Intelligent Agricultural Information Monitoring Using Data Mining Techniques

Md. Tahmid Shakoor	**13101046**
Karishma Rahman	**13101284**
Sumaiya Nasrin Rayta	**13141004**

Supervisor: Dr. Amitabha Chakrabarty

Department of Computer Science and Engineering

School of Engineering and Computer Science

BRAC University

Submitted On: 18th April, 2017

DECLARATION

Thesis Submission to the Department of Computer Science and Engineering, BRAC University, Dhaka, submitted by the authors for the purpose of obtaining the degree of Bachelor of Science in Computer Science, and the degree of Bachelor of Engineering in Computer Science and Engineering. We hereby announce that the results of this thesis are entirely based on our research. Resources taken from any research conducted by other researchers are mentioned through reference. This thesis either in whole or in part, has not been previously submitted for any degree.

Signature of Supervisor:

____

Dr. Amitabha Chakrabarty
Assistant Professor
Department of Computer Science and Engineering
BRAC University

Signature of Authors:

1. __________________________
Md. Tahmid Shakoor

2. __________________________
Karishma Rahman

3. __________________________
Sumaiya Nssrin Rayta

ABSTRACT

Farmers usually select crops for cultivation based on their previous experiences, the better the profit margin of a crop achieved in the past, probability of choosing that crop increases. However, the lack of information about scientific factors that can affect the output and precise knowledge about cultivation, they end up cultivating crops that do not meet the desired profit margin. To help the farmers take decisions that can make their farming more efficient and profitable, this research tries to establish an intelligent information prediction analysis on farming in Bangladesh. Also, it provides an interface to this analysis for the farmers through an android app which also provides necessary information on cultivation procedure, irrigation and fertilization process. The research suggests area based beneficial crop rank before the cultivation process. It indicates the crops that are cost effective for cultivation for a particular area of land. To achieve these results, we are considering six major crops which are Aus, Aman, Boro rice, Potato, Jute and Wheat. The prediction is based on analyzing a static set of data using Supervised Machine Learning techniques. This static dataset contains previous years' data taken from the Yearbook of Agricultural Statistics and Bangladesh Agricultural Research Council of those crops according to the area. The research intents to do a comparative analysis on Decision Tree Learning, K-Nearest Neighbors and Multiple Linear Regression algorithms to obtain these predictions. The past ten years (2004-2013) of Bangladesh have been considered making this dataset to ensure learning and training of the algorithms and increasing the accuracy rate of the prediction and for testing we used three years (2014-2015) for computing accuracy.

Index Terms: Data Mining, Supervised Machine Learning, Iterative Dichotomiser 3, k-Nearest Neighbour Regression, Multiple Linear Regression, Decision Tree Learning, agriculture, prediction, comparative analysis, dependent variable, independent variable, android application, learning dataset, test dataset, error percentage, prediction accuracy.

ACKNOWLEDGEMENT

We came up with the idea behind this thesis project with the guidance of Dr. Amitabha Chakrabarty, Assistant Professor of the School of Computer Science and Engineering of BRAC University. He encouraged us to contribute to our country's agricultural growth by introducing state of the art technology in this sector. After some research, the three of us- Tahmid Shakoor, Sumaiya Nasrin and Karishma Rahman decided to move ahead with this idea as our thesis project. Every member of this group had one objective in mind- building a fully operational prototype. All the members have shown remarkable dedication to this project throughout the past year and now we are finally presenting our completed work.

We are grateful to our supervisor, Dr. Chakrabarty for his confidence in us and for sharing his valuable insights, suggestions and guidance for the completion of this thesis work. This project would have never been completed successfully without his constant supervision and critical evaluation. He urged us to contribute to our highest potentials. We are honored to work under his supervision.

We are also thankful to Talha Siddique, Dipro Barua and Zannatul Ferdous who helped us in our tough times in this project. Moreover, the support we got from our families are immeasurable. Throughout the whole process our families have supported us both mentally and financially. Our work would have never been completed without their persistent support and encouragement.

We are thankful to Agriculture Information Service (AIS) for sharing the data necessary for the completion of our research. Especially, we would like to thank Mohammed Zakir Hasnat, Information Officer, Plant Protection of AIS for sharing his valuable time to guide us in understating the project scope. We appreciate him for helping us to collect the necessary data from Bangladesh Agricultural Research Council and Bangladesh Bureau of Statistics.

Lastly, we would like to express our sincere gratitude to all the faculty members of Computer Science and Engineering department for enriching our minds with all the knowledge throughout our undergraduate life with their outstanding wisdom which undoubtedly aided us to execute this thesis. We dedicate our work to them for their contribution in shaping our mind.

Table of Contents

Table of Figures

Chapter 1

INTRODUCTION

In this chapter, we will discuss about the functionality and features of the Farmer's Guide app, which implements data mining techniques for intelligent agricultural information monitoring. It gives a detailed overview of the system and how the implemented techniques were used to obtain the solution.

1.1 Introduction

In this era of technology, applying scientific methods and automated learning in problem solving has come a long way from being a trend to becoming a necessity. This realization has made the first world countries more advanced in fields like agriculture, medicine, education etc. The use of technology resulted in a booming economy [1] and made them richer each day, whereas developing countries like Bangladesh, lags behind for turning a blind eye to it by giving the excuse of the process being too costly. Though agriculture is a crucial part of our economy, due to the lack of use of technology and scientific methods to farming, farmers in most cases do not get the preferred output. Success in agricultural growth, it is believed, is based on an ecologically adapted and economically viable agricultural technology which involves a continuous adaptation to available resources [1]. The objective of this research is to provide a learning agent that can aid in taking decisions to make the farming more efficient and profitable through technology. The output of this research, an app called Farmer's Guide, provides a list of profitable crops in a particular area using decision making algorithms [2]. This prediction will be made using the Decision Tree Learning, ID3 (Iterative Dichotomiser 3) algorithm, which will analyze the production output of different crops in previous years. Also, a comparative analysis with two major regression algorithms in data mining- k-Nearest Neighbor (KNNR) and Multiple Linear Regression (MLR) was done to see the difference in error percentages of the predictions which is discussed later in this report. The goal of this research is to help the farmers maximize their profit margin by providing predictions on crops that will give the maximum output in a particular area. The research only focuses on the six major crops for ten region, Dhaka, Tangail, Barishal, Jessore, Sylhet, Mymenshing, Bogra, Patuakhali, Faridpur and Comilla, stored in database system. The dataset contains of details on crops' yield per hectare (M.Ton), average of minimum and maximum temperature, rainfall, year range, and region. Analyzing these data, the algorithm gives the result which predicts the most profitable crop in a particular region at a point in time. For the accuracy of the prediction, data of past twelve years is being used by the algorithms for the purpose of learning and result analysis. The app also provides basic instructions on the cultivation process, irrigation, fertilization and harvesting for each crop cited by Department of Agricultural Extension, Bangladesh which will facilitate the farmers in using scientific farming methodologies and reaching the preferred output mark.

1.2 Motivation

Since Bangladesh as a 3rd world country is still in the early demographic transition stage, the emphasis on the use of technology in production has been seen only in recent years. The digitalization of different sectors that have economic value has now become a necessity. With an intention of ruling out the traditional farming methodologies and introduce the use of scientific farming techniques, and also to contribute to the digitalization process of the agricultural industry, our team has worked day and night to develop this automated system. Our goal is to provide a model through this app that can provide the farmers with accurate agricultural production output predictions that can facilitate their crop selection decisions, alongside the instructions on cultivating these crops in a scientific way to ensure maximum production. In a country where three fifths of the population earns their bread through agriculture [4], the use of traditional farming techniques have affected the production of different crops by not being aware of the effect of natural factors like rain, temperature etc in farming, whereas the developed countries have effective weather extensions that can provide farmers with accurate climate predictions through which they can decide on the crops that are most profitable in that environment [5]. The objective of our research was to draw a pattern between these variables and the production output of six major crops in Bangladesh to identify the most profitable crop in a given environment. As an output of this research, we are providing the farmers with an android application with an easy to user interface that can help the farmers in choosing the most beneficial crop given the effects of the mentioned variables as well as proper farming instructions which will result in maximum profit. This will not only increase our total production of agriculture but also result in economic growth [4].

1.3 Thesis Contribution

The objective of this thesis was to eliminate the limitations farmers face when making crop selection decisions and to show them the efficient way of farming through the instructions provided. The Farmer's Guide application contributes to Bangladesh's agriculture sector by being one of the best automated prediction based system in terms of the error percentage of the predictions obtained, which incorporates data mining and android application technology to achieve these results. Alongside that, the result of the comparative study that was done in this thesis provides an insight on the performance of Multiple Variable Linear Regression, k-Nearest Neighbour Regression and Iterative Dichotomiser 3 algorithms which can be proven helpful to other researchers who are working to solve similar real life problems using machine learning [2] techniques. The industry standards were maintained in every step of the design process of the system. The main challenge was to develop a system that can easily suit the network infrastructure available in rural areas. Also, the technical knowledge of the farmers and the use of smartphones in rural areas were a concern. These problems, and the necessity and applicability of the system in areas where technology like smartphones and internet were just recently introduced, were identified and dealt with during the technological and economical feasibility study of the system. The organizational feasibility study was left out as the target market of this app is only the farmers. Then our group discussed with different officers from the Department of Agricultural Extension Bangladesh to gather knowledge about the needs of the farmers to identify the functional and non-functional requirements of the system. Ease of use of the application was the main concern during the development phase of the system and thus, effective offline use of the app was the objective rather than the online use as most of the farmers do not have the technical knowledge nor do they have the economic condition to operate the app if the mobile data was made necessary. As a result of this analysis, we were able to develop a fully functional system that addresses the issues mentioned and works around those issues to be effective in its use. The functionalities provided by this app facilitate agricultural problem solving in a manner that is absent in the farming methodologies followed by any developing country like Bangladesh. We consider it as an achievement not only for us but the whole country in digitalizing the agriculture sector.

1.4 Problem Statement

In this research, nearly thousand data instances containing values like crop name, year, temperature, rainfall and production rate per hectare have been analysed with the help of the MLR, KNNR and ID3 algorithms to obtain the predictions. Collecting this huge amount of data was a hassle due to a lack of digital approach in recording these information in the local level of the Bureau of Statistics of Bangladesh. Most of these regional data provided to us by the Agricultural Information Service were in hard copy and were very much scattered. These data were then converted to digital copies and were analyzed manually to identify the outliers to keep track of the cause behind the irregularities in the predictions. Extensive research was done to gain basic understanding of the algorithms which turned out to be very difficult for us due to the lack of any prior knowledge on these subject matters. As a result, we had to go through significant number of trial and error cycle during the development phase of the application. Some major changes had to be made in default accessing methodology used to fetch the data from the cloud to ensure effective offline use.

1.5 Solutions

From the comparative study, it was found that Iterative Dechotomiser 3 [11] algorithm provides the best prediction than MLR [6] and KNNR [7]. Without omitting any outlier from our dataset we achieved better accuracy with less error percentage in prediction from ID3. This prediction procedure was used to generate the crop rank for each region and suggest profitable crops for a specific area to the user.

1.6 Methodology

Many researchers in the past have used algorithms like KNNR, MLR and ID3 to implement pattern recognition and predict on results like share prices, property value, land capability and other agricultural applications [3] similar to our research. The resultant Farmer's Guide app of this thesis incorporates data mining, machine learning and android application technology to predict on the crop yield rate of a given year by implementing these three above mentioned algorithms and analysing past ten years average (min + max) temperature and average rainfall throughout the cropping period of major crops to define a classification model for the data set. Then, the current temperature and rainfall values can be tested against this model to obtain the prediction. The system takes the time period in which the farmer wants to start cultivating a crop as user input. Taking account the farmer's location, the system then decides on the crops that can be cultivated during that time period and provides a rank of crops that will be profitable based on the yield rate predictions obtained in that particular region at that given time. The application provides an easy sign up and sign in feature which facilitates in storing previously selected crop and the cost of cultivating that crop information for each user. Alongside that, the app provides scientific cultivation instructions for each of the crop through cultivation, fertilization and irrigation suggestion feature as well as the schedule for those activities as crop calendar. The cost calculator feature can help the farmer to keep track of his expenses and in making crop selection decisions by providing previous cultivation cost history. For efficient access to these information, the database was designed to be distributed between the local and cloud server.

1.7 Summary of Contributions

The prototype that was developed during this thesis work introduces efficient scientific farming techniques in Bangladesh with the help of computational technology. It provides a methodology to increase the agricultural production which works at the intersection of statistics, data mining and artificial intelligence. Upon more investigation on this research and variables that have major influence on economic outcome in various fields, this model can be replicated to predict the target result in fields like automated disease identification, industrial production output prediction, predicting fluctuation in currency value etc. Also, incorporating more predictors in this prototype will result in more accurate prediction as well expansion of the functionality of the system to predict factors land capabilities, fertilization etc. which will result in increase the mass agricultural production.

1.8 Goals

Our main objective was to implement a machine learning algorithm that gives better prediction of suitable crop for the corresponding region and crop season in our country. We have used average rainfall and average (max + min) temperature throughout the cropping period as dependent variables and yield rate per unit area of each year as independent variable to implement the ID3 algorithm [11]. This algorithm has successfully provided the approximate prediction values for the crops. Another objective was to introduce our country's agriculture sector to technological revolution with an automated farming process and make our farming sector more technologically progressive. A prototype of the system was built which can be helpful to serve the farmers of our country. It was also our intention to contribute to the economic growth of Bangladesh by proposing this model which can help increase the mass agricultural production which will in turn affect the GDP growth positively [4]. In the future, the successful implementation of this machine learning algorithm [11] can be done by following this model to predict other factors than crop yield rate, such as- prediction for diseases, natural disasters by analyzing weather forecast and so on. As result, the sector of crop farming using machine learning technology will open more possible research opportunities. The mobile application platform is revolutionizing the society with social media apps which also brings attention to utility apps that has functionalities to solve real world problems. Thus, if the app can be shown as capable enough to assist in agricultural decision making, it will encourage the other 3rd world countries to follow our path and invest in the idea of reshaping the traditional farming techniques with the help of technology. With ample assistance from our government, we can flourish the possibilities of being the light bearer in this technological revolution.

1.9 Thesis Outline

Chapter 1 introduces the reader to our research idea, the mechanism behind the implementation, and our objectives and contribution.

Chapter 2 discusses background study of our project, the algorithms used for the analysis, key terms that are useful to understand the methodologies behind these algorithms and existing works that implemented similar ideas.

Chapter 3 talks about the system structure, functionalities and the task flow within the system.

Chapter 4 discusses the results obtained from the comparative study and analyzes the accuracy of the algorithms in terms of the error percentage of the predictions.

Chapter 5 discusses the functional and non-functional requirements of the system and investigates whether the system is feasible.

Chapter 6 draws conclusion from the research and talks about improvements that can be made in the future.

Chapter 2

LITERATURE REVIEW

In this chapter, a detailed explanation of the mechanism behind the algorithms used in this research as well as a brief overview of the existing works that used these algorithms is given. Many researches are reaching out the field of automated farming. However, Bangladesh is new to this area of research and the datasets are not well structured which makes it difficult to predict the precise result. In India, Vinciya and A. Valarmathi analysed the classification between organic, inorganic and real estate data to make a prediction [3]. In this research for extracting useful information and to predict they used data mining technology, for the selected region they used multiple linear regression [3]. Decision Tree algorithm was used for prediction which is supervised learning algorithm and multiple linear regression which is generalized prediction model [3]. To predict the structured object instead of discrete or real values they used Decision Tree algorithm [3]. The expected loss was calculated from training data from classification [3]. To find the difference between actual values and the fitted values there was an equation for prediction and the estimation of the residuals in multiple linear regressions. Its use was to calculate the mean-squared error [3].An attempt has been made to research the influence made by decision tree induction technique of climatic parameters on soybean productivity [8]. For better understanding by the end users the findings of Decision tree were framed into different rules [8]. In predicting/forecasting the crop yield in advance for market dynamics, the study findings will help the researchers, policymakers and farmers [8]. The decision tree is one of the most popular classification algorithms currently in machine learning and data mining [9]. Land capability classification using data of 38 soil series of Wardha district, Maharashtra was evaluated by Iterating Dichotomizer 3 (ID3) - A classical decision tree (DT) algorithm [9]. For land capability classification soil depth, slope, drainage, texture, erosion, and permeability were selected as attributes [9]. Ashwani Kumar and her co-author used a new algorithm named “Agro Algorithm” in Hadoop platform to predict the crop yield and the best crop suggestion and used Hadoop framework for handling large data [10]. They considered the soil type and weather [10]. Firstly, from the Hadoop distributed file system, the data is accessed and traversed through line by line to perform the normalization among the data [10]. After that, classification is performed which gives the prediction about soil and crop from classification on the basis of disease. [10]. Using of Artificial Neural Network (ANN) for crop yield prediction was discussed by Snehal and Sandeep, where they used feed forward back propagation neural network [12]. The pattern is identified by feed forward and for comparison with the actual and gained output, back propagation is used. Random weight was chosen in the beginning of train data in order to minimize the error rate, [12]. For accurate prediction they used the number of parameters like PH, Nitrogen, Temperature, and Rainfall [12]. The authors tried to discuss about the various data mining approach in agricultural field in Data Mining in Agriculture on Crop Price Prediction: Techniques and Applications [13]. K-Means, K-Nearest Neighbor, Artificial Neural Networks, Support Vector Machines and the necessity of Price Prediction of crops according to the current market price policy was discussed in here [13]. So according to our countries networking management and technology the stated methods either used too many variables or complex method to give the prediction which is sometimes inefficient and error prone. The paper, Application of Classification Technique in Data Mining for Agricultural Land analyses 10 cities of India, their data of arable land area, rural labor and the gross output value of agriculture based on the decision tree[14]. To discrete data during the process of data mining compared to the traditional classification methods they also implement clustering analysis method [14]. At last, from the results of the classification they carry out generalization conceptual process [14].

The following discussion focuses on the techniques used to obtain agricultural output prediction in this study.

2.1 Data Mining

A new discipline that is defining the methodologies being used recently in the field of statistics, database technology, pattern recognition, machine learning, and other areas is called Data Mining [15]. It comprises methods from the field of artificial intelligence, machine learning, statistics and database system that implements pattern recognition in large dataset [16]. The objective of data mining is to transform bulk amount of data into an understandable structure which can be used in different analysis [16].

2.2 Supervised Machine Learning

From a specified set of predictors (independent variables), target variable (dependent variable) are predicted by these forms of algorithms [2]. We generate a function that maps inputs to preferred outputs using these set of variables. The training process remains till the model attains an anticipated level of precision on the training data.

2.2 Multiple Linear Regression (MLR)

Simple linear regression is a supervised machine learning technique [17] that evaluates the relationship between a dependent variable and an independent variable given a dataset containing observations on these variables for a particular population sample [18]. Multiple linear regression differs in the fact that it evaluates the relationship between a dependent variable and two or more independent variables [18].

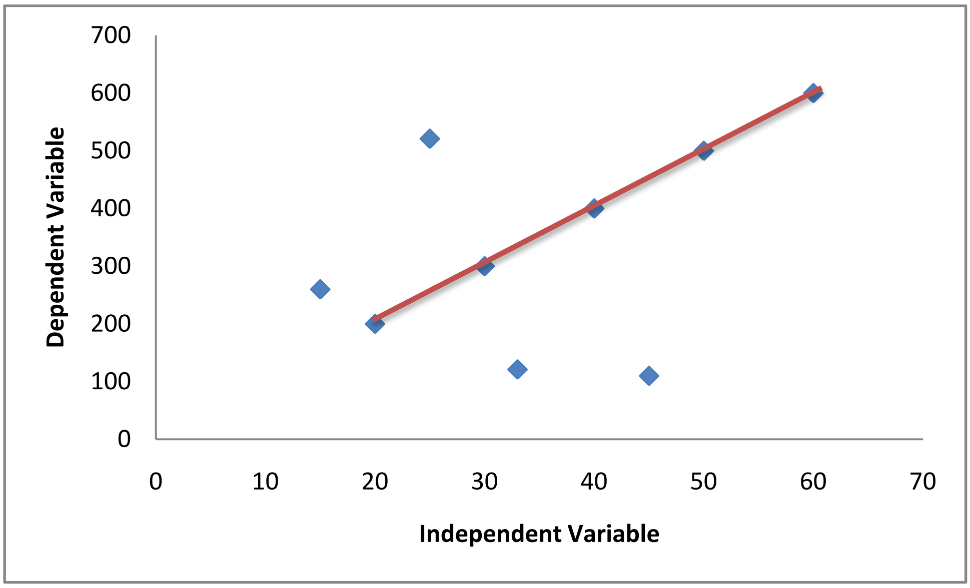

Figure 2.2: Regression Model

In the given regression model graph, a best fit line is drawn to establish a relationship between the independent and dependent variables. The line is obtained from the following equation [18],

$$y = \beta_0 + \beta_1 * x$$

Where,

y = Dependent variable,

x = Independent variable,

β_0 = Intercept,

β_1 = Slope of the line.

Since, multiple linear regression has multiple independent variables, the equation to obtain the best fit line is following [18],

$$y = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \cdots + \beta_n * x_n$$

Where,

y = Dependent variable, x_i = Independent variables, β_0 = constant term, β_i = coefficients relating y and x_i. It is assumed that error term has a mean value of 0 [18].

2.3 K Nearest Neighbors Regression (KNNR)

In *k*-Nearest Neighbors Regression algorithm, a similarity measure between cases in the learning data set and the test data set is done to identify the classification of the test cases. For instance, Euclidean distances of all cases can be measured and then the k nearest learned data instances votes to identify the class in which the test cases belongs [19]. It finds the adjacent possible values based on the number of K which is optimal.

Euclidean Distance Formula:

$$d[i] = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2} = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

Where, x and y are the two predictors in the model.

2.4 Decision Tree Learning

Data mining techniques are concerned with pattern extraction and classification of huge and uncertain data [11]. Decision tree algorithm is a frequently used technique in data mining to classify large amount of data and extract dataset that has similar patterns [11]. Decision tree represents classification model for a data set in the form of a tree structure, breaking down the dataset into smaller subsets. The root node is chosen based on the information gain of attributes in a dataset where an entropy calculation mechanism is applied to obtain the gain information.

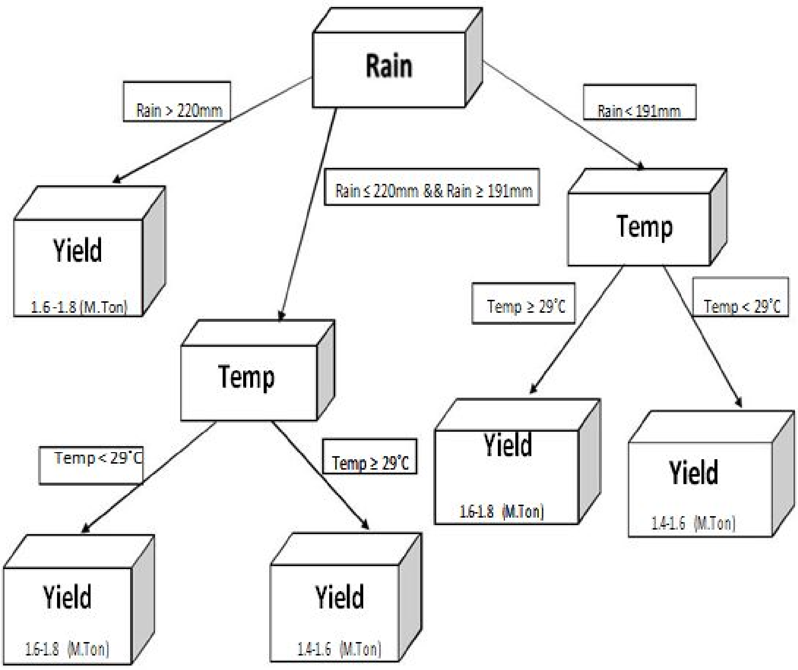

Figure 2.4: A decision tree model obtained from the comparative analysis

2.5 Iterative Dichotomiser 3

Among all the decision tree algorithms, Iterative Dichotomiser 3 or ID3 algorithm proposed by QUINLAN in 1986 [11] is used in this research to classify the dataset. The ID3 algorithm analyzes information gain to determine the attributes that classify the data set effectively [11]. The entropy, which represents the number of uncertain data present in the database [11], is calculated using the following equation,

$$Entropy(S) = -\sum p(X)\, log_2\big(p(X)\big) \ldots \ldots \ldots (1)$$

where S represents the dataset for which the entropy is calculated and the set of classes in S are represented by X [11]. In (1), p(X) is the ratio of the number of components in class X to the number of components in the set S [11]. The following equation-

$$IG(S) = I(S) - \sum p(t) * I(t) \ldots \ldots \ldots (2)$$

calculates the information gain which represents the amount of uncertainty reduced after splitting the dataset over an attribute Y [11]. In (2), I(S) is the entropy of the dataset and I(t) is the entropy of the subset t, that has been created by splitting S over attribute Y and p(t) is the probability of members in a class to the total number of members of set S[11]. The attribute that has the best information gain is chosen as the root of the decision tree and the algorithm continues to build the tree in a recursive process by creating sub-trees which also have the attribute with best information gain as root.

2.6 Pseudo code for Multiple Variable Linear Regression

```
Regression ( double [][] x, double [] y){
  Matrix x ← new Matrix (x);
 //create matrix from vector
  Matrix y ← new Matrix (y, N);

  //find the least squared solution
  QRDecomposition qr = new QRDecomposition (x);

  beta ← qrsolve(y);

  //mean of y[] values

    For ( i = 0 to N){

         Sum += y[i];

         Mean ← Sum / N;

    }

    //total variation to be accounted for
    For ( i = 0 to N){

         Dev ← y[i] – mean;

         Sst += Dev * Dev;

     }

// variation accounted for
        Matrix residuals = x.times(beta).minus(y);
        SSE = residuals.norm2() * residuals.norm2();
}
```

2.7 Pseudo code for *k*-Nearest Neighbor Regression

```
// calculating the Euclidean Distance
For ( j=0 to Yield .length) {

D[j]=distance (rainfall, temp, rainfall[i], temp[i])

}

//Finding the indices of ‘k’ no. mean Euclidean
Distances

FindMinIndex (k, d, minIndex);

//calculating the predicted yield

Yield ← calculateAverage(y, minIndex)
Error ← percentageErrors (y1, yield)  //Calculate percentage if error
```

2.8 Pseudo code for Iterative Dichotomiser 3

```
ID3 (Examples, Target_Attribute[Yield_Rate], Attributes[avg_Temp,avg_Rain])
  Create a root node for the tree;
  If number of predicting attributes is empty, then Return the root;
Otherwise do
    Y= Choose the best attribute with highest information gain(Gain[Attributes]);
    Root = Y.
    For each possible value, vi, of Y,
      Add a new tree branch below Root, corresponding to the test Y = vi;
      Let Examples(vi) be the subset of examples that have the value vi for Y;
      if Examples(vi) is empty
        Then below this new branch add a leaf node with label = most common target value in the examples; //yield_rate
      Else below this new branch add the subtree ID3 (Examples(vi), Target_Attribute[Yield_Rate], (Attributes[avg_Temp, avg_Rain] – Y));
  Return Root
```

Chapter 3

ALGORITHM AND SYSTEM ARCHITECTURE

In this chapter, we will analyse the Farmer's Guide system in terms of its structure, functionality and workflow. Alongside that, we will look at how the prediction algorithm was chosen from the comparative study and how it contributes to the core functionalities of the system. This chapter follows the system analysis and design rulebook by using use case, flow and class diagram to depict a picture of the system structure and workflow.

3.1 Sample Area Definition

The dataset being analysed by this system consists of monthly average (max+min) temperature, rainfall [20] and yearly production rate of six major crops [21] of the ten major regions in Bangladesh. Due to the unavailability of complete and reliable dataset consisting of these attributes for different regions, only the following 10 areas in Bangladesh were considered for this analysis,

1. Dhaka,
2. Jessore,
3. Tangail,
4. Barishal,
5. Sylhet,
6. Mymenshing,
7. Bogra,
8. Patuakhali,
9. Faridpur and
10. Comilla.

The dataset obtained was divided into learning sets and testing sets to identify the error percentage of the predictions. The algorithms rely on the learning set examples for analysing test cases and generate yield rate predictions. The predictions are then evaluated against the true yield rate using relative error methodology to determine the accuracy of each algorithm.

3.2 Crop Selection

As Bangladesh is an agrarian country [21], a large number of crops are cultivated all throughout the region every year. Thus, it raises the need of a statistical analysis on the total agricultural production and at present this analysis is being conducted by Bangladesh Bureau of Statistics which identifies the following six crops as the major GDP earner,

1. Aus,
2. Aman,
3. Boro,
4. Wheat,
5. Jute and
6. Potato.

Considering the dominant characteristics of them in the country's economy, and also due to the availability of the complete and reliable dataset for these crops, these six major ones were considered for this analysis.

3.3 Defining Predictors

The factors that affect selected crops' production the most are temperature, rainfall, land capability, soil composition etc. Though the analysis provides a more accurate prediction when the dataset being analysed contains all these attributes, due to the lack of reliable and consistent records of these information, we had to work only with temperature and rainfall to generate predictions. Using these attributes, we often achieved results with almost a 100% accuracy rate which shows the possibilities this app have when other factors can be brought into the analysis. The model proposed in this research has a lot of potential to predict exact output if more predictors that have an effect on the crop yield rate can be incorporated. For example, during data collection, we obtained information on data attributes like relative humidity, sunshine brightness [20], soil composition etc. However, no government or private organization has any reliable data collection methodology or data storage for these information. These attributes can have great impact on the accuracy of this model but due to the mentioned reason, we were discouraged to use them in our analysis.

3.4 System Work Flow

The following diagram describes the flow of tasks within the system and also depicts the picture of the whole system structure,

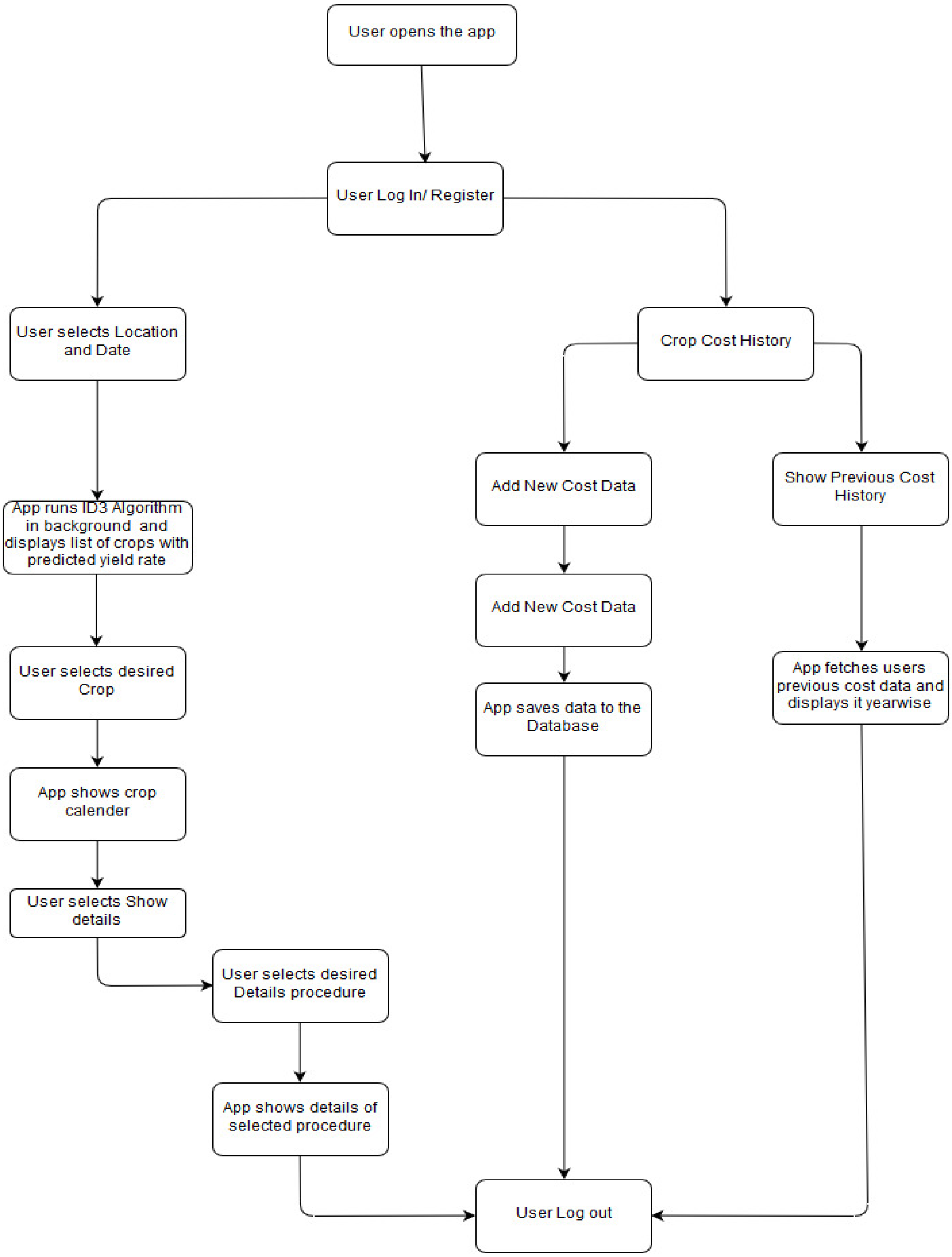

Figure 3.4: Flow Diagram of Farmer's Guide App

A new user will follow the task sequence given below,

1. User opens the app,
2. User has to register/log in to access the app,
3. User selects the location and date,
4. App will run ID3 Algorithm with the given location and date and suggest the crops with their predicted yield rate,
5. User will choose one crop from the suggested ones,
6. App will show the calendar output for the selected crop in Bangla,
7. User clicks details to see the detailed procedures for cropping,
8. On selecting each procedure, details of that procedure will be displayed,
9. User can also click on Crop Cost History after login,
10. App gives two options, Add new cost data and show history, after login,
11. Add new data takes all cost data, calculates profit/loss and saves it to the database,
12. Show previous history will fetch the cost data for the user and display it year wise,
13. User logs out of the app.

3.5 Use Case Diagram

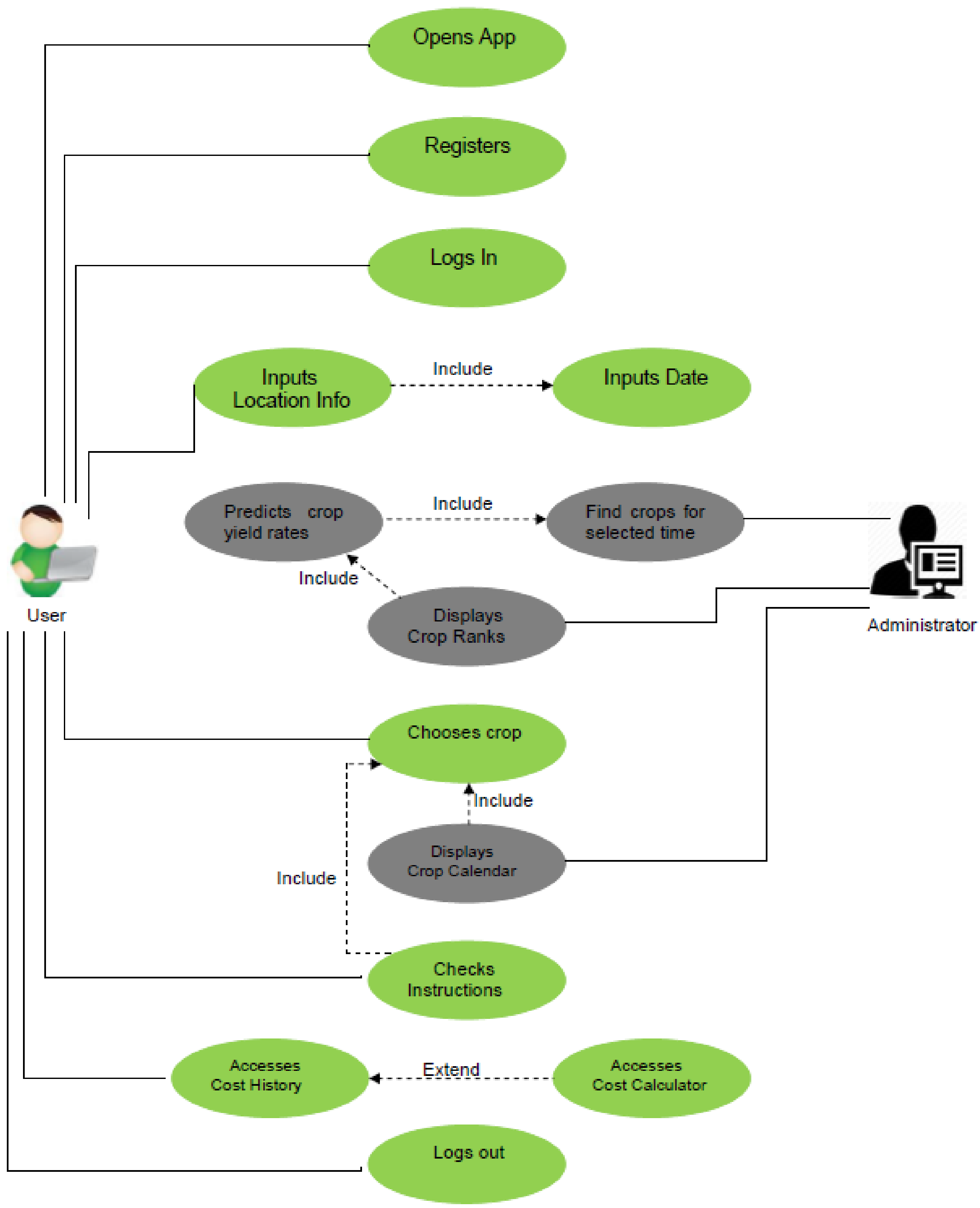

Figure 3.5: Use Case Diagram of Farmer's Guide System

Use Case Name: Farmer’s Guide Prediction Application

Primary Actor: User

Secondary Actor: Administrator

Brief Description: The use case describes how a user operates the farmer’s guide application and how it generates prediction and helps user with instructions.

Relationships:

- Include: “Inputs Location Info” includes “Inputs Date”, “Predicts crop yield rates” includes “Find crops for selected time”, “Display crop ranks” includes “Predicts crop yield rate”, “Display crop calendar” and “Checks instructions” include “Chooses crop”.
- Extend: “Accesses Cost Calculator” extends “Accesses Cost History”

Entry Condition: The user has registered and inserted valid email address and password in the sign in page.

Exit condition: The user has logged out.

Normal Flow of Events:

1. User opens the app.
2. User registers using email address and password.
3. User logs in using the registered email address and password.
4. User inputs his location and probable date from which he wants to start cultivating.
5. App finds the appropriate crops for that region and time period and generates yield rate predictions for each of those crop.
6. App displays a rank of crops based on the predictions obtained.
7. User chooses a crop.
8. App displays crop calendar for the chosen crop.
9. User checks instructions on cultivation/irrigation/fertilization.
10. User accesses cost history.
11. User logs out.

Sub Flows:

Adding Cost Information

1. User calculates cost using the provided cost calculator.
2. The calculated results stored in history.

Exceptional Case:

1. No crop available for the given time period.
2. Database connection issue when retrieving history from cloud.

3.6 Class Diagram

The class diagram representation of the Farmer's Guide app looks like the following,

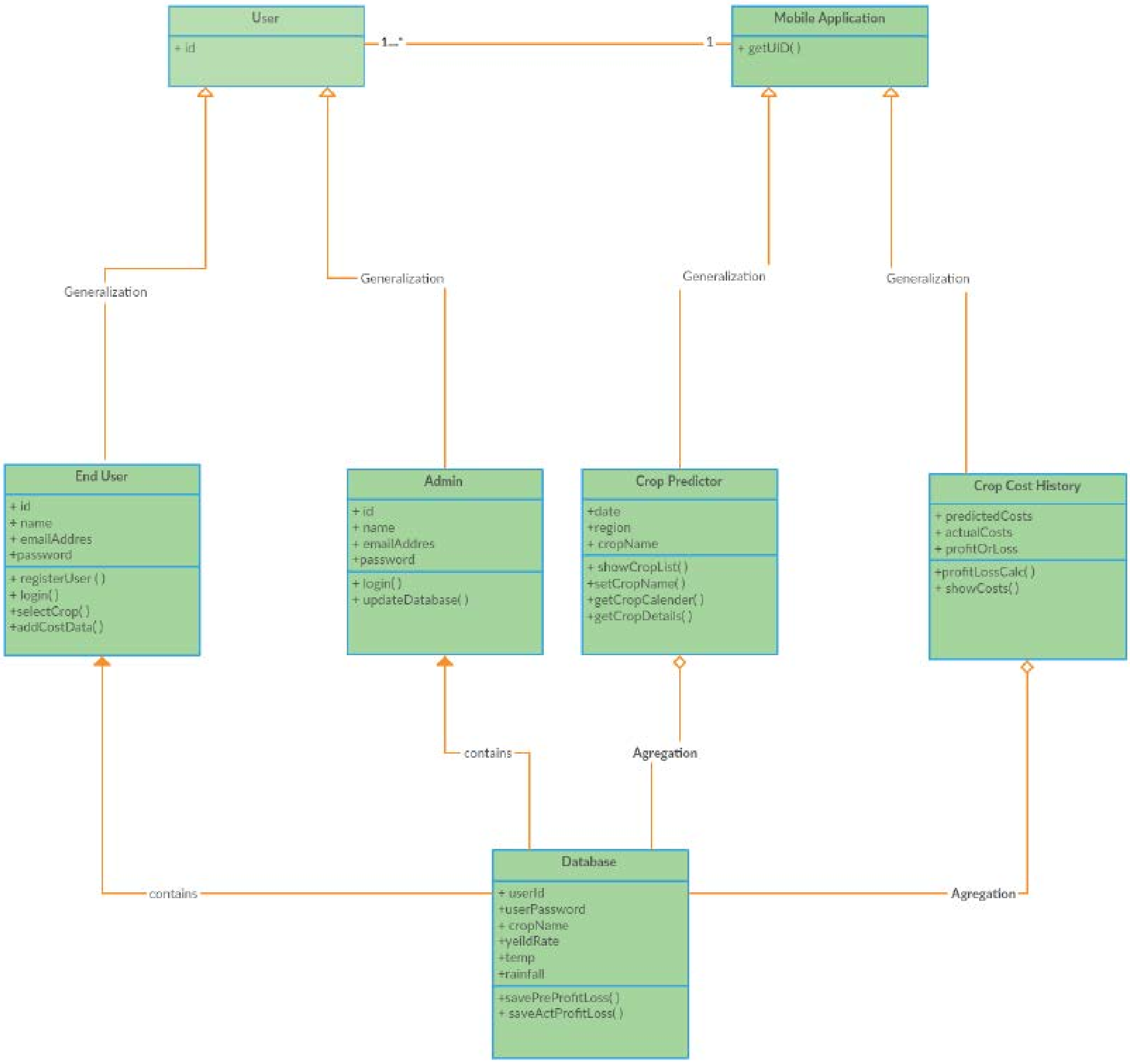

Figure 3.6: Class Diagram of Farmer's Guide App

A class diagram is static structure diagram which illustrates the structure of a system by presenting the system's classes, their attributes, operations or methods, and the relationships among objects. The classes in this system are identified as:

A. User
B. Mobile App
C. Database

A. User Class:

This class includes all the users in general. It has an attribute - id. User class has two User subclasses which has 2 specializations –

1. End User
2. Admin

1. End User:

This subclass handles all the users that create an account to run this app and gets benefit from it. The attributes of this subclass are:

- id – The number that describes a generalized serial of the users.
- name – The username that users will use while creating a new account.
- emailAddress– the email address the user use to register to the app.
- password – The password a user will create for his/her account.

The attributes are handled by few methods. The methods which are used to create the End User subclass are:

- registerUser() – The function that registers the user to the app to authenticate.
- login() – The function gets the user logging in to the app.
- selectCrop()– The function that gets the selected crop by the user'.
- addCostData()– The function that adds the new cost data entered by the user.

2. Admin:

This subclass handles all the administrative body people who have access to the database and can update any information anytime. The attributes of this subclass are:

- id – The number that describes a generalized serial of the users.
- name – The username that users will use while creating a new account.

- emailAddress– the email address the user use to register to the app.
- password – The password a user will create for his/her account.

The attributes are handled by few methods. The methods which are used to create the Admin subclass are:

- login() – The function gets the user logging in to the app.
- updateDatabase() – The function which allows the admins to make any modifications to the database.

B. Mobile Application:

This class mainly portrays the app's front-end. It has a method-

- getUID() – The function gets ID if the current user logged in to the application.

Mobile Application class has two User class has 2 specializations –

1. Crop Predictor
2. Crop Cost Calculator

1. Crop Predictor:

This subclass main portrays the prediction part of user interface. The attributes of this subclass are:

- date – The date user entered to cultivate the crops.
- region – The region that user intend to cultivate the crop.
- cropName– The crop user selects finally to cultivate.

The attributes are handled by few methods. The methods which are used to create the Crop Predictor subclass are:

- showCropList() – The function that shows the list of crops according to user's information given.
- setCropName() – The function that saves the name of the crop once the user selects it.
- getCropCalender()– The function that gets calendar for the selected crop.
- getCropDetails()– The function that gets the details procedure for the selected crop.

2. Crop Cost History:

This subclass main portrays the Crop cost and saving history part of user interface. The attributes of this subclass are:

- predictedCosts– The user enters the cost values he/she estimated.

- actualCosts – The user enters the cost values he/she actually spent throughout the cultivation process.
- profitOrLoss– The value for profit and loss depending on the sales.

The attributes are handled by few methods. The methods which are used to create the Crop Cost History subclass are:

- profitLossCalc() – The function that calculates the profit or loss for the crops.
- showCosts() – The function that displays the previous history of the crop.

C. Database:

This class symbolizes the backend data support system of our application. The attributes of this class are:

- userId- The unique id generated for each user.
- userPassword- The password user entered while registering.
- cropName – Name of different crop types.
- yieldRate – Yield rate per year for different regions of different crop types.
- aveRainfall – Average rainfall of the cropping time-period months for any chosen crop.
- aveTemperature – Average temperature of the cropping time-period months for any chosen crop.
- region – The chosen 10 regions.
- month- The monthly timing each crop needs to be cultivated.
- year – The learning data and testing data, total of 15 years.

The attributes are handled by few methods. The methods which are used to create the Database class are:

- savePreProfitLoss() – The function that stores all the predicted profit loss data into the data support system.
- saveActProfitLoss() – The function that stores all the actual profit loss data into the data support system.

Chapter 4
RESULT ANALYSIS

This chapter explains the results obtained from the comparative analysis of the three algorithms- Iterative Dichotomiser 3, *k*-Nearest Neighbor Regression and Multiple Linear Regression, and analyses the reason behind the difference in accuracy. The data containing three attributes– average (max + min) temperature and rainfall per month throughout the cropping period and crop yield rate per year of six major crops from ten major regions of Bangladesh was used for this analysis. The regions were considered are, Dhaka, Tangail, Barishal, Jessore, Sylhet, Mymenshing, Bogra, Patuakhali, Faridpur and Comilla. The considered crops were Aus, Aman, Boro, Wheat, Jute and Potato. Total of 12 data instances per region– that is the data of 12 years from each region was used to obtain the prediction. Data from year 2004 – 2013 were considered for learning cases and the data from year 2014 – 2015 were used for accuracy analysis. The results were obtained without removing any outliers to reflect the capability of the Farmer's Guide system without modifying the environment being analysed. This comparative analysis compares the performance of the above mentioned three algorithms in terms of the error percentage of the crop yield rate predictions obtained from each of them. The error was calculated using relative error finding methodology where the formula used is as following,

$$\%\ Error = \frac{True\ Value - Predicted\ value}{True\ Value} * 100$$

Also, this section investigates the reason behind the fluctuation in error percentage for each of the algorithm and draws conclusion on best performing one. The result of this analysis is proven to reliable as in many of the cases the resultant error was below 10%. The failure to predict a close value in other cases was due to inconsistency in the predictors and also due to performing the analysis without omitting any outliers. This shows that further modification can make this prediction model more accurate and can be made trustworthy.

The graphs in the next pages discuss the results from this comparative analysis.

4.1 Result Analysis for Tangail Region

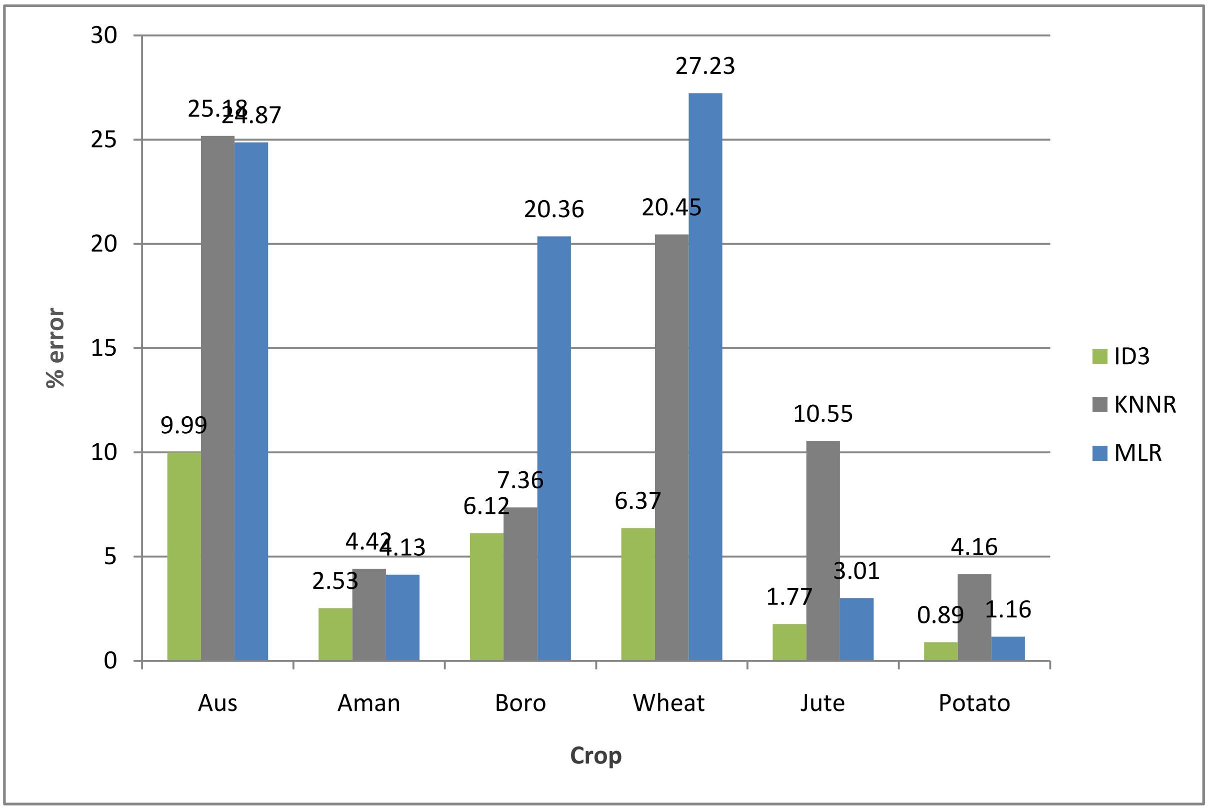

Figure 4.1(a): Resulting percentage error for Tangail region in year 2014

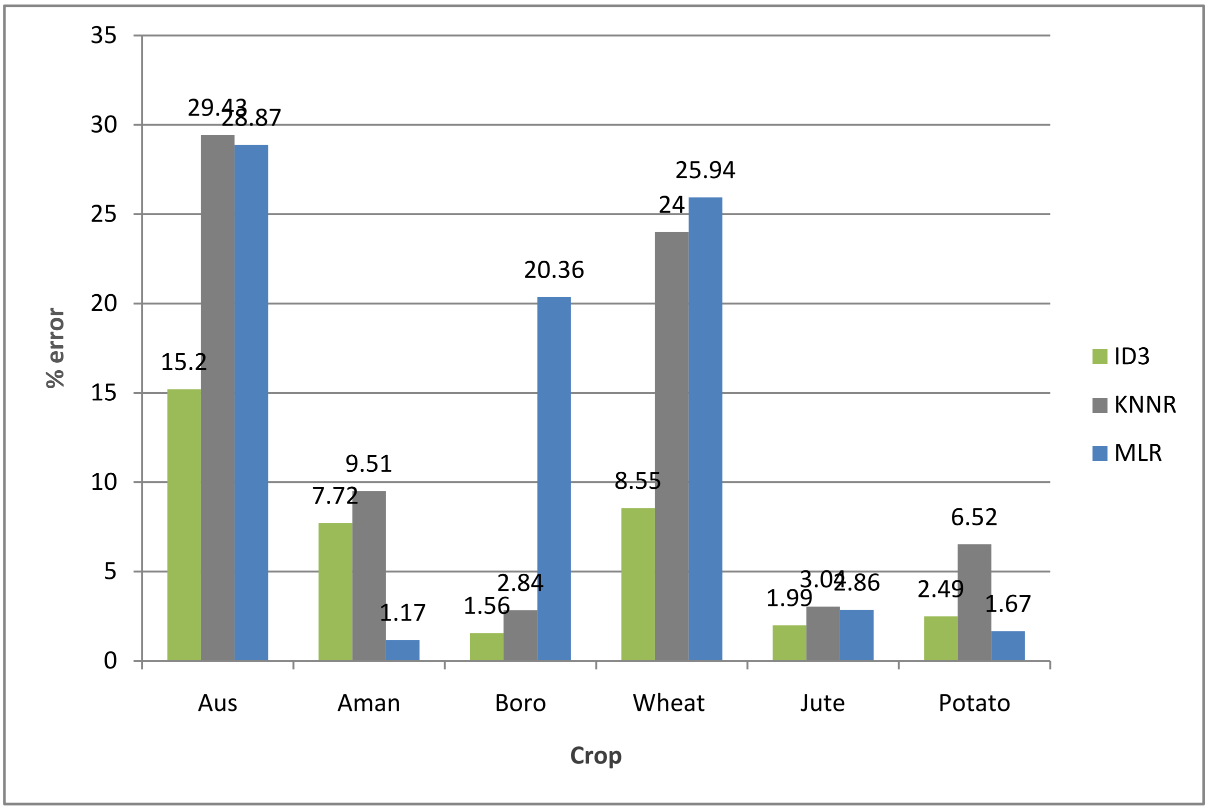

Figure 4.1(b): Resulting percentage error for Tangail region in year 2015

From the given graphs, we can see that in year 2014, ID3 gives better prediction for all crops than KNNR and MLR. Also, KNNR dominates in terms of accuracy over MLR in predicting yield rate of Boro and Wheat. The alternate case is also seen when predicting Aus, Aman, Jute and Potato.

In the year 2015, MLR wins on couple of occasions during the predictions (Aman and Potato), though ID3 comes out as the best predictor having the least error percentage when predicting Aus, Boro, Wheat and Jute. KNNR fails to win in any of the cases in terms of accuracy.

The decision tree has the better prediction methodology as it can single out any unusual cases in a separate branch. That is why we obtain better prediction from ID3 in most of the cases. KNNR and MLR fail to give the best prediction on many occasions because their performance depends on the consistency of the predictors. Due to the variation in temperature and rainfall in the learning years, we see MLR and KNNR to obtain predictions with error rate often as high as 20% or more. However, there are some cases in which ID3 has higher error percentage. This is because of the variation in crop yield rates in the learning years.

4.2 Result Analysis for Jessore Region

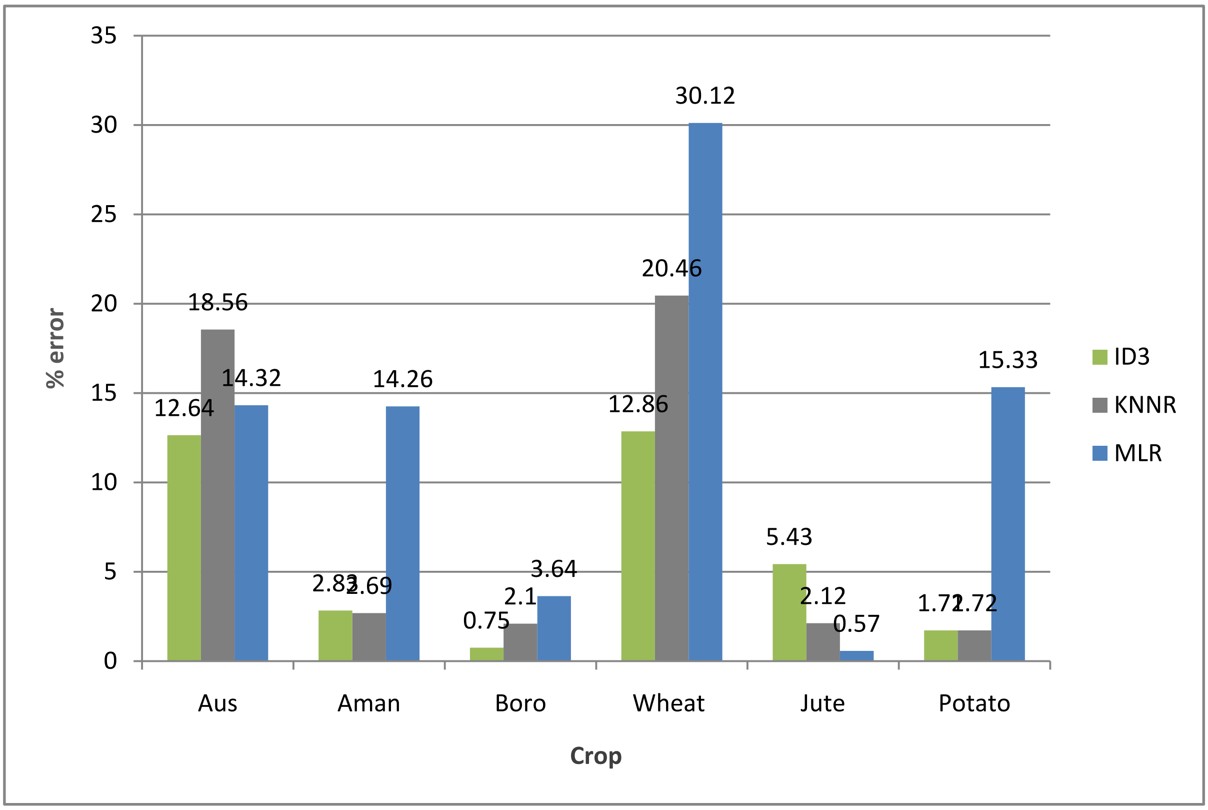

Figure 4.2(a): Resulting percentage error for Jessore region in year 2014

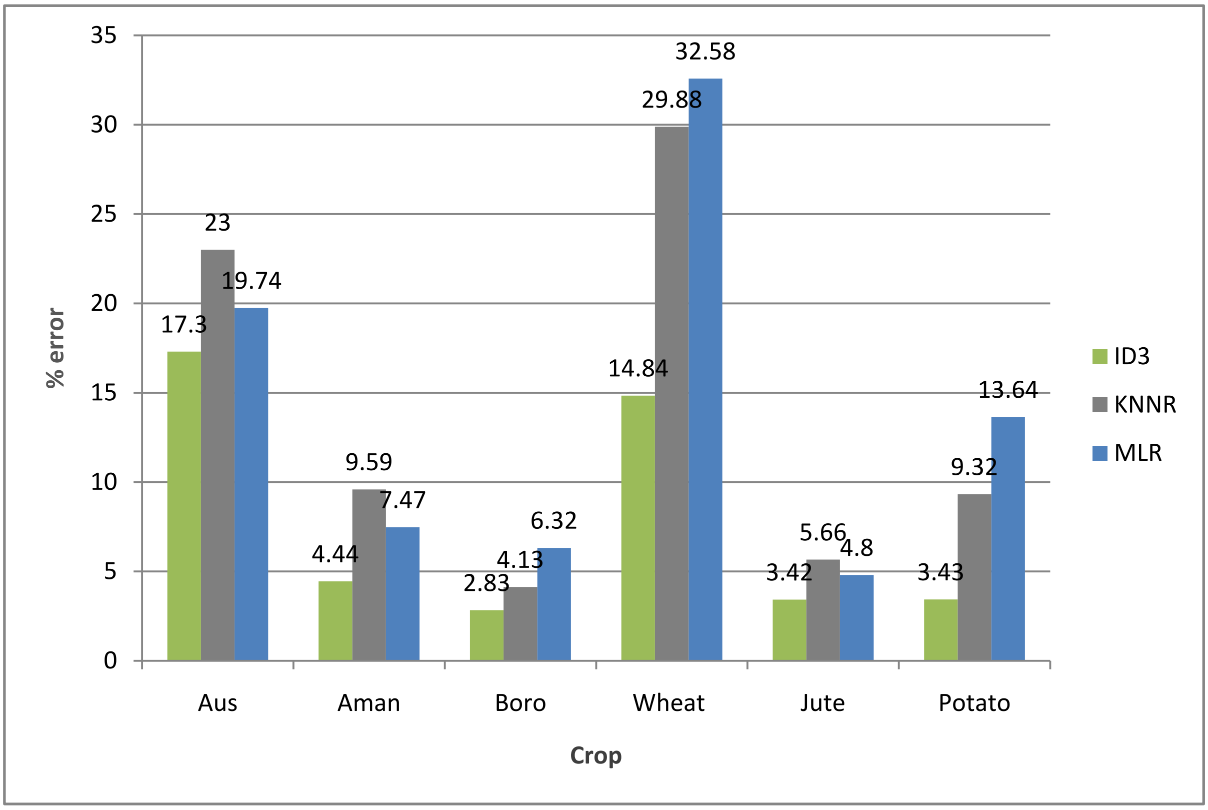

Figure 4.2(b): Resulting percentage error for Jessore region in year 2015

From the given result analysis graphs, we see that in the year 2014, ID3 algorithm obtained more accurate predictions of Aus, Boro, and Wheat than MLR and KNNR. Surprisingly, when predicting yield rate of Potato, both ID3 and KNNR obtained the same accuracy rate. MLR and KNNR obtained better results for Jute and Aman respectively.

In the year 2015, ID3 wins as a better predictor for all the crops. MLR dominates over KNNR in terms of error percentage when predicting Aus, Aman and Jute. The alternate case can be seen when KNNR predicts Boro, Wheat and Potato.

The decision tree has the better prediction methodology as it can single out any unusual cases in a separate branch. That is why we obtain better prediction from ID3 in most of the cases. KNNR and MLR fail to give the best prediction on many occasions because their performance depends on the consistency of the predictors. Due to the variation in temperature and rainfall in the learning years, we see MLR and KNNR to obtain predictions with error rate often as high as 20% or more. However, there are some cases in which ID3 has higher error percentage. This is because of the variation in crop yield rates in the learning years.

4.3 Result Analysis for Barishal Region

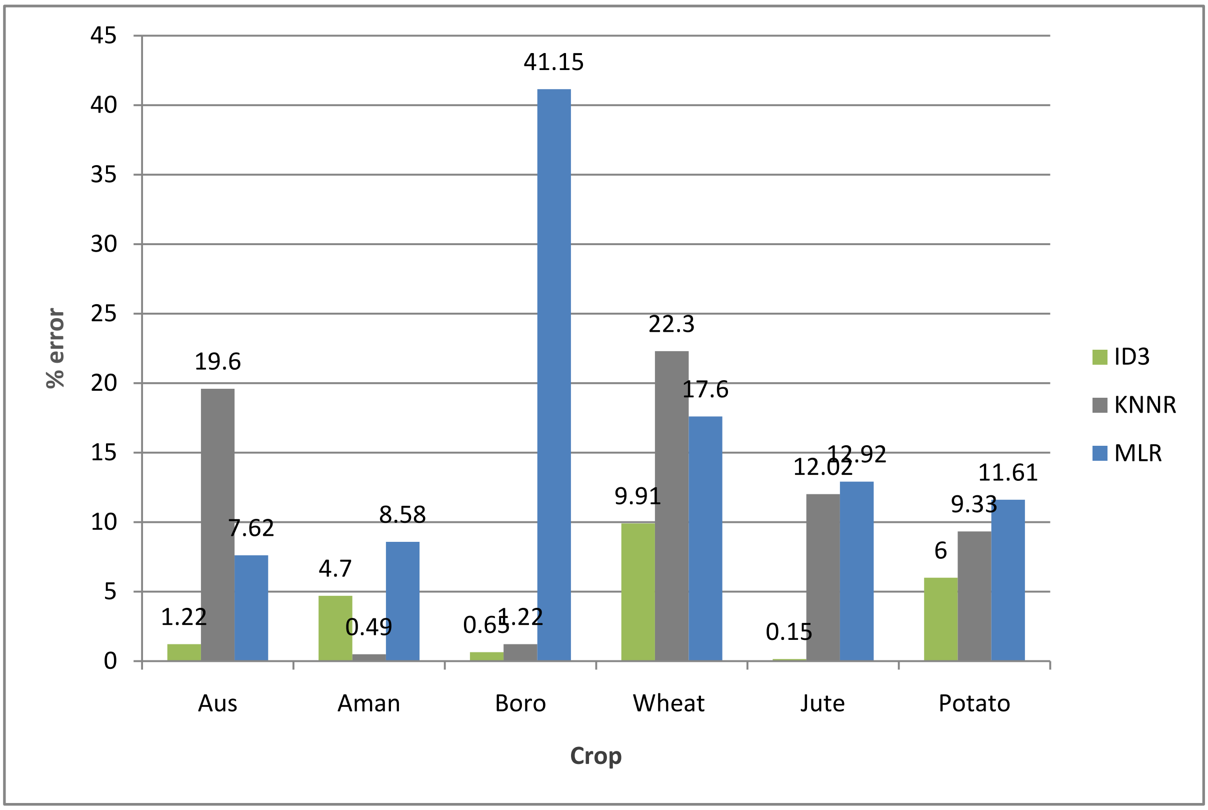

Figure 4.3(a): Resulting percentage error for Barishal region in year 2014

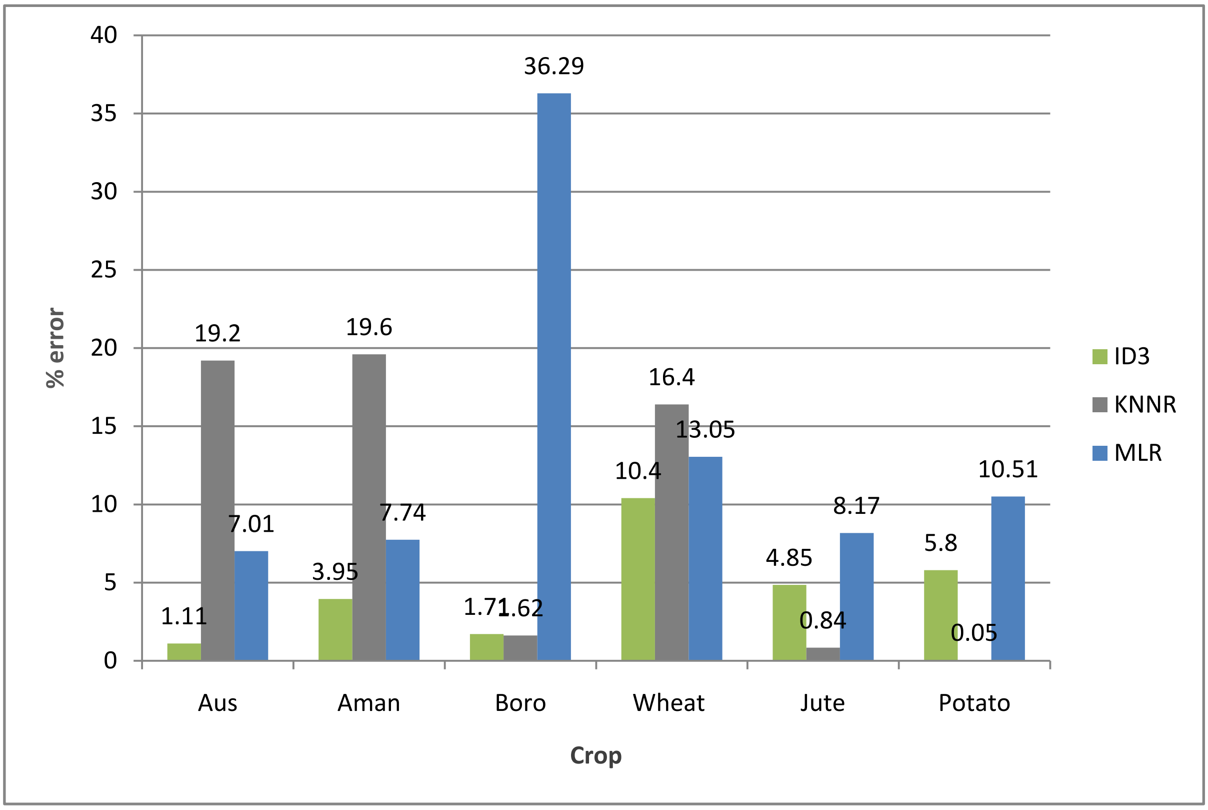

Figure 4.3(b): Resulting percentage error for Barishal region in year 2015

From the given result analysis graphs, we see that in the year 2014, ID3 algorithm obtained more accurate predictions of Aus, Boro, Wheat, Jute and Potato than MLR and KNNR. In the case of Aman, KNNR has the best prediction with an error rate near 0%. MLR dominates over KNNR when predicting Boro, Jute and Potato.

In the year 2015, KNNR wins on couple of occasions during the predictions (Boro, Jute and Potato). Also, ID3 comes out as the best predictor having the least error percentage when predicting Aus, Aman, and Wheat. MLR fails to win in any of the cases in terms of accuracy.

The decision tree has the better prediction methodology as it can single out any unusual cases in a separate branch. That is why we obtain better prediction from ID3 in most of the cases. KNNR and MLR fail to give the best prediction on many occasions because their performance depends on the consistency of the predictors. Due to the variation in temperature and rainfall in the learning years, we see MLR and KNNR to obtain predictions with error rate often as high as 20% or more. However, there are some cases in which ID3 has higher error percentage. This is because of the variation in crop yield rates in the learning years.

4.4 Result Analysis for Mymenshing Region

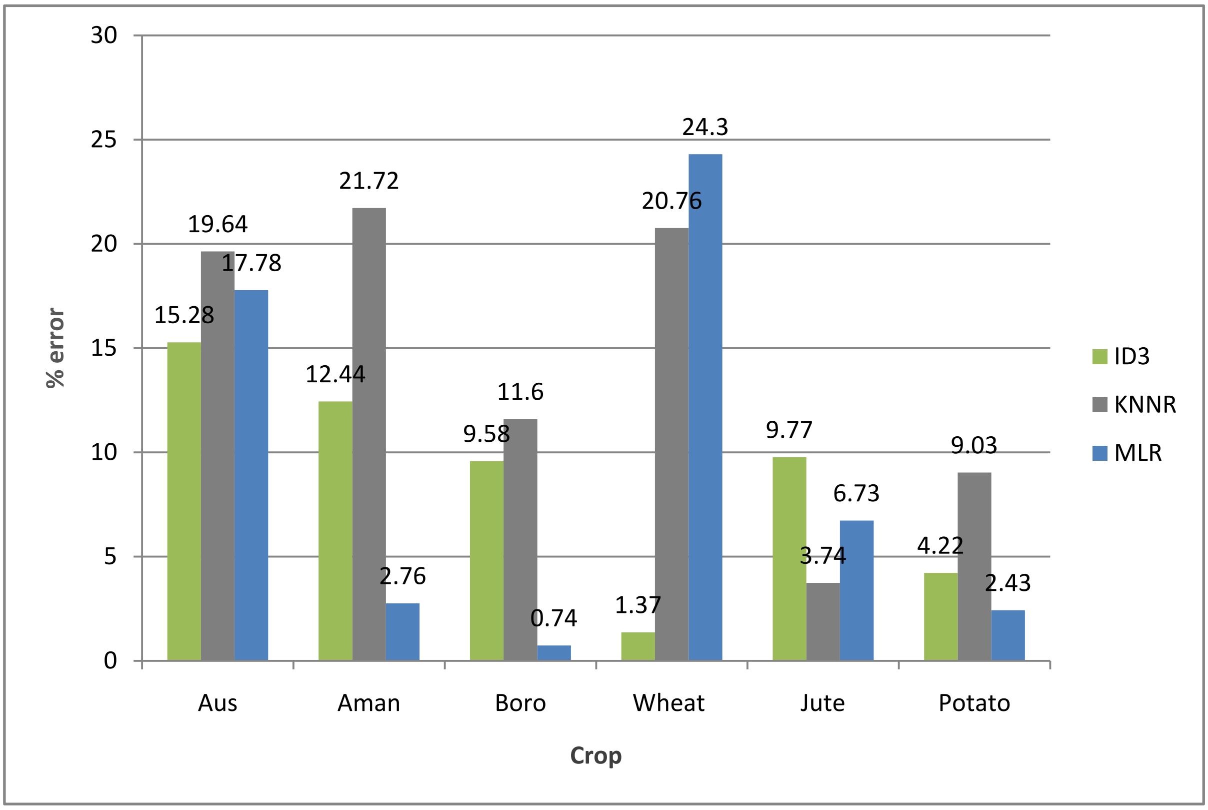

Figure 4.4(a): Resulting percentage error for Mymenshing region in year 2014

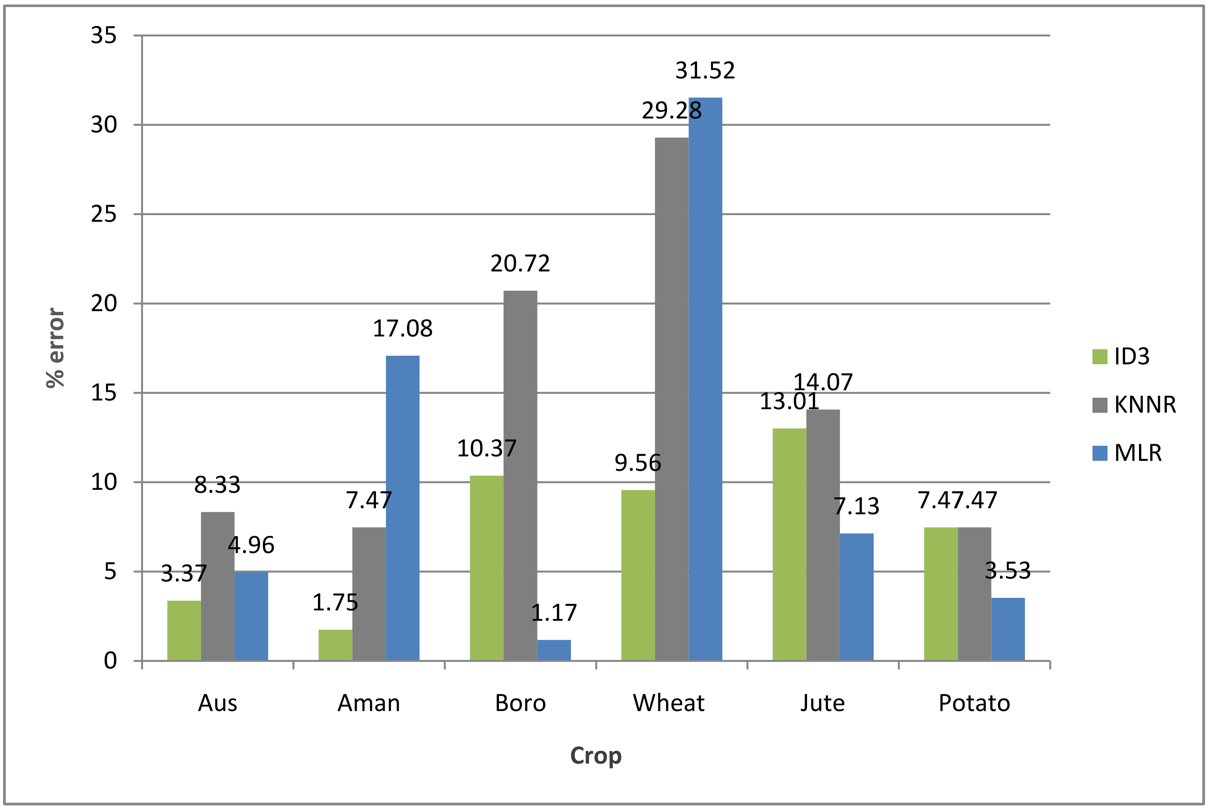

Figure 4.4(b): Resulting percentage error for Mymenshing region in year 2015

In the given result analysis graph, we can see that in year 2014, ID3 comes out as the best predictor only in two occasions (Aus and Wheat). MLR is the dominator in this analysis in terms of error percentage as it gives the best result when predicting Aman, Boro (below 1%) and Potato. KNNR predicts the best yield rate for Jute.

In the year 2015, MLR and ID3 both predicts 3 results that have the least error. ID3 comes out on top when predicting Aus, Aman and Wheat. MLR gives the better prediction with an error rate of <10% for Boro, Jute and Potato. KNNR fails to obtain any prediction with least error rate between three algorithms.

The results reflect the variations in yield rate in the learning dataset due to which the ID3 algorithm only were able to produce couple of predictions with the least error percentage. MLR’s performance in the year 2014 shows that the temperature and rainfall information in Mymenshing region were consistent throughout the ten years (2004 - 2013) that were considered for the learning cases. KNNR fails to produce better results due to the same reason ID3 algorithm struggled. The variation in crop yield rate over the years could happen due to any natural calamities which caused drop in production in some of the years considered.

4.5 Result Analysis for Dhaka Region

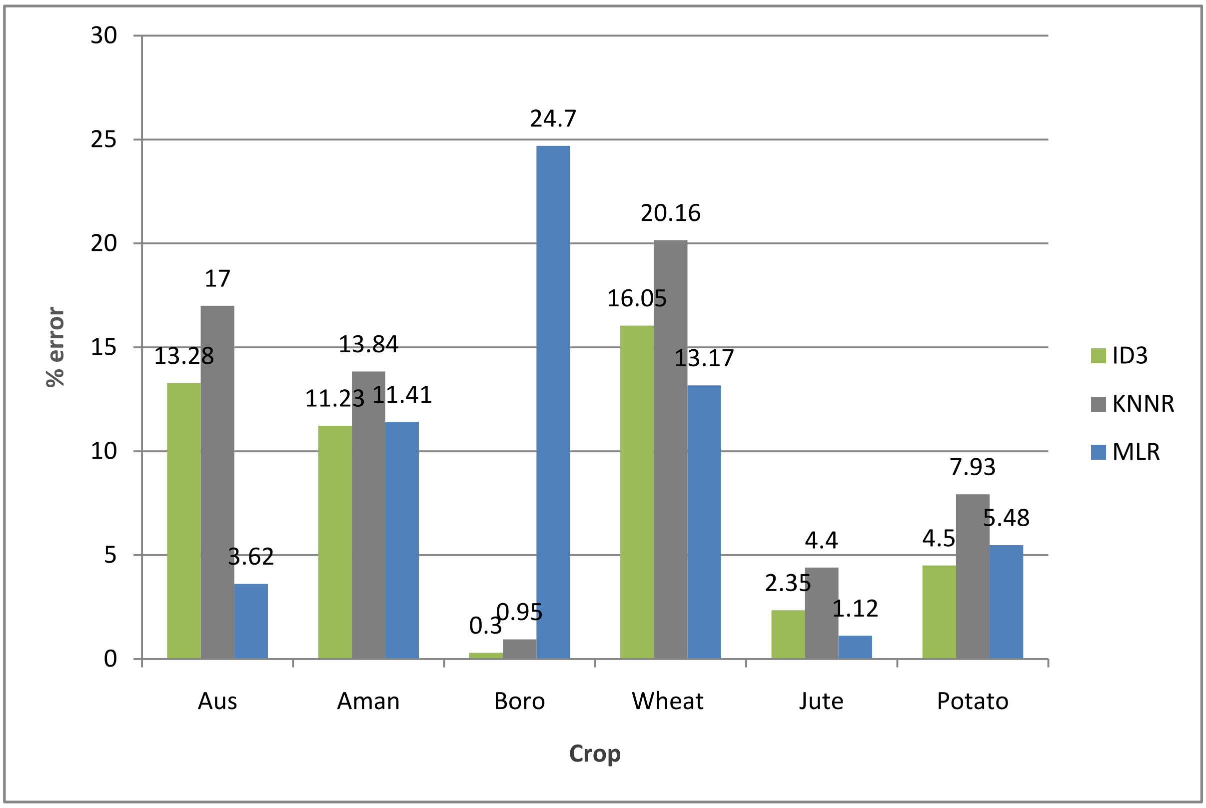

Figure 4.5(a): Resulting percentage error for Dhaka region in year 2014

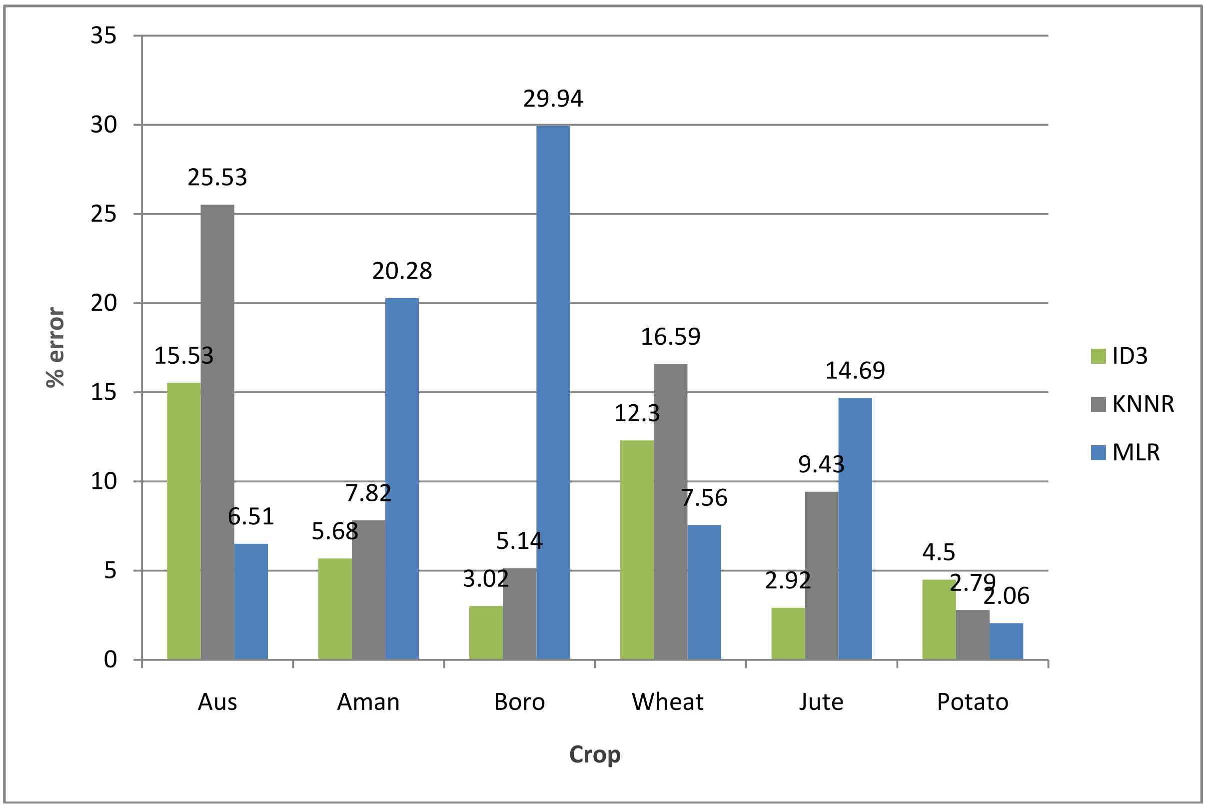

Figure 4.5(b): Resulting percentage error for Dhaka region in year 2015

From the given result analysis graphs for Dhaka region, we can see that ID3 gives better prediction for Aman, Boro and Potato. MLR dominates over KNNR on five out of six prediction results.

In the year 2015, ID3 wins over MLR and KNNR in predictions of Aman, Boro and Jute. MLR dominates over KNNR in terms of error percentage when predicting Aus, Wheat and Potato. The alternate case can be seen when KNNR predicts Boro, Aman and Jute.

In both years, MLR and ID3 have three predictions with least error percentage which reflects the performance of both of these algorithms in situations when predictor values are not consistent. The fact that both of them produced better result in 50% of the cases concludes that the prediction model being used in this analysis is somewhat consistent.

The results reflect the variations in yield rate in the learning dataset due to which the ID3 algorithm only were able to produce couple of predictions with the least error percentage. MLR's performance in the year 2014 shows that the temperature and rainfall information in Dhaka region were consistent throughout the ten years (2004 - 2013) that were considered for the learning cases. KNNR fails to produce better results due to the same reason ID3 algorithm struggled. The variation in crop yield rate over the years could happen due to any natural calamities which caused drop in production in some of the years considered.

4.6 Result Analysis for Faridpur Region

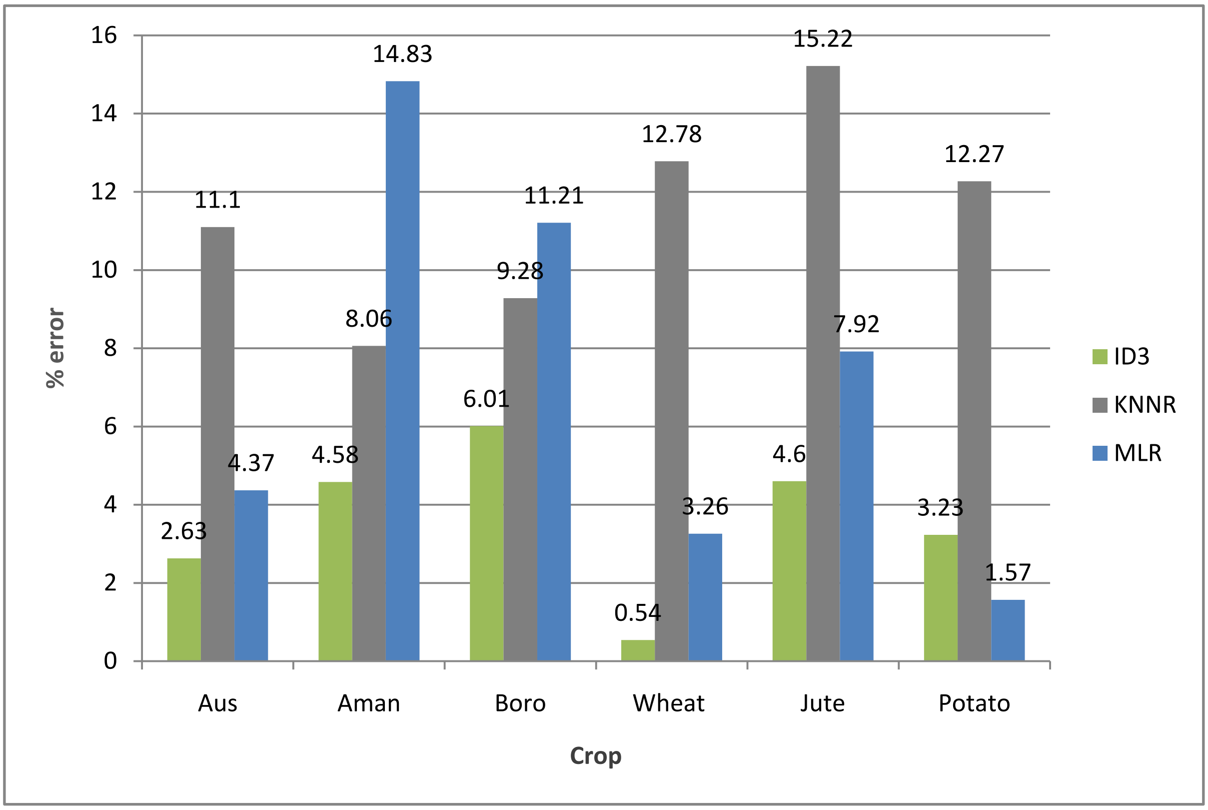

Figure 4.6(a): Resulting percentage error for Faridpur region in year 2014

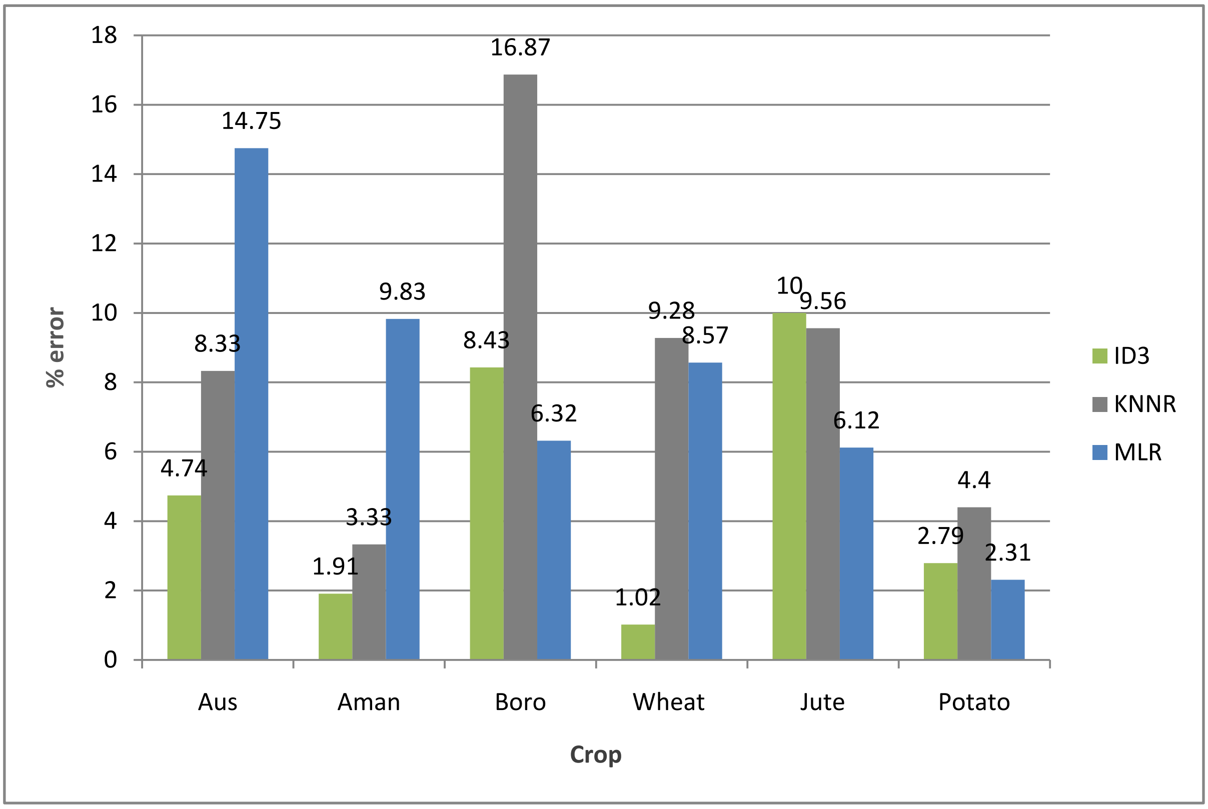

Figure 4.6(b): Resulting percentage error for Faridpur region in year 2015

From the given graphs for Faridpur region, we can see that in year 2014, ID3 gives better prediction for all crops than KNNR and MLR. Also, KNNR dominates in terms of accuracy over MLR in predicting yield rate of Boro and Aman. The alternate case is also seen when predicting Aus, Wheat, Jute and Potato.

In the year 2015, MLR wins on couple of occasions during the predictions (Boro, Jute and Potato). ID3 comes out as the best predictor having the least error percentage when predicting Aus, Aman, and Wheat. KNNR fails to win in any of the cases in terms of accuracy.

The decision tree has the better prediction methodology as it can single out any unusual cases in a separate branch. That is why we obtain better prediction from ID3 in most of the cases. KNNR and MLR fail to give the best prediction on many occasions because their performance depends on the consistency of the predictors. Due to the variation in temperature and rainfall in the learning years, we see MLR and KNNR to obtain predictions with error rate often as high as > 15%. However, there are some cases in which ID3 has higher error percentage. This is because of the variation in crop yield rates in the learning years.

4.7 Result Analysis for Comilla Region

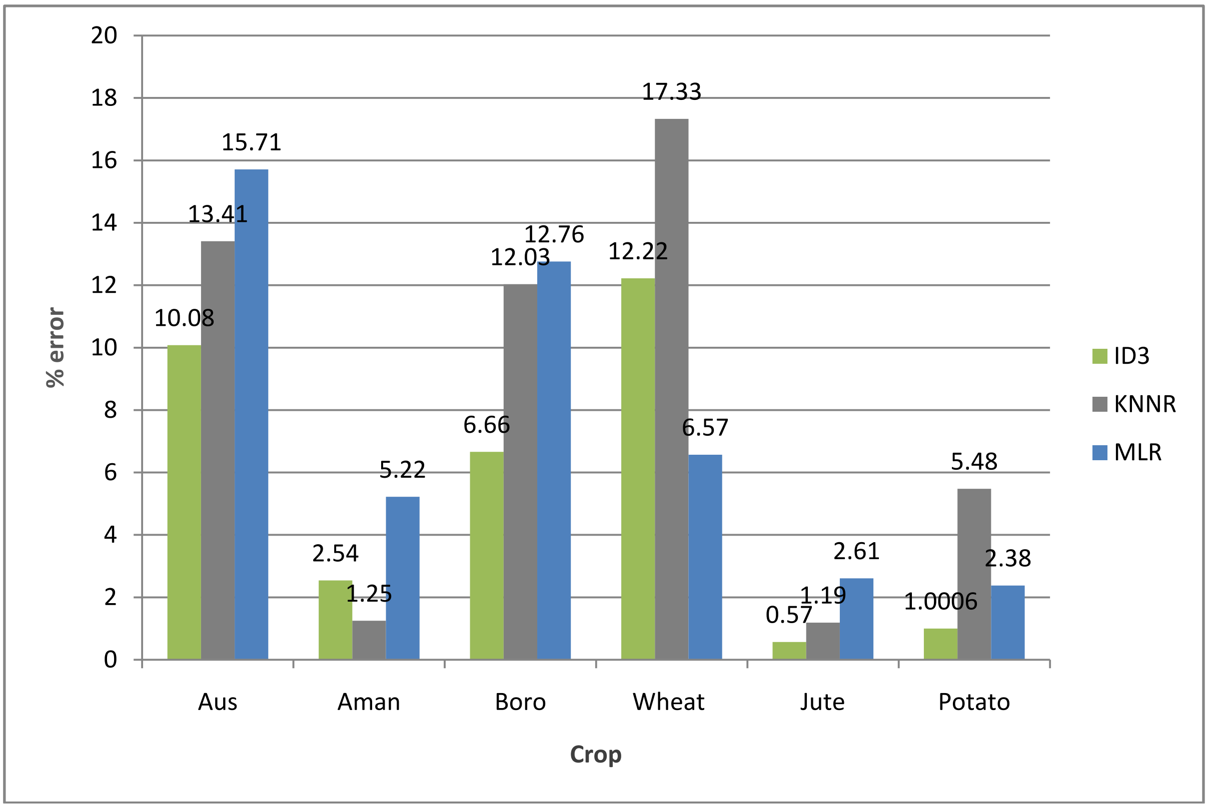

Figure 4.7(a): Resulting percentage error for Comilla region in year 2014

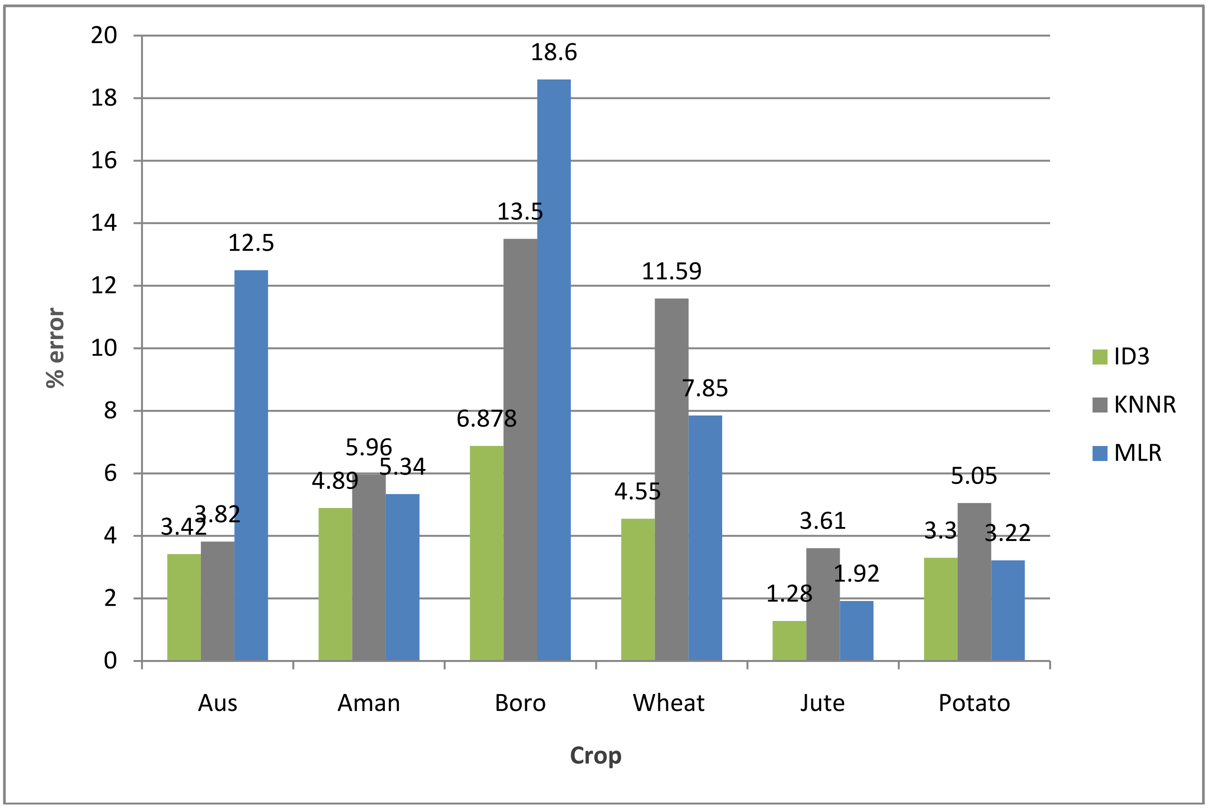

Figure 4.7(b): Resulting percentage error for Comilla region in year 2015

From the given graphs for Comilla region, we can see that in year 2014, ID3 gives better prediction for Aus, Boro, Jute and Potato than KNNR and MLR. Also, KNNR dominates in terms of accuracy over MLR in predicting yield rate of Aus, Aman, Boro and Jute. The alternate case is also seen when predicting Wheat and Potato.

In the year 2015, ID3 gives better prediction for all crops than KNNR and MLR. Also, MLR dominates in terms of accuracy over KNNR in predicting yield rate of Aman, Jute, Wheat and Potato. The alternate case is also seen when predicting Aus and Boro.

The decision tree has the better prediction methodology as it can single out any unusual cases in a separate branch. That is why we obtain better prediction from ID3 in most of the cases. KNNR and MLR fail to give the best prediction on many occasions because their performance depends on the consistency of the predictors. Due to the variation in temperature and rainfall in the learning years, we see MLR and KNNR to obtain predictions with error rate often as high as > 15%. However, there are some cases in which ID3 has higher error percentage. This is because of the variation in crop yield rates in the learning years.

Chapter 5

MOBILE APPLICATION

This chapter analyzes our mobile application in terms of its feasibility based on the product itself and its market. Also the requirements of the app, how much importance they bear, screenshots of our app and what function they are significant, features of our app, limitations while developing this app– all will be discussed in this chapter.

5.1 Feasibility Analysis

We did feasibility analysis of our project for different sectors:

a. Product Feasibility:

   The main feature of our app is to provide an efficient crop list to cultivate according to seasonal changes. It will benefit the young and new farmers who have a very little knowledge about farming as well as the experienced ones as the scientific analysis in this field was not introduced before. The concept of this kind of app is new for our country. There are few agricultural apps which describe the cultivation process, pest control and fertilization. Our app introduces a new concept and it also has few features of other agricultural apps. As it has a unique feature of generating efficient crop list based on real time, it has low competition in the mobile app market. Since our app is a one of a kind, there will be high demand on the market for this app. The websites related to agriculture and production can be benefited from our app and vice-versa.

b. Technological Feasibility:

   Our app consists of all the latest features. It is user-friendly in such a way that even people with very little knowledge can easily learn to use this app. All the user will need is a phone with Android support. Apart from that, no other additional hardware will be needed. Due to the cheap widespread availability of mobile data, the internet connectivity also is not a concern. The software components included in this app are as follows- the backend algorithms, the app itself and the information database. All of these features are backed up by latest technology. The Graphical interface is smooth and Bangla is used for the crop calendar and cultivation procedure in this app to make the use of it feasible and easy for the people of our country.

c. Economical Feasibility:

   Our app is easy to reach and economically feasible for all. We are releasing the Alpha version of our app for free in the Play Store. It may seem not economically feasible for us, but in the long run, we will release better versions of the app with more features, updates, better algorithms to increase accuracy based on the user demands. The cost will be suited in a feasible way for the developers. All the necessary data for our app was collected free of cost from AIS (Agricultural Information Service), BARC (Bangladesh Agricultural Research Centre) and many other government organizations. We have worked in this area previously and no additional programmer was required. Therefore, no additional cost was incurred in those sectors as well.

d. Market Feasibility:

   No doubt the mobile app development organizations have the widest range of customers; anyone who owns any handheld devices, such as smart phones / mobile phones, is a potential customer for our mobile apps. Our target customers are the farmers. Smartphones are now available around the world. So it will not be difficult for a farmer to get an access to our app. As it will be a unique system that offers features that are not conventional, there will be no competition in the mobile app market for now.

5.2 Requirement Analysis

Analyzing the requirements even before the development phase starts, is a necessity in system development as it can identify the functional (necessary for operational system) and non-functional (optional) requirements and helps to elicit dependent and independent functionalities. The results of the requirement analysis of the farmer's guide application are as follows,

a. Functional Requirements:

   - The users will need a smartphone with Android for the app to function.
   - Data connectivity is required in the smartphone.
   - User authentication is a must to use the app. Unauthenticated user will not be able to login into the app.
   - Our app requires a very good database support. The Firebase database is made up of that developers can mix and match to feed their needs.

b. Non-Functional Requirements:

- Any kind of problems with the app should be recovered as fast as possible. A crash report can be helpful for the user and provide reliability.
- We expect high traffic and so it is evident that multiple users are able to use the app simultaneously. We need to develop efficient ways to provide such service.
- The apps’ data processing speed should be as fast as possible. Local database (mobile storage) should not need to do too much work.
- The application currently uses basic, free for all database system – Firebase.

- We have plans to develop more user-friendly UIs in the future for this app. The initial prototype did not focus on UIs much.

5.3 User Interface

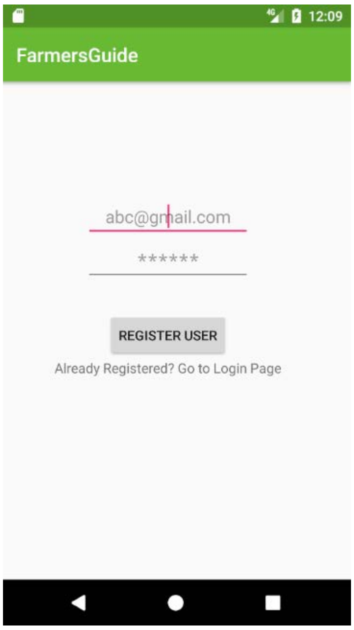

Figure 5.3 (a): Register Activity

Figure 5.3 (b): Login Activity

Figure 5.3(a) and Figure 5.4(b) shows the register and sign in activity of the farmer's guide app respectively. A registered user needs to enter email address and password only once to log in. To register to a new id, one has to input a valid email address and a password which must be of any numbers of characters. Once the user is registered she/he will be directed to the Login UI of the app. After login, the user will be directed to their Welcome UI (Figure 5.3(c)). Here, user will be provided with 10 options for region to choose from and a drop down menu from where the user will choose the year and the month that reflects probable cropping procedure beginning time. Once a user has chosen a valid cropping time, the app will direct the user to a new UI.

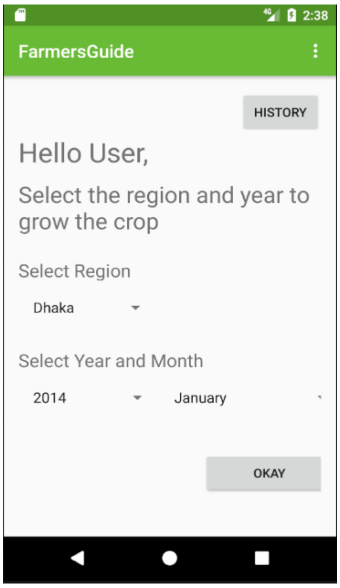

Figure 5.3 (c): Welcome UI

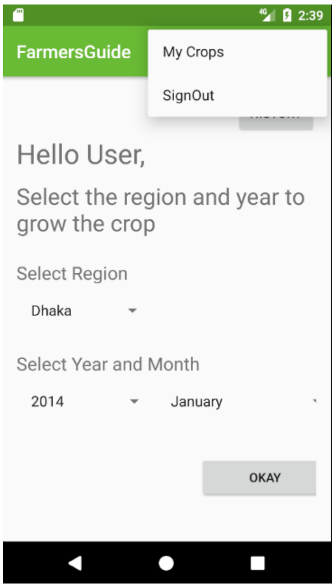

Figure 5.3 (d): App menu

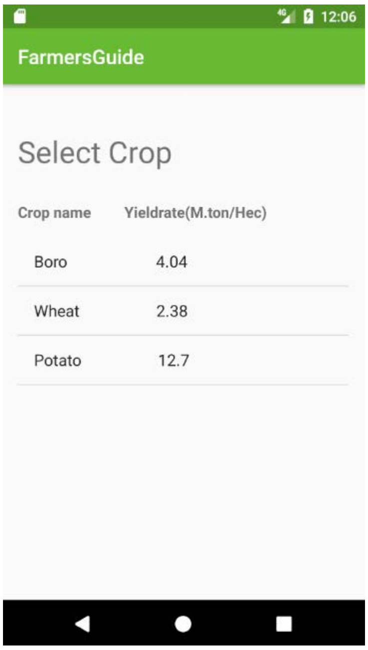

Figure 5.3 (e): Prediction Result UI

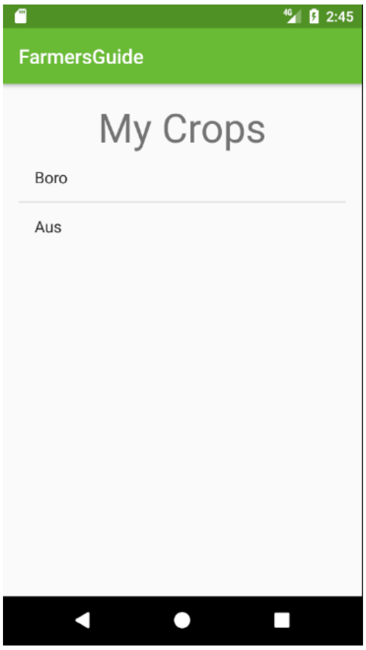

Figure 5.3 (f): Previous history activity

The user is now at prediction result activity (Figure 5.3 (e)). In this UI, the app shows the rank of possible crop types in terms of predicted yield rate that can be cultivated in the given time and chosen region by the user. For some cases, there can be multiple crop suggestions and in many others, there can be only one. The unit of yield rate per unit area is set to metric ton/ Hectare. When there is only one crop type possible to cultivate, there will be no option for the user to choose from. But when there are multiple options to choose from, user will have full independence to choose any of the suggested crops.

In the welcome page (Figure 5.3(d)) there is a dropdown menu which provides sign out and view history options where the latter, upon choosing, will direct to a new page (Figure 5.3(f)) which contains information on all the selected crops by the user in the past. This will also help the user to recall his previous experience with the crops in terms of cost, actual yield rate obtained and profit margin achieved. According to the option user has chosen, app will then proceed to the next UI.

Figure 5.3 (g): Crop Calendar Activity

Figure 5.3 (h): Cultivation Instruction Activity

Figure 5.3 (i): Suggestion activity 1

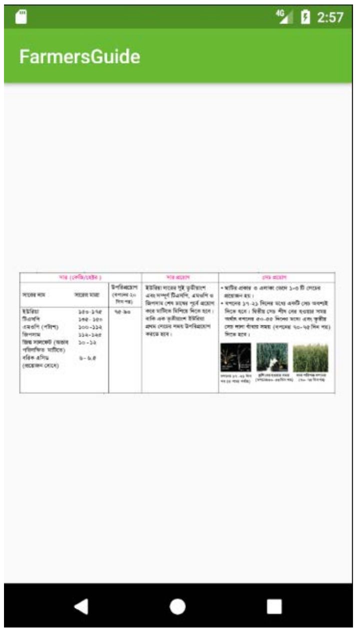

Figure 5.3 (j): Suggestion activity 2

Upon selecting a crop from the result menu, user will be able to see the crop calendar (details about cultivation, irrigation, fertilization schedule) of the selected crops (Figure 5.3(g)). There is also a DETAILS button in the bottom of the crop calendar representation which, upon pressing, will direct the app to a new page. The user now will be able to see all the points of cropping procedure of his/her chosen crop type (Figure 5.3(h)). The name of the procedure will be in button form. When to sow the seed, when to plant the seedling, when to irrigate and how much, when to apply fertilizer and the amount of fertilizer to be applied, all of these information will be available to access through these array of buttons (Figure 5.3(i), 5.3(j)).

Figure 5.3(k): Crop History Activity

Figure 5.3(l): Crop Estimated Cost Input Activity

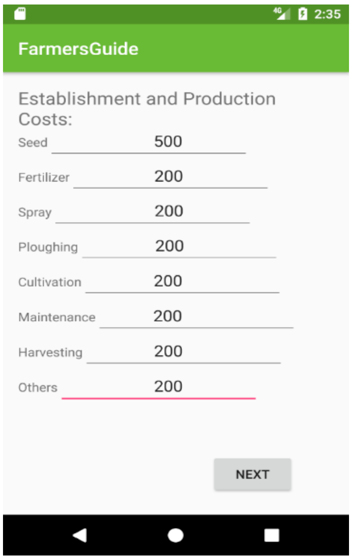

Figure 5.3(m): Crop Actual Cost Input Activity

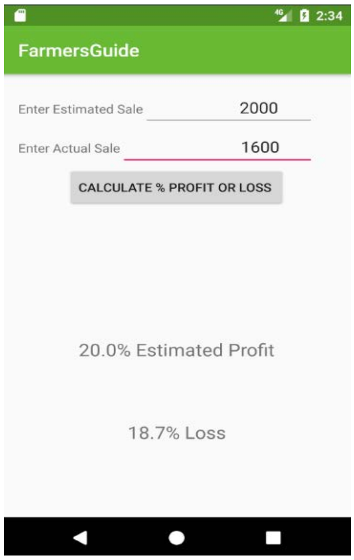

Figure 5.3(n): Profit/Loss Percentage Calculator

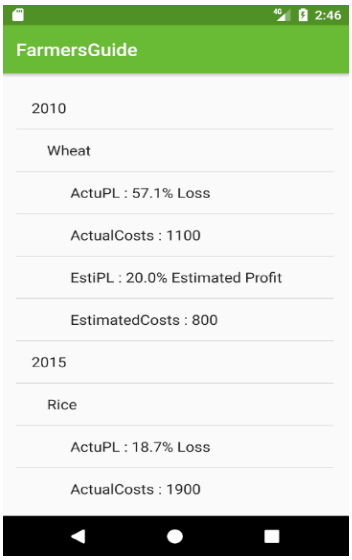

Figure 5.3(o): User History Activity

The cost calculator is accessible through the menu option in the welcome page (Figure 5.3(d)). The menu contains a option called CROP COST HISTORY which directs to a new activity (Figure 5.3(k)). In this activity, we can see two button ADD NEW COST DATA and SHOW PREVIOUS COST HISTORY. Upon clicking on the ADD NEW COST DATA button, a new activity will appear (Figure 5.3(l)). This is the estimated cost page where the user enters the cost values he/she estimated before the cultivation procedure. The NEXT button in the bottom of the display directs to a new activity which is the actual cost calculator (Figure 5.3(m)). Here the user enters the actual cost values he/she spent for the cultivation procedure. There is another NEXT button in the bottom of the page which directs to a new activity which is the estimated and actual sale comparator (Figure 5.3(n)). Here the user enters the estimated and actual sale values he/she predicted and achieved during cultivation and after harvesting. By clicking the button CALCULATE PROFIT/LOSS the page shows the Estimated Profit/Loss and the Actual Profit/Loss percentage (Figure 5.3(n)). The Total cost and sale values and the profit/loss values are then stored in the Firebase real time database. From the CROP COST HISTORY activity (Figure 5.3(k)), if the user clicks on the SHOW PREVIOUS COST HISTORY button, the user will be directed to an activity containing the data of the cost, sale and profit/loss values for the crops that the user had entered through the previous year's (Figure 5.3(o)).

5.4 Features

- Register and Login - Users needs register first. Then she/he can log in. The user data is stored against the user account information.
- Location – The app takes user's location as input to generate region specific result.
- Date - The app takes the cropping beginning date as input and identifies appropriate crops according to that date.
- Machine Learning Algorithm – The app analyzes the information given to decide which crops can be processed through a machine learning algorithm ID3.
- Prediction - By running the machine learning algorithm in the backend, it predicts possible yield rate in metric ton per hectare for each crop.
- Choose Crop - App then allows the user to choose one of the multiple crops (if available).
- Crop Calendar - After user chooses a certain crop, app then shows the crop calendar of that crop.
- Farming Procedure - The app can show detailed planning of the farming procedure, fertilization, pest control, irrigation, harvesting for the chosen crop if user wishes.
- Crop Cost History - Our app calculates the estimated profit/loss and actual profit/loss based on estimated cost, actual cost, estimated sale and actual sale. It stores all these information in the Firebase real time database and can show the farmer's previous history of crop cost.
- Uniqueness - The app is unique as it is the only one available in our country currently that proposes the mentioned features.
- Database - Our app consist a database with the crop information data of the past 12 years in the backend to make more accurate predictions.
- Internal Server - The app stores the data in a cache memory and runs the machine learning algorithm in the backend and generates prediction for the future
- Less Local Processing - We tried to minimize the involvement of local database and processing functions by storing the whole data in the cloud from which the app only retrieves the region specific information for processing so that it could process the predictions faster and run smoothly.
- Additional Instructions: Our app not only predicts on the crop yield rate to facilitate in taking crop selection decisions, but also provides necessary scientifically proven cultivation, irrigation and fertilization instructions as well as their schedule to ensure maximum agricultural output.

5.5 Limitations

- Prototype: We are launching a prototype of the app which will be updated with more features and more time and effort will be contributed in the future to make fully operational. Basically, we are developing a standard for others to follow and make it better with gradual function integration.
- Limited data size: The data size we have collected is inadequate for generating accurate predictions. The data of the past 12 years (starting from 2004 to 2015) has been used in this analysis which is not enough to generate a prediction model. A project like ours needs a huge amount of data in the backend to generate a pattern and draw a conclusion based on them. Addition of more data instances (15 – 20 years) is necessary to make utilize this prototype to the fullest.
- Lack of relevant knowledge: Due to a lack of deep understanding on this subject matter, the reason behind fluctuation in the predictions could not be identified. As a result, the algorithms are modified enough to generate results with less than 10% error rate whether removing the mentioned outliers could reduce the percentage by a significant amount.
- Less area coverage: Utilizing the app to its full extent to see actual agricultural growth of Bangladesh is not possible as it currently covers only 10 regions of the country. To see actual improvement in this sector, inclusion of more regions is necessary.

- Language limitation: The app user interface in not friendly enough for the farmer of the Bangladesh as it does not have the Bangla interface.

Chapter 6

CONCLUSION AND FUTURE WORKS

This chapter draws a conclusion on the work that have done till date and gives a picture of the future possibilities the Farmer's Guide application holds. It describes the functionalities which, upon addition, will convert the prototype developed in this research to a fully operational application.

6.1 Future Works

1. More area coverage: The application currently can predict on major crops of 10 districts in Bangladesh. Data of other 54 districts can be added in the future to facilitate farmers all over the country with their crop selection decision and cropping methods.
2. Increasing data size: The more the number of data instances can be added for analysis, the better the prediction obtained. Thus, our objective is to either collect and add more data of the late 1900s if possible or rely on the addition of data for each passing year.
3. Addition of crops: Currently, the app can predict the yield rate of six major crops from Bangladesh, where a large number of crop variation are being cultivated each year. Adding more crops in the data sample will help the farmers who rely on crops that are not considered major contributor to the economy but hold better profit margin for them.
4. Location tracker: Implementation of location tracking through technologies like global positioning system is possible to allow the app to identify user's location automatically instead of taking it as user input.
5. Automated land measurement: Implementation of automatic land area measurement feature by using gyroscope and GPS is possible where the user will just have to walk around the land to let the app identify the location and land area.
6. Disease detection using image processing: Our initial plan included the function where the camera sensor of the smartphone and a image processing algorithm is incorporated to identify diseases. However, due to the magnitude of the work and considering the time constraint, it is added to futures scope of our application.
7. Increased security: Currently, the app provides secured environment for information access to the user using the default security protocols offered by the android

application technology. Later on, we will be adding better encryption protocols to ensure maximum security for the user data.

8. Bangla language package: Since the app was built considering the farmers of Bangladesh as the target market, addition of Bangla language was necessary. However, due to the limited time constraint, we were unable to add it in our prototype. Thus, future objectives include addition of Bangla language to facilitate the local farmers.
9. User interface improvement: Since the app developed during this research is a prototype, the objective was to make it simple and operational and thus, less emphasis was put on the design of the user interface. Therefore, future scope includes development of an interactive UI which will not only be easy to operate but also will be informative.
10. Multi-platform app: Considering the economic condition of the farmers and since the only possible smartphone options they can afford come with preloaded android OS, the app was built to be run on android platform. This bias view will be eliminated with the development of the app on ios and windows platform.
11. Schedule notifications: Adding this function will allow the app to remind the farmer about irrigation and fertilization according to the respective crop schedule.
12. Cost estimation model: Using the provided cost calculator, the calculated cost of each crop for different user can be accumulated and analysed to predict the cost of cultivation. This cost attribute could also be incorporated with the existing prediction model to predict the beneficial crop not only in terms of maximum output but also in terms of maximum profit.

6.2 Conclusion

Investing huge amount of money in cultivating a crop selected by basing one's judgment on intuition should be considered as a medieval act in this era of technology. This app provides a solution to this problem which was much needed for farmers in Bangladesh. Though the features provided are limited, the future ahead promises addition of many features that will not only predict the profitable crops, but also provide a hub for the community to share their knowledge and experience which can be analyzed with machine learning techniques to generate predictions on more important factors like cultivation cost, soil composition etc. These analyses will result in more profits and invention of advanced farming techniques that will in turn, improve our economy and will help us stand out as a technologically advanced country on the map.

Reference

[1] Y. Hayami and V. W. Ruttan, Agricultural development: an international perspective, 1st ed. Baltimore, London: The Johns Hopkins Press, 1971.

[2] S. Ray, "Essentials of Machine Learning Algorithms (with Python and R Codes)," in Analytics Vidhya, 2015 [Online]. available at: https://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/ Accessed: Nov. 29, 2016.

[3] P. Vinciya, Dr. A. Valarmathi, "Agriculture Analysis for Next Generation High Tech Farming in Data Mining," International Journal of Advanced Research in Computer Science and Software Engineering, vol. 6, issue. 5, pp. 481-488, May.2016.

[4] Trading Economics. (1994-2016) Trading Economics. [Online]. Available at: http://www.tradingeconomics.com/bangladesh/gdp-growth (Accessed: 15 August 2016).

[5] IRRI. (2012) IRRI. [Online] available at: http://irri.org/news/media-releases/feeding-rice-just-got-easier-with-smartphones (Accessed: 12 December 2016).

[6] Algorithm (Multiple Linear Regression). [Online]. Available at: http://www.originlab.com/doc/Origin-Help/Multi-Regression-Algorithm (Accessed:12 December 2016).

[7] K-Nearest Neighbors Algorithm [Online]. Available at: https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm#k-NN_regression (Accessed: 12 December 2016)

[8] S. Veenadhari, D. Bharat Mishra, and D. C. Singh, "Soybean productivity Modelling using decision tree Algorithms," International Journal of Computer Applications, vol. 27, no. 7, pp. 11–15, Aug. 2011.

[9] N. KUMAR, G. P. O. REDDY, S. CHATTERJI, and D. SARKAR, "An application of ID3 Decision Tree Algorithm in land capability classification," Nagpur 440033, India, 2012, pp. 35–42.

[10] Ashwani Kumar, Sweta Bhattachrya, "Crop yield prediction using Agro Algorithm in Hadoop," International Journal of Computer Science and Information Technology & Security (IJCSITS), vol. 5, no. 2, pp. 271-274, April.2015.

[11] Entropy: M. Somvanshi and P. Chavan, "A review of machine learning techniques using decision tree and support vector machine," 2016 International Conference on Computing Communication Control and automation (ICCUBEA), Pune, India, 2016, pp. 1-7.doi: 10.1109/ICCUBEA.2016.7860040

[12] Miss.Snehal, Dr.Sandeep, "Agricultural Crop Yield Prediction Using Artificial Neural Network Approach," International Journal of Innovative Research In Electrical, Electronics, Instrumentation And Control Engineering (ijireeice), vol. 2, Issue 1, pp. 683-686, Jan.2014.

[13] Manpreet, Heena, Harish, "Data Mining in Agriculture on Crop Price Prediction: Techniques and Applications," International Journal of Computer Applications, vol. 99, no. 12, pp.1-3, August.2014.

[14] M. C. S. Geetha, "Application of classification technique in data mining for agricultural land," IJARCCE, pp. 352–355, Mar. 2015.

[15] D. J. Hand, "Data Mining: Statistics and More?," *The American Statistician*, vol. 52, no. 2, pp. 112–112, 1998.

[16] ACM SIGKDD, "Data Mining Curriculum: A Proposal," in KDD, 2006. [Online]. Available: http://www.kdd.org/curriculum/index.html. Accessed: Nov. 02, 2016.

[17] S. Yang, “A Search Algorithm and Data Structure for an Efficient Information System,” University of Wisconsin Madison, Wisconsin.

[18] M. Tranmer and M. Elliot. "Multiple linear regression." The Cathie Marsh Centre for Census and Survey Research (CCSR) (2008).

[19] K-Nearest Neighbors Algorithm [Online]. Available at: https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm#k-NN_regression Accessed: 27 Nov. 2016.

[20] "Climate Information Management System", Climate.barcapps.gov.bd, 2017. [Online]. Available: http://climate.barcapps.gov.bd/. [Accessed: 16- Aug- 2016].

[21] Yearbook of Agricultural Statistics - (2008-2015) [Online]. Available at: http://bbs.gov.bd Accessed: 27 Nov. 2016.