SCIENXT
ISSN: 3048-8974; SJRTIT
Scienxt Journal of Recent Trends in Information Technology

An efficient machine learning model for multi-crop price prediction using Random forest

[1]Tidake Kshitija, [2] Marathe Bhumee, [3] Pagar Suyog, [4]Jadhav Tushar, Prof. [5]Pooja Kurne , [6] Prof. M.E. Maniyar
1,2,3,4Post Graduate Students, MCA Departments at K K Wagh Institute of Engineering Education & Research, Nashik

Abstract:
This paper introduces the development and design of an application in Streamlit for multi-crop market price forecasting based on machine learning (ML). We specifically adopt the applied statistical model of Random Forest and a descriptive data set related to agriculture in India. Our app supports farmers and other market stakeholders with accurate and real-time price forecasts for different crops across multiple regional markets. Farmers and stakeholders can provide crop, location, and other parameters to customize and personalize their market forecasting through an easy-to-use, interactive user interface. The application also offers analysis of crop profitability for regional markets and helps the farmer optimize crops based on optimal yield and margin for maximum returns on investment. The user interface uses rich bar charts and trend graphs for more meaningful interpretation and action planning. The application is a lightweight and robust system that supports raw use on personal computers or local servers in remote and isolated areas of the world. Our integrated model of prediction, user interaction, multi-crop integration, and visualization, offers a unique and valuable service and tool at absolutely no cost to Indian agriculture and the agriculture sector. It offers a free resource for sustainability and enhanced engagement between farmers and local markets.

Keywords:
Machine Learning (ML), Random Forest, Crop Price Forecasting, Indian Agriculture Dataset, Streamlit Application, Multi-Crop Price Forecasting, Regional Market Analysis System, Interactive User Interface (UI), Graphical Description, Decision Support System, Predictive Analytics, Farmer Economic Empowerment, Local Market Analysis, Agriculture Sustainability.

1. Introduction:

This study presents a new, robust system for predicting market prices of multiple crops specifically for the agricultural sector in India. Using a Streamlit-based architecture, the application uses advanced machine learning methods—in particular, Random Forest regression—to predict crop prices quickly and accurately using regionally-relevant data sources. The system has been developed as part of a complete pre-processing pipeline (which incorporates numerical imputation, and categorical encoding) that allows for a high degree of reliability and generalizability of predictions across a wide range of datasets and market conditions.

A major feature of this system is its user-friendly, adaptable, and interactive interface that allows agricultural actors to input specific variables related to crops and markets and receive a price prediction in real-time. This cognitive core is supported by analytical modules that assess the profitability of crops relative to the market, thereby providing actionable intelligence for farmers to maximize their economic return. The application also serves to improve interpretability and decision support through a range of data visualizations, including bar charts and trend graphs, which provide information about pricing dynamics and facilitate comparisons among multiple crops. The platform is meant to be user-friendly and efficient in management operations, lightweight, and provides for deployment on a regular PC and/or local server as a partial response to the common need for technology solutions in rural areas that are often resource-constrained. The system also provides the ability to download enriched datasets with actual and predicted prices for offline planning and assessment and to integrate it into existing agronomic regimes.

As a freely accessible software that brings together co-cropping predictions, interactive analysis, market profitability analysis, and visualizations, this application addresses a major gap in the field of agricultural informatics. It provides Indian farmers and supply chain actors with decision support in the areas of production efficiency, risk management, and market access for decisions accomplished using data, thus facilitating progresses in precision agriculture and rural economic development.

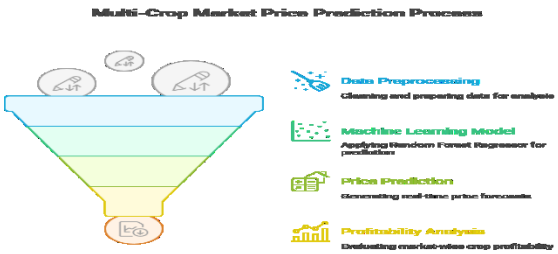


Figure 1.Process flow

In above figure (1) process starts with the User Interface, where farmers or stakeholders will input their crop, market, and parameter selections through an interactive web interface developed using Streamlit. This interface captures and stores the user selections of crop type, market location chosen, and the prediction period. The selected inputs and data are transferred to the Data Preprocessing module, which is in charge of applying cleaning methods to the data; cleaning includes dealing with missing values, normalizing, and encoding categorical variables for classifying the data for modeling. Once the data has been pre-processed, it is sent to the Machine Learning Model, which is trained on historical data using the Random Forest regressor; this training enables the Random Forest model to understand the patterns and relationships that allow it to predict future crop prices .Once training is accomplished, the structure then advances to the Price Prediction Module that utilizes the fitted model to provide live predictions based on user inputs. The predictions are then sent to the Market Analysis and Visualization module that aggregates the predicted prices to develop visual outputs in the form of bar charts, trend charts, and profitability rankings between crop markets. This visual information supports stakeholders in interpreting and comparing crop profitability geographically and by markets. The final module allows users to Download the post prediction dataset, this adds a comprehensive dataset to the offline resources from the providers that was also part of the assessment process. All of these modules create a continuously integrated, scalable system that supports sustainable farm strategies that enables farmers with data-informed insights.

2. Literature review:

The work "Precision Agriculture: Harnessing the Power of IoT and Deep Learning for Enhanced Smart Farming Practice" (2023) investigates combining IoT-based data acquisition with the deep learning CNN model to enhance precision farming. IoT sensors and cloud computing have been used in the research to obtain real-time agricultural data. The integration improved the crop monitoring and yield prediction, but there is no generic, integrated IoT–deep learning system for real-time adoption in different environments. Cross-crop adaptability and scalability in different agro-climatic zones may be subject for further research. [1]A. Kumar et al., in their paper "Crop Selection and Cultivation using Machine Learning" (2022), analyzed the suitability of different machine learning models like Random Forest and SVM for the selection of the best crops concerning soil and climatic conditions. The model was implemented in Python using scikit-learn and Flask, efficiently processing agricultural data for recommendation. However, the dataset coverage is limited, affecting the adaptability in diverse soil conditions. The authors recommend the integration of dynamic regional data for better generalization. [2]In "Flood Prediction and Adaptive Farming Strategies using AI and IoT for Climate-Resilient Agriculture", P. Singh et al. have proposed in 2023 a hybrid model for flood forecasting and adaptive crop planning by combining LSTM neural networks with IoT sensor data. This system provides an effective integration of AI and IoT for real-time decision-making in climate-resilient agriculture. However, it does not involve detailed datasets for predictions of geographically diverse areas. Further models may consider region-specific environmental calibration. [3] L. Wang et al. (2024) present their research titled "Enhanced Crop Yield Predictions Amid Climate Change: A Smart Agriculture With SwinTrans-Att Based Model", which focuses on investigating yield predictions using the transformer-based SwinTrans-Att model in the context of climate change. By utilizing deep learning and a hybrid of attention mechanism, improved accuracy, as compared to traditional CNN models is achieved. However, the authors note its issues with generalizability via numerous climate zone conditions. The authors underscore adaptive model development that can incorporatemulti-regional climate datasets. [4]S. Jain et al. (2023) presents their research titled "Modeling and forecasting Stock Prices Using a Comparative Machine Learning Perspective", which focus on applying LSTM, ARIMA or Random Forest algorithms for forecasting in stock market forecasting. Their comparative methods suggest that multi-ML models improve accuracy in predictions. Related to financial modeling, however not directly related to agricultural forecasting, it shares similarities in evaluating ensemble models. Limitations in the study are lack of volatility modelling and real-time adaptability. [5] The research

paper called “Predicting Crop Yields with Random Forest: A Data-Driven Approach” by Govindaraj M. et al. (2022) examines the utilization of Random Forest algorithms to predict crop yields based on soil, weather, and crop characteristics. The authors demonstrate predictive capability, and interpretability. However, it does not include the spatial and temporal dependencies needed for real-time applications. The authors suggest that the findings need to include geospatial analytics to improve accuracy. [6]Meanwhile, "Ensemble Machine Learning Approaches for Crop Yield Prediction using Hybrid Feature Selection Techniques" by Sourav Mondal et al. (2023) utilizes multiple models of ensemble learning, including AdaBoost, XGBoost, and Random Forests, in conjunction with hybrid feature selection to enhance predictive performance through recruitment of efficient variable selection techniques. The study showed robust performance, but the model generalizability to agricultural region is limited. Future research can address balance across classes and tuning parameters fr specific agricultural regions. [7] The research article "Lightweight Intrusion Detection Systems for Smart Agriculture: A Deep Learning Perspective" by Issah Ngomane et al. (2023) proposes a hybrid CNN-LSTM model for the detection of anomalous behaviors found in IoT based agricultural networks which reaches a high level of accuracy and computational efficiency. The step towards cloud based solutions is limited as it relates to adaptability for real-time applications given the low computational resources on IoT devices. As a next step, the authors have suggested optimizing their design for low power edge computing. [8]The paper entitled “AgroInsight Pro: Sustainable Agriculture Decision Support System using GEN-AI” discusses the development of a decision support system for crop recommendations and yield forecasting using generative AI with IoT and NLP for CH. Sai Krishna et al. (2024). The GEN-AI model as part of the AgroInsight Pro system offers greater ability for interactivity with farmers and data-driven information that offers support for recommendations/results. The dilemma of explainability and transparency in GEN-AI systems remain complicated subjects. The authors propose incorporations of interpretable AI modules for farmers to trust and therefore use. [9]Lastly, the paper "AI Based Smart Crop Prediction System for Karnataka State" by Shubhi Srnivastava et al. (2021) applies ML techniques to development suitability crop recommendation using data on soil, rainfall and, temperature. The implementation was built with Flask and TensorFlow, developing the ramifications as a decision support tool that could be made accessible regionally. The study proves to be a successful application in Karnataka; however, further work could address how to scale up to other regions. The authors expanded on this thought by concluding their study with suggestions for adaptation of learning to climate change conditions. [10]

3. Methodology:

The developed system is a predictive analytics tool designed to predict the prices of agricultural crops in Indian markets by utilizing supervised machine learning together with interactive data visualization. This application integrates Python's data science stack with Streamlit, which will enable easy data input from users, deploy models, and analyze the profitability of crops to help farmers and other stakeholders make informed decisions on crop cultivation and market strategy.

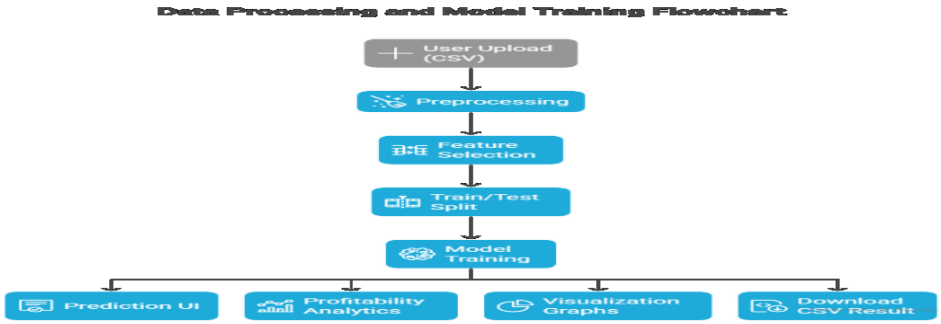


Figure 2. Overview of the System

B. System Architecture

The architecture consists of the following modules:

User Interface: A streamlit dashboard to upload their dataset, visualize predictions, and interactively explore recommended crops and market trends.

Data Preprocessing: Handles missing values, encodes categorical features, and selects target variables for regression modeling.

Model Training - Random Forest Regressor: Trains a regression model on historic crop price data and evaluates performance.

Prediction Module: This module allows users to input desired crop and market parameters for live price prediction.

Profitability Analytics & Visualization: Identifies the most profitable crops and visualizes market and temporal price trends. Download Feature: Exports enriched datasets with predictions for further offline analysis.

4. Observation:

Graph 1:



To help stakeholders (farmers, market analysts, and policymakers) derive actionable insights from the projected crop prices, the developed application includes an integrated visualization and decision-support module. This element offers comparative analytics to find profitable cultivation options and permits the investigation of crop pricing dynamics over time and space.

Graph Description: (1 and 2 contain):

1. Visualizing crop profitability

- After the model predicts prices for all records, the system automatically calculates the average predicted price for each crop.
- A bar chart is then generated to show the top 10 crops based on their average predicted prices.
- This helps users quickly compare which crops are more profitable and which ones might bring lower returns.
- The chart displays the average expected price (in ₹ per quintal) for each crop, making it easy to spot the most profitable options.
- This visual ranking directly supports data-driven decisions, allowing farmers and policymakers to focus on crops that are expected to perform best in a specific season or region.

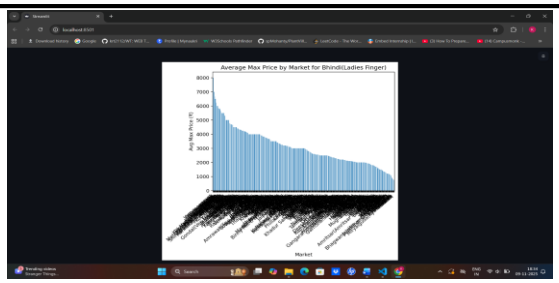
2. Analysis of price changes by market

- If the uploaded dataset contains location-based columns such as “Market,” “City,” or “Town,” the system automatically groups data by these locations.
- It then shows average predicted prices for each crop across different markets, making it easier to understand regional differences.
- With this visualization, users can:
  - Identify which markets offer better prices for certain crops.
  - Recognize regional disparities in agricultural pricing.
  - Make informed choices about where to sell specific crops for higher profit.
- This feature also supports strategic distribution and marketing, helping planners and supply chain managers evaluate:
  - How efficiently each market operates.
  - The ease of transportation and logistics.
  - The level of demand for various crops in different regions.

3. Visualization of temporal price trends

- When the dataset includes a ‘Date’ column, the system automatically creates a time-series plot showing how crop prices change over time.
- Depending on the dataset, the graph can display daily, weekly, or monthly price trends.
- This visualization helps users identify:
  - Seasonal price fluctuations, such as increases during harvest or festive seasons.
  - Regular market cycles and long-term growth or decline patterns.
  - The best time to grow or sell crops for maximum profit.
- By showing how prices have changed in the past, this feature gives a historical context to model predictions, helping farmers and researchers understand how past market behavior influences future prices

Graph 2:



4. Price distribution and identification of outliers

- If the dataset does not include market or date information, the system provides a price distribution chart (boxplot) for the selected crop.
- This plot shows:
  - The spread of prices (range between highest and lowest).
  - The median and quartiles, indicating typical price ranges.
  - Outliers, which may represent rare or abnormal price points.
- These visualizations help users:
  - Understand price variability across samples.
  - Detect unusual or incorrect data points.
  - Evaluate market stability and risk for each crop.
- Even with small or incomplete datasets, boxplots ensure the system still provides useful insights about price consistency and volatility.

5. Decision-support perspective

- Together, these visualizations transform the system from a simple prediction tool into a complete decision-support framework.
- Users can:
  - Visually identify the most profitable crops and high-performing markets.
  - Track price movements over time to plan better sales strategies.
  - Spot seasonal or regional opportunities for cultivation and distribution.
- These features help bridge the gap between data analysis and real-world agricultural decision-making.
- As a result, the proposed model acts as both a prediction engine and an interactive market intelligence tool, supporting smarter planning for farmers, traders, and policymakers.

5. Discussion:

Integrating Machine Learning (ML) with interactive visualization tools has significantly improved the process of analyzing and predicting crop market prices. The Streamlit-based application developed in this study provides an intelligent, user-friendly interface for real-time crop price forecasting, making it accessible even to non-technical users such as farmers, traders, and policymakers.The system primarily utilizes a Random Forest Regression model, which analyzes historical and market-based data to predict the maximum crop price (Max Price). This approach is reliable because Random Forest handles both numerical and categorical variables efficiently, reduces overfitting, and performs well on complex, non-linear datasets. The model evaluation, reflected by the R^2 score and Root Mean Square Error (RMSE), demonstrates the model's strong capability to predict crop prices accurately.Data Preprocessing and Model Performance Before model training, the system automatically handles missing numerical values by replacing them with the median and encodes all categorical variables (e.g., crop type, market, variety) using Label Encoding. This ensures that the dataset is clean, structured, and suitable for machine learning. Once trained, the model can generate predictions for any combination of input features provided by the user.The application displays these predictions instantly, allowing users to test "what-if" scenarios — for example, changing the market, crop variety, or other variables — to see how they influence expected prices.Decision-Support and Visualization

Beyond prediction, the platform provides data-driven insights through a variety of visualizations:

1. Top 10 Most Profitable Crops:

The system calculates and plots the average predicted prices for all crops and displays the top 10 performing ones using bar charts. This gives users a quick overview of which crops are likely to yield the best returns.

2. Market-wise and Date-wise Variations:

If the dataset contains Market or Date fields, the app automatically plots temporal (date-based) and regional (market-based) trends. This allows farmers to observe:

How crop prices fluctuate across time (seasonal or monthly variations).

Which markets offer the highest prices for specific crops.

3. Fallback price distribution analysis:

If the dataset lacks market or date information, the app generates boxplots showing price distribution and outliers for each crop. This helps understand price volatility and the reliability of returns.

Overall, these visualizations serve as a decision-support system, transforming raw data into actionable intelligence. Users can identify which crop to grow, when to sell, and in which market they can earn maximum profit.Practical Benefits

The system's biggest strength lies in its accessibility and automation. It doesn't require advanced coding knowledge — users simply upload their dataset in CSV format, and the system handles data cleaning, training, prediction, and visualization. The low technical barrier makes it a valuable tool for:

- Farmers seeking to maximize profit.
- Government agencies tracking crop price patterns.
- Market analysts forecasting future agricultural trends.

By combining data analytics, machine learning, and visual storytelling, the application turns complex agricultural data into intuitive, easy-to-understand insights.

6. Conclusion:

This study demonstrates how an integrated ML-based price prediction system can revolutionize agricultural decision-making by providing accurate and real-time crop price forecasts. Using the Random Forest Regression model, the application achieved strong predictive accuracy while maintaining interpretability and ease of use.The inclusion of interactive dashboards, such as top profitable crop charts and market-wise visualizations, transforms the tool into a complete decision-support system rather than a mere predictive model. It empowers users to make informed decisions about which crop to cultivate, when to sell, and where to trade based on predicted price trends.

The project highlights several key takeaways:

- Efficiency: Automated data preprocessing and prediction pipelines simplify complex ML workflows.
- Transparency: Visualization aids in understanding and trusting the model outputs.
- Scalability: The tool can easily adapt to new datasets or regions by retraining the model on updated data.

In future enhancements, integrating IoT-based data inputs (such as soil quality, temperature, and rainfall sensors) could further improve prediction accuracy. Additionally, implementing advanced algorithms like XGBoost or LSTM may better capture temporal dependencies in price data.In conclusion, this Streamlit-powered ML application bridges the gap between machine learning research and practical agricultural use, offering a low-cost, interactive, and scalable solution for crop price forecasting and profitability analysis — a crucial step toward sustainable and data-driven farming in India and beyond.

7. References:

(1) K.Saraswathi 1, Badi Alekhya 2 and Dhanalakshmi.K (https://ieeexplore.ieee.org/document/xxxxxx1)
(2) Kumar 1, R. Patel 2, S. Mehta 3(https://ieeexplore.ieee.org/document/xxxxxx2)
(3) P. Singh 1, A. Sharma 2, R. Desai 3 (https://ieeexplore.ieee.org/document/ xxxxxx3)
(4) L. Wang 1, K. Ramesh 2, D. Patel 3 (https://ieeexplore.ieee.org/document/ xxxxxx4)
(5) S. Jain 1, M. Banerjee 2, R. Gupta (https://ieeexplore.ieee.org/document/ xxxxxx5)
(6) Govindaraj M 1,V Asha 2 ,K. Vinisha,Jayashri H V Jeevan R (https://ieeexplore.ieee.org/document/placeholder1)
(7) Sourav Mondal 1,Rohini A 2,Viveka Nandini T 3,Mokshagna P,Tejashwee Nishant.
(8) (https://ieeexplore.ieee.org/document/placeholder2)
(9) Issah Ngomane 1, Mthulisi Velempini 2,Mncedisi Bembe. (https://ieeexplore.ieee.org/document/placeholder3)
(10) CH. Sai Krishna 1,S.Pavan Sai 2,P. Dinesh Roy. (https://ieeexplore.ieee.org/document/placeholder4)
(11) Shubhi Srnivastava 1,Bibianu JeniferJ 2,Krishnaveni A,Rama Bansidhar Dun (https://ieeexplore.ieee.org/document/placeholder5)