A DATA SCIENCE TECHNIQUE FOR CROPPING LOCALIZATION FROM THE WEATHER DATASET.

BY

Md. MINHAJUL ABEDIN
ID: 181-15-11140

Md. NURUL ISLAM
ID: 181-15-11160
AND

JOY RAY CHOWDHURY
ID: 181-15-11163

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

Supervised By
Dr. Fizar Ahmed
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised By
Sharmin Akter
Lecture
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2022

APPROVAL

This Project titled "**A data science technique for cropping localization from weather data set**", submitted by **Md. Minhajul Abedin**, **Md. Nurul Islam** and **Joy Ray Chowdhury** 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 03-01-22.

BOARD OF EXAMINERS

Dr. S.M Aminul Haque **Chairman**
Associate Professor and Associate Head
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Naznin Sultana **Internal Examiner**
Assistant Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Raja Tariqul Hasan Tusher **Internal Examiner**
Senior Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Farid

____

Dr. Dewan Md. Farid
Professor
Department of Computer Science and Engineering
United International University

External Examiner

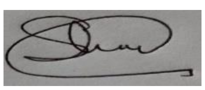
DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Dr. Fizar Ahmed,** Assistant Professor, **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:

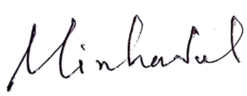
Dr. Fizar Ahmed
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised by:



Sharmin Akter
Lecture
Department of CSE
Daffodil International University

Submitted by:



Md. Minhajul Abedin
ID: -181-15-11140
Department of CSE
Daffodil International University

MD. Nurul Islam.

Md. Nurul Islam
ID: -181-15-11160
Department of CSE
Daffodil International University

Joy Ray

Joy Ray Chowdhury
ID: -181-15-11163
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 **Supervisor Dr. Fizar Ahmed**, **Assistant Professor**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Data Science*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts 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**, **Professor 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 an agricultural country, and it is the backbone of our nation. Agriculture employs over half of Bangladesh's people, and crops occupy more than 70% of the country's territory. More than 12 percent of revenue comes from the agriculture sector. But our lands are limited, and our population is growing day by day. That's why we need more food and increase the demand for crops continuously. So, it is a huge challenge to increase crop production. But our farmers face different types of problems during cultivation. They cannot justify which crop should be cultivated. Because of this, they did not get the expected yield. We know that machine learning plays a vital role in agricultural prediction. Crop prediction is a complex process. A massive amount of data is needed, like temperature, humidity, precipitation, wind speed, dew etc. In this system, we applied different types of machine learning algorithms and checked which algorithm gives us better accuracy. We get Random Forest gives us the best accuracy. So, we applied it to our dataset. We collect weather data from NASA Power Access Viewer. And crop data from different sources. Then we apply training and testing to this dataset. We took 80% data for training and 20% for testing. We get 91% accuracy from the Random Forest algorithm, which will help the farmer to decide which crop should be cultivated. It will increase crop production. It Removes hunger and poverty from our country.

TABLE OF CONTENTS

CONTENTS **PAGE**

CHAPTER **PAGE**

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Introduction

Since independence, Bangladesh's agriculture has grown at a phenomenal rate. This is true in terms of both yield and output growth in the cereal industry. Bangladesh's food output has kept up with the country's population increase. Despite continued population expansion, the government has achieved aggregate food self-sufficiency and improved calorie availability. Access to food has been enhanced on average.

In the Global Food Security Index, Bangladesh was rated 83rd out of 113 nations in 2019. Bangladesh has the lowest ranking of all the South Asian countries. Bangladesh's results in terms of food affordability and availability have been good on a more disaggregated level. To ensure food security, we have to increase production. For this, we should use smart agriculture system.

We know that our soil is very fertile for cultivation. We can easily grow any crop. This is blessing for us. If we properly utilize our land then we can ensure our food security. For this we have to digitalize our agricultural system. Most of our farmer face problems in selecting crop. If they cultivate right crop in right time, we get much better output than before. If we use digital technology on this, we can easily predict that which crop should be cultivate in which area in a particular time.

If the agronomist knows the accurate information about the crop it increases the production and minimizes the loss. In this project we make a system which will predict which crop is suitable for particular location.

1.2 Motivation

Bangladesh is an agricultural county. Most of our people are connected with agriculture. They do agriculture for their living. But most of the farmers are illiterate and poor. They don't know about the right way of cultivation. They follow the ancient path. Because of that, they didn't get the expected output. They face a huge loss every year because of this. We know that most of our farmers borrow money for cultivation. If they cannot get the proper output, they have to suffer the most. Because of it, sometimes they commit suicide.

Bangladesh is the largest delta island in the world. There is the Bay of Bengal in the south part of our country. Being coastal, we face many natural disasters every year. Every we face floods, cyclones, tidal waves, and drought. Because of all of these, our farmers have to face considerable losses every year.

As our farmers are illiterate, they cannot select the right time of cultivation. We know that if we don't cultivate properly, we cannot get the best production. Also, if we don't produce in proper time, there will be a huge chance of damage for a natural disaster. Every year we see that happen, and our farmers suffer.

Bangladesh is an overpopulated county. But our land is limited. So, if we want to ensure food security for our growing population, we have to increase crop production. It is not possible if we don't adopt an intelligent agricultural system. So that's why we decided to make this crop localization system.

1.3 Objective

Most of the people in Bangladesh are involved in agricultural work. This agriculture sector provides us the largest part of food supplies and gives us many ecosystem services. The agricultural process is a complex interconnected matrix of soil, plants, animal equipment, power, labor, capital, and other inputs that is managed in part by farm families and impacted by a variety of political, economic, institutional, and social influences at several levels. The main objective of our project is:

- If we complete this project, we can increase the yield of the crop.
- It will help our farmers to select the right crop.
- This system will be a substitute for traditional farming methods.
- It removes poverty from our society.
- It will increase our food security.
- It will increase our GDP.

1.4 Expected Outcome

We have worked on weather and some crop data in three areas in this project. Based on the accuracy we have obtained using some algorithms in these data, we can say which of the three areas has better climate and which crop yields will be higher.

1.5 Report layout

This research paper contains the following contents as given below:

- Chapter **one** explains the introduction of the research with its motivation Rationale of the study, research questions, and expected outcome.
- Chapter **two** discusses related works, research summary and challenges.
- Chapter **three** contains the workflow of this research, data collection procedure, data processing and statistical analysis and feature implementation.
- Chapter **four** covers experimental results and some relevant discussions, the accuracy model of research via numerically and graphically, algorithm comparison and prediction.
- Chapter **five** contains a summary of this research work along with the limitation and future work.

CHAPTER 2

BACKGROUND

2.1 Related Works & research summary

Different strategies and methodologies have evolved into an effective crop prediction. A lot of references are drawn from different case studies, also it discusses the many crops forecast techniques in use. The majority of the concepts deal with crop-producing approaches including sensors, support vector machines, and big data. These concepts are utilized to anticipate the crop properly.

In [1] This research focuses on crop prediction and yield computation, With the assistance of machine learning algorithms. They work on past Indian Weather data and get best accuracy using Random Forest algorithm. Which will help increase the crop production rate of that area.

In [2] this project the estimating crop production using machine learning technique depend on location, temperature and season. This project is done with 30 districts of India. They get about 98% accuracy. It gives details by mentioning whether or not the crop is lucrative. Which will help the farmer to choose right crop.

In This [3] project work on downfall, ground wetness, temperate environmental pressure to predict which crop will be suitable. They use Naïve bayes and Random Forest algorithm.

In this [4] discovered that Machine Learning algorithms might predict a goal using Supervised Learning. The goal of this research is to estimate crop yields using supervised learning systems. To get the required outcomes, you'll need to build an appropriate function consisting of a set of variables that can transform the input variable into the intended output. According to the study, the forecasts might be generated using the Random Forest ML approach, which promises to produce the best accurate crop forecast with the lowest number of models.

In [5] To anticipate crop yield, they have used Regression analysis like a speculate modeling approach. They made a system so that farmers yield better and get more profit. Here they applied different machine learning algorithms on a specific area, average temperature and area. According to find, they speculate which crop will produce more.

In. [6] They found when a dataset with more characteristics is utilized, the accuracy rate improves. Compared to other algorithms such as Decision tree and Linear regression, they find Random Forest is better, and they apply it for speculation. This dataset has a large number of variables which allows for many precise predictions.

In This [7] work assists agronomists in determining which crop to produce in a specific location at a particular time and determining is economical or non-economical. It goes into great depth on whether or not the crop is economical. Consequently, this technology helps agronomists make better decisions and saves them time.

In [8] The research offered multiple machine learning techniques for forecasting crop production, On the basis of temperature, rainfall, season, and area. They did an experiment on their dataset and found that a Random Forest Algorithm is the best for predicting crop yield. It will help agronomists for getting much production.

2.2 Challenges

Our first hurdle was collecting weather data. But I have faced even bigger problems in collecting crop data and working on some crop data. Then we had another hurdle in determining which algorithms to use in these data. We needed a lot of resources to work which we had to face a lot of problems to find. We had to get the job done with very little resources.

CHAPTER 3
RESEARCH AND METHODOLOGY

3.1 Introduction

This chapter contains all of our research's theoretical data. Anyone who reads our thesis will have a clear idea of what we believe. We quickly gathered some vital facts to make it roomier to comprehend. Data is critical when it comes to learning machine learning or data mining. So, let's take a quick look at the data collecting process. Statistical analysis also includes computations and visualizations of all data sets to provide a more logical explanation. We will present a detailed picture of our crop prediction approach in the context of Workflow. Apart from that, this chapter concludes by interpreting the implementation requirements in detail.

3.2 Data Collection

We need weather and crop data for crop localization. Through which we can determine which crop is better in which place. That's why we took weather data from NASA Power Access Viewer and collected data on certain crops from different sources. We have collected data from three areas for this work. We have collected a total of 12,000 weather data and data of 5 crops from different sources for the last 10 years in each area. We collected our data based on the following factors:

- Precipitation
- Relative Humidity
- Dew
- Wind Speed
- Minimum Temperature
- Maximum Temperature
- Temperature

3.3 Research Subject and Instrumentation

Machine Learning is the greatest technology for providing a more effective actual solution of agricultural productivity. To selecting an algorithm for use, first we applied different machine learning algorithm and compare which one will fit and give us best result on our dataset. We used Decision Tree, K-Nearest Neighbor, Naïve Bayes, Support Vector Machine and Random Forest algorithm. We used python as programming language. We also used Pandas, Seaborn, NumPy and Matplotlib. We used Google Collaboratory and Microsoft Excel as research tool.

3.4 Proposed Methodology

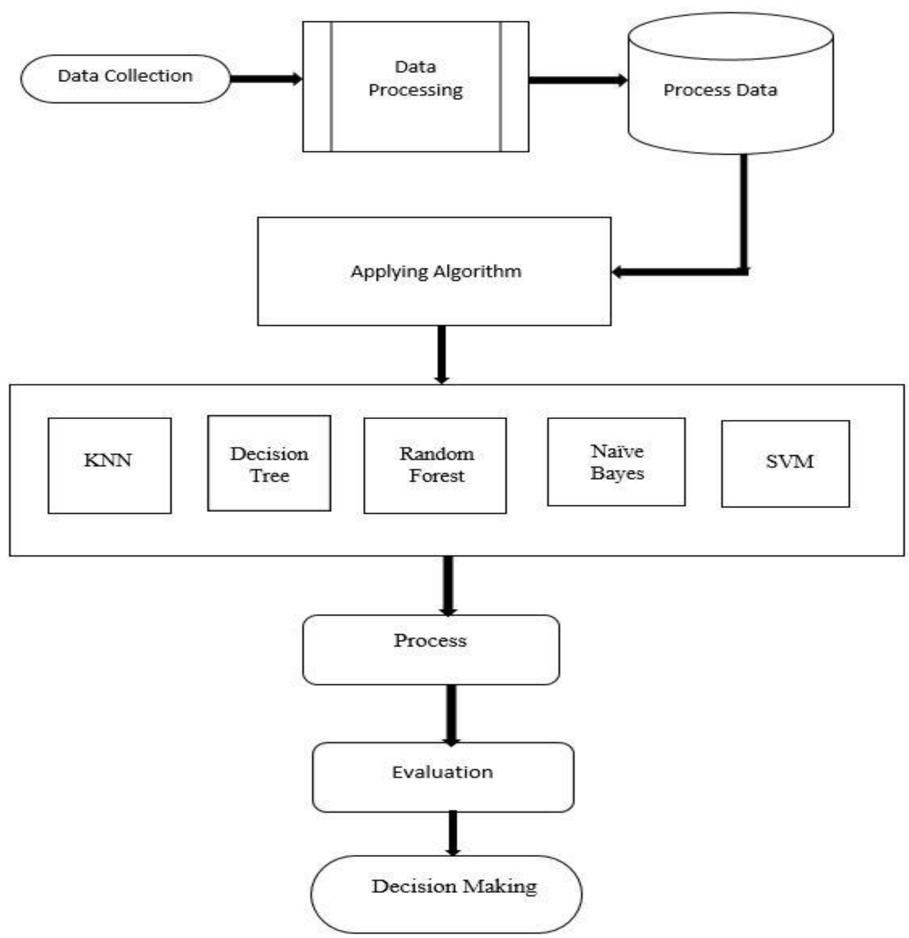


Figure 3.1: Proposed Methodology

3.5 Data Preprocessing

The method we used to convert the raw data into clean data is called data preprocessing. We know that real-world data have a lot of noise, missing value, null value, unorganized format. We can't use it directly into the machine learning model. That's why we need to preprocess data so that we can used it on machine learning. So, this clean data accepted by machine learning and give us better accuracy and efficiency. So, we replaced missing value from our dataset. We removed all the null values and made it understandable.

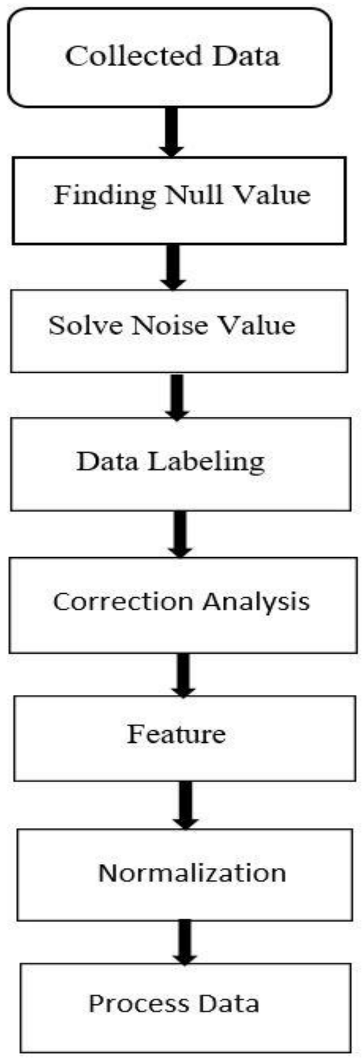


Figure 3.2: Data Preprocessing.

3.6 Statistical Analysis

CROP DATA

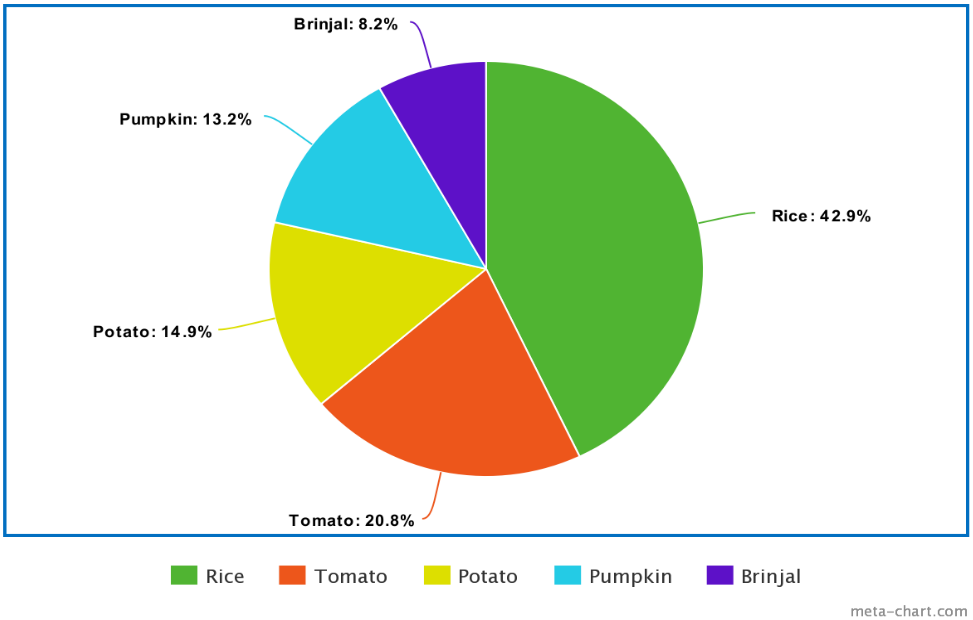


Figure 3.3: Pie chart of Crop Data.

PRECIPITATION

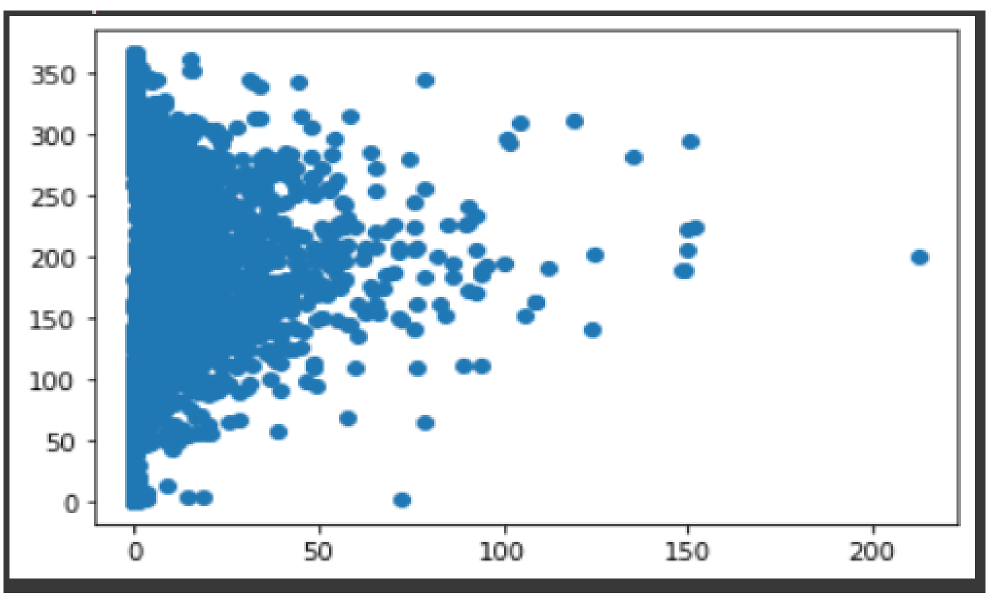


Figure 3.4: Scatter Plot of Precipitation.

DEW

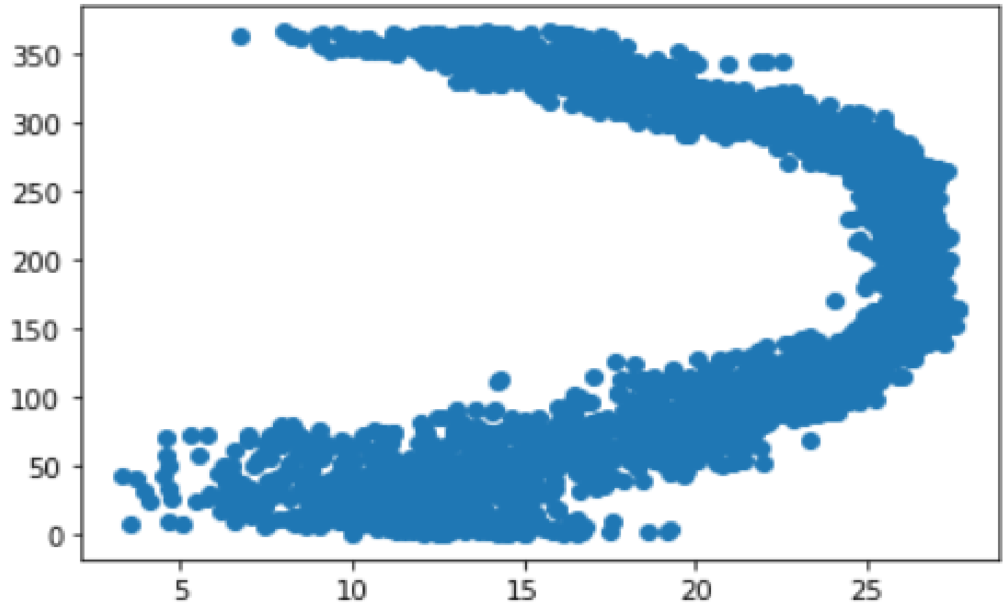


Figure 3.5: Scatter Plot of Dew

HUMIDITY

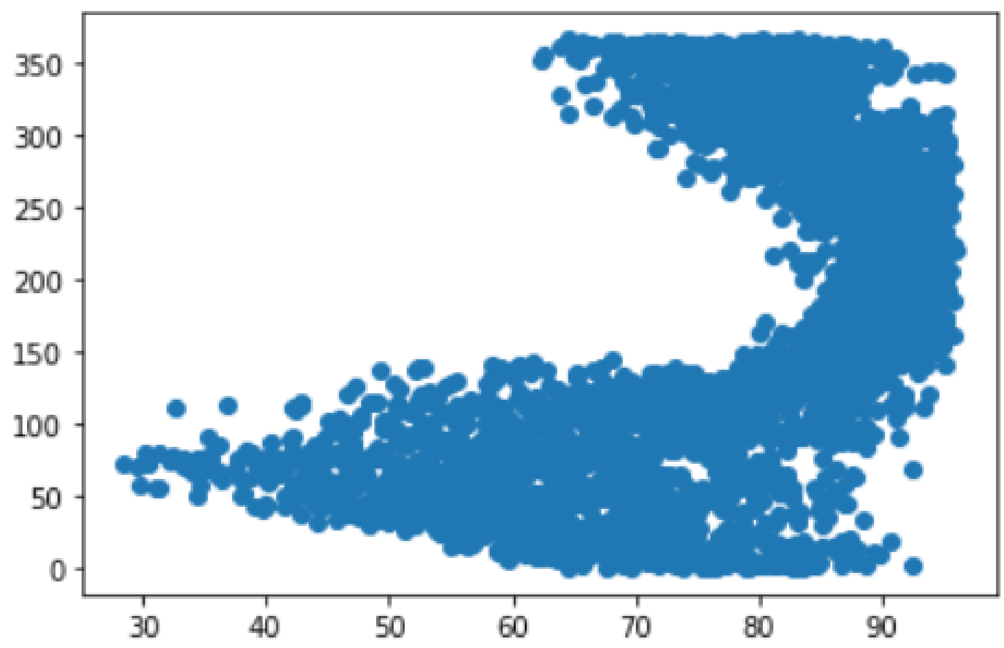


Figure 3.6: Scatter Plot of Humidity.

WIND SPEED

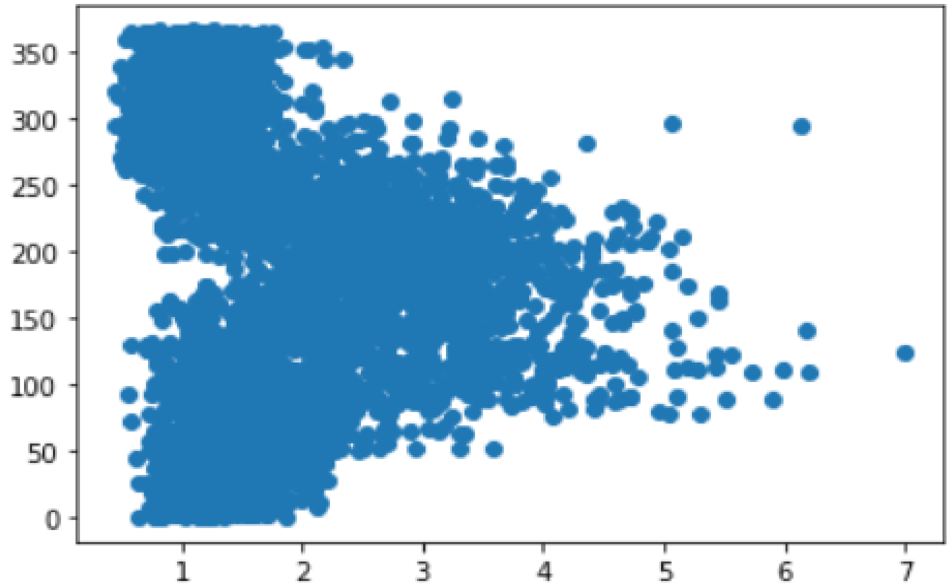


Figure 3.7: Scatter Plot of Wind Speed.

TEMPERATURE

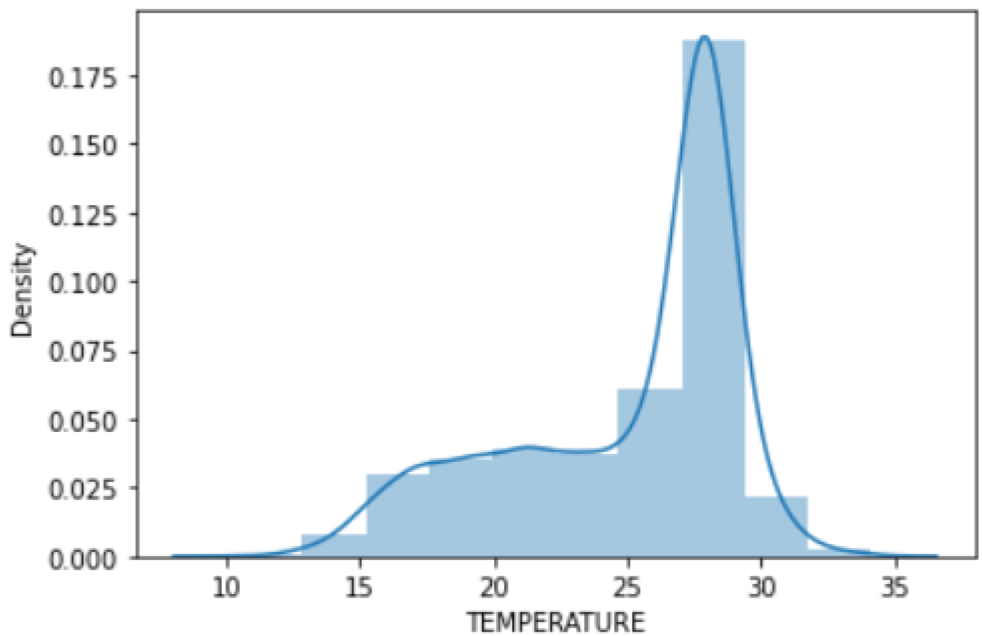


Figure 3.8: Dist Plot of Temperature.

3.7 Implementation Requirements

After collecting all the data and analyzing it theoretically, some tools are needed to get the job done. Which helps in crop localization through the algorithms used for our work. We will need some tools for this unique research work. Here are some of the tools we need:

Hardware/Software Requirements:

- Operating System (Windows 8.1)
- Hard Disk (128 GB)
- Ram (2 GB) s
- Anaconda Navigator / Google Colab
- Microsoft Excel

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Introduction

In Chapter 3, we discuss the dataset and data processing process. After this data processing, we apply some machine learning algorithms in this dataset, and the result of that algorithm is discussed in this chapter. We used Decision Tree, K-Nearest Neighbor, Naïve Bayes, Support Vector Machine and Random Forest algorithm. We found the accuracy of the dataset for analysis to see which algorithm is better for this dataset. We discuss this analysis in the sector and give details of the work.

4.2 Experimental Results & Analysis

We have divided our dataset into three district-based sections. We used different algorithms on the weather and crop data of the three districts and compared the accuracy of the algorithms. Now the results of that experiment have been analyzed based on accuracy.

We divided the dataset into training set and testing set. We divided the dataset into ratio,

- Training set 60% and testing set 40%
- Training set 80% and testing set 20%

4.3 Accuracy model

We will demonstrate the confusion matrix for our model for the classifier (KNN, SVM, Decision tree, Random Forest, Naïve Bayes) that we apply in our study in this part. The confusion matrix is a table used to describe the performance of a classification model or Classifier.

- **KNN Classifier**

K-Nearest Neighbor is most widely used supervised machine learning algorithm. At first, it takes the number of k. then its calculated k numbers of neighbors by using Euclidean distance. It selects the nearest neighbor of k based on Euclidean distance. From this k nearest neighbor, it counts the number of data points of every category. After that, it assigns the new data point to the category for which the neighbor is maximum.

Dataset of Cumilla:

- For 80% training data and 20% testing data.

Table 4.1: Confusion Matrix of 80% training data for KNN classifier

Predicted/ True Class	**Rice**	**Tomato**	**Potato**	**Pumpkin**	**Brinjal**
Rice	**370**	**0**	**0**	**3**	**0**
Tomato	**1**	**93**	**4**	**14**	**0**
Potato	**0**	**5**	**100**	**1**	**5**
Pumpkin	**71**	**18**	**0**	**76**	**0**
Brinjal	**0**	**1**	**8**	**0**	**58**

Accuracy of KNN classifier of 80% training set = 84.00%

Table 4.2: Testing data of (20%) accuracy for KNN classifier

Crop Name	Precision	Recall	f1-score
Rice	84	99	91.00
Tomato	79	83	81.00
Potato	89	90	90.00
Pumpkin	81	46	59.00
Brinjal	92	87	89.00
Total/Average	85	81	82.00

- For 60% training data and 40% testing data.

Table 4.3: Confusion Matrix of 60% training data for KNN classifier

Predicted /True Class	**Rice**	**Tomato**	**Potato**	**Pumpkin**	**Brinjal**
Rice	**690**	**0**	**0**	**12**	**0**
Tomato	**5**	**200**	**10**	**29**	**0**
Potato	**0**	**17**	**200**	**1**	**13**
Pumpkin	**160**	**22**	**0**	**150**	**0**
Brinjal	**0**	**1**	**29**	**0**	**120**

Accuracy of KNN classifier of 60% training set = 82.00%

Table 4.4: Testing data of (40%) accuracy for KNN classifier

Crop Name	Precision	Recall	f1-score
Rice	80	98	88.00
Tomato	83	82	83.00
Potato	84	86	85.00
Pumpkin	78	45	57.00
Brinjal	90	79	84.00
Total/Average	83	78	79.00

Dataset of Sathkhira:

- For 80% training data and 20% testing data.

Table 4.5: Confusion Matrix of 80% training data for KNN classifier

Predicted/ True Class	Rice	Tomato	Potato	Pumpkin	Brinjal
Rice	**430**	**0**	**2**	**0**	**0**
Tomato	**0**	**100**	**5**	**9**	**0**
Potato	**55**	**9**	**60**	**0**	**0**
Pumpkin	**0**	**11**	**0**	**82**	**0**
Brinjal	**0**	**0**	**0**	**7**	**55**

Accuracy of KNN classifier of 80% training set = 88.00%

Table 4.6: Testing data of (20%) accuracy for KNN classifier

Crop Name	Precision	Recall	f1-score
Rice	89	100	94.00
Tomato	84	88	86.00
Potato	90	48	63.00
Pumpkin	84	88	86.00
Brinjal	100	89	94.00
Total/Average	89	83	85.00

- For 60% training data and 40% testing data.

Table 4.7: Confusion Matrix of 60% training data for KNN classifier

Predicted /True Class	Rice	Tomato	Potato	Pumpkin	Brinjal
Rice	**840**	**0**	**7**	**0**	**0**
Tomato	**5**	**200**	**12**	**24**	**0**
Potato	**100**	**22**	**110**	**0**	**0**
Pumpkin	**0**	**15**	**0**	**190**	**3**
Brinjal	**0**	**0**	**0**	**16**	**110**

Accuracy of KNN classifier of 60% training set = 88.00%

Table 4.8: Testing data of (40%) accuracy for KNN classifier

Crop Name	Precision	Recall	f1-score
Rice	89	99	94.00
Tomato	85	83	84.00
Potato	85	47	60.00
Pumpkin	83	91	87.00
Brinjal	97	87	92.00
Total/Average	88	82	83.00

Dataset of Jassore:

- For 80% training data and 20% testing data.

Table 4.9: Confusion Matrix of 80% training data for KNN classifier

Predicted/ True Class	Rice	Tomato	Potato	Pumpkin	Brinjal
Rice	**400**	**0**	**5**	**0**	**0**
Tomato	**1**	**100**	**3**	**10**	**0**
Potato	**49**	**5**	**82**	**0**	**0**
Pumpkin	**0**	**10**	**1**	**75**	**4**
Brinjal	**0**	**0**	**0**	**11**	**68**

Accuracy of KNN classifier of 80% training set = 88.00%

Table 4.10: Testing data of (20%) accuracy for KNN classifier

Crop Name	Precision	Recall	f1-score
Rice	89	99	94.00
Tomato	88	88	88.00
Potato	90	60	72.00
Pumpkin	78	83	81.00
Brinjal	94	86	90.00
Total/Average	88	83	85.00

- For 60% training data and 40% testing data.

Table 4.11: Confusion Matrix of 60% training data for KNN classifier

Predicted /True Class	Rice	Tomato	Potato	Pumpkin	Brinjal
Rice	**770**	**0**	**15**	**0**	**0**
Tomato	**5**	**200**	**16**	**12**	**0**
Potato	**120**	**13**	**140**	**0**	**0**
Pumpkin	**0**	**19**	**1**	**170**	**13**
Brinjal	**0**	**0**	**0**	**34**	**130**

Accuracy of KNN classifier of 60% training set = 85.00%

Table 4.12: Testing data of (40%) accuracy for KNN classifier

Crop Name	Precision	Recall	f1-score
Rice	86	98	92.00
Tomato	86	86	86.00
Potato	82	52	64.00
Pumpkin	79	84	81.00
Brinjal	91	80	85.00
Total/Average	85	80	82.00

➢ SVM Classifier

SVM means Support Vector Machine. In supervised machine learning algorithm support vector machine is mostly used. In high-dimensional spaces, it works efficiently. SVM classifier creates a model that assigns incoming instances to one of two categories and making it non probabilistic binary linear classifier.

Dataset of Cumilla:

- For 80% training data and 20% testing data.

Table 4.13: Confusion Matrix of 80% training data for SVM classifier

Predicted/ True Class	**Rice**	**Tomato**	**Potato**	**Pumpkin**	**Brinjal**
Rice	**360**	**0**	**0**	**16**	**0**
Tomato	**6**	**90**	**5**	**11**	**0**
Potato	**2**	**6**	**97**	**0**	**7**
Pumpkin	**53**	**15**	**0**	**97**	**0**
Brinjal	**1**	**0**	**7**	**0**	**59**

Accuracy of SVM classifier of 80% training set = 84.00%

Table 4.14: Testing data of (20%) accuracy for SVM classifier

Crop Name	**Precision**	**Recall**	**f1-score**
Rice	85	96	90.00
Tomato	81	80	81.00
Potato	89	87	88.00
Pumpkin	78	59	67.00
Brinjal	89	88	89.00
Total/Average	85	82	83.00

- For 60% training data and 40% testing data.

Table 4.15: Confusion Matrix of 60% training data for SVM classifier

Predicted/ True Class	Rice	Tomato	Potato	Pumpkin	Brinjal
Rice	**670**	**0**	**0**	**32**	**0**
Tomato	**11**	**200**	**7**	**31**	**0**
Potato	**9**	**14**	**190**	**0**	**13**
Pumpkin	**110**	**17**	**0**	**210**	**0**
Brinjal	**6**	**0**	**24**	**0**	**120**

Accuracy of SVM classifier of 60% training set = 83.00%

Table 4.16: Testing data of (40%) accuracy for SVM classifier

Crop Name	Precision	Recall	f1-score
Rice	83	95	89.00
Tomato	86	80	83.00
Potato	86	84	85.00
Pumpkin	77	61	68.00
Brinjal	90	79	84.00
Total/Average	84	80	82.00

Dataset of Sathkhira:

- For 80% training data and 20% testing data.

Table 4.17: Confusion Matrix of 80% training data for SVM classifier

Predicted/ True Class	Rice	Tomato	Potato	Pumpkin	Brinjal
Rice	**420**	**0**	**12**	**0**	**0**
Tomato	**5**	**100**	**6**	**8**	**0**
Potato	**35**	**6**	**83**	**0**	**0**
Pumpkin	**5**	**4**	**0**	**82**	**2**
Brinjal	**1**	**0**	**0**	**6**	**55**

Accuracy of SVM classifier of 80% training set = 89.00%

Table 4.18: Testing data of (20%) accuracy for SVM classifier

Crop Name	Precision	Recall	f1-score
Rice	90	97	94.00
Tomato	91	84	87.00
Potato	82	67	74.00
Pumpkin	85	88	87.00
Brinjal	96	89	92.00
Total/Average	89	85	87.00

- For 60% training data and 40% testing data.

Table 4.19: Confusion Matrix of 60% training data for SVM classifier

Predicted /True Class	**Rice**	**Tomato**	**Potato**	**Pumpkin**	**Brinjal**
Rice	**830**	**0**	**19**	**0**	**0**
Tomato	**17**	**200**	**13**	**13**	**0**
Potato	**61**	**13**	**160**	**0**	**0**
Pumpkin	**6**	**10**	**0**	**180**	**7**
Brinjal	**3**	**0**	**0**	**16**	**110**

Accuracy of SVM classifier of 60% training set = 89.00%

Table 4.20: Testing data of (40%) accuracy for SVM classifier

Crop Name	**Precision**	**Recall**	**f1-score**
Rice	91	98	94.00
Tomato	90	82	83.00
Potato	83	69	75.00
Pumpkin	86	89	88.00
Brinjal	94	85	89.00
Total/Average	89	84	86.00

Dataset of Jessore:

- For 80% training data and 20% testing data.

Table 4.21: Confusion Matrix of 80% training data for SVM classifier

Predicted/ True Class	Rice	Tomato	Potato	Pumpkin	Brinjal
Rice	**390**	**0**	**14**	**0**	**0**
Tomato	**5**	**99**	**5**	**12**	**0**
Potato	**32**	**2**	**100**	**0**	**0**
Pumpkin	**5**	**6**	**0**	**78**	**1**
Brinjal	**3**	**0**	**0**	**11**	**65**

Accuracy of SVM classifier of 80% training set = 88.00%

Table 4.22: Testing data of (20%) accuracy for SVM classifier

Crop Name	Precision	Recall	f1-score
Rice	90	97	93.00
Tomato	93	82	87.00
Potato	84	75	79.00
Pumpkin	77	87	82.00
Brinjal	98	82	90.00
Total/Average	88	84	86.00

- For 60% training data and 40% testing data.

Table 4.23: Confusion Matrix of 60% training data for SVM classifier

Predicted /True Class	Rice	Tomato	Potato	Pumpkin	Brinjal
Rice	**760**	**0**	**28**	**0**	**0**
Tomato	**13**	**200**	**17**	**12**	**0**
Potato	**87**	**9**	**180**	**0**	**0**
Pumpkin	**9**	**13**	**0**	**170**	**13**
Brinjal	**4**	**0**	**0**	**24**	**140**

Accuracy of SVM classifier of 60% training set = 86.00%

Table 4.24: Testing data of (40%) accuracy for SVM classifier

Crop Name	Precision	Recall	f1-score
Rice	87	96	91.00
Tomato	90	82	86.00
Potato	80	65	71.00
Pumpkin	82	83	83.00
Brinjal	92	83	87.00
Total/Average	86	82	84.00

➢ Decision Tree Classifier

Decision Tree Classifier is a tree based supervised algorithm which is used for both classification and regression. It is best for classification problems First of all, this algorithm chose a target attribute. Then it calculates the information gain of that target attribute. After that, it computes the entropy of other attributes using the formula of entropy. Subtract entropy from the information gain of each attribute to find out the gain. The value of entropy maintains zero to one. If the value is zero, it will we leaf node, and if the value is one, then basically says that it is completely subset.

Dataset of Cumilla:

- For 80% training data and 20% testing data.

Table 4.25: Confusion Matrix of 80% training data for Decision Tree Classifier

Predicted/ True Class	**Rice**	**Tomato**	**Potato**	**Pumpkin**	**Brinjal**
Rice	**360**	**0**	**0**	**21**	**0**
Tomato	**0**	**100**	**2**	**9**	**0**
Potato	**0**	**4**	**100**	**0**	**6**
Pumpkin	**43**	**21**	**0**	**100**	**0**
Brinjal	**0**	**0**	**15**	**0**	**52**

Accuracy of Decision Tree Classifier of 80% training set = 85.00%

Table 4.26: Testing data of (20%) accuracy for Decision Tree Classifier

Crop Name	Precision	Recall	f1-score
Rice	89	94	92.00
Tomato	80	90	85.00
Potato	86	91	88.00
Pumpkin	77	61	68.00
Brinjal	90	78	83.00
Total/Average	84	83	83.00

- For 60% training data and 40% testing data.

Table 4.27: Confusion Matrix of 60% training data for Decision Tree classifier

Predicted /True Class	Rice	Tomato	Potato	Pumpkin	Brinjal
Rice	**690**	**0**	**0**	**14**	**0**
Tomato	**0**	**220**	**8**	**21**	**0**
Potato	**0**	**11**	**210**	**0**	**6**
Pumpkin	**120**	**35**	**0**	**180**	**0**
Brinjal	**0**	**0**	**34**	**0**	**110**

Accuracy of Decision Tree classifier of 60% training set = 85.00%

Table 4.28: Testing data of (40%) accuracy for Decision Tree classifier

Crop Name	Precision	Recall	f1-score
Rice	85	98	91.00
Tomato	82	88	85.00
Potato	83	93	88.00
Pumpkin	84	53	65.00
Brinjal	95	77	85.00
Total/Average	86	82	83.00

Dataset of Satkhira:

- For 80% training data and 20% testing data.

Table 4.29: Confusion Matrix of 80% training data for Decision Tree classifier

Predicted/ True Class	Rice	Tomato	Potato	Pumpkin	Brinjal
Rice	**420**	**0**	**17**	**0**	**0**
Tomato	**0**	**110**	**3**	**2**	**0**
Potato	**14**	**3**	**110**	**0**	**0**
Pumpkin	**0**	**4**	**0**	**89**	**0**
Brinjal	**0**	**0**	**0**	**7**	**55**

Accuracy of Decision Tree classifier of 80% training set = 94.00%

Table 4.30: Testing data of (20%) accuracy for Decision Tree classifier

Crop Name	Precision	Recall	f1-score
Rice	97	96	96.00
Tomato	94	96	95.00
Potato	84	86	85.00
Pumpkin	91	96	93.00
Brinjal	100	89	94.00
Total/Average	93	93	93.00

- For 60% training data and 40% testing data.

Table 4.31: Confusion Matrix of 60% training data for Decision Tree classifier

Predicted /True Class	**Rice**	**Tomato**	**Potato**	**Pumpkin**	**Brinjal**
Rice	**820**	**0**	**35**	**0**	**0**
Tomato	**0**	**230**	**10**	**6**	**0**
Potato	**20**	**8**	**210**	**0**	**0**
Pumpkin	**0**	**5**	**0**	**190**	**15**
Brinjal	**0**	**0**	**0**	**13**	**110**

Accuracy of Decision Tree classifier of 60% training set = 93.00%

Table 4.32: Testing data of (40%) accuracy for Decision Tree classifier

Crop Name	**Precision**	**Recall**	**f1-score**
Rice	98	96	97.00
Tomato	95	93	94.00
Potato	82	88	85.00
Pumpkin	91	90	91.00
Brinjal	88	90	89.00
Total/Average	91	91	91.00

Dataset of Jessore:

- For 80% training data and 20% testing data.

Table 4.33: Confusion Matrix of 80% training data for Decision Tree classifier

Predicted/ True Class	Rice	Tomato	Potato	Pumpkin	Brinjal
Rice	**370**	**0**	**32**	**0**	**0**
Tomato	**0**	**120**	**1**	**0**	**0**
Potato	**10**	**5**	**120**	**0**	**0**
Pumpkin	**0**	**13**	**0**	**77**	**0**
Brinjal	**0**	**0**	**0**	**13**	**66**

Accuracy of Decision Tree classifier of 80% training set = 91.00%

Table 4.34: Testing data of (20%) accuracy for Decision Tree classifier

Crop Name	Precision	Recall	f1-score
Rice	97	92	95.00
Tomato	87	99	93.00
Potato	79	89	83.00
Pumpkin	86	86	86.00
Brinjal	100	84	91.00
Total/Average	90	90	89.00

- For 60% training data and 40% testing data.

Table 4.35: Confusion Matrix of 60% training data for Decision Tree classifier

Predicted /True Class	Rice	Tomato	Potato	Pumpkin	Brinjal
Rice	**760**	**0**	**25**	**0**	**0**
Tomato	**0**	**220**	**12**	**6**	**0**
Potato	**62**	**7**	**200**	**0**	**0**
Pumpkin	**0**	**16**	**0**	**180**	**6**
Brinjal	**0**	**0**	**0**	**29**	**140**

Accuracy of Decision Tree classifier of 60% training set = 90.00%

Table 4.36: Testing data of (40%) accuracy for Decision Tree classifier

Crop Name	Precision	Recall	f1-score
Rice	92	97	95.00
Tomato	91	92	91.00
Potato	85	75	79.00
Pumpkin	84	89	86.00
Brinjal	96	83	89.00
Total/Average	89	87	88.00

➢ Naïve Bayes Classifier

For implementing analysis, Naive Bayes is a primary but surprisingly strong method. It is divided into two parts: Knave and bias. No Naive based model is simple to construct and is very beneficial when dealing with significant amounts of data.

Dataset of Cumilla:

- For 80% training data and 20% testing data.

Table 4.37: Confusion Matrix of 80% training data for Naïve Bayes classifier

Predicted/ True Class	Rice	Tomato	Potato	Pumpkin	Brinjal
Rice	350	0	0	23	0
Tomato	0	92	12	8	0
Potato	0	1	110	0	5
Pumpkin	64	24	0	77	0
Brinjal	0	0	11	0	56

Accuracy of Naive Bayes classifier of 80% training set = 82.00%

Table 4.38: Testing data of (20%) accuracy for Naive Bayes classifier

Crop Name	Precision	Recall	f1-score
Rice	85	94	89.00
Tomato	79	82	80.00
Potato	82	95	88.00
Pumpkin	71	47	56.00
Brinjal	92	84	88.00
Total/Average	82	80	80.00

- For 60% training data and 40% testing data.

Table 4.39: Confusion Matrix of 60% training data for Naive Bayes classifier

Predicted /True Class	Rice	Tomato	Potato	Pumpkin	Brinjal
Rice	**670**	**0**	**0**	**38**	**0**
Tomato	**0**	**190**	**29**	**23**	**0**
Potato	**0**	**3**	**210**	**0**	**13**
Pumpkin	**130**	**43**	**0**	**170**	**0**
Brinjal	**0**	**0**	**24**	**0**	**120**

Accuracy of Naive Bayes classifier of 60% training set = 82.00%

Table 4.40: Testing data of (40%) accuracy for Naive Bayes classifier

Crop Name	Precision	Recall	f1-score
Rice	84	95	89.00
Tomato	81	79	80.00
Potato	80	93	86.00
Pumpkin	73	50	59.00
Brinjal	90	84	87.00
Total/Average	82	80	80.00

Dataset of Satkhira:

- For 80% training data and 20% testing data.

Table 4.41: Confusion Matrix of 80% training data for Naive Bayes classifier

Predicted/ True Class	Rice	Tomato	Potato	Pumpkin	Brinjal
Rice	**420**	**0**	**17**	**0**	**0**
Tomato	**0**	**100**	**1**	**16**	**0**
Potato	**35**	**17**	**172**	**0**	**0**
Pumpkin	**0**	**4**	**0**	**86**	**3**
Brinjal	**0**	**0**	**0**	**3**	**59**

Accuracy of Naive Bayes classifier of 80% training set = 88.00%

Table 4.42: Testing data of (20%) accuracy for Naive Bayes classifier

Crop Name	Precision	Recall	f1-score
Rice	92	96	94.00
Tomato	83	86	94.00
Potato	80	58	67.00
Pumpkin	82	92	87.00
Brinjal	95	95	95.00
Total/Average	86	85	86.00

- For 60% training data and 40% testing data.

Table 4.43: Confusion Matrix of 60% training data for Naive Bayes classifier

Predicted /True Class	**Rice**	**Tomato**	**Potato**	**Pumpkin**	**Brinjal**
Rice	**800**	**0**	**49**	**0**	**0**
Tomato	**0**	**200**	**10**	**31**	**0**
Potato	**54**	**33**	**150**	**0**	**0**
Pumpkin	**0**	**7**	**0**	**190**	**13**
Brinjal	**0**	**0**	**0**	**9**	**120**

Accuracy of Naive Bayes classifier of 60% training set = 88.00%

Table 4.44: Testing data of (40%) accuracy for Naive Bayes classifier

Crop Name	**Precision**	**Recall**	**f1-score**
Rice	94	94	94.00
Tomato	84	83	83.00
Potato	71	63	67.00
Pumpkin	82	90	86.00
Brinjal	90	93	91.00
Total/Average	84	85	84.00

Dataset of Jessore:

- For 80% training data and 20% testing data.

Table 4.45: Confusion Matrix of 80% training data for Naive Bayes classifier

Predicted/ True Class	Rice	Tomato	Potato	Pumpkin	Brinjal
Rice	**380**	**0**	**23**	**0**	**0**
Tomato	**0**	**100**	**2**	**17**	**0**
Potato	**44**	**10**	**82**	**0**	**0**
Pumpkin	**0**	**4**	**0**	**81**	**5**
Brinjal	**0**	**0**	**0**	**7**	**72**

Accuracy of Naive Bayes classifier of 80% training set = 87.00%

Table 4.46: Testing data of (20%) accuracy for Naive Bayes classifier

Crop Name	Precision	Recall	f1-score
Rice	90	94	92.00
Tomato	88	84	86.00
Potato	77	60	67.00
Pumpkin	77	90	83.00
Brinjal	94	91	92.00
Total/Average	85	84	84.00

- For 60% training data and 40% testing data.

Table 4.47: Confusion Matrix of 60% training data for Naive Bayes classifier

Predicted /True Class	**Rice**	**Tomato**	**Potato**	**Pumpkin**	**Brinjal**
Rice	**730**	**0**	**50**	**0**	**0**
Tomato	**0**	**210**	**8**	**21**	**0**
Potato	**82**	**24**	**160**	**0**	**0**
Pumpkin	**0**	**9**	**0**	**180**	**18**
Brinjal	**0**	**0**	**0**	**17**	**150**

Accuracy of Naive Bayes classifier of 60% training set = 86.00%

Table 4.48: Testing data of (40%) accuracy for Naive Bayes classifier

Crop Name	**Precision**	**Recall**	**f1-score**
Rice	90	94	92.00
Tomato	86	88	87.00
Potato	74	61	67.00
Pumpkin	82	87	84.00
Brinjal	89	90	90.00
Total/Average	84	84	84.00

➢ Random Forest Classifier

Random Forest is a supervised algorithm. Random Forest algorithm combines several decision trees. This algorithm is an ensemble classifier-based classification algorithm. The dataset will be separated into two sets: training data and testing data. The decision tree is built using a larger training dataset. The model will create a decision tree using training data and extract the inferior node from the training data to generate an excellent model. Each training dataset will yield a decision tree, which will be followed by a random forest. The bagging method's basic premise is that aggregating mastering output will improve the final outcome. During training, the random forest algorithm creates several decision trees. The results of various decision trees' predictions will be compiled, and the ultimate output will be the one with the most votes. Jupiter notebook is a tool for building trained models with the Random Forest Algorithm.

Dataset of Cumilla:

- For 80% training data and 20% testing data.

Table 4.49: Confusion Matrix of 80% training data for Random Forest classifier

Predicted/ True Class	**Rice**	**Tomato**	**Potato**	**Pumpkin**	**Brinjal**
Rice	**350**	**0**	**0**	**22**	**0**
Tomato	**0**	**100**	**2**	**10**	**0**
Potato	**0**	**4**	**100**	**0**	**5**
Pumpkin	**42**	**16**	**0**	**110**	**0**
Brinjal	**0**	**0**	**11**	**0**	**56**

Accuracy of Random Forest classifier of 80% training set = 87.00%

Table 4.50: Testing data of (20%) accuracy for Random Forest classifier

Crop Name	Precision	Recall	f1-score
Rice	89	94	92.00
Tomato	82	89	86.00
Potato	89	92	90.00
Pumpkin	77	65	70.00
Brinjal	92	84	88.00
Total/Average	86	85	85.00

- For 60% training data and 40% testing data.

Table 4.51: Confusion Matrix of 60% training data for Random Forest classifier

Predicted /True Class	Rice	Tomato	Potato	Pumpkin	Brinjal
Rice	**660**	**0**	**0**	**41**	**0**
Tomato	**0**	**210**	**6**	**25**	**0**
Potato	**0**	**12**	**200**	**0**	**17**
Pumpkin	**88**	**23**	**0**	**230**	**0**
Brinjal	**0**	**0**	**20**	**0**	**130**

Accuracy of Random Forest classifier of 60% training set = 86.00%

Table 4.52: Testing data of (40%) accuracy for Random Forest classifier

Crop Name	Precision	Recall	f1-score
Rice	88	94	91.00
Tomato	86	87	87.00
Potato	88	87	88.00
Pumpkin	78	67	72.00
Brinjal	88	86	87.00
Total/Average	86	84	85.00

Dataset of Satkhira:

- For 80% training data and 20% testing data.

Table 4.53: Confusion Matrix of 80% training data for Random Forest classifier

Predicted/ True Class	Rice	Tomato	Potato	Pumpkin	Brinjal
Rice	**420**	**0**	**16**	**0**	**0**
Tomato	**0**	**120**	**1**	**2**	**0**
Potato	**14**	**4**	**110**	**0**	**0**
Pumpkin	**0**	**3**	**0**	**88**	**2**
Brinjal	**0**	**0**	**0**	**7**	**55**

Accuracy of Random Forest classifier of 80% training set = 94.00%

Table 4.54: Testing data of (20%) accuracy for Random Forest classifier

Crop Name	Precision	Recall	f1-score
Rice	97	96	97.00
Tomato	94	97	96.00
Potato	86	85	86.00
Pumpkin	91	95	93.00
Brinjal	96	89	92.00
Total/Average	93	93	93.00

- For 60% training data and 40% testing data.

Table 4.55: Confusion Matrix of 60% training data for Random Forest

Predicted /True Class	**Rice**	**Tomato**	**Potato**	**Pumpkin**	**Brinjal**
Rice	**820**	**0**	**28**	**0**	**0**
Tomato	**1**	**230**	**8**	**9**	**0**
Potato	**29**	**15**	**190**	**0**	**0**
Pumpkin	**0**	**5**	**0**	**200**	**6**
Brinjal	**0**	**0**	**0**	**13**	**110**

Accuracy of Random Forest classifier of 60% training set = 93.00%

Table 4.56: Testing data of (40%) accuracy for Random Forest

Crop Name	**Precision**	**Recall**	**f1-score**
Rice	96	97	97.00
Tomato	92	93	92.00
Potato	84	81	83.00
Pumpkin	90	95	92.00
Brinjal	95	90	92.00
Total/Average	91	91	91.00

Dataset of Jessore:

- For 80% training data and 20% testing data.

Table 4.57: Confusion Matrix of 80% training data for Random Forest classifier

Predicted/ True Class	Rice	Tomato	Potato	Pumpkin	Brinjal
Rice	**390**	**0**	**17**	**0**	**0**
Tomato	**0**	**110**	**1**	**9**	**0**
Potato	**20**	**2**	**110**	**0**	**0**
Pumpkin	**0**	**7**	**0**	**81**	**2**
Brinjal	**0**	**0**	**0**	**12**	**67**

Accuracy of Random Forest classifier of 80% training set = 92.00%

Table 4.58: Testing data of (20%) accuracy for Random Forest classifier

Crop Name	Precision	Recall	f1-score
Rice	95	96	95.00
Tomato	93	92	92.00
Potato	86	84	85.00
Pumpkin	79	90	84.00
Brinjal	97	85	91.00
Total/Average	90	89	90.00

- For 60% training data and 40% testing data.

Table 4.59: Confusion Matrix of 60% training data for Random Forest classifier

Predicted /True Class	Rice	Tomato	Potato	Pumpkin	Brinjal
Rice	**750**	**1**	**36**	**0**	**0**
Tomato	**0**	**220**	**8**	**7**	**0**
Potato	**38**	**11**	**220**	**0**	**0**
Pumpkin	**0**	**15**	**0**	**180**	**10**
Brinjal	**0**	**0**	**0**	**26**	**140**

Accuracy of Random Forest classifier of 60% training set = 91.00%

Table 4.60: Testing data of (40%) accuracy for Random Forest classifier

Crop Name	Precision	Recall	f1-score
Rice	95	95	95.00
Tomato	89	94	91.00
Potato	83	82	83.00
Pumpkin	84	88	86.00
Brinjal	93	85	89.00
Total/Average	89	89	89.00

4.4 Compare Algorithm

- **Cumilla**

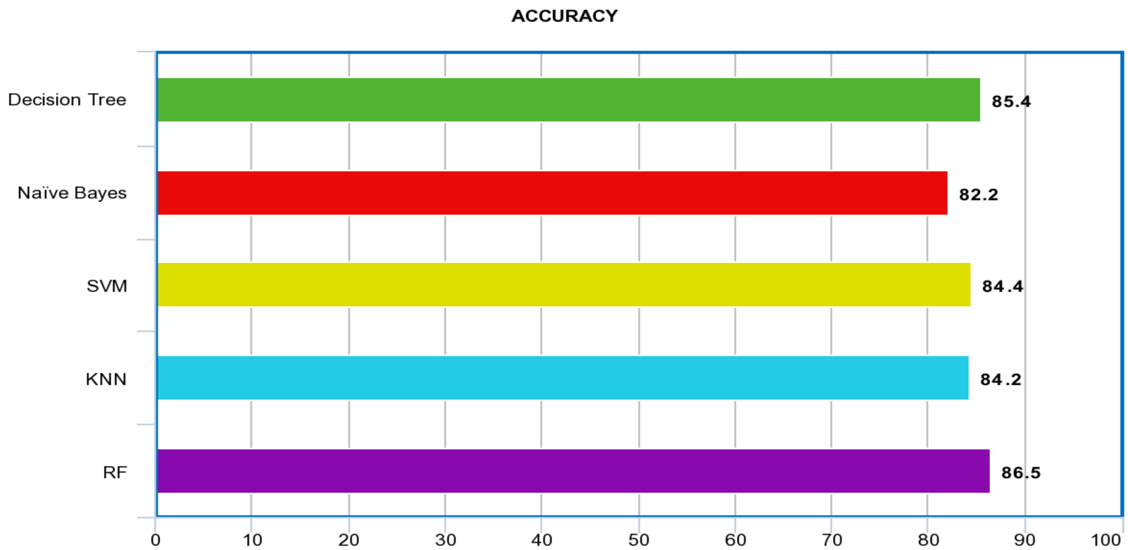


Figure 4.1: Accuracy of 80% training data (Cumilla)

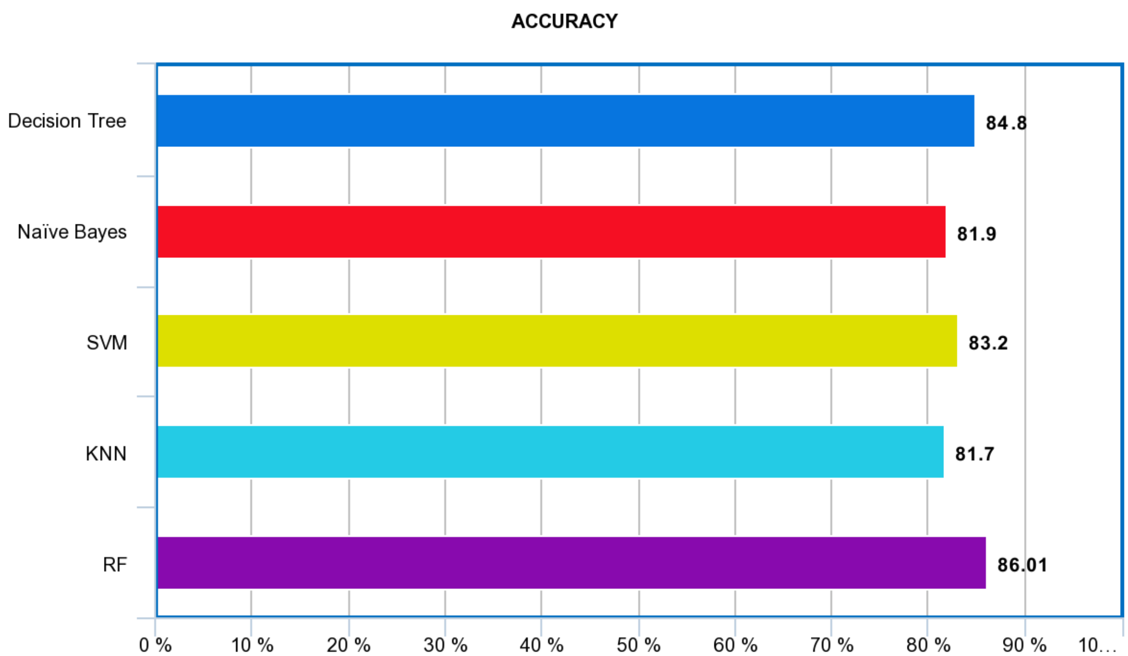


Figure 4.2: Accuracy of 60% training data (Cumilla)

- **Satkhira**

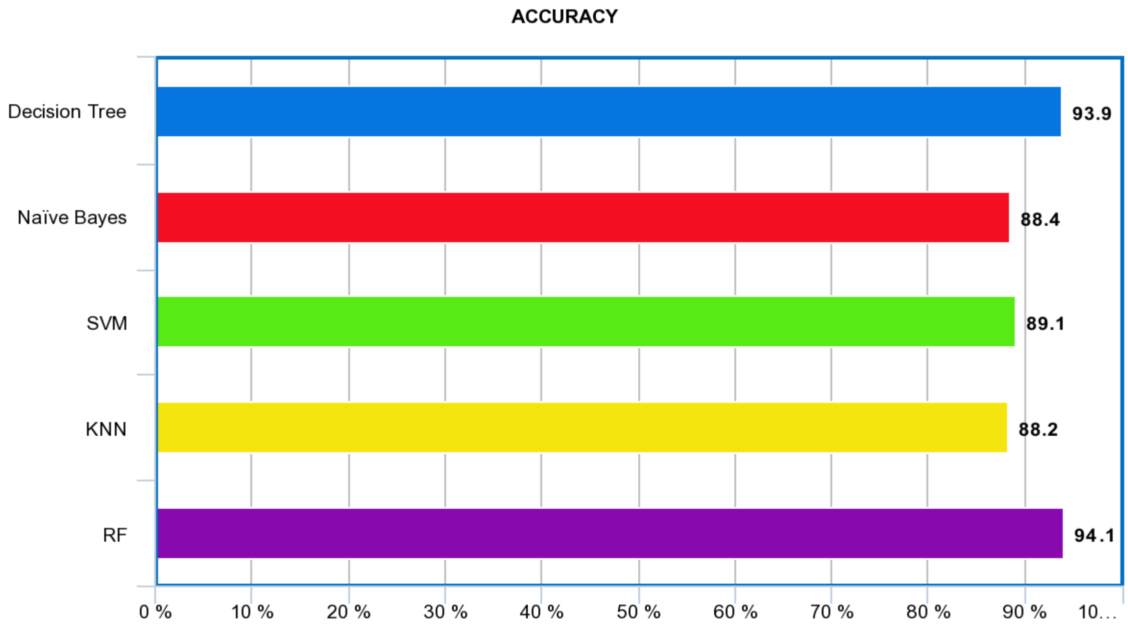


Figure 4.3: Accuracy of 80% training data (Satkhira)

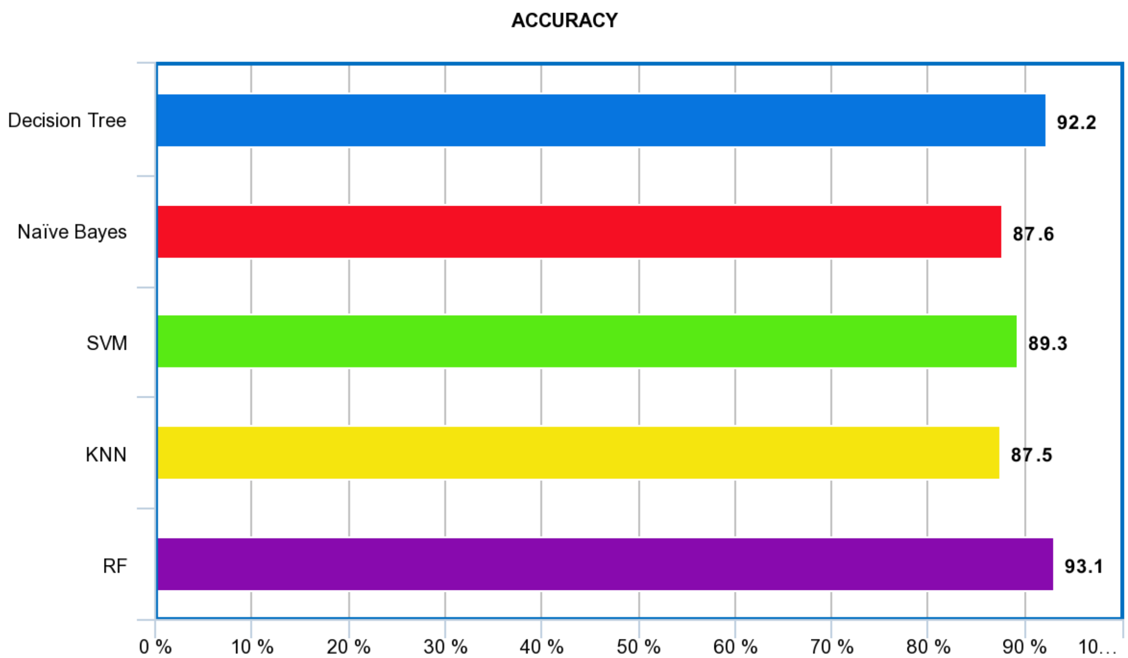


Figure 4.4: Accuracy of 60% training data (Satkhira)

- **Jessore**

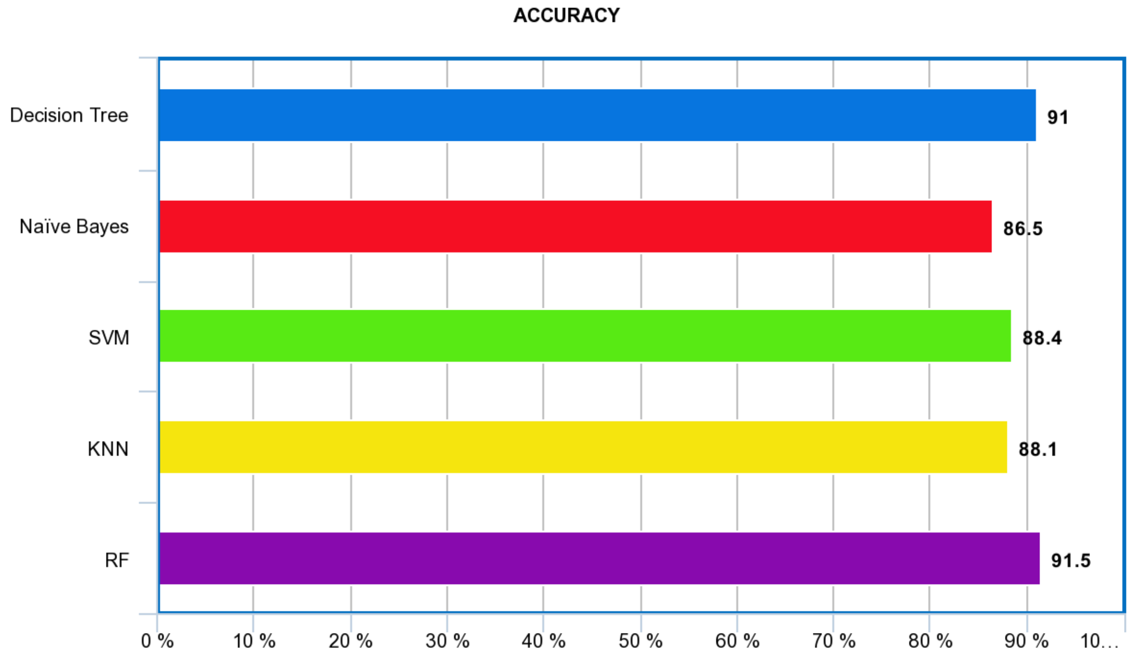


Figure 4.5: Accuracy of 80% training data (Jessore)

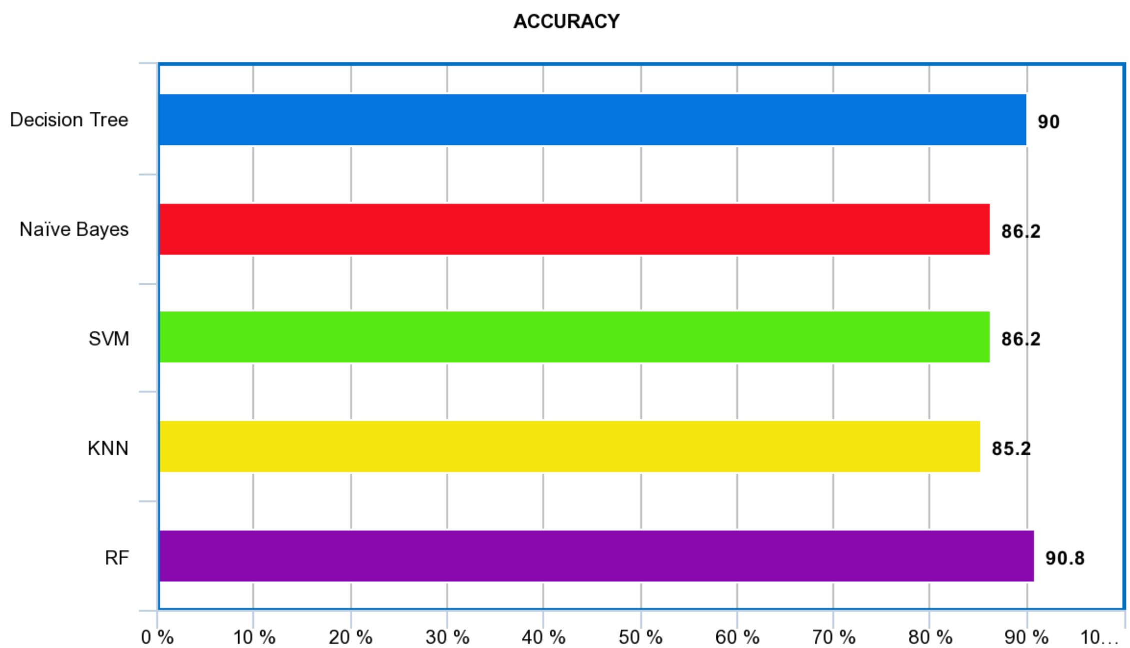


Figure 4.6: Accuracy of 60% training data (Jessore)

4.5 Descriptive Analysis

We split our dataset into two types of ratios. First, we take 80% training data and 20% testing data then apply 5 algorithms. Again, we divided the ratio into 60% training and 40% testing. So that we compare the accuracy ratio.

Here we can see that the difference is less than 2%. In every dataset, we can see that Random Forest is giving us better accuracy. So, we used Random Forest algorithm for prediction.

4.6 Prediction

In this section we discuss about the prediction. After using algorithm, we find that Random Forest algorithm is the best for this dataset. If we give the data Precipitation, Humidity, Dew, Wind speed and Temperature then it will give the results which crop is suitable for this weather.

```
data = np.array([[0.6,2.3, 60.8, 15.65,25.88, 18.23,13.55]])
prediction = RF.predict(data)
print(prediction)
```

```
[5]
```

Figure 4.7: Crop prediction

We select five crops for this work. Those are-

- Rice means 1
- Tomato means 2
- Potato means 3
- Pumpkin means 4
- Brinjal means 5

So, we can see that above figure Brinjal is the suitable crop for this weather.

CHAPTER 5
IMPLECATION FOR FUTURE RESEARCH
&
CONCLUSION

5.1 Discussion and Conclusion

In the Machine learning technique, temperature, season and location are the prime subject matters for approximate Crop production. It can be used to decide which crop is suitable for a particular district's temperature, nature and climate. The process of getting the forecast right can be improved by including more characteristics in the dataset. The random forest method is a more helpful prediction than others algorithms. Here we took five crops like rice, tomato, pumpkin, brinjal and potato. Then we apply random forest algorithms to our dataset and get results on which crop is best for that area. Our dataset has a lot of variables from which we can understand that accurate prediction is possible. At the same time, by this initiative, farmers will be able to reduce their losses. On the other hand, their production will be increased. This technique assists in selecting the finest crop for the upcoming season and aids in bridging the technical and agricultural differences. Agronomists would produce their expected crops through this technique.

5.2 Limitations

The digital agriculture system has a lot of benefits. It’s making farmer life more manageable and increase crop production. But it has some limitations also. In this project, we work with only three district data. We should take all sixty-four districts’ data for better result. Beside this we took only five crops. If we take more crop, we can make it better project. Also, we don’t make any Website or Application right now that’s why we only can predict using data.

5.3 Scope for Further Developments

- We took only five crops. In future we will work with more crop.
- For better accuracy we will take more data.
- In future we will work with more algorithms.
- We will make an android application so that people can easily use it.

References

[1] A. Nigam, S. Garg, A. Agrawal, and P. Agrawal, "Crop yield prediction using machine learning algorithms," in 2019 Fifth International Conference on Image Information Processing (ICIIP), 2019, pp. 125–130.

[2] P. Kamath, P. Patil, Shrilatha, Sushma, and Sowmya, "Crop yield forecasting using data mining," Global Transitions Proceedings, vol. 2, no. 2, pp. 402–407, 2021.

[3] V. Geetha, A. Punitha, M. Abarna, M. Akshaya, S. Illakiya, and A. P. Janani, "An effective crop prediction using random forest algorithm," in 2020 International Conference on System, Computation, Automation and Networking (ICSCAN), 2020, pp. 1–5.

[4] Y. J. N. Kumar, V. Spandana, V. S. Vaishnavi, K. Neha, and V. G. R. R. Devi, "Supervised machine learning approach for crop yield prediction in agriculture sector," in 2020 5th International Conference on Communication and Electronics Systems (ICCES), 2020, pp. 736–741.

[5] K. Sivanand, A. M. Sai, and N. J. Chaitra, "Web application development for site specific crop prediction using machine learning," IJMA, vol. 10, no. 2, pp. 3538–3549, 2021.

[6] R. Vidhya, P. Mathur, and S. S. Valluri, "Crop Yield Prediction using Random Forest," Int. j. adv. sci. technol., vol. 29, no. 9s, pp. 3084–3086, 2020.

[7] Akshatha, S. S. Shailesh, A. P. James, A. M. Saseendran, and C. M. Poojary, "Crop Analysis and Profit Prediction using Data Mining Techniques (Id:39)," Int. J. Eng. Res. Technol. (Ahmedabad), vol. 7, no. 8, 2019.

[8] A. Nigam, S. Garg, A. Agrawal, and P. Agrawal, "Crop yield prediction using machine learning algorithms," in 2019 Fifth International Conference on Image Information Processing (ICIIP), 2019, pp. 125–130.

[9] G. Duveiller, M. Donatelli, D. Fumagalli, A. Zucchini, R. Nelson, and B. Baruth, "A dataset of future daily weather data for crop modelling over Europe derived from climate change scenarios," Theor. Appl. Climatol., vol. 127, no. 3–4, pp. 573–585, 2017.

[10] V. S. Konduri, T. J. Vandal, S. Ganguly, and A. R. Ganguly, “Data science for weather impacts on crop yield,” Front. Sustain. Food Syst., vol. 4, 2020.

[11] T. Iizumi and N. Ramankutty, “How do weather and climate influence cropping area and intensity?” Glob. Food Sec., vol. 4, pp. 46–50, 2015.

[12] A. Suresh, P. Ganesh Kumar, and M. Ramalatha, “Prediction of major crop yields of Tamilnadu using K-means and Modified KNN,” in 2018 3rd International Conference on Communication and Electronics Systems (ICCES), 2018, pp. 88–93.

[13] P. S. Vijayabaskar, R. Sreemathi, and E. Keertanaa, “Crop prediction using predictive analytics,” in 2017 International Conference on Computation of Power, Energy Information and Commuincation (ICCPEIC), 2017, pp. 370–373.

[14]“ShieldSquare Captcha,” Iop.org. [Online]. Available: https://iopscience.iop.org/article/10.1088/1748-9326/ab5ebb/meta. [Accessed: 01-Dec-2021].

[15] A. Suruliandi, G. Mariammal, and S. P. Raja, “Crop prediction based on soil and environmental characteristics using feature selection techniques,” Math. Comput. Model. Dyn. Syst., vol. 27, no. 1, pp. 117–140, 2021.

[16] J. You, X. Li, M. Low, D. Lobell, and S. Ermon, “Deep Gaussian process for crop yield prediction based on remote sensing data,” Aaai.org. [Online].

[17] R. Gupta et al., “WB-CPI: Weather based crop prediction in India using big data analytics,” IEEE Access, vol. 9, pp. 137869–137885, 2021.

[18] S. S. Gumaste and A. J. Kadam, “Future weather prediction using genetic algorithm and FFT for smart farming,” in 2016 International Conference on Computing Communication Control and automation (ICCUBEA), 2016, pp. 1–6.

CROP LOCALIZATION FROM WEATHER DATASET

ORIGINALITY REPORT

19%	15%	9%	12%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	7%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	3%
3	"International Conference on Innovative Computing and Communications", Springer Science and Business Media LLC, 2021 Publication	1%
4	Pallavi Kamath, Pallavi Patil, Shrilatha S, Sushma, Sowmya S. "Crop yield forecasting using data mining", Global Transitions Proceedings, 2021 Publication	1%
5	dokumen.pub Internet Source	1%
6	Vaibhavi Vanarase, Vrushali Mane, Harsha Bhute, Ankita Tate, Simran Dhar. "Crop Prediction Using Data Mining and Machine Learning Techniques", 2021 Third International Conference on Inventive	<1%