International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 11 Issue 10 ǁ October 2023 ǁ PP. 89-98

Crop Exhortation Using Machine Learning Approach

G. SARANYA[1], G. ABINAYA[2]
[*1] *Department of Computer Science & Engineering, Sir Issac Newton College of Engg & Tech,Nagapattinam*
[2]*Department of Artificial Intelligence & Data Science, Sir Issac Newton College of Engg & Tech, Nagapattinam*
PG Student[1], Assistant Professor[2],

Abstract
When cultivating crops, farmers must overcome a number of obstacles, including erratic irrigation and subpar soil. The majority of farmers, particularly in India, lack the expertise necessary to choose the right crops and fertilizers. Additionally, both the producers and the customers suffer a great deal when crops fail due to disease. Despite recent advancements in the automated detection of certain diseases using Machine Learning techniques, the application of Deep Learning has not yet been completely investigated. The high-quality data used in their training, the shortage of processing capacity, and the models' limited generalizability all contribute to the difficulty in using such models. In an effort to address some of these problems, we developed an open-source, user-friendly online application.

Keywords: *Crop Yield Prediction, Machine Learning, Random Forest, Crop Recommender System*.

Date of Submission: 02-10-2023 Date of acceptance: 13-10-2023

I. INTRODUCTION

Our farmers are at the vanguard of the agricultural business, which is incredibly risky. Farmers deal with a variety of issues, such as crops that are afflicted by illnesses and soil that is not nutrient-rich enough for plants to flourish on. The combined yield is decreased by each of these elements. Agribusiness is the primary source of income for more than 70% of Indians. Diseases cause more than 15% of India's crops to be lost, making it one of the most pressing issues that need to be addressed. Crop diseases pose a serious threat to crops, but because many regions of the world lack the requisite infrastructure, it is still difficult to quickly identify them.

The development of deep learning-enabled breakthroughs in computer vision has prepared the way for web-assisted disease detection. This is due to the combination of rising smartphone penetration in rural areas of the world and these developments. For a plant, there are a number of ways to look for illness, including growth, roots, and leaves. Numerous technologies already in use employ mobile applications to identify disease using leaf data. These programs, however, function with photographs of leaves that have a flat, all-black background.

Additionally, choosing the right crops and fertilizers for the soil is quite important because it enables farmers to maximize crop yield and boost sales. Nitrogen, phosphorus, and potassium are the three macronutrients that plants require. In a plant, N promotes the development of the leaves, P the growth of the root, flower, and fruit, and K aids in the general operation of the plant. Knowing a fertilizer's NPK values can assist a farmer choose a fertilizer that is suitable for the kind of plant they are growing. In this project, we suggest a system that aids farmers in the detection of plant illness, suggests the appropriate crop for their soil, and suggests fertilizers so they may obtain the most yield possible. We employ the Efficient Net deep learning model for plant disease detection, which on the dataset of our choice achieves 99.8% validation accuracy, the Random Forest model for crop recommendation based on soil (N, P, K, and pH) and weather features, and a rule-based classification scheme for fertilizer recommendation.

Additionally, in order for people to better comprehend our model, we do explanations of our disease detection model using a well-known interpretability technique called LIME. We do this by scanning news websites for information about plants.

1.1 LITERATURE SURVEY

A. Deep Learning in Computer Vision: Since the development of the ImageNet dataset and the ILSVRC challenge, deep learning in computer vision has made considerable strides. The widely used method for addressing computer vision issues with a shortage of data is to pre-train deep learning models using ImageNet. The majority of WordNet's 80,000 synsets will be filled with an average of 500–1000 crisp, full-resolution images thanks to ImageNet. Numerous deep convolutional neural networks have been developed since the beginning to take on the problem. The VGG network has 19 layers, but AlexNet only has 5 convolutional layers. The Vanishing/Exploding gradient problem was solved by ResNet employing residual connections. MobileNet is made for embedded and

mobile vision applications while considering the limited resources. In order to scale up CNNs more methodically, Efficient Net suggests a model scaling technique that makes use of a very effective compound coefficient.

B. Plant Disease Detection: Over the years, a variety of different techniques have been developed, with the most recent ones utilizing deep learning techniques. Plant Disease Detection has been a very active area of research. To feed an artificial neural network, features like color, texture, and morphology were retrieved. Additionally, they test using data retrieved from Bing and Google Search and depict activations. Furthermore, they draw attention to the fact that these programs perform badly on photographs taken in the actual world, which typically contain several leaves and backgrounds that are highly varied.

C. Crop Recommendation: Using a variety of techniques, including Support Vector Machines, Artificial Neural Nets, and Random Forests, researchers have determined that Random Forests perform the best for their dataset when recommending crops. Additionally, they develop a mobile application system that uses GPS position data to collect crop production predictions for certain crops as well as crop recommendations depending on inputs like area and soil quality. The majority vote method for crop recommendation is similar, using an ensemble of CHAID, Naive Bayes, K-NN, and Random Trees.

D. Fertilizer Recommendation: Numerous studies on fertilizer recommendations have been conducted, and the majority of them make use of the soil's N, P, K, and pH values, sometimes in addition to depth, temperature, weather, location, and precipitation. Rule-based categorization is the standard method; however, some methods include cluster data on fertilizers using K-Means and Random Forests for suggestion. E. Deep Learning Interpretability the LIME approach is a straightforward interpretability method that substitutes a local linear regression model for the original model. On masked versions of the image, the linear model is trained using the original model predictions. Based on their appropriate weights, the scores for the image segments are calculated. Positive and higher-scoring segments are crucial for the projected class, but lower-scoring segments reduce the model's level of confidence. The Grad CAM performs a linear combination to identify the regions with a positive influence on a given class, weighting the activation maps at that layer using the average of gradients at the last convolutional layer of a CNN-based model. This offers a heatmap of key areas leading to a prediction.

II. PROBLEM DEFINITION

The utilization of diverse composts is also uncertain due to changes in climatic conditions and significant resources like soil, water, and air. In this circumstance, the harvest production rate is steadily declining. Projecting harvest yields is difficult in the horticultural industry. Each rancher tries to assess crop productivity and whether it matches their expectations, forecasting the yield based on past harvesting experiences. The main factors affecting horticulture yields are the weather, annoyances, and job planning. Accurate information on crop history is needed to make judgments regarding horticulture risk.

2.1 EXISTING SYSTEM.

One technique that was previously utilized for determining the type of land was image-based analysis.

ALGORITHM DISADVANTAGES

Because soil conditions are not considered in this procedure, which is based on image analysis, the process's outcomes are not accurate.

The process of processing images takes time.

2.2 PROPOSED SYSTEM

Python's programming language is advantageous for applying a variety of methods for agricultural production prediction depending on the input data set when using machine learning, a cutting-edge technology. Support Vector Machine (SVM), Logistic Regression, and the Random Forest method are employed in this procedure to make predictions. In this research, testing training is carried out using a text dataset that comprises features for soil and temperature conditions and labels for crop kind.

ADVANTAGES

A textual dataset is used to forecast crop production, and any user can evaluate which crop type best suits their growing environment and receive crop recommendations.

III. SYSTEM ANALYSIS

Planning, analysis, design, deployment, and maintenance are just a few of the processes that make up the systematic process of developing a system. It is a method of resolving issues that enhances the system and guarantees that every part functions effectively to provide its intended function. Analysis lays out the system's proper course of action. System design is the process of designing the components or modules of a new business system or updating an existing system in order to meet the necessary requirements. Prior to making any plans, it is essential to have a complete understanding of the current setup and decide how to make the most effective use

of computers. System design focuses on how to achieve the system's goal. The system design outlines the system specifications, operating environment, system and subsystem architecture, file and database design, input and output formats, layouts for the human-machine interface, detailed design, processing logic, and external interface. programme design comprises additional technical tasks, including design coding implementation testing after the programme requirements have been examined and described.This phase's design activities are of the utmost importance because it is during these that choices are made that will ultimately determine how well the software will be implemented and how simple it will be to maintain.

3.1 USE CASES AND USE CASE DIAGRAM

A dynamic or behaviour is shown in a use case diagram. employ case diagrams employ actors and use cases to represent the functionality of a system. Use cases are a collection of tasks, offerings, and operations the system must handle. A "system" in this sense refers to something that is being created or run, like a website. The "actors" are individuals or groups acting in certain capacities inside the system. The use case diagram's main goal is summarised as follows:

1. Used to collect system requirements.
2. Applied to obtain a system's external perspective.
3. Identify the system-influencing external and internal elements.

The development team can locate potential issues during a transaction with the aid of a use case document and fix them. The use case diagram for the analysis of comments and tweets is shown below. We can quickly recognise the modules in this diagram by understanding what they represent. Only the user and the system may take part in this activity to ensure that the procedure may be carried out. The actors (users) are shown in figure 4.1 together with the functional requirements the system meets for them. The system has two actors: the end user (the farmer) and the administrator. Ovals are used to depict the system's functionality. The interdependence and visibility of the functionalities are shown by the arrows.

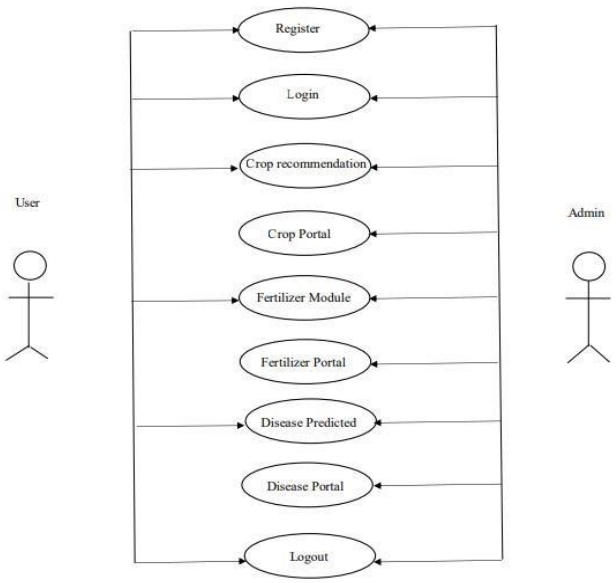

Fig 4.1 Use Case Diagram

3.2 SEQUENCE DIAGRAM

A sequence diagram only shows how things interact with one another sequentially, or the order in which these interactions occur. A sequence diagram can also be referred to as event diagrams or event scenario. Sequence diagrams show the actions taken by the components of a system in chronological order. Businesspeople and software engineers frequently use these diagrams to record and comprehend the requirements for new and current systems.

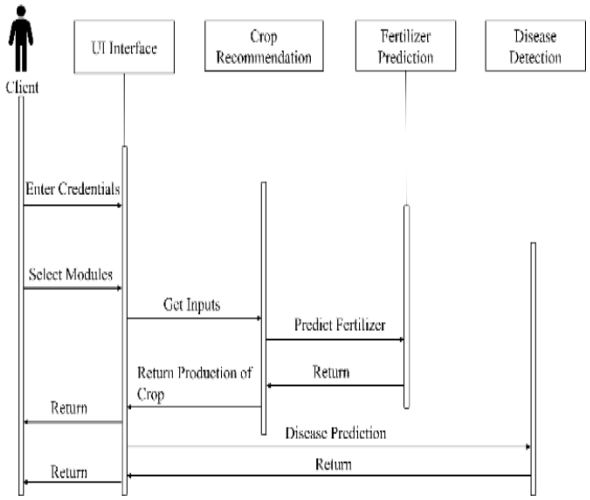

Fig 4.2 Sequence Diagram

3.3 ACITIVITY DIAGRAM

Another crucial UML diagram for describing the system's dynamic elements is the activity diagram. In essence, it is a flowchart that shows how one activity leads to another. The action might be referred to as a system operation. One operation leads to the next in the control flow. This flow may be parallel, contemporaneous, or branched. By utilising various parts like fork, join, etc., activity diagrams cope with all kinds of flow control. It captures the system's dynamic behaviour.

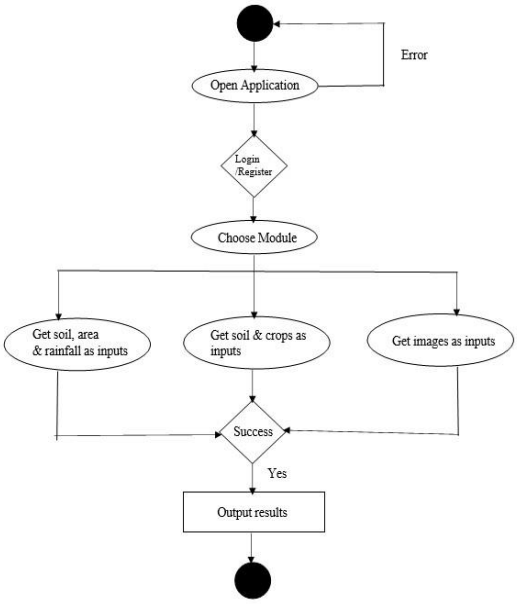

Fig 4.3 Activity Diagram

3.4 CLASS DIAGRAM

Static diagrams include class diagrams. It represents the application's static view. Class diagrams are used to create executable code for software applications as well as for visualizing, explaining, and documenting various elements of systems. The characteristics and functions of a class are described in a class diagram, along with the restrictions placed on the system. Because they are the only UML diagrams that can be directly mapped with object-oriented languages, class diagrams are frequently employed in the modeling of object-oriented systems.

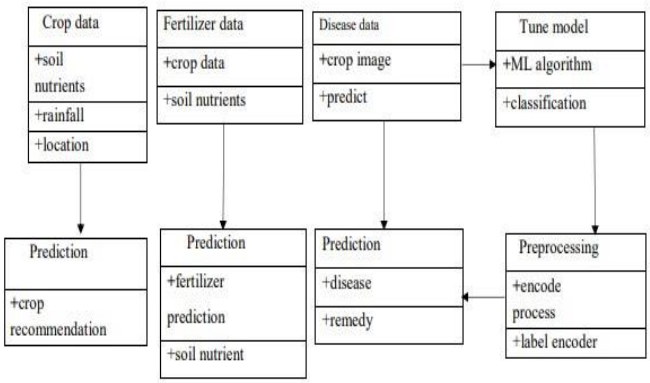

Fig 4.4 Class Diagram

IV. SYSTEM DESIGN

A design is a useful technical depiction of a future construction. It is the most important stage of a system's development. The process of translating requirements into a software representation is called software design. at software engineering, design is encouraged at the design phase. The new system must be designed based on user requirements and a thorough analysis of the current system. The system design phase is currently underway. The only approach to faithfully translate a customer need into a finished software solution is through design. Design develops a representation or model and offers information about the architecture, interfaces, and components required to implement a system.

4.1 ARCHITECTURAL DESIGN

The overall logical structure of the project is divided into processing modules and a conceptual data structure is defined as Architectural Design.

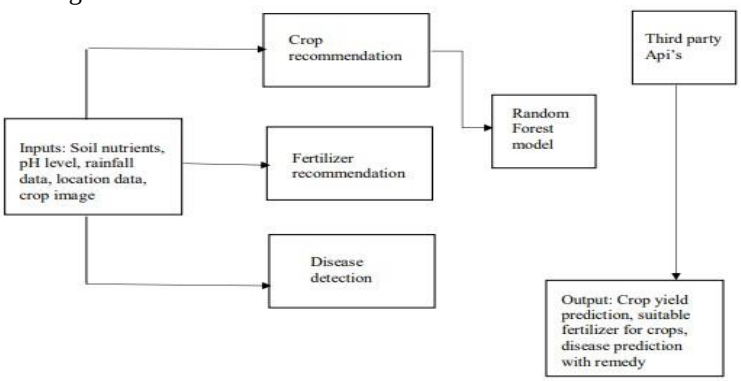

Fig 4.1 Architectural Design

4.2 DATA DESIGN

Agricultural Dataset - The majority of the research publications analysed took climatic factors including area, temperature, precipitation, and humidity into consideration. Seasonal variations and some soil agronomical factors, such as chalky, clay, loamy, and so on, are included. These variables' data were provided as input. In the beginning, a dataset is compiled that includes variables such properties like State Name, District Name, Humidity, Temperature, Yield, etc. Any crops that will be planted in the area should be considered. This dataset was compiled in csv format.

Pre-processing - For the data mining application, a sizable dataset is required. The data gathered from various sources is frequently in an unprocessed state. It might contain facts that are incorrect, out-of-date, or contradictory. Therefore, in this process, such redundant data should be filtered. The data needs to be normalised. Many of the 'NA' values in the provided data collection have been filtered in Python. Test and Train Model - Datasets will be split into training datasets and testing datasets during the pre-processing step. This is a crucial stage in the creation of a model. A model is trained using the training data, and it is then tested using the testing data. We therefore use training data to fit the model and testing data to test it. Classification Algorithm - After data splitting, the classification algorithm's next step is to create and train a model using scikit-learn. To learn the pattern, a machine learning algorithm and training set are needed during the training process. Here, we are utilising the well-known supervised learning algorithm Random Forest, which utilises the bagging technique. The Random

Forest Algorithm combines various decision trees. This approach uses an ensemble classifier as its foundation for classification. The dataset will be split into training and testing data. The decision tree is constructed using additional training dataset. During training, the random forest algorithm creates several decision trees.Predict Yield: Using the previously trained model, one may forecast the output from fresh input. Here, the trained model was saved in a file so that it could be predicted using the new input. Accuracy is one of the measures used to assess classification model accuracy. By dividing the total number of forecasts by the number of right predictions, accuracy is calculated. Since we were able to obtain a 98% accuracy rate, this model is effective at forecasting yield.

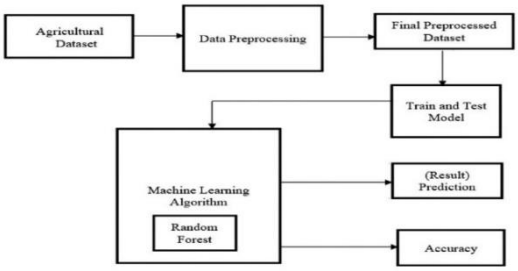

Fig 4.2 Data Design

4.3 E-R DIAGRAM

Entity relationship diagrams are used in software engineering during the planning stages of the software project. They help to identify different system elements and their relationships with each other.

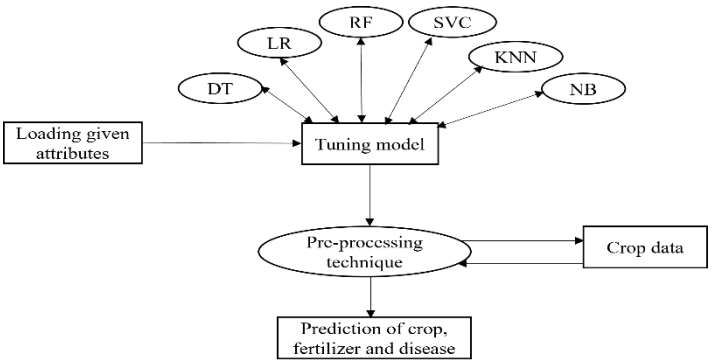

Fig 4.3 ER Diagram

V. IMPLEMENTATION

Implementation is the most crucial stage in achieving a successful system and giving the user's confidence that the new system is workable and effective. Implementation of a modified application to replace an existing one, this type of conversation is relatively to handle, provide there are no major changes in the system. Each program is tested individually at the time of development using the data and has verified that this program linked together in the way specified in the program specification, the computer system and its environment is tested to the satisfaction the computer system and its environment is tested to the satisfaction of the user. The system that has been developed is accepted and proved to be satisfactory for the user. And so, the system is going to be implemented very soon. Initially as a first step the executable from of the application is to be created and loaded in the common server machine which is accessible to the entire user and the final stage is to document the entire system which provides components and the operating procedure of the system.

MODULES

1. Login / Register
2. Crop Recommendation
3. Fertilizer Recommendation
4. Disease Prediction

5.1 LOGIN / REGISTER In this model, first the user should register with the following credentials:

- Username
- User mail
- User password

After registration, the user should login to the account for the further access of the model. For login the user should use the registered credentials such as:

- Username
- User mail
- User password

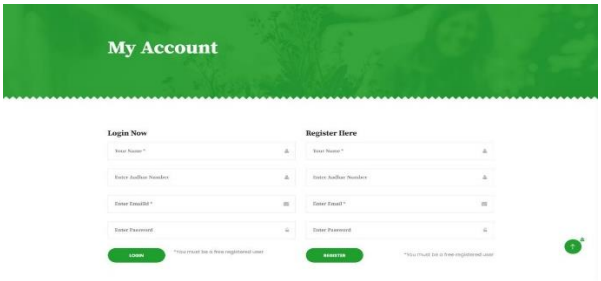

Fig 5.1 Login / Register

5.2 CROP RECOMMENDATION

On entering the values of Nitrogen, phosphorus, and Potassium a post request is made to the flask API. After the model runs an HTTP response is sent to the front-end which tells the best crop a farmer can grow in the soil in order to get the best out of the land.

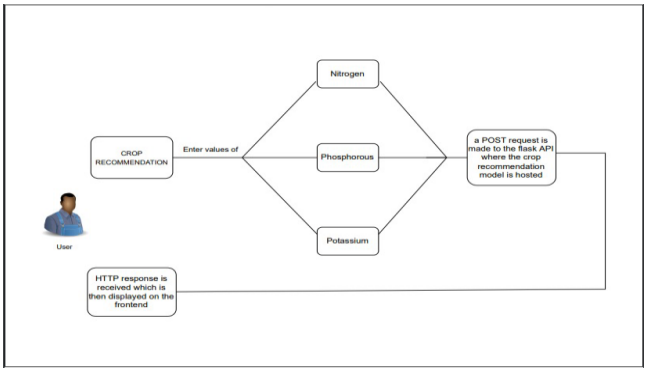

Fig 5.2 Flow Diagram of Crop Recommendation

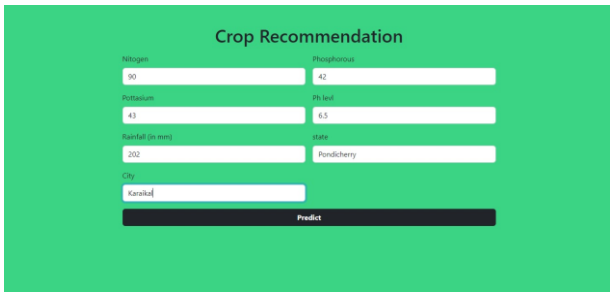

Fig 5.3 Crop Recommendation Module

Fig 5.4 Crop Recommendation Result

5.3 FERTILIZER RECOMMENDATION

The user has to enter the Nitrogen, Phosphorus, Potassium values along with the crop Name. A POST request is made to the flask API. Over here the fertilizer recommendation classifier is hosted. An HTTP response is sent to the front-end and in turn on the front-end the user gets recommendation to fertilizer

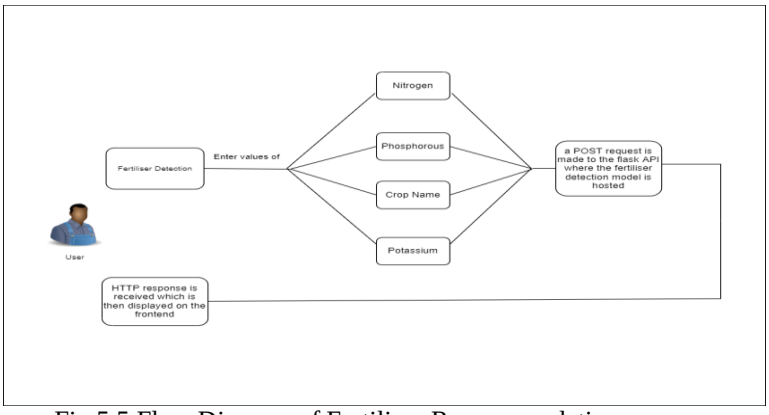

Fig 5.5 Flow Diagram of Fertilizer Recommendation

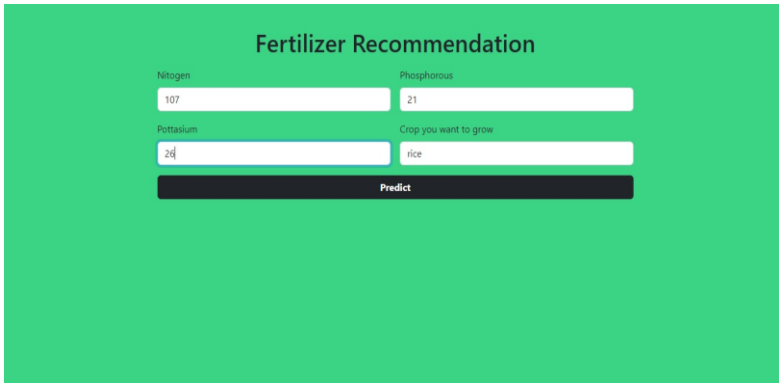

Fig 5.6 Fertilizer Recommendation Module

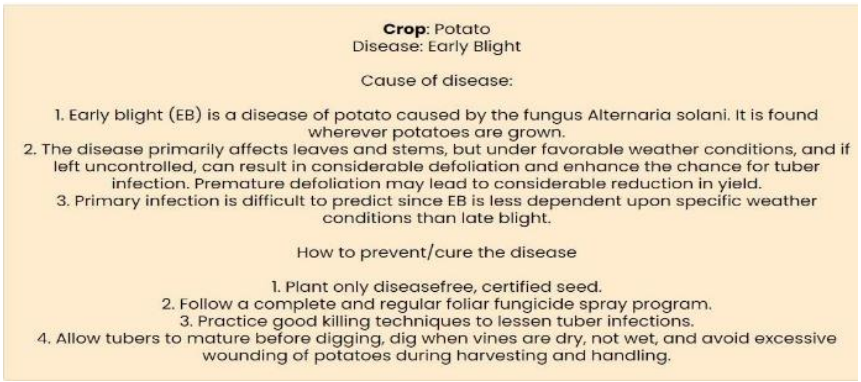

Fig 5.7 Disease Detected Result

5.4 DISEASE PREDICTION

In disease detection the user has to click an image or directly upload it. The image is sent to the back end and is processed by the model. After the image has been processed an HTTP response is sent to the front-end. The user receives the disease the plant has and its remedies. The disease portal provides a detailed view of various plant diseases and the kinds of products that may be bought to cure the plants of the disease

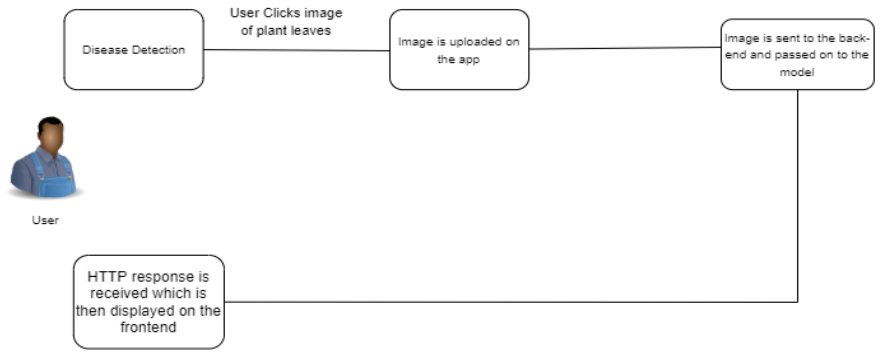

Fig 5.8 Flow Diagram of Disease Detection

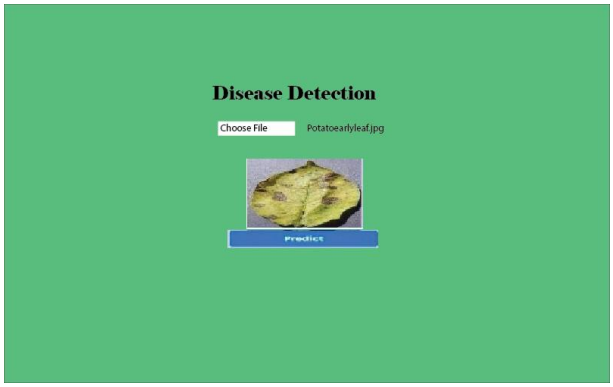

Fig 5.9 Disease Detection Module

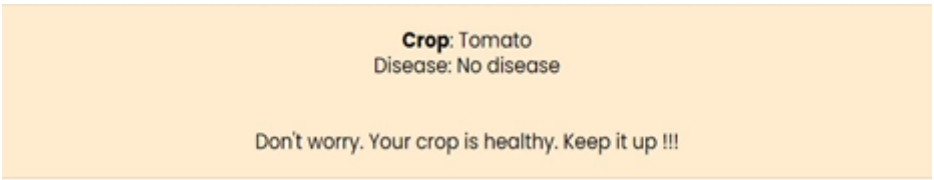

Fig 5.10 Disease Detected Result

VI. CONCLUSION AND FUTURE ENHANCEMENT

6.1 CONCLUSION

In this project, 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.

6.2 FUTURE ENHANCEMENT

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. Another improvement that can be done with fertilizer recommendation is that we want to be able to find data on various brands and items available based on the N,P,K values. In future, we want to be able to use complex machine learning systems to provide finer recommendations.

REFERENCERE

[1]. Predicting Yield of the Crop Using Machine Learning Algorithm by P. Priya, U. Muthaiah & M. Balamurugan, International Journal of Engineering Sciences & Research Technology (IJESRT), April 2018.

[2]. Machine learning approach for forecasting crop yield based on climatic parameters by S.Veenadhari, Dr.Bharat Misra &Dr.CD Singh, International Conference on Computer Communication and Informatics (ICCCI -2014), Jan, 2014.

[3]. A Survey on Crop Prediction using Machine Learning Approach by Sriram Rakshith.K, Dr.Deepak.G, Rajesh M, Sudharshan K S, Vasanth S & Harish Kumar, International Journal for Research in Applied Science & Engineering Technology (IJRASET) Volume 7, Issue IV, Apr 2019

[4]. Heuristic Prediction of Crop Yield using Machine Learning Technique by S. Pavani, Augusta Sophy Beulet P, International Journal of Engineering and Advanced Technology (IJEAT) Volume-9, December 2019 Sadal O.I.,Marwa S.M.,Wala T.A. (2012) "Linear Alkylbenzene Production from Kerosene" Seminar presented to the Department of Chemical engineering University of Khartoum.

[5]. Kale, Shivani S., and Preeti S. Patil. "A Machine Learning Approach to Predict Crop Yield and Success Rate." In 2019 IEEE Pune Section International Conference (PuneCon), pp. 1-5. IEEE, 2019. .

[6]. Kumar, Y. Jeevan Nagendra, V. Spandana, V. S. Vaishnavi, K. Neha, and V. G. R. R. Devi. "Supervised Machine learning Approach for Crop Yield Prediction in Agriculture Sector." In 2020 5th International Conference on Communication and Electronics Systems (ICCES), pp. 736-741. IEEE, 2020..

[7]. Nigam, Aruvansh, Saksham Garg, Archit Agrawal, and Parul Agrawal. "Crop yield prediction using machine learning algorithms." In 2019 Fifth International Conference on Image Information Processing (ICIIP), pp. 125-130. IEEE, 2019.