SAVITRIBAI PHULE PUNE UNIVERSITY

A PROJECT REPORT ON

”CROP SELECTION METHOD BASED ON VARIOUS ENVIRONMENTAL FACTORS USING MACHINE LEARNING”

SUBMITTED TOWARDS THE
PARTIAL FULFILLMENT OF THE AWARD OF

BACHELOR OF ENGINEERING (INFORMATION TECHNOLOGY)

BY

Nishit Jain	**B121028534**
Amit Kumar	**B121028502**
Sahil Garud	**B121028511**
Vishal Pradhan	**B121028562**

Under The Guidance of
Mrs. Prajakta Kulkarni

Sinhgad Institutes

DEPARTMENT OF INFORMATION TECHNOLOGY
R.M.D. SINHGAD SCHOOL OF ENGINEERING
PUNE - 411058

2016-2017

A PROJECT REPORT

ON

CROP SELECTION METHOD BASED ON VARIOUS ENVIRONMENTAL FACTORS USING MACHINE LEARNING

BY

Nishit Jain	B121028534
Amit Kumar	B121028502
Sahil Garud	B121028511
Vishal Pradhan	B121028562

Under The Guidance of

Mrs. Prajakta Kulkarni

BACHELOR OF ENGINEERING

DEPARTMENT OF INFORMATION TECHNOLOGY

R.M.D. SINHGAD SCHOOL OF ENGINEERING

PUNE - 411058

SAVITRIBAI PHULE PUNE UNIVERSITY

2016-2017

Sinhgad Institutes

R.M.D. SINHGAD SCHOOL OF ENGINEERING
DEPARTMENT OF INFORMATION TECHNOLOGY

CERTIFICATE

This is to certify that the project entitled

Crop Selection Method Based On Various Environmental Factors Using Machine Learning

Submitted by

Nishit Jain	**B121028534**
Amit Kumar	**B121028502**
Sahil Garud	**B121028511**
Vishal Pradhan	**B121028562**

is a bonafide work carried out by Students under the supervision of Mrs. Prajakta Kulkarni and it is submitted towards the partial fulfillment of the requirement of Bachelor of Engineering (Information Technology) Project.

Date:

Place:

Mrs. Prajakta Kulkarni	**Dr. C.B. Bangal**	**Mrs. Sweta Kale**
(Internal Guide)	(Principal)	(Head of Department)

This report has been examined by us as per the Savitribai Phule Pune University, Pune requirements at RMD Sinhgad School of Engineering, Warje, Pune-58.

External Examiner: ____________________

Acknowledgments

It gives us great pleasure in presenting the preliminary project report on ***'Crop Selection Method Based On Various Environmental Factors Using Machine Learning'****.*

We would like to take this opportunity to thank our internal project guide ***Mrs. Prajakta Kulkarni*** *for giving us all the help and guidance we needed. We are really grateful to her for her kind support. Her valuable suggestions were very helpful.*

We are also grateful to ***Mrs. Sweta Kale****, Head of Information Technology Department, RMDSSOE for her indispensable support and suggestions. And, finally last but not the least, our respected Principal* ***Dr. C.B. Bangal****, for providing a wonderful infrastructure.*

Nishit Jain

Amit Kumar

Sahil Garud

Vishal Pradhan

Abstract

Agricultural planning plays an important role where economic growth of country like India and food supplies are considered. In a scenario, where crop yield rate is falling consistently, there is a need of a smart system which can solve the problem of decreasing crop yield rate. For farmers, its such a complex when there are more than one crops to sow. Therefore, to eliminate this problem, we propose a system which provides Crop Selection based on weather, regional water availability, soil type, market requirements, yield rate etc. to reap the maximum benefit out of it for the farmers which will subsequently help meet the elevating demands for food supplies.

INDEX

List of Figures

List of Tables

Chapter 1

Introduction

1.1 PROBLEM STATEMENT

To develop an web based software system using machine learning techniques with Java Servlets which provides a crop selection and sequencing method to maximize yield.

1.2 TECHNICAL KEYWORDS

1. Machine Learning
   (a) Machine Learning Algorithms
      i. WEKA machine learning API
         A. Binary Classifier
         B. Decision Trees

1. Learning Models
   (a) Training Model
      i. Training Data Set
         A. Test Data Set
         B. Classifiers

1.3 INTRODUCTION

- India is an agriculture-based economy whose major GDP growth is calculated on the basis of agriculture.
- Agriculture is the mainstay of a developing economy like India. Majority of its population depends on agriculture for their income. With depleting resources, reducing

land sizes and increase in input and labor costs, combined with the uncertainty of various factors like weather, market prices etc, agriculture in India has become a profession which is full of risks. The advancements in technology must be worked upon across various disciplines and it has already shown dramatic improvements in many fields. However, agriculture has not benefitted much from such advancements. Smart farming is the need of the hour of the Indian economy. Machine learning is an imminent field of computer science which can be applied to the farming sector quite effectively. It can facilitate the up-gradation of conventional farming techniques in the most cost-friendly approach. The purpose of this paper is to broaden the farming horizon by listing and evaluating the different applications of machine learning in Indian agriculture and to help the farmers advance their work up by many notches.

- Machine learning is the branch of computer science which is used to construct algorithms which exhibit self-learning property i.e. learning which is done by the machine itself hence the term ”Machine Learning”. It is considered to be one of the major areas under Artificial Intelligence. For a machine to become intelligent like a human mind, it has to first think and learn like a human. Human mind learns from past data and experiences that it is exposed to and based on that it takes decisions in future. Any conventional computer algorithm works as it is programmed by its developer. In other words, it will only follow the instructions given by its controller. For a machine to exhibit intelligence, it has to interpret and analyze the input and result data apart from simply following the instructions on that data. This is what the machine learning algorithms do.

- Earlier systems could only analyze the soil type or irrigation patterns or detect crop diseases but we propose a system which by summing up the analysis of all the affecting parameters, can suggest the most suitable crop(s) which will maximize yield.

1.4 MOTIVATION

- Degrading crop production in India.

- Rate of increasing farmer suicides.

- Integrating machine learning with agriculture can help maximize yield.

1.5 GOALS AND OBJECTIVES

- To implement machine learning algorithms to analyze various crop affecting parameters.
- To suggest suitable crop(s)by summing up analysis obtained from various data sets using various machine learning algorithms.

1.6 PRE-REQUISITES

Project Domain: Java Servlets MVC - Machine Learning

- In-depth knowledge of Java Programming language.
- Basic machine learning knowledge.
- Familiarity with the WEKA API.
- Knowledge of Data Analytics.

1.7 SOFTWARE AND HARDWARE REQUIREMENTS

1.8 HARDWARE RESOURCES REQUIRED

Table 1.1: Hardware Requirements

Sr. No.	Parameter	Minimum Requirement
1	CPU Speed	2 GHz
2	RAM	3 GB
3	Storage	2 GB

Software requirements:

- Windows 7/8/10.
  - Windows 7/8/10 both operating systems are todays most advance featured operating systems provided by microsoft Windows 8/10 introduces what Microsoft described as ”universal apps”; expanding on Metro-style apps, these apps can be designed to run across multiple Microsoft product families with nearly identical codeincluding PCs, tablets, smartphones, embedded systems. These qualities of those operating systems makes creating such machine learning based systems in Java that works on multiple devices very easy and convenient.

- JDK 1.7 or later version installed.
  - The JDK is the Java Development Kit. It contains the JRE as well as a lot of other useful stuff for developing Java applications. That includes the compiler obviously (which is also contained in the JRE for some good reason, but you can ignore this fact now), the JAR utility to create .jar files, many tools for "decompiling" class files, inspect .jar files, repackage them, etc.
  - Different versions of Java (1.0, 1.1, etc. all the way to 1.7, also known as Java 7) usually contain improvements to both the JVM and the standard library, so the two usually need to run together, and are packaged together in the JRE.
- WEKA Machine Learning Java API (for decision tree learning, K-Nearest Neighbor and other greedy tree algorithms).
  - Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes.
  - Found only on the islands of New Zealand, the Weka is a flightless bird with an inquisitive nature. The name is pronounced like this, and the bird sounds like this.
  - Weka is open source software issued under the GNU General Public License.
- Any IDE well integrated with Java (e.g. NetBeans IDE).
  - NetBeans IDE is the official IDE for Java 8. With its editors, code analyzers, and converters, you can quickly and smoothly upgrade your applications to use new Java 8 language constructs, such as lambdas, functional operations, and method references.

Chapter 2

Literature Survey

2.1 LITERATURE REVIEW

1. *Rakesh Kumar, M.P. Singh, Prabhat Kumar and J.P. Singh,* ***Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique*** *2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), 6 - 8 May 2015. pp.138-145.*

   - Rakesh Kumar, M.P. Singh, Prabhat Kumar and J.P. Singh were the students of Dept. of CSE, National Institute of Technology, Patna.
   - They have discussed various machine learning methods to maximize crop yield using machine learning using the yield criteria of each crop.
   - They portrayed very well the application of machine learning algorithms to conclude upon the best possible crop which can provide maximum yield.
   - They used sowing data sets of all the crops and compared their yields in kg per hectare.
   - **Algorithms used:**
     - SVM - Support Vector Machine
     - ANN - Artificial Neural Network

2. *K. Krishna Kumar, K. Rupa Kumar, R. G. Ashrit, N. R. Deshpande and J. W. Hansen,* ***Climate Impacts on Indian Agriculture.*** *International Journal of climatology, 24: 13751393, 2004.*

- K. Krishna Kumar, K. Rupa Kumar, R. G. Ashrit andN. R. Deshpande are meteorologists at Indian Institute of Tropical Meteorology and J. W. Hansen is a research scientist at International Research Institute for Climate Prediction, located in Palisades, USA.
- They have discussed various climatic parameters having impact on agriculture and the extent of the effect. They have elaborated these effects as well.
- They conducted various studies in the past ranging from the data of 1960s to 2000s and have predicted the effect of climate on various crops in terms of yield.
- They studied various weather patterns to understand the reasons behind fluctuations in crop yield.
- They analyzed cropclimate relationships for India, using historic production statistics for major crops (rice, wheat, groundnut and sugarcane) and for aggregate food grain, cereal, pulses and oilseed production.

3. *Karandeep Kaur,* ***Machine Learning: Applications in Indian Agriculture.*** *International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 4, April 2016.*

    - Karandeep Kaur is an assistant professor in Department of Computer Science, Guru Nanak Dev University located at Amritsar.
    - It explains applications of various machine learning techniques in agricultural field and how they can be used to leverage benefits out of it.
    - It also summarizes the techniques used by various other publishers as well as scientists in parts of agriculture to leverage yield, predict crop diseases, smarter irrigation systems etc.
    - Better illustrates and classifies which machine learning technique is suitable for exploring which part of the agricultural field.
    - Lists and evaluates the different applications of machine learning in Indian agriculture and to help the farmers advance their work up by many notches.

2.2 LIMITATIONS

- Could not evaluate the crop yield by considering the environmental factors such as soil chemical composition, temperature, humidity etc. which play a vital role.
- Used only statistical model in analyzing climate impact on agriculture whereas data mining could have provided a way better analysis.
- Some of the methods suggested for the analysis do have a better substitute. For example, the review paper suggests to use classification algorithms to predict yield of the crops whereas Back Propagation Feed Forward Neural Network is a better option.

2.3 EXISTING SYSTEM

- There are systems in which physical sensors are used to detect soil composition, field humidity content, temperature etc. But, they are so expensive which makes it impractical to be fitted in every field.
- Also, such systems only provide information about the quality of soil etc. They do not suggest crop(s) which can be planted over a season to gain maximum yield.
- There are systems which analyze all the information according to the field and provide a detailed analysis but they do not suggest a crop by integrating the analysis from both data i.e. weather data as well as yield information.

2.4 PROPOSED SYSTEM

- Crop Selection Method (by predicting crop(s) based on weather data, soil data, humidity and temperature data etc. and yield information to suggest the final crop).
- Crop Sequencing Method (predicting set of crop(s) which can be planted in a specific sequence over the year to reap the maximum yield out of it.

Chapter 3

Project Description

3.1 PROJECT IDEA

The idea of the project is to build a system that would help the farmers find an optimal solution to the problem of selection of crop(s) to grow in one season. Also, it would help them with selection of the sequence of crop(s) to be grown in a single agricultural year.

3.2 PROJECT AIM

- To Maximize yield of the crop(s).

3.3 INPUT AND OUTPUT OF THE SYSTEM

3.3.1 Crop Selection Method

- **Input required:**
  - Average rainfall of the area.
  - Average Temperature of the area.
  - Average humidity Content of the area.
  - Soil Chemical Composition (e.g. phosphate, nitrogen, water etc.) of the field soil.
- **Output generated:**
  - A single crop which can provide the maximum yield according to the input.

3.3.2 Crop Sequencing Method

- **Input required:**
  - Different set(s) of crop(s).
- **Output generated:**
  - A single set of crp(s) with crops arranged in a sequence which will provide maximum yield.

3.3.3 Statement of scope

Software will be compact in size, minimum number of input required from user and the output generated will be according to the preferences set by the user.

Description:

- Input will be set through forms.
- Bounds of input will be minimum.
- Input validation will be applicable on the Input form according to the method used.
- No/minimum input dependency.

The Scope of the project is to find a crop which will provide maximum yield according to the inpu(s) supplied and it will also provide a mechanism to get a specific sequence for the whole growing season.

3.4 SOFTWARE CONTEXT

We do different feature transformations on available data sets for training the crop selection and sequencing model to generate a good trained model.

3.5 MAJOR CONSTRAINTS

Since the project build is on the Java platform using machine learning WEKA API, user will have to supply the input in a specific form to maximise accuracy of the system and avoid errors.

3.6 OUTCOME

- A Single crop providing maximum yield.
- A set of crop in a sequence which will provide maximum yield for the whole growing year.
- Performing various data mining operations on crop statistics of India to study various yield patterns over the time in various parts of the country.

3.7 APPLICATIONS

- Major application of the project is for the farmers and agriculturists to solve the confsion about the type of crop(s) they should grow and gain maximum yield.
- Studying various various yield fluctuation patterns in various regions of the country helps us gain deeper knowledge about the types of major crops for various regions of the country.
- Crop Sequencing method can help farmers gain proper knowledge about proper sequence of the crops of various season time.

Chapter 4

Project Design

4.1 INTRODUCTION

4.1.1 Purpose and Scope of Document

- This document contains detailed Software Development Specification of our project - ”CROP SELECTION BASED ON VARIOUS ENVIRONMENTAL FACTORS USING MACHINE LEARNING”. Document will be helpful to the developer for implementation of project, to testers for testing of this project and also to others who wants to do any changes in this system.
- Also, it describes how we have performed KDD on various datasets for knowledge extraction.

4.1.2 Overview of responsibilities of Developer

- Coordinate Project Guides and Project coordinators and other contributors on current programming tasks.
- Collaborate with other programmers to design and implement features.
- Quickly produce well organized,optimized and documented source code.
- Contribute to technical design documentation
- Continuously learn and improve skills
- Collaborating with other software developers,business analysts and software architects to plan, design,develop,test and maintain web based applications built on various web technologies.

4.2 USAGE SCENARIO

This section provides various usage scenarios for the system to be developed.

4.2.1 User profiles

- **User**:- Authenticate, supply required input in specified format.
- **Data Scientist**:- Training the crop selector model on various data set rigorously.
- **Model**:- Predicting crop on the basis of test data supplied.

4.2.2 Use Case View

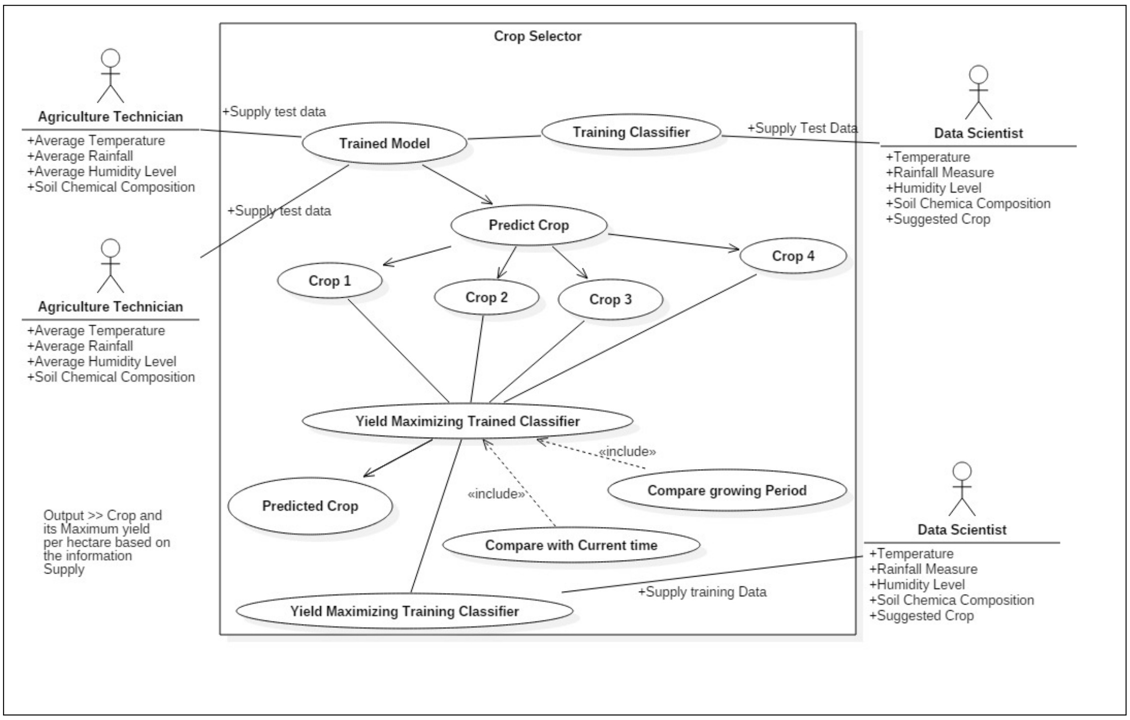


Figure 4.1: Use case diagram - Crop Selector

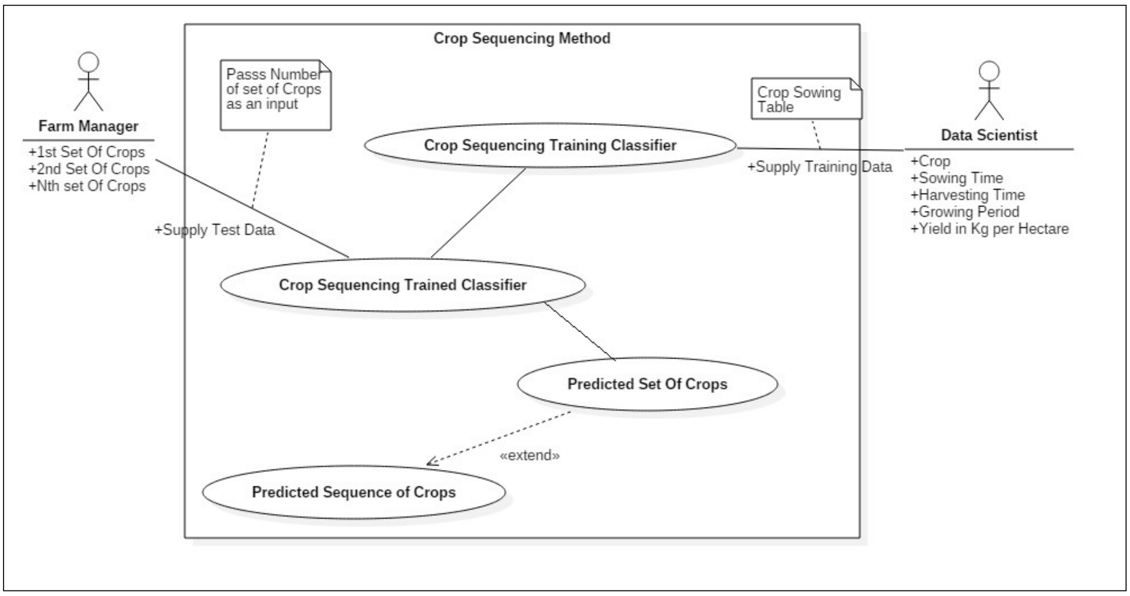


Figure 4.2: Use case diagram - Crop Sequencer

4.2.3 Activity Diagram

Figure 4.3: Activity diagram - Crop Selector

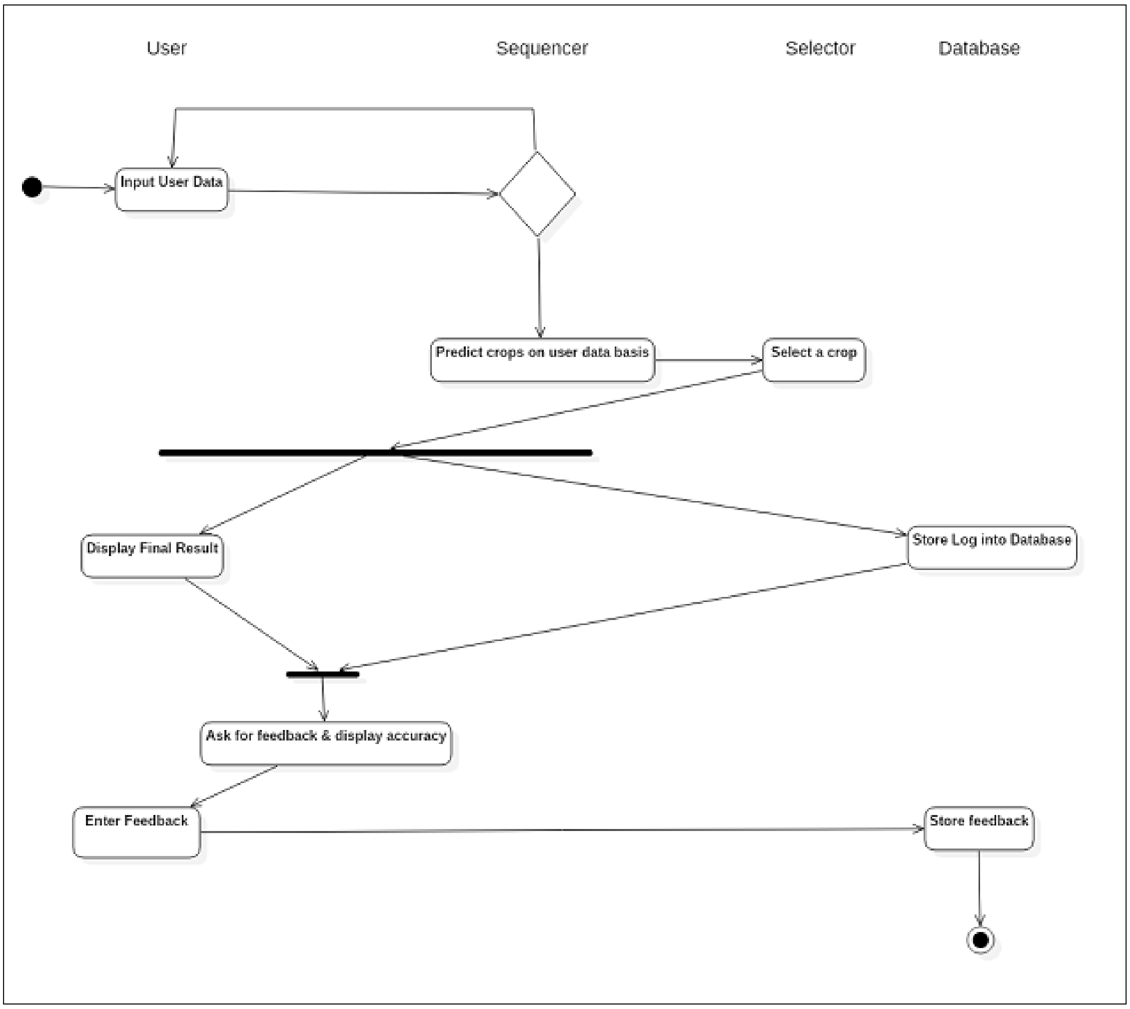
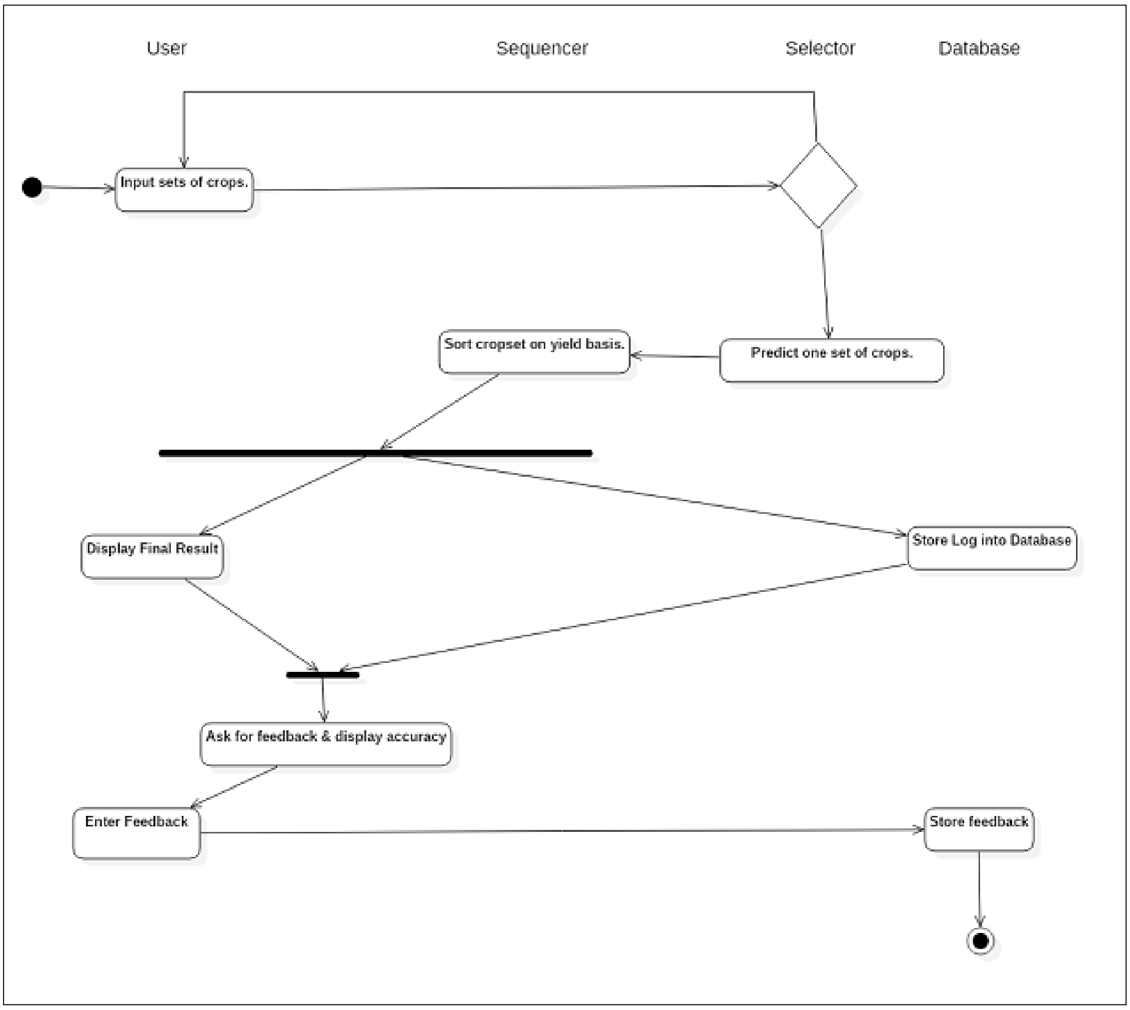


Figure 4.4: Activity diagram - Crop Sequencer



4.2.4 Class Diagram

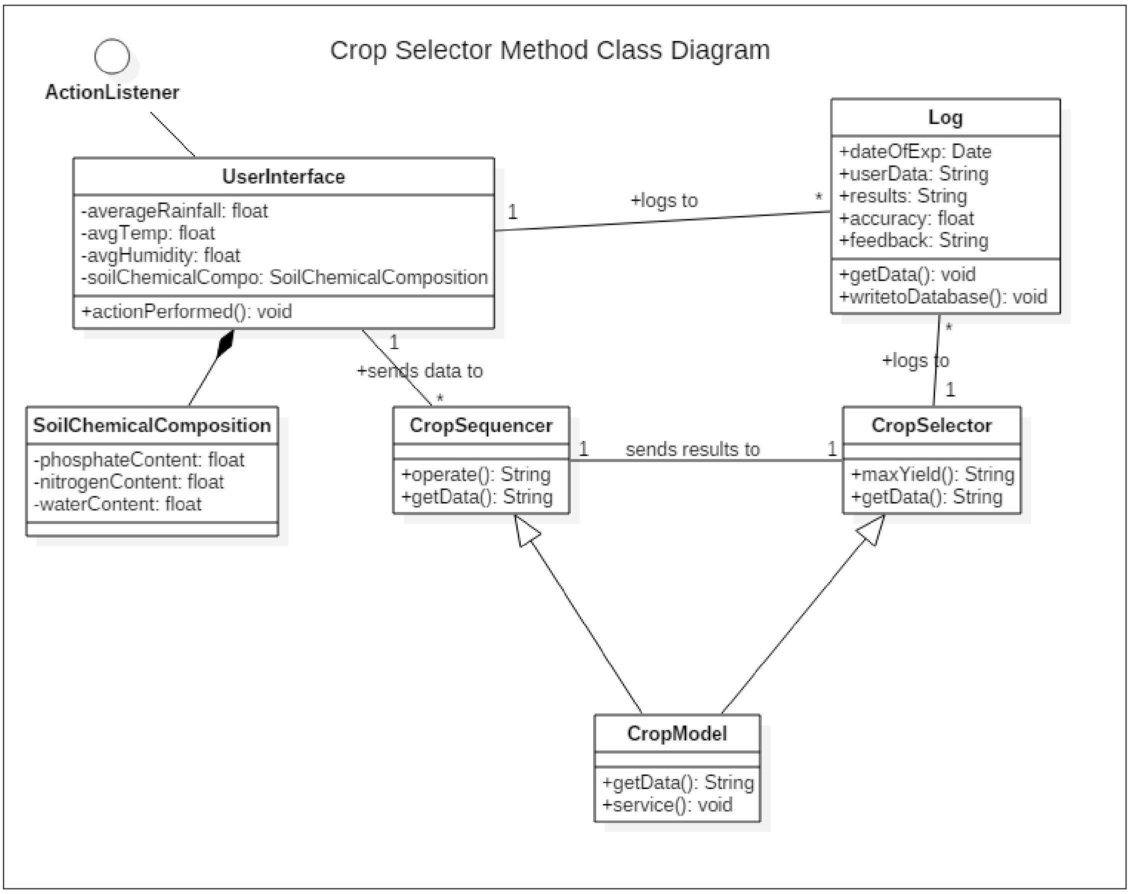


Figure 4.5: Class Diagram - Crop Selector

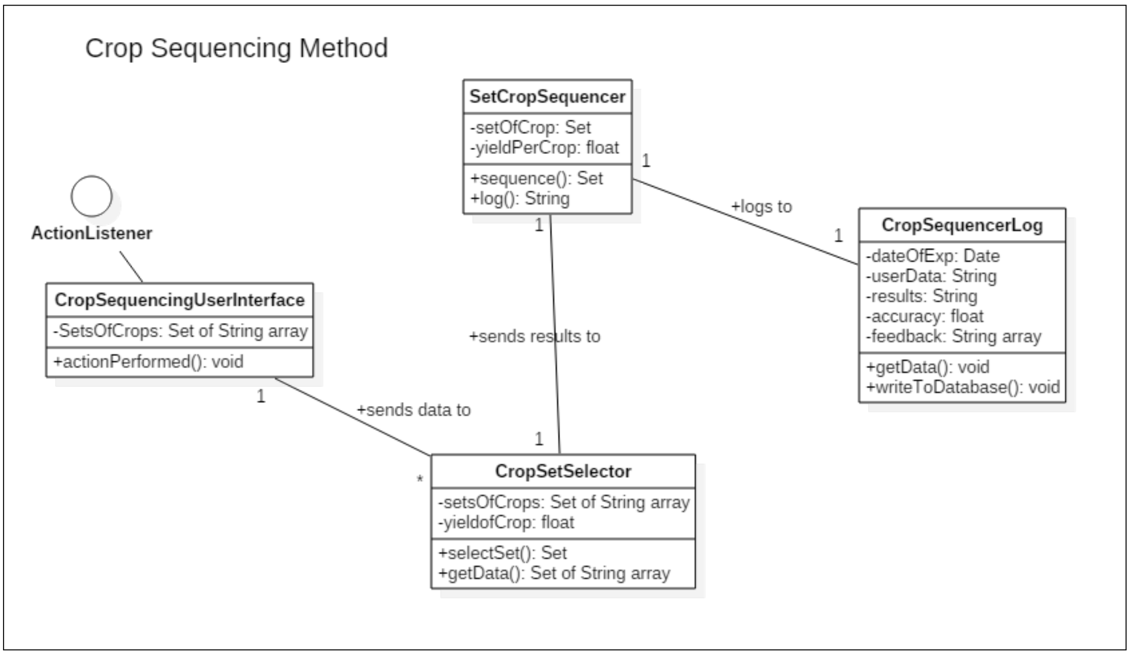


Figure 4.6: Class Diagram - Crop Sequencer

4.2.5 State Diagram

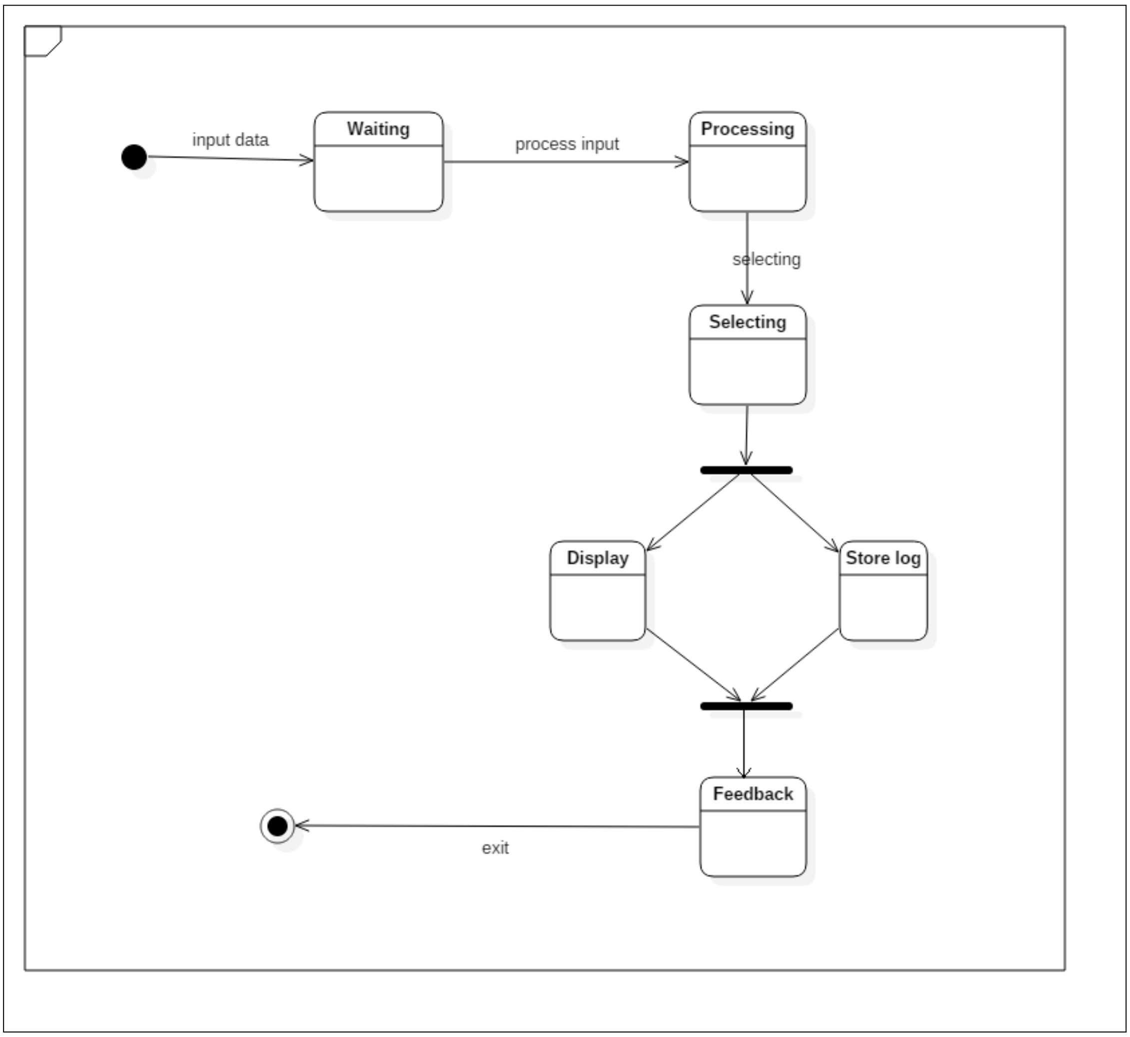


Figure 4.7: State Diagram

4.2.6 Sequence Diagram

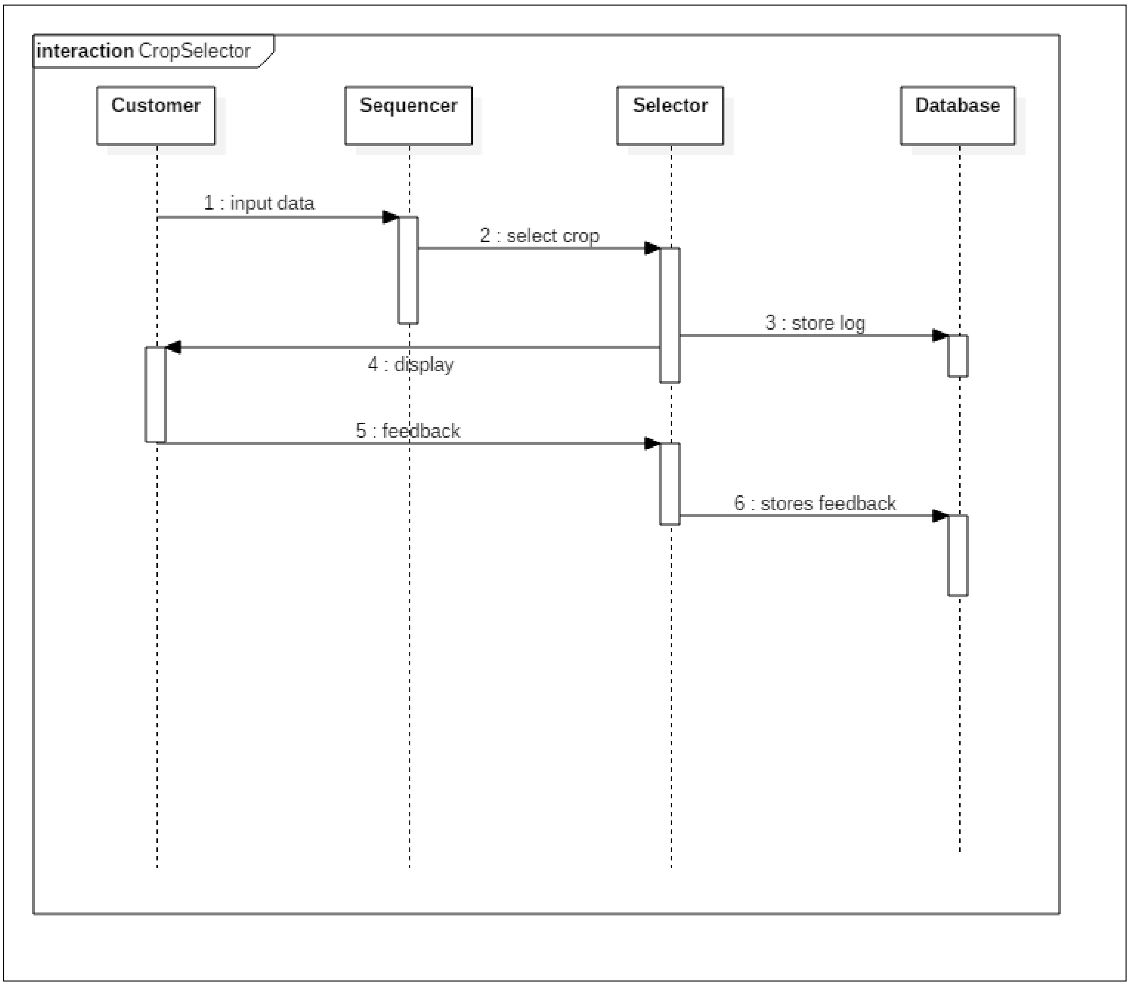


Figure 4.8: Sequence Diagram - Crop Selector

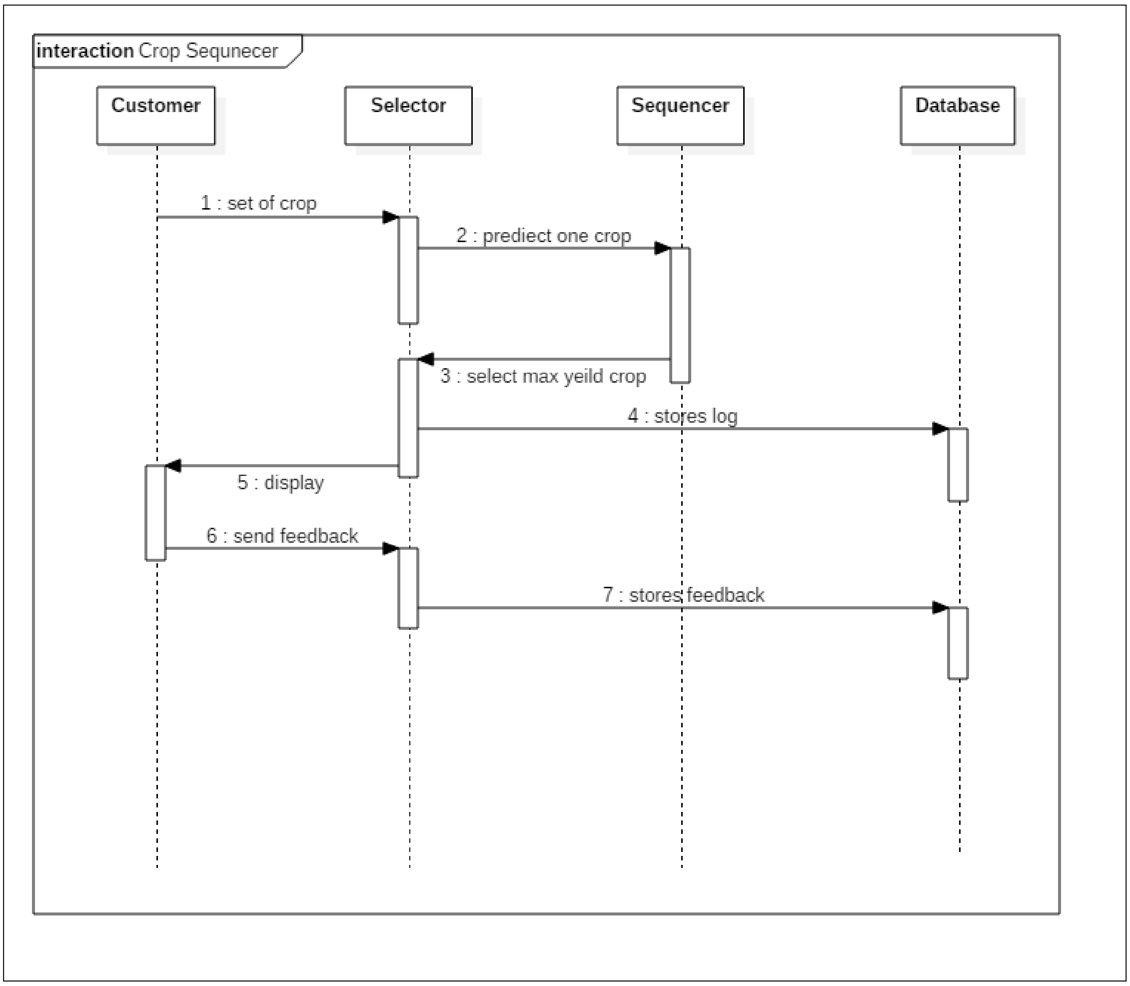


Figure 4.9: Sequence Diagram - Crop Sequencer

Chapter 5

Project Plan

5.1 PROJECT ESTIMATES

Our system can be developed in few phases. The result of earlier phase acts as key point for the next phase. Inter software engineering, we can define as a set of activities whose goal is development or evolution of software. When a process is defined a particular model gets associated with that process so as to implement that result in an orderly manner and user gets the expected output. Our project supports iterative model which is a kind of generics of software process model.This model works in 5 phases as described below and is shown in fig 4.1.

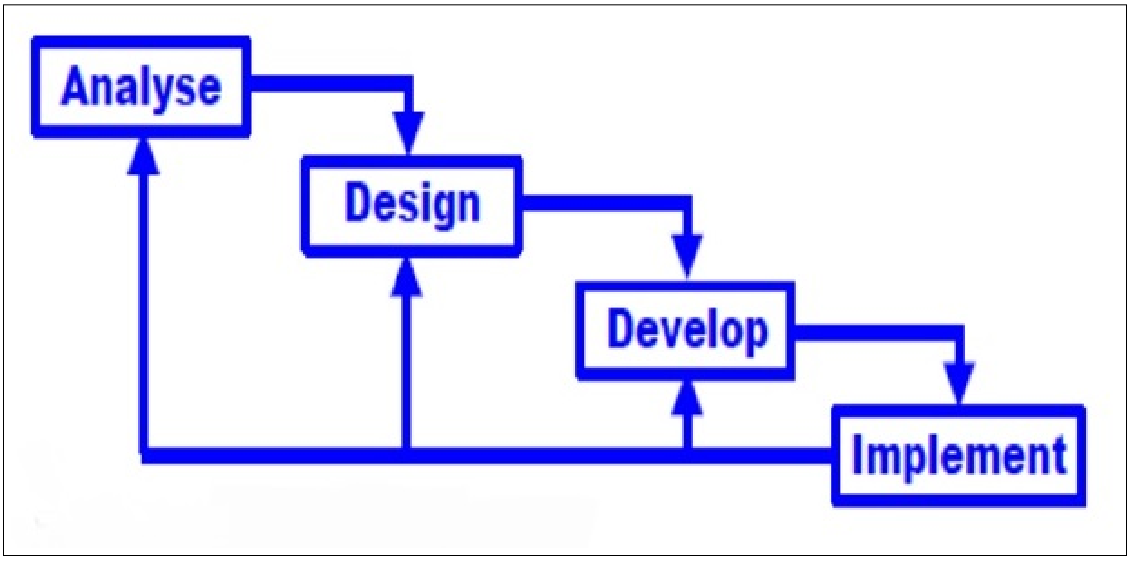


Figure 5.1: Iterative Waterfall Model

- **Analyse**:- In this stage all the required documents are collected and use cases are prepared. All the research work is done here.

- **Design**:- In this stage system architecture is created. A basic idea of what is to be done is formulated.
- **Develop**:-This is the stage where all the coding is done to create the software. Data storage and retrieval tasks are carried on.
- **Implement and Maintenance**:- Installation of the software is done along with testing and debugging. Error checking is done. All the capabilities of the project are optimized if possible.
- **Iteration**:- After completion of each cycle, the model is iterated again and again to re-evaluate, fix the subltle errors and optimize he performane of the system. Iterative model helps in modifying the requirements even if there is a change at a later stage.

5.1.0.1 Cost Estimate

The cost required to develop the application can be evaluated as, Internet cost, literature survey cost, material cost and various other resources cost. If some cost are not identified and if any miscellaneous costing occurs than the cost may get increased drastically

5.1.0.2 Time Estimates

The time required can be evaluated as, time for literature survey, planning and designing time, time for collecting the right data sets, time for analyzing error or success report and finding alternate solutions for the same, Construction and development time.

5.2 RISK IDENTIFICATION

For risks identification, review of scope document, requirements specifications and schedule is done. Answers to questionnaire revealed some risks. Each risk is categorized as per the categories mentioned in [**?**].

1. Power failure leading to data and state loss.
2. When the server stops working due to some hardware problem.
3. Malicious user intervention
4. Database Failure may occur
5. Operational failure may occur if wrong input is given to the system.
6. Interdependent modules may stop working.

7. Release of similar product by another team.

8. Unavailability of hardware/software tools.

9. Network failure.

5.2.1 Risk Analysis

Table 5.1: Risk Analysis

Sr. No.	Risk Desription	Probability
1	Network Inconsistencies	Low
2	Server Unresponsive	Low
3	System Failure	Low
4	Database Failure	Low
3	Wrong Input to the System	Medium

Table 5.2: Risk Analysis (Impact On Schedule and Quality

Sr. No.	Risk Desription	Schedule	Quality
1	Network Inconsistencies	Low	High
2	Server Unresponsive	Low	High
3	System Failure	High	High
4	Database Failure	High	High
3	Wrong Input to the System	Low	High

Table 5.3: Risk Analysis (Overall Effect)

Sr. No.	Risk Desription	Overall Effect
1	Network Inconsistencies	High
2	Server Unresponsive	High
3	System Failure	High
4	Database Failure	High
3	Wrong Input to the System	High

5.3 PROJECT SCHEDULE

5.3.1 Project task set

Major Tasks in the Project stages are:

- Task 1: Topic Search:
  Search the topic for project in the area of interest by studying the recent developments in various domains of Informtion Technology.

- Task 2: Literature Survey:
  Study of existing methodologies and research conducted on project topic and their drawbacks or improvements they can have.

- Task 3: Risk Analysis:
  Probable risks involved in project are analysed and accordingly actions are taken to avoid occurrence of risks while project is being developed.

- Task 4: Design:
  Designing of Software architecture and various modules. Modelling the flow of data and identifying the process flow and activities involved in the project.

- Task 5: Prototyping:
  Prototype or replica is a preliminary version of how software will be. It is made so as to get deep idea of all the required modules.

- Task 6: Implementation:
  Construction of all the modules, coding and development phase, integrating all the modules to develop different builds of the software using spiral model of SDLC.

- Task 7: Testing:
  Testing of the built software my writing different test suites and executing on the software to uncover bugs if any. Debugging the software to remove bugs.

- Task 8: Documentation and deployment:
  Preparing the final document of each phase in this project. Starting from requirement gathering phase through all the phases of SDLC. Writing complete report for this project including conclusion and future scope.

Table 5.4: Project schedule

Phase	Task	Description
Phase 1	Topic Search	Search topic for project from area of interest
Phase 2	Literature survey	Collect raw data and elaborate on literature surveys
Phase 3	Analysis	Analyze all the information on the selected topic
Phase 4	Design	Assign the module and design the process flow control.
Phase 5	Prototyping	Prototype of the final project to get deep idea of all modules.
Phase 6	Implementation	Implement the code for all the modules and integrate all the modules.
Phase 7	Testing	Test the code
Phase 8	Documentation	Prepare the doucment for this project with conclusion

5.4 TEAM ORGANIZATION

The manner in which staff is organized and the mechanisms for reporting are noted.

5.4.1 Team structure

Table 5.5: Team Structure

Sr No.	Name	Role	Responsibilities
1	Mrs. Prajakta Kulkarni	Project Guide	Guidance for selection of project platform and topic. Monitoring weekly status and reviews. Feedback and suggestions for each module. Solving queries and doubts
2	Nishit Jain	Team Member	Project idea, Requirement gathering, UML diagrams, GUI development, Algorithm development, Project coding, Project report and Project paper preparation.
3	Amit Kumar	Team Member	Requirement gathering, UML diagrams, GUI development, Algorithm development, Project coding, Project report and Project paper preparation.
4	Sahil Garud	Team Member	Requirement gathering, Project report and Project paper preparation.
5	Vishal Pradhan	Team Member	Requirement gathering, UML diagrams, Project report and Project paper preparation.

Chapter 6

Project Testing

6.1 INTRODUCTION TO WEKA CLASSIFIERS

The WEKA classifiers that we have used in our project are:

- J48 Classifier
- Naive Bayes Classifier
- Sequential Minimal Optimization

6.1.1 J48 Classifier

J48 is an open source Java implementation of the C4.5 algorithm in the Weka data mining tool.C4.5 builds decision trees from a set of training data in the same way as ID3, using the concept of information entropy. At each node of the tree, C4.5 chooses the attribute of the data that most effectively splits its set of samples into subsets enriched in one class or the other. The splitting criterion is the normalized information gain (difference in entropy). The attribute with the highest normalized information gain is chosen to make the decision. The C4.5 algorithm then recurs on the smaller sublists.

This algorithm has a few base cases.

- All the samples in the list belong to the same class. When this happens, it simply creates a leaf node for the decision tree saying to choose that class.
- None of the features provide any information gain. In this case, C4.5 creates a decision node higher up the tree using the expected value of the class.
- Instance of previously-unseen class encountered. Again, C4.5 creates a decision node higher up the tree using the expected value.

6.1.1.1 Pseudocode

In pseudocode, the general algorithm for building decision trees is:

1. Check for the above base cases.
2. For each attribute a, find the normalized information gain ratio from splitting on a.
3. Let 'abest' be the attribute with the highest normalized information gain.
4. Create a decision node that splits on 'abest'.
5. Recur on the sub lists obtained by splitting on 'abest', and add those nodes as children of node.

6.1.2 Naive Bayes Classifier

Naive Bayes is a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. It is not a single algorithm for training such classifiers, but a family of algorithms based on a common principle: all naive Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable. For example, a fruit may be considered to be an apple if it is red, round, and about 10 cm in diameter. A naive Bayes classifier considers each of these features to contribute independently to the probability that this fruit is an apple, regardless of any possible correlations between the color, roundness, and diameter features. For some types of probability models, naive Bayes classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for naive Bayes models uses the method of maximum likelihood; in other words, one can work with the naive Bayes model without accepting Bayesian probability or using any Bayesian methods.

6.1.2.1 Algorithm Explanation

Naive Bayes model is easy to build and particularly useful for very large data sets. Along with simplicity, Naive Bayes is known to outperform even highly sophisticated classification methods.

Bayes theorem provides a way of calculating posterior probability P(c—x) from P(c), P(x) and P(x—c). Look at the equation below:

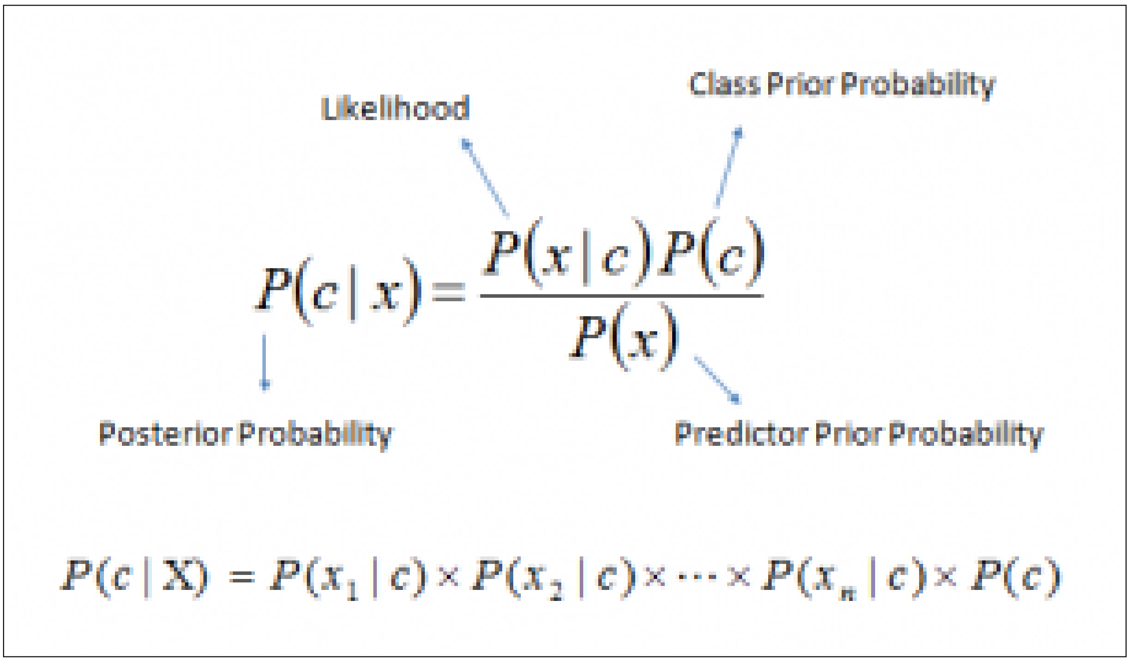


Figure 6.1: Naive Bayes Algorithm

6.1.2.2 Working

1. Convert the data set into a frequency table.

2. Create Likelihood table by finding the probabilities like Overcast probability = 0.29 and probability of playing is 0.64.

Weather	Play
Sunny	No
Overcast	Yes
Rainy	Yes
Sunny	Yes
Sunny	Yes
Overcast	Yes
Rainy	No
Rainy	No
Sunny	Yes
Rainy	Yes
Sunny	No
Overcast	Yes
Overcast	Yes
Rainy	No

Frequency Table		
Weather	No	Yes
Overcast		4
Rainy	3	2
Sunny	2	3
Grand Total	5	9

Likelihood table				
Weather	No	Yes		
Overcast		4	=4/14	0.29
Rainy	3	2	=5/14	0.36
Sunny	2	3	=5/14	0.36
All	5	9		
	=5/14	=9/14		
	0.36	0.64		

Figure 6.2: Naive Bayes Algorithm Example Data

3. Now, use Naive Bayesian equation to calculate the posterior probability for each class. The class with the highest posterior probability is the outcome of prediction.

6.1.3 Sequential Minimal Optimization

Sequential minimal optimization (SMO) is an algorithm for solving the quadratic programming (QP) problem that arises during the training of support vector machines. It was invented by John Platt in 1998 at Microsoft Research. SMO is widely used for training support vector machines and is implemented by the popular LIBSVM tool. The publication of the SMO algorithm in 1998 has generated a lot of excitement in the SVM community, as previously available methods for SVM training were much more complex and required expensive third-party QP solvers.

6.2 UNIT TESTING

Unit Testing is a level of software testing where individual units/ components of a software are tested. The purpose is to validate that each unit of the software performs as designed.

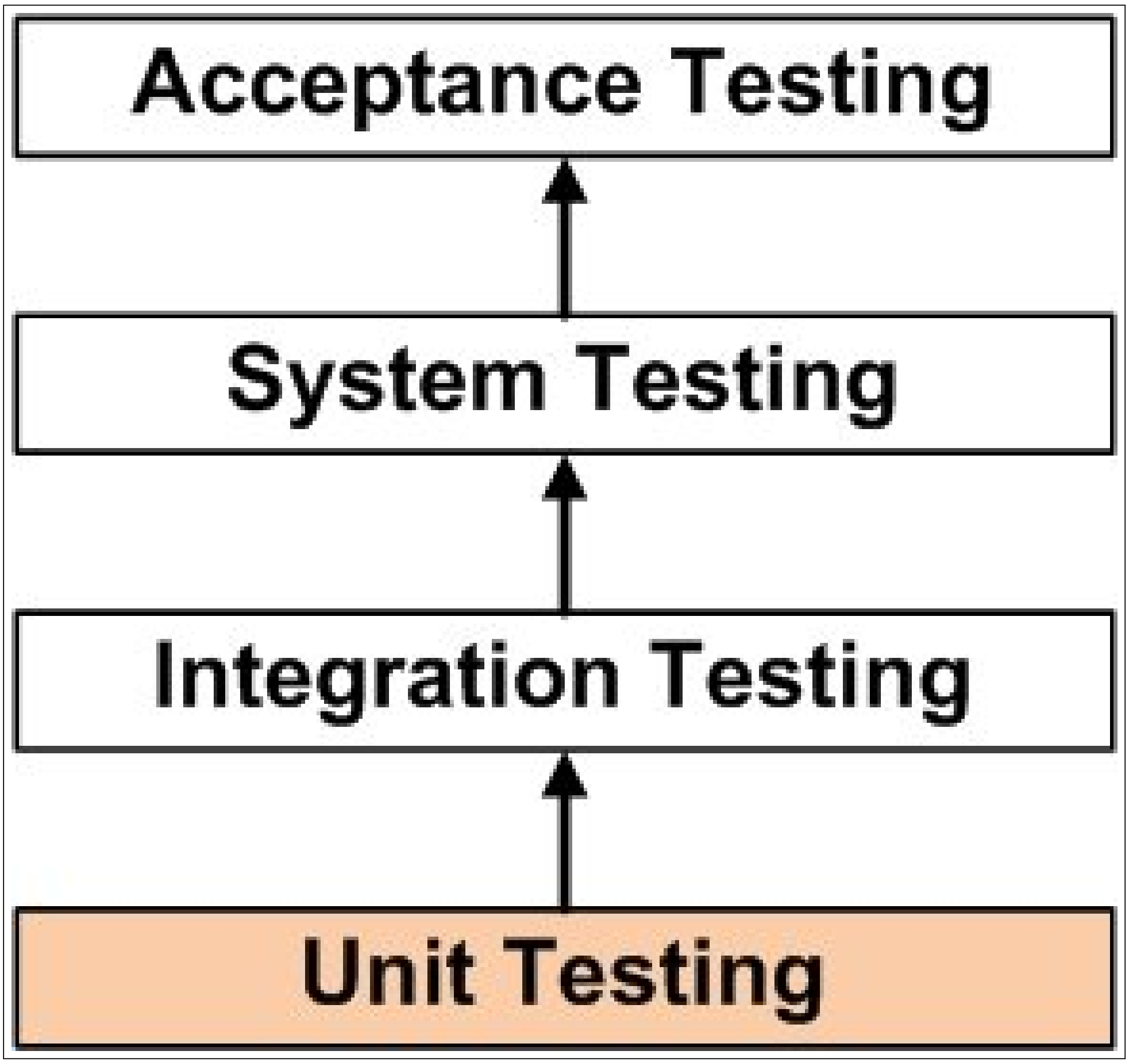


Figure 6.3: Unit Testing

A unit is the smallest testable part of 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.)

6.2.1 Test Cases

The application was tested for the following test cases:

1. Test Case : Opening Crop Selecton Method Window.

    - Test Data : Click on the button ”Crop Selection Method”.
    - Expected Result : Loads Crop Selection Method multi-tabbed window.
    - Action Result : Window loading successful.
    - Test Case result : Pass

2. Test Case : Loading Training Data in ’J48’ and ’Naive Bayes’ tab.

    - Test Data : Click on ’Load Training Data’ to load the file in the given textarea.
    - Expected Result : Loads ’cropnametrain.arff’ training dataset into textarea successfully from the specified location.
    - Action result : Training Dataset loading successful.
    - Test Case Result : Pass

3. Test Case : Loading Test Data in ’J48’ and ’Naive Bayes’ tab.

    - Test Data : Click on ’Load Test Data’ to load the file in the given textarea.
    - Expected Result : Loads ’cropnametest.arff’ training dataset into textarea successfully from the specified location.
    - Action result : Training Dataset loading successful.
    - Test Case Result : Pass

4. Test Case : Classifying the test data and loading the classified output (with the specified file name) in the provided textarea.

    - Test Data : Click on ’Classify’ to generate the classified output file with the specified file name.

- Expected Result : Classifies and loads classified output file.
- Action Result : Output file loading successful.
- Test Case Result : Pass

5. Test Case : Opening Crop Sequencing Method Window.
    - Test Data : Click on the button ”Crop Sequencing Method”.
    - Expected Result : Loads Crop Sequencing Method window.
    - Action Result : Window loading successful.
    - Test Case result : Pass

6. Test Case : Loading data in JTable.
    - Test Data : Click on the button ”Run Sequencer”.
    - Expected Result : Loads desired data in JTable.
    - Action Result : Data loading successful.
    - Test Case result : Pass

Hence, we have successfully tested our project on every possible test case.

6.3 PROJECT SCREENSHOTS

1. Home Screen

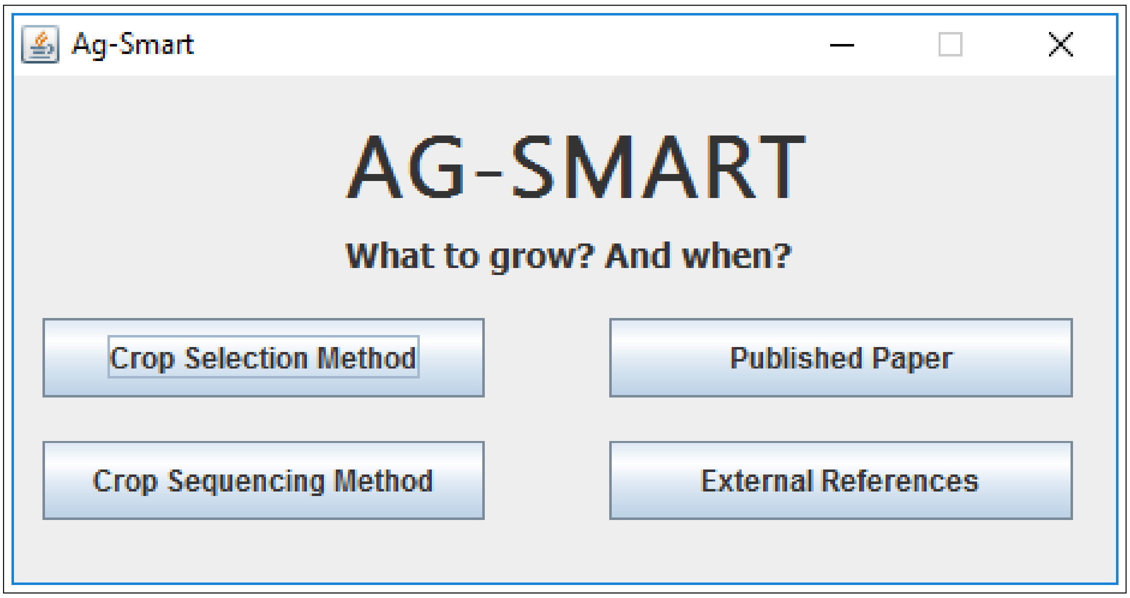


Figure 6.4: Home Screen

2. Crop Selection Method multi-tabbed window - J48 Classifier (Loading Training Data).

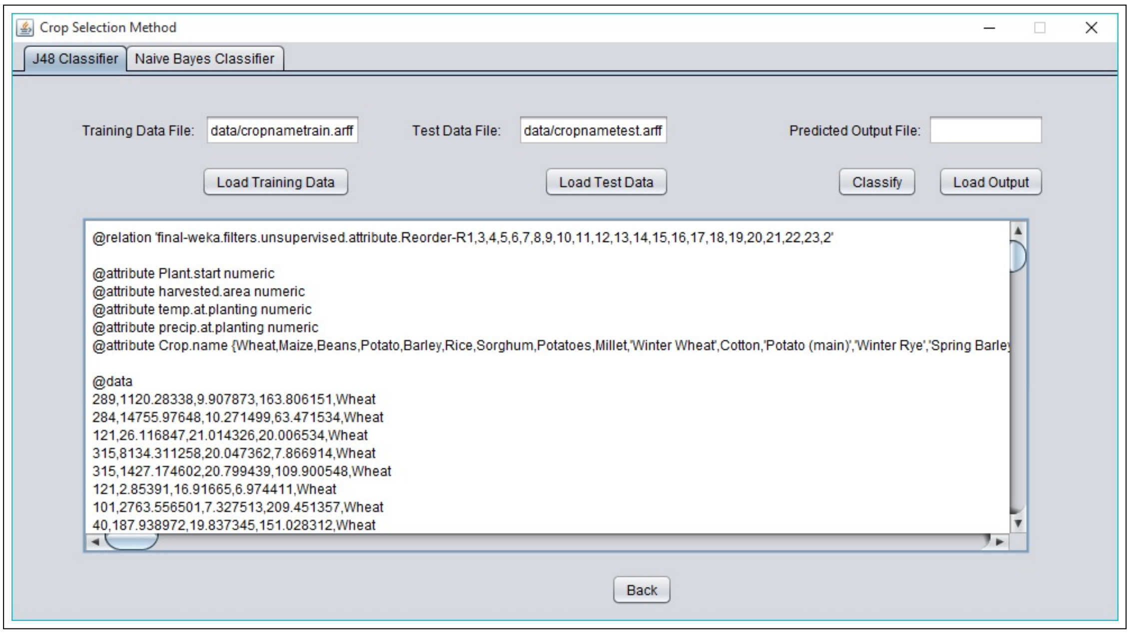


Figure 6.5: Loading Training Data - J48 Classifier

3. Crop Selection Method multi-tabbed window - J48 Classifier (Loading Test Data).

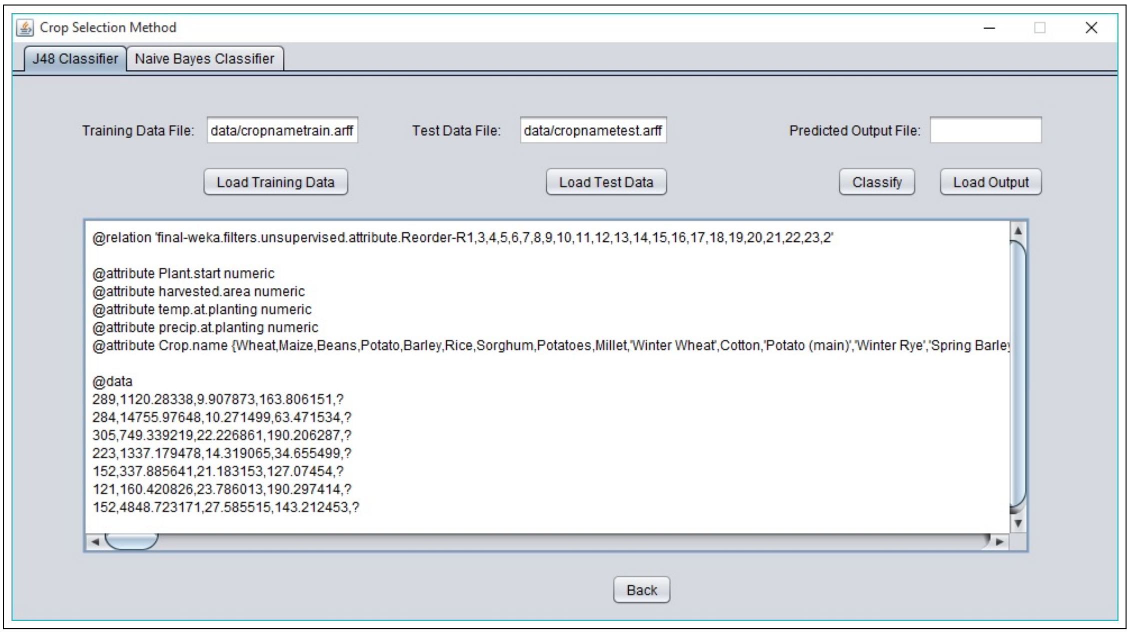


Figure 6.6: Loading Test Data - J48 Classifier

4. Crop Selection Method multi-tabbed window - J48 Classifier (Classifying Test Data).

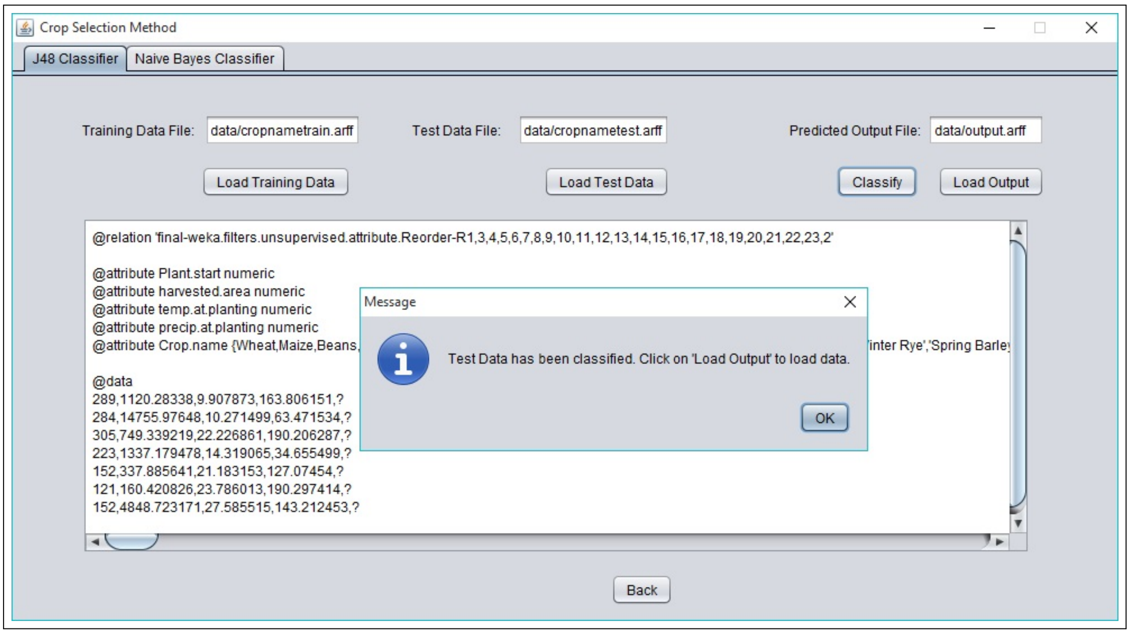


Figure 6.7: Loading Test Data - J48 Classifier

5. Crop Selection Method multi-tabbed window - J48 Classifier (Loading classified test Data).

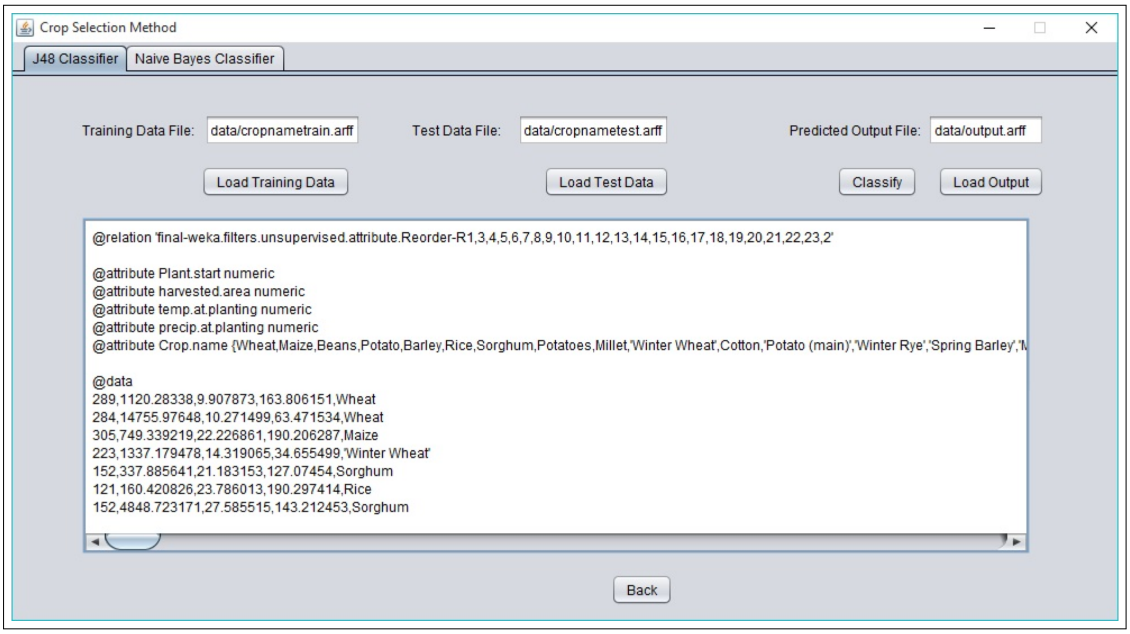


Figure 6.8: Loading classified test Data - J48 Classifier

6. Crop Selection Method multi-tabbed window - Naive Bayes Classifier (Loading Training Data).

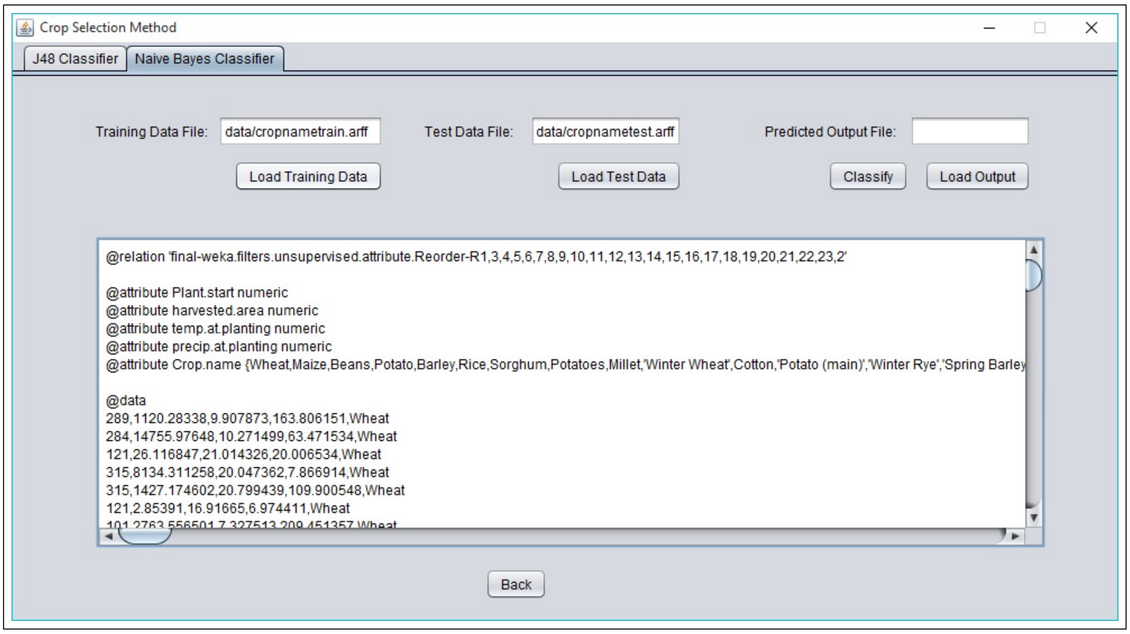


Figure 6.9: Loading Training Data - Naive Bayes Classifier

7. Crop Selection Method multi-tabbed window - Naive Bayes Classifier (Loading Test Data).

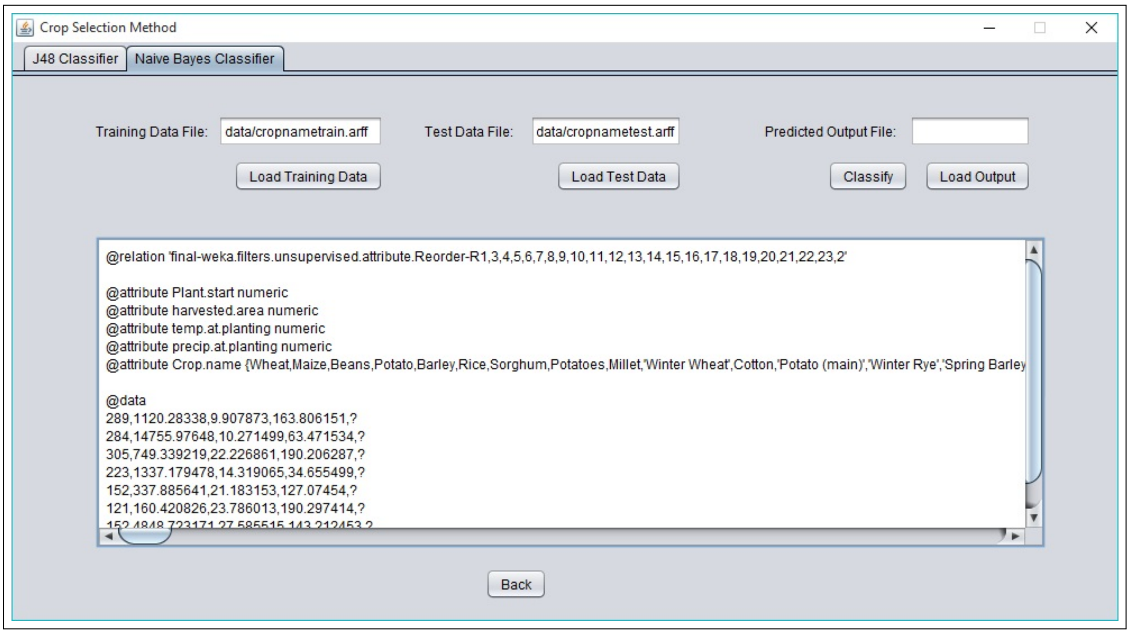


Figure 6.10: Loading Test Data - Naive Bayes Classifier

8. Crop Selection Method multi-tabbed window - Naive Bayes Classifier (Classifying Test Data).

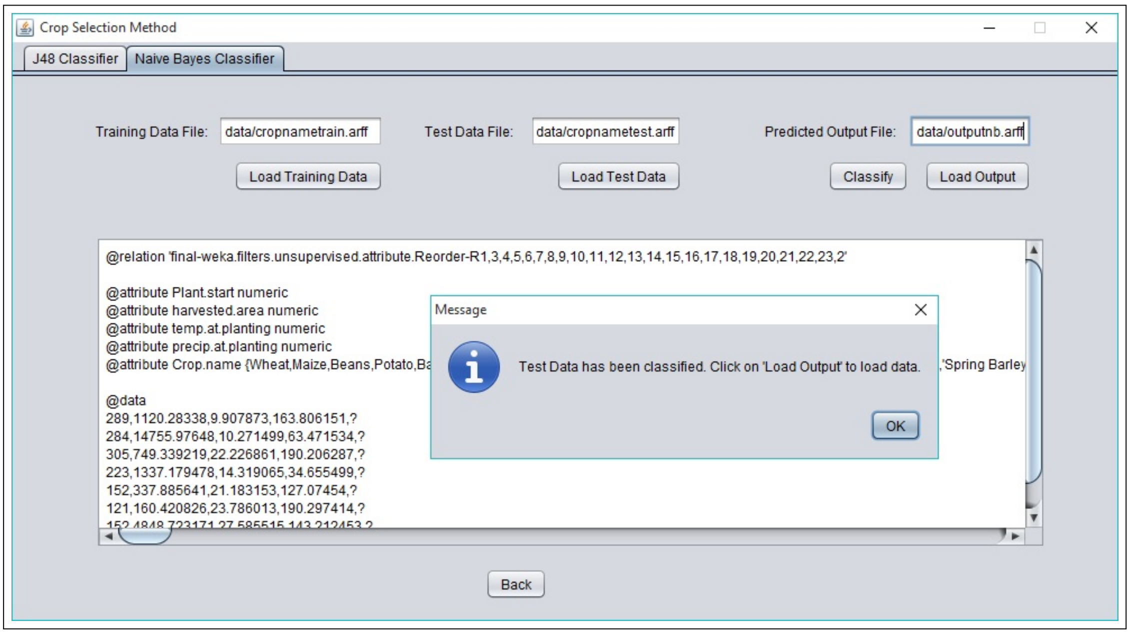


Figure 6.11: Loading Test Data - Naive Bayes Classifier

9. Crop Selection Method multi-tabbed window - Naive Bayes Classifier (Loading classified test Data).

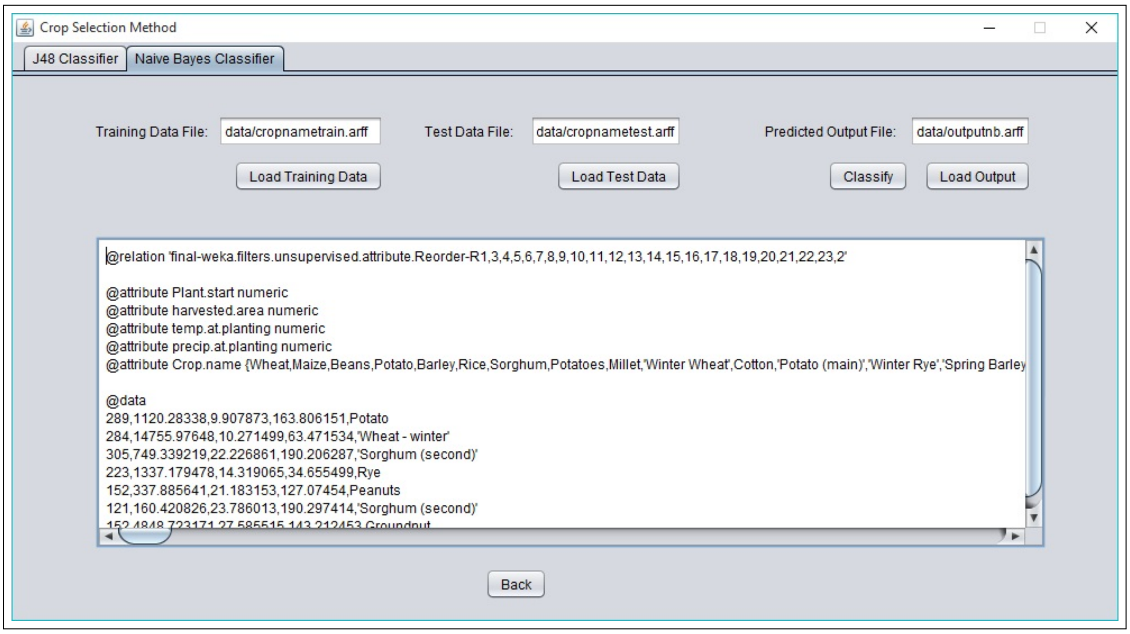


Figure 6.12: Loading classified test Data - Naive Bayes Classifier

10. Crop Sequencing Method window - (Crop Sequencing Method Interface).

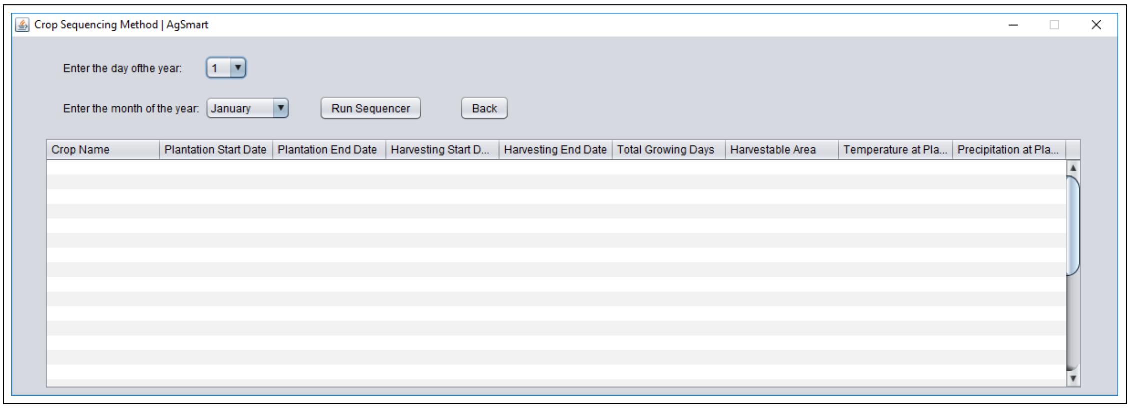
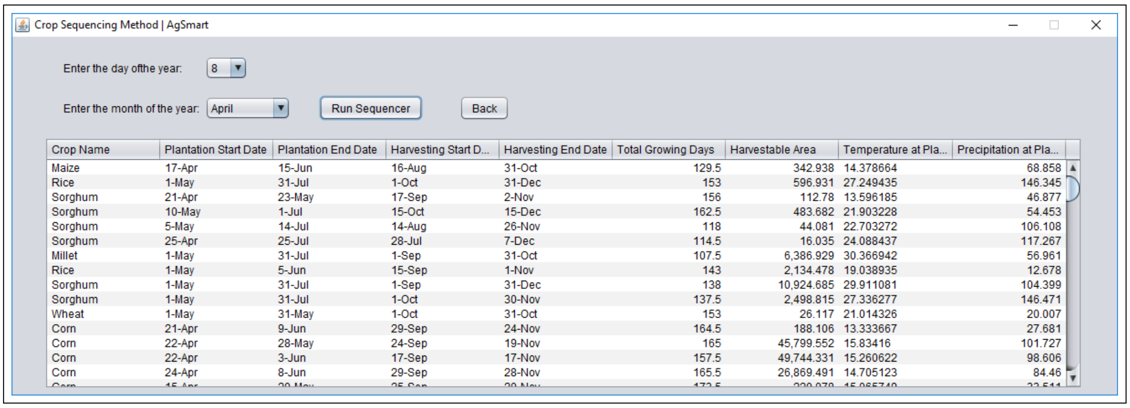


Figure 6.13: Crop Sequencing Method Interface

11. Crop Sequencing Method window - (Loading the Data).



Crop Name	Plantation Start Date	Plantation End Date	Harvesting Start D...	Harvesting End Date	Total Growing Days	Harvestable Area	Temperature at Pla...	Precipitation at Pla...
Maize	17-Apr	15-Jun	16-Aug	31-Oct	129.5	342.938	14.378664	68.858
Rice	1-May	31-Jul	1-Oct	31-Dec	153	596.931	27.249435	146.345
Sorghum	21-Apr	23-May	17-Sep	2-Nov	156	112.78	13.596185	46.877
Sorghum	10-May	1-Jul	15-Oct	15-Dec	162.5	483.682	21.903228	54.453
Sorghum	5-May	14-Jul	14-Aug	26-Nov	118	44.081	22.703272	106.108
Sorghum	25-Apr	25-Jul	28-Jul	7-Dec	114.5	16.035	24.088437	117.267
Millet	1-May	31-Jul	1-Sep	31-Oct	107.5	6,386.929	30.366942	56.961
Rice	1-May	5-Jun	15-Sep	1-Nov	143	2,134.478	19.038935	12.678
Sorghum	1-May	31-Jul	1-Sep	31-Dec	138	10,924.685	29.911081	104.399
Sorghum	1-May	31-Jul	1-Oct	30-Nov	137.5	2,498.815	27.336277	146.471
Wheat	1-May	31-May	1-Oct	31-Oct	153	26.117	21.014326	20.007
Corn	21-Apr	9-Jun	29-Sep	24-Nov	164.5	188.106	13.333667	27.681
Corn	22-Apr	28-May	24-Sep	19-Nov	165	45,799.552	15.83416	101.727
Corn	22-Apr	3-Jun	17-Sep	17-Nov	157.5	49,744.331	15.260622	98.606
Corn	24-Apr	8-Jun	29-Sep	28-Nov	165.5	26,869.491	14.705123	84.46

Figure 6.14: Crop Sequencing Method Interface - Loading the Data

Chapter 7

Conclusion

7.1 SUMMARY

If we have a system which can suggest the farmers or agriculturists the crop(s) according to the environmental conditions available instead of going through the dilemma or the confusion of crop selection which may lead to selection of wrong crop(s) leading to lesser yield of crop(s), we can get maximum yield using machine learning techniques. We saw various advantages of integrating machine learning with agriculture. Using such techniques can let farmers utilize the potential of the soil to its maximum. Greater yield marks greater agricultural growth leading to eradication of farmer suicides in the country which has been a prominent issue for a long time. We, by our system want to eliminate all such problems and help our Indian agriculture to grow for the better.

7.2 FUTURE SCOPE

We have extracted and studied various crop yield patterns as well as performed various machine learning tasks on the various data sets to maximize the yield. But, in future, still there is scope to optimize this model and make these machine learning tasks more accurate. Also, various new functionalities can be added such as predictions of sale of the crop and their demand in various markets so that the crop produce reaches the right market and provides right price to its farmers.

7.2.0.1 Further Possible Improvements

Some of the improvements which can be expected further are mentioned below:

- Data : The major hurde we faced during the successful completion of the project was the collection of apt and accurate datasets. The type of dataset required for the completion of this projects were hardly available and hence, in future, more proper datasets maty result in improving the efficiency of this model.

- The Price Factor: Since, we were not able to integrate the prices of the crop(s) with the availabel model as it was hard generalizing the price factor for all the crop(s), we expect some further developments in the model on the basis of including the price factor in the model too.

7.3 CONCLUSION

We reviewed various papers which has previously worked on this topic and studied them thoroughly. There were some papers which worked only on embedded sensors in agriculture. Some papers showed the works wherein the embedded was used to display the current state of crop(s) and other information on field. All this works can prove to be very useful in improving the Indian agriculture if implemented properly. But there were some shortcomings like none of the paper used machine learning and analytical techniques to gain better insights about agriculture and provide farmers with better crop suggestions. So in this project of 'Crop Selection Method Based On Various Environmental Factors Using Machine Learning' the outcome will give the farmers the optimal solution to the problem of crop selection.

Chapter 8

References

1. Rakesh Kumar, M.P. Singh, Prabhat Kumar and J.P. Singh, **Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique** 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), 6 - 8 May 2015. pp.138-145.

2. K. Krishna Kumar, K. Rupa Kumar, R. G. Ashrit, N. R. Deshpande and J. W. Hansen, **Climate Impacts on Indian Agriculture.** International Journal of climatology, 24: 13751393, 2004.

3. Karandeep Kaur, **Machine Learning: Applications in Indian Agriculture.** International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 4, April 2016 .

4. **www.cs.waikato.ac.nz/ml/weka/** - Weka 3: Data Mining Software in Java

5. **https://www.youtube.com/** - WEKA API Tutorial : How to use WEKA in Java.

Chapter 9

Paper Publication

Name of Pubication: IRJET - International Research Journal for Engineering and Technology.
Impact Factor for 2016: 5.181
Paper Issue: Volume: 04 Issue: 02 — February- 2017
Paper Title: Crop Selection Method Based On Various Environmental Factors Using Machine Learning.
Paper Publication website: www.irjet.net
(Go to the end of the book for reading the paper)

Thank You.