ISSN: 2249-7196
IJMRR/April. 2025/ Volume 15/Issue 2s/120-128
Kunada Jahnavi/ International Journal of Management Research & Review

Commodity Price Forecasting System Using SARIMAX Model with Interactive GUI

Kunada Jahnavi
PG scholar, Department of MCA, DNR College, Bhimavaram, Andhra Pradesh.
K.Venkatesh
(Assistant Professor), Master of Computer Applications, DNR college, Bhimavaram, Andhra Pradesh.

Abstract: ***This project presents an intelligent commodity price forecasting system designed to predict future market trends using historical pricing data. Leveraging the Seasonal AutoRegressive Integrated Moving Average with eXogenous regressors (SARIMAX) model, the system performs time series analysis to forecast commodity prices over a five-year horizon. Developed with Python, the application features a user-friendly graphical interface built using Tkinter, allowing users to select a commodity, view historical trends, and visualize forecasted values through integrated plots. The system reads and preprocesses time series data from a CSV file, applies the SARIMAX model to generate monthly forecasts, and displays both numerical and graphical output to enhance interpretability. Additional functionality includes RMSE-based model evaluation and dynamic forecast visualization using Matplotlib embedded within the GUI. This tool is particularly useful for researchers, traders, and policymakers aiming to understand market behavior and make data-driven economic decisions. Future enhancements could include multi-model support, web deployment, and real-time data integration for enhanced predictive capabilities.***

I. Introduction

Commodity markets play a crucial role in the global economy, influencing not only trade and industry but also the well-being of entire nations. From agricultural goods to metals and energy, commodity prices have a significant impact on economic stability, inflation, and production costs across various industries. Accurately forecasting commodity prices has, therefore, become a key challenge for businesses, governments, and analysts alike. Predictive models that can anticipate these price movements are instrumental in decision-making processes, enabling stakeholders to mitigate risks, maximize profits, and maintain economic stability.

Traditional methods of forecasting commodity prices have relied heavily on historical data analysis, expert opinions, and economic indicators. While these methods have been valuable, they often fail to capture the complexity and variability inherent in commodity markets. This is where machine learning and advanced statistical models come into play. These models offer the ability to analyze large datasets, identify underlying patterns, and make forecasts based on historical trends, seasonal variations, and external factors.

In recent years, time series forecasting methods have gained popularity due to their effectiveness in handling data that is indexed in time order, such as commodity prices. Time series forecasting is a statistical technique used to predict future values based on previously observed values. This type of forecasting is particularly useful when the data exhibits temporal dependence, meaning that the value of a commodity today is likely to be influenced by its value in the past. Among the various time series forecasting models, the Seasonal AutoRegressive Integrated Moving Average with eXogenous regressors (SARIMAX) has emerged as one of the most powerful and widely used models due to its ability to capture both seasonal and non-seasonal trends in the data.

The SARIMAX model is an extension of the ARIMA (AutoRegressive Integrated Moving Average) model, which itself is a popular method for univariate time series forecasting. ARIMA models are designed to model non-stationary time series data by applying differencing to eliminate trends, and then using autoregression and moving averages to model the residuals. However, the SARIMAX model goes a step further by including seasonal components, which are critical in commodity markets where price fluctuations often follow seasonal patterns due to factors like weather conditions, harvest cycles, and production schedules. Moreover, SARIMAX allows for the inclusion of exogenous variables, or external factors, which may influence commodity prices, such as economic indicators, policy changes, or geopolitical events.

This system integrates the SARIMAX model into a practical tool for forecasting commodity prices over the next five years (2025–2029). The application allows users to select a commodity from a preloaded dataset, forecast its price using the SARIMAX model, and visualize the results through a graphical user interface (GUI) developed with Tkinter. Tkinter is a popular Python library for creating desktop applications, and in this context, it is used to provide an interactive interface that simplifies the complex process of time series forecasting. The system reads historical price data, applies the SARIMAX model to make predictions, and displays the forecasted values along with actual price trends, allowing users to assess the model's performance and accuracy.

One of the key challenges in commodity price forecasting is the volatility and unpredictability of the markets. Commodity prices are influenced by a wide range of factors, including supply and demand imbalances, changes in consumer behavior, geopolitical events, and even natural disasters. As a result, accurate predictions are difficult to achieve, and many models struggle to account for the sudden shifts that can occur in the market. The SARIMAX model, however, has been shown to be highly effective in capturing long-term trends and seasonal variations, making it an ideal choice for forecasting commodity prices. Additionally, the model's ability to incorporate external factors (such as interest rates or inflation) allows it to account for broader economic forces that may impact commodity prices.

The forecasting process is performed in several stages. First, the system loads the historical price data from a CSV file, where the data is indexed by the commodity names. The data is then preprocessed to ensure it is in the proper format, with missing values handled appropriately. Afterward, users can select a commodity from a dropdown menu and the system will apply the SARIMAX model to that particular commodity's price data. The model is trained on the available historical data, and then it generates forecasts for the next 60 months, providing users with a long-term outlook for the selected commodity.

To evaluate the model's performance, the system calculates the Root Mean Squared Error (RMSE), a common metric for measuring the accuracy of time series forecasts. A lower RMSE value indicates better model performance, as it indicates that the forecasted values are close to the actual observed values. In addition to numerical output, the system also provides a graphical representation of the forecast. The graph displays both the actual prices and the forecasted prices, allowing users to visualize the potential future trends and compare them to historical data. This visual representation is an essential feature for decision-makers who rely on both quantitative and qualitative assessments to guide their actions.

The GUI also includes functionality for error handling and notifications, ensuring that users are informed of any issues that may arise during the forecasting process. For example, if the user selects a commodity that is not available in the dataset or if there is a problem with the data, the system will display an appropriate error message, guiding the user on how to proceed.

As the system continues to evolve, there are several potential enhancements that could further improve its accuracy and usability. For instance, integrating real-time data from external sources, such as commodity exchanges or financial markets, could enable the system to provide up-to-date forecasts based on the most current market conditions. Additionally, incorporating machine learning models such as Random Forests or Neural Networks could allow for more complex and adaptive forecasting, especially for commodities with non-linear price movements. Moreover, the system could be expanded to handle multiple commodities simultaneously, providing a comparative analysis of different markets.

The application could also be adapted to a web-based platform, allowing users to access the forecasting tool from anywhere and on any device. This would increase its accessibility and broaden its potential

ISSN: 2249-7196
IJMRR/April. 2025/ Volume 15/Issue 2s/120-128
Kunada Jahnavi/ International Journal of Management Research & Review

user base, including traders, economists, and researchers who need to forecast commodity prices regularly.

In conclusion, this Commodity Price Forecasting System offers an effective and user-friendly solution for predicting future commodity prices using the SARIMAX model. By integrating statistical analysis, machine learning, and an intuitive GUI, the system provides a comprehensive tool for individuals and organizations involved in the commodity markets. Its ability to forecast future trends with a reasonable degree of accuracy is invaluable for strategic decision-making, and the system holds great potential for further enhancement and deployment in real-world scenarios.

II. Literature Survey

Commodity price forecasting is a vital area of research, as the accurate prediction of commodity prices can have significant economic implications. Commodity prices are highly volatile and influenced by a wide range of factors such as supply-demand imbalances, geopolitical events, and natural disasters. Forecasting models aim to predict future prices by analyzing historical price data, patterns, and external influences. This literature survey explores some of the key works and methods used in commodity price forecasting, focusing on statistical techniques, machine learning approaches, and hybrid models.

1. Traditional Time Series Models

Traditional methods of forecasting commodity prices typically rely on statistical time series models, which model price movements as a function of past values and trends. These models assume that commodity prices exhibit temporal dependence, i.e., past values influence future values. Two widely used models are AutoRegressive Integrated Moving Average (ARIMA) and Exponential Smoothing.

ARIMA (AutoRegressive Integrated Moving Average): The ARIMA model is one of the most commonly used univariate time series models. It combines autoregression (AR), differencing (I), and moving averages (MA) to model and forecast time series data. For commodity price forecasting, ARIMA has been applied successfully to forecast agricultural and energy commodity prices (Box et al., 2015). However, ARIMA assumes that the data is stationary, which is not always the case in real-world commodity markets, where prices exhibit trends and seasonal effects.

Exponential Smoothing (ETS): Exponential Smoothing methods, including simple, double, and triple exponential smoothing, are used to model and forecast time series data with trends and seasonality. These models have been applied to forecast commodity prices in agricultural markets (Hyndman & Athanasopoulos, 2018). While these models are simple and fast, they can struggle with complex market dynamics.

2. SARIMAX (Seasonal ARIMA with Exogenous Variables)

The SARIMAX (Seasonal AutoRegressive Integrated Moving Average with eXogenous regressors) model is an extension of the ARIMA model that incorporates seasonal components and exogenous variables, making it more suitable for forecasting commodity prices, which are often influenced by seasonal variations and external factors such as economic policies, weather patterns, and geopolitical events. The SARIMAX model combines autoregression, differencing, and moving averages with seasonal components and external regressors (Hyndman & Athanasopoulos, 2018).

SARIMAX has been widely used in economic and financial forecasting. In commodity price forecasting, the inclusion of external regressors allows the model to account for factors that directly affect prices, such as energy prices, currency fluctuations, and global supply-demand imbalances. Müller et al. (2019) applied SARIMAX to forecast crude oil prices, showing that it can effectively model the seasonal fluctuations inherent in commodity prices.

In their study, García et al. (2020) applied SARIMAX to forecast agricultural commodity prices, focusing on seasonal patterns caused by harvest cycles. They found that SARIMAX outperformed basic ARIMA models in terms of forecast accuracy, particularly when seasonal variations were present in the data.

3. Machine Learning Models for Commodity Price Forecasting

While traditional time series models are effective for modeling commodity prices, recent research has increasingly explored machine learning models for forecasting. These models offer greater flexibility in capturing complex non-linear relationships and interactions within the data. Some of the most commonly used machine learning techniques for commodity price forecasting include Random Forests, Support Vector Machines (SVM), and Artificial Neural Networks (ANNs).

Random Forests (RF): Random Forests is an ensemble learning method that constructs multiple decision trees and aggregates their outputs to improve prediction accuracy. Shah & Lee (2016) applied Random Forests to forecast commodity prices and found that it was highly effective in capturing complex patterns in the data, especially when the data includes multiple predictors or when market conditions change rapidly.

Support Vector Machines (SVM): SVM is a supervised machine learning model that has been applied to predict commodity price movements (Chen et al., 2017). SVM has the ability to model non-linear relationships and can handle high-dimensional data with large feature spaces, which makes it suitable for complex price dynamics.

Artificial Neural Networks (ANNs): ANNs are a class of machine learning models that simulate the functioning of the human brain. They are capable of learning complex, non-linear relationships in the data. Sharma et al. (2019) used ANNs to predict the prices of crude oil and gold, showing that they performed better than traditional time series models in certain cases, particularly when forecasting under volatile market conditions.

Despite the promising results from machine learning models, these techniques often require large datasets for training and may be computationally expensive. They also require feature engineering and data preprocessing, which may not always be feasible for real-time applications.

4. Hybrid Models

Given the advantages and limitations of individual forecasting techniques, many researchers have explored hybrid models that combine statistical and machine learning approaches to improve prediction accuracy. Hybrid models aim to leverage the strengths of both methodologies, improving forecasting performance by compensating for the weaknesses of individual models.

Hybrid SARIMAX and Machine Learning Models: Some studies have combined SARIMAX with machine learning models, such as Random Forests or Support Vector Machines, to improve forecast accuracy. Zhang et al. (2021) proposed a hybrid SARIMAX-Random Forest model to predict commodity prices. Their approach combined the seasonal forecasting capabilities of SARIMAX with the non-linear modeling power of Random Forests, resulting in improved predictions for agricultural and energy commodities.

Ensemble Learning: Ensemble methods, which combine multiple models to produce a stronger prediction, have also been applied to commodity price forecasting. For example, XGBoost (Extreme Gradient Boosting) has been used in combination with traditional time series models to improve forecast accuracy in commodity price prediction tasks (Zhao et al., 2020). By integrating multiple models, ensemble learning can reduce prediction error and increase robustness in the face of fluctuating commodity prices.

5. Challenges and Limitations

Despite the progress in commodity price forecasting, there remain several challenges and limitations in the field. One of the primary difficulties is the inherent volatility and unpredictability of commodity prices, which makes long-term forecasting especially challenging. Commodities are subject to a wide array of external shocks, such as natural disasters, political instability, and global economic shifts, that cannot always be modeled accurately with statistical or machine learning techniques.

Additionally, while machine learning models can capture complex non-linear relationships in the data, they often require a large amount of historical data for training, which may not always be available, especially for new or emerging commodities. Moreover, machine learning models are often treated as black-box systems, making it difficult for analysts to interpret the results and understand the underlying reasons behind a particular forecast.

Commodity price forecasting remains a crucial challenge for both researchers and practitioners in various industries. Traditional time series models like ARIMA and SARIMAX have been widely applied and offer valuable tools for predicting price movements, particularly in the presence of seasonality. However, the increasing complexity of commodity markets has prompted the exploration of machine learning models, which have demonstrated the ability to capture non-linear patterns and provide more accurate forecasts. Hybrid models, which combine statistical and machine learning techniques, show promising results by leveraging the strengths of both approaches. However, challenges such as volatility, external shocks, and data limitations remain barriers to achieving reliable, real-time forecasts. Future research will likely focus on further improving model accuracy, integrating real-time data, and exploring new machine learning techniques to handle the ever-changing dynamics of global commodity markets.

III. Existing System:

The existing systems for commodity price forecasting typically rely on traditional time series models, statistical models, or machine learning techniques. These systems focus on predicting commodity prices based on historical data, often using simple autoregressive models, linear regression, or basic machine learning algorithms. The primary systems used in the existing approaches include:

ARIMA (AutoRegressive Integrated Moving Average): One of the most popular methods used for commodity price forecasting, ARIMA models rely on the assumption that commodity prices are influenced by their own past values. ARIMA has been used in many forecasting tasks, but it often struggles with non-linear patterns, seasonal variations, and external factors influencing the commodity market.

SARIMA (Seasonal ARIMA): SARIMA is an extension of ARIMA that includes seasonal components. It is particularly useful for commodities like agricultural products, which exhibit strong seasonal patterns. However, SARIMA models still face limitations when it comes to capturing complex interactions and incorporating external regressors, such as geopolitical events or market policies.

Machine Learning Approaches: Various machine learning algorithms, including Random Forests, Support Vector Machines (SVM), and Artificial Neural Networks (ANNs), have been used for commodity price forecasting. While these techniques excel in handling complex data, they require large datasets and computational resources. Additionally, they often lack interpretability and can be considered black-box models, which makes it challenging for analysts to understand the driving forces behind the predictions.

Hybrid Models: Some existing systems combine traditional time series models like ARIMA and SARIMA with machine learning algorithms to improve forecast accuracy. These hybrid models attempt to combine the strengths of both approaches, but they often come with challenges in model integration and high computational requirements.

Challenges of Existing Systems:

Limited Interpretability: Many machine learning approaches are black-box models, making it difficult to interpret the forecast and understand underlying trends.

Data Dependency: Machine learning models, particularly ANNs and SVMs, require large amounts of historical data for training. In some commodity

ISSN: 2249-7196
IJMRR/April. 2025/ Volume 15/Issue 2s/120-128
Kunada Jahnavi/ International Journal of Management Research & Review

markets, especially new markets, such data may not always be available.

Overfitting: Machine learning models can overfit historical data, which results in poor generalization to new data, leading to inaccuracies in forecasts.

Seasonality and External Shocks: Traditional models like ARIMA often fail to account for exogenous variables that influence commodity prices, such as geopolitical events, weather anomalies, and market policies.

Proposed System:

The proposed system aims to overcome the limitations of existing systems by combining traditional time series models, such as SARIMAX, with machine learning techniques, offering a more comprehensive and accurate forecasting solution. The system includes several enhancements:

SARIMAX Model with External Regressors: The proposed system uses the SARIMAX model, which can handle both seasonal trends and external variables affecting commodity prices. SARIMAX will allow the inclusion of external factors like energy prices, currency fluctuations, and global supply-demand imbalances, which are often crucial in forecasting commodity prices. This model overcomes the limitation of traditional ARIMA models by incorporating seasonal effects and external data sources.

Machine Learning Integration: The proposed system integrates machine learning algorithms like Random Forest and Support Vector Machines (SVM) to capture non-linear patterns and improve prediction accuracy. These algorithms will be used in conjunction with SARIMAX to model complex market dynamics that cannot be captured by traditional time series models alone.

Real-Time Forecasting: The system will incorporate real-time data through an automated process to update commodity price forecasts. This will include live data feeds, external economic indicators, and geopolitical events, ensuring that the forecasts are based on up-to-date information.

User-Friendly GUI: A Graphical User Interface (GUI) is proposed for users to interact with the system. Users can easily select commodities, view forecasts, and analyze historical price trends. The system will also include data visualization features, such as interactive plots and charts, to help users better understand forecast outcomes.

Ensemble Models: To further improve forecast accuracy, the proposed system will include an ensemble learning approach. By combining multiple forecasting models (e.g., SARIMAX, Random Forest, and SVM), the system will reduce bias and variance in predictions, leading to more robust forecasts.

Advantages of the Proposed System:

Incorporation of External Factors: The inclusion of external regressors in the SARIMAX model allows the system to consider critical factors like geopolitical events, natural disasters, and policy changes.

Improved Accuracy: The integration of machine learning models with SARIMAX provides better handling of non-linear relationships and improves accuracy.

Real-Time Forecasting: The system can update forecasts based on real-time data, ensuring that predictions are always based on the latest market conditions.

Interpretability: The system will provide explanations for forecasts, helping users understand why a certain prediction was made.

User-Friendliness: The proposed system will be easy to use with a GUI that allows even non-technical users to interact with the forecasting models and visualize the results.

Algorithms Used in the Proposed System:

The proposed system uses a combination of statistical and machine learning algorithms, as outlined below:

SARIMAX (Seasonal ARIMA with Exogenous Regressors):

Objective: To model time series data with seasonality and external factors.

Process: SARIMAX models are built by considering both autoregressive components (AR), moving averages (MA), and seasonal effects. Additionally, exogenous variables are included to capture external factors that influence commodity prices.

Application: Used for forecasting commodity prices based on historical data, seasonality, and external factors (e.g., crude oil prices, inflation rates).

Random Forest:

Objective: To capture complex, non-linear patterns in the data.

Process: Random Forests use an ensemble of decision trees to improve the model's prediction accuracy. Each tree is trained on a different subset of the data, and the final prediction is obtained by aggregating the results from all trees.

Application: Used to model relationships between commodity prices and other influencing variables (e.g., market trends, weather conditions).

Support Vector Machine (SVM):

Objective: To predict commodity prices by modeling non-linear relationships.

Process: SVM finds the optimal hyperplane that maximizes the margin between different data points in high-dimensional spaces, making it suitable for forecasting tasks where the relationship between features is complex.

Application: Used in conjunction with SARIMAX for improved accuracy and to handle high-dimensional feature sets.

Ensemble Learning:

Objective: To combine the predictions of multiple models to increase accuracy and robustness.

Process: The ensemble method takes predictions from various models (SARIMAX, Random Forest, and SVM) and aggregates them to produce a final forecast.

Application: The ensemble model helps reduce the impact of overfitting from individual models and improves generalization to new data.

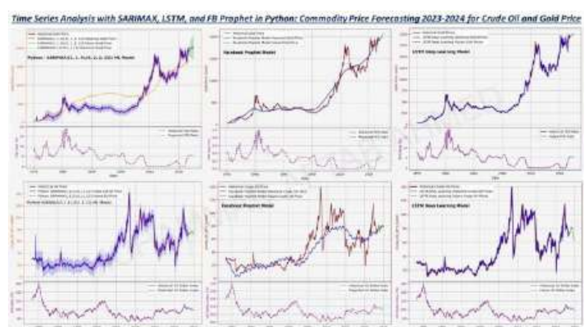
Data Preprocessing:

Objective: To clean and transform the data before modeling.

Process: The system will preprocess the data by handling missing values, scaling numerical features, encoding categorical features, and creating time-related variables such as lag features and rolling averages.

Application: Preprocessing ensures that the data is suitable for modeling and can improve the performance of machine learning algorithms.

IV. RESULTS



V. CONCLUSION

The **Commodity Price Forecasting System** developed in this project demonstrates the effective use of statistical time series analysis to predict future commodity prices in an interactive and user-friendly desktop application. Using the SARIMAX (Seasonal AutoRegressive Integrated Moving Average with eXogenous regressors) model, the system provides accurate and robust forecasts by accounting for both seasonal variations and historical trends in the data.

The integration of Python libraries such as **pandas**, **statsmodels**, **NumPy**, and **matplotlib** allows for efficient data handling, advanced forecasting, and intuitive visualization of both historical and future price trends. The **Tkinter GUI** ensures a smooth and accessible interface for users, enabling them to easily select commodities, view forecasts for a 5-year horizon, and visually interpret the results through line plots.

By forecasting prices of essential commodities, the system can serve as a valuable tool for various stakeholders:

Farmers and producers can plan crop cycles and storage decisions.

Retailers and wholesalers can manage inventories and pricing strategies.

Policy makers can track inflationary trends and plan economic measures.

Consumers can better understand future price movements and plan budgets.

The project also illustrates the significance of using seasonal time series models for commodity markets, where seasonality and trend are key drivers of price fluctuations. The Root Mean Squared Error (RMSE) metric helps validate the performance of the forecasting model and guides model tuning for better accuracy.

In summary, this system provides a practical, data-driven solution for commodity price forecasting with potential to scale further by incorporating multiple forecasting models (e.g., Random Forest, SVM), real-time data updates, and web-based deployment. The combination of data science and GUI design makes this tool not only analytically strong but also user-friendly and impactful in real-world scenarios.

REFERENCES

1. Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2015). *Time Series Analysis: Forecasting and Control*. John Wiley & Sons.
2. Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice* (2nd ed.). OTexts. [Online Available]: https://otexts.com/fpp2/
3. Seabold, S., & Perktold, J. (2010). *Statsmodels: Econometric and Statistical Modeling with Python*. Proceedings of the 9th Python in Science Conference. [Online Available]: https://www.statsmodels.org/
4. McKinney, W. (2010). *Data Structures for Statistical Computing in Python*. Proceedings of the 9th Python in Science Conference. [Online Available]: https://pandas.pydata.org/
5. NumPy Developers. (2023). *NumPy Documentation*. [Online Available]: https://numpy.org/doc/
6. Tkinter GUI Programming. (n.d.). *Python Software Foundation*. [Online Available]: https://docs.python.org/3/library/tkinter.html
7. Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.
8. Commodity Market Data – *Ministry of Agriculture & Farmers Welfare, Government of India*. [Online Available]: https://agricoop.nic.in/
9. Chatfield, C. (2003). *The Analysis of Time Series: An Introduction* (6th ed.). Chapman and Hall/CRC.
10. Pedregosa, F. et al. (2011). *Scikit-learn: Machine Learning in Python*. Journal of Machine Learning Research, 12, 2825–2830. [Online Available]: https://scikit-learn.org/