PREDICTION OF PRODUCTION RATE FOR CULTIVABLE LAND CORRELATING CROP COLOR, SOIL QUALITY PARAMETERS AND WEATHER CONDITION

By

Mehedi Hassan
17101402
S.M. Mushfiqur Rahman
16101040
Shadman Sobhani
16301064

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

Department of Computer Science & Engineering
BRAC University
July, 2023

Declaration

It is hereby declared that

1. The thesis submitted is my/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. I/We have acknowledged all main sources of help.

Student's Full Name & Signature:

Mehedi Hassan
17101402

S.M. Mushfiqur Rahman
16101040

Shadman Sobhani
16301064

Approval

The thesis/project titled "Prediction of Production Rate for Cultivable land Correlating Soil Quality Parameters and Weather Condition submitted by

1. Mehedi Hassan (17101402)
2. S.M. Mushfiqur Rahman (16101040)
3. Shadman Sobhani (16301064)

of Fall, 2021 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Computer Science & Engineering on 03 October, 2021.

Examining Committee:

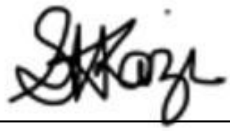
Supervisor:
(Member)

A.M.Esfar-E-Alam
Senior Lecturer
Department of Computer Science and Engineering
Brac University
Email: esfar.alam@bracu.ac.bd

Program Coordinator:
(Member)

Dr. Md. Golam Rabiul Alam
Associate Professor
Department of Computer Science and Engineering
Brac University

Departmental Head:
(Chair)



Sadia Hamid Kazi, PhD
Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Ethics Statement

1. **Data Privacy and Consent:** The project adhered to strict data privacy principles, ensuring that any personally identifiable information collected from individuals, such as farmers or landowners, was anonymized and kept confidential. Informed consent was obtained from participants before collecting any data, and they had the right to withdraw their consent at any time.

2. **Data Collection and Usage**: The project collected data on crop color, soil quality parameters, and weather conditions through appropriate and ethical means, ensuring that the data collection methods didn't harm the environment or infringe upon the rights of individuals or communities. The data was being used solely for the purpose of predicting production rates and improving agricultural practices.

3. **Transparency and Accountability:** The project strived for transparency in its methodologies and findings, providing clear explanations of the data analysis techniques used and the assumptions made. Any limitations or uncertainties in the predictions were explicitly communicated to stakeholders. The researchers were accountable for the accuracy and integrity of the data and analysis conducted.

4. **Beneficence and Non-Maleficence:** The project aimed to contribute to the agricultural sector by providing valuable insights for optimizing crop production and land management practices. Efforts were made to ensure that the project's outcomes benefit farmers and promote sustainable agricultural practices. Potential negative impacts, such as increased resource exploitation or environmental degradation, was minimized and mitigated.

5. **Fairness and Equity:** The project was conducted in a manner that respects the principles of fairness and equity. Bias and discrimination was actively avoided in data collection, analysis, and interpretation, taking into account diverse agricultural systems and considering the needs and perspectives of different stakeholders, including marginalized communities and small-scale farmers.

Abstract

In the modern time due to increasing global warming, the change of weather patterns and increase in pollution, the amount of crop production has been adversely affected. While on the other hand farmers trying to prevent this have been starting to use harmful fertilizers, insecticides and better growing crops. While doing that farmers are decreasing the soil fertility at an unnatural fast rate and making agricultural products that are becoming more and more harmful for human consumption in the long run. This has been having harmful effects on the human body like heart, liver and pancreatic problems or decrease in both male and female fertility. Bangladesh is a small riverain country with the staple food production of rice. It is the most commonly consumed food among the people. We believe that by analyzing the weather pattern of the previous years and analyzing the soil fertility with modern technologies it is possible to tackle the problems while making a cheaper investment by the farmers in the land on natural fertilizers and less pesticides.

This paper intends to focus on the development of crop production in the agricultural field through comparing soil test data and weather conditions. The analysis of parameters can be proven to be accurate and less time-consuming as compared to traditional approaches. Moreover, the prediction can help the farmer to know the number of crops that can be grown in the next session.

Keywords: Weather patterns, Pollution, Crop production, Harmful fertilizers, Insecticides, Soil fertility, Agricultural products, Human consumption, Health effects, Bangladesh, Rice production, Weather pattern analysis, Soil test data, Natural fertilizers, Prediction.

Table of Content's

List of Tables

List of Figures

List of Acronyms

Sl No.	Acronyms	Full Form
01.	pH	Potential of Hydrogen
02.	i.e.	That is
03.	SVM	support vector machine
04.	RFC	Random Forest Classifier
05.	Ec	Emulsifiable Concentrate
06.	OC	Organic Carbon
07.	OM	Organic Matter
08.	N,P,K	nitrogen, phosphorus, and potassium
09.	CSV	Climate Smart Village

Chapter- 1

I. Introduction

The development of human civilization and the development of agriculture have gone hand-in-hand since ancient times. Those who have been in power have always been the civilization or country with the most advanced agricultural system or The Civilization with the most advanced Agricultural system. If we analyze the Indus Civilization or Egyptian civilization, they were built based on agriculture.

Even the most powerful Nation at the present time the U.S.A. was able to be in its present condition thanks to agriculture and the mass production of corn as an important crop and with the help of the development of Technology, they were able to mass-produce Huge amount of crops.

Bangladesh is an agrarian society and still more than 40% of the people in our country are farmers in rural areas. The main crop that we cultivate here is rice. Still, even with the amount of farmers in our land We had to import 800 tons of rice in the year 2020[1]. We believe that importing rice even though it is the main crop of the farmers in our country is due to the lack of implementation of Technologies in agriculture.

Even though tractor is a common cultivating tool throughout the world since the 80s, Bangladesh was finally able to introduce the practice to 95% of the population to tractors in around 2020. It took more than 30 years for this technology to be available to everybody. It occurs due to the poor conditions of the farmers. If you cultivate one bigha (33 decimals) of land using a plow, it will cost Tk 2,000 (23.59 USD). The cost with a power-tiller is Tk 1,500 (17.70 USD). If you do it with a tractor, it will cost you Tk 600 (7.08 USD). In the case of harvesting paddy, it costs Tk 2,000 on one bigha of land. It costs Tk 500 (5.90 USD) if the harvest is done with a combined harvester. In planting the paddies, it takes Tk 2,200 (25.95 USD) on one bigha. It costs about Tk 500 to transplant and saves time as well [2]. So it is evidently clear that using technologies in the field makes agriculture cheaper and more accessible to people.

We believe that by implementing soil minerals including pH level, and data analysis of those data along with the collection of weather patterns of the previous years we can make a prediction of the cultivable lands. With the advancement and happening of technology around 41% of Bangladeshi people are able to use smart phones and at present 26% of the population are able to use the internet without any problem (Mobile-Internet-Connectivity-South-Asia-Fact-Sheet 2020).

So by implementing these technologies we believe that farmers can increase the production of rice with almost zero cost and decrease the import of rice.

1.1 Research Problem

With the development of technology, the internet has become available throughout the whole world and especially in Bangladesh. With the help of mobile networking, it has become easier for the people of our country, especially the rural area, to have access to the internet. But even though the internet has become easier to access, the majority of our people still cannot use the internet due to illiteracy and lack of knowledge. but things are changing as the literacy rate of our country is increasing. The younger generations of our country are using the internet through smart phones, computers, and other accessible media. At present only 26% of our population can use the internet. We believe that Even though 26% of the population are accessing the internet which is much increased from the previous 13% in 2010 most of the people especially in the rural areas where we plan to implement our program are using it for Facebook and YouTube only. They lack the proper knowledge on how to use the internet for daily uses outside this too mainstream media.

One of the biggest hurdles we will face with this thesis is the lack of a weather forecast. Even if you can forecast the weather of the targeted fields of work for a week we cannot go beyond that.. the Chaos Theory dictates that We cannot predict the weather accurately for a long time.

Chaos emerges in a prediction when a small unknown or error amplifies. It's somewhat like a spacecraft sent to the moon on a slightly off-course trajectory, eventually sailing past and into the black ether of space. [4]

The earth's atmosphere is constantly changing and evolving with the rising sinking and whirling air which is a prime example of chaos. We cannot fully predict the weather because we cannot know when a small atom in the air is traveling and how fast it is thanks to the laws of Physics and all the uncertainty principles. Even if on a large scale the climate model can be calculated through a super computer[5] we cannot still calculate Such a meticulous thing like the weather on a small scale. If we could predict the weather on a large scale the data analysis for the production of our targeted crop could have been even more accurate than calculating it with only the weather up to one single week.

Other than this one of the problems we might face and we believe that we might face is the accurate collection of data of soil testing for the last few years. Even though we have an infrastructure to collect the soil data of land fertility by the Department of Land Record and Survey, their land record collection and fertility recording surveys are irregular and have a lack of proper infrastructure. We believe we might face a problem while trying to collect the data from them. Their irregular surveys Mike cause a problem in data analysis.

We might also face a problem as the farmers even though they fertilized the land they don't properly measure or keep a record of the number of fertilizers they use on their land. They also don't keep a proper record of how much of the targeted crops were cultivated throughout the years accurately. They are not also digitized.

We believe that throughout our Research and this thesis we might be finally able to make a proper database of information that was lacking in the past to our farmers easily, while we create a prediction algorithm that might help the farmers take steps to get the maximum amount of production of rice in their fields.

1.2 Research Objective

This research aims to develop a prediction algorithm, which will predict the maximum production rate of our targeted crop (i.e. rice), that will grow in the following season, on the test land. With this in mind, the procedure will include weather condition analysis, soil minerals, and fertility analysis, and image processing of land. Usually, only soil test data analysis is used as a standard for the cultivation and prediction of the amount of cultivation that might yield but there are more aspects to a proper prediction of the cultivation. We plan to not only use the rate of soil taste data but to include the color of the crops or rather how much the color of the crop is varied from its desired state at that particular state of production, the weather pattern like if it will rain or there would be a lack of water etc. With the help of the above-mentioned data collection and tackling the problems through them we believe the growth of crops can be manipulated to the desired level. So the objectives of this research are:

1. To develop the agricultural sector.

2. To develop an algorithm for predicting the amount of rice yield for the next season.

3. To evaluate the weather conditions to accommodate the farmer with a sustainable situation.

4. To offer the farmer folk more production with less expense.

5. To apply the most efficient amount of fertilizer in the field.

Chapter- 2

Literature Review

Modern technology as we know it is developing rapidly as a result of the introduction of smart devices that have the ability to overcome the problem of humans within a short amount of time. In the recent publication of daily star- the average share of agriculture as a proportion of GDP growth for the past decade ranged from 17.5 percent (2009) to 12.68 percent (2019).

In comparison, the share of the manufacturing sector steadily rose from 25.3 percent (2009) to 29.65 percent (2019), while the share of the services sector remained almost constant—53.2 percent in 2009 and 58.8 percent in 2019

One question may arise, what can be possible enough for the drastic change in the agricultural sector for such a development. The answer is simple, the development of equipment as well as methods used in production. According to the GSMA, only 41 percent of people use smart devices in Bangladesh. Among them only 30 percent of the farmers use smartphones, so the ability to predict the future crop production in a particular area will benefit the farmer.

2.1 Soil Minerals Data Analysis

As the population of Bangladesh is increasing rapidly, the crisis of sufficient food is rising. Farmers are trying hard to fulfill their needs but with the usage of old technology, the production rate is low. Farmers are applying fertilizers to the land which may hinder the fertility of the land. Soil fertility is an important factor for farmers which determines whether the land is for cultivation or not. For plants, the essential micronutrients are boron, sodium, chlorine, copper, zinc, etc. But the farmers are using organic, inorganic, and bio-fertilizer without knowing what amount is needed for the plant. As a result, the production rate for the crops is not yielding high. The table below shows the soil Nutrient status in agricultural land at Kalihati Upazila, Tangail.

Soil nutrients	Land types			
	HL (n=6)	MHL (n=9)	MLL (n=9)	LL (n=6)
Soil pH	6.00±0.54	5.82±0.40	5.53±0.23	5.16±0.16
Status	SlA	SlA	StA	StA
OM (%)	2.00±0.12	2.15±0.26	2.40±0.49	2.64±0.71
Status	M	M	M	M
TN (%)	0.10±0.01	0.11±0.01	0.12±0.02	0.13±0.03
Status	L	L	L	L
AP (μg g^{-1})	8.20±4.96	3.26±1.62	1.88±0.83	2.04±1.00
Status	OP	VL	VL	VL
Zn (μg g^{-1})	2.80±1.98	3.80±1.22	3.02±1.35	3.39±1.17
Status	VH	VH	VH	VH
Fe (μg g^{-1})	168.10±82.50	192.80±43.70	288.80±45.19	365.10±50.59
Status	VH	VH	VH	VH
Mn (μg g^{-1})	25.60±11.21	15.74±8.70	25.25±15.10	18.49±9.93
Status	VH	VH	VH	VH
B (μg g^{-1})	0.10 ±0.03	0.17±0.05	0.18±0.04	0.15±0.05
Status	VL	L	L	VL
K (meq $100g^{-1}$)	0.10 ±0.02	0.15±0.05	0.21±0.06	0.28±0.05
Status	L	L	M	OP
Ca (meq $100g^{-1}$)	4.80 ±1.12	5.22±1.15	6.59±1.49	6.25±1.07
Status	OP	OP	H	H
Mg (meq $100g^{-1}$)	2.00±0.43	2.05±0.37	2.11±0.27	2.12±0.25
Status	VH	VH	VH	VH

Note: H=High, L=Low, M=Medium, N=Neutral, OP=Optimum, SlA=Slightly Acidic, StA=Strongly Acidic, VH=Very High, VL=Very Low.

Table 01: soil Nutrient status in agricultural land at Kalihati Upazila, Tangail.

In the above chart, we can see the nutrients present in the soil, and by analyzing the data the farmer can use the current amount of fertilizer for the land. Moreover, the PH value is also needed because it regulates plant nutrient availability by controlling the chemical form of the different nutrients and also influences their chemical reaction.

Standard of soil pH		Standard of soil organic matter (%)	
Value	Soil reaction class	Value	Status
<4.5	Very strongly acidic	<1.00	Very low
4.5-5.5	Strongly acidic	1.00-1.70	Low
5.6-6.5	Slightly acidic	1.71-3.40	Medium
6.6-7.3	Neutral	3.41-5.50	High
7.4-8.4	Slightly alkaline	>5.50	Very high
8.5-9.0	Strongly alkaline	-	-
>9.0	Very strongly alkaline	-	-

Table 02: pH value of agricultural land at Kalihi Upazila, Tangail. [7]

Soil water Moisture Level

Soil moisture is determined in the first meter below the surface. The water contained in soil may vary from season to season. During the dry season, the soil may contain 3% to 10% water. On the other hand, 20% to 40% in the rainy season. According to the research,

Soil Minerals Data Analysis- Investigation of soil properties and nutrients in agricultural practiced land in Tangail, Bangladesh M.S. Islam*, T. Nusrat, M.R. Jamil, F. Yesmin, M.H. Kabir and R.H. Rimi Received 10 August 2020, Revised 23 November 2020, Accepted 24 December 2020, Published online 31 December 2020

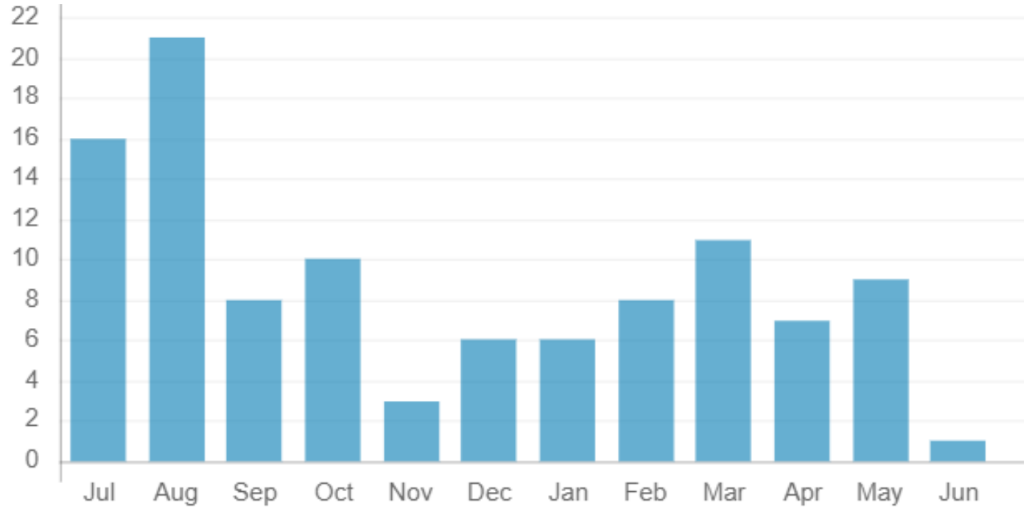


Figure 01: Soil moisture levels during the year in Bangladesh

The above chart shows the soil moisture levels during the year in Bangladesh.

2.2 Weather Condition

Bangladesh has a tropical monsoon climate which causes rainfall, high temperature, and high humidity. Moreover, the geographical location of Bangladesh is suitable for farming as a result 70 percent of the land is suitable for cultivation and 50 percent of Bangladesh people are farmers. The occurrence of erratic weather is beyond human control so adaptability is the only thing that humans can do. Weather forecasting for agriculture is an essential requirement.“Weather forecast is an expression of the probability of a particular future state of the atmospheric system in a given point or territory. Late forecasts calling for a season may cause departure for normal agronomic practices at the land. Therefore, the pre-season forecast must have a validity period of 1 week not less than that. During different seasons many pests, diseases may attack the crop and damage the production rate which can be prevented by anticipating the weather. Moreover, the soil water moisture level can also hamper crop production.

Elements for agricultural weather forecast:

a) Amount of cloud cover
b) Maximum, minimum, and dewpoint temperature
c) Relative humidity
d) Wind speed and direction

e) Solar radiation
f) Bright hours of sunshine
g) Pan evaporation etc.

Sometimes for special crops or cost-saving special weather forecasts are needed such that the necessary meteorological information can aid the farmer. Sometimes thunderstorms, tornadoes, wind gales, and floods can damage the production which is considered a natural disaster. But the farmers with adequate measures should be adopted to avoid these hazards.

Chapter- 3

Methodology

The process of this prediction method starts with identifying the preferred location, of which we would want to know the cultivation state. Firstly, we check if the selected location has any cultivable land or not. This would be possible with the database/set we are going to use that has land usage information. Afterward, the whole process will be divided into three different processes and they are, retrieving soil test data, collecting satellite imagery, and checking with a production rate of the year prior. These data are analyzed and then compared with the reference data. Finally, a decision will be made about the cultivation state of the selected land.

Primarily, we are relying on the weather prediction and soil fertility rate as these two factors are the determiners for crop production. To determine and predict our production rate, we are going
only with the production of Aush rice in Bangladesh [8]. For weather prediction and forecast history, our parameters are limited to only rainfall and temperature. It is because Aush is highly dependent on rainfall. Based on the past weather history, a linear regression model is used here to predict the future weather forecast. We have collected and generated some datasets on rainfall and temperature based on the past ten years. With that we have generated a predicted value using the linear regression approach.

Linear regression method is a very well known and efficient method for predictions. With this simple approach, events based on weather prediction can easily be determined [9]. In the datasets, average rainfall and temperature of each month starting from 1971 all the way to 2015 [8]. After that, these are being analyzed and a prediction of next season is generated.

As for soil fertilization, we have used three datasets here. The datasets consist of pH levels, minerals levels, along with a constant value which has two parameters- fertile and non fertile. The reason behind it is that we have used the lands that match a particular parameter [10], for the ease of making the procedure simpler and that we can easily come up with a successful prediction. As it happens, we are doing this only for a better implementation.

$$\frac{\partial J(\theta)}{\partial \theta_0} = \frac{1}{m} \sum_{i=1}^{m} (h(x^i) - y^i)$$

$$\frac{\partial J(\theta)}{\partial \theta_1} = \frac{1}{m} \sum_{i=1}^{m} (h(x^i) - y^i) x_1^i$$

(a)

Y(pred) = b0 + b1*x

The values b0 and b1 must be chosen so that they minimize the error. If sum of squared error is taken as a metric to evaluate the model, then goal to obtain a line that best reduces the error.

$$\text{Error} = \sum_{i=1}^{n} (actual\_output - predicted\_output) ** 2$$

Figure 2: Error Calculation

If we don't square the error, then positive and negative point will cancel out each other.

For model with one predictor,

$$b_0 = \bar{y} - b_1 \bar{x}$$

Figure 3: Intercept Calculation

$$b_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Figure 4: Co-efficient Formula

(b)

(c)

$$\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{m} \sum_{i=1}^{m} (h(x^i) - y^i) x_j^i$$

We can compute the partial derivatives for all parameters at once using

$$\begin{bmatrix} \frac{\partial J(\theta)}{\partial \theta_0} \\ \frac{\partial J(\theta)}{\partial \theta_1} \\ \vdots \\ \frac{\partial J(\theta)}{\partial \theta_n} \end{bmatrix} = \frac{1}{m} x^T (h(x) - y)$$

where *h(x)* is

$$h(x) = \theta_0 + \theta_1 x_1 + ... + \theta_n x_n$$

Fig 2 : Linear Regression theorem(a, b, c)

With the above mentioned datasets, we have used SVM and Random Forest Classifier(RFC) to determine the soil fertility rate. RFC is a supervised learning algorithm. It can work for both classification and regression. It creates decision trees by fetching data from random data and gives predictive output and from those predictions, chooses the best result. Therefore, this algorithm is best suited for our prediction.

After compiling the model, the next phase is efficient computation. The model executed on some data. Train and fit our model on our loaded data by calling the fit function on the model. Training occurs over epochs and each epoch is split into batches. Epoch is the data that undergoes all of the rows within the training dataset. Batch is one or more samples considered by the model within an epoch before weights are updated. One epoch consists of one or more batches, based on the chosen batch size and the model is fit for many epochs. The training process ran for a fixed number of iterations through the dataset called epochs. We also set the numbers of dataset rows that are considered before the model weights are updated within each epoch as called the batch size and set using a batch argument.

3.1 Input data

From the dataset, the past weather history of temperature and rainfall values are run through a linear regression model. Here, the standard linear regression method is being used. The model uses total rainfall each year as inputs and in the same way, average temperature of each year is taken as inputs.

Algorithm 1. Standard linear regression

Input: data set $\boldsymbol{D} = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \cdots, (\boldsymbol{x}_n, y_n)\}$ and threshold t
Output: $\boldsymbol{\theta} = [w_1, w_2, \cdots, w_d, b]$

1: Define the loss function $J(\boldsymbol{\theta}) = (\frac{1}{2n}) \sum_{i=1}^{n} (\boldsymbol{\theta}^T \boldsymbol{x}_i - y_i)^2$
2: Generating the $\boldsymbol{\theta}^0$ randomly
3: **repeat**
4: $\boldsymbol{\theta}^k = \boldsymbol{\theta}^{k-1} - \alpha \frac{\partial J(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$, where $\boldsymbol{\theta}^k$ is the value of k^{th} iteration and α is the iteration step.
5: **until** $J(\boldsymbol{\theta}^{k+1}) - J(\boldsymbol{\theta}^k) < t$
6: **return** $\boldsymbol{\theta}$

Fig 3: *Linear Regression Algorithm*

3.2 Data pre-processing

Due to a lack of sufficient infrastructure, processing 720 45-year readings can be computationally challenging. As a result, the fertilizer from the readings was kept constant and the rest of the data was stored in the “New Aush.csv”.csv file. Furthermore, since the suggested model is focused on detecting abnormalities at the datasheet's applications layer, irrelevant features (i.e. columns) such as date and time of reading were eliminated.

Anomalies were also manually added to the incoming data (i.e. 454 records). By manually adding records with anomalies in the dataset, employed the same method to modify input data. The advantages of using this method are that (1) the input dataset will become qualified for anomaly detection and meet the research's aim. (2) addressing the difficulty of acquiring a highly accurate Agricultural dataset of our country

Temperature, Rainfall, Area and production would be the multiple attributes (i.e. columns) in the generated input dataset, which would have 720 entries. As seen in figure, the value of each characteristic in each record is real.

Chapter- 4

Implementation and Results

4.1 Implementation

In order to get valid predictions we have used Weka and linear regression approach. The inputs in Weka are the outputs of the analyzed data provided by the weather and soil prediction. From it we were able to find the linear regression and the rate at which the per yield crop productions are seen.

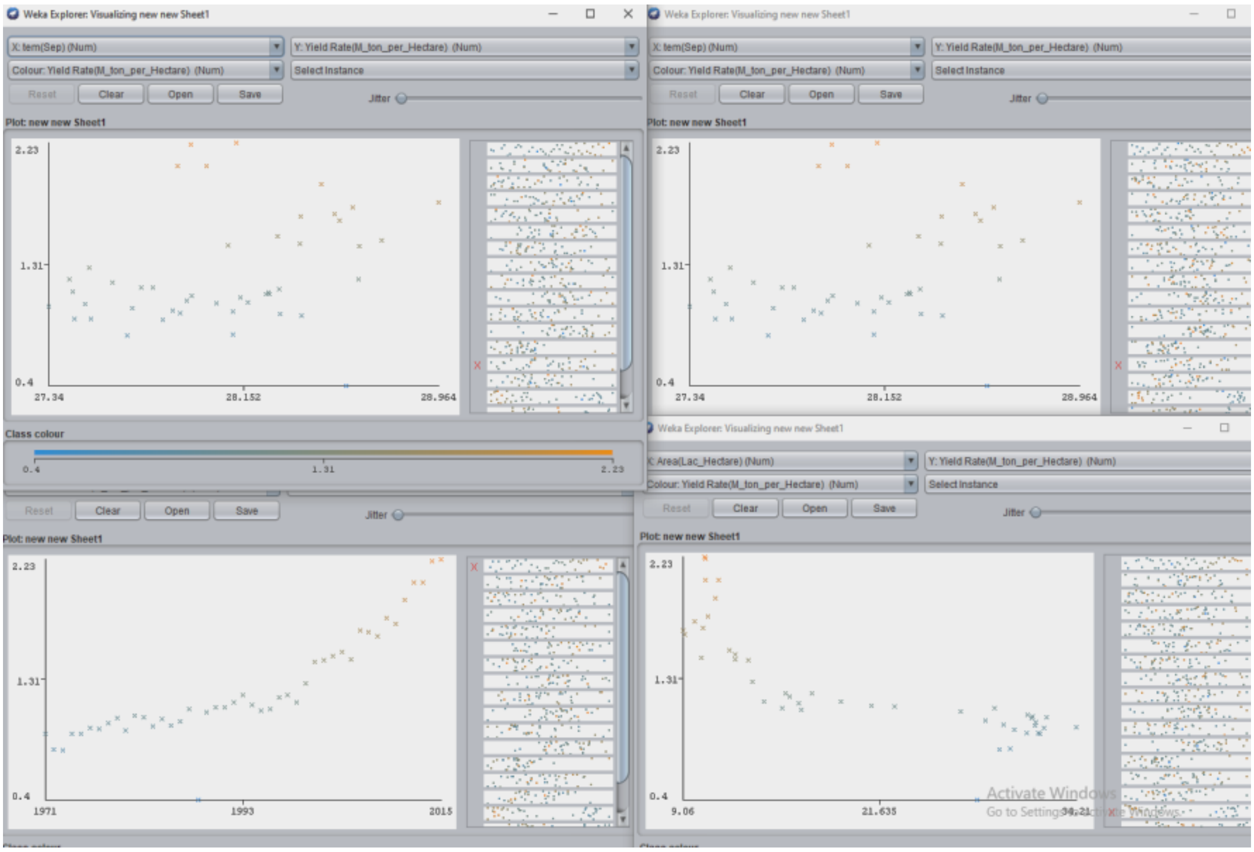


Figure 04: (Temperature v Yield) Plotting with Weka

4.2 Soil Fertility Prediction

We improve the number of data by data synthesis method and then we analyze the data and divide the data according to the variables. Here the dependent variable is Fertility of the soil which is a dichotomous classification. We train, test and split the data. We used the Random Forest Model to find the accuracy of the evaluation.

Classification if the soil is Fertile or Not:

i) Very attributes are taken such as Ph, Ec,OC,OM,N,P,K etc to determine if the soil is fertile or not.

ii) we take SVC and the accuracy has come around 0.82.

iii) We tried multinomial naive bayes model and the accuracy has come around 0.65.

iv) Finally we tried a Random forest and the accuracy came up to 100%.

```
In [18]:
final_l = []

for index, row in res.iterrows():
    if row['Output']=="Non-Fertile":
        val = 0
        new_p = 0
        l = []
        for inner_loop in range(15):
            val, row = Fertilize(row)
            if val!=0:
                l.append(val)

            n_test = row
            new = pd.DataFrame(n_test[0:16].values.reshape(1,-1))
            new.columns = ['pH','EC','OC','OM','N','P','K','Zn','Fe','Cu','Mn','Sand','Silt','Clay','CaCO3','CEC']
            new_p = rf.predict(new)
        final_l.append(l)
for i in range(len(final_l)):
    print(i,final_l[i])

0 ['OC: 0.24', 'N: 108', 'K: 45', 'Zn: 0.32', 'Fe: 1.9', 'Mn: 0.2']
```

Fig . Final output of fertile land

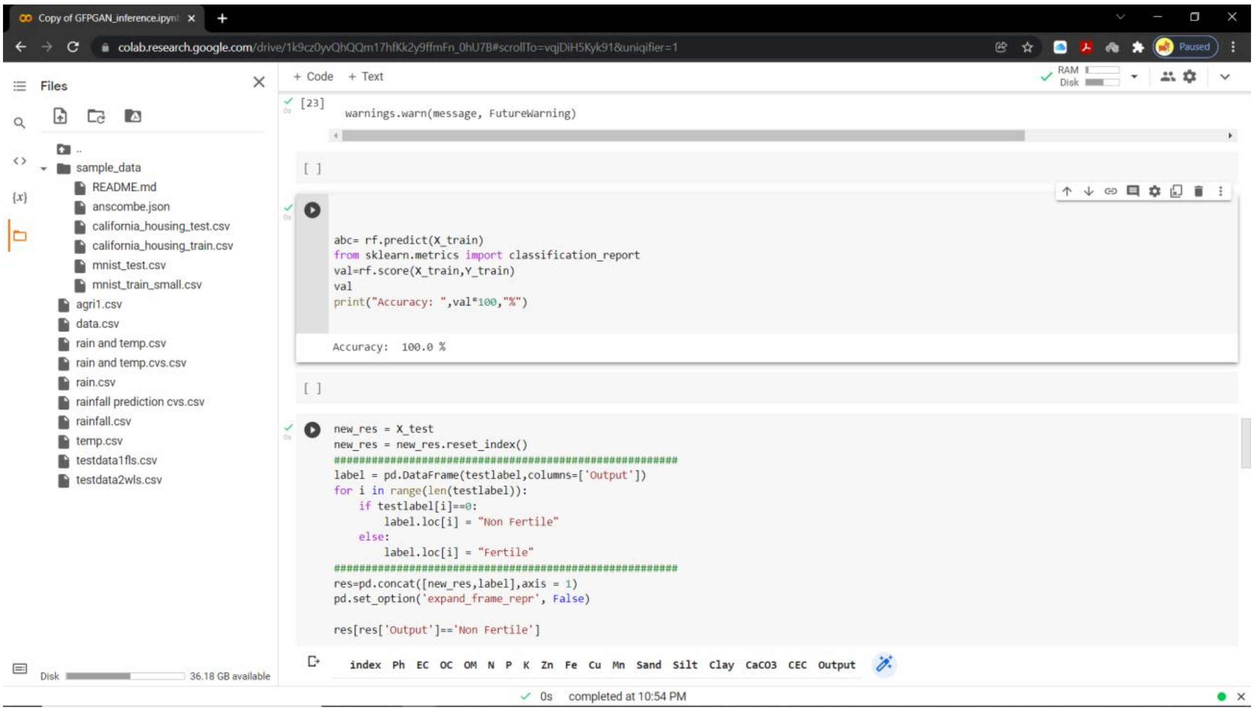
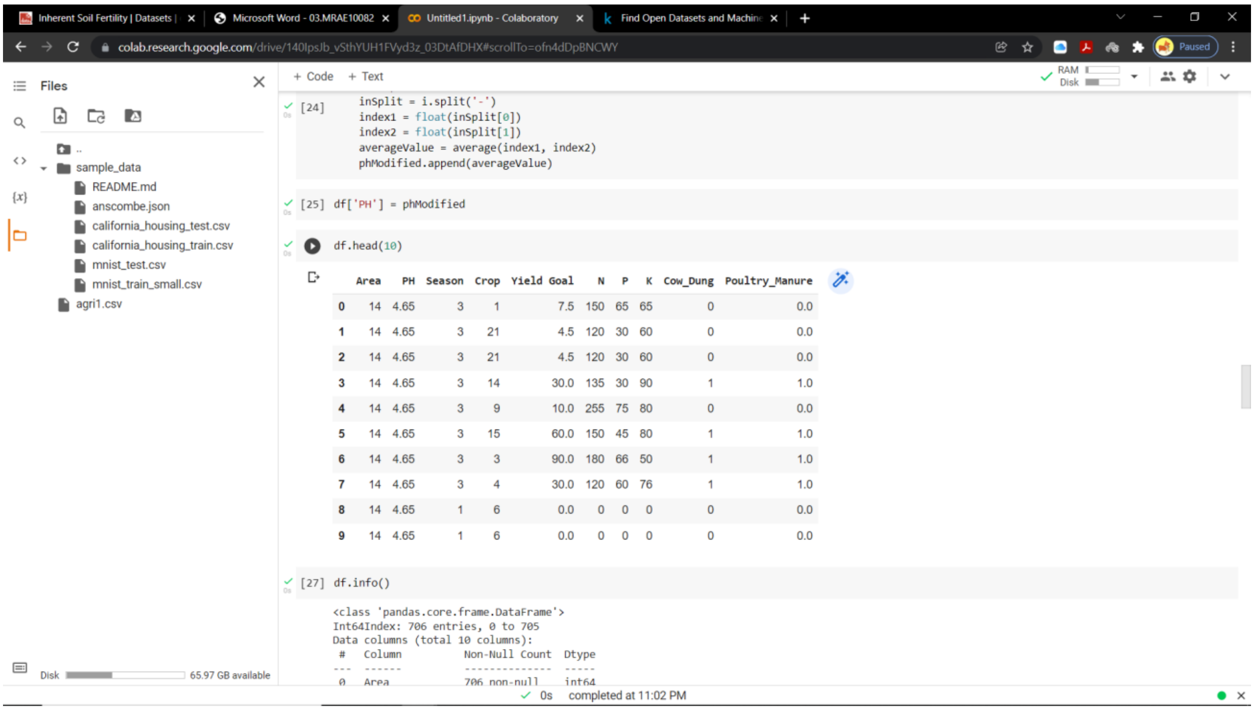


Fig Accuracy of Random Forest Classifier

Crop Yield Prediction

To Predict the yield of a crop, we have taken a 'yield dataset' with various attributes such as N, P, k, Organic carbon, pH, and temperature determining the yield of a crop. As the yield of a crop is continuous we adopt regression techniques to predict the yield of the crop.

We used Random Forest model and SVC for maximum accuracy. But the SVC model is the best model for yield prediction of the crop.



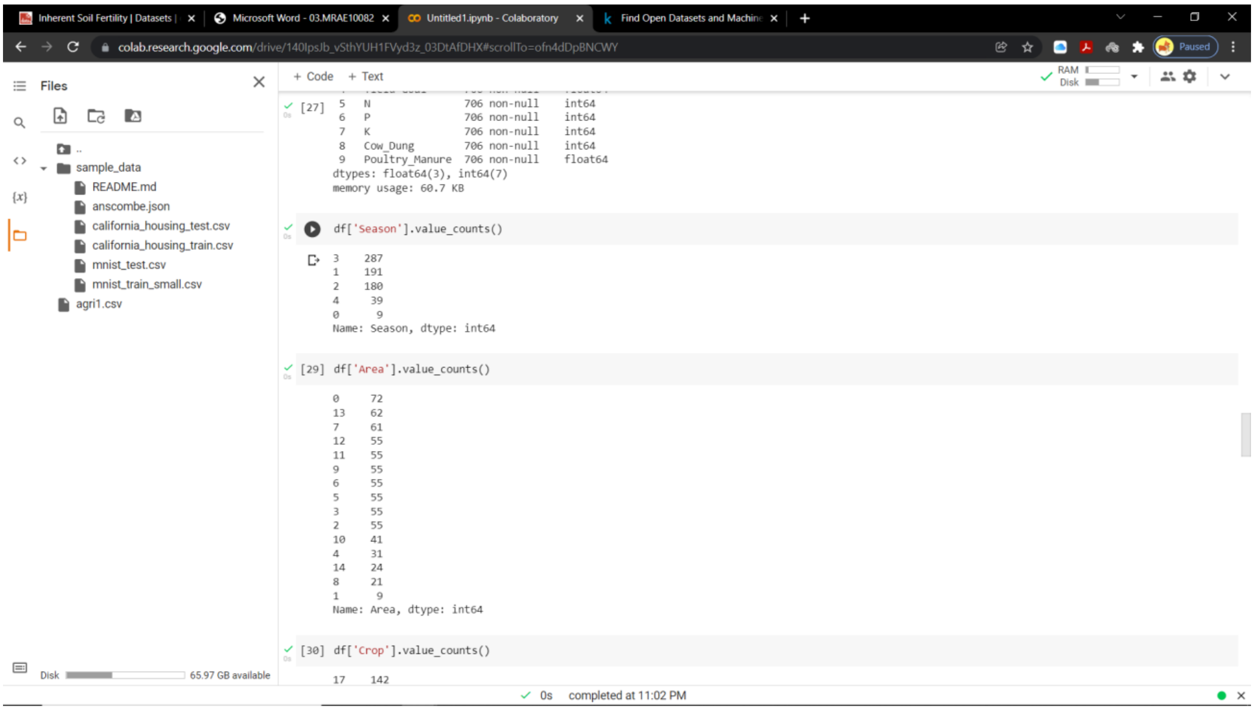


Fig Data of area 14

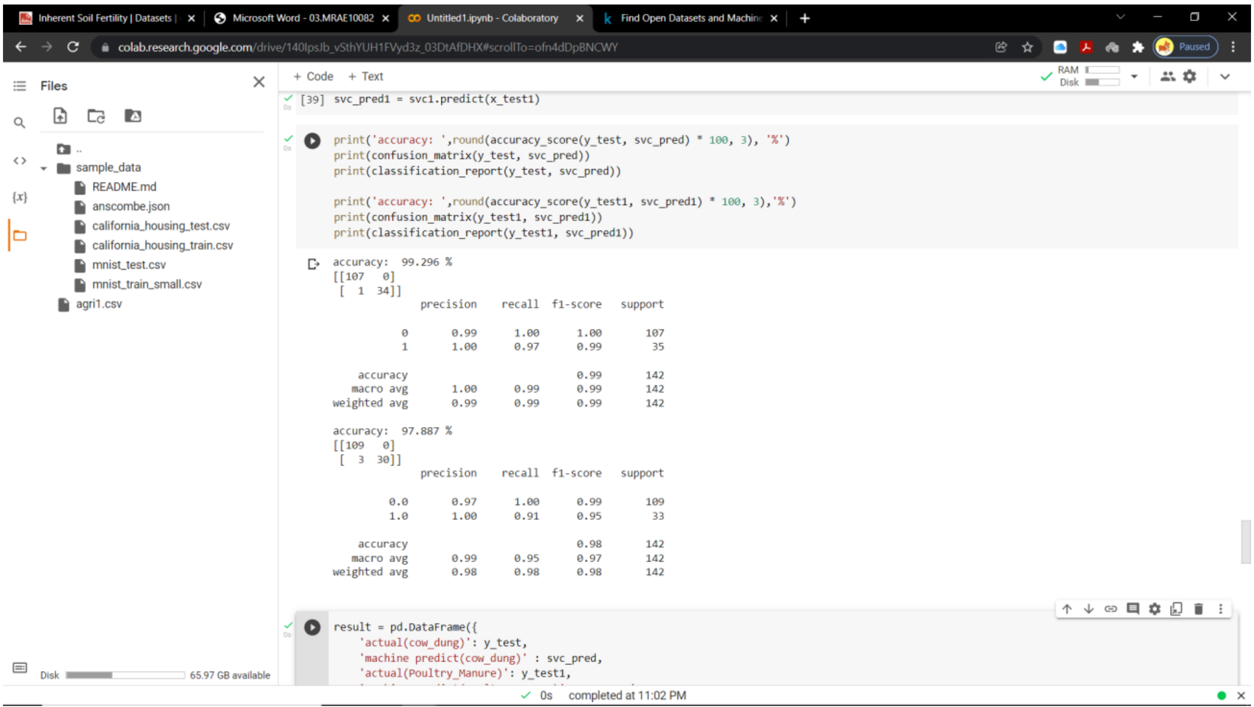


Fig Accuracy of yield prediction.

The accuracy of the yield goal comes up to 97.296% and 97.887% . Here we also keep natural fertilizer an option for being more accurate for soil fertility and crop production.

4.3 Results

After using Weka for making the cumulative prediction through multiple jossing with intentional error and without any error we get the graph ratios in Fig-4.

Also we did normalize the multiple average rain and temperature attributes while removing the year, the area of land and total production and kept only the per area production to see the linear regression formula along with its various variables in the Fig - 5.

```
Test mode:    20-fold cross-validation

=== Classifier model (full training set) ===

Linear regression on tem(Jul)

0.7 * tem(Jul) + 0.86

Predicting 0 if attribute value is missing.


Time taken to build model: 0 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient                  0.1658
Mean absolute error                      0.3136
Root mean squared error                  0.4089
Relative absolute error                 91.075  %
Root relative squared error             97.321  %
Total Number of Instances               45
```

(a)

```
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

Linear regression on tem(Jul)

0.7 * tem(Jul) + 0.86

Predicting 0 if attribute value is missing.


Time taken to build model: 0 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient                  0.0893
Mean absolute error                      0.3331
Root mean squared error                  0.4219
Relative absolute error                 97.7081 %
Root relative squared error            100.8113 %
Total Number of Instances               45
```

(b)

Figure 04: Prediction of crop yield using Weka

Chapter- 5

Conclusion

With the advancement of technologies, the world is increasing in population also. It is predicted that the human population in the year 2050 will be around 9 billion and the food production in the world must increase by 70% at least to meet that demand. Begint the ninth-largest population in the world with such limited area Bangladesh will evidently have to face that crisis head-on. But we are still lagging in agricultural production and making proper use of our resources because we lack advanced technologies and proper infrastructure. Our thesis hopes to build a proper database and a base technology with easy accessibility which can be used by others to create a better infrastructure and more complex technologies to help in the growth of the agriculture of our country.

Reference

1. Knoema. (2021). Bangladesh - Rice imports quantity. knoema. https://knoema.com/atlas/Bangladesh/topics/Agriculture/Trade-Import-Quantity/Rice-imports-quantity
2. Seraj, S. (2020, July 10). Agricultural machinery: Where is Bangladesh heading? *The Daily Star*.
3. Etherington, D. (2017, February 4). Google selling Terra Bella's satellite imaging business to Planet. *Tech Crunch*. https://techcrunch.com/2017/02/03/google-selling-terra-bella-satellite-imaging-business-to-planet/
4. Kaufman, M. (2020, September). The Dominance Of Chaos. Mashable. https://in.mashable.com/science/16814/the-dominance-of-chaos
5. Judt, F. (2018, May 01). Atmospheric Predictability of the Tropics, Middle Latitudes, and Polar Regions Explored through Global Storm-Resolving Simulations. Journal of the Atmospheric Sciences, 75(5), 1477–1497. ametsoc. https://doi.org/10.1175/JAS-D-17-0343.1
6. Asokan, A., Jude, A., Ciobanu, M., Naaji, A., & Hemanth, D. J. (2020, June 19). Image Processing Techniques for Analysis of Satellite Images for Historical Maps Classification—An Overview. Applied Science, 10(12). https://doi.org/10.3390/app10124207
7. Investigation of soil properties and nutrients in agricultural practice land in Tangail, Bangladesh M.S. Islam*, T. Nusrat, M.R. Jamil, F. Yesmin, M.H. Kabir and R.H. Rimi Received 10 August 2020, Revised 23 November 2020, Accepted 24 December 2020, Published online 31 December 2020
8. Bangladesh Bureau of Statistics (BBS). (2018, January). 45 years Agriculture Statistics of Major Crops (Aus, Amon, Boro, Jute, Potato & Wheat) (Statistics and Informatics Division (SID) & Ministry of Planning, Eds.). Ministry of Planning. http://bbs.portal.gov.bd/sites/default/files/files/bbs.portal.gov.bd/page/16d38ef2_2163_4252_a28b_e65f60dab8a9/45%20years%20Major%20Crops.pdf
9. Garafola, N. (2015, MAY). Event Detection Using Linear Regression and Historical Weather Data. https://core.ac.uk/display/37750759?recSetID=
10. Biswas, J. C., Kalra, N., Maniruzzaman, M., Haque, M. M., Naher, U. A., Ali, H., Kabir, W., & Rahnamayan, S. (2019, July). Soil Fertility Levels in Bangladesh for Rice Cultivation. *Asian Journal of Soil Science and Plant Nutrition*. 10.9734/ajsspn/2019/v4i430051