A Machine Learning and Web-based Autonomous Agricultural system for Crop and Fertilizer Recommendation and Yield prediction

BY

Silvia Kh Shuvessa
ID: 191-15-12261

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

Supervised By

Md. Sazzadur Ahamed
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised By

Md. Ferdouse Ahmed Foysal
Lecturer
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2023

APPROVAL

This Project titled "**A Machine Learning and Web-based Autonomous Agricultural system for Crop and Fertilizer Recommendation and Yield prediction**", submitted by Silvia Kh Shuvessa 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 4 January 2023.

BOARD OF EXAMINERS

Professor Dr. Touhid Bhuiyan — **Chairman**
Professor and Head
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Dr.Fizar Ahmed — **Internal Examiner**
Associate Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Tasmima Ferdaus Shuva — **Internal Examiner**
Assistant Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Dr. Md Sazzadur Rahman — **External Examiner**
Associate Professor

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Md**. Sazzadur Ahamed**, Assistant Professor, and 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.

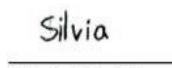
Supervised by:

Md. Sazzadur Ahamed
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised by:

Md. Ferdouse Ahmed Foysal
Lecturer
Department of CSE
Daffodil International University

Submitted by:



Silvia Kh Shuvessa
ID: -19-15-12261
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, I express my heartiest thanks and gratefulness to Almighty Allah for His divine blessing which makes me possible to complete the final year project/internship successfully.

I really grateful and wish my profound indebtedness to **Md**. **Sazzadur Ahamed, Assistant Professor,** Department of CSE, Daffodil International University, Dhaka, deep knowledge & keen interest of my supervisor in the field of Machine Learning to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this project.

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

Finally, I must acknowledge with due respect the constant support and patients of my parents.

ABSTRACT

In the world, agriculture is recognized as a crucial economic factor since it meets one of civilization's most basic necessities, namely food. In the majority of nations, it is recognized as the primary source of revenue. Bangladesh is one of many countries that still use agricultural practices. In FY 2020–21, agricultural products accounted for about 12.5% of Bangladesh's GDP, while 40% of the population worked in agriculture. Therefore, the lack of acreage and the depletion of natural resources make it even more urgent to achieve optimal precision agriculture technologies, such as crop selection, fertilizer suggestion, water management, crop yield prediction, etc. This web-based solution uses machine learning to forecast production while also recommending crops and fertilizers. The accuracy of the machine learning models is remarkable. The maximum accuracy in the dataset for both crop recommendation and yield prediction are 99%, and it is 98% for fertilizer recommendations. We developed a farmer-friendly system with a graphical user interface to help farmers solve agricultural difficulties. This technology is very crucial to the country and can improve the traditional agricultural system. Along with using this strategy, the farmer will be able to benefit even further from farming.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

1.1 Introduction

Agriculture is regarded as a key economic dimension in the world and provides for one of civilization's fundamental needs, primarily food. It is regarded as the main source of income in the majority of countries. Many nations, including Bangladesh, continue to practice agricultural methods. Approximately 12.5 percent of Bangladesh's GDP was formed from farm commodities in FY 2020–21, while 40 percent of the country's population was engaged in the sector. The majority of Bangladesh's agricultural production is classified as traditional subsistence farming. Bangladesh produces a wide range of agricultural goods, including wheat, rice, corn, legumes, flowers, fruits, vegetables, and others. The urgency of constructing innovative precision agriculture, such as crop selection, fertilizer recommendation, water management, crop yield prediction, etc., is increased by the lack of farmland and the depletion of natural resources. Besides Many factors affect agriculture, including soil types, weather, rainfall, humidity, temperature, fertilizer, etc. Farmers there are hesitant to adopt cutting-edge systems due to a major lack of knowledge, high costs, or an absence of awareness of the benefits of such technologies. Lack of understanding about soil types, yield, crop prediction, weather analysis, inappropriate pesticide use, irrigation issues, and incorrect harvesting resulted in the loss of farmers or increased costs. New issues arise as a result of a lack of information at each level of agriculture. An extremely high percentage of agricultural decisions, from crop selection through yield prediction are made erroneously altogether.

A major agricultural concern is predicting crop yield. It is among the more difficult chores in agriculture. It is crucial to decision-making at the worldwide, regional, and local levels. Agricultural, soil, climatic, ecological, and other characteristics are used to estimate crop yield. For making decisions regarding agricultural risk assessment and generating forecasts for the upcoming, it is vital to understand that agricultural productivity is mainly affected by weather conditions (rain, temperature, etc.), pesticides, and reliable information regarding past crop productivity. Successful agricultural production depends on wise crop selection. It is a need that must be fulfilled before beginning a farming endeavor. The crop to be cultivated can indeed be chosen even without a set location and farm site based mostly on soil and meteorological factors. Varied crops require different weather conditions and soil types. For example, carrots, potatoes, radishes, lettuce, tomatoes, zucchini, collard

greens, corn, watermelon, beans, asparagus, and cucumber prefer sandy soil, whereas wheat, sugarcane, cotton, jute, pulses, and oilseeds thrive in a loamy soil. Temperature, humidity, and rainfall are the key climatic variables for crop selection, while the most important soil components for crop selection are nitrogen, phosphorus, potassium, and PH. By adding the proper balance of nutrients to the soil, fertilizers have the important responsibility of increasing yield and ensuring healthy products. To obtain a great harvest, it is crucial to recommend the right fertilizer based on the characteristics of the soil. Using the information on soil elements and climate, we can forecast the dosage of fertilizer.

1.2 Problem Statement

Agriculture is an important problem in the current circumstances since agricultural lands are shrinking while the global population is growing. In addition, many nations around the world lack adequate agricultural knowledge. They are unable to anticipate yield, apply the appropriate fertilizer, or cultivate suitable crops. Consequently, agricultural output is not increasing. It is a challenge to meet the desire for food for this population as the world's population grows. Particularly in Bangladesh, where there is a large population and little farmland. Aside from that, the agricultural system is conventional. It is a significant issue for our nation. Considering this, we suggested an autonomous system that uses machine learning to forecast yield and suggest crops and fertilizers based on soil type and climate.

1.3 Research Objectives

- To boost agricultural production by applying machine learning to estimate crop yield, fertilizer recommendations, and crop selection.
- To develop an autonomous system for the prediction models.
- To improve the old agriculture system and make recommendations and suggestions to farmers for optimal yield.

1.4 Research Questions

- How might the agriculture system be improved using machine learning?
- How may machine learning be used to boost agricultural crop yield?

1.5 Report Layout

1. The research introduction, objectives, and major research questions are presented in Chapter 1.
2. A comprehensive analysis of the relevant literature is emphasized in Chapter 2.
3. Chapter 3 provides a complete explanation of the proposed methodology.
4. The result analysis and comparison with prior work are explained in Chapter 4.
5. Chapter 5 concludes the existing study and makes recommendations for future effort.

CHAPTER 2

Literature Review

One of civilization's core necessities, namely food, is met by agriculture, which is regarded as a major economic factor worldwide. In the majority of nations, it is recognized as the primary source of revenue. In this section, we'll review a selection of earlier research works on this topic.

One of the earliest studies created a specific web app to evaluate the effect of weather conditions on crops production in the designated Madhya Pradesh. This approach is defined by the collection of a soil database from farms, the provision of crop information by agriculture experts, and the accomplishment of factors such as soil through a soil testing lab dataset. The recommendation system will leverage the data from the soil testing lab dataset to build an ensemble model with majority voting, employing support vector machines and artificial neural networks (ANN) as learners, to make highly accurate and effective crop recommendations for site-specific parameters. [3] The crop-growing region was taken into consideration when choosing the districts. The five most important districts with the most amount of cropland was selected based on these factors. The crops have chosen for this study were based on the most common crops in the chosen districts. The crops chosen were maize, wheat, soybean, and paddy, and the yields for a continuous 20-year period of knowledge were tabulated for each. For the selected crops, the built model's accuracy ranged from 76% to 90%, with an average accuracy of 82%. Another significant piece of research evaluates the soil's quality and forecasts crop yields together with an appropriate fertilizer suggestion [4].

This model's inputs included the PH value and the user-provided location. To predict the weather and temperature for the present location, an API was used. The system compares the outcomes of supervised and unsupervised machine learning methods. A classifier that predicts agricultural yield using a greedy technique was put out. [5] Numerous ML algorithms used in the agriculture industry, and a lot of work has been done. Increasing

farm output and providing it to consumers at the best price and quality is the largest issue in agriculture. Additionally, it has been noted that at least 50% of farm food is lost and never consumed. The suggested approach offers suggestions for reducing farm produce waste.

A model that predicts crop yield using KNN algorithms and clustering is presented in one of the more recent papers. KNN clustering has been demonstrated to perform significantly better than SVM or regression [6]. Crop yield prediction system implementation involves supervised learning. The technique is able to boost crop productivity owing to the correlation that has been established between many historical attributes [7]. Rainfall and temperature were key elements that affect crop output. These time series data were subjected to Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) algorithms to improve accuracy [8]. This study uses Nave Bayes, Decision Trees, and a hybrid approach of Nave Bayes and Decision Trees to investigate the nutrients nitrogen, phosphorus, potassium, calcium, magnesium, sulfur, iron, zinc, and other soil supplements. Based on accuracy and execution time, the classification algorithms' performances are contrasted.[9] This paper proposes a method for predicting the category of yield based on the status of macronutrients and micronutrients in a dataset using various machine-learning approaches. The Krishi Bhawan (Talab-Tillo) Jammu dataset used for crop yield prediction was obtained. Macro-nutrients (ph, Oc, Ec, N, P, K, S) and Micro-nutrients (Zn, Fe, Mn, Cu) found in samples taken from various parts of the Jammu District are the parameters included in the data. Following examination, machine learning algorithms are used to forecast the yield category. The category, as foreseen, will specify the crop yield. Different classifier methods are utilized to formulate the crop yield prediction problem as a classification problem. [10]

CHAPTER 3
Materials and Methods

3.1 Working Process

The entire process is subdivided into five steps. These are enumerated below:

a) Data Collection and Data Analysis
b) Data processing
c) Data partitioning
d) Apply Machine Learning Algorithms
e) Analysis of prediction results
f) Implementation of the website

The entire working process is shown in figure 1 and is thoroughly explained in the sections as follows.

3.2 Data Collection and Data Analysis

The yield prediction, fertilizer recommendation, and crop recommendation datasets are all taken from KAGGLE [11]. There are 2200 items in the crop recommendation dataset, 28243 entries in the yield prediction dataset, and 100 entries in the fertilizer recommendation dataset. Nitrogen, potassium, phosphorus, temperature, humidity, PH, rainfall, and crops are the data columns for crop recommendations. Rice, maize, chickpeas, black gram, lentil, pomegranates, bananas, mango, grapes, watermelon, muskmelon, apple, orange, papaya, coconut, cotton, jute, and coffee are among the crops. The columns for Yield Prediction are Area (country), Crops, Yield, Rainfall, Pesticides, and Temperature. The crops include maize, potato, rice, paddy, sorghum, soybeans, wheat, sweet potato, cassava, yams, and plantain leaf. The columns for fertilizer recommendations include temperature, humidity, soil type, crops, nitrogen, potassium, and phosphorus, Fertilizer. Sandy, red, black, clayey, and loamy is the different soil textures. Maize, sugarcane, cotton,

tobacco, paddy, barley, wheat, millet, oil seed, pulses, and ground nuts are among the crops.

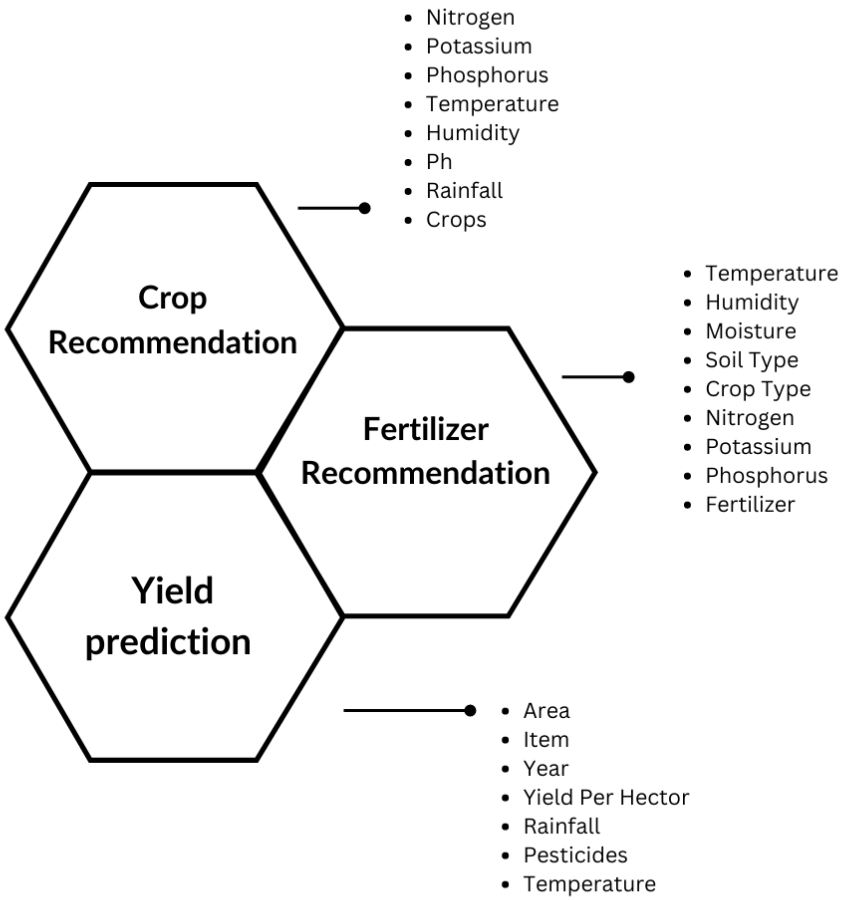


Figure 1: Datasets

3.3 Data Preprocessing

For effectively training the algorithms, we initially preprocessed all datasets. As part of the filtering procedure, we eliminated the null data and redundant columns from the datasets. By using a label encoder to preprocess crop data from the crop recommendation dataset, we transform categorical variables into numerical variables. We also utilized a label encoder for the crop type, soil type, and fertilizer columns since these were categorical elements in the fertilizer recommendation dataset. We also utilized a label encoder for the crop type, soil type, and fertilizer columns since these were categorical fields in the fertilizer recommendation dataset. We encoded the Area using a label encoder for the yield prediction dataset. The "Area" and "Item" values are then modified to generate dummy or indicator variables. To normalize the range of independent variables or data characteristics, feature scaling was also performed successfully.

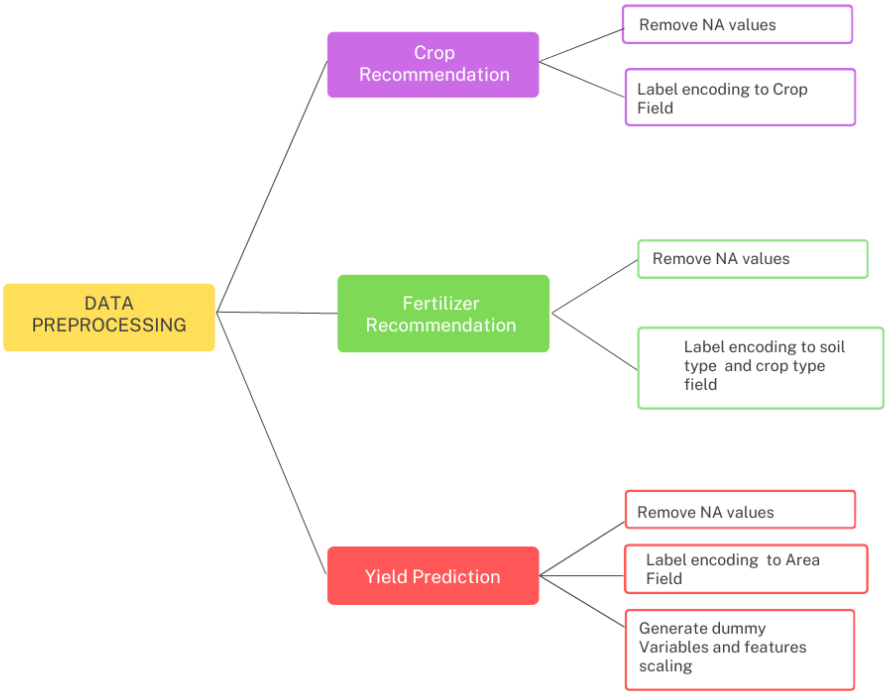


Figure 2: Data Preprocessing

3.4 Data partitioning

Data partitioning is frequently used in machine learning to avoid over fitting. In that situation, a machine learning model fits its training data too well and is unable to consistently fit new data. In this study, the datasets were divided into two sets. For crop recommendation, 80% of the data was used for training and 20% for testing. We used 70% of the data for training and 30% for testing for both the yield prediction dataset and the fertilizer recommendation dataset.

3.5 Machine learning algorithms

Linear regression

Linear regression is one of the most basic and popular Machine Learning algorithms. It is a technique for undertaking predictive analysis. For continuous/real/numeric variables like temperature, humidity, rainfall, and productivity, among others, predictions are created using linear regression. There are two types of linear regression; in this instance, multiple regression was used.

Gradient Boosting Regression:

Gradient boosting, a machine learning algorithm, is used, among other things, for problems involving regression and classification. It provides a prediction model in the form of decision tree aggregation, which is similar to inadequate estimating techniques. When a decision tree fails to learn, the resulting method, known as gradient-boosted trees, frequently beats random forest. A gradient-boosted trees model is built in the same stage-by-stage fashion as prior boosting algorithms, but it generalizes those techniques by allowing the optimization of any discrete weight vector.

Decision Tree:

The decision tree algorithm is a form of supervised learning strategy. The decision tree technique, unlike other supervised techniques, can handle both classification and

regression jobs. The purpose of employing a Decision Tree is to develop a training model that can be used to forecast the class or value of the target variable by learning fundamental decision rules generated from time series (training data).

Random Forest:

In the training stage of the random forest approach, which is used for classification, regression, and other tasks, a massive percentage of decision trees are constructed. The result of the random forest is the class that most of the trees in a classification task selected. For regression tasks, the average prediction generated by a specific tree is returned. Random decision forests are appropriate for this since decision trees frequently over fit their training set. Random forests often outperform decision trees, but they are less precise than gradient-enhanced trees. However, data attributes might affect how effective they are.

KNN:

K-Nearest Neighbor is a simple ML technique based on the supervised learning method. The K-NN algorithm assumes that the new and existing cases are comparable, and it assigns the new instance to the category that is most similar to the existing categories. After all of the previous data has been stored, a new data point is categorized using the K-NN algorithm based on similarity. This indicates that by employing the K-NN approach, new data can be promptly and accurately classified. Although the K-NN method is most used for classification problems, it can also be used for regression. Because K-NN is a non-parametric approach, no assumptions about the underlying data are made. It is also known as a lazy learner algorithm since it keeps the training dataset instead of learning from it right away. When classifying data, it instead acts on the dataset. The KNN approach simply keeps the information during the training phase and then categorizes it into a category that is quite like the new data when it receives fresh data.

Gaussian Naive Bayes

A stochastic machine learning approach called Naive Bayes can be applied to a variety of categorization problems. Naive Bayes is frequently used for document classification, spam filtering, prediction, and other tasks. This method takes its name from Thomas Bayes' discoveries, on which it is based. The method combines features into its model that is independent of one another, hence the term "Nave." Any changes to the values of one parameter of the algorithm have no direct effect on the value of any other feature.

The Naive Bayes algorithm's key benefit is that it is an easy-to-use yet effective method. It is built on a probabilistic model, making predictions fast and in real time with an easily code-able algorithm. Since it can be adjusted to rapidly reply to user demands, this algorithm is typically used to solve problems in the real world. Before learning more about Nave Bayes and Gaussian Nave Bayes, however, we first understand what conditional probability is.

SVM

SVMs, or support vector machines, are widely used to solve classification issues. The term "Support Vector Regression" is used to describe these models (SVR). Because of SVR's flexibility, we may specify how much error in our model is acceptable, and it will find an appropriate line (or distance measure in higher dimensions) to fit the data. With the help of the potent SVR algorithm, we may adjust the overall tolerance for errors by defining an acceptable error margin and tuning the overall tolerance for exceeding that acceptable error rate.

3.6 Implementation of System

3.6.1 Layout Design

Razor

Visual Basic and C# server-based code can be embedded into web pages using the markup syntax known as Razor. While a web page is being delivered to the browser, server-based programming has the ability to generate dynamic web content instantly. Razor is an ASP.NET programming syntax that is used to create dynamic web pages using the C# or VB.NET programming languages. It is simpler to use and to learn, yet it has the same capability as conventional ASP.NET markup. This web application uses Razor to construct the web page's view.

Bootstrap

A front-end programming structure called Bootstrap is used to create open-source, free websites and web applications. Bootstrap is a collection of template design terminology that allows you to create a personalized website that prioritizes mobile consumers. Developers merely must place the code into a pre-defined grid system because Bootstrap includes the fundamentals for developing responsive websites as a framework. Cascading style sheets (CSS), JavaScript, and HTML are the foundational technologies of the Bootstrap system. Without having to spend time thinking about fundamental commands and functionalities, web developers can create websites considerably more quickly when utilizing Bootstrap.

HTML

Standard markup for documents intended to be viewed in a web browser is called HTML, or Hyper Text Markup Language. Technologies like Cascading Style Sheets (CSS) and scripting languages like JavaScript can help. HTML documents are downloaded from a web server or local storage by web browsers, who then turn them into multimedia web pages. HTML originally featured cues for the document's design and semantically explains the structure of a web page. The foundation of HTML pages is HTML components. Images and other objects, like interactive forms, may be embedded within the produced page using HTML techniques. By indicating structural semantics for text elements like headings, paragraphs, lists, links, and quotes, HTML offers a way to create structured documents.

CSS

The purpose of CSS is to make it possible to separate content from overview, including layout, colors, as well as fonts. By putting relevant CSS in a separate.css file, developers

can reduce complexity and redundancy in structure content, multiple web pages can share formatting, that can increase accessibility, flexibility, and control in the specification of presentation qualities, as well as page load speed amongst pages that share the file and its formatting.

3.6.2 User Interaction

ASP.NET

A web development platform called ASP.NET offers a programming model, a thorough software architecture, and a variety of services needed to create reliable web applications for PCs and mobile devices. Asp.net version 6 was used to construct the user interaction in this web application.

3.6.3 Deploy Machine Learning Models

ML.net

C#, one of the most widely used languages today, is employed in several applications. Microsoft developed a package named ML.NET that offers all the core machine learning features, to enable the usage of machine learning in C #. Python and R are the first two programming languages that come to mind when discussing data science and machine learning. These languages may be used for practically any machine learning problem and provide all widely used machine learning algorithms, preprocessing methods, and other features.

There are situations when the Developer is unable to use Python or R or does not want to. This could be for a variety of reasons, such as having existing code in another language or lacking Python or R familiarity. C#, one of the most widely used languages today, is employed in several applications. Microsoft developed a package called ML.NET, which

offers all the fundamental Machine Learning features, to enable the usage of Machine Learning in C#.

3.6.4 Overview of the Website

Homepage

The sign-in and crop prediction buttons are located on the homepage. Users must click both the sign-in and crop prediction buttons to sign up and use the crop prediction feature.

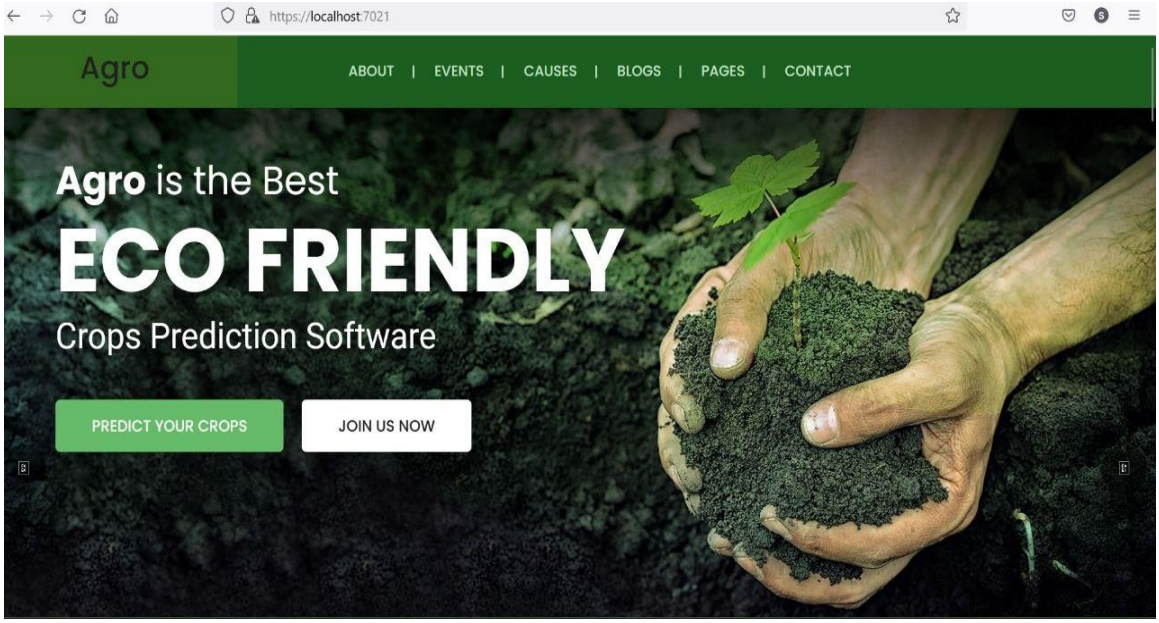


Figure 3: Homepage

Prediction Page

Prediction page has been designed for the main functional activity. Three Buttons are available on the prediction page. Crop, Fertilizer, and Yield. Click on Crop to get crop assistance, Fertilizer to get fertilizer recommendations, and Yield to get yield forecast.

There will be input areas when you make your choice. The user will move forward to acquire the prediction result after providing all the inputs.

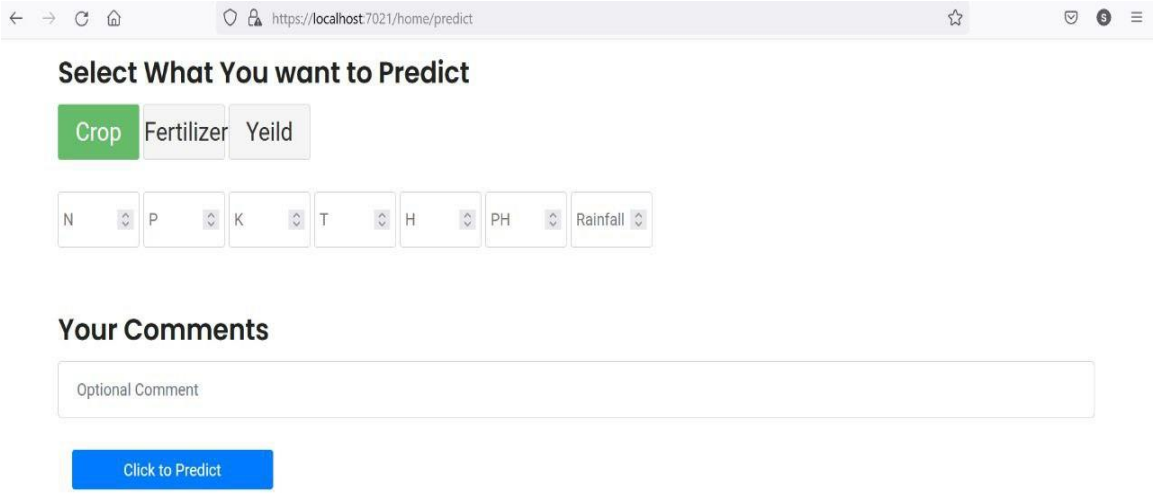


Figure 4: Prediction Page

Result Page

Prediction result page will show the prediction result and result summary. This is basically the system's output page.

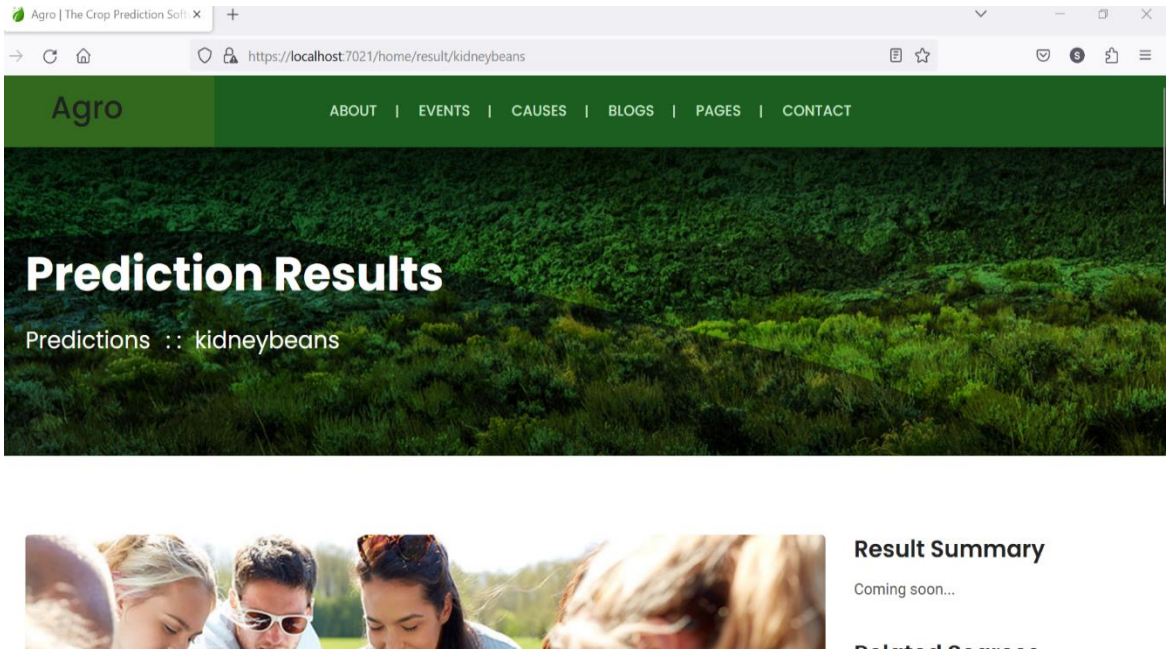


Figure 5: Result Page

Demo Prediction

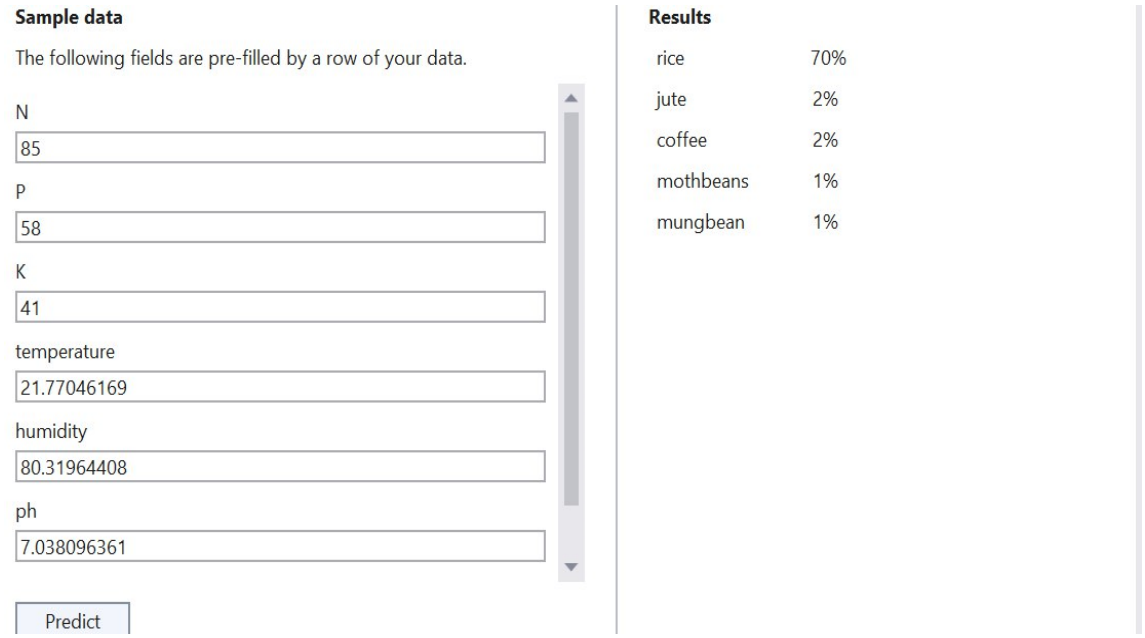


Figure 6: Demo Prediction of Crop Recommendation

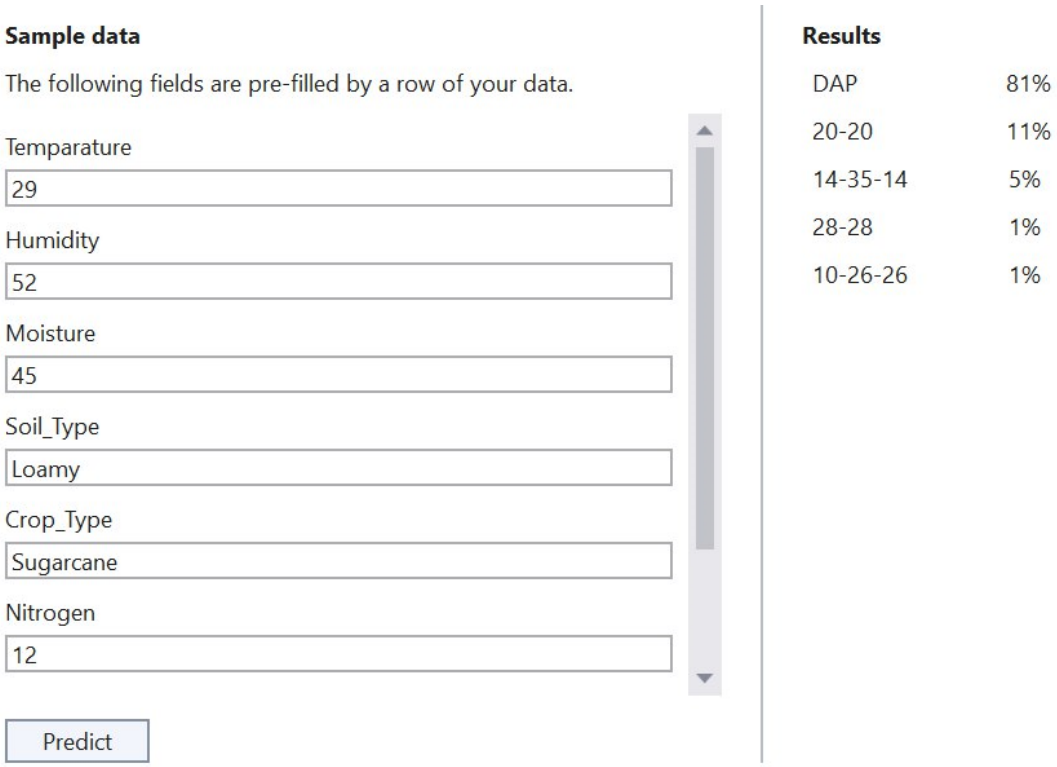


Figure 7: Demo Prediction of Fertilizer Recommendation

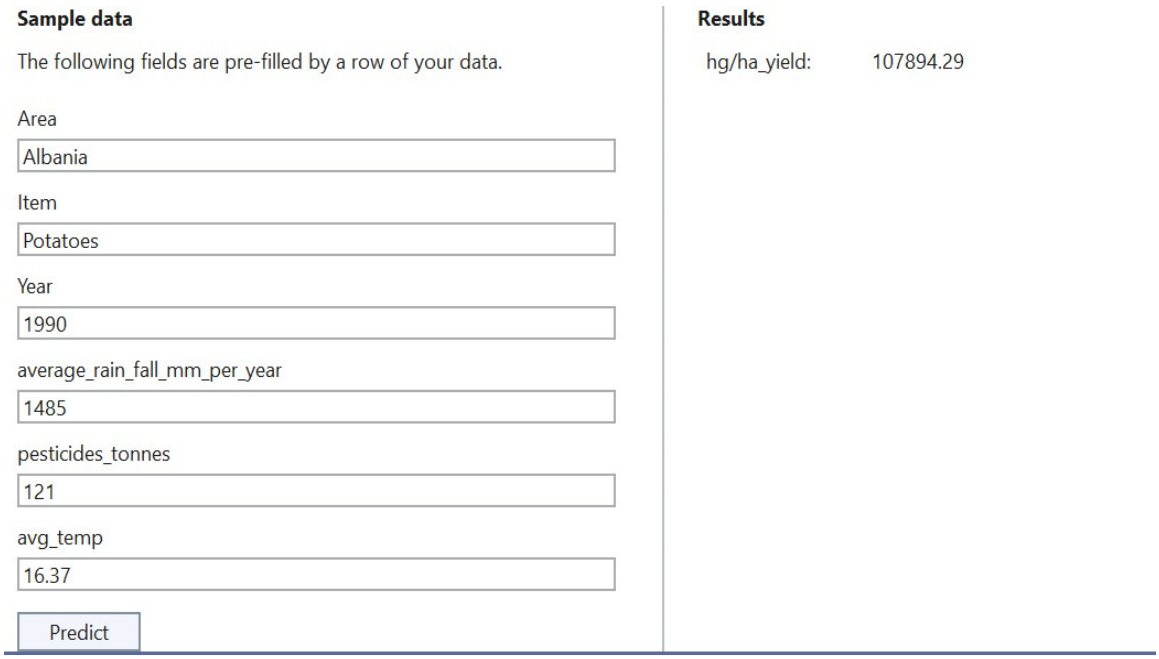


Figure 8: Demo Prediction of Yield prediction

CHAPTER 4

Results and Discussion

4.1 Prediction Accuracy

The crop prediction accuracy for the crop recommendation dataset is summarized in Figure 9.

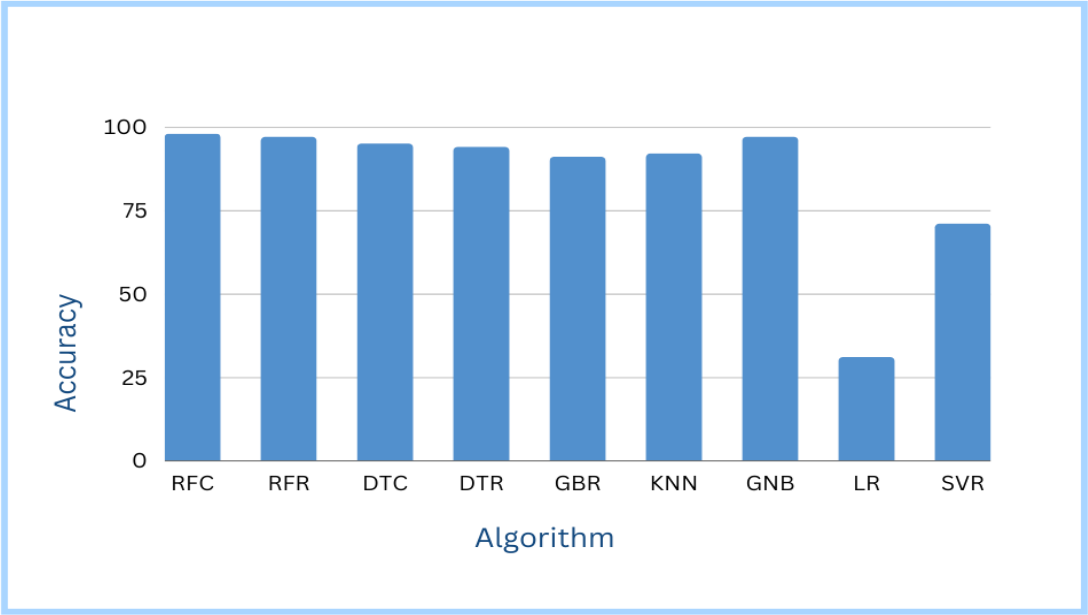


Figure 9: Accuracy of Crop Recommendation

The Random Forest classification method achieved the highest result, which is 98%, as can be shown. Both the Gaussian naive Bayes and the random forest regression have an accuracy of 97%. The results from DTC, DTR, GBR, and KNN are also satisfactory at 95%, 94%, 91%, and 92%. SVR reported an accuracy of 71%, while LR reported the lowest accuracy of 31%.

Figure 9 shows the fertilizer prediction accuracy for the fertilizer recommendation model.

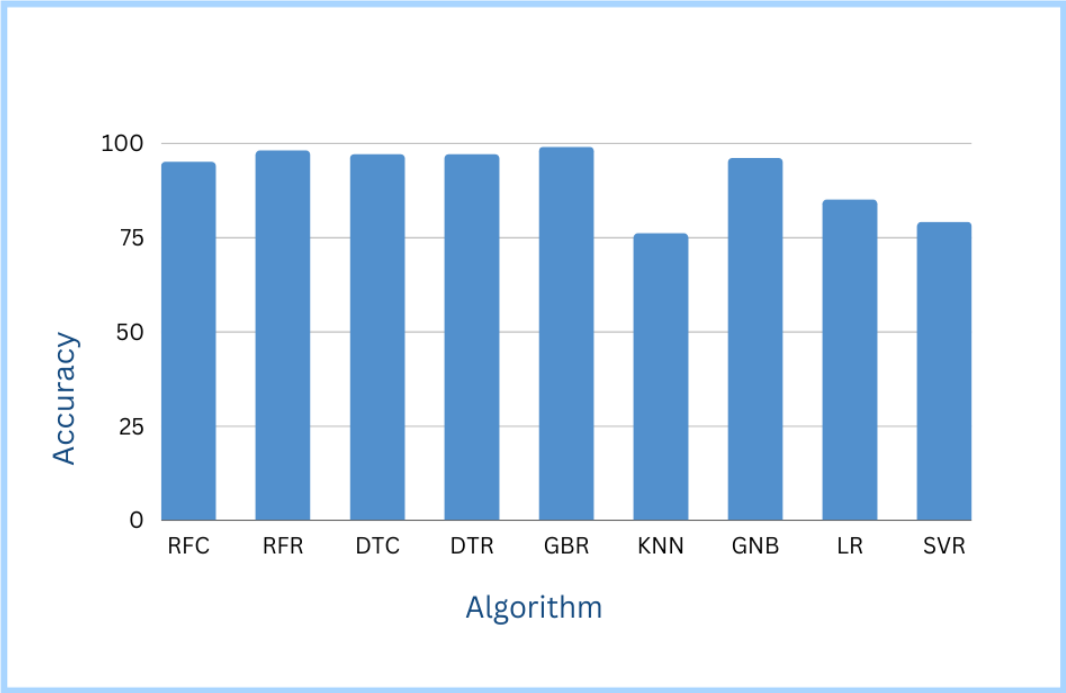


Figure 10: Accuracy of Fertilizer Recommendation

The fertilizer recommendation outcome for the fertilizer recommendation dataset is listed in Figure 10. We can observe that the highest result obtained by GBR is 99%. Both the DTC and the DTR provide accuracy ratings of 97%. The results from RFC, RFR, GNB, and LR are also decent at 95%, 98%, 96%, and 85% respectively. For this dataset, SVR has an accuracy rating of 79%, while KNN has the lowest accuracy rating of 76%.

The results of the yield prediction are shown in figure 11.

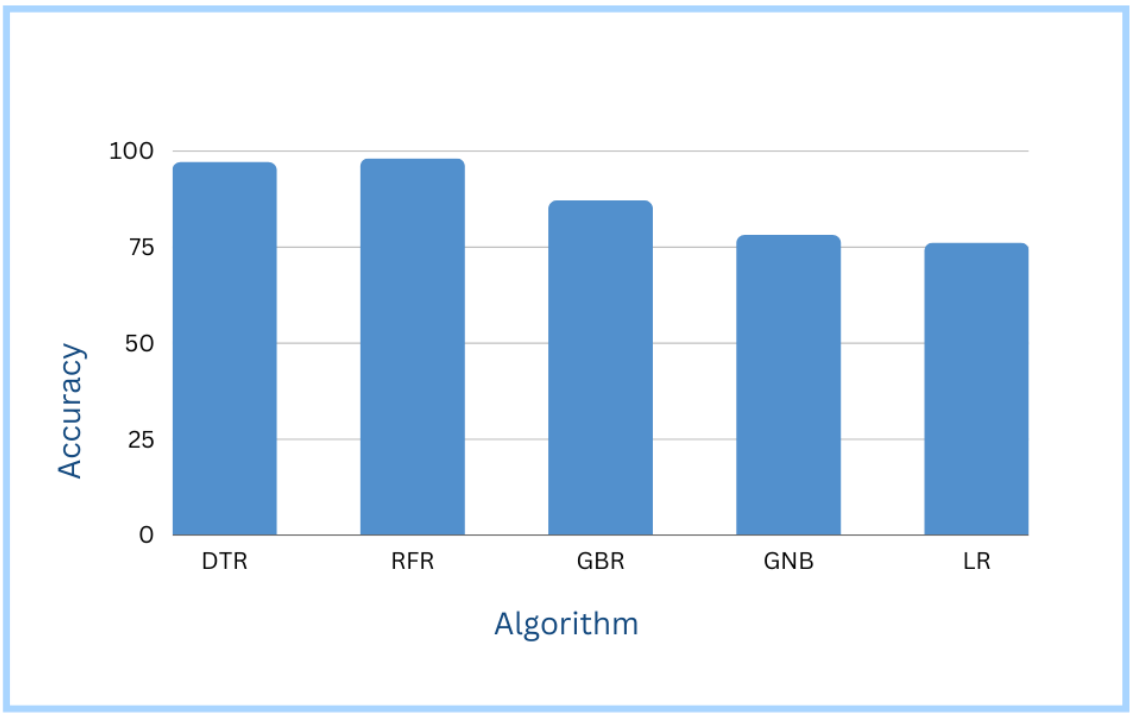


Figure 11: Accuracy of yield Prediction

RFR achieves the highest result of 98% for this dataset. GBR, GNB, and DTR all report accuracy rates of 97%, 87%, and 78% respectively. For this dataset, LR has the lowest accuracy at 76%.

4.2 Evaluation Metrics

The evaluation metrics used in regression analysis are mean squared error (MSE) and root mean squared error (RMSE). A statistical metric known as the coefficient of determination is the R square score. Always between 0 and 1, the r square score's range, a higher score indicates a better goodness of fit for the data points. Mean Squared Error

(MSE) for each model is shown in Table 1,2 and 3 below for Decision Tree, Random Forest, Gradient Boosting, KNN, GNB, Linear Regression, and SVR.

Table 1 shows a comparison of the crop prediction error score for the crop recommendation dataset.

TABLE 1: ERROR SCORRE OF CROP RECOMMENDATION

Algorithms	**MAE**	**MSE**	**RMSE**	**Accuracy**
RFC	0.054	0.654	0.809	0.98
RFR	0.384	1.594	1.262	0.97
DTC	0.086	0.95	0.974	0.95
DTR	0.3	3.718	1.92	0.94
GBR	1.298	4.064	2.015	0.91
KNN	0.168	1.93	1.38	0.92
GNB	0.081	0.98	.99	0.97
LR	4.159	27.50	5.24	0.31
SVR	2.148	10.066	3.17	0.71

A comparison of the Fertilizer prediction error score for the fertilizer recommendation dataset is presented in Table 2.

TABLE 2: ERROR SCORE OF FERTILIZER RECOMMENDATION

Algorithm	MAE	MSE	RMSE	R2 Score
RFC	0.25	0.45	0.67	0.95
RFR	0.06	0.018	0.134	0.98
DTC	0.35	0.65	0.836	0.97
DTR	0.30	0.67	0.806	0.97
GBR	0.33	0.60	0.77	0.99
KNN	0.55	0.35	1.16	0.76
GNB	0.4	0.8	0.89	0.96
LR	0.71	0.62	0.78	0.85
SVR	0.75	0.92	0.96	0.79

A comparison of the yield forecast error scores for each model can be found in Table 3.

TABLE 3: ERROR SCORE OF YIELD PREDICTION

Algorithms	MAE	MSE	RMSE	R2 Score
DTR	3920.05	161091114.90	12692.16	0.97
RFR	3718.35	97368158.8	9867.53	0.98
GBR	19599.244	938540076.17	30635.60	0.87
GNB	13208.68	1596069180.8	39950.83	0.78
LR	29338.02	1773035233.1	42107.42	0.76

4.3 Confusion Matrix

Analyzing the confusion matrix is a far more effective technique to assess a classifier's performance. It is a way to evaluate how well a machine learning classification algorithm performs when the output can include two or more classes. It is a table with various combinations of values that were predicted and actual. The values given diagonally in figures 1 and 2 represent the actual prediction. We can observe that the created models have provided accurate results that are satisfactory.

Figure 12 and 13 depicts the confusion matrix for each model (crop prediction, fertilizer recommendation). The mismatch between the actual label and the predicted label was displayed in each matrix.

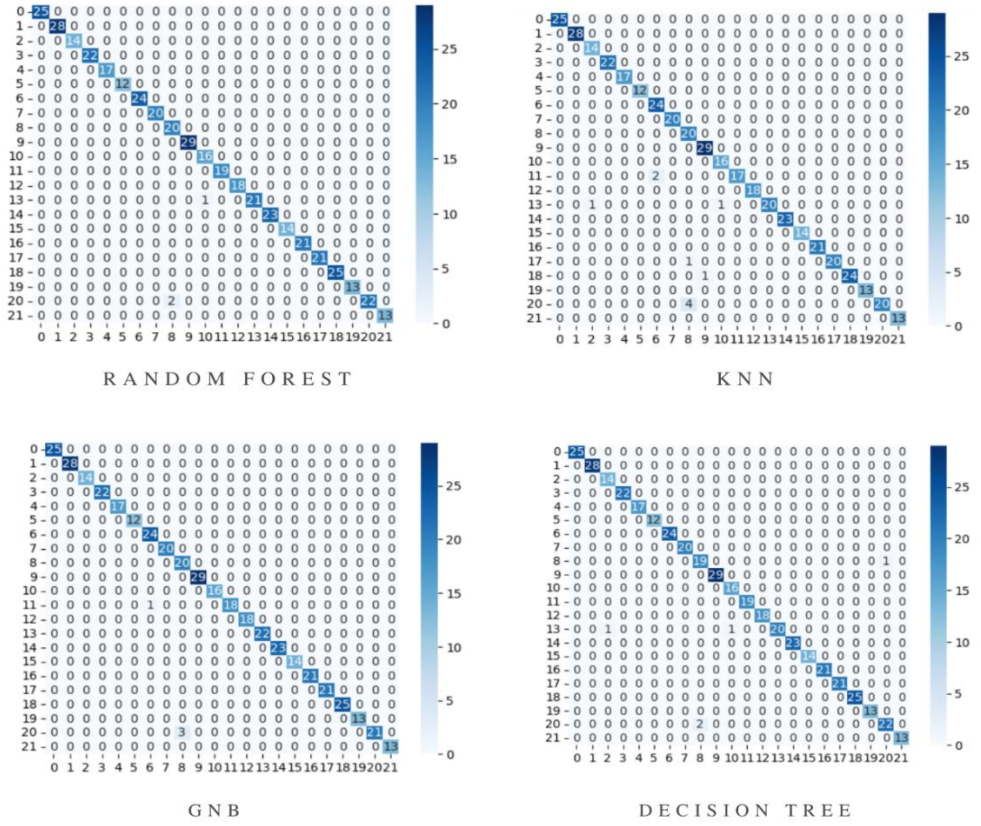


Figure 12: Confusion Matrix of Crop Recommendation

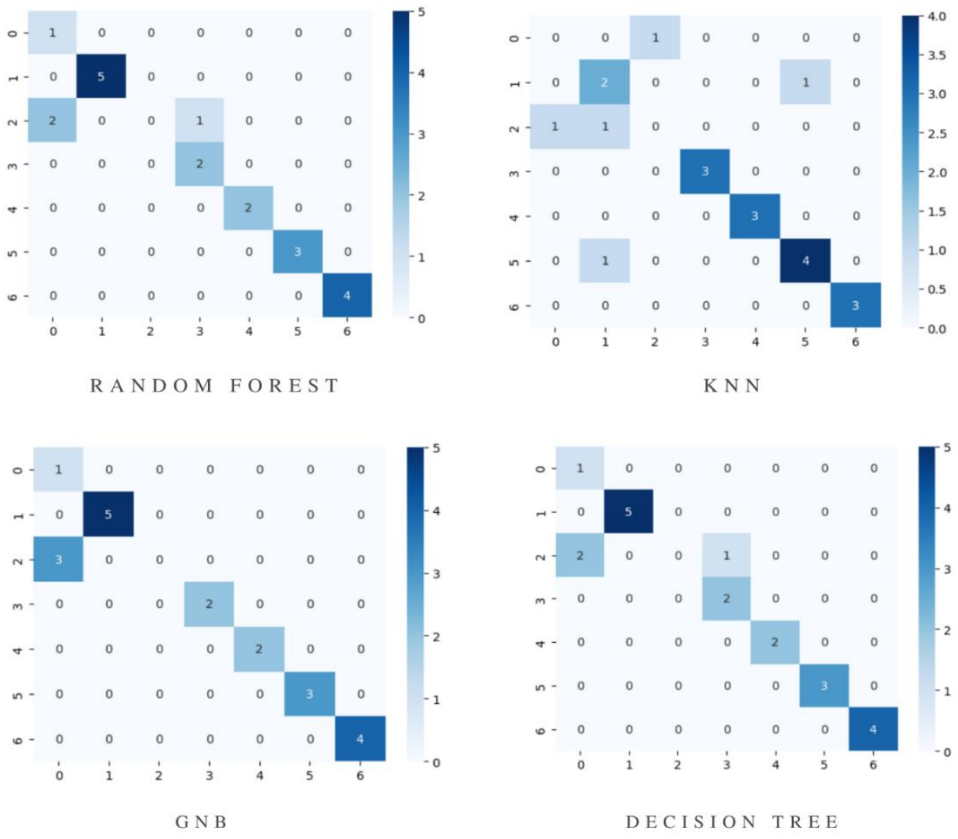


Figure 13: Confusion Matrix of Fertilizer Recommendation

CHAPTER 5

Conclusion and Future Work

5.1 Conclusion

Agriculture has long been important to our country, but there hasn't been much interplay across technologies and agriculture. Because our farmers are now not employing technology and analysis properly, there is a risk that they may choose the wrong crop, use the improper fertilizer, and anticipate the wrong yields for cultivation, which would lower their income. We created a farmer-friendly system with a GUI to anticipate which crop would be the greatest fit for a specific plot of land in order to prevent those types of losses. According to our research, crop advice, fertilizer recommendation, and accurate yield forecasts may be helpful to many farmers and other consumers. This approach will assist farmers in selecting the best crop for their plot of land and in applying the proper fertilizer to get the highest yield. We used machine learning techniques to create three models—a crop recommendation model, a fertilizer recommendation model, and a yield prediction model—in order to establish this system. The results from all three models are satisfactory. The random forest regression has a 98% accuracy rate when recommending crops. The gradient boosting regression has a 99% accuracy rate for fertilizer advice. The random forest regression has the highest accuracy (98%) when it comes to yield prediction. In addition, we developed a web-based application to provide a useful platform and user-friendly system for farmers. The simple and developer-friendly frameworks asp.net core and ml.net core was used to create this web application. These three recommendations and predictions are available in this application. The right input will be provided by the user, who will then receive their estimation. We may therefore conclude that this method is highly significant to the country and can enhance the conventional agriculture system. The farmer will be able to gain extra benefits from cultivation in addition to adopting this approach.

5.2 Limitations and Future Work

This system is constrained in some ways. The dataset for fertilizer recommendations is very limited. The yield prediction model has produced satisfactory results for a few

algorithms. The yield prediction model has produced satisfactory results for a few algorithms.

We can expand the dataset with new crops and update it with more significant attributes. Additionally, we can add additional estimation options to the web application, like disease detection and pesticide advice.

References

1. Agriculture in Bangladesh (2022) Wikipedia. Wikimedia Foundation. Available at: https://en.wikipedia.org/wiki/Agriculture_in_Bangladesh.
2. Bangladesh - agriculture sectors, International Trade Administration | Trade.gov. Available at: https://www.trade.gov/country-commercial-guides/bangladesh-agriculture-sectors.
3. Veenadhari S, Misra B, Singh CD, “Machine learning approach for forecasting crop yield based on climatic parameters”, In 2014 International Conference on Computer Communication and Informatics, 2014 Jan 3 (pp. 1-5). IEEE.
4. Ghadge R, Kulkarni J, More P, Nene S, Priya RL, “Prediction of crop yield using machine learning”, Int. Res. J. Eng. Technol. (IRJET), 2018 Feb;
5. Priya P, Muthaiah U, Balamurugan M, “Predicting yield of the crop using machine learning algorithm”, International Journal of Engineering Sciences & Research Technology, 2018 Apr;7(1):1-7.
6. S. Pavani, Augusta Sophy Beulet P., “Heuristic Prediction of Crop Yield Using Machine Learning Technique”, International Journal of Engineering and Advanced Technology (IJEAT), December 2019, pp (135-138)
7. Kumar, Y. Jeevan Nagendra, 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), pp. 736-741. IEEE, 2020.
8. Nigam, Aruvansh, Saksham Garg, Archit Agrawal, and Parul Agrawal. "Crop yield prediction using machine learning algorithms." In 2019 Fifth International Conference on Image Information Processing (ICIIP), pp. 125-130. IEEE, 2019
9. Rohit Kumar Rajak, Ankit Pawar, Mitalee Pendke, Pooja Shinde, Suresh Rathod, and Avinash Devare (2017). Crop Recommendation System to maximize Crop yield using Machine Learning. Issue 12 IRJET.
10. E. Manjula, and S. Djodiltachoumy (2017). Data Mining Technique to Analyze Soil Nutrients based on Hybrid Classification. IJARCS.
11. Your machine learning and Data Science Community (no date) Kaggle. Available at: https://www.kaggle.com/ (Accessed: January 28, 2023).

Defence_final_report.pdf

ORIGINALITY REPORT

23%	20%	11%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	12%
2	www.jatit.org Internet Source	1%
3	roi4cio.com Internet Source	1%
4	www.ijirset.com Internet Source	1%
5	S Iniyan, V Akhil Varma, Ch Teja Naidu. "Crop yield prediction using machine learning techniques", Advances in Engineering Software, 2023 Publication	1%
6	Kajal P Parmar, Tejas Bhatt. "Crop Yield Prediction based on Feature Selection and Machine Learners: A Review", 2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS), 2022 Publication	1%