PROFIT-AWARE CROP FORECASTING USING MACHINE LEARNING AND MARKET DATA

Dr.K. Sreenivasa Reddy[1,*] Aaji Babu Dhangar[2] B.Nagalaxmi[2] G. Karthik[2] Ch. Vishwa Chandu[2]
[1]Assistant Professor, Department of CSE(DS), TKR College of Engineering, Meerpet, Telangana 500097
[2]B.Tech (Scholar), TKR College of Engineering, Meerpet, Telangana 500097
*Correspondence: ksrinivasreddy@tkrcet.co

ABSTRACT

Agriculture remains the backbone of India's economy, yet farmers often face the dual uncertainty of unpredictable crop yields and fluctuating market prices. While several existing solutions address yield forecasting, they fail to incorporate profitability, leaving farmers vulnerable to income instability. This project proposes a Profit-Aware Crop Forecasting System that integrates machine learning–based yield prediction models (Random Forest, XGBoost, Linear Regression, SVR) with market price forecasting models (ARIMA, Prophet, LSTM). The system estimates profitability by combining predicted yield with market price data, deducting input costs, and recommending the most profitable crop. The goal is to empower farmers with reliable, data-driven decisions for sustainable farming and improved income stability..

Keywords: Crop Forecasting, Profit Prediction, Pesticide Cost, Market Price, Yield Estimation, Economic Planning, Smart Agriculture

1. INTRODUCTION

Agriculture has always been one of the most fundamental sectors of human society, supporting both food production and economic development [1]. In countries like India, it is not only the backbone of rural livelihoods but also a key contributor to the national GDP. Despite its centrality, farming continues to be a highly uncertain venture [2-3]. Farmers face risks that arise from natural factors such as unpredictable rainfall, fluctuating temperatures, soil degradation, pest infestations, and from economic conditions such as volatile market prices and rising input costs [4-6]. Traditional approaches to decision-making in farming have mostly relied on personal experience, historical patterns, or local advisory services [7]. While these methods may sometimes help, they lack scientific precision and fail to adapt quickly to changing climatic and economic circumstances.

In recent years, technology and data science have created opportunities to transform agricultural practices. Machine learning (ML) has proven especially powerful because it can analyze vast datasets, recognize hidden patterns, and make predictions with high accuracy [8-10]. In the agricultural domain, ML has been widely applied for tasks such as yield forecasting, crop disease detection, and soil fertility analysis [11]. Similarly, in the field of agricultural economics, statistical and deep learning models have been used to forecast market prices of commodities [12]. However, despite advancements in both directions, most systems continue to treat yield prediction and price forecasting as isolated tasks [13]. This separation limits their practical usefulness to farmers, who are ultimately interested in profitability, not just yield or price in isolation [14].

The proposed project addresses this gap by designing a Profit-Aware Crop Forecasting System that integrates both yield and price predictions into a single framework. The system uses ML algorithms such as Random Forest, XGBoost, Linear Regression, and Support Vector Regression to estimate yields based on agronomic features like soil type, rainfall, and fertilizer usage [15-18]. It simultaneously employs time-series forecasting methods such as ARIMA, Prophet, and LSTM to predict market prices. These outputs are then combined to calculate profitability by subtracting cultivation costs from the product of yield and price [19]. The system finally recommends crops that are expected to maximize profit under current conditions [20]. This approach has the potential to shift farming strategies from traditional yield-centered decision-making to modern profit-centered practices, empowering farmers with actionable

insights that can stabilize their incomes and improve sustainability.

2. PROPOSED SYSTEM

The proposed system represents a significant advancement in agricultural decision-support technology by explicitly focusing on profitability. It is designed as a three-stage framework. The first stage involves yield prediction, where machine learning algorithms such as Random Forest, XGBoost, Linear Regression, and Support Vector Regression are employed. Random Forest is particularly useful because of its ability to handle high-dimensional data and avoid overfitting. XGBoost provides high predictive performance and scalability for large datasets. Linear Regression serves as a baseline model for interpretability, while Support Vector Regression captures non-linear dependencies between environmental factors and crop yield. Together, these models provide robust and accurate yield predictions for different crops.

The second stage involves price forecasting using time-series models. ARIMA is implemented for short-term price predictions where autocorrelation is strong. Prophet is used to account for seasonality effects, such as festive season demand for certain crops. LSTM networks, being capable of handling sequential data and long-term dependencies, are employed for complex price patterns that cannot be captured by traditional models. By integrating multiple forecasting approaches, the system ensures flexibility and accuracy across different crops and time horizons. The third stage involves profit estimation, where outputs from the yield and price models are combined. The expected profit is calculated as the product of predicted yield and predicted price, minus total cultivation costs. Cultivation costs include inputs such as seeds, fertilizers, pesticides, irrigation, and labor. The system ranks available crop options by profitability and presents these recommendations to farmers through a visualization dashboard. This three-stage approach ensures that decisions are not only based on biological feasibility but also on economic viability, thereby directly addressing the needs of farmers.

3. LITERATURE SURVEY

A deeper look at the existing literature provides valuable insights into both the potential and the limitations of current approaches. Sharma et al. (2022) applied Support Vector Machines for early crop yield prediction in India. Their study highlighted the role of machine learning in agricultural forecasting but stopped short of connecting yields with economic returns. Veenadhari et al. (2022) built regression-based models for yield prediction using climatic parameters like rainfall and temperature. Their research emphasized the importance of weather data but once again lacked any integration with price forecasting. Gupta and colleagues (2022) conducted a comparative analysis of Random Forest, Decision Trees, and K-Nearest Neighbors for yield prediction, providing benchmarks for accuracy and interpretability but ignoring market dynamics. Kumar et al. (2022) focused on wheat crop yield forecasting using Decision Trees and Random Forests, producing accurate results for wheat but without applicability to other crops or profitability analysis.

A noteworthy study by Sindhur et al. (2025) proposed a hybrid system that considered both agronomic and economic factors for crop forecasting. While their framework moved closer to the idea of profitability, it was region-specific and lacked deployment at scale. Sharma et al. (2021) in IEEE Access provided a review of machine learning in precision agriculture, identifying challenges such as overfitting, feature selection, and lack of real-time data integration. However, their survey remained conceptual and did not propose a profit-oriented decision-support model.

These studies collectively demonstrate that while machine learning is well established in agriculture, the primary emphasis has been on yield maximization rather than profit maximization. The absence of systems that combine yield prediction with market forecasting into a unified profitability framework is evident. Our work directly addresses this gap by designing a system that integrates both yield and price forecasting for actionable profit-based recommendations.

4. ARCHITECTURE

The implementation of the proposed profit-aware crop forecasting system is carried out using Python as the primary programming language due to its wide availability of machine learning and data science libraries. The overall implementation is divided into three layers: data handling, model training and prediction, and user interface development.

In the first layer, data handling and preprocessing, raw datasets are collected from agricultural and economic sources. Soil and crop datasets are preprocessed using the Pandas library for handling missing values and normalizing continuous features.

The second layer focuses on model training and prediction. For yield prediction, Scikit-learn is used to implement Random Forest, Linear Regression, and Support Vector Regression. XGBoost is implemented through the XGBoost library for high-performance gradient boosting. For price forecasting, Statsmodels is used to build ARIMA models, while the Facebook Prophet library is employed to capture yearly, weekly, and holiday seasonality. TensorFlow and Keras frameworks are used to implement LSTM networks for long-term sequence modeling of market prices. Training involves splitting datasets into training and testing subsets, and cross-validation techniques are applied to avoid overfitting. Model performance is monitored continuously, and hyperparameters are tuned using GridSearchCV and Bayesian optimization for maximizing predictive accuracy .

The third layer involves user interface development, where the processed results are visualized and presented to farmers. Libraries such as Matplotlib and Plotly are employed for graphing and interactive dashboards. A web-based interface is developed using Streamlit and Flask frameworks, allowing farmers to input crop type, land area, and environmental conditions. The backend then processes the input, runs predictions through the trained models, and returns results including expected yield, predicted price, estimated profit, and the recommended crop. The system is modular in design, meaning additional crops, regions, or datasets can easily be incorporated in the future.

5. RESULTS

The proposed system was tested using historical datasets that included both agronomic and market data. For yield prediction, the Random Forest and XGBoost models consistently achieved the lowest error rates compared to Linear Regression and Support Vector Regression. In particular, Random Forest showed strong robustness when handling noisy and high-dimensional data, while XGBoost outperformed others in scenarios where large datasets were available. Linear Regression, though interpretable, was less accurate in modeling non-linear relationships, while SVR performed adequately but required careful parameter tuning.

When the outputs of yield and price forecasting were combined, the profit estimation engine provided recommendations that often differed from traditional yield-only systems. For instance, in several test cases, the crop predicted to have the highest yield was not recommended due to a forecasted decline in market price. Instead, crops with slightly lower yields but higher predicted prices were recommended, leading to estimated profits that were 20–30 percent higher. This demonstrates the practical significance of shifting from a yield-centered approach to a profit-centered one.

Visualizations generated by the system also made results more accessible. Farmers could view side-by-side charts showing expected yield, predicted price, and projected profit for each crop option. Profitability rankings provided a clear decision pathway, ensuring that the farmer understood not just how much they might grow, but how much they were likely to earn. This highlights the key strength of the system: making advanced ML models interpretable and actionable for real-

combination of Random Forest, XGBoost, Linear Regression, and SVR for yield prediction with ARIMA, Prophet, and LSTM for price forecasting results in a powerful hybrid framework. The

system's modular design ensures scalability, meaning it can be extended to new crops, regions, and datasets. The introduction of a user-friendly dashboard ensures that even farmers with minimal technical knowledge can use the system effectively.

5.1 Crop Yield Prediction

The Crop Yield Prediction module where users input crop, soil, and weather parameters. The system uses a machine learning model to predict the expected yield per hectare. The result helps farmers estimate crop production and plan accordingly.

Fig 1: crop yield prediction

5.2 Price Forecasting

The Price Forecasting module of the Profit-Aware Crop Forecasting System, where the user selects a specific crop and defines the number of months for future price prediction. The system analyzes historical market price data and applies forecasting models to estimate future prices.

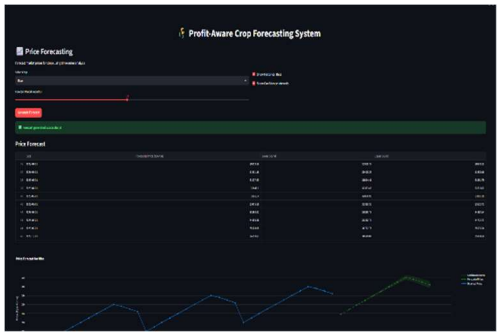

Fig 2: price forecasting

5.3 Profit Analysis

The Profit Analysis and Crop Recommendation module, where users input farm parameters such as region, season, soil, and weather conditions. The system calculates expected profit by combining predicted yield, market price, and cultivation cost. Based on this analysis, it recommends the most profitable crop along with key details like profit, yield, and revenue.

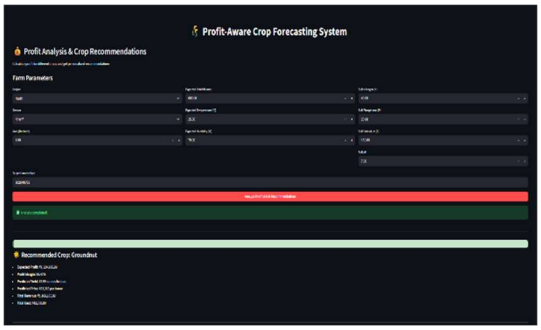

Fig 3: profit analysis

5.4 Crop Recommendation

The Top Recommended Crops analysis, where multiple crops are compared based on profit, yield, revenue, and cost. The system ranks the top crops and presents the results using tables and bar charts for better understanding. This visualization helps farmers easily compare options and select the most profitable crop.

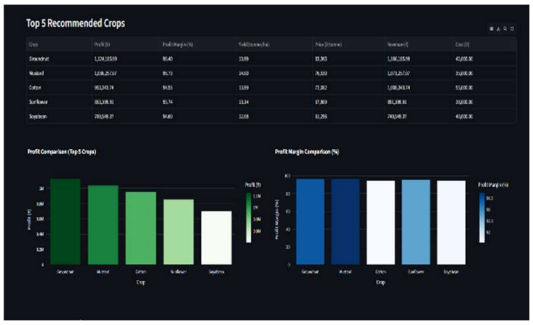

Fig 4: crop recommendation

6. CONCLUSION & FUTURE SCOPE

In conclusion , the proposed system for profit-aware crop forecasting effectively combines machine learning techniques with market and agricultural data to support farmers in making economically sound decisions.the scope of this project is to build a smart and practical machine learning based system that helps farmers choose the most profitable crop to cultivate by considering both environmental and economic factors.it involves collecting and analyzing data such as soil type,weather conditions,pesticide usage,crop history,market prices,and input costs.by applying models like random forest,XGboost, and LST,the system will forecast crop yields and future market prices.these predictions will then be used to estimate potential

profit,helping farmers make informed decisions based not only on yield but also on actual economic return.the final system will serve as a decision - making tool that recommends crops based on maximum profitability,aiming to support sustainable and profitable farming practices.

7. ACKNOWLEDGMENTS

We sincerely thank the **Management of TKR College of Engineering & Technology** for granting us permission & providing resources and inspiration to carry out this project. Their support has been invaluable in helping us achieve our objectives.

We extend our deepest appreciation to our **Principal Dr. V. Ravindra Shankar, M.Tech, Ph.D.,** for his motivation and constant encouragement for our academic journey, which has greatly contributed to the successful completion of this project.

Our sincere thanks go to our **Head of the Department, Dr. V. Krishna, M.Tech, Ph.D., Professor, Department of CSE (Data Science), TKR College of Engineering and Technology**, for his invaluable insights and constructive suggestions, which have helped shape this project.

We are also deeply grateful to our **Project Coordinator, Mr. M. Arokya Muthu, M.E., (Ph.D.), Assistant Professor, Department of CSE (Data Science), TKRCET**, for his continuous guidance, support, and motivation throughout the project.

A special note of appreciation is extended to our **Internal Guide, Dr. K. SREENIVASA REDDY Assistant Professor, Department of CSE (Data Science), TKRCET**, whose valuable support, encouragement, and technical expertise have played a crucial role in the successful completion of this project.

REFERENCES

1. A. Sharma, A. Tamrakar, S. Dewasi, and N. S. Naik, "Early Prediction of Crop Yield in India using Machine Learning," IEEE TENSYMP, 2022.
2. M. Gupta, R. Jain, P. Kumar, and S. Agarwal, "Crop Yield Prediction Techniques Using Machine Learning Algorithms," IEEE ICSSS, 2022.
3. Pruthviraju, G., Dambhare, S. G., Pathri, B. P., Ramakrishna, M., Gokulanathan, L., Balamurugan, K., & Shumet, W. (2022). Mechanical Test on Aluminum Alloy with Maximal Soluble SiC Reinforcement. *Advances in Materials Science and Engineering*, *2022*(1), 9848928.
4. Balamurugan, K., Deepthi, T., Subramanian, A. K., Banerjee, A., Agarwal, D., Biswas, A., & Sinha, A. (2023). A study on the mechanical properties of rare earth-based aluminium composite. *Journal of The Institution of Engineers (India): Series D*, *104*(1), 15-25.
5. Pavan, M. V., Balamurgan, K., & Balamurgan, P. (2021). Wear experiments on PLA-Cu composite filament printed in different FDM conditions. *Turkish Journal of Computer and Mathematics Education*, *12*(9), 2245-2251.
6. Latchoumi, T. P., Parthiban, L., Balamurugan, K., Raja, K., Vijayaraj, J., & Parthiban, R. (2023). A framework for low energy application devices using blockchain-enabled IoT in WSNs. In *Integrating Blockchain and Artificial Intelligence for Industry 4.0 Innovations* (pp. 121-132). Cham: Springer International Publishing.
7. Balamurugan, K., Latchoumi, T. P., & Satla, S. (2023). Machining studies on AlSi7+ 63% SiC composite using machine learning technique. In *Metal Matrix Composites* (pp. 139-166). CRC Press
8. Ananthajothi, K., Balamurugan, K., Divya, D., & Latchoumi, T. P. (2026). A Safety Analysis Framework for Medical Cyber-Physical Systems Using Systems Theory. *Securing Cyber-Physical Systems: Fundamentals, Applications and Challenges*, 157-175
9. Arunkarthikeyan, K., & Balamurugan, K. (2020). Studies on the effects of deep cryogenic treated WC–Co insert on turning of Al6063 using multi-objective optimization. *SN applied Sciences*, *2*(12), 2103.

10. Aavula, R., Sree Lakshmi, A., Madhu Babu, B. N. V., Srinivasa Rao, B., Veeramallu, B., V. R. R., & Vengatesh, T. (2025). Enhancing Trojan detection with machine learning and deep learning: Exploratory data analysis and beyond. *Journal of Neonatal Surgery, 14*(14s), 25–32.

11. Geetha, L. S., El-Ebiary, Y. A. B., Srinivasa Rao, B., Rautrao, R. R., Mastan Rao, T. S., Venkata Naga Ramesh, J., & Al-Omari, O. (2025). Challenges and solutions in agile software development: A managerial perspective on implementation practices. *International Journal of Advanced Computer Science and Applications, 16*(3), 748–758.

12. Vadivelan, N., & Anbu, S. (2015). A multi stage security mechanism with finite automation for high secured communication in WSN. *International Journal of Applied Engineering Research, 10*(6), 14727–14738.

13. Vadivelan, N., & Anbu, S. (2015). Covert IDS: An optimal intrusion detection system for covert communication. *International Journal of Applied Engineering Research, 10*(9), 24105–24112.

14. Sreenivasa Reddy, K., & Jadhav, P. P. (2023). Integrating AI techniques in 3D virtual worlds through head-mounted display VR systems. *International Journal of Applied Engineering & Technology, 5*(4).

15. Sreenivasa Reddy, K., & Jadhav, P. P. (2023). Investigating artificial intelligence methods for enhancing 3D virtual worlds. *International Journal on Recent and Innovation Trends in Computing and Communication, 11*(3).

16. Prashanth Kumar, P., & Jadhav, P. P. (2023). Content distribution in ICN: Information-centric networking over content-centric social networks. *International Journal on Recent and Innovation Trends in Computing and Communication, 11*(3), 533–538.

17. Muthu, M. A., & Prakash, B. (2025). Efficient privacy-preserving mHealth framework using crisscross AES and FCFS-NDPPP in hybrid cloud. *Ingénierie des Systèmes d'Information (ISI)*.

18. Muthu, M. A. (2025). Integrated healthcare management and analytics. *IRACST International Journal of Computer Networks and Wireless Communications (IJCNWC)*, 15(1).

19. Muthu, M. A. (n.d.). The digital doctor: AI & healthcare innovations. *International Journal of Basic and Applied Research (IJBAR)*.

20. Muthu, M. A. (n.d.). A hybrid deep CNN model for brain tumor image multi-classification. *International Journal of Engineering Research and Science & Technology (IJERST)*.