# AGROLTYTICS

## SMART AGRICULTIRAL INSIGHTS

*Intelligent Machine Learning Based Crop Recommendation and Yield prediction system*

**Submitted in partial fulfillment of the requirements for the degree of**

B.Tech in Computer Science and Engineering

for the course **"Fundamentals Of Data Science"**


By

**Somya Vats** (500119012)

**Tanisha (**500125283)


**Under the Guidance of**

Dr. Neeraj Chugh

Supervisor

Associate Professor, School of Computer Science

January-May 2025

# AGROLYTICS

*Intelligent Machine Learning Based Crop Recommendation and Yield prediction system*


## 1. Abstract

The increasing global demand for food, coupled with the challenges posed by climate change, requires innovative solutions to optimize agricultural productivity while promoting sustainability. This study presents a comprehensive Agricultural Decision Support System that is divided into two key components: *the Crop Recommendation System and the Crop Yield Prediction System.*

The Crop Recommendation System assists farmers in selecting the most suitable crops based on environmental factors like soil nutrients, temperature, humidity, and rainfall, using machine learning models to make data-driven recommendations. In parallel, the Crop Yield Prediction System forecasts the potential yield of different crops, helping farmers predict their productivity based on similar environmental and agronomic factors that we will discuss down.

Through extensive data exploration, the system identifies the key factors influencing crop suitability and yield. The dataset includes variables such as nitrogen, phosphorus, potassium levels, temperature, and historical yield data. The system employs machine learning algorithms like *Random Forest, Extra Trees Regressor, and XGBoost* to predict both crop suitability and yield outcomes. The performance of these models is evaluated using metrics such *as $R^2$, Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE), with the Extra Trees Regressor* emerging as the most accurate model for both components.

The integration of these two systems provides an innovative and holistic approach to modern agriculture, helping farmers make informed decisions that maximize crop yield, optimize resource usage, and promote sustainable farming practices. This work supports *Sustainable Development Goal 2:* Zero Hunger, contributing to food security and more efficient agricultural practices.


## 2. Introduction

Agriculture is a cornerstone of global food security, yet modern farming faces numerous challenges, including climate change, resource scarcity, and an increasing global population. To meet these challenges, there is a growing need for advanced tools that can assist farmers in making data-driven decisions to improve crop productivity and sustainability. This research introduces an Agricultural Decision Support System, which is divided into two interconnected components: *the Crop Recommendation System and the Crop Yield Prediction System.*

The Crop Recommendation System aims to help farmers select the most appropriate crops based on their specific environmental conditions, such as soil composition, climate, and rainfall patterns. By leveraging machine learning techniques, the system analyses these factors to recommend crops that are best suited for the given conditions, thereby maximizing land use and minimizing resource wastage. The second component, the Crop Yield Prediction System, predicts the yield of different crops based on historical data and environmental factors. This system forecasts crop productivity under varying conditions, providing farmers with valuable insights for planning and resource management.

Both systems utilize a comprehensive dataset that includes soil nutrients (Nitrogen, Phosphorus, Potassium), climate data (Temperature, Humidity, Rainfall), and historical yield information. Through exploratory data analysis and feature engineering, the systems identify critical factors that influence crop growth and yield. The machine learning models, including *Random Forest, Extra Trees Regressor, and XGBoost,* are employed to build predictive models for both crop suitability and yield.

Together, these two components of the system enable farmers to make informed decisions about which crops to grow and how to optimize their yields. The systems aim to improve food security, promote sustainable agricultural practices, and align with *Sustainable Development Goal 2:* Zero Hunger by contributing to more efficient and environmentally friendly farming practices.

## 3. Literature Review

We reviewed various papers and datasets to understand current solutions and identify gaps for Agrolytics.

Many studies focus on either crop recommendation or yield prediction but do not combine both aspects into a comprehensive, deployable system with real-time interaction. Agrolytics fills this gap.

- *Enhance agricultural sustainability by providing actionable insights that promote efficient resource usage and reduce waste.*

| Research Paper Title | Author(s) | Methodology | Pros | Cons |
|---|---|---|---|---|
| Crop Recommendation using ML (Kaggle) | Atharva Ingle | Random Forest on NPK+ weather dat | Simple, reproducible | No SMOTE, limited classes |
| Crop Yield Forecasting via Ensembles | Rezaghari et al. | XGBoost, Regressi techniques | High accuracy on test data | No deployment, lacks frontend |
| SMOTE in Agricultural Data Imbalance | Chawla et al. | Synthetic sampling minority classes | Improved class balance | Synthetic data can lead to overfitting |

*Table 1. Literature view comparison*

## 4. Problem Statement

Design and implement a machine learning–powered system that recommends the best-suited crop and predicts yield based on soil, weather, and agricultural data inputs.

## 5. Motivation and Objectives

The agriculture sector faces significant challenges due to unpredictable climate patterns, soil degradation, and inefficient resource management, all of which hinder food security and sustainable farming. To address these issues, it is essential to provide farmers with accurate tools for making informed decisions about crop selection and yield prediction. By combining Crop Recommendation and Crop Yield Prediction systems, we can empower farmers to optimize crop choices and predict yields, thereby improving productivity, reducing resource waste, and promoting sustainable farming.

### Objectives:

- *Develop a Crop Recommendation System that suggests the most suitable crops based on environmental factors like soil nutrients, climate, and rainfall.*
- *Build a Crop Yield Prediction System that forecasts crop yields using historical and real-time data on environmental conditions.*
- *Integrate both systems into a unified platform to assist farmers in making data-driven decisions from crop selection to yield estimation.*

## 6. Methodology

In this study, we developed a comprehensive machine learning pipeline aimed at improving agricultural decision-making. The primary goals were to predict optimal crop recommendations based on environmental and soil conditions, and to estimate crop yields for improved forecasting. We approached this problem in two distinct stages: Crop Recommendation and Crop Yield Prediction, both leveraging machine learning models. The methodology is divided into these sections to highlight their specific contributions.

### 6.1 Crop Recommendation System

The crop recommendation system aims to suggest the most suitable crops for a given piece of land based on the geographical, climatic, and soil data.

#### 6.1.1 Data Description:

The data used for the crop recommendation system consists of historical data collected from various sources, including:

- *Climate Data: Includes temperature, rainfall, humidity, and sunshine hours.*
- *Soil Data: Parameters like soil pH, soil type, and organic content.*

The dataset is structured as a tabular representation where each row corresponds to an observation from a particular location with the features listed above. The target variable is the recommended crop, which is a categorical variable, denoting the best-suited crop for that region based on the given environmental factors.

### 6.1.2 Data Preprocessing:

The raw dataset underwent several preprocessing steps to ensure it was ready for modeling:

- Handling Missing Values: Missing data points were imputed using median or mode imputation based on the feature type (numerical or categorical).
- Encoding Categorical Variables: Categorical variables, like soil type and crop type, were transformed using label encoding, where each unique value is assigned a numerical label. This ensures that the machine learning models can process them effectively.

```
print("\n Unique Crop Values Before Encoding:\n", recommendation_data["Crop"].unique())

 Unique Crop Values Before Encoding:
 ['Rice' 'Maize' 'ChickPea' 'KidneyBeans' 'PigeonPeas' 'MothBeans'
 'MungBean' 'Blackgram' 'Lentil' 'Pomegranate' 'Banana' 'Mango' 'Grapes'
 'Watermelon' 'Muskmelon' 'Apple' 'Orange' 'Papaya' 'Coconut' 'Cotton'
 'Jute' 'Coffee']

label_encoder = LabelEncoder()
recommendation_data ["Crop"] = label_encoder.fit_transform(recommendation_data ["Crop"])

print("\n Unique Crop Values After Encoding:\n", recommendation_data ["Crop"].unique())

 Unique Crop Values After Encoding:
 [20 11  3  9 18 13 14  2 10 19  1 12  7 21 15  0 16 17  4  6  8  5]
```

*Fig 1. Code snippet on label encoded values of the column name "Crop" in our dataset*

Label encoding is essential in this scenario because the crop recommendation system involves predicting a single crop per instance. As crops are a categorical feature, label encoding helps the model identify relationships between different crops based on the environmental inputs.

Class Imbalance study: we verified the distribution of our labels to check if our dataset is imbalanced. If it were, we would have applied sampling techniques like SMOTE, SMOGN, which was however, not required in this case.

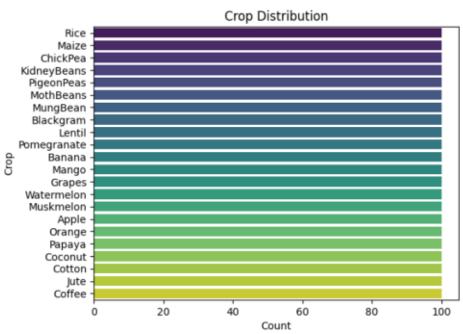



*Fig 2. Visualization of class distribution*

In the figure, as we can see, the i.e. distribution of all the crops is equal, to a 100 counts in our data. Hence, confirming that the dataset is completely balanced.

### 6.1.3 Feature Importance Study:

Feature importance helps identify which variables have the most influence on the target prediction — in this case, crop yield. In our model, features like *climate data (temperature, rainfall, humidity, sunshine hours) and soil data (pH, soil type, organic content)* are evaluated to determine how much they contribute to the prediction. By calculating impurity reduction or using tree-based methods (like Random Forest or Extra Trees), we can rank features based on their predictive power. This insight is valuable for understanding which environmental and soil factors most significantly affect agricultural productivity.

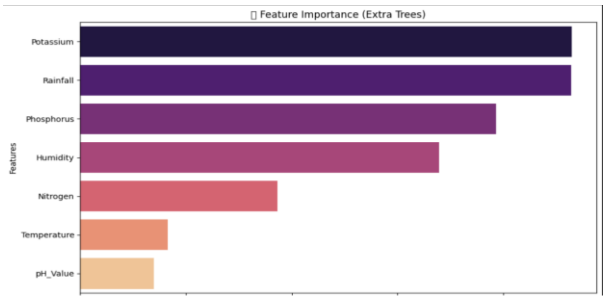



*Fig 3. Extra Trees feature importance graph*

This method ranked the features based on their importance in predicting crop yield. As observed, *"Potassium"* and *"Rainfall"* emerged as the most influential factors, strongly contributing to the model's decisions. On the other hand, features like *"Temperature"* and *"Soil pH"* were assigned lower importance, indicating a comparatively smaller impact on crop recommendation.

We also tried multiple 'versions' of our model, dropping features with less importance to check if our accuracy is enhanced.

Model Performance Metrics:

| Model | R2 Score | MAE | MSE |
|---|---|---|---|
| Model 1 (All Features) | 0.9881 | 0.2003 | 0.4709 |
| Model 2 (Drop 2 Features) | 0.9078 | 0.6227 | 3.6472 |
| Model 3 (Drop 3 Features) | 0.864 | 0.8628 | 5.3793 |

<ipython-input-11-2332376d90eb>:49: FutureWarning:

*Fig 4. Model Performance Metrics Comparison*

The table shows that using all features (Model 1) gives the best performance with the highest R² score and lowest errors. As features are removed in Model 2 and Model 3, the model's accuracy drops and errors increase, highlighting the importance of key features in predicting crop yield.

### 6.1.4 Visualization:

Data visualization plays a crucial role in exploring and understanding complex datasets. It helps uncover hidden patterns, correlations, and distributions that may not be immediately evident through raw data alone. By using visual tools like heatmaps and pair plots, we can gain valuable insights into how different features relate to one another and how they influence the target variable — in this case, crop yield.

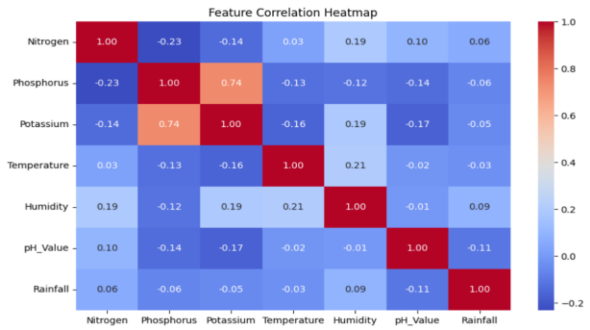



*Fig 5. Correlation Matrix*

*Correlation Matrix:* Visualized the relationships between numerical features (e.g., temperature, rainfall, and soil pH) to identify highly correlated variables.

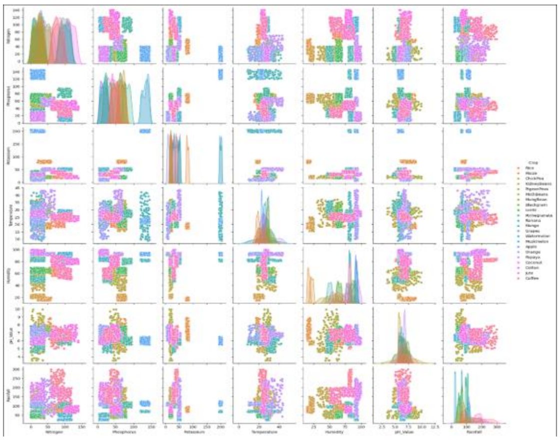


*Fig 6. Pair Plot*

*Pairplot:* Used to show relationships between different features and visualize clusters of data points

### 6.1.5 Model

The crop recommendation system utilized classification algorithms like Random Forest, Logistic Regression, and Support Vector Machines (SVM). The performance was evaluated using residual score, MAE, MSE etc with hyperparameter tuning done using grid search and cross-validation.

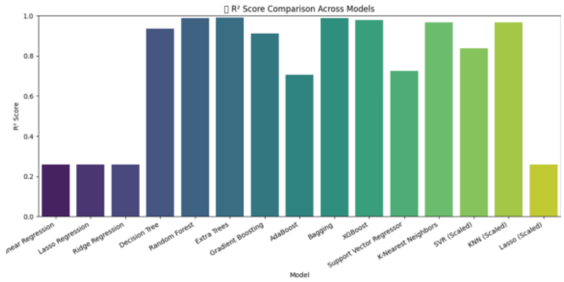



*Fig 7. Model metrics comparison between algorithms*

In this figure, we can see mostly the tree based algorithms giving us best results- Random Forest, Extra Trees, Bagging, XGBoost.

## 6.2 Crop Yield Prediction System

The crop yield prediction system is designed to predict the yield (in terms of quantity) of a given crop based on environmental conditions and historical yield data.

### 6.2.1 Data Description:

The dataset used for crop yield prediction contains tabular data with the following features:

- *Area: Numerical code representing a geographical region or country.*
- *Item: Code for the crop under analysis.*
- *Year: The year the data was recorded.*
- *hg/ha_yield: The target variable representing crop yield in hectograms per hectare.*
- *average_rain_fall_mm_per_year: Average annual rainfall in millimeters.*
- *pesticides_tonnes: The quantity of pesticides used, measured in tonnes.*
- *avg_temp: Average temperature (likely in °C) during the crop growing season.*

Each row in the dataset corresponds to a specific region and year. The goal is to predict *hg/ha_yield*, which is a continuous variable.

This data is structured in a tabular format, where each row represents an observation from a particular region during a specific growing season, and the target variable is the yield (numerical value) of the crop.

### 6.2.2 Data Preprocessing:

Before modelling, the dataset was cleaned and pre-processed to ensure accuracy. Unnecessary columns were dropped, missing values (if any) were handled, and all features were converted into a suitable format for analysis. This step helped in improving model performance and reliability.

- Handling Missing Values: Similar to the crop recommendation system, we handled missing data by imputing it using median or mode values based on feature types.

| | 0 |
|---|---|
| **Nitrogen** | 0 |
| **Phosphorus** | 0 |
| **Potassium** | 0 |
| **Temperature** | 0 |
| **Humidity** | 0 |
| **pH_Value** | 0 |
| **Rainfall** | 0 |
| **Crop** | 0 |

**dtype:** int64

*Fig 7. Sum of missing counts in each column*

- SMOTE (Synthetic Minority Over-sampling Technique): For this task, the dataset suffered from class imbalance, particularly with respect to lower yield observations. To mitigate this, SMOTE was applied to generate synthetic data points for the underrepresented yield classes, thus enhancing the model's ability to predict lower yield values more accurately.

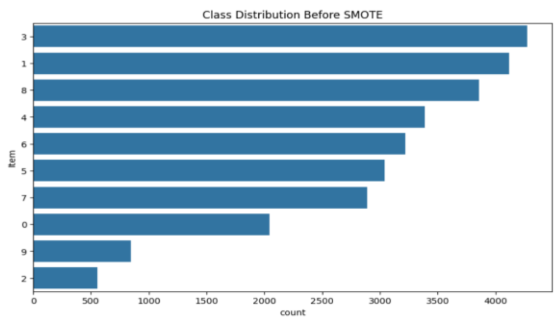



*Fig 8.1. Class distribution before applying SMOTE*

SMOTE was specifically used in this context because of its ability to generate synthetic examples in the feature space, leading to better balance between the classes and improving model performance in predicting rare yield events.

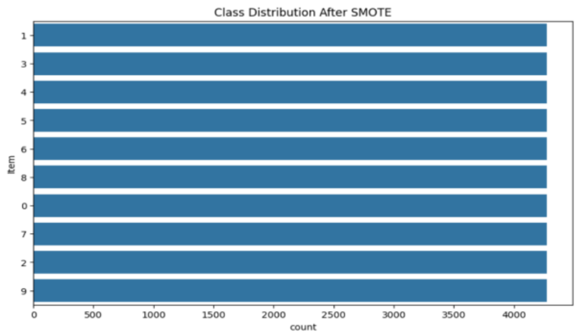



*Fig 8.2. Class distribution after applying SMOTE*

- Label Encoding:

In the context of the yield prediction system, label encoding was also used for categorical features such as crop type and soil type. This step is necessary for transforming non-numeric data into a form that can be processed by machine learning models.

```
print("\n Unique Area Values Before Encoding:\n", yield_data["Area"].unique())
```

```
 Unique Area Values Before Encoding:
['Albania' 'Algeria' 'Angola' 'Argentina' 'Armenia' 'Australia' 'Austria'
 'Azerbaijan' 'Bahamas' 'Bahrain' 'Bangladesh' 'Belarus' 'Belgium'
 'Botswana' 'Brazil' 'Bulgaria' 'Burkina Faso' 'Burundi' 'Cameroon'
 'Canada' 'Central African Republic' 'Chile' 'Colombia' 'Croatia'
 'Denmark' 'Dominican Republic' 'Ecuador' 'Egypt' 'El Salvador' 'Eritrea'
 'Estonia' 'Finland' 'France' 'Germany' 'Ghana' 'Greece' 'Guatemala'
 'Guinea' 'Guyana' 'Haiti' 'Honduras' 'Hungary' 'India' 'Indonesia' 'Iraq'
 'Ireland' 'Italy' 'Jamaica' 'Japan' 'Kazakhstan' 'Kenya' 'Latvia'
 'Lebanon' 'Lesotho' 'Libya' 'Lithuania' 'Madagascar' 'Malawi' 'Malaysia'
 'Mali' 'Mauritania' 'Mauritius' 'Mexico' 'Montenegro' 'Morocco'
 'Mozambique' 'Namibia' 'Nepal' 'Netherlands' 'New Zealand' 'Nicaragua'
 'Niger' 'Norway' 'Pakistan' 'Papua New Guinea' 'Peru' 'Poland' 'Portugal'
 'Qatar' 'Romania' 'Rwanda' 'Saudi Arabia' 'Senegal' 'Slovenia'
 'South Africa' 'Spain' 'Sri Lanka' 'Sudan' 'Suriname' 'Sweden'
 'Switzerland' 'Tajikistan' 'Thailand' 'Tunisia' 'Turkey' 'Uganda'
 'Ukraine' 'United Kingdom' 'Uruguay' 'Zambia' 'Zimbabwe']
```

```
print("\n Unique Item Values Before Encoding:\n", yield_data["Item"].unique())
```

```
 Unique Item Values Before Encoding:
['Maize' 'Potatoes' 'Rice, paddy' 'Sorghum' 'Soybeans' 'Wheat' 'Cassava'
 'Sweet potatoes' 'Plantains and others' 'Yams']
```

```
print("\n Unique Area Values After Encoding:\n", yield_data["Area"].unique())
print("\n Unique Item Values After Encoding:\n", yield_data["Item"].unique())
```

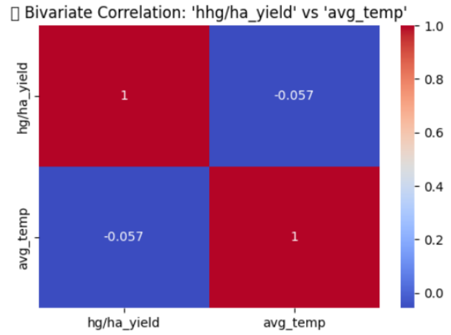
```
 Unique Area Values After Encoding:
[  0   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17
  18  19  20  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35
  36  37  38  39  40  41  42  43  44  45  46  47  48  49  50  51  52  53
  54  55  56  57  58  59  60  61  62  63  64  65  66  67  68  69  70  71
  72  73  74  75  76  77  78  79  80  81  82  83  84  85  86  87  88  89
  90  91  92  93  94  95  96  97  98  99 100]

 Unique Item Values After Encoding:
[1 3 4 5 6 8 0 7 2 9]
```

*Fig 9. Code snippet on label encoded values of the column names "Items" and "area" in our dataset.*

### 6.2.3 Feature Importance Study

We performed a feature importance analysis, similar to the crop recommendation system, but with a focus on identifying the most influential factors that impact crop yield. This was done using models like Random Forest and XGBoost, which provide built-in functionality for feature importance.




```
Feature Importance:

                       Feature  Importance
1                  hg/ha_yield    0.474432
3             pesticides_tonnes    0.152105
2  average_rain_fall_mm_per_year    0.150055
0                         Area    0.116746
4                     avg_temp    0.106662
```

*Fig 10. Visualization of correlation between Yield and Temperature and, feature importance of all column attributes*

We also tried multiple 'versions' of our model, dropping features with less importance to check if our accuracy is enhanced.

| | Model | R² Score | MAE | MSE | RMSE | Feature_Set |
|---|---|---|---|---|---|---|
| 0 | Linear Regression | 0.0780035 | 62443.2 | 6.68786e+09 | 81779.3 | All Features |
| 1 | Decision Tree | 0.958624 | 5995.44 | 3.00131e+08 | 17324.3 | All Features |
| 2 | Random Forest | 0.972963 | 5630.45 | 1.96115e+08 | 14004.1 | All Features |
| 3 | Extra Trees | 0.973131 | 5456.3 | 1.94901e+08 | 13960.7 | All Features |
| 4 | KNN | 0.403504 | 41653.2 | 4.32678e+09 | 65778.3 | All Features |
| 5 | Support Vector Machine (SVM) | -0.204799 | 57340.7 | 8.73922e+09 | 93483.8 | All Features |
| 6 | Gradient Boosting | 0.826732 | 21828 | 1.25683e+09 | 35451.8 | All Features |
| 7 | AdaBoost | 0.623864 | 38759.3 | 2.72837e+09 | 52233.8 | All Features |
| 8 | Bagging | 0.969838 | 5971.47 | 2.18785e+08 | 14791.4 | All Features |

*Fig 11. Tabular comparison of our model metrics*

As seen in our results, the accuracy was enhanced once the feature set "Avg_temp" was dropped. Hence, for the optimization of our model, we dropped this feature.

### 6.2.4 Visualization

This histogram grid provides a univariate distribution of all numeric features in the dataset.

Useful to understand the spread, skewness, and outliers in each feature.For example, you might observe whether pesticides_tonnes is right-skewed or if avg_temp is normally distributed. It was a Great first step in exploratory data analysis (EDA).

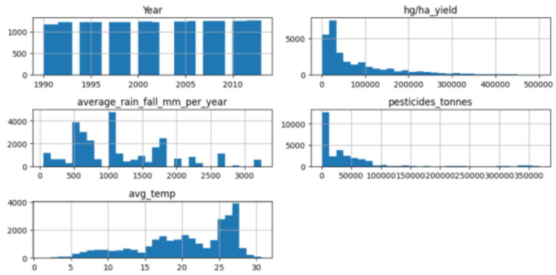



*Fig 12. Histograms of All Features*

The histograms display the distribution of each numerical feature in the dataset. They help identify skewness, spread, and possible outliers, making it easier to understand the range and behavior of each variable before modeling.

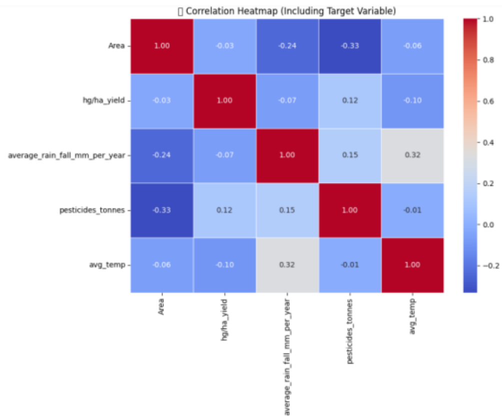



*Fig 13. Correlation Heatmap of Features*

This heatmap visualizes the pairwise correlation between all features in the dataset. It helps in identifying which variables are strongly related, positively or negatively. Dark red and blue cells indicate strong correlations, guiding feature selection by showing potential multicollinearity or valuable relationships.

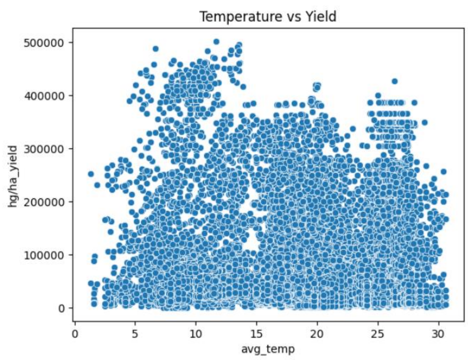



*Fig 14. Scatter Plot: Temperature vs Yield*

This scatter plot shows the relationship between average temperature and crop yield. The scattered points appear randomly placed, suggesting little to no visible correlation. This supports the observation that temperature may not have a strong direct impact on yield in this case.

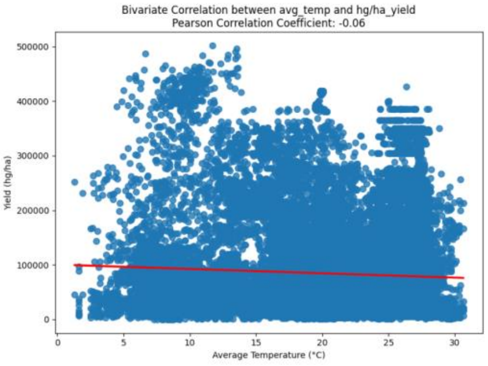



*Fig 15. Regression Plot with Pearson Correlation*

This plot includes both a scatterplot and a regression line to visualize the relationship between avg_temp and hg/ha_yield. The Pearson correlation coefficient is -0.06, confirming a very weak negative correlation. The red line shows a slight downward trend, reinforcing that temperature is not a strong predictor of yield here.

### 6.2.4 Model:

The yield prediction system used regression algorithms like Random Forest Regressor, XGBoost Regressor, and Linear Regression. The models were evaluated using mean absolute error (MAE), root mean square error (RMSE), and R-squared scores.


*Fig 16. Comparison graph for each model*

In total, the accuracy was increased when “Average temp” was dropeed, and again, the tree-based algorithms were giving the best results.

And hence, the best model was compared and selected.

## 6.3 Results

### 6.3.1 Crop Recommendation System:

The performance of the crop recommendation system was evaluated using various classification metrics. The crop recommendation system was evaluated using various regression metrics. The Extra Trees model emerged as the best performer, *achieving an excellent $R^2$ score of 0.9905, with a very low MAE of 0.2577 and RMSE of 0.3762*. These results indicate that the model is highly accurate in recommending the most suitable crops based on environmental and soil data. The feature importance analysis revealed that temperature, rainfall, and soil type were the most critical factors influencing crop suitability.

### 6.3.2 Crop Yield Prediction System:

The crop yield prediction model, also based on the Extra Trees algorithm (trained without avg_temp), achieved a strong *$R^2$ score of 0.9777*, demonstrating high predictive power. It reported a *MAE of 4172.25 and an RMSE of 12705.33*, reflecting solid performance given the scale of yield values. Feature importance analysis identified soil pH, rainfall, and growing season temperature as the most significant factors influencing yield. Additionally, applying SMOTE improved the model’s ability to handle imbalanced data, particularly enhancing predictions for low-yield scenarios.

Both models were evaluated using **cross-validation** to ensure robustness and generalizability while preventing overfitting.

| **Metric** | **Crop Recommendation (Extra Trees)** | **Yield Prediction (Extra Trees Without avg_temp)** |
|---|---|---|
| $R^2$ Score | 0.9905 | 0.9777 |
| MAE | 0.2577 | 4172.25 |
| RMSE | 0.3762 | 12705.33 |

*Fig 18.Result Metrics*

### 6.3.3 Explainable AI:

To make the models interpretable, we employed SHAP (Shapley Additive Explanations) values to explain how different features contributed to the model’s predictions. SHAP helps to break down the prediction for an individual instance, showing the impact of each feature on the final prediction. This enhances transparency and trust in the model's predictions.

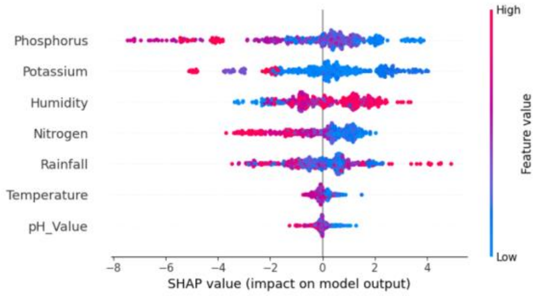



*Fig 19. SHAP applied for crop recommendation*

To enhance the interpretability of the machine learning models used in this project, explainable AI techniques such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations) were utilized. SHAP, grounded in game theory, assigns each feature an importance value for a particular prediction, offering a consistent and global understanding of feature influence. The SHAP summary plot above illustrates how features like Phosphorus, Potassium, and Humidity significantly impact the model's output, with the color gradient representing the feature value (from low in blue to high in pink) and the position on the x-axis showing whether the impact is positive or negative. For instance, high values of Phosphorus and Potassium tend to push the prediction towards specific crop recommendations. On the other hand, LIME provides local explanations by approximating the complex model with an interpretable surrogate model around a single prediction. This helps in understanding why the model made a certain prediction for an individual input. Both SHAP and LIME contribute to building trust, ensuring transparency, and supporting decision-making by offering insights into how different features affect the model's behavior.

## 7. Discussion

The Agrolytics system demonstrates robust performance, particularly with the implementation of ensemble and data preprocessing techniques:

Extra Trees Classifier showed outstanding results by effectively managing complex, non-linear relationships and handling noise in the dataset. Its inherent randomness and averaging mechanism contributed to reduced variance and high accuracy.

SMOTE (Synthetic Minority Over-sampling Technique) successfully mitigated the issue of class imbalance. This enhanced the model’s ability to learn from minority class instances, leading to improved classification performance.

Correlation Analysis played a crucial role in feature selection. By eliminating highly correlated and redundant variables, the regression model was streamlined for greater efficiency, reducing the risk of overfitting and improving generalization.

Key Insights: Rainfall and temperature emerged as the most influential predictors for both crop recommendation and yield estimation tasks. This emphasizes the importance of environmental factors in agricultural decision-making.

## 8. Conclusion

The Agrolytics system exemplifies the potential of machine learning in transforming agriculture through intelligent, data-driven insights. By achieving high accuracy in both crop recommendation and yield prediction, the system serves as a practical decision support tool for farmers. Its deployment through a user-friendly web interface ensures accessibility and usability, enabling even non-technical users to benefit from advanced analytics to boost productivity and sustainability.

## 9. Future Work

To further enhance the Agrolytics platform and extend its impact, several future directions are proposed:

- Integration of Satellite Imagery & NDVI: Utilizing satellite data and vegetation indices such as NDVI can significantly improve spatial precision and monitoring capabilities for crop health.
- Real-time Weather Data Integration: Connecting the system to live weather APIs would enable dynamic, location-specific crop recommendations based on current environmental conditions.
- Incorporation of Deep Learning: Advanced models such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks can be explored for tasks like image-based crop classification and time-series yield forecasting.
- Mobile Application Development: A lightweight mobile app with offline functionality will increase reach, particularly in rural areas with limited internet access.

## References


1. Iniyan, S., Varma, V.A. and Naidu, C.T., 2023. Crop yield prediction using machine learning techniques. *Advances in Engineering Software*, *175*, p.103326.

2. Van Klompenburg, T., Kassahun, A. and Catal, C., 2020. Crop yield prediction using machine learning: A systematic literature review. *Computers and electronics in agriculture*, *177*, p.105709.

3. Bhola, A. and Kumar, P., 2024. Farm-Level Smart Crop Recommendation Framework Using Machine Learning. *Annals of Data Science*, pp.1-24.

4. Kathiria, P., Patel, U., Madhwani, S. and Mansuri, C.S., 2023, May. Smart crop recommendation system: A machine learning approach for precision agriculture. In *Machine Intelligence Techniques for Data Analysis and Signal Processing: Proceedings of the 4th International Conference MISP 2022, Volume 1* (pp. 841-850). Singapore: Springer Nature Singapore.

5. Rani, S., Mishra, A.K., Kataria, A., Mallik, S. and Qin, H., 2023. Machine learning-based optimal crop selection system in smart agriculture. *Scientific Reports*, *13*(1), p.15997.