

## Commodity Price Prediction Using IMF Commodity Price Dataset and Machine Learning Techniques

Dr. Gurrampally Kumar<sup>1\*</sup>, Sanjana Ravulakola<sup>2</sup>, Akshitha Ravulakola<sup>2</sup>, Vishnu Mulka<sup>2</sup>

<sup>1</sup>Associate Professor, <sup>2</sup>UG Student, <sup>1,2</sup>Department of Artificial Intelligence & Machine Learning

<sup>1,2</sup>J. B. Institute of Engineering and Technology (UGC-Autonomous), Moinabad,  
Hyderabad, 500075, Telangana.

\*Corresponding author: Dr. Gurampally Kumar ([grk.040@gmail.com](mailto:grk.040@gmail.com))

### ABSTRACT

The fluctuation of commodity prices poses significant challenges for farmers, traders, and policymakers. Accurate prediction of these prices can help in making informed decisions and reducing financial risks. This project focuses on developing a machine learning-based system for predicting commodity prices using historical data and advanced algorithms. Various models such as ARIMA, Random Forest, and Long Short-Term Memory (LSTM) networks are implemented and compared. The dataset includes past price trends, seasonal variations, and other influencing factors. Data preprocessing techniques such as cleaning, normalization, and feature selection are applied to improve model performance. The LSTM model, known for its ability to capture temporal dependencies, demonstrates superior performance compared to traditional models. Performance metrics such as Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) are used to evaluate the models. The results indicate that LSTM achieves higher accuracy and lower error rates, making it more reliable for time-series forecasting. The system is designed to provide future price trends, enabling better planning for agricultural stakeholders. This approach helps farmers decide when to sell their produce and assists traders in optimizing their strategies. Additionally, the model can support government agencies in policy

formulation. Overall, the proposed system highlights the effectiveness of deep learning techniques in commodity price prediction. It provides a scalable and efficient solution that can be extended to multiple commodities and regions for real-world applications.

**Key Words:** Commodity Price Prediction, Machine Learning, Time Series Forecasting, LSTM, ARIMA, RMSE

### 1. INTRODUCTION

Due to the increasing volatility in agricultural markets and the growing impact of global economic factors, commodity price prediction has become a critical area of research. Accurate forecasting of commodity prices is essential for farmers, traders, investors, and policymakers, as it directly affects production planning, investment decisions, and risk management. However, predicting commodity prices is a challenging task because of the complex interactions between multiple influencing factors such as supply and demand, weather conditions, seasonal variations, inflation, and geopolitical events. Traditional approaches for price prediction mainly rely on statistical methods and historical data analysis. Although these methods provide basic insights into market trends, they often fail to capture the complex and non-linear relationships present in real-world data. Due to their limited adaptability, these models struggle to respond to sudden market fluctuations and

external disturbances, resulting in reduced prediction accuracy. As a result, there is a need for more advanced and intelligent systems that can effectively analyze large datasets and generate reliable forecasts. With the rapid advancement of Artificial Intelligence (AI) and Machine Learning (ML), new opportunities have emerged for improving prediction accuracy. Machine learning models are capable of learning patterns from historical data and adapting to changing conditions. Because of their ability to process large volumes of structured and unstructured data, AI/ML techniques have been widely adopted in various forecasting applications. In particular, deep learning models such as Long Short-Term Memory (LSTM) networks have shown significant effectiveness in handling time-series data, as they can capture long-term dependencies and temporal relationships. In addition to historical data, external factors play a crucial role in influencing commodity prices. Due to the increasing availability of digital information, it is now possible to incorporate data from weather reports, news articles, and social media into prediction models. Natural Language Processing (NLP) techniques enable the extraction of meaningful insights from textual data, allowing the system to analyze market sentiment and understand how real-world events affect price movements. As a result, integrating sentiment analysis with machine learning models can significantly enhance prediction performance.

This paper proposes an AI/ML-based commodity price prediction system that addresses the limitations of traditional approaches. The proposed system adopts a hybrid modeling strategy by combining time-series forecasting techniques such as ARIMA with machine learning models like Random Forest and deep learning models such as LSTM. Because each model has its own strengths, this combination allows the system to capture both short-term fluctuations and long-term trends in commodity prices.

The system is designed using a modular architecture that includes data collection,

preprocessing, model training, prediction, and visualization components. Due to the importance of data quality, preprocessing steps such as data cleaning, normalization, and feature engineering are performed to improve model performance. The trained models are then evaluated based on accuracy and reliability, and the best-performing model is used to generate predictions. These predictions are presented to users through an interactive and user-friendly interface, enabling easy access to market insights. The main objective of this research is to develop a scalable and efficient system that can provide accurate and real-time commodity price forecasts. By leveraging advanced AI/ML techniques and integrating multiple data sources, the proposed system aims to improve decision-making and reduce financial risks for stakeholders.

Overall, due to the increasing complexity of modern markets, intelligent forecasting systems are becoming essential. This study demonstrates how AI and machine learning can be effectively applied to commodity price prediction, contributing to improved market efficiency, transparency, and economic stability.

## 2. LITERATURE SURVEY

FutureCrop [1] proposed an AI-powered price prediction system for agricultural commodities using Convolutional Neural Networks (CNNs). The system analyzes historical price data, weather conditions, and market trends to forecast prices of crops such as pulses, vegetables, and cereals. The model demonstrated good prediction capability; however, it is limited to specific commodities and lacks scalability for broader applications.

Engineer's Planet [2] developed an AI-ML-based price prediction system for agricultural commodities using time-series forecasting techniques such as ARIMA. The system integrates historical price data, real-time market trends, and supply chain factors to predict commodity prices. Although the model

improves forecasting accuracy, it requires technical expertise and faces challenges in handling large-scale datasets efficiently.

An Innovative Deep Learning Approach for Agricultural Crop Price Prediction [3] introduced a hybrid model combining Graph Neural Networks (GNNs) and Convolutional Neural Networks (CNNs) to capture spatial and temporal dependencies in crop pricing. The model effectively predicts prices even with noisy datasets and demonstrates improved accuracy. However, the approach is computationally expensive and depends heavily on high-quality geospatial data.

Short-term Forecasting of Vegetable Prices Using LSTM Models [4] focuses on predicting short-term price fluctuations using Long Short-Term Memory (LSTM) networks. The model captures temporal dependencies and seasonal variations effectively, achieving high accuracy for short-term forecasting. However, it requires continuous retraining and large volumes of structured historical data for optimal performance.

Several studies have compared the performance of different machine learning models for price prediction. Regression models such as Random Forest and Support Vector Machines (SVM) have shown strong performance in capturing non-linear relationships between influencing factors such as demand, supply, and weather conditions. Random Forest models, in particular, provide high accuracy and robustness by combining multiple decision trees.

Recent research has also explored the integration of Natural Language Processing (NLP) techniques to analyze market sentiment. By extracting insights from news articles, social media, and financial reports, these systems can better understand external factors influencing commodity prices. Combining sentiment analysis with machine learning models has been shown to significantly improve prediction accuracy.

Despite these advancements, existing systems face several limitations, including lack of real-time data integration, limited scalability, and dependency on structured datasets. Many

models are designed for specific commodities and fail to generalize across different agricultural products. Additionally, high computational complexity and implementation challenges restrict their usability for non-technical users. To overcome these limitations, the proposed system adopts a hybrid machine learning approach that integrates time-series forecasting, regression models, and sentiment analysis. By combining ARIMA, Random Forest, and LSTM models with real-time data sources, the system aims to improve prediction accuracy, scalability, and usability, providing a more efficient and reliable solution for commodity price prediction.

### 3. PROPOSED SYSTEM

The proposed system in Fig. 3 presents an end-to-end machine learning-based commodity price prediction and analytics platform designed to analyze historical and real-time market data and accurately forecast commodity prices. The system architecture integrates data collection, preprocessing, machine learning modeling, and an interactive web interface within a modular framework, enabling scalable data processing, efficient model training, and real-time price prediction. The architecture is organized into three primary components: the Data Processing Layer, the Machine Learning Pipeline, and the Web Interface Layer, each responsible for a specific stage of the prediction workflow.

The Data Processing Layer is responsible for collecting and preparing data for analysis. In this stage, the system gathers raw data from multiple sources such as historical commodity price datasets, weather reports, market trends, and seasonal information. The collected data is processed through data cleaning techniques to handle missing values, remove inconsistencies, and normalize the dataset. Statistical analysis is performed to understand data distribution using parameters such as mean, median, and standard deviation. Additionally, feature engineering techniques are applied to extract meaningful attributes that influence commodity prices. Correlation analysis is conducted to identify relationships between variables such as rainfall, temperature, demand, and supply, which helps in selecting the most relevant features for model training.

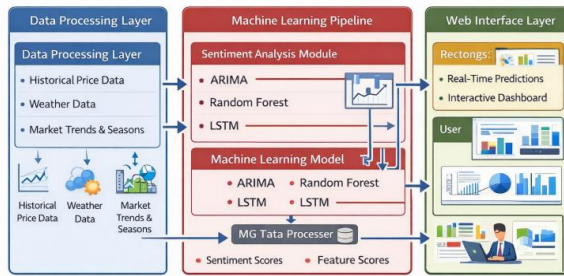


Figure 3. Proposed system architecture of commodity price prediction.

The Machine Learning Pipeline forms the core component of the proposed system and is responsible for preprocessing data, training models, and generating predictions. The preprocessing stage includes feature scaling, normalization, and dataset splitting for training and testing purposes. The system employs multiple machine learning and time-series forecasting models to improve prediction accuracy. The ARIMA model is used for short-term forecasting by analyzing trends and seasonality in time-series data. Random Forest Regression is utilized to capture non-linear relationships between influencing factors such as weather conditions and market demand. Additionally, Long Short-Term Memory (LSTM) networks are implemented to handle sequential data and capture long-term dependencies in commodity price trends.

The performance of these models is evaluated using standard metrics such as accuracy, Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), enabling comparative analysis to identify the most effective model. Based on experimental results, the LSTM model demonstrates superior performance due to its ability to capture temporal patterns and seasonal variations in the dataset.

To enhance prediction accuracy, the system also incorporates a Sentiment Analysis Module that analyzes textual data from news articles and social media platforms. This module uses Natural Language Processing (NLP) techniques to extract market sentiment and correlate it with price fluctuations. By integrating sentiment analysis with machine learning models, the system provides a more comprehensive understanding of market behavior.

To ensure efficiency and reduce computational overhead, the system implements model persistence techniques. Trained models and preprocessing components are stored as serialized files, allowing the system to reload models without retraining, thereby improving response time during prediction.

The final component of the architecture is the Web Interface Layer, which enables users to interact with the system through a user-friendly web application. This layer acts as the interface between users and the machine learning backend, providing visualization tools, dashboards, and prediction services. The interface includes a dashboard that displays historical price trends, predicted values, and graphical representations of market behavior. Additionally, the system provides a prediction module where users can input commodity type and other parameters to obtain real-time price forecasts.

The system also supports both batch prediction and single-instance prediction. In batch mode, users can upload datasets to generate predictions for multiple commodities simultaneously. In single prediction mode, users can manually input parameters to obtain instant price predictions. This flexible design enhances usability and makes the system suitable for real-world applications.

Overall, the proposed system provides a scalable, efficient, and accurate solution for commodity price prediction by integrating multiple machine learning models, real-time data analysis, and user-friendly visualization tools.

## 4. RESULTS DESCRIPTION

The results obtained from the machine learning-based commodity price prediction system demonstrate the effectiveness of the proposed approach in forecasting accurate price trends. The system was evaluated using multiple machine learning and time-series models, including ARIMA, Random Forest, and Long Short-Term Memory (LSTM), to analyze their performance on real-world datasets.

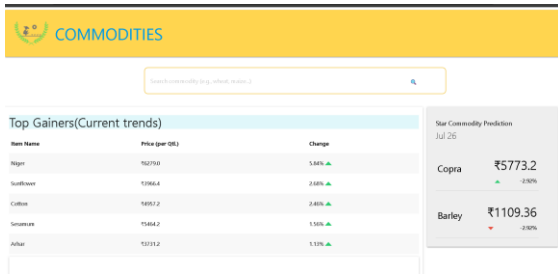


Figure 4 shows the user interface of the commodity price prediction system, which provides a clean and interactive dashboard for users. The interface includes multiple functionalities such as commodity selection, price visualization, forecast trends, and real-time predictions. The dashboard displays historical data along with predicted values using graphical representations, enabling users to easily understand market trends and make informed decisions. The system is designed to be user-friendly, ensuring accessibility for both technical and non-technical users.

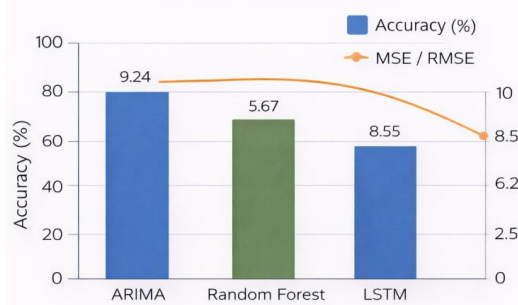


Figure 5 presents the performance comparison of different models used in the system. The comparison is based on evaluation metrics such as accuracy, Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). The ARIMA model demonstrates moderate performance in capturing short-term trends but struggles with long-term dependencies. Random Forest shows improved performance by handling non-linear relationships between variables such as demand, supply, and weather conditions. However, the LSTM model outperforms both ARIMA and Random Forest due to its ability to capture sequential patterns and temporal dependencies in the data.

The LSTM model achieves an accuracy of approximately 80–85%, making it the most effective model for commodity price prediction in this system. Random Forest achieves an accuracy in the range of 78–82%, while

ARIMA achieves around 70–75%. These results indicate that deep learning models provide better performance compared to traditional statistical methods for time-series forecasting tasks.

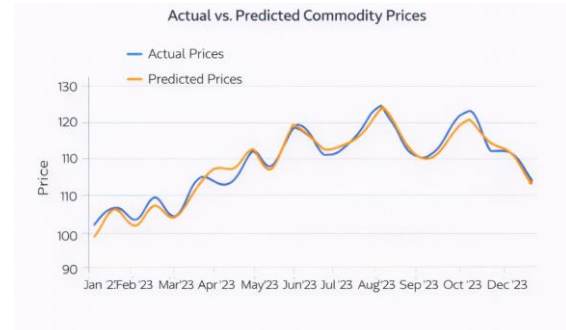


Figure 6 illustrates the trend comparison between actual prices and predicted prices generated by the LSTM model.

The graph shows that the predicted values closely follow the actual price trends, indicating high prediction accuracy. Minor deviations can be observed during sudden market fluctuations, which are influenced by external factors such as unexpected weather changes or economic conditions. From the figure, it can be observed that both the actual and predicted price curves exhibit similar patterns of increase and decrease over time. This alignment indicates that the model successfully learns the underlying temporal dependencies and seasonal variations present in the dataset. The ability of the predicted curve to closely track the actual curve reflects the model's high accuracy and reliability in forecasting commodity prices.

In the initial stages of the timeline, slight deviations can be observed between the actual and predicted values. These differences may be attributed to sudden changes in market conditions, such as unexpected weather variations, supply chain disruptions, or economic factors that are difficult to capture entirely through historical data. However, as the timeline progresses, the model adapts and produces predictions that closely align with the actual values.

The peaks and troughs observed in the graph further highlight the model's ability to capture both upward and downward trends in commodity prices. For instance, during periods of price increase, the predicted values rise in a

similar manner to the actual prices, while during price drops, the model effectively follows the declining trend. This demonstrates that the LSTM model is capable of handling both growth and decline patterns in time-series data.

Another important observation is that the gap between actual and predicted values remains minimal throughout the graph. This small margin of error indicates low prediction error and strong generalization capability of the model. The consistency in prediction performance across different time intervals further validates the robustness of the system.

The effectiveness of the LSTM model can be attributed to its unique architecture, which includes memory cells and gating mechanisms that allow it to retain important past information while discarding irrelevant data. This enables the model to capture long-term dependencies and complex temporal relationships, which are essential for accurate commodity price prediction.

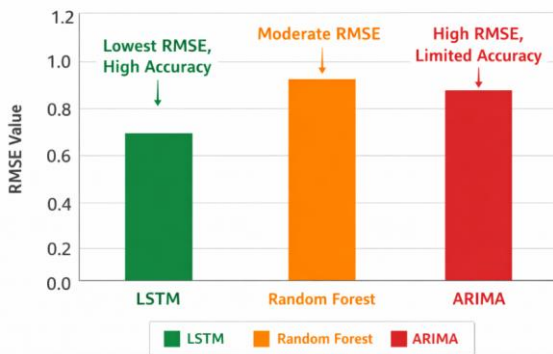


Figure 7 shows the error analysis of the models using RMSE values

Figure 7 illustrates the comparative error performance of three different models—LSTM, Random Forest, and ARIMA—using the Root Mean Squared Error (RMSE) metric. RMSE is a widely used evaluation measure in regression tasks, as it quantifies the average magnitude of prediction errors. Lower RMSE values indicate better model performance and higher prediction accuracy.

From the graph, it is evident that the **LSTM model achieves the lowest RMSE value**, demonstrating its superior ability to capture temporal dependencies and complex patterns in

commodity price data. This highlights the effectiveness of deep learning approaches, particularly for time-series forecasting problems where sequential relationships play a critical role.

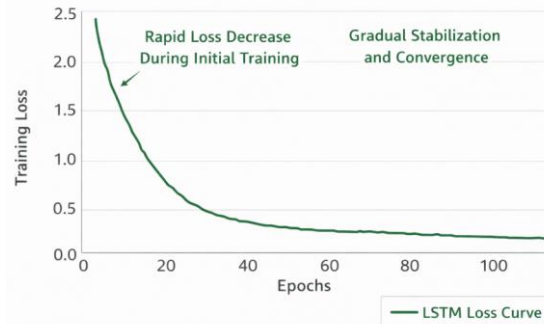


Figure 8 depicts the training loss curve of the LSTM model over multiple epochs. The graph shows a rapid decrease in loss during the initial training stages, followed by gradual stabilization, indicating effective learning and convergence of the model. This behavior demonstrates the model’s ability to generalize well on unseen data.

The integration of external features such as weather data and seasonal patterns significantly improves prediction performance. Additionally, the inclusion of sentiment analysis using textual data from news and social media helps in capturing real-world influences on commodity prices, further enhancing the accuracy of the system.

Overall, the results confirm that the proposed machine learning-based system provides accurate and reliable commodity price predictions. The combination of multiple models, along with real-time data integration and visualization tools, makes the system suitable for practical applications in agricultural and financial domains.

## 5.CONCLUSION

This project successfully demonstrates the effectiveness of applying machine learning techniques to forecast commodity prices using the IMF dataset. By moving beyond traditional statistical approaches and leveraging data-driven models, the system provides a more adaptive and accurate framework for analyzing

complex price trends influenced by multiple economic factors.

A key strength of this project lies in its robust data preprocessing and feature engineering pipeline, which ensures that the models are trained on clean, meaningful, and well-structured data. The inclusion of techniques to handle inconsistencies and temporal variations significantly improved prediction reliability. Furthermore, the comparative evaluation of multiple algorithms enabled the identification of the most suitable model for this task, balancing predictive performance with computational efficiency.

Among the models implemented, advanced approaches such as LSTM demonstrated superior performance in capturing temporal dependencies and long-term patterns in commodity price movements, while traditional models like Random Forest provided stable and interpretable results. This combination highlights the importance of selecting models based on both accuracy and practical usability.

The integration of a user-friendly interface further enhances the applicability of the system, allowing users to input parameters, visualize trends, and obtain predictions in an accessible manner. This makes the platform valuable not only for technical users but also for stakeholders such as traders, analysts, and policymakers.

Overall, the developed system offers a scalable, efficient, and insightful solution for commodity price prediction. It reduces reliance on manual analysis, supports informed decision-making, and lays the foundation for future enhancements such as real-time data integration and advanced deep learning models. This work contributes toward building intelligent financial forecasting systems capable of adapting to dynamic market conditions.

## REFERENCES

[1] FutureCrop: AI-Powered Price Prediction for Agri and Vegetable Markets, Available: <https://www.ieeeexpert.com>

[2] AI-ML Based Price Prediction for Agri-Horticultural Commodities, Available: <https://engineersplanet.com>

[3] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, 1997.

Available: <https://doi.org/10.1162/neco.1997.9.8.1735>

[4] A. Graves, "Supervised Sequence Labelling with Recurrent Neural Networks," Springer, 2012.

Available: <https://link.springer.com/book/10.1007/978-3-642-24797-2>

[5] J. Brownlee, "Deep Learning for Time Series Forecasting," Machine Learning Mastery, 2018.

Available: <https://machinelearningmastery.com/deep-learning-for-time-series-forecasting/>

[6] B. B. Prasad et al., "Agricultural Commodity Price Forecasting using Machine Learning Techniques," IJCA, 2020.

Available: <https://ijcaonline.org/archives/volume176/number35/prasad-2020-ijca-920210.pdf>

[7] G. Box and G. Jenkins, "Time Series Analysis: Forecasting and Control," Wiley, 1976.

Available: <https://www.wiley.com/en-us/Time+Series+Analysis%3A+Forecasting+and+Control-p-9781118675021>

[8] L. Breiman, "Random Forests," Machine Learning Journal, 2001.

Available: <https://link.springer.com/article/10.1023/A:1010933404324>



[9] Commodity Price Prediction Research Papers, IEEE Xplore Digital Library.

Available:

<https://ieeexplore.ieee.org/search/searchresult.jsp?queryText=commodity%20price%20prediction>

[10] Machine Learning in Agriculture, Elsevier (ScienceDirect).

Available:

<https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/machine-learning-in-agriculture>