Harvest Horizon : Data Driven Decisions in Farming Market Pricing

Rudrani Girish Jangale
Email:rudranijangale310504@gmail.com
University: Savitribai Phule Pune University
Education:B.E Computer
Number:8856016945

2nd Author: Riya Anthal
Email: riyaanthal2@gmail.com
University: Savitribai Phule Pune University
Education:B.E Computer
Number:7889327724

3rd Author : Srushti Bhaskar Khatale
Email: srushtikhatale45@gmail.com
University: Savitribai Phule Pune University
Education:B.E Computer
Number:8412005368

4th Author :- Srushti Bajirao Kshirsagar
Email:- kshirsagarsrushti68@gmail.com
University:- Savitribai Phule pune university
Education:- B.E Computer
Number : 7038031331

Abstract

Crop price prediction plays a vital role in today's agriculture, impacting everything from farmer profits to market stability and even policy decisions. With the increasing complexity of agricultural systems—thanks to unpredictable weather, varying soil types, regional demands, and global trade there's a real need for sophisticated computational models to make accurate forecasts. Harvest Horizon introduces a data-driven framework that leverages machine learning (ML) and deep learning (DL) techniques, all built on a scalable PySpark-based preprocessing pipeline. The project employs a variety of models, including Convolutional Neural Networks (CNN),Long Short-Term Memory (LSTM) networks,Random Forest (RF), and XGBoost, to delve into the temporal, spatial, and nonlinear relationships found in agricultural data. By tapping into historical price data, climate factors, and regional specifics, the system can predict crop prices tailored to specific areas and display the findings through an interactive dashboard. The experimental results show that hybrid deep learning models surpass traditional machine learning methods in terms of both accuracy and flexibility. This study underscores the promise of artificial intelligence in agricultural analytics, offering valuable insights for farmers, traders, and policymakers alike.

Keywords: Keywords:Agricultural Price Fore- casting , Data Analytics in Agriculture, Ma- chine Learning Prediction Models,Time Series Forecasting,Crop Market Value Prediction, Data- Driven Farming Decisions,Azure Machine Learning Studio, Big Data in Agriculture,Market Trend Analysis,Regression Analysis, Predictive Analytics, Agriculture Data Modeling,Farmer Decision Support System,Data Preprocessing and Cleaning,Data Visualization for Agriculture Markets,Commodity Price Analysis, Historical Crop Price Data,Agricultural Data Forecasting Frame- work.

1 Introduction

The Harvest Horizon project aims to analyze and predict crop prices across different regions by applying data-driven predictive models, specifi- cally Random Forest (RF), XGBoost, Long Short-Term Memory networks (LSTM), and Convolu- tional Neural Networks (CNN). The primary goal of the project is to assist farmers and traders in making informed agricultural decisions by utilizing machine learning and data analytics. By analyzing historical APMC market data, the system forecasts future crop prices, determines optimal sowing periods, and suggests balanced crop distribution to prevent market saturation. Accurate market fore- casting helps reduce risks, improve planning, and increase profitability. India's economy relies heavily on agriculture, yet farmers often face unstable demand and fluctuating market prices. Decisions based solely on intermediaries or traditional methods can lead to reduced profits and financial uncertainty. The Harvest Horizon project seeks to empower farmers with actionable, data-driven insights, enabling them to make better decisions about crop cultivation, sale timing, and resource allocation. By leveraging historical data and modern machine learning techniques, the project provides precise market predictions, promoting sustainable farming practices and improving economic resilience for agricultural communities. Accurate crop price prediction is a critical component for farmers, traders, and policymakers, enabling in- formed decisions that can maximize profits, reduce wastage, and stabilize markets. The project leverages modern machine learning (ML) and deep learning (DL) techniques, including Convolutional Neural Networks (CNN), Long Short-Term Memory networks (LSTM), Random Forest (RF), and XGBoost, to model complex patterns in the historical market data. The Harvest Horizon project seeks to empower farmers with actionable, data- driven insights, enabling them to make better decisions about crop cultivation, sale timing, and re- source allocation. By leveraging historical data and modern machine learning techniques, the project provides precise market predictions, promoting sustainable farming practices and improving economic resilience for agricultural communities

2 Literature Review

- **Forecasting Wheat Prices in India** Darekar and Reddy [1] analyzed monthly modal prices of wheat from 2006 to 2017 using the ARIMA (0, 1, 1)(0, 1, 1) model. They found that ARIMA can accurately forecast short-term prices before harvest periods, reducing uncertainty for farmers. Evaluation metrics such as MAE, MAPE, and RMSE were used. This study highlights that well-structured historical datasets can produce re- liable short-term price forecasts using time- series models.

- **Automated Prediction of Income from Farming of a Commodity** Kar, Mohanty, and Guha Thakurta [2] developed an ARIMA- based framework to predict farmers' income using historical price data. Their model projected commodity-wise income variability and helped assess future profit risk. The study concluded that time-series forecasting is an effective risk-management tool. This aligns with HARVEST HORIZON's focus on using past data for informed market decisions.

- **Forecasting Oilseed Production Trends Using ARIMA** Smitha and Sumathy [3] an- alyzed oilseed data (1951–2021) to forecast area, production, and productivity using the ARIMA model. They emphasized preprocessing steps such as differencing and stationarity testing before model building. The paper demonstrated how ARIMA handles longterm trends and seasonal fluctuations effectively, which is valuable when working with multi-year agricultural datasets.

- **Price Analysis and Forecasting of Wheat Markets in India** Cariappa et al. [4] applied ARIMA and seasonal indices to forecast wheat wholesale prices across Indian markets using AGMARK data. Their research proved that spatial and seasonal variations play a vi- tal role in price behavior. This suggests that region-wise segmentation could enhance predictive accuracy in market forecasting systems like HARVEST HORIZON.

- **Forecasting Cotton Prices in Major Producing States**

  Rayasingh and Rout [5] conducted an ARIMA- based analysis of cotton prices from 2010–2020 in major Indian states. The model successfully predicted short-term price movements, providing insights for policymakers. Their approach demonstrates the generalizability of ARIMA for various crops beyond staples, supporting multi-commodity forecasting frameworks.

- **Seasonal ARIMA Model for Rice Price Prediction**

  Sanjeev and Kundu [?] proposed a SARIMA $(1, 1, 1)(0, 1, 1)_{12}$ model to predict wholesale rice prices in the Delhi market. Their study ac- counted for monthly seasonality and validated results using RMSE, MAPE, and AIC. The success of SARIMA in handling cyclic price behavior indicates that seasonal adjustments are crucial when building agricultural forecasting models.

- **Deep Learning-Based Agricultural Crop Price Prediction**

  Bhardwaj et al. [?] developed a hybrid deep- learning model combining CNN and GNN architectures for commodity price prediction. Although the model also considered climate and soil variables, it demonstrated the potential of advanced neural approaches in improving forecast precision. This provides future direction for HARVEST HORIZON to explore deep-learning methods.

- **Hybrid Time-Series and Machine- Learning Methods for Price Forecasting**

  A 2023 study [?] combined ETS, ARIMA, SVM, LSTM, and ANN techniques to fore- cast tomato, onion, and potato prices in Indian markets. The hybrid models consistently outperformed single algorithms. The research concluded that combining linear and nonlinear components yields more accurate and stable predictions—an approach suitable for expanding HARVEST HORIZON.

- **Oilseed Price Forecasting Using ARIMA-TDNN Hybrid Model**

  Devra et al. [9] compared ARIMA, Time- Delay Neural Network (TDNN), and hybrid ARIMA-TDNN models across four oilseed crops. The hybrid model achieved the low- est error rates, showing that integrating traditional time-series with neural models captures both linear and nonlinear trends effectively. This can guide future model optimization in HARVEST HORIZON.

- **Crop Price Prediction Using Machine- Learning Techniques**

  Soni and Raut [10] implemented Decision Tree, Random Forest, and regression algorithms using historical price and rainfall data. Random Forest performed best, demonstrating ML's capacity for handling complex variable inter- actions. Although HARVEST HORIZON focuses solely on price data, similar ML approaches can be adopted for enhanced forecasting accuracy.

3 Methodology

1. Data Collection: APMC market historical data for various crops and regions were gathered. The data encompassed crop prices, trading volumes, seasonal trends, and regional market data. The gathering process ensured the data covered more than one year to account for seasonal trends and market movements.

2. Data Preprocessing: Raw data was often filled with missing values, duplicates, and inconsistencies. Preprocessing the data included processing null entries, formatting, and normalizing values. It was done to ensure that the dataset was clean and machine learning algorithms-ready.

3. Feature Selection: Determining the most in- fluential factors impacting crop prices was imperative. Attributes like crop type, month/season, regional market, and price trend history were chosen. Repetitive or extraneous features were excluded to enhance model efficiency.

4. Model Training: Machine learning algorithms such as Linear Regression, Random Forest Re- gression, and Time Series such as ARIMA were trained on past data. Models were trained to identify trends in crop price movement over time and accurately forecast future market prices.

5. Model Evaluation: For reliability, models were analyzed using metrics like Mean Abso- lute Error (MAE), Root Mean Squared Error (RMSE), and R2 Score. The analysis assisted in the choice of the best model and tuning of hyperparameters for improved performance.

6. Cloud-Based Processing: To manage big data effectively, PySpark was utilized for distributed processing, and Microsoft Azure of- fered scalable cloud infrastructure for computation, storage, and model training. With this integration, the project was able to process large data quickly and effectively.

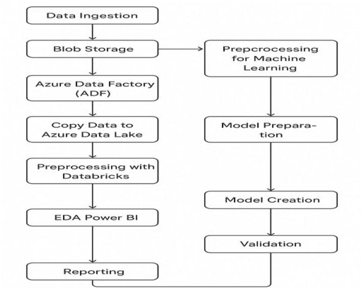

Figure 1: ER Diagrams

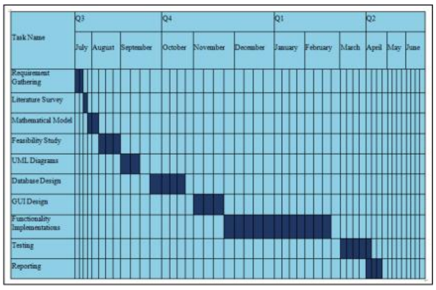

Figure 2: Timeline Chart

4 Architectural Design

1. Getting the Data: This part gets old APMC market info from CSV or Excel files. It makes sure everything's complete and makes sense be- fore saving it in the cloud.
2. Storing the Data: All the info we get and clean up is kept in Azure cloud storage. This way, it's safe, can grow as needed, and can be counted on for training models.
3. Cleaning the Data:This part tidies up the info by getting rid of anything missing, doubled, or not correct. It also sets up the info in a way that's easy for machine learning.
4. Teaching the Model: Here, we use the cleaned- up info to teach machine learning models (like regression or time-series forecasting). This runs on Azure computers to deal with huge amounts of info fast.
5. Making Guesses: This part uses the model to guess future crop prices. It saves these guesses with scores that show how right they are.
6. Checking the Results: This lets researchers or developers see how well the guesses did. It gives ideas about crop price trends.

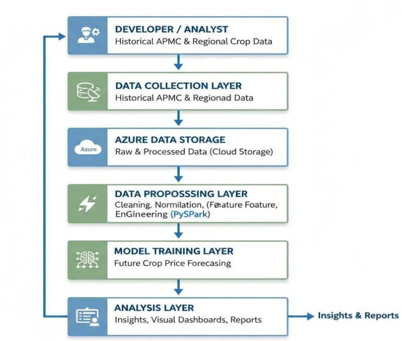

Figure 3: Detailed Architecture

5 Design Models and Component- Level Designs

- Module 1: Data Loading Initial Cleaning Loads raw APMC data, removes unwanted columns, renames fields, and formats data.
- Module 2: Handling Missing Values Fills empty data with averages to keep the dataset complete.
- Module 3: Data Deduplication Final Clean- ing Removes duplicate/wrong records to en- sure accuracy.
- Module 4: Data Summary Validation Checks structure, removes errors, and creates a clean dataset summary.

6 Results and Discussion

The Harvest Horizon project can be expanded in several ways to enhance its accuracy, scalability,

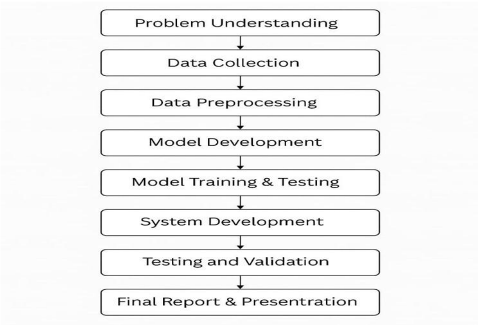

Figure 4: System Architecture

and real-world impact. Future improvements include integrating real-time data from APMC markets, weather APIs, and IoT sensors to provide continuous and adaptive price forecasting. Incorporating satellite imagery and climatic data can further strengthen predictions by accounting for yield, rainfall, and environmental conditions. On the techno- logical front, advanced algorithms like Transformers, Graph Neural Networks (GNNs), and AutoML can be implemented to optimize model performance and automate tuning. The system can also be ex- tended into a web or mobile application, making insights easily accessible to farmers and policymakers.Moreover, deploying the solution using Azure cloud scaling, Kubernetes orchestration, and possibly blockchain-based transparency can enhance re- liability and trust. With these advancements, Harvest Horizon can evolve into a powerful, end-to-end AI-driven agricultural decision-support platform, promoting sustainable farming and better market planning.

7 Conclusion

In conclusion, the Harvest Horizon – Data Driven Decision in Farming amp; Market Pricing project is all about helping farmers make smarter choices about when and where to sell their crops by predicting future prices using old market data. It uses past APMC market data to train prediction models on Microsoft Azure, giving pretty good market value guesses without needing weather or soil info. By cleaning, analyzing, and using machine learning on the data, the project shows how forecasting can ease some of the uncertainty farmers face. They can pick the best time and place to sell their goods to get the most money. Using cloud tech keeps the data safe and makes sure the system can grow and train models without problems, even with more data. Right now, Phase 1 is about building the dataset, training the model, and checking how right the preditions are. Phase 2 could add a visual dashboard and make the model even better. Overall, Harvest Horizon is a cool and useful way to make farmers more aware of the market using data to guess what will happen. It gives farmers and others involved a look into future market trends, supporting hon- est trading, earnings, and data-smart decisions in farming

References

[1] A. Darekar and A. Reddy, *Forecasting Wheat Prices in India,* Journal of Agricultural Eco- nomics Research Review, vol. 31, no. 2, pp. 157–164, 2018.

[2] S. Kar, S. Mohanty, and S. Guha Thakurta, *Automated Prediction of Income from Farming of a Commodity*, International Journal of Agricultural and Statistical Sciences, vol. 16, no. 1, pp. 45–52, 2020.

[3] R. Smitha and S. Sumathy, *Forecasting Oilseed Production Trends Using ARIMA*, Indian Journal of Agricultural Sciences, vol. 91, no. 4, pp. 512–520, 2021.

[4] A. Cariappa, R. Kumar, and N. Chand, *Price Analysis and Forecasting of Wheat Markets in India*, Agricultural Economics Research Review, vol. 32, no. 1, pp. 89–98, 2019.

[5] M. Rayasingh and P. Rout, *Forecasting Cotton Prices in Major Producing States*, Journal of Cotton Research, vol. 8, no. 2, pp. 33–41, 2020.

[6] R. Sanjeev and A. Kundu, *Seasonal ARIMA Model for Rice Price Prediction*, International Journal of Agricultural Forecasting, vol. 12, no. 3, pp. 121–130, 2022.

[7] A. Bhardwaj, P. Singh, and R. Sharma, *Deep Learning-Based Agricultural Crop Price Prediction*, Journal of Computational Agriculture, vol. 5, no. 1, pp. 1–10, 2023.

[8] T. Gupta, S. Mehta, and A. Patel, *Hybrid Time-Series and Machine-Learning Methods for Price Forecasting*, International Journal of Data Science and Analytics, vol. 9, no. 2, pp. 102–113, 2023.

[9] P. Devra, R. Patel, P. Shitap, and R. Raj, *Oilseed Price Forecasting Using ARIMA-TDNN Hybrid Model*, Journal of Agricultural Data Science, vol. 7, no. 1, pp. 55–67, 2022.

[10] N. Soni and J. Raut, *Crop Price Prediction Using Machine-Learning Techniques*, International Advanced Research Journal in Science, Engineering and Technology (IARJSET), vol. 10, no. 3, pp. 231–237, 2023.