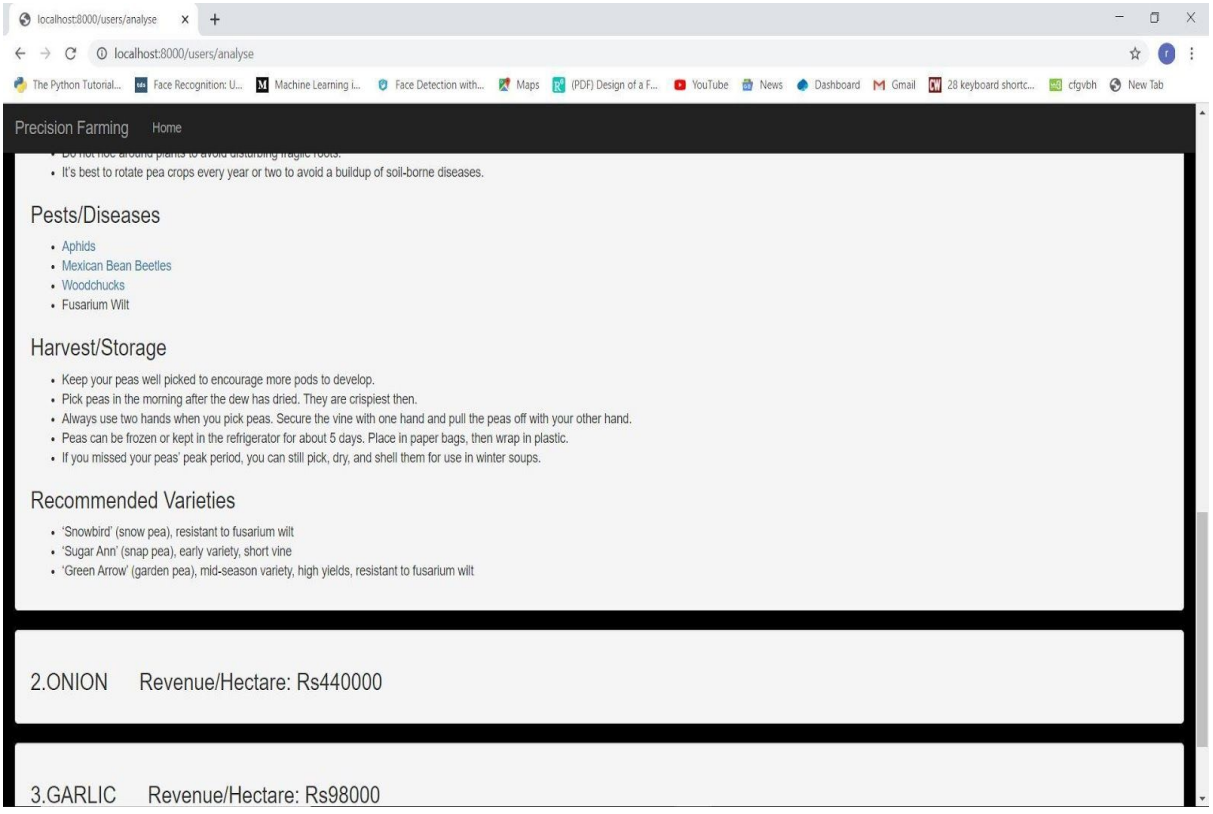


Figure 13 View Predicted values

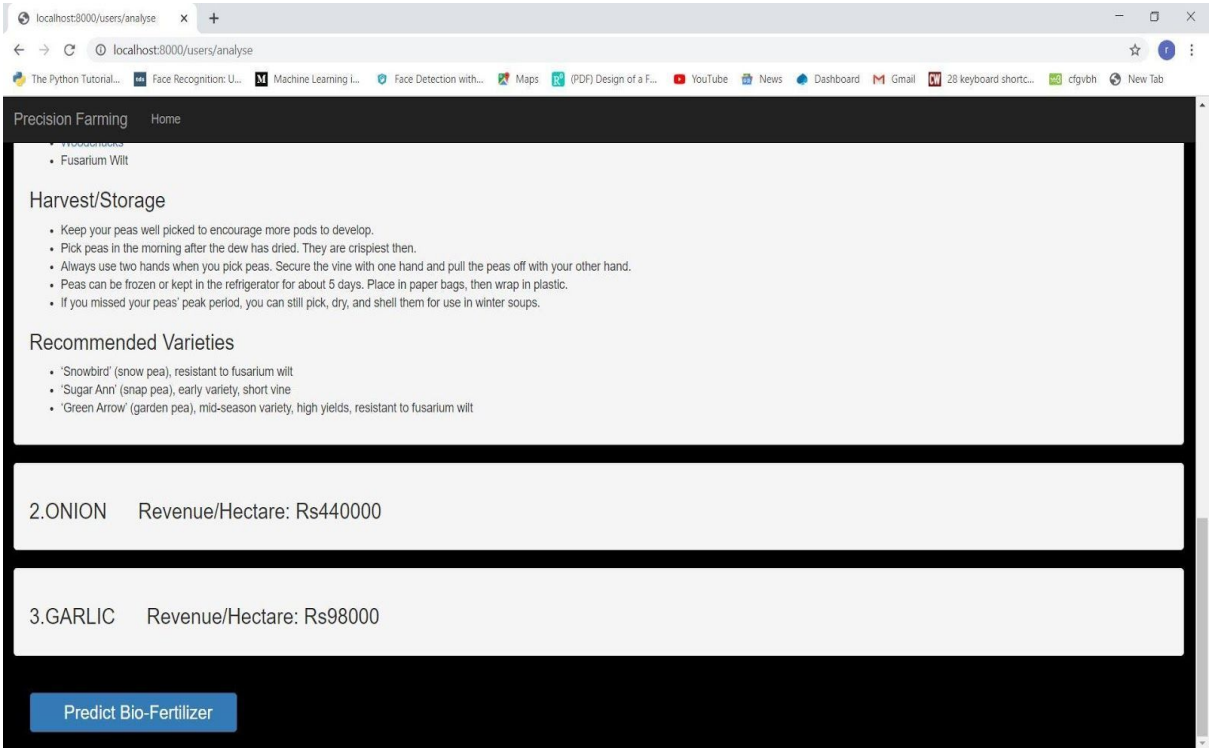


Figure 14 View Predicted values

FERTILIZER :

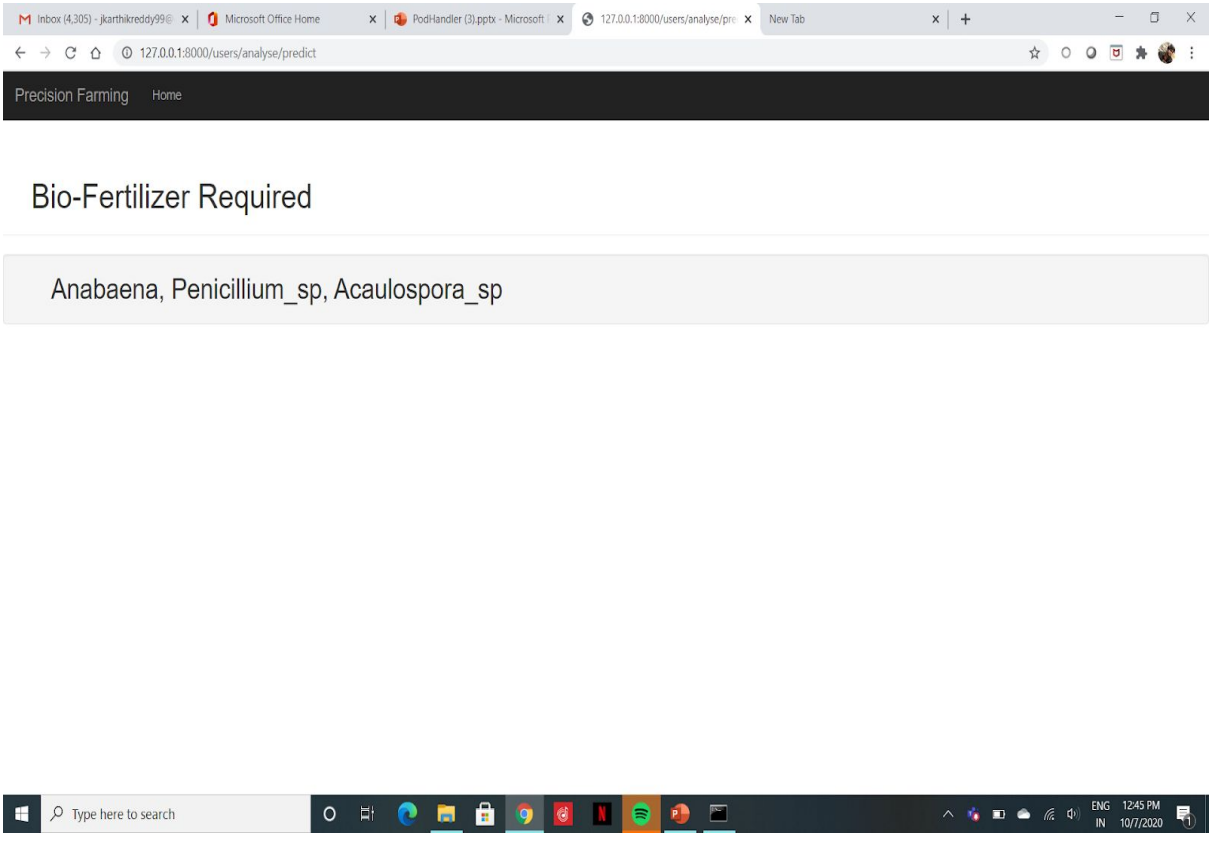


Figure 15 Fertilizer

6.TESTING AND RESULTS

The main motivation behind testing is to detect the flaws. Testing is the method toward attempting to find each possible flaw or drawback of a work item. It gives us an approach to check the effectiveness of parts, sub congregations, gatherings as well as the completed item. It is one of the ways toward practicing programming with the purpose of assurance that the

Programming framework lives up to its prerequisites and the client wishes and doesn't flop in a dissatisfying way. There are different kinds of testing. Each testing type tends to have a particular testing prerequisite.

6.1 Design of test cases and scenarios:

Test Case Id	Test Case Name	Test Case Desc.	Test Steps			Test Case Status	Test Priority
			Step	Expected	Actual		
01	Upload the tasks dataset	Verify either file is loaded or not	If dataset is not uploaded	It cannot display the file loaded message	File is loaded which displays task waiting time	High	High

02	Upload patients dataset	Verify either dataset loaded or not	If dataset is not uploaded	It cannot display dataset reading process completed	It can display dataset reading process completed	low	High
03	Preprocessing	Whether preprocessing on the dataset applied or not	If not applied	It cannot display the necessary data for further process	It can display the necessary data for further process	Medium	High
04	Prediction Random Forest	Whether Prediction algorithm applied on the data or not	If not applied	Random tree is not generated	Random tree is generated	High	High

05	Recommendation	Whether predicted data is displayed or not	If not displayed	It cannot view prediction containing patient data	It can view prediction containing patient data	High	High
06	Noisy Records Chart	Whether the graph is displayed or not	If graph is not displayed	It does not show the variations in between clean and noisy records	It shows the variations in between clean and noisy records	Low	Medium

6.2 Test Results and Validation: All the test cases mentioned above passed successfully. No defects encountered.

7.CONCLUSION

7.1 PROJECT CONCLUSION:

This project handover a machine learning framework for the mission of crop prediction, based on an affordable remote sensing data. It allows for real time prediction throughout the year and is suitable world-wide, especially for developing countries where field evaluations are hard to conduct. We are the first to use the latest representation learning ideas for crop prediction, and successfully learn and understand much more virtual characteristics from unprocessed data compared with the hand-crafted features that are popularly used. We present a dimensionality reduction approach which is based on histograms that successfully removes spatially correlated error, which might inspire other applications in remote sensing and figures livable. The model provides us with state-of-the-art prediction accuracy which has a great impact in sustainable agriculture and food security.

7.2 FUTURE ENHANCEMENT:

It is preposterous to expect to build a framework that makes all the prerequisites of the client. Client essentials continue to change as the framework is being made use of. A portion of the things that can make improvements that should be possible for this framework are:

- As innovation develops, it is believable to update the framework and can be adaptable to necessary conditions.
- Based on future security issues, it can be improved utilizing rising advances like single sign-on.

8. REFERENCES

1. DAAC, N. L. 2015. The modis land products. http://lpdaac.usgs.gov. Dodds, F., and Bartram, J. 2016. The Water, Food, Energy and Climate Nexus: Challenges and an Agenda for Action. Routledge.

2. Ermon, S.; Xue, Y.; Toth, R.; Dilkina, B.; Bernstein, R.; Damoulas, T.; Clark, P.; DeGloria, S.; Mude, A.; Barrett, C.; and Gomes, C. 2015. Learning large-scale dynamic discrete choice models of spatio-temporal preferences with application to migratory pastoralism in East Africa. In AAAI Conference on Artificial Intelligence.

3. FAO. 2018. The state of food insecurity in the world. Meeting the 2018 international hunger targets: Taking stock of uneven progress.