SDG 2: Predicting Agricultural Yield Using Machine Learning(Contributing to Sustainable Agriculture and Zero Hunger)

Symbiosis Institute Of Technology, Nagpur

Student names-
Tanya Singh
Nakshatra Sawarkar

Problem Statement

- **Brief Overview:**
- Our project aims to analyze and predict crop yield (Yield_tons_per_hectare) based on multiple variables such as environmental factors (Rainfall, Temperature, Weather), farming inputs (Soil_Type, Fertilizer_Used, Irrigation_Used), and time to harvest.
- This analysis can guide effective agricultural practices by helping farmers and policymakers make data-driven decisions. Our case study aligns strongly with SDG Goal 2: Zero Hunger by focusing on optimizing agricultural productivity and resource management to ensure food security.

Source : www.freepik.com/

Problem Statement

- **Key Objectives:**

1.Crop Yield Prediction: Develop machine learning model to predict crop yield (Yield_tons_per_hectare) based on various environmental, agricultural factors.

2. Regional Insights for Precision Agriculture: Tailor recommendations for specific regions and soil types to enhance local productivity and promote sustainable agriculture.

3. Weather and Climate Adaptation: Use weather and climate data to guide farming strategies and help farmers adapt to changing environmental conditions.

Source : www.freepik.com/

Dataset Overview(Optional)

The dataset is publicly available on Kaggle and is designed for crop yield prediction.

Size-

Number of Records: 1,000,000 entries.

Number of Features (Columns): 10 columns capturing various factors affecting crop yield.

Key Features-

Region: Geographic area where the data is collected.

Soil_Type: Classification of soil types.

Crop: Type of crop being cultivated (e.g., rice, wheat, maize).

Rainfall_mm: Total rainfall received in millimeters.

Temperature_Celsius: Average temperature during the growing season.

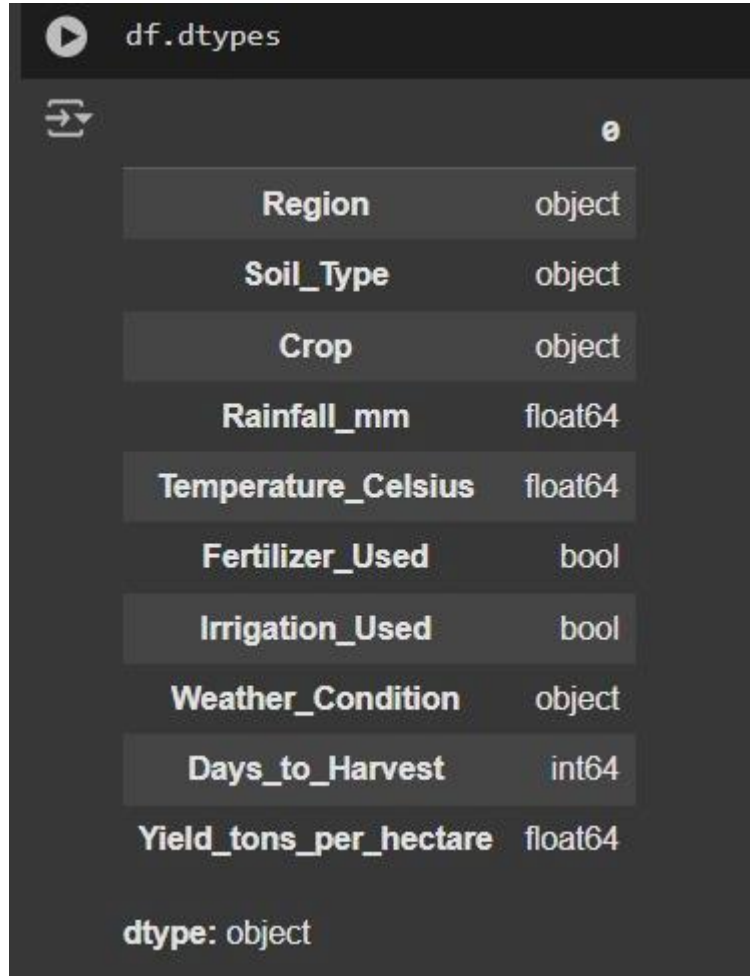
Fertilizer_Used: Quantity or type of fertilizer applied.

Irrigation_Used: Whether irrigation is used (Yes/No).

Weather_Condition: General weather patterns (e.g., sunny, rainy, cloudy).

Days_to_Harvest: Time taken for crops to reach maturity.

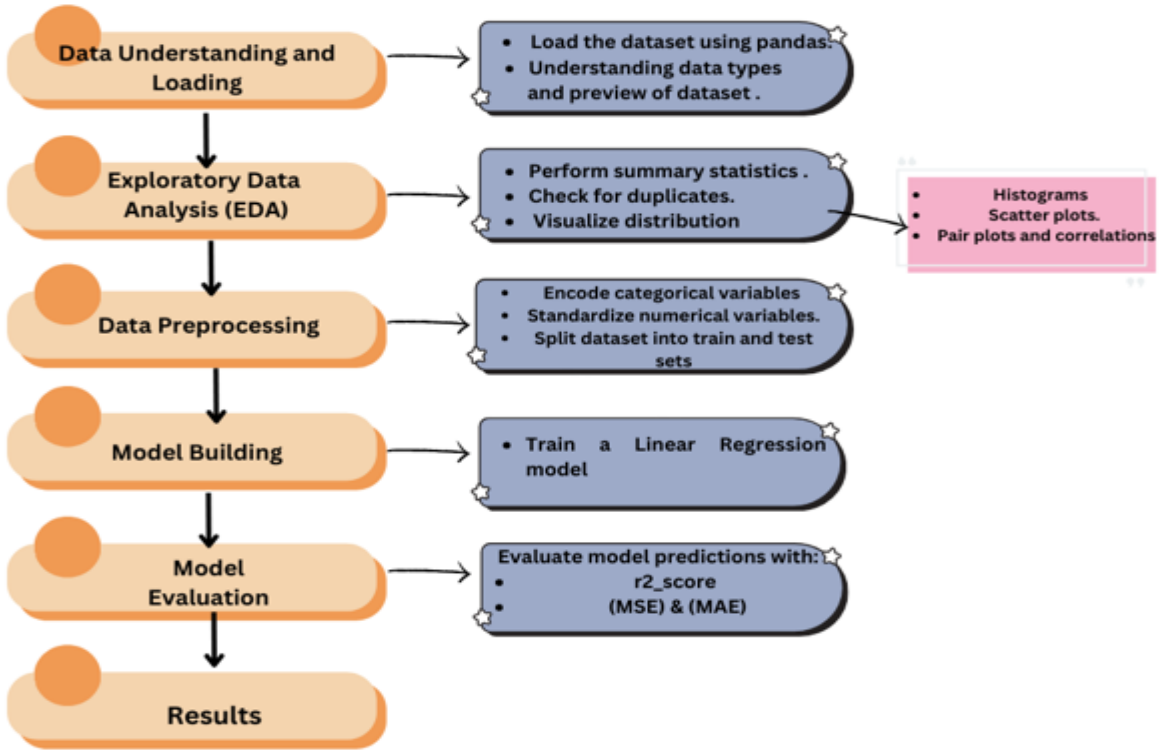
Yield_tons_per_hectare: The **target variable**, representing crop yield in tons per hectare



Source : www.freepik.com/

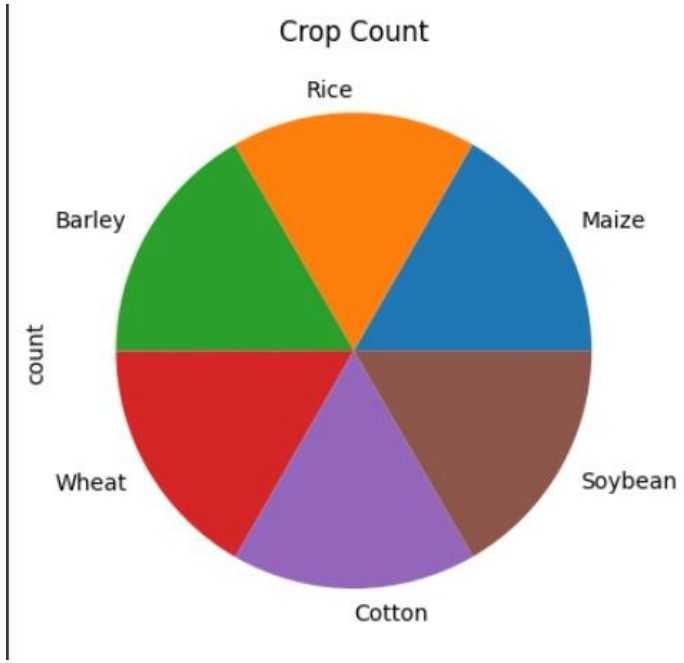
Methodology

- Approach:

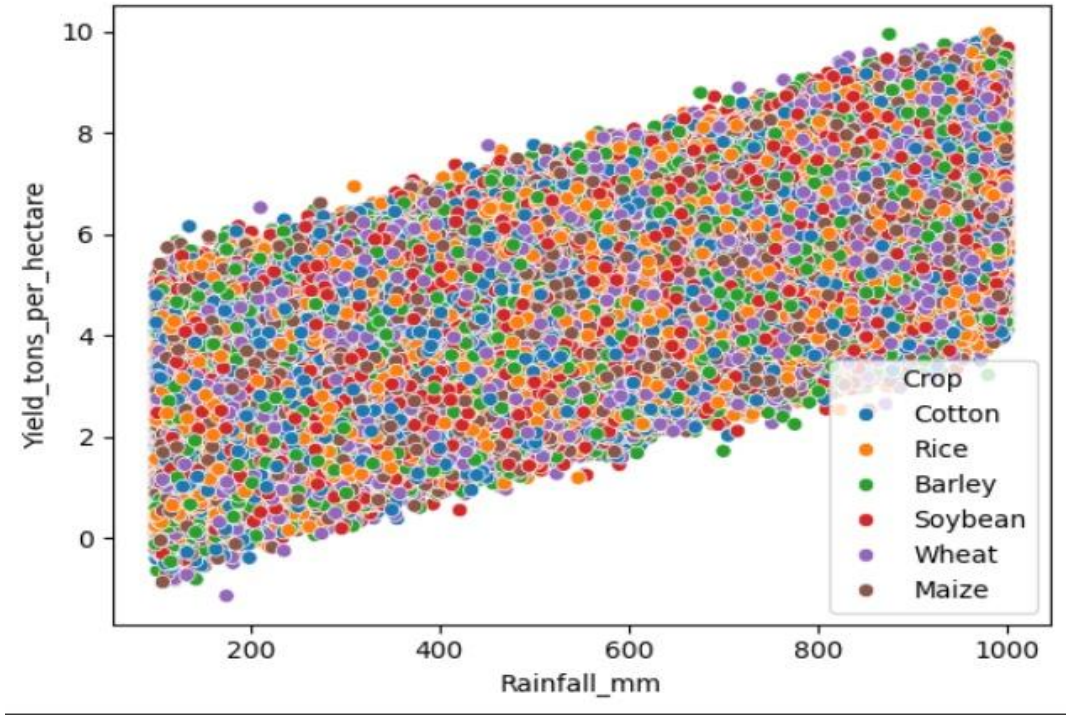


Source : www.freepik.com/

Methodology



The pie chart shows the distribution of crops in the dataset, highlighting Maize and Rice as the most prevalent.



- **Positive Correlation:** Higher rainfall leads to higher crop yields.
- **Yield Range:** Mostly between **2–8 tons/hectare**.
- **Crop Spread:** All crops are evenly distributed across rainfall and yield.

Source : www.freepik.com/

Methodology

Algorithms Used:

We have chosen Linear Regression as the primary algorithm for predicting crop yields. Linear Regression is chosen for its simplicity and effectiveness in modeling the relationship between dependent and independent variables. It helps understand how various factors like rainfall, temperature, and soil type influence crop yield.

To evaluate the performance of the model, the following metrics were employed:

- **R^2 Score**: Measures the proportion of the variance in the target variable that is predictable from the features.
- **Mean Squared Error (MSE):** Evaluates the average squared difference between the predicted and actual values.
- **Mean Absolute Error (MAE):** Measures the average magnitude of prediction errors, providing an easily interpretable metric.

Source : www.freepik.com/

Conclusion

- Summary:
- The case study successfully demonstrates the application of Linear Regression for predicting crop yields based on key agricultural and environmental factors.
- By leveraging features like Rainfall, Temperature, and soil_type, the model achieves an impressive **R^2 Score of 0.9127**, indicating a strong ability to explain crop yield variability.
- The **low MSE and MAE of 0.2512 and 0.4001** further validate its reliability in making accurate predictions.
- This approach highlights the potential of data-driven methods to support agricultural planning, optimize resource usage, and enhance productivity. The solution is particularly effective in identifying key factors influencing yields, making it a valuable tool for both farmers and policymakers aiming to improve agricultural outcomes.

Source : www.freepik.com/

Conclusion

Future Work:

INCORPORATION OF ADDITIONAL FEATURES

- We can include factors like pest attacks, disease prevalence, and crop management practices to improve model accuracy.

SCALING & DEPLOYMENT

- We can deploy the model as a user-friendly application for farmers and policymakers, offering actionable insights to enhance crop planning and resource management.

Source : www.freepik.com/

References

Kaggle Dataset Link-

https://www.kaggle.com/datasets/samuelotiattakorah/agriculture-crop-yield

Thank You