

**CROPS PRODUCTION PREDICT USING MACHINE LEARNING ON
PERSPECTIVE OF BANGLADESH**

BY

SIBLI SHADIK SHAON

ID: 181-15-10569

MD. FERDOUS SARKER ZIHAD

ID: 181-15-10541

AND

ABDULLAH AL MAMUN

ID: 181-15-10623

This Report Presented in Partial Fulfillment of the Requirements for the
Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Zerin Nasrin Tumpa

Lecturer

Department of CSE

Daffodil International University

Co-Supervised By

Abdus Sattar

Assistant Professor

Department of CSE

Daffodil International University



**DAFFODIL INTERNATIONAL UNIVERSITY
DHAKA, BANGLADESH
January 2022**

APPROVAL

This Project/Internship titled “**Crops Production Predict Using Machine Learning on Perspective of Bangladesh**”, submitted by **Sibli Shadik Shaon**, ID No: **181-15-10569**; **Md. Ferdous Sarker Zihad**, ID No: **181-15-10541**; **Abdullah Al Mamun**, ID No: **181-15-10623**; to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 6 **January 2021**.

BOARD OF EXAMINERS



Dr. Touhid Bhuiyan

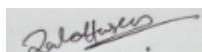
Professor and Head

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University

Chairman



Zahid Hasan (ZH)

Associate Professor

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University

Internal Examiner



Mohammad Monirul Islam (MMI)

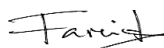
Senior Lecturer

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University

Internal Examiner



Dr. Dewan Md. Farid

Professor

Department of Computer Science and Engineering

United International University

External Examiner

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Zerin Nasrin Tumpa, Lecturer, Department of CSE Daffodil International University**. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:



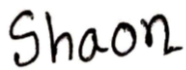
Zerin Nasrin Tumpa
Lecturer
Department of CSE
Daffodil International University

Co-Supervised by:



Abdus Sattar
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:



Sibli Shadik Shaon
ID: 181-15-10569
Department of CSE
Daffodil International University



Md. Ferdous Sarker Zihad
ID: 181-15-10541
Department of CSE
Daffodil International University



Abdullah Al Mamun
ID: 181-15-10623
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project/internship successfully.

We really grateful and wish our profound our indebtedness to **Zerin Nasrin Tumpa**, Lecturer, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of “*Machine Learning*” to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior draft and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to **Dr. Touhid Bhuiyan and Head**, Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

Bangladesh is predominantly agricultural country and the agriculture sector is crucial to the country's economic development. To maintain long-term food security for humans, it is necessary to developing a productive, stable, and ecologically friendly agricultural system. Researchers in agriculture are working hard to develop techniques that will boost livestock and crop yields, improve farmland productivity and minimize the diseases and insect damage. They strive to increase overall food quality and construct more efficient equipment. This research work proposed a system for estimating the production of twelve distinct crops cultivated in Bangladesh's Cumilla and Chandpur districts. This system takes a few simple characteristics as input (district, sub district, crop name, and area) and predicts crop yield as output. Random Forest Regression and Decision Tree Regression are used in this study as advanced regression systems. People will be able to determine which crop will produce the maximum yield on their land by using this approach.

TABLE OF CONTENTS

CONTENTS	PAGES
Approval	i
Declaration	ii
Acknowledgement	iii
Abstract	iv
 CHAPTER	
CHAPTER 1: INTRODUCTION	1-2
1.1 Introduction	1
1.2 Motivation	2
1.3 Rationale of the Study	2
1.4 Expected Output	2
 CHAPTER 2: BACKGROUND	3-6
2.1 Introduction	3
2.2 Related Works	3-5
2.3 Research Summary	5
2.4 Scope of the Problem	5
2.5 Challenges	5
 CHAPTER 3: RESEARCH METHODOLOGY	6-9
3.1 Introduction	6
3.2 Research Subject and Instrumentation	6
3.3 Data Collection Procedure	6
3.4 Statistical Analysis	7-8
3.5 Implementation Requirements	8-9

CHAPTER 4: EXPERIMENTAL RESULTS AND DISCUSSION	10-17
4.1 Introduction	10
4.2 Experimental Results	10-15
4.3 Descriptive Analysis	15-17
4.4 Summary	17
CHAPTER 5: SUMMARY, CONCLUSION, RECOMMENDATION AND IMPLICATION FOR FUTURE RESEARCH	18-19
5.1 Summary of the Study	18
5.2 Conclusions	18
5.3 Recommendations	18-19
5.4 Implication for future scope	19
REFERENCES	19-20
APPENDICES	21-23

LIST OF FIGURES

FIGURES	PAGE NO
Figure 3.4.1 Line Chart	7
Figure 3.4.2 Scatter Chart	7
Figure 3.5.1 Research Methodology	8
Figure 3.5.2 System Architecture	9
Figure 4.2.1 Accuracy Score Chart	12
Figure 4.2.2 Font End Diagram	12
Figure 4.2.3 Web View Home Page	13
Figure 4.2.4 Web View District Select	13

Figure 4.2.5 Web View Sub-District Select	14
Figure 4.2.6 Web View Crop Name Selection	14
Figure 4.2.7 Web View Area Select	15
Figure 4.3.1 Bar chart of 24sub-district's production	15
Figure 4.3.2 Bar chart Aman Rice Production	16
Figure 4.3.3 Til Production	17

LIST OF TABLES

TABLE	PAGE NO
Table 3.4.1 District Distribution	8
Table 4.2.1 Accuracy of algorithms	11

CHAPTER 1

INTRODUCTION

1.1 Introduction

Bangladesh is mainly an agricultural country, and the agricultural sector is crucial to the country's economic prosperity. Agriculture is the major sector of the Bangladeshi economy because of its importance in ensuring food security, creating jobs, and sustaining livelihoods. About 80% of Bangladesh's population directly or indirectly depend on Agriculture. Agriculture currently accounts for 21% of GDP and employs nearly half of the working population. Agriculture productivity increases are an essential policy goal in developing countries because agriculture is a big financial sector. A significant fraction of the population in all developing countries, particularly in rural areas where poverty is more visible, depends on agriculture for a living, both overt and covert. As a result, an expanding agricultural sector aids overall growth as well as poverty reduction. Cumilla and Chandpur have long been a district where commercial agricultural is practiced. Unfortunately, population pressure continues to put a strain on production capacity, resulting in a food shortage. Our country's cultivable land is dwindling day by day. The loss of agricultural land and the decrease in the variety of crops and livestock produced are two of the most serious issues in agriculture. Besides, farmers have not proper knowledge about scientific method in Agriculture. That's why they are unable to produce sufficient crops. In this paper we worked with crop data of 24 upazilas of Cumilla and Chandpur districts. Our goal is to aid farmers by developing a model for predicting crop production using Machine Learning. Farmers will be able to identify which crop will generate the most on their land as a result of this.

1.2 Motivation

Many people in Bangladesh are unfamiliar of scientific farming methods. They have no idea what kind of crops they should plant on their land to generate more profit. They don't know which crops will flourish on their property. That is why our country has so many unused cropland.

If we can solve these difficulties, it will be extremely beneficial to Bangladesh.

1.3 Rationale of the Study

1. We collect crops data from Cumilla and Chandpur district of Bangladesh.
2. Using algorithm there will be given prediction about the crop production.
3. Let farmer can know their crops production.

1.4 Expected Output

1. Find out crop production prediction with best accuracy.
2. Knowing the growth rate of crops where grow well
3. Increase awareness about agricultural production.

CHAPTER 2

BACKGROUND

2.1 Introduction

In our country maximum people doing farming about 75% people. A vast population's only one income source. But most of them don't know scientific cultivation method. They can't understand which crops will be profitable for them, which will grow well. On the other hand, Bangladesh is a low riparian country. For this reason we have prepared this Machine Learning based system to predict crops production basis on district, sub district, crops name and area perspective on Bangladesh.

2.2 Related Works

In this paper, they proposed a user friendly system to justify which crop can be grown in a particular region. They are performing crops prediction for district level by using production details of past 10-12 years also details about climatic parameters and soil parameters like rainfall, temperature, moisture, soil contents etc. They use KNN algorithm to predict rainfall or weather report and Decision tree algorithm to predict crop. Finally, they have found accuracy for Decision-Tree: 76.8% and accuracy for KNN: 89.4%.[1]

They create a strategy for calculating agricultural yield based on historical methodology in this research. There are 3101 instances in the data collection that are related to the local data. The five factors or qualities are warmth, acid levels, humid, monsoon, and crop identification. They intended to train the model using decision trees, but discovered that using the random forest technique minimized overfitting and enhanced accuracy. Finally, they trained the model using the random forest approach. [2]

This approach predicts crop yield using soil type, pH value, weather, temperature, and fertilizer as inputs. For soil equity, the system uses a Support vector and a Relevant vector Machine in this paper. Decision trees, Naive Bayes, and KNN are also used. Beyond everything, this technique suggests the best fertilizer for increased crop yield. [3]

The goal of this research is to predict the production of various crops. There is a dataset that include data from 1998 to 2014. This paper only look at district and state levels, not area levels, and don't include fields less than 100 hectors. Their accuracy is affected by the district and the weather. They apply KNN Algorithm and get an accuracy of 82 percent. This system includes the state, district, weather, pH, and crop yield. Above all, at the state level, this method recommends agricultural output and success rate. [4]

This technique is based on a supervised data collection and works with six crops that grow in Bangladesh's 12 districts. KNN, a decision tree method, was employed in this study. This system predicts crop growth rate based on average rainfall, tem, and average yield rate. [5]

They used hardware software to collect data in a real-life experiment in this research. It has an easy-to-use interface for farmers to analyze. They employ temperature, humidity, and moisture content as inputs and forecast successful growth as outputs using the naive bayes supervised learning method, which has a 97 percent accuracy. [6]

They employed Random Forests Algorithm to predict crop yield in this paper. This data pertains to the Indian state of Maharashtra. Implementation with a web-based graphics user interface, inputting data such as rainfall, temperature, and area receives a crop production output. The accuracy of projections in all crops to predict crop yield is more than 75%. [7]

For soil equity, the system uses a Support vector and a Relevance vector Machine in this paper. They also use KNN, Naive Bayes, and Decision Tree. This approach predicts crop yield using soil type, pH value, weather, temperature, and fertilizer as inputs. [8]

In this work, they develop a system that uses input parameters such as district, crop name, and area to forecast output. They plan to create an app in the future so that farmers can utilize it in their native language. The accuracy levels of supervised learning approaches for classification of agricultural data vary. The accuracy of SVM-based categorization for farm data is higher. [9]

They proposed a system uses some parameters such as crop name, pH, Nitrogen, Depth, Temperature, and Rainfall to predict yield as an output using an ANN algorithm. [10]

According to the literary works, the forecasting model takes into account weather inputs such as rainfall, temperature, and humidity, as well as non-weather inputs such as soil moisture, pH, salts in soil (nitrogen, phosphate, potassium, organic carbon, calcium, magnesium, sulphur, and so on), crop type, and seed variety. As a result, it is recommended to examine how big data tools might help improve crop yield predictions. [11]

2.3 Research Summary

In modern era of science Machine Learning is very popular technique to analyze any kind of dataset. In our work we collect data from agricultural sector to predict the production of some crops which grow in Bangladesh. By using Machine learning system, we get accuracy 99%. First of all, we collect data from Kaggle and we make a survey for collecting data. Then we pre-process our data set. After that we handle Null value of data set. Then we use Random Forest and Decision Tree algorithm for prediction.

2.4 Scope of the Problem

This research is on agriculture based. We live in an agricultural country. That's why we have a great opportunity to collect data in this field. In any kinds of research there can be a scope of problem. So, there is no exception in our work too. We train our data so many times to get better accuracy.

2.5 Challenges

Any kinds of task can't be accomplished without challenges. When we try to doing any work, challenges could come in any time in any steps of the work. In our research the main challenge we had to face while collecting data from the sub district office and farmers. That means in our research surveying data from sub district & farmers and collecting the information is the main challenge on this pandemic situation. Collection of proper and appropriate data is needed from farmers for setup a better dataset. That's why, we can say the main challenge was collecting data which should be more appropriate for our research to find the better output.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

Research methodology refers to the techniques or tactics used to find, select, process, and analyze information about a topic. A scholarly article's research approach allows the reader to critically evaluate the study's overall legitimacy and trustworthiness. It is the set of rules that will help us do our research in a more manageable, efficient, and effective manner. We built a dataset containing primary data that was manually collected for this paper. After that, we finish data pre-processing and deal with null values in the dataset. Finally, we use two different machine learning algorithm over the dataset to predict crops production.

3.2 Research Subject and Instrumentation

We work in our project for welfare of formers so that they can produce more and more crops. In our work we use different kinds of algorithms. We work with many famers, sub-district agriculture officer in two district named Cumilla and Chandpur in Bangladesh. We make a crop production prediction system based on machine learning which is very helpful for user. We use google colab to execute our dataset about crop production to give better accuracy. And the whole system performed by python language system.

3.3 Data Collection Procedure

When we make a plan to do something in agricultural sector by using Machine Learning; First of all, we start to read some research paper from google scholar then we try to find out some data set from kaggle, CNI, Reddit etc. We also try to read many kinds of magazines, websites, journals etc. Then we went to meet with our sub district Agriculture Officer for collecting data. After that we go to field to talk with farmer to acquire knowledge about crops. But in covid - 19 pandemic situation it was very difficult for us.

3.4 Statistical Analysis

After preparing our dataset we explain our whole data set graphically.

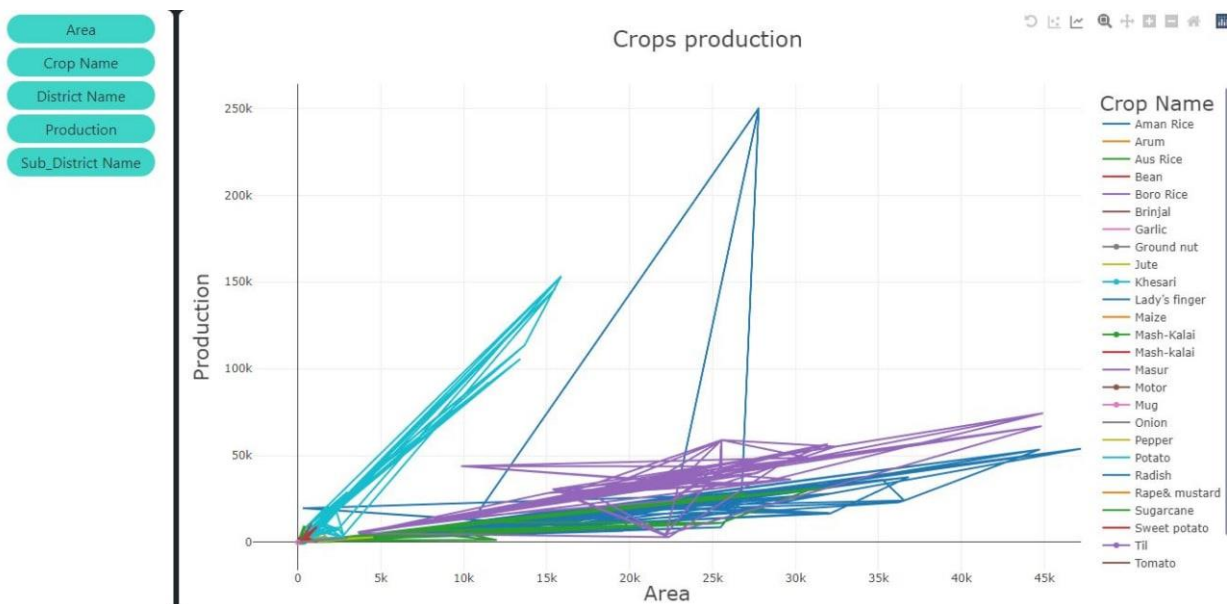


Figure 3.4.1 Line Chart

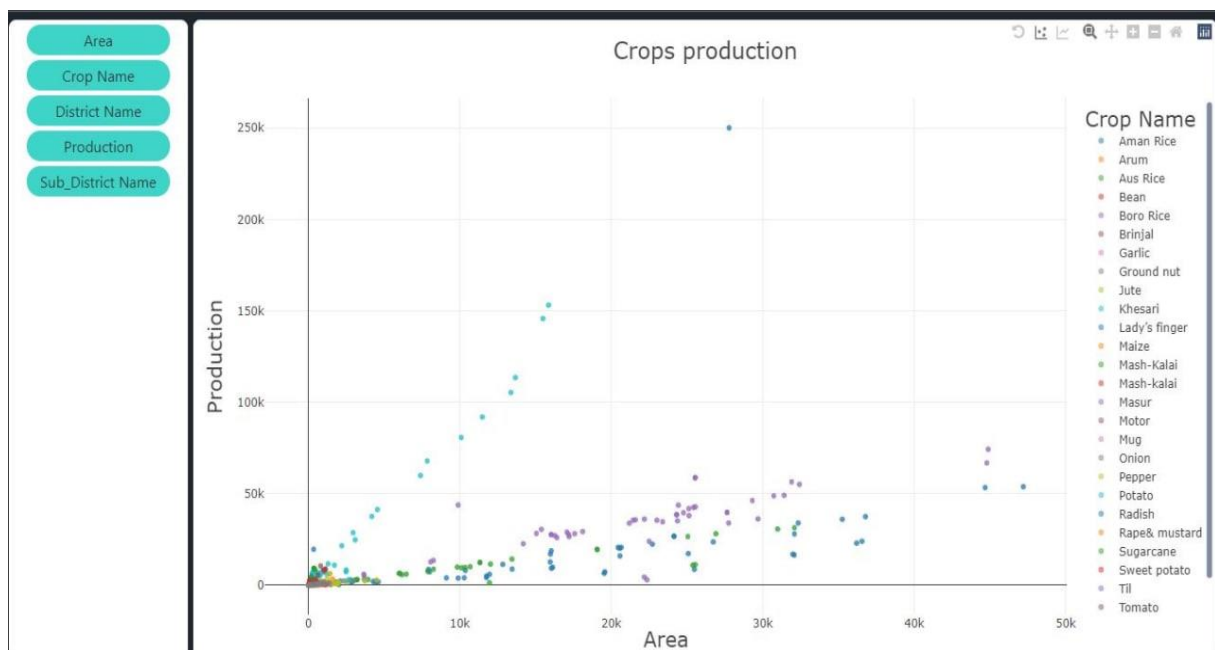


Figure 3.4.2 Scatter Chart

From this graph we can see the production (unit of metric_ton) basis of area(per_acor). From this chart we can see the crops named Aman Rice, Aus Rice, Boro Rice, Wheat, Mug, Musur, Til, Sugarcane, Maskalai, Jute, Potato, Sweet-potato etc.

In our system we consider two District which shown in table:

District	Sub District
Cumilla	Remain 16 Sub District
Chandpur	Remain 8 Sub District

Table 3.4.1 District Distribution

3.5 Implementation Requirements

Here we can the research methodology step by step

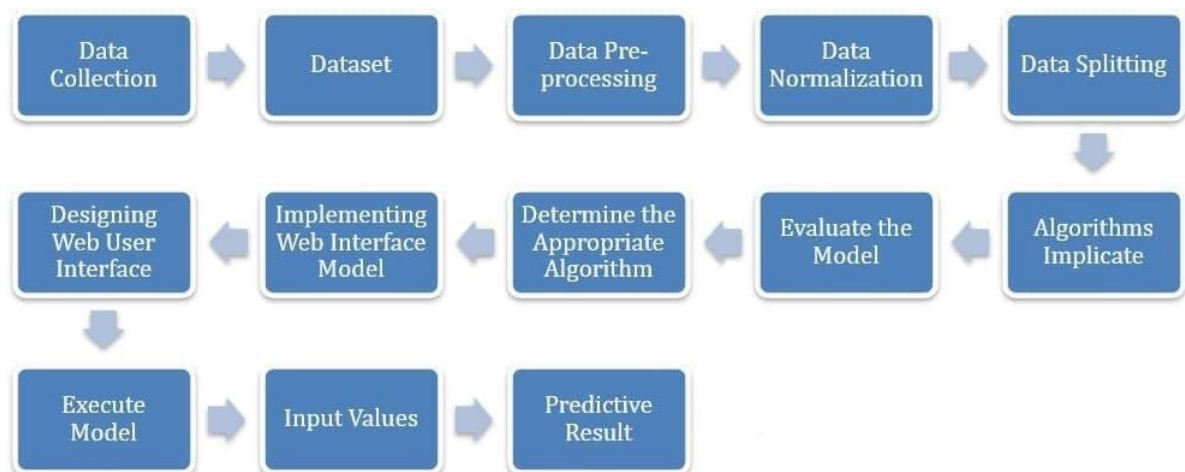


Fig 3.5.1 Research Methodology

We can target a good producible crops by use our system and it'll be profitable for farmers.

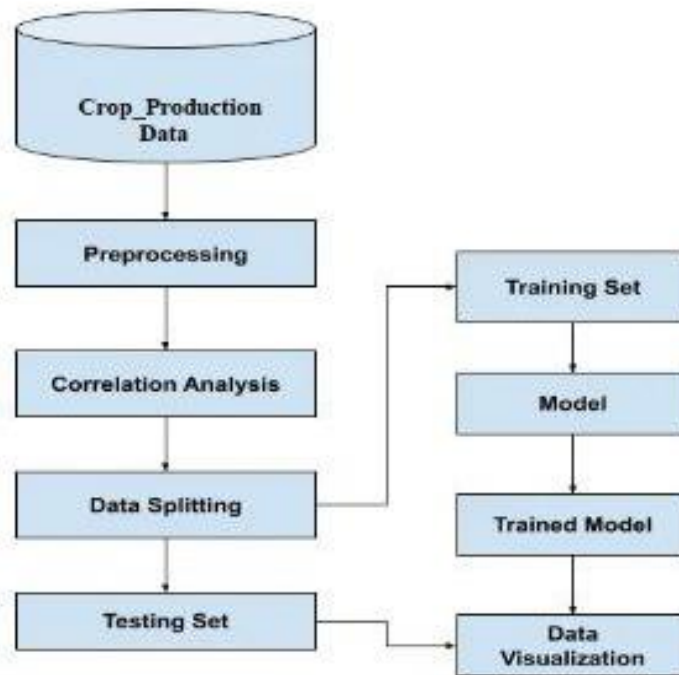


Fig 3.5.2 System Architecture

We use two kinds of algorithms

1. Random Forest Algorithm
2. Decision Tree Regression.

In Random Forest Train data accuracy 95.32% and test data accuracy 85.36%

In other hand Decision Tree Regression Train data accuracy 99.97% and test data accuracy 99.27%

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
from sklearn.metrics import confusion_matrix
```

We've used pandas, NumPy library in our project. Pandas is used for data frame. Sklearn.model_selection is used for importing the models or algorithm which we want to use and also used for importing train_test_split function for train and test the dataset.

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Introduction

To building up our model we collect many kinds of data from farmers, upazila agriculture Officers from two district about Cumilla and Chandpur which contain around 24 sub-districts. Then we make a dataset for our proposed working model. The dataset contains some parameter like district, sub-district, years, crop-name, production. To analyze the dataset and get result accuracy we use Decision Tree Algorithm and Random Forest Algorithm. From our system, before cultivating any crop field user can easily acquire knowledge about which crop will grow well in their land.

From our Machine Learning research, we also get a good idea about suitable crops for suitable land and user can produce best crops in big production and get good profit.

4.2 Experimental Results

Output of Decision Tree Regressor:

```
# import the decision_tree regressor
from sklearn.tree import DecisionTreeRegressor

# create a regressor object
tree = DecisionTreeRegressor(random_state = 0)

# fit the regressor with X and Y data
tree.fit(vs_train_norm, vs_target_train)

#score
print("Score Train", tree.score(vs_train_norm, vs_target_train)*100)
print("Score Test", tree.score(vs_test_norm, vs_target_test)*100)
```

```
Score Train 99.97536187272661
Score Test 99.27665484710585
```

Output of Random Forest Regressor:

```
# import the random forest regressor
from sklearn.ensemble import RandomForestRegressor

# create regressor object
randomfst = RandomForestRegressor(n_estimators = 10, random_state = 0)

# fit the regressor with x and y data
randomfst.fit(vs_train_norm, vs_target_train)

#score
print("Score Train", randomfst.score(vs_train_norm, vs_target_train))
print("Score Test", randomfst.score(vs_test_norm, vs_target_test))
```

Score Train 0.953270066060619
Score Test 0.8536671578501648

Accuracy of Algorithms:

Algorithm Name	Accuracy Score
Decision Tree	99.27%
Random Forest	85.36%

Table 4.2.1 Accuracy of Algorithms

Prediction result:

Actual Vs Predicted Table for Decision Tree

	Actual	Predicted
946	70.0	70.0
363	44.0	72.1
1095	7100.0	6141.7
678	36.0	80.4
311	2098.0	5612.1

Actual Vs Predicted Table for Decision Tree

	Actual	Predicted
946	70.0	70.0
363	44.0	44.0
1095	7100.0	7100.0
678	36.0	36.0
311	2098.0	2098.0

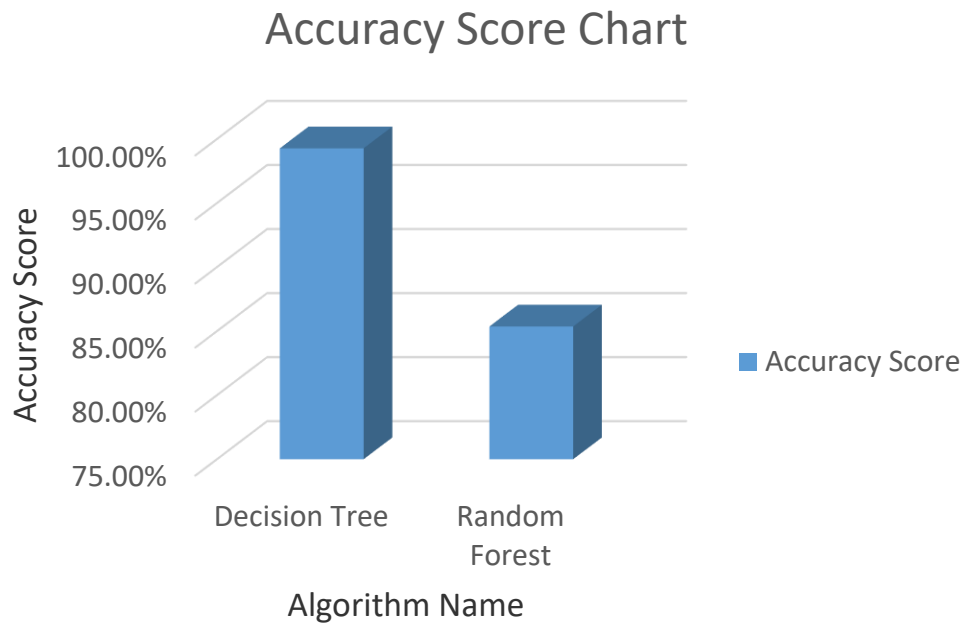


Figure 4.2.1 Accuracy Score Chart

We implement a website to see our prediction result

Front End Diagram:

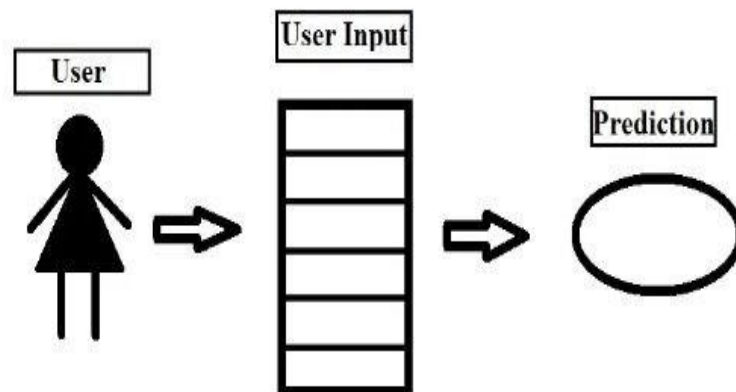


Figure 4.2.2 Front End Diagram

From this diagram we can see user have to provide district name, sub district name, crops name, area as input to get output (prediction of crops production).

Home page: There are three option first of all Home second is Prediction and third one is Contact Us. Now, we can see the work of prediction option.

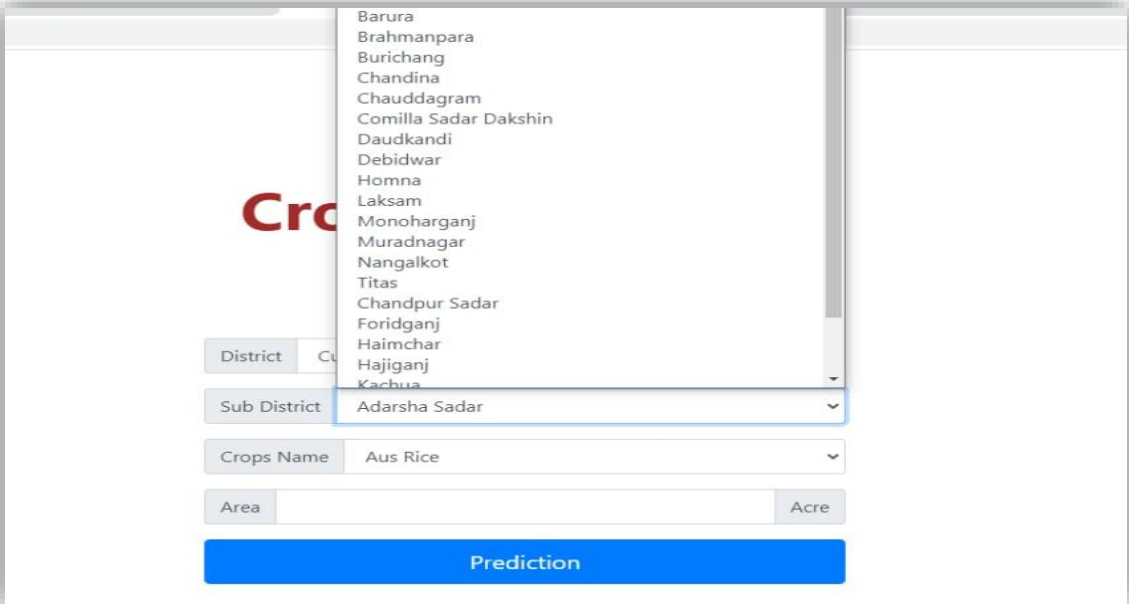


Figure 4.2.3 Web View Home Page

Step1: Select District: Here we can see two district. 1st of all we've to select one district.

Figure 4.2.4 Web View District Select

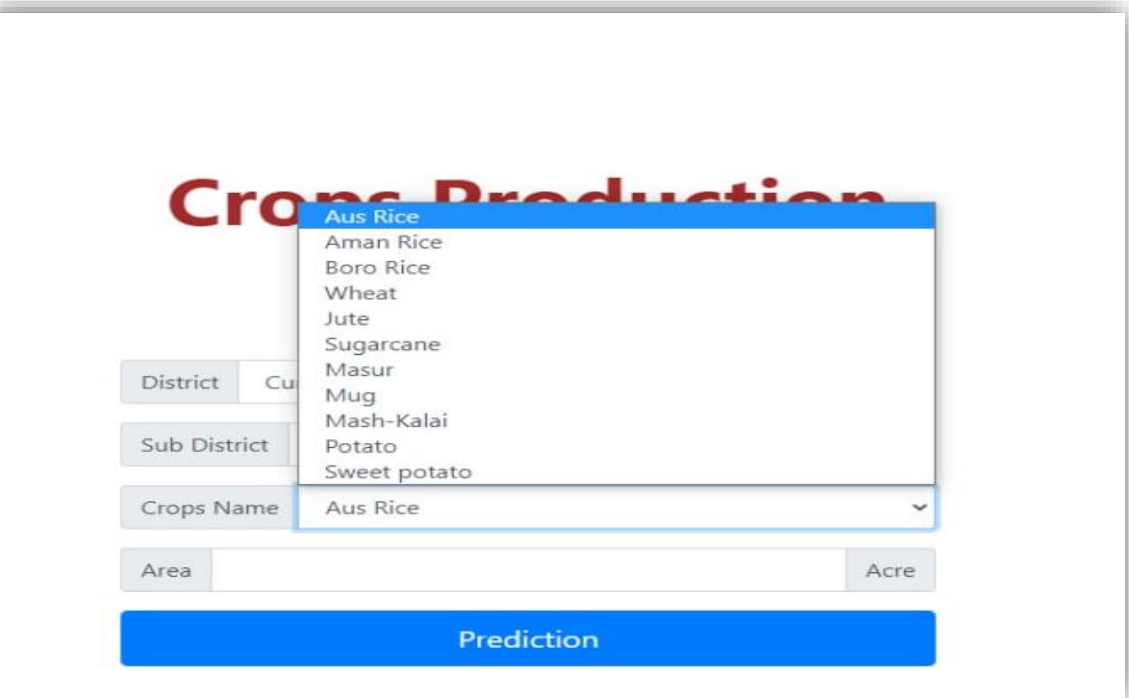
Step2: Select Sub-District: In the step2 there are around 32 sub districts and we've to select sub district name



The screenshot shows a web form titled "Crops Production". It has several input fields: "District" (set to "Comilla"), "Sub District" (open dropdown showing a list of 32 sub-districts including Barura, Brahmanpara, Burichang, Chandina, Chauddagam, Comilla Sadar Dakshin, Daudkandi, Debidwar, Homna, Laksam, Monoharganj, Muradnagar, Nangalkot, Titas, Chandpur Sadar, Foridganj, Haimchar, Hajiganj, Kachua, and Adarsha Sadar), "Crops Name" (set to "Aus Rice"), and "Area" (empty input field with "Acre" as the unit). A blue "Prediction" button is at the bottom.

Figure 4.2.5 Web View Sub-District Select

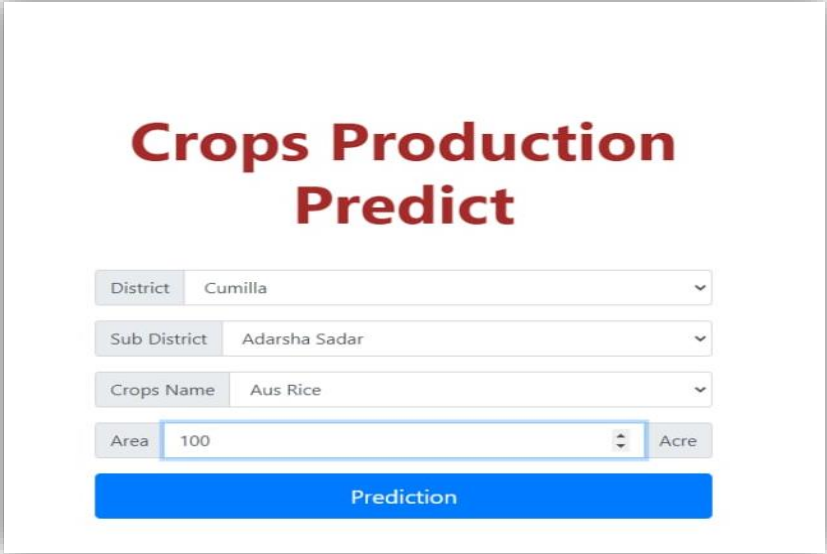
Step3: Select Crop Name: In this step there are many kinds of crops and we've to select one.



The screenshot shows the same web form as Figure 4.2.5, but with the "Crops Name" dropdown menu open. The list of crops includes Aus Rice, Aman Rice, Boro Rice, Wheat, Jute, Sugarcane, Masur, Mug, Mash-Kalai, Potato, and Sweet potato. The "District" dropdown is set to "Comilla", and the "Sub District" dropdown is set to "Adarsha Sadar". The "Area" input field is empty, and the unit is "Acre". A blue "Prediction" button is at the bottom.

Figure 4.2.6 Web View Crop Name Selection

Step3: Select Area: At the last step we've to select amount of area



Crops Production Predict

District: Cumilla

Sub District: Adarsha Sadar

Crops Name: Aus Rice

Area: 100 Acre

Prediction

Figure 4.2.7 Web View Area Select

4.3 Descriptive Analysis

We collect data from two district which containing 24 sub-districts.

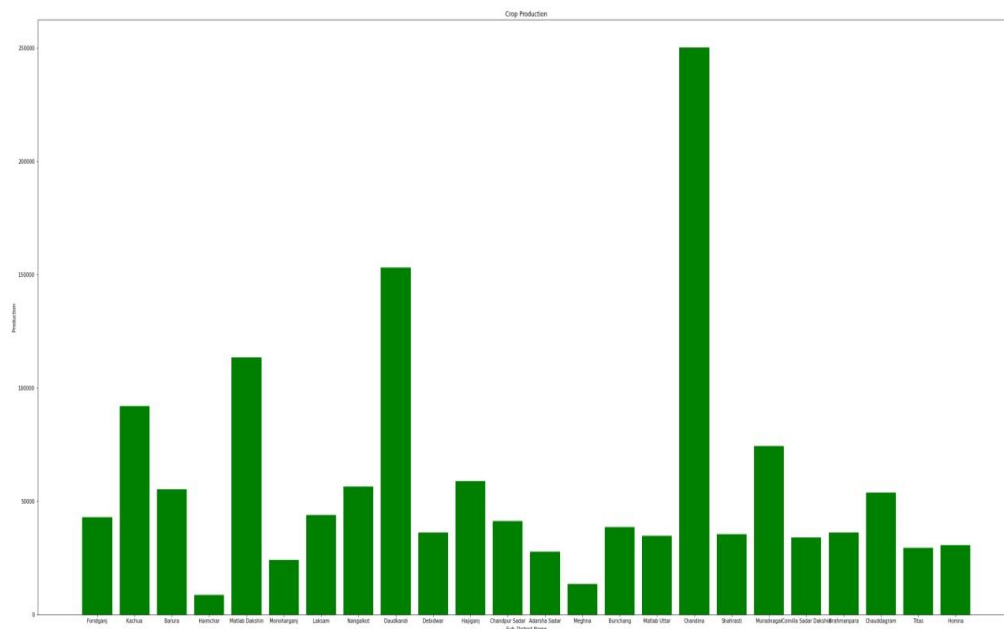


Figure 4.3.1 Bar chart of 24sub-district's production

From this graph we can see the different kinds of crop production of 24 sub-districts. And there is a large amount of production of crops contain Chandpur, Cumilla.

To analyze this graph user can easily understand which kinds of crops will grow in their own area.

Aman Rice Production:

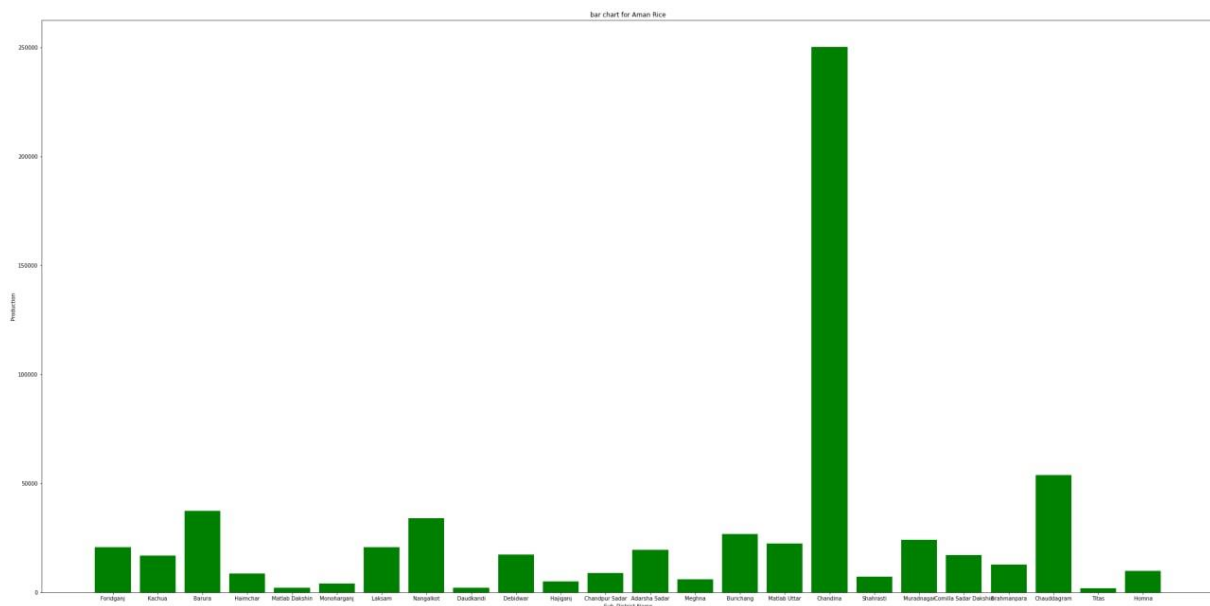


Figure 4.3.2 Bar chart Aman Rice Production

From this graph we can understand that Aman rice production very high in Chandina, Cumilla.

And low production in Titadi & Daudkandi.

To analyze this graph user can easily understand where aman rice will grow well.

Til Production:

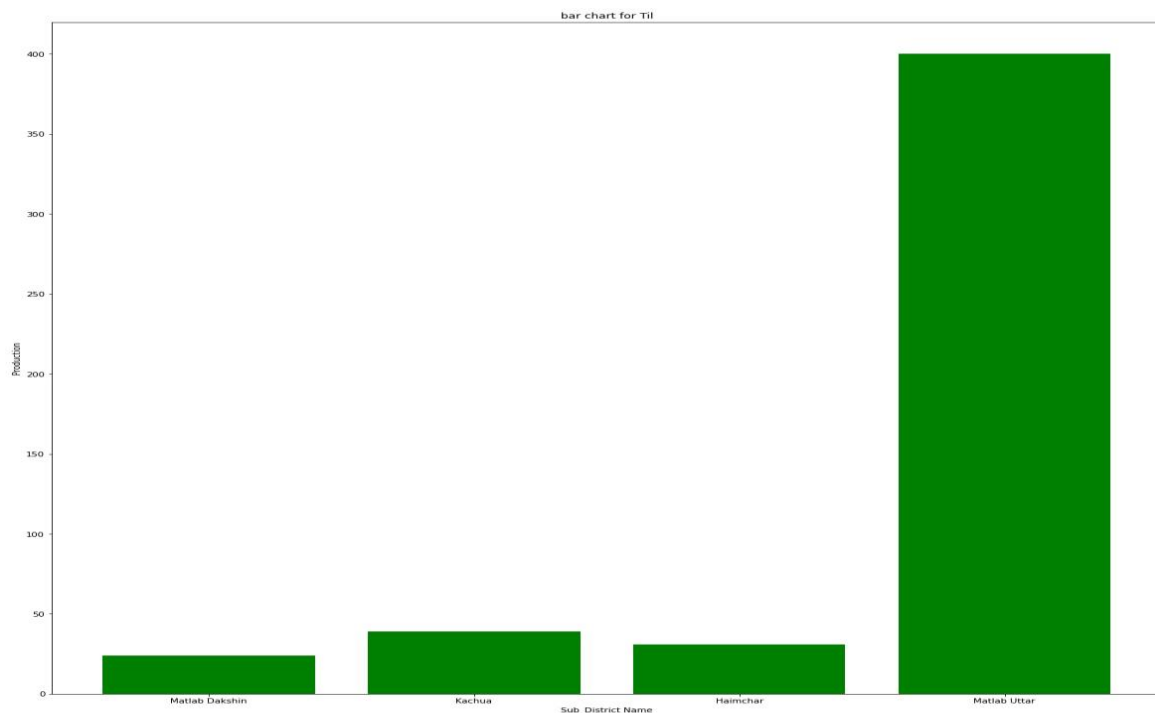


Figure 4.3.3 Til Production

From this graph we can understand that Til production very high in Motlob Uttar, Chandpur

And low production in Motlab Dakshin, Chandpur

To analyze this graph user can easily understand where Til will grow well.

4.4 Summary

In this paper by using Machine Learning Algorithm, we can predict the crop production before starting cultivation the crop field. We use two Machine Learning Algorithm one is Random Forest Regression and another one is Decision Tree Regression for Our dataset. The output of Random Forest Regression and Decision Tree Regression have given below

Decision Tree Regression Accuracy:

```
Score Train 99.97536187272661
Score Test 99.27665484710585
```

Random Forest Regression Accuracy:

```
Score Train 0.953270066060619
Score Test 0.8536671578501648
```

CHAPTER 5

SUMMARY, CONCLUSIONS, RECOMMENDATIONS AND IMPLICATION FOR FUTURE RESEARCH

5.1 Summary of the Study

Between 1983 and 1996, Bangladesh lost roughly 1 million hectares of agricultural land (BBS, 1999). That equates to around 80,000 acres of agricultural land being taken out of production each year. Land loss is mostly caused by urbanization, human settlement, infrastructure construction, and river erosion. The loss is undoubtedly alarming, and it must be addressed as soon as possible. In this scenario, we must cultivate our land systematically so that farmers can harvest as much as possible. Besides, the farmers have to be benefitted economically. That is why we initially work with crop data of 24 upazilas of our district town Cumilla and adjoining district Chandpur. So we have prepared a model for predicting crops production using some different machine learning algorithm. Decision Tree and Random Forest Classifier have been used for predicting the crops production over our collected data and our dataset. Data collection was the main challenge for our research work.

5.2 Conclusions

Now-a-days our country is developing modern technology. As a Digital Bangladesh, The Republic Government of Bangladesh, The Ministry of Agriculture have already taken some good necessary step for welfare of farmers so that they can cultivate their land in proper way and get best production. In our model, we use Machine Learning technology to predict the crops production based on area. Besides we also implement a website where we can easily predict the crop production.

5.3 Recommendations

To create our model for predicting crop production and to demonstrate which is better in prediction over our collected dataset, we applied two different machine learning methods. Random forest algorithm and Decision Tree are the two algorithms. When we conducted our

research on our dataset, we discovered that the accuracy rates of these various algorithms were nearly identical. Though algorithm accuracy varies from dataset to dataset, the accuracy rate for the algorithms tested in our dataset is remarkably similar. The accuracy of Decision Tree is higher than Random Forest when we increase or reduce our test value. As a result, we suggest that you apply or attempt Decision Tree in predicting research work.

5.4 Implication for Further Study

We are attempting to complete our entire project in Python in order to predict agriculture production. Random Forest and Decision Tree are two Machine Learning Algorithms that we use. Due to the Covid-19 pandemic, we have been collecting data from two districts, Cumilla and Chandpur. However, in the future, we would like to collaborate with the entire Chittagong division. We are currently developing a website to forecast crop production, but we intend to develop a mobile application for this project in the future, which will aid farmers in prediction crop production. We can operate on a broad scale to make our agricultural sector more digital if we have data for the entire Bangladesh.

References

- [1] P Patil, V Panpatil, PS Kokate “Crop Prediction System using Machine Learning Algorithms”, [Online]. Available:
<https://www.irjet.net/archives/V7/i2/IRJET-V7I2163.pdf>. [02 | Feb 2020]
- [2] YJN Kumar, V Spandana, VS Vaishnavi “Supervised Machine learning Approach for Crop Yield Prediction in Agriculture Sector”. [Online]. Available:
<https://ieeexplore.ieee.org/abstract/document/9137868>. [Accessed 10 July 2020].
- [3] Rushika Ghadge, Juilee Kulkarni “Prediction of Crop Yield using Machine Learning”, [Online]. Available:
<https://d1wqtxts1xzle7.cloudfront.net/56030057/IRJET-V5I2479->. [02| Feb 2018]
- [4] SS Kali, Ps Patil “A Machine Learning Approach to Predict Crop Yield and Success Rate”. [Online]. Available:
https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=A+Machine+Learning+Approach+to+Predict+Crop+Yield+and+Success+Rate&btnG=. [Accessed 2 June 2020].
- [5] Md. Tahmid Shakoor, Karishma Rahman, Sumaiya Nasrin Rayta, Amitabha Chakrabarty “Agricultural Production Output Prediction Using Supervised Machine Learning Techniques”, [Online]. Available:
https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Agricultural+Production+Output+Prediction+Using+Supervised+Machine+Learning+Techniques.+&btnG=. [24 August 2017]
- [6] Thomasvan Klompenburga, AyalewKassahuna, Cagatay Catal “Crop prediction using machine learning in this paper”. [Online]. Available:
<https://ieeexplore.ieee.org/abstract/document/9214190>. [Accessed 06 October 2020].
- [7] Aruvansh Nigam, Saksham Garg, Archit Agrawal, Parul Agrawal “Crop yield prediction using machine”, [Online]. Available: Learning.International journal of science and research(IJSR)
<https://www.researchgate.net/publication/340594772>. [April 2020]
- [8] R Medar, VS Rajpurohit “Prediction of Crop Yield using Machine Learning”. [Online]. Available:
https://www.academia.edu/36137300/Prediction_of_Crop_Yield_using_Machine_Learning?from=cover_page. [Accessed 02 Feb 2018].
- [9] PS Nishant, PS Venkat, BL Avinash “Crop Yield Prediction based on Indian Agriculture using Machine Learning”, [Online]. Available:
<https://ieeexplore.ieee.org/abstract/document/9154036>. [03 August 2020]
- [10] SS Dahikar, SV Rode “Agricultural Crop Yield Prediction Using Artificial Neural Network Approach”. [Online]. Available:
<https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.429.1195&rep=rep1&type=pdf>
[Accessed 10 July 2020].
- [11] “ AN APPROACH FOR PREDICTION OF CROP YIELD USING MACHINE LEARNING AND BIG DATA TECHNIQUES”, [Online]. Available:
<https://d1wqtxts1xzle7.cloudfront.net/60341957/ijcet.10.3.2019->. [03, May-June 2019]

APPENDICES

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
from sklearn.metrics import confusion_matrix
```

```
from google.colab import files
uploaded = files.upload()
```

```
import io
vstable = pd.read_csv(io.BytesIO(uploaded['Crops production.csv']))
```

```
#Remove Null Value
prod = vstable['Production']
pro = prod.mean()
import math
integer_value = math.floor(pro)
prod.fillna(integer_value, inplace = True)
vstable.isnull().sum()
vstable.shape
```

```
vs_records = vstable[['District Name', 'Sub_District Name', 'Crop Name', 'Area', 'Production']]
```

```
vs_target= vs_records.Production
```

```
vs_target[0:5]
```

```
vs_matrix = pd.get_dummies(vs_records[['District Name', 'Sub_District Name', 'Crop Name', 'Area', 'Production']])
print(vs_matrix)
```

```
from sklearn.model_selection import train_test_split
vs_train, vs_test, vs_target_train, vs_target_test = train_test_split(vs_matrix, vs_target, test_size=0.20, random_state=0)
```

```
len(vs_train)
```

```
len(vs_test)
```

```

from sklearn import preprocessing
min_max_scaler = preprocessing.MinMaxScaler().fit(vs_train)
vs_train_norm = min_max_scaler.transform(vs_train)
vs_train_norm = pd.DataFrame(vs_train_norm, columns=vs_train.columns)

vs_test_norm = min_max_scaler.transform(vs_test)
vs_test_norm = pd.DataFrame(vs_test_norm, columns=vs_test.columns)

```

```

print(vs_train_norm)

```

```

# import the random_forest regressor
from sklearn.ensemble import RandomForestRegressor

# create regressor object
randomfst = RandomForestRegressor(n_estimators = 10, random_state = 0)

# fit the regressor with x and y data
randomfst.fit(vs_train_norm, vs_target_train)

#score
print("Score Train", randomfst.score(vs_train_norm, vs_target_train))
print("Score Test", randomfst.score(vs_test_norm, vs_target_test))

```

```

plt.figure(figsize=(10,5))
plt.title('Crop Production')
plt.ylabel('Crop Name')
plt.xlabel('Production')
sns.scatterplot(vstable['Production'],vstable['Crop Name'])
plt.show()

```

```

vs_records.hist(figsize=(15,30),layout=(9,3))

```

```

plt.figure(figsize=(10,5))
plt.xlim(0,100000)
plt.xlabel('Area')
plt.ylabel('Production')
sns.kdeplot(vs_records['Area'],shade=True)
plt.show()

```

```
import matplotlib.pyplot as plt
from matplotlib.pyplot import figure
import pandas as pd
```

```
data = pd.read_csv('Crops production.csv')
```

```
df = pd.DataFrame(data)
```

```
X = list(df.iloc[:, 1])
Y = list(df.iloc[:, 4])
```

```
# Plot the data using bar() method
plt.bar(X, Y, color='g')
plt.title("Crop Production")
plt.xlabel("Sub_District Name")
plt.ylabel("Production")
# figure(figsize=(8, 6), dpi=80)
# Show the plot
plt.rcParams['figure.figsize'] = [40,20]
plt.savefig("High resolution.png",dpi=300)
plt.show()
```

```
# bar chart for Aman Rice
X = list(df.iloc[0:97, 1])
Y = list(df.iloc[0:97, 4])
```

```
# Plot the data using bar() method
plt.bar(X, Y, color='g')
plt.title("bar chart for Aman Rice")
plt.xlabel("Sub_District Name")
plt.ylabel("Production")
# figure(figsize=(8, 6), dpi=80)
# Show the plot
plt.rcParams['figure.figsize'] = [20,20]
plt.savefig("Amon bar chart.png",dpi=300)
plt.show()
```

plagiarism report

Submission date: 30-Dec-2021 11:56AM (UTC+0600)

Submission ID: 1736397858

File name: Project-ID-FL20D170_report.pdf (1,020.59K)

Word count: 4149

Character count: 22659

ORIGINALITY REPORT

25%

SIMILARITY INDEX

23%

INTERNET SOURCES

7%

PUBLICATIONS

20%

STUDENT PAPERS

PRIMARY SOURCES

1

Submitted to Daffodil International University

Student Paper

12%

2

Submitted to Intercollege

Student Paper

3%

3

dspace.daffodilvarsity.edu.bd:8080

Internet Source

2%

4

iaeme.com

Internet Source

1%

5

Rui Wu, Jose T. Painumkal, John M. Volk, Siming Liu, Sushil J. Louis, Scott Tyler, Sergiu M. Dascalu, Frederick C. Harris. "Parameter estimation of nonlinear nitrate prediction model using genetic algorithm", 2017 IEEE Congress on Evolutionary Computation (CEC), 2017

Publication

1%

6

Submitted to Asia Pacific University College of Technology and Innovation (UCTI)

Student Paper

1%

7

www.ijcseonline.org