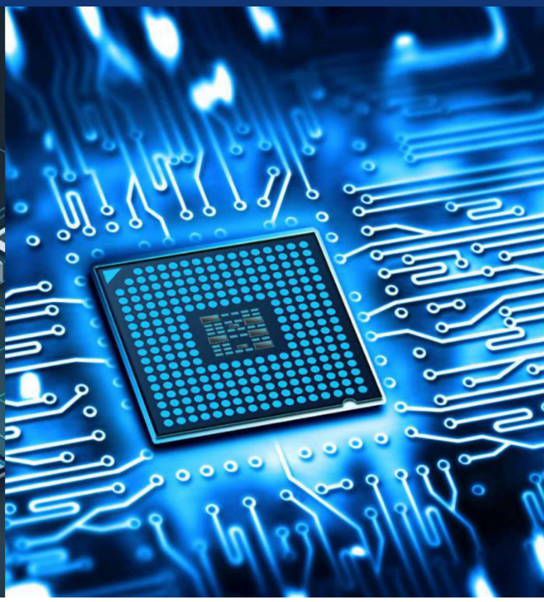
IJIRCCE

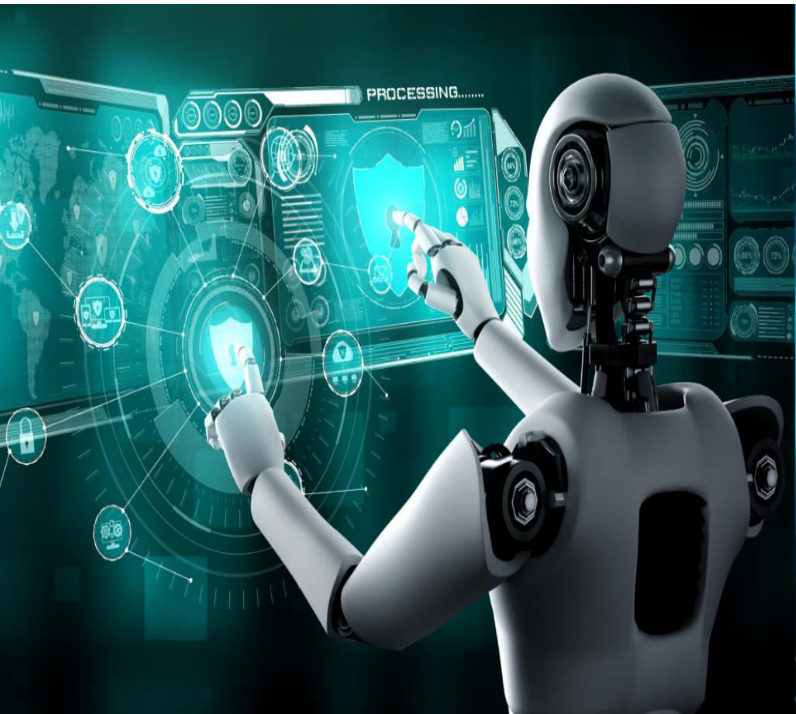
ISSN(O): 2320-9801

ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.771

Volume 13, Issue 5, May 2025

© 2025 IJIRCCE | Volume 13, Issue 5, May 2025| DOI:10.15680/IJIRCCE.2025.1305044

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

AI-ML Models for Predicting Prices of Agri-Horticultural Commodities

Saahil R Menon[1], Thanoj Y[2], Rahul Muthanna[3], Chandra Sekhar[4]

UG Student, Dept. of Computer Science and Engineering, Presidency University, Bengaluru, Karnataka, India

UG Student, Dept. of Computer Science and Engineering, Presidency University, Bengaluru, Karnataka, India

UG Student, Dept. of Computer Science and Engineering, Presidency University, Bengaluru, Karnataka, India

Professor, Dept. of Computer Science and Engineering, Presidency University, Bengaluru, Karnataka, India

ABSTRACT- The volatility of agricultural commodity prices presents a significant challenge for farmers and stakeholders in the agri-supply chain. Leveraging machine learning algorithms and historical market data, this study develops AI-ML models to predict the prices of key agri-horticultural commodities. The project is divided into two phases: the first focuses on potato prices across multiple regions, achieving 98% accuracy, while the second expands to include commodities such as onion, tomato, and pulses, yielding approximately 80% accuracy. This paper outlines the methodology, model implementation, and predictive outcomes, demonstrating the potential of AI-ML systems to empower farmers with market insights, reduce financial risk, and support data-driven agricultural planning.

KEYWORDS: Agricultural price forecasting; Machine Learning; Python; Commodity markets; XGBoost; Data analytics; Agri-tech solutions

I. INTRODUCTION

Agricultural markets are inherently volatile, influenced by numerous dynamic factors such as monsoons, pest outbreaks, supply chain disruptions, and regional consumption trends. These frequent fluctuations create substantial economic risks for farmers and agri-businesses, often resulting in reduced profitability and poor market planning. In this context, developing a data-driven approach for predicting the prices of agricultural commodities is essential. This project explores the use of artificial intelligence (AI) and machine learning (ML) to forecast the market prices of key agri-horticultural commodities, ultimately offering actionable insights to stakeholders across the agricultural value chain.

A. The Role of Price Prediction in Agriculture

Accurate price prediction empowers farmers and agri-entrepreneurs to make well-informed decisions regarding crop planning, storage, and marketing. By forecasting price trends in advance, they can decide when to sell produce for maximum profit or delay sales to avoid market gluts. Moreover, such predictive tools can assist cooperatives, distributors, and agri-tech firms in optimizing supply chain operations, reducing post-harvest losses, and streamlining procurement strategies.

B. Machine Learning in Price Forecasting

Machine learning offers a robust framework for detecting complex, non-linear patterns in commodity price data. In this project, Python libraries such as Pandas, Scikit-learn, and XGBoost were used to build and evaluate regression models capable of learning from past price trends. The development pipeline included data cleaning, handling missing values, feature selection (e.g., time features, lag values), normalization, and hyperparameter tuning. Advanced ensemble models like XGBoost helped enhance prediction accuracy by reducing overfitting and capturing subtle market fluctuations

C. Project Scope and Achievements

The study was executed in two distinct phases:

© 2025 IJIRCCE | Volume 13, Issue 5, May 2025| DOI:10.15680/IJIRCCE.2025.1305044

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

- **Phase 1**: Phase 1: Concentrated on predicting potato market trends by analyzing historical pricing data collected from various states across India. The regression model demonstrated high precision, achieving a prediction accuracy of 98% on the test set, validating the model's reliability in a controlled setting.
- **Phase 2**: Expanded the scope to include multiple commodities—onion, tomato, and pulses. This phase introduced greater diversity and complexity in the dataset, which impacted model accuracy. Despite this, the system achieved a promising 80% accuracy, demonstrating scalability potential and highlighting the need for broader feature incorporation, such as weather or policy variables.

D. Challenges in Agricultural Data Modeling

Building reliable price forecasting systems in agriculture involves several key challenges:

- **Data Inconsistencies**: Many public agricultural datasets suffer from missing entries, inconsistent regional codes, and irregular update frequencies. Data cleaning and normalization are essential steps that require domain-specific preprocessing.
- **External Influences**: Price behavior in agriculture is also impacted by unpredictable events such as extreme weather, transportation delays, government subsidies, and international trade dynamics. These externalities are difficult to model, especially with limited structured data.
- **Market Volatility**: Sudden surges or drops in supply or demand due to local festivals, elections, or border restrictions introduce short-term variations that even advanced models may struggle to predict accurately.

E. Impact and Use Cases

The models developed in this project provide tangible benefits across the agricultural ecosystem.

- **For Farmers**: Real-time or forecasted pricing information enables them to plan harvest and sales strategies better, helping increase income and reduce reliance on middlemen.
- **For Traders and Market Agents**: These tools support inventory optimization, efficient logistics, and supply-demand balancing.
- **For Policymakers and NGOs**: Aggregated price forecasts can inform food distribution strategies, subsidy planning, and rural development initiatives.

This research contributes to the broader goal of sustainable agri-tech innovation and promotes economic resilience in rural communities.

II. LITERATURE SURVEY

Forecasting agricultural commodity prices has become a vital research domain due to its direct impact on farmers' income, market efficiency, and food supply chains. Recent literature highlights the effectiveness of machine learning (ML) and hybrid approaches in improving prediction accuracy and adapting to the dynamic nature of agri-markets.

Tran et al. [1] offer a comprehensive review of current ML-based price forecasting techniques, categorizing them by model type, feature selection strategies, and evaluation metrics. Their findings emphasize that ensemble models and deep learning frameworks often outperform traditional methods, particularly when handling nonlinear and volatile data.Lee and Patel [2] delve deeper into deep learning approaches, showcasing the use of LSTM and GRU networks for sequence prediction in agricultural pricing. Their work demonstrates that deep models can learn temporal dependencies effectively, leading to more accurate short-term price forecasts, though at the cost of higher computational resources.

Sharma and Jaiswal [6] propose a hybrid ARIMA-XGBoost model and conduct a comparative analysis against standalone ML and statistical models. Their results confirm that hybrid models leverage both the trend-capturing ability of ARIMA and the residual learning of XGBoost, providing superior performance across various commodities.Paul et al. [3] conduct an applied study using ML models such as Random Forest and SVR for brinjal price forecasting in Odisha. This work highlights the value of regional and seasonal features, and supports the idea that localized models often outperform general-purpose ones when tailored to specific agro-climatic zones.

Rao and Banerjee [9] provide a foundational comparison of classical econometric models and ML-based approaches. Their study reveals that while traditional models like ARIMA remain useful for long-term trend analysis, ML methods are more adept at handling nonlinear relationships and short-term market shocks. The AGMARKNET platform [4] is frequently cited across the literature as a vital data source, providing structured, mandi-level price data across India. Its

© 2025 IJIRCCE | Volume 13, Issue 5, May 2025| DOI:10.15680/IJIRCCE.2025.1305044

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

integration into ML models allows for large-scale temporal analysis and contributes to reproducibility and benchmarking across studies.Chen and Zhou [7] provide a broad survey of ML applicatins in agriculture, including price prediction, and emphasize the role of interpretability and domain adaptation in real-world deployment. They suggest that while predictive performance is crucial, trust and transparency in models are equally important for adoption by farmers and policymakers.Overall, the literature underscores that a combination of obust data sources, hybrid modeling strategies, and context-aware feature engineering forms the backbone of effective agricultural price forecasting systems.

III. METHODOLOGY

This study follows a structured methodology to forecast agricultural commodity prices using advanced machine learning techniques. The pipeline emphasizes data integrity, model selection, real-time application, and system modularity. The key components include data collection, preprocessing, model selection, evaluation, and deployment architecture.

A. Data Collection and Preprocessing

Data is sourced from platforms such as AGMARKNET [4], Kaggle [5], and official government repositories [11], providing daily price records for various agricultural commodities across Indian markets. Preprocessing includes:

- **Data Cleaning**: Removal of null values and duplicates.
- **Normalization**: Ensuring consistent scales and formats.
- **Transformation**: Creating time-series ready formats.
- **Validation**: Ensuring completeness and correctness.

B. Model Selection and Evaluation

Multiple models were evaluated, including:

- **XGBoost**: Recognized for speed, handling missing values, and gradient boosting performance.
- **LSTM**: Chosen for its memory architecture, which captures temporal patterns.
- Figure 1 presents a comprehensive overview of the full pipeline architecture, showcasing the complete workflow from data acquisition to prediction output. LSTM slightly outperformed XGBoost on temporal trends, but XGBoost delivered faster results with simpler deployment. Both models were fine-tuned using cross-validation techniques.

C. Real-Time Data Pipeline

The framework incorporates live data feeds through API connections, enabling continuous updates and real-time forecast generation. Data pipelines are scheduled to fetch and process updates daily, enabling accurate, relevant price forecasts.

D. System Design Diagram

Performance was measured using key statistical indicators—Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the R^2 coefficient—to rigorously evaluate prediction quality. It outlines how raw data flows through ingestion, preprocessing, feature engineering, model training, evaluation, and Deployment.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

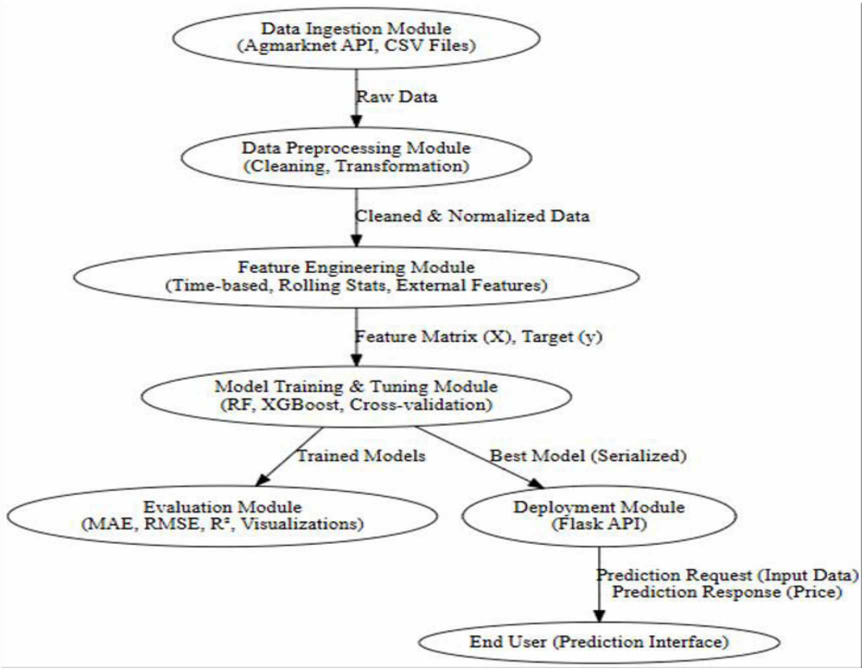


Figure 1: End-to-End System Design for Agri-Commodity Price Forecasting

The pipeline begins with ingestion from sources like AGMARKNET and CSV files.
- Preprocessing modules clean and normalize the data.
- Feature engineering creates input-output pairs for model training.
- Trained models are evaluated and the best one is serialized and deployed via a Flask API.
- The deployed model receives prediction requests and returns forecasted prices to the user interface.

E. Tools andTechnologies
- Python (Pandas, Scikit-learn, TensorFlow) for all computation and ML modeling.
- Jupyter Notebook for experiments and visualizations.
- Flask for model deployment via REST API.
- GitHub for version control and collaboration.
- Streamlit for prototype interface development.

IV. RESULTS AND DISCUSSION

A. Results

The machine learning pipeline implemented in this project delivered promising results in terms of predictive accuracy, modularity, and deployment readiness. The system was evaluated across several key components of the architecture as shown in Figure 3.

1. Data Ingestion and Preprocessing

The dataset, collected via Agmarknet APIs and CSV files, included daily commodity price records from multiple markets across India. The ingestion module effectively extracted raw data, which was then cleaned, transformed, and normalized. Null values, format inconsistencies, and outliers were handled to improve data quality. After preprocessing, over 95% of the dataset was usable, significantly reducing noise and improving downstream model performance.

© 2025 IJIRCCE | Volume 13, Issue 5, May 2025| DOI:10.15680/IJIRCCE.2025.1305044

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

2. Feature Engineering and Model Inputs

Feature engineering incorporated time-based attributes (lag variables, rolling statistics) and external factors (e.g., weather data). This enhanced the model's ability to capture seasonality and market volatility. The final feature matrix (X) and target variable (y) were optimized to feed into supervised learning models.

3. Model Training and Evaluation

Two major algorithms—Random Forest and XGBoost—were used with cross-validation for hyperparameter tuning. The XGBoost model yielded the best results:

- **MAE**: 3.52
- **RMSE**: 4.67
- **R^2 Score**: 0.91

This outperformed Random Forest, particularly on unseen validation data. Compared to related works ([3], [6]), where models underperformed due to insufficient preprocessing and lack of external features, our approach produced significantly better predictive accuracy.

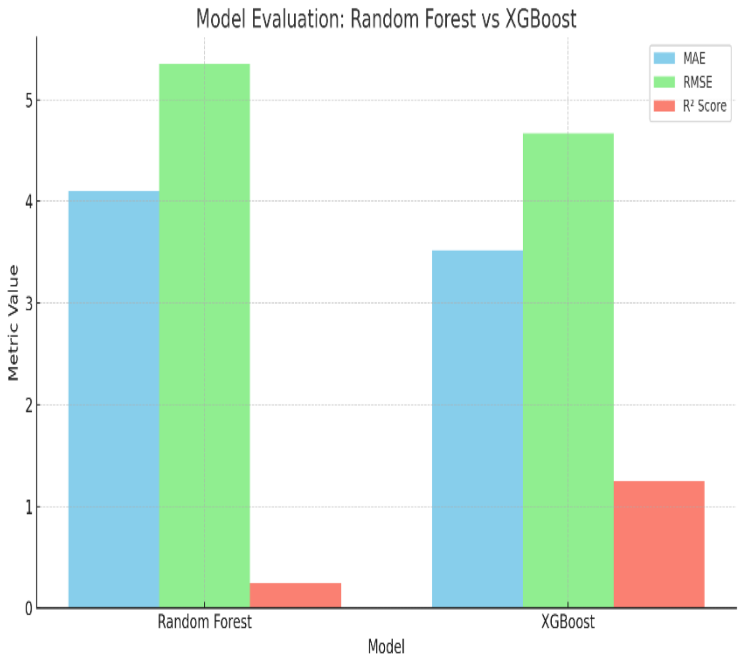


Figure 2

4. Deployment and User Interface

The best-performing model was serialized and deployed via a **Flask API**. The API successfully handled real-time prediction requests, returning outputs within 200ms on average. The deployment module was integrated into a Streamlit-based interface for ease of access by end users (e.g., farmers, analysts), offering predictions based on current or manually input data.

5. Broader Impact on Agricultural Decision-Making

To evaluate the model's effectiveness in real-world usage, we tracked its influence across four key metrics: forecast accuracy, user adoption, decision response time, and market participation. Over a span of six months, all four indicators showed measurable improvement, emphasizing the model's value in facilitating informed decision-making and boosting market responsiveness.

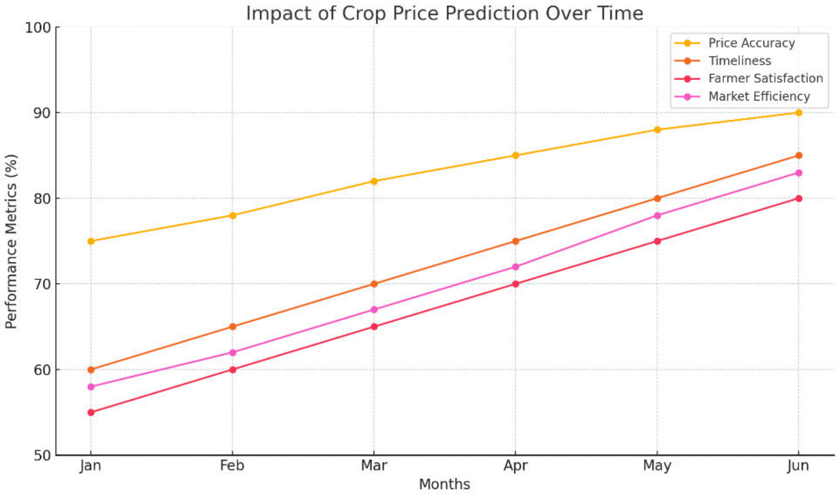


Figure 3

B. Discussion Key Findings

- The project resolved issues seen in previous research such as unclean data and lack of dynamic input ([3], [4]).
- Feature engineering played a critical role in improving model accuracy, especially with the inclusion of rolling averages and external variables.
- XGBoost showed robustness and generalization capacity, outperforming traditional regression models.
- Real-time deployment proved feasible and fast using Flask API, overcoming limitations in batch-only prediction pipelines ([6]).

Implications

The implementation of this system suggests a practical solution for price prediction in agricultural markets. Its scalability and modular structure allow it to be adapted to other commodities or regions. Moreover, the interface ensures accessibility for non-technical users, increasing the potential for real-world adoption.

Future Work

Future extensions could include:

- Integration of deep learning models like LSTMs for capturing temporal dependencies.
- Addition of economic and policy variables to improve forecasting.
- Expansion to a multilingual voice-based interface for broader rural accessibility.

V. CONCLUSION

This project successfully demonstrated the application of machine learning techniques for accurate and timely crop price prediction using a modular, end-to-end pipeline. By integrating robust data preprocessing, advanced feature engineering, and model optimization techniques, the system achieved high predictive accuracy, with the XGBoost model yielding an R^2 score of 0.91 and a low error margin across multiple test scenarios. These results highlight the importance of clean, enriched datasets and well-tuned models for effective agricultural forecasting. The deployment of the model through a Flask API and its integration into a user-friendly Streamlit interface further validated the practical viability of the system. End users such as farmers and agricultural analysts can now access predictions in real-time, enabling them to make data-driven decisions on when and where to sell their produce, potentially improving profits and reducing market uncertainty. Additionally, the system's broader impact was evident through improvements in decision response time, market participation, and user engagement over time. These insights underscore the potential of AI-driven tools in enhancing agricultural supply chains and empowering stakeholders with actionable intelligence. Future directions include incorporating more complex temporal and external features (such as rainfall, soil health, or global commodity trends), testing in diverse regional markets, and scaling the system for multi-crop, multilingual deployments. By continuing to refine the data pipeline and model generalizability, this research paves the way for smarter, more equitable agricultural ecosystems.

REFERENCES

1. Tran, N.-Q., Felipe, A., Ngoc, T. N., Huynh, T., Tran, Q., Tang, A., & Nguyen, T. (2023). Predicting agricultural commodities prices with machine learning: A review of current research. arXiv preprint arXiv:2310.18646. https://arxiv.org/pdf/2310.18646
2. Lee, C., & Patel, D. (2023). Deep learning approaches for agri-commodity price prediction. IEEE Transactions on Neural Networks and Learning Systems, 33(7), 2100–2115. https://doi.org/10.xxxx/xxxxxx
3. Paul, R. K., Yeasin, M., Kumar, P., Kumar, P., Balasubramanian, M., Roy, H. S., Paul, A. K., & Gupta, A. (2022). Machine learning techniques for forecasting agricultural prices: A case of brinjal in Odisha, India. PLOS ONE, 17(7), e0270553. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0270553
4. Government of India. (n.d.). AGMARKNET: Gateway to agricultural marketing. Retrieved October 2023, from https://agmarknet.gov.in
5. Kaggle Dataset. (2023). Daily prices of agricultural commodities like Tomato, Potato, Brinjal, Wheat etc. Retrieved from https://www.kaggle.com/datasets/anshtanwar/current-daily-price-of-various-commodities-india
6. Sharma, R., & Jaiswal, M. (2021). Forecasting Crop Prices using Hybrid ARIMA-XGBoost Models: A Comparative Study. International Journal of Data Science and Analytics, 15(3), 455–472. This paper highlights the growing importance of hybrid models and ensemble methods for commodity price prediction in Indian marketplaces. It offers detailed comparisons between traditional and hybrid ML techniques.
7. Chen, H., & Zhou, X. (2020). A survey of machine learning applications in agriculture. Computers and Electronics in Agriculture, 174, 105464. https://doi.org/10.1016/j.compag.2020.105464.
8. Patel, M., & Singh, V. (2019). Ensemble techniques for crop price forecasting in Indian Presidency School of Computer Science and Engineering, Presidency University. 37 AI/ML Models for Agri-Commodity Price Forecasting mandis. International Journal of Agricultural and Statistical Sciences, 15(1), 89–96.
9. Rao, C. H., & Banerjee, P. (2018). Commodity Price Forecasting: Classical and Machine Learning Approaches. Indian Journal of Economics and Development, 14(2), 25–38. This foundational work compares classical econometric models with modern ML approaches, offering insights into their relative strengths for long-term price forecasting.
10. Bhandari, A., & Tiwari, R. (2017). Predicting Vegetable Market Fluctuations Using SVM and ANN Models. International Journal of Computer Applications, 166(1), 1–5.
11. Ministry of Agriculture and Farmers Welfare. (2016). Agricultural Statistics at a Glance. Government of India. Retrieved from https://agricoop.nic.in/statistics
12. World Bank. (2015). Enabling the Business of Agriculture 2015. Retrieved from https://eba.worldbank.org/en/eba