*International Journal of Scientific Research and Engineering Development-– Volume 7 Issue 6, Nov-Dec 2024*

*Available at **www.ijsred.com***

RESEARCH ARTICLE OPEN ACCESS

# Commodity Trend Analysis and Prediction Using Machine Leaning

Mr. Mohit Wani*, Mr. Yash Patil**, Dr. Varsha Kshirsagar***

*(Electronics and Telecommunication, Savitribai Phule Pune University, Pune, Maharashtra, India
Email: wanimohit121@gmail.com)

**(Electronics and Telecommunication, Savitribai Phule Pune University, Pune, Maharashtra, India
Email : yashpkp34@gmail.com)

***(Electronics and Telecommunication, Savitribai Phule Pune University, Pune, Maharashtra, India
Email : varshak.phd@gmail.com)

----------------------------------------************************----------------------------------

## Abstract:

This literature survey examines the evolution of prediction methods for agricultural commodity prices, assessing various approaches and highlighting current trends and challenges. Despite their widespread use, traditional statistical techniques such as linear regression and ARIMA models often struggle to handle the non-linear characteristics, cyclical patterns, and complex interrelationships typical of agricultural markets. Machine learning techniques, including Support Vector Regression (SVR), Extreme Learning Machines (ELM), and neural network architectures, particularly Long Short-Term Memory (LSTM), have emerged as promising alternatives to capture these complexities. Combination models, which integrate multiple forecasting methods, have also been explored to enhance prediction accuracy.

Web crawling techniques to gather data from online sources, such as market platforms and news websites, expanding the range of potential predictive features. By combining this information with conventional datasets, including past price records and climate statistics a more thorough insight into market forces can be achieved. Sentiment analysis derived from social media and news articles as additional features, reflects the growing influence of public opinion. The review reports the impact of external factors, like global events and climate change, on agricultural prices. Incorporating these factors into forecasting models is crucial for accurate predictions and informed policy decisions.

*Keywords* — **Machine Learning, Optimization, Deep Learning, Prediction, Agriculture, Commodity.**

----------------------------------------************************-------------------------------
---

## I. INTRODUCTION

Agriculture is the backbone of many economies and countries, especially in developing countries, where a significant portion of the population relies on farming for their livelihood. The prices of agricultural commodities play an important role in determining the economic stability of the region Predicting these prices is essential for various stakeholders, including farmers, traders, policymakers, and consumers. Accurate price predictions can help farmers make informed decisions about planting, harvesting, and selling their crops. Traders can create profitable trading strategies while policymakers would be able to design effective policies to stabilize markets and ensure food security. Consumers can prepare for potential price hikes and adjust their spending accordingly. However, the volatile nature of agricultural commodity prices presents significant challenges for accurate forecasting. Prices are influenced by a complex interplay of factors like Supply and demand dynamics (i.e. Production levels, inventory levels, import and export volumes), Meteorological conditions (Temperature, rainfall, and natural disasters can significantly impact crop yields, influencing prices), Economic factors (Global economic growth, currency exchange rates, energy prices, and inflation), Government policies (Trade policies, subsidies, and price controls can influence market dynamics) and also Market sentiment (Speculation and investor behavior can contribute to price volatility). Effective storage management is crucial for mitigating losses due to spoilage and optimizing market timing. By correlating predicted price trends with storage capacity and conditions, we can provide stakeholders with actionable insights on when to store, sell, or release commodities to maximize profits.

## II. MOTIVATION

The primary motivation behind this review paper is to analyse the current technologies and models in the domain. So, further work would benefit to reduce food wastage by providing more accurate predictions of commodity prices, allowing farmers and suppliers to optimize their production and distribution processes. Additionally, promoting a stable exchange rate for commodities, ensuring that price volatility does not harm local economies or global trade. By empowering farmers with actionable insights on market trends, the project aims to increase the median income of farmers, creating a more sustainable and profitable agricultural sector, which in turn could help prevent extreme outcomes for farmer by reducing financial stress and uncertainty.

## III. LITERATURE SURVEY

### A. Comparative Study of EMD based Modelling Techniques for Improved Agricultural Price Forecasting.

Agricultural commodity price forecasting is challenging due to non-linear and non-stationary nature of the data. This research aims to improve agricultural price prediction using Empirical Mode Decomposition (EMD) modeling techniques. The researchers analyzed monthly wholesale potato prices from three Indian markets between January 2005 and December 2020. They evaluated various models, including ARIMA, TDNN, SVR, EMD-SVR, and EMD-TDNN. Results showed that EMD-based models are better than traditional methods like ARIMA, especially in markets such as Agra and Bangalore where non-linear price patterns exist. However, TDNN is expected to handle non-linear data well but its inability to manage non-stationarity had limited its effectiveness. So, it highlights the necessity of considering data characteristics in selecting forecasting models.

### B. Automated Agriculture Commodity Price Prediction System with Machine Learning Techniques

The paper works upon the lack of research on the Malaysian commodities market compared to the US and China. It proposes a system which is integrated web-based model for predicting agricultural commodity prices in Malaysia. The proposed system predicts prices using machine learning techniques like ARIMA, SVR, XGBoost, and LSTM. The system was implemented using Python and Google Colab, a free cloud service that supports GPU, enabling the development of deep learning applications. The authors stress the importance of selecting the optimal algorithm for prediction, taking into account factors like accuracy and the ability to handle increasing data volumes. Future work aims to incorporate location-specific price analysis to facilitate trading among farmers. It also mentions the use of ARIMA for forecasting agricultural prices and discusses statistical methods for time series analysis.

### C. Agricultural Product Price Forecasting Methods

This extensive review article provides a comprehensive overview of agricultural product price forecasting methods, propagating the advancements from traditional methods like regression analysis to intelligent methods like machine learning and deep learning. The article discusses various traditional forecasting methods like linear regression, along with their limitations and suitability for specific forecasting problems. It also examines integration methods for combining multiple models, such as equal weighting, minimum variance, dominance matrix, and least squares estimation. The points or stages the authors stressed upon were feature selection, model development and optimization for precise results. They suggest that future research should focus on incorporating external factors, improving data quality and accessibility, and exploring hybrid models. The article discusses the use of various forecasting methods in the context of agricultural prices, including ARIMA, exponential smoothing, support vector machines, Bayesian networks, and neural networks. It also explores the use of data from different sources, including web-based sentiment analysis and text-based data, to improve forecasting accuracy.

### D. Forecasting Agricultural Commodity Prices Using Dual Input Attention LSTM

This paper points out the challenges of predicting the agricultural commodity prices, particularly focusing on cabbage and radish in Korea. The study introduces a Dual Input Attention LSTM (DIA-LSTM) model that utilizes both input and temporal attention mechanisms to improve forecasting precision. The authors compare the performance of their proposed model against several benchmark models, including simple LSTM, GCN-LSTM, STL-ATTLSTM, and DSA-LSTM. Their findings indicate that DIA-LSTM yields the lowest RMSE and MAPE for both cabbage and radish, showcasing its superior predictive capabilities. The research underscores the significance of attention mechanisms in capturing temporal dependencies and input variable characteristics for agricultural commodity price forecasting. It cites references supporting the use of LSTM for addressing the vanishing gradient problem and explores the effectiveness of various deep learning and statistical techniques for time series forecasting.

### E. Forecasting-Commodity-Prices

This research delves into the application of machine learning and linear regression techniques for predicting commodity prices, emphasizing the importance of accurate price forecasting in agricultural decision-making. Daily life and agricultural productivity are directly influenced by the price fluctuation of agricultural commodities in the market. The study examines linear regression as a statistical method for modelling the relationship between dependent and independent variables. The authors also emphasize the importance of data visualization and exploratory data analysis in understanding key relationships in data.

### F. Agricultural Price Fluctuation Model Based on SVR

This research presents a Support Vector Regression (SVR) model for examining and forecasting agricultural product price variations in China. The authors points that analyses of price fluctuations using traditional models is hard and SVR model is proposed to overcome these difficulties. Various factors like influencing agricultural prices, including meteorological factors, input factors, and resident consumption levels are taken into consideration in this study. The authors analysed daily, monthly, and annual price data for major agricultural products in China over a 15-year period. The study concludes that the SVR model is a viable approach for predicting agricultural product prices and suggests further research to improve the accuracy of machine learning models for this purpose.

### G. Various optimized machine learning techniques to predict agricultural commodity prices

This study examines the effectiveness of various optimized machine learning techniques for predicting agricultural commodity prices, addressing the increasing global food demand. The authors propose using Auto-ARIMA, GA-LSTM, and GA-ELM models for forecasting commodity prices. They analyze daily price data for wheat, corn, sugar, soybeans, rice, coffee, cotton, cocoa, oat, lumber, and orange juice, obtained from reputable sources like CBOT and ICE-US. The study evaluates the performance of these models using various metrics, including RMSE, MAE, and R-squared. It found that the GA-ELM model consistently outperforms the other two models. The authors emphasize the importance of considering worldwide events and their impact on commodity price forecasting. The study cites numerous research articles discussing various aspects of agricultural commodity price forecasting, including the use of machine learning techniques, deep learning models, and optimization algorithms.

### H. An Empirical Study on Spot and Futures Market Price of Soybean, Soybean Oil and Soybean Meal in China

This empirical study focuses on analyzing the spot and futures market prices of soybean, soybean oil, and soybean meal in China. The researchers review existing literature on international and domestic soybean market studies. They explore the price relationships between soybean and its derivatives, such as soybean meal and soybean oil. The study investigates the price discovery function of the futures market and its impact on spot prices. The authors use various statistical methods, including correlation analysis and co-integration tests, to analyze the relationships between spot and futures prices.

### I. The Impact of Excess Liquidity on the Domestic Food Price Inflation:

This study examines the impact of excess liquidity on domestic food price inflation in China, using a structural VAR method to analyze the influence of various factors on food prices. It explores the consequences of U.S. excess liquidity, global commodity prices, and China's internal liquidity on food price inflation. The researchers argued that lenient monetary policies contribute to inflationary pressure on agricultural prices The study emphasized that understanding the monetary policy transmission mechanism in the agricultural market is crucial for controlling rising agricultural prices. The authors suggest further research on the adjustment process of agricultural product prices and the elasticity of different agricultural products.

### J. Forecasting Agricultural Commodity Prices Using Model Selection Framework with Time Series Features and Forecast Horizons

This study introduces a model selection framework for predicting agricultural commodity prices, incorporating time series features and forecast horizons to enhance prediction accuracy. The authors use twenty-nine features to describe agricultural commodity prices and compare the performance of three intelligent models: artificial neural network (ANN), support vector regression (SVR), and extreme learning machine (ELM). They used random forest (RF) and support vector machine (SVM) as classifiers to learn the relationships between features and model performance. The study utilizes the minimum redundancy and maximum relevance approach (MRMR) to reduce feature redundancy and enhance forecast accuracy. The experimental results show that the proposed model selection framework outperforms the optimal candidate model and simple model averaging. The authors stress the significance of feature reduction and the role of forecast horizons in model selection. The study cites numerous studies on agricultural commodity price forecasting using various methods, including statistical models, intelligent models, and meta-learning approaches.

### K. The Crawling and Analysis of Agricultural Products Big Data based on Jsoup:

This research presents a platform for crawling, extracting, and analysing agricultural product big data from the internet to tackle the challenge of balancing agricultural production and market demand. The authors used Jsoup (HTML collector), as the main technology for data crawling and extraction. They describe the process of data crawling, including modeling internet web pages as a directed graph and employing a breadth-first crawling strategy. The study employs correlation analysis and regression analysis to study factors influencing price changes and market trends. The authors highlight the importance of their platform in providing valuable market information to consumers and supporting the growth of agricultural businesses and e-commerce industries.

### L. Decision Making Support System for Prediction of Prices in Agricultural Commodity

  

This study focuses on developing a decision-making support system for predicting agricultural commodity prices in India, aiming to minimize agribusiness risk and assist farmers in making informed decisions. The necessity of predicting the agricultural commodities is put to light for better insights. The study proposes using soft computing techniques for predicting agricultural commodity market prices. The authors aim to provide farmers with a system that allows them to access historical price data and make informed decisions for bidding on their crops.

### M. Discovery_and_Prediction_of_Stock_Index_Pattern_via_Three-Stage_Architecture_of_TICC_TPA-LSTM_and_Multivariate_LSTM-FCNs

This study presents a three-stage architecture for discovering and predicting stock index patterns using multivariate time series analysis. The authors contended that traditional econometric models like ARIMA are inadequate for capturing the complex, non-stationary nature of stock index data. The proposed architecture combines Toeplitz Inverse Covariance-Based Clustering (TICC), Temporal Pattern Attention and Long-Short-Term Memory (TPA-LSTM), and Multivariate LSTM-FCNs (MLSTM-FCN and MALSTM-FCN) to discover and predict stock index patterns. The study uses the Hangseng Stock Index and eleven industrial sub-indices to demonstrate the effectiveness of the proposed architecture. The authors compare the performance of their approach with various machine learning methods, including Naive Bayes Classifier (NB), Support Vector Machine Classifier (SVM), Random Forest (RF), and XGBoost (XGB), and find that their method outperforms these benchmarks.

### N. Multivariate Financial Time Series Prediction with Certified Robustness

This study proposes a robust forecasting framework called DP-MAELS for predicting agricultural commodity futures prices, addressing the limitations of deep neural networks (DNNs) in handling multivariate time series data. They addressed these challenges by incorporating improved deep neural networks and a certified noise injection mechanism. The DP-MAELS framework utilizes a multimodal variational autoencoder for feature extraction, Long- and Short-Term recurrent neural networks for forecasting, and a differential privacy-inspired noise injection mechanism for enhancing robustness and accuracy. Using real-world agricultural commodity futures price time series and external data, the authors evaluated and showed promising performance compared to the other benchmark methods.

## IV. BLOCK DIAGRAM

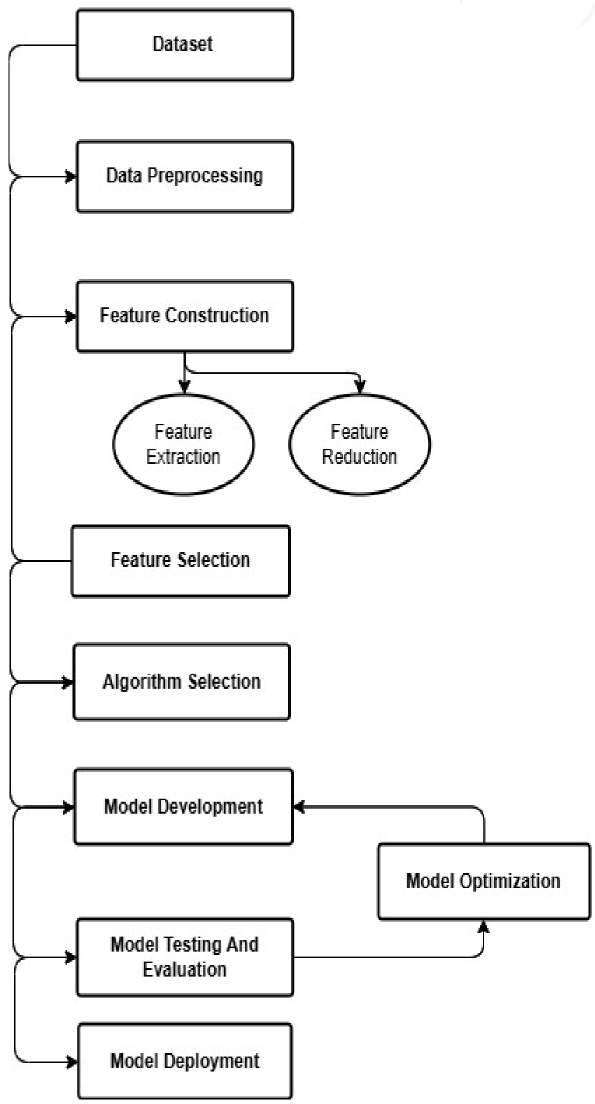


Fig. 1 Block Diagram

  

## V. METHODOLOGY

To ensure a comprehensive and systematic review, we searched a variety of academic databases and repositories. The primary sources included Google Scholar, IEEE Xplore, etc. Emphasis on studies published between 2010 and 2023 to ensure the inclusion of the most recent research and also studies conducted with empirical data or real-world applications of forecasting models. We initially searched and identified 45 articles. After analysing, we conducted a title and abstract screening to assess relevance. In total, 14 studies were selected for a full-text review. From the selected studies, we extracted and analysed the key points like research methodology, key findings and commodity focus (the agricultural commodities analysed). For understanding and explaining complex models, such as machine learning algorithms and econometric techniques, we utilized generative AI tools (e.g., Gemini). These tools assisted in clarifying the procedures, summarizing model descriptions, and providing simplified explanations of various forecasting methodologies used in the reviewed studies.

Also based on the following points and parameters for the structured review paper:

1. Statistical methods for agricultural commodity price prediction can be limited but acted as benchmark for further development

   **a.** ARIMA (Autoregressive Integrated Moving Average): ARIMA models can capture linear trends and seasonality in time series.

   b. SARIMA (Seasonal Autoregressive Integrated Moving Average): SARIMA models, which build upon the ARIMA model, account for both trend and seasonal components by using a seasonal difference operator.

2. Emphasis on Machine Learning and Deep Learning Techniques

   The importance and increasing use of machine learning and deep learning methods for predicting agricultural commodity prices. These techniques offer advantages over traditional statistical methods like ARIMA in handling complex, non-linear relationships and capturing long-term dependencies in time series data.

3. Explore Hybrid and Ensemble Models
   It can be seen that hybrid (combining different models) can increase the accuracy and robustness of models. For example:

   a. EMD-based Hybrid Models: Using Empirical Mode Decomposition (EMD) to break down the price series into Intrinsic Mode Functions (IMFs) and a residue. Each component is then predicted using either an Artificial Neural Network (ANN) or Support Vector Regression (SVR), and the final forecast is obtained by summing up the individual predictions.
   b. Dual Input Attention LSTM: Dual Input Attention Long Short-Term Memory (LSTM) model incorporates both price data and relevant external factors, like meteorological data.
   c. Optimized Hybrid Models: Suggestion of the use of optimization techniques like Genetic Algorithm (GA) to tune the parameters of machine learning models like ELM and LSTM for improved prediction accuracy.

4. Evaluation of Model Performance

   a. Appropriate Metrics: Using combination of evaluation metrics, such as Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), and R-squared, to assess model accuracy and compare different forecasting methods.
   b. Comparative Analysis: The sources consistently emphasize the importance of comparing different models and techniques to identify the best-performing approach for specific

  

agricultural commodities and forecast horizons.

5. The accountability and responsibility of commodity price prediction: Taking into factor the potential impact of prediction models on market manipulation, food security, crop failure and fair pricing practices.

6. The specific agricultural commodities of interest: Various variety of commodities, including soybeans, wheat, corn, sugar, rice, cabbage, and radish. We may funnel down the focus on commodities of most importance to the region, demand/supply, data available and volatility.

## VI. CONCLUSION

In conclusion, predicting agricultural commodity prices remains a complex but essential challenge in the agricultural and economic sectors. This review examined a range of methodologies, spanning traditional statistical models like ARIMA to more sophisticated machine learning techniques like LSTM, SVR, and hybrid approaches. Further development would be focus on improving model robustness, refinement, incorporating more diverse and real-time data sources, and addressing the practical application of these models for farmers, traders, and policymakers considering the core Indian markets and its challenges.

## VII. REFERENCES

[1] Ghose, B., Pandit, P., Mazumder, C., Sinha, K., & Sahu, P. K., "Comparative Study of EMD based Modelling Techniques for Improved Agricultural Price Forecasting," *Journal of the Indian Society of Agricultural Statistics, 78(1), 53-62.,* New Delhi, 2024.

[2] Chen, Zhiyuan, Howe Seng Goh, Kai Ling Sin, Kelly Lim, Nicole Ka Hei Chung, and Xin Yu Liew, "Automated Agriculture Commodity Price Prediction System with Machine Learning Techniques.," *Advances in Science, Technology and Engineering Systems Journal (ASTES),* Malaysia, 2021.

[3] Feihu Sun, , Xianyong Meng, Yan Zhang, Yan Wang, Hongtao Jiang, and Pingzeng Liu, "Agricultural Product Price Forecasting Methods," *Agriculture (MDPI),* Taian, China, 2023.

[4] Yeong Hyeon Gu, Dong Jin, Helin Yin, Ri Zheng, Xianghua Piao, and Seong Joon Yoo, "Forecasting Agricultural Commodity Prices Using Dual Input Attention LSTM," *Agriculture (MDPI),* Seoul, Korea, 2022.

[5] G M Vinay Sreekar Reddy, Sahana Gowda Kr, Bhavana V, Ankitha More B, Poojitha G, "Forecasting Commodity Prices," *International Journal Of Scientific Research In Engineering And Management(IJSREM),* Bangalore, India, 2024..

[6] Wang Shengwei, Li Yanni, Zhuang Jiayu, and Liu Jiajia, "Agricultural Price Fluctuation Model Based on SVR," *9th International Conference on Modelling, Identification and Control (ICMIC),* Kunming, China, 2017.

[7] Murat Sari, Serbay Duran, Huseyin Kutlu, Bulent Guloglu, Zehra Atik, "Various optimized machine learning techniques to predict agricultural commodity prices," *Neural Computing and Applications, Istanbul, Turkey,* 2024.

[8] L. -h. Li and Q. Zhang, "An Empirical Study on Spot and Futures Market Price of Soybean, Soybean Oil," *IEEE 18th International Conference on Industrial Engineering and Engineering Management,* Changchun, China, 2011.

[9] L. Yang and D. Jiang, "The impact of excess liquidity on the domestic food price inflation," *ICSSSM,* Tianjin, China, 2011.

[10] D. Zhang, S. Chen, L. Liwen and Q. Xia, "Forecasting Agricultural Commodity Prices Using Model Selection Framework With Time Series Features and Forecast Horizons," *in IEEE Access,* Guangdong, China, 2020.

[11] Jie Wang, Shuo Yang, Yuezhi Wang and Cheng Han, "The crawling and analysis of agricultural products big data based on Jsoup," *12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD),* Zhangjiajie, China, 2015.

[12] A. Vohra, N. Pandey and S. K. Khatri, "Decision Making Support System for Prediction of Prices in Agricultural Commodity," *2019 Amity International Conference on Artificial Intelligence (AICAI),* Dubai, United Arab Emirates, 2019.

[13] H. Ouyang, X. Wei and Q. Wu, "Discovery and Prediction of Stock Index Pattern via Three-Stage Architecture of TICC, TPA-LSTM and Multivariate LSTM-FCNs," *in IEEE Access,* Wuhan, China, 2020.

[14] H. Li, Y. Cui, S. Wang, J. Liu, J. Qin and Y. Yang, "Multivariate Financial Time-Series Prediction With Certified Robustness," *in IEEE Access,* Beijing, China, 2020.