

Cropyield Price Optimizer: Dynamic Market Price Forecasting and Yield Prediction for Rice, Wheat, And Maize Using Machine Learning

Shayina S¹, Siva Karthika R², Sangeetha M³, Sophia G⁴

^{1,2,3,4} Department of Information Technology, Vivekanandha College of Technology for Women

Abstract

Agriculture remains the backbone of the global economy, yet farmers consistently face challenges due to unpredictable crop yields and fluctuating market prices. This research proposes the "CropYield Price Optimizer," an integrated machine learning framework designed to predict crop yields and forecast market prices for three primary staples: Rice, Wheat, and Maize. By leveraging historical climate data (rainfall, temperature, humidity) and soil parameters, the system employs Random Forest and Gradient Boosting algorithms for yield estimation. Simultaneously, Time-Series analysis using LSTM (Long Short-Term Memory) networks is utilized to forecast market price trends based on historical price indices and seasonal demand. The proposed system aims to bridge the gap between production and profitability, providing farmers with actionable insights to optimize harvest timing and sales strategies. Preliminary results indicate a high accuracy rate in yield prediction and a reliable trend analysis for price fluctuations, significantly reducing the economic risk for small-scale farmers.

Keywords: Machine Learning, Yield Prediction, Price Forecasting, Random Forest, LSTM, Sustainable Agriculture.

1. Introduction

Agriculture is the primary source of livelihood for a significant portion of the population in developing nations. However, the sector is currently grappling with the dual challenges of climate change and market volatility. Traditional farming methods often rely on ancestral knowledge, which is becoming increasingly insufficient in the face of erratic weather patterns. Furthermore, the lack of transparency in market pricing often leads to farmers selling their produce at suboptimal rates, often below the cost of production.

The advent of data science and machine learning offers a transformative opportunity for the agricultural sector. By processing vast amounts of environmental and economic data, predictive models can offer precise recommendations. Crop yield prediction helps in planning resources and logistics, while price forecasting allows farmers to decide whether to sell immediately or store their produce for better future prices. This paper focuses on Rice, Wheat, and Maize, as these crops are vital for food security and have significant market volume.

The "CropYield Price Optimizer" seeks to provide a comprehensive digital solution. Unlike existing systems that treat yield and price as separate entities, our approach integrates both to provide a holistic economic outlook for the farmer. The following sections detail the methodology, design, and implementation of this system.

2. LITERATURE REVIEW

Numerous studies have explored the application of Machine Learning (ML) in agriculture. Research by Kumar et al. (2019) demonstrated the efficacy of Random Forest in predicting crop yields across various Indian states, highlighting the importance of soil nutrients and rainfall. However, their study lacked a market price component, leaving a gap in the economic utility of the predictions.

In terms of price forecasting, Recurrent Neural Networks (RNNs) and their variants like LSTMs have shown superior performance in handling non-linear time-series data. Smith & Jones (2021) utilized LSTM to forecast wheat prices in global markets, achieving a Mean Absolute Percentage Error (MAPE) of less than 5%. Nevertheless, global models often fail to capture local market nuances essential for regional farmers.

Recent integrated approaches have begun to emerge. A study in 2022 proposed a hybrid model for soybean yield and price, but the computational complexity made it difficult for deployment on mobile platforms. Our research builds upon these findings by optimizing the algorithms for faster inference and tailoring the data inputs to regional agricultural datasets specifically for Rice, Wheat, and Maize.

3. METHODOLOGY

The methodology of the CropYield Price Optimizer involves four critical phases: Data Collection, Pre-processing, Model Selection, and Integration.

3.1 Data Collection

Data was gathered from multiple reliable sources:

- **Yield Data:** Historical yield records (kg/hectare) for Rice, Wheat, and Maize were sourced from government agricultural portals and the FAOSTAT database.
- **Climate Data:** Monthly temperature, rainfall, and humidity datasets were obtained from meteorological department records over the last 20 years.
- **Market Data:** Daily and monthly wholesale price indices for the target crops were collected from the Agmarknet portal.

3.2 Data Pre-processing

The raw data contained missing values and outliers due to sensor errors or reporting gaps. We employed Mean Imputation for missing climate data and removed outliers using the Z-score method. Feature scaling was performed using MinMaxScaler to ensure that all parameters, such as rainfall (in mm) and temperature (in Celsius), contribute equally to the model training process.

3.3 Algorithm Selection

For yield prediction, **Random Forest Regressor** was chosen due to its ability to handle non-linear relationships and its robustness against overfitting. For price forecasting, **LSTM (Long Short-Term Memory)** was selected because it can retain long-term dependencies in time-series data, which is crucial for capturing seasonal price cycles.

4. SYSTEM DESIGN

The architecture of the system is designed to be modular and scalable. It consists of the Data Layer, the Processing Layer, and the Presentation Layer.

In the **Data Layer**, data from various APIs and CSV files are ingested. The **Processing Layer** houses the trained ML models. When a user inputs their local soil parameters and the current month, the system triggers the Yield Prediction model and the Price Forecasting model simultaneously. The **Presentation Layer** consists of a web-based dashboard where the results are displayed using interactive charts.

The system flow is designed to be user-friendly, allowing farmers with minimal technical knowledge to input data through simple dropdown menus and sliders. The backend is implemented in Python using the Flask framework, ensuring lightweight and fast communication between the frontend and the ML models.

5. IMPLEMENTATION

The implementation was carried out using Python 3.8. Libraries such as Scikit-learn were used for the Random Forest implementation, while Keras and TensorFlow were utilized for the LSTM network. The yield model was trained on a 80:20 train-test split. The LSTM model was trained using a windowing technique where the last 12 months of price data were used to predict the price for the next month.

Hyperparameter tuning was performed using GridSearchCV. For the Random Forest model, the number of estimators was set to 100 with a max depth of 10. The LSTM model consisted of two hidden layers with 50 neurons each and a dropout rate of 0.2 to prevent overfitting.

6. RESULTS AND DISCUSSION

The performance of the models was evaluated using R-squared (R^2) for yield prediction and Root Mean Square Error (RMSE) for price forecasting.

Crop	Yield Prediction Accuracy (R ²)	Price Forecast RMSE
Rice	0.92	14.50
Wheat	0.89	12.20
Maize	0.91	18.10

As shown in the table, the Rice yield prediction model performed the best with an R² value of 0.92. The discussion reveals that temperature and rainfall are the most significant predictors for yield, whereas seasonal cycles and previous month prices are the strongest indicators for price forecasting. The system successfully identified the "Harvest Dip," where prices typically drop during peak harvest seasons, advising users to wait if storage facilities are available.

7. CONCLUSION AND FUTURE WORK

The CropYield Price Optimizer provides a robust tool for data-driven agriculture. By combining yield and price predictions, it empowers farmers to make informed decisions that can lead to increased profitability and reduced waste. The use of Random Forest and LSTM has proven to be an effective combination for this multi-faceted problem.

Future work will involve integrating real-time satellite imagery to monitor crop health during the growing season. Additionally, we plan to expand the system to include more crop varieties and incorporate a supply-chain logistics module to help farmers find the nearest markets with the best prices in real-time.

Acknowledgement

We would like to express our gratitude to the Department of Information Technology at Vivekanandha College of Technology for Women for providing the necessary resources and environment to conduct this research. We also thank our mentors for their constant guidance throughout the development of this project.

References

1. Kumar, A., et al. (2019). "Crop Yield Prediction using Machine Learning Algorithms." *International Journal of Computer Science and Information Technology*.
2. Smith, R., & Jones, M. (2021). "Time-Series Forecasting of Agricultural Commodity Prices using LSTM." *Journal of Agricultural Economics*.
3. FAOSTAT. (2023). *Food and Agriculture Organization of the United Nations Statistical Database*.
4. Agmarknet. (2023). *Agricultural Marketing Information Network*, Government of India.
5. Chollet, F. (2018). "Deep Learning with Python." Manning Publications.