N.P.K. BASED CROP SUGGESTING MODEL

by

Asif Iftekher
14201062
Tomalika Ghosh
12301028

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
Brac University
December 2019

Declaration

It is hereby declared that

1. The thesis submitted is our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Asif Iftekher	Tomalika Ghosh
14201062	12301028

Approval

The thesis/project titled "N.P.K. Based crop Suggesting Model" submitted by

1. Asif Iftekher (14201062)
2. TomalikaGhosh (12301028)

Of fall, 2019 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on December 26, 2019.

Examining Committee:

Supervisor:
(Member)

Dr. Md. Ashraful Alam
Assistant Professor
Department
Institution

Head of Department:
(Chair)

Mahbub Alam Majumdar
Professor
Department of Computer Science and Engineering
Brac University

Executive Summary

Bangladesh is a country having an area of 1, 47,570 square kilometers in which roughly 70.63 percent is agricultural lands [1]. As an agricultural country we are mostly dependent on soil. There are 3 most important nutrients in any soil, it's known as the primary macro nutrients: Nitrogen (N), Phosphorus (P), and Potassium (K). Each of the primary nutrients is very essential in plant nutrition, serving a critical role in growth and reproduction of the plant. The purpose of this project is to make a N.P.K. Based Crop Suggesting Model by using machine learning which will determine the best crop to grow in a particular soil based on some major criteria. This model will play a vital role in our agricultural sectors to fulfill the needs of our country by reaching the highest level of efficiency and ensure the best use of our arable lands.

Keywords: N.P.K. based, Machine Learning, Crop Suggesting model

Acknowledgement

Firstly all praise to our great almighty for whom our thesis have been completed without any major interruption.
Secondly to our supervisor Dr. Md. Ashraful Alam sir, who guided and helped us in everywhere we needed his help.
Thirdly to our thesis panel, their feedback on our thesis draft helped us to finish the paper properly.
And finally to our parents without their support maybe it was not possible. With their kind and prayer we are now on the verge of our graduation.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Bangladesh is a country having about 14 million hectares of land space. Among this the arable land is about 8 mha [2]. Soil of a crop land is different in each part of the country. The general soil types have been described with reference to their occurrence, geographic distribution, morphology, land use and cropping pattern. A total number of 21 general soil types have been recognized that are distributed on three units. 14 general types identified on floodplain area, 6 on terraces and 1 on hilly area. Every year the number of arable land is reducing by 1 percent for urbanization and industrialization. Moreover very good agricultural lands account for less than 50 percent of the total arable land. In all those different soil types the proportion of nutrients fluctuates based on their type. As an agricultural country we have to fully focus on the amount of nutrients in our soil to enhance the soil quality and the production of plants while making the best use of the arable lands we get. Every crop requires a certain level of nutrients which helps a crop to have maximum growth and good health. There are various kinds of nutrients which a plant needs to grow up. Factors like weather, temperature, rainfall and other factors plays a vital role in cultivating crops in different areas but we will only focus on the main three nutrients which are Nitrogen (N), Phosphorus (P) and Potassium (K). As it's almost a pilot project to come up with an idea and make it work in real time, we focused on a particular upozila that is Ullapara Upozila from Shirajganj to overcome huge load of data and more complexity. If we get satisfied with the accuracy of this model we can apply it for the rest of the places in Bangladesh. In this project we are proposing a crop suggesting model using machine learning algorithms to determine the best suitable crops for any particular soil sample. This model will make it easier to decide which crop farmers have to cultivate in order to gain maximum efficiency and earn maximum range of profit.

1.1 Motivation

At first we planned about doing a project on soil and crops using machine learning where we will taste the soil to find its main nutrient values and take some inputs from the user and cross match the inputs with the value we found from soil test and give the best crop to cultivate in a particular land based one some criteria. As we started gathering information related to our project, we came to know that testing soil samples can be too complicated as it requires high-tech machines and also requires a vast knowledge on soils science which we don't have. We found

out that governments of Bangladesh have data of every single soil samples exist in Bangladesh but it was almost impossible task for a farmer to put the nutrient value of a particular soil as an input .To overcome this critical condition we had to improvise some new ideas to make it more easier for the end users. That leaded us to this present idea of this N.P.K Based Crop Suggesting Model which will help the users to know about the best crops can be cultivated in a particular land in a very short period of time. Our model just requires some common inputs such as name of the district; union and upozila, land type and our model will suggest the best 3 crops which can be cultivated on that particular land with full efficiency and a good profit can be earned through it.

1.2 Thesis overview

It's a soil based project. The purpose of this project is to make proper use of the arable land we have and ensure maximum profit by cultivating the exact suitable crop in a particular land. It's a very easy process for the user to determine the best crop. This model will ask some input such as district, union name of the sample soil, what is the type of that soil, is it high or low land and lastly the investment that an user can make. Based on the input this model will suggest best suitable crop for that particular soil sample by using machine learning algorithms which will be feasible with their investment and will make them be able to earn maximum profit out of it. The output will be generated by keeping in concern of various economic factors as well like market price, production cost and loss factors etc.

1.3 Major Contributions

We already discussed about the total land space we have and how much of it is arable lands. But the volume of arable lands in our country is not fixed .It's getting lesser day by day with the increasing number of population. Currently we have 160 millions of population and the number will definitely increase with time. As a result there will be a huge shortage of arable lands in near future. As our country is mainly an agricultural country where the overall economy depends mostly on agriculture, so shortage of arable lands will be so alarming for us.Though we have shortage of arable lands but even then we are not making proper use of it. Farmers often have lands and investment but failed to cultivate their desired crops with a good margin of profit. Choosing the right land or soil for any particular crop is so important. Farmer has to be fully aware about the soil nutrients of his land, how much fertilizer is needed based on that nutrient values and finally which crop can give him the best outcome and he can earn a good marginal profit. Our project will bring up the best solution of these problems. Farmers just have to know a very minimum level of information about the land where they want to cultivate a crop. Not necessary that he has to know the nutrient values or anything of his land. He just has to give the name of the area specifically where his land exist, types of soil and land and the amount of money he can invest to grow any particular crop. Our model will suggest the best crop he can cultivate on that particular soil and can make a good profit. It will make a life of a farmer easier than before. They can fully focus on the cultivation process rather than thinking about what crops can be cultivated in

their lands. It will help to make the best use of arable lands we have and also play a vital role in farmer's economic condition as our model is suggesting to cultivate a crop according to the investment he can make and also concern about the margin of profit he can make after selling the crop. So it's becoming almost a sure shot for them cause they will know the end result from the very beginning if no unwanted natural calamities take place. So we believe that our N.P.K. Based Crop Suggesting Model will keep a major contribution in agricultural sectors and in Farmer's day to day life.

Chapter 2

Related Work

2.1 Background study

Ullahpara is an upozila situated in the south-west part of shirajganj. Its total volume is 415 square kilometers, among this 5.81 square kilometers are water source such as pond rivers etc. It consists of 1 pouroshova, 14 unions and 263 villages. Average highest and lowest temperatures are 13.5 and 34.9 degree and annual rainfall is roughly 698 mm. Two types of soil mainly constructed ullapara upozila those are Tista Polol Vumi and Ganga Polol Vumi. But most soils are Tista polol Vumi.Both Tista , Ganga Polol Vumi consists some soil groups such as Manda, Jamun,Kaunia,Loshkora etc. from Tista type soil and only Gheur and Batra are from Ganga Polol Vumi.Based on the soil types Ullapara upozila is divided into 8 Manchitro Ekok. Soil types and nutrients are totally different from each other in every single place. Roughly 91 types of crop is cultivated here by the farmers. There are two kinds of crops, one is Upland Crop and one is Wet Land crops. Every single crops requires certain level of nutrient values such as Nitrogen, Potassium, zinc, Boron, phosphorus, iron etc. but among these the Nitrogen, Phosphorus, potassium are the main nutrient values which play a significant role in the growth of a particular crop. So we are only focusing on the N.P.K. values for the soils samples and for the crops as well. Necessary data for this project about Ullapara Upozila and the union under it is shown in data set part bellow.

2.2 Algorithm

We implemented four different algorithms in order to get correct accuracy and to predict the best crop for any particular soil. The algorithm we used in these projects is SVM, Ada-boost, Random Forest and Logistic Regression. We will discuss about the algorithms in this section of our project.

2.3 Logistic Regression

Logistic regression is a statistical method to analyze one or more independent variables in a data set to determine an output, which will be a dichotomous variable. To predict a binary outcome we used this supervised machine learning algorithm [3].

Logistic regression is a special case of linear regression. So before we go in how logistic regression works, we have to go through from linear regression Fig: 2.1 and for better understanding polynomial regression Fig: 2.2 with multiple degrees as well. Both in this linear and polynomial regression we try to find the best fit hypothesis which differentiates the classes properly Let's assume some data points where 'x' is the feature and 'y' is the label values. Now let's find the hyper plane which will be the best fit for these data points we assumed.

Figure 2.1: Linear Regression

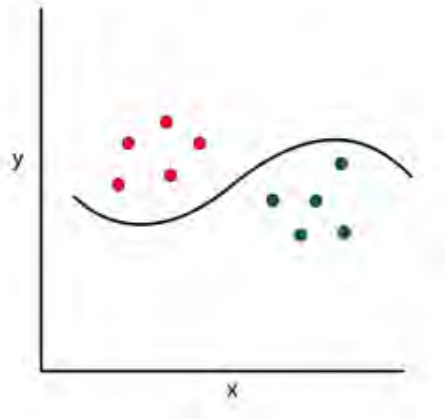

Figure 2.2: Polynomial Regression

But in real time the data points will not always be in so well organized form. This will be a problem in classification. To overcome this problem we use logistic regression which uses a sigmoid function to find out the best fit for any data points. Sigmoid function is the best function to predict probability as an outcome which is in binary format 0 or 1.The sigmoid function is given below.

$$SigmoidFunction = 1/1 + e^{-}x \tag{2.1}$$

Now let's see how our sigmoid function works in logistic regression for some random data points where 'x' will be feature and 'y' will be label as before. So with the help of the sigmoid function we come with a hypothesis which shows the best fit for this random data points in Fig: 2.3. That's how logistic regression helps in classification with independent variable with a binary outcome for nay data set.

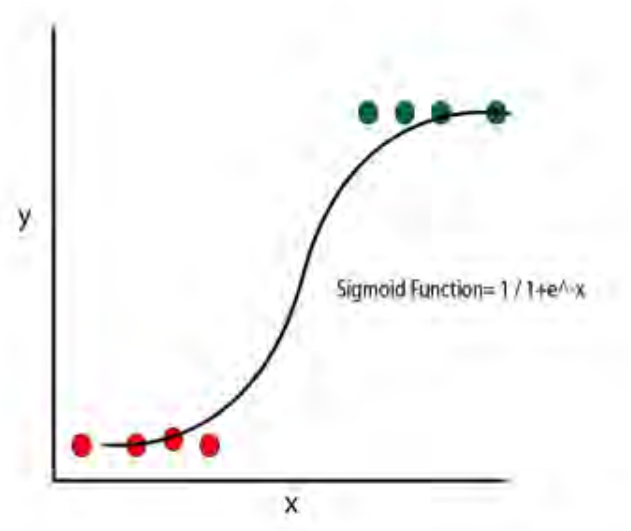

Figure 2.3: Logistic Regression

2.4 SVM

SVM is an also supervised machine learning algorithm which we can use for classification and regression problems as well but mostly used for classification. SVM [4].or the support vector Machine tries to draw a hyper plane between different classes in a way so that the hyper plane maintains the maximum distance from the closest points of the class. As we have already seen in logistic regression algorithm that we use sigmoid function to determine the hypothesis to find the best fit. But there is a problem remains, if we have some random data points shown in Fig 2.4, then what will be the exact hyper plane in this case. As we can see there is two types

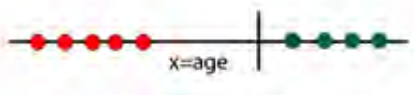

Figure 2.4: Threshold

observation, one is green and another one is red. In Fig: 2.5 we have to determine if the person is old or not. The red points are he/she is not old and the greens are he/she is, depending on his/her age is above the threshold. Here we selected the threshold based on the observations. After finishing the training when a new

observation came and as it is smaller than our observation, so we can denote it as a red observation but it won't be a good classification because this observation is much closer to the green observation than the red one.

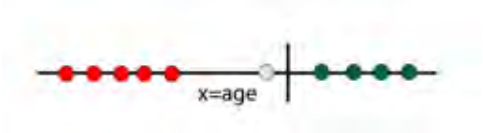

Figure 2.5: Threshold

Now we are keeping the threshold just in the middle of these two observations when it has the same distance from the last observations from each side Fig: 2.6. This process is called maximum margin classifier. So again in term of testing a new observation is here and as the age is below our threshold and closer to the red observation, we can denote it as a red observation. So we are able to classify in this case.

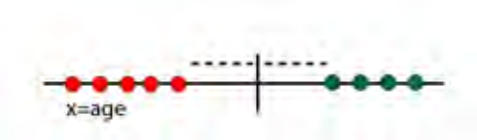

Figure 2.6: Maximum Margin Classifier

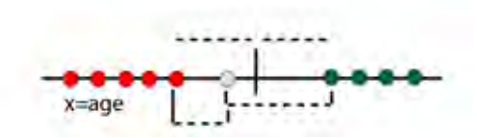

Figure 2.7: Maximum Margin Classifier

But if the training data points are like this shown in Fig: 2.8 where an observation which is denoted as red observation but it's too close to the green observations. To fix this it uses maximum margin classifier again which goes super close to green observation and keeps so far from red observation to provide a good classification. So now if there is any other observation like Fig: 2.9 then it will easily denote it properly and the give best classification.

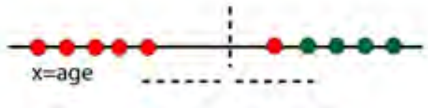

Figure 2.8: Maximum Margin Classifier

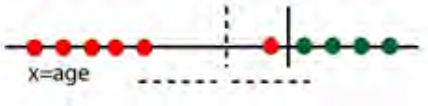

Figure 2.9: Maximum Margin Classifier

2.5 Random Forest

It's another assemble learning model that creates a set of tree structure and then using bagging to create a strong classifier out of them. Random Forest [5] models have raised its popularity significantly for some good reasons. It's so powerful and can implement quickly out of the box. This versatile machine learning model is capable of performing classification tasks and regression as well. The algorithm undertakes treats missing values, outlier values and dimensional reduction methods and it's so good in its job. It's a type of ensemble learning methods. Random Forest basically works with decision trees to determine the prediction. The trees are easy to build and use but in practice trees have an aspect that prevents that form being the ideal tool for prediction. Tree works great when it's time to create them but to classify new samples they are not so flexible. In Random Forest we make a bootstrapped data set with the random values taken from the original data set. It's completely a random process to choose the values and duplicate input is allowed in bootstrapped data set. Then we start making trees from the bootstrapped data set. We will continue the process by choosing every feature we have and will have a number of trees in hand. Now when a new sample arrives we will run it over each and every tree we made so far and count the label values. The process of bootstrapping the data and aggregate to make decisions is called Bagging. So after going through all the decision trees we made from bootstrapped data set with the sample data we surely have the label count values to determine the prediction of the end result.

2.6 Ada-boost

Ada-boost [6]is assemble learning model that use boosting to combine a bunch of weak classifier to create a strong classifier that can classify multiple class properly. It's a supervised model for classification which works with decision tree and that is the best way to use Ada-boost algorithm. There are three main concepts of Ada-boost algorithm which are related with decision trees and Random Forest algorithm. To explain these three concepts we will compare it with Random Forest algorithm here. 1. In Random Forest we create each of the tree with its full size where the size can be different but have no predetermine maximum depth .But on the other hand the trees made by Ada-boost just have one node and two leaves. This is called

stump. In a full size tree we can use all the variables to make a decision. But in stump we can use only one variable to make a decision. 2. In Random Forest each of the trees has equal vote in final classification. But in Ada-boost some stumps can have more say in than the other stumps. 3. In Random Forest each tree is independent from others, order of the tree creation is ignorable but in Ado-boost the order is a crucial thing. Because the error made by the first stump will influence the next stump creation. It's a continuous process. Now let's discuss about how the algorithm works. Assume a random data set and we have to predict the decision based on the variables in the dataset. We are giving a sample weight to every sample in the data set to represent how important it is to be well classified. At first all the samples will have same weight values which are 1 / n. where 'n' is number of samples. So now all the samples are equally important. But after making the first stump the weights values will automatically change then the weight will guide the next stump creation. Now we will take a feature and will keep count how much correctly and incorrectly it is classifying the samples. After doing it for all the features, we will use Gini Index to for all of them. Gini index mainly works with inaccuracy. The feature having the lowest Gini index value will be the first Stump. The incorrectly classified samples in first stump will get increased in weight to get more emphasize to be more accurate. We will add a normalize weight value in the data set. Then we will use weighted Gini Index which will emphasize on the maximum weight to make it more correctly classified. This process will keep continue to find the best classification.

2.7 Data Set

Overall Volume of lands in the unions of UllaparavUpozila

Union Name	Volume (hectare)
Ramkrishno	2,862
Solonga	3,425
Hatikumrul	3,472
Borohor	3,264
Ponchokroshi	2,909
Ullahpara	1,381
Pouroshova	1,270
Purnimagati	2,491
Bangala	3,294
Udhunia	3,846
BoroPangashi	3,006
Koyra	1,574
Mohonpur	3,664
Durganogor	2,609
Solop	2,421

Table 2.1: Overall Volume of lands in the unions of UllaparaUpozila

Village code number under each union

Union Name	Volume (hectare)
Ramkrishno	1-22
Solonga	23-38
Hatikumrul	39-55
Borohor	56-73
Ponchokroshi	74-86
Ullahpara	87-96
Pouroshova	97-106
Purnimagati	107-127
Bangala	128-152
Udhunia	153-175
BoroPangashi	176-193
Koyra	194-204
Mohonpur	205-223
Durganogor	224-249
Solop	250-265

Table 2.2: Overall Village code number under each union

Volume of Upper Bunot in Different Soil Group

MrittikaBunot	Volume (hectare)
Bele	186
Beledoash	1052
Doash	5720
EtelDoash	20,003
Others(Home,river,pond)	5,774

Table 2.3: Volume of Upper Bunot in Different Soil Group

Volume and Percentage of Different types of land

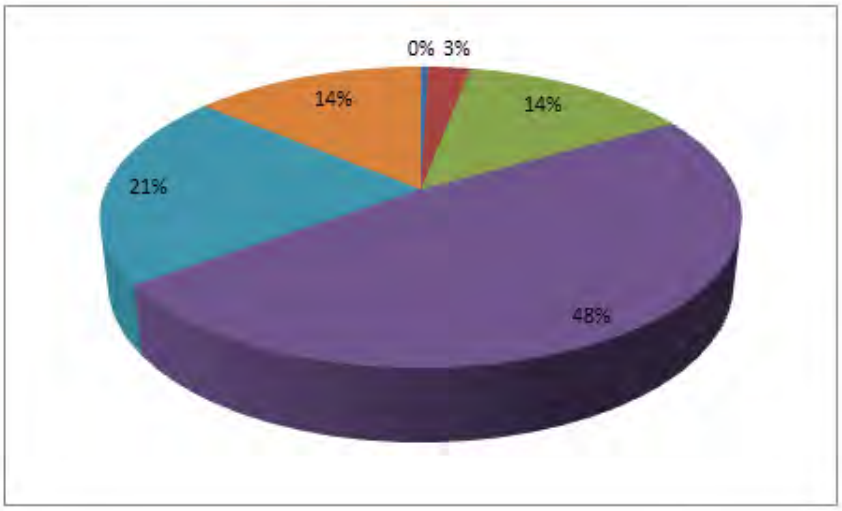

Figure 2.10: Percentage of Upper Bunot in Different Soil Group

Volume andPercentage of Different types of land

Land Type	Volume (hectare)
High Land	1,194
Medium High Land	18,342
Medium Low land	8,541
Low Land	6,404
Very Low Land	1,206
Others(Home,river,pond)	5,778

Table 2.4: Volume andPercentage of Different types of land

Volume and Percentage of soil depending on different land type

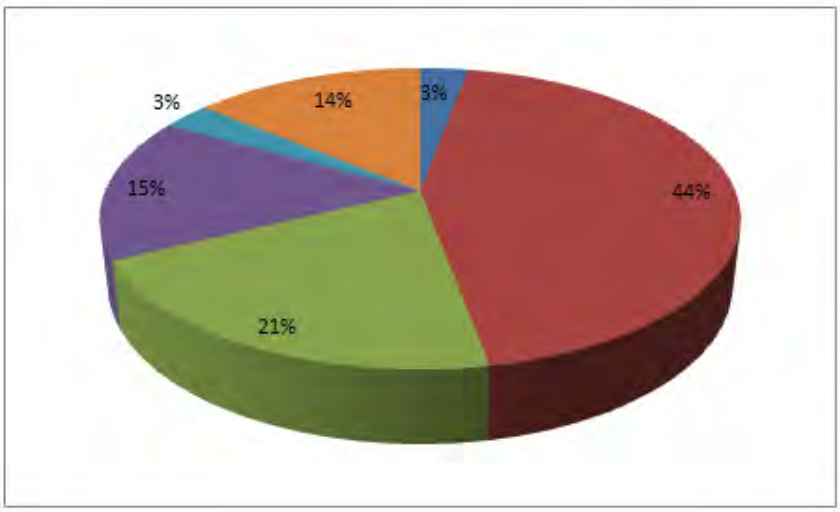

Figure 2.11: Percentage of Different types of land

2.8 Divisions of ullapara upozila

UllahPara upozila is divided into 8 regions depending on their soil types . The soil types in these areas are Jamun, GongaChora, Kaunia,Loshkora,Manda,Tista Polly soil,Tista sandy soil,Ghure,Batra. Based on these soil types the whole upozila is divided into these regions. The division of the whole upozila is shown bellow

Soil	soil Group	land type	Volume(hectare)
Tista Polol	jamun	High land	683
Tista Polol	GongaChora	Medium High Land	3,963
Tista Polol	Kaunia	Medium High Land	13,252
Tista Polol	Kaunia	Medium Low Land	5,341
Tista Polol	Kaunia	Low Land	1,410
Tista Polol	Loshkora	Medium Low Land	2,601
Tista Polol	Loshkora	Low Land	4,477
Tista Polol	Loshkora	Very Low Land	743
Tista Polol	Manda	Medium High Land	1,052
Tista Polol	Tista Polly soil	Medium High Land	75
Tista Polol	Tista Polly soil	Medium Low Land	413
Tista Polol	Tista Polly soil	Low Land	75
Tista Polol	Tista sandy soil	Medium Low Land	186
Ganga Polol	Ghure	Medium High Land	100
Ganga Polol	Batra	low land	342
Ganga Polol	Batra	very low land	463
Viti-Mati		High land	511
Others(Home,river,pond)			5774

Table 2.5: Volume and Percentage of soil depending on different land type

2.9 Region 1

This region consists of 13658 hectare lands under it. Among them 12,975 hectare land is arable here.5 percent of it are high lands nad rest 90 percent are medium high land. High lands soils are 'jamun' and medium high lands are 'GangaChora' and 'kaunia' in type.
Region 1

Union Name	Volume (hectare)
Ramkrishno	508
Solonga	1005
Hatikumrul	1465
Borohor	1271
Ponchokroshi	1073
Ullahpara	1006
Pouroshova	809
Purnimagati	1165
Bangala	771
Udhunia	19
BoroPangashi	717
Koyra	539
Mohonpur	500
Durganogor	1416
Solop	1394

Table 2.6: Region 1

2.10 Region 2

This region has 6158 hectare lands, almost all of it is arable and among these 80 percent are medium high land and rest 20 percent are medium low land. Medium high lands are 'GangaChora' and 'kaunia' and medium low lands are 'kaunia' and 'loshkora'
Region 2

Union Name	Volume (hectare)
Ramkrishno	992
Solonga	435
Hatikumrul	1260
Borohor	648
Ponchokroshi	207
Ullahpara	134
Pouroshova	165
Purnimagati	126
Bangala	660
Udhunia	592
BoroPangashi	246
Koyra	539
Mohonpur	500
Durganogor	199
Solop	490

Table 2.7: Region 2

2.11 Region 3

It contain 8389 hectare lands, all the lands are arable.80 percent of it is medium low land and 20 percent is low land.The medium low lands are 'kaunia' and 'loshkora' and the low lands are the same soil type.
Region 3

Union Name	Volume (hectare)
Ramkrishno	343
Solonga	1509
Hatikumrul	237
Borohor	191
Ponchokroshi	
Ullahpara	111
Pouroshova	
Purnimagati	756
Bangala	784
Udhunia	323
BoroPangashi	1440
Koyra	847
Mohonpur	1162
Durganogor	475
Solop	211

Table 2.8: Region 3

2.12 Region 4

It contains4952 hectare lands, all the lands are arable.15 percent of it is very low land and 85 percent is low land. The low lands are 'kaunia' and 'loshkora' and the very low lands are 'loshkora' soil type.
Region 4

Union Name	Volume (hectare)
Ramkrishno	279
Solonga	36
Hatikumrul	
Borohor	
Ponchokroshi	
Ullahpara	6
Pouroshova	83
Purnimagati	
Bangala	679
Udhunia	2594
BoroPangashi	381
Koyra	
Mohonpur	869
Durganogor	25
Solop	

Table 2.9: Region 4

2.13 Region 5

It contains 1503 hectare lands, all the lands are arable.75 percent of it is medium high land and 25 percent is medium low land. The medium low lands are 'manda' and 'tista poli-mati' and the medium low lands are 'tista poli-mati' soil type.
Region 5

Union Name	Volume (hectare)
Ramkrishno	
Solonga	
Hatikumrul	52
Borohor	629
Ponchokroshi	814
Ullahpara	
Pouroshova	3
Purnimagati	
Bangala	
Udhunia	5
BoroPangashi	
Koyra	
Mohonpur	
Durganogor	
Solop	

Table 2.10: Region 5

2.14 Region 6

It contains 373 hectare lands, 298 hectare lands are arable.60 percent of it is medium low land and 20 percent is low land. The medium low lands are 'tista bele-mati' and 'tista poli-mati' and the low lands are 'tista poli-mati' soil type.
Region 6

Union Name	Volume (hectare)
Ramkrishno	
Solonga	
Hatikumrul	67
Borohor	25
Ponchokroshi	243
Ullahpara	9
Pouroshova	5
Purnimagati	
Bangala	
Udhunia	5
BoroPangashi	
Koyra	
Mohonpur	
Durganogor	14
Solop	10

Table 2.11: Region 6

2.15 Region 7

It contains 401 hectare lands, all the lands are arable.15 percent of it is very low land and 85 percent is low land. The very low lands are 'batra' and the low lands are 'Gheur' and 'Batra' soil type.
Region 7

Union Name	Volume (hectare)
Ramkrishno	
Solonga	
Hatikumrul	
Borohor	
Ponchokroshi	
Ullahpara	
Pouroshova	
Purnimagati	
Bangala	
Udhunia	5
BoroPangashi	
Koyra	
Mohonpur	401
Durganogor	
Solop	

Table 2.12: Region 7

2.16 Region 8

It contains 5041 hectare lands, all the lands are arable.20 percent of it is low land and 80 percent is very low land. The very low lands and the low lands are both 'Batra' soil type.
Region 8

Union Name	Volume (hectare)
Ramkrishno	
Solonga	
Hatikumrul	
Borohor	
Ponchokroshi	
Ullahpara	
Pouroshova	
Purnimagati	
Bangala	
Udhunia	
BoroPangashi	
Koyra	
Mohonpur	
Durganogor	
Solop	504

Table 2.13: Region 8

2.17 Train and Test Data

These are some sample of our train and test data .Nitrogen, Phosphorus, Potassium for some particular soil types is given below. Here "Bunot" 1, 2,3,4,5 is Doash, Etel Doash, Etel, Bele doash and Bele type. "h" represents High, "vh" represents Very High,"m" represents Medium,"mo" represents Moderate, l represents Low, "vl" represents Very Low.

MrittikaDol	land Types	Bunot	N	P	K	NPK for Upland crops	NPK for Wetland crops
Jamun	h	1	0.11	9.85	0.5	l-l-l	l-m-l
Jamun	h	1	0.1	22.36	0.11	l-l-l	l-m-l
Jamun	h	1	0.09	5.07	0.07	l-l-l	l-m-l
Jamun	h	1	0.1	12.74	0.08	l-l-l	l-m-l
Jamun	h	1	0.09	13.22	0.1	l-l-l	l-m-l
Jamun	h	1	0.07	10.16	0.06	l-l-l	l-m-l
Jamun	h	1	0.13	3.75	0.09	l-l-l	l-m-l
Jamun	h	1	0.1	3.97	0.1	l-l-l	l-m-l
Jamun	h	1	0.12	7.28	0.19	l-l-l	l-m-l
Jamun	h	1	0.1	8.16	0.1	l-l-l	l-m-l
Jamun	h	1	0.09	4.15	0.08	l-l-l	l-m-l
Jamun	h	1	0.11	2.76	0.08	l-l-l	l-m-l
Jamun	h	1	0.1	1.17	0.06	l-l-l	l-m-l
Jamun	h	1	0.11	3.31	0.07	l-l-l	l-m-l
GongaChora	h	1	0.08	2.55	0.07	l-l-l	l-mo-l
GongaChora	mh	1	0.12	3.44	0.11	l-m-l	l-mo-l
GongaChora	mh	1	0.11	8.88	0.12	l-m-l	l-mo-l
GongaChora	mh	1	0.12	21.93	0.15	l-m-l	l-mo-l
GongaChora	mh	1	0.1	23	0.13	l-m-l	l-mo-l
GongaChora	mh	1	0.14	2.86	0.1	l-m-l	l-mo-l
GongaChora	mh	1	0.14	8.36	0.08	l-m-l	l-mo-l
GongaChora	mh	1	0.11	8.42	0.09	l-m-l	l-mo-l
GongaChora	mh	1	0.11	9.27	0.01	l-m-l	l-mo-l
GongaChora	mh	1	0.12	8.3	0.09	l-m-l	l-mo-l
GongaChora	mh	1	0.14	9.41	0.09	l-m-l	l-mo-l
GongaChora	mh	1	0.14	10.68	0.12	l-m-l	l-mo-l
GongaChora	mh	1	0.14	44.84	0.1	l-m-l	l-mo-l
GongaChora	mh	1	0.14	46.46	0.1	l-m-l	l-mo-l
GongaChora	mh	1	0.11	19.6	0.07	l-m-l	l-mo-l
GongaChora	mh	1	0.1	8.52	0.12	l-m-l	l-mo-l
GongaChora	mh	1	0.14	10.99	0.13	l-m-l	l-mo-l
GongaChora	mh	1	0.14	9.71	0.11	l-m-l	l-mo-l
GongaChora	mh	1	0.1	13.97	0.07	l-m-l	l-mo-l
GongaChora	mh	1	0.14	20.23	0.8	l-m-l	l-mo-l
Kaunia	mh	2	0.12	5.25	0.08	l-l-l	l-l-l
Kaunia	mh	2	0.14	1.79	0.08	l-l-l	l-l-l

Figure 2.12: Train and test Data

Kaunia	mh	2	0.13	1.96	0.07	l-l-l	l-l-l
Kaunia	mh	2	0.14	2.19	0.08	l-l-l	l-l-l
Kaunia	mh	2	0.14	3.75	0.12	l-l-l	l-l-l
Kaunia	mh	2	0.1	1.88	0.13	l-l-l	l-l-l
Kaunia	mh	2	0.08	1.21	0.06	l-l-l	l-l-l
Kaunia	mh	2	0.07	1.29	0.08	l-l-l	l-l-l
Kaunia	mh	2	0.09	10.16	0.07	l-l-l	l-l-l
Kaunia	mh	2	0.13	10.39	0.12	l-l-l	l-l-l
Kaunia	mh	2	0.14	7.4	0.12	l-l-l	l-l-l
Kaunia	mh	2	0.1	22.36	0.11	l-l-l	l-l-l
Kaunia	mh	2	0.08	10.36	0.09	l-l-l	l-l-l
Kaunia	mh	2	0.09	10.93	0.08	l-l-l	l-l-l
Kaunia	mh	2	0.09	9.93	0.08	l-l-l	l-l-l
Kaunia	mh	2	0.13	3.42	0.11	l-l-l	l-l-l
Kaunia	mh	2	0.12	7.89	0.07	l-l-l	l-l-l
Kaunia	mh	2	0.12	10.7	0.07	l-l-l	l-l-l
Kaunia	mh	2	0.12	6.07	0.11	l-l-l	l-l-l
Kaunia	mh	2	0.09	7.16	0.2	l-l-l	l-l-l
Kaunia	mh	2	0.1	1.81	0.13	l-l-l	l-l-l
Kaunia	mh	2	0.1	1.82	0.11	l-l-l	l-l-l
Kaunia	mh	2	0.01	1.98	0.11	l-l-l	l-l-l
Kaunia	mh	2	0.08	1.2	0.12	l-l-l	l-l-l
Kaunia	mh	2	0.07	6.43	0.15	l-l-l	l-l-l
Kaunia	mh	2	0.13	12.02	0.13	l-l-l	l-l-l
Kaunia	mh	2	0.14	11.28	0.1	l-l-l	l-l-l
Kaunia	mh	2	0.09	9.46	0.11	l-l-l	l-l-l
Kaunia	mh	2	0.07	8.37	0.13	l-l-l	l-l-l
Kaunia	mh	2	0.09	17.4	0.09	l-l-l	l-l-l
Kaunia	mh	2	0.09	15.07	0.08	l-l-l	l-l-l
Kaunia	mh	2	0.08	14.91	0.08	l-l-l	l-l-l
Kaunia	mh	2	0.14	4.12	0.1	l-l-l	l-l-l
Kaunia	mh	2	0.13	4.63	0.1	l-l-l	l-l-l
Kaunia	mh	2	0.11	3.36	0.08	l-l-l	l-l-l
Kaunia	mh	2	0.08	9.45	0.13	l-l-l	l-l-l
Kaunia	mh	2	0.09	14.96	0.07	l-l-l	l-l-l
Kaunia	mh	2	0.08	19.07	0.06	l-l-l	l-l-l
Kaunia	mh	2	0.14	6.3	0.11	l-l-l	l-l-l
GongaChora	mh	1	0.12	5.72	0.07	l-l-l	l-l-l
GongaChora	mh	1	0.12	6.61	0.07	l-l-l	l-l-l

Figure 2.13: Train and test Data

GongaChora	mh	1	0.13	8.16	0.07	l-l-l	l-l-l
GongaChora	mh	1	0.11	7.06	0.1	l-l-l	l-l-l
GongaChora	mh	1	0.1	3.66	0.11	l-l-l	l-l-l
GongaChora	mh	1	0.09	8.19	0.19	l-l-l	l-l-l
GongaChora	mh	1	0.12	7.17	0.08	l-l-l	l-l-l
GongaChora	mh	1	0.1	5.53	0.08	l-l-l	l-l-l
GongaChora	mh	1	0.11	8.31	0.09	l-l-l	l-l-l
Kaunia	mh	2	0.14	11.45	0.08	l-l-l	l-l-l
Kaunia	mh	2	0.14	10.35	0.07	l-l-l	l-l-l
Kaunia	mh	2	0.14	12.74	0.07	l-l-l	l-l-l
Kaunia	mh	2	0.12	6.87	0.12	l-l-l	l-l-l
Kaunia	mh	2	0.12	5.31	0.1	l-l-l	l-l-l
Kaunia	mh	2	0.09	3.93	0.11	l-l-l	l-l-l
Kaunia	mh	2	0.11	3.63	0.12	l-l-l	l-l-l
Kaunia	mh	2	0.11	4.6	0.1	l-l-l	l-l-l
Kaunia	mh	2	0.14	2.43	0.13	l-l-l	l-l-l
Kaunia	mh	2	0.14	2.02	0.1	l-l-l	l-l-l
Kaunia	mh	2	0.13	3.15	0.14	l-l-l	l-l-l
Kaunia	mh	2	0.12	2.37	0.11	l-l-l	l-l-l
Kaunia	mh	2	0.13	3.13	0.15	l-l-l	l-l-l
Kaunia	ml	2	0.12	3.92	0.12	l-l-l	l-l-l
Kaunia	ml	2	0.13	5.09	0.08	l-l-l	l-l-l
Kaunia	ml	2	0.13	10.88	0.1	l-l-l	l-l-l
Kaunia	ml	2	0.13	4.38	0.1	l-l-l	l-l-l
Kaunia	ml	2	0.12	7.1	0.07	l-l-l	l-l-l
Kaunia	ml	2	0.13	9.24	0.1	l-l-l	l-l-l
Kaunia	ml	2	0.09	10.54	0.17	l-l-l	l-l-l
Loshkora	ml	3	0.09	6.49	0.06	l-l-l	l-l-l
Loshkora	ml	3	0.12	7.63	0.07	l-l-l	l-l-l
Loshkora	ml	3	0.13	7.43	0.25	l-l-l	l-l-l
Kaunia	ml	2	0.07	3.47	0.07	l-l-l	l-m-l
Kaunia	ml	2	0.09	13.47	0.11	l-l-l	l-m-l
Kaunia	ml	2	0.08	4.36	0.08	l-l-l	l-m-l
Kaunia	ml	2	0.07	3.1	0.07	l-l-l	l-m-l
Kaunia	ml	2	0.12	12.24	0.12	l-l-l	l-m-l
Kaunia	ml	2	0.12	4.57	0.1	l-l-l	l-m-l
Kaunia	ml	2	0.1	3.85	0.1	l-l-l	l-m-l
Kaunia	ml	2	0.11	11.89	0.14	l-l-l	l-m-l
Kaunia	ml	2	0.1	8.66	0.12	l-l-l	l-m-l

Figure 2.14: Train and test Data

Kaunia	ml	2	0.12	8.36	0.12	l-l-l	l-m-l
Kaunia	ml	2	0.12	7.38	0.23	l-l-l	l-m-l
Kaunia	ml	2	0.1	8.69	0.13	l-l-l	l-m-l
Kaunia	ml	2	0.09	6.35	0.11	l-l-l	l-m-l
Kaunia	ml	2	0.08	18.71	0.09	l-l-l	l-m-l
Kaunia	ml	2	0.08	21.18	0.1	l-l-l	l-m-l
Kaunia	ml	2	0.12	9.46	0.12	l-l-l	l-m-l
Kaunia	ml	2	0.11	7.4	0.13	l-l-l	l-m-l
Kaunia	ml	2	0.11	8.82	0.13	l-l-l	l-m-l
Kaunia	ml	2	0.07	17.26	0.09	l-l-l	l-m-l
Kaunia	ml	2	0.13	9.45	0.14	l-l-l	l-m-l
Kaunia	ml	2	0.13	10.16	0.15	l-l-l	l-m-l
Kaunia	ml	2	0.11	10.37	0.18	l-l-l	l-m-l
Kaunia	ml	2	0.08	11.63	0.1	l-l-l	l-m-l
Kaunia	ml	2	0.09	10.5	0.11	l-l-l	l-m-l
Loshkora	ml	3	0.11	4.08	0.19	l-l-l	l-m-l
Loshkora	ml	3	0.11	4.33	0.12	l-l-l	l-m-l
Loshkora	ml	3	0.13	5.16	0.14	l-l-l	l-m-l
Loshkora	ml	3	0.12	3.71	0.13	l-l-l	l-m-l
Loshkora	ml	3	0.14	11.16	0.15	l-l-l	l-m-l
Loshkora	ml	3	0.14	8.7	0.14	l-l-l	l-m-l
Loshkora	ml	3	0.12	7.64	0.12	l-l-l	l-m-l
Loshkora	ml	3	0.13	8.27	0.14	l-l-l	l-m-l
Loshkora	ml	3	0.13	10.59	0.15	l-l-l	l-m-l
Loshkora	ml	3	0.12	11.52	0.15	l-l-l	l-m-l
Loshkora	ml	3	0.13	9.27	0.12	l-l-l	l-m-l
Loshkora	ml	3	0.14	8.17	0.19	l-l-l	l-m-l
Loshkora	ml	3	0.13	9.09	0.14	l-l-l	l-m-l
Loshkora	ml	3	0.1	6.92	0.14	l-l-l	l-m-l
Loshkora	ml	3	0.1	6.88	0.12	l-l-l	l-m-l
Kaunia	l	2	0.1	9.69	0.12	l-l-l	l-m-m
Kaunia	l	2	0.12	9.86	0.2	l-l-l	l-m-m
Loshkora	l	3	0.14	10.72	0.14	l-m-l	l-m-l
Loshkora	l	3	0.14	9.39	0.15	l-m-l	l-m-l
Loshkora	l	3	0.14	12.04	0.11	l-m-l	l-m-l
Loshkora	l	3	0.12	9.53	0.17	l-m-l	l-m-l
Loshkora	l	3	0.11	10.92	0.12	l-m-l	l-m-l

Figure 2.15: Train and test Data

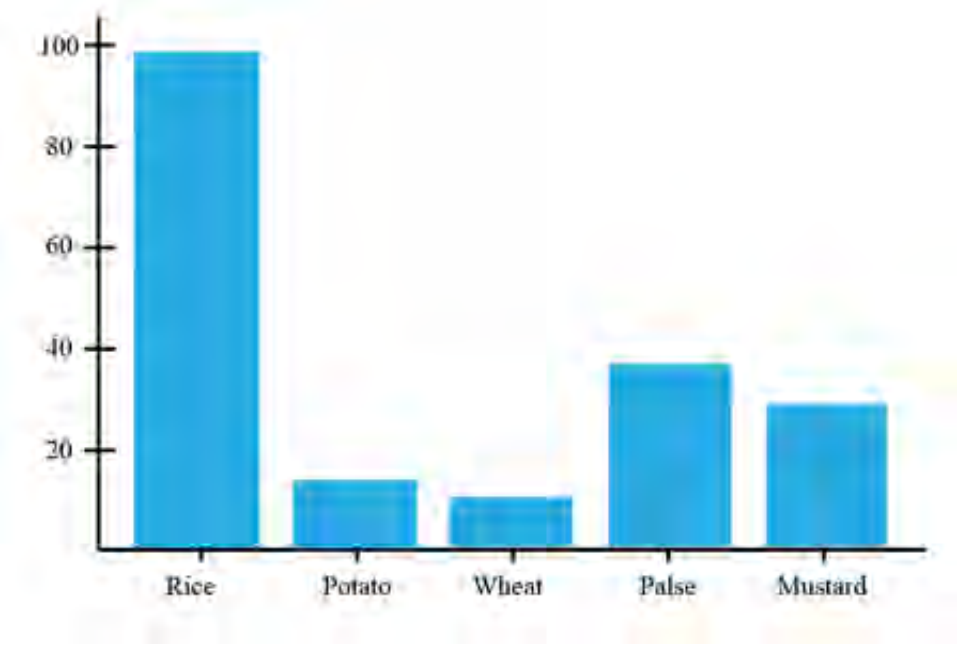

Figure 2.16: Distribution of crops are shown by a graph given above

Chapter 3

Literature Review

We have gone through some research, journals, conference paper and project papers to understand the concept and process for N.P.K. Based Crop Suggesting Model. Some of the existing papers and related works regarding crop suggesting model are as follows.

Crop Recommendation and Fertilizer Purchase System [7] a project done by some students of Computer Department, Marathwada Mitra Mandal's College Of Engineering, Pune,and Maharashtra, India. In their project they make an android app which can be used by the farmer through his mobile phone. They tried to recommend crops and as well as the required fertilizer for cultivating that particular crop. They suggested a term for the farmers named "Crop Rotation" which means to cultivate crops in rotation. Cultivating the same crop again and again cause damage to the nutrient values of any particular land. So the suggested to grow different sort of crops in a regular interval cause it is so essential to keep up the nutrient value of that soil. To recommend a crop they tried with 3 different algorithms to find the best accuracy and efficiency. At first they used The Naive Bayes algorithm which has 40 percent efficiency, then they used the ID3 algorithm which has 60 percent and finally the Random forest algorithm which turned out the best algorithm to sort out this problem and suggest the best crop to the farmers. For the fertilizer recommendation part they used an algorithm which is basically a data mining algorithm for suggestion of frequently purchased item sets.

Jeetendra Sheenoy, Prof. YogeshPingle [8] this paper focusing on the solution of reducing the transportation cost and other middle agents between the farmers and the end users by using an IOT based approach. That helped farmers in many ways to earn their desired profit by cultivating any particular crops. Reducing the middle hops and agents helped farmers to get the price what they have calculated already with per unit production. This paper helped us a lot in planning the economic part of our project. In our project we will suggest crops to the farmers considering their economic conditions as well. Transportation cost will be included in production investment before a farmer can sell the crop. So less production cost will always lead to earn a good profit if the middle hops are reduced and the farmers get the fair amount of selling price. This paper became the motivation of our project. We incorporate the mechanisms used in this paper to suggest the best crop with maximum profit.

Monali Paul, Santosh K. Vishwakarma, Ashok Verma [9] In this paper in order to predict the best suitable crop for any particular soil type they actually analyzed the

crop and based on the analysis they categorized the crops of a particular area. They used data mining process to categorize the crops. Moreover this paper focused on various classification methods such as Naive Bayes, K-Nearest Neighbor methods to ensure proper classification of the crops. Using this paper we analyze the classification methods and determined which classification will be appropriate for the data set we are using in our project.

Abdullah Na, William Isaac,ShashankVarshney, EkramKhan [10] this paper is about a smart phone based application which can measure PH value of the soil, temperature , humidity in real time .They have used many blocks like micro controller block, sensing block and communication block. The sensor is employed in farms which can communicate with a smartphone via Bluetooth. This paper provides the facility of remote analysis of soils through various techniques. It encouraged us in looking for different techniques by which we can transfer the data obtained from sensors for processing and eventually generating the best output.

S.Nagini, Dr. T. V. RajiniKanth, B.V.Kiranmayee [11] this paper discussed about various data analysis process and prediction models. They took a simple set of data and put apply many regression techniques in order to analyze and identify the properties of each of it. The various regression techniques this paper focused on are Linear, Multiple Linear, non-Linear, Logistic, Polynomial and Ridge regression. We came to know about many comparative data analytic algorithm after going through this paper. This paper helped us to make judgment about which algorithm best suits our proposed system.

Awanit Kumar,Shiv Kumar [12]in this paper they proposed a system of prediction of the production of a particular crop. In order to predict the crop they used a data mining algorithm named K-Means .It also used a prediction mechanism in from of Fuzzy Logic. Its rule based prediction logic where a set of rules are applied on any land for farming, rainfall and production of crops. This paper give a clear concept about how K-Means can be used to analyze data sets is obtained.We got a clear idea of applying the set of rules to predict which crop maximum the profit based on previous years will cost of crops and current soil and weather data.

The paper [13] discussed about precise farming and what are the requirements are and planning needed to build a successful software model to ensure precise farming even for a small farmer. In this paper precise farming is deeply analyzed from the very basic level and then developed a model which will work. The model they are proposing that is Precise Agriculture (PA).The main goal of this model is to deliver advisory service even to the smallest farmer with the smallest amount of land he has via mostly used technologies like sms or email. The model they are proposing designed for Kerala state which is in India but also can be applicable for the rest parts of India by bringing some small changes.

The paper [14] emphasize on crop yield prediction and how what's the necessity of it. According to the author prediction of crop yield can be so much helpful for a country economy in agricultural sectors. He developed a frame work named extensible Crop Yield Prediction Framework (XCYPF) which offers flexible inclusion different process regarding crop yield prediction. To help people predicting the crop yields of any particular crop with dependable or independent variables a tool was also developed.

The paper [15] discuss about the necessity of crop selection and about the factors deciding the best crop for any particular land. The main factors of selecting any crops

are production cost, government policies and market price. The author proposed a model named Crop Suggesting Model (CMS) which will suggest the best crop to cultivate on a particular land depending on some criteria like session, weather, water fall, soil type, crop type etc. The accuracy of this model depends on the predicted value of the parameter. The paper [16] discoursed about different kinds of classification method in order to classify the liver disease data set.it focused on accuracy of data set because efficiency and accuracy depends directly on the data set and the learning algorithms. This paper discussed about many algorithms like J48, Naive Bayes, ANN, ZeroR, 1BK and VFI and also compares correction rate and efficiency among them. All the algorithms where judged on the basis of accuracy and completing time and come to an end with a conclusion that NaviseBaise algorithm shows the improved prediction and Multi layer perception shows the most accuracy among all the algorithms that is proposed

The paper [17]deeply emphasized on crop prediction which plays a vital role in agriculture. The main reason that is mentioned behind not having enough productivity is that he farmers are not aware of the prediction before cultivating any crop. It offers a solution which is named Precise Agriculture model. This model is highly accurate and has great efficiency. The model use four learners to predict the best suitable crop for a particular land based on parameters of soils. The learners of this Precise Agriculture (PA) models are random tree, CH-AID, K-Nearest Neighbor and Naive Bayes which will recommend the best suitable crop.

This paper [18] is about the broad implication of crop yield prediction in country ecology, economy and in human welfare as well. As there are a huge number of factors that is responsible for a crop growth, so predicting the suitable crop gets difficult. The author discussed about Forecasting Yield Prediction and the difficulties of it because it determines the future crop yield based on the previous condition. As crop yield depends on various factor so it gets more challenging. Two types of modeling method are mentioned here in the paper which is the Statistical and the process model. The purpose of statistical model is to define a relationship based on previous data and use the results in future prediction of crop yield. The process models mentioned here are the main fundamental models of predicting crop yield because the incorporate knowledge of specific characteristics of any particular plant also keeps in consideration of some environmental variables such as soil properties. Process models can be helpful because they work on the basis of known principles and predict the best suitable crop and can explain clearly how different kind of variables can affect a crop yield. This paper mentioned about some process models such as CERES-Maize model, CROPGRO-soybean model, GAEZ63 model and SALUS model but the models require a large number of data which are often unavailable.

This paper [19] is completely a crop and soil research paper in which the author discussed about eleven types of crop simulation models which can be used in prediction of crop yield. These models were tested for three years based on various nitrogen fertilizer rates. The widely used 11 crop simulations model are APSIM, CERES, CROPSYST, COUP, DAISY, EPIC, FASSET, HERMES, MONICA, STICS and WOFOST. This is the best paper to make a standard crop model inter comparison under different level of nitrogen supply. From this paper we came to know about various kinds of crop simulation models which can be used in predicting the most suitable crop for any particular land type. It also helped us to know about the advantages and limitations of those crop simulation models discussed in this paper

and make the work easy for us to choose the best models among these and to come up with the maximum and accuracy and efficiency level in predicting crops for the farmer's welfare.

We have gone through a book [20] published by Soil and Research Development Institute (SRDI) in Bangladesh. They have branches in almost every district in our country. They are mainly assigned to look after the overall production of crop, research on the nutrient values in all arable lands in Bangladesh and to develop better idea for increasing the nutrient values of the soil to make it more productive. The have the data of every single district exists in our country and data they have are very well organized on the basis of soil types and its nutrient values. Their working process is very much attractive as they divide a particular union or upozila in a format of Manchitro Ekok depending on the soil type and soil groups which make the data more specific to the end user. They collect all the soil samples from a particular place and analyze it and find out the main nutrient values of that particular soil suck as nitrogen, phosphorus, potassium, zinc, magnesium and so on which are mainly responsible for plant growth. The have the same necessary nutrient values for any crops which can be cultivated on that particular place. They proposed a Fertilizer Suggesting Model in this book. Which takes a number of inputs from the user and cross match it with the database the created and suggest fertilizer for a particular crop shortly. This book gave us a clear concept about the project we are working on. The data that this book includes in is very much helpful for us in making a project based on n, p, and k values to predict the best suitable crops.

Chapter 4

Proposed Model

The flowchart shown above represents our overall working process throughout the system. First we gathered the N.P.K values of the soil samples exists in that area .Then we collected the required N.P.K values to grow any particular crop with a good production rate and efficiency. After gathering all the N.P.K values to determine the best suitable crops we used Euclidean Distance method .In which the Lowest Euclidean Distance is considered as the best suitable crop for any particular crop in any sort of soil samples. We put the lowest Euclidean distance as 'y' the output of our system and the input we are taking from the user is our 'x' values respectively. We have taken 70 percent of our data sample as train data and rest 30 percent is our test data. We used several algorithms like SVM, Ada-boost, Random Forest and Logistic Regression, to make the best prediction about the most suitable crop in a particular soil type.

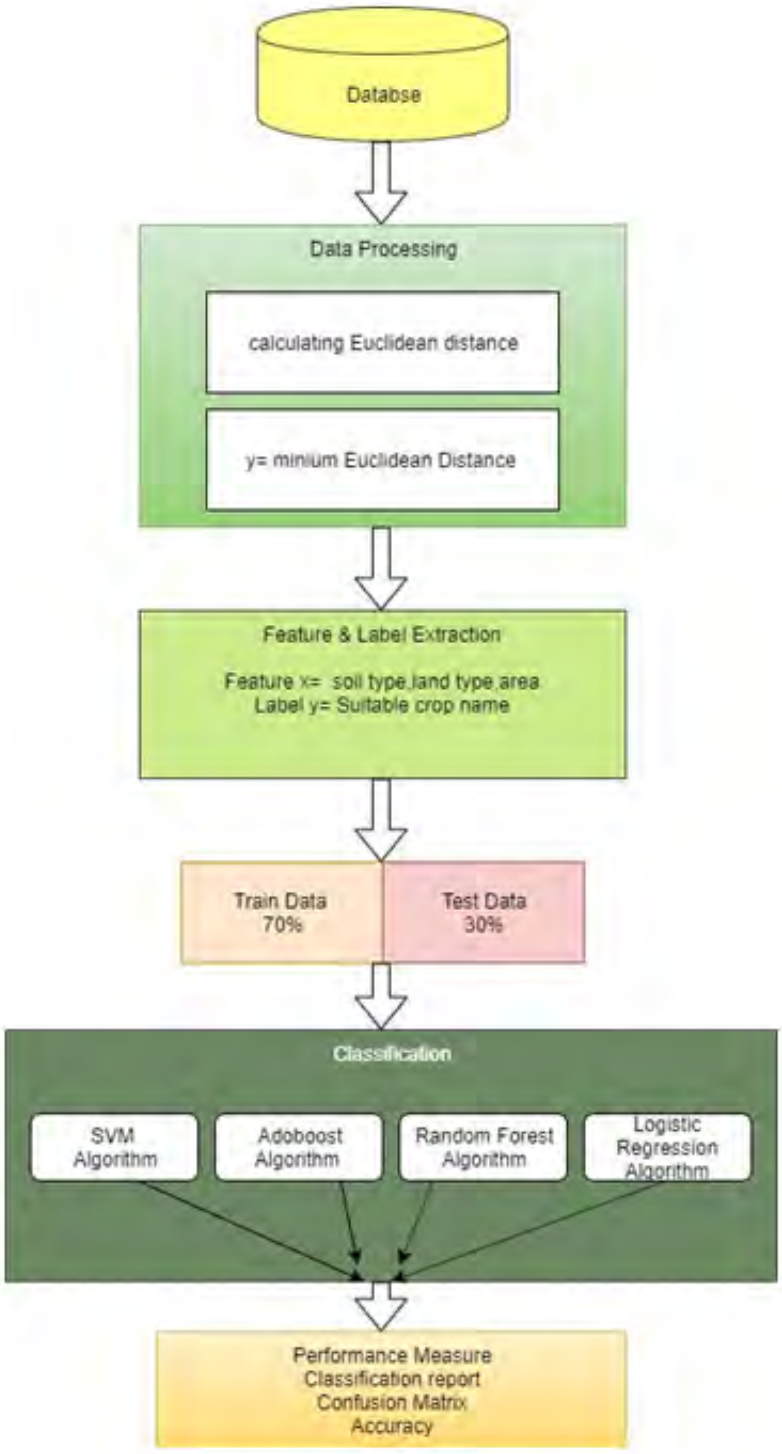

Figure 4.1: Proposed Model

Chapter 5

Implementation and Result Analysis

We implemented several algorithms on our data set. Those are SVM, Ada-boost, Random Forest, Logistic Regression, and Neural Network algorithms which have given different precision, recall, f1-score, and support values with different level of accuracy. At first let's focus on the formula of precision, recall, f1-score, support and how to determine the values

5.1 Precision and Recall Formula

The fraction of relevant instances among the retrieved instances is called precision also known as positive predictive value in pattern recognition, information retrieval and classification where recall known as sensitivity. The percentage of total relevant results correctly classified by your algorithm is referred by Recall. To find the precision and recall ratio we use the formula shown bellow

$$Precision = TruePositive/(TruePositive + FalsePositive) \tag{5.1}$$

$$Recall = TruePositive/(TruePositive + FalseNegative) \tag{5.2}$$

5.2 F1-score Formula

Combining precision and recall relative to a specific positive class is done by F1 score .it can be interpreted as a weighted average of the precision and recall.the F1 score (F-score/ F measure) is a measure of a test's accuracy In statistical analysis of binary classification.to find the F1 score we are using this formula

$$F1 - score = (Precision.Recall)/(Precision + Recall) \tag{5.3}$$

5.3 Confusion Matrix

Confusion matrix Fig:5.1 is used in describing the performance of a classification model on the set of test data in which there exist unknown true values.

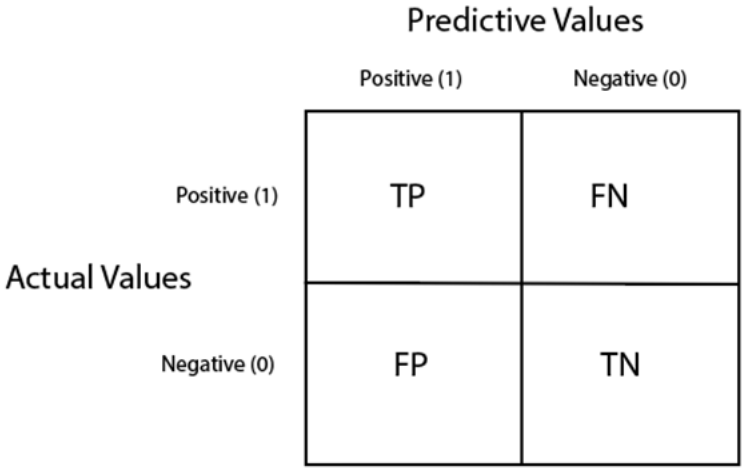

Figure 5.1: Confusion Matrix

The implementation and outcomes of these algorithms are discussed below.

5.4 Result Analysis

SVM Algorithm

	Precision	recall	f1-score	support
Dhan	1.00	1.00	1.00	30
Alu	1.00	1.00	1.00	4
Gom	0.67	1.00	0.80	2
Mugh	1.00	0.91	0.95	11
Shorisha	1.00	1.00	1.00	11
Accuracy			0.98	58
Macro avg	0.93	0.98	0.95	58
Weighted avg	0.99	0.98	0.98	58

Figure 5.2: SVM Algorithm

The accuracy of this algorithm is 98 percent and the precision, recall, f1-score and support values are shown below.

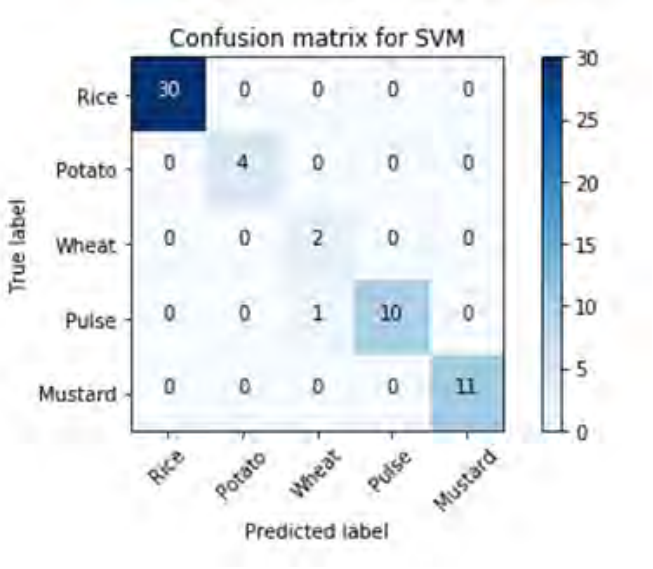

Figure 5.3: SVM Confusion Matrix

Ada-boost Algorithm

The accuracy of this algorithm is 91 percent and the precision, recall, f1-score and support values are shown below

	Precision	recall	f1-score	support
Dhan	0.94	1.00	0.97	30
Alu	1.00	1.00	1.00	7
Gom	1.00	1.00	1.00	1
Mugh	1.00	0.50	0.67	10
Shorisha	0.77	1.00	0.87	10
Accuracy			0.91	58
Macroavg	0.94	0.90	0.90	58
Weightedavg	0.93	0.91	0.90	58

Figure 5.4: Ada-boost Algorithm

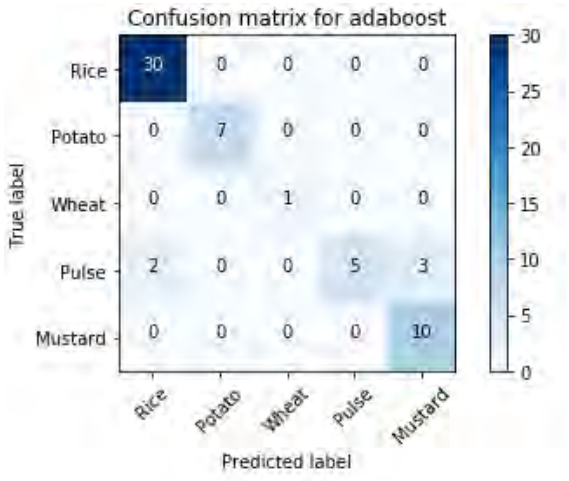

Figure 5.5: Ada-boost Confusion Matrix

Random Forest Algorithm

The accuracy of this algorithm is 98 percent and the precision, recall, f1-score and support values are shown below.

	Precision	recall	f1-score	support
Dhan	0.92	1.00	0.96	24
Alu	1.00	1.00	1.00	6
Gom	0.67	0.50	0.57	4
Mugh	1.00	0.92	0.96	13
Shorisha	1.00	1.00	1.00	11
Accuracy			0.95	58
Macroavg	0.92	0.88	0.90	58
Weighted avg	0.95	0.95	0.94	58

Figure 5.6: Random Forest Algorithm

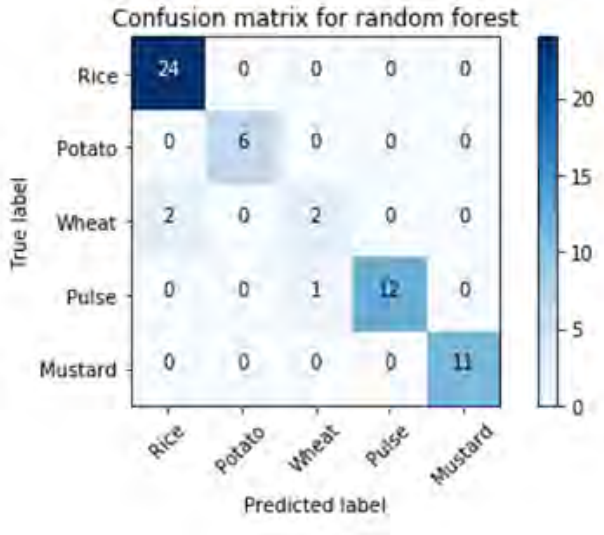

Figure 5.7: Random Forest Confusion Matrix

Logistic Regression Algorithm

The accuracy of this algorithm is 88 percent and the precision, recall, f1-score and support values are shown above

	Precision	recall	f1-score	support
Dhan	0.91	1.00	0.95	30
Alu	1.00	1.00	1.00	5
Gom	0.00	0.00	0.00	2
Mugh	1.00	0.62	0.76	13
Shorisha	0.73	1.00	0.84	8
Accuracy			0.88	58
Macroavg	0.73	0.72	0.71	58
Weighted avg	0.88	0.88	0.87	58

Figure 5.8: logistic regression Algorithm

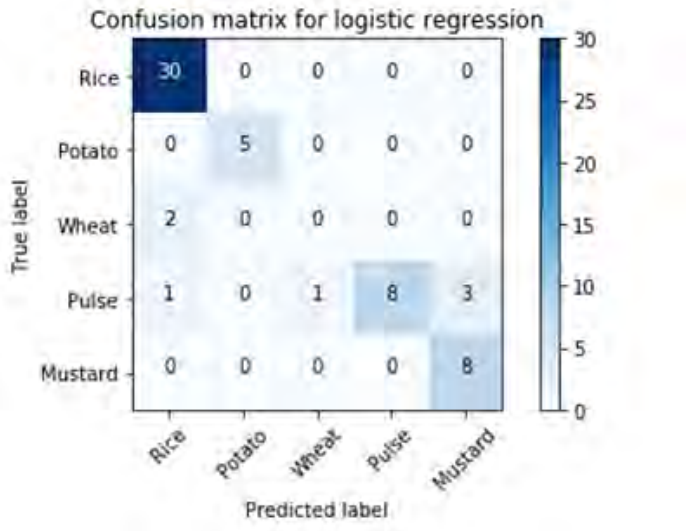

Figure 5.9: logistic regression Confusion matrix

5.5 Accuracy of the Algorithms

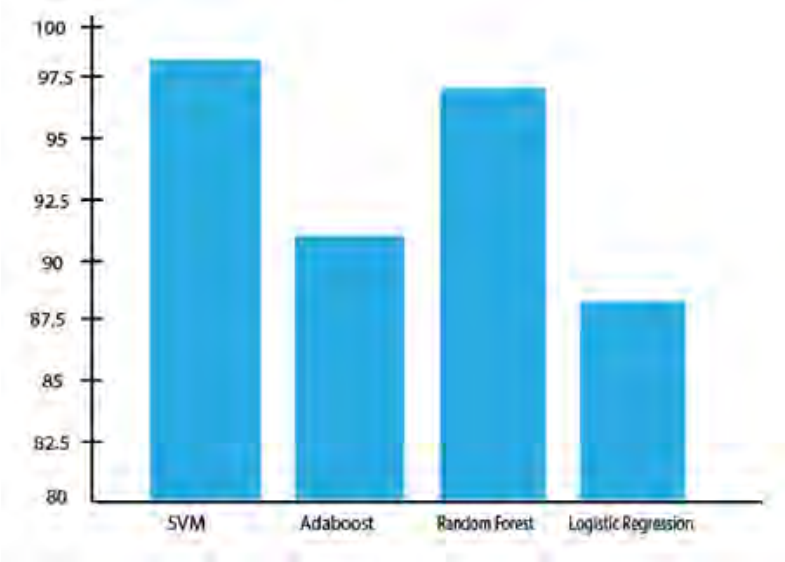

Figure 5.10: Accuracy of the Algorithms

As we can see in the accuracy graph that the SVM algorithm shows the best accuracy and the logistic regression algorithm shows the minimum accuracy. The accuracy of the SVM algorithm is almost 98 percent and the lowest accuracy is almost 88 percent which is logistic regression. The Ada-boost algorithm shows 91 percent accuracy on our data set

Chapter 6

Conclusion

To conclude the paper we are presenting is a model of "N.P.K Based Crop Suggesting model" for the farmers or the people related to agricultural sectors. We already mentioned about that the system, which would be very helpful for people even for the illiterate people. It aims to help both the less educated and educated people by giving assistance to grow the best suitable crops in a particular land. So the model is not only a model for mechanism but also effective and helpful for various purposes. As we are already in a shortage of arable lands, our project will reduce the chances of growing the wrong crop in a land without judging the overall outcome and will increase the possibility of producing exactly the most suitable crop which will enhance the overall production and outcome will be as expected The recognition results produced by our system showed more than 85 percent every time, which is satisfactory. So we believe our system will be a effective one for the end users

6.1 Future Work

As we already mentioned that we are working on a particular upozila which is UllahPara and we come up with a good accuracy and best prediction of crops. So we can easily elaborate our system by increasing our data set consists of the rest of the divisions, zila ,upozila. We can apply the same formula or procedure for the rest of the areas and can determine the best suitable crops for any particular land in whole Bangladesh.

6.2 Limitations

There are some limitations in our system. These are the amount of sample was not adequate enough to accommodate the number of features we have. We have 5 features and the data sample was almost 192.So the system could work more efficiently without this limitation. Class distribution of data set was extremely diverse, as we know our land is mostly suitable for cultivating rice, so the majority of the class distribution was taken by only this. The variance of class distribution may produce some biased result.

Bibliography

[1] N. Stern, *A strategy for development*. The World Bank, 2002.

[2] Q.-z. GU, X.-y. YANG, B.-h. SUN, and Z.-c. GENG, "Analyses of main nutrients' concentration and its accumulation characteristics in vegetable greenhouse soils [j]", *Journal of Northwest A & F University (Natural Science Edition)*, vol. 3, 2008.

[3] J. M. Hilbe, *Logistic regression models*. Chapman and hall/CRC, 2009.

[4] H. Drucker, D. Wu, and V. N. Vapnik, "Support vector machines for spam categorization", *IEEE Transactions on Neural networks*, vol. 10, no. 5, pp. 1048–1054, 1999.

[5] M. Pal, "Random forest classifier for remote sensing classification", *International Journal of Remote Sensing*, vol. 26, no. 1, pp. 217–222, 2005.

[6] R. E. Schapire, "Explaining adaboost", in *Empirical inference*, Springer, 2013, pp. 37–52.

[7] M. Shinde, K. Ekbote, S. Ghorpade, S. Pawar, and S. Mone, "Crop recommendation and fertilizer purchase system", *International Journal of Computer Science and Information Technologies*, vol. 7, no. 2, pp. 665–667, 2016.

[8] J. Shenoy and Y. Pingle, "Iot in agriculture", in *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, IEEE, 2016, pp. 1456–1458.

[9] M. Bendre, R. Thool, and V. Thool, "Big data in precision agriculture: Weather forecasting for future farming", in *2015 1st International Conference on Next Generation Computing Technologies (NGCT)*, IEEE, 2015, pp. 744–750.

[10] M. Paul, S. K. Vishwakarma, and A. Verma, "Analysis of soil behaviour and prediction of crop yield using data mining approach", in *2015 International Conference on Computational Intelligence and Communication Networks (CICN)*, IEEE, 2015, pp. 766–771.

[11] N. Suma, S. R. Samson, S. Saranya, G. Shanmugapriya, and R. Subhashri, "Iot based smart agriculture monitoring system", *International Journal on Recent and Innovation Trends in computing and communication*, vol. 5, no. 2, pp. 177–181, 2017.

[12] N. Hemageetha, "A survey on application of data mining techniques to analyze the soil for agricultural purpose", in *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, IEEE, 2016, pp. 3112–3117.

[13] S. Babu, "A software model for precision agriculture for small and marginal farmers", in *2013 IEEE Global Humanitarian Technology Conference: South Asia Satellite (GHTC-SAS)*, IEEE, 2013, pp. 352–355.

[14] A. Manjula and G. Narsimha, "Xcypf: A flexible and extensible framework for agricultural crop yield prediction", in *2015 IEEE 9th International Conference on Intelligent Systems and Control (ISCO)*, IEEE, 2015, pp. 1–5.

[15] R. Kumar, M. Singh, P. Kumar, and J. Singh, "Crop selection method to maximize crop yield rate using machine learning technique", in *2015 international conference on smart technologies and management for computing, communication, controls, energy and materials (ICSTM)*, IEEE, 2015, pp. 138–145.

[16] T. R. Baitharu and S. K. Pani, "Analysis of data mining techniques for healthcare decision support system using liver disorder dataset", *Procedia Computer Science*, vol. 85, pp. 862–870, 2016.

[17] D. A. Reddy, B. Dadore, and A. Watekar, "Crop recommendation system to maximize crop yield in ramtek region using machine learning", 2019.

[18] Y. Cai, K. Moore, A. Pellegrini, A. Elhaddad, J. Lessel, C. Townsend, H. Solak, and N. Semret, "Crop yield predictions-high resolution statistical model for intra-season forecasts applied to corn in the us", in *2017 Fall Meeting*, 2017.

[19] T. J. Salo, T. Palosuo, K. C. Kersebaum, C. Nendel, C. Angulo, F. Ewert, M. Bindi, P. Calanca, T. Klein, M. Moriondo, *et al.*, "Comparing the performance of 11 crop simulation models in predicting yield response to nitrogen fertilization", *The Journal of Agricultural Science*, vol. 154, no. 7, pp. 1218–1240, 2016.

[20] *Soil research and development institute bangladesh*, Available at http://www.srdi.gov.bd//.