Optimizing Agricultural Resource Management Through Machine Learning

BU425 – Professor Michael Pavlin

Yash Harshal Barve - 211544920
Amy Freij Camacho – 211912090
Wang Qianyi – 169036969
Nicholas Rebello – 210572430

Project Overview

Agriculture faces a critical challenge in managing crop yield. This challenge stems from unpredictable environmental conditions, inefficient resource use, and economic constraints. Machine learning offers a data-driven means to optimize resource allocation and enhance decision-making, enabling farmers to improve productivity while minimizing waste.

This project explores how machine learning can address these challenges by investigating key agriculture datasets. These datasets have environmental factors such as rainfall, temperature, pesticide usage, and historical crop yield data. Leveraging models like Lasso Regression, Random Forest, and LSTM, the team aimed to predict crop yields and optimize farming practices. Specifically, the models focus on resource allocation and risk mitigation.

Through careful training and testing of these models, we seek to demonstrate how predictive analytics can revolutionize agricultural resource management and risk mitigation, ultimately contributing to increased yields, lower costs, and improved resilience against climate condition variabilities.

Business Problem & Motivation

The agricultural sector faces significant challenges in optimizing crop yields due to the uncertainty caused by environmental and agronomic variability. Farmers often struggle to allocate resources, which results in crop loss due to drought or nutrient deficiencies, poor decision-making on planting schedules, and increased costs from overuse of water, fertilizers, and pesticides. All of these elements contribute to a reduction in profitability. These challenges not only impact the financial sustainability of farmers but also affect the broader agricultural supply chain, making it crucial to find solutions that enable better resource allocation, and effective risk management, and ensure that food remains available for families.

The motivation for addressing these challenges comes from the need to improve agricultural productivity while minimizing resource wastage. Farmers can reduce the risks associated with environmental uncertainties and agronomic factors by improving resource allocation, which increases crop yields and profitability. Using these more efficient practices also promotes food security and contributes to environmental sustainability, which ensures that farmers can meet the demands of the future without using up essential resources.

Dataset & Methodology

Resource Allocation and Yield Prediction:

To predict crop yield and identify key drivers of performance, we used a combination of Random Forest for feature selection and Lasso Regression for model development. We began by loading and preprocessing the dataset, which included features such as average rainfall, average temperature, pesticide usage, and crop types across different regions. Categorical variables (Area, Item) were converted into numerical format using one-hot encoding to ensure compatibility with machine learning algorithms. Numerical features (average_rain_fall_mm_per_year, pesticides_tonnes, avg_temp) were standardized using StandardScaler to normalize the range of values and improve model stability. To enhance the model's ability to capture interactions and efficiency, we engineered additional features: rainfall_temp_interaction to reflect how rainfall and temperature jointly impact yield, yield_per_rainfall and yield_per_pesticide to represent resource efficiency.

A Random Forest Regressor was used to estimate feature importance and select the most relevant predictors. We retained those with importance > 0.01, including crop type, rainfall, temperature, pesticide use, and derived efficiency features. Using these selected features, we trained a Lasso regression model with hyperparameter tuning via GridSearchCV. Lasso Regression provides a sparse, interpretable model while preventing overfitting. The best regularization strength (alpha) was found to be 0.001.

By combining these techniques, we developed a highly interpretable and robust model that provides actionable insights into the factors influencing agricultural yield. The model not only predicts crop yield with accuracy but also identifies key drivers of performance, such as rainfall, temperature, pesticide use, and crop type. This enables a deeper understanding of how changes in these variables impact yield, offering data-driven guidance for optimizing resource allocation.

Risk Mitigation:

This section focuses on predicting crop yields and evaluating risk factors for agricultural production. The process began with data collection and preprocessing to ensure data completeness. Missing values were addressed using both forward and backward filling methods. A log transformation was applied to correct the left-skewed distribution of the crop yield variable. Additionally, numerical features were normalized using a MinMaxScaler to bring them to a consistent scale, which enhanced both the quality of the data and the performance of the machine learning models.

Subsequently, feature engineering was performed, where we created derived features such as rolling averages, year-on-year changes, and lag features to improve the predictive power of the models. Two machine learning models were employed: Random Forest Regressor and LSTM. The Random Forest Regressor was used to analyze feature importance, and hyperparameter tuning was conducted using GridSearchCV to optimize model parameters.

An LSTM model was implemented to capture temporal dependencies in sequential data. Sequences of five years were used for training, validation, and testing. The model consisted of two LSTM layers with dropout regularization and a dense output layer for regression. The test set achieved an MSE of approximately 1.2996.

To assess risk, threshold values for each feature were established based on historical data. These thresholds alongside the LSTM model's yield predictions were used to develop a risk score function. Finally, the distributions of predicted yields and risk scores were visualized using histograms and scatter plots to facilitate interpretation and analysis.

Results & Key Findings

Resource Allocation and Yield Prediction:

Our analysis shows that machine learning models such as Random Forest and Lasso Regression can effectively identify and quantify the influence of key agricultural features on crop yield. The Random Forest model showed that crop type Item_Potatoes (0.3693), yield per rainfall (0.2615), and yield per pesticide (0.1495) were the most impactful features, emphasizing the role of both input efficiency and crop selection. While potatoes and sweet potatoes exhibited the strongest positive effects on yield, environmental inputs such as rainfall, pesticide use, and average temperature, though less dominant, still contribute meaningfully to yield variation. Rainfall, pesticide use, and average temperature were intentionally selected as input features as they are measurable, directly impact crop development, and can be actively managed by farmers. Although crop type is not a controllable input, it was included in the model as a categorical feature to capture how different crops respond under varying environmental conditions. The Lasso Regression model, built using the most impactful features, achieved an R^2 score of 0.4914 and an RMSE of approximately 60,738. These error margins are acceptable given the high variability in agricultural data, where external factors like soil quality, disease, and farming practices introduce significant noise. The model explains about 49.1% of the variance in crop yield, which is considered reasonable in real-world scenarios. By highlighting inputs with the greatest influence on yield, the model supports more efficient resource allocation, helping farmers prioritize inputs that drive the highest returns. The model provides a practical, interpretable framework for improving resource allocation, optimizing crop decisions, and mitigating climate-related risks.

Our model revealed that rainfall has a strong positive impact, with yield increasing steadily as rainfall rises from 800 mm/year to 1800 mm/year. In contrast, average temperature showed a negative effect, suggesting heat stress reduces productivity. Higher temperatures from 15°C to 35°C drastically reduce yield from 50,000 hg/ha to 20,000 hg/ha. Pesticide use shows only a minimal positive effect, which indicates diminishing returns beyond a certain application level. Overall, rainfall is a key driver of productivity, while high temperatures pose a major challenge, and pesticide use has a limited impact.

Risk Mitigation:

Our analysis indicates that most data points in the test dataset, comprising approximately 3,700 points, exhibit risk scores ranging between 0.4 and 0.5, suggesting a moderate risk level. A subset of data points shows risk scores as low as 0.1, while only a few points exceed a risk score of 0.65. This pattern points to several false positives, which can be attributed to both the limitations of the risk_score() function and issues within the dataset, such as missing values and inaccuracies.

Additionally, the mean difference between the actual and predicted yields in the test dataset is approximately -6000. Considering the scale of the values and the dataset size, this result is quite satisfactory. The Root Mean Squared Error (RMSE) for the LSTM model is 1.1400, demonstrating strong performance. This outcome was achieved by applying several optimization techniques, including log transformation for dataset smoothing, dropout regularization, and using the Adam optimizer in the LSTM model.

To further enhance model interpretability, we employed Random Forest Regression to identify the most influential features within the dataset. With this, we selected the top five features, collectively accounting for nearly 80% of the total feature importance.

Recommendations & Business Implications

Based on our analysis, we recommend farmers adopt a resource efficiency framework to optimize crop yields and reduce resource waste. Key variables such as yield per rainfall and yield per pesticide demonstrate that resource efficiency is proportionally critical to the amount of inputs applied. Agribusinesses should establish Key Performance Indicators (KPIs) to measure yield per input unit.[1] Additionally, adopting variable-rate application technologies, such as smart sprayers, will allow for more precise input distribution, reducing waste and improving resource efficiency. [2]To address temperature-related yield losses, we recommend farmers focus on irrigation management, select heat-resistant crop varieties, and implement protective practices such as mulching or shading.[3] These strategies will help optimize resource use, enhance productivity, and reduce environmental impact by promoting more sustainable farming practices.

By implementing these strategies, farmers and agribusinesses can boost yields while using fewer resources, which promotes sustainability and profitability. They can save wasteful expenses without compromising productivity by emphasizing input efficiency over quantity. In an industry that is becoming more challenged by resource scarcity and climate variability, these approaches enable agricultural businesses to remain competitive, increase operational resilience, and increase return on investment.

To enhance agricultural productivity and mitigate risks, impacted stakeholders should leverage predictive modelling to optimize resource allocation. By identifying key drivers of crop yield such as rainfall, temperature, and pesticide usage, stakeholders can make data-driven decisions to support and improve yield stability and sustainability. A critical feature of this risk assessment model is that it is dynamic: variables can continuously evolve as more data is incorporated into the dataset, allowing for better adaptability and increased precision of the risk scores.

The model provides a simple function whereby farmers can input key variables like expected rainfall, planned pesticide usage, and average temperature estimations. This approach can predict crop yield and generate suggestions about environmental factors affecting the overall risk. Farmers using this model can create ideal environments, mitigate preventable risks, and improve long-term crop performance.

The risk assessment model allows for early detection of potential yield declines, enabling users to consider strategic interventions such as adjusting irrigation strategies or modifying pesticide applications. Businesses in the agricultural industry can use these insights to develop smarter supply chain strategies, reduce waste, and enhance food security. For future development, stakeholders can refine risk mitigation efforts by integrating these models with real-time weather and soil data, ultimately making agriculture more resilient to climate variability.

Concluding Remarks

By leveraging machine learning, we can help farmers make smarter decisions, boost efficiency, maximize crop yields, and reduce the financial risks of unpredictable weather patterns. While the models provide valuable insights, there is still room for improvement. In the future, there can be a focus on expanding the dataset, using more variables and refining the model’s accuracy. This would help make predictions even more reliable and widely applicable.

Full Dataset Link:
https://www.kaggle.com/datasets/patelris/crop-yield-prediction-dataset

References

Mischke M. 2025 Agriculture & Farm Key Performance Indicators (KPIs). PerformYard. Last updated January 21, 2025. Accessed March 24, 2025. https://www.performyard.com/articles/agriculture-performance-indicators

Sairone. Sairone | Innovating Agriculture with AI/ML-Driven Drones. Sairone. Published December 1, 2024. Accessed March 24, 2025. https://saiwa.ai/sairone/blog/smart-spraying-technology-in-agriculture/

BRANDT Europe. How to protect crops from extreme heat in summer. BRANDT Europe. Published August 3, 2023. Accessed March 24, 2025. https://brandteurope.com/en/how-to-protect-crops-from-extreme-heat-in-summer/