Machine Learning Approach for Crop Yield Prediction

Submitted in partial fulfillment of the requirement of the award of the Degree of

BACHELOR OF ENGINEERING
In
Information Science and Engineering

Submitted by
Deepthi G N (01JST17IS015)
Madhurya Shankar (01JST17IS025)
Sinchana Hathwar (01JST17IS054)
Namratha M (01JST17IS058)

Supervisor Name
Dr.B.S.Harish
(Professor
Dept of ISE
JSSSTU)

JSS Science and Technology University
Mysuru

2021

DECLARATION

We declare that this written submission represents our ideas in our own words and where other's ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academics honestly and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission.

Deepthi G N
01JST17IS015

Madhurya Shankar
01JST17IS025

Sinchana Hathwar
01JST17IS054

Namratha M
01JST17IS058

Educate Elevate Enlighten

JSS Mahavidyapeetha
JSS Science And Technology University
(Established Under JSS Science and Technology University Act No. 43 of 2013)
(Formerly Known as SJCE)

JSS SCIENCE AND TECHNOLOGY UNIVERSITY MYSURU

JSS SCIENCE AND TECHNOLOGY UNIVERSITY

Department of Information Science and Engineering

CERTIFICATE

This is to certify that the following students Deepthi G N, Madhurya Shankar, Sinchana Hathwar and Namratha M of JSS SCIENCE AND TECHNOLOGY UNIVERSITY, Department of Information Science and Engineering have successfully completed the project work titled "Machine Learning Approach for Crop Yield Prediction" during the year 2021

Name of the students	University Seat Number
Deepthi G N	01JST17IS015
Madhurya Shankar	01JST17IS025
Sinchana Hathwar	01JST17IS054
Namratha M	01JST17IS058

Internal Examiner External Examiner Head of the Department

Acknowledgement

We especially wish to acknowledge the following:

We would like to thank the whole management of the department of Information Science and Engineering for having given us an opportunity to carry out a project on our own by trusting and acknowledging our abilities. We have great pleasure in expressing our deep sense of gratitude to our institution JSS Science and Technology University, Mysore. We wish to express our sincere gratitude to Dr. S B Kivade, Principal, SJCE and Dr. B S Mahanand, Head of the Department, Information Science and Engineering, SJCE for the facilities, guidance and support provided throughout the project work. We are profoundly grateful to our project guide and mentor Dr.B.S.Harish for his suggestions, valuable support, encouragement and guidance throughout the project. We would like to thank all the teaching and non-teaching staff of the Information Science and Engineering Department. We also convey our gratitude to all those who have contributed to this project directly or indirectly.

Abstract

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

Table of Contents

List of Figures

Chapter 1

Introduction

1.1 Problem Statement

Predicting the crop yield based on parameters like temperature, rainfall and other soil features (pH, N, P, K etc.) using machine learning approaches

1.2 General Introduction:

Agriculture is the process of producing food, feed, fibre and many other desired products by the cultivation of certain plants. India is a country where agriculture and agriculture related industries are the major source of living for the people. India being an agricultural country, its economy predominantly depends on agriculture yield growth and agro-industry products. The agriculture sector is the backbone of an economy which provides the basic ingredients to mankind and raw material for industrialization.

Agriculture plays a vital role in the Indian economy. Over 70 per cent of the rural households depend on agriculture. Agriculture is an important sector of Indian economy as it contributes about 17% to the total GDP and provides employment to over 60% of the population. In 2020, 41.49 percent of the workforce in India were employed in agriculture, while the other half was almost evenly distributed among the two other sectors, industry and services. While the share of Indians working in agriculture is declining, it is still the main sector of employment. Agriculture is the basic source of food supply of all the countries of the world whether underdeveloped, developing or even developed. Due to heavy pressure of population in underdeveloped and developing countries and its rapid increase, the demand for food is increasing at a fast rate. If agriculture fails to meet the rising demand of food products, it is found

to affect adversely the growth rate of the economy. Raising supply of food by agricultural sector has, therefore, great importance for economic growth of a country.
Various important industries in India find their raw material from agriculture sector -cotton and jute textile industries, sugar, vanaspati, etc. are directly dependent on agriculture. Handlooms, spinning oil milling, rice thrashing, etc. are various small scale and cottage industries, which are dependent on agriculture sector for their raw material. This highlights the importance of agriculture in industrial development of the nation.

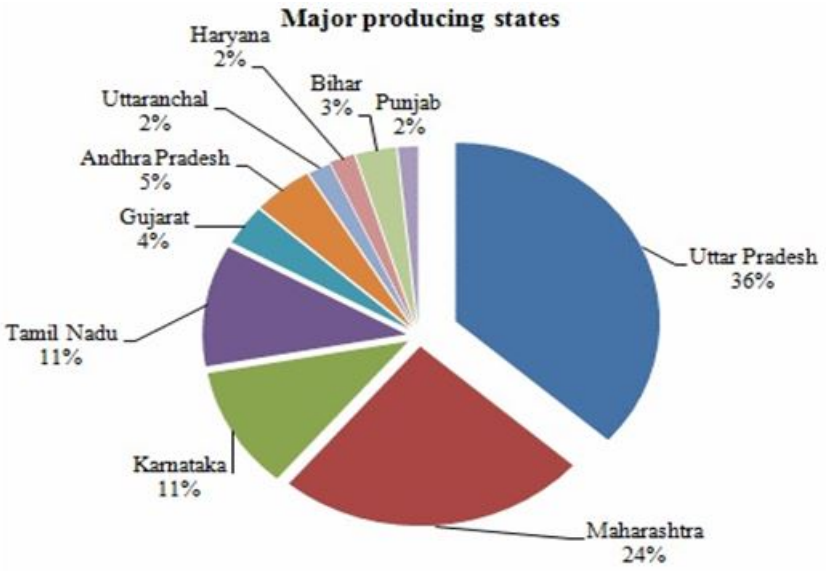


Figure 1.1: Major crop producing states of India.

(Courtesy : Ministry of Agriculture,GoI)

Agriculture is a major source of the economy of the country. Current agriculture sector has challenges in several areas such as lack of mechanization, insufficient storage facilities, coping with climate change, maintaining fertilizers, to meet the demand of the market etc. Prediction of crop yield helps in improving risk management issues related to crop production. With the current parameters for growing crops of a type may fail to give satisfactory yield, hence the estimation of yield is very demanding and required.

Till now only the past experience of the farmer was used which consisted of randomly counting the number of seed buds that a plant is having and within each seed bud the number of seeds it holds. Further based on experience the farmer used to predict the crop yield. Predicting the crop yield well in advance prior to its harvest can help the farmers and Government organizations to make appropriate planning like storing, selling, fixing minimum support price, importing/exporting etc. Information

technology can be used to avert the risk associated with agriculture and it can also be used to predict the crop yield more accurately prior to harvest. Food is perhaps the most important aspect of life, without it, humankind can't survive. Hence the estimation of food is very demanding. So, we attempt to forecast the yield of crops based on agriculture parameters.

1.2.1 Machine Learning

Machine learning is a process of studying a system based on data. It is a part of data science where we use machine learning algorithms to process data. There is continuous improvement in Machine Learning algorithms and are capable of learning from the data we provide. There is automation for everything and hence takes less time to process data. It provides more accurate results and faster decision making. Machine learning also supports a wide range of applications. Machine learning approach can be used in many domains like evaluation of human behaviour, prediction of the market, supermarkets etc.

Machine learning is a decision support tool which plays an important role in crop yield prediction. It can provide better yield prediction based on several parameters like meteorological data, Agri-meteorological, soil (pH, N, P, K) data, remotely sensed data, agricultural statistics etc. They will determine different patterns and analyse the methods and features. Precision is the greatest challenge during prediction in machine learning. It needs more datasets as crop yield is dependent on many factors. This can show that the process is quite complicated and is not a trivial task

1.2.2 Supervised Learning Technique

It's a predictive model used for the tasks where it involves prediction of one value using other values in the data-set. Supervised learning will have predefined labels. It classifies an object based on the parameters to one of the predefined set of labels. We have many algorithms to build model in supervised learning such as KNN, Naive bayes, Decision Tree, ID3, Random Forest, SVM , Regression techniques etc. Depending of the requirement, labels, parameters and data-set we select the appropriate algorithm for predictions. Algorithm is used to build a model that makes predictions based on evidence in the presence of uncertainty.

Machine learning algorithms applied to analyze data and to crop yield. Datasets collected from agriculture departments. Rainfall, temperature, PH value, nitrogen, potassium, zinc, phosphorous, iron etc.. all these parameters are used for prediction.

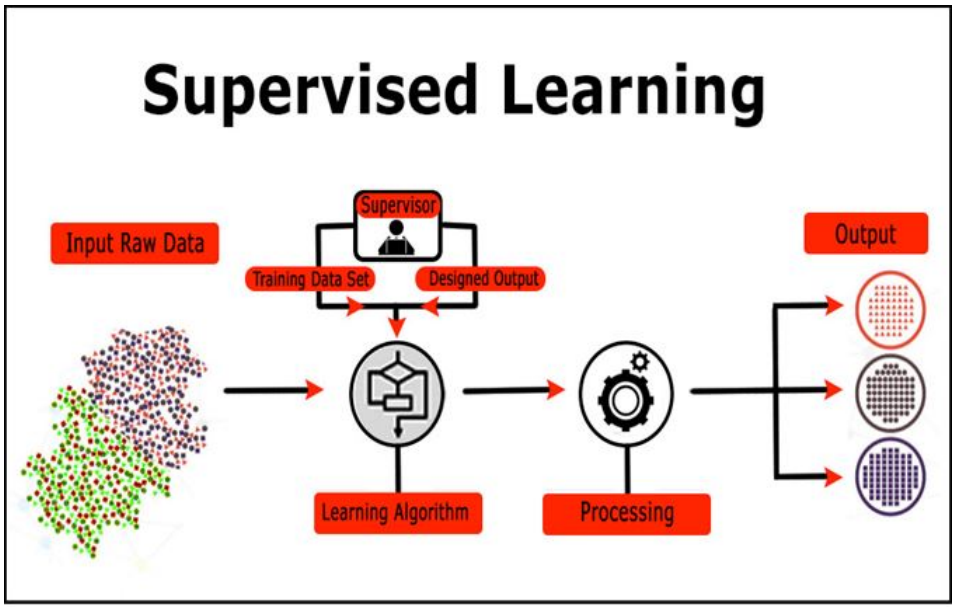


Figure 1.2: Supervised learning Technique

(Courtesy:https://echoinnovateit.com/types-of-machine-learning-algorithms/)

Crop yield is a highly complex trait determined by multiple factors such as rainfall, temperature and soil features (pH, N, P, K, Zn, S, Fe, C). Accurate yield prediction requires fundamental understanding of the relationship between yield and these factors, to achieve this it requires both comprehensive datasets and powerful algorithms.

1.3 Challenges of Existing System

- Current system is manual where we compare the previous results with the present. Based on the previous experiences and results we come to know how much crop yield will be produced.
- There is no automation to predict the crop yield using the constraints temperature, rainfall, humidity, area and region and other factors.
- The existing system uses Manual Processing which makes it Time consuming, Less Reliable and Less Efficient thereby providing Less User Satisfaction

1.4 Motivation

- Agriculture is one of the most important occupation practiced in our country .It is the broadest economic sector and plays an important role in overalls development of the country

- About 60% of the land in the country is used for agriculture in order to suffice the needs of the 1.2 billion population . Also ,the crop yield has a direct impact on the national and international economies annually

- Also, in the last decade India has seen serious natural calamities like drought or flood. Due to such disasters there is a huge loss to crop production and ultimately to the farmers . Due to such financial loss many farmers are committing sucide .If natural calamities are not present then there may be a sudden pest attack destroying the crop. In any case the farmer and the crop are always at the edge of risk . So predicting the crop yield well in advance prior to its harvest can help the farmers and the government organisation to make appropriate planning like storing ,selling, exporting etc.

- Currently agriculture field is facing many problems . One of the major problems the farmers are facing is less profit . Farmers grow the crop but are not getting proper yield which leads to lesser profit levels . Thus yield prediction plays a vital role and hence modernisation of agriculture is very important and will lead the farmers of our country towards profit.

1.5 Objectives

- To develop a model by considering various parameters like temperature, rainfall, PH value, nitrogen, potassium, iron, zinc etc.

- To use Machine Learning algorithms to classify and predict the crop yield.

- To test the developed model in real time.

Chapter 2

Literature Survey

Yield prediction is a very important agricultural problem that remains to be solved based on the available data. The problem of yield prediction can be solved by employing Data Mining techniques. Ramesh et al.,[2] used k nearest neighbor and k-means algorithm to predict the average yield production . The drawback is that K-Means algorithm is able to partition the samples in clusters, but no considerations were made on the compounds that were responsible for this partition and it only supported numerical data. It requires huge data for good results. Machine learning algorithms are also used for comparative analysis on soil properties and to predict fertility and crop yield. Monali et al., [3] used K-Nearest Neighbor and Naive Bayes for prediction. The system helped the soil analysts and farmers to decide sowing in which land may result in better crop production. It was observed that the category having maximum confidence value was predicted as the category of that particular soil. Only the soil features were used for yield prediction, so it was less efficient. Mistry et al., [4] used Classification techniques such as Linear Regression, K - nearest neighbour and Clustering Techniques such as K- means clustering, Self organised maps which allowed farmers to plant different crops in different districts based on simple predictions. The results indicated that sunshine hours and daily temperature range play critical roles in rice yield variability in the current study area. The drawbacks are that it is not suitable for real time application and gave less accurate results. Gandge et al., [5] used a classification algorithm to predict crop yield per acre with some recommendation . It is observed that the algorithm which is used by most of the authors does not use a unified approach where in all the factors affecting the crop yield can be utilized simultaneously for predicting the crop yield. Sally et at.,[1] used Weka classifier, Association Rules and Clustering Algorithms to predict the classification of a new data instance by using a decision tree or set of rules. The WEKA (Waikato Environment for Knowledge Analysis) system provides a comprehensive suite of facilities for applying data mining techniques to large data sets. It was observed that Weka was a significant step in the transfer of machine learning technology into the workplace.

The drawbacks are it requires a huge amount of data, ready tools for prediction and more time for data processing. Tahmid et al.,[6] used k-Nearest Neighbor, Decision Tree algorithm, ID3(Iterative Dichotomis) algorithm for prediction. The result shows that Decision Tree Learning- ID3 algorithm gives a less value for percentage error than the K-Nearest Neighbor algorithm without omitting the outliers of the dataset. Though the research is limited to some fixed dataset, the future ahead promises addition of more data that can be analysed with more machine learning techniques to generate crop predictions with better precision. Machine learning algorithms are also used to find the association among agriculture parameters. Kuljit et al., [8]used the Apriori Algorithm to predict the growth of paddy yield considering various parameters such as rainfall and temperature. It was observed that with increase in rainfall the paddy yield also increased. During the reproductive phase rainfall and temperature did not influence but during maturation phase, paddy yield was better expected at lower temperature, worse at high temperature and paddy yield was found to be high at low rainfall and low during high rainfall. Nidhi et al.,[9] developed a crop recommendation system that uses the ensemble technique of machine learning. Random Forest, Naive Bayes, and Linear Support Vector Machine were the independent base learners used in the ensemble model for prediction. This system recommends suitable crops for the farmers but it is not suitable for crop yield prediction since it uses less parameters for crop recommendations and less amount of datasets which leads to less accurate results and takes more time processing data. Instead of the conventional approach of using single remotely sensed imagery for the target crop classification, a machine learning based classification algorithm is used while keeping in view the phonological cycle of the target tobacco crop. Waleed et al., [10] used Regression and Support Vector Machine for prediction of the yield of crop tobacco. They used parameters related to tobacco and the datasets are not suitable for other crop yield prediction. The drawback is that it is not suitable for real time. It concentrates on a single crop, cannot be applied for multiple crops and not suitable for multiple regions. Jie Sun et al., [11]used Convolutional neural network (CNN) - Long Short-Term Memory (LSTM) model to predict the yield of crop corn by using the parameters related to corn. It is not suitable for real time. It concentrates on a single crop, cannot be applied for multiple crops and not suitable for multiple regions. Neha Rale et al., [12] used Linear regression with polynomial features, and support-vector regression using a Radial Basis Function (RBF) kernel for prediction.The drawback is Linear regression and support vector regression generates outputs graphically which is difficult to analyze and it is not suitable in real time. It uses a small dataset for prediction. Bhanumathi et al., [13] used the Random Forest and Back propagation algorithm for implementation. Based on fertilizers, the system predicts crop yield. But it does not consider all agriculture parameters and uses less parameters for yield

prediction. Elavarasan et at., [14] developed a Deep Recurrent Q-Network model which is a Recurrent Neural Network deep learning algorithm over the Q-Learning reinforcement learning algorithm to forecast the crop yield. The limitation was that it requires huge data. Jeevan et al., [15] predicted by a machine learning algorithm called Random Forest. The drawbacks are that it is not suitable for real time, takes more time for prediction and gives less accurate results. Nageswararao et al., [16] used morphological operations and MATLAB for yield prediction. The accuracy of the image processing technique used is 92.7 and error rate is 7.3. This method has several drawbacks. As this method uses images for yield prediction, it requires more time .This even produced less efficient results and this is applicable for only Chili crops. This can't be used as a generic system where it can be applied for all crops and Image processing may lead to more error prone and less accuracy. Haowei Muet et al.,[17] used histogram dimensionality reduction and time series fusion to generate the input layer of Convolutional Neural Network (CNN). Convolutional Neural Network (CNN) is able to determine more relevant information and has higher robustness. It also provides a technical reference for estimating large-scale crop yield. This is applicable for only wheat crops. The limitations of using Convolutional Neural Networks methods is, it requires more time for image processing, huge amounts of data and also produces less accurate results. Jayantrao et al., [18] used spatial estimation of rice yield by assimilation of parameters derived from Synthetic Aperture RADAR (SAR) data from Sentinel-1 satellite into a process-based Oryza crop growth simulation model. The comparison of simulated and actual yield showed Normalized Root Mean Squared Value (NRMSE) of 9.21%. The overall agreement between actual and simulated yield is 83-89%. Satellites were used for yield prediction. This method produces less accurate results, requires more data and is suitable for only rice crops. Shivani et al., [19] developed a model using a Multilayer perceptron neural network. The dataset is filtered using Python Pandas and Pandas Profiling tools to retrieve data for Maharashtra state. Python was used for yield prediction. Since ready libraries were used , it is not suitable to work with real time applications and ANN algorithms were used which leads to less efficient results. Porandla et al., [20] used the method of regression analysis. Data mining techniques used to predict yield. This requires a very huge amount of data. Graphical outputs are generated as they have used regression algorithms. Anil et al.,[21] used deep neural networks for crop yield prediction. Convolutional Neural Network (CNN) and Long-Short Term Memory Network (LSTM) were proposed. Convolutional Neural Network is not suitable to process text data as it will generate less accurate results. This proposed method is applicable for only soybean and not suitable for other crops. Hulya et al.,[22] used a deep learning architecture method to estimate crop yield in field images .This approach was experimented on sunflower image sequences collected from four different parcels and

obtained promising results. As this method uses images for yield prediction, it is difficult to analyze images.Debra et al.,[23] developed an AI recommender system (RS) which learns data analysis choices from user behavior for predicting agricultural production responses to rainfall and learns to identify classes of agroecosystem responses to alternative climate scenarios.It also makes use of the large amounts of data and scientific expertise in the agricultural enterprise to predict agroecosystem dynamics under changing environmental conditions. Haowei et al.,[24] designed a CNN network structure to extract the features of winter wheat growth from multitemporal MODIS images for yield estimation in North China.It was able to mine more relevant information and had higher robustness and also provided a technical reference for estimating large-scale crop yield.Bhimanpallewar et el.,[25]designed a hybrid machine learning algorithm works better than the existing supervised classification techniques. Existing techniques can be worked for the attributes which generally captured continuously and every attribute value recorded with the fixed interval of time.But regardless of the variety in environment, soil and the crop-pattern this hybrid approach will work appropriately, as it is completely depending on dataset. If the dataset is not proper or if the training and testing datasets are belonging to different region then the accuracy get reduced.

Chapter 3

System Analysis

3.1 Existing System

Prediction of agriculture yield is essential to deal with storage of crops, transportation decisions and risk management issues related to crops. Current system is manual where we compare the previous results with the present. Based on the previous experiences and results we come to know how much crop yield will be produced. There is no automation to predict the crop yield using the constraints temperature, rainfall, humidity, area and region.

Limitations of the existing system:
1. Manual Process
2. Time Consuming
3. Less Reliable
4. Less Efficient
5. Less User Satisfaction

3.2 Proposed System

Proposed system is an agriculture application which analyzes the previous data related to rainfall, temperature, humidity and crop yield. Proposed system makes use of data mining in agriculture for decision making. The research is conducted taking under consideration the various constraints such as temperature, rainfall, humidity, region, area and other constraints. System uses *"Classification Rules"* technique - *Naive Bayes Algorithm* for crop yield prediction. Data collected from government sector can be used to predict crop yield.

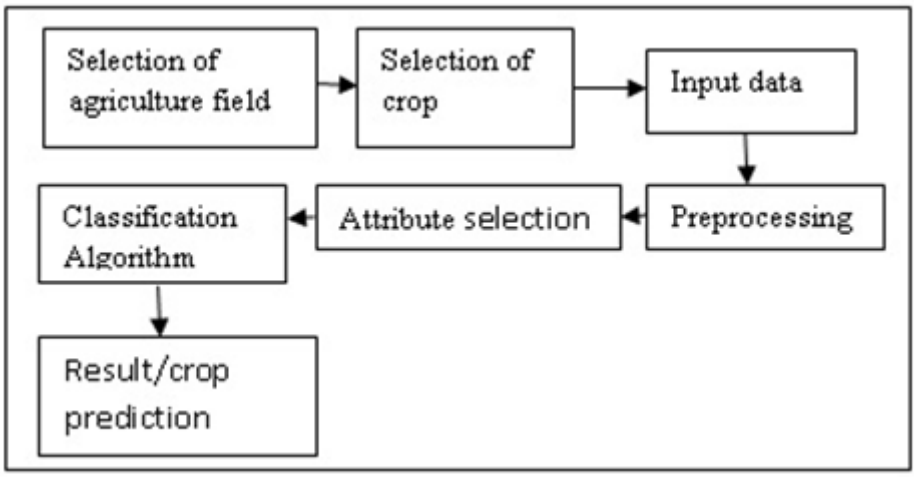


Figure 3.1: Overview of the Crop Prediction System

3.2.1 Brief overview of the Crop Prediction System

1. Selection of agriculture field: Consider any agriculture field for the crop prediction system.
2. Selection of crop: consider any crop of choice which will be sown in that field.
3. Input data: Data may include information regarding soil (Nitrogen (N),Phosphorus(P) content), Micronutrients present in soil, Moisture in soil etc which is collected over some period of time.
4. Preprocessing: Data which is collected should be preprocessed redundant data, inconsistent should be taken care.
5. Attribute Selection: Important Features have to be extracted.
6. Classification Algorithm: An appropriate and efficient algorithm should be employed.
7. Result: prediction or recommendation can be provided to the farmers based on the results obtained.

3.2.2 Scope and Objectives

1. System objective is to estimate or forecast the yield of crop.
2. System makes use of data collected from agriculture department.
3. Proposed system makes use of data mining in agriculture for decision making.
4. System uses “Classification Rules” technique - Naive Bayes Algorithm for crop yield prediction.
5. System makes use of SQL Serve to store the previous agriculture data.
6. System works for dynamic data and System is a agriculture application useful for farmers.

3.2.3 Input and Output

1. Input – Previous year's agriculture data which includes temperature, rainfall, humidity and other constraint.
2. Output – predicts crop yield using different constraints such as region, temperature, rain, humidity and yield prediction based on year wise and location wise.

3.3 Proposed Method

Step 1: Collection Data
This is the first step in the Crop yield prediction process, where we collect agriculture data. Agriculture data is collected from the region "Mysore" which contains agriculture parameters, crop details, farmers' details and parameter details. Agriculture parameters include rainfall, temperature, soil features such as PH, nitrogen, potassium, iron etc.

Step 2: Data Preparation
Here agriculture data is analyzed and only relevant data are extracted. The data required for processing is extracted and segmented according to the different regions. Only the required data is extracted because entire agriculture data is not required for processing.

• Data cleaning – to remove noise and irrelevant data.
Data quality is a main issue in quality information management. Data quality problems occur anywhere in information systems. These problems are solved by data cleaning. Data cleaning is a process used to determine inaccurate, incomplete or unreasonable data and then improve the quality through correcting of detected errors and omissions. Generally data cleaning reduces errors and improves the data quality. Correcting errors in data and eliminating bad records can be a time consuming and tedious process but it cannot be ignored.

• Data integration – where multiple data sources are combined.
Data integration involves combining data from several disparate sources, which are stored using various technologies and provide a unified view of the data. Data integration becomes increasingly important in cases of merging systems of two companies or consolidating applications within one company to provide a unified view of the company's data assets. The later initiative is often called a data warehouse.

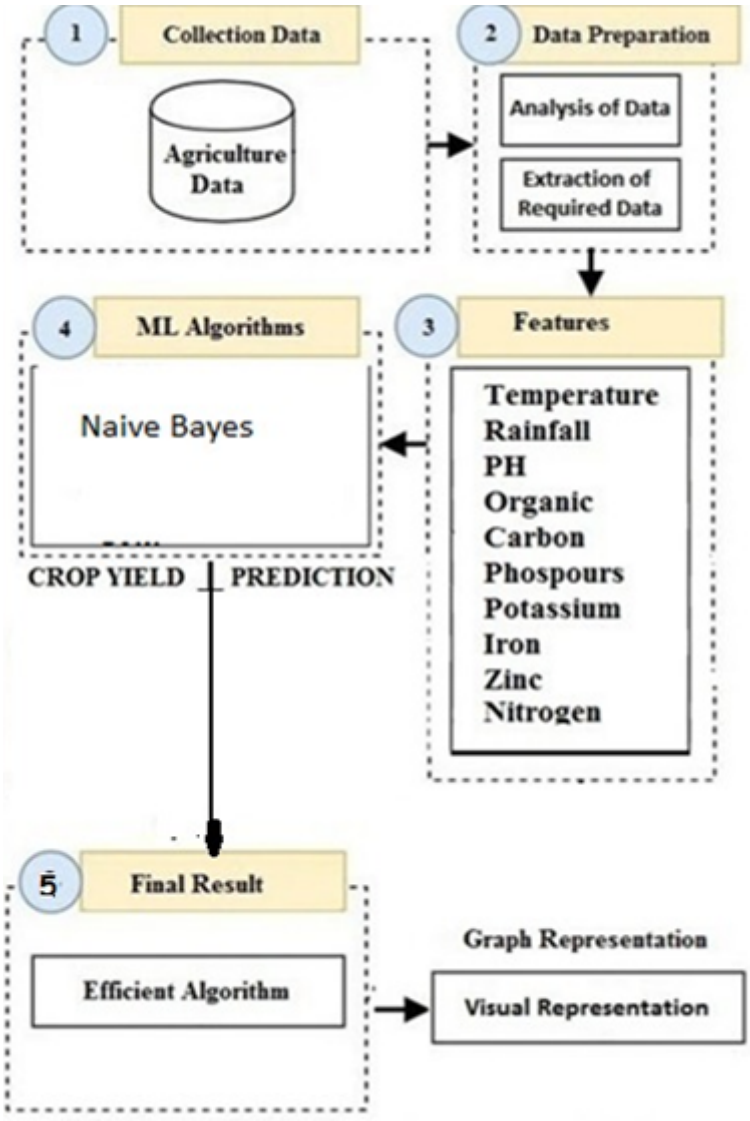


Figure 3.2: General block diagram for crop yield prediction

• Data selection – to retrieve only the relevant data from the database for the analysis.
• Data transformation – where data are transformed or consolidated into appropriate forms for processing.

Step 3: Features

Agriculture parameters used for crop yield prediction like temperature, rainfall, PH, nitrogen, organic carbon, zinc, iron, sulphur, potassium, phosphorus are fetched.

Step 4: ML Algorithms

In Supervised learning, you train the machine using data which is well labelled. It means some data is already tagged with the correct answer. We know the correct answers, the algorithm iteratively makes predictions on the training data. Supervised learning will have predefined labels. It classifies an object based on the parameters

to one of the predefined set of labels. We have many algorithms to build models in supervised learning, depending on the requirement, labels, parameters and dataset we select the appropriate algorithm for predictions. We use Supervised Learning algorithm such as the Naive Bayes Algorithm for crop yield prediction. The System predicts the yield for the selected crop based on the agriculture parameters using machine learning algorithms.

Step 5: Final Result
Algorithm results are analyzed and an efficient algorithm is identified. Crop yield is displayed for the farmers on GUI. When a user logs into the application, the system recommends suitable and high profit crops for the farmers on a GUI.

3.4 Approach Adapted

According to Software Engineering the approach adopted to develop this project is the Iterative waterfall Model. The iterative waterfall Model is a systematic approach that begins at the feasibility study phase and progress through analysis, design, coding, testing, integration and maintenance. Feedback paths are there in each phase to its preceding phase as show in the fig to allow the correction of the errors committed during a phase that are detected in later phase.

Feasibility study:
The main aim of this phase is to determine whether it would be financially and technically feasible to develop the product. The feasibility study activity involves the analysis of the problem and collection of all the relevant information relating to the product such as different data items which would be input to the system, processing required to be carried out on these data, the output data required to be produced by the system, as well as constraints on the behavior of the system.
Requirement Analysis and Specification:
The main aim of this phase is to understand the exact requirements of the customer and to document them properly.

Design:
The goal of design phase is to transform the requirements specified in the SRS document into a structure that is suitable for implementation in some programming language. In technical terms, during the design phase the software architecture is derived from SRS document. Two distinctly different design approaches are available: the traditional approach and the object oriented approach. We have adopted traditional design to develop the product.

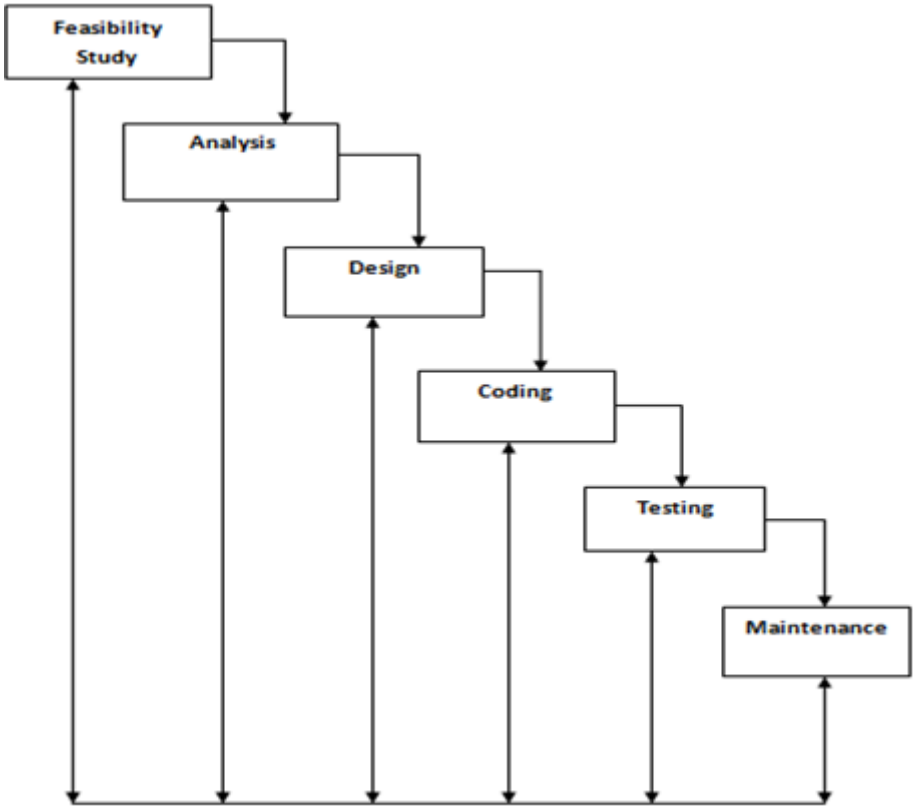


Figure 3.3: Waterfall model for Crop yield prediction

Coding:
Once design is complete, goal of the coding phase is to translate the design of the system into code in a given programming language. For a given design, the aim in this phase is to implement the design in the best possible manner. We have coded the design using C language to develop the product.

Testing:
Testing is the major quality control measure employed during software development. Its basic function is to detect errors in the software.

Maintenance:
Maintenance is not a part of software development. It is an extremely important activity in the life of software product. Maintenance involves performing any one or more of the following kinds of activities:

• Correcting errors that were not discovered during the product phase. This is called corrective maintenance.

• Improving the implementation of the system, and enhancing the functionalities of the system according to the customer's requirements. This is called perfective maintenance.
• Porting the software in a new environment. This is called adaptive maintenance.

Feasibility Study

1. Economical Feasibility:
When the economical feasibility of the project being developed it was found that the income for the administrator through the proposed system(faculty rating system) was more than the development cost of the system (investment). Thus the proposed project is economically feasible.

2. Operational Feasibility:
The system provides attractive and easy graphical user interface. The actors on this are mainly administrator, faculty and student and the application is designed in a better manner and user friendly and thus the system is easily accessible by them. Therefore the project is operationally feasible.

3. Technical Feasibility:
To develop this system one should have the knowledge of visual studio framework, ASP.NET 4.0, C, SQL server. From the user prospective it just requires a web browser and an efficient internet connection. This makes the project technically feasible.

Chapter 4

Requirement Specification

Software Requirement Specification is the starting point of the software development activity. It includes an introduction that gives the purpose, scope and an overview of the system. This needs requirement by talking to the people and understanding their needs. It also includes a general description of the product perspective, product function and certain user characteristics of the system. It also specifies the overall functional requirements, performance requirements and design constraints. The SRS is a means of translating the idea in the mind of the clients (the input), into a formal document (the output of the requirement phase). The Software Requirement Specification document is organized in such a manner it aids validation and system design.

4.1 Brief overview of the Crop Prediction System

1. Administrator
Administrator is a person who maintains the entire application. System contains only one administrator.

2. Location Incharger /Staff
Staff is an actor who receives the services from the application. Staff is a person who is incharge of a particular location. Staffs created by the administrator.

3. Visitors/Farmers
Visitors are the users who can access the basic agriculture information.

4.2 System Requirements

Software Requirements :

1. OS : Windows 2000/NT/XP/Higher
2. Back End : SQLSERVER
3. Designed Tool Kit : Visual Studio 2010
4. Front End : ASP.NET 4.0
5. Programming Language : C

Hardware Requirements :

1. Intel P4 +
2. 2.4 GHz or above
3. 2 GB RAM +
4. 500GB + HDD Minimum

4.3 Functional requirements

This section of Software Requirement Specification describes all general factors of the product and its requirements.

• System used by administrator, staffs and farmers.

• System is browser based application which predicts rice yield based on the soil test results and temperature , rainfall.

• System makes use of data mining technique for rice yield prediction.

• System makes use of previous dataset for the yield prediction

• System makes use of navie bayes algorithm for rice yield prediction

• System generates accurate results based on the size of the dataset.

• Administrator of the system creates the staffs and sets the unique Id and password for each staff.

4.4 Non Functional Requirements

• **Usability-** our application will be useful to government sector where system is an automation for rice yield prediction using the previous agriculture dataset, here we make use of real time data for prediction. System is implemented for a government sector and which helps farmers. As it's a browser based application it can be accessed worldwide.

• **Reliable-** our application provides the services according the users satisfaction and interest, and designed as per users requirements and more user friendly, so the application is more reliable.

• **Maintainability-** as we update the software regularly it will be easy to maintain it. Application is designed in such a way that future modifications and enhancements can be done easily.

• **Efficiency-** The application provides the efficient results as it uses data mining technique for crop yield prediction.

• **Re-usability**– The system is a web based application, once the user creates an account; user can access the system multiple times.

• **Quality of Service-** System provides the services as per the user requirements. System is more user friendly. Once user registered into the system, user can access the system multiple times. As we use data mining techniques, system generates more accurate results.

4.5 Communication Interface Requirements

• Hyper Text Transfer Protocol (HTTP) is used to transmit documents around network.
• Mainly the website is developed using ASP.NET as front and SQL Server 2005 as backend.
• The website is based on three tier architecture with data server, application server and a client.

Chapter 5

System design

The purpose of the design phase is to plan a solution of the problem specified by the requirements document. This phase is the first step in moving from the problem domain to the solution domain. In other words, starting with what is needed, design takes us toward how to satisfy the needs. The design of a system is perhaps the most critical factor affecting the quality of the software; it has a major impact on the later phases particularly testing and maintenance.
The design activity often results in three separate outputs –
• Architecture design.
• High level design.
• Detailed design.

Architecture Design:
Architecture focuses on looking at a system as a combination of many different components, and how they interact with each other to produce the desired result. The focus is on identifying components or subsystems and how they connect. In other words, the focus is on what major components are needed.

High Level Design:
In high level design identifies the modules that should be built for developing the system and the specifications of these modules. At the end of system design all major data structures, file format, output formats, etc., are also fixed. The focus is on identifying the modules. In other words, the attention is on what modules are needed.

Detailed Design:
In the detailed design the internal logic of each of the modules is specified. The focus is on designing the logic for each of the modules. In other words how modules can be implemented in software is the issue.
A design methodology is a systematic approach to creating a design by application of

a set of techniques and guidelines. Most methodologies focus on high level design.

5.1 Architectural design

In this project three tier architecture is used.

5.1.1 Introduction

As a developer, the .NET framework and Visual Studio present many choices for choosing the right architecture, from placing the data access code directly in the UI through datasets and data source controls, to creating a data access layer that talks to the database, all the way to creating an n-tier architecture approach that consists of multiple layers, and use data-transfer objects to pass data back and forth.

Layer: A layer is a reusable portion of code that performs a specific function. In the .NET environment, a layer is usually setup as a project that represents this specific function. This specific layer is in charge of working with other layers to perform some specific goal. In an application where the presentation layer needs to extract information from a back-end database, the presentation would utilize a series of layers to retrieve the data, rather than having the database calls embedded directly within itself. Now we will look briefly at the latter situation first.

5.1.2 Three-Tier Architecture

Three tier architecture consists of three layers. They are:

a) The Data Layer:

The key component to most applications is the data. The data has to be served to the presentation layer somehow. The data layer is a separate component (often setup as a separate single or group of projects in a .NET solution), whose sole purpose is to serve up the data from the database and return it to the caller. Through this approach, data can be logically reused, meaning that a portion of an application reusing the same query can make a call to one data layer method, instead of embedding the query multiple times. This is generally more maintainable.

b) Business Layer:

Though a web site could talk to the data access layer directly, it usually goes through another layer called the business layer. The business layer is vital in that it validates the input conditions before calling a method from the data layer. This

ensures the data input is correct before proceeding, and can often ensure that the outputs are correct as well. This validation of input is called business rules, meaning the rules that the business layer uses to make "judgments" about the data. One of the best reasons for reusing logic is that applications that start off small usually grow in functionality. The business layer helps move logic to a central layer for "maximum re-usability."

c) Presentation Layer:

The ASP.NET web site or windows forms application (the UI for the project) is called the presentation layer. The presentation layer is the most important layer simply because it's the one that everyone sees and uses. Even with a well structured business and data layer, if the presentation layer is designed poorly, this gives the users a poor view of the system.

The **presentation tier** contains the UI (User Interface) elements of the site, and includes all the logic that managers the interaction between the visitor and the client's business. (ASP.NET Web Forms, Web User Controls, ASP.NET Master Pages)
The **business tier** receives requests from the presentation tier and returns a result to the presentation tier depending on the business logic it contains. (C Classes)
The **data tier** is responsible for storing the application's data and sending it to the business tier when requested. (SQL Server Stored Procedures).

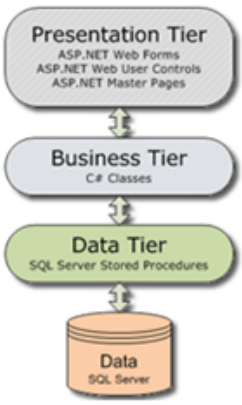


Figure 5.1: Three tier Architecture

5.2 High level design

5.2.1 Data flow diagram

A data flow diagram (DFD) is a graphical representation of the "flow" of data through an information system. DFDs can also be used for the visualization of data processing (structured design).
On a DFD, data items flow from an external data source or an internal data store to an internal data store or an external data sink, via an internal process.
A DFD provides no information about the timing of processes, or about whether processes will operate in sequence or in parallel. It is therefore quite different from a flowchart, which shows the flow of control through an algorithm, allowing a reader to determine what operations will be performed, in what order, and under what circumstances, but not what kinds of data will be input to and output from the system, nor where the data will come from and go to, nor where the data will be stored (all of which are shown on a DFD).

Symbols used in DFD's:

Processes A process transforms data values. The lowest processes are our functions without side effects.

Data Flows A data flow connects the output of an object or process to the input of another object or process. It represents the intermediate data values within the computation. It is draws as an arrow between the procedure and the consumer of the data value. The arrow is labeled with the description of the data, usually its name or type.

Actors An actor is an active object that drives the data flow graph by producing or consuming values. Actors are attached to the inputs and the outputs of a dataflow graph. In sense, the actors lie on the boundary of the flow graph but terminate the flow of data as sources and sinks of data, and so are sometimes called terminators.

Data Store A data store is a passive object within a data flow diagram that stores data for later access. Unlike an actor, a data store does not generate any operations on its own but merely responds to requests to store and access data.

Context data flow diagram It is common practice to draw a context-level data flow diagram first, which shows the interaction between the system and external agents which act as data sources and data sinks. On the context diagram (also known as the

'Level 0 DFD') the system's interactions with the outside world are modeled purely in terms of data flows across the system boundary. The context diagram shows the entire system as a single process, and gives no clues as to its internal organization.

This context-level DFD is next "exploded", to produce a Level 1 DFD that shows some of the detail of the system being modeled. The Level 1 DFD shows how the system is divided into sub-systems (processes), each of which deals with one or more of the data flows to or from an external agent, and which together provide all of the functionality of the system as a whole. It also identifies internal data stores that must be present in order for the system to do its job, and shows the flow of data between the various parts of the system.

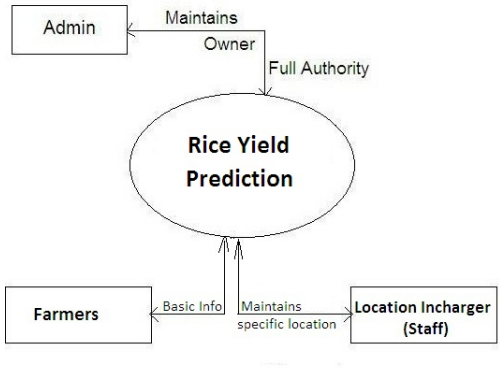


Figure 5.2: Context Data Flow Diagram. (Level 0)

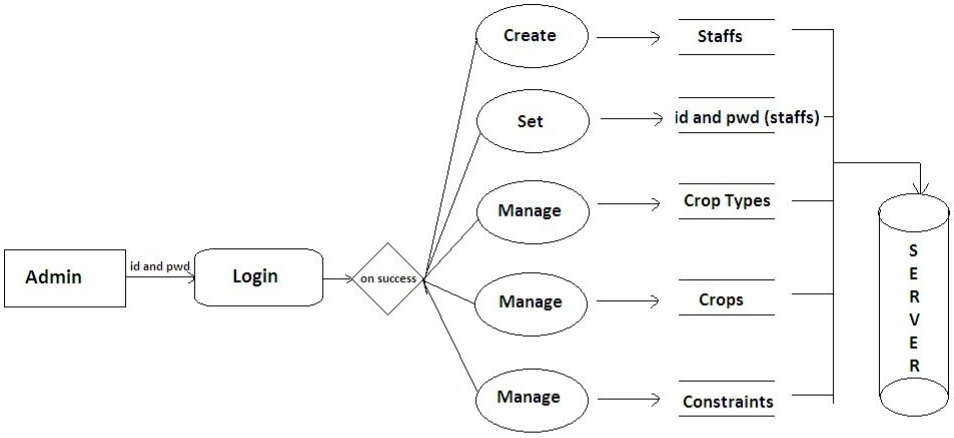


Figure 5.3: Data Flow Diagram - admin. (Level 1)

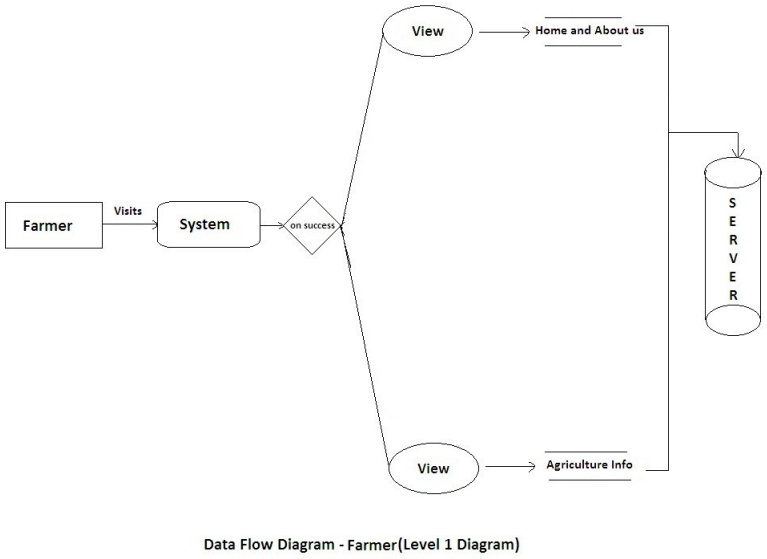


Figure 5.4: Data Flow Diagram - farmer (Level 1)

5.3 Detailed Design

5.3.1 Use Case Diagram

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a

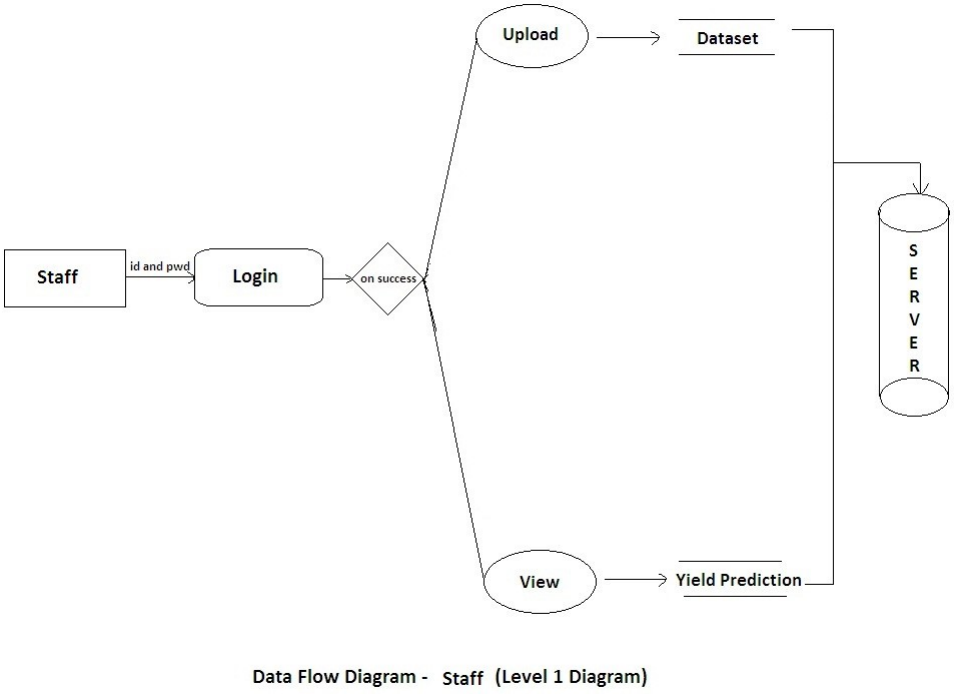


Figure 5.5: Data Flow Diagram - staff (Level 1)

graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases.
The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted. Interaction among actors is not shown on the use case diagram. If this interaction is essential to a coherent description of the desired behavior, perhaps the system or use case boundaries should be re-examined. Alternatively, interaction among actors can be part of the assumptions used in the use case.

Use cases

A use case describes a sequence of actions that provide something of measurable value to an actor and is drawn as a horizontal ellipse.

Actors

An actor is a person, organization, or external system that plays a role in one or more interactions with the system.

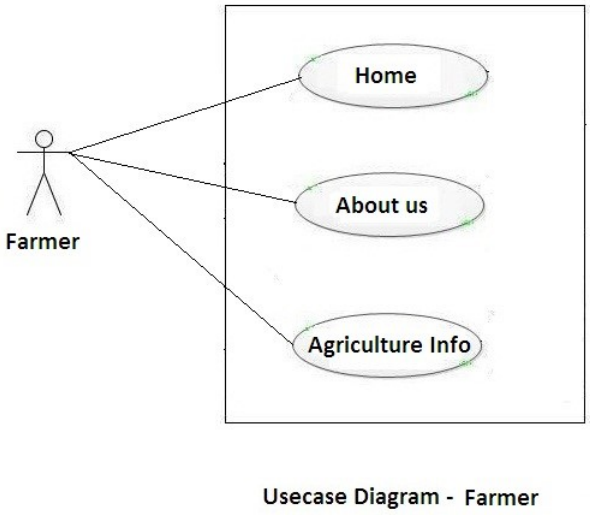


Figure 5.6: Usecase diagram - Farmer

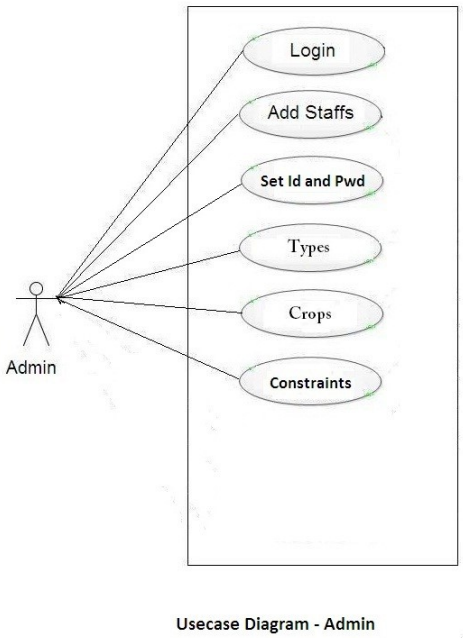


Figure 5.7: Usecase diagram - Admin

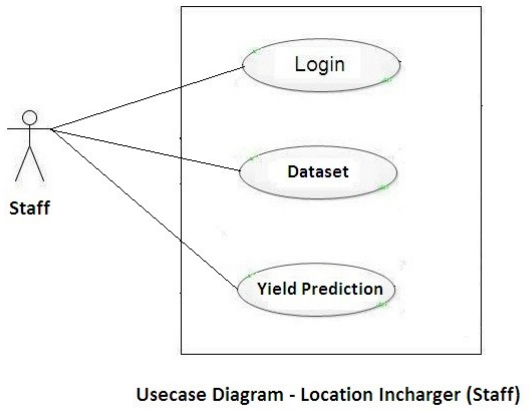


Figure 5.8: Usecase diagram - Staff

System boundary boxes

A rectangle is drawn around the use cases, called the system boundary box, to indicate the scope of system. Anything within the box represents functionality that is in scope and anything outside the box is not.

Sequence Diagram

A sequence diagram in Unified Modelling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart.

Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.

A sequence diagram shows, as parallel vertical lines (lifelines), different processes or objects that live simultaneously, and, as horizontal arrows, the messages exchanged between them, in the order in which they occur. This allows the specification of simple runtime scenarios in a graphical manner.

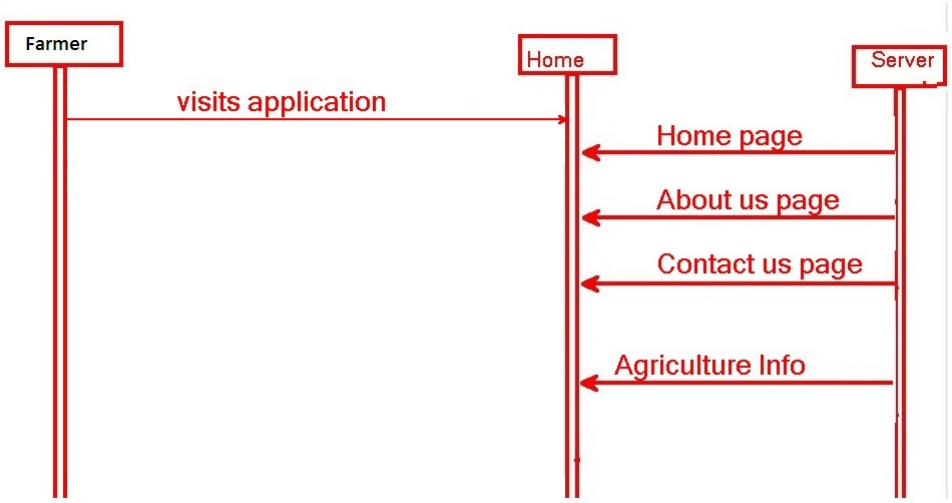


Figure 5.9: Sequence diagram - farmer/visitor

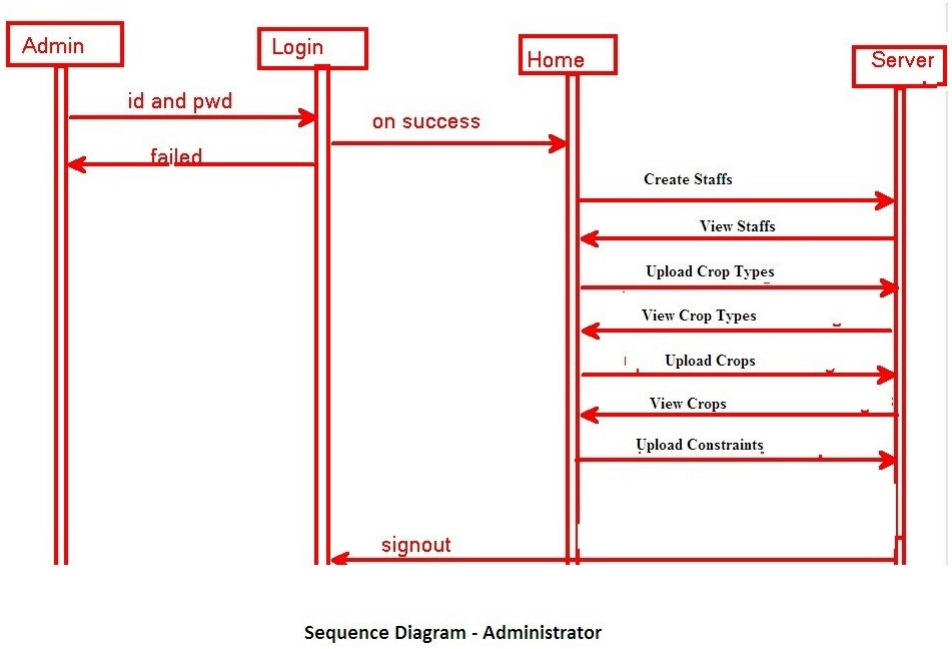


Figure 5.10: Sequence diagram - administrator

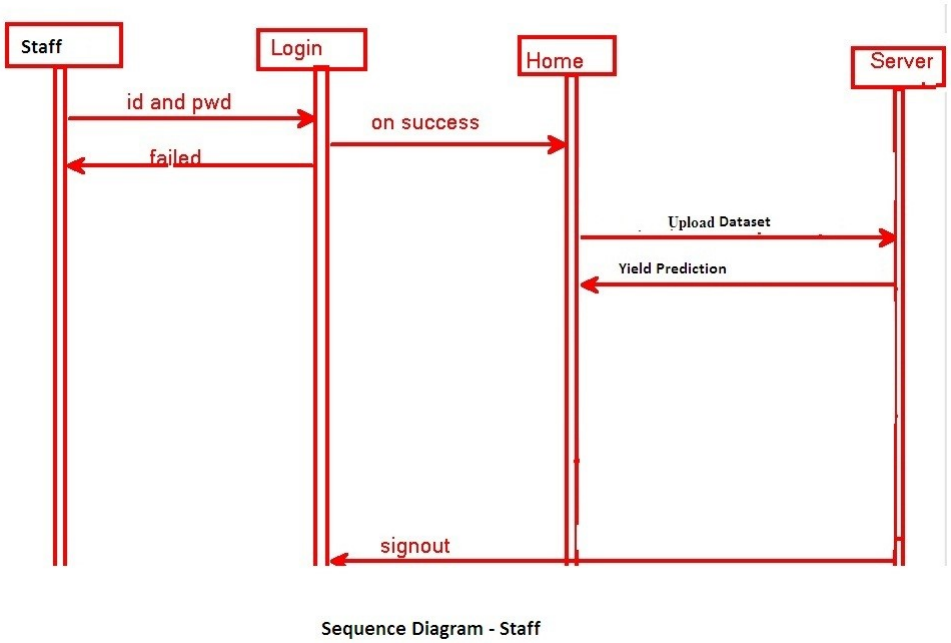


Figure 5.11: Sequence diagram - staff

5.4 System Architecture Diagram

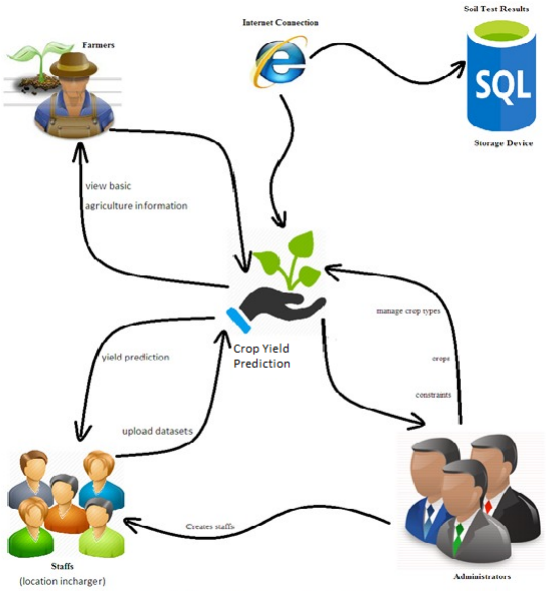


Figure 5.12: crop yield prediction

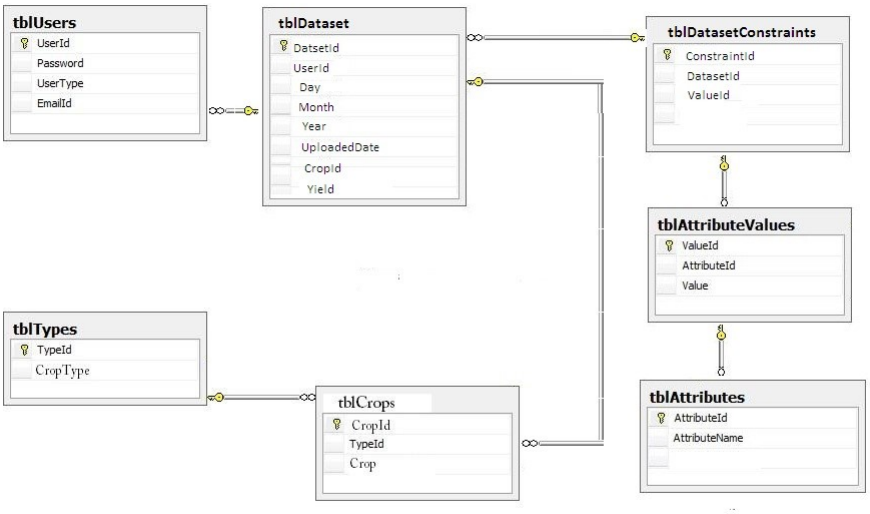


Figure 5.13: Storage structure

Chapter 6

Implementation and Experimental Analysis

Implementation can be described as realization of an application, or execution of a plan, idea, model, design, specification, standard, algorithm, or policy. In computer science, an implementation is explained as realization of a technical specification or algorithm as a program, a software component, or any other computer system through computer programming and deployment. Many implementations may exist for a given specification or standard.

Supervised Learning

Classification Rules
Classification is a process of finding a model (or function) that describes and distinguishes data classes or concepts. The model is derived based on the analysis of a set of training data (i.e., data objects for which the class labels are known). The model is used to predict the class label of objects for which the class label is unknown.

Example: Suppose the sales manager of All Electronics want to classify a large set of items in the store, based on three kinds of responses to sales campaign: good response, mild response and no response. The model for each of these three classes is derived based on the descriptive features of the items, such as price, brand,type and category. The resulting classification should maximally distinguish each class from the others, presenting an organized picture of the data set.

6.1 Naive Bayes Algorithm Steps

Step 1: Scan the dataset

Step 2: Calculate the probability of each attribute value. [n, nc, m, p]

Step 3: Apply the formulae P(attributevalue(ai)/subjectvalue(vj))=(nc + mp)/(n+m) where:
n = the number of training examples for which v = vj
nc = number of examples for which v = vj and a = ai
p = a priori estimate for P(aijvj)
m = the equivalent sample size

Step 4: Multiply the probabilities by p

Step 5: Compare the values and classify the attribute values to one of the pre-defined set of class.

Sample Example

Parameters – Temp,Humidity,Area [m=3]
Outcome – 100T, 200T [p=1/2=0.5]

Training Dataset

Soil Type	Temp(L,M,H)	Humidity(L,M,H)	Area(100, 200,300)	Yield(subject)
2013	L	L	100	100T
2014	M	L	200	100T
2015	L	M	100	200T
2016	M	L	300	200T
2017	H	M	200	100T

2018 Parameters – L, M, 200 Which Yield- ?

P=[n_c + (m*p)]/(n+m)

100T	200T
L P=[n_c + (m*p)]/(n+m) n=2, n_c=1,m=3,p=0.5 p=[1+(3*0.5)]/(2+3) p=0.5	L P=[n_c + (m*p)]/(n+m) n=2, n_c=1,m=3,p=0.5 p=[1+(3*0.5)]/(2+3) p=0.5
M P=[n_c + (m*p)]/(n+m) n=2, n_c=2,m=3,p=0.5 p=[2+(3*0.5)]/(2+3) p=0.7	M P=[n_c + (m*p)]/(n+m) n=2, n_c=2,m=3,p=0.5 p=[2+(3*0.5)]/(2+3) p=0.3
200 P=[n_c + (m*p)]/(n+m) n=2, n_c=1,m=3,p=0.5 p=[1+(3*0.5)]/(2+3) p=0.5	200 P=[n_c + (m*p)]/(n+m) n=2, n_c=1,m=3,p=0.5 p=[1+(3*0.5)]/(2+3) p=0.5

100T = 0.7 * 0.7 * 0.5 * 0.5 (p) 200T = 0.5 * 0.3 * 0.5 * 0.5(p)

=0.1225 =0.0225

Since 100T >200T

So this new SoilInput6 is classified to 100T

EXPERIMENTAL AND RESULT ANALYSIS

DATASETS

We have collected the datasets for the crops Ragi and Paddy from the agriculture department for the region “Mysuru”. The Paddy dataset has around 1500 records and Ragi dataset has around 1200 records. It contains the parameters required for the crop yield prediction such as Ph, Organic Carbon(Oc) ,Nitrogen(N) ,Phosphorus(P), Potassium(K) , Sulphur(S), Zinc(Zn) Iron(Fe), Temperature, Rainfall, Year ,Crop Yield.

6.2 Snapshot of datasets

Paddy Dataset

N509 fx Paddy(rice)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Name	PH	organic c	nitrogen(	phosphor	potassiun	sulphur(s	zinc(zn)	iron(fe)	Tempera	Rainfall	Region	Year	Crop	Yield
2	Sunanada	5.7	7	188.43	7	7.15	176	127.4	5.7	26.12	68.21	Belagunda	2012	Paddy(rice	15
3	Kumar	5.78	7.6	190.1	7.6	6.4	160.2	126.3	5.78	25.28	105.43	Belagunda	2013	Paddy(rice	10
4	Rajappa	5.83	7.24	240.6	7.24	6.2	140	123	5.83	25.75	105.14	Belagunda	2014	Paddy(rice	15
5	Ramappa	5.9	7.15	184.4	7.15	6	126.2	128.6	5.9	26.66	88.69	Belagunda	2015	Paddy(rice	10
6	Nangunda	5.8	6.4	177	6.4	5.9	124.4	124.8	5.8	26.07	41.71	Belagunda	2016	Paddy(rice	25
7	Prakash	5.76	6.2	161.2	6.2	5.7	120.6	127.6	5.76	25.95	140.6	Belagunda	2017	Paddy(rice	24
8	Suresh	7.2	6	127.4	6	6.3	110	158	7.2	26.12	68.21	Nagarle	2012	Paddy(rice	20
9	Mahadev	7.4	5.9	126.3	5.9	6.4	177.8	126.3	7.4	25.28	105.43	Nagarle	2013	Paddy(rice	18
10	Mahadeva	7	5.7	123	5.7	6.2	127	123	7	25.75	105.14	Nagarle	2014	Paddy(rice	16
11	Narayana	7.6	6.3	128.6	6.3	7	116.4	128.6	7.6	26.66	88.69	Nagarle	2015	Paddy(rice	15
12	Mahadeva	7.24	6.4	124.8	6.4	7.6	105.55	124.8	7.24	26.07	41.71	Nagarle	2016	Paddy(rice	28
13	Ramesh	7.15	6.2	127.6	6.2	7.24	122.4	127.6	7.15	25.95	140.6	Nagarle	2017	Paddy(rice	28
14	Siddaraju	6.4	7	158	7	7.15	191.2	158	6.4	26.12	68.21	Nagarle	2012	Paddy(rice	18
15	Ramappa	6.2	7.6	143.3	7.6	6.4	143.4	143.3	6.2	25.28	105.43	Nagarle	2013	Paddy(rice	19
16	Siddaraju	6	7.24	137.4	7.24	6.2	140.2	137.4	6	25.75	105.14	Nagarle	2014	Paddy(rice	12
17	Gowdru	5.9	7.15	122	7.15	6	116.4	122	5.9	26.66	88.69	Nagarle	2015	Paddy(rice	10
18	Siddaraju	5.7	6.4	124.1	6.4	6.2	105.55	124.1	5.7	26.07	41.71	Nagarle	2016	Paddy(rice	8
19	Siddaraju	6.3	6.2	125.6	6.2	6	122.4	125.6	6.3	25.95	140.6	Nagarle	2017	Paddy(rice	12
20	Sushila	6.4	6	110	6	5.9	191.2	110	6.4	26.12	68.21	Haniyamb	2012	Paddy(rice	15
21	Naganaya	6.2	5.9	161.2	5.9	5.7	143.4	161.2	6.2	25.28	105.43	Haniyamb	2013	Paddy(rice	10
22	Malamma	6.2	5.7	184.4	5.7	6.3	140.2	184.4	6.2	25.75	105.14	Haniyamb	2014	Paddy(rice	15
23	Naganaya	6.1	6.3	155	6.3	6.4	155.8	155	6.1	26.66	88.69	Haniyamb	2015	Paddy(rice	10
24	Kumari	5.9	6.4	161.2	6.4	6.2	154.2	161.2	5.9	26.07	41.71	Haniyamb	2016	Paddy(rice	25

Ragi Dataset

	Name	PH	organic c	nitrogen(	phosphor	potassiun	sulphur(s	zinc(zn)	iron(fe)	Tempera	Rainfall	Region	Year	Crop	Yield
2	Mahadev	7.4	5.9	126.3	5.9	6.4	177.8	126.3	7.4	25.28	105.43	Nagarle	2013	Wheat	18
3	Mahadeva	7	5.7	123	5.7	6.2	127	123	7	25.75	105.14	Nagarle	2014	Wheat	16
4	Narayana	7.6	6.3	128.6	6.3	7	116.4	128.6	7.6	26.66	88.69	Nagarle	2015	Wheat	15
5	Mahadeva	7.24	6.4	124.8	6.4	7.6	105.55	124.8	7.24	26.07	41.71	Nagarle	2016	Wheat	28
6	Ramesh	7.15	6.2	127.6	6.2	7.24	122.4	127.6	7.15	25.95	140.6	Nagarle	2017	Wheat	28
7	Siddaraju	6.4	7	158	7	7.15	191.2	158	6.4	26.12	68.21	Nagarle	2012	Wheat	18
8	Ramappa	6.2	7.6	143.3	7.6	6.4	143.4	143.3	6.2	25.28	105.43	Nagarle	2013	Wheat	19
9	Siddaraju	6	7.24	137.4	7.24	6.2	140.2	137.4	6	25.75	105.14	Nagarle	2014	Wheat	12
10	Gowdru	5.9	7.15	122	7.15	6	116.4	122	5.9	26.66	88.69	Nagarle	2015	Wheat	10
11	Ramappa	5.9	7.15	184.4	7.15	6	126.2	128.6	5.9	26.66	88.69	Belagunda	2015	Wheat	10
12	Nangunda	5.8	6.4	177	6.4	5.9	124.4	124.8	5.8	26.07	41.71	Belagunda	2016	Wheat	25
13	Prakash	5.76	6.2	161.2	6.2	5.7	120.6	127.6	5.76	25.95	140.6	Belagunda	2017	Wheat	24
14	Suresh	7.2	6	127.4	6	6.3	110	158	7.2	26.12	68.21	Nagarle	2012	Wheat	20
15	Mahadev	7.4	5.9	126.3	5.9	6.4	177.8	126.3	7.4	25.28	105.43	Nagarle	2013	Wheat	18
16	Mahadeva	7	5.7	123	5.7	6.2	127	123	7	25.75	105.14	Nagarle	2014	Wheat	16
17	Narayana	7.6	6.3	128.6	6.3	7	116.4	128.6	7.6	26.66	88.69	Nagarle	2015	Wheat	15
18	Mahadeva	7.24	6.4	124.8	6.4	7.6	105.55	124.8	7.24	26.07	41.71	Nagarle	2016	Wheat	28
19	Ramesh	7.15	6.2	127.6	6.2	7.24	122.4	127.6	7.15	25.95	140.6	Nagarle	2017	Wheat	28
20	Siddaraju	6.4	7	158	7	7.15	191.2	158	6.4	26.12	68.21	Nagarle	2012	Wheat	18
21	Ramappa	6.2	7.6	143.3	7.6	6.4	143.4	143.3	6.2	25.28	105.43	Nagarle	2013	Wheat	19
22	Siddaraju	6	7.24	137.4	7.24	6.2	140.2	137.4	6	25.75	105.14	Nagarle	2014	Wheat	12
23	Gowdru	5.9	7.15	122	7.15	6	116.4	122	5.9	26.66	88.69	Nagarle	2015	Wheat	10
24	Amir	6.73	6.2	38.5	6.2	6.4	155.8	38.5	6.73	25.95	140.6	Belagunda	2017	Wheat	10

6.3 Code Snippet for Naive Bayes Algorithm:

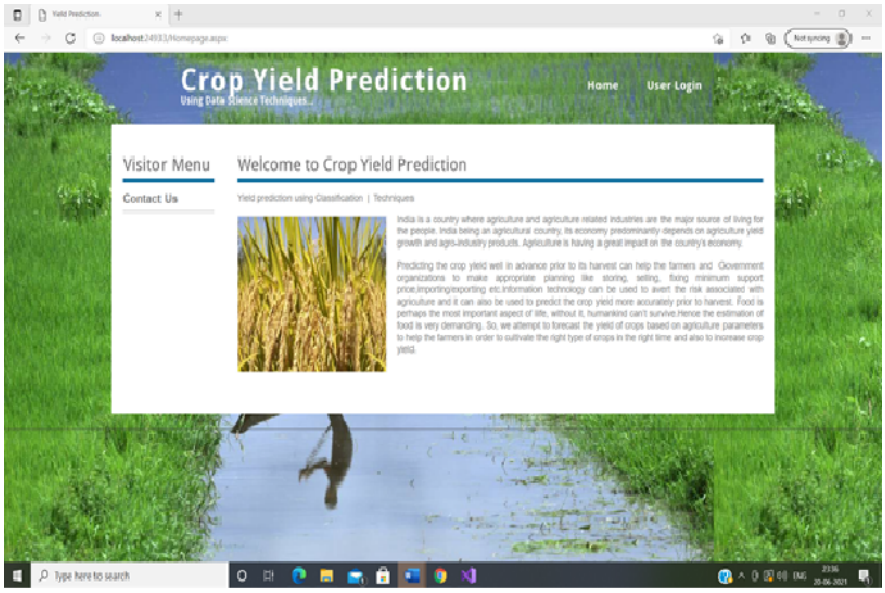
```
double pi;
    int nc, n;
    double result;
    ArrayList output = new ArrayList();
    ArrayList mul = new ArrayList();

    //function which contains the algorithm steps
    private string NaiveBayes(string[] values)
    {
      ArrayList s = new ArrayList();
      output.Clear();

      //try
      //{
      s = GetSubject();

      int m = 10;
      double numer = 1.0;
      double dino = double.Parse(s.Count.ToString());
      double p = numer / dino;

      string[] features = { "PH","organic carbon(oC)", "nitrogen(n)" , "phosphorus(p)", "potassium(k)",
                  "sulphur(s)","zinc(zn)","iron(fe)","Temperature","Rainfall" };

      for (int i = 0; i < s.Count; i++)
      {
        mul.Clear();

        for (int j = 0; j < features.Length; j++)
        {
          n = 0;
          nc = 0;

          for (int d = 0; d < dt.Rows.Count; d++)
          {
            if (dt.Rows[d][j + 1].ToString().Equals(values[j]))
            {
              ++n;

              if (dt.Rows[d][m + 4].ToString().Equals(s[i]))

                ++nc;
            }
          }

          double x = m * p;
          double y = n + m;
          double z = nc + x;

          pi = z / y;
          mul.Add(Math.Abs(pi));
        }

        double mulres = 1.0;
```

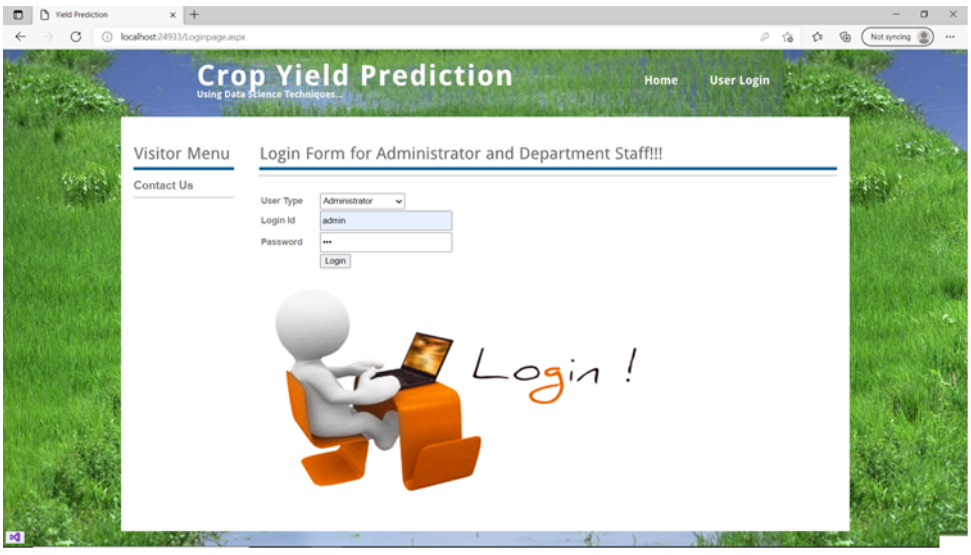
```
        for (int z = 0; z < mul.Count; z++)
        {
            mulres *= double.Parse(mul[z].ToString());
        }

        result = mulres * p;
        output.Add(Math.Abs(result));
    }

    ArrayList list1 = new ArrayList();

    for (int x = 0; x < s.Count; x++)
    {
        list1.Add(output[x]);
    }

    list1.Sort();
    list1.Reverse();

    string _output = null;

    for (int y = 0; y < s.Count; y++)
    {
        if (output[y].Equals(list1[0]))
        {
            _output = s[y].ToString();
            return _output;
        }
    }
    //}
    //catch
    //{

    //}

    return _output;
}
```

6.4 GUI:

Login Module

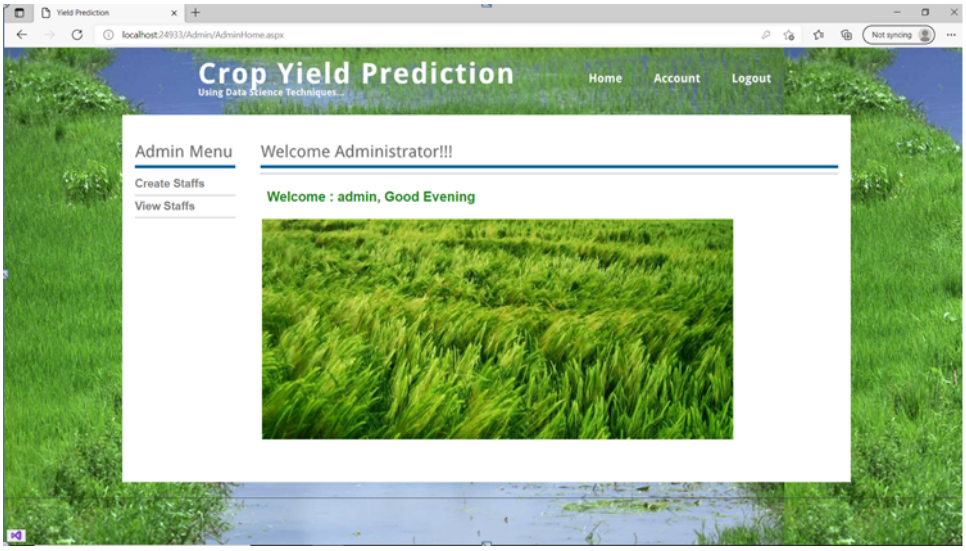
Description: interface designed using 2 textboxes, submit button, validiators and label to display the error message. Initially user will enter the user id and password and clicks on the submit button, on click on submit button following instructions gets executed and check the user authentication, if user authentication is valid user will be redirected to the home page or else system displays a error message to the user saying user id and password incorrect.

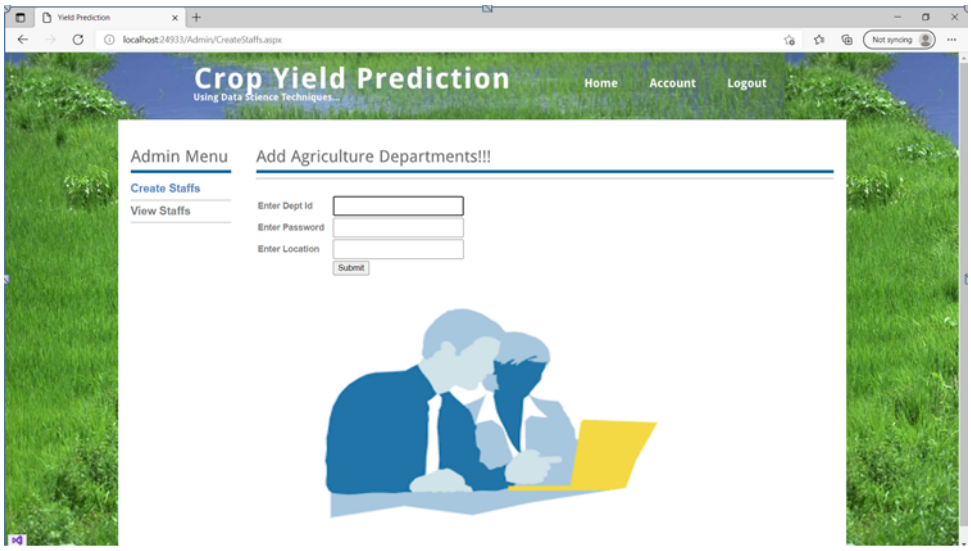




Admin - Create Agriculture Departments or Staffs Module

Description: Administrator of the application will register the staffs (agriculture departments). Proposed system is an generic application useful for the multiple regions. Administrator of the system sets the unique id and password for every staff and registers to the system. Interface designed using 3 textboxes, 1 submit button, validators and label to display the error message. Admin will enter unique id and password and location and clicks on the submit button. On click of submit button the following code gets executed and registers the staff into the system. If staff id is already exists, system will display "staff id" already exists.



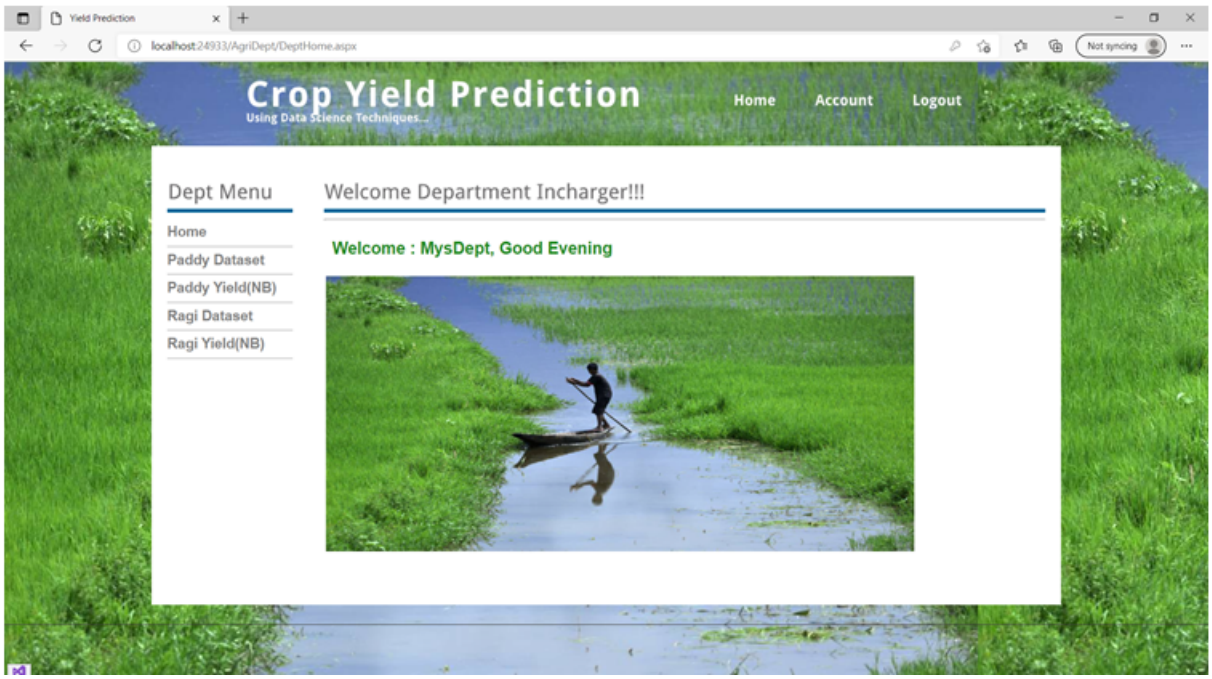


Admin - Update Password Module

Description: Here administrator updates the password and interface designed using 3 textboxes to enter old password, new password and confirm password and 1 submit button and validators. Admin will enter the old password , new password and confirm password and clicks on submit button, on click the following code gets executed and initially checks whether inputted old password is correct or not, if it is correct system will update the new password or else error message displayed for the user saying "Old password is not correct".

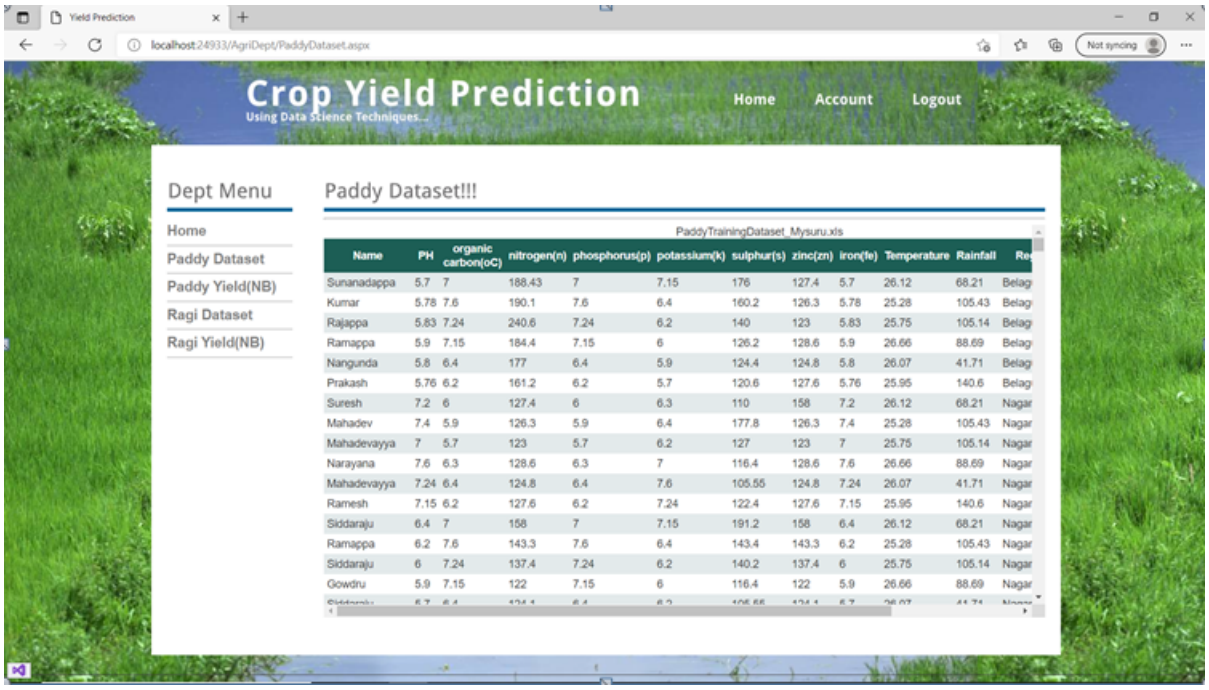
Agriculture Department Home Page

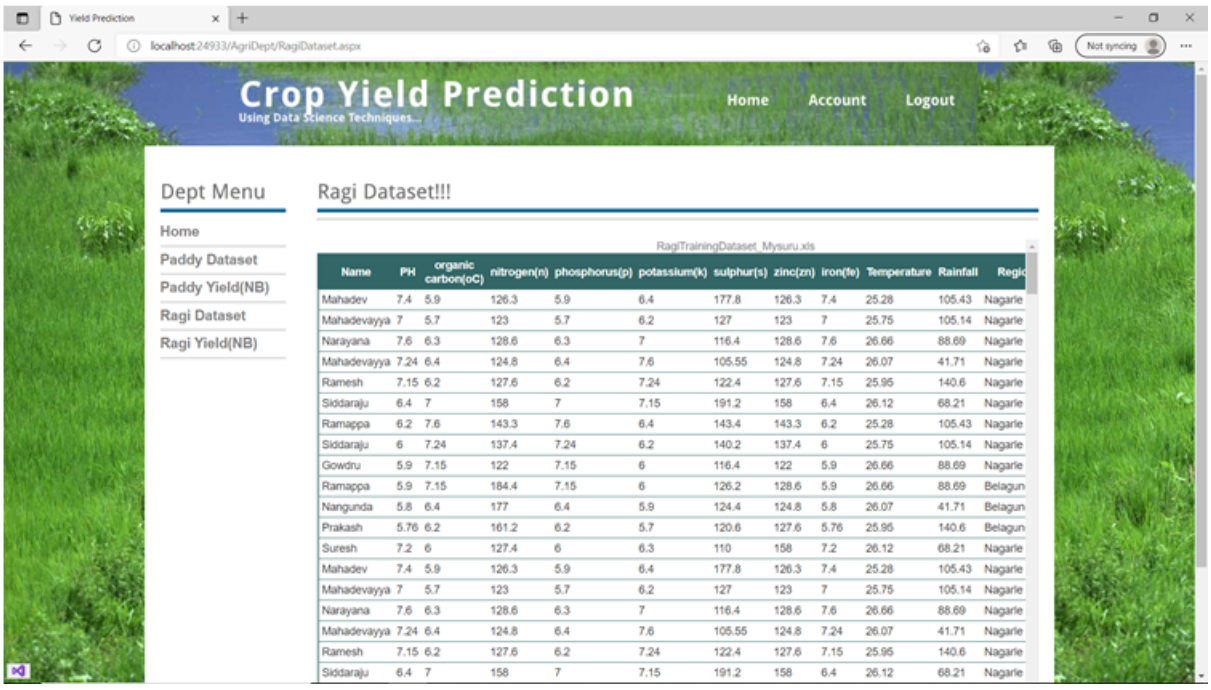
Description: This is the staff home page, a welcome page designed using label, when user enters the home page, user will be displayed and good morning or good afternoon message based on the system date time.



Agriculture Department Datasets Module

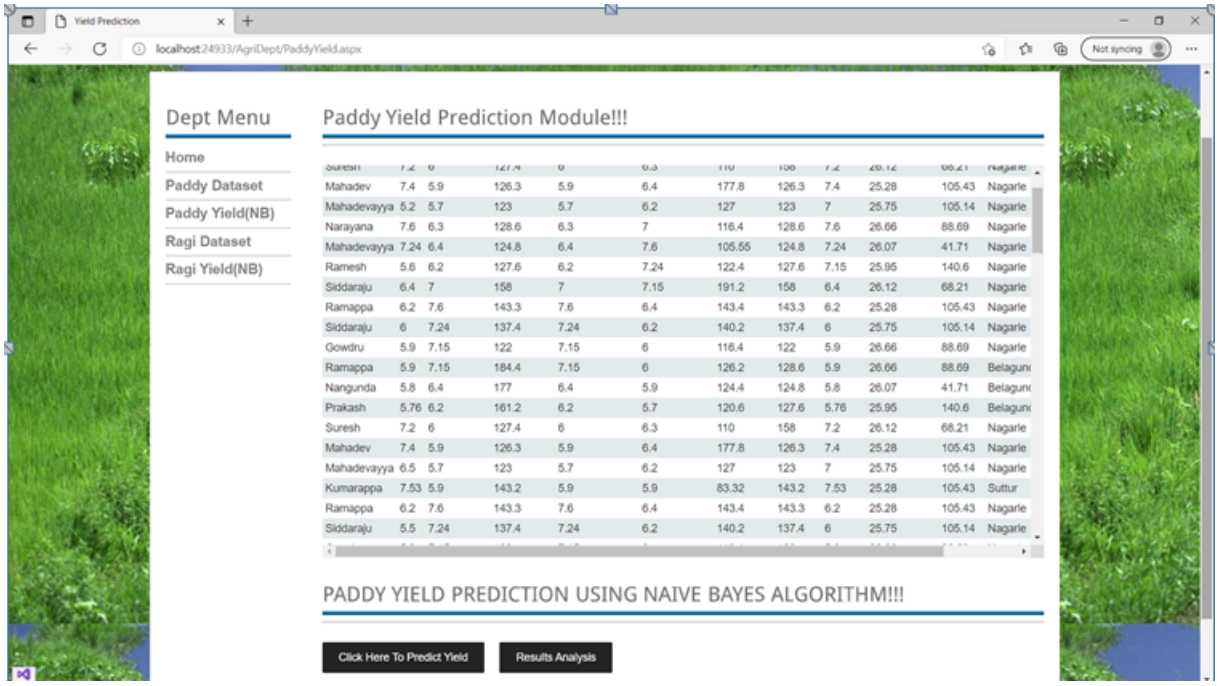
Description: This is the training datasets module where interface contains a grid-view. Training datasets stored using excel sheets and following code gets executed to import data from excel sheet and display on gridview.

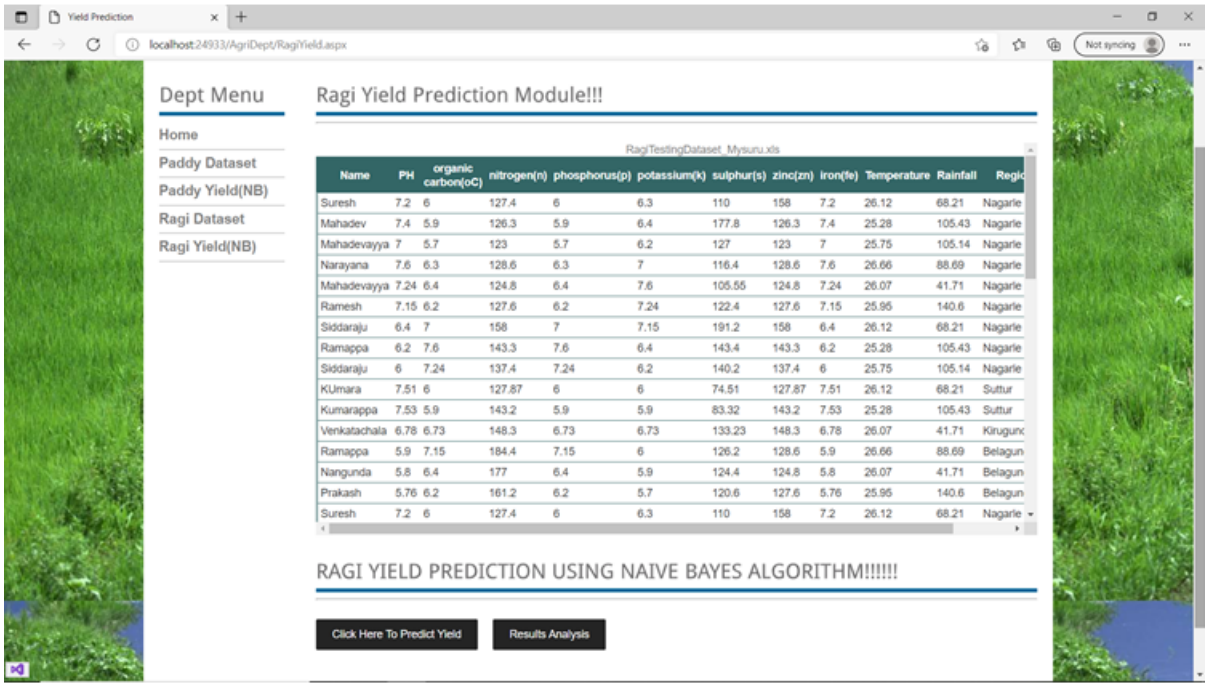




Agriculture Department - Yield Prediction Module

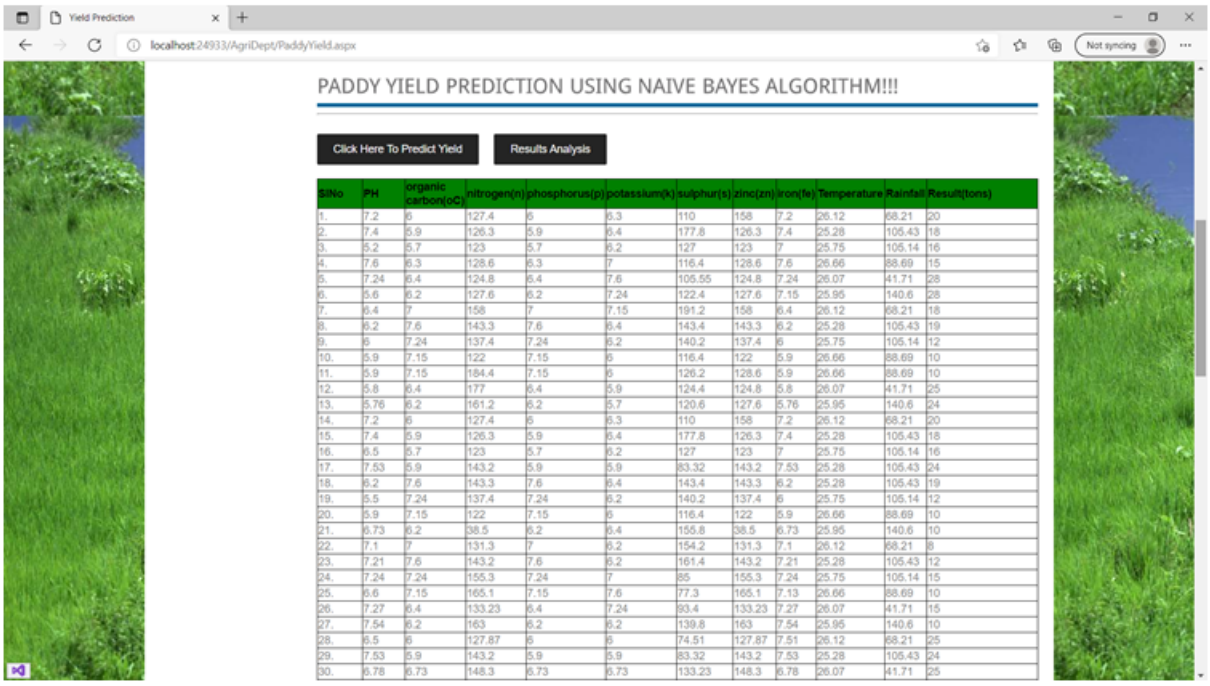
Description: This is the main module in the project which predicts crop yield using machine learning algorithms. Interface build using gridviews and submit buttons. On click of predict submit following code gets executed and results will be predicted and displayed on gridview.



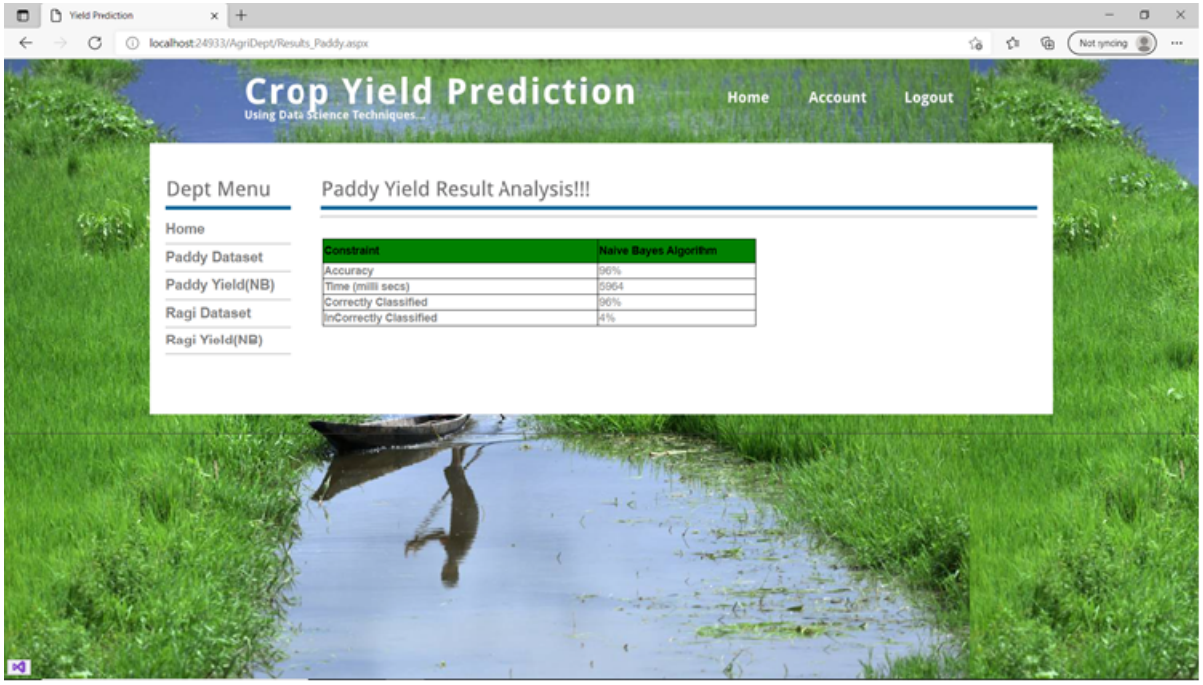


Agriculture Department - Result Analysis Module

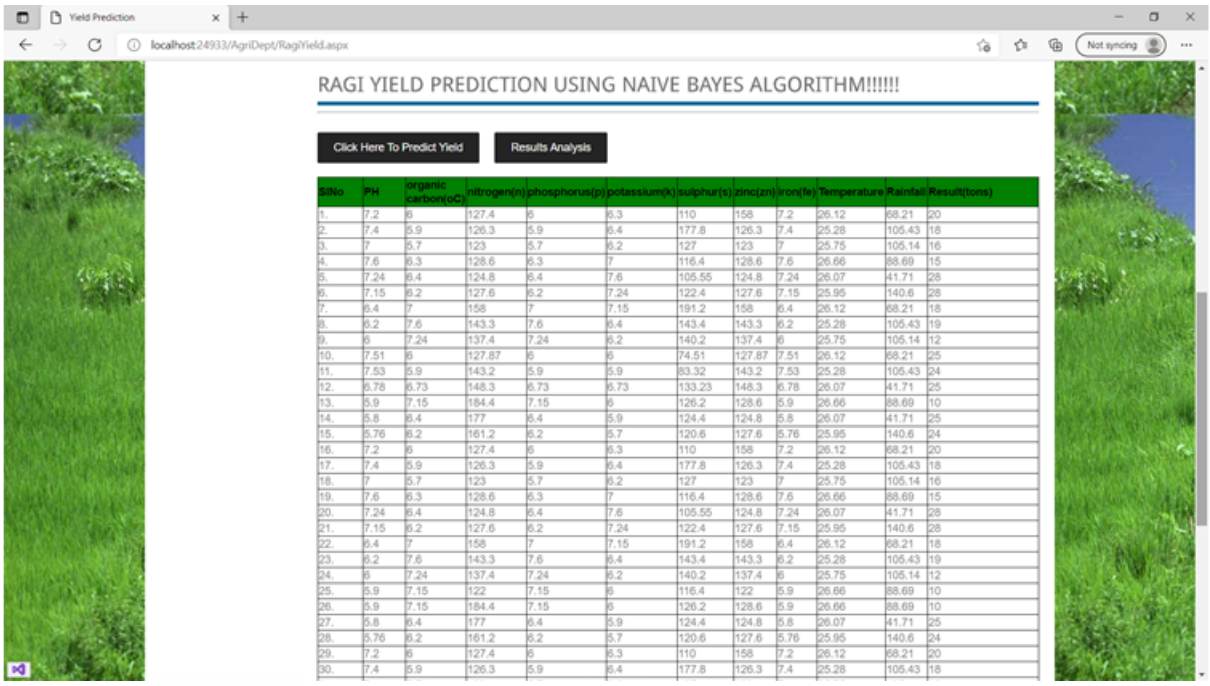
Description: This is the result analysis module where we display the results of the algorithm such as accuracy, efficiency, precision and recall. Interface contains a dynamic table and the following code enclosed on page load function , so when page gets executed the following instructions gets executed and results of the algorithm predicted and displayed on dynamic table.



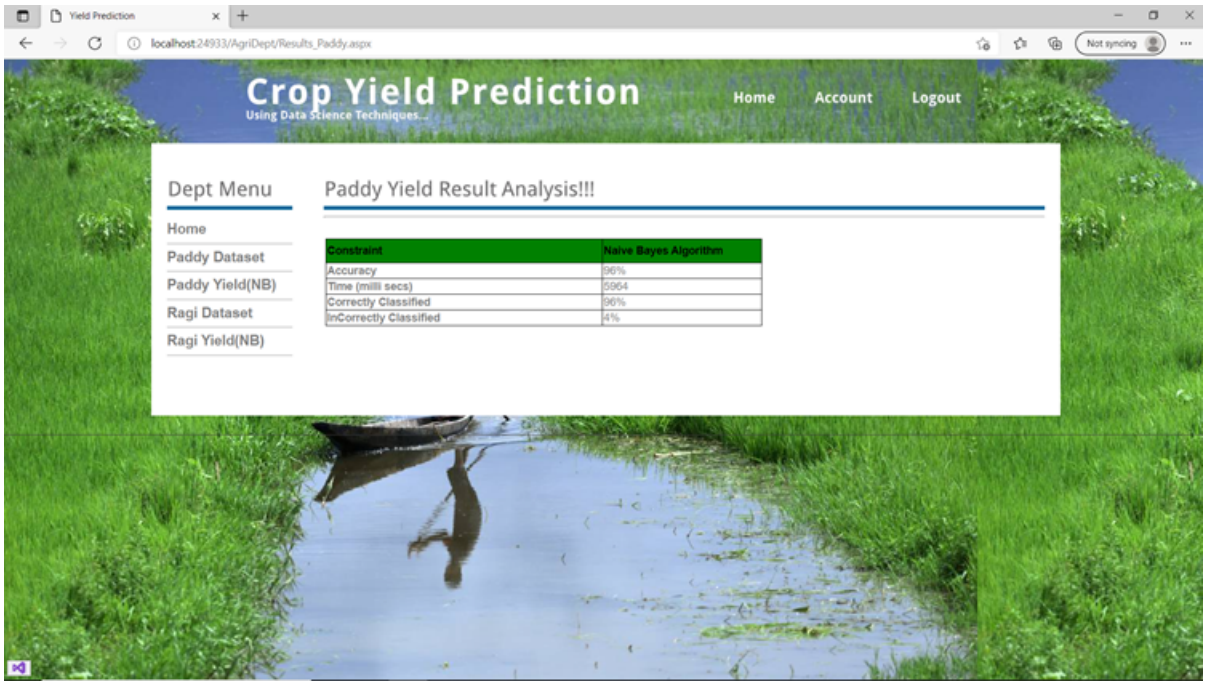
SlNo	PH	organic carbon(oC)	nitrogen(n)	phosphorus(p)	potassium(k)	sulphur(s)	zinc(zn)	iron(fe)	Temperature	Rainfall	Result(tons)
1.	7.2	6	127.4	6	6.3	110	158	7.2	26.12	68.21	20
2.	7.4	5.9	126.3	5.9	6.4	177.8	126.3	7.4	25.28	105.43	18
3.	5.2	5.7	123	5.7	6.2	127	123	7	25.75	105.14	16
4.	7.6	6.3	128.6	6.3	7	116.4	128.6	7.6	26.66	88.69	15
5.	7.24	6.4	124.8	6.4	7.6	105.55	124.8	7.24	26.07	41.71	28
6.	5.6	6.2	127.6	6.2	7.24	122.4	127.6	7.15	25.95	140.6	28
7.	6.4	7	158	7	7.15	191.2	158	6.4	26.12	68.21	18
8.	6.2	7.6	143.3	7.6	6.4	143.4	143.3	6.2	25.28	105.43	19
9.	6	7.24	137.4	7.24	6.2	140.2	137.4	6	25.75	105.14	12
10.	5.9	7.15	122	7.15	6	116.4	122	5.9	26.66	88.69	10
11.	5.9	7.15	184.4	7.15	6	126.2	128.6	5.9	26.66	88.69	10
12.	5.8	6.4	177	6.4	5.9	124.4	124.8	5.8	26.07	41.71	25
13.	5.76	6.2	161.2	6.2	5.7	120.6	127.6	5.76	25.95	140.6	24
14.	7.2	6	127.4	6	6.3	110	158	7.2	26.12	68.21	20
15.	7.4	5.9	126.3	5.9	6.4	177.8	126.3	7.4	25.28	105.43	18
16.	6.5	5.7	123	5.7	6.2	127	123	7	25.75	105.14	16
17.	7.53	5.9	143.2	5.9	5.9	83.32	143.2	7.53	25.28	105.43	24
18.	6.2	7.6	143.3	7.6	6.4	143.4	143.3	6.2	25.28	105.43	19
19.	5.5	7.24	137.4	7.24	6.2	140.2	137.4	6	25.75	105.14	12
20.	5.9	7.15	122	7.15	6	116.4	122	5.9	26.66	88.69	10
21.	6.73	6.2	38.5	6.2	6.4	155.8	38.5	6.73	25.95	140.6	10
22.	7.1	7	131.3	7	6.2	154.2	131.3	7.1	26.12	68.21	8
23.	7.21	7.6	143.2	7.6	6.2	161.4	143.2	7.21	25.28	105.43	12
24.	7.24	7.24	155.3	7.24	7	85	155.3	7.24	25.75	105.14	15
25.	6.6	7.15	165.1	7.15	7.6	77.3	165.1	7.13	26.66	88.69	10
26.	7.27	6.4	133.23	6.4	7.24	93.4	133.23	7.27	26.07	41.71	15
27.	7.54	6.2	163	6.2	6.2	139.8	163	7.54	25.95	140.6	10
28.	6.5	6	127.87	6	6	74.51	127.87	7.51	26.12	68.21	25
29.	7.53	5.9	143.2	5.9	5.9	83.32	143.2	7.53	25.28	105.43	24
30.	6.78	6.73	148.3	6.73	6.73	133.23	148.3	6.78	26.07	41.71	25



Constraint	Naive Bayes Algorithm
Accuracy	96%
Time (milli secs)	5964
Correctly Classified	96%
InCorrectly Classified	4%



SlNo	PH	organic carbon(oC)	nitrogen(n)	phosphorus(p)	potassium(k)	sulphur(s)	zinc(zn)	iron(fe)	Temperature	Rainfall	Result(tons)
1.	7.2	6	127.4	6	6.3	110	158	7.2	26.12	68.21	20
2.	7.4	5.9	126.3	5.9	6.4	177.8	126.3	7.4	25.28	105.43	18
3.	7	5.7	123	5.7	6.2	127	123	7	25.75	105.14	16
4.	7.6	6.3	128.6	6.3	7	116.4	128.6	7.6	26.66	88.69	15
5.	7.24	6.4	124.8	6.4	7.6	105.55	124.8	7.24	26.07	41.71	28
6.	7.15	6.2	127.6	6.2	7.24	122.4	127.6	7.15	25.95	140.6	28
7.	6.4	7	158	7	7.15	191.2	158	6.4	26.12	68.21	18
8.	6.2	7.6	143.3	7.6	6.4	143.4	143.3	6.2	25.28	105.43	19
9.	6	7.24	137.4	7.24	6.2	140.2	137.4	6	25.75	105.14	12
10.	7.51	6	127.87	6	6	74.51	127.87	7.51	26.12	68.21	25
11.	7.53	5.9	143.2	5.9	5.9	83.32	143.2	7.53	25.28	105.43	24
12.	6.78	6.73	148.3	6.73	6.73	133.23	148.3	6.78	26.07	41.71	25
13.	5.9	7.15	184.4	7.15	6	126.2	128.6	5.9	26.66	88.69	10
14.	5.8	6.4	177	6.4	5.9	124.4	124.8	5.8	26.07	41.71	25
15.	5.76	6.2	161.2	6.2	5.7	120.6	127.6	5.76	25.95	140.6	24
16.	7.2	6	127.4	6	6.3	110	158	7.2	26.12	68.21	20
17.	7.4	5.9	126.3	5.9	6.4	177.8	126.3	7.4	25.28	105.43	18
18.	7	5.7	123	5.7	6.2	127	123	7	25.75	105.14	16
19.	7.6	6.3	128.6	6.3	7	116.4	128.6	7.6	26.66	88.69	15
20.	7.24	6.4	124.8	6.4	7.6	105.55	124.8	7.24	26.07	41.71	28
21.	7.15	6.2	127.6	6.2	7.24	122.4	127.6	7.15	25.95	140.6	28
22.	6.4	7	158	7	7.15	191.2	158	6.4	26.12	68.21	18
23.	6.2	7.6	143.3	7.6	6.4	143.4	143.3	6.2	25.28	105.43	19
24.	6	7.24	137.4	7.24	6.2	140.2	137.4	6	25.75	105.14	12
25.	5.9	7.15	122	7.15	6	116.4	122	5.9	26.66	88.69	10
26.	5.9	7.15	184.4	7.15	6	126.2	128.6	5.9	26.66	88.69	10
27.	5.8	6.4	177	6.4	5.9	124.4	124.8	5.8	26.07	41.71	25
28.	5.76	6.2	161.2	6.2	5.7	120.6	127.6	5.76	25.95	140.6	24
29.	7.2	6	127.4	6	6.3	110	158	7.2	26.12	68.21	20
30.	7.4	5.9	126.3	5.9	6.4	177.8	126.3	7.4	25.28	105.43	18



Constraint	Naive Bayes Algorithm
Accuracy	96%
Time (milli secs)	5964
Correctly Classified	96%
InCorrectly Classified	4%

Agriculture Departement - Update Password

Description: Here staff updates the password and interface designed using 3 textboxes to enter old password, new password and confirm password and 1 submit button and validators. Staff will enter the old password , new password and confirm password and clicks on submit button, on click the following code gets executed and initially checks whether inputted old password is correct or not, if it is correct system will update the new password or else error message displayed for the user saying "Old password is not correct".

6.5 RESULTS

Description: We build a model to predict the crop yield. System predicts yield for 2 different crops paddy and ragi. We use machine learning algorithm to predict yield with more accuracy and efficient results. We use training datasets to predict the crop yield. We divide the ratio 80:20 training datasets and testing datasets. For this ratio we got 96System took 2346 milli seconds for paddy yield prediction and 594 milli seconds for ragi yield prediction and precision factor for paddy is 96recall factor is 4

Paddy Crop Yield

Constraint	Naive Bayes Algorithm
Accuracy	96%
Time (milli secs)	2346
Correctly Classified (precision)	96%
InCorrectly Classified (Recall)	4%

Ragi Crop Yield

Constraint	Naive Bayes Algorithm
Accuracy	98%
Time (milli secs)	594
Correctly Classified (precision)	98%
InCorrectly Classified (Recall)	2%

Chapter 7

Conclusion and Future scope

7.1 Conclusion

India is a country where agriculture and agriculture related industries are the major source of living for the people. Agriculture is a major source of economy of the country. India is one of the countries which suffers from major natural calamities like drought or flood which damages the crop. This leads to huge financial loss for the farmers thus leading to the suicide. This proposed method is useful for society as it is a real time application for crop yield prediction. It helps farmers to make the right decision at the right time and helps them in preventing the losses.

7.2 Future enhancement

- In the proposed system we implemented machine learning algorithms for just 2 different crops such as "Paddy" and "Ragi", in the future the same concept can be improvised to predict yield for different crops.
- In the proposed work, we predicted crop yield for the region "Mysuru", in future we can enhance the work to predict crop yield for different regions.
- The proposed method only gives the yield of the crops, in future enhancement we can try to give the chemical combination that can produce better yield with the given land and weather conditions.

References

[1] Cunningham, Sally Jo, and Geoffrey Holmes. "Developing innovative applications in agriculture using data mining." The proceedings of the Southeast Asia regional computer confederation conference, 1999.

[2] D Ramesh, B Vishnu Vardhan, "Data Mining Techniques and Applications to Agricultural Yield Data", International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 9, September 2013.

[3] Monali Paul, Santosh K. Vishwakarma, Ashok Verma," Analysis of Soil Behaviour and Prediction of Crop Yield using Data Mining Approach", International Conference on Computational Intelligence and Communication Networks, DOI 10.1109/CICN.2015.156, IEEE 2015.

[4] Ami Mistry, Vinita Shah, Vallabh Vidyanagar, "Brief Survey of data mining Techniques Applied to applications of Agriculture", International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), vol. 5, pp. 301-304, Issue 2, February 2016.

[5] Yogesh Gandge, Sandhya, "A study on various data mining techniques for crop yield prediction", International conference on computer and optimization techniques (ICEECCOT), IEEE 2017.

[6] Md. Tahmid Shakoor, Karishma Rahman, Sumaiya Nasrin Rayta, Amitabha Chakrabarty "Agricultural production output prediction using Supervised Machine Learning techniques." Next Generation Computing Applications (NextComp), 2017 1st International Conference on. IEEE, 2017.

[7] Umid Kumar Dey, Abdullah Hasan Masud, Mohammed Nazim Uddin, "Rice Yield Prediction Model Using Data Mining", International Conference on Electrical, Computer and Communication Engineering (ECCE), pp. 321- 326, 2017.

[8] Kuljit Kaur, Kanwalpreet Singh Attwal, "Effect of temperature and rainfall on paddy yield using data mining", 7th International conference on cloud computing, data science engineering-confluence, IEEE 2017.

[9] Nidhi H Kulkarni, Dr. G N Srinivasan, Dr. B M Sagar, Dr. N K Cauvery, "Improving Crop Productivity Through a Crop Recommendation System Using Ensembling Technique", 3rd IEEE International Conference on Computational Systems and Technology for Sustainable Solutions, 2018.

[10] Waleed Khan, Nasru Minallah,Imran Ullah Khan, Zahid Wadud,Muhammad, Zeeshan,Suhail Yousaf and Abdul Baseer Qazi , "On the Performance of Temporal Stacking and Vegetation Indices for Detection and Estimation of Tobacco Crop", IEEE Access,8,103020 - 103033, 2020.

[11] Jie Sun, Zulong Lai, Liping Di,Ziheng Sun,Jianbin Tao,Yonglin Shen,"Multilevel Deep Learning Network for County-Level Corn Yield Estimation in the U.S",IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (Volume: 13), 2020.

[12] Neha Rale, Raxitkumar Solanki, Doina Bein, James Andro-Vasko, Wolfgang Bein. " Prediction of Crop Cultivation" , IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC),2019.

[13] S. Bhanumathi, M. Vineeth and N. Rohit, "Crop Yield Prediction and Efficient use of Fertilizers," 2019 International Conference on Communication and Signal Processing (ICCSP), pp. 0769- 0773,2019.

[14] D.Elavarasan P.M.D.Vincent " Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian Applications". IEEE Access, 8, 86886-86901.2020.

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

[16] Nageswararao Naik Bhookya, R. Malmathanraj and P. Palanisamy, "Yield Estimation of Chilli Crop using Image Processing Techniques", 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2020.

[17] Haowei Mu, Liang Zhou, Xuewei Dang, Bo Yuan, "Winter Wheat Yield Estimation from Multitemporal Remote Sensing Images based on Convolutional Neural Networks",10th International Workshop on the Analysis of Multitemporal Remote Sensing Images, 2019.

[18] Jayantrao Mohite, Suryakant Sawant, Mariappan Sakkan, Praveen Shivalli, Krishnaiah Kodimela, Srinivasu Pappula, "Spatialization of rice crop yield using

Sentinel-1 SAR and Oryza Crop Growth Simulation Model", 8th International Conference on Agro-Geoinformatics, 2019.

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

[20] Porandla Srinivas, P Santhuja, "Utilization of Data Mining Methods to Investigate Crop Yield Forecast", International Conference on Emerging Trends in Science and Engineering (ICESE), 2019.

[21] Anil Suat Terliksiz, D. Turgay Altılar, "Use Of Deep Neural Networks For Crop Yield Prediction: A Case Study Of Soybean Yield in Lauderdale County, Alabama, USA", 8th International Conference on Agro-Geoinformatics (Agro-Geoinformatics), 2019.

[22] Hulya Yalcin, "An Approximation for a Relative Crop Yield Estimate from Field Images Using Deep Learning", 8th International Conference on Agro-Geoinformatics, 2019.

[23] Debra P. C. Peters, Heather M. Savoy, Geovany A. Ramírez, Haitao Huang," AI Recommender System With ML for Agricultural Research",IEEE , 2020

[24] Haowei Mu, Liang Zhou, Xuewei Dang, Bo Yuan, "Winter Wheat Yield Estimation from Multitemporal Remote Sensing Images based on Convolutional Neural Networks", 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images,2019.

[25] R N Bhimanpallewar, M R Narasingarao,"Alternative approaches of Machine Learning for Agriculture Advisory System",10th International Conference on Cloud Computing, Data Science Engineering (Confluence)", 2020.