A PROJECT REPORT

on

"AgroClimate Driven Fertilizer Recommendation"

Submitted to

KIIT Deemed to be University

In Partial Fulfilment of the Requirement for the Award of

BACHELOR'S DEGREE IN
INFORMATION TECHNOLOGY

BY

Dona Choudhury	2206022
Adarsh Kumar	2206151
Eshita Yadav	2206340
Gulshan Kumar	2206344
Manish Kumar	2206353
Shreyasi Saha	2206381

UNDER THE GUIDANCE OF
Dr. Supriyo Madal

SCHOOL OF COMPUTER ENGINEERING

KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY

BHUBANESWAR, ODISHA - 751024

April 2025

A PROJECT REPORT

on

"AgroClimate Driven Fertilizer Recommendation"

Submitted to

KIIT Deemed to be University

In Partial Fulfilment of the Requirement for the Award of

BACHELOR'S DEGREE IN
INFORMATION TECHNOLOGY

BY

Dona Choudhury	2206022
Adarsh Kumar	2206151
Eshita Yadav	2206340
Gulshan Kumar	2206344
Manish Kumar	2206353
Shreyasi Saha	2206381

UNDER THE GUIDANCE OF
Dr. Supriyo Madal

SCHOOL OF COMPUTER ENGINEERING

KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY

BHUBANESWAR, ODISHA - 751024

April 2025

KIIT Deemed to be University
School of Computer Engineering
Bhubaneswar, ODISHA 751024

CERTIFICATE

This is certify that the project entitled

“AgroClimate Driven Fertilizer Recommendation”

Submitted by

Dona Choudhury	2206022
Adarsh Kumar	2206151
Eshita Yadav	2206340
Gulshan Kumar	2206344
Manish Kumar	2206353
Shreyasi Saha	2206381

is a record of Bonafide work carried out by them, in the partial fulfilment of the requirement for the award of Degree of Bachelor of Engineering (Information Technology) at KIIT Deemed to be university, Bhubaneswar. This work is done during year 2024 2025 under our guidance.

Date: 08/04/2025

(Dr. Supriyo Mandal)
Project Guide

Acknowledgements

We are profoundly grateful to **Dr. Supriyo Mandal** of **Assistant Professor II** for his expert guidance and continuous encouragement throughout to see that this project rights its target since its commencement to its completion.

Dona Choudhury
Adarsh Kumar
Eshita Yadav
Gulshan Kumar
Manish Kumar
Shreyasi Saha

ABSTRACT

Sustainable agriculture depends on the application of fertiliser in a manner that is timely and efficient, since the rainfall and other agroclimatic factors will determine nutrient availability and loss. Moderate rainfall aids in fertilizer dissolution and its movement into the root zones, while excessive rainfall takes away the nutrients in solution from the soil through leaching, thus degrading the soil structure and contaminating the environment.

On the other hand, the study proposes a complex Random Forest algorithm that, using time-series data on temperature, humidity, and rainfall, suggests the timely administration of N, P, and K. The influence of micronutrients like Mn and B is also taken into account in order to maximise crop development and maintain soil fertility in balance.

This technique offers an opportunity for farmers to optimize fertilizer use with respect to rainfall patterns, yield optimization with least nutrient losses therefore enhanced environmental management. This data-driven approach helps improve soil fertility and lessens nitrogen contamination of water bodies thereby setting up an environmental concern. Providing targeted recommendations and thus able to support decision-making for farmers, the system guarantees that the fertilizers will be used as efficiently as the climate allows.

Keywords: Sustainable agriculture, fertilizer efficiency, Random Forest algorithm, agroclimatic factors, nutrient leaching, soil fertility optimization**.**

Contents

List of Figures

Chapter 1

Introduction

Agriculture produces food for all and supports the survival of millions of people to create worldwide food supply. The current farming system faces major hurdles because farmers misuse and waste fertilizers which harms the soil quality and pollutes water bodies while affecting environmental health. High agriculture requirements demand enhancing fertilization processes that boost production without hurting the environment. Farmers adopt typical fertilization methods from standard plans that neglect real-time environmental factors and plant needs which creates a waste of plant nutrients and adds unnecessary farming expenses. Our research develops an intelligent data-based method that helps farmers protect the environment during fertilization.

Need for the Project

Farmers mainly base their fertilizer usage on past learnings or basic guidelines. The common fertilizer guidance does not include exact soil types or changes related to weather and crop nutrient needs that vary between specific situations. Farmers end up using either too much or too little fertilizer when precision levels remain low. This (MPI) results in soil pollution and produces fewer harvests. The use of too many fertilizer products turns into nitrous oxide emissions that create planet-wide temperature increases. Our farming systems require precise agricultural automation to provide correct fertilizer application decisions at the ideal moment.

Farmers use modern AI technology to design and release new farming assistance platforms faster. The weather data system connects with forecasting tools to help farmers use their fertilizer supplies wisely. The project develops a fertilizer method by connecting optimization and machine learning processes with soil examination to enhance farming results and keep harmony with nature.

Gaps in Existing Solutions

Current smart farming tools experience problems as they receive more widespread use even though farmers keep adopting new technology. The regular farming equipment cannot work with live weather data because it cannot adjust to new conditions. Modern systems that blend satellite data need professional equipment few smallholder farmers can afford or handle properly. Many people dislike using modern decision-support systems because they are hard to understand and that prevents farmers from using them on actual farms.
Present farming technology mainly focuses on fixed datasets but lacks systems that receive updated information regularly. Smarter systems should change fertilizer recommendations based on environment conditions since basic systems lack this learning ability. This project creates a new eco-fertilization system to address these technology and data problems.

Our system contains machine learning algorithms with strong data predictions that farmers can use with ease farmers and other agricultural stakeholders can easily use our web platform to access its services. Most available systems depend on fixed data sets instead of continuously updating models. Current fertilizer recommendations systems need to adapt to changing environmental conditions which many systems today do not provide. Our initiative builds an eco-fertilization model using machine learning algorithms with excellent prediction results available through a simple web tool for farmers and agribusiness participants.

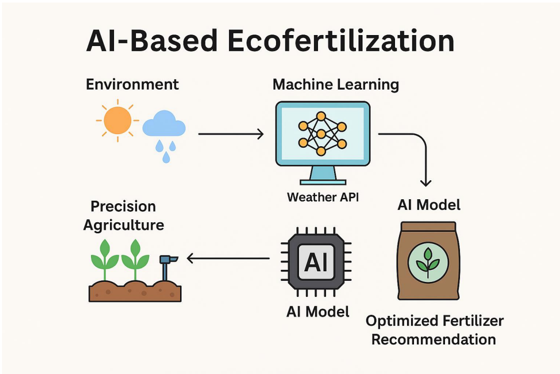

Figure 1.1: Basic model

Chapter 2

Basic Concepts

To improve land food production and environmental protection farmers use eco-fertilization methods that manage fertilizer amounts. The scientific methods of eco-fertilization prevent soil damage and chemical overuse because its practices include soil tests and organic fertilizer replacements. Actual crop conditions get tracked and managed using real-time data sensors as well as remote sensing technology under precision agriculture methods. Farmers can enhance their fertilizer usage with ML tools that deliver future fertilization forecasts. We study environmental signs of temperature, moisture and soil through Random Forest analysis to forecast fertilizer application requirements.

The system uses Decision Support System (DSS) software to take weather API and soil data straight from the source while suggesting exact fertilizer amounts. The online system lets farmers base their choices on data to prevent loss and work toward environmentally friendly practices. Heavy use of synthetic fertilizers causes serious damage to soils and pollutes water while adding more carbon to the environment. The system reduces environmental risk by using AI recommendations and organic choices to apply only required fertilizer at optimal times. The initiative uses computer systems to enhance farming practices making fertilization systems more effective and better for both our environment and Earth.

Literature Review

The central nature of sustainable agriculture establishes equilibrium between environmental safeguards and agricultural production prerequisites. The precise management of fertilizers serves as essential for both keeping soil health and achieving high agricultural production levels. Excessive fertilizer application and improper distribution methods generate environmental issues between water pollution from nutrient runoff and seepage into the ground and soil destruction. Researchers analyze data through agroclimatic elements such as rainfall, temperature, and humidity for enhancing fertilizer recommendations as part of their solution to these challenges.

The fertilizer application optimization together with nitrogen depression analysis in chili plants by using image processing methods. Analysis of leaf images for plant health assessment uses the histograms as a classification tool for healthy and deficient plants. An exact fertilizer recommendation is generated through area evaluation to help reduce unnecessary fertilizer usage while maximizing crop production. The study demonstrates vital importance of nitrogen for plant health while developing affordable farming technology for everyday farmer use. Subsequent developments will focus on developing the system for detecting various crop nutritional deficiencies using advanced machine learning approaches.[1]

By offering predictive analytics for disease detection, crop yield prediction, and fertilizer optimization, machine learning (ML) has completely changed agricultural decision-making (Liakos et al., 2018) [2]. To maximize the use of agricultural resources, machine learning algorithms like 2 decision trees, support vector machines, and neural networks have been used. Due to its ability to handle intricate, nonlinear relationships between numerous variables, the Random Forest algorithm has become well-liked among these (Gupta et al., 2020) [3].

Using different activation functions in Extreme Learning Machines enables this study to perform classification analyses on soil fertility indices and pH levels within Kerala's North Central Laterite region. Soil nutrient classification demonstrated the best outcomes using the Gaussian Radial Basis Function as its activation method yet the hyperbolic tangent function provided superior performance in pH classification. Soil sustainability along with optimal fertilizer use and environment protection results from this model implementation. The model aims for future development by expanding its functionality to various nutrients across different regions and by generating soil maps for enhanced fertilizer suggestions.[4]

A study examines how to use IoT together with data analytics and machine learning to forecast apple scab disease outbreaks inside Kashmir apple orchards. IoT sensors within a predictive framework acquire current information about environmental conditions from temperature and humidity measurements to detect disease outbreaks. Through this study IoT demonstrates its ability to boost crop health tracking and irrigation management and pest prevention operations. Various obstacles exist mainly from high implementation cost together with farmer ignorance about modern agricultural methods and technical system issues. Python-based IoT applications show promise to enhance both crop production numbers and minimize agricultural product wastage according to research findings.[5]

The analysis utilizes data mining methods to forecast future crop output of Indian major agricultural products through a historical data analysis spanning 1950–2013. The analysis of rainfall, temperature along with irrigation area, cultivated area, production and yield utilized Multiple Linear Regression (MLR), Random Forest Regression (RFR) and Multivariate Adaptive Regression Splines (Earth) for regression modeling. The prediction model that achieved the best results for Rice and Wheat analytics was Earth but the assessment of Maize yield proved most accurate through MLR. The research results demonstrate that agricultural precision forecasting can be achieved with data mining thus enabling better agricultural decisions.[6]

The paper analyzes how machine learning algorithms particularly Random Forest predict crop yields throughout India. Through the evaluation of rainfall together with temperature and soil conditions the model enables farmers to select appropriate crops and predict their yields. The research demonstrates the necessity of technological solutions for handling issues involving climate change combined with declining agricultural output. The implemented system makes better decisions possible through precise forecast predictions which allows farmers to enhance resource efficiency while boosting their yield amounts.[7]

The research presents an upgraded genetic algorithm (IGA) which recommends suitable nutrient quantities (Nitrogen, Phosphorus, Potassium) for agricultural cultivation through analysis of time-dependent sensor information. The model enhances soil fertility and crop output while fixing the problems caused by fertilizer usage deficiencies or excesses. The research demonstrates that using this system leads to yield improvement through effective nutrient management. Evolutionary computation demonstrates its capability to enhance sustainability and profitability in farming through precision agricultural approaches according to the research results.[8]

The authors examine how Controlled Traffic Farming (CTF) affects grain sorghum productivity together with nitrogen use efficiency (NUE) and rainfall use efficiency (RUE). CTF produces significant soil compaction decreases while improving water penetration and yielding better crop harvests than conventional systems without CTF. Results from APSIM simulations demonstrate that Controlled Traffic Farming boosts both RUE by 65% and NUE by 45% which results in improved profitability together with reduced environmental harm in rainfed agricultural systems. The study confirms the use of CTF contributes to agricultural practices that are environmentally friendly.[9]

The research analyzed N and P and K fertilizer requirements of irrigated rice throughout various Chinese regions through examination of 3,896 field data points. The study showed that N and K fertilization needs rise as temperature and solar radiation levels increase and rainfall amounts increase yet P needs decrease under such conditions until P fertilization rates increase. The northern territory of China needed less N and K but required higher P application rates compared to its southern regions because of dissimilar soil-climate patterns. The research objectives seek to develop better nutrient management strategies that support sustainable rice farming across China.[10]

The authors employed Kernel Ridge and Lasso and Elastic Net (ENet) regression techniques to forecast Indian crop yields using basic information about state location and district, season available region and crop category alongside farming area. The model prediction quality increased drastically through the combination of multiple prediction systems. The research platform enables easy yield prediction through a web application since it excludes sophisticated elements such as soil nutrients and climate information. The forthcoming developments for this system will concentrate on developing mobile applications using regional languages to reach more farmers across the country.[11]

A fuzzy logic-based decision support system was developed to optimize fertilizer use through improved crop productivity research spread across two agro-climatic zones during three years. The system combines analyzes of the soil with meteorological information and professional guidance to maximize outputs while reducing fertilizer usage. Sustainable agricultural practices received support from a system that raised productivity by 30 to 50% using lowered fertilizer amounts. The research demonstrates how fuzzy logic could advance farming precision yet proposes that handheld devices should be created for agricultural purposes.[12]

Researchers analyzed how excessive rainfall affects nitrate discharge rates between tilled and no-till farming operations in the U.S. Midwest region. The research findings indicated that increased rainfall rates resulted in higher nitrate leaching levels within tilled fields however no-till areas were not significantly affected because their unique soil structures allowed better macropore drainage. The practice of tilled soils led to greater nitrate losses however no-till management protected the soil by routing water beneath the nitrate. This research proves that no-till farming stands as an environmentally advantageous practice as climate patterns transform.[13]

Chapter 3

Problem Statement

Farmers encounter major difficulties in measuring their crops' nutrient needs due to the combined effects between environmental elements including temperature, rainfall, humidity and weather conditions. Soil fertility together with crop development are strongly affected by these environmental elements which hinders farmers from correctly measuring the amounts of Nitrogen (N), Phosphorus (P), and Potassium (K) needed. Accurate methods to evaluate nutrient requirements must be accessible to farmers since both under-feeding and over-feeding nutrients hamper yields along with damaging the future health of their land.

3.1 Project Planning

The procedure was followed with great care in developing a crop recommendation system. The initial stage required defining all guidelines and requirements for the project while selecting a suitable agricultural dataset. About the dataset we did profound assessments to evaluate its reliability for prediction tasks. To receive user data about crop type and location the system incorporated an interface. The system received real-time weather data about temperature and humidity and rainfall through Weatherbit API integration. Currently we use Python along with fundamental data science libraries to build the system where Logistic Regression serves with Decision Tree, Random Forest, KNN, and SVM as classification models that help achieve nutrient level predictions with accuracy. The evaluation process revealed Random Forest as the most effective model because it presented the best performance metrics among multiple classification models tested. The system featured two parts including a heavy rainfall warning system that also produced predictions about optimal Nitrogen Phosphorus and Potassium levels. A user-friendly notebook environment served as the platform for developing the whole system to maximize human interaction and future deployment needs. The development involved continuous testing in addition to feedback collection to guarantee system effectiveness and reliability in real agricultural conditions.

3.2 Project Analysis

The system underwent through a detailed analysis for risk and ambiguity identification after requirement acquisition and problem statement completion. A systematic evaluation analyzed how users provide crop and location information before the system retrieves weather data using the Weather API and proceeds to determine rainfall conditions. McMaster University & Saginaw Valley State University analyzed the Weatherbit API data reliability to ensure dependable and accurate results for nutrient predictions since small weather data variations would impact the prediction results. The process of evaluating heavy rainfall and warning protocol underwent thorough examination to eliminate both warning lag time and incorrect signals that cause farmer confusion. The predictive modeling process tested numerous algorithms until the adoption of Random Forest because of its high accuracy alongside robust performance. The application required strategies to address initial potential problems in user input and API data incompleteness or missing information. The analysis verified the project process had logical organization while demonstrating potential future growth and delivering dependable fertilizer suggestions that boost agricultural soil nutrition levels.

3.3 System Design

3.3.1 Design Constraints

Software Requirements:

- **Programming Language**: Python
- **Libraries**: NumPy, Pandas, Seaborn, Matplotlib
- **Machine Learning Tools**: Scikit-learn (StandartScaler, LabelEncoder)
- **Visualisation Tools**: Searborn, Matplotlib
- **Development Environment:** Jupyter Notebook, VSCode
- **Data Collection:** Kaggle, Weatherbit API

Hardware Requirements:

- **Processor:** Multi-core CPU / GPU
- **RAM:** Minimum 8 GB
- **Storage:** SSD recommended for faster data processing

- **Operating System:** Windows/ macOS/ Linux

Experimental Setup:

- **Data Sources:** Weather API for real-time temperature, humidity, and rainfall data
- **Preprocessing**: Data standardization, encoding, and imputation
- **Visualization**: Display of data trends and relationships using Seaborn and Matplotlib
- **Model Training:** Predictive modeling for soil nutrient (NPK) content prediction
- **Deployment:** Local machine for testing, scalable to cloud platforms (AWS, Google Cloud) for larger datasets

3.3.2 System Architecture/Block Diagram

The proposed system designs a fertilizer optimization process which relies on data-driven techniques for making agroclimatic condition-based fertilizer recommendations. The system architecture contains the components shown in Fig.

1) User Input:

• The system requires users to supply crop information together with location details as entry data.

2. Data Collection:

• The system obtains agroclimatic information including temperature and humidity as well as rainfall levels through its integration with a Weather API.

3. Heavy Rainfall Assessment:

• Rainfall evaluation by the system allows identification of heavy rainfall scenarios in the environment.

• Heavy rainfall alerts are displayed to users through the system when detection systems reveal such conditions.

4. Predictive Modeling:

• The predictive modeling system uses collected data to provide sustainable recommendations about N P and K fertilization amounts for farming purposes.

5. Output Generation:

• Predictive analysis determines appropriate fertilizer levels for which the system provides as output.

The block diagram outlines the steps which compose a crop recommendation and soil nutrient prediction system process. The system begins by asking users to enter crop type as well as location data. Temperature and humidity data and rainfall information obtain their values through a Weather API connection. When assessing heavy rainfall situations, the system produces warnings which users must review. Predicative modeling techniques enable the system to examine collected data for soil nutrient forecasting of nitrogen, phosphorus and potassium content which supports farmer decision-making.

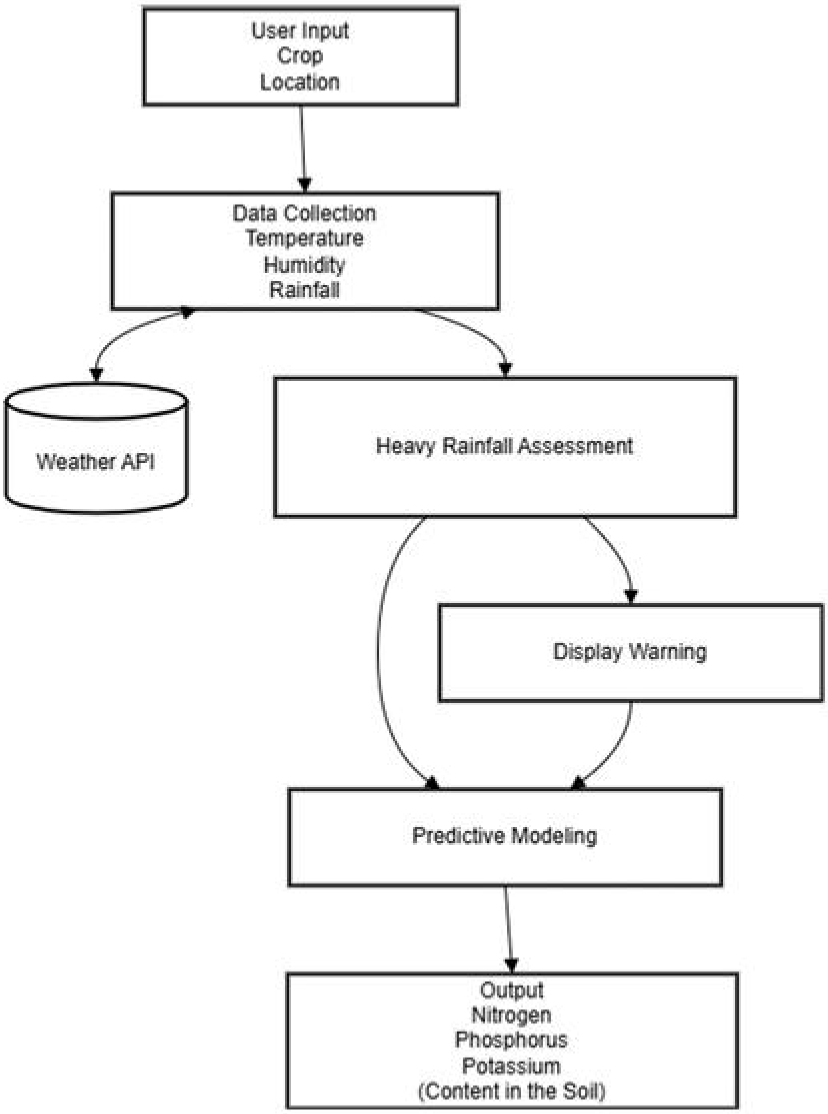

Fig 3.1: Block diagram of our project

Chapter 4

Implementation

The predictions coupled with weather-based advisory system development process uses specified software methodologies that empower agricultural choices by incorporating data analytics. A major objective of this system exists to generate accurate assessments for agricultural experts and farmers about their soil nutrients (Nitrogen, Phosphorus, Potassium) and weather conditions which influence crop health. Python's extensible framework allows the system to use several reliable libraries and frameworks that handle data preparation as well as predictive modeling and visualization tasks. Real-time weather API data collection enables the system to get environmental information about temperature and humidity and rainfall which makes it adaptive to shifting climatic conditions. By issuing tailored warnings in a timely manner the heavy rainfall assessment module helps businesses reduce the threats that come from heavy rains. The predictive models which leverage Scikit-learn and optional TensorFlow use these collected data to establish precise soil nutrient content estimations. The system enables running on local machines and cloud platforms like AWS and Google Cloud which delivers high flexibility to handle testing needs from small projects and deployment requirements at every scale. The visualization tools Matplotlib and Seaborn alongside each other allow users to view data trends clearly thus supporting better and easier decision-making processes. Specified integrated infrastructure components work as one system to build a usable solution which helps improve soil management while ensuring sustainable agricultural practices.

4.1 Methodology OR Proposal

1. Dataset Description:

A total of 2200 rows with seven columns composed the dataset that provided agroclimatic conditions with corresponding fertilizer recommendations. This dataset includes subsequent features for input and output as its characteristics.

Input Features:

I. The dataset includes a categorical Crop variable to define the current agricultural product whether maize rice or wheat.

II. The average temperature measurements in Celsius act as the numerical value for this variable.
III. The variable Humidity gives a numerical value to measure air moisture content in percentages.
IV. The rainfall conditions of the area are evaluated through a numerical variable which measures precipitation in millimeters.

Output Features (Target Variables):
I. The N_label field indicates the suggested Nitrogen treatment concentrations through categorical values.
II. The output features include the P_label which presents categorical recommendations for Phosphorus levels.
III. The K_label indicates the Potassium recommendation category as a variable type.

2. Data Preprocessing:

The dataset received multiple processing steps for data preprocessing both before and during training:

I. Missing Value Imputation:
There were missing values in the numerical as well as categorical features. For numeric features (Temperature, Humidity, Rainfall), SimpleImputer with strategy=mean was used to impute the missing values. As for the categorical feature "Crop", its missing values were imputed using the SimpleImputer with the most_frequent strategy.

II. Label Encoding: The output features (N_label, P_label, and K_label) and the Crop feature were categorical variables. Label Encoding was employed to convert these categorical features into a numerical format that can be utilized by machine learning algorithms. This turned any individual category into a number.

III. Feature Scaling: To ensure that all features contribute equally to the model training process and to prevent features with larger scales from dominating, StandardScaler was applied to the numerical features (Temperature, Humidity, Rainfall). This scaled the features to have zero mean and unit variance.

IV. Train-Test Split: The pre-processed dataset underwent training and testing split procedures according to which 80% of data became training data while testing consumed 20% of the dataset. The model training used substantial data parts while testing occurred on an independent dataset in order to evaluate its generalization capabilities.

3. Exploratory Data Analysis:

I. Heatmap:

A correlation heatmap (Fig:4.1) was generated to observe the correlation between the features in the dataset.

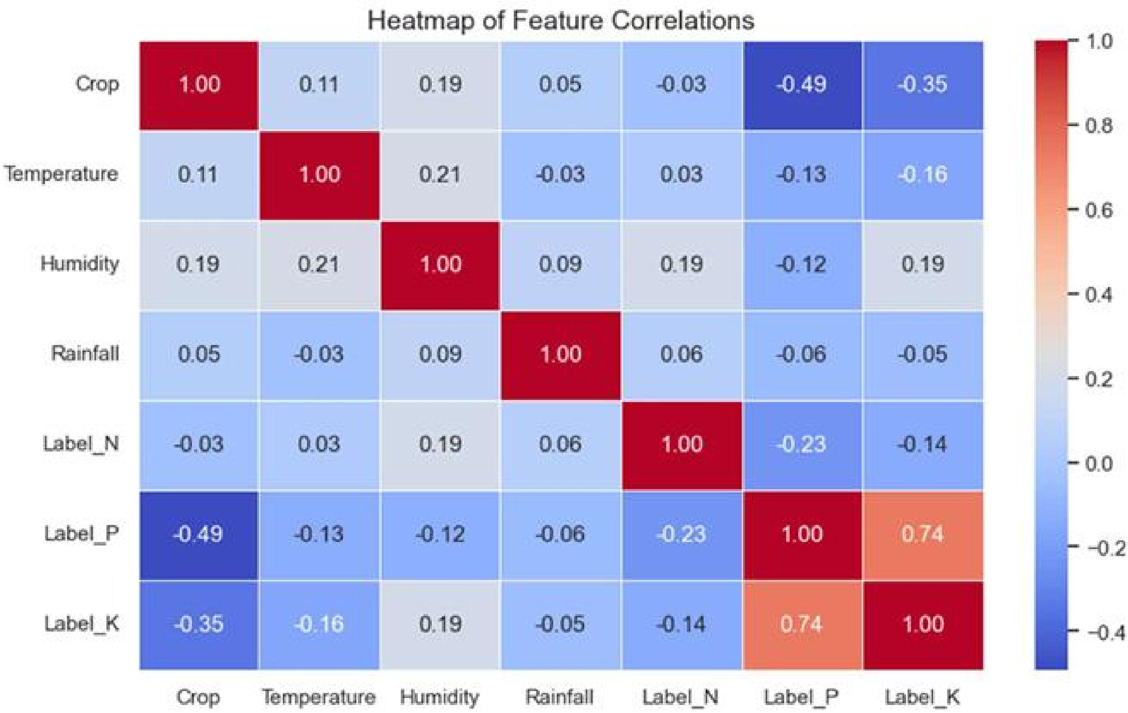

Fig 4.1: Heatmap correlation

The heatmap shows a few significant correlations:

Positive Correlations:

- Highly positive correlation (0.74) between Label_P and Label_K: As we can see, there's a strong relationship between the recommended levels of Phosphorus and Potassium which indicates they are needed together.
- A weak positive correlation between Humidity and Label_N (0.19).
- A small positive correlation between Humidity and Temperature (0.21).

Negative Correlations:

- A moderate negative correlation between Crop and Label_P (-0.49) and Crop and Label_K (0.35). The research shows Phosphorus and Potassium recommended levels exist at reduced amounts for particular plant varieties. Research should conduct an investigation to learn both which crops show these results and the reasons behind their associations.
- A slight negative correlation between Temperature and Label_P (-0.13) and Temperature and Label_K (-0.16).
- A slight negative correlation between Crop and Label_N (-0.03).

- A negative correlation between Label_P and Label_N (-0.23) and Label_K and Label_N (0.14).

Weak Correlations:

- Rainfall demonstrates a weak relationship with other features because most of its correlations remain low.

II. Box-plot:

The box plot is used to visualize the distribution of key agricultural parameters like temperature, humidity, and pH. It helps in detecting outliers and understanding data variability, which improves the accuracy of crop recommendation and yield prediction models.

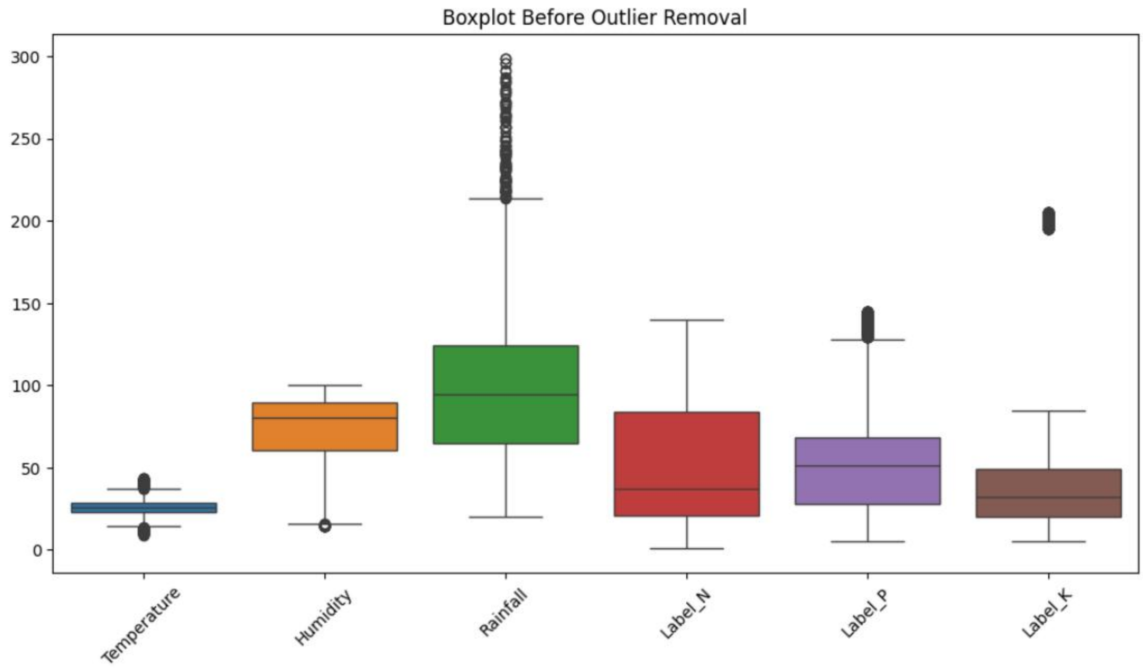

Fig 4.2: Box plot After outlier removal

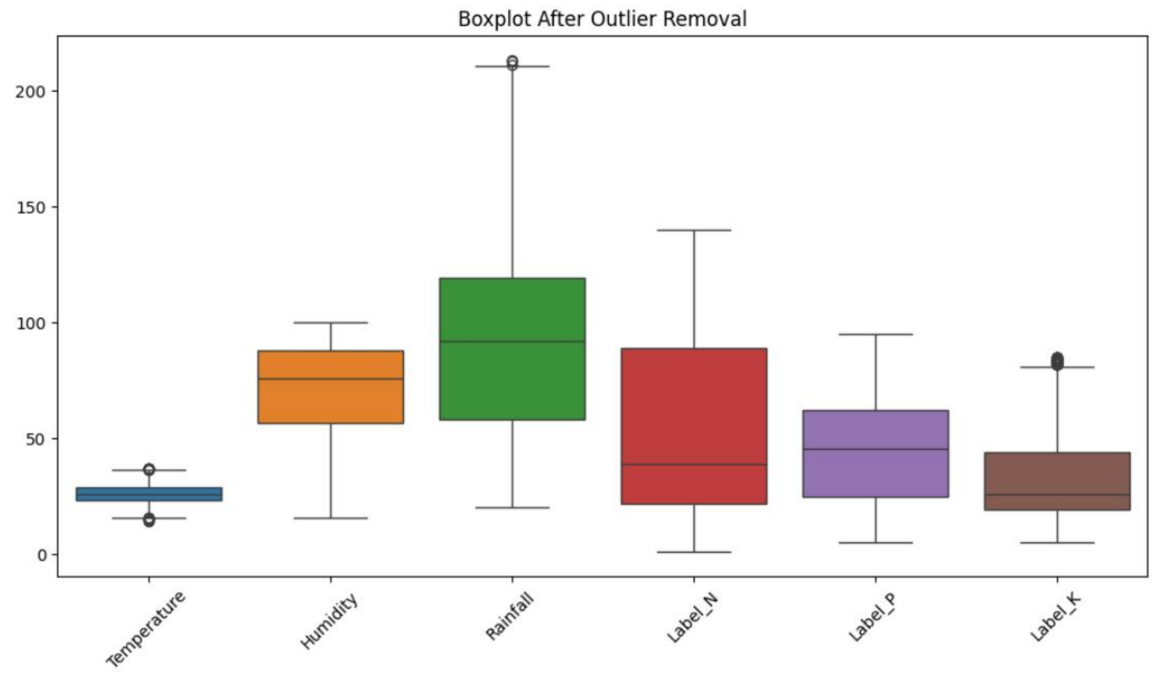

Fig 4.3: Box plot After outlier removal

4. Model Performance Metrics:

Machine learning models require performance evaluation to understand their predictive power and achieve correct results. These assessment metrics help researchers choose the best predictive model for heating and cooling load estimates while allowing them to understand their predictive span for new untrained data. The following assessment measures were used for model evaluation:

I. **R^2 Score:** The R^2 Score reveals how much variability the model explains in dependent variables. The R^2 score represents both names as the measurement that demonstrates how well independent variables explain the dependency variable changes. The model captures most of the target variable values when the R2 value approaches 1.

II**. Mean Squared Error (MSE):** finds the average amount of squared difference between predicted and observed values. Model performance quality rises when the MSE value decreases. MSE has increased sensitivity to outliers because it intensifies its penalty for large deviation distances above minor deviation distances.

III. **Root Mean Squared Error (RMSE):** delivers MSE data in a more understandable form. MSE shows larger errors because its vulnerability to significant deviations surpasses that of RMSE. This technique helps users determine the magnitude of errors in the model effectively.

IV. **Mean Absolute Error (MAE):** serves as a performance evaluation tool for regression models by calculating the mean values of absolute prediction-actual value disparities.

5. Model Development:

A. Classification Models:

- **Logistic Regression:**
  A linear model that predicts the probability of each crop class. It works well as a baseline but has limitations with complex, non-linear patterns.
- **Decision Tree Classifier:**
  Splits data into branches based on feature values, making decisions easy to understand and handling non-linear relationships effectively.
- **K-Nearest Neighbors (KNN):**
  Classifies crops based on the closest data points in the feature space. It is simple and effective when data distribution is clear.
- **Support Vector Machine (SVM):**
  Constructs hyperplanes in high-dimensional space to separate different crop classes, providing good performance with well-separated data.
- **Random Forest Classifier:**
  An ensemble of multiple decision trees that improves accuracy and reduces overfitting by averaging the results. It performed best in the crop recommendation task.

B. Regression Models:

- **Linear Regression:**
  Establishes a linear relationship between environmental features and crop yield. It provides a simple, interpretable approach but may miss complex trends.
- **Decision Tree Regressor:**
  Splits the data into regions with similar output values. It handles non-linearity well and is easy to visualize.
- **Random Forest Regressor:**
  Combines predictions from multiple decision trees to provide more reliable and accurate yield predictions by reducing variance.
- **Gradient Boosting Regressor:**
  Builds models sequentially, each correcting errors from the previous one, which results in highly accurate predictions for crop yield.

4.2 Testing

Test Id	Crop	State	City	Value of N	Value of P	Value of K
1	Rice	Jharkhand	Bokaro	61.28	52.62	19.7
2	Maize	Haryana	Rohtak	62.3	51.56	22.84
3	Mango	Bihar	Arrah	61.96	48.8	20.46
4	Jute	West Bengal	Asansol	57.14	63.98	20.96

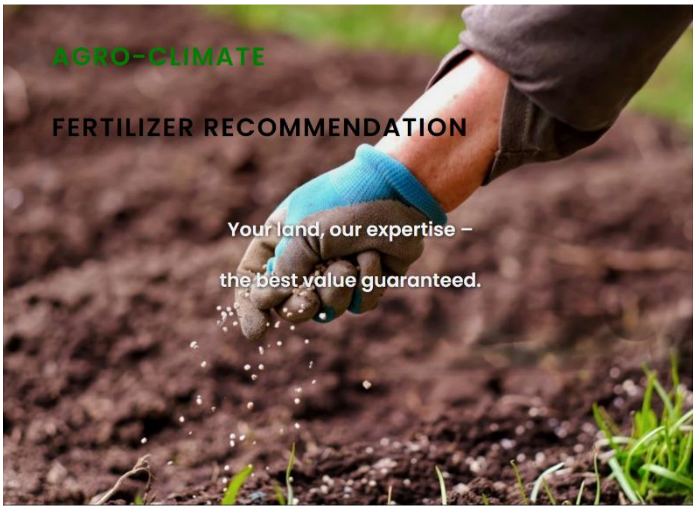

Fig 4.4: Website image

Fig 4.5: Website image

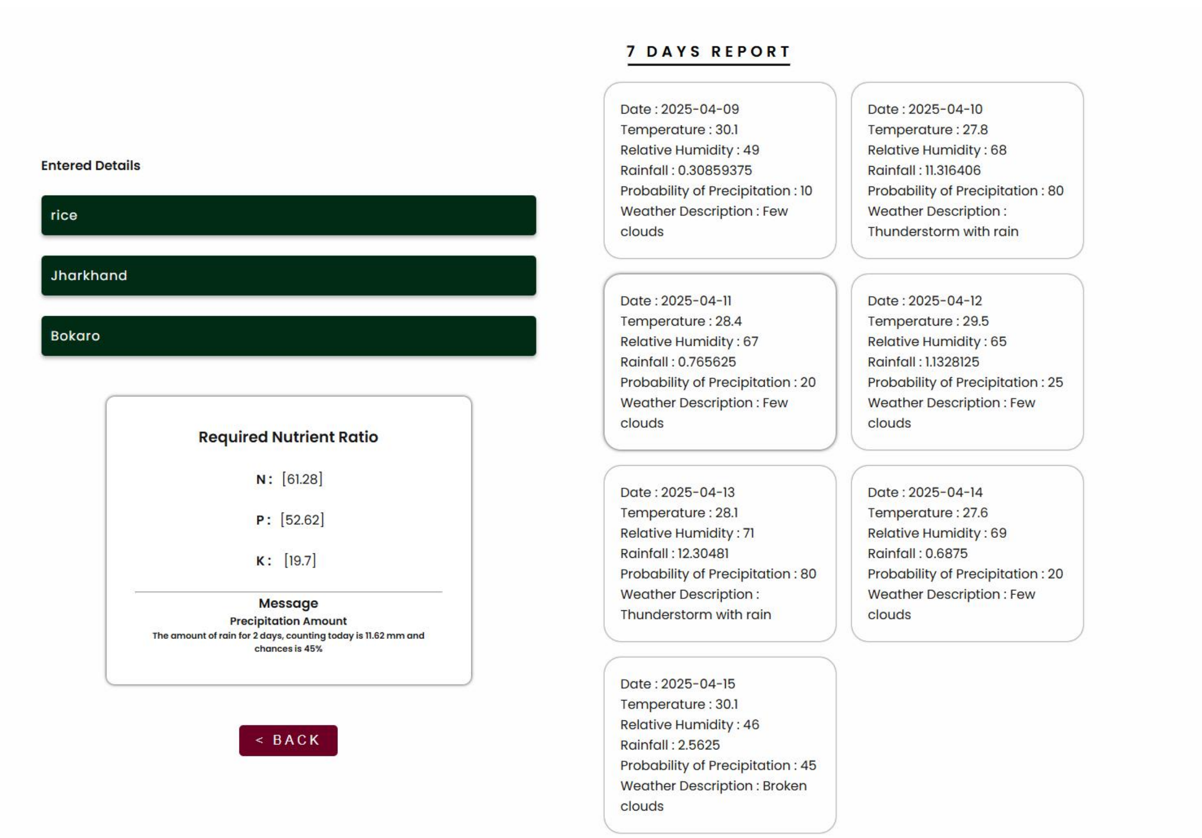

Fig 4.6: Testing result of rice from website

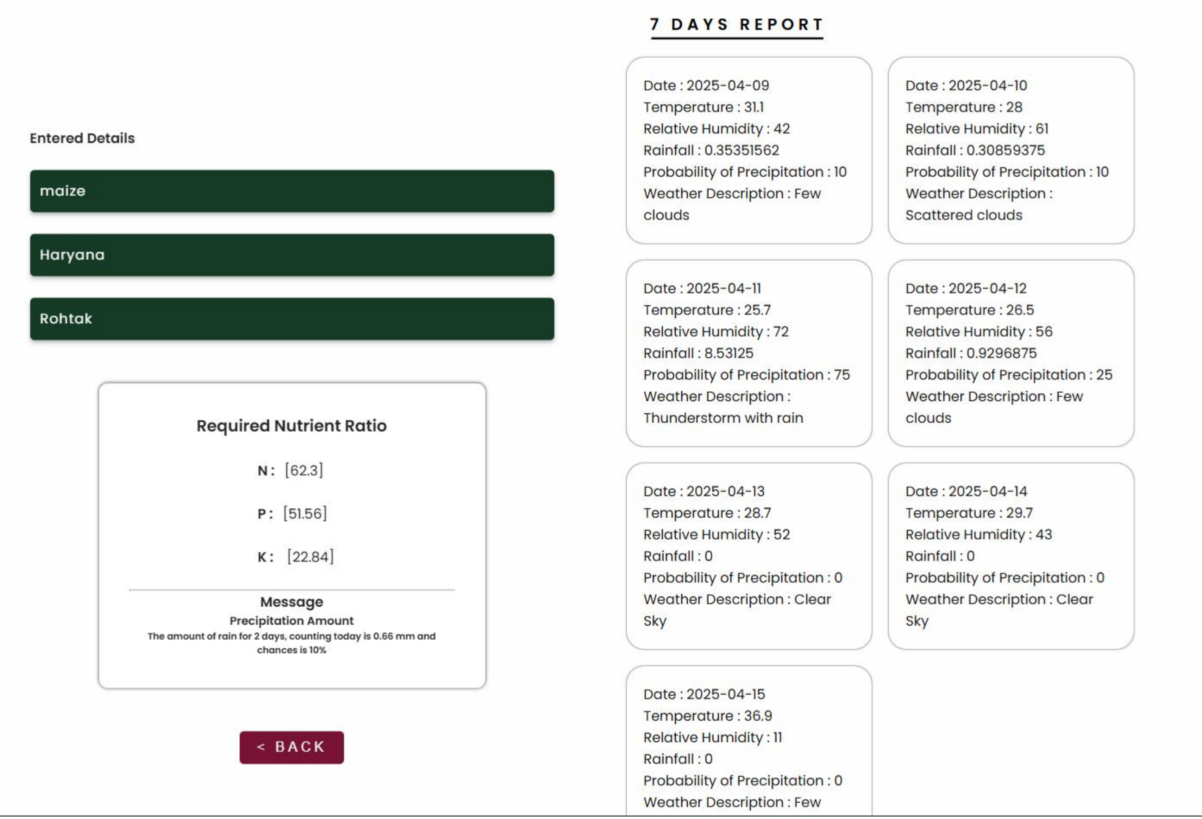

Fig 4.7: Testing result of maize from website

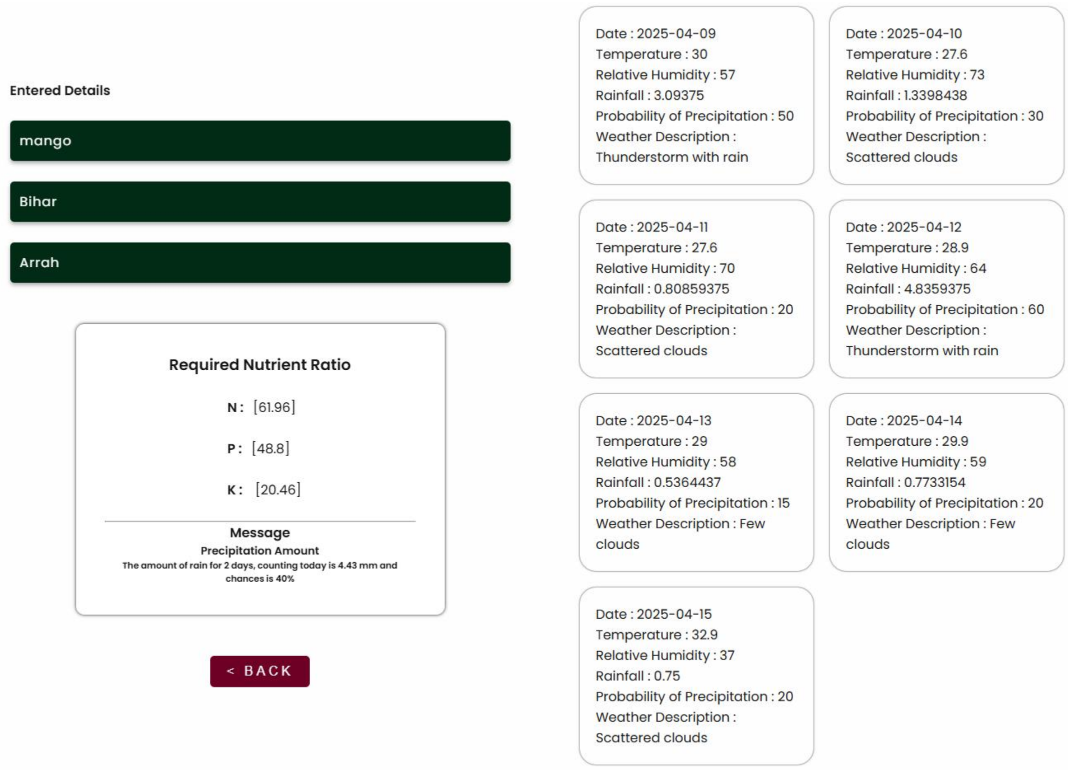

Fig 4.8: Testing result of mango from website

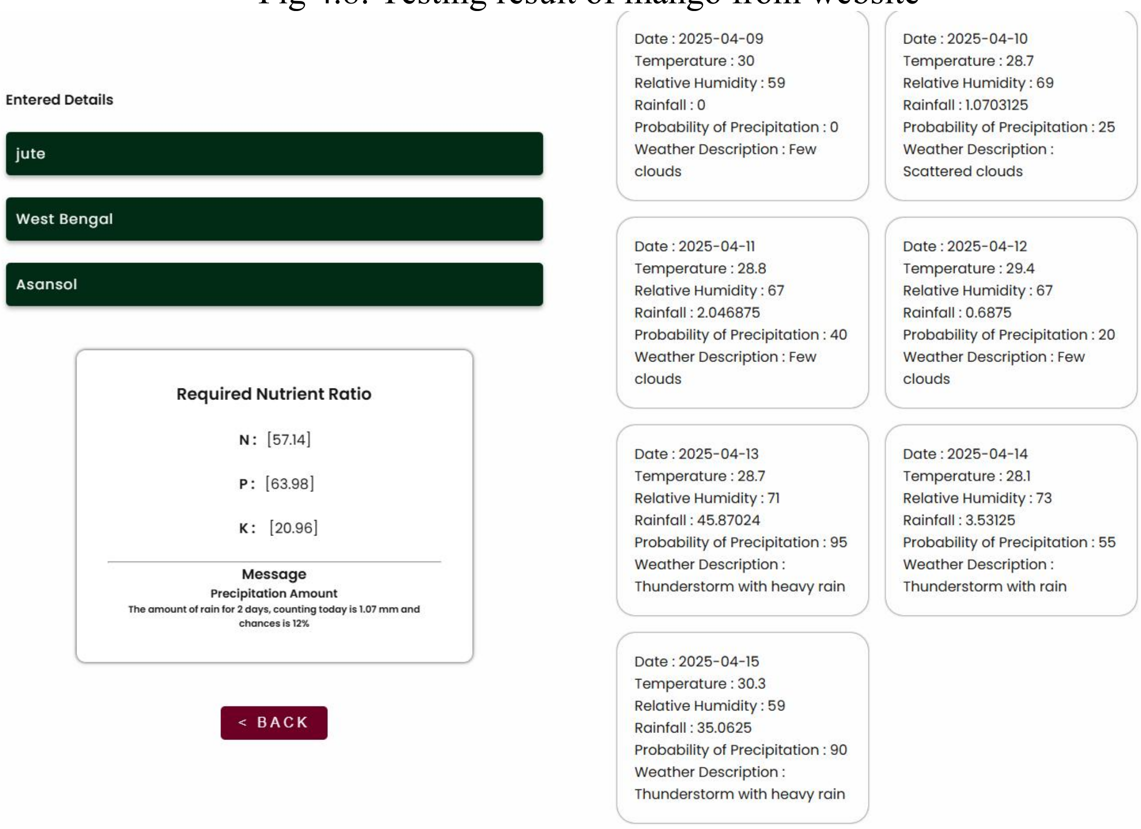

Fig 4.9: Testing result of jute from website

4.3 Result Analysis

Regression models:

Models used	MAE	MSE	RMSE	R^2
Multiple Linear Regression	19.31	608.20	24.6617	0.1552
KNN	7.1996	98.8379	9.9418	0.8686
Decision Tree	8.7826	148.6461	12.1921	0.8047
SVR	7.9896	118.5553	10.8883	0.8480
XGBoost	7.1126	92.4307	9.6141	0.8777
Random Forest	6.5851	78.0155	8.8326	0.908

Classification models:

Models used	Accuracy_N	Accuracy_N	Accuracy_N	Average Accuracy
KNN	0.59	0.69	0.71	0.67
Random Forest	0.56	0.7	0.7	0.65
XGBoost	0.57	0.66	0.71	0.65
Decision Tree	0.52	0.65	0.7	0.62

Confusion Matrix:

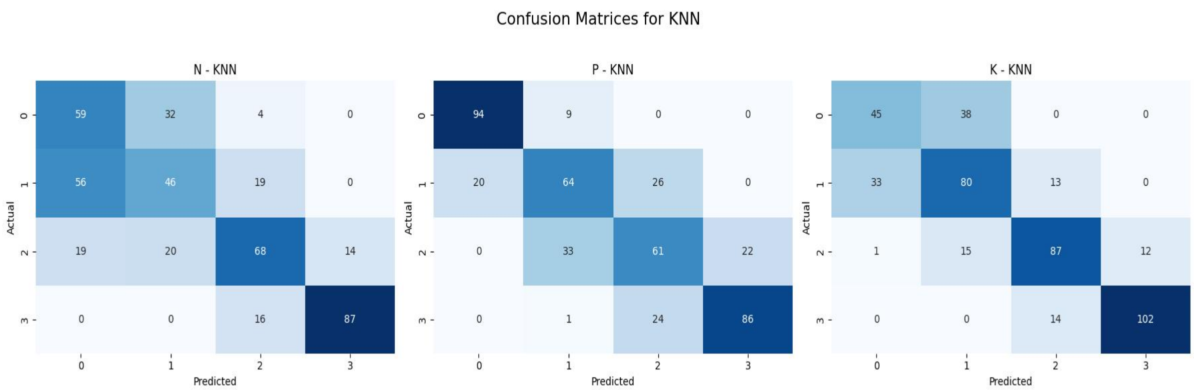

Fig 4.10: Confusion matrix of KNN

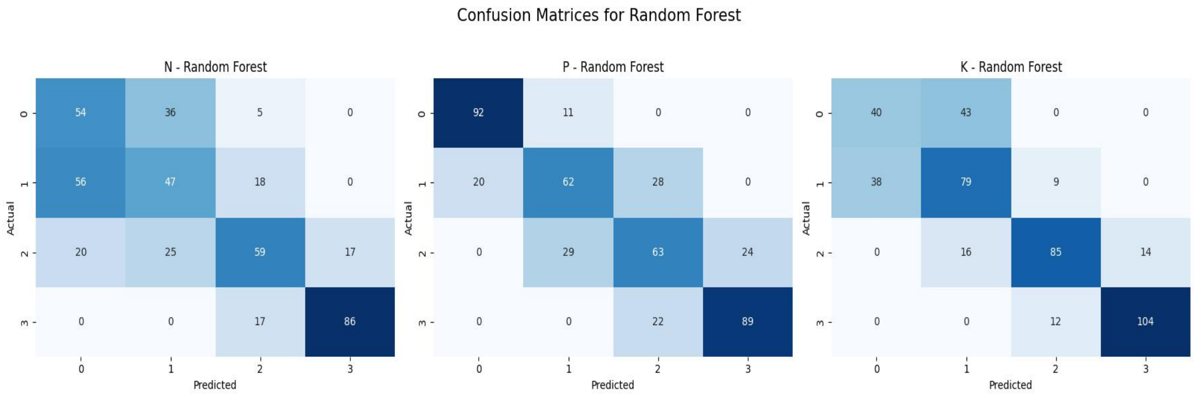

Fig 4.11: Confusion matrix of Random Forest

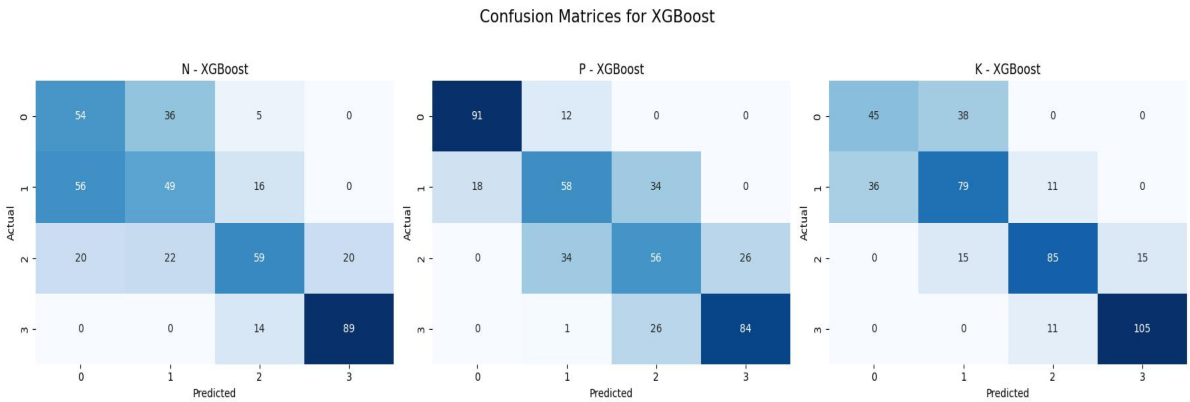

Fig 4.12: Confusion matrix XGBoost

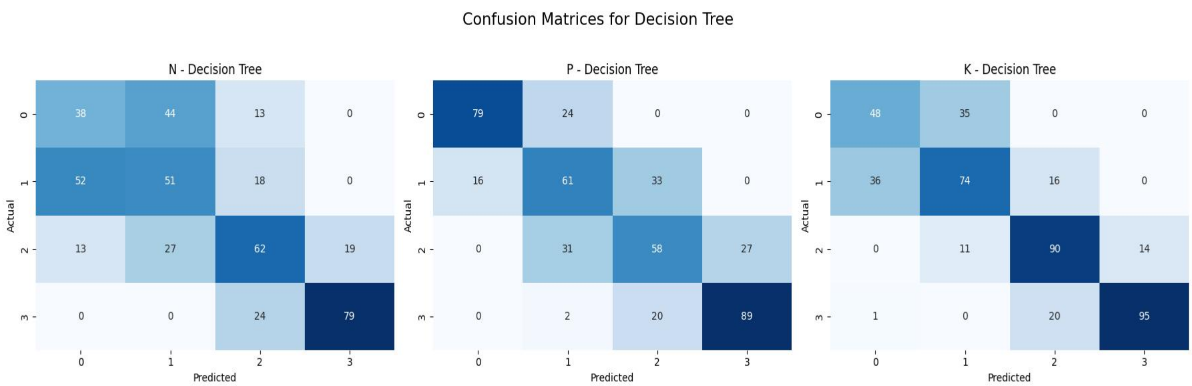

Fig 4.13: Confusion matrix Decision Tree

From the above table it is quite clear that Random Forest Regressor has highest accuracy with 90.8% to correctly identify the correct amount of N, P and K values.

4.4 Quality Assurance

Quality assurance strategies that comprised multiple techniques were utilized to build a system with accuracy and reliability throughout the classification and regression model development of this project. Data validation eliminated incomplete values along with data anomalies and outliers which preserved the overall dataset integrity. Cross-validation provided stability assessments to avoid overfitting and we used accuracy, precision, recall, F1-score along with MAE, MSE, RMSE because these metrics measured model effectiveness. The software execution performed grid search optimization methods to achieve the best achievable model outcomes. The data distribution and detection of inconsistencies used visual assessment tools including box plots and correlation heatmaps. The developed codebase received frequent inspections while tests were performed on different data instances to verify system reliability and growth capabilities. The project implemented well-known libraries such as Scikit-learn Pandas Matplotlib together with TensorFlow to uphold developmental standards. The project implemented standard data science practices from industry regulations to maintain its quality despite absence of external verification.

Chapter 5

Standards Adopted

5.1 Design Standards

Block diagrams served to display the system architecture together with data flow during project design since they provided straightforward visual explanations of the crop recommendation and fertilizer prediction processes. Python code implements standards of software development through PEP 8 guidelines which enhance both code readability and maintenance. The implementation of data normalization and encoding alongside data visualization methods through Pandas, NumPy, Seaborn and Matplotlib libraries achieved best practices for data preprocessing. The data science process followed CRISP-DM methodology for its workflow stages which included data understanding and preparation together with modeling and evaluation along with deployment readiness. The designed visualizations focused on clarity in order to enhance effective interpretation of model outcomes.

5.2 Coding Standards

A scheme of descriptive and meaningful naming conventions improved the code understanding by being applied to all variables and functions and classes. The logic of the program became straightforward because naming conventions matched the specific functions of project components such as *temperature and humidity and ph_value and nitrogen.*

The codebase used a modular structure that divided sections into separate functions which carried out single distinct tasks. The approach enhanced both future development ease and code debugging efficiency while enabling easier reuse of code blocks. The entire codebase had uniform indentation to present a visual representation of program logic which was most important for loops and conditional statements.

Development process benefits greatly from both the use of comments and documentation. The code received inline comments together with docstrings which explained intricate parts to make understanding simpler for both present and future contributors. The development process was characterized by using trusted libraries alongside *Pandas, NumPy, Matplotlib and Seaborn and Scikit-learn* while also preventing redundant code lines.

5.3 Testing Standards

The software quality assurance testing procedures aligned with established standards monitored both accuracy and reliability of the developed models. The framework followed IEEE 829 (Standard for Software Test Documentation) to produce structured test cases which verified every component from data handling through model training to prediction results. Unit testing was applied consistently to data loading mechanisms and prediction processing along with preprocessing functions thus ensuring error management and proper functionality verification. The testing activities along with documentation followed the principles found in ISO/IEC/IEEE 29119 to achieve uniformity. Model testing procedures known as cross-validation were applied to those classification and regression models to improve their robustness through data subset evaluations. The generated model predictions received verification through accuracy calculations in addition to mean squared error computations and classification reports. The verification process received additional help from visualization tools which utilized confusion matrices and regression plots to give clear performance and reliability insights about the model.

Chapter 6

Conclusion and Future Scope

6.1 Conclusion

The project successfully demonstrates integration of machine learning models with agroclimatic data for creating optimized fertilizer recommendations that advance sustainable agricultural practices. The Random Forest algorithm produced better results than other evaluation models when forecasting the best Nitrogen (N), Phosphorus (P), and Potassium (K) nutrient levels from temperature and humidity readings alongside rainfall statistics. The model accomplishes two environmental goals by both controlling nutrient overflow and ensuring soil health improvement. These real-time recommendations from the web application help farmers make better decisions and protect their crops from heavy rainfall-induced nutrient leaching. The model's precision farming abilities are confirmed through its high R^2 scores which show its predictive accuracy. The model faces two main challenges which include inconsistent weather data from APIs and missing essential soil microbiological properties for more accurate predictions. Research should investigate deep learning methods to enhance future span predictions connected to fertilizer consumption. The research outcomes connect traditional agricultural practices with data-powered precision farming methods which lead to enhanced yield results alongside lower environmental strain while supporting better soil conditions for sustainable agricultural practices in the future.

6.2 Future Scope

Machine learning integration for agriculture receives strong support through this project while additional capabilities can be developed in numerous ways. Real-time weather data combined with satellite imagery represents an effective way to generate dynamic and precise soil moisture measurements as well as temperature readouts and rainfall conditions. Such enhancement of weather data collection would lead to better crop recommendation accuracy along with improved yield prediction precision. A holistic improvement of the model would result from adding soil microbiological properties alongside pest/disease data to the existing dataset.

Proficient deep learning algorithms including Convolutional Neural Networks (CNN) for image assessment and Recurrent Neural Networks (RNN) for time-based analysis would make the model perform at a higher level. Farmers will receive precise real-time decisions through improved decision support due to Internet of Things sensors that collect data perpetually in agricultural fields. A mobile-friendly application and a multilingual interface development project enables broader accessibility of the system while addressing farmers who have limited access to digital platforms located in rural areas.

A cloud platform migration of the system will enable the processing of extensive datasets and multiple user access at once to make it viable for government organizations and larger institutions. System evolution into a robust smart agriculture tool is achievable through continuous updates and new data training which establishes it as a sustainability tool for farming and effective resource management and productivity enhancement.

References:

[1] Krutika Hampannavar, Vijay Bhajantri, Shashikumar G. Totad "Prediction of Crop Fertilizer Consumption," Fourth International Conference on Computing Communication Control and Automation (ICCUBEA),2018.

[2] Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. Sensors, 18(8), 2674.

[3] Gupta, A., Sharma, A., & Kumar, R. (2020). Predictive analytics for smart agriculture: Crop and soil health monitoring. IEEE Access, 8, 177115-177132.

[4] M.S. Suchithra, Maya L. Pai, "Improving the prediction accuracy of soil nutrient classification by optimizing extreme learning machine parameters," Information Processing in Agriculture, vol-7, 2022, PP. 1-11

[5] Raves Akhtar, Shabbir Ahmad Sofi, "Precision agriculture using IoT data analytics and machine learning," Journal of King Saud University - Computer and Information Sciences, 2021, PP. 1-17

[6] Benny Antony, "Prediction of the production of crops with respect to rainfall.," Environmental Research, vol-202, 2021, PP. 1-5

[7] Tony Yang, Kadambot H.M., Siddique, Kui Liu, "Cropping systems in agriculture and their impact on soil health," Global Ecology and Conservation, vol-23, year, PP. 1-13

[8] Usman Ahmed, Jerry Chun-Wei Lin, Gautam Srivastava, Youcef Djenouri, "A nutrient recommendation system for soil fertilization based on Evolutionary Computation," Computers and Electronics in Agriculture, vol-189, 2021, PP. 1-7

[9] A.Hussein, Diogenes L. Antille, Shreevatsa Kodur, GuangnanChen, Jeff N.Tullberg, “Controlled traffic farming effects on productivity of grain sorghum, rainfall and fertilizer nitrogen use efficiency,” Journal of Agriculture and Food Research, vol-3, 2021, PP. 1-17

[10] Yulong Yin, Hao Ying, Huifang Zhen, Q ingsong Zhang, Y anfang Xue, Zhenling I, “Estimation of NPK requirements for rice production in diverse Chinese environments under optimal fertilization rate,” Agricultural and Forest Meteorology, vol-279, 2019, PP. 1-6

[11] Potnuru Sai Nishant,Pinapa Sai Venkat,Bollu Lakshmi Avinash,B. Jabber, “Crop Yield Prediction Based on Indian Agriculture using Machine Learning,” 2020 International Conference for Emerging Technology (INCET),

[12] G. Prabakaran, D. Vaithiyanathan, Madhavi Ganesa, “Fuzzy decision support system for improving the crop productivity and efficient use of fertilizers,” Computers and Electronics in Agriculture, vol-150, 2018, PP. 88-97

[13] Laura J.T. Hess, Eve-Lyn S. Hinckley, G. Philip Robertson, Pamela A. Matson, “Rainfall intensification increases nitrate leaching from tilled but not no-till cropping systems in the U.S. Midwest,” Agriculture, Ecosystems & Environment, vol-290, 2020, PP. 1-10

Individual Contribution

Dona Choudhury:	Classification models
Adarsh Kumar:	Report + website + regression model
Eshita Yadav:	Dataset conversion + classification
Gulshan Kumar:	Dataset conversion + classification
Manish Kumar:	Regression model+ website
Shreyasi Saha:	Report