

An Artificial Intelligence-Enabled Crop Recommendation System

by

Raiyan Uddin
18201172

Mrinmoy Barua
18201208

Mohammed Hossain Kabir
19101099

Muhammed Nufayel
19101341

Abu Sadman Sajid
19101528

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

Department of Computer Science and Engineering
Brac University
January 2023

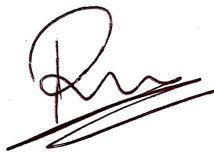
© 2023. Brac University
All rights reserved.

Declaration

It is hereby declared that

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

Student's Full Name & Signature:



Raiyan Uddin

18201172



Mohammed Hossain Kabir

19101099



Mrinmoy Barua

18201208



Muhammed Nufayel

19101341



Abu Sadman Sajid

19101528

Approval

The thesis titled “AN ARTIFICIAL INTELLIGENCE-ENABLED CROP RECOMMENDATION SYSTEM” submitted by

1. Raiyan Uddin (18201172)
2. Mrinmoy Barua (18201208)
3. Mohammed Hossain Kabir (19101099)
4. Muhammed Nufayel (19101341)
5. Abu Sadman Sajid (19101528)

Of Fall, 2022 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on January, 2023.

Examining Committee:

Supervisor:
(Member)



Dr. Md. Ashraful Alam

Assistant Professor
Department of Computer Science and Engineering
Brac University

Program Coordinator:
(Member)

Dr. Golam Robiul Alam

Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Sadia Hamid Kazi

Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Ethics Statement

The thesis submitted is our original work while completing the graduate program at BRAC University. The thesis only includes previously published or written third-party material if it is properly cited in the references. The thesis does not include any content already approved or submitted for consideration for another degree or diploma at a university or other institution. In light of earlier and current research, the results are appropriately framed. Each author will accept full responsibility for the paper's content because they actively and personally contributed to the substantial work that led to it. All significant sources of assistance have been acknowledged.

Abstract

Our country is overburdened with a population of more than 160 million. So fulfilling the need for food for the whole population can be overwhelming. This research has been conducted to ensure maximum efficiency in agriculture to overcome this problem. It is a known fact that Nitrogen (N), Phosphorus (P), and Potassium (K) are the three most essential micro-nutrients of any soil. Together it is called N.P.K. Soils also contain Sulfur (S), Zinc (Zn), and Boron (B). Together we can call it S.Zn.B., which are also important micro-nutrients. Different soils have different amounts of these essentials. Based on their values, an Automated system can suggest crops for a particular land to maximize production and profitability. We propose an AI-enabled crop recommendation system that will determine the best crops based on the soil type and its N.P.K and S, Zn, and B values through a Machine Learning Model. In our research, we use a comparative analysis among some existing Machine Learning Models to identify the most efficient model for our system. This system can effectively and accurately suggest the best suitable crops for a particular land. We used Random Forest which gave us 98%, Decision Tree which gave us 98%, Naive Bayes which gave us 89%, Ensemble Model which gave us 99% of accuracy and implemented Explainable-AI. As most farmers cannot select suitable crops for their land following their soil type, the agricultural sector is facing considerable losses. To minimize this, the efficiency of crop cultivation needs to be increased. Therefore, our system can revolutionize this sector by providing effective and suitable crops for land more accurately.

Keywords: Nitrogen, Phosphorus, Potassium, Sulfur, Zinc, Boron, NPK, SZB.

Dedication

This paper is wholeheartedly dedicated to the parents of the researchers, who have been the biggest pillar of support during the study period. This paper is also dedicated to the faculties who have guided us through their guidance in completing the final document. To the students who might find the paper helpful in the future. We would also like to thank Brac University for allowing us to complete our study. We dedicated this book to the Almighty God; thank you for the guidance, strength, power of the mind, protection, and skills and for giving us a healthy life. Lastly, the students behind this paper for their sacrifice and cooperation in the research of this paper.

Acknowledgement

We want to express our sincerest gratitude to our supervisor, Professor Dr. Md. Ashraful Alam, for his unwavering support, guidance, and advice throughout the research and writing of this thesis report. His invaluable insights and guidance were instrumental in helping us accomplish our goals. His willingness to provide feedback and answer our inquiries with great patience and understanding was greatly appreciated. We are also immensely thankful to our family and friends for their constant encouragement and motivation throughout this journey. With their support and understanding, this thesis report was possible. We are deeply grateful to all those who provided us with invaluable knowledge and insight during our research. We would particularly like to thank our research assistant Shakib Mahmud for his help in completing this project. His expertise in the field was a great asset to our research. We also thank the CVIS (Computer vision and Intelligence System Research) Lab for its assistance in completing the research by helping us with the technical support we needed. Finally, we would like to thank everyone who has been part of this journey for their support and encouragement. They have been a great source of motivation throughout the process, and their assistance has been deeply appreciated.

Table of Contents

Declaration	i
Approval	ii
Ethics Statement	iv
Abstract	v
Dedication	vi
Acknowledgment	vii
Table of Contents	viii
List of Figures	x
List of Tables	xi
Nomenclature	xii
1 Introduction	1
1.1 Executive Summary	1
1.2 Problem Statement	1
1.3 Research Objective	2
2 Background Study	3
2.1 Existing Works	3
2.2 Model and description	9
3 Research Methodolog	13
4 Dataset	15
4.1 Description	15
4.2 Maps of the sub-districts:	16
4.3 Sample Data	24
4.4 Experimental Setup	25
4.5 Data Visualization	26
4.6 Data Preprocess	28
4.7 Data Split	29

5	Performance Matrices	31
5.1	Descriptions	31
6	Result Analysis	34
6.1	Random Forest	34
6.2	Decision Tree	37
6.3	Naive Bayes	40
6.4	Ensemble	43
6.5	Accuracy Comparison	46
6.6	Explainable-AI	47
7	Future work and Conclusion	48
8	References	49

List of Figures

2.1	Random Forest	10
2.2	Decision Tree	11
3.1	workflow	13
4.1	Map of Dacope sub-district	16
4.2	Map of Bheramara sub-district	17
4.3	Map of Sonagazi sub-district	18
4.4	Map of Galachipa sub-district	19
4.5	Map of Jhenaigati sub-district	20
4.6	Map of Nilphamari-Shodor sub-district	21
4.7	Map of Tala sub-district	22
4.8	Map of Tarash sub-district	23
4.9	Crop amount	26
4.10	Correlation Heatmap	26
4.11	Recommended Crops	29
4.12	Train and Test data amount	30
6.1	Confusion Matrix for Random Forest	34
6.2	ROC for Random Forest	35
6.3	Classification for Random Forest	36
6.4	Confusion Matrix for Decision Tree	37
6.5	ROC for Decision Tree	38
6.6	Classification for Decision Tree	39
6.7	Confusion Matrix for Naive Bayes	40
6.8	ROC for Naive Bayes	41
6.9	Classification for Naive Bayes	42
6.10	Confusion Matrix for Ensemble	43
6.11	ROC for Ensemble	44
6.12	Classification for Ensemble	45
6.13	Lime Implementation	47

List of Tables

4.1	Sample Soil Data	24
4.2	Crop Data	24
4.3	Sample Soil Data with crop	28
6.1	Accuracy Graph	46

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

σ *EuclideanDistance*

b_c *BoronOfCrop*

b_s *BoronOfSoil*

FN *FalseNegative*

FP *FalsePositive*

k_c *PotassiumOfCrop*

k_s *PotassiumOfSoil*

n_c *NitrogenOfCrop*

n_s *NitrogenOfSoil*

p_c *PhosphorusOfCrop*

p_s *PhosphorusOfSoil*

s_c *SulfurOfCrop*

s_s *SulfurOfSoil*

TN *TrueNegative*

TP *TruePositive*

z_c *ZincOfCrop*

z_s *ZincOfSoil*

Chapter 1

Introduction

1.1 Executive Summary

Bangladesh is exclusively dependent on agriculture and still has less farm productivity. Farmers must raise productivity so that with less effort they can earn more money from the same amount. Farmers can do this with precision agriculture. Also, precision farming is known as the application of a precise and suitable crop at the right moment to maximize production and yields. Precision agriculture does not come up with the best results every time. However, in farming, the recommendations that must be created are correctly associated with Nursing precisely because just in case of errors should cause serious loss. Several analyses are being carried out to create a cost-effective and reliable model that predicts crops more efficiently. This research will offer a system that uses the Nitrogen (N), Phosphorus (P), Potassium (K), Sulfur (S), Zinc (Zn), and Boron (B) values and the correct model to recommend the optimum crop for the land.

1.2 Problem Statement

In our paper we see the largest task is making the AI-derived insights useful. Farmers need to undertake AI, however, if it is no longer not used in the fields the farmers are not going to waste their time. AI powers new generation tools. Same as every tool, they will not be used if they are no longer useful. Agri-tech offerings want to pay attention closely to make analytic actionable in a manner that enables farmers to prioritize the troubles of their fields in an extra significant manner. In order to count the appropriateness of AI, agriculture still remains hard to statistically quantify, despite advances in the agricultural field. As said by Joseph Byrum, a 2017 weblog post, situations continuously change inside a single field. The statistical analyses can be affected by adjustments in soil quality, unpredictable weather and also crop disorder. At the same time, ecosystems range vastly from one continent to another, which means such things as seed and fertilizer applications are not often regular globally. Therefore, AI and system learning are nonetheless a long way from being capable of expecting vital results in agriculture merely through the cognitive ability of machines.

One of the primary issues is connectivity. Many underdeveloped regions of Bangladesh are far away from telecommunications systems, putting some farmers at a disadvan-

tage. Another hassle is getting farmers to believe in the system and educating them about its proper implementation. This may be tough for plenty of reasons, consisting of farmers being worried over data ownership. According to farmers, AI is more useful in the virtual than in the real world. However, they may need to consider how AI assists them while painting the whole terrain. This is not always because they are conservative or cautious of the unknown. Their resistance results from a need for more know-how of the sensible utility of AI tools.

New technology regularly appears confusing and unreasonably pricey because AgriTech vendors need to explain why their answers are beneficial and how they should be implemented. This is what occurs with synthetic intelligence in agriculture. Although AI may be beneficial, there is nevertheless a variety of work to be accomplished with the aid of using generation vendors to assist farmers in putting in force properly.

In this sector, two main issues have hindered the widespread use of machine learning despite precise agriculture growth for the past three decades. Firstly, the utilities of precision-Ag have yet to be adopted by many growers and farmers. Secondly, collecting critical information using any other medium for prediction is not the main aim of precision-Ag technologies. From a technical point of view, this could result in a breaking point for the usability of AI. Because of some environmental factors that are not under our control, most agricultural activities result in inherent uncertainty(fundamental error in AI). This results in ongoing collaboration around data collection procedures between data scientists, having domain knowledge on one side, and agronomists, farmers, growers, and consultants on the other. This shared knowledge has paved the way for developing models with sustainable utilities for farmers and growers.

1.3 Research Objective

This research aims to develop an AI-based crop recommendation system using N, P, K, S, Zn, and B values and various models. Our system will run by statistical analysis based on production rate, production cost, real-time market value, market demand, and loss factor to recommend the best suitable crops for the given land. So there are some objectives we have to fulfill:

- Deeply understand N, P, K, and S, Zn, and B detection techniques.
- Evaluate Data set.
- Deeply understand the used models, algorithms, and detection techniques.
- Develop machine learning model for crop recommendation based on N, P, K, S, Zn, and B values.
- Evaluate the model and result.
- Ensemble the chosen models.
- Offer recommendations on improving the model.

Chapter 2

Background Study

2.1 Existing Works

Vikas and Deepak (2022) [1], state an architecture proposed to develop a platform for data integration, processing, and analytic. All the components included in the platform can help in real-time data processing and storage. Furthermore, these real-time data can be used for mining key insights. Using ML, AI, and Data science algorithms, the insights derived from the data can be visualized on the dashboard. Regression and matrix factorization is used to understand better the coefficients to recommend the varieties. Different regression techniques were used for model selection, and two ensemble techniques were found to perform best: Random Forest Regressor and Gradient Boosted Regressor. In the future, an end-to-end solution is aimed to be developed with available input data using the proposed architecture. By Using this architecture, a PoC completion is planned with the integrated data from multiple stakeholders, which will help in performance and scalability evaluation.

Ashutosh and Rajendra (2021) [2], states in their research that, they suggested a smart irrigation system prototype. Environmental attributes and Regression of soil are included in the proposed recommendation system, which is improved further with the help of AC. Their experiment reveals that for selecting suitable pesticides and herbicides SVM model out matched other classification models.

Nischitha K(2020) [3], introduced us to a machine learning system that can predict crops based on soil composition and environmental data such as temperature, humidity, soil PH, and rainfall. The proposed algorithm will estimate the best crop for a given piece of land. They used data from different resources like govt websites, VC Farm Mandya, APMC website. For rainfall prediction, they used the SVM algorithm, and for crop production, they used Decision Tree Algorithm. They have created a system with the help of GUI to predict crops and minerals. Therefore, this helps farmers in choosing the appropriate crop. They also presented their future scope of developing this idea by implementing GPS location with the help of a rain forecasting system. This would help them predict crop production only by knowing the GPS location. Framework, which is a setback for the adoption of smart technologies and establishing smart, sustainable agricultural practices.

Rohit Kumar Rajak, and Ankit Pawar (2017) [4], talked about maximizing crop

yield using a machine learning technique. Firstly, they created a dataset based on minerals. They trained the model with the help of an assembling technique known as the Majority Voting technique. The learners used in the projects are Support Vector Machine, NAïVE Bayes, Multi-layer Perceptron (Artificial Neural Network), and random forest. The mode aims to reduce chemical use and get maximum production of crops. The improved dataset will give more precious recommendations about maximum crop yield.

Gaurav Chauhan and Alka Chaudhary [5], stated a model to predict which crop will be suitable for which land. The paper used Regression, Decision tree, SVM, Naive Bayes, Knn, and K-means to predict the predictor's accuracy value. They studied the agriculture recommendation system using the machine learning method and got 94.78% accuracy, and the majority voting method was applied. The system gave 95% accuracy using SVM a& artificial neural network. As a agriculture dependent country we need to boost our yield and to make it possible Random Forest showed the best result.

Nidhi H Kulkarni, Dr. B M Sagar, Dr. G N Srinivasan, and Dr.N K Cauvery [6], tried to improve the growth of crops. They used a majority voting technique to predict the accuracy of the recommendation system. Their analysis was based on soil type, NPK values, pH value, soil porosity, rainfall, and temperature. From 9000 samples, they used 6750 samples for training and 2250 samples for testing. These values were tested and trained using SVM, Naive Bayes, and Random Forest and got a 99.1% accuracy average.

Priyadharshini A, Swapneel Chakraborty, Aayush Kumar, and Omen Rajendra Pooniwala [7], trained and tested several data to achieve good accuracy for their Intelligent Crop Recommendation System using Machine Learning. They proposed and tested several models for their system. They analyzed the data, preprocessed it, and proposed the mechanisms. They used Decision tree, K nearest neighbor, Linear Regression Model, Naive Bayes, Neural Network, and SVM and got an accuracy of %, 85%, 88.26%, 82%, 89.88%, and 78% respectfully. They analyzed profit analysis, crop recommender, and crop sustainability predictor and built the data to achieve accuracy. As farming is one of the country's main sectors, it is important to yield our soils fully.

Modi and Sutagundar (2021) [8], states in their research that despite being the world's largest agricultural producers, India's farm productivity remains low. As they fail to select appropriate crops for their soil, they face problems right from the beginning. This research aimed to design a recommendation using machine learning for precision in crop selection and improved crop productivity. The system will also consider the soil N, P, K, and pH values to determine the most productive crops which can be cultivated under such conditions. This research will help the farmers to determine which crop will give maximum productivity amongst all the available crops in their area by listing. Also, this study talks about an artificial neural network used for pattern comparison and statistical analysis for better running productivity. In this paper, they have used the SVM model. For evaluation, confusion matrix is used. Considering all the parameters for evaluating the model, the overall accuracy

is considered as 97.65%. By improving this technique, real-time soil values can be obtained.

According to S. Ramanujam(2017) [9], recommendation farming is an approach that uses research data on the soil and its type to give farmers a fair recommendation. In their research paper, a recommender system which is an ensemble model that uses Naive Bayes, KNearest Neighbor, Random Tree, and CHAID as students to recommend a structure for a particular site with parameters with great precision. The data set comprises the attributes of soils which are specific. Also, they have used an ensemble technique called the Majority Voting technique. They have used models like CHAID, known as Chi-squared Automatic Interaction Detection, K-Nearest Neighbor, and Naive Bayes. This model has a prediction accuracy of 88 percent. Soon, this model will need to be more precise to use in real-life scenarios. So, the farmers can be recommended the best suitable crop increasing their capitulation and increasing the country's productiveness.

V.K. and Mukhopadhyay(2019) [10], state in the paper about crop suitability and fertilizer recommendation based on the creation of ontology. Based on the region in Maharashtra, India, and the soil type, the method forecasts potential crops for the field under consideration. The paper describes a system that advises farmers how the best crops for a given region based on the region's crop yield history over the last three years and the best fertilizers for certain crops based on soil data works. A crop recommendation system has been used where Random forest algorithm has been used for accuracy because it was much bigger the ID3 algorithm for a particular data set. An unsupervised learning technique was also utilized to locate fertilizers with NPK values closest to the needs of the selected crops when suggesting fertilizer. Relevant crops should be picked if experts have recommended them. The system provides recommendations based on region, type of crop, and fertilizer recommendation based on NPK content of the soil, available on their mobile phones.

According to Jung, J.(2021) [11], new research avenues have been opened through the integration of genomics and UAS-based phenomics so that the dissection of complex agronomic traits and identifying the genes governing these traits can be performed, which can increase the efficiency, size and genetic gain of breeding programs. The research aims to improve the use of different sensors and data resources so that the advancement of agricultural systems can be ensured to meet the world's current requirements. Therefore, the paper presents the importance of using different remote sensors through trained datasets and ML to develop sustainable crop management. Then again, apart from plant maintenance and growth, advanced genomics also offer analytical tools to establish crop breeding programs so that the molecular basis of complex traits can be comprehended. The work presents UAS-based HTP systems with AI, crop simulation models, and remote sensing to develop the overall status of Digital Agriculture. By observing the research paper, we can understand that they have presented a thorough review of different types of remote sensors via AI and ML.

Suchithra(2020) [12] states that hi-tech computer systems are used to calculate many characteristics, such as crop prediction, crop quality, weed detection, yield detec-

tion, and other machine-learning techniques. Applications such as IoT, ML, ANN in agriculture, and other models which help in precision farming are discussed in this paper. This paper outlines a proposal for developing a system that uses sensors, ML, and IoT to mechanize traditional processes of agriculture. The usage of deep learning is considered an advantage over ML as it adds depth to ML.

Yiannis Ampatzidis, Victor Partel, and Lucas Costa (2020) [13], proposed a paper on the importance of UAVs (Unmanned Aerial Vehicles) in agriculture as there needs to be more availability of personnel trained for field scouting. The UAVs can be equipped with sensors that can pick up various data resulting in a boost of efficiency and accuracy, and also can cut huge amounts of costs. They mentioned a program ‘Agroview’ which was created for analyzing data of UAV and other platforms. This tool can geo-locate, identify, and count plants and plant gaps. Also the canopy sizes and can create plant health or stress maps.

Beloiev et al (2021) [14], presented a paper that focuses on the design, building process, key characteristics, and assessment of a prototype of a robot that may be utilized for some of the most basic tasks in agricultural operations. The robot has been developed as an end-user autonomous mobile system that can map or examine a specific farming region and self-localize. Artificial intelligence (AI) algorithms underlie the robot’s decision-making abilities and enable it to choose a course of action depending on the situation and the environment. Although the given prototype is still in its early stages of development and evaluation, the article concludes with ideas for potential platform upgrades in the future. The widely accessible Raspberry Pi 3 Model B+ microcontroller was employed to accomplish computer vision tasks using the created small-scale agricultural robot (Raspberry 3, 2020). It is the heart of the robot’s top-level hardware layer.

This study presented by Vincent et al (2019) [15], proposes an expert system for assessing agricultural land suitability that combines sensor networks with Artificial Intelligence systems like neural networks and multi-layer perceptrons (MLP) to suggest an expert system for evaluating the suitability of agricultural land . Using four decision classes—more appropriate, suitable, somewhat suitable, and unsuitable—this suggested method will help farmers assess agricultural land for cultivation. This evaluation is based on the data obtained from the numerous sensor devices utilized to train the system. Compared to other current models, the results produced utilizing MLP with four hidden layers are determined to be successful for the multiclass classification system. This trained model will evaluate subsequent evaluations and classify the land following each crop. The algorithm used in this paper was the MLP Algorithm for the Machine Learning model. The datasets were collected from four areas: Climate, Soil, and Groundwater, and then they were sectioned into four classes according to suitability for land sustainability. The model has presented an accuracy of 99

Talaviya(2020) [16], narrates that his research focuses on the agriculture sector where AI, like weeding, irrigation, and spraying with the help of sensors. This study compiles the findings of several academics to provide a quick summary of the present state of the agriculture sector where we use automation, including systems of weed-

ing that use drones and robots. It states different models that assist in irrigation and soil moisture sensors. It uses unusual lighting situations, such as radiant and shady, which were mosaicked to see if GAHSI could be used to determine the location of zones. The clock, the ADC, the UART sequential I/O, and the outer intrudes on the source are four P80C592 that interfere in the framework that we use to overcome these contentions and consider each of the undertakings.

Environmental factors and weather greatly affect crop diseases. Potential economic losses are mainly caused by weather fluctuation. In this paper, Mallocci(2020) [17] stated that a new AI to find the disease of potatoes. Data regarding the following factors, such as temperature, rainfall, wind speed, humidity, and solar radiation, have been collected from different locations in the past 4 years to experiment. They used a model which was a forecasting model based on AI. Due to the lack of ability to learn in general-purpose supervised predictions without being strictly programmed, it has been adopted with satisfactory results in some factors over the past years in the literature. The model showed an accuracy of 82 percent when humidity, temperature, wind, and solar radiation were considered and 78 percent when the factors mentioned above, along with rainfall, were considered. However, estimation of disease risk outbreak was impossible as there was no label in the dataset. To overcome this, in collaboration with the LAORE sardinia agency, a new technique has been performed, which was proposed to label the dataset. The LANDS Decision Support System would integrate the forecasting model in the future to forecast the stages of the diseases for each monitored culture in real time.

Singh et al (2020) [18], presented their AI and IoT. based monitoring system, which helps increase the Yield in Crop Production. With the help of IoT, they first stored data in the cloud. Machine learning algorithms are used to find the best condition for maximum crop production based on the data set. Their proposed architect is to take saplings from a target plant and monitor them continuously through various sensors. These sensors have been specially created to measure the targeted parameters, such as temperature and humidity, soil temperature and moisture, and light intensity. With the help of image processing, data on plant height and width are stored in CSV form to train the model and detect the growth rate. Richa Singh and Sarthak Srivastava implemented these techniques for marigolds and got 83.33

In this paper, Alreshidi, E. (2019) [19], gives a brief review of the development within SSA, which provides a smart, sustainable agriculture platform as a solution by providing an IoT/AI architecture. The review includes the enhancement of the study through informal interviewing of experienced farmers. The aim was to establish an IoT/AI SSA architecture and use IoT and AI as the core to build an SSA platform. However, due to fewer studies in this area, the current research is insufficient to establish an IoT/AI framework, which is a setback for adopting smart technologies and establishing smart, sustainable agricultural practices.

Bhuvneshwar and Saravanan (2022) [20], present their architecture of smart agricultural systems based on AI, Wireless Networks, and contemporary IoT communication. Farmers can evaluate and gather valuable data using IoT agriculture sensors. Optical, weather measuring and NPK sensors are used for data collection from the

environment and process automation. Bi-LSTM prediction deep learning model has been used for the recommendation, which gives a remarkable accuracy of 92%. The farmers are compelled to satisfy the demand of the increasing. Recent research aids farmers in recommending the optimal amount of fertilizer to be used in terms of chemical properties. This proposed approach is investigated and analyzed as an.

Vaishnavi and Shobana (2021) [21], stated in their research that due to uncertainty Tamil Nadu faces a loss of production in agriculture. With the help of data mining, farmers can carry on their crop cultivation using a recommendation system. Based on season and productivity, the dataset has been tested and recommended. Machine learning techniques develop a well-defined model with the data for attaining predictions. In this paper, the concepts of recommendation have been given to crop to sow. They have mentioned the SNN model for having an analysis with crop estimation. By considering the parameters like season, production and recommendations related to this can be given to the farmers.

Dharani(2021) [22] narrates that three different classifications were performed in the hybrid network, including multivariate regression, which gives a yield range of about 85 percent with different algorithms based on the need. The research study mentions hybrid and recurrent neural networks in the agricultural sector. The approach was also used using reinforcement learning multiple networks, which resulted in an accurate crop yield range of roughly 90 percent, with an actual yield of around 89 percent. Artificial neural networks and convolutional neural networks like these need to stand out in terms of performance in front of this model. From the above discussion, the RNN and hybrid networks beat the other full networks with an accuracy of up to 90 percent. It contains the CNN-failed feedback loop, which results in an average estimation of 83–89 percent with the highest accuracy.

According to Aitkenhead (2003) [23], firstly, they used a characteristic semantic measure which is leaf-shaped. It had variable efficiency in the middle of 52 to 74 percent in perceptive two different plants. Variations depend on the size of the plant. To automate agriculture systems and reduce the use of chemicals by localized spray, the improvement of a visual method to distinguish between crops and weeds. The main drawback of the plant analysis method is that individual clusters had to be analyzed, but it took a lot of work. It does not guarantee the isolation of the plant. Also, another has involved a self-organizing neural network that is biologically more plausible than many NN methods. The system learns and differs between species and achieves the accuracy of 75 percent of the necessity of predefined plant descriptions.

Alshanbari, R. (2021) [24], stated that a release mechanics system had been reported based on crop row detection. The field navigation-based machine vision needs this as most plants are usually cultivated in a row. The release mechanics system releases multi-sensory devices that are lightweight and flexible on top of each plant so that humidity and temperature can be measured. For research purposes, they used an image data set to simulate robot movement in the field and for research purposes, image processing to speed up computation and improve real-time algorithm detection using the hough transform and filtering, and Kalman filter, respectively. The

complexity of the computation required for the Hough transform and the execution release mechanic system-based crop row detection algorithm present a challenge in this type of work. Future work should focus on making the vehicle autonomously follow the crop row while also further optimizing the releasing mechanism, collecting sensor data wirelessly, and steering the vehicle using navigation control strategies.

2.2 Model and description

Random Forest: It is a supervised machine-learning algorithm that is based on the concept of decision trees. It is a collection of decision trees, where each tree is trained on a different subset of the data. The final output of the algorithm is the average of the outputs of all the decision trees in the forest. The Random Forest algorithm is highly flexible and accurate and is particularly useful in cases where the relationship between the input variables and the output variable is non-linear. The algorithm begins by randomly selecting a subset of the data to train each decision tree. This is done using a technique called bootstrapping, randomly selecting data samples with replacement. Once the subsets are selected, the algorithm grows a decision tree for each subset. This is done by recursively splitting the data into subsets based on the value of a randomly selected input variable. The split that results in the highest reduction in impurity is chosen as the split point. This process is repeated until a stopping criterion is met, such as a maximum tree depth or a minimum number of samples in a leaf node. The final output of the Random Forest algorithm is determined by averaging the outputs of all the decision trees in the forest. The averaging process is done by taking a majority vote for classification problems and taking the average for regression problems. Using multiple decision trees improves the algorithm's overall performance by reducing the variance and increasing the model's generalization ability.

In summary, Random Forest is an ensemble learning algorithm that utilizes a collection of decision trees to make predictions. It works by training multiple decision trees on randomly selected subsets of the data and then averaging the outputs of all the trees to produce the final result. The algorithm is known for its high accuracy, robustness, and ability to handle non-linear relationships.

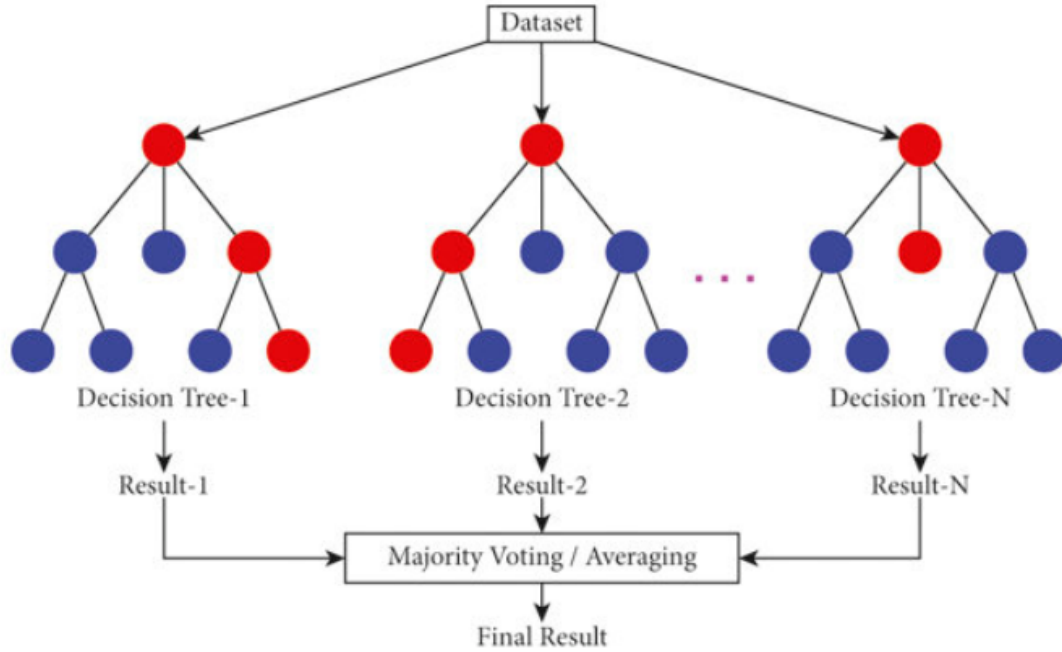


Figure 2.1: Random Forest

Decision Tree: A supervised machine-learning approach for both classification and regression tasks is decision tree. This method builds a tree model with choices of potential outcomes with each internal node representing feature or attribute of input data, each branch indicating a decision or rule based on the attribute's value, and each leaf node representing the output or prediction. The decision tree algorithm is based on entropy, a measure of impurity or randomness in a given data set. The algorithm constructs the tree by recursively partitioning the data into subsets based on the values of the input features and selecting the feature that results in the most significant reduction in entropy as the root node. The process is repeated for each branch until a stopping criterion is reached, such as a minimum number of samples per leaf or a maximum tree depth. The final tree can then be used to make predictions by traversing the tree from the root to a leaf node and returning the output associated with that leaf.

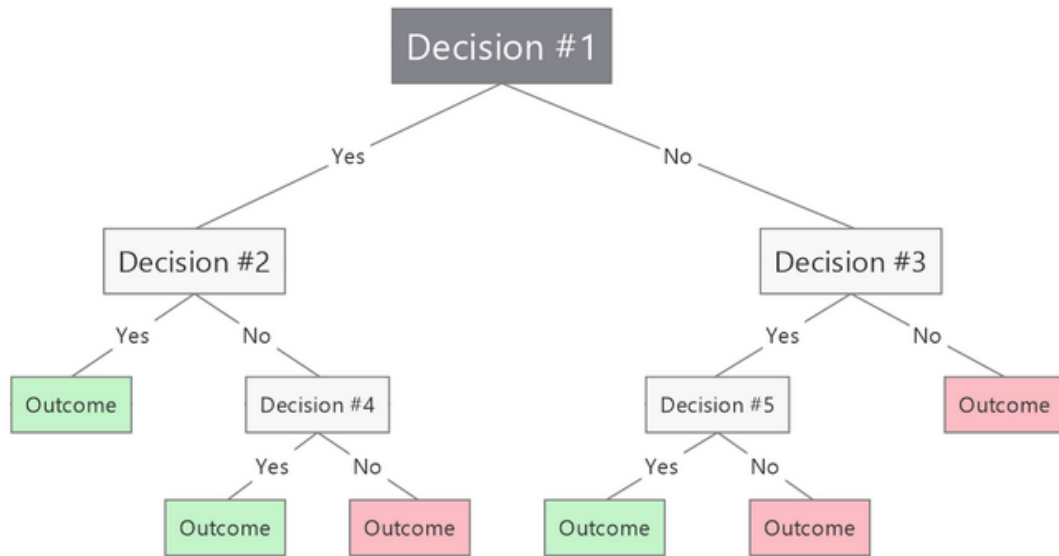


Figure 2.2: Decision Tree

Naive Bayes: To produce predictions based on prior knowledge and to observe the evidence, a probabilistic method that uses Bayes' theorem is Naive Bayes. In real-world data, all features within a data set are dependent with one another where Naive Bayes does not fit in, and that's why it is considered "naive". Naive Bayes is primarily used for classification tasks, where the goal is to delegate a new observation to one of several standard categories. The algorithm calculates the probability of each category given the observed feature values and selects the category with the highest probability as the prediction. There are several variations of Naive Bayes, including Gaussian Naive Bayes, which is used for continuous data, and multinomial Naive Bayes, which is used for discrete data such as text. It can also be used for regression and Multi-class classification tasks. In summary, it is a simple yet effective method that utilizes Bayes' theorem to make predictions based on prior knowledge and observed evidence, despite assuming that features are independent, making it widely used in various classification tasks.

Ensemble Method: Ensemble methods in machine learning involve combining the predictions of multiple models to improve the model's overall performance. The idea behind ensemble methods is that by combining the predictions of multiple models, which are often trained using different algorithms or on different subsets of the data, the resulting ensemble model will be more robust and have a lower chance of overfitting than any other regular model. There are a variety of ensemble methods available, including (1) Bagging, in which multiple models are trained using many data subsets, and their predictions are combined through a voting mechanism. (2) Boosting, in which multiple models are trained sequentially, with each model attempting to correct the mistakes of the previous model. (3) Stacking, in which multiple models are trained, and their predictions are used as input to a meta-model, which makes the final prediction.

Ensemble methods are widely used in machine learning and have been shown to improve performance on a broad variety of tasks, including speech recognition, image

natural language processing, and image classification. It is essential to notice that, to create an ensemble model, it is necessary to have a good set of base models to combine. Moreover, the way that these base models are combined can significantly impact the final performance.

Explainable-AI: Explainable AI, also known as XAI, refers to the ability of artificial intelligence systems to provide clear and interpretable explanations of their decision-making processes and outcomes. This is achieved through techniques such as transparency, interpretability, and accountability, which allow stakeholders to understand how the AI system arrived at its conclusions and evaluate the trustworthiness of its predictions. The goal of XAI is to create AI systems that are not only accurate and efficient but also trustworthy and transparent, allowing them to be used in high-stakes and critical applications such as healthcare, finance, and transportation. Furthermore, it also helps to mitigate the risk of bias and increase the accountability of the AI system.

Lime: LIME (Local Interpretable Model-agnostic Explanations) is an algorithm for providing interpretable explanations of the predictions made by any black-box machine learning model. The algorithm approximates the behavior of the complex model locally by fitting a simpler, interpretable model to the predictions of the complex model in the vicinity of a specific instance. The key idea behind LIME is to understand the predictions of a complex model by approximating it locally with an interpretable model. In order to achieve this, LIME generates a set of synthetic data instances around the instance of interest. Then it fits an interpretable model to the predictions of the complex model for these synthetic instances. The interpretable model is then used to explain the predictions of the complex model for the instance of interest. The explanations generated by LIME are model-agnostic, meaning that they can be applied to any black-box machine learning model and can be used to understand the model's predictions for any individual instance. In summary, LIME is an algorithm for explaining predictions made by any black-box machine-learning model by locally approximating the complex model with an interpretable model in the vicinity of a specific instance. This approach allows us to understand the model's predictions for any individual instance.

Chapter 3

Research Methodolog

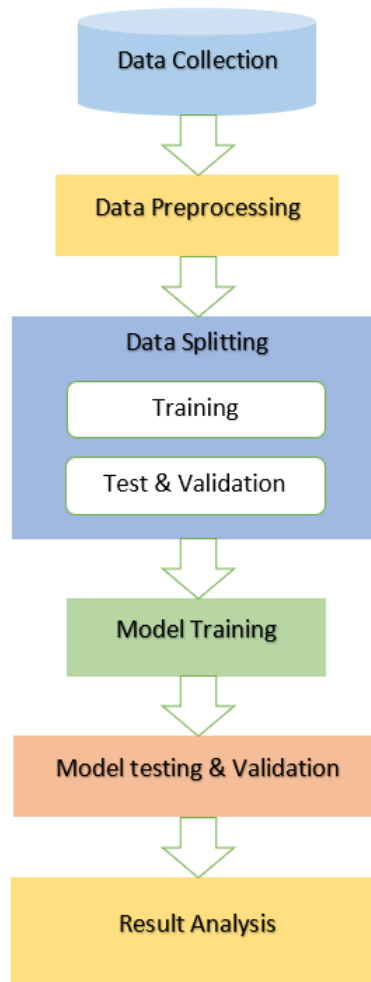


Figure 3.1: workflow

Here the above flowchart shows our standard operating procedure for the whole system. Firstly, we did collect the data of the N, P, K, S, Zn, and B values for certain areas' soil and some necessary information. We have worked on five different

crops, which are Garlic (Rosun), Grass peas (Kheshari), Mung Bean (Mugg Dal), Cotton (Tula), and Soybean oil. N, P, K, S, Zn, and B values were required for the optimal development of these crops. Following that, we looked at the crop with the lowest N, P, K, S, Zn, and B values as the most appropriate crop for the relevant soil. Finally, we divided all the data in two categories: train and test to run the models.

Chapter 4

Dataset

4.1 Description

We collected data books from "Soil Resource Development Institution". We digitized all the data from the books and got 1056 data of different lands and its values. Our dataset contains N.P.K, S, Zn, B values of 8 districts which are: Satkhira District, Patuakhali District, Khulna District, Feni District, Nilphamari District, Sherpur District, Kushtia District, Sirajganj District. We have collected one sub-district from each district. Those are:

- Dacope sub-district (Khulna District)
- Bheramara sub-district (Kushtia District)
- Sonagazi sub-district (Feni District)
- Galachipa sub-district (Patuakhali District)
- Jhenaigati sub-district (Sherpur District)
- Nilphamari-Shodor sub-district (Nilphamari District)
- Tala sub-district (Satkhira District)
- Tarash sub-district (Sirajganj District)

4.2 Maps of the sub-districts:

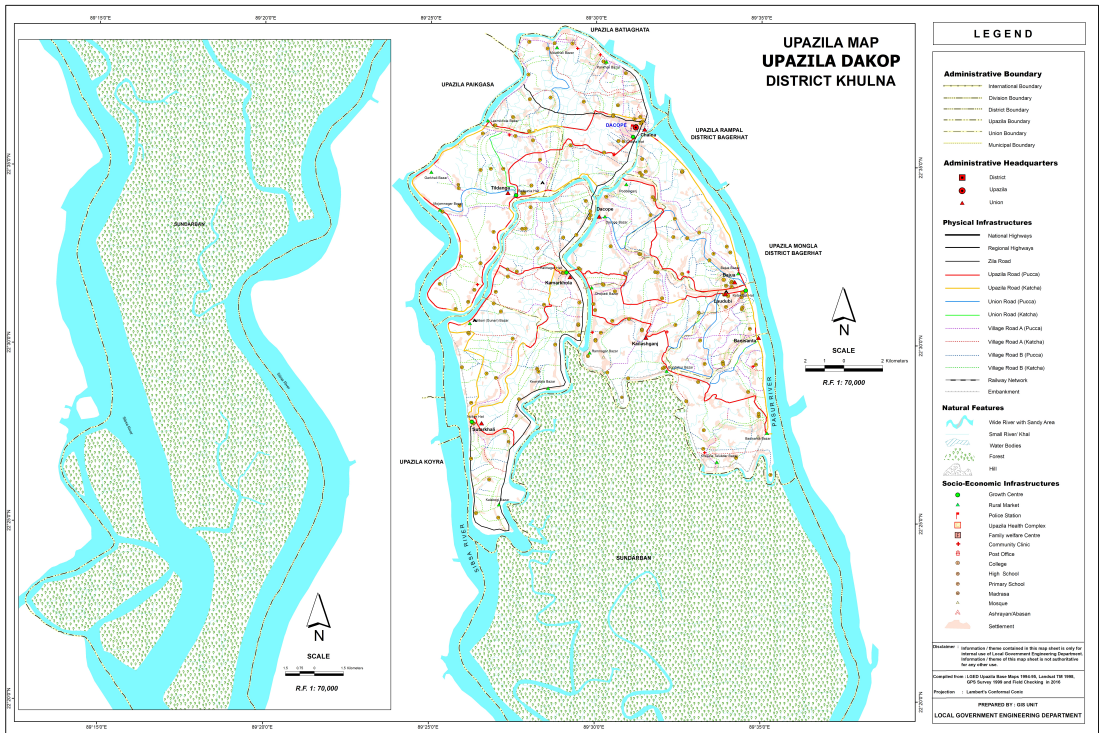


Figure 4.1: Map of Dacope sub-district

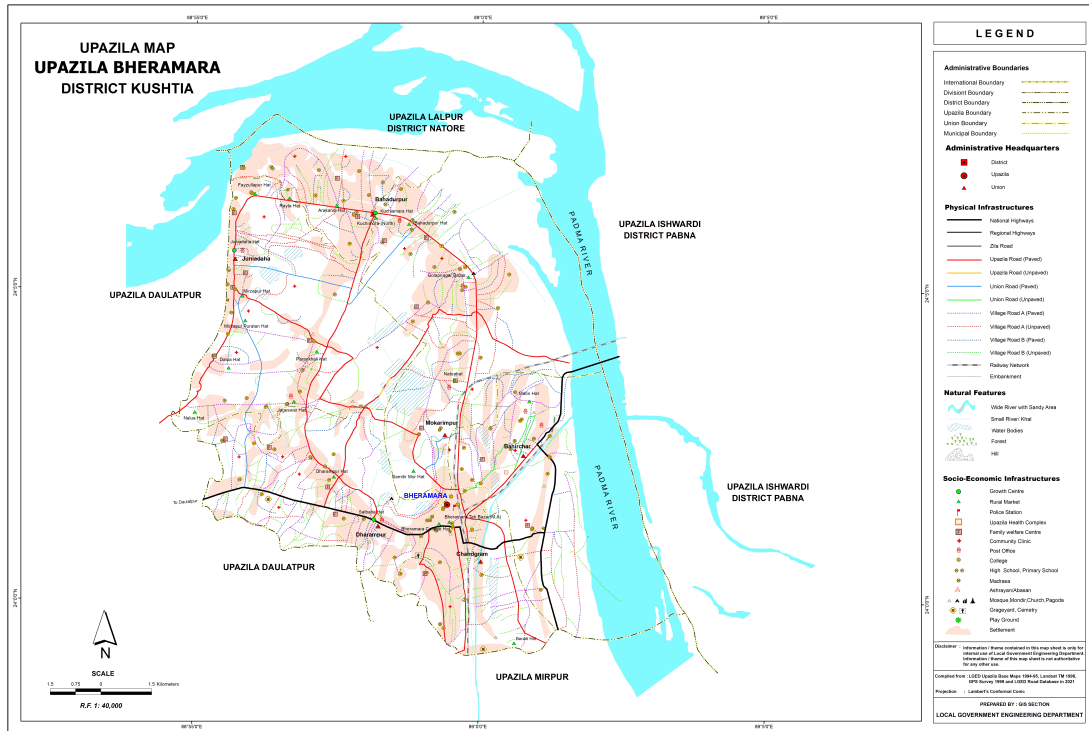


Figure 4.2: Map of Bheramara sub-district

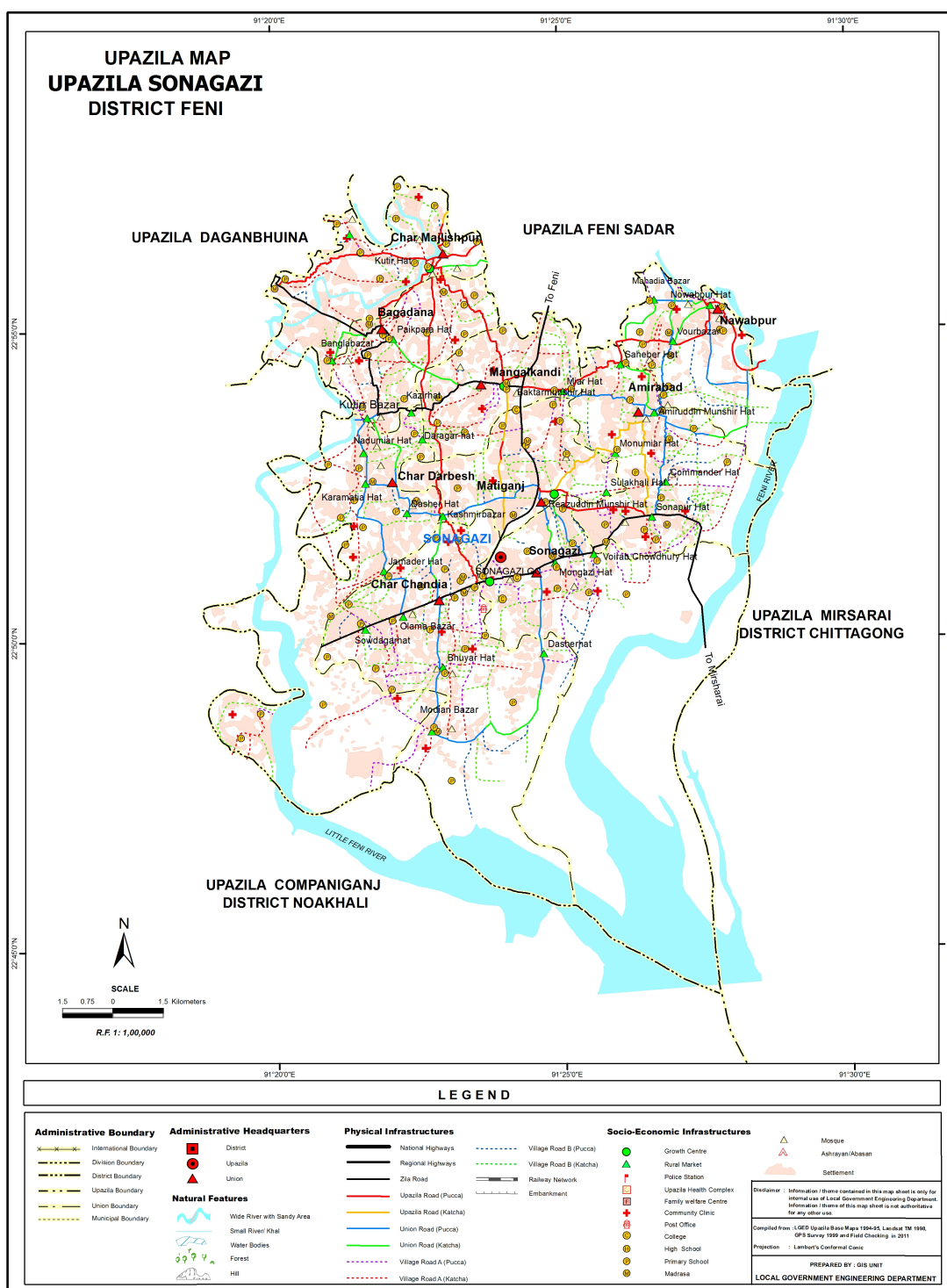


Figure 4.3: Map of Sonagazi sub-district



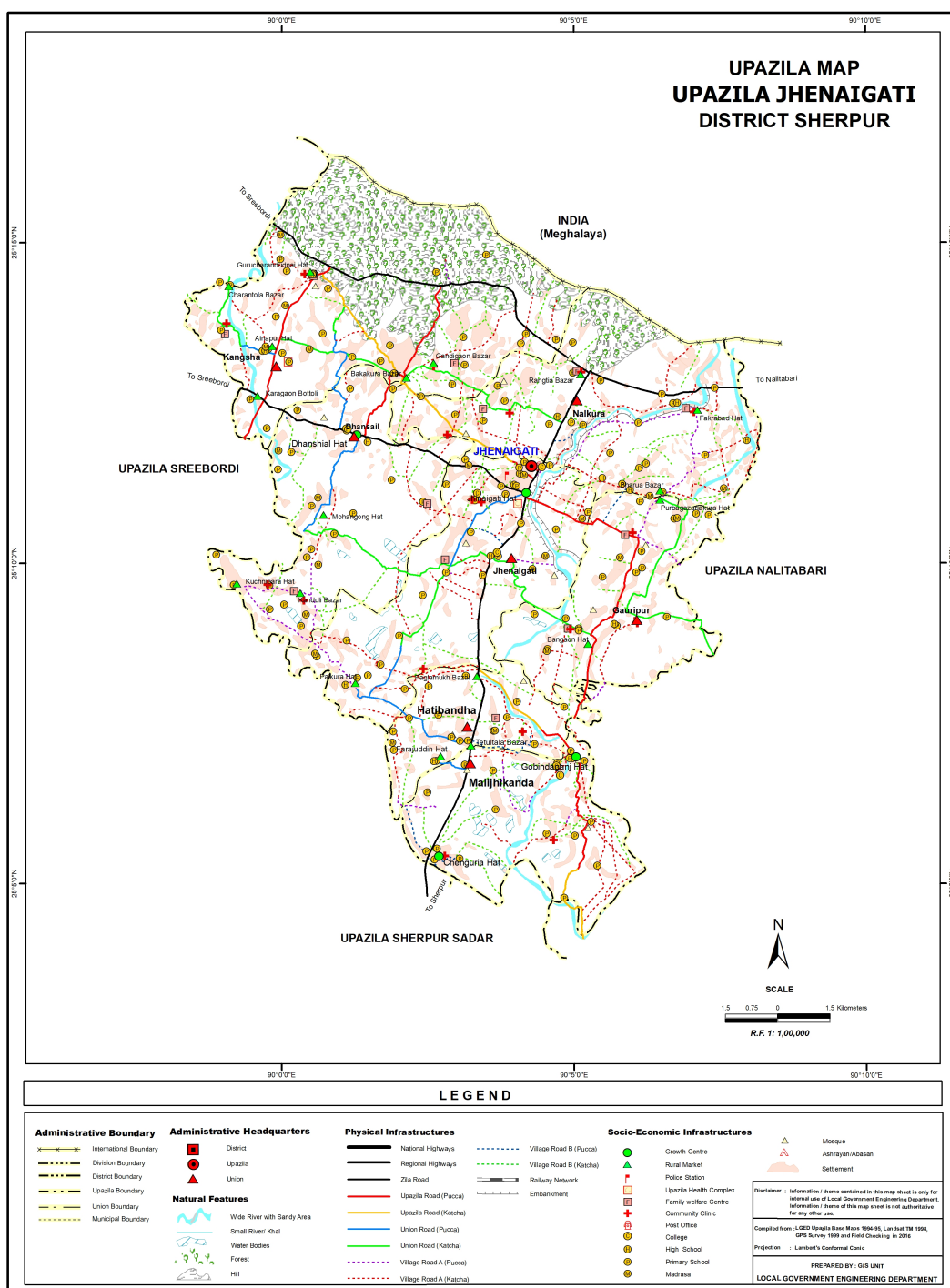


Figure 4.5: Map of Jhenaigati sub-district

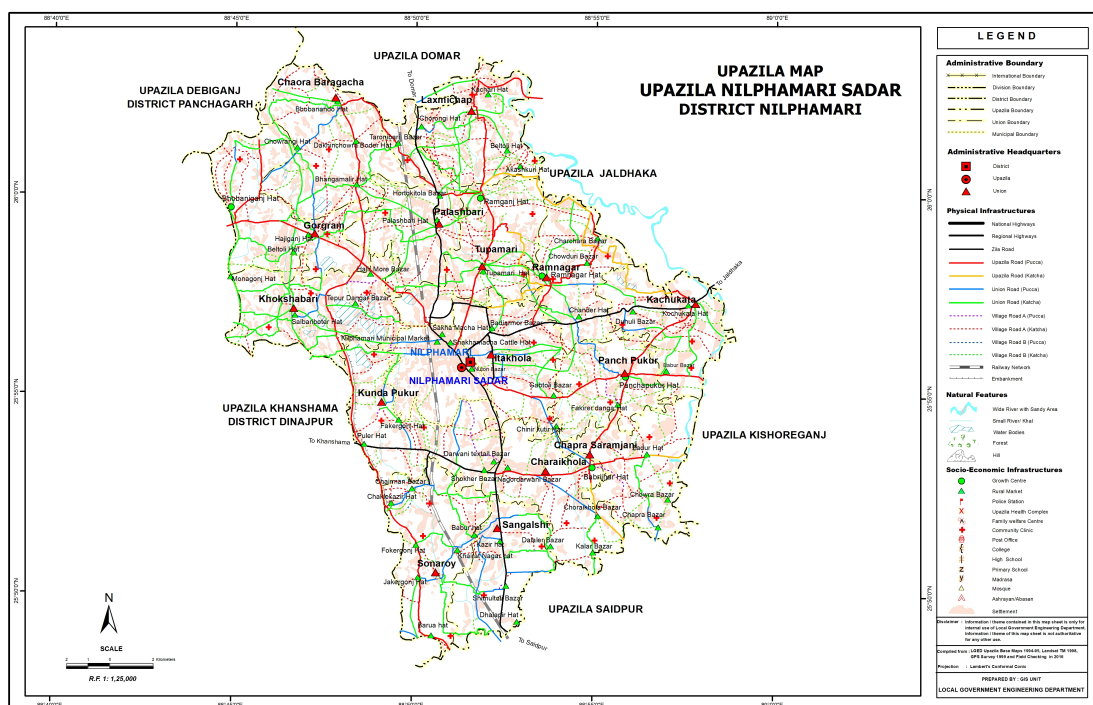


Figure 4.6: Map of Nilphamari-Shodor sub-district

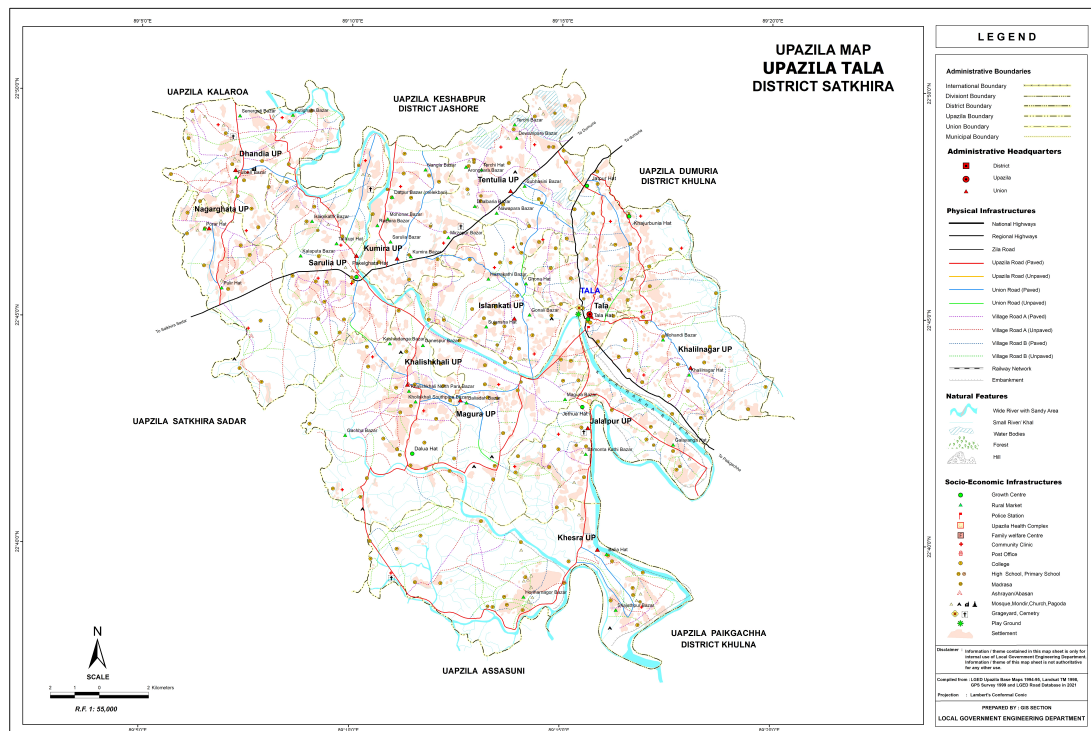


Figure 4.7: Map of Tala sub-district

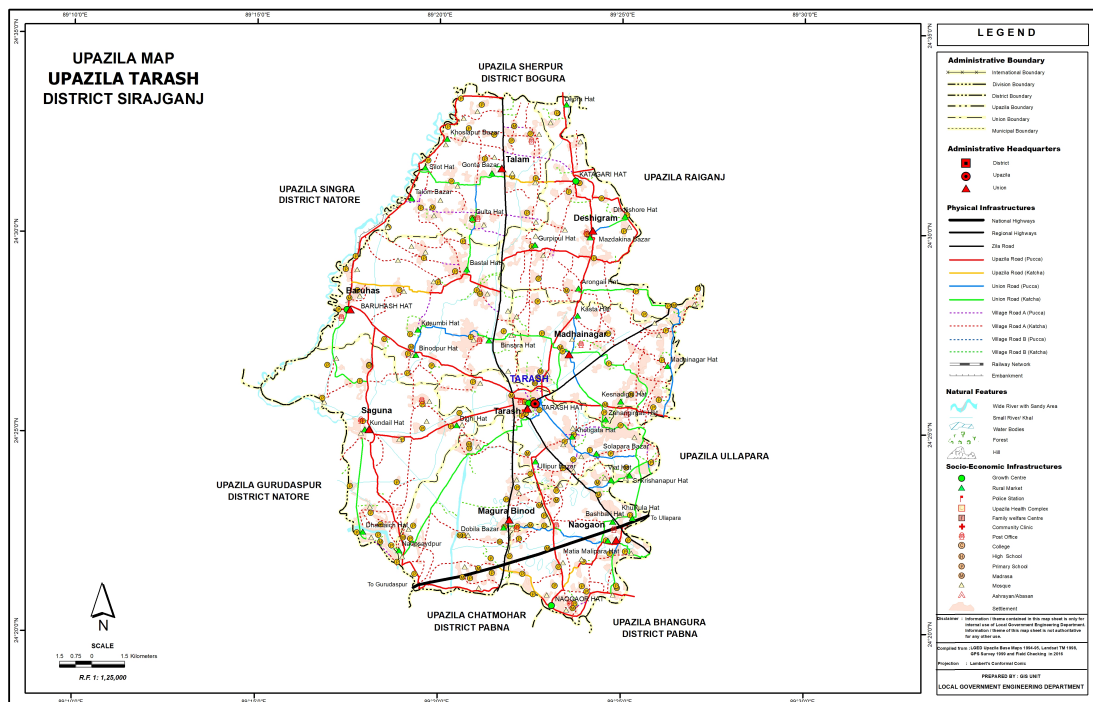


Figure 4.8: Map of Tarash sub-district

4.3 Sample Data

Mrittika Dol	Bhumi	Bunot	N	P	K	S	Z	B
Bajowa	mh	2	0.093	2.72	0.97	208.35	0.52	1.15
Bajowa	mh	2	0.09	2.28	0.96	315.65	1.65	1.01
Bajowa	mh	2	0.144	2.42	0.82	246.17	1.19	1.05
Bajowa	mh	2	0.133	2.85	1.08	222.4	0.47	0.95
Bajowa	mh	2	0.151	5.88	0.82	177.56	0.47	0.95
Bajowa	mh	2	0.134	13.01	0.6	265.37	1.01	0.73
Bajowa	mh	2	0.134	13.55	0.58	225.22	1	0.56
Bajowa	mh	2	0.107	5.74	0.78	170.35	0.49	0.81
Bajowa	mh	2	0.12	4.84	0.86	143.37	0.46	0.78
Jhalkathi	mh	3	0.12	5.22	0.71	57.58	0.57	13.84
Jhalkathi	mh	3	0.179	4.41	0.68	41.16	0.79	10.26
Jhalkathi	mh	3	0.086	4	0.7	52.62	0.63	11.6
Amnura	ml	1	0.18	3.74	0.25	9.55	0.59	0.24
Amnura	ml	1	0.12	5.14	0.16	7.67	2.62	0.13
Ekdala	ml	2	0.12	1.93	0.4	15.64	1.27	0.13
Ekdala	ml	2	0.13	2.5	0.3	10.29	0.92	0.75
Ekdala	ml	2	0.15	2.15	0.26	4.81	1.23	0.22
Ekdala	mh	2	0.18	13.91	0.24	22.65	1.08	0.72
Ekdala	mh	2	0.17	23.48	0.27	6.58	1.29	0.71
Ekdala	mh	2	0.17	8.81	0.19	16.59	0.24	0.61
Ishwardi	h	3	0.1	10.55	0.5	115.24	0.92	0.51
Gopalpur	mh	2	0.1	14.24	0.32	2.88	0.72	0.91
Ishwardi	ml	3	0.11	40.72	0.27	24.57	1.27	0.32

Table 4.1: Sample Soil Data

The data shows us the soil type, its name, classification and most importantly the N,P,K and S,Zn,B values.

Crop	N	P	K	S	Z	B
Rosun	19	8	15	8	0.5	0.5
Kheshari	3	3	3	2	0	0
Soabin	4	5	10	3	0	0
Tula	15	6	13	3	0	0
Mugg	3	5	6	3	0	0

Table 4.2: Crop Data

Also we have chosen 5 crops from the books to see from these 5 crops which crops are most suitable for the lands. We gathered the suitable N,P,K and S,Zn,B values of the crops which are needed to grow them.

4.4 Experimental Setup

In Table 1, "h" stands for High, "mh" stands for Medium High, "ml" stands for Medium Low, and "l" is for Low. Bunot, on the other hand, is an indicator of soil classification. Doash, EtelDoash, Etel, Bele doash, and Bele are represented by the letters "Bunot" 1,2,3,4,5. We gathered the N,P,K, and S,Zn,B of existing soil samples, as well as different crop needs in the region, to determine the crop best suited for a certain soil sample. We utilized the Euclidean Distance approach to find the best crops after collecting the N,P,K, and S,Zn,B data. The Euclidean Distance with the lowest value denotes the best crop for a certain soil sample.

Between two points in Euclidean space, the Euclidean distance (or, more generally, in a real inner product space) is the length of the shortest path between the two points, measured along a straight line. Formally, given two points, x and y , in a real inner product space, the Euclidean distance between them is defined as the square root of the inner product of the difference between the two points with itself.

$$\sigma = \sqrt{(n_s - n_c)^2 + (p_s - p_c)^2 + (k_s - k_c)^2 + (s_s - s_c)^2 + (z_s - z_c)^2 + (b_s - b_c)^2}$$

We have a soil sample and a crop sample with six variables (e.g. N,P,K,S,Zn,B). This will give us a single number in terms of the considered variables representing the distance between the soil and crop samples. In terms of these variables the soil and crops samples will be similar if the distance is short.

4.5 Data Visualization

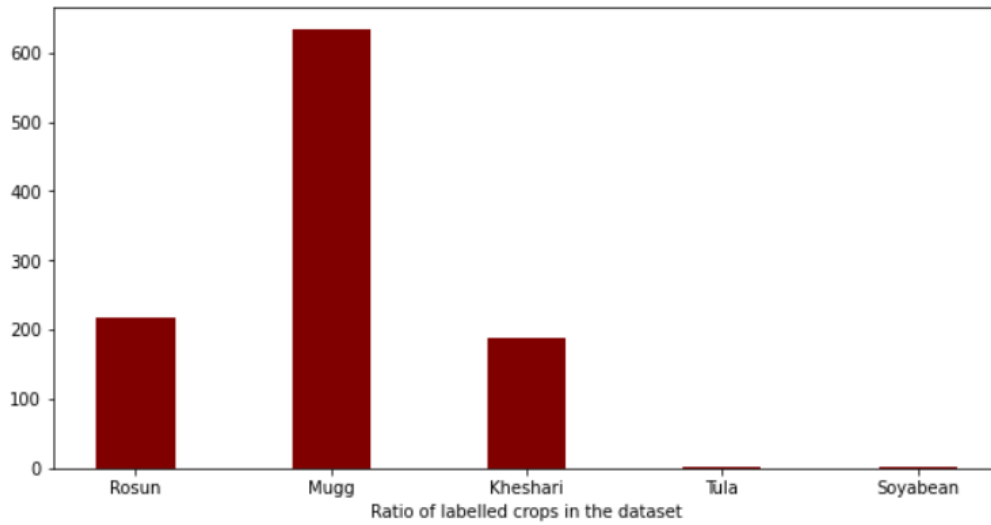


Figure 4.9: Crop amount

After calculating the euclidean distance between soil and crop, we have got the above crop recommendation for soil. We have got 634 Mugg, 212 Rosun, 186 Kheshari, 8 Tula, 7 Soyabean. The amount of Mugg was high from other crops, and the amount of Tula and Soyabean was negligible.

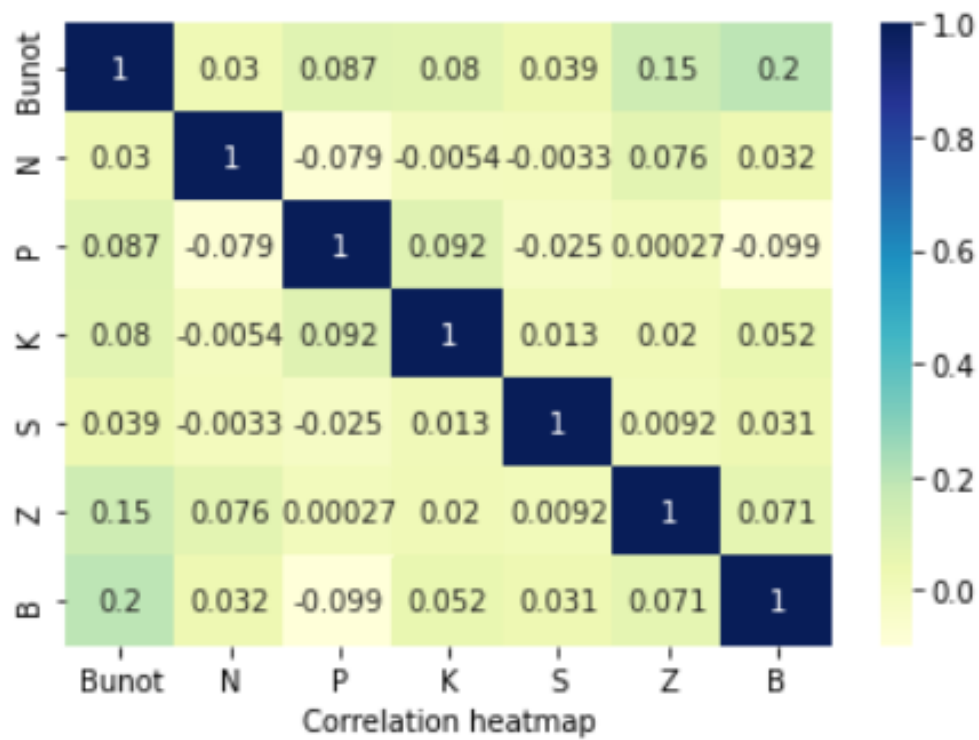


Figure 4.10: Correlation Heatmap

A correlation heatmap shows the correlation coefficients between various variables in a dataset graphically. It is a valuable tool for exploratory data analysis, as it makes it possible to see the direction and strength of linear correlations between variables. The heatmap typically uses a color-coding scheme, where warm colors (such as red or yellow) indicate a positive correlation, cool colors (such as blue) indicate a negative correlation, and neutral colors (such as white or gray) indicate no correlation. The degree of intensity of the color can indicate the degree of correlation. This visualization technique can facilitate the identification of potential confounding variables and aid in selecting features for modeling. There is a positive correlation between P and K. When two variables change in the opposite directions. When one rises, the other falls and there is a negative correlation. For instance, P and N have a negative correlation.

4.6 Data Preprocess

Mruttika Dol	Bhumi	Bunot	N	P	K	S	Z	B	Crop
Bajowa	mh	2	0.093	2.72	0.97	208.35	0.52	1.15	Rosun
Bajowa	mh	2	0.09	2.28	0.96	315.65	1.65	1.01	Rosun
Bajowa	mh	2	0.144	2.42	0.82	246.17	1.19	1.05	Rosun
Bajowa	mh	2	0.133	2.85	1.08	222.4	0.47	0.95	Rosun
Bajowa	mh	2	0.151	5.88	0.82	177.56	0.47	0.95	Rosun
Bajowa	mh	2	0.134	13.01	0.6	265.37	1.01	0.73	Rosun
Bajowa	mh	2	0.134	13.55	0.58	225.22	1	0.56	Rosun
Bajowa	mh	2	0.107	5.74	0.78	170.35	0.49	0.81	Rosun
Bajowa	mh	2	0.12	4.84	0.86	143.37	0.46	0.78	Rosun
Jhalkathi	mh	3	0.12	5.22	0.71	57.58	0.57	13.84	Mugg
Jhalkathi	mh	3	0.179	4.41	0.68	41.16	0.79	10.26	Mugg
Jhalkathi	mh	3	0.086	4	0.7	52.62	0.63	11.6	Mugg
Amnura	ml	1	0.18	3.74	0.25	9.55	0.59	0.24	Kheshari
Amnura	ml	1	0.12	5.14	0.16	7.67	2.62	0.13	Kheshari
Ekdala	ml	2	0.12	1.93	0.4	15.64	1.27	0.13	Kheshari
Ekdala	ml	2	0.13	2.5	0.3	10.29	0.92	0.75	Kheshari
Ekdala	ml	2	0.15	2.15	0.26	4.81	1.23	0.22	Kheshari
Ekdala	mh	2	0.18	13.91	0.24	22.65	1.08	0.72	Mugg
Ekdala	mh	2	0.17	23.48	0.27	6.58	1.29	0.71	Mugg
Ekdala	mh	2	0.17	8.81	0.19	16.59	0.24	0.61	Mugg
Ishwardi	h	3	0.1	10.55	0.5	115.24	0.92	0.51	Rosun
Gopalpur	mh	2	0.1	14.24	0.32	2.88	0.72	0.91	Mugg
Ishwardi	ml	3	0.11	40.72	0.27	24.57	1.27	0.32	Mugg

Table 4.3: Sample Soil Data with crop

The decision was made to eradicate the Tula and Soyabean crops due to low recommendations. Eradicating these crops may also have been a strategic move to allocate resources toward more profitable or sustainable options. This decision was likely made after thorough analysis and consideration and was likely carried out in a formal and sophisticated manner. (Fig.4.1).

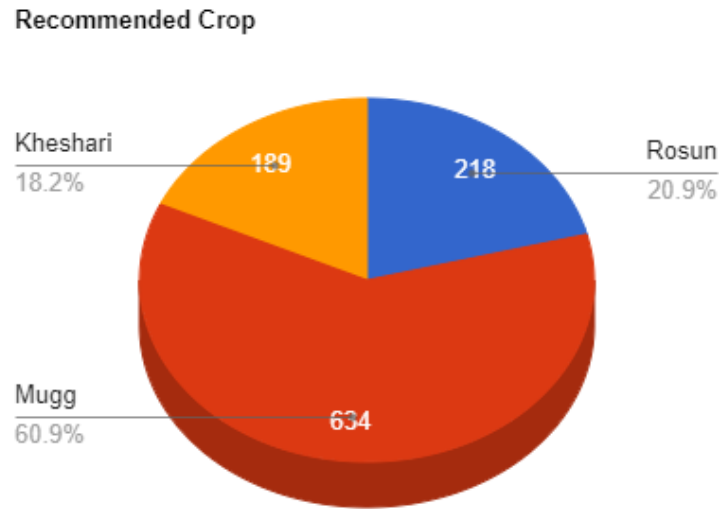


Figure 4.11: Recommended Crops

During our analysis, we have identified a total of 218 occurrences of the letter 'Rosun', 634 occurrences of the letter 'Mugg' and 189 occurrences of the letter 'Kheshari'. (Fig.4.3)

4.7 Data Split

We have divided the data into an 80/20 split, where we used 80% of the data for training, and 20% is used for testing. This allows for a robust evaluation of the model's performance, as the model is tested on unseen data. This method is widely used in the machine learning community, and it is considered a best practice to have a significant portion of the data set aside for testing.

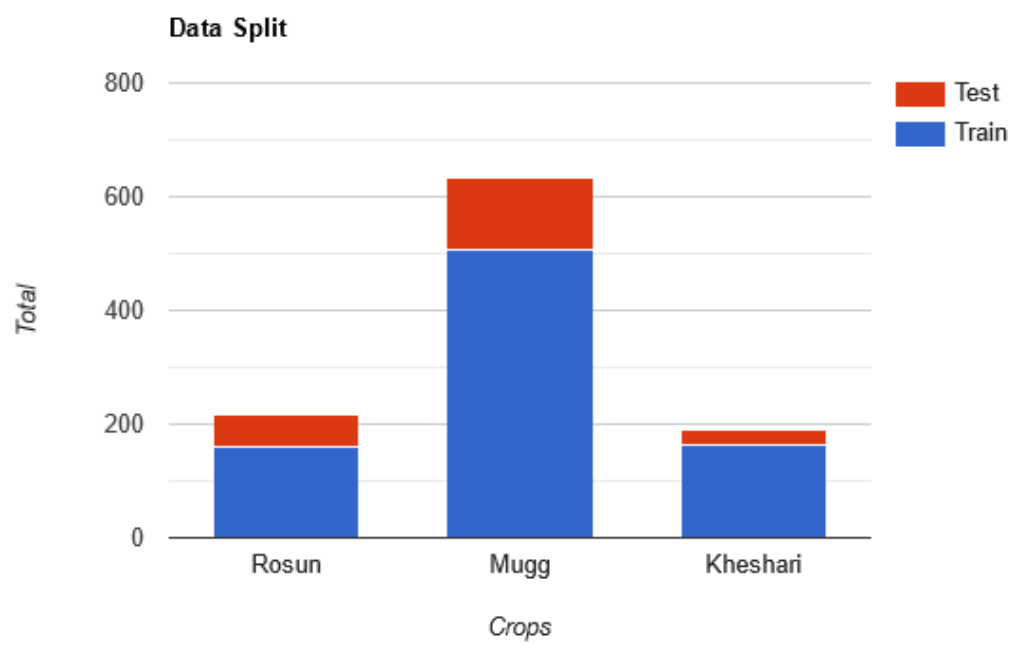


Figure 4.12: Train and Test data amount

Chapter 5

Performance Matrices

5.1 Descriptions

1) *TruePositive* (**TP**): A true positive is an instance in which a model correctly identifies a positive sample or class. This type of outcome is typically denoted as "TP" and it is one of the essential measures used to assess the effectiveness of the classification model. It represents the number of instances in which the model's prediction aligns with the actual class label and is a vital indicator of the model's ability to identify positive samples accurately.

TrueNegative (**TN**): A true negative (TN) is an instance in which the model correctly identifies a negative class. Specifically, a true negative occurs when the model predicts the hostile class, and the actual class of the instance is indeed the negative class. This outcome indicates the model's ability to identify instances that do not belong to the positive class.

FalsePositive (**FP**): A False Positive (FP) refers to a situation where a model incorrectly classifies an instance as belonging to a specific class despite belonging to another class. This error can be particularly detrimental in specific applications, such as medical diagnosis or fraud detection, as it may lead to unnecessary treatment or investigation. A False Positive rate, the proportion of instances incorrectly classified as positive out of all actually negative instances, is often used to measure a model's performance. In order to mitigate the impact of False Positives, it is essential to carefully evaluate the trade-offs between model accuracy and the rate of False Positives and to select an appropriate threshold for classifying instances.

FalseNegative (**FN**): A false negative (also known as a type II error) is a situation where a model incorrectly classifies an instance as belonging to a negative class when it belongs to a positive class. This type of error can have significant consequences, particularly in applications such as medical diagnosis or fraud detection, where the cost of misclassifying a positive instance can be high. Formally, a false negative can be defined as the number of instances that are positive but are predicted to be negative, divided by the total number of actual positive instances. It is an important performance metric to consider when evaluating the effectiveness of a machine-learning model.

2) **Accuracy**: Accuracy is a statistic that measures the percentage of correct pre-

dictions made by the machine-learning model. Formally, the accuracy of a model can be defined as:

$$Accuracy = \frac{TruePositive + TrueNegative}{TruePositive + TrueNegative + FalsePositive + FalseNegative}$$

where True Positives represent the number of instances that are correctly classified as positive, True Negatives represent the number of instances that are correctly classified as negative, and Total Predictions represent the total number of instances that are classified by the model.

3) **Precision:** In machine learning, precision refers to the percentage of true positive predictions (correctly identified positive instances) among all positive predictions made by a model (false positives + true positives). It is a measure of a model's ability to identify positive instances correctly and can be used to assess a binary classification model's efficiency. A model with high precision has a low false positive rate, meaning it only occasionally identifies a negative instance as positive.

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive}$$

4) **Recall:** In machine learning, recall is a metric that measures the proportion of relevant instances that were correctly identified by the classifier. It is characterized as the ratio of the number of true positives (appropriately classified occurrences as being in the positive class) to the sum of the true positives and the false negatives (instances that were actually positive but were incorrectly classified as negative). In other words, recall measures how many of the instances that should have been identified as positive were identified as such by the classifier.

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative}$$

5) **F1-Score:** A well-known metric used to assess the performance of a binary classification model is the F1-score which is also known as the F-measure or F-score. It is the harmonic mean of precision and recall, where precision is the proportion of true positive predictions among all positive predictions, and recall is the proportion of true positive predictions among all actual positive instances. It denotes improved performance when the F1-score ranges from 0 to 1. It is beneficial when the distribution of classes in the target variable is imbalanced, as it places equal weight on both precision and recall, whereas metrics such as accuracy can be misleading in such situations.

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

6) **Confusion Matrix:** A confusion matrix is used as a machine-learning tool to assess how better the classification method works. It is also known as a contingency

table or as an error matrix. By summarizing the true negative and positive, also the false negative and positive proportion, the table defines the representation of a classification method. The rows of the matrix represent the actual class, while the columns represent the predicted class. The number of observations known to belong to a particular class and those predicted to belong to a particular class make up the entries in the matrix. The number of accurate predictions is represented by the entries inside the diagonal of the matrix, while the number of inaccurate predictions is represented by the entries outside the diagonal. Overall, the confusion matrix provides a clear and concise summary of the performance of a classification algorithm, allowing for easy identification of areas where the algorithm may be improved.

7) **Auc-Roc Curve:** The performance of a binary classifier system when the discrimination threshold is changed is graphically represented by the Receiver Operating Characteristic (ROC) curve. How well a model can distinguish between positive and negative classes is measured by the area under the ROC curve (AUC ROC). AUC ROC is basically a value between 0 and 1, a random classifier is represented by value 0.5, and a perfect classifier is represented by value 1. In general, a higher AUC ROC value indicates better classification performance.

Chapter 6

Result Analysis

6.1 Random Forest

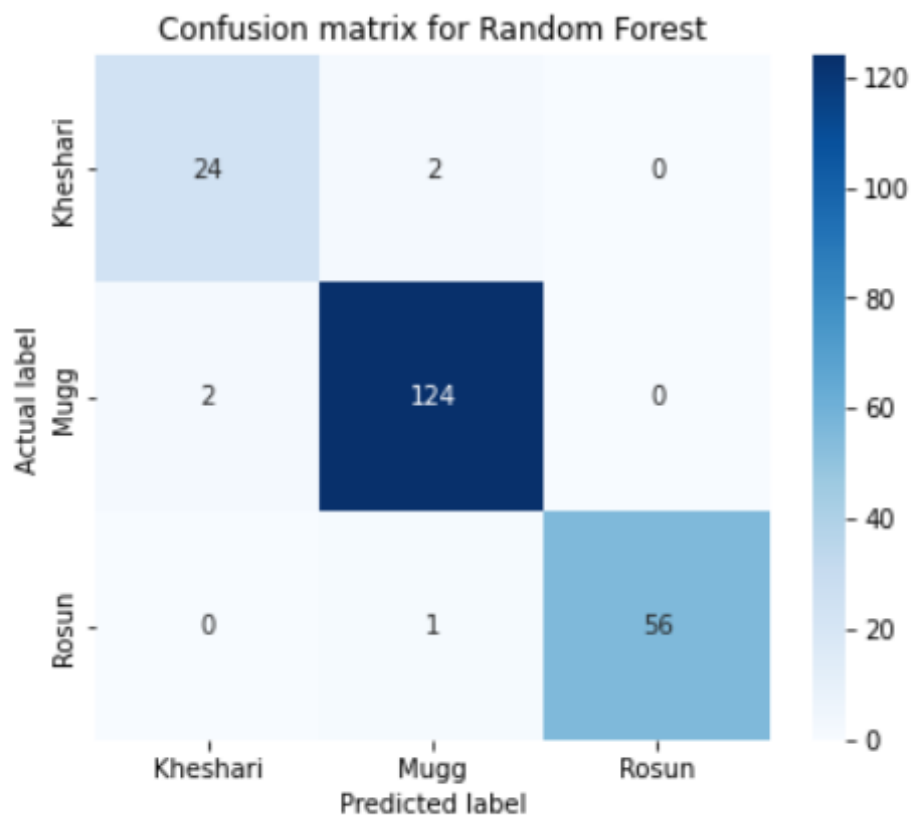


Figure 6.1: Confusion Matrix for Random Forest

As we can see in the Random Forest algorithm, the actual Kheshari was 26, but the algorithm predicted 24 kheshari, and the remaining two were Mugg. The actual Mugg was 126, but the predicted amount of mugg was 124, and the remaining two were Kheshari. The actual Rosun was 57, but the predicted value was 56, and the remaining one was mugg.

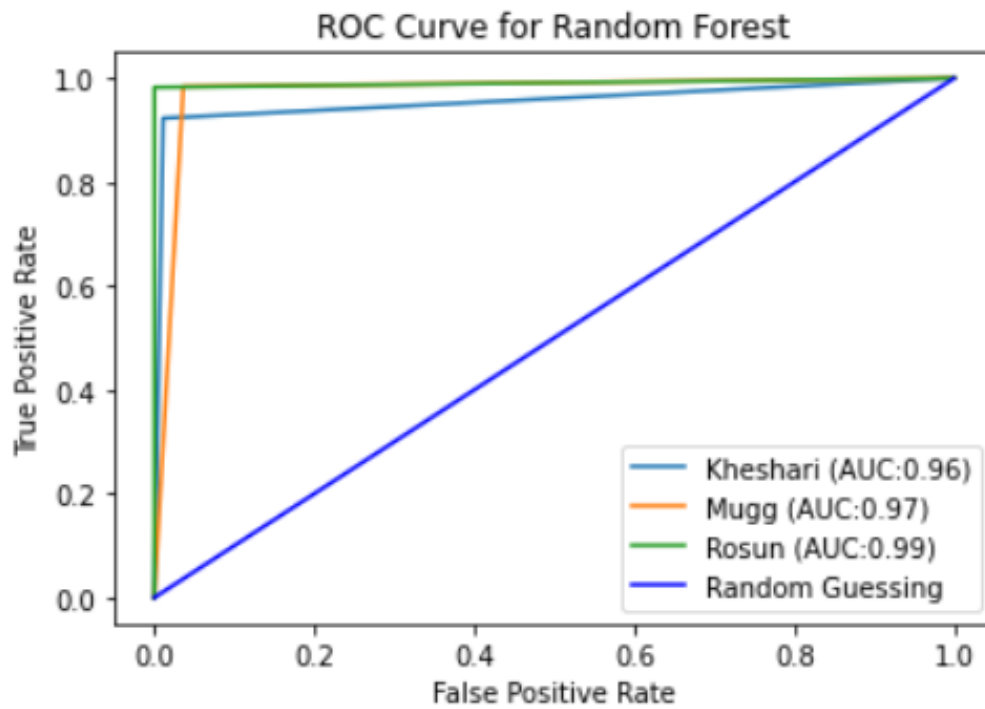


Figure 6.2: ROC for Random Forest

In the Roc Curve of Random Forest we saw that the AUC score of Rosun was the highest among all which was 0.99. This simply means that Random Forest algorithm was able to predict the amount of Rosun with highest accuracy, followed by Mugg (0.97) and kheshari (0.96).

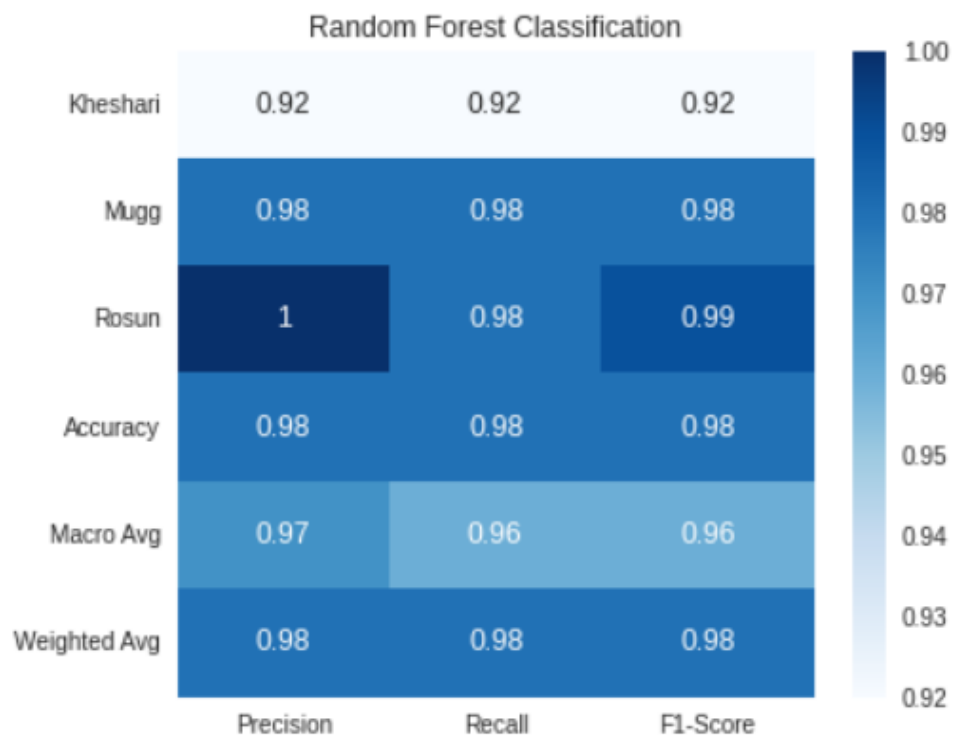


Figure 6.3: Classification for Random Forest

After using the Random forest Algorithm we have got the above values for precision, recall, f1-score, accuracy, macro avg and weighted avg. We achieved a 98% accuracy.

6.2 Decision Tree

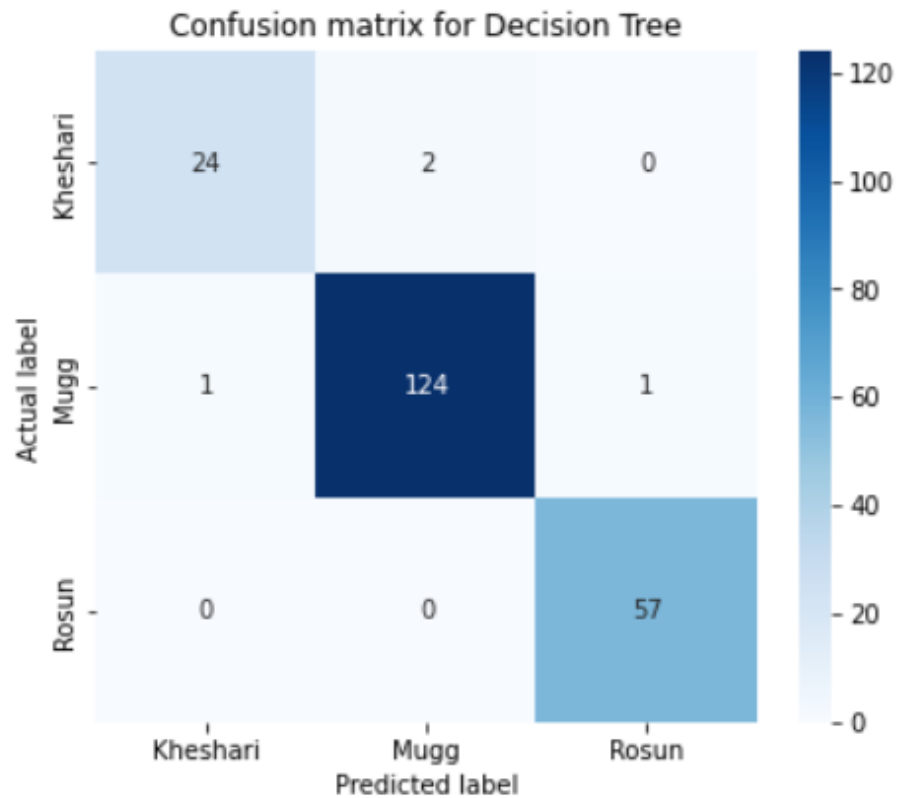


Figure 6.4: Confusion Matrix for Decision Tree

As we can see in the Decision tree algorithm, the actual Khashari was 26, but the algorithm predicted 24 khashari, and the remaining two were Mugg. The actual Mugg was 126, but the predicted amount of mugg was 124, and the remaining two were Khashari and Rosun. The actual Rosun was 57 and the predicted value was 57.

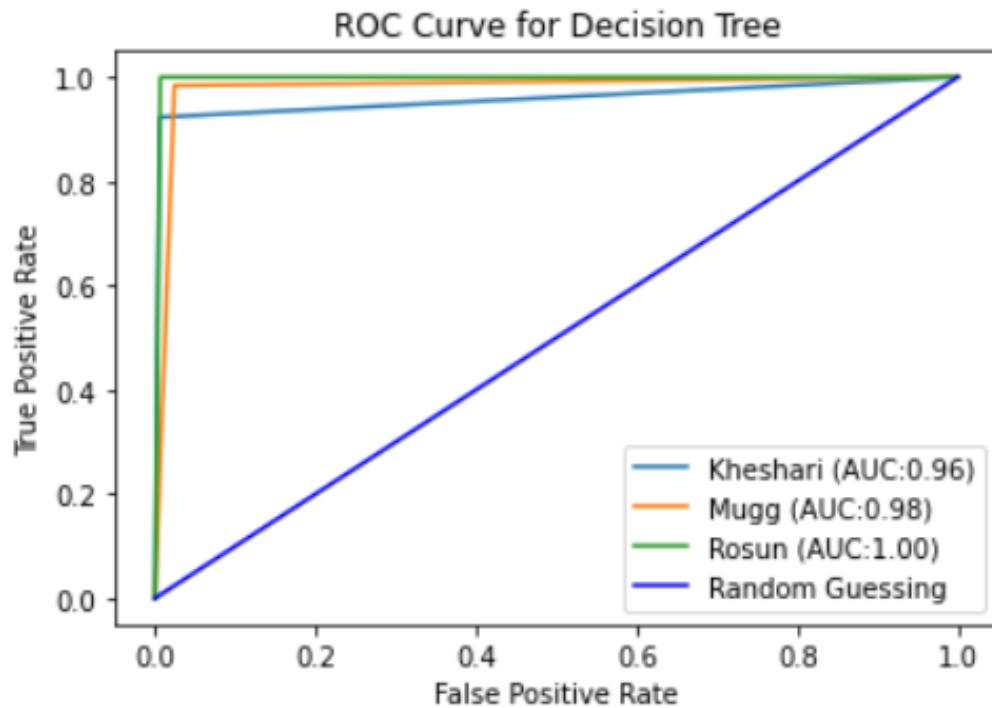


Figure 6.5: ROC for Decision Tree

In the Roc Curve of Decision Tree we saw that the AUC score of Rosun was the highest among all which was 1. This simply means that Decision Tree algorithm was able to predict the amount of Rosun with highest accuracy, followed by Mugg (0.98) and khesari (0.96).

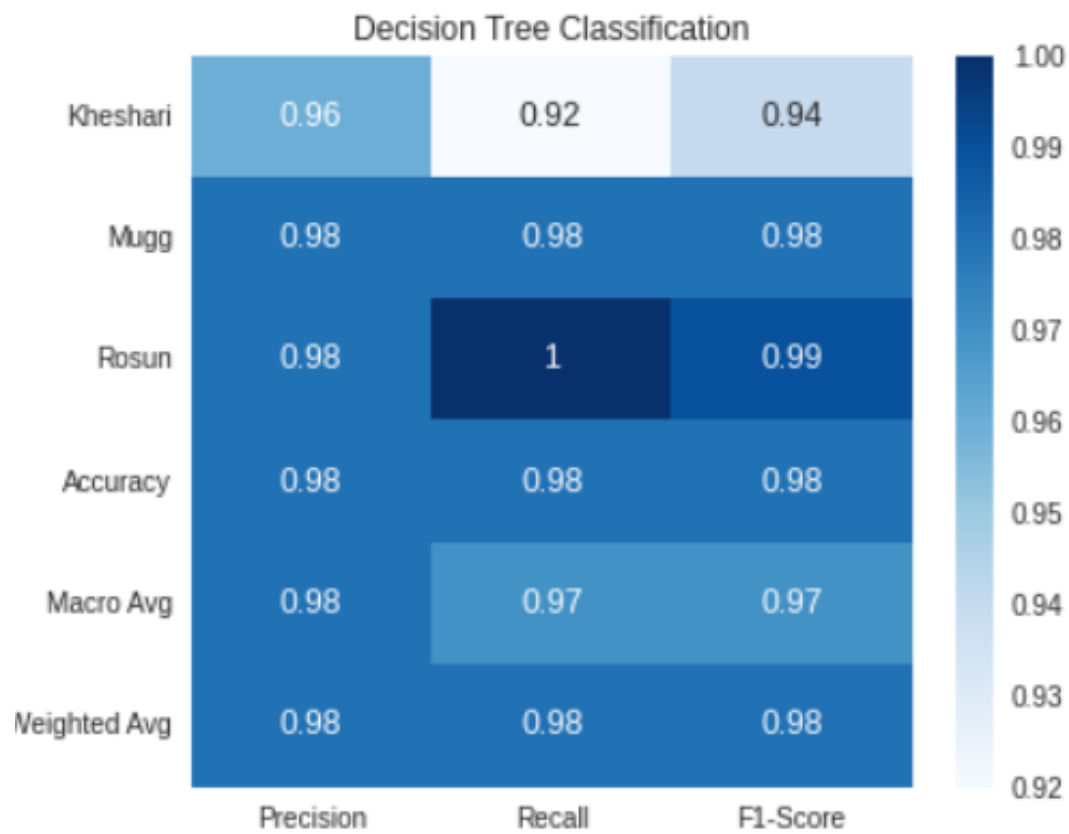


Figure 6.6: Classification for Decision Tree

After using the Decision Tree Algorithm we have got the above values for precision, recall, f1-score, accuracy, macro avg and weighted avg. We achieved a 98% accuracy.

6.3 Naive Bayes

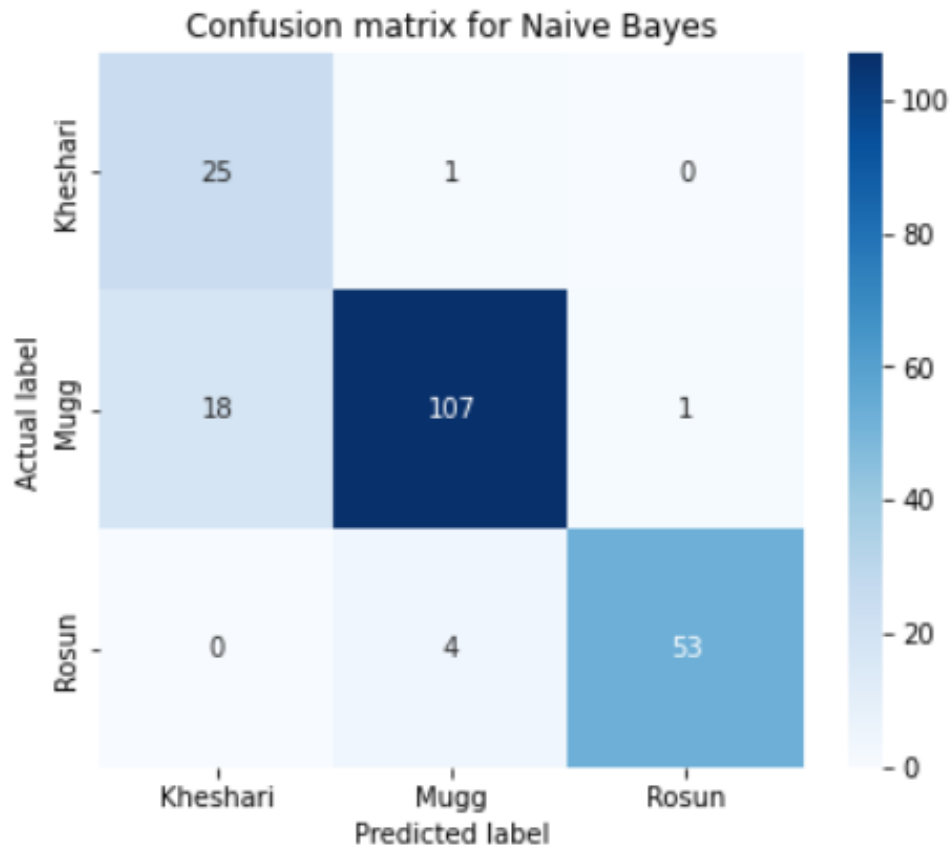


Figure 6.7: Confusion Matrix for Naive Bayes

As we can see in the Naive Bayes algorithm, the actual Khashari was 26, but the algorithm predicted 25 khashari, and the remaining one was Mugg. The actual Mugg was 126, but the predicted amount of mugg was 107, and from the remaining 18 were Khashari and 1 was Rosun. The actual Rosun was 57 and the predicted value was 53 and the remaining 4 are Mugg.

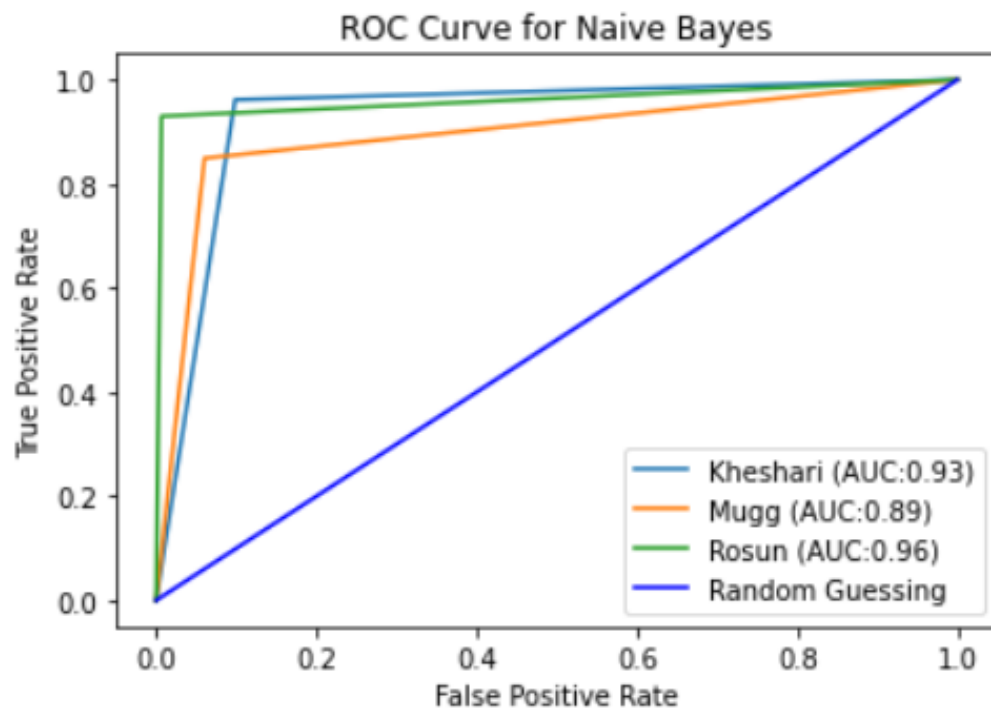


Figure 6.8: ROC for Naive Bayes

In the Roc Curve of Naive Bayes we saw that the AUC score of Rosun was the highest among all which was 0.96. This simply means that Naive Bayes algorithm was able to predict the amount of Rosun with highest accuracy, followed by khashari (0.93) and Mugg (0.89).

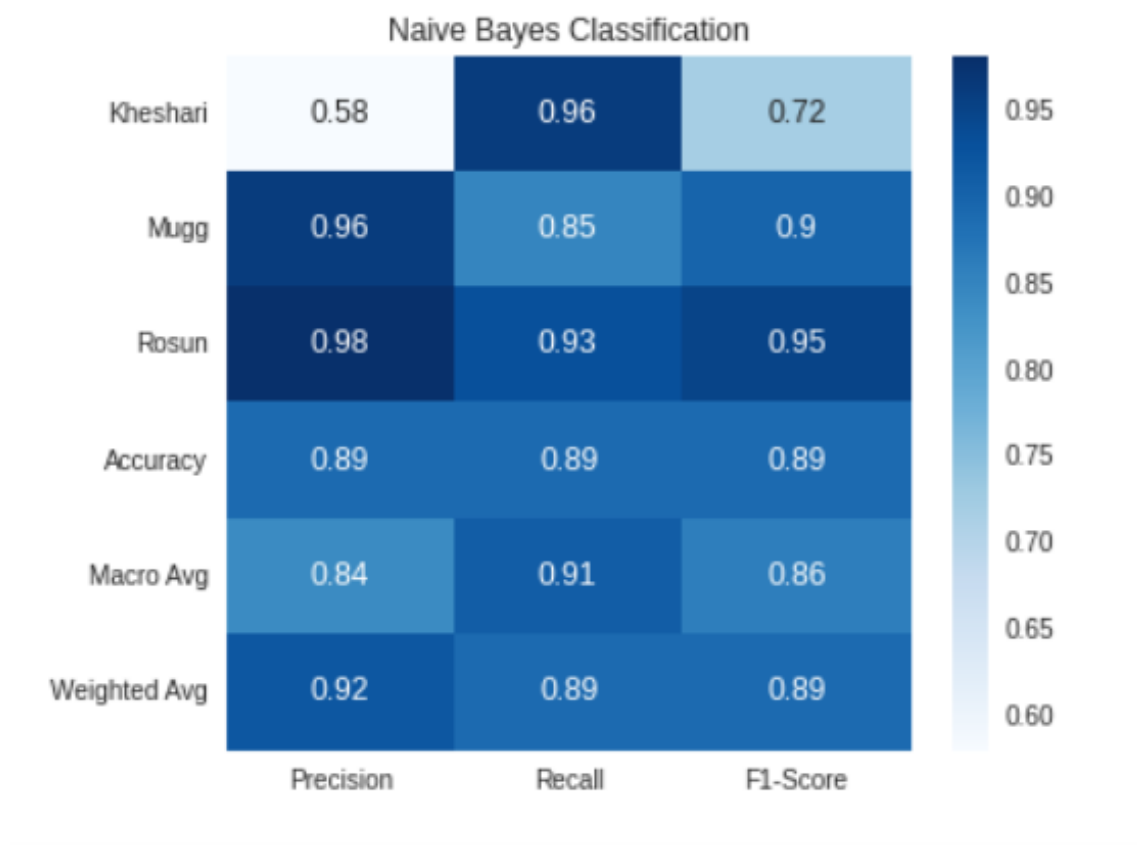


Figure 6.9: Classification for Naive Bayes

After using the Naive Bayes Algorithm we have got the above values for precision, recall, f1-score, accuracy, macro avg and weighted avg. We achieved a 89% accuracy.

6.4 Ensemble

In the Ensemble Model we combined three algorithms which are Random Forest, Decision Tree and Naive Bayes.

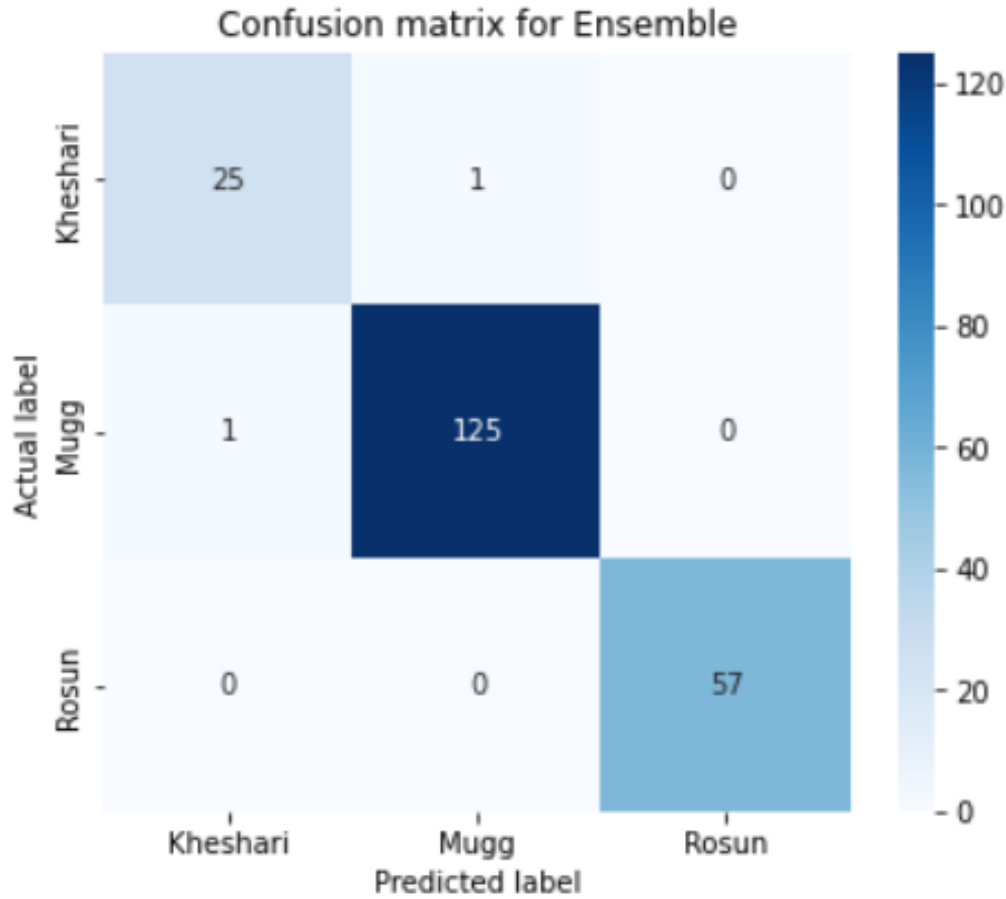


Figure 6.10: Confusion Matrix for Ensemble

As we can see in the Ensemble Model, the actual Kheshari was 26, but the Model predicted 25 kheshari, and the remaining one was Mugg. The actual Mugg was 126, but the predicted amount of mugg was 125 and the remaining 1 was Kheshari. The actual Rosun was 57 and the predicted value was 57.

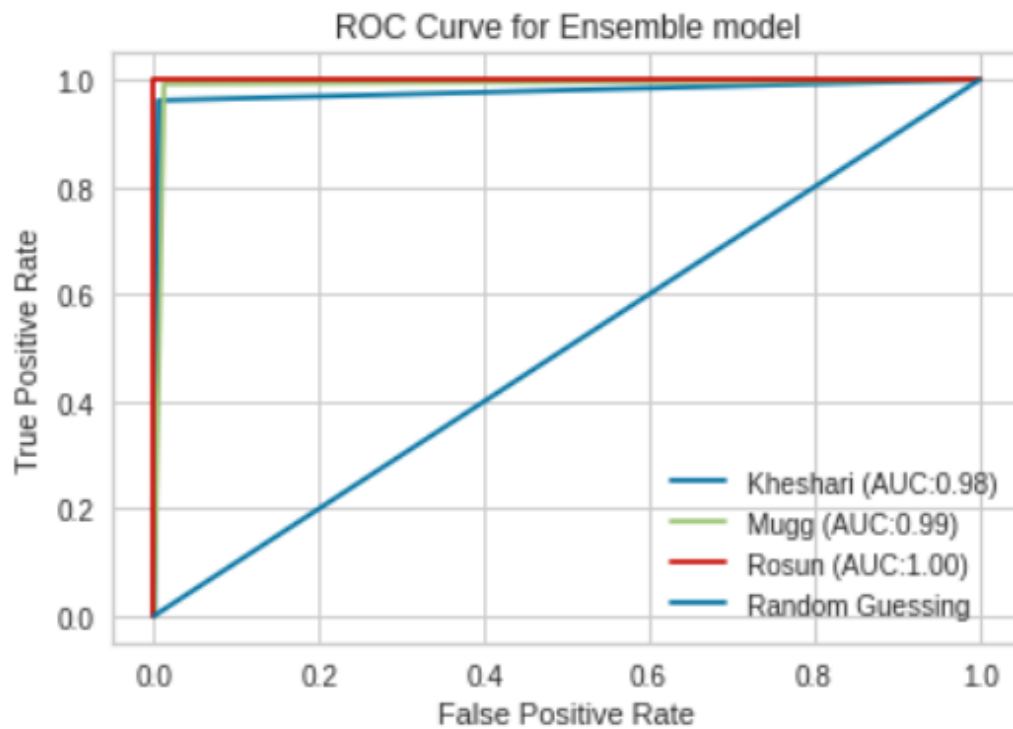


Figure 6.11: ROC for Ensemble

In the Roc Curve of Ensemble model we saw that the AUC score of Rosun was the highest among all which was 1. This simply means that Ensemble model was able to predict the amount of Rosun with highest accuracy, followed by Mugg (0.99) and khesari (0.98).

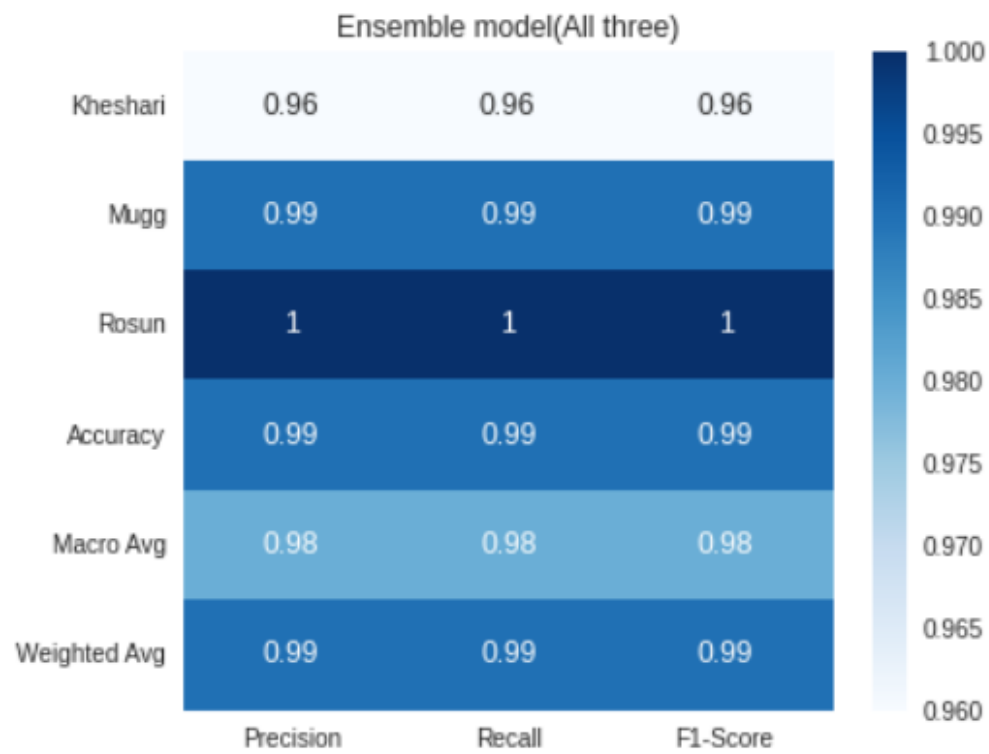


Figure 6.12: Classification for Ensemble

After Ensemble all three algorithms (Decision tree, Naive Bayes and Random Forest), we have got the above values for precision, recall, f1-score. We achieved a 99% accuracy.

6.5 Accuracy Comparison

Random Forest	Accuracy	0.98	0.98	0.98
	Macro avg	0.97	0.96	0.96
Decision Tree	Accuracy	0.98	0.98	0.98
	Macro avg	0.98	0.97	0.97
Naive Bayes	Accuracy	0.89	0.89	0.89
	Macro avg	0.84	0.91	0.86
Ensemble model	Accuracy	0.99	0.99	0.99
	Macro avg	0.98	0.98	0.98
		Precision	Recall	F1-score

Table 6.1: Accuracy Graph

As we can see in the accuracy table, the Decision Tree and Random Forest algorithm shows the best accuracy and the Naive Bayes shows the minimum accuracy among the Algorithms. The accuracy of the Decision Tree algorithm and Random Forest algorithm is almost 98%. After Ensemble the three algorithms we achieved an accuracy of 99% .

After analyzing all the values we saw a slightly lower scores for Kheshari in all the algorithm which is because of the data limitation on our dataset. The amount of actual and predicted Kheshari is not a high a value which is the reason for this lower output.

6.6 Explainable-AI

We used Lime to implement XAI to explain how our Ensemble Model performs and predicts. The explanations generated by LIME will show which features of the input data are most important for the ensemble model’s predictions and how they contribute to the final prediction.

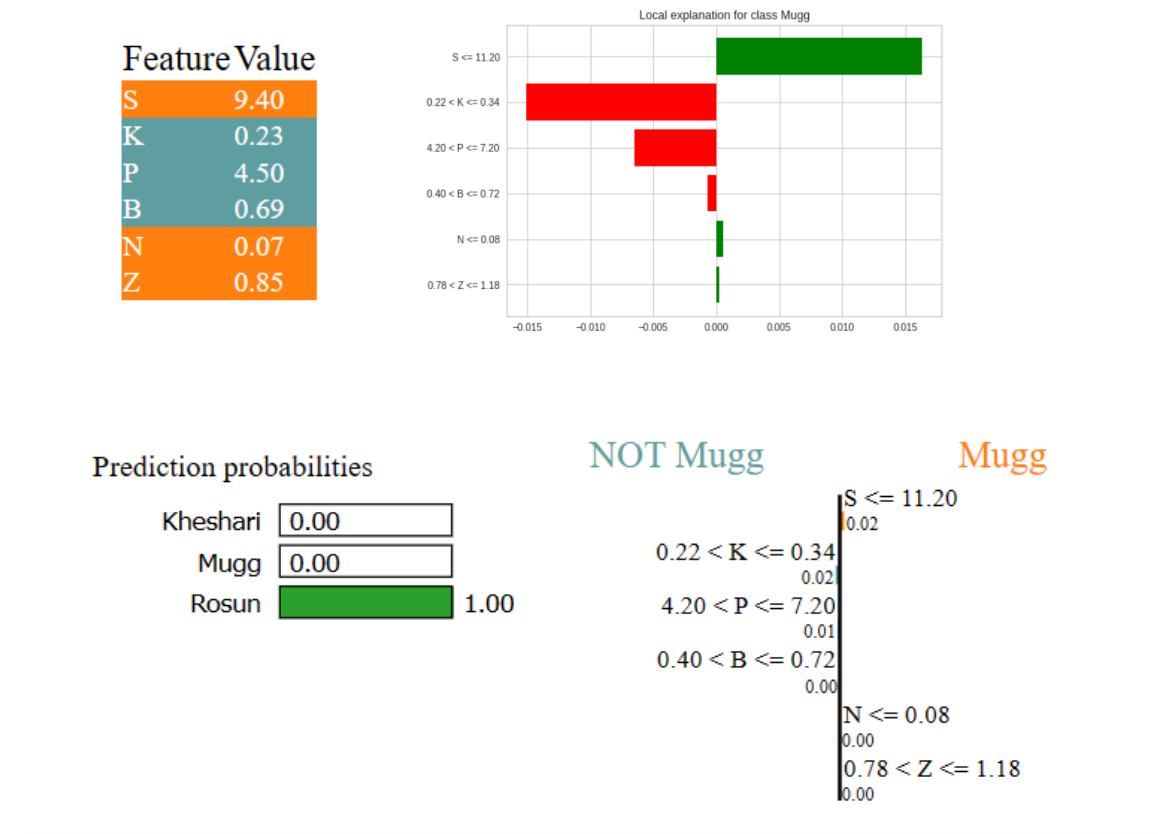


Figure 6.13: Lime Implementation

We considered a column where the label is Rosun and this is what we get after implementing lime. The diagrams show the contribution of the features (N,P,K,S,Zn,B) in determining the ideal crop (Rosun). The contribution of Sulfur(S) is the highest, whereas the Nitrogen has the lowest.

Chapter 7

Future work and Conclusion

Bangladesh ordinarily depends on farmers to manufacture the vegetation since the bulk of the meals is produced here. As a growing country, we are dropping our arable lands daily, due to which we need to apply our arable lands to use its potential. New ways to increase efficiency with a constant decline of arable lands are a necessity in the current state of the country. And that is where Artificial Intelligence will prove to be of immeasurable assistance. By measuring N,P,K and S,Zn,B values and using extraordinary fashions, we can construct a system that can endorse the ideal crop to provide in a selected land. So we collected dataset from Soil Resource Development Institute, Bangladesh and run some machine learning models and implemented Explainable-AI. The models are Random forest, Decision Tree, Naive Bayes and Ensembled them in a model. This models gave us 98%, 98%, 89% and 99% accuracy accordingly. There is a huge scope to further improve this by collecting a big data and run analysis on that data to find the suitable crop for any land. Also a performing software can be made which will calculate the amount of crops will be produced in a certain land, loss chances because of rain or other disaster. Also the financial position of a farmer is important and the software can recommend a crop which will bring profit to the farmer. Thus, the farmers will be able to select suitable crops for their lands, maximizing the overall efficiency and production of crops. Therefore, the solution plays a vital role in accelerating the state of the current agricultural scenario of Bangladesh from an economic standpoint. As a result, the system will benefit the farmers and the government.

Chapter 8

References

- [1] Kumar, V., Hiremath, D., and Chaudhary, S. (2022). An Architecture for a Quality Centric Crop Production System. In *Data Science in Agriculture and Natural Resource Management* (pp. 127-141). Springer, Singapore.
- [2] Swaminathan, B., Palani, S., Kotecha, K., Kumar, V., and Subramaniaswamy, V. (2022). IoT Driven Artificial Intelligence Technique for Fertiliser Recommendation Model. *IEEE Consumer Electronics Magazine*.
- [3] Bhoi, A., Nayak, R. P., Bhoi, S. K., Sethi, S., Panda, S. K., Sahoo, K. S., and Nayyar, A. (2021). IoT-IIRS: Internet of Things based intelligent-irrigation recommendation system using machine learning approach for efficient water usage. *PeerJ Computer Science*, 7, e578.
- [4] Singh, R., Srivastava, S., and Mishra, R. (2020, February). AI and IoT Based Monitoring System for Increasing the Yield in Crop Production. In *2020 International Conference on Electrical and Electronics Engineering (ICE3)* (pp. 301-305). IEEE.
- [5] Nischitha, K., Dhanush Vishwakarma, M. N., and Ashwini, M. M. (2020). Crop prediction using machine learning approaches. *Int. J. Eng. Res. Technol.(IJERT)*, 23-26.
- [6] Rajak, R. K., Pawar, A., Pendke, M., Shinde, P., Rathod, S., and Devare, A. (2017). Crop recommendation system to maximize crop yield using machine learning technique. *International Research Journal of Engineering and Technology*, 4(12), 950-953.
- [7] G. Chauhan and A. Chaudhary, "Crop Recommendation System using Machine Learning Algorithms," 2021 10th International Conference on System Modeling and Advancement in Research Trends (SMART), 2021, pp. 109-112, doi: 10.1109/SMART52563.2021.9676210.
- [8] N. H. Kulkarni, G. N. Srinivasan, B. M. Sagar and N. K. Cauvery, "Improving Crop Productivity Through A Crop Recommendation System Using Ensembling Technique," 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS), 2018, pp. 114-119, doi: 10.1109/CSITSS.2018.8768790.
- [9] P. A, S. Chakraborty, A. Kumar and O. R. Pooniwala, "Intelligent Crop Recommendation System using Machine Learning," 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), 2021, pp. 843-848, doi: 10.1109/ICCMC51019.2021.9418375.

- [10] Yiannis Ampatzidis, Victor Partel, Lucas Costa, Agrovie: Cloud-based application to process, analyze and visualize UAV-collected data for precision agriculture applications utilizing artificial intelligence, *Computers and Electronics in Agriculture*, Volume 174, 2020, 105457, ISSN 0168-1699
- [11] Ivan Beloev, et al. "Artificial Intelligence-driven Autonomous Robot for Precision Agriculture." *Acta technologica agriculturae*, v. 24 ,.1 pp. 48-54. doi: 10.2478/ata-2021-0008
- [12] Vincent DR, Deepa N, Elavarasan D, Srinivasan K, Chauhdary SH, Iwendi C. Sensors Driven AI-Based Agriculture Recommendation Model for Assessing Land Suitability. *Sensors (Basel)*. 2019;19(17):3667. Published 2019 Aug 23.
- [13] D. Modi, A. V. Sutagundar, V. Yalavigi and A. Aravatagimath, "Crop Recommendation Using Machine Learning Algorithm," 2021 5th International Conference on Information Systems and Computer Networks (ISCON), 2021, pp. 1-5, doi: 10.1109/ISCON52037.2021.9702392.
- [14] S. Vaishnavi., M. Shobana., R. Sabitha. and S. Karthik., "Agricultural Crop Recommendations based on Productivity and Season," 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), 2021, pp. 883-886, doi: 10.1109/ICACCS51430.2021.9441736.
- [15] Pudumalar, S., Ramanujam, E., Rajashree, R.H., Kavya, C., Kiruthika, T. and Nisha, J., 2017, January. Crop recommendation system for precision agriculture. In 2016 Eighth International Conference on Advanced Computing (ICoAC) (pp. 32-36). IEEE.
- [16] Chougule, A., Jha, V.K. and Mukhopadhyay, D., 2019. Crop suitability and fertilizers recommendation using data mining techniques. In *Progress in Advanced Computing and Intelligent Engineering* (pp. 205-213). Springer, Singapore.
- [17] Alshanbari, R., 2021. Artificial-Intelligence-Enabled Robotic Navigation Using Crop Row Detection Based Multi-Sensory Plant Monitoring System Deployment (Doctoral dissertation).
- [18] Talaviya, T., Shah, D., Patel, N., Yagnik, H. and Shah, M., 2020. Implementation of artificial intelligence in agriculture for optimisation of irrigation and application of pesticides and herbicides. *Artificial Intelligence in Agriculture*, 4, pp.58-73.
- [19] Alreshidi, E., 2019. Smart sustainable agriculture (SSA) solution underpinned by internet of things (IoT) and artificial intelligence (AI). arXiv preprint arXiv:1906.03106.
- [20] Jung, J., Maeda, M., Chang, A., Bhandari, M., Ashapure, A. and Landivar-Bowles, J., 2021. The potential of remote sensing and artificial intelligence as tools to improve the resilience of agriculture production systems. *Current Opinion in Biotechnology*, 70, pp.15-22.
- [21] Aitkenhead, M. J., Dalgetty, I. A., Mullins, McDonald, A. J. S., & Strachan, N. J. C. (2003). Weed and crop discrimination using image analysis and artificial intelligence methods. *Computers and electronics in Agriculture*, 39(3), 157-171
- [22] Fenu, G. and Mallocci, F.M., 2020, June. Artificial intelligence technique in crop disease forecasting: a case study on potato late blight prediction. In *International Conference on Intelligent Decision Technologies* (pp. 79-89). Springer, Singapore.
- [23] Dharani, M. K., Thamilselvan, R., Natesan, P., Kalaivaani, P. C. D., & San-

thoshkumar, S. (2021, February). Review on Crop Prediction Using Deep Learning Techniques. In Journal of Physics: Conference Series (Vol. 1767, No. 1, p. 012026). IOP Publishing.

[24] Suchithra, M.S. and Pai, M.L., 2020. Improving the prediction accuracy of soil nutrient classification by optimizing extreme learning machine parameters. Information processing in Agriculture, 7(1), pp.72-82.

[25-32] <https://oldweb.lged.gov.bd/ViewMap.aspx>

[33] <https://blog.mindmanager.com/decision-tree-diagrams/>

[34] P. A, S. Chakraborty, A. Kumar and O. R. Pooniwala, "Intelligent Crop Recommendation System using Machine Learning," 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2021, pp. 843-848, doi: 10.1109/ICCMC51019.2021.9418375.

[35] Automated Prediction of Good Dictionary EXamples (GDEX): A Comprehensive Experiment with Distant Supervision, Machine Learning, and Word Embedding-Based Deep Learning Techniques - Scientific Figure on ResearchGate. Available from: <https://www.researchgate.net/figure/Illustration-of-random-forest-trees-fig4-354354484> [accessed 17 Jan, 2023]