K.L.N. COLLEGE OF ENGINEERING

(An Autonomous Institution, Affiliated to Anna university, Chennai)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

PREDICTIVE MODELING AND FORECASTING OF STOCK PRICES USING MACHINE LEARNING

DOMAIN NAME :

MACHINE LEARNING

GUIDE :

Ms. R.Nivethitha

Assistant Professor 2

Department of Computer Science and Engineering

TEAM MEMBERS :

MADHU PRIYAA S [910622104058]

NIVETHA R S [910622104068]

ABSTRACT

- ❖Stock market prices are highly volatile, and traditional models like ARIMA and Moving Averages fail to handle real-time fluctuations.
- ❖Deep learning methods like LSTM can predict sequences but are slow, resource-heavy, and difficult to interpret.
- ❖XGBoost combined with feature engineering (lags, moving averages, returns, volatility) is used to develop a faster, scalable, and more accurate forecasting model.
- ❖The system is plug-and-play, supports multiple stocks, updates with live data, and provides explainable results for real-world financial use.
- ❖The model enhances decision-making by providing investors and analysts with actionable insights in real time.

INTRODUCTION

- Traditional statistical models like ARIMA and Moving Averages often fail to capture the nonlinear and volatile nature of stock price movements.
- The stock market is highly dynamic and influenced by numerous unpredictable factors such as demand, supply, and investor sentiment.
- Machine Learning techniques enable more accurate forecasting by analyzing large datasets and identifying hidden patterns in market trends.
- The proposed system utilizes the XGBoost algorithm along with feature engineering to enhance prediction accuracy and adaptability.
- This approach aims to assist investors and analysts in making data-driven, real-time financial decisions with improved reliability.

LITRATURE SURVEY:

S.No	Paper title	Authors and publications	Methodology	Merits & demerits
1.	A Novel Market Sentiment Analysis Model for Forecasting Stock and Cryptocurrency Returns	IEEE transactions on systems, man, and cybernetics: systems, vol. 54, no. 9, september 2024 Shivam Bansal, Subhashis Banerjee.	Proposed a sentiment analysis–driven model for financial forecasting. Collected data from social media, news articles, and financial reports. Applied Natural Language Processing (NLP) techniques to extract sentiments. Integrated sentiment scores with machine learning models for return prediction. Evaluated performance on both stock and cryptocurrency markets.	❖Merits : First to combine sentiment analysis with price forecasting for both stocks and crypto. Captures market psychology and investor behavior.Outperforms models that rely only on historical numerical data. ❖Demerits : Sentiment data is noisy and unstructured, requiring heavy preprocessing. Strong dependence on data sources (Twitter, news APIs, etc.).

LITRATURE SURVEY:

S.No	Paper title	Authors and publications	Methodology	Merits & demerits
2	An Approach Toward Stock Market Prediction and Portfolio Optimization in Indian Financial Sectors.	IEEE transactions on computational social systems, vol. 12, no. 1, february 2025 Manjula A. B., Sumathi R.	Applied machine learning and deep learning techniques for stock prediction. Used LSTM (Long Short-Term Memory) to handle sequential time-series data. Collected historical stock data from Indian financial markets. Preprocessed the data with normalization and feature extraction. Compared accuracy of ML models against traditional forecasting methods. Integrated results with portfolio optimization to guide investors.	❖Merits: LSTM captures long-term dependencies in stock price sequences. Better accuracy compared to basic statistical models like ARIMA. ❖Demerits: High computation time – deep learning requires large resources.

LITRATURE SURVEY:

S.No	Paper title	Authors and publications	Methodology	Merits & demerits
3	Forecasting Bitcoin Prices Using Deep Learning for Consumer-Centric Industrial Applications.	IEEE transactions on consumer electronics, vol. 70, no. 1, february 2024 S. S. Manivannan, A. Sivasankar.	Applied deep learning models to forecast cryptocurrency prices (Bitcoin). Collected historical Bitcoin price data from public repositories. Used LSTM networks for time-series forecasting. Compared with other models to evaluate prediction accuracy. Focused on Bitcoin as a consumer-centric financial product. Performance validated using error metrics like MAE and RMSE.	❖Merits: Demonstrated the use of deep learning in cryptocurrency forecasting. Bitcoin prices. Showed better accuracy than basic ML/statistical models. ❖Demerits : LSTM requires large datasets and high computing power.

LITRATURE SURVEY:

S.No	Paper title	Authors and publications	Methodology	Merits & demerits
3.	Discovering Predictable Latent Factors for Time Series Forecasting.	IEEE transactions on knowledge and data engineering,vol.36,no.10,october2024 Yong Liu, Yu Zheng, Defu Lian, Lei Chen.	Proposed a latent factor model for time-series forecasting. Aimed at discovering hidden predictable factors from historical time-series data. Integrated matrix factorization with temporal patterns. Compared latent factor forecasting with traditional statistical models. Evaluated on real-world time-series datasets for accuracy and scalability.	❖Merits : Introduced a novel approach for uncovering hidden structures in time-series data. More scalable and efficient compared to standard time-series models. ❖Demerits : Requires large datasets to properly learn latent factors. Model is mathematically complex.

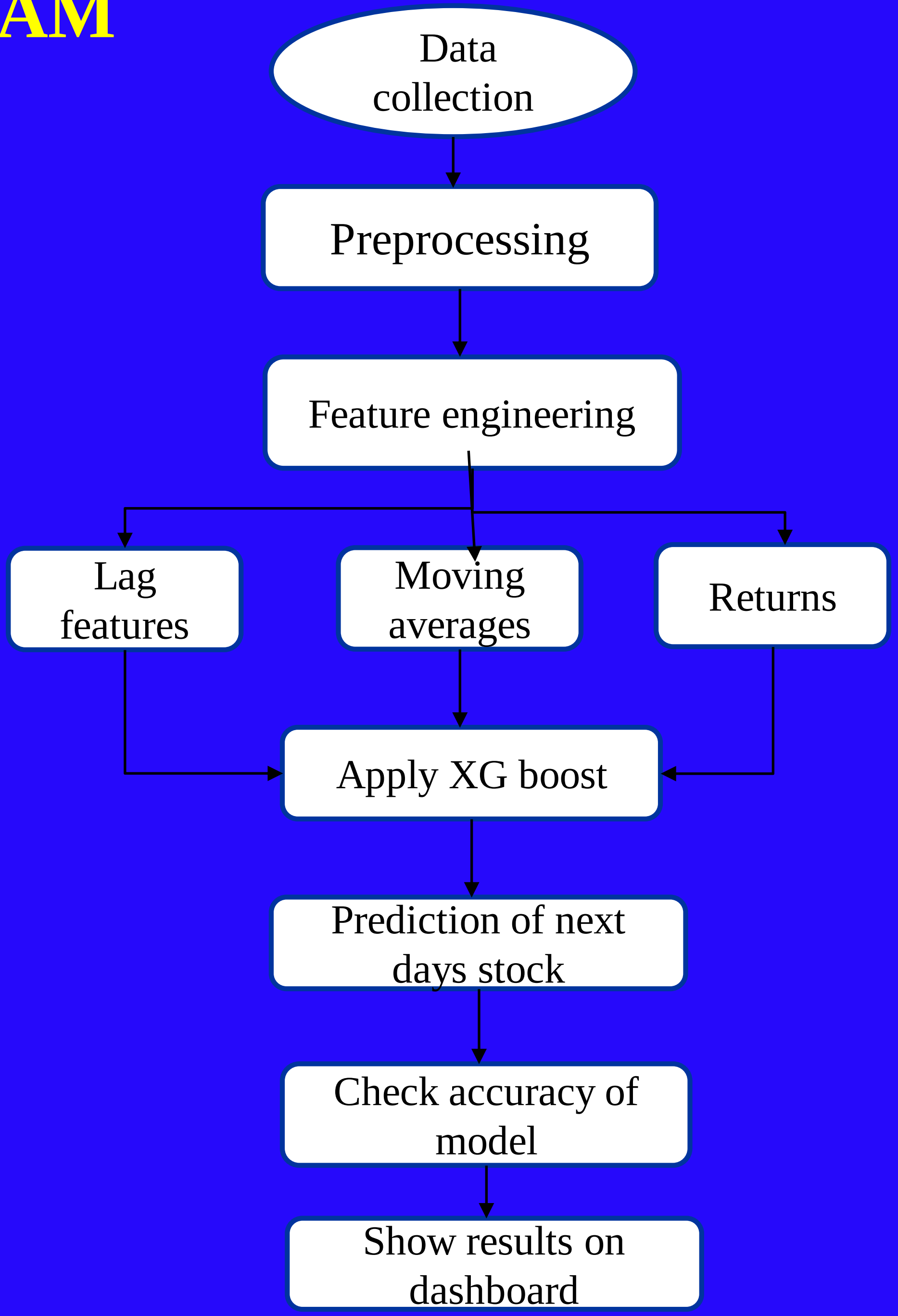
EXISTING SYSTEM

- ❖Existing methods like LSTM/GCN-LSTM/DR2TNet are resource-intensive, slow to train, and unsuitable for real-time applications.
- ❖Deep learning models lack interpretability, making it difficult to understand or explain predictions to investors.
- ❖Prior models rely on predefined or dynamically learned stock relations, but often fail when data is incomplete or noisy.
- ❖Many models emphasize stock-to-stock relationships but ignore multivariate features like volume, moving averages, or volatility.
- ❖Existing systems are not easily adaptable to live data streams and financial dashboards, restricting real-world usability.

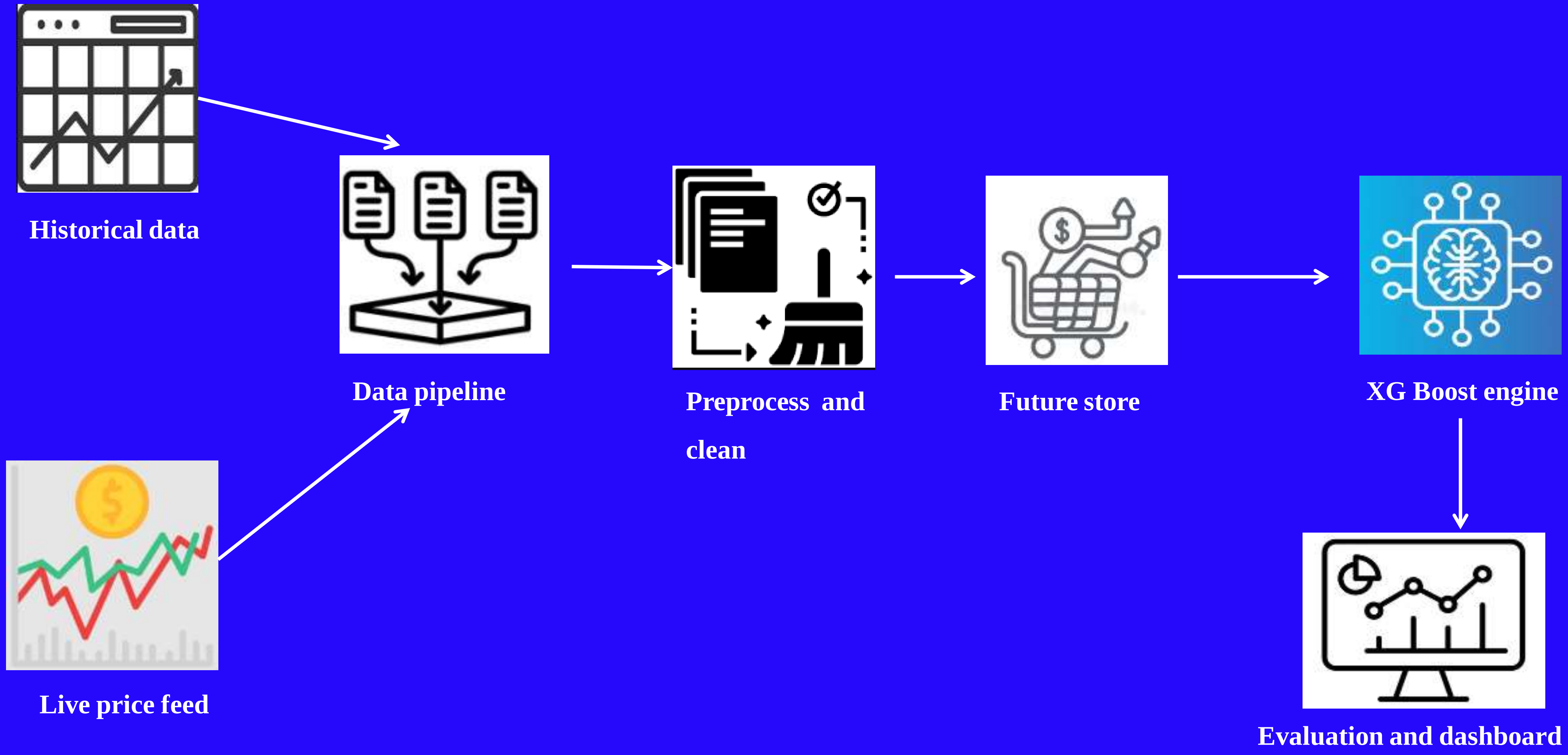
PROPOSED SYSTEM

- The proposed system uses the XGBoost algorithm, which is faster and more accurate than traditional deep learning models for stock forecasting.
- Stock data such as Open, High, Low, Close, and Volume is collected automatically from the yfinance API, ensuring up-to-date inputs.
- Advanced feature engineering is applied by adding lag values, moving averages, returns, and volatility indicators to capture hidden patterns.
- Unlike deep learning, XGBoost offers explainable predictions by showing feature importance, helping investors understand influencing factors.
- The system is flexible and adapts easily to multiple stocks, making it scalable across various companies without major reconfiguration.

DATA FLOW DIAGRAM



ARCHITECTURE DIAGRAM



SYSTEM REQUIREMENTS

Hardware Requirements

- A system with at least Intel i5 processor (or equivalent) to handle training and execution efficiently.
- 8 GB RAM (minimum) is required to manage datasets and model training smoothly.
- 500 GB storage or more to store historical stock datasets, logs, and results.
- A stable internet connection for fetching live stock data using APIs.

Software Requirements

- Operating System: Windows / Linux / MacOS with Python support.
- Programming Language: Python 3.x for implementation.
- Libraries/Packages: XGBoost, Scikit-learn, Pandas, Numpy, Matplotlib, Seaborn.
- Data Source: yfinance API (or equivalent) for collecting live stock market data.
- IDE/Tools: Jupyter Notebook / PyCharm / VS Code for development and testing.

MODULES

- Data Collection
- Data Preprocessing
- Feature Engineering
- Model Training
- Prediction
- Performance Evaluation

REAL TIME CAPTURE AND SYNCHRONIZATION

❖A robust synchronization process keeps the Open, High, Low, Close, and Volume values updated instantly without delays.

❖The system continuously captures live stock market data from trusted APIs such as yfinance to ensure accuracy and timeliness.

❖Real-time pipelines ensure that incoming data is cleaned, validated, and preprocessed before being used by the prediction model.

❖The synchronized data stream allows the XGBoost model to react quickly to sudden price movements or volatility.

❖Continuous synchronization guarantees consistency between stored historical data and new market updates.

❖Enables seamless integration of predictions into dashboards, reports, or advisory platforms for immediate decision-making.

PROCESSING AND CONTEXTUAL UNDERSTANDING

❖The system leverages XGBoost’s advanced processing power to learn complex, non-linear relationships in stock data.

❖Contextual understanding is enhanced by analyzing multiple features simultaneously, including price, volume, volatility, and returns.

❖AI interprets how different market indicators interact and collectively influence stock price movements.

❖The model provides explainable predictions, highlighting which features contribute most to forecasting accuracy.

❖Context-aware learning ensures the system can adapt to different stocks and time frames without reconfiguration.

❖By embedding contextual insights, the system enables data-driven decision-making for investors and analysts.

KNOWLEGDE MANAGEMENT AND STORAGE

❖The system maintains a centralized repository for storing historical stock data and processed features.

❖Knowledge is organized systematically, allowing efficient retrieval of datasets, models, and results when required.

❖A storage framework is implemented to preserve trained XGBoost models along with their evaluation metrics.

❖Continuous storage of past predictions enables trend analysis and long-term performance monitoring.

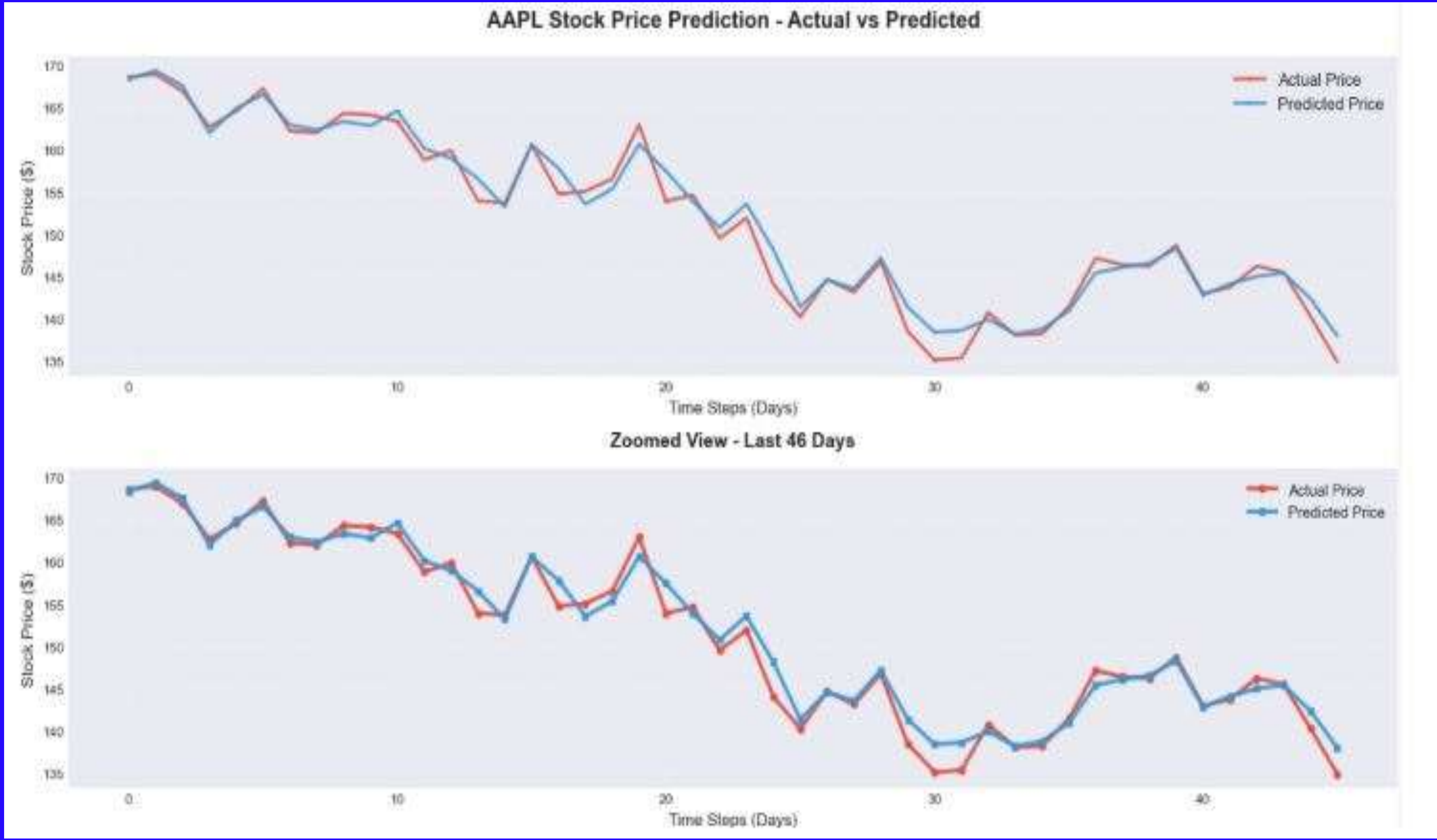
❖Proper knowledge management ensures the system can learn from past errors and improve prediction accuracy over time.

❖Secure and structured storage supports scalability, transparency, and collaborative access for future research or integration.

USER INTERACTION AND COLLABORATION

- ❖The system offers a user-friendly interface where investors can easily access stock predictions and insights.
- ❖Users can customize stock selection, time frames, and feature preferences, making the system highly flexible.
- ❖Interactive visualizations allow users to compare actual vs predicted values in an intuitive way.
- ❖Collaboration is enabled by integrating the system with financial advisory platforms for team-based decision-making.
- ❖Real-time updates ensure that users always work with the most recent and accurate market forecasts.
- ❖The platform encourages knowledge sharing and collaborative analysis, improving decision outcomes for groups or organizations.

SAMPLE OUTPUT



SAMPLE OUTPUT

```
📊 Performance Metrics:
============================================================
TRAINING SET:
MAE:   $0.03
RMSE:  $0.05
R²:    1.0000
MAPE:  0.02%
============================================================
TEST SET:
MAE:   $1.21
RMSE:  $1.62
R²:    0.9747
MAPE:  0.82%
============================================================
✅ Model generalization looks good!
```

SAMPLE OUTPUT

```
🔮 Forecasting next 1 days...
✅ Predicted price for next trading day: $137.89
   Current price: $134.81
   Expected change: $3.08
```

CONCLUSION

❖XGBoost proves to be a powerful approach for handling the volatility of stock prices with higher accuracy.

❖Compared to ARIMA and LSTM, it delivers faster predictions, scalability, and better interpretability.

❖Feature engineering with lags, moving averages, and returns significantly enhances forecasting performance.

❖The model supports real-time updates, multiple stocks, and adapts well to dynamic financial data.

❖This system creates a strong foundation for future work, including sentiment analysis and intelligent dashboards.

FUTURE ENHANCEMENT

- Integration of real-time sentiment and news data will capture market psychology and improve forecasting precision.
- The system can evolve to suggest investment strategies and portfolio adjustments based on predicted stock trends.
- Advanced dashboards with dynamic visualizations will enable real-time monitoring and comparative stock analysis.
- Hosting the system on cloud platforms will ensure scalability, faster computation, and seamless multi-user access.
- Automated data pipelines and live model updates will maintain consistent accuracy and adaptability to market changes.

REFERENCES

1. Patel, M., Jariwala, K., and Chattopadhyay, C., "An Approach Toward Stock Market Prediction and Portfolio Optimization in Indian Financial Sectors," IEEE Transactions on Computational Social Systems, vol. 12, no. 1, February 2025.
2. Roy, P. K., Kumar, A., and Singh, A., "Forecasting Bitcoin Prices Using Deep Learning for Consumer-Centric Industrial Applications," IEEE Transactions on Consumer Electronics, vol. 70, no. 1, pp. 1351, February 2024.
3. Hou, J., Dong, Z., Zhou, J., and Liu, Z., "Discovering Predictable Latent Factors for Time Series Forecasting," IEEE Transactions on Knowledge and Data Engineering, vol. 36, no. 10, October 2024.
4. Rosa, M. J. A., Souza, M. R., Machado, C. L. S., Rigo, S. J., and Barbosa, J. L. V., "Brazilian Stock Market Forecast with Heterogeneous Data Integration for a Set of Stocks," IEEE Latin America Transactions, vol. 23, no. 7, July 2025.

REFERENCES

5. Doroslovački, K., and Gradojevic, N., "A Novel Market Sentiment Analysis Model for Forecasting Stock and Cryptocurrency Returns," IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 54, no. 9, September 2024.
6. Zhu, Q., Li, J., and Liu, S., "Decomposition-Based Dynamic Inductive Graph Embedding Learning Method to Forecast Stock Trends," IEEE Transactions on Computational Social Systems, published August 30, 2024.
7. Li, S., Tang, G., Chen, X., and Lin, T., "Stock Index Forecasting Using a Novel Integrated Model Based on CEEMDAN and TCN-GRU-CBAM," IEEE Access, published July 29, 2024.
8. Alam, K., Bhuiyan, M. H., Haque, I. U., Monir, M. F., and Ahmed, T., "Enhancing Stock Market Prediction: A Robust LSTM-DNN Model Analysis on 26 Real-Life Datasets," IEEE Access, published July 29, 2024.

THANK YOU