A PROJECT REPORT ON

AGRICULTURAL ANALYSIS BASED ON ENVIRONMENTAL FACTORS

SUBMITTED TO SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF

BACHELOR OF ENGINEERING (Computer Engineering)

SUBMITTED BY

VEDHAS KHARCHE	Exam No: 37
ANIKET NEEMA	Exam No: 38
DHRUV CHAUDHARY	Exam No: 39
SHUBHAM KOTHAWADE	Exam No: 40

Department of Computer Engineering
K. K. Wagh Institute of Engineering Education & Research
Hirabai Haridas Vidyanagari, Amrutdham, Panchavati,
Nashik-422003
Savitribai Phule Pune University

2019-20

K. K. Wagh Institute of Engineering Education and Research
Department of Computer Engineering

CERTIFICATE

This is to certify that the Project Titled

AGRICULTURAL ANALYSIS BASED ON ENVIRONMENTAL FACTORS

Submitted by

VEDHAS KHARCHE	Exam No:37
ANIKET NEEMA	Exam No:38
DHRUV CHAUDHARY	Exam No:39
SHUBHAM KOTHAWADE	Exam No:40

is a bonafide work carried out by Students under the supervision of Prof. R.H.Jadhav and it is approved for the partial fulfilment of the requirement of Savitribai Phule Pune University, for the award of Bachelor of Engineering (Computer Engineering).

Prof. R.H.Jadhav
Internal Guide
Dept. of Computer Engg.

Prof. Dr. S.S.Sane
H.O.D
Dept. of Computer Engg.

Dr. K. N. Nandurkar
Principal

Place: Nashik

Date:

ACKNOWLEDGEMENT

It gives us great pleasure in presenting the preliminary project report on ***'Agricultural Analysis Based On Environmental Factors'****.*

We would like to take this opportunity to thank our internal guide ***Prof. R.H.Jadhav*** *for giving us all the help and guidance we needed. We are really grateful to them for their kind support. Their valuable suggestions were very helpful.*

We are also grateful to ***Prof. Dr.S.S.Sane****, Head of Computer Engineering Department, K.K.Wagh Institute of Engineering Education And Research for his indispensable support, suggestions.*

In the end our special thanks to ***Prof. D.M.Kanade*** *for providing various resources such as laboratory with all needed software platforms, continuous Internet connection, for our Project.*

Vedhas Kharche
Aniket Neema
Dhruv Chaudhary
Shubham Kothawade
(B.E. Computer Engg.)

Abstract

Agricultural statistics and forecast is an important resource that the government has not explored extends to its impact. The proposed system is to make this process computerized by implementing principles of data mining and analytics . More specifically, system aims at targeting the social issue of drought, analyzing data based on crop produce, amount of rainfall, agricultural inputs, irrigation, and similar factors for every crop in the state of Maharashtra. Data is mined and analyzed to find various trends and relations, such as contrast between total irrigation area and type of crop; total principal and non-principal crop amount versus district-wise rainfall etc. The end result of the project will be specifying these trends, studied and analyzed from data taken over the past few years. Actions to minimize the damage of drought will also be suggested.

INDEX

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

1.1 MOTIVATION

1. Maharashtra is divided into five geographical regions, comprising six administrative divisionsKonkan, Pune, Nashik, Marathwada (Aurangabad) and Vidarbha (Amravati and Nagpur). As per the Government report Maharashtra saw 12,006 farmer sucide from 2015 to 2018.

2. The farmer suicides, which have remained unstoppable for past few years in eight districts of Marathwada, have crossed the staggering 400-mark in just over a four-month period in 2016. Compared to 2015, as many as 92 more farmers have embraced deaths in the first four and half months of 2016, highlighting the failure of the government schemes launched in August to curb the spate of suicides.

3. In past few months, 1,548 distressed farmers have been reported dead in the Marathwada region which is witnessing fourth successive years of drought with wells, rivers and dams having

4. Looking at these staggering numbers, one can easily figure out that there is an undying need to formulate solutions for farmers, techniques that can be used by the farmers to produce crops and forecasts that can help farmers be better prepared for droughts. The farming sector has been hit hard by the recent situation of rainfall scarcity, and this is affecting the farmers in almost all regions of the state of Maharashtra. This truly, is a motivation behind developing a system that can learn about various attributes and provide answers instead of only storing raw data. Merging the current information system with a result-oriented application using Data Mining techniques to alleviate the effects of drought is the motivation of the system.

1.2 PROBLEM STATEMENT AND OBJECTIVES

1.2.1 Problem Statement

To develop a system for predicting yield of crop by applying data mining technique in agricultural field and provide a Agricultural Consultancy Firm.

1.2.2 Objectives

1. To develop a system for predicting the yield of crop.
2. To develop a system that can give suggestion for tackling the damages from drought.
3. To develop a system for ease of farmers to carry out transactions with distributor.
4. Type of farming will be tackled.
5. To develop a Agricultural Consultancy firm for Latest machines and innovations.

1.3 PROJECT SCOPE AND LIMITATION

1.3.1 Project Scope

1. The proposed system promises to help farmers to select appropriate crops according to the environmental conditions.
2. The system mainly revolves around collection and cataloguing of data in its raw form. This data can first be standardized and the analyzed to find various trends that can help build solutions for farmers.
3. In India, states like Maharashtra, Madhya Pradesh, Gujarat etc. have been facing similar situations in terms of droughts and agricultural produce.Thus, this issue must be utmost importance to the government.
4. To help reduce the sufferings of farmers, governments of various states and local bodies of a number of districts can make use of this system to generate reports and find solutions for their farmers.

5. Government should help the farmers about how to use the system to farmers so that the farmers can take the advantage of the proposed system.

1.3.2 Limitation

1. The proposed system needs large datasets and basic understanding of system.
2. The system needs to be trained on regular basis.

1.4 METHODOLOGIES OF PROBLEM SOLVING

In Phase One,

- The proposed system requires analyzing all dataset from repositories
- The System will make use of SVM and Random forest for classifying the elements.

In Phase Two,

- All the comparisons needs to be done which will affect the conditions , So the main challenge is to keep all comparison and perform prediction on it.
- Transactions are also done by farmer and user after prediction is done and then he want to sell that particular gross material.
- Agricultural Consultancy Firm is to be suggested.

CHAPTER 2

LITERATURE SURVEY

The following resources are required for the proposed system -

1.”Classification of Agricultural Land Soils: A Data Mining Approach”[1]:

The data for soil classification was acquired from the FAO/UNESCO and the ORSTROM/INRA systems, which are based on the USDA soil taxonomy.

2. "Data Mining And Analysis Of Our Agriculture Based On The Decision Tree"[2]:

The experimental data is a total output value of the aforementioned parameters, for 30 provinces and cities in China. Cluster analysis, using association method, on this data, with Euclidean distance as a parameter, is used, to label the data. This labelled data is used with the ID3 algorithm, and an inverted tree, which gives specific rules for predicting crop productivity, is generated.

3. ”Application of Spatial Data mining for Agriculture”[3]:

It illustrates the basic steps of data mining, which are: data cleaning, data integration, data selection, data transformation, data mining and knowledge representation. It further goes on to explain how spatial data mining is potentially useful for extracting patterns from large datasets, and also explains the steps involved in k-means clustering. The paper concludes by providing visualizations of the cluster analysis for temperature and rainfall, and that the results can be improved by considering more parameters.

4. ”A Study of Crop Yield Forecasting Using Classification Techniques”[4]:

The proposed system suggests the use of C4.5, Naïve Bayes, ANNs and Decision trees for production of rules and classification of test data. The system will help in improving agriculture by making use of predictions based on historical data of the aforementioned parameters, depending on the season a crop is being grown in, and the weather at particular period of time.

5. ”A Brief survey of Data Mining Techniques Applied to Agricultural Data”[5]:

It is analysis of data mining techniques that can be used, and have been used in the past for agricultural datasets. It further provides an overview of the following data mining techniques. Classification, Clustering, Association Rule Mining, and Regression. Support Vector Machines can be used in case of crop classification, in case of changing weather scenarios.

6. "Crop Prediction System using Machine Learning"[6]:

The system provides a solution by monitoring the agricultural field which assists the farmers in increasing productivity to a great extent.It gives overview for which crops to be cultivated in a particular area.

CHAPTER 3

SOFTWARE REQUIREMENT SPECIFICATION

3.1 ASSUMPTION AND DEPENDENCIES

1. It is assumed that basic agricultural data will be made available for the system in some phase of its completion. Until the, test data will be used for providing the demo for the presentations.

2. It is assumed that the user is familiar with an internet browser and how to operate mobile.

3. Since the system is web based there is a need for the internet browser.

4. It will be assumed that the users will possess decent internet connectivity.

5. Various python packages and libraries are used such as sklearn,spyder,etc.

3.2 FUNCTIONAL REQUIREMENT

1. Classification of Districts: The system will need a preliminary classification of districts based on the area of cultivation, average rainfall, crop produce and agricultural inputs, in the form of training labels. The system will then learn using these training labels, and should be then able to classify the given districts into separate classes for individual crops.

2. Crop suggestion: The system should be able to suggest new crops for different districts, so as to maximize crop output based on that district's parameters. It should also show the degree of crop favourability for a particular district based on the parameters in terms of 'High', 'Medium', and 'Low'.

3. Season-wise predictions: The system should be able to identify suitable seasons for every crop, and the corresponding trend for that crop in a given district. For example, the system should be able to suggest Kharif crops such as Rice, Maize, Bajra if the rainfall in a given district is similar to the typical conditions observed in the Kharif Season.

4. Rainfall Prediction: The main aim of the system is to alert the end-users, the farmers, of any upcoming drought like situations. The system should be

able to analyse the rainfall data, and predict any erratic changes in the rainfall, or any major deviations from the standard amount for that particular month, for that particular district.

5. Consultancy Firm: The different types of machines and innovations as per requirement are given in the firm according to the Agriculture basic techniques.

3.3 EXTERNAL INTERFACE REQUIREMENT

3.3.1 User Interfaces

- Website

3.3.2 Hardware Interfaces

3.3.3 Software Interfaces

- PHP with HTML
- Python 3 with Pycharm platform

3.3.4 Communication Interfaces

- Website

3.4 NON-FUNCTIONAL REQUIREMENT

3.4.1 Performance Requirement

The system should not take up a lot of resources for the training or testing sessions, to generate the output in form of prediction, and should run smoothly on the platform that it is installed on. Also, it should give accurate results.

3.4.2 Safety Requirement

- PCI Compliance: The PCI Security Standards Council is a global group — whose founding members include American Express, Discover Financial Services, JCB International, MasterCard and Visa Inc. — formed to develop, enhance and maintain security standards for payment account security.
- SSL Certificate: The SSL Certificate — also mandatory per PCI — also works to ensure that the sensitive information that is sent over the internet is encrypted and secure.
- Use HTTPS: HTTPS should just be used on pages that collect and store data so that site visitors customers can feel secure sending their information.
- DoS and DDOS protection.
- Use a Firewall: As the name suggests, a firewall is a hardware or software system that essentially works as a wall or gateway between two or more networks, permitting authorized traffic and blocking unauthorized or potentially malicious traffic from accessing a network or system.
- Proxy Firewalls.
- Application Gateways.

3.4.3 Security Requirement

- Penetration Testing: In Network security, we often call those "Black Hat Hackers" who hack into computer networks with despicable purpose. When an organization needs to think beyond hackers and for that white hats (ethical computer hacking) is required and Network penetration testing is part of it.
- PCI compliance and SSL certificates: The (PCI DSS) Payment Card Industry Data Security Standard is a set of standards that has set certain rules for merchants those who transact payments online. In order to be in conformity, you need to give assurance to cardholder's confidential data perform

strong access control steps. You must ensure PCI compliance even if you are using payment modules.

- Ensure about Data Storage: For email and marketing campaigns, you will need a record of person's name, mail, and contact numbers including addresses of your clients. Therefore, you should also think carefully about how much risky data you amass on your server with the registered customers.

- Manifold levels of security: The security layers are an essential part to prevent cybercrime. The security layers starts with firewalls, which prevent attackers from gaining access over the network. From there you just need to add layers of security on contact forms and secure passwords for logins, including the search queries.

- Regularly patch the system: Do not wait for installation of security patch after it has published. Whether your ecommerce runs on Magento, WordPress, or use any third party platforms like Java, Python, you need to fix a security flaw in each of them. Major Websites that have breached in the past were running old software versions. If your site is not well known that does not mean that it will remain safe.

3.4.4 Software Quality Attributes

- Reliability: The system, in case of failure, should not experience a loss of data, or any of the acquired learning from past training sessions.

- Availability: The system should be in a specified state before the start of an operation at any random time.

- Scalability: The system is going to be used for the state of Maharashtra first, but should be scalable to accommodate other states facing similar drought conditions in later stages, and should also account for the change in the agricultural practices of the state.

- Maintainability: On the occurrence of a failure in the system, the time required to restore it to its former state should be less.

3.5 SYSTEM REQUIREMENT

3.5.1 Database Requirement

The system requires the datasets as input.All the features based on factors are extracted and processed.

- No of samples: 1548
- Data: Rainfall,temperature

3.5.2 Software Requirement

Platform :

1. Python and its various libraries such as, NumPy, MatPlotLib, Seaborn, Pandas, BeautifulSoup
2. Microsoft Excel
3. Scikit-learn

3.5.3 Hardware Requirement

Sr. No.	Parameter	Minimum Requirement
1	Intel Core II Duo or higher	1.5 GHz
2	RAM	2 GB

Table 3.1: Hardware Requirements

3.6 ANALYSIS MODELS: EXTREME PROGRAMMING(XP)

- The XP programming consists the steps Requirements,Design,Code reviews, Implementation,Feedback.
- Requirement gathering consists the datasets and all basic consultancy information.(Rainfall and Temperature datasets)
- Design the system format with its page documentation and overview the structure of frontend.Design the whole prediction pages with its frontend.
- Code reviews must be done among 4 members of the team as the XP model consists very few people.All SVM,ID3 and RFC algorithms must be discussed among members.
- Implement the system after code reviews are done among members.Implementation of algorithms,frontend and backend is completed.
- Deploy the system to the inbound user only and take the feedback from him if required to some changes.Here,inbound user may be any farmer and feedback is to be taken from him and accordingly prototype is made.

CHAPTER 4

SYSTEM DESIGN

4.1 SYSTEM ARCHITECTURE

Refer fig.C.2 for Architectural design.

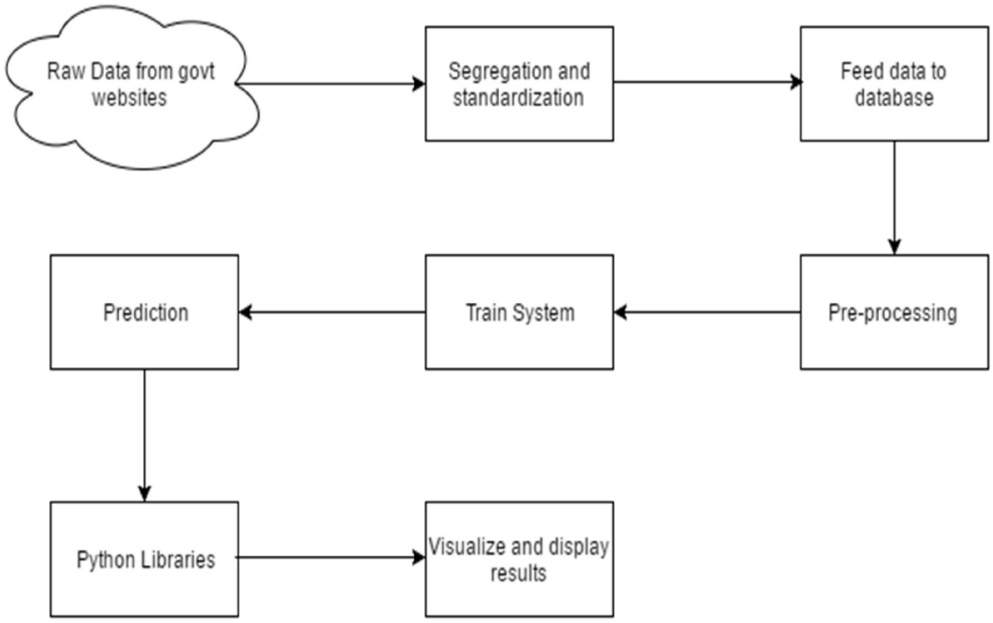

Figure 4.1: Architecture diagram

4.2 MATHEMATICAL MODEL

4.2.1 System Description

- Let S be the proposed system

  Where,

  S={I,F,O}

  I={I1,I2,I3,I4}

  F={F1,F2,F3,F4}

  O={O}

  Where,

  I : Input given to system

  I1 : Rainfall Data

  I2 : Pressure Data

I3 : Tempreature data

I4 : Crop Data

O : Output of the system

F1 : Data Collector

F2 : Data preprocessor

F3 : Data Wrangling

F4 : Predictor

4.2.2 Input

- The input to the system is Raw Data from various government and private repository.

4.2.3 Functions

4.2.3.1 F1 : Data Collector

Let F1 be the function which collect raw data from various government and private websites.It return collected data to process further.

- **Input:** Gathering data from Sources
- **Processing :**
  F1(rain,humidity,temperature,crop statistics):
  rain := Search(rain data)
  humidity := Search(humidity data)
  temperature := Search(temperature)
  crop := Search(crop statistics)
  return rain,humidity,temperature,crop

- **Output :**Raw Data collected from different sources

4.2.3.2 F2 : Data preprocessor

Let F2 be the function which process data generated by F1 and convert it into standardize format.It returns Data in Standard format.

- **Input :** Raw Data from F1
- **Processing :**
  F2(Raw data):
  standard :=process(Raw data)
  return standard
- **Output :** return standardize data and feed into database.

4.2.3.3 F3 : Data Wrangling

Let F3 be the function which use F2 generated data

- **Input :**Standard data from F2
- **Processing :**
  F3(data):
  data:=label(data)
  data:=Clustering(data)
  data:Integrate(data)
  data:Classify(data)
  return data
- **Output:**return data which use as input to algorithms for prediction

4.2.3.4 F4 : Predictor

Let F4 be the function which uses F3 generated data with algorithms to predict.

- **Input :** Classified and labeled data from F3
- **Processing :**
  F4(data):
  data:=fit(data)
  report:=predict(data)

return report

- **Output:** return predicted crop to yield.

4.2.4 Venn Diagram

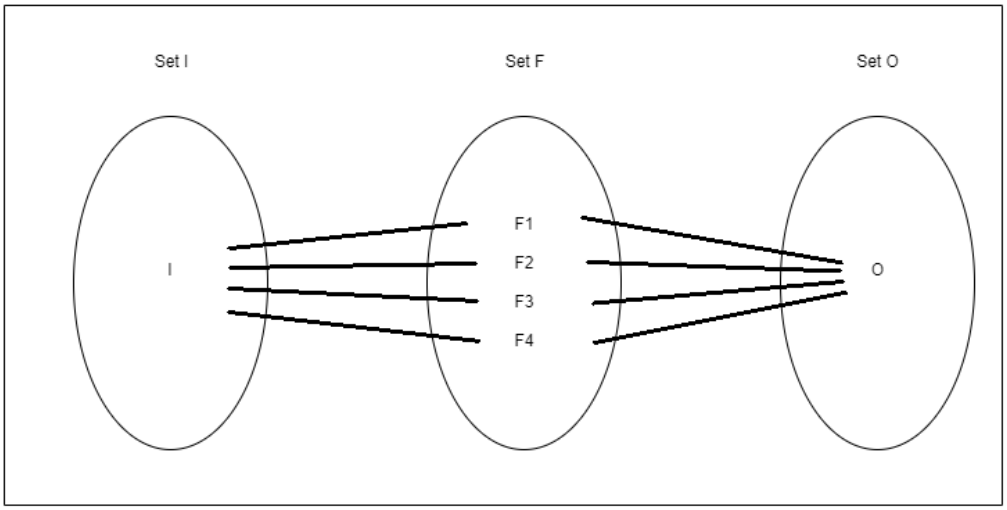

Figure 4.2: Venn Diagram

4.2.5 Output

- The System outputs a report which shows the best crop to yield in a particular season based on several affecting factors.

4.3 DATA FLOW DIAGRAM

4.3.1 Level 0 Data Flow Diagram

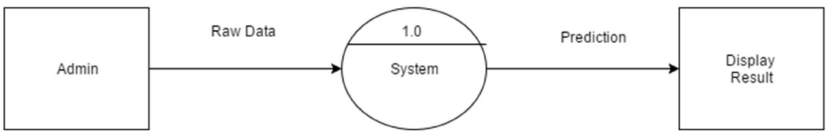

Figure 4.3: Dataflow 0 diagram

4.3.2 Level 1 Data Flow Diagram

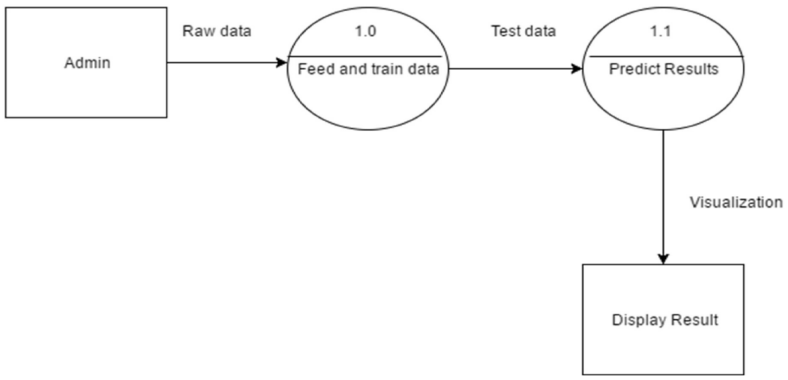

Figure 4.4: dataFlow1 diagram

4.4 ENTITY RELATIONSHIP DIAGRAM

Refer Fig. 4.5 for Entity relationship Diagram

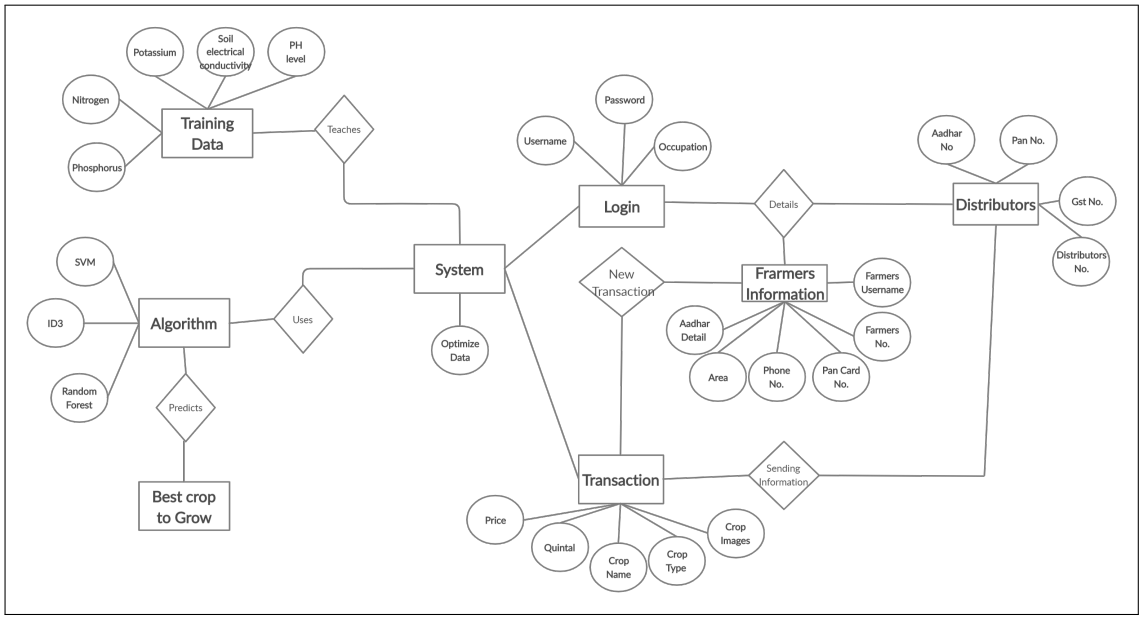

Figure 4.5: entity relationship diagram

4.5 UML DIAGRAM:

4.5.1 State Transition Diagram:

Refer Fig. 4.6 for state transition diagram.

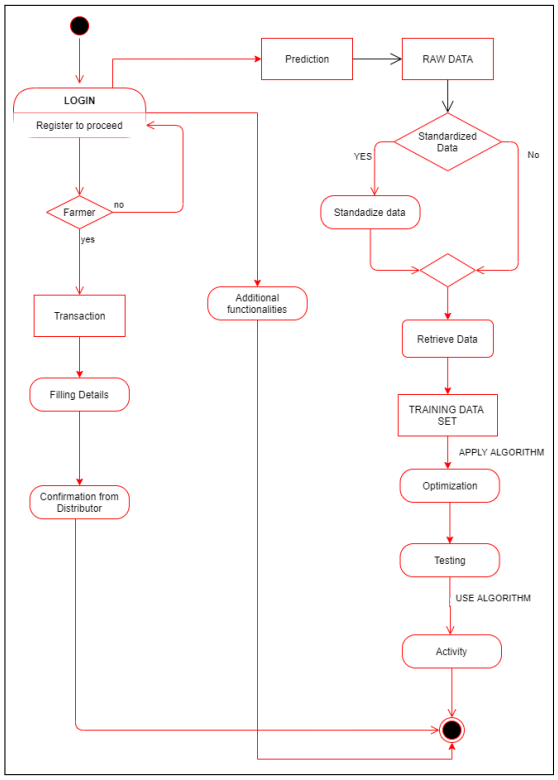

Figure 4.6: State transition diagram

4.5.2 Use Case Diagram:

Refer fig. 4.7 for Use Case diagram.

4.5.3 Activity Diagram:

Refer fig. 4.8 for Activity diagram.

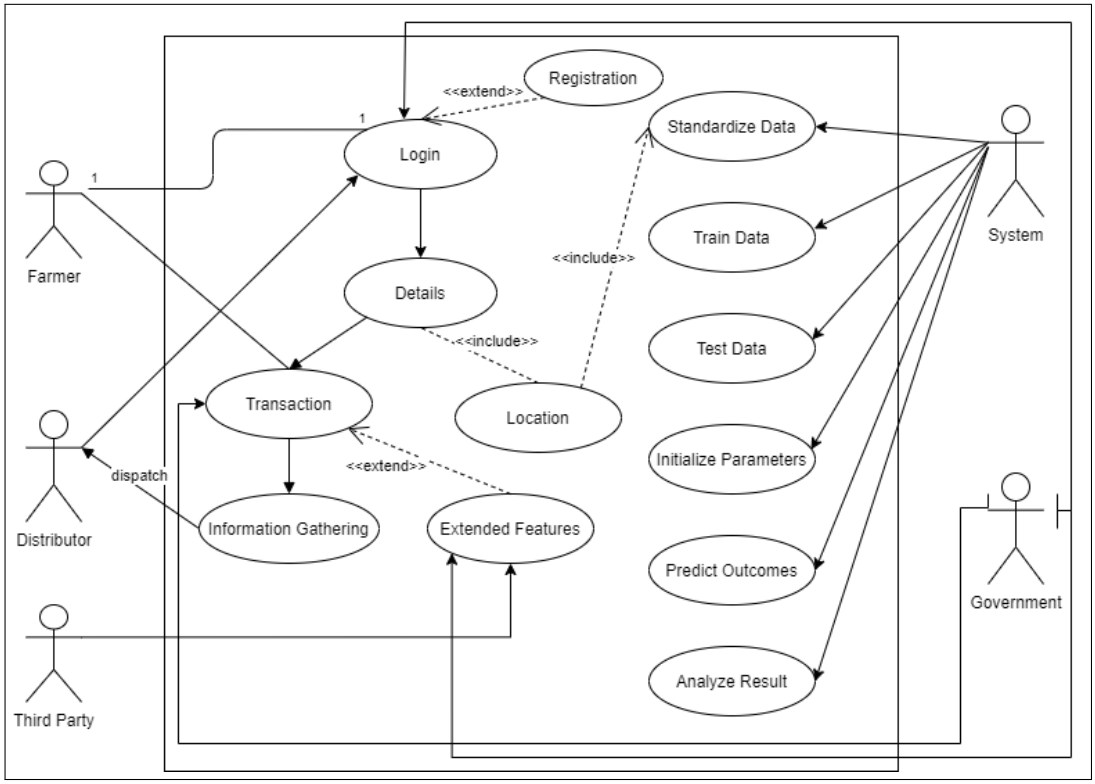

Figure 4.7: Use case diagram

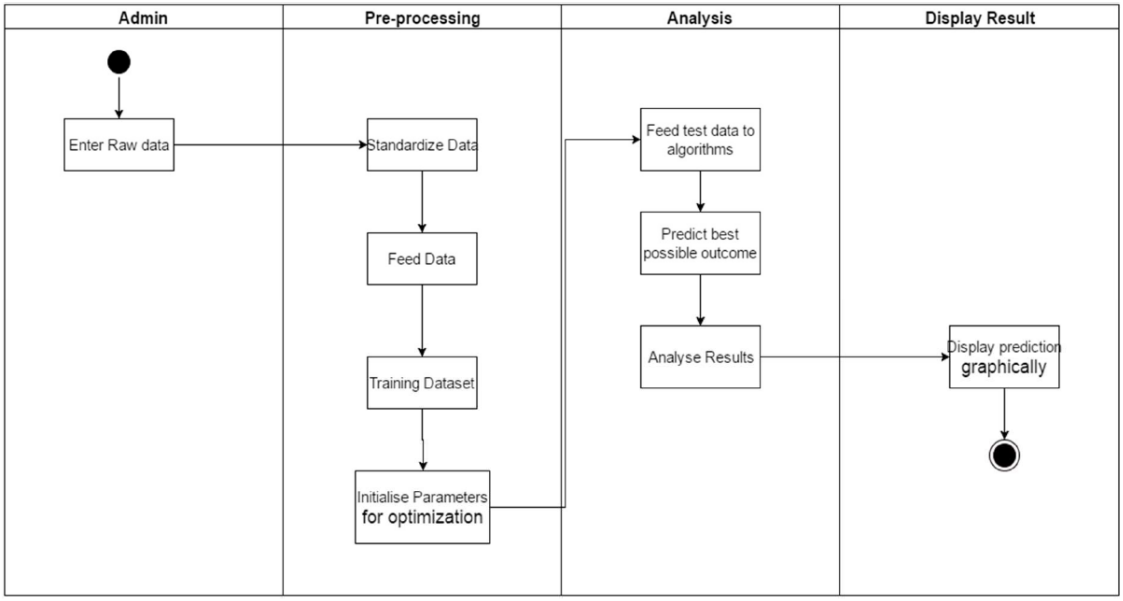

Figure 4.8: Activity diagram

CHAPTER 5

PROJECT PLAN

5.1 PROJECT ESTIMATES

5.1.1 Reconciled Estimates

The model followed for estimating the cost is the Basic COCOMO model for estimating the cost and effort for the project. for This project, the estimate is calculated from lines of code. Actual Kloc = 9.7

5.1.1.1 Time Estimate

The effort is calculated using Formula: $E = a(KLOC)^b$
Where E = effort in person - months
for a semi- detached project, a =3.0 , b = 1.12
Effort(E) in person- months = 38.22

Duration can be calculated from formula :$D = c(E)^d$
Where D = Duration c = 2.5 , d = 0.38 Duration(D) in months = 8.94 Months

5.1.1.2 Cost Estimates

The cost can be calculated from formula Cost = E (Cp) D
Where E=effort in person-month, Cp=Cost per person, D=Duration Here Cp=25rs
Cost = 8,452 rs

5.1.2 Project Resources

Hardware:

- Processor: Intel Core II Duo or higher.
- Required RAM: 32 MB min.
- Required Memory: 16 GB min.

Software:

- Python(NumPy, MatPlotLib, Pandas)
- Microsoft Excel
- Scikit-learn

5.2 RISK MANAGEMENT

5.2.1 Risk Identification

The Proposed project is subjected to various risks. This risk involve the development time required , infrastructure failure. The risk identified are as follows:

1. The system may be prone to virus attacks.
2. The system may suffer from downtime resulting in unavailability of services.
3. The system may not be available during retraining of model.
4. The system is prone to power outage due to power grid failure.
5. The system not accumulated proper dataset.

5.2.2 Risk Analysis

The risks for the Project can be analyzed within the constraints of time and quality

ID	Risk Description	Probability	Impact		
			Schedule	Quality	Overall
1	Virus attacks	Medium	Low	Medium	Medium
2	Server Down	Medium	Low	High	Medium
3	Delay in retraining	Low	High	Low	Low
4	Power failure	High	Low	High	High
5	Impure dataset	Low	Low	Medium	Low

Table 5.1: Risk Table

Probability	Value	Description
High	Probability of occurrence is	> 75%
Medium	Probability of occurrence is	26 – 75%
Low	Probability of occurrence is	< 25%

Table 5.2: Risk Probability definitions [**?**]

Impact	Value	Description
Very high	> 10%	Schedule impact or Unacceptable quality
High	5 – 10%	Schedule impact or Some parts of the project have low quality
Medium	< 5%	Schedule impact or Barely noticeable degradation in quality Low Impact on schedule or Quality can be incorporated

Table 5.3: Risk Impact definitions [**?**]

5.2.3 Overview of Risk Mitigation, Monitoring, Management

Following are the details for each risk.

Risk ID	1
Risk Description	Virus attacks
Category	Development Environment.
Source	Cyber Attack
Probability	Medium
Impact	Medium
Response	Mitigate
Strategy	server Down time
Risk Status	Identified

Risk ID	2
Risk Description	Server Down
Category	Requirements
Source	Software Design Specification documentation review.
Probability	Medium
Impact	Medium
Response	Monitoring
Strategy	Server restarting
Risk Status	Identified

Risk ID	3
Risk Description	Delay in retraining
Category	Development Environment
Source	Software Design Specification documentation review.
Probability	Low
Impact	Low
Response	Mitigate
Strategy	Stop retraining
Risk Status	Identified

Risk ID	4
Risk Description	Power grid failure
Category	Development Environment
Source	Power failure
Probability	High
Impact	High
Response	Monitoring
Strategy	Turn On generator
Risk Status	Identified

Risk ID	5
Risk Description	Impure Data set
Category	Deployment Environment
Source	Software Design Specification documentation review.
Probability	Low
Impact	Low
Response	Management
Strategy	Standardizing data
Risk Status	Occurred

5.3 PROJECT SCHEDULE

5.3.1 Project task set

Major Tasks in the Project stages are:

- Task 1: Study of an existing system.
- Task 2: Literature Survey.
- Task 3: Designing and Planning.
- Task 4: System flow.
- Task 5: Implementation.
- Task 6: Testing and Documentation.

5.3.2 Task network

Refer fig.5.1 for Task Dependency Graph.

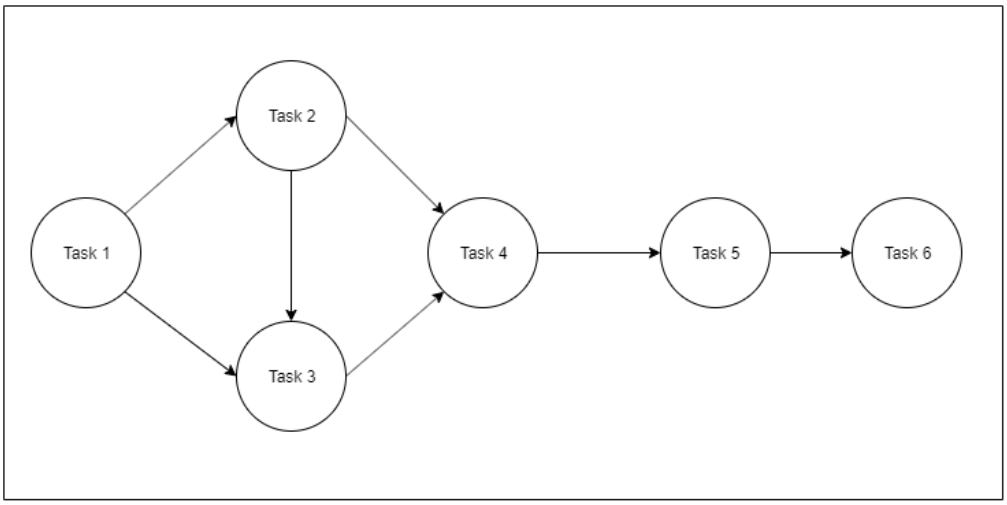

Figure 5.1: Task Dependency Graph

Phase	Task	Descriptions
1	Analysis	Analyze information from external Sources
2	Literature Survey	Collect Raw data and elaborate on literature survey
3	Design	Design the process flow
4	Implementation	Implement the code for System
5	Testing	Test the code for overall process
6	Documentation	Prepare the Document with conclusion and future scope

Table 5.4: Plan Description

5.3.3 Timeline Chart

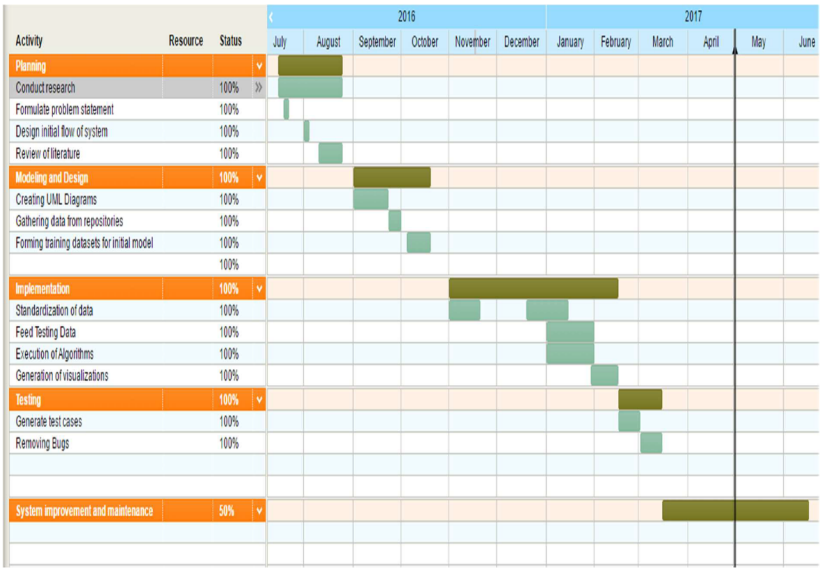

Figure 5.2: Time Line Chart

5.4 TEAM ORGANIZATION

The team consists of four members which discusses progress of the system with guide and updates logsheet on Thursday and Saturday of every month.

5.4.1 Team Structure

The team consists of 4 members, where task related to project are roughly distributed equally.The team members are assigned following responsibilities:

- Vedhas Kharche: Algorithm and techniques
- Aniket Neema: GUI Designer
- Dhruv Chaudhary: Cor-relational Analysis and Extracting Features
- Shubham Kothawade: Documentation

5.4.2 Management Reporting and Communication

Management Reporting and Communication includes:

- Status - Completed
- Costs incurred - 6k to 7k
- Working time spent - 6 months
- Report Format - PDF
- Rules Used:
    1. Should be done using LaTex format.
    2. Report must be of 50-60 pages.
- Communication methods and techniques - Online Platform
- Reporting system for capturing, storing and distributing performance reports - Online

CHAPTER 6

PROJECT IMPLEMENTATION

6.1 OVERVIEW OF PROJECT MODULES

- Raw data is collected.
- Rainfall and Temperature data is collected.
- The training data teaches the system.
- The system uses the following modules to predict final results:
    1. SVM
    2. ID3
    3. Random Forest
- It will predict the best crop to be grown in a particular area.
- It will analyze the result outcomes to provide viable solutions to end user.
- Transactions will also be generated.
- Consultancy firms are to be shown for machine innovations according to requirements.

6.2 TOOLS AND TECHNOLOGY USED

- The proposed system will use tree data structure used for decision making.
- Temporary buffer are proposed to store the incoming labeled data and output of thesystem before sending to the end user.
- A large datasets of soil and weather is to be used for agricultural decision-making by making use of classifiers like Naive Bayes and Random Forests.
- The data for soil classification has been acquired from the FAO/UNESCO and the ORSTROM/INRA systems, which are based on the USDA soil taxonomy.

- Data mining will be performed on a relevant subset of the soil database, collected from the various regions , taking 1500 samples from a total of 2045 samples.
- Tools:
    1. Python(NumPy, MatPlotLib, Pandas)
    2. Microsoft Excel
    3. Scikit-learn

6.3 ALGORITHM DETAILS

6.3.1 Algorithm 1: ID3

Following are the steps to be followed to induct a decision tree using ID3:
1. Calculate the entropy of each and every attribute.
2. Splitting the dataset into subsets using the calculated entropy for each and every attribute.Calculated value should be minimum and IG is maximum.
3. Construct a decision tree (for node) which contains that attribute.
4. Recurse on subsets using remaining attributes.

6.3.2 Algorithm 2: SVM

1.A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplane.
2.The operation of the SVM algorithm is based on finding the hyperplane which gives the largest minimum distance for the given trained examples.This distance gives the important name of margin. Hence, the obtained optimal separating hyperplane will maximize the margin of trained data.
3.In this system, it make use of the Gaussian Radial Basis Function kernel, a popular kernel function used in various kernelized algorithms.

6.3.3 Algorithm 3: Random Forest

1.The random forest is an ensemble approach that can also be thought of as a form of nearest neighbor predictor. It is a method for classification, regression and other tasks that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.
2.Random decision forests correct for decision trees' habit of overfitting to their training set. Random forests does not overfit.
3.Random Forest is one of the algorithms used to classify the training instances at the both the aforementioned phases.

6.4 USE-CASES

All use-cases for the software are presented. Description of all main Use cases using use case template is to be provided.

Sr No.	Use Case	Description	Actors	Assumptions
1	Retrieve data	data is retrieved from repository	Admin	data retrieved
2	train data	training system for prediction	System	system is trained
3	corelation analysis	analysis performed using algorithm	System	analysis generated
4	prediction of crop	prediction of crop based on rainfall temperatuere	system	predicted crop

Table 6.1: Use Cases

CHAPTER 7

SOFTWARE TESTING

7.1 TYPES OF TESTING

7.1.1 Unit Testing

- White Box Testing is performed on the various modules while implementation.
- Modules such as data.py,scrape.py,etc on which Unit testing is performed.

7.1.2 Functionality Testing

- Test Forms(in transaction part of system) are working as expected.
- Message is shown in case of any blank details .
- After submission of form, the data is to be given in main database.

7.1.3 Database Testing

- While executing queries no errors are reported.
- Data Integrity is also maintained while creating, updating or deleting data in data.csv database.
- Test data retrieved from data.csv database is shown accurately in website.

7.1.4 Performance Testing

- Website response times at different connection speeds is about 10-15 seconds.
- If a crash occurs due to peak load, the site recover from such an event is tested.

7.2 TEST CASES AND TEST RESULTS

Test Case Id	Test Case	Expected Result	Actual Result	Status
1	Login	Farmer should be login into system	Farmer is logged in	Pass
2	Prediction	Crop should be correctly predicted	Crop is correctly predicted	Pass
3	Transaction	Farmer details should be entered	Farmer details	Pass
4	Government Id	Farmer and distributor should enter correct Id	Farmer and distributor entered wrong Id	Fail
5	Prediction Value	Farmer should enter appropriate probable value	Farmer entered wrong value	Fail

Table 7.1: Test Cases

CHAPTER 8

RESULTS

8.1 OUTCOME

- The proposed system will generate corelations between various attributes.
- Prediction of crops based on corelation data.
- Transaction options on the system is generated.
- Agricultural consultancy firm and types of farming are suggested.

8.2 SCREENSHOTS

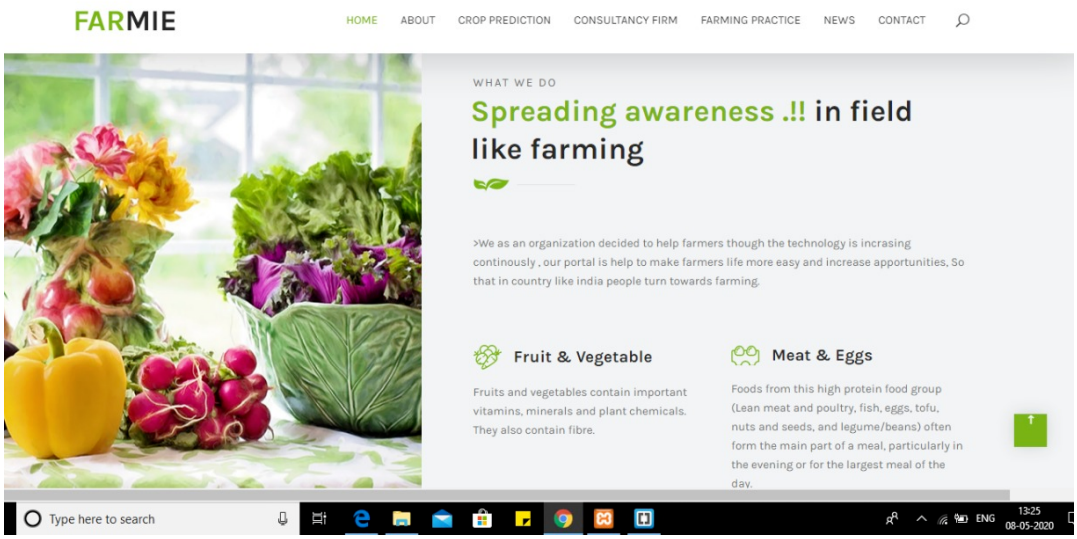

Figure 8.1: Screenshot1

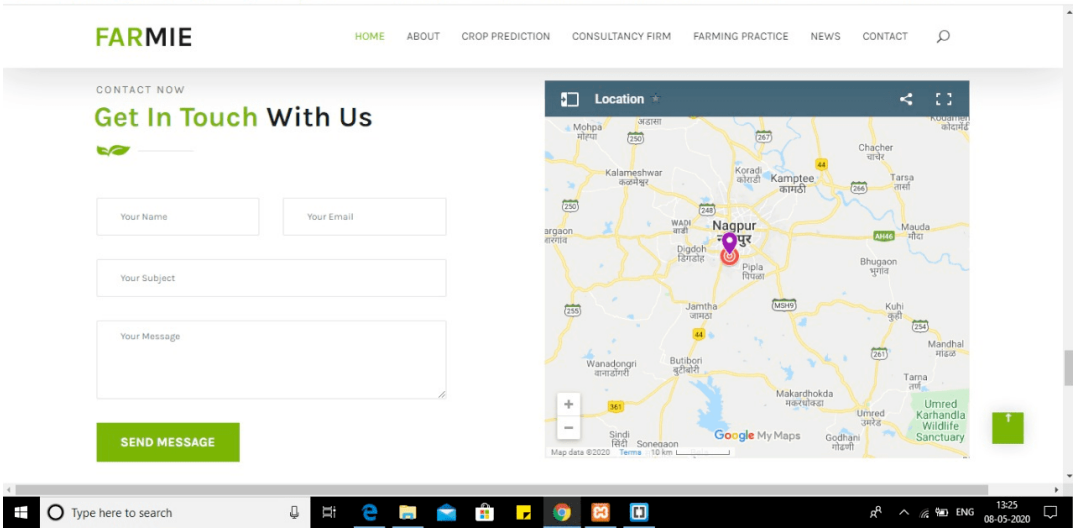

Figure 8.2: Screenshot2

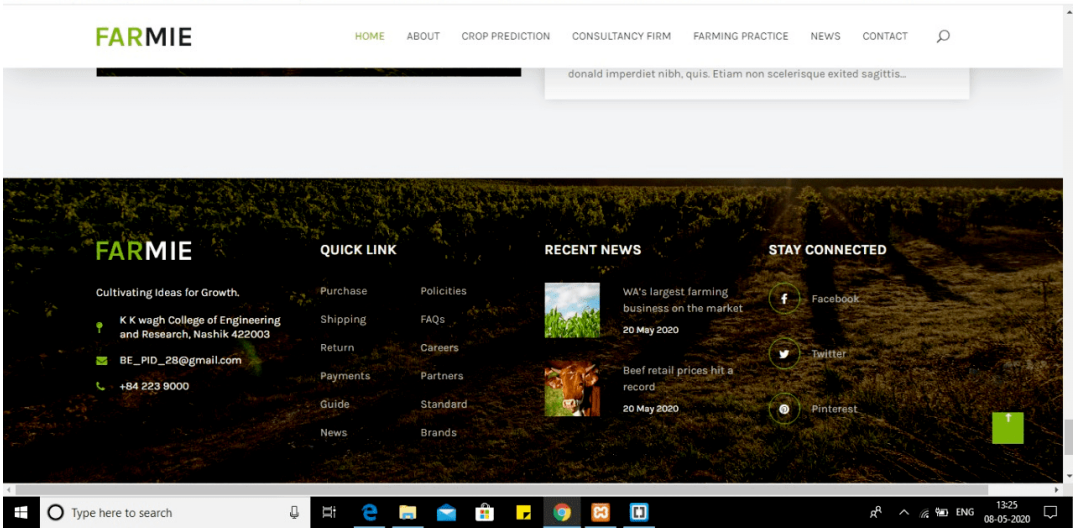

Figure 8.3: Screenshot3

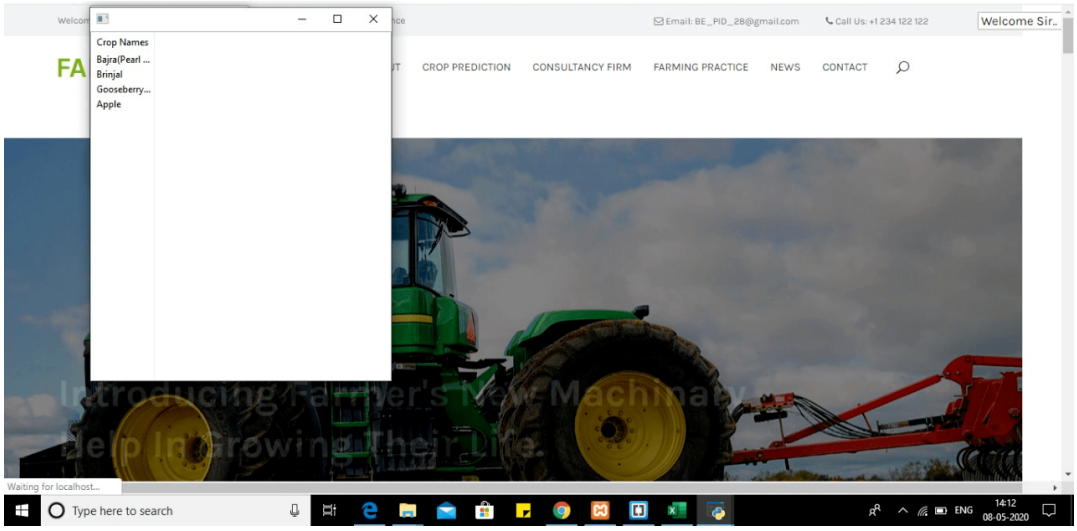

Figure 8.4: Screenshot4

CHAPTER 9

CONCLUSIONS

9.1 CONCLUSION

The System has been proposed to help the farmers for good yield by deciding which crop will be best to be grown in the particular area based on the environmental factors.And this might also reduce the suicide rate which has increased amongst the farmers and the ultimately lead to growth in economy.Consultancy firm leads to machine innovations suggestions.

9.2 FUTURE WORK

- In future parameters can be added such as soil quality,agricultural inputs,soil nutrients,irrigated area.
- The Soil Health Card Scheme by the Ministry of Agriculture and Farmers Welfare, which aims at providing Integrated Nutrient Management (INM), when incorporated along with this idea, will improve the efficiency of this system.
- In future, this system can also be implemented in various other states and also countries, only to alleviate the suffering caused to farmers.

9.3 APPLICATIONS

1. The system can be used by farmers for predicting the best crop to grow based on rainfall,temperature.
2. The system can be used by the government for tackling the damages from drought.
3. The system leads to best use of cashless economy while transactions between farmer and distributors.

ANNEXURE A

SATISFIABILITY ANALYSIS

The implemented project is the NP-Complete problem as the algorithms used for prediction are NP-Complete.

NP is the set of decision problems that can be solved by Non-deterministic Turing Machine in Polynomial time.P is the subset of NP.Any problem that can be solved by deterministic machine in polynomial time can also be solved by non-deterministic machine in polynomial time.

- The SVM, ID3 and Random Forest algorithms used for prediction consists the decision tree like structure which is NP-Complete.
- These algorithms accessing large datasets. As the size of datasets will increase, the algorithms take non-polynomial time for prediction tends to be NP-Complete problem.

ANNEXURE B

ANNEXURE C

PLAGIARISM REPORT

Sr.No.	Chapter	Uniqueness(%)	Plagarism(%)
1	Introduction	95	5
2	Literature Survey	97	3
3	Software Requirement Specification	100	0
4	System Design	100	0
5	Project Plan	93	7
6	Project Implementation	96	4
7	Software Testing	100	0
8	Results	100	0
9	Conclusions	100	0
	Total	97.88	2.12

Table C.1: Plagarism Report

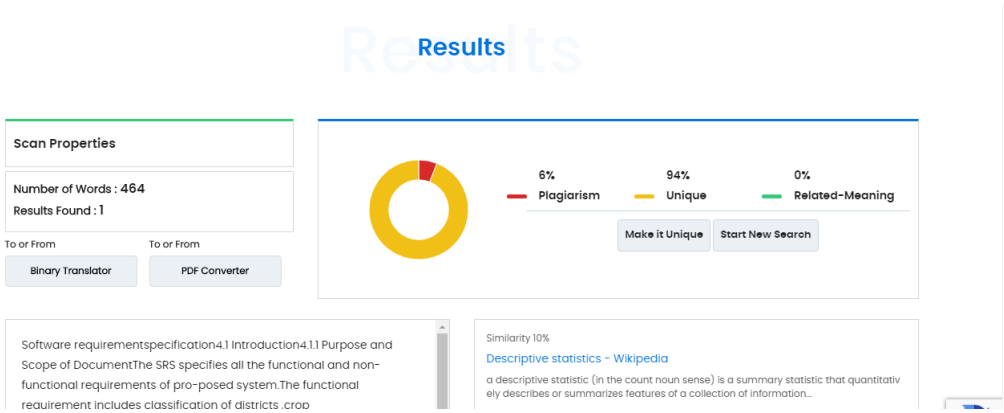

Figure C.1: snapshot1

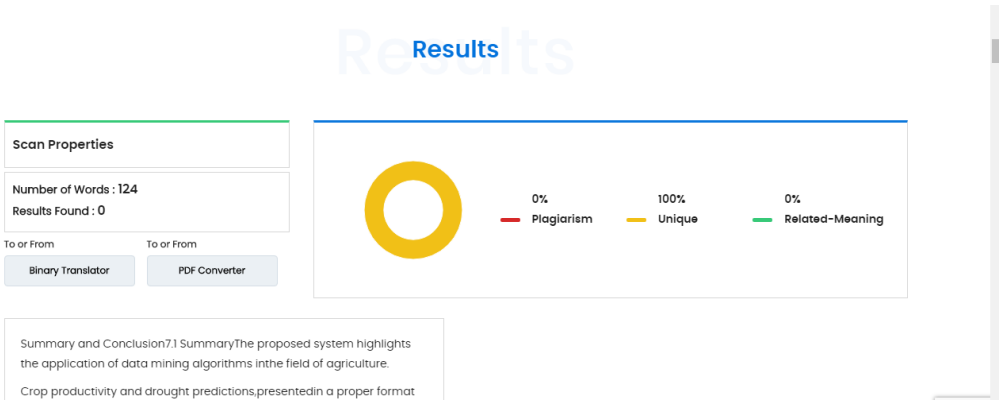

Figure C.2: snapshot2

REFERENCES

[1] V. Ramesh and K. Ramar, “Classification of agricultural land soils: A data mining approach,” Agricultural Journal, 2011.

[2] G. Yi-Yang and R. Nan-ping, “Data mining and analysis of our agriculture based on the decision tree,” International Colloquium on Computing, Communication, Control, and Management, 2009.

[3] D. Rajesh, “Application of spatial data mining for agriculture,” International Journal of Computer Applications, 2011.

[4] R. Sujatha and P. Isakki, “A study of crop yield forecasting using classification techniques,” International Conference on Computing Technologies and Intelligent Data Engineering, 2016.

[5] H. Patel and D. Patel, “A brief survey of data mining techniques applied to agricultural data,” International Journal of Computer Applications, 2014.

[6] N. M. S. G. C. M. Prof. D.S. Zingade, Omkar Buchade, “Crop prediction system using machine learning,” International Journal of Advance Engineering and Research Development, 2017.