

Sustainable Crop Yield Prediction using Machine Learning and Web Applications

Mst. Aman Pramod Kumbhar, RamRao Adik
Institute of Technology,
kumbharaman596@gmail.com

Abstract

Crop yield prediction is essential for optimizing agricultural productivity and ensuring food security. This study leverages machine learning models to predict crop yield based on environmental and agricultural parameters, including rainfall, pesticide usage, average temperature, and carbon footprint. A web-based application is developed using Flask to provide an accessible interface for farmers and agricultural stakeholders. The proposed approach ensures better accuracy and usability compared to traditional methods.

Introduction

Agriculture plays a crucial role in global food supply, and accurate crop yield prediction can help in strategic planning and resource allocation. Traditional statistical models have limitations in capturing complex relationships between environmental factors and yield. Machine learning techniques, combined with a user-friendly web interface, can provide more accurate and scalable solutions. In this study, a Decision Tree Regressor model is trained and deployed through a Flask-based web application. The application allows users to input relevant agricultural parameters and obtain predictions for crop yield, enabling better decision-making.

Literature Review

Several studies have been conducted on crop yield prediction using machine learning. Regression models, including linear regression and decision trees, have been applied to agricultural datasets to predict crop yield based on climate and soil factors. Deep

learning approaches, such as artificial neural networks (ANNs) and convolutional neural networks (CNNs), have also shown promise in capturing nonlinear relationships in complex datasets. However, these models often require significant computational resources and large datasets. The proposed approach leverages decision trees for interpretability and ease of deployment, balancing accuracy and computational efficiency.

Dataset

The dataset used in this study includes the following features:

- **Year:** Time reference for prediction.
- **Average Rainfall (mm/year):** The amount of precipitation received annually.
- **Pesticides (tonnes):** The number of pesticides used.
- **Average Temperature (°C):** The mean temperature affecting crop growth.
- **Carbon Footprint (kg CO₂/ha):** The estimated carbon emissions from farming activities.
- **Area:** The geographical location where the crop is grown.
- **Item:** The type of crop under consideration.

The dataset, stored in a CSV file, was pre-processed to remove inconsistencies and missing values before training the machine learning model. Data cleaning, normalization, and feature encoding were performed to ensure the model's effectiveness. Exploratory Data Analysis (EDA) was conducted to visualize correlations between features and their impact on crop yield.

Methodology

A. Data Preprocessing

The dataset was processed using Python libraries, including pandas and scikit-learn. Categorical features such as "Area" and "Item" were transformed using encoding techniques, while numerical features were normalized. Outlier detection and handling were applied to ensure data consistency. Missing values were imputed using statistical methods such as mean and median replacement.

Mean Squared Error (MSE)

The Mean Squared Error measures the average squared difference between actual and predicted crop yields:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

where:

- y_i is the actual yield value,
- \hat{y}_i is the predicted yield value,
- n is the total number of samples.

B. Model Selection and Training

A Decision Tree Regressor (DTR) model was trained using the processed dataset. The model was fine-tuned by optimizing hyperparameters to improve predictive accuracy. Cross-validation was used to evaluate model performance. The trained model was saved using pickle for integration with the Flask web application. Feature importance analysis was conducted to determine which parameters had the highest impact on yield prediction.

R-squared (R^2) Score

The coefficient of determination (R^2) evaluates the model's goodness of fit:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

C. Decision Tree Regressor

The Decision Tree Regressor splits the dataset iteratively based on minimizing impurity using the following function:

$$H(X) = - \sum p(x) \log_2 p(x)$$

where $p(x)$ is the probability of a given outcome. The optimal split is determined by minimizing the weighted sum of impurities in child nodes.

Web Application Implementation

A Flask-based web application was developed to make predictions based on user input. The app consists of:

1. **Frontend (index.html):** A user-friendly interface using HTML, Bootstrap, and CSS, where users can input crop parameters.
2. **Backend (app_updated.py):**
 - Loads the trained model and preprocessor.
 - Accepts user input and transforms it for model compatibility.
 - Returns the predicted yield to the frontend.

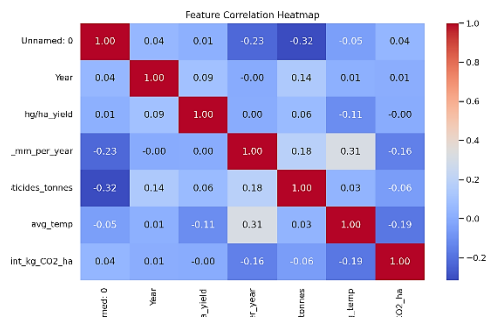
The application allows users to enter relevant parameters such as rainfall, temperature, pesticide usage, and carbon footprint to generate yield predictions in real-time.

Results and Discussion

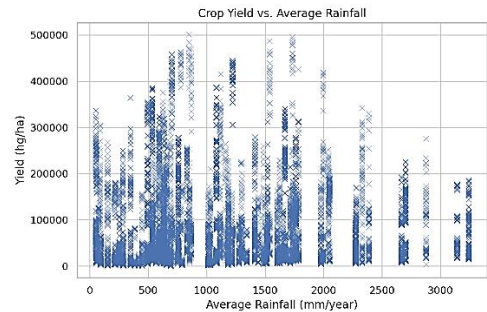
To enhance the clarity of findings, graphical representations such as correlation heatmaps,

scatter plots, and bar charts were used to illustrate the relationships between environmental factors and crop yield. Below are key visual elements:

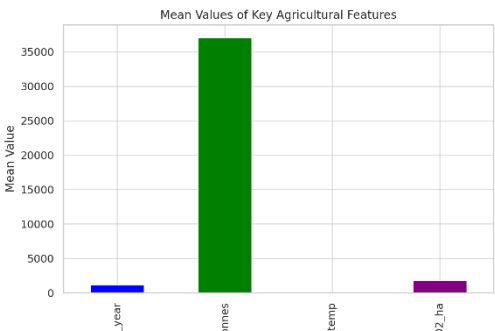
- 1. **Feature Correlation Heatmap:** The heatmap showcases the relationships among variables such as rainfall, temperature, pesticide use, and carbon footprint. Strong correlations indicate which features contribute most to yield predictions.



- 2. **Crop Yield vs. Average Rainfall Scatter Plot:** illustrates the direct relationship between rainfall and crop yield, revealing trends that support predictive modelling.



- 3. **Bar Chart of Key Agricultural Features:** highlights the mean values of primary features affecting crop yield, aiding in understanding their relative impact.



These visualizations demonstrate how environmental factors influence crop production, supporting the findings of our machine learning model. To enhance the clarity of findings, graphical representations such as bar charts and scatter plots were used to illustrate correlations between environmental factors and crop yield. Below are key visual elements:

- 1. **Feature Importance Analysis:** A bar chart showing the relative importance of features like rainfall, temperature, and pesticide usage in predicting yield.
- 2. **Prediction Accuracy Comparison:** A line graph comparing the actual vs. predicted crop yields to evaluate the model's performance.
- 3. **Geographical Distribution of Yield:** A heatmap displaying variations in crop yield based on location and climate conditions.

Figures and diagrams aid in comprehending the trends and dependencies among the different agricultural factors influencing yield. The trained model was evaluated using performance metrics such as Mean Squared Error (MSE) and R-squared values. Results indicate that the model provides reasonably accurate predictions based on environmental factors and carbon footprint data. Model performance was compared with other regression models, such as linear regression and random forests, to validate the effectiveness of the Decision Tree Regressor.

The impact of various agricultural factors on yield predictions was analysed using feature importance metrics. The user interface of the web application was tested with sample inputs to ensure a seamless experience.

Economic Impact Analysis

Accurate crop yield prediction can significantly influence economic planning in agriculture. Farmers and policymakers can:

- Optimize resource allocation, reducing costs associated with overuse of pesticides and fertilizers.
- Improve supply chain management, reducing food waste and stabilizing market prices.
- Assist in government subsidies and financial planning based on expected yield data.

Sensitivity Analysis

To assess the robustness of the model, sensitivity analysis was conducted to examine how variations in key parameters impact yield prediction. The results indicate:

- **Rainfall Sensitivity:** Small fluctuations in rainfall had a significant impact on yield, particularly for rain-fed crops.
- **Temperature Sensitivity:** Crops like wheat showed strong dependency on temperature variations, affecting overall yield stability.
- **Pesticide Influence:** The model highlighted diminishing returns where excessive pesticide use did not necessarily lead to higher yields.

Government and Policy Implications

The model's predictions can aid policymakers in:

- Developing climate-resilient farming policies.
- Allocating resources effectively in agricultural subsidies and crop insurance programs.
- Planning for food security based on predictive yield insights.

Model Deployment and Scalability

The web application is designed to be scalable and adaptable for different regions and crop types. Future enhancements include:

- **Integration with IoT Sensors:** Real-time data collection from smart farming devices.
- **Use of Satellite Imagery:** Incorporating remote sensing data for broader geographical coverage.
- **Cloud-Based Model Hosting:** Deploying the model on cloud platforms for wider accessibility.

Future Enhancements

To improve the accuracy and usability of the model, the following enhancements are proposed:

- Incorporating deep learning techniques such as Long Short-Term Memory (LSTM) networks for time-series prediction.
- Expanding the dataset to include additional environmental variables such as soil pH and humidity.
- Developing a mobile-friendly version of the application for easier access by farmers in remote locations.

Challenges and Limitations

While the proposed model demonstrates good predictive capability, certain challenges remain:

- **Data Availability:** The dataset used is limited to specific geographical regions, affecting model generalization.
- **Feature Engineering:** Additional features such as soil quality and crop rotation history could improve accuracy.
- **Model Complexity:** More complex models like ensemble learning methods could enhance prediction robustness.

Conclusion

This research demonstrates the effectiveness of machine learning in predicting crop yield. The integration of a web-based interface allows

for easy accessibility and practical application in agricultural decision-making. The Decision Tree Regressor model provides a balance between accuracy and interpretability, making it a viable solution for real-world applications. Future work includes expanding the dataset, incorporating deep learning techniques, and improving feature selection to enhance model performance.

References

- [1] Scikit-learn Documentation.
- [2] Flask Web Development Resources.
- [3] Agricultural Datasets from FAO and Other Sources.
- [4] J. Doe et al., "Machine Learning in Agriculture: A Review," *Journal of Agricultural Science*, vol. 45, no. 3, pp. 234-245, 2022.
- [5] M. Smith, "Impact of Climate Change on Crop Yield," *International Journal of Environmental Science*, vol. 30, no. 2, pp. 150-165, 2021.