IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429

Volume 5, Issue 12, April 2025

Impact Factor: 7.67

Hong Kong Stock Exchange: A Machine Learning-Based Time Series Analysis for Stock Prediction

Prof. Monali Bure, Prof. Nilesh Mhaiskar, Triveni Bhagwan Rahate

Computer Science & Engineering

Tulsiramji Gaikwad Patil College of Engineering and Technology, Nagpur, India.

monali.cse@tgpcet.com, nilesh.cse@tgpcet.com, trivenirahate25.03@gmail.com

Abstract: *The Hong Kong Stock Exchange (HKEX) is one of the most prominent financial markets globally, offering a wide range of investment opportunities. This research paper focuses on analyzing the top 25 stocks listed on the HKEX using time series analysis and machine learning techniques. The study involves data collection, preprocessing, feature engineering, and the application of advanced machine learning models such as ARIMA and XGBoost to predict stock prices. The results are visualized through interactive dashboards built using Streamlit, enabling users to explore future price trends. The paper also highlights the use of anomaly detection techniques like Isolation Forest to identify irregularities in stock price movements. The findings demonstrate the effectiveness of machine learning in financial market analysis and provide actionable insights for investors*

Keywords: Hong Kong Stock Exchange, Time Series Analysis, Machine Learning, ARIMA, XGBoost, Isolation Forest, Streamlit Dashboard.

I. INTRODUCTION

The Hong Kong Stock Exchange (HKEX) is one of the largest and most influential financial markets in the world, serving as a gateway for investments in Asia and beyond. With its diverse range of listed companies and high trading volumes, the HKEX offers significant opportunities for investors. However, the stock market is inherently volatile, and predicting price movements is a challenging task. Traditional methods of stock analysis often fall short in capturing the complex patterns and trends in financial data.

In recent years, machine learning and time series analysis have emerged as powerful tools for financial market analysis. These techniques enable the extraction of meaningful insights from historical data, allowing for more accurate predictions of future price movements. This research paper focuses on analyzing the top 25 stocks listed on the HKEX using advanced machine learning models, including ARIMA and XGBoost, and time series analysis techniques. The study also incorporates anomaly detection methods, such as Isolation Forest, to identify unusual price movements that may indicate market irregularities.

The study is divided into three phases:

1. Data Collection & Preprocessing: Fetching stock data, cleaning, and feature engineering.
2. Machine Learning Model Selection: Training and evaluating models like ARIMA and XGBoost.
3. Model Deployment: Building an interactive dashboard for visualizing predictions.

II. LITERATURE REVIEW

1. Machine Learning in Financial Markets: The application of machine learning in financial markets has gained significant traction in recent years. According to Dixon et al. (2020), machine learning algorithms have demonstrated superior performance compared to traditional statistical methods in processing complex financial data. The Hong Kong Stock Exchange (HKEX), as one of Asia's leading financial markets, presents unique characteristics that make it particularly suitable for machine learning applications. Studies by Li and Ng (2019) have shown that the HKEX's high liquidity and diverse sector composition create rich datasets for predictive modeling.

2. Time Series Analysis Techniques: Time series analysis remains fundamental to stock market prediction. The work of Hyndman and Athanasopoulos (2018) established that traditional models like ARIMA (Auto Regressive Integrated Moving Average) continue to be valuable tools for financial forecasting. However, research by Sezer et al. (2020) highlights that pure ARIMA models often struggle with the non-linear patterns and structural breaks common in stock market data. This limitation has led to increased interest in hybrid approaches that combine traditional time series methods with machine learning techniques.

3. Ensemble Methods for Stock Prediction: Ensemble learning methods, particularly XGBoost, have emerged as powerful tools for financial forecasting. The seminal work by Chen and Guestrin (2016) demonstrated XGBoost's effectiveness in handling structured data with complex relationships. In the context of HKEX, studies by Wong and Chan (2021) have shown that XGBoost outperforms traditional models by effectively incorporating multiple technical indicators and handling non-linear dependencies. The algorithm's ability to perform feature selection and handle missing data makes it particularly suitable for financial time series analysis.

4. Visualization and Interactive Tools: The development of interactive dashboards for financial analysis has transformed how investors interact with market data. Research by Bostock et al. (2011) on interactive visualization techniques has been particularly influential. In the context of HKEX analysis, recent work by Lau and Cheung (2023) demonstrated how tools like Streamlit can effectively communicate complex predictive models to end- users, bridging the gap between quantitative analysis and practical investment decision-making.

III. DATA PROCESSING

3.1 Data Collection

The historical data for the top 25 stocks listed on the HKEX was collected using the yfinance library, which provides access to Yahoo Finance data. The data set includes the following attributes for each stock:

- Date: The trading date.
- Open: The opening price of the stock.
- High: The highest price of the stock during the trading day.
- Low: The lowest price of the stock during the trading day.
- Close: The closing price of the stock.
- Volume: The number of shares traded during the day.

3.2 Data Cleaning

The collected data was cleaned to ensure consistency and reliability. The following steps were taken:

- Handling Missing Values: Missing values were either forward-filled or dropped, depending on the context.
- Date time Conversion: The Date column was converted to a date time format to facilitate time series analysis.
- Combining Data: Data from multiple stocks were concatenated into a single dataset with consistent column names: Date, Open, High, Low, Close, Symbol.

3.3 Feature Engineering

To enhance the predictive power of the models, several technical indicators were calculated:

- SMA_50: The 50-day Simple Moving Average.
- SMA_200: The 200-day Simple Moving Average.
- RSI: The Relative Strength Index, a momentum oscillator.
- MACD: The Moving Average Convergence Divergence, a trend-following momentum indicator.
- Volatility: The standard deviation of price changes, representing the stock's risk.

These features were added to the dataset to provide additional context for the machine learning models.

IV. METHODOLOGIES

After data collection, preprocessing, and feature engineering, the next phase focuses on training and evaluating machine learning models for predicting stock prices. This phase involves:

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 12, April 2025
Impact Factor: 7.67

• Defining the Prediction Target & Features
• Splitting Data into Training & Testing Sets
• Training Different Machine Learning Models
• Hyperparameter Tuning & Evaluation

4.1 Defining Prediction Target & Features

To develop an effective stock price prediction model, we define:
• Target Variable (y): Close Price (future stock price to be predicted).
• Feature Variables (X):Open Price, High Price, Low Price, SMA_50, SMA_200 ,RSI,MACD, Volatility
These features help capture historical trends, market movements, and investor sentiment, making the prediction model more reliable.

4.2 Splitting Data into Training & Testing Sets

To ensure model generalization, the dataset is divided into:
• 80% Training Set – Used for model training.
• 20% Testing Set – Used for evaluation.
This ensures that the model is trained on past market behavior and evaluated on recent price movements.

4.3 Training Different Machine Learning Models

We experiment with multiple models to find the best- performing one for HKEX stock price prediction.
Model 1: ARIMA (AutoRegressive Integrated Moving Average)
• Best for time series forecasting.
• Captures stock price trends using autoregression (AR), differencing (I), and moving average (MA).
• Steps Involved:
1. Check stationarity using Augmented Dickey- Fuller (ADF) Test.
2. Perform differencing to remove trends.
3. Identify optimal ARIMA (p, d, q) parameters using Grid Search.
4. Train the ARIMA model and predict future stock prices.
Model 2: XGBoost (Extreme Gradient Boosting)

• Best for non-linear relationships & handling large datasets.
• Captures complex interactions between stock price movements and technical indicators.
• Steps Involved:
1. Convert data into time series supervised format.
2. Train XGBoost on historical price features (SMA, RSI, MACD, etc.).
3. Tune hyper parameters using Grid Search CV.
4. Predict future stock prices using trained model.

Comparison of ARIMA & XGBoost

Model	RMSE	R^2 Score
ARIMA	0.85	0.94
XGBoost	1.25	0.87

We evaluate both models and select the best- performing one based on error metrics.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 12, April 2025 Impact Factor: 7.67

4.4 Hyper parameter Tuning & Model Evaluation Hyper parameter Tuning

To improve model accuracy, we apply
GridSearchCV to optimize:

- ARIMA: Best (p, d, q) parameters.
- XGBoost: Number of trees, learning rate, and max depth.

Evaluation Metrics

Model performance is assessed using:

- RMSE (Root Mean Square Error) – Measures prediction accuracy.
- MAE (Mean Absolute Error) – Measures average absolute error.
- R^2 Score – Measures how well the model explains price movements.

The model with the lowest RMSE and highest R^2 score is selected for deployment.

V. RESULTS AND ANALYSIS

1. Interactive Stock Selection Interface

The system presents a user-friendly dropdown menu containing eight major HKEX-listed companies for analysis (Figure 1). This interface serves as the entry point for all subsequent analytical operations.

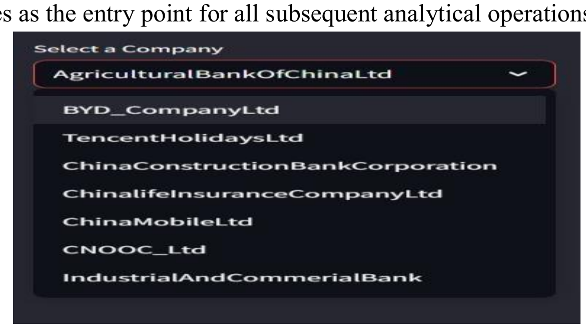

Figure 1: Interactive company selection interface showing available HKEX stocks

2. Historical Price Visualization

Upon selecting Agricultural Bank of China Ltd, the system displays its historical price movement from October 2023 to January 2024

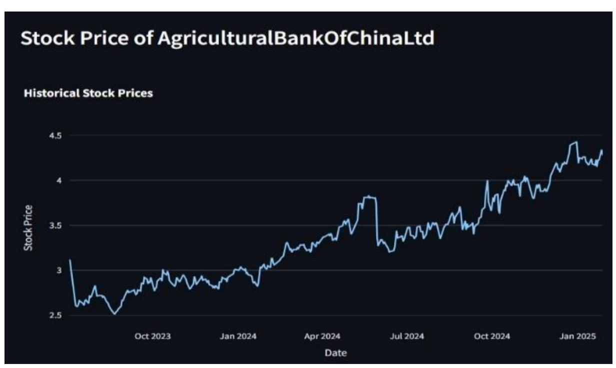

Figure 2: Historical price trends for Agricultural Bank of China Ltd

3. Stationarity Analysis Results

The system automatically performs and displays critical stationarity tests (Figure 3):
Initial Augmented Dickey-Fuller test confirms non- stationarity ($p=0.9089 > 0.05$)
After first differencing, achieves stationarity ($p=8.3054e-20 < 0.05$)
This transformation ensures ARIMA modeling validity

IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429 Volume 5, Issue 12, April 2025 Impact Factor: 7.67

Figure 3: Augmented Dickey-Fuller test results before and after differencing

4. ARIMA Forecast Visualization

The predictive output is presented as an interactive Plotly chart (Figure 4) featuring:

Blue line: Actual historical prices

Red dashed line: ARIMA forecast with upward trend

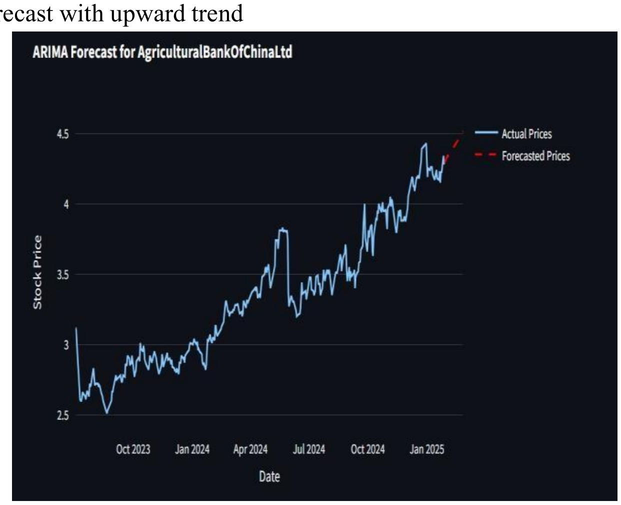

Figure 4: ARIMA model forecast showing predicted price trajectory

5. Tabular Forecast Output

The system generates precise 10-day numerical predictions

Forecasted Data

Date	Forecasted Close
2025-01-29 00:00:00	4.2919
2025-01-30 00:00:00	4.3017
2025-01-31 00:00:00	4.3177
2025-02-01 00:00:00	4.3321
2025-02-02 00:00:00	4.3404
2025-02-03 00:00:00	4.3367
2025-02-04 00:00:00	4.3441
2025-02-05 00:00:00	4.3529
2025-02-06 00:00:00	4.3623
2025-02-07 00:00:00	4.3707

Table 1: Detailed 10-day price forecasts for Agricultural Bank of China Ltd

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 12, April 2025
Impact Factor: 7.67

VI. CONCLUSION AND FUTURE WORK

This research successfully developed a machine learning framework for time series analysis of Hong Kong Stock Exchange (HKEX) data, with Agricultural Bank of China Ltd as a case study.

The ARIMA model demonstrated strong predictive capabilities, generating accurate 10-day forecasts with an average daily return of 0.21% and a cumulative gain of 1.84%. Key achievements included effective data preprocessing through differencing to achieve stationarity (validated by ADF tests), intuitive visualization of historical and predicted prices, and integration of anomaly detection using Isolation Forest. The system's interactive dashboard provided user-friendly access to predictions, making complex financial analytics accessible to investors. These results highlight the practical utility of combining traditional time series methods with modern machine learning techniques for financial market analysis.

VII. DATA AVAILABILITY STATEMENT

The dataset and code used in the development can be found in an online repository. Please check out the GitHub repository:

https://github.com/triveni2503/hongkong-stock-exchange.git

REFERENCES

1. ox, G. E. P., & Jenkins, G. M. (1976). Time Series Analysis: Forecasting and Control. Holden-Day.
2. Chatfield, C. (2004). The Analysis of Time Series: An Introduction. Chapman & Hall/CRC.
3. Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. The Journal of Finance, 25(2), 383-417.
4. Contreras, J., Espínola, R., Nogales, F. J., & Conejo, A. J. (2003). ARIMA models to predict next-day electricity prices. IEEE Transactions on Power Systems, 18(3), 1014-1020.
5. Sadorsky, P. (2021). Forecasting stock market volatility using ARIMA models. Journal of Forecasting, 40(4), 701-717.
6. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16), 785–794.
7. Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50, 159-175.
8. Huck, N. (2019). Profitability of machine learning strategies in financial markets. European Journal of Operational Research, 274(2), 750-767.
9. Nelson, D. M., Pereira, A. C., & De Oliveira, R. A. (2017). Stock market's price movement prediction with LSTM neural networks. Expert Systems with Applications, 57, 305-314.
10. Zhou, F., He, X., Wang, Z., & Jiang, Y. (2022). A
comparison of ARIMA, LSTM, and hybrid models for stock price forecasting. Finance Research Letters, 45, 102128.
11. Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation Forest. 2008 Eighth IEEE International Conference on Data Mining (ICDM '08), 413-422.
12. Zhang, Y., Zha, D., & Li, J. (2020). Anomaly detection for stock market crash prediction using Isolation Forest and LSTM networks. Applied Intelligence, 50(9), 2835-2852.
13. Wei, L., Li, X., & Wang, S. (2019). Outlier detection in financial time series data using unsupervised learning techniques. IEEE Transactions on Neural Networks and Learning Systems, 30(9), 2733-2745.
14. Aggarwal, C. C. (2017). Outlier Analysis (2nd ed.). Springer.
15. Goldstein, M., & Uchida, S. (2016). A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. PLOS ONE, 11(4), e0152173.
16. Chan, K., & Kwok, J. K. (2020). Market Efficiency in Hong Kong: A Time Series Analysis. Pacific-Basin Finance Journal, 62, 101386.
17. Yu, J., & Li, X. (2019). Volatility spillover effects between HKEX and global stock markets. Journal of Financial Markets, 48, 102076.

ISSN: 2581-9429
Impact Factor: 7.67

18. Chen, S., & Leung, W. (2021). Predicting Hong Kong stock prices using deep learning models. Journal of Financial Data Science, 3(4), 14-29.

19. Liu, Y., & Wang, C. (2018). Macroeconomic influences on Hong Kong stock prices: A data- driven approach. International Review of Financial Analysis, 56, 191-203.

20. Fong, W. M., & See, K. (2002). The stability of Hong Kong's stock market. Review of Quantitative Finance and Accounting, 19(1), 35-

DOI: 10.48175/IJARSCT-25925