A PROJECT REPORT ON

AN APPLICATION OF DATA ANALYTICS IN AGRICULTURE SECTOR FOR MULTI-ADVICE GENERATOR IN NATIVE LANGUAGE
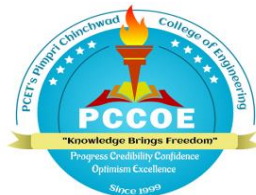
SUBMITTED TO THE SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE
IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE

OF

BACHELOR OF ENGINEERING (COMPUTER ENGINEERING)

SUBMITTED BY

MOHIT DALVI	**Exam No :71710743E**
SAMIHAN DESHMUKH	**Exam No :71710761C**
DEVESH DHANNAWAT	**Exam No :71710772J**
PRATIK GAWLI	**Exam No : 71818855B**

DEPARTMENT OF COMPUTER ENGINEERING

PCET'S PIMPRI CHINCHWAD COLLEGE OF ENGINEERING

SECTOR NO. 26, PRADHIKARAN, NIGDI, PIMPRI-CHINCHWAD, PUNE 411044

SAVITRIBAI PHULE PUNE UNIVERSITY

2019 -2020

CERTIFICATE

This is to certify that the project report entitles

**AN APPLICATION OF DATA ANALYTICS IN AGRICULTURE SECTOR FOR
MULTI-ADVICE GENERATOR IN NATIVE LANGUAGE**

Submitted by

MOHIT DALVI	**Exam No** :**71710743E**
SAMIHAN DESHMUKH	**Exam No** :**71710761C**
DEVESH DHANNAWAT	**Exam No**: **71710772J**
PRATIK GAWLI	**Exam No** :**71818855B**

is a bonafide student of this institute and the work has been carried out by him/her under the supervision of **Prof. S R Vispute** and it is approved for the partial fulfillment of the requirement of Savitribai Phule Pune University, for the award of the degree of **Bachelor of Engineering** (Computer Engineering).

(Prof. Sushma R Vispute)   **(Prof. Dr. K. Rajeswari)**
Guide   Head,
Department of Computer Engineering   Department of Computer
Engineering

(Dr. G. V. Parishwad)
Principal,
Pimpri Chinchwad College of Engineering Pune – 44

Place: Pune
Date :14/06/2020

ACKNOWLEDGEMENT

ABSTRACT

Farmer suicide is one of the major concerns of India. Around 60% to 70% population is directly or indirectly dependent on Agriculture as their source of income. Most of the farmers in India fail to produce sufficient yield due to various uncertainties and problems faced in agriculture. This is the reason farmers take loans, fall into debts and end up giving up their lives. Technological advancements have not been able to make a significant impact in guiding farmers in improving soil health as per crop requirements, crop selection, fertilizers selection, pesticide selection, etc.in order to ensure proper yield. Also, the current systems generate advice based only on single feature of agriculture. To overcome the limitations of the present system, this paper proposes a multi advice generator that takes into account various soil health parameters and crop district based on which multiple advice related to soil health, fertilizer and pesticide selection for respective crops is provided to farmers in native language. Various machine learning algorithms are studied and implemented for accurate soil health and advice generation.

TABLE OF CONTENTS

1 INTRODUCTION

The main objective of the proposed work is to develop a multi advice generator that will assist farmers in their farming. This system will generate multiple advice in native language of farmers by taking into account various soil conditions and crop peculiarity of the respective area.

There are various technological advances coming up in agriculture but a very few have been able to make a significant impact to farming in order to assure proper crop yield. A major problem faced by Indian farmers is lack of proper assistance in farming that is customized to the soil and climatic conditions of their specific locality. Many farmers in India fail to generate sufficient crop produce due to lack of timely guidance and various uncertainties which arise during the course of farming.

Around 70% of Indian population is directly or indirectly dependent on agriculture for their livelihood. Agriculture majorly depends on 2 factors; one is soil fertility and the other is climatic conditions. Our system takes into account the soil conditions like the NPK values and its pH, district and the crop calendar. Based on this the system generates multiple advice in farmers' native language suggesting him deficient amount of NPK content in soil, pesticides and fungicides for a particular crop.

1.1 OVERVIEW

The main objective of the proposed work is to develop a multi advice generator that will assist farmers in their farming. This system will generate multiple advice in native language of farmers by taking into account various soil conditions and crop peculiarity of the respective area.

Various systems have been made which give farmers advice based only on one particular feature. We have made an attempt to take into account various features and develop a multiple advice generator to assist farmers that would have a significant impact to their crop produce. With the help of various machine learning algorithms, our system is able to predict soil NPK content with accurate precision and suggest various advises like selecting appropriate fertilizer and pesticides for respective crops. This system will cater to the needs of various farmers in rural areas. Apart from this the resulting collection of database will serve as a valuable resource for future use.

1.2 MOTIVATION

Farmer suicide in India has always continued to be a major area of concern. This national catastrophe is an outcome of various farmers committing suicide since the 1990s, often by drinking pesticides, due to their inability to repay loans mostly taken from landlords and banks. As of 2014, in Maharashtra alone, more than 60,000 suicides had taken place, with an average of 10 suicides every day. The National Crime Records Bureau of India reported that a total 296,438 Indian farmers had committed suicide since 1995. Of these, 60,750 farmer suicides were in the state of Maharashtra since 1995, with the remainder spread out in Odisha, Telangana, Andhra Pradesh, Madhya Pradesh, Gujarat and Chhattisgarh, all states with loose financial and entry regulations. India is an agrarian country with around 70% of its people depending directly or indirectly upon agriculture. Farmer suicides account for 11.2% of all suicides in India. Activists and scholars have offered a number of conflicting reasons for farmer suicides, such as high debt burdens, poor government policies, corruption in subsidies, crop failure, public mental health, personal issues and family problems.

1.3 PROBLEM DEFINITION

The main objective of the proposed system is to come up with a multi advice generator in native language that will generate advice related to soil health, crops, fertilizers and pesticides in order to help farmers grow their crops efficiently.

1.4 PROJECT SCOPE AND LIMITATION
- Generate multiple advice system customized to farmer's soil health ad crop peculiarity.
- To learn, implement and test various machine learning models in accordance to the data.
- Generate advice in native language and get rid of the language barrier.
- To make the model accessible from remote places.
- Make the system more accurate and precise as system learns from data.
- To be able to generate advice for variety of soil and crop features.

1.5 METHODOLOGIES OF PROBLEM SOLVING

There are various methodologies to solve this particular problem. The most suitable and efficient methodology pertaining to the following problem is used.

Iterative Approach:

In an iterative model, initial development takes place by starting with some of the software specifications and gathering some requirement. After the first version changes are made iteratively and the model is improved. Every release of the Iterative Model is a new version which is better than the previous version. The Iterative Model allows the accessing earlier phases, in which the variations made respectively.
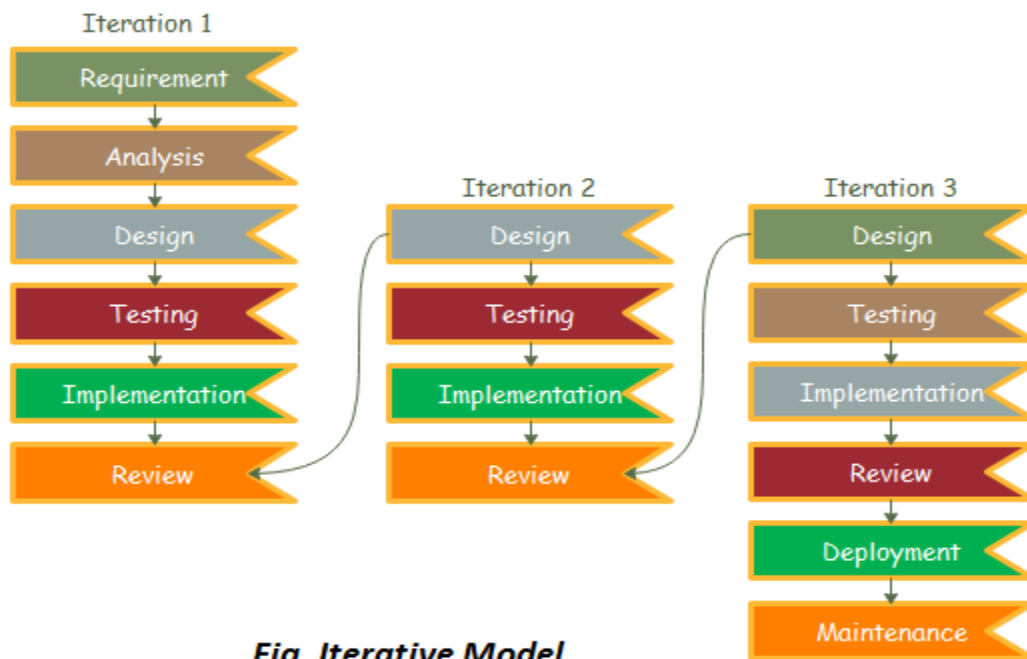
Fig. Iterative Model

The various phases of Iterative model are as follows:

1. Requirement gathering & analysis: In this phase, requirements are gathered and check by an analyst whether requirements will fulfill or not. Analyst checks that need will achieve within budget or not. After all of this, the software team skips to the next phase.
2. Design: In the design phase, team design the software by the different diagrams like Data Flow diagram, activity diagram, class diagram, state transition diagram, etc.
3. Implementation: In the implementation, requirements are written in the coding language and transformed into computer programmes.

4. Testing: After completing the coding phase, software testing starts using different test methods. There are many test methods, but the most common are white box, black box, and grey box test methods.

5. Deployment: After completing all the phases, software is deployed to its work environment.

6. Review: In this phase, after the product deployment, review phase is performed to check the behaviour and validity of the developed product. And if there are any error found then the process starts again from the requirement gathering.

7. Maintenance: In the maintenance phase, after deployment of the model in the working environment there may be some bugs, some errors or new updates required. Maintenance involves debugging and new addition options.

After end of every cycle, a decision is made a decision is made as to whether the model produced by the cycle should be kept or discarded, or kept as a starting point of next cycle(known as incremental prototyping).Eventually a point is reached where there is no further enhancements needed for the model.

The key to a successful use of an iterative model life cycle is proper validation of requirements, and verification, testing of each version of the software against those requirements within each cycle of the model.

2 LITERATURE SURVEY

Sr.No	Paper Name	Author Name	Description
1	Soil Moisture Prediction Using Machine Learning	Shikha Prakash	Use of machine learning algorithm like multiple linear regression, support vector regression and recurrent neural networks for prediction of soil moisture for 1 day, 2 days and 7 days ahead. The performance was compared with R 2 and RMSE. The comparison shows multiple linear regression is better in providing RMSE and R2 of 0.14 and 0.975 for day1, 0.353 and 0.939 for day 2, 1.59 and 0.786 for day 7 ahead.
2	The prediction model for soil water evaporation based on Back Propagation neural network	LiliMa et al	Use of BP neural network for prediction which gives high accuracy and high stability. Average temperature, relative humidity, net radiation these factors are considered for finding the output.
3	Soil Fertility Grading With Bayesian Network Transfer Learning	Hai-yang Jia et al	It describes methods for grading soil into different types so that suitable crops can be grown in it. They have used Bayesian Network based transfer learning algorithm. Algorithm includes structural learning parameter learning, considering both similarities between learning task and the geographical position of land square. Empirical experiment results show a significant improvement in terms of structure and parameters when transfer knowledge between similar soil fertility grading task.
4	Predicting Soil Heavy Metal Based On Random Forest Model	Weibo Ma et al	It predicts heavy metal content of soil. Use of three machine learning algorithms SVM, RF, ELM and compared with PLS method. Here ELM and RF performs better than SVM. The concentration of metals will affect the prediction of ELM. Stability of RF is best among

			these three models. RF algorithm has higher accuracy for inversion of soil heavy metal research.
5	Soil Data Analysis Using Classification Techniques and Soil Attribute Prediction	Jay Gholap et al	It describes analysis of the soil data using different algorithms and prediction technique. In spite the fact that the least median squares regression is known to produce better results than the classical linear regression technique, from the given set of attributes, the most accurately predicted attribute was "P" (Phosphorous content of the soil) and which was determined using the Linear Regression technique in lesser time as compared to Least Median Squares Regression. We have demonstrated a comparative study of various classification algorithms i.e. Naïve Bayes, J48 (C4.5), JRip with the help of data mining tool WEKA. J48 is very simple classifier to make a decision tree, but it gave the best result in the experiment.

The proposed work focuses on experimenting with various algorithms in predicting the soil contents accurately.

3 ALGORITHMIC SURVEY

Model Name (Soil NPK Prediction)	Accuracy
Decision Tree	0.984
K Nearest Neighbors	0.959
SVM(Support Vector Machine)	0.912
Naive Bayes	0.695
Logistic Regression	0.876
Stochastic Gradient Descent	0.629

Various algorithms were used for predicting N, P, K contents in soil accurately.
Accuracy of decision tree for predicting soil contents was highest.

Model Name (Advice Generation)	Accuracy
CN2 Rule Inducer	0.94
Adaboost	0.89
KNN (K-Nearest Neighbors)	0.82
Random Forest	0.85
SVM(Support Vector Machine)	0.80
Logistic Regression	0.79
Naive Bayes	0.81
SGD (Stochastic Gradient Descent)	0.86

Various algorithms were also experimented for accurately generating advice related to crops,
fertilizers and pesticides.CN2 Rule Inducer has the highest accuracy for advice generation.

4 SOFTWARE REQUIREMENT SPECIFICATION

4.1 Assumptions and Dependencies
- End application will be a web application
- Web application is designed using django framework using python
- Training data is a CSV file

4.2 Functional Requirements
- Machine learning algorithms are used for predicting soil health
- Based on the soil health various machine learning algorithms are used to generate multiple advice like crop, fertilizer, pesticide and fungicide selection

4.3 External Interface Requirements

4.3.1 User Interfaces
User interface is designed in python using DJANGO (a python framework)

4.3.2 Software Interfaces

WSGI Server:

A Web Server Gateway Interface (WSGI) server implements the web server side of the WSGI interface for running Python web applications.
A traditional web server does not understand or have any way to run Python applications. In the late 1990s, a developer named Grisha Trubetskoy came up with an Apache module called mod python to execute arbitrary Python code. For several years in the late 1990s and early 2000s, Apache configured with mod python ran most Python web applications.

Therefore the Python community came up with WSGI as a standard interface that modules and containers could implement. WSGI is now the accepted approach for running Python web applications.
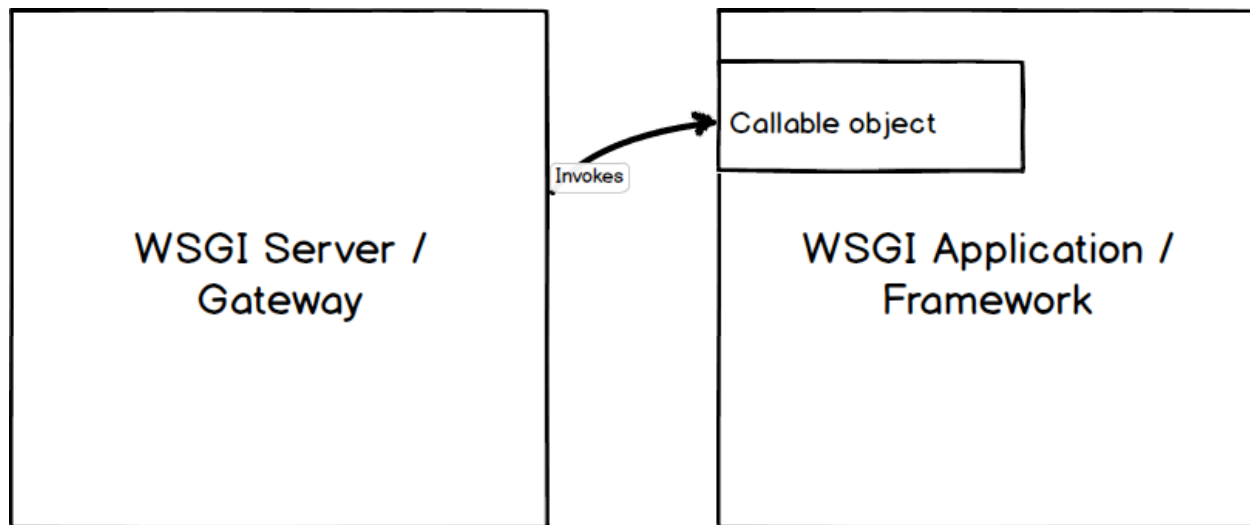
As shown in the above diagram, a WSGI server simply invokes a callable object on the WSGI application as defined by the PEP 3333 standard.

4.4 Performance Requirements

- Every module must perform properly in conjunction with other module
- Algorithms used must give the maximum possible accuracy

4.5 Safety Requirements

- The application is designed using modules where identifying and fixing errors can be easily done.
- Due to this there is ease in implementing various functionalities according to requirement

4.6 Software quality requirements

- **Adaptability**: This software is adaptable by all users
- **Availabilit**y: This software is freely available to all users. The availability of the software is easy for everyone
- **Maintainability**: After the project deployment 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 reliability of the software
- **User Friendless**: Since, the software is a GUI application; output generated is much user friendly in 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

14

- **Testability**: The software will be tested considering all aspects

4.7 Software Requirements

- Operating system: Linux.
- Coding Language: Python
- Web Server: WSGI Server
- Back End: DJANGO and Python

## 4.8 ANALYSIS MODELS: SDLC MODEL TO BE APPLIED

Agile Model:

The meaning of Agile is swift or versatile. "Agile process model" refers to a software development approach based on iterative development. Agile methods break tasks into smaller iterations, or parts do not directly involve long term planning. The project scope and requirements are laid down at the beginning of the development process. Plans regarding the number of iterations, the duration and the scope of each iteration are clearly defined in advance.
Each iteration is considered as a short time "frame" in the Agile process model, which typically lasts from one to four weeks. The division of the entire project into smaller parts helps to minimize the project risk and to reduce the overall project delivery time requirements. Each iteration involves a team working through a full software development life cycle including planning, requirements analysis, design, coding, and testing before a working product is demonstrated to the client.

Phases of Agile Model:

Following are the phases in the Agile model are as follows:

1. Requirements gathering
2. Design the requirements
3. Construction/ iteration
4. Testing/ Quality assurance
5. Deployment
6. Feedback

1. Requirements gathering: In this phase, you must define the requirements. You should explain business opportunities and plan the time and effort needed to build the project. Based on this information, you can evaluate technical and economic feasibility.

2. Design the requirements: When you have identified the project, work with stakeholders to define requirements. You can use the user flow diagram or the high-level UML diagram to show the work of new features and show how it will apply to your existing system.

3. Construction/ iteration: When the team defines the requirements, the work begins. Designers and developers start working on their project, which aims to deploy a working product. The product will undergo various stages of improvement, so it includes simple, minimal functionality.

4. Testing: In this phase, the Quality Assurance team examines the product's performance and looks for the bug.

5. Deployment: In this phase, the team issues a product for the user's work environment.

6. Feedback: After releasing the product, the last step is feedback. In this, the team receives feedback about the product and works through the feedback.
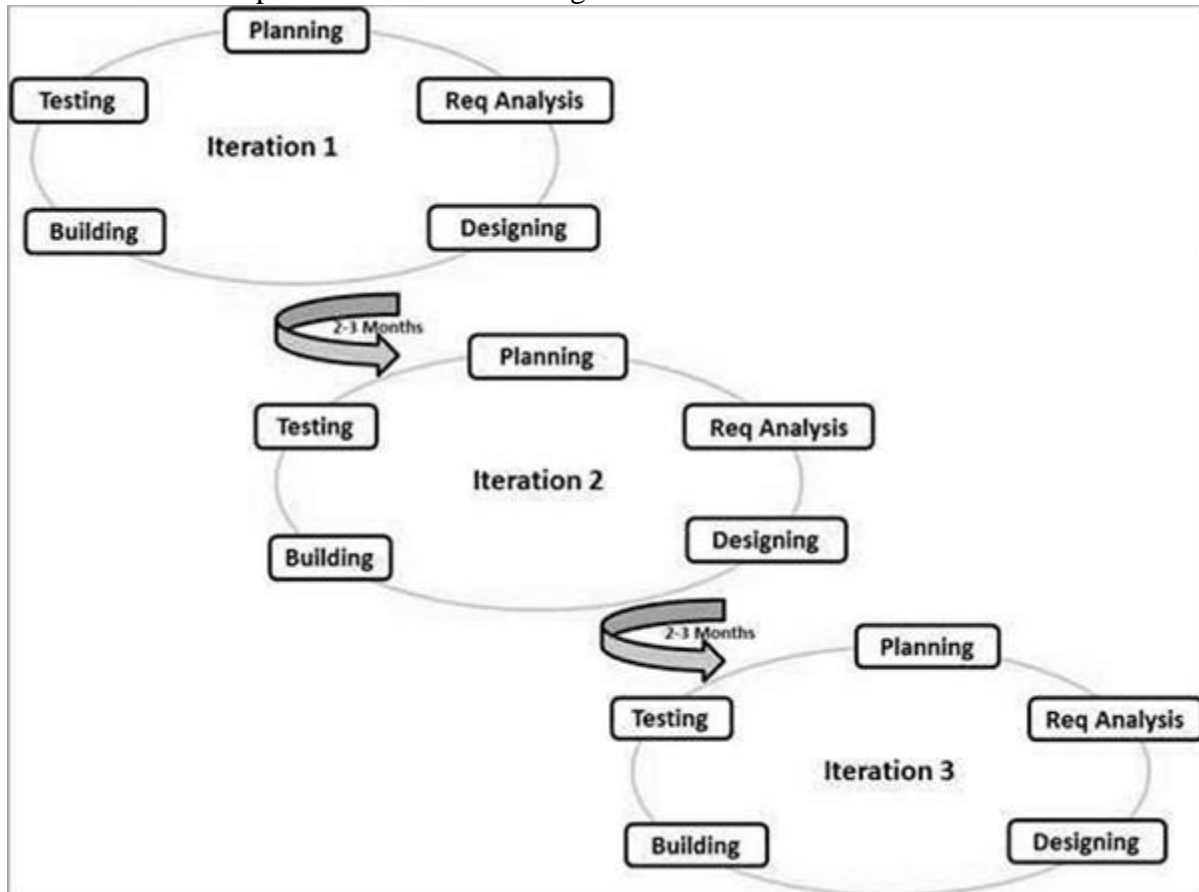
5 MATHEMATICAL MODEL:

M= {s, e, X, Y, DD, Fme}
Where,
s= {start state}
e= {end state}
X= {NPK input values, Ph, District where crop is grown}
Y= {Output as crops, fertilizers to be used, pesticides}
DD=Deterministic data
Fme= {functions used: Decision Tree () for soil health and CN2 classifier () for multi advice generator}

Procedure:

1. User will input data containing soil nitrogen, phosphorous and potassium values along with crop district
2. Decision tree algorithm is used to calculate soil health
3. Further CN2 classifier generates crop advice, pesticides, fungicides and insecticide advice.

6 SYSTEM DESIGN

6.1 System architecture diagram
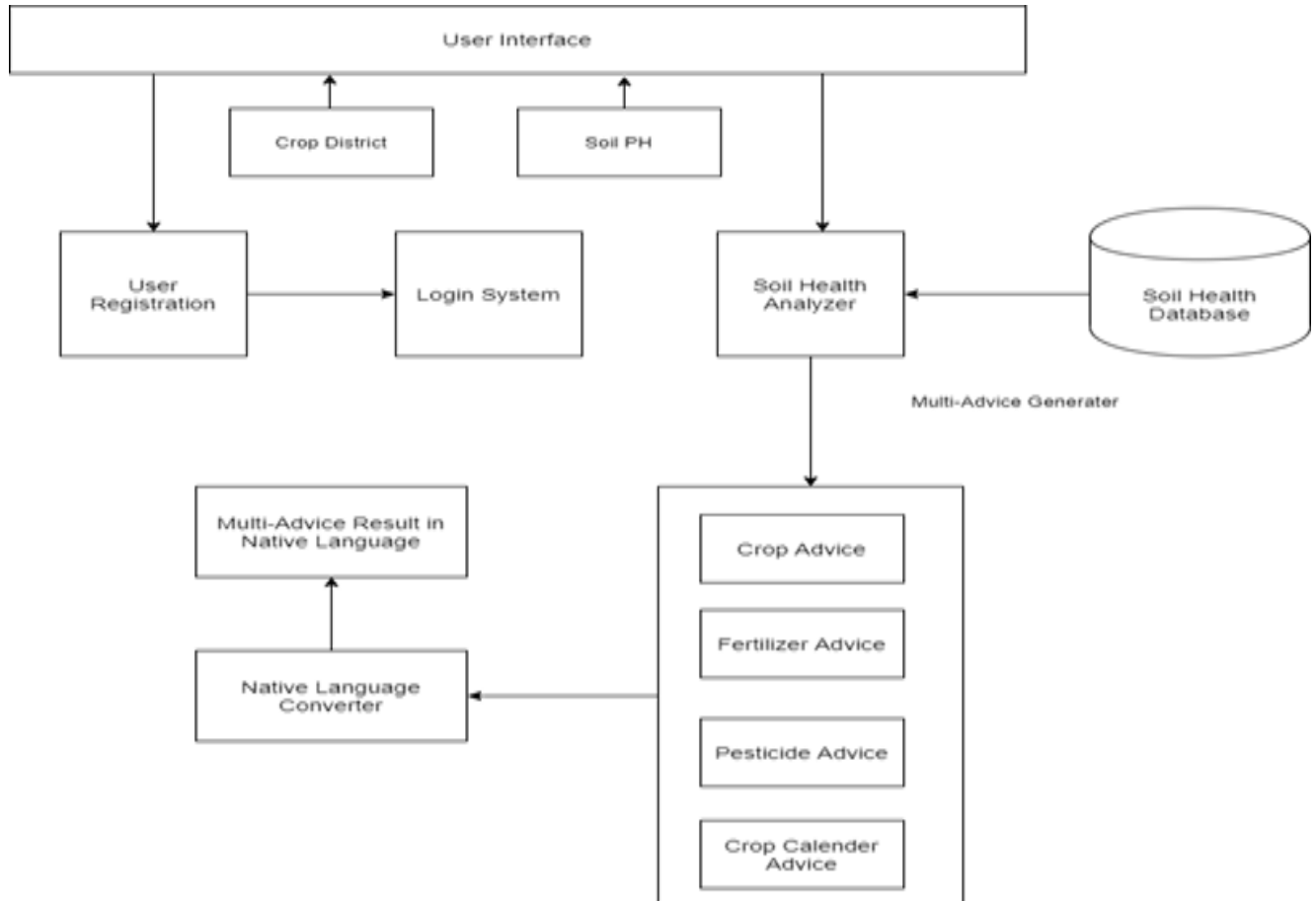
Description

First the user has to log in to the system by registering. User will give N, P, K, crop district and soil ph as input. This input is then used to predict the soil quality. Further the system generates multiple advices - crop recommendation, fertilizer, pesticide, insecticide to be used .Also, the advice generated is in farmer's native language.

6.2 Dataflow diagram
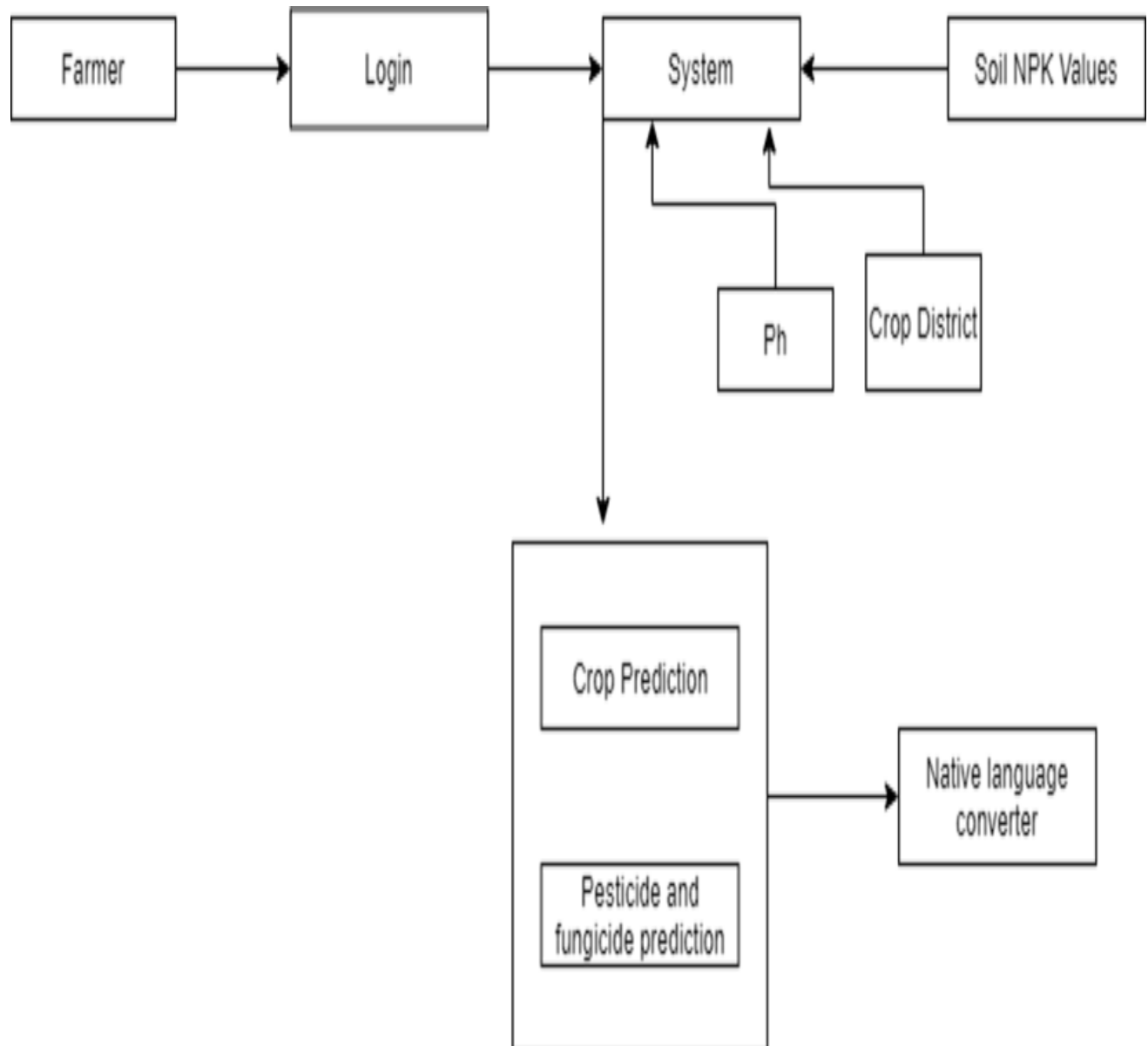
6.3 Use case diagram

6.4 Class diagram

User

+name
+address
+location
+module_no

+login_to_system()
+enter_soil_credentials()
+enter_crop_district()

System

+predict_soil_NPK_content()
+predict_crop()
+recommend_fertilizers()
+recommend_pesticides()

system
database

Native language Multi Advice

+ generate_advice_in_native
_language()
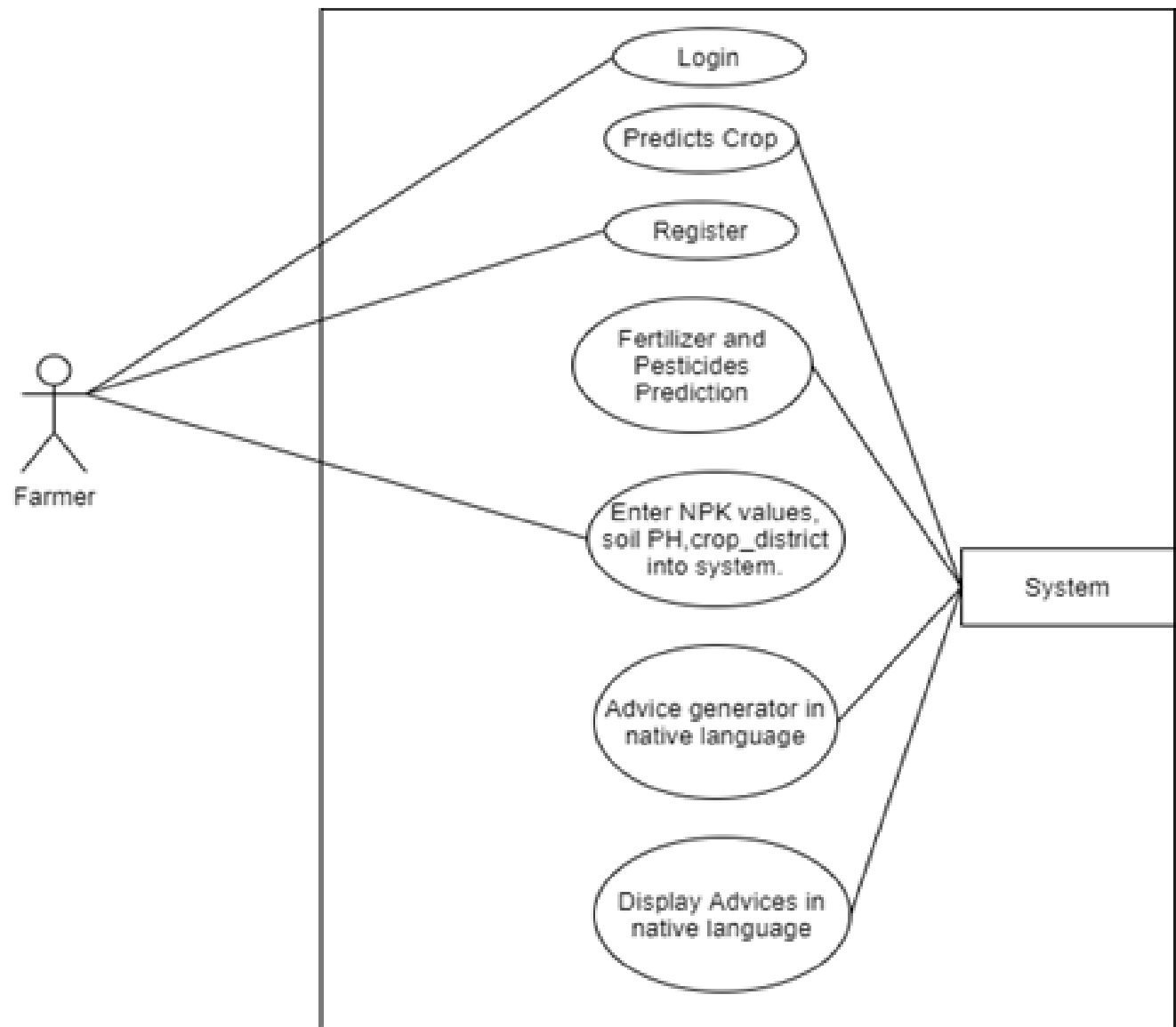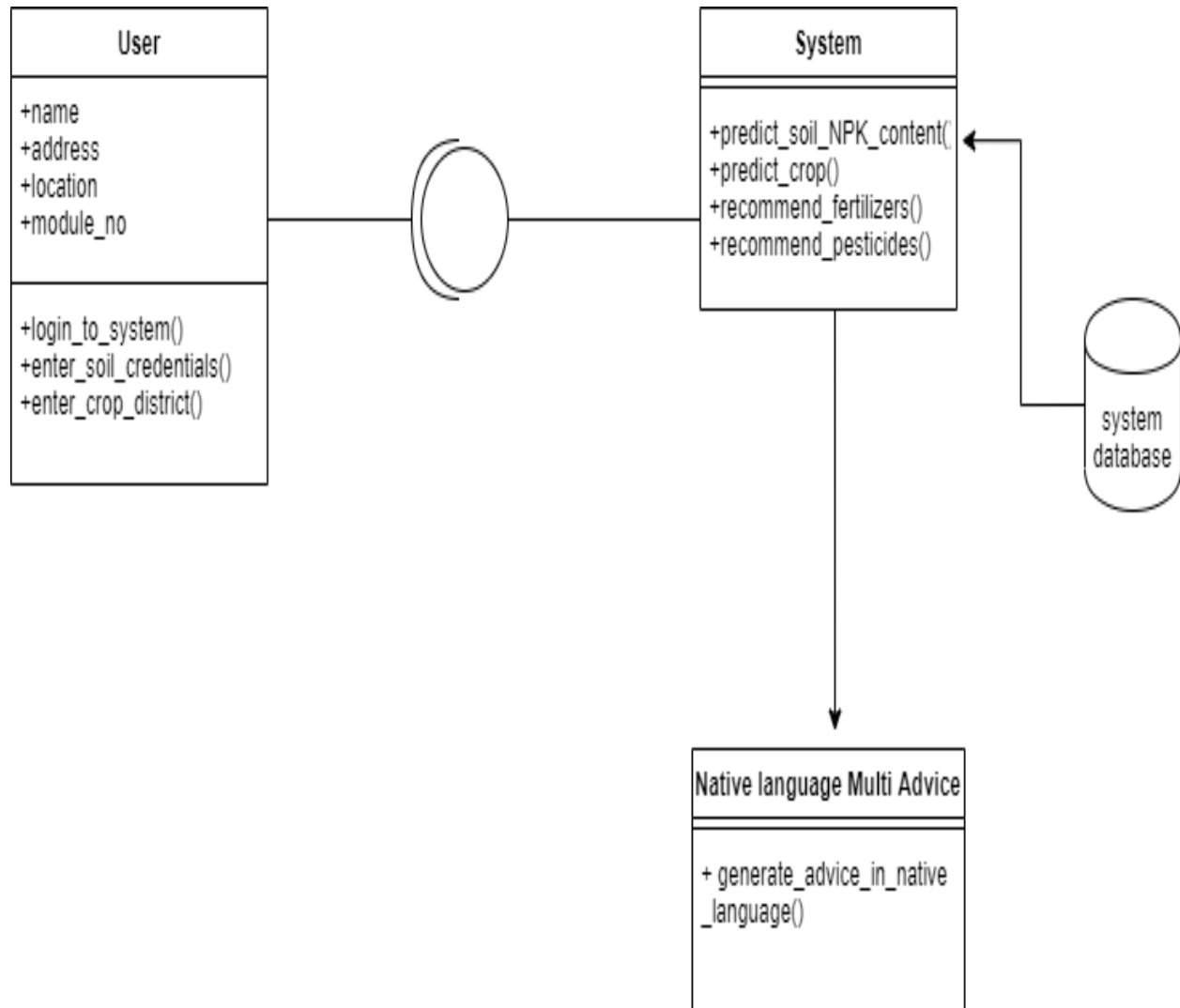
7 PROJECT PLAN

7.1 Project resources
Laptop, Internet Connection, Django, Python

7.2 Risk management

1.Have top software and customer managers formally committed to support the project?
Ans-Yes

2.Are end users enthusiastically committed to the project and the system/product to be built?
Ans-Yes

3.Are requirements fully understood by the software engineering team and its customers?
Ans-Yes

4.Have customers been involved fully in the definition of requirements?
Ans-Moderate

5.Do end users have realistic expectations?
Ans-Yes

6.Does the software engineering team has the right mix of skills?
Ans-Yes

7.Are project requirements stable?
Ans-Yes

8.Is the number of people on the project team adequate to do the job?
Ans-Yes

9.Do all customer/user constituencies agree on the importance of project and on the requirement

7.2.1 Risk Mitigation

- Risk mitigation is a strategy to prepare for and lessen the effects of threats faced by a data center. Comparable to risk reduction, risk mitigation takes steps to reduce the negative effects of threats and disasters
- Mitigate those causes that are under your control before the project starts
- Organize project teams so that information about each development activity is widely dispersed
- Define work product standards and establish mechanisms to be sure that all models and documents are developed in a timely manner
- Conduct peer reviews of all work
- Assign a backup staff member for every critical technologist

7.2.2 Risk Monitoring

As project goes on risk monitoring activities commence. The project manager monitors factors that may provide an indication of whether the risk is becoming more or less likely. In the case of high staff turnover, the general attitude of team members based on project pressures, the degree to which the team has jelled, interpersonal relationships among team members, potential problems with compensation and benefits, and the availability of jobs within the company and outside it are all monitored.

7.2.3 Risk Management

Project risk management is the process of identifying, analyzing and then responding to any risk that arises over the life cycle of a project to help the project remain on track and meet its goal. Risk management isn't reactive only; it should be part of the planning process to figure out risk that might happen in the project and how to control that risk if it in fact occurs.

A risk is anything that could potentially impact your project's timeline, performance or budget. Risks are potentialities, and in a project management context, if they become realities, they then become classified as "issues" that must be addressed. So risk management, then, is the process of identifying, categorizing, prioritizing and planning for risks before they become issues.

Risk management can mean different things on different types of projects. On large-scale projects, risk management strategies might include extensive detailed planning for each risk to ensure mitigation strategies are in place if issues arise. For smaller projects, risk management might mean a simple, prioritized list of high, medium and low priority risks.

Major tasks in project stages are:
- Correctness
- Availability
- Integrity

Task Network
Correctness -> Availability ->Integrity

7.2.4 Team Organisation:

Schedule Plan

Month Scheduled	Phase	Name of group members	Task Performed
June	Topic Searching	Samihan,Mohit,Devesh,Pratik	Topic Searched
August-September	Project Confirmation	Samihan,Mohit,Devesh,Pratik	Project Confirmed
August-September	Literature Survey	Samihan,Mohit,Devesh,Pratik	Literature Survey Done
September-October	Requirement Analysis	Samihan,Mohit,Devesh,Pratik	Requirement Analysis Done
October	Requirement Analysis	Samihan,Mohit,Devesh,Pratik	Requirement Gathering Done
November-December	Designing	Samihan,Pratik	Designing Done
November-December	Designing Test	Samihan	Designing Test Done
December-January	Database Creation	Samihan,Mohit,Devesh,Pratik	Database created
January	Coding	Samihan,Pratik	Coding Done
February	Result Analysis	Samihan,Devesh	Result Analysed
March	Database and modules connectivity	Samihan,Pratik	Connectivity Done
April	Testing	Samihan	Testing Done

Team Structure

Team of 4 people is formed and roles are assigned. All members are contributing in all phases of project

Management Reporting and communication:
Well planning mechanisms are used for progress reporting and intr/intra team communication are identified as per project requirement

24

8 PROJECT IMPLEMENTATION

8.1 PROJECT MODULES OVERVIEW:

8.1.1 Software Module:

Software module consists of a web application. It takes into account the N, P, K values of soil, crop calendar from farmers. At the backend various processes take place. Based on the input values quality of soil is checked on which crops fit for the soil quality are predicted. Also, various fertilizers, pesticides and insecticides are recommended for the crops. Apart from this the advice generated is in the native language of the farmer.

8.2 TOOLS AND TECHNOLOGIES:

- **PYTHON**

Python is a general purpose and high level programming language. You can use Python for developing desktop GUI applications, websites and web applications. Also, Python, as a high level programming language, allows you to focus on core functionality of the application by taking care of common programming tasks.

Python is widely used because of the following reasons-

1) Readable and Maintainable Code
2) Multiple Programming Paradigms
3) Compatible with Major Platforms and Systems
4) Robust Standard Library
5) Many Open Source Frameworks and Tools
6) Simplify Complex Software Development
7) Adopt Test Driven Development

- **PYTHON WEB FRAMEWORK-DJANGO**

Django is an open-source python web framework used for rapid development, pragmatic, maintainable, clean design, and secures websites. ... involved in web development; enables users to focus on developing components needed for their application.

With Django, you can take Web applications from concept to launch in a matter of hours. Django takes care of much of the hassle of Web development, so you can focus on writing your app without needing to reinvent the wheel. It's free and open source.

Advantages of Django are as follows:

Ridiculously fast.

Django was designed to help developers take applications from concept to completion as quickly as possible.

Fully loaded.

Django includes dozens of extras you can use to handle common Web development tasks. Django takes care of user authentication, content administration, site maps, RSS feeds, and many more tasks — right out of the box.

Reassuringly secure.

Django takes security seriously and helps developers avoid many common security mistakes, such as SQL injection, cross-site scripting, cross-site request forgery and clickjacking. Its user authentication system provides a secure way to manage user accounts and passwords.

Exceedingly scalable.

Some of the busiest sites on the planet use Django's ability to quickly and flexibly scale to meet the heaviest traffic demands.

Incredibly versatile.

Companies, organizations and governments have used Django to build all sorts of things — from content management systems to social networks to scientific computing platforms.

Created by Ryan Nevius
http://www.ryannevius.com
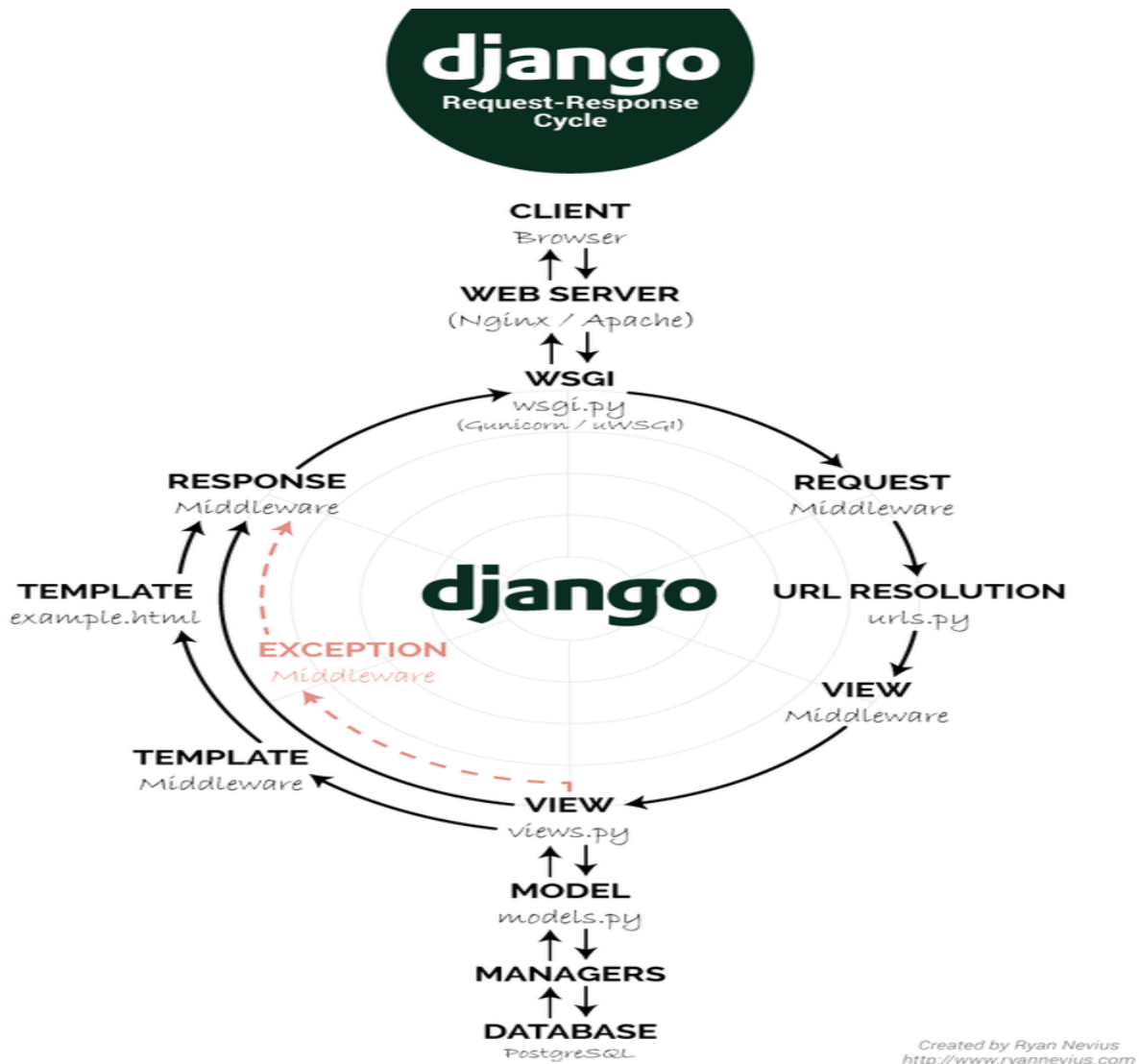
- **HTML**

HTML stands for Hypertext Markup Language, and is used to describe the visual appearance of a document to be displayed by an internet browser. ... HTML documents are saved in text format and are designed to be viewed or edited on any operating system that is able to connect to the Internet.

Used for:

- **Web pages development**: HTML is heavily used for creating pages that are displayed on the world wide web. Every page contains a set of HTML tags including hyperlinks which are used for connecting to other pages. Every page that we witness, on the world wide web, is written using a version of HTML code.

27

- **Web document creation:** Document creation on the internet is dominated by HTML and its basic concept via tag and DOM i.e. document object model. HTML tags are inserted before and afterward or phrases to locate their format and location on the page. A web document consists of three sections: title, head, and body. Head includes the information to identify the document, including title and any other important keyword. A title can be seen on the browser's bar and body section is the main portion of the website visible to the viewer. All the three segments are designed and created by the uses of HTML tags. Every section has their own specific set of tags, which are dedicatedly rendered keeping the head, title and body concepts in a loop.
- **Internet navigation**: This is one of the most important uses of HTML which is revolutionary. This navigation is possible by utilizing the concept of Hypertext. It is basically a text which refers to other page and when user click on it, would navigate to referenced text or page. HTML is heavily used to embed the hyperlink within the web pages. A user can easily navigate within the web pages and between websites as well, which are located on different servers.

- **CSS:**
  CSS is the language for describing the presentation of Web pages, including colors, layout, and fonts. It allows one to adapt the presentation to different types of devices, such as large screens, small screens, or printers. CSS is independent of HTML and can be used with any XML-based markup language.

8.3 Algorithms Used:

- **Decision tree**
  Decision tree classifiers are utilized as a well-known classification technique in different pattern recognition issues, for example, image classification and character recognition (Safavian & Landgrebe, 1991). Decision tree classifiers perform more successfully, specifically for complex classification problems, due to their high adaptability and computationally effective features. Besides, decision tree classifiers exceed expectations over numerous typical supervised classification methods (Friedl & Brodley, 1997).
  In particular, no distribution assumption is needed by decision tree classifiers regarding the input data. This particular feature gives to the Decision Tree Classifiers a higher adaptability to deal with different datasets, whether numeric or categorical, even with missing data. Also, decision tree classifiers are basically nonparametric. Also, decision trees are ideal for dealing with nonlinear relations among features and classes. At long last, the classification procedure through a tree-like structure is constantly natural and interpretable.
  Generally, a decision tree comprises of three basic segments including a root node, a few hidden nodes, and a lot of terminal nodes (known as leaves). An illustrative case of a decision tree structure is depicted in Fig. 2.21. As demonstrated, for each hidden and terminal node (known as child node), there should exist a parent node

28

demonstrating the data source. In the interim, with respect to the root node and each hidden node (known as parent hub), at least two child nodes will be created from these parent nodes dependent on different decision rules.
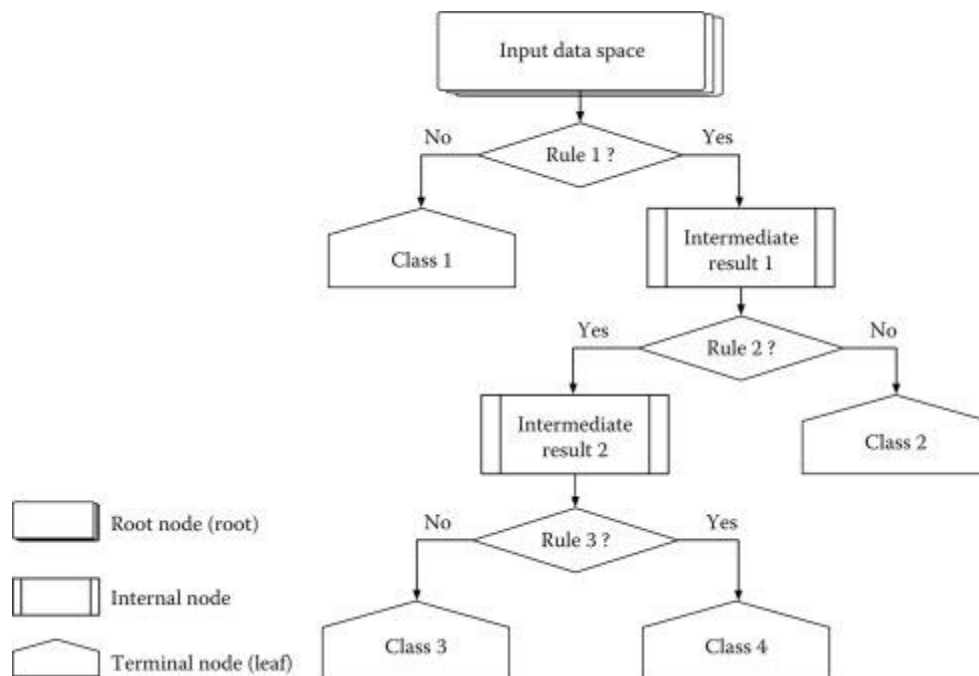
- **CN2 Classifier:**

The representation for rules output by CN2 is an ordered set of if-then rules, also known as a decision list (Rivest, 1987). CN2 uses a heuristic function to terminate search during rule construction, based on an estimate of the noise present in the data. This results in rules that may not classify all the training examples correctly, but that perform well on new data.

8.4 Goslate

Goslate provides you free python API to google translation service by querying google translation website. It is: Free: get translation through public google web site without fee. Fast: batch, cache and concurrently fetch.

9 SOFTWARE TESTING:

9.1 Types of Testing

GUI TESTING

GUI testing is a testing technique in which the application's user interface is tested whether the application performs as expected with respect to user interface behaviour.

GUI Testing includes the application behaviour towards keyboard and mouse movements and how different GUI objects such as toolbars, buttons, menubars, dialog boxes, edit fields, lists, behavior to the user input.

Proposed system is tested for various input parameters in different modules. Validation for this is done. GUI testing involves:

1.Check elements for size, position, width, length. U may be able to provide input and input fields.

2.Functionality of modules is checked.

3.Error messages are displayed.

4.Element positioning is checked.

Unit Testing

Here, individual software units are tested. The purpose is to validate that each unit of the software performs as designed. A unit is the smallest testable part of any software. It usually has one or a few inputs and usually a single output. In procedural programming, a unit may be an individual program, function, procedure, etc. In object-oriented programming, the smallest unit is a method, which may belong to a base/ super class, abstract class or derived/ child class. (Some treat a module of an application as a unit. This is to be discouraged as there will probably be many individual units within that module.) Unit testing frameworks, drivers, stubs, and mock/ fake objects are used to assist in unit testing.

Integration Testing
It is a level of software testing where individual units are combined and tested as a group. The purpose of this level of testing is to expose faults in the interaction between integrated units. Test drivers and test stubs are used to assist in Integration Testing.

Testing Strategy

1.White Box Testing

WHITE BOX TESTING (also known as Clear Box Testing, Open Box Testing, Glass Box Testing, Transparent Box Testing, Code-Based Testing or Structural Testing) is a software

testing method in which the internal structure/design/implementation of the item being tested is known to the tester. The tester chooses inputs to exercise paths through the code and determines the appropriate outputs. Programming know-how and the implementation knowledge is essential. White box testing is testing beyond the user interface and into the nitty-gritty of a system. This method is named so because the software program, in the eyes of the tester, is like a white/transparent box; inside which one clearly sees.

2.Black Box Testing

BLACK BOX TESTING, also known as Behavioral Testing, is a software testing method in which the internal structure/design/implementation of the item being tested is not known to the tester. These tests can be functional or non-functional, though usually functional.
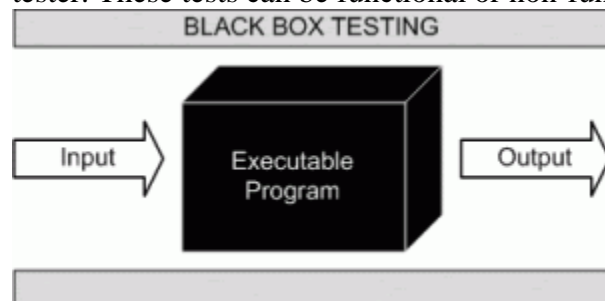
This method is named so because the software program, in the eyes of the tester, is like a black box; inside which one cannot see. This method attempts to find errors in the following categories

3.Grey Box Testing

Grey Box testing is testing technique performed with limited information about the internal functionality of the system. Grey Box testers have access to the detailed design documents along with information about requirements.

Grey Box tests are generated based on the state-based models, UML Diagrams or architecture diagrams of the target system.

9.2 TEST CASES AND RESULTS:

SYSTEM TEST CASES
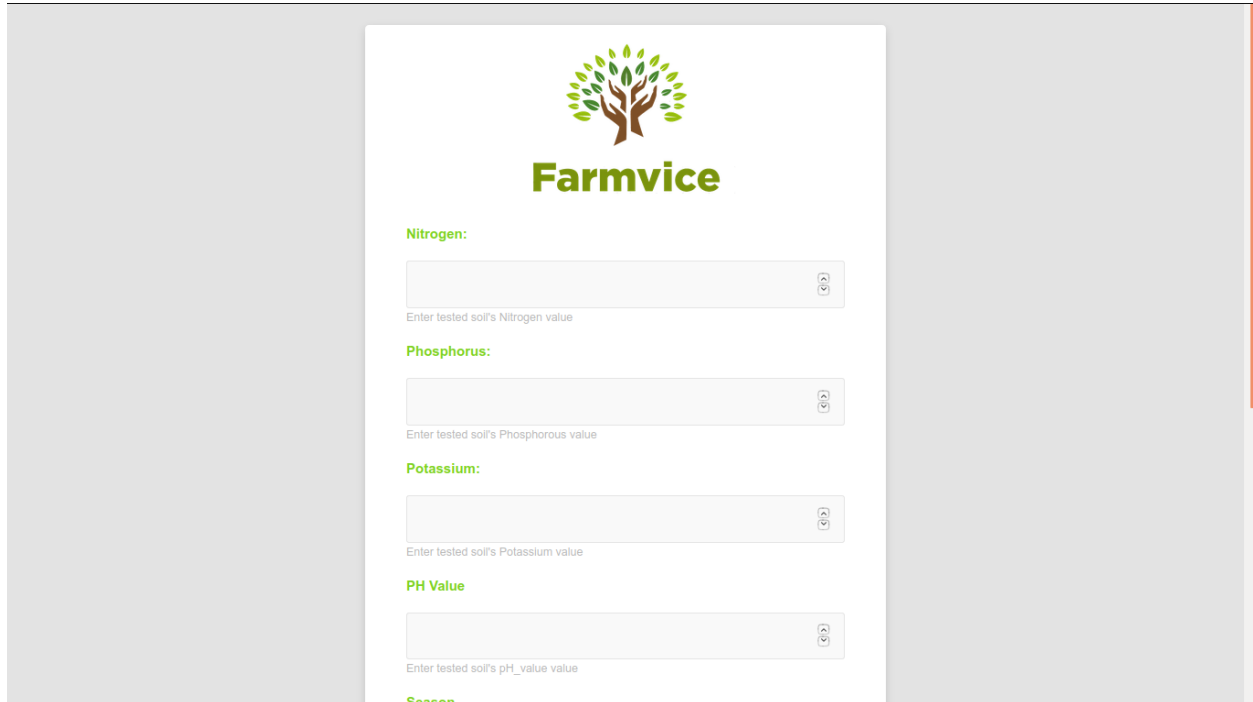
Test ID	Test Case Objective	Test Case Prerequisite	Input	Display	Status
TC_001	Enter Ph,NPK values properly	Values should be numeric	a,b	a,b	Fail
TC_002	NPK value acceptance	Value must be positive	5.7	5.7	Success
TC_003	PH accepted range	Values must Be in between 4 to 8.	7	7	Success
TC_004	Enter valid season	Season must be either Rabi or Kharif	Rabi	Rabi	Success
TC_005	Enter valid district	District must be in Maharashtra	Ahmednagar	Ahmednagar	Success

10 RESULTS

10.1 OUTCOME
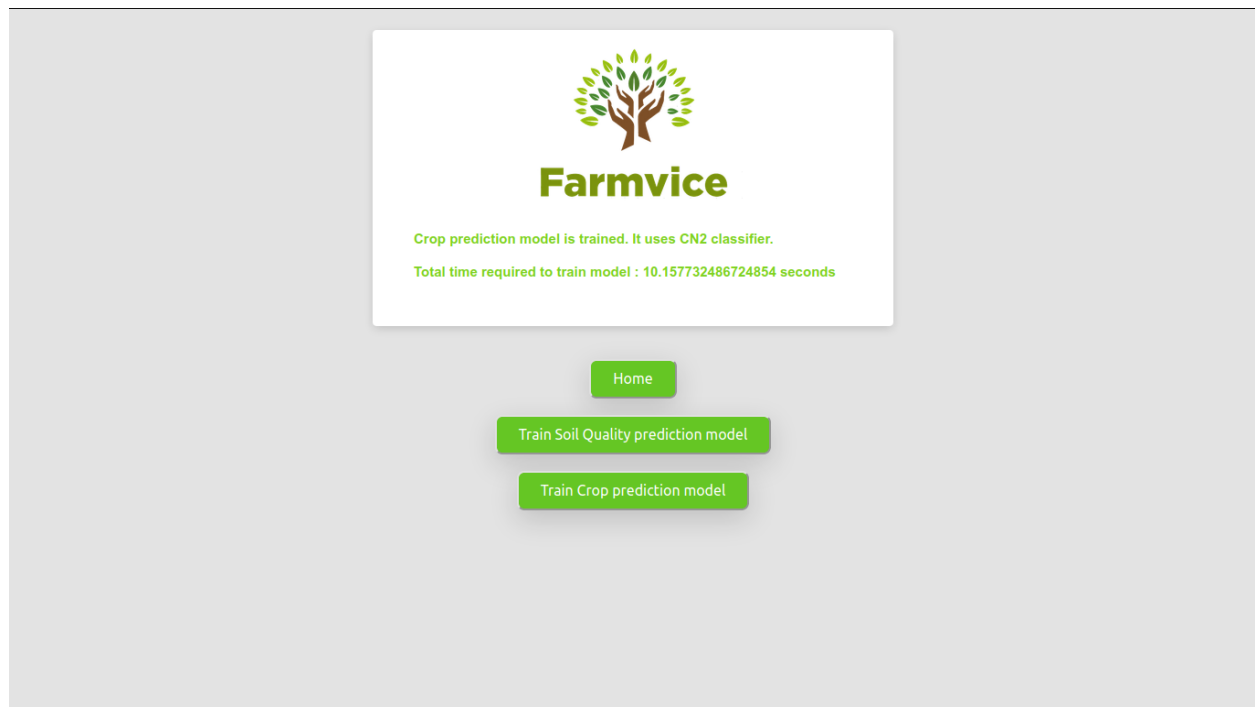
We have developed a multi advice crop generator. Various crops, insecticides, pesticides, fungicides are suggested based on soil nitrogen, phosphorous and potassium content. Advice is generated in native language of farmer.

10.2 SCREEN SHOTS

Interface for input parameters

Farmvice

Crop prediction model is trained. It uses CN2 classifier.

Total time required to train model : 10.157732486724854 seconds

Home

Train Soil Quality prediction model

Train Crop prediction model

Result after applying CN2 algorithm for crop prediction

Farmvice

Soil Quality prediction model is trained. It uses Decision tree classifier.

Total time required to train model : 0.00981283187866211 seconds

Home

Train Soil Quality prediction model

Train Crop prediction model
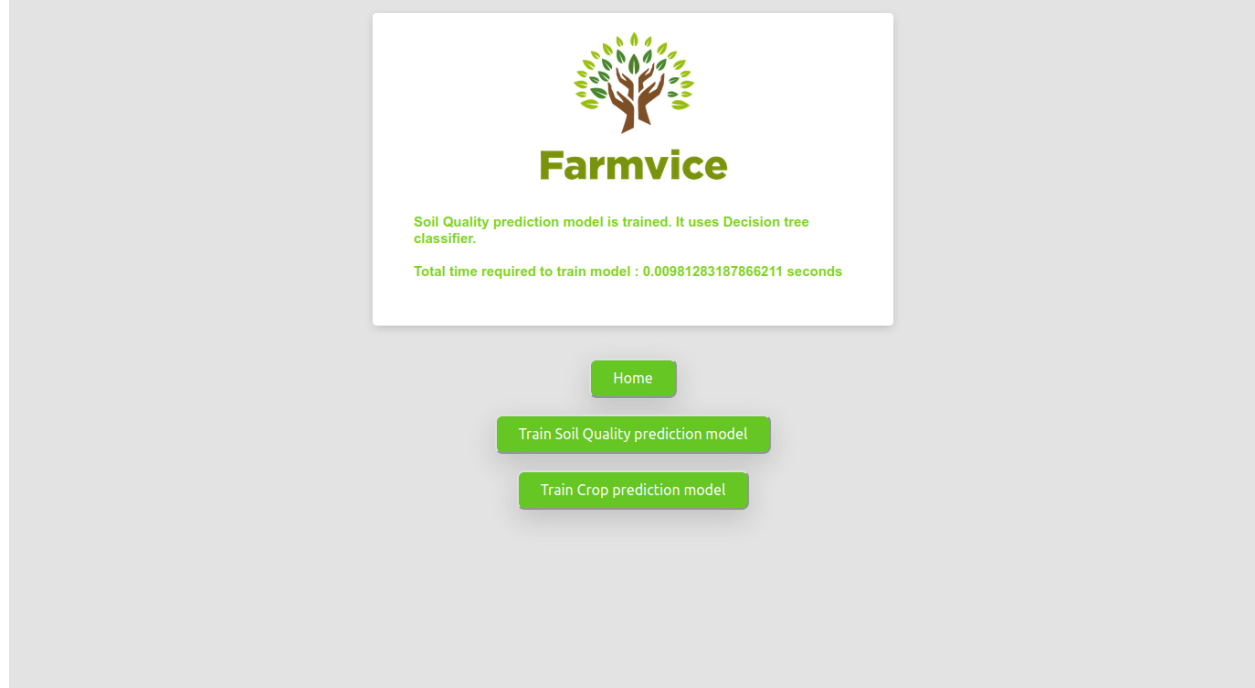
Result after applying Decision Tree algorithm for soil health prediction

Farmvice

Advice:

मातीची गुणवत्ता कमी आहे. आपण ज्वारी पेरले पाहिजे.

Nitrogen: 353.0

Phosphorus: 6.0

Potassium: 764.0

pH value: 6.7

Season: Rabi

District: AHMEDNAGAR

Soil Quality: Low

Suggested crop: Jowar

Suggested Pesticides:

Pesticide	Dieses
Mancozeb 75% WP	Leaf spot
Zineb 75% WP	Red leaf spot

Advice generated in native language (Marathi) using Goslate

Suggested crop: Jowar

Suggested Pesticides:

Pesticide	Dieses
Mancozeb 75% WP	Leaf spot
Zineb 75% WP	Red leaf spot
Carboxin 17.5% + Thiram 17.5% FF	Leaf spot
Propiconazole 13.9% + Difenoconazole 13.9% EC	Leaf blight

Suggested Insecticides:

Insecticide	Pests
3% Carbofuran, 5% Disulfoton,10% Phorate granule	Shoot fly
apply 4% Endosulfan, carbaryl granules, spray 0.05% Lindane	Stem borers
0.05% Endosulfan, 01% Carbaryl, 0.05% Lindane sprays or dusts	Sorghum midge
0.02% Phosphamidon, 0.04% Diazinon Methyl-demeton , Dimethoate	Aphids
dust hoppers and adults with 5 and 10% BHC, respectively.	Deccan wingless grass hopper
5% BHC or 4% Carbaryl,0.1% Carbaryl	Earhead bug
0.03% Phosphamidon Dimethoate Diazinon, Monocrotophon, Methyl demeton	Sorghum shoot bug
10% BHC	Cut worm
10% BHC	Hairy caterpillars
5% BHC or 4% Carbaryl,spray 0.1% Carbaryl	Earhead caterpillars
Dust sulphur, spray 0.05% wettable sulphur	Mites

Advice generated suggesting pesticides and fungicides according to the crop

11 CONCLUSION AND FUTURE WORK

In our work, we have attempted to develop a multi-advice generator in native language with a motive to help farmers in agriculture in order to ensure healthy crop produce. We have studied and experimented with various machine learning techniques to achieve maximum possible accuracy in predictions based on crop and soil health. Amongst variety of algorithms tested, Decision Tree Classifier predicts soil health-nitrogen, phosphorous and potassium content with highest accuracy of 98.4 %. These soil heath parameters are then fed as input for predicting crops. Experiments show that CN2 predicts crops with highest accuracy of 94 % which is significantly more than other algorithms tested for crop prediction.

This project can be further developed to generate more valuable advice by taking into account various other parameters affecting crop health. For instance, Real time weather conditions can be taken into account along with present parameters while making predictions and generating advice. The more parameters influencing yield will be considered for predictions, the more real and correct will be the advices generated. Apart from this, initially the system makes predictions based on a limited dataset but with successive predictions the system learns from input making the dataset richer resulting in increased accuracy in predictions with time.

12 References.

[1] S. R. Vispute and M. A. Potey, " Automatic text categorization of Marathi documents using clustering technique, " 2013 15th International Conference on Advanced Computing Technologies (ICACT), Rajampet, 2013, pp. 1-5

[2] Singh, Ravinder & Kaur, Rubleen & Singla, Anshu. (2018). Two phase fuzzy based prediction model to predict Soil nutrient
.
[3] Vispute, S. R, Patil, S., Sangale, S., Padwal, A., & Ukarde, A. (2015). Parallel Processing System for Marathi Content Generation. 2015 International Conference on Computing Communication Control and Automation.

[4] H. Nie, L. Yang, X. Li, L. Ren, J. Xu and Y. Feng, "Spatial Prediction of Soil Moisture Content in Winter Wheat Based on Machine Learning Model, " 2018 26th International Conference on Geoinformatics Kunming, 2018, pp. 1-6

[5] W. Ma, K. Tan and P. Diu, " Predicting soil heavy metal based on Random Forest Model, " 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, 2016, pp. 4331-4334

[6] Prakash, Shikha & Sharma, Animesh & Sahu, Sitanshu. (2018). Soil Moisture Prediction Using Machine Learning. 1-6

[7] S. Rose, S. Nickolas and S. Sangeetha, " Machine Learning and Statistical Approaches used in Estimating Parameters that Affect the Soil Fertility Status: A Survey, " 2018 Second International Conference on Green Computing and Internet of Things (ICGCIoT), Bangalore, India, 2018, pp. 381-385

[8] Vispute, S. R, Patil, S., Sangale, S., Padwal, A., & Ukarde, A. (2015). Parallel Processing System for Marathi Content Generation. 2015 International Conference on Computing Communication Control and Automation.

[9] E. Claverie, J. Lecoeur, V. Letort and P. Cournède, " Modeling soil temperature to predict emergence, "

2016 IEEE International Conference on
Functional-Structural Plant Growth Modeling,
Simulation, Visualization and Applications (FSPMA),
Qingdao. 2016, pp. 28-37

[10] H. Nie, L. Yang, X. Li, L. Ren, J. Xu and Y. Feng, "Spatial
Prediction of Soil Moisture Content in Winter Wheat
Based on Machine Learning Model, " 2018 26th
International Conference on Geoinformatics.,
Kunming, 2018, pp. 1-6

[11] Al Zaminur Rahman, Sk & Chandhra Mitra, Kaushik &
Islam, S.M. (2018). Soil Classification Using Machine
Learning Methods and Crop Suggestion Based on Soil
Series. 1-4

[12] L. Ma, Y. Wu, J. Ji and C. He, " The Prediction Model for
Soil Water Evaporation Based on BP Neural Network, "
2011 International Conference Computer Distributed
Control and Intelligent Environmental Monitoring,
Changsha, 2011, pp. 641-645

[13] H. Nie, L. Yang, X. Li, L. Ren, J. Xu and Y. Feng, "Spatial Prediction
of Soil Moisture Content in Winter Wheat Based on Machine
Learning Model, " 2018 26th International Conference on
Geoinformatics, Kunming, 2018, pp. 1-6

[14] S. Rose, S. Nickolas and S. Sangeetha, " Machine Learning and
Statistical Approaches used in Estimating Parameters that Affect the
Soil Fertility Status: A Survey, " 2018 Second International
Conference on Green Computing and Internet of Things (ICGCIoT),
Bangalore, India, 2018, pp. 381-385
[15] W. Ma, K. Tan and P. Du, " Predicting soil heavy metal based on
Random Forest Model, " 2016 IEEE International Geoscience and
Remote Sensing Symposium (IGARSS), Beijing, 2016, pp. 4331-
4334

[16] Al Zaminur Rahman, Sk & Chandra Mitra, Kaushik & Islam, S.M.
(2018). Soil Classification Using Machine Learning Methods and
Crop Suggestion Based on Soil Series. 1-4

[17] Singh, Ravinder & Kaur, Rubleen & Singla, Anshu. (2018). Two
phase fuzzy based prediction model to predict Soil nutrient. 80-85
[6] Prakash, Shikha & Sharma, Animesh & Sahu, Sitanshu. (2018). Soil
Moisture Prediction Using Machine Learning. 1-6

[18] L. Ma, Y. Wu, J. Ji and C. He, " The Prediction Model for Soil Water
Evaporation Based on BP Neural Network, " 2011 International
Conference on Computer Distributed Control and Intelligent
Environmental Monitoring, Changsha, 2011, pp. 641-645

[19] H. Jia, J. Chen, H. Yu and D. Liu, " Soil fertility grading with
Bayesian Network transfer learning, " 2010 International Conference
on Machine Learning and Cybernetics, Qingdao, 2010, pp. 1159-1163

[20] S. R. Vispute and M. A. Potey, " Automatic text categorization of
Marathi documents using clustering technique, " 2013 15th
International Conference on Advanced Computing Technologies
(ICACT), Rajampet, 2013, pp. 1-5

[21] E. Claverie, J. Lecoeur, V. Letort and P. Cournède, " Modeling soil
temperature to predict emergence, " 2016 IEEE International
Conference on Functional-Structural Plant Growth Modeling,
Simulation, Visualization and Applications (FSPMA), Qingdao, 2016,
pp. 28-37

[22] Y. Ge, Z. Liu and B. Wang, " Soil Water Prediction of Moving Dune
Based on BP Neural Network Model in Northwest Liaoning Sandy
Lindquist; 2010 Second International Conference on Computer Modeling
and Simulation, Sanya, Hainan, 2010, pp. 215-219

[23] P. A. Harlianto, T. B. Adji and N. A. Setiawan, " Comparison of
machine learning algorithms for soil type classification, " 2017 3rd
International Conference on Science and Technology - Computer
(ICST), Yogyakarta, 2017, pp. 7-10

[24] M. Paul, S. K. Vishwakarma and A. Verma, " Analysis of Soil
Behaviour and Prediction of Crop Yield Using Data Mining
Approach, " 2015 International Conference on Computational
Intelligence and Communication Networks (CICN), Jabalpur, 2015,
pp. 766-771

[25] S. Peng, T. Li and F. Wang, " Time Series Prediction Model of Soil
Moisture Based on Wavelet De-Noising, " 2009 International
Conference on Management and Service Science, Wuhan, 2009, pp.
1-4

[26] E. Santi et al., " Estimating Soil Moisture from C and X Band Sar
Using Machine Learning Algorithms and Compact Polarimetry, "
IGARSS 2018 - 2018 IEEE International Geoscience and Remote
Sensing Symposium, Valencia, 2018, pp. 1426-1429

[27] D. Smith and W. Peng, " Machine learning approaches for soil classification in a multi-agent deficit irrigation control system, " 2009 IEEE International Conference on Industrial Technology, Gippsland, VIC, 2009, pp. 1-6

[28] M. Veres, G. Lacey and G. W. Taylor, " Deep Learning Architectures for Soil Property Prediction, " 2015 12th Conference on Computer and Robot Vision, Halifax, NS, 2015, pp. 8-15

[29] Y. Liu, X. Pan, R. Shi, Y. Li, C. Wang and Z. Li, " Predicting Soil Salt Content Over Partially Vegetated Surfaces Using Non-Negative Matrix Factorization, " in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 8, no. 11, pp. 5305-5316, Nov. 2015

[30] Yong He, Hai-Yan Song, Pereira and Gomez, " Study on spectral feature characterizing and soil nitrogen content Prediction, " 2005 International Conference on Machine Learning and Cybernetics, Guangzhou, China, 2005, pp. 5424-5428 Vol. 9

[31] Yue Liu, Long Mei and Su Ki Ooi, " Prediction of soil moisture based on Extreme Learning Machine for an apple orchard, " 2014 IEEE 3rd International Conference on Cloud Computing and Intelligence Systems, Shenzhen, 2014, pp. 400-404

[32] M. Zhang, S. Wang, S. Li, J. Yi and P. Fu, " Prediction and map-making of soil organic matter of soil profile based on imaging spectroscopy: A case in Hubei China, " 2011 19th International Conference on Geoinformatics, Shanghai, 2011, pp. 1-5

[33] Mr. Harshad Bhujbal et al, "An Application For E-Marketing Of Agricultural Commodities and Analysis of Marketable Surplus" International Journal of Advanced Research in Computer and Communication Engineering, Vol. 7, Issue 4, April 2018

[34] Prachi S., Kirti P., Pooja N., Alisha S., Sushma V. (2018) Automatic Integration and Clustering of Marathi Documents in Different Formats for Effective Information Retrieval. In: Tiwari B., Tiwari V., Das K., Mishra D., Bansal J. (eds) Proceedings of International Conference on Recent Advancement on Computer and Communication. Lecture Notes in Networks and Systems, vol 34. Springer, Singapore

[35] Vinit Savant et al "A Survey on Farmer's Need and Feedback Analysis System" International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3, Issue 11,

November 2015

[36] R. Vispute, Sushma & Patil, Shrikant & Sangale, Sagar & Padwal,
Akshay & Ukarde, Avinash. (2015). Parallel Processing System for
Marathi Content Generation. Proceedings - 1st International
Conference on Computing, Communication, Control and Automation,
ICCUBEA 2015. 575-579. 10.1109/ICCUBEA.2015.118.

[37] Sirsat, Manisha & Cernadas, E & Fernandez-Delgado, M & Barro, S.
(2018). Automatic prediction of village-wise soil fertility for several
nutrients in India using a wide range of regression methods.
Computers and Electronics in Agriculture. 154.120-
133.10.1016/j.compag.2018.08.003.

[38] Vispute, S. R., Patil, S., Sangale, S., Padwal, A., & Ukarde, A.
(2015). Parallel Processing System for Marathi Content Generation.
2015 International Conference on Computing Communication
Control and Automation.

13 CERTIFICATE