HINDUSTHAN INSTITUTE OF TECHNOLOGY,COIMBATORE

DEPARTMENT OF INFORMATION TECHNOLOGY

Project Based Experiential Learning Program

(Nalaiya Thiran)

FERTILIZER RECOMMENDATION SYSTEM FOR DISEASE PREDICTION

BATCH:2019-2023 **YEAR:**FINAL YEAR

SEM :VII

TEAM MENTOR : Mrs. ELAKKIYA.U

TEAM LEADER : Mr. ARAVIND.T (720819205006)

TEAM MEMBERS : Mr. DIPSIJO.M.P (720819205013)

Mr. PACHAIYAPPAN.P (720819205030)

Mr. VIJAY.A (720819205051)

TEAM ID :B4-4M6E

TABLE OF CONTENT

ABSTRACT:

The agriculture field contains so many challenging applications for yield prediction. In Tamil Nadu, farmers are mostly depend upon spinach production for their livelihoods and it is advantageous over the Nation also. Therefore, productiongrowth should be proper even with some uncertainties like weather disturbance, pests' attack, virus attacks, bacterial and fungal attacks etc Machine Learning is thedomain that uses past data for predicting. Machine Learning is the understandingof computer system under which the Machine Learning model learn from data and experience. The machine learning algorithm has two phases:1)Training2)Testing. To predict the disease from a patient's symptoms and from the history of the patient, machine learning technology is struggling from past decades.Detection and recognition of plant diseases using machine learning are very efficient in providing symptoms of identifying diseases at its earliest. Plant pathologists can analyze the digital images using digital image processing for diagnosis of plant diseases. Application of computer vision and image processing strategies simply assist farmers in all of the regions of agriculture. Generally, the plant diseases are caused by the abnormal physiological functionalities of plants. Therefore, the characteristic symptoms are generated based on the differentiation between normalphysiological functionalities and abnormal physiological functionalities of the plants. Mostly, the plant leaf diseases are caused by Pathogens which are positioned on the stems of the plants. These different symptoms and diseases of leaves are predicted by different methods in image processing. These different methods include different fundamental processes like segmentation, feature extraction and classification and so on. Mostly, the prediction and diagnosis of leafdiseases are depending on the segmentation such as segmenting the healthy tissues.

CHAPTER-1

INTRODUCTION

1.1 PROJECT OVERVIEW:

The Agriculture Sector remained resilient even after the pandemic in India. It effectively met rising global food demands while maintaining a continuous supply chain of vital food goods across the country. India's agriculture sector employs a large number of people and is second after China in terms of producing fruits and vegetables. Traditional farming methods, on the other hand, are ineffective. It fails to make proper use of all available resources. Because the primary focus is on production, traditional methods frequently result in soil nutrient depletion and weariness. By producing only certain crops, the earth is depleted. The ideal pH range for plant development varies depending on the crop. Most plants thrive in soil pH ranges of 6.0-7.5, as this is where the majority of nutrients are accessible. By combining a soil sample with water and measuring the resultant aqueous solution, the pH of the soil may be calculated. pH 7.0 represents neutral, Acidic is below 7.0, and basic or alkaline is above 7.0. The availability of nutrients for plant development is influenced by the pH of the soil. Aluminum and manganese can become more accessible and hazardous to plants in very acidic soil, whereas calcium, phosphorus, and magnesium are less available. Phosphorus and other micronutrients are present in small amounts in more alkaline soil. It's a good ideato evaluate the soil pH before planting a new crop since various plants flourish in different pH ranges. The pH of the soil might indicate if it is acceptable for the plants to be cultivated or whether it has to be altered in order to achieve optimal plant development. When paired with other criteria, pH can aid in the recommendation of fertilizers and the cultivation of the appropriate soil type for

the region. For the correct sort of soil growth, the right proportion of nitrogen, potassium, and phosphorus is crucial.

1.2.PURPOSE

In India, The Agriculture industry is extremely vital and crucial for economic and social development and jobs. In India, the agricultural sector provides a living for almost 48% of the population. As per the 2019-2020 economic survey, an Indian farmer's median wage in 16 states is Rupees 2500. Most of the Indian population depends on agriculture for their livelihood. Agriculture gives an opportunity of employment to the village people to develop a country like India on large scale and give a push in the economic sector. The majority of farmers face the problem of planting an inappropriate crop for their land based on a conventional or non-scientific approach. This is a challenging task for a country like India, where agriculture feeds approximately 42% of the population. And the outcomes for the farmer of choosing the wrong crop for land is moving towards metro city for livelihoods, suicide, quitting the agriculture and give land on lease To industrialist or use for the non-agriculture purpose. The outcome of wrong crop selection is less yield and less profit

CHAPTER-2

LITERATURE SURVEY

1. Agricultural crop recommendation systems are available in the market which consider various parameters like weather at the time the crop is to be planted, soil type, topography of the region, temperature and rainfall in the region, market prices of the crop duration, etc. Research has been carried out in this field and the following papers have been referred for the purpose of research and study. Prof. Rakesh Shirsath and other co-author in paper [1] proposed a system which helps the users to make decisions regarding the crop to be planted. The system used is a subscription-based system which would have personalized information of every farmer registered. The system includes a module which maintains the information of the previous crops planted collected from various sources and shows a matching crop that can be planted. The whole process is done with the help of artificial neutral networks. At the end a feedback system is provided so thatthe developer can make changes required if the farmer finds some difficulty while using the system.

2. Plant Disease Detection and Classification using CNN Model with Optimized Activation Function S. Yegneshwar Yadhav; T. Senthilkumar; S. Jayanthy; J. Judeson Antony Kovilpillai This research discusses the application of Convolutional Neural Networks (CNN) algorithms for the optimum real-time detection of diseases that impact the plant and the afflicted area, so that proper fertilizers can be employed to prevent additional harm to plants from pathogenic viruses. The activation function is at the heart of the CNN model since it combines non-linearity to create a true artificial intelligence system for classification. ReLu is one of the best activation functions, however it has the problem that its

derivative is 0 for negative values, resulting in neuronal necrosis.To increase the accuracy and performance of the system using a TensorFlow framework, a new mathematical activation function is constructed and compared with current activation functions.Experiment findings on trained databases demonstrate that the created activation function increased CNN model accuracy and performance by 95%.The suggested optimizer improves the training speed of the CNN model by 83 percent when implemented in an ARM processor. A further area impacted by illness is computed using the K - means clustering approach for fertilizer optimization.

2.1. EXISTING SYSTEM:

- Deep learning algorithms were used in this study to develop a novel way for automatically categorizing and detecting plant illnesses using leaf pictures. The developed computer could detect the presence of leaves and distinguish between healthy leaves and 13 abnormalities that could be seen visually. In the end, the trained model's overall accuracy was 96.3 percent. Because the suggested approach had not been applied in the field of plant disease detection, there was no comparison with similar findings obtained using the exact process. This study will be expanded to include the development of a whole system composed of server-side components including a trained model and an application for smart mobile devices capable of identifying diseases captured by a smartphone camera. The authors anticipate that by expanding this research, they will have a substantial impact on sustainable development, influencing crop quality for future generations.

- Plant Leaf Disease Detection and Classification Based on CNN with LVQ Algorithm - Melike Sardogan, Adem Tuncer, Yunus Ozen. Early disease detection is critical in agriculture for efficient crop yield. The diseases bacterial spot, late blight, septoria leaf spot, and yellow curved leaf have an impact on tomato crop quality. Automatic plant disease classification methods also aid in taking action after detecting symptoms of leaf diseases. This paper describes a Convolutional Neural Network (CNN) model and Learning Vector Quantization (LVQ) algorithm-based method for detecting and classifying tomato leaf disease. The dataset contains 500 images of tomato leaves with four disease symptoms. We created a CNN model for automated feature extraction and classification. Color information is being extensively employed in plant leaf disease research .The filters in our model are applied to three channels depending on RGB components. For training the network, the LVQ was fed the output feature vector of the convolution component. The experimental findings show that the proposed approach accurately detects four forms of crop leaf diseases.

2.2.REFERENCE:

[1] Reyes Angie .K, Juan C. Caicedo, and Jorge E. Camargo, "Fine-tuning Deep Convolutional Networks for Plant Recognition", In CLEF (Working Notes), 2015.

[2] Hamrouni .L, Aiadi .O, Khaldi .B and Kherfi .M.L, "Plants Species Identification using Computer Vision Techniques", Revue des Bioressources 7, no. 1, 2018.

[3] Dimitrovski, Ivica, GjorgjiMadjarov, DragiKocev, and PetreLameski, "Maestra at LifeCLEF 2014 Plant Task: Plant Identification using Visual Data", In CLEF (Working Notes), pp. 705-714, 2014.

[4] Naresh, Y. G., and H. S. Nagendraswamy, "Classification of medicinal plants: an approach using modified LBP with symbolic representation", Neurocomputing 173, pp: 1789-1797, 2016.

[5] Sue Han, CheeSeng Chan, Paul Wilkin, and Paolo Remagnino, "Deep-plant: Plant identification with convolutional neural networks", In Image Processing (ICIP), 2015 IEEE International Conference on, pp. 452-456, IEEE, 2015.

[6] Niketa G, Leisa A.J., Owaiz Petkar, and Amiya K.T.(2016). Rice Crop Yield Prediction in India using Support Vector Machines 978-1-5090-2033- 1/16/$31.00 ©2016 IEEE.

[7] Website DAVIS, L. E..25 1943. MEASUREMENTS OF pH WITH THE GLASS ELECTRODE AS AFFECTED BY SOIL MOISTURE Soll Sel. 56: 405-422, Illus.

[8] James. N. Mugo, Nancy N. Karanja, Charles K. Gachene, Klaus Dittert, Shadrack O. Nyawade,and Elmar Schulte-Geldermann - Assessment of soil fertility and potato crop nutrient status in central and eastern highlands of Kenya, 8th May, 2020. https://www.ncbi.nlm.nih.gov/pmc/articles/PM C7210878/.

2.3.PROBLEM STATEMENT DEFINITION:

The solution to the problem is Machine learning, which is one of the applications of Artificial Intelligence, is being used to implement the proposed system. Crop recommendation is going to recommend you the best crop you can grow in your land as per the soil nutrition value and along with as per the climate in that region. And recommending the best fertilizer for every particular crop is also a challengingtask. And the other and most important issue is when a plant gets caught by heterogeneous diseases that effect on less amount of agriculture production and

compromises with quality as well. To overcome all these issues this recommendation has been proposed . Nowadays a lot of research and work is being implemented in the smart and modern agriculture domain. Crop recommendation is characterized by a soil database comprised of Nitrogen, Phosphorus, potassium. The ensembles technique is used to build a recommendation model that combines the prediction of multiple machine learning. Models to recommend the right crop based on soil value and the best fertilizer to use.

CHAPTER-3

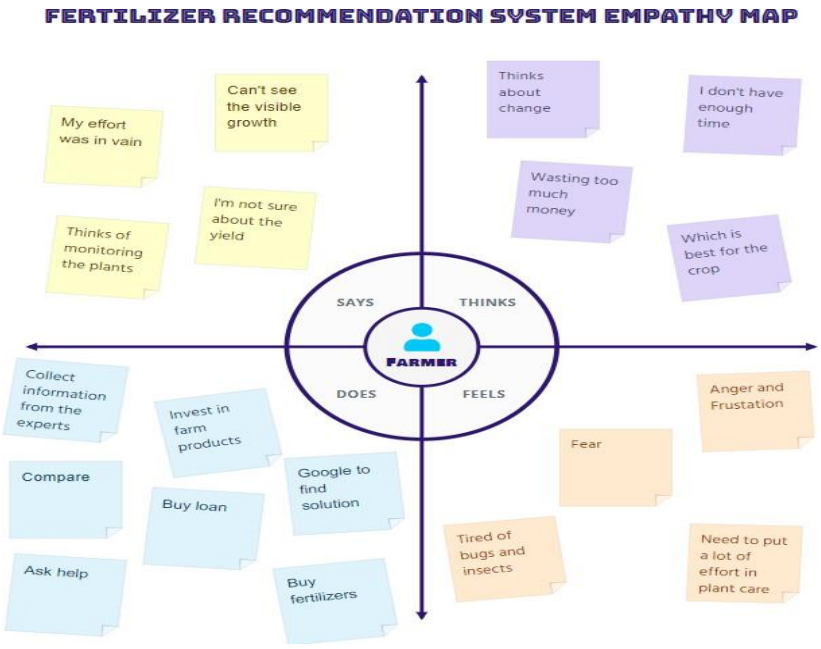
IDEATION & PROPOSED SOLUTION

3.1. EMPATHY MAP CANVAS:

In this diagram, we know about the description of the farmers' thinking.

- ❖ Says.
- ❖ Thinks.
- ❖ Does.
- ❖ Feels.

Nowadays a lot of research and work is being implemented in the smart and modern agriculture domain. Crop recommendation is characterized by a soil database comprised of Nitrogen, Phosphorus, potassium. The ensembles technique is used to build a recommendation model that combines the prediction of multiple machine learning. Models to recommend the right crop based on soil value and the best fertilizer to use.



3.2. IDEATION AND BRAINSTORMING:

Modern Technology is increasing and optimizing the Performance of the Artificial Intelligences (AI) Model. In the Based Crop Yield Disease Prediction System, it will be helpful for farmers to prevent the crop from the various diseases which can identify the Disease within a process of capturing the Image at the plant and Machine Learning Algorithm will give affected Disease Name. In this Project Milestone will be given the Best Solution for the farmer using the complete friendly and simple user interface web application to fetch the solution by own. In addition, we are planning to add a valid Module that is a Fertilizer recommendation for the Specific Disease. It can give both Artificial fertilizer and Natural Fertilizer in a suggested manner.

Activity List:

In Project Management Planning is an Important task to scheduling the phrase of the project to the Team Member. In this Activity can shows the various activity are allocated and Done by the Team Members! In Project we can Split into the Four Step of Phrases are :

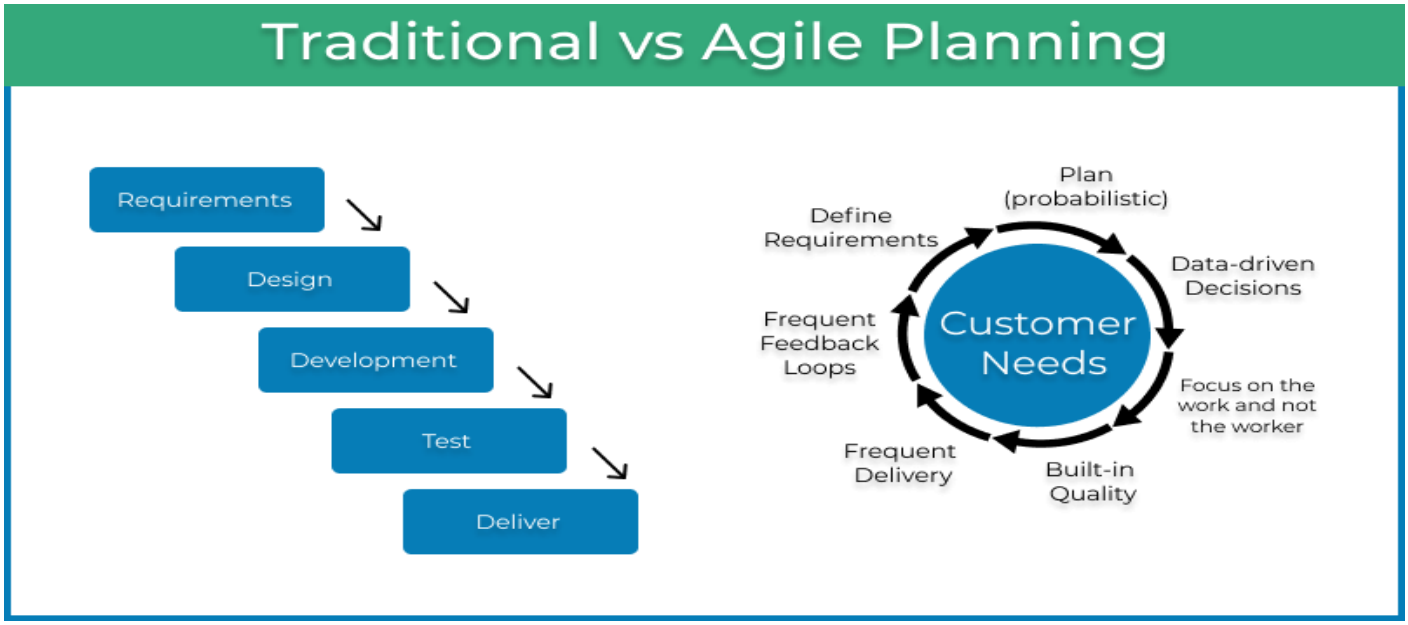
Phrase 1: Information Collection and Requirement Analysis

Phrase 2: Project Planning and Developing Modules

Phrase 3: Implementing the High Accuracy Deep Learning Algorithm to Perform

Phrase 4: Deploying the Model on Cloud and Testing the Model and UI

Performance Agile Methodology for Activity Planning.



Agile Methodology for Activity Planning.

3.3. PROPOSED SOLUTION:

The aim of the proposed system is to help farmers to cultivate crops for better yield. The dataset of crop yield is collected from the last 5 years from different sources. There are 3 steps in proposed work.

Crop Yield Prediction: Crop Yield Prediction can be done using crop yield data, nutrients and location data. These inputs are passed to Random Forest and Support Vector Machine algorithms. These algorithms will predict crop based on present inputs.

Fertilizer Recommendation: Fertilizer Recommendation can be done using fertilizer data, crop and location data. In this part suitable crops and required fertilizer for each crop is recommended. SENSOR for pH the hydrogen-ion in the soil is measured using a PH meter, an electric instrument. A voltage test is used by apH meter to detect hydrogen ion concentration and consequently pH. It is used to determine the soil's acidity or alkalinity. A solution with more H+ ions will stay acidic, whereas a solution with OH- ions would remain alkaline. The pH of 1 soil is exceedingly acidic, whereas the pH of 14 soil is excessively alkaline. Pure water, being a neutral solution, has a pH of 7. The pH of various soil samples is determinedusing a pH meter. It is more precise than using pH strips.

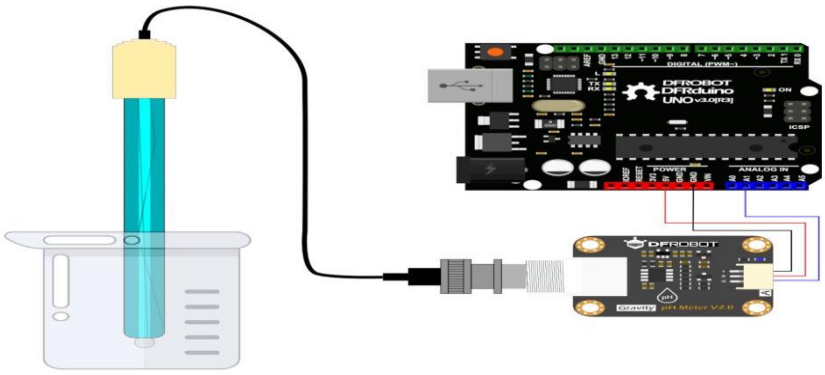


Fig. pH probe & Sensor

PROPOSED MODEL FRAME:

The proposed system is used to determine the nutrient quantity of soil through NPK Ratio and predict various diseases crops may be infected with. As we know all the nutrients present in the soil but what amount of nutrients are present in the particular field. Every soil has different micronutrients. But to measure the amount

of nutrient available in the soil we are going to design a device which will give proper reading of the micronutrient and that can be used to predict crops, fertilizers and crop diseases. Following are the main objectives of the proposed systemDesign and develop a microcontroller-based sensor interfacing for reading soil parameters (NPK value). Converting the sensor value which is analog signal todigital signal for further processing Sending all reading to the system using USB ports available. Developing a website application for displaying the result and generating the report. Based on the result give suggestions to improve the quality of soil.

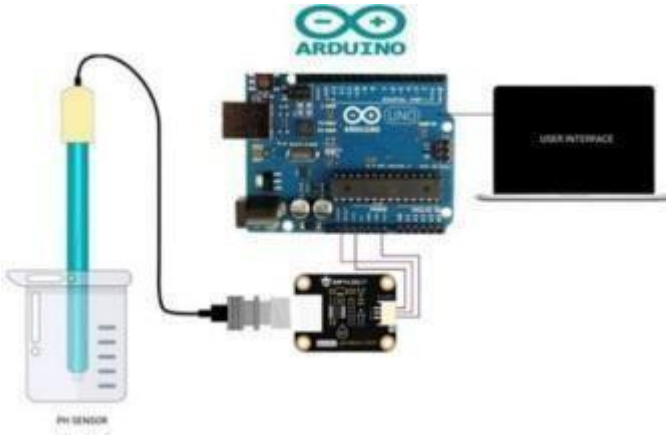
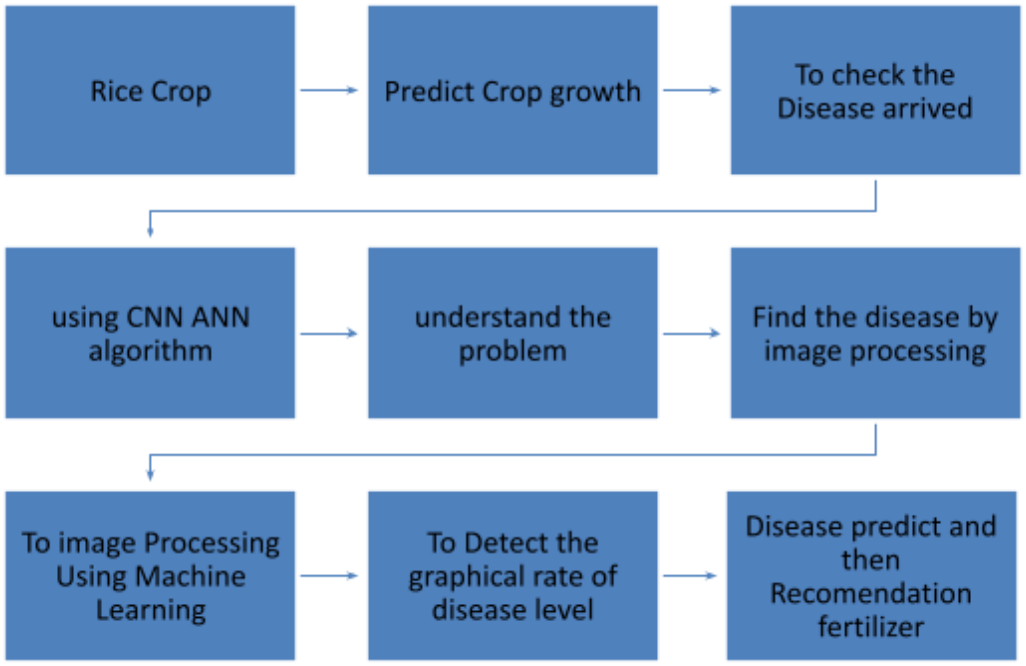


Fig: System design

3.4. PROBLEM SOLUTION FIT:

The solution to the problem is Machine learning, which is one of the applications of Artificial Intelligence, is being used to implement the proposed system. Crop recommendation is going to recommend you the best crop you can grow in your land as per the soil nutrition value and along with as per the climate in that region. And recommending the best fertilizer for every particular crop is also a challengingtask. And the other and most important issue is when a plant gets caught by heterogeneous diseases that effect on less amount of agriculture production and

compromises with quality as well. To overcome all these issues this recommendation has been proposed . Nowadays a lot of research and work is being implemented in the smart and modern agriculture domain. Crop recommendation is characterized by a soil database comprised of Nitrogen, Phosphorus, potassium. The ensembles technique is used to build a recommendation model that combines the prediction of multiple machine learning. Models to recommend the right crop based on soil value and the best fertilizer to use.



CHAPTER-4

REQUIREMENT ANALYSIS

4.1. FUNCTION REQUIREMENTS:

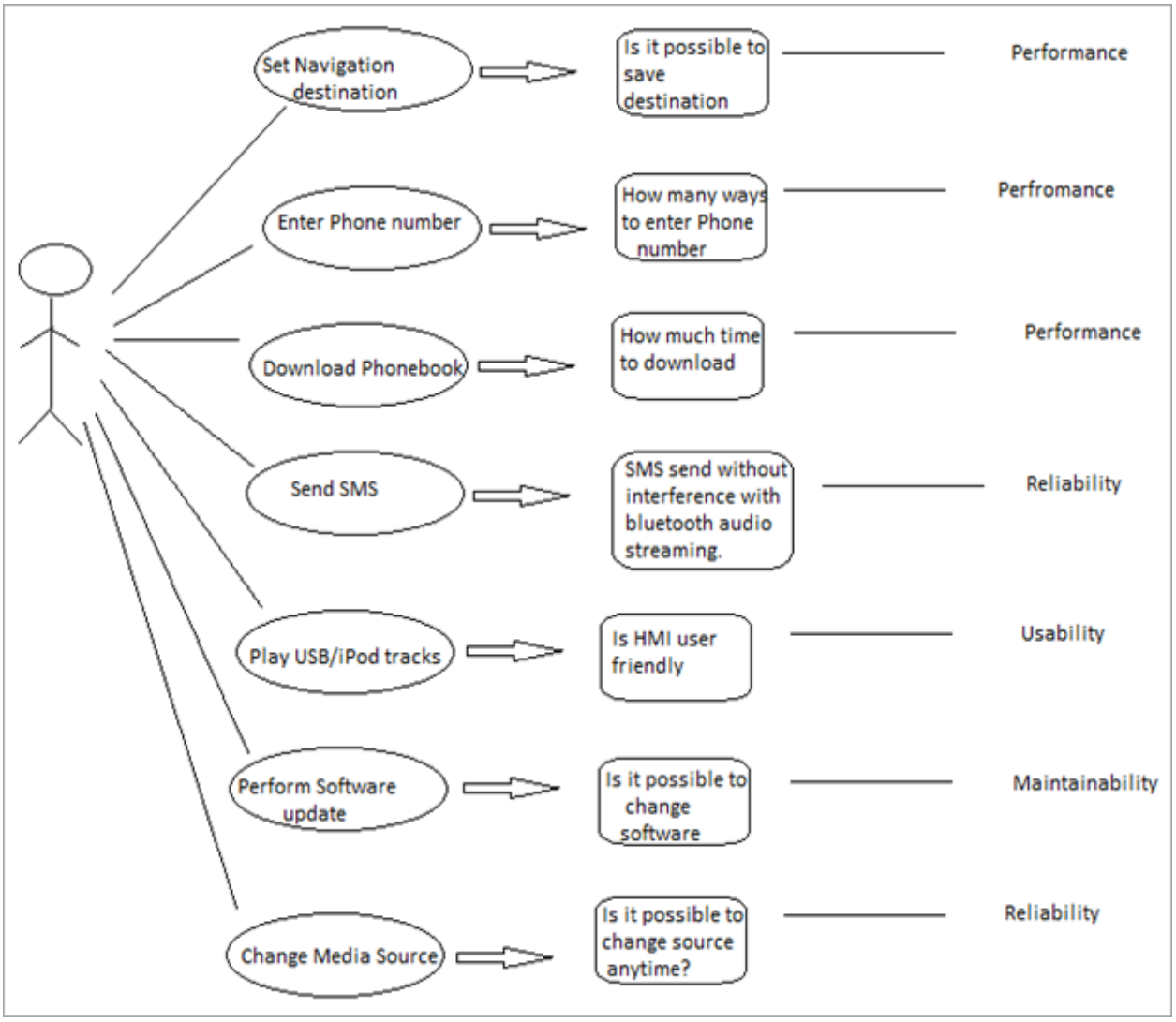
FR.NO	FUNCTIONAL REQUIREMENT(E-pic)	SUb-REQUIREMENT (story/sub-task)
FR-1	User Registration	Registration through farm.
FR-2	User conformation	Confirmation via Email.
FR-3	User Profile	Filing the profile page after logging in.
FR-4	Upload DataSet	Image of the leaves are to be uploaded.
FR-5	Requesting Solution	Uploaded images is compared with thepre-defined model and solution is generated.
FR-6	Downloading Solution	The solution in PDF format which contains the recommendation of fertilizer and the possible diseases.

MATERIAL AND METHODS:

A digital camera or similar devices are used to take images of different types, and then those are used to identify the affected area in leaves. Then different types of image-processing techniques are applied to them, the process those images, to get different and useful features needed for the purpose of analyzing later-Plant leaf disease identification is especially needed to predict both the quality and quantity of the First segmentation step primarily based on a mild polygonal leaf model is first achieved and later used to guide the evolution of an energetic contour. Combining global shape descriptors given by the polygonal model with local curvature based features, the leaves are then classified overleaf datasets. In this

research work introduce a method designed to deal with the obstacles raised by such complex images, for simple and plant leaves. A first segmentation step based on graph-cut approach is first performed and later used to guide the evolution of leaf boundaries, and implement classification algorithm to classify the diseases and recommend.

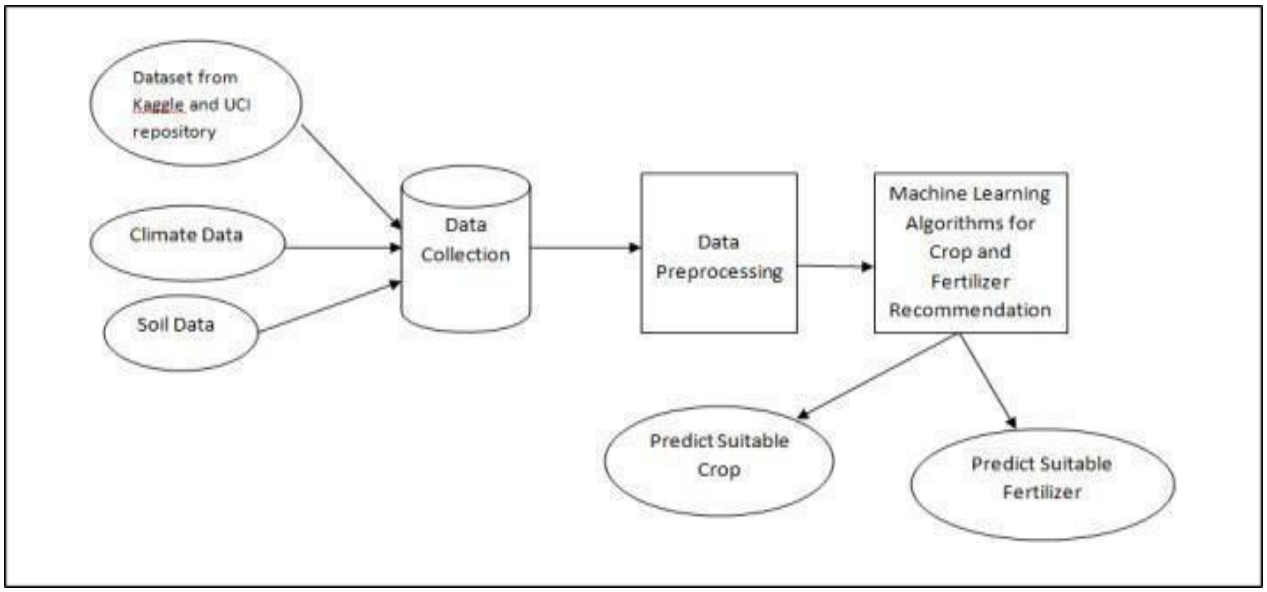
4.2. NON- FUNCTION REQUIREMENT:



CHAPTER-5

PROJECT DESIGN

5.1. DATA FLOW DIAGRAM:



Data Collection:

DatasetLink:https://www.kaggle.com/code/emmarex/plant-disease-detection-usin g-keras

Dataset Splitting:

The historical weather data obtained were divided into two sets, namely the training set and training set. Initially, the artificial intelligence (AI) model is built on a dataset called a training set and the built model is tested on a new set called the test set. The newly developed ML model is applied to test the dataset to measure the performance. This dataset was further split in two different ways. As per the bifurcation of the dataset, if 70–30% split is considered then the data valuesfor the year field that are less than equal to 2010 and greater than equal to 1989 will be considered for training. The data values for the year field that are greater than 2010 and equal to 2019 will be considered for testing. If 80–20% split is considered, then the data values for the year field that are less than equal to 2013 and greater than equal to 1989 will be considered for training and the data values for the year field that are greater than 2013 and equal to 2019 will be considered for testing. Table 1, shows these two variations in dataset division.

Table 1

Two cases of data split considered in the manuscript for implementation, summary of activation functions, MAE values for two cases of data split for four activation functions, comparison based on overall accuracy of the model for different activation functions.

% of train–test dataset split	Total weather instances	Training instances	Testing instances
70–30%	1,643	1,150	493
80–20%	1,643	1,313	329

5.2. SOLUTION AND TECHNICAL ARCHITECTURE:

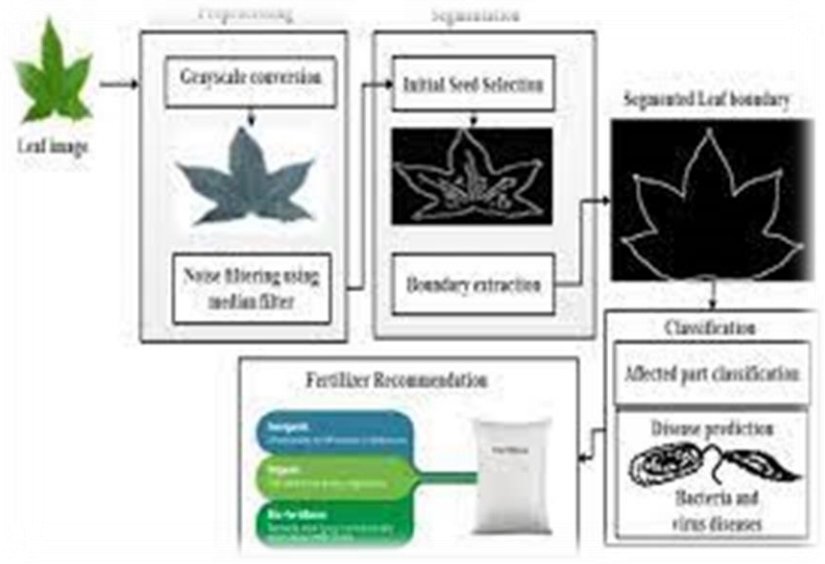


Fig.solution architecture

Technical Architecture:

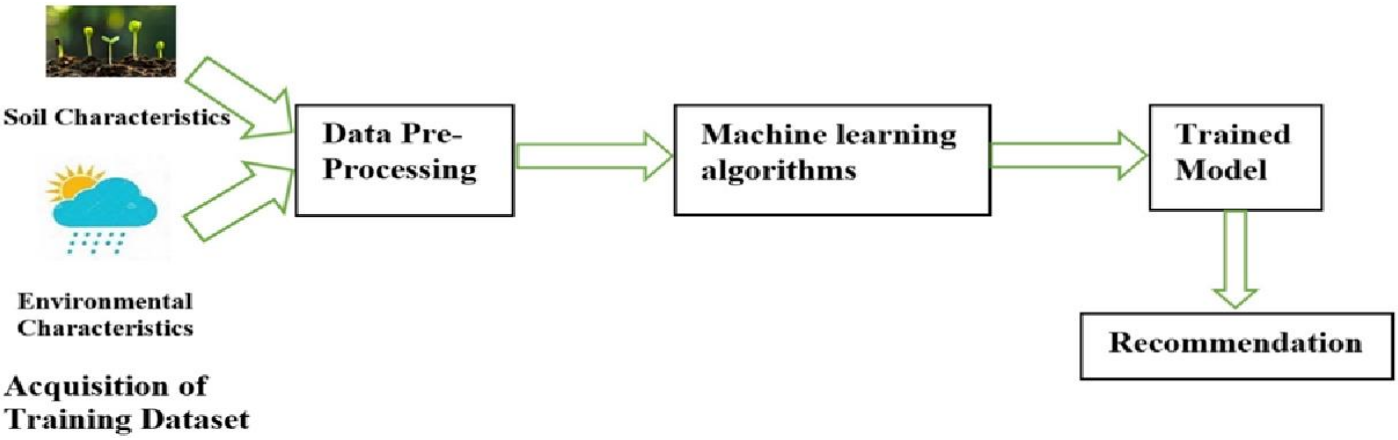


Fig:Technical architecture

Technology Architecture:

Components & Technologies

S.No	Component	Description	Technology
1	User Interface	user interacts with the application .To depict the human computer interaction and communication.	HTML, CSS,JSP
2	Application Logic 1	A page to upload images as input	Python

5.3. USER STORIES:

- Farmer
- Common People
- Seller
- Buyer
- Employees
- Industrial People

Value for society :

Consumer Farming is one of the major sectors that influences a country's economic growth. In country like India, majority of the population is dependent on

agriculture for their livelihood. Many new technologies, such as Machine Learning and Deep Learning, are being implemented into agriculture so that it is easier for farmers to grow and maximize their yield

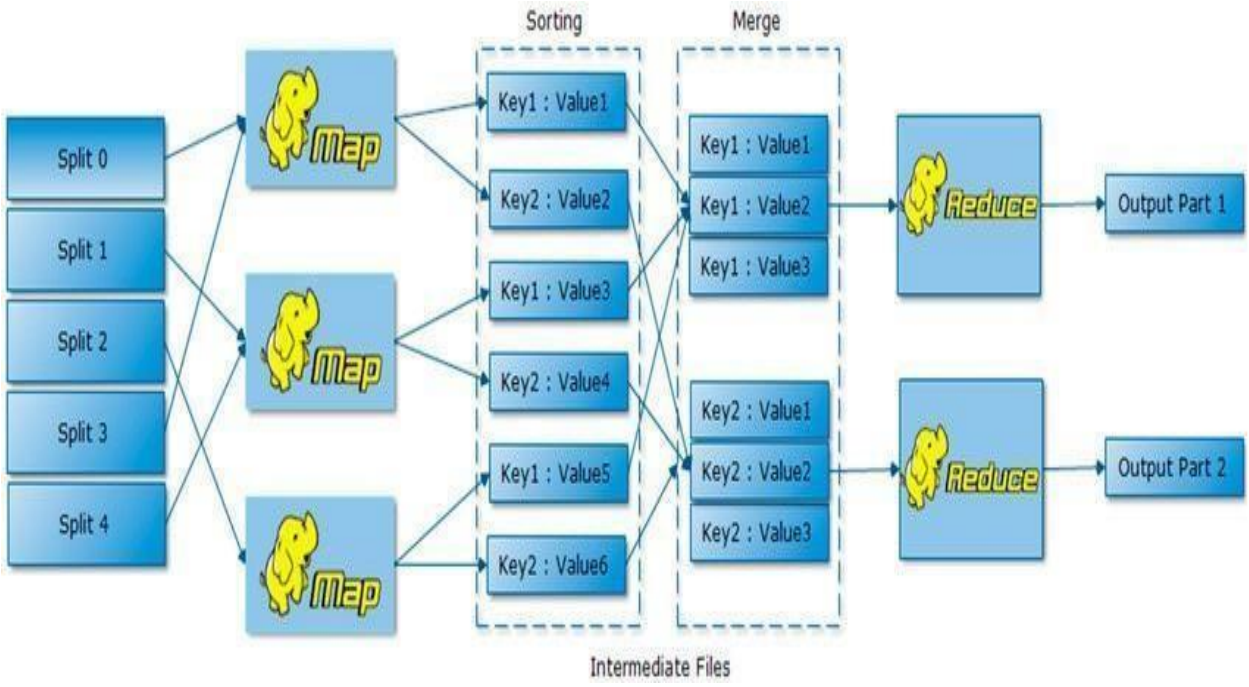
Value for Environment :

- In the crop recommendation application, the user can provide the soil data from their side and the application will predict which crop should the user grow.
- For the fertilizer recommendation application, the user can input the soil data and the type of crop they are growing, and the application will predict what the soil lacks or has excess of and will recommend improvements.
- For the last application, that is the plant disease prediction application, the user can input an image of a diseased plant leaf, and the application will predict what disease it is and will also give a little background about the disease and suggestions to cure it.
- These all are to improve Agriculture, which slightly reduces poverty, climatic conditions, soil erosion etc..

CHAPTER-6

PROJECT PLANNING & SCHEDULING

6.1. SPRINT PLANNING & ESTIMATION:

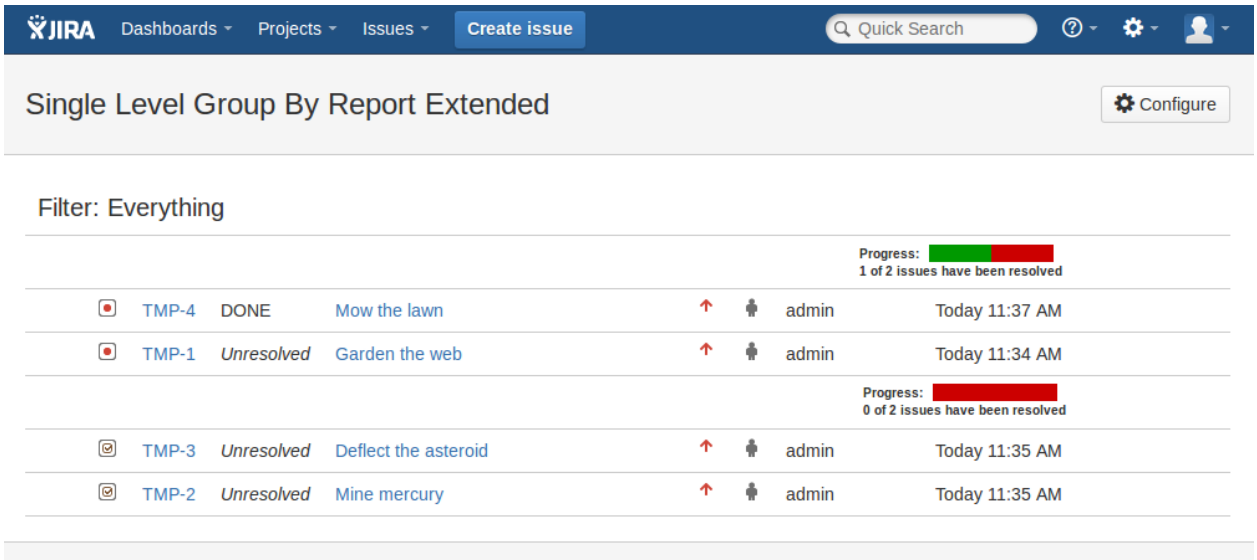


At present, due to the unavailability of natural resources, society should take the maximum advantage of data, information, and knowledge to achieve sustainability goals. In today's world condition, the existence of humans is not possible without the essential proliferation of plants. In the photosynthesis procedure, plants use solar energy

6.2. SPRINT DELIVERY SCHEDULE:

Crops	Actual	RF	SVM	RF_Err	SVM_Err	RF_Ac	SVM_Ac
Rice	119	120	119	0.41841	0	99.5815	100
Jowar	125	121	125	1.62602	0	98.3739	100
Wheat	126	124	126	0.8	0	99.2	100
Soybeans	110	111	110	0.45249	0	99.5475	100
Sunflower	125	114	125	4.60251	0	95.3974	100
Cotton	146	146	146	0	0	100	100
Sugarcane	98	102	98	2.0	0	98	100
Tobacco	99	119	110	9.17431	5.26315	90.825	94.7368
Onion	147	130	147	6.13718	0	93.862	100
Dry chilli	186	186	186	0	0	100	100
Avg Accuracy						**97.48**	**99.47**

6.3. Report for JIRA:



CHAPTER-7

CODING & SOLUTIONING

7.1. Feature 1:

In this system the micro- controller device is connected to the system through USB ports available and transmits the data from the device to the system. The result is generated from the received data and suggestions are given.

CROP RECOMMENDATION:

The NPK value of the soil is calculated using the pH value supplied from the instrument. An API is used to collect temperature and humidity. The pH value, NPK value, temperature, humidity, and rainfall are the characteristics that are used to forecast the best crop to grow in a given place. The crop is predicted using a machine learning model called XGBoost, which has a 99 percent accuracy rate.

FERTILIZER RECOMMENDATION:

Based on the NPK value acquired from the device for a certain soil, a suitable fertilizer is advised for the crop. Proper recommendations for increasing soil fertility are presented (NPK).

DISEASE DETECTION:

The visual data collected from the user is used to detect crop-based illnesses. Deep learning techniques and CNN models are used to forecast if the crop is affected which disease, and a viable remedy is then offered to the user.

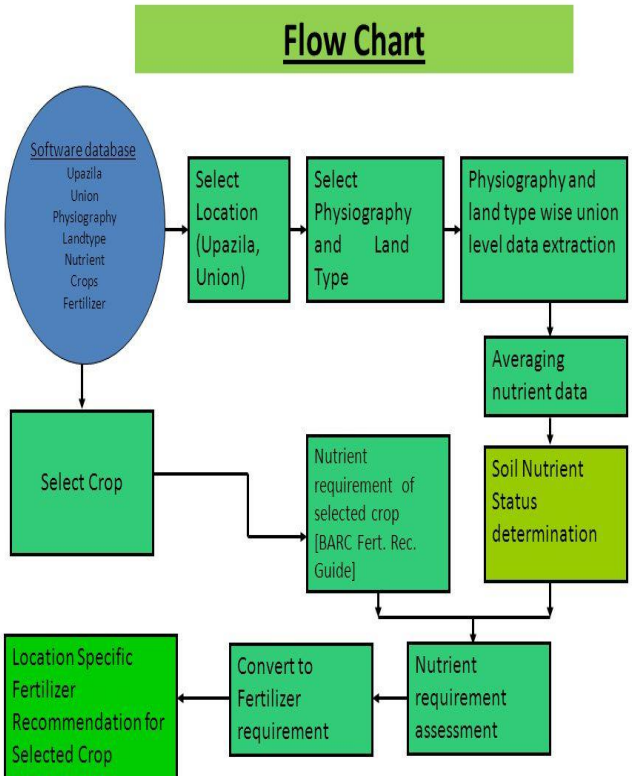


Fig: Flowchart of System

7.2. Feature 2:

The proposed model is a combination of regression and classification. ANN model is used for implementation. Following are the algorithmic steps of the model.

Step 1: Collect week-wise data of agro-meteorological parameters for disease forecasting.

Step 2: Load data by using the pandas library

Step 3: Perform data pre-processing on the instances of the dataset.

Step 4: Build a model based on ANN using Keras.

Step 5: Divide the dataset into two parts: training dataset and test dataset. The dataset is split into two cases (70–30%, 80–30%) and results are verified accordingly.

Step 6: Training the network by utilizing the dataset.

Step 7: Prediction of future values of climatic parameters.

Step 8: Evaluation of the prediction model

Step 9: Prediction of crop disease occurrence. The final output will be in the form of five classes namely Healthy, Rice Blast, Blight, Brown Spot, and False Smut.

Step 10: Evaluation of the classification model.

Class 1—Healthy, Class 2—Rice Blast, Class 3—Bacterial Blight, Class 4—Brown Spot, Class 5—False Smut.

to determine the nutrient quantity of soil through NPK Ratio and predict various diseases crops may be infected with. As we know all the nutrients present in the soil but what amount of nutrients are present in the particular field. Every soil has different micronutrient. But to measure the amount of nutrient available in the soil we are going to design a device which will give proper reading of the micronutrient and that can be used to predict crops, fertilizers and crop diseases. Following are the main objectives of the proposed systemDesign and develop a microcontroller-based sensor interfacing for reading soil parameters (NPK value). Converting the sensor value which in analog signal to digital signal for further processing Sending all reading to system using USB ports available. Developing

website application for displaying the result and generating the report. Based on the result give suggestion to improve the quality of soil.

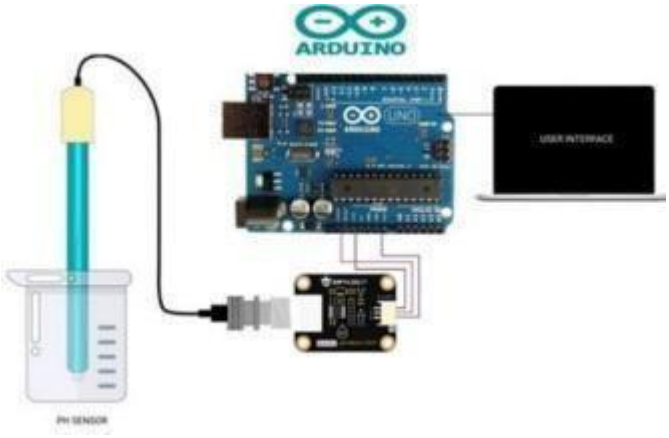


Fig: System design

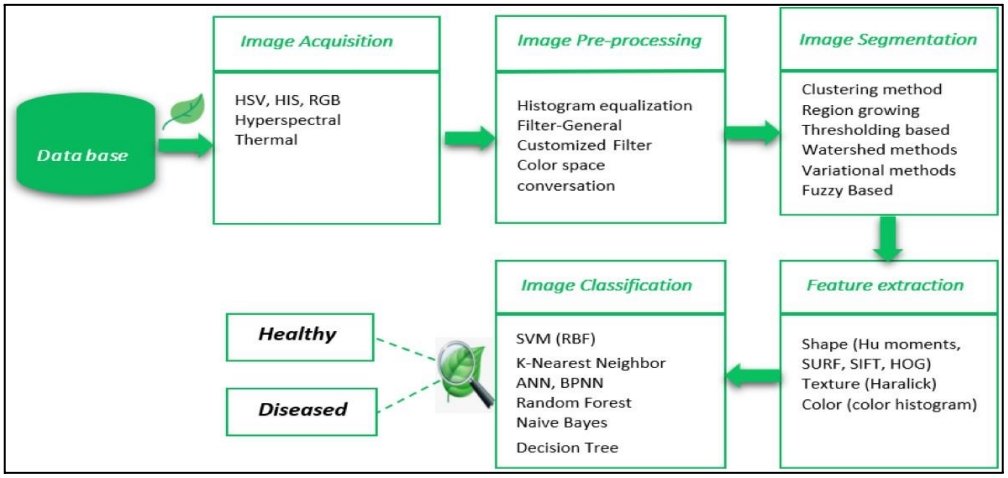
CHAPTER-8

TESTING

8.1. TEST CASE:

The primary step in identifying diseases is the acquisition of images. In most cases,images can be fetched either from a digital camera or an imaging system. As raw images tend to containoise, removing these impurities is required. As a result, the second step is known as image preprocessing, and involves the removal of unwanted distortions, in addition to contrast enhancement, to clarify and brighten the image features. For example, a Gaussian function that creates soft blur is commonly used to lessen the noise in the image. Subsequently, image segmentation is the third step in which the image is segmented from its background, whereas the region of interest (ROI) is partitioned to emphasize the prominent features. The fourth step is feature extraction [which unveils the information and details of an image. As a side note, the leaf features usually include shape, texture, and color, which are used to diagnose the crop. Thus, these chosen features form an input feature vector which is then fed into the classifier. Using this vector, it is possible to discriminate one class of objects from another. The final step is classification Note that the choice of a suitable classifier depends on the specific problem. The classifier's aim is to recognize the images by sorting them into several predefined classes based on the resulting feature vector obtained in the fourth step. For this purpose, the classification task contains two phases, namely, training and testing. The training operation trains the classifier on a training dataset; thus, the greater the number of training sets, the better the accuracy obtained. It should be noted that the result, which is the crop's healthy

state or diseased state associated with the species name, must be achieved as swiftly as possible.



8.2. USER ACCEPTANCE TESTING:

8.2.1. IMAGE CLASSIFICATION:

i. **Image acquisition:** To get the image of a leaf so that evaluation in the direction of a class can be accomplished.

ii. **Preprocessing:**

The purpose of image preprocessing is improving image statistics so that undesired distortions are suppressed and image capabilities which are probably relevant for similar processing are emphasized. The preprocessing receives an image as input and generates an output image as a grayscale, an invert and a smoothed one.

iii. Segmentation:

Implements Guided active contour method. Unconstrained active contours applied to the difficult natural images. Dealing with unsatisfying contours, which would try and make their way through every possible grab cut in the border of the leaf. The proposed solution is used the polygonal model obtained after the first step not only as an initial leaf contour but also as a shape prior that will guide its evolution towards the real leaf boundary.

iv. Disease Prediction:

Leaves are affected by bacteria, fungi, virus, and other insects. Support Vector Machine (SVM) algorithm classifies the leaf image as normal or affected. Vectors are constructed based on leaf features such as color, shape, textures. Then hyperplane constructed with conditions to categorize the preprocessed leaves and also implement multiclass classifier, to predict diseases in leaf image with improved accuracy.

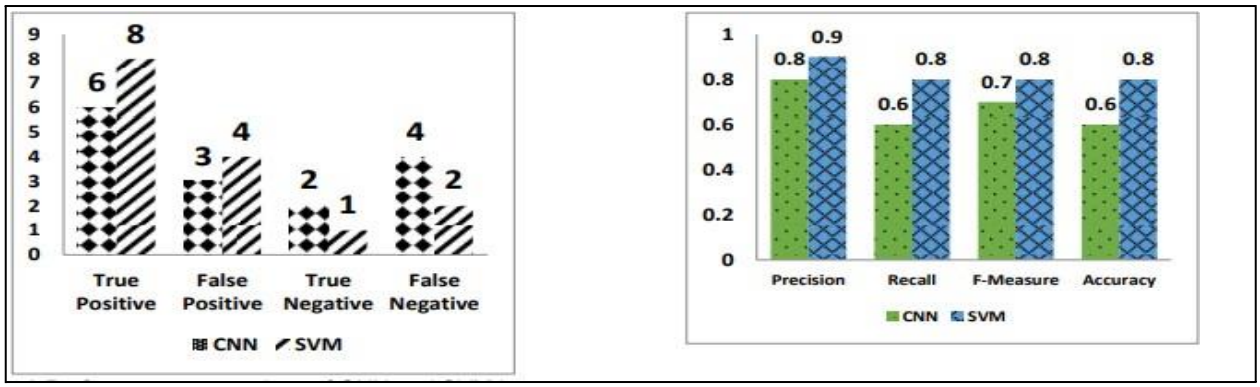
v. Fertilizer Recommendation:

Recommend the fertilizer for affected leaves based on severity level. Fertilizers may be organic or inorganic. Admin can store the fertilizers based on disease categorization with severity levels. The measurements of fertilizers suggested based on disease severity.

8.2.2. SVM Classification Algorithm:

Support Vector Machine(SVM) SVM is a binary classifier to analyze the data and recognize the pattern for classification. The main goal is to design a hyperplanethat classifies all the training vectors in different classes. The objective of SVM is to identify a function Fx which obtain the hyper-plane. Hyperplane separates two

classes of data sets. The linear classifier is defined as the optimal separating hyperplane. The data sets can be separated in two ways: linearly separated or nonlinearly separated. The vectors are said to be optimally separated if they are separated without error and the distance between the two closest vector points is maximum. For linear separable data sets, training vectors of a different class of pairs (am, bm), where m = 1,2,3,4 …,t am ϵ Rn(Reference Vector) bm ϵ { +1, -1} The decision boundary is placed using a maximal margin between the closest points. w is being a vector perpendicular median to the street. am be the unknown of to be positioned especially elegance according to the decision boundary, and hyperplane (w. a) + c =0 with c as constant For classification (w . am) + c0 ≥ 1, ∀ bm = +ve samples (1) (w . am) + c0 ≤ -1, ∀ bm = −ve samples (2) where (w.am) has a dot product of w and am. The inequalities if added i.e multiplying equations (1) and (2) with +1, −1 and bm. Suppose bm such that bm = 1 for +ve samples bm = -1 for -ve samples it results, bm [(w.am) + c0] ≥ 1 bm [(w.am) + c0] ≥ -1 Therefore rearranging the above equations bm (w.am) + c0 – 1 ≥ 0 for points into dataset to in the gutter i.e on the decision boundary bm (w.am) + c0 – 1 = 0.



Performance comparison of CNN and SVM in chart for terms of True Positive, False Positive, True Negative and False Negative

Precision, Recall, F-Measure and Accuracy comparison CNN and SVM

Precision, Recall ,E-Measure and Accuracy Value of CNN and SVM :

Classifiers	Pre	Re	F-M	Acc
ANN	0.8	0.6	0.7	0.6
SVM	0.9	0.8	0.8	0.8

3. Random Forest :

Random forest is a supervised machine learning algorithm based on ensemble learning. Ensemble learning is a type of learning where you join different types of algorithms or same algorithm multiple times to form a more powerful prediction model. The random forest algorithm combines multiple algorithm of the same type. Random Forest algorithm can be used for classification and regression problems.

Advantages:

1) The random forest algorithm is not biased, since, there are multiple trees and each tree is trained on a subset of data.
2) Random Forest algorithm is stable if a new data point is introduced in the dataset the overall algorithm is not affected.

CHAPTER -9

RESULT

9.1.PERFORMANCE METRICS:

The complete system is designed using Python. Different datasets like crop, crop yield dataset, Location, soil and crop nutrients, fertilizer datasets are gathered from other sources like agricultural books, agricultural websites.

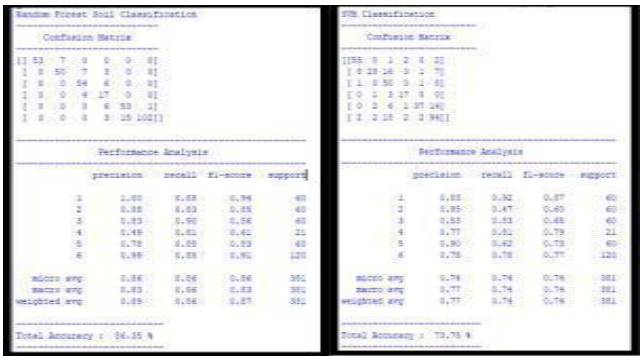


Fig 1. RF and SVM Classification

Figures (Fig 1) shows soil classification using Random Forest algorithm and Support Vector Machine. The output of these algorithms shows confusion matrixas a summary of algorithms different parameters like Precision, Recall averages and accuracy in percentage.

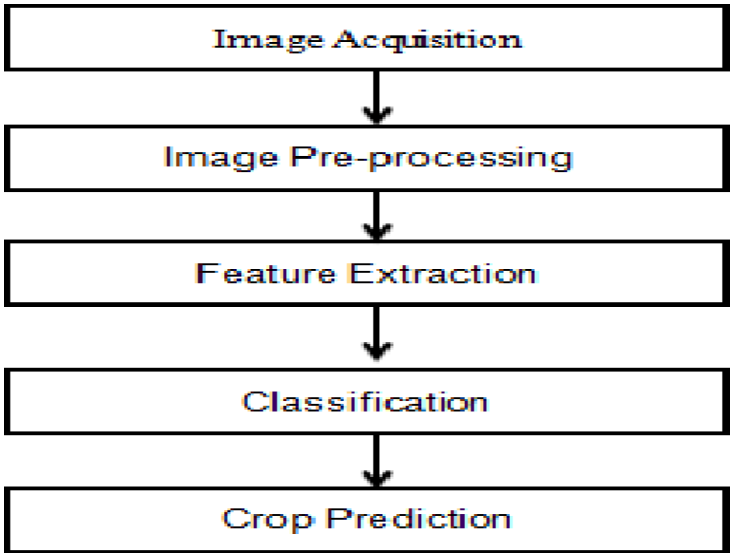


Fig 2.Soil Classification and Crop Yield Analysis

Figures (Fig 2) shows Soil Classification and Crop Yield analysis by graphical representations. The accuracy of Random Forest is 86.35% and Support Vector Machine is 73.75% so Random Forest algorithm is good for Soil Classification. The accuracy of SVM algorithm is 99.47% for yield prediction and RF accuracy is 97.48%. So, for crop yield prediction SVM algorithm is good.

Fig 3.Yield Prediction using RF and SVM

Figures (Fig 3)shows Yield Predicted by both algorithms like if Jowar is selected then Random Forest predicts the “Jowar Predicted Yield 121 Q/H” and SVM predicts “Jowar Predicted Yield is 125Q/H”.

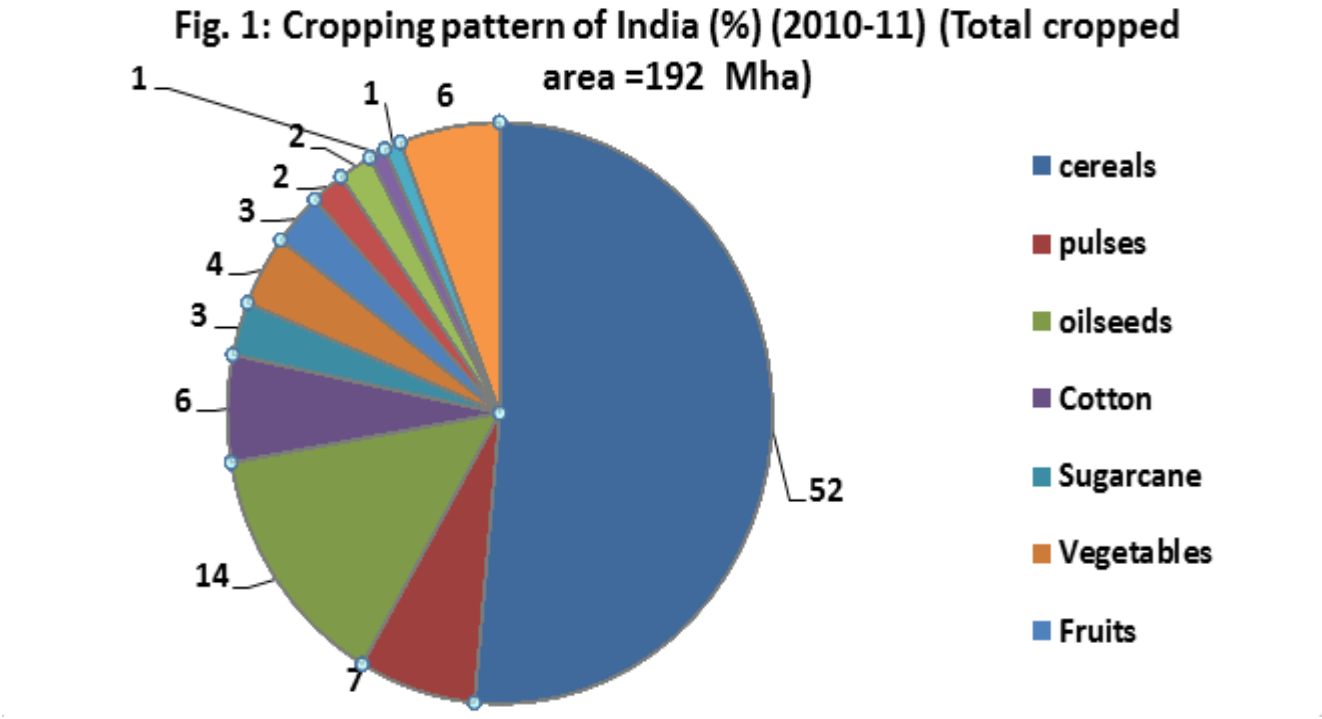


Fig 4.Crop Wise Yield Analysis

Figures (Fig 4)Shows Crop wise yield analysis where each crop yield is taken into consideration.

CHAPTER-10

ADVANTAGES & DISADVANTAGES

To improve the productivity of plants and boost their growth a chemical substance is added to soil or plant tissues is known as Fertilizers. It can be organic or inorganic and both can be used for providing basic nutrients to plants. Most fertilizers consist of nitrogen, potassium, phosphorus and some contain other nutrients like lime, zinc, magnesium, etc. which are necessary for plant growth. It is beneficial for agricultural production. so basically three types of fertilizers are commonly used;

- Chemical
- Organic
- biofertilizers

Chemical fertilizers: The macronutrients nitrogen, phosphorus, and potassium are mainly present in chemical or inorganic fertilizers. This chemical fertilizer ismostly used because it is easily soluble and boosts the growth of the crop instantly.

Organic fertilizers: Organic fertilizers are mainly made by elements that occur in nature or end product of natural material. Like most of the other fertilizers, it also provides basic nutrients of the plant which are potassium, nitrogen, and phosphorus. Organic fertilizers increase a lot of soil organisms which helps toboost plant growth.

Biofertilizers: Microbial inoculants or biofertilizers help fix the soil elements by non-usable to usable form through the biological process. These fertilizers are microorganisms like bacteria, algae, and fungi which help in fixing the soil atmosphere and convert potassium phosphate into a soluble form so that soil can absorb it to boost plants.

- Advantages of Fertilizers
- Disadvantages of Fertilizers
- Comparison Table for Advantages and Disadvantages of Fertilizers
- FAQ' s on Pros and Cons of Fertilizers

Fertilizers come with some advantages and a few disadvantages .let us discuss its advantages:

Advantages of Fertilizers:

Full of nutrients: All three types of fertilizers provide important nutrients to the plants which are helpful in the growth of plants and crops. Potassium, nitrogen, phosphorus are basic elements to boost the yield which is easily provided by all fertilizers in different ways. Chemical fertilizers are made artificially and it has fast production but it boosts quickly the plants so it is used widely whereas organic fertilizers are made by natural process and take time to be prepared but they are good for plants also gives all the necessary nutrient to the plant without harming the nature and human .bio fertilizers are also made by a natural process so it takes time to be processed but it helps soil to fix its problems naturally and convert it into efficient for plants. So in any form fertilizers are the main ingredient of plants that help them to grow properly.

Fast absorption: Sometimes plants need a quick fix to survive, in this type of cases fertilizers play a vital role to improve plants' health. plants need nutrients that can be absorbed quickly which is fulfilled by fertilizers. They are easily soluble and fastly absorbed by plants and as soon as possible it helps to regain and boost plant health.

Easily Available: As the fertilizers are very helpful in fast improvement in plants and best for agricultural production its demand also increases but many factories are working constantly to reach the demand. Thus now all the types of fertilizers are easily available in the market.

Enhance Metabolism: Fertilizers are food for plants that promote their growth. It can only be possible when a proper metabolic activity is processed. Fertilizers are easily digestible by plants and thus increase their metabolism rate to enhance plant growth.

Beneficial for large productions: As the population is increasing, there is a huge demand for food, so good yield is required to fulfill the demand. Here fertilizers become helpful for the good production of crops due to their numerous benefits which promote the fast and healthy growth of plants. For large production, fertilizers become compulsory.

Let us discuss some Disadvantages of Fertilizers:

Expensively: Fertilizers are man-made so they need production in factories which makes them costlier than naturally made manure. But it is important for plant nutrients so it is in demand and thus it has high value.

Over usage can damage plants: Fertilizers are used in moderate quantities if we use excessive fertilizers it surely damages the roots of plants and their tissues and thus plants can die. fertilizers are used according to the need of the plant. Unnecessary use of them can affect the plant's health specially if plants have good fertile soil.

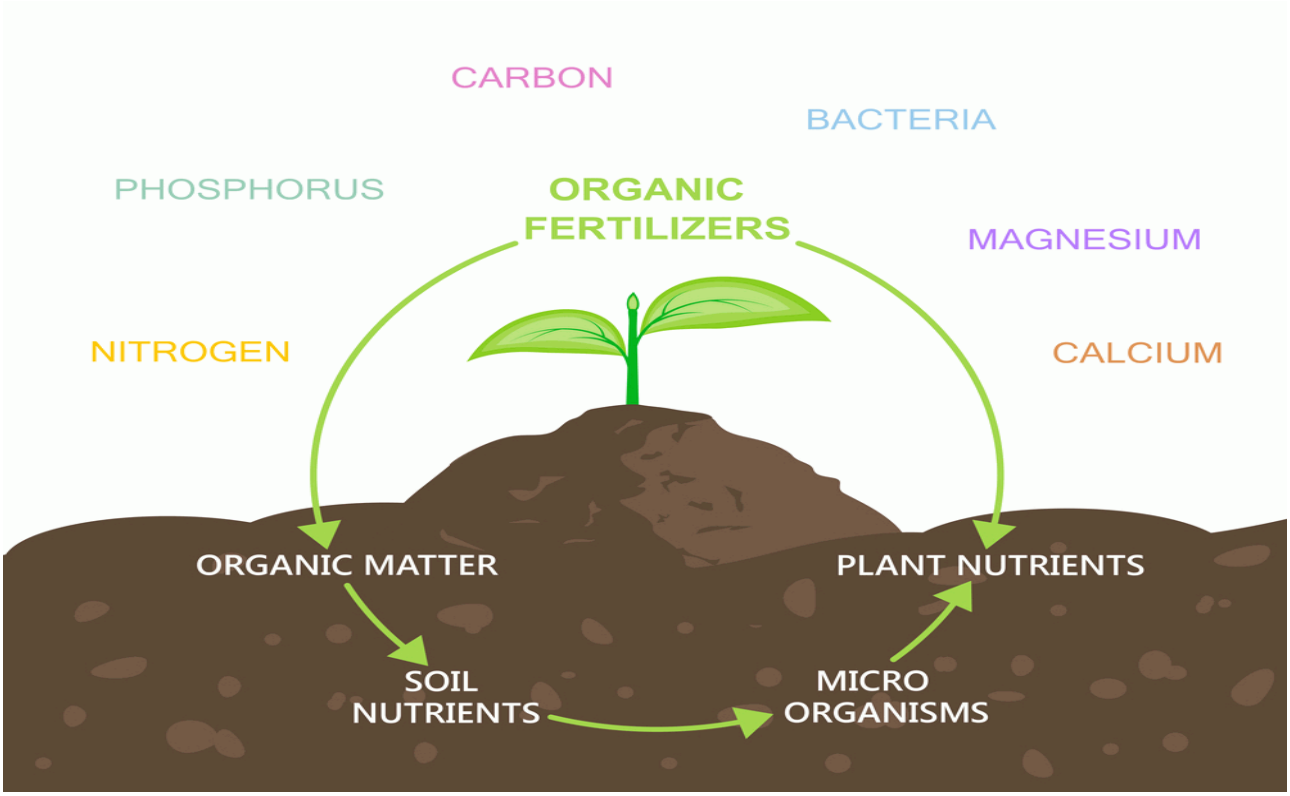
Toxic: There are many types of fertilizers in the market, some of them are chemically made. These chemical fertilizers are harmful to humans and plants also. Skin irritation, respiratory problems commonly occur due to fertilizers. Can pass harmful chemical in our food which affects

Effects of the environment: Indirect or indirect ways fertilizers are affecting the environment. by soil pollution, groundwater pollution and also can affect thegrowth of plants by faster growth instead of normal. It can affect the ecosystem.

Reduce soil quality: fertilizers can reduce the quality of soil and can harm microorganisms in the soil. Long-term use disturbs the pH of the soil and also reduces the microbial activities which are naturally good for plants.

WHAT IS A FERTILIZER ?

A fertilizer is any material of natural or synthetic origin that is applied to soils or to plant tissues to supply one or more plant nutrients essential to the growth of plants.



Comparison Table for Advantages and Disadvantages of Fertilizers

Advantages of fertilizers	**Disadvantages of fertilizers**
Fertilizers have all nutrients required for plants growth	Fertilizers are more expensive than manure.
It is soluble and easily absorbed by plants	Overfertilization can damage the plants
It enhances the metabolism of plants	It is toxic and can harm humans
It is easily available in the market	It affects the environment and ecosystem
Highly needed for large production	Long term use reduce soil quality

Comparison Table for Advantages and Disadvantages of DAP

Advantages

- It has high nutrient content
- DAP is free-flowing and dust-free fertilizer
- DAP is water-soluble
- It is not subjected to leaching losses.

Disadvantages

- Overuse of DAP fertilizer fails to resolve the nutrient deficiency problem in the soil
- It may cause leaf chlorosis
- Excessive use may harm beneficial microorganisms living in soil
- DAP may cause high probability of water pollution

AplusTopper.com

CHAPTER-11

CONCLUSION

CUSTOMER JOURNEY:

In this paper, we propose a user-friendly web application system based on machine learning and web-scraping called the ‘Farmer’s Assistant’. With our system, we are successfully able to provide several features - crop recommendation using Random Forest algorithm, fertilizer recommendation using a rule based classification system, and crop disease detection using EfficientNet model on leaf images. The user can provide the input using forms on our user interface and quickly get their results. In addition, we also use the LIME interpretability method to explain our predictions on the disease detection image, which can potentially help understand

why our model predicts what it predicts, and improve the datasets and models using this information. While our application runs very smoothly, we have several directions in which we can improve our application. Firstly, for crop recommendation and fertilizer recommendation, we can provide the availability of the same on the popular shopping websites, and possibly allow users to buy the crops and fertilizers directly from our application.

CHAPTER-12

FUTURE SCOPE

This further research is implementing the proposed algorithm with the existing public datasets. Also, various segmentation algorithms can be implemented to improve accuracy. The proposed algorithm can be modified further to identify the disease that affects the various plant organs such as stems and cropsThe vast potential of Indian agriculture remains unexplored, and we still have a long way to go in this field of study, as we need to make the device more compact, lightweight, and inexpensive to farmers. The technology will assist farmers by providing required advice on crops, their growth, and other basic information. It will also offer the location of the nearest store where farmers can purchase fertilizer and other materials. It would also assist farmers in selling their commodities to merchants by providing accurate information on market prices and merchant details. The device can also help farmers calculate crop MSP. The disease detectionfeature can also be improved by adding dedicated cameras to the device, whichwill improve the device's accuracy even further.

COMPARATIVE STUDY:

The Several prototypes have been suggested in this sector that are assisting in the resolution of agricultural challenges. A comparable model is presented in paper [16], and the principle is partly similar, but we tried to cut the cost substantially, and our model's accuracy is also higher due to the usage of the XG Boost model. Our system also has certain extra characteristics, such as crop disease prediction, which has boosted our project and has significant promise in tackling important agricultural concerns. vast potential of Indian agriculture remains unexplored, and we still have a long way to go in this field of study, as we need to make the device more compact, lightweight, and inexpensive to farmers. The technology will assist farmers by providing required advice on crops, their growth, and other basic information. It will also offer the location of the nearest store where farmers can purchase fertilizer and other materials. It would also assist farmers in selling their commodities to merchants by providing accurate information on market prices and merchant details. The device can also help farmers calculate crop MSP. Thedisease detection feature can also be improved by adding dedicated cameras to the device, which will improve the device’s accuracy even further.

CHAPTER-13

APPENDIX

SOURCE CODE:

 # Importing libraries

1.from__future__import print_function

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.metrics import classification_report

from sk learn import metrics

from sk learn import

tree import warnings

warnings.filterwarnings('ignore')

2.df = pd**.**read_csv('../Data-processed/crop-recommendation.csv')

3.df**.**head()

4. df**.**tail

() **5.**df**.**size

6.df**.**shape

7.df**.**column

s

8.df['label']**.**unique()

9.df**.**dtypes

10,sns**.**heatmap(df**.**corr(),annot=**True**)

Separating features and target label

11.features = df[['N', 'P','K','temperature', 'humidity', 'ph', 'rainfall']]
target = df['label']

12.#features = df[['temperature', 'humidity', 'ph', 'rainfall']]
labels = df['label']

13. # Initializing empty lists to append all model's name and corresponding name acc = []
model = []
# Splitting into train and test data

14. **from** sklearn.model_selection **import** train_test_split
Xtrain, Xtest, Ytrain, Ytest = train_test_split(features,target,test_size = 0.2,random_state =2)

Decision Tree

15. **from** sklearn.tree **import** DecisionTreeClassifier
DecisionTree = DecisionTreeClassifier(criterion="entropy",random_state=2,max_depth=5)
DecisionTree**.**fit(Xtrain,Ytrain)
predicted_values = DecisionTree**.**predict(Xtest)
x = metrics**.**accuracy_score(Ytest, predicted_values)
acc**.**append(x)
model**.**append('Decision Tree')
print("DecisionTrees's Accuracy is: ", x*****100)

16. print(classification_report(Ytest,predicted_values))

17. **from** sklearn.model_selection **import** cross_val_score

18. # Cross validation score (Decision Tree)

score = cross_val_score(DecisionTree, features, target,cv=5)

19. score

Saving trained Decision Tree model

20. **import** pickle

Dump the trained Naive Bayes classifier with Pickle

DT_pkl_filename = '../models/DecisionTree.pkl'

Open the file to save as pkl file

DT_Model_pkl = open(DT_pkl_filename, 'wb')

pickle**.**dump(DecisionTree, DT_Model_pkl)

Close the pickle instances

DT_Model_pkl**.**close()

Guassian Naive Bayes

21. **from** sklearn.naive_bayes **import**

GaussianNB NaiveBayes = GaussianNB()

NaiveBayes**.**fit(Xtrain,Ytrain)

predicted_values = NaiveBayes**.**predict(Xtest)

```
x = metrics.accuracy_score(Ytest, predicted_values)
acc.append(x)
model.append('Naive Bayes')
print("Naive Bayes's Accuracy is: ", x)
```

22. print(classification_report(Ytest,predicted_values))

22.# Cross validation score (NaiveBayes)

23.score = cross_val_score(NaiveBayes,features,target,cv=5)

24. score

Saving trained Guassian Naive Bayes model

25. **import** pickle

```
# Dump the trained Naive Bayes classifier with Pickle
NB_pkl_filename = '../models/NBClassifier.pkl'
# Open the file to save as pkl file
NB_Model_pkl = open(NB_pkl_filename, 'wb')
pickle.dump(NaiveBayes, NB_Model_pkl)
# Close the pickle instances
NB_Model_pkl.close()
```

Support Vector Machine (SVM)

26. **from** sklearn.svm **import** SVC

```
# data normalization with sklearn
from sklearn.preprocessing import MinMaxScaler
```

```
# fit scaler on training data
norm = MinMaxScaler().fit(Xtrain)
X_train_norm = norm.transform(Xtrain)
# transform testing dataabs
X_test_norm = norm.transform(Xtest)
SVM = SVC(kernel='poly', degree=3, C=1)
SVM.fit(X_train_norm,Ytrain)
predicted_values = SVM.predict(X_test_norm)
x = metrics.accuracy_score(Ytest, predicted_values)
acc.append(x)
model.append('SVM')
print("SVM's Accuracy is: ", x)
```

27. `print(classification_report(Ytest,predicted_values))`

28.
```
# Cross validation score (SVM)
score = cross_val_score(SVM,features,target,cv=5)
score
```

29.
```
#Saving trained SVM model
import pickle
# Dump the trained SVM classifier with Pickle
SVM_pkl_filename = '../models/SVMClassifier.pkl'
# Open the file to save as pkl file
SVM_Model_pkl = open(SVM_pkl_filename, 'wb')
```

```
# Close the pickle instances
SVM_Model_pkl.close()
```

Logistic Regression

```
30.  from sklearn.linear_model import
LogisticRegression LogReg =
LogisticRegression(random_state=2)
LogReg.fit(Xtrain,Ytrain)
predicted_values = LogReg.predict(Xtest)
x = metrics.accuracy_score(Ytest, predicted_values)
acc.append(x)
model.append('Logistic  Regression')
print("Logistic Regression's Accuracy is: ", x)
print(classification_report(Y test,predicted_values))
```

```
31.  # Cross validation score (Logistic Regression)
score =
cross_val_score(LogReg,features,target,cv=5)
```

32.score

Saving trained Logistic Regression model

```
# Dump the trained Naive Bayes classifier with Pickle
LR_pkl_filename = '../models/LogisticRegression.pkl'
# Open the file to save as pkl file
  LR_Model_pkl = open(DT_pkl_filename, 'wb')
  pickle.dump(LogReg, LR_Model_pkl)
  # Close the pickle
  instances
  LR_Model_pkl.close()
  Random Forest
  33. from sklearn.ensemble import RandomForestClassifier
  RF = RandomForestClassifier(n_estimators=20, random_state=0)
  RF.fit(Xtrain,Ytrain)
  predicted_values = RF.predict(Xtest)
  x = metrics.accuracy_score(Ytest, predicted_values)
  acc.append(x)
  model.append('RF')
  print("RF's Accuracy is: ",
  x)

  34. print(classification_report(Y test,predicted_values))
```

```
35.  # Cross validation score (Random Forest)
score =
cross_val_score(RF,features,target,cv=5) score
```

Saving trained Random Forest model

```
36. import pickle
# Dump the trained Naive Bayes classifier with Pickle
RF_pkl_filename = '../models/RandomForest.pkl'
# Open the file to save as pkl file
RF_Model_pkl = open(RF_pkl_filename,
'wb') pickle.dump(RF, RF_Model_pkl)
# Close the pickle instances
RF_Model_pkl.close()
```

XGBoost

```
37.  import x gboost as
x gb XB =
xgb.XGBClassifier()
XB.fit(Xtrain,Ytrain)
```

39.predicted_values = XB**.**predict(Xtest)

40.x = metrics**.**accuracy_score(Ytest, predicted_values)

acc**.**append(x)

model**.**append('XGBoost')

print("XGBoost's Accuracy is: ", x)

41. print(classification_report(Y Test,predicted_values))

42. # Cross validation score (XGBoost)

score = cross_val_score(XB,features,target,cv=5)

score

Saving trained XGBoost model

43. import pickle

Dump the trained Naive Bayes classifier with Pickle

XB_pkl_filename = '../models/XGBoost.pkl'

Open the file to save as pk l file

```
XB_Model_pkl = open(XB_pkl_filename, 'wb')

pickle.dump(XB, XB_Model_pkl)

# Close the pickle instances

XB_Model_pkl.close()
```

Accuracy Comparison

```
44.plt.figure(figsize=[10,5],dpi = 100)

plt.title('Accuracy Comparison')

plt.xlabel('Accuracy')

plt.ylabel('Algorithm')

sns.barplot(x = acc,y = model,palette='dark')



45.accuracy_models = dict(zip(model, acc))

46.for k, v in accuracy_models.items():

   print (k, '-->', v)
```

Making a prediction

```
47.data = np.array([[104,18, 30, 23.603016, 60.3, 6.7, 140.91]])

prediction =

RF.predict(data)

print(prediction)
```

48.data = np**.**array([[83, 45, 60, 28, 70.3, 7.0, 150.9]])

prediction = RF**.**predict(data): print(prediction)

GitHub & Project Demo Link:

https://github.com/IBM-EPBL/IBM-Project-32848-1660212539.git

ASSIGNMENTS:

ASSIGNMENT 1:

Basics of python codes in jupyter notebook.

ASSIGNMENT 2:

Data Visualization and Pre-processing

Perform Below Tasks to complete the assignment:-

Tasks:-

1. Download the dataset: Dataset
2. Load the dataset.
3. Perform Below Visualizations.
   - Univariate Analysis
   - Bi - Variate Analysis
   - Multi - Variate Analysis
4. Perform descriptive statistics on the dataset.
5. Handle the Missing values.
6. Find the outliers and replace the outliers
7. Check for Categorical columns and perform encoding.
8. Split the data into dependent and independent variables.
9. Scale the independent variables
10. Split the data into training and testing

ASSIGNMENT 3:

Problem Statement :- Build CNN Model for Classification Of Flowers

Perform Below Tasks to complete the assignment:-

- Download the Dataset : Dataset
- Image Augmentation
- Create Model
- Add Layers (Convolution,MaxPooling,Flatten,Dense-(Hidden Layers),Output)
- Compile The Model
- Fit The Model
- Save The Model
- Test The Model

ASSIGNMENT 4:

Problem Statement :- SMS SPAM Classification

Problem Statement:
Over recent years, as the popularity of mobile phone devices has increased, Short Message Service (SMS) has grown into a multi-billion dollar industry. At the same time, reduction in the cost of messaging services has resulted in growth in unsolicited commercial advertisements (spams) being sent to mobile phones. Due to Spam SMS, Mobile service providers suffer from some sort of financial problems as well as it reduces calling time for users. Unfortunately, if the user accesses such Spam SMS they may face the problem of virus or malware. When SMS arrives at mobile it will disturb mobile user privacy and concentration. It may lead to frustration for the user. So Spam SMS is one of the major issues in the wireless communication world and it grows day by day.

Perform the Below Tasks to complete the assignment:-

- Download the Dataset:- Dataset
- Import required library
- Read dataset and do pre-processing
- Create Model
- Add Layers (LSTM, Dense-(Hidden Layers), Output)
- Compile the Model
- Fit the Model
- Save The Model
- Test The Model