International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

Gold Price Prediction System Using Machine Learning and Deep Learning Techniques

Abhisek Purohita
2201298288
Dept. of Computer Science & Engineering
GIFT Autonomous, Bhubaneswar

Subham Sahoo
2201298485
Dept. of Computer Science & Engineering
GIFT Autonomous, Bhubaneswar

Under the Guidance of **Prof. Mohapatra Girashree Sahoo** *— Dept. of CSE, GIFT Autonomous, Bhubaneswar*
Bachelor of Technology — Computer Science & Engineering | BPUT, Odisha | Batch 2022–2026

Abstract— *Gold has long been considered one of the most reliable and valuable investment assets across the world. People invest in gold not only for financial security but also as a safeguard against inflation and economic uncertainty. However, the price of gold is highly dynamic and influenced by a variety of factors such as global economic conditions, inflation rates, currency fluctuations, geopolitical events, and market demand. This project presents a Gold Price Prediction System that forecasts future gold prices based on historical data and current market trends using advanced time-series forecasting techniques. Traditional statistical models like ARIMA (Auto Regressive Integrated Moving Average) are applied to identify linear patterns while modern deep learning models such as LSTM (Long Short-Term Memory) neural networks capture complex non-linear relationships. The system compares model performances using evaluation metrics and presents both historical and predicted prices through data visualizations, providing investors and financial analysts with an efficient, data-driven decision-support platform.*

Keywords— *Gold Price Prediction; ARIMA; LSTM; Machine Learning; Time Series Forecasting; Python; Data Visualization; FinTech.*

I. INTRODUCTION

Gold has always been one of the most valuable and trusted assets in the world. It is widely used as an investment option, a hedge against inflation, and a symbol of financial security. In countries like India, gold also holds deep cultural and economic importance.

The price of gold is highly dynamic and keeps changing due to factors such as global economic conditions, inflation rates, currency exchange rates, interest rates, and geopolitical events. Because of these fluctuations, predicting gold prices has become a crucial task for investors, traders, and financial analysts. Accurate predictions can help individuals and organizations make better investment decisions, reduce risks, and maximize profits.

With the advancement of machine learning and data science, it has become possible to analyze large amounts of historical data and identify hidden patterns and trends. The Gold Price Prediction System is designed to use historical gold price data along with current economic indicators to forecast future price trends. Models such as ARIMA are used for statistical analysis, while LSTM neural networks capture complex patterns in the data.

In addition to prediction, the system provides visual representations of price trends through graphs and charts, making it easier for users to understand and interpret results. The goal of this project is to create a reliable and user-friendly system that assists in making informed financial decisions.

II. LITERATURE REVIEW

A. Traditional Prediction Methods

Existing gold price prediction systems rely mainly on manual analysis, historical comparison, and basic statistical techniques. Investors and financial analysts depend on past price trends, market news, and personal experience to estimate future prices. These methods provide only a general understanding of the market and are not always accurate due to the complex and dynamic nature of gold price movements.

B. Statistical Models — ARIMA

ARIMA (Auto Regressive Integrated Moving Average) is a widely used statistical model for time-series forecasting. It is capable of identifying trends and seasonal patterns in data. However, ARIMA is limited to linear relationships and struggles with sudden fluctuations and non-linear patterns commonly present in financial data [1].

C. Deep Learning Models — LSTM

Long Short-Term Memory (LSTM) is a type of Recurrent Neural Network (RNN) specifically designed for sequential data. LSTM models can remember long-term dependencies, making them highly effective for financial time-series data. Several studies have demonstrated that LSTM outperforms traditional statistical models in capturing complex non-linear patterns [2].

D. Existing Tools and Limitations

Tools like Bloomberg Terminal, Yahoo Finance, and various ML-based platforms provide partial prediction support but are either too expensive, lack integrated pipelines, or do not

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

provide clear visualizations for non-technical users. There is a clear need for an integrated, automated, and user-friendly gold price prediction system that combines statistical and deep learning approaches within a single platform.

E. Research Gap

Current systems focus on either statistical or deep learning models but rarely combine both within a unified framework. Most lack automated data collection, multi-factor input consideration, integrated model evaluation, and rich data visualization. The proposed system addresses all these limitations.

III. PROPOSED SYSTEM

The proposed Gold Price Prediction System is an integrated, data-driven platform designed to automate the complete gold price forecasting lifecycle. Unlike traditional methods, this system combines statistical modeling (ARIMA) with deep learning (LSTM) within a single unified environment, along with automated data collection, preprocessing, model evaluation, and visualization.

A. Data Collection Module

Historical gold price data is gathered from reliable online sources such as Yahoo Finance (via the yfinance API), Kaggle datasets, and financial data portals. The system can also collect related economic indicators such as inflation rates, interest rates, USD exchange rates, crude oil prices, and stock market indices. Automated web scraping using BeautifulSoup and Selenium ensures continuous data updates.

B. Data Preprocessing Module

Raw data is cleaned and transformed by removing missing values, handling duplicates, filtering outliers, and performing normalization using MinMaxScaler. Feature engineering extracts time-based features and selects the most relevant variables to improve model accuracy.

C. ARIMA Model

The ARIMA model performs statistical time-series forecasting by identifying linear trends, seasonality, and autocorrelation patterns. It uses the ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) to select optimal p, d, q parameters. ARIMA serves as the baseline model for comparison.

D. LSTM Model

The LSTM neural network captures complex, non-linear relationships and long-term dependencies in sequential gold price data. The architecture includes Bidirectional LSTM layers, Dropout layers for regularization, and Dense output layers. The model is trained using historical price windows of 60 days to predict the next day's price.

E. Model Evaluation and Comparison

Both models are evaluated using MAE (Mean Absolute Error), RMSE (Root Mean Square Error), MAPE (Mean Absolute Percentage Error), and R^2 score. The best-performing model is selected for final predictions.

TABLE I. COMPARISON OF EXISTING VS PROPOSED SYSTEM

Feature	Traditional Methods	ARIMA	LSTM	Proposed System
Linear Trend Detection	Partial	Yes	Yes	Yes
Non-linear Patterns	No	No	Yes	Yes
Automated Data Collection	No	No	No	Yes
Multi-factor Input	No	No	Partial	Yes
Visualization	No	Partial	Partial	Yes
User-Friendly Interface	No	No	No	Yes

IV. SYSTEM ARCHITECTURE

The architecture of the Gold Price Prediction System follows a 3-tier client-server model designed for scalable time-series forecasting. The system is divided into three primary layers:

A. Presentation Layer (Frontend)

The Presentation Layer provides the user interface through which users interact with the system. It includes web pages or dashboards developed using HTML, CSS, and JavaScript. Users can upload datasets, select prediction models, view forecasts, and analyze results through interactive graphs and charts. Visual representations include line graphs, comparison charts, and trend analysis plots.

B. Application Layer (Business Logic)

The Application Layer handles all processing and logic. It is implemented in Python using libraries such as Pandas, NumPy, TensorFlow, Keras, Statsmodels, and Scikit-learn. This layer manages data preprocessing, model training (ARIMA and LSTM), prediction generation, and performance evaluation. The primary IDE used is Jupyter Notebook and VS Code.

C. Data Layer

The Data Layer uses MySQL for structured storage of historical gold prices, economic indicators, model metadata, prediction results, and user information. Proper normalization (1NF, 2NF, 3NF) is applied to reduce redundancy and maintain data integrity. CRUD operations manage data through the complete system lifecycle.

V. METHODOLOGY

The methodology follows a structured SDLC-based approach divided into five major stages:

1. Data Collection & Preprocessing

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

Historical gold price data and economic indicators are collected via the yfinance API and Kaggle. The data is then cleaned, normalized using MinMaxScaler, and formatted as time-series input. Missing values are handled using forward-fill and interpolation techniques.

2. Exploratory Data Analysis

Trend analysis, correlation analysis, and distribution plots are used to understand the dataset. Key relationships between gold prices and external factors (USD exchange rate, inflation, interest rates, crude oil prices) are identified and visualized using heatmaps and line graphs.

3. Model Training

ARIMA model parameters are selected using ACF and PACF plots. The LSTM model uses a sliding window of 60 time steps with a Bidirectional LSTM architecture. Both models are trained on 80% of historical data and validated on the remaining 20%.

4. Evaluation & Selection

Models are evaluated using MAE, RMSE, MAPE, and R^2 metrics. Cross-validation is applied to ensure generalization. The model with the lowest MAPE and highest R^2 is selected for deployment. Regular retraining ensures the model remains accurate as new data becomes available.

5. Visualization & Reporting

Prediction results are presented as 'Actual vs Predicted' line graphs, future forecast trend lines, and downloadable CSV/PDF reports. The visualization module uses Matplotlib and Seaborn to generate all graphical outputs.

VI. IMPLEMENTATION

A. Technology Stack

The system is implemented using Python as the primary language. Pandas and NumPy handle data manipulation. Matplotlib and Seaborn provide visualization. Scikit-learn handles preprocessing and evaluation. TensorFlow and Keras implement the LSTM model. Statsmodels implements ARIMA. MySQL stores all persistent data. The frontend, where applicable, uses HTML/CSS/JavaScript.

B. LSTM Architecture

The LSTM model architecture consists of: (1) Input Layer — 60 time steps × 11 features, (2) Bidirectional LSTM layer — 256 units, (3) Dropout layer — 0.2, (4) LSTM layer — 64 units, (5) Dropout layer — 0.2, (6) Dense layer — 32 units, (7) Dense output layer — 1 unit. Total trainable parameters: 229,164 (approximately 891.25 KB).

C. Project Scheduling

The project follows a PERT and Gantt-based schedule. The critical path is: Data Collection (5 days) → Data Preprocessing (4 days) → Feature Selection (3 days) → Model Selection (2 days) → Model Training (5 days) → Model Evaluation (3 days) → Prediction & Visualization (4 days). Total estimated duration: 26 days.

VII. RESULTS AND EVALUATION

The Gold Price Prediction System was tested using historical gold price data spanning from 2010 to 2024 (4,041 data points). Both ARIMA and LSTM models were trained and evaluated on the same dataset split (80% train / 20% test).

TABLE II. MODEL PERFORMANCE COMPARISON

Metric	ARIMA	LSTM
MAE	₹43,166.45	₹7,955.58
RMSE	₹47,168.85	₹12,717.53
R^2	0.9003	0.7998
MAPE	17.02%	7.74%

The LSTM model significantly outperformed ARIMA in terms of MAPE (7.74% vs 17.02%), indicating it captures non-linear patterns in gold price data more effectively. The R^2 score of 0.7998 for LSTM confirms acceptable predictive accuracy. ARIMA, while achieving a higher R^2 (0.9003), produced larger absolute errors due to its inability to model non-linear relationships.

Key system metrics observed during testing:

- Latest gold price tracked: ₹1,42,170.28 per 10g
- All-Time High tracked: ₹1,65,305.62
- All-Time Low tracked: ₹15,844.42
- Total data points processed: 4,041
- Forecast generation time: < 2 seconds per prediction

The visualization module successfully generated historical trend charts, actual vs predicted comparison graphs, and future forecast displays, providing users with clear and actionable insights.

VIII. FUTURE SCOPE

The future scope of the Gold Price Prediction System is very wide and can be improved with advanced technologies and extended functionalities:

- Advanced AI Models: Integrate Transformer models, GRU, and hybrid deep learning models for better accuracy.
- Real-Time Data Integration: Add live gold price updates using APIs for dynamic predictions.
- Sentiment Analysis (NLP): Analyze news articles and social media to understand market sentiment.
- Mobile Application: Convert the system into a mobile app for anytime access.

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

- Cloud Deployment: Host on AWS, Azure, or Google Cloud for improved scalability.
- Automated Trading Integration: Connect with trading platforms for auto buy/sell decisions.
- Blockchain Integration: Use blockchain for secure, tamper-proof financial data handling.
- Commodity Expansion: Extend predictions to silver, crude oil, and cryptocurrencies.

IX. CONCLUSION

This paper presented the Gold Price Prediction System, an intelligent and data-driven platform that automates the complete gold price forecasting lifecycle. The system successfully integrates ARIMA and LSTM models within a unified environment, achieving a MAPE of 7.74% with the LSTM model, demonstrating strong prediction capability for time-series financial data.

The platform addresses major limitations of existing systems by combining automated data collection, multi-factor preprocessing, dual-model prediction, comparative evaluation, and interactive visualization within a single framework. Experimental results confirm that deep learning-based LSTM models significantly outperform traditional ARIMA models for non-linear gold price patterns.

Overall, the Gold Price Prediction System provides an efficient, scalable, and user-friendly solution for modern financial forecasting, demonstrating the growing potential of machine learning and deep learning techniques in the FinTech domain.

REFERENCES

[1] G. Box, G. Jenkins, G. Reinsel, and G. Ljung, Time Series Analysis: Forecasting and Control, 5th ed. Wiley, 2015.

[2] F. Chollet, Deep Learning with Python. Manning Publications, 2018.

[3] R. J. Hyndman and G. Athanasopoulos, Forecasting: Principles and Practice, 3rd ed. OTexts, 2021.

[4] G. James, D. Witten, T. Hastie, and R. Tibshirani, An Introduction to Statistical Learning. Springer, 2013.

[5] S. Hochreiter and J. Schmidhuber, 'Long Short-Term Memory,' Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[6] TensorFlow Documentation, https://www.tensorflow.org/ (accessed 2024).

[7] Pandas Documentation, https://pandas.pydata.org/ (accessed 2024).

[8] yfinance Library — Gold Price Data Collection, https://pypi.org/project/yfinance/ (accessed 2024).

[9] Kaggle Gold Price Dataset, https://www.kaggle.com/ (accessed 2024).

[10] Scikit-learn Documentation, https://scikit-learn.org/ (accessed 2024).