A PROJECT REPORT ON

# **CROP RECOMMENDATION SYSTEM**

SUBMITTED TO THE SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE

IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

AWARD OF THE DEGREE OF

**BACHELOR OF ENGINEERING (COMPUTER ENGINEERING)**

**SUBMITTED BY:**

**Tushar Pangare**  Exam No: B150374253

**Sujit Hingane** Exam No: B150374226

**Rohan Dhore** Exam No: B150374219

**Gaurav Bhore** Exam No: B150374210



**DEPARTMENT OF COMPUTER ENGINEERING**

**JSPM’s RAJARSHI SHAHU COLLEGE OF ENGINEERING TATHAWADE, PUNE 411033**

**SAVITRIBAI PHULE PUNE UNIVERSITY 2021-22**



# **CERTIFICATE**

This is to certify that the project report entitles

**CROP RECOMMENDATION SYSTEM**

**Submitted by:**

**Tushar Pangare** Exam No: B150374253

**Sujit Hingane** Exam No: B150374226

**Rohan Dhore** Exam No: B150374219

**Gaurav Bhore**  Exam No: B150374210

|  |  |
| --- | --- |
| **Prof. A. H. Hingmire** | **Dr. S. V. Kedar** |
| Internal Guide Dept. of Computer Engineering | Head  Department of Computer Engineering |
| **Dr. R. K. Jain**  Principal  JSPM’s RSCOE | |

are bonafide students of this institute and the work has been carried out by them under the supervision of Prof. **A.H. Hingmire** and it is approved for the fulfillment of the requirement of Savitribai Phule Pune University, for the award of the degree of **Bachelor of Engineering (Computer Engineering).**

Place :Pune

Date: 8/7/2022

**Acknowledgments**

It gives us great pleasure in presenting the preliminary project report on **‘Crop Recommendation System’**. I would like to take this opportunity to thank my internal guide **Prof. Amruta Hingmire** for giving me all the help and guidance I needed. I am really grateful to them for their kind support. Their valuable suggestions were very helpful. I am also grateful to **Dr. S. V. Kedar**, Head of Computer Engineering Department, JSPM’s Rajarshi Shahu College of Engineering for his indispensable support, suggestions, and laboratory with all needed software platforms, continuous internet connection, for our Project.

**Tushar Pangare**

**Sujit Hingane**

**Rohan Dhore**

**Gaurav Bhore**

**Abstract**

In general, agriculture is the backbone of India and also plays an important role in Indian economy by providing a certain percentage of domestic product to ensure the food security. But now-a-days, food production and prediction is getting depleted due to unnatural climatic changes, which will adversely affect the economy of farmers by getting a poor yield and also help the farmers to remain less familiar in forecasting the future crops. This research work helps the beginner farmer in such a way to guide them for sowing the reasonable crops by deploying machine learning, one of the advanced technologies in crop prediction. Random Forest, a supervised learning algorithm puts forth in the way to achieve it. The seed data of the crops are collected here, with the appropriate parameters like temperature, humidity and moisture content, which helps the crops to achieve a successful growth. In addition as the software, a web application is being developed. The users are encouraged to enter parameters like temperature, rainfall, etc. that will be taken automatically in this application in order to start the prediction process.

**Keywords: -** Crop Recommendation, Random Forest, Bootstrapping.

# INDEX

1. Introduction………………………………………………………..1
   1. Motivation………………………………………………………2
   2. Problem Statement……………………………………………...2
2. Literature Survey…………………………………………………..3
   1. Study Of Research Paper . . . . . . . . . . . . . . . . . . . . . . ………….3
3. Software Requirements Specification…………………………….14
   1. Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . ……………..14
      1. Project Scope . . . . . . . . . . . . . . . . . . . . . . . . ………………...14
      2. User Classes and Characteristics . . . . . . . . . . . . . . ………….14
      3. Assumptions and Dependencies . . . . . . . . . . . . . . …………..14
   2. Functional Requirements ………………………………………14
      1. Functional Specification: . . . . . . . . . . . . . . . . . . …………….14
      2. Dependency and Constraints: . . . . . . . . . . . . . . ……………...14
   3. EXTERNAL INTERFACE REQUIREMENT . . . . . . . . . . . . ...15
      1. User Interface . . . . . . . . . . . . . . . . . . . . . . . . ………………...15
      2. Hardware Interfaces: . . . . . . . . . . . . . . . . . . . . ………………15
      3. Software Interfaces . . . . . . . . . . . . . . . . . . . . . ……………….15
   4. NON FUNCTIONAL REQUIREMENT . . . . . . . . . . . . . . ……15
      1. PerformanceRequirements . . . . . . . . . . . . . . . . . ……………..15
      2. Safety Requirement . . . . . . . . . . . . . . . . . . . . . ……………….16
      3. Software Quality Attributes . . . . . . . . . . . . . . . . ……………16
   5. SYSTEM REQUIREMENT . . . . . . . . . . . . . . . . . . . . ………17
      1. Database Requirements : . . . . . . . . . . . . . . . . . . ……………...17
      2. Hardware Requirements: . . . . . . . . . . . . . . . . . . ……………...17
      3. Software Requirements : . . . . . . . . . . . . . . . . . . ………………17
   6. Analysis Models: SDLC Model to be applied . . . . . . . . . . . ……18
   7. System Implementation Plan . . . . . . . . . . . . . . . . . . . . ………….20
4. System Design………………………………………………………21
   1. system Architecture . . . . . . . . . . . . . . . . . . . . . . . . . …………….21
   2. Data Flow Diagram . . . . . . . . . . . . . . . . . . . . . . . . . . …………..22
   3. ER Diagrams . . . . . . . . . . . . . . . . . . . . . . . . . . . . . …………..23
   4. UML DIAGRAMS . . . . . . . . . . . . . . . . . . . . . . . . . . ………….23
5. OTHER SPECIFICATION…………………………………………27
   1. Advantages . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . ………….27
   2. Limitation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . ………….27
   3. APPLICATION . . . . . . . . . . . . . . . . . . . . . . . . . . . ……………..27
   4. Reconciled Estimates…………………………………………….27

6 Project Implementation……………………………………………...29

7. Software Testing…………………………………………………….34

7.1 Type of Testing………. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . … 34

7.2 Test Cases and Test Results………………………………………...35

8. Conclusion and Future Scope………………………………………..40

8.1 Conclusion………………………………………………………….40

9.1 Annexure A…………………………………………………………41

9.2 References …………………………………………………………44

­

List of Figures

3.1 Waterfall Model . . . . . . . . . . . . . . . . . . . . . . . . . . . ……19

* 1. system Architecture . . . . . . . . . . . . . . . . . . . . . . . . 21
  2. Data Flow(0) diagram . . . . . . . . . . . . . . . . . . . . . . . . . 22
  3. Data Flow(1) diagram . . . . . . . . . . . . . . . . . . . . . . . . . 22
  4. Data Flow(2) diagram . . . . . . . . . . . . . . . . . . . . . . . . . 23
  5. ER Diagram . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 23
  6. Classs Diagram . . . . . . . . . . . . . . . . . . . . . . . . . . . . 23
  7. Use case Diagram . . . . . . . . . . . . . . . . . . . . . . . . . . . 24
  8. Activity Diagram . . . . . . . . . . . . . . . . . . . . . . . . . . . 25
  9. Sequence Diagram . . . . . . . . . . . . . . . . . . . . . . . . 26

6.1 Random Forest Algorithm Working…………………….30

6.2 Crop Suggestion ………………………………………...32

6.3 Prediction of crop………………………………………..32

7.1 Confusion Matrix………………………………………..37

7.2 Heatmap…………………………………………………38

7.3 Accuracy Comparison ………………………………….39

**INTRODUCTION**

Weather plays an important role in agriculture production. For optimal productivity at given location crops must be such that their weather requirement match the Current weather system. So we need to plan them according to the weather conditions and soil fertility. This is a prototype for a crop recommendation algorithm in Python using Machine Learning and Data Analytics.

This work presents a system, in form of a website. The business logic in Python uses Machine Learning techniques in order to predict the most profitable crop in the forecasted weather and soil conditions such as pH, nitrogen, phosphorous, potassium, humidity of the soil.  
 The proposed system will integrate the data obtained from soil, crop repository, weather department and by applying machine learning algorithm.

This provides a farmer with variety of options of crops that can be cultivated. Crop prediction is one of the challenging problems in precision agriculture, and an abundance of models are proposed and validated to date.

This problem requires the utilization of several datasets since crop yield depends on many different factors like climate, weather, soil, use of fertilizer, etc. To develop a Crop prediction system. Being a complete software solution, it doesn't allow maintenance factors to be considered much. Also, the accuracy level of the model will be higher when compared to hardware-based solutions, because the components like soil composition, rainfall value, pH value, and weather all get in the picture during the prediction process of the crop.

### **1.1 MOTIVATION**

Modern technologies have enabled human society to produce enough food to feed more than 7 billion people. However, food security is still jeopardized due to a variety of factors such as climate change, pollinator decline, crop plant diseases, and others. Crop Plant diseases not only pose a global threat to food security, but they can also have disastrous consequences for smallholder farmers whose livelihoods rely on healthy crops. Furthermore, the majority of hungry people (50 percent) live in smallholder farming households, making smallholder farmers particularly vulnerable to pathogen-related disruptions in food supply.

### 1.2 **PROBLEM STATEMENT**

Crop prediction is one of the challenging problems in precision agriculture, and many models have been proposed and validated so far. This problem requires the use of several datasets since crop yield depends on many different factors such as climate, weather, and soil, use of fertilizer etc. To develop Soil detection and Crop predication system.

## **2. LITERATURE SURVEY**

### 2.1 STUDY OF RESEARCH PAPER

In Crop Yield Analysis Using Machine Learning Algorithms

Authors Fatin Farhan Haque, Ahmed Abdelgawad, Venkata Prasanth Yanambaka, Kumar Yelamarthi stated that,

Agriculture is not only a huge aspect of the growing economy, but it’s essential for us to survive. Predicting crop yield is not an easy task, as it depends on many parameters such as water, ultra-violet (UV), pesticides, fertilizer, and the area of the land covered for that region. In this paper, two different Machine Learning (ML) algorithms are proposed to analyze the crops’ yield. These two algorithms, Support Vector Regression (SVR) and Linear Regression (LR), are quite suitable for validating the variable parameters in the predicting the continuous variable estimation with 140 data points that were acquired.

The parameters mentioned above are key factors affecting the yield of crops. The error rate was measured with the help of Mean Square Error (MSE) and Coefficient of Determination (R2), where MSE gave out approximately 0.005 and R2 gave around 0.85. The same dataset has been used for quick comparison between the algorithms’ performances.[1]

Prof. A. V. Deorankar in his paper An Analytical Approach for Soil and Land Classification System using Image Processing studied about land mapping.

In the last few decades researchers are interested in land mapping and its classification due to various reasons. The reasons for an increase in the focus of the research community are, the increasing demand for agricultural land and soil health analysis, as the health of the soil, is essential for the healthy production of crops. Image classification is one such approach for soil and land health analysis. It is a complex process having the effects of various factors. This paper has proposed the study of current researches, the problems it addressed, and its prospects. The emphasis is focused on the analytical study of various advanced and efficient classification mechanisms and techniques. Here, it has been attempted to study the factors these approaches have addressed to improve the accuracy of the classification. Proper utilization of the number of features of remotely sensed data and selecting the best suitable classifier are most important for improving the accuracy of the classification. The knowledgebased classification or Non-parametric classifiers like decision tree classifier or neural network have gained more popularity for multisource data classification in recent times. However, there is still the scope of further research, to reduce uncertainties in the improvement of accuracy of the Image classification mechanisms.[2]

Ramesh Medar in Crop Yield Prediction using Machine Learning Techniques says agriculture is the field which plays an important role in improving our countries economy. Agriculture is the one which gave birth to civilization. India is an agrarian country and its economy largely based upon crop productivity. Hence we can say that agriculture can be backbone of all business in our country. Selecting of every crop is very important in the agriculture planning. The selection of crops will depend upon the different parameters such as market price, production rate and the different government policies. Many changes are required in the agriculture field to improve changes in our Indian economy. We can improve agriculture by using machine learning techniques which are applied easily on farming sector. Along with all advances in the machines and technologies used in farming, useful and accurate information about different matters also plays a significant role in it. The concept of this paper is to implement the crop selection method so that this method helps in solving many agriculture and farmers problems. This improves our Indian economy by maximizing the yield rate of crop production.[3]

Author Yogesh Gandge has published a paper, “A Study on Various Data Mining Techniques for Crop Yield Prediction”, in which he has studied the different affecting factors on crops in Indian soil.[4]

India is a country where agriculture and agriculture related industries are the major source of living for the people. Agriculture is a major source of economy of the country. It is also one of the country which suffer from major natural calamities like drought or flood which damages the crop. This leads to huge financial loss for the farmers thus leading to the suicide. Predicting the crop yield well in advance prior to its harvest can help the farmers and Government organizations to make appropriate planning like storing, selling, fixing minimum support price, importing/exporting etc. Predicting a crop well in advance requires a systematic study of huge data coming from various variables like soil quality ,pH ,EC,N,P,K etc. As Prediction of crop deals with large set of database thus making this prediction system a perfect candidate for application of data mining. Through data mining we extract the knowledge from the huge size of data. This paper presents the study about the various data mining techniques used for predicting the crop yield. The success of any crop yield prediction system heavily relies on how accurately the features have been extracted and how appropriately classifiers have been employed. This paper summarizes the results obtained by various algorithms which are being used by various authors for crop yield prediction, with their accuracy and recommendation.[4]

In “CROP PREDICTION USING PREDICTIVE ANALYTICS”

Author P. S. Vijayabaskar, Sreemathi.R, Keertanaa.E showed the effects of fertilizers on soil.[5]

This work is to construct a model for testing the soil fertility. It also suggests the crop which has to be planted depending upon the value obtained from the sensor. It also provides the regional wise information about the crop in the form of graph. We have farmer chat where the farmers can share and get idea from the expert by registering in this application. It also suggests the fertilizer which has to be added to the soil in order to increase the crop productivity. It helps the farmer to analyze the fertility of their yard and plant the better crop to increase their productivity and profit. It also provides the information about the fertilizer to be added in the soil and also provide the information about the nearby fertilizer shop.

In paper Real-Time Monitoring of Agricultural Land with Crop Prediction and Animal Intrusion Prevention using Internet of Things and Machine Learning at Edge author Nikhil R studied the dependency of crops on soil parameters.[6]

Agriculture is considered as a foundation of life, since it is the primary source of food and other raw materials. It plays a crucial part in the country’s economic development. Sadly, many of our farmers cultivate their land using the conventional methods. we should replace these obsolete techniques of farming with advanced techniques. The proposed system describes how the use of the IOT and ML techniques can be combined to make the irrigation smart. The proposed system saves time avoiding problems like constant vigilance over the field by using IOT devices, crop prediction helps the farmers to grow suitable crops depending on the soil parameters by the use of machine learning techniques and it also helps in prevention of the intruders like wild animals into the field. It also helps in water conservation by supplying the plants / field with minimal amount of water automatically through the help of sensors depending on the water requirements. and finally, SMS and Email notifications will be sent to the farmer mobile phone during the abnormal conditions of his farm. The proposed system can be used for taking edge decisions in real time.[6]

Author Javier E. Sanchez-Gal in “Supervised Classification of Spectral Signatures from Agricultural Land-Cover in Panama Using the Spectral Angle Mapper Algorithm” studied the supervised classification of the agricultural.

In this article the development of a database of referenced spectral signatures from agricultural land-cover for the Republic of Panama is presented. This database consists of reflectance spectra measured on crops and low vegetation, such as: rice, chili, onion, watermelon, maize and bare soil and of satellite images of their plots. Details of the integration process of the database and software developed for the manipulation of spectral signatures, are described. The Spectral Angle Mapping algorithm (SAM) is used for the supervised classification of the agricultural coverages in the database. On the one hand, results indicate the possibility of using this classification technique for the automatic determination of crops and even different phenological stages in a crop via a satellite image. On the other hand, results highlight the limitations of using this technique on recently planted crops and soil flooded by rain or with soil cultivated with a low agricultural cover crop. We foresee the use of this methodology and database for agricultural land surveys, crop management or used in the general organization of the territory. [7]

In “Soil Classification and Crop Suggestion using Image Processing” authors T. Abimala, S. Flora Sashya and K. Sripriya researched on use of image processing in crop prediction using feature extraction.

This paper is intended to support agriculture by classifying 7 different types of soils like Clay, Clayey Peat, Clayey Sand, Humus Clay, Peat, Sandy Clay and Silty Sand, and in suggesting suitable crops that could be grown in those particular soils using image processing. Pre-processing is done by using Low Pass filter. HSV, GLCM, Gabor Wavelet algorithms are used for feature extraction. HSV, GLCM are used to perform colour based feature extraction. Gabor filters are used to perform texture based feature extraction. The features obtained from the test image are then compared with the features obtained from the images in the dataset. Matching of image features is achieved by training the Decision Tree classifier with statistical measurements like mean, standard deviation, skew and kurtosis. Finally the soil is predicted with the help of segmented images that are given as input for simulation using Matlab R2018a and is followed by crop suggestion.[8]

Author Dr. Y. Jeevan Nagendra Kumar1 in their paper “Supervised Machine learning Approach for Crop Yield Prediction in Agriculture Sector” the impact of different factors such as temperature, ph, rainfall on crop forecasting.

Machine learning (ML) is a crucial perspective for acquiring real-world and operative solution for crop yield issue. From a given set of predictors, ML can predict a target/outcome by using Supervised Learning. To get the desired outputs need to generate a suitable function by set of some variables which will map the input variable to the aim output. Crop yield prediction incorporates forecasting the yield of the crop from past historical data which includes factors such as temperature, humidity, ph, rainfall, crop name. It gives us an idea for the finest predicted crop which will be cultivate in the field weather conditions. These predictions can be done by a machine learning algorithm called Random Forest. It will attain the crop prediction with best accurate value. The algorithm random forest is used to give the best crop yield model by considering least number of models. It is very useful to predict the yield of the crop in agriculture sector. [9]

Shilpa Pande, Prem Kumar Ramesh, and Anmol developed a crop detector using a Machine Learning approach. They tried and tasted different Machine Learning algorithms like Support Vector Machine (SVM), Artificial Neural Network (ANN), Random Forest (RF), Multivariate Linear Regression (MLR), and K-Nearest Neighbour (KNN). Among these used algorithms they got the best results with random forest. [10]

**3. SOFTWARE REQUIREMENTS SPECIFICATION**

### 3.1 INTRODUCTION

#### 3.1.1 Project Scope

Crop yields and crop yield forecasts directly affect the annual national and international economy and play a major role in the food economy. Crop yields are highly dependent on irrigation and climate data. More irrigation does not necessarily increase yield, and therefore, optimization of irrigation and more efficient irrigation systems are critical. Predicting yield based on different types of irrigation is one way to optimize the process.

#### 3.1.2 User Classes and Characteristics

User Can give the inputs in our system then predict the suggested crops based on the various factors.

#### 3.1.3 Assumptions and Dependencies

Assumptions : We have used Python Technique. Input as Dataset and Weather.

Dependencies : We have used python libraries like Numpy, Pandas, and Sklearn. Output to detect Suggested Crops .

### 3.2 FUNCTIONAL REQUIREMENTS

3.2.1 Functional Specification:

* The application is user friendly.
* It provides an easy interface to user.
* The accessibility or response time of the application should be fast.

3.2.2 Dependency and Constraints:

* End User application will be developed in Windows OS.
* All scripts shall be written in Python
* Application design pattern shall be Singleton.

### 3.3 EXTERNAL INTERFACE REQUIREMENT

#### 3.3.1 User Interface

Application Based on Soil based crop prediction and weather forecasting.

3.3.2 Hardware Interfaces:

RAM : 8 GB

As we are using Machine Learning Algorithm and Various High Level Libraries Laptop.

RAM minimum required is 8 GB.

Hard Disk : 40 GB

Data Set of Crops is to be used hence minimum 40 GB Hard Disk memory

is required.

Processor: Intel i5 Processor

#### 3.3.3 Software Interfaces

Operating System: Windows 10 (64 Bit)

IDE: VS Code

Programming Language: Python

### 3.4 NON-FUNCTIONAL REQUIREMENT

#### 3.4.1 Performance Requirements

The performance of the functions and every module must be well. The overall performance of the software will enable the users to work efficiently. Performance of encryption of data should be fast. Performance of the providing virtual environment should be fast Safety Requirement

The application is designed in modules where errors can be detected and xedeasily. This makes it easier to install and update new functionality if required.

#### 3.4.2 Safety Requirement

The application is designed in modules where errors can be detected and fixed easily.

This makes it easier to install and update new functionality if required.

#### 3.4.3 Software Quality Attributes

Our software has many quality attribute that are given below:Adaptability: This software is adaptable by all users.

Availability: This software is freely available to all users. The availability of the software is easy for everyone.

Maintainability: After the deployment of the project if any error occurs then it can be easily maintained by the software developer.

Reliability: The performance of the software is better which will increase the reliabilityof the Software.

User Friendliness: Since, the software is a GUI application; the output generated is much user friendly in its behavior.

Integrity: Integrity refers to the extent to which access to software or data by unauthorized persons can be controlled.

Security: Users are authenticated using many security phases so reliable security is provided.

Testability: The software will be tested considering all the aspects.

### 3.5 SYSTEM REQUIREMENT

3.5.1 Database Requirements :

* Database : DBSqlite

3.5.2 Hardware Requirements:

* Hardware : intel core
* Speed : 2.80 GHz
* RAM : 8GB
* HardDisk : 500 GB
* Key Board: Standard Windows Keyboard

3.5.3 Software Requirements :

* Operating System: Windows 10(64 Bit)
* IDE: Visual Studio Code
* Programming Language : python version 3.7,3.8.,HTML,CSS
* Libraries: Numpy,Pandas,Sklearn

### 3.6 ANALYSIS MODELS: SDLC MODEL TO BE APPLIED

SDLC Models stands for Software Development Life Cycle Models. In this article, we explore the most widely used SDLC methodologies such as Agile . Each software development life cycle model starts with the analysis, in which the Also, here are defined the technologies used in the project, team load. One of the basic notions of the software development process is SDLC models which stands for Software Development Life Cycle models. SDLC – is a continuous process, which starts from the moment, when it’s made a decision to launch the project, and it ends at the moment of its full remove from the exploitation. There is no one single SDLC model. They are divided into main groups, each with its features and weaknesses.

1. Requirement Analysis - Requirement Analysis is the most important and necessary stage in SDLC. The senior members of the team perform it with inputs from all the stakeholders and domain experts or SMEs in the industry. Planning for the quality assurance requirements and identifications of the risks associated with the projects is also done at this stage. Business analyst and Project organizer set up a meeting with the client to gather all the data like what the customer wants to build,

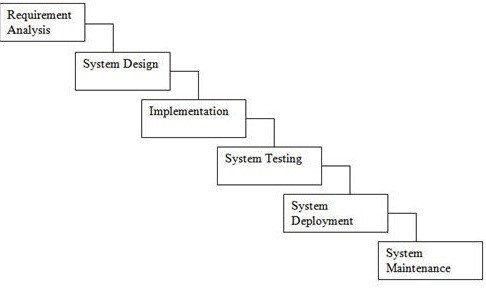


Figure 3.1: Waterfall Model

who will be the end user, what is the objective of the product. Before creating a product, a core understanding or knowledge of the product is very necessary.

1. System Design - The next phase is about to bring down all the knowledge of requirements, analysis, and design of the software project. This phase is the product of the last two, like inputs from the customer and requirement gathering.
2. Implementation - In this phase of SDLC, the actual development begins, and the programming is built. The implementation of design begins concerning writing code. Developers have to follow the coding guidelines described by their management and programming tools like compilers, interpreters, debuggers, etc. are used to develop and implement the code.
3. Testing - After the code is generated, it is tested against the requirements to make sure that the products are solving the needs addressed and gathered during the requirements stage. During this stage, unit testing, integration testing, system testing, acceptance testing are done.
4. Deployment - Once the software is certified, and no bugs or errors are stated, then it is deployed. Then based on the assessment, the software may be released as it is or with suggested enhancement in the object segment. After the software is deployed, then its maintenance begins.
5. Maintenance - Once when the client starts using the developed systems, then the real issues come up and requirements to be solved from time to time. This procedure where the care is taken for the developed product is known as maintenance.

### 3.7 SYSTEM IMPLEMENTATION PLAN

The System Implementation plan table, shows the overall schedule of tasks compilation and time duration required for each task.

|  |  |  |  |
| --- | --- | --- | --- |
| Sr. No. | Name/Title | Start Date | End Date |
| 1 | Preliminary Survey | 18th June 20201 | 25th June 20201 |
| 2 | Introduction and Problem Statement | 26th June 2021 | 4th July 2021 |
| 3 | Literature Survey | 4th August 2021 | 7th Nov 2021 |
| 4 | Project Statement | 10th Nov 2021 | 15 Nov 2021 |
| 5 | Software Requirement And Specification | 16th Nov 2021 | 28 Nov 2021 |
| 6 | System Design | 1 Dec 2021 | 26 Dec 2021 |
| 7 | Partial Report Submission | 5th Jan 2022 | 6th Jan 2022 |
| 8 | Architecture Design | 7th Jan 2022 | 15 Jan 2022 |
| 9 | Implementation | 16 Jan 2022 | 1st March 2022 |
| 10 | Testing | 2nd March 2022 | 15 March 2022 |
| 11 | Deployment | 16th March 2022 | 20th March 2022 |
| 12 | Paper Publish | 28 March 2022 | 6 April 2022 |
| 13 | Report Submission | 7 April 2022 | 23 July 2022 |

## **4. SYSTEM DESIGN**

### 4.1 SYSTEM ARCHITECTURE

Here we have crop dataset. The dataset is first cleaned using preprocessing techniques. After that feature extraction is done. Model is trained using different decision trees. All these decision trees are trained using separate dataset using bootstrapped dataset where duplication is allowed. In the testing phase random forest algorithm takes majority voting for prediction purpose in the prediction. In such a way system works.

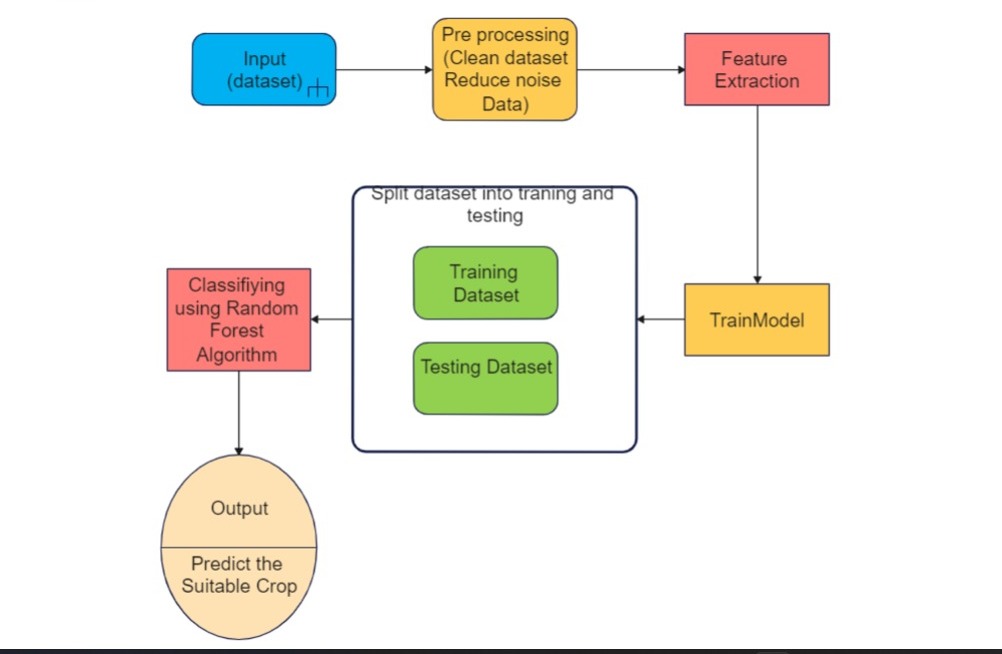


Figure 4.1: system Architecture

### 4.2 DATA FLOW DIAGRAM

In Data Flow Diagram,we Show that flow of data in our system in DFD0 we show that base DFD in which rectangle present input as well as output and circle show our system,In DFD1 we show actual input and actual output of system input of our system is text or image and output is rumor detected like wise in DFD 2 we present operation of user as well as admin.



Figure 4.2: Data Flow(0) diagram

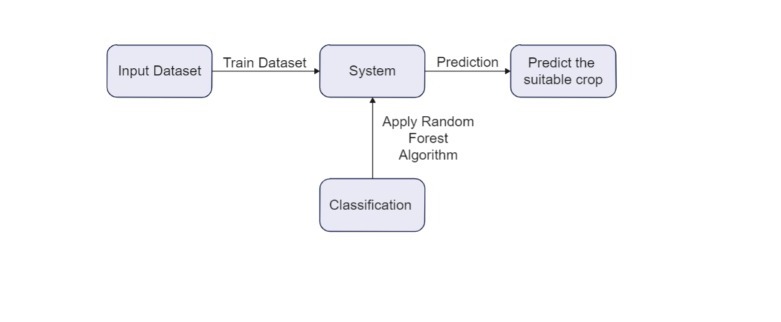


Figure 4.4: Data Flow(1) diagram

### 4.4 UML DIAGRAMS

Unified Modeling Language is a standard language for writing software blueprints. The UML may be used to visualize,specify,construct and document the artifacts of a softwareintensive system.UML is process independent,although optimally it should be used in process that is use case driven,architecture-centric,iterative,and incremental.The Number of UML Diagram is available.

4.4.1 Class Diagram:

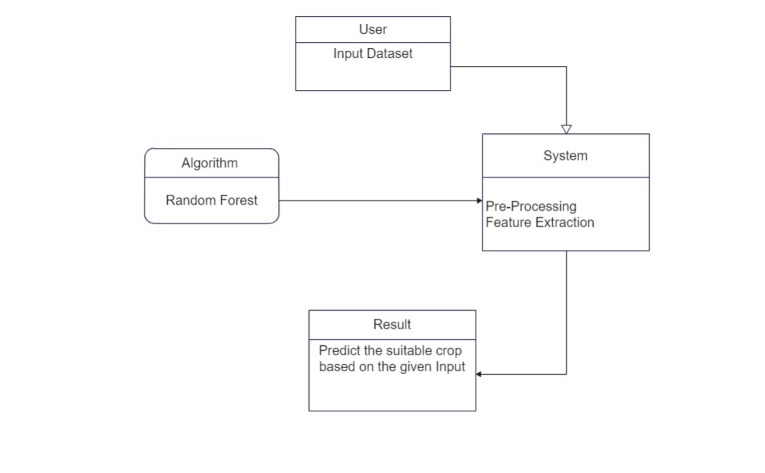


Figure 4.6: Class Diagram

4.4.2 Use Case Diagram:

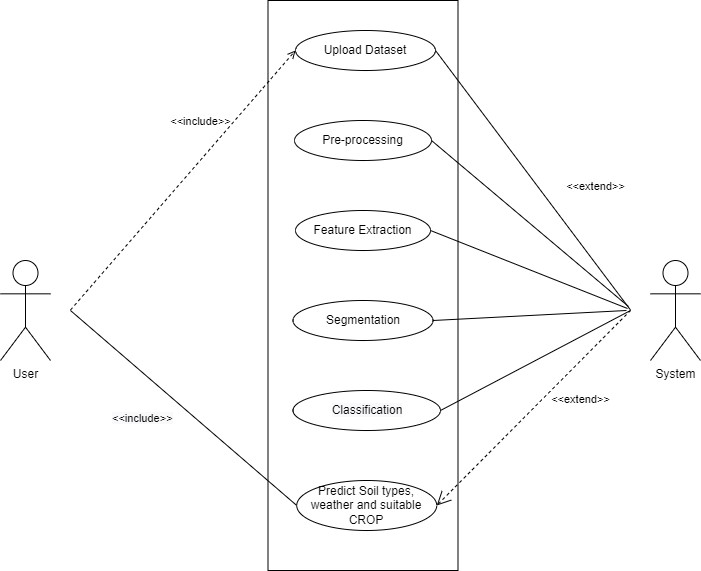


Figure 4.7: Use case Diagram

4.4.3 Activity Diagram:

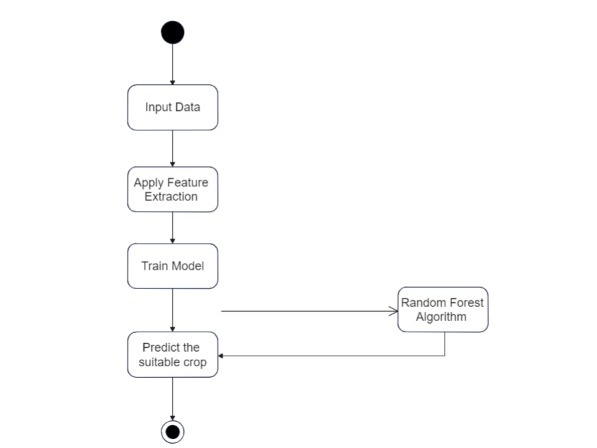


Figure 4.8: Activity Diagram

4.4.4 Sequence Diagram:

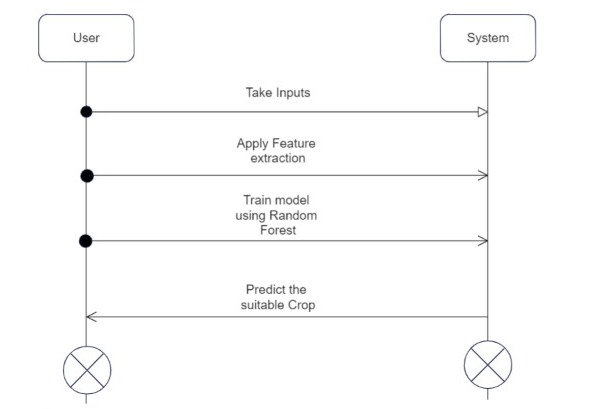


Figure 4.9: Sequence Diagram

## **5. OTHER SPECIFICATION**

### 5.1 ADVANTAGES

Predicting productivity of crop in various climatic conditions can help farmer and other partners in essential basic leadership as far as agronomy and product decision. This model can be used to select the most excellent crops for the region and also its yield thereby improving the values and gain of farming also.

### 5.2 LIMITATION

Results showed that barriers to effective simulations exist because, in most instances, the input data, like climate, soil, farm management practices, and cultivar characteristics, were generally incomplete, poor in quality, and not easily accessible or usable.

### 5.3 APPLICATION

yield prediction is an essential task for the decision-makers at national and regional levels (e.g., the EU level) for rapid decision-making. An accurate crop yield prediction model can help farmers to decide on what to grow and when to grow. There are different approaches to crop yield prediction.

5.4 RECONCILED ESTIMATES

Project is based on software, it is required to implement on our personal system. Cost estimation of project beforehand plays an important role in project management lifecycle. While designing our project we keep following things in mind:

1. What is the size of product to be developed?
2. Number of hours to be spend by persons on the project.
3. What should be the schedule of project in calendar months?
4. What are goals for the product to be delivered?
5. What is time and space complexity?
6. How is project broken into small chunks

## **6.PROJECT IMPLEMENTATION**

## 6.1: Algorithm Details: Random Forest

Random forest is a *Supervised Machine Learning Algorithm* that is *used widely in Classification and Regression problems*. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression.

It uses bagging technique. Bagging, also known as *Bootstrap Aggregation* is the ensemble technique used by random forest.Bagging chooses a random sample from the data set. Hence each model is generated from the samples (Bootstrap Samples) provided by the Original Data with replacement known as *row sampling*. This step of row sampling with replacement is called*bootstrap*.

Now each model is trained independently which generates results. The final output is based on majority voting after combining the results of all models. This step which involves combining all the results and generating output based on majority voting is known as *aggregation*.

Steps involved in random forest algorithm:

Step 1: In Random Forest n number of random records are taken from the data set having k number of records.

Step 2: Individual decision trees are constructed for each sample.

Step 3: Each decision tree will generate an output.

Step 4: Final output is considered based on *Majority Voting or Averaging*for Classification and regression respectively.

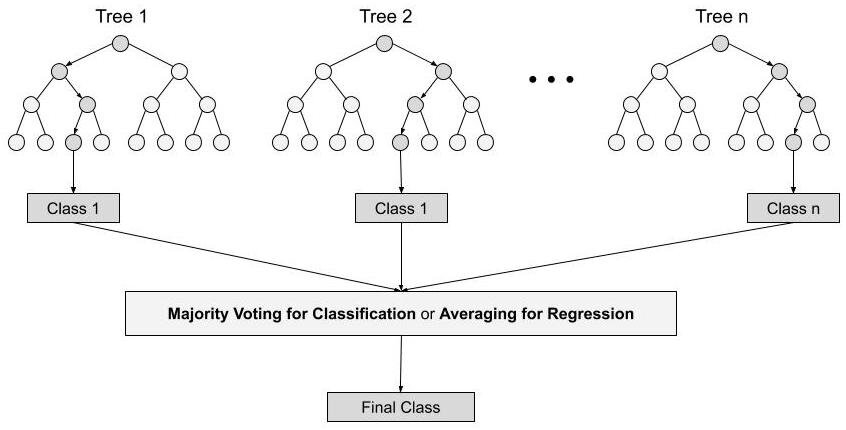


Fig 6.1: Random Forest Algorithm Working

## 6.2 Modules:

## 6.2.2 Module 1: Here we have taken the raw dataset and then using pandas duplicate values and null values are removed for optimal accuracy of the algorithm. Here missing data is handled. Missing Data can occur when no information is provided for one or more items or for a whole unit. Missing Data is a very big problem in a real-life scenarios Here duplicate data is removed using the Pandas library which is available in Machine Learning. Having duplicate data can give rise to Inefficiency and Lack of Productivity. Sometimes it will have inaccurate reporting and less informed decisions.

## 6.2.2 Module 2: Train Model

## Here to recommend a suitable crop we have used Random Forest Algorithm. It could be a popular machine learning algorithm that belongs to the supervised learning technique. Feature Extraction(Similar features are extracted from the dataset for Training Model). The different features used for the algorithm are Nitrogen, Potassium, Phosphorous, Temperature, Humidity, pH, and Rainfall value. Using these features random forest classifier is trained. Training Model and splitting it into Training Dataset and Testing Dataset. Now using this training dataset Random Forest could be a classifier that is trained on a variety of decision trees on various subsets of the given dataset and takes the typical to boost the predictive accuracy of that dataset.

6.3 Module 3: Suggestion of Crop

Classifying model using Random Forest Algorithm and Predicting the suitable crop. We were able to successfully predict the best suitable crop for that particular soil and weather conditions using Random Forest Algorithm. Using this algorithm we got an accuracy of more than 95%.

Diagram

Description automatically generated

Fig 6.2 : Crop Suggestion

6.4 Output Screenshots:

## 

Fig 6.3: taking user input

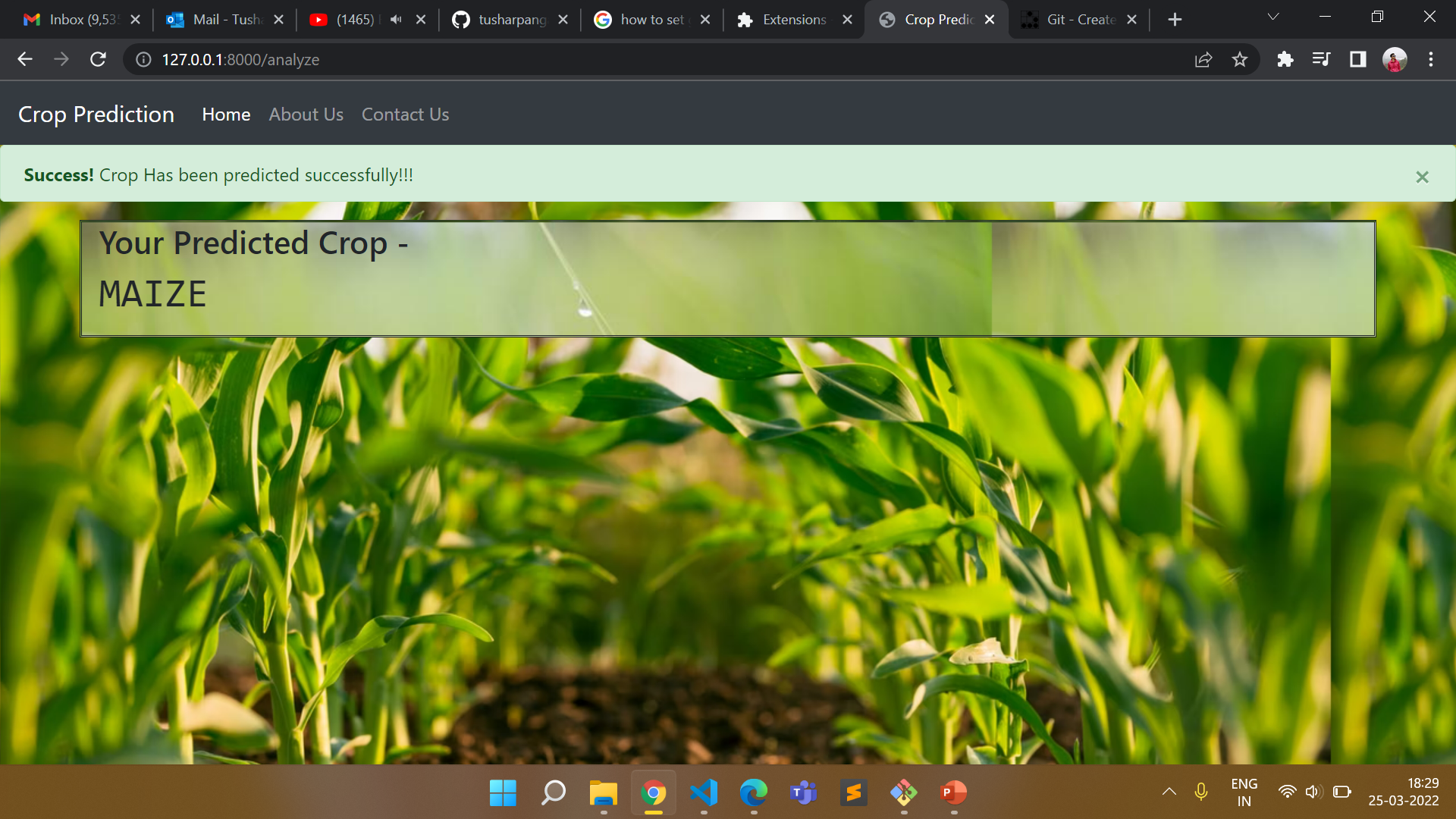


Fig 6.4: prediction of the crop

## **7. Software Testing**

7.1 Type of Testing

1. Unit testing: validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done afterthe completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

2. Integration testing: Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

3. Functional test: Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals. Functional testing is centered on the following items:

• Valid Input : identified classes of valid input must be accepted.

• Invalid Input: identified classes of invalid input must be rejected.

• Functions: identified functions must be exercised.

• Output: identified classes of application outputs must be exercised.

• Systems/Procedures: interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic overage pertaining to identify Business process flows; data fields, pre- defined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

4. System Test: System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration- oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

5. White Box Testing: White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure andlanguage of the software, or at least its purpose. It is purpose. It is used to test areas that cannot be reached from a black box level.

6. Black Box Testing: Black Box Testing is testing the software with-out any knowledge of the inner workings, structure or language of themodule being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirementsdocument. It is a testing in which the software under test is treated, as a black box. You cannot “see” into it. The test provides inputs and responds to outputs without considering how the software works.

7.2 Test Case and Test Results

1)Unit Testing: Unit testing is usually conducted as part of a combined code and unit test phase of the software lifecycle, although it is not un-common for coding and unit testing to be conducted as two distinct phases. Test strategy and approach: Field testing will be performed manually and functional tests will be written in detail.

Test objectives :

• All field entries must work properly.

• Pages must be activated from the identified link.

• The entry screen, messages and responses must not be delayed.

Features to be tested :

• Verify that the entries are of the correct format

• No duplicate entries should be allowed

• All links should take the user to the correct page.

2)Integration Testing: Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects. The task of the integration test is to check that components or software applications, e.g. components in a software system or one step up software applications at the company level interact without error. Test Results: All the test cases mentioned above passed successfully. No defects encountered.

3)Acceptance Testing: User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

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

Test cases: Testing of project problem statement using generated test data (using mathematical models, GUI, Function testing principles, if any) selection and appropriate use of testing tools, testing of UML diagram’s reliability.

**Confusion Matrix:**

Confusion matrix is calculated for the actual and the predicted class and which evaluates the performance of our model.

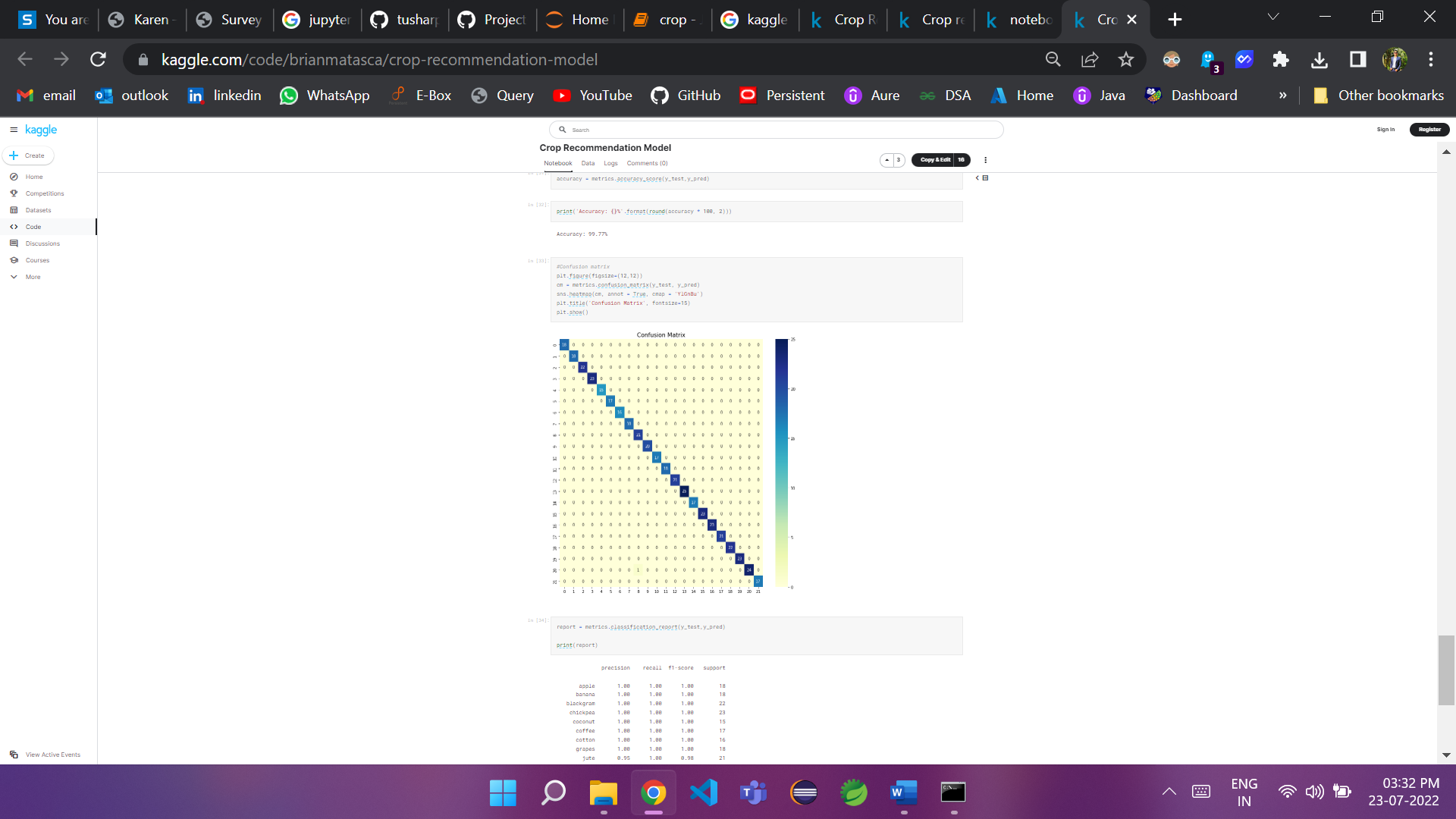


Fig 7.1: Confusion Matrix

## **Heatmap:**

Heatmap is a way to show some sort of matrix plot. To use a heatmap the data should be in a matrix form. By matrix we mean that the index name and the column name must match in some way so that the data that we fill inside the cells are relevant. Lets look at an example to understand this better.

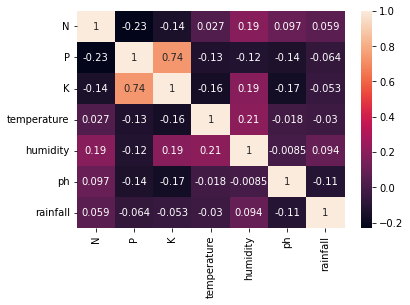


Fig 7.2 Heatmap

**Accuracy:**

We tested different algorithms and got better accuracy with random forest. Also, it handles huge amount of data fast.



Fig 7.3: Accuracy Comparison

## **8. CONCLUSION AND FUTURE SCOPE**

### 8.1 CONCLUSION

A model is proposed for predicting and providing suitable crop yield suggestion for that specific soil and weather. The model has been tested by applying different kinds of Deep algorithm. RANDOM FOREST shows highest accuracy in soil classification and suggests crops with less time. It gives us more accuracy as compared to existing system and gives more benefit to farmers.

In reference to rainfall can depict whether extra water availability is needed or not. This research work can be enhanced to higher level by availing it to whole India. Crop diseases detection using Image Processing where users can upload picture of diseased crop and get pesticides recommendations. Implementation of Smart Irrigation System to monitor weather and soil conditions, plant water usage etc. to automatically alter watering schedule.

8.2 FUTURE SCOPE

The pesticide provider as well as soil nutrition provider can use our system to detect the crops in their area and plan their sells accordingly.

ANNEXUREA

NP-Hard NP-Complete:

What is P?

* P is set of all decision problems which can be solved in polynomial time by

a deterministic.

* Since it can be solved in polynomial time, it can be verified in polynomial

time.

* Therefore P is a subset of NP.

P : To identify road condition or road survey requires more man power, time and money. To resolve these problems we need effective system.

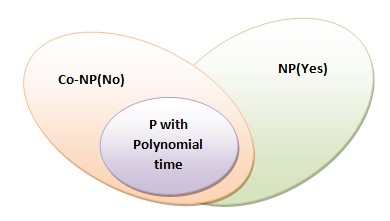


Figure A.1: P Problem

What is NP?

NP means we can solve it in polynomial time if we can break the normal rules of step-by-step computing.

What is NP Hard?

A problem is NP-hard if an algorithm for solving it can be translated into one for solving any NP-problem (nondeterministic polynomial time) problem. NP-hard therefore means ”at least as hard as any NP-problem,” although it might, in fact, be harder.

NP-Hard:

Propose system analyze the road condition and road surface. It identify bad road patches and gives notification to navigation system. For that we used inbuilt accelerometer sensor and gyroscope sensor. To improve the system result we use decision tree algorithm. Propose system has self-managing database which collect data from vehicle drivers smart phones. This data update . Application utilizes this data to inform other application users about road condition.

So here in this case the ‘P’ problem is NP hard. i.e. P=NP-Hard

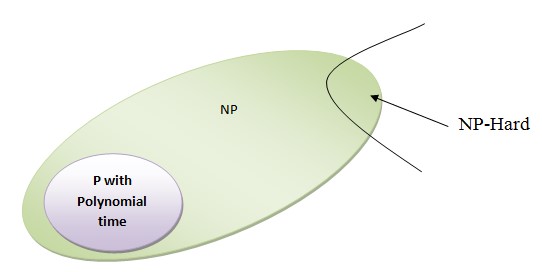


Figure A.2: NP Problem

What is NP-Complete?

* Since this amazing ”N” computer can also do anything a normal computer can, we know that ”P” problems are also in ”NP”.
* So, the easy problems are in ”P” (and ”NP”), but the really hard ones are \*only\* in ”NP”, and they are called ”NP-complete”.
* It is like saying there are things that People can do (”P”), there are things that Super People can do (”SP”), and there are things \*only\* Super People can do (”SP-complete”).

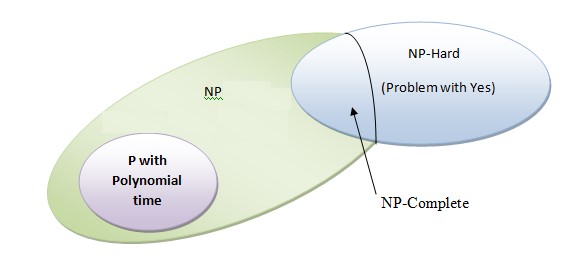


Figure A.3: NP Complete Problem

NP-Complete:

We have used inbuilt mobile sensor to identify road conditions.

Hence the ‘P’ is NP-Complete in this case

## **REFERENCES**

1. Arun Kumar, Naveen Kumar, Vishal Vats, “Efficient Crop Yield Prediction Using

Machine Learning Algorithms”, International Research Journal of Engineering &

Technology (IRJET)- e-ISSN: 2395-0056, pISSN:2395-0072, Volume: 05 Issue: 06

— June-2018

1. Nithin Singh saurabh chaturvedi, “Weather Forecasting Using Machine Learning”, 2019 International Conference on Signal Processing and Communication (ICSC) Volume: 05 — DEC-2019.
2. Aakash Parmar Mithila Sompura, ”Rainfall Prediction using Machine Learning”,

2017 International Conference on (ICIIECS) at Coimbatore Volume: 3 — March 2017.

1. Sachee Nene Priya, R “Prediction of Crop yield using Machine Learning”, International Research Journal of Engineering and Technology (IRJET) Volume: 05 Issue: 02 — Feb-2018.
2. Ramesh Medar Anand M. Ambekar, “Sugarcane Crop prediction Using Supervised Machine Learning” published in International Journal ofIntelligent Systems and Applications Volume: 3 — August 2019.
3. Andrew Crane Droesch, “Machine learning methods for crop yield prediction and climate change impact assessment in agriculture”, Published by IOP Publishing Ltd Volume: 05 — OCT -2018.
4. Vinita Shah Prachi Shah, ”Groundnut Prediction Using Machine Learning Techniques “,published in IJSRCSEIT. UGC Journal No : 64718 — March-2020.
5. Renuka Sujata Terdal, ”Evaluation of Machine Learning Algorithms for Crop Prediction” Published in International Journal of Engineering and Advanced Technology (IJEAT) Volume-8 — August 2019.
6. P. Vinciya, Dr. A. Valarmathi, “Agriculture Analysis for Next Generation High

Tech Farming in Data Mining” IJARCSSE,vol. 6, Issue 5, 2016

1. Shivnath Ghosh,Santanu Koley, “Machine Learning for Soil Fertility and Plant

Nutrient Management using Back Propagation Neural Networks” IJRITCC, vol. 2, Issue 2,292-297,2014.