

From Data To Decision: Empowering Companies and Investors With Hybrid AI Stock Prediction Method

Norjiah Muslim*, Rosita Binti Hussin, Fatin Fasiah Binti Johari

Faculty of Business and Accountancy, Universiti Selangor, Malaysia

*Corresponding Author

DOI: <https://dx.doi.org/10.47772/IJRISS.2025.90600046>

Received: 26 May 2025; Accepted: 27 May 2025; Published: 28 June 2025

ABSTRACT

This research presents a hybrid Artificial Intelligence (AI) model for stock price prediction, combining several advanced techniques to enhance the accuracy and reliability of financial forecasting. The model integrates neural network methods such as artificial Neural Networks (ANN), in conjunction with a sliding window approach and hierarchical clustering. The sliding window method segments historical stock data into fixed intervals, enabling the model to detect localized temporal trends, while hierarchical clustering groups similar historical patterns to improve forecasting relevance. A comprehensive literature review was conducted to evaluate existing hybrid AI approaches and identify research gaps. Feature selection was performed using stepwise regression and leverage analysis to refine the dataset before model training. The hybrid model demonstrated superior performance compared to traditional methods, based on evaluation metrics such as RMSE and MAE, both in backtesting and real-time simulation scenarios. The results confirm the model's ability to generate timely and accurate predictions, supporting more informed investment decisions. This study also recommends future enhancements such as sentiment analysis integration, broader market validation, and real-time deployment capabilities, affirming the strong potential of hybrid AI models in financial forecasting.

Keywords: hybrid Artificial Intelligence (AI) model, artificial Neural Networks (ANN), a sliding window approach, hierarchical clustering.

INTRODUCTION

Financial markets are dynamic and turbulent, which makes it extremely difficult for businesses and investors to make well-informed decisions. Traditional stock prediction techniques, which frequently depend on linear models and trends in historical data, have not been able to adequately capture the complicated, nonlinear patterns that characterize market behaviors. Transformative powers brought about by the development of artificial intelligence (AI) have made it possible to analyze enormous datasets and uncover complex patterns that were previously difficult to find. (Guan, Y., & Zong, Z. 2024). This gap is filled by artificial intelligence (AI), which transforms raw data into actionable insights. The Hybrid AI Stock Prediction Method combines several types of AI approaches, such as statistical modeling, machine learning, and deep learning, to increase the precision of stock market forecasts. Recent developments in artificial intelligence have produced hybrid models, which integrate many algorithms to improve prediction accuracy. For example, combining Convolutional Neural Networks (CNN) with attention mechanisms and Long Short-Term Memory (LSTM) networks has demonstrated potential for capturing temporal and spatial relationships in stock data. (Lin, et al. 2022). Prediction models have been significantly improved by combining machine learning techniques with conventional financial indicators, providing a more comprehensive understanding of market dynamics. (Olubusola, et al. 2024).

To improve the accuracy of stock market projections, the Hybrid AI Stock Prediction Method integrates several AI techniques, including statistical modelling, machine learning, and deep learning. Recent advancements in AI have led to the development of hybrid models that combine various algorithms to enhance predictive accuracy. For instance, integrating Convolutional Neural Networks (CNN) with Long Short-Term

Memory (LSTM) networks and attention mechanisms has shown promise in capturing both spatial and temporal dependencies in stock data. According to Jin et al. (2023), despite these developments, there are still difficulties in accurately simulating the temporal features of stock data. To solve this problem, the sliding window approach has been used, which entails examining fixed-size chunks of data over time. Finding the ideal window size, however, is still a crucial issue because it has a big impact on the model's capacity to identify pertinent patterns without overfitting.

Despite these advances, a major challenge remains in modeling the temporal evolution of stock data accurately. To address this, the sliding window technique is employed to divide time series data into fixed-length segments, allowing the detection of localized patterns over time. However, determining the appropriate window size is critical, as it affects the model's ability to generalize without overfitting. To further enhance pattern recognition, we incorporate hierarchical clustering, which enables the grouping of historical patterns based on similarity in their price trajectories. By identifying clusters of similar behaviors, the model can retrieve relevant reference patterns for comparison and forecasting. When integrated with deep learning and feature selection, this approach results in a robust framework capable of delivering more accurate and interpretable predictions of stock market trends

Research Problem

A gap still exists in the efficient integration of temporal alignment techniques with hybrid AI frameworks, despite the widespread use of AI-based models in stock prediction. The advantages of integrating DTW and sliding window techniques are frequently overlooked by current models, which results in less-than-ideal temporal pattern detection. Furthermore, the usefulness and dependability of these models in actual trading situations are restricted by the absence of a cohesive strategy that combines both sophisticated AI algorithms and conventional financial indicators. In order to overcome these obstacles, this study will create a hybrid AI stock prediction model that combines deep learning algorithms, conventional financial analytics, and Hierarchical Clustering and sliding window approaches. The goal is to improve the model's capacity to identify complex temporal trends and offer more precise and trustworthy stock forecasts, enabling businesses and investors to make wise choices in the ever-changing financial market environment.

This study aims to propose a new complex methodology geared towards identifying the optimal historical dataset with similar patterns by utilizing various algorithms for each stock item. The core intention is to enhance the accuracy of daily stock price predictions, bridging the gap between data and decisive action for companies and investors. In light of this, the research seeks to empower stakeholders with actionable insights derived from a Hybrid AI Stock Prediction Method, meticulously crafted and validated to ensure reliability and real-time applicability.

To realise the ambitious goal of evolving the landscape of stock price prediction, the following SMART objectives are articulated:

1. Conduct a comprehensive review of existing AI techniques in stock prediction, focusing on hybrid models that integrate various algorithms and traditional financial methods
2. To investigate the impact of the sliding window method on investment decision making by applying the sliding window method to historical stock data and analysing its influence on predicting stock trends and aiding investment decisions.
3. To evaluate the integration of Hierarchical Clustering with a sliding window for enhanced prediction by, aiming at enhanced pattern recognition and accuracy in stock predictions.
4. To develop and test a hybrid AI stock prediction model by creating a model that combines AI techniques, Hierarchical Clustering, the sliding window method, and traditional financial analytics for improved stock predictions.
5. Deploy the developed hybrid AI model in real-time trading scenarios to assess its applicability, performance, and reliability.

LITERATURE REVIEW

Hybrid AI Techniques in Stock Prediction

Stock price prediction has evolved significantly with the rise of AI and machine learning (ML) technologies. Once dominant, traditional statistical models like ARIMA and GARCH have gradually been supplanted by AI-driven techniques capable of capturing nonlinear and dynamic market behaviour (Nikou et al., 2019; Chen, 2023). Deep learning (DL) models, including Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Convolutional Neural Networks (CNN), have shown remarkable performance in modelling complex temporal dependencies (Zhang et al., 2023; Mehtab & Sen, 2022).

Recent studies highlight the success of hybrid models that combine multiple AI techniques. For instance, Zhang et al. (2023) proposed a CNN-BiLSTM-Attention model, integrating spatial, temporal, and attention mechanisms to enhance stock price prediction accuracy. Similarly, Kim and Kim (2019) demonstrated that feature fusion using LSTM-CNN improved the generalisation ability across different datasets. Generative models like GANs have also been explored. Kumar et al. (2022) combined GANs with enhanced root mean square error metrics to predict stock price movements, demonstrating the potential of generative approaches in financial forecasting. Furthermore, transformer-based models, known for their attention mechanisms, have been applied to capture long-term dependencies in stock data. Muhammad et al. (2023) demonstrated the efficacy of a transformer-based deep learning model in handling data in sequence by using it to forecast stock prices in the Bangladeshi stock market.

Furthermore, research has emphasized the importance of attention mechanisms and feature engineering in improving the performance of hybrid AI models. For example, the integration of external market signals such as news sentiment, macroeconomic indicators, and trading volumes has significantly improved model interpretability and accuracy (Soni et al., 2022). The hybridisation is appropriate for real-time forecasting environments since it guarantees generalizability across various stock markets in addition to increasing predictive accuracy. (Kumbure et al., 2022).

Impact of Sliding Window Method on Investment Decision Making

The sliding window method is pivotal for segmenting time series data into meaningful subsequences, which allows models to detect local patterns (Dai et al., 2022). Jeon et al. (2018) showed that window-based tracking increased the accuracy of forecasting by enhancing pattern graph recognition in financial datasets. By generating several overlapping training samples, the technique makes model training easier and works especially well in dynamic settings like financial markets. The window size selection is important because it affects the model's capacity to identify appropriate patterns without overfitting. (Pahwa et al., 2017).

Mittal and Nagpal (2022) revealed the usefulness of the sliding window technique in segmenting time series data for focused analysis by using it to find dependable stocks for mid- and long-term investments. Researchers can analyse the impact of particular periods on stock trends by using the sliding window method on historical stock data, which helps investors make better decisions. Because of the technique's flexibility, different window sizes can be examined to maximise model performance and capture relevant temporal dynamics.

Optimising window sizes to balance model complexity and performance has been the focus of recent research. According to Ampomah et al. (2021), using the sliding window method stabilises model learning in short-term prediction tasks and enhances data stationarity. Because it continuously updates predictions based on new data, its application fosters real-time decision-making. Further study has shown that altering the window size can have an immense impact on the model's accuracy, indicating the necessity of optimisation according to particular market circumstances.

To improve reliability, the latest study also recommends combining ensemble learning methods with sliding windows. For example, boosting or bagging algorithms combined with window-based LSTM can enhance model stability and lower variance (Han & Fu, 2023). More reliable investment signals are produced by this

integrated approach, particularly during periods of market fluctuations. 3.2.3 Integration of Dynamic Time Warping (DTW) with Sliding Window for Enhanced Prediction

Hierarchical Clustering for Pattern Matching

Hierarchical clustering has been widely explored as an unsupervised learning technique for pattern recognition in financial time series, particularly in the context of stock price prediction. Its strength lies in the ability to organize data into a tree-like structure (dendrogram), enabling flexible grouping based on the similarity of temporal patterns. Unlike flat clustering methods such as k-means, hierarchical clustering does not require the pre-specification of the number of clusters, making it especially suitable for dynamic financial environments where the optimal number of patterns is unknown or evolves over time (Mao et al., 2019).

In stock forecasting, pattern matching involves identifying historical sequences that resemble current market behavior. Hierarchical clustering has been used to cluster historical price patterns based on distance metrics such as Euclidean or Dynamic Time Warping, though recent studies increasingly favor vectorized window-based sequences for efficiency (Jeon et al., 2020). Once clusters are formed, the most recent pattern (e.g., today's price trend) can be matched to a similar group, and the subsequent movement of matched patterns is used to predict future prices.

Studies by Zhang and Zhou (2021) demonstrate that integrating hierarchical clustering into the stock prediction pipeline improves accuracy by filtering out dissimilar historical data, thereby refining the input space for machine learning models. Moreover, combining hierarchical clustering with feature selection methods (e.g., Lasso or stepwise regression) ensures that only the most predictive features are retained for model training, enhancing interpretability and performance. Therefore, hierarchical clustering offers a powerful mechanism for uncovering latent structure in temporal financial data, making it a valuable preprocessing step for hybrid AI forecasting models.

Development and Testing of a Hybrid AI Stock Prediction Model

The goal of hybrid AI models is to create systems that are accurate and comprehensible by integrating the best features of traditional financial analytics and different artificial intelligence techniques. Although nonlinear dependencies in financial time series can be captured by deep learning models like CNN or LSTM, their interpretability and robustness in a variety of market conditions are improved when combined with statistical or rule-based indicators (e.g., RSI, MACD) (Nikou et al., 2019; Mittal & Nagpal, 2022).

Data segmentation and clustering before model training greatly enhanced model generalisation across asset classes, as demonstrated by Li et al. (2023). Furthermore, researchers like Jang and Lee (2020) demonstrated that combining technical indicators with DL architectures produced better price direction and fluctuation prediction. Another important direction in hybrid model design is multimodal data integration, combining numerical stock data, textual sentiment from news and social media, and macroeconomic variables. For example, Schumaker and Chen (2009) demonstrated that integrating financial news text with numerical data using machine learning improved stock return predictions.

Metrics like Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R^2 are commonly used in both walk-forward validation and backtesting settings for performance testing of such models (Rouf et al., 2021; Mehtab & Sen, 2022). The concern over interpretability is growing in addition to accuracy. To visualize the significance of input features and give users confidence in predictions, recent models employ SHAP values or LIME techniques (Zhang et al., 2023).

Deployment in Real-Time Trading Scenarios

Real-world deployment is the last stage of validating AI models. Accuracy alone is not enough for real-time stock trading systems, they also need speed, stability, and flexibility. The model needs to run fast, adapt to changes in the market, and integrate easily with data pipelines. This necessitates effective infrastructure design in addition to strong architectures (Kumar et al., 2022; Muhammad et al., 2023). When implementing hybrid models, delay and scaling are two major performance bottlenecks. In high-frequency trading, where even

millisecond delays can result in opportunity loss, low-latency execution is particularly crucial. As a result, researchers are now investigating edge AI, cloud computing, and GPU acceleration in real-world implementations (Awan et al., 2021; Moukalled, 2019).

Online learning methods that enable models to update gradually as new data becomes available are also advantageous for real-time systems. In order to maintain model relevance without requiring total retraining, Bhat and Ahmad (2021) proposed an LSTM-based framework that retrains with each batch of new data. Reinforcement learning for autonomous trading agents is another developing field in which models pick the best trading tactics through interactions with the environment that are based on rewards. Research has shown that integrating RL with risk-adjusted metrics and financial constraints results in more profitable and realistic strategies (Yang et al., 2020). Finally, user interface integration, such as trading signals, dashboards, or alerts, is necessary for real-time deployment. To facilitate human-in-the-loop decision-making, interpretability tools and scenario analysis features are necessary because usability is crucial for investor adoption (Shahi et al., 2020).

Outline of The Proposed Model

In this section, we describe the overall framework of the proposed stock price prediction model. The process begins with data preprocessing to generate structured, continuous time series data from raw FBMKLCI intraday transactions. The preprocessed data is then segmented into fixed-length patterns using a sliding window approach, allowing the extraction of temporal trends.

Next, hierarchical clustering is applied to group historical patterns that exhibit similar behaviors. From these clustered patterns, relevant features are identified using Lasso regression, which selects only the most significant predictors (such as open and high prices) for each cluster.

Finally, the selected features are used to neural network prediction, including Artificial Neural Networks (ANN). These models are evaluated based on their ability to accurately predict the stock's closing price in the short term. The entire pipeline is designed from a data analysis and processing perspective to maximize prediction accuracy while maintaining model interpretability.

Aggregation of Stock Data

The original dataset consists of tick-by-tick transaction data from the FBMKLCI index, where each entry represents an individual trade event. As a result, the data is inherently non-continuous, with irregular timestamps and potential gaps in pricing when no transaction occurs at a specific time, as illustrated in Figure 1(a). This irregularity makes it challenging to model price movements effectively or apply time series forecasting techniques directly.

To address this, the raw tick data is aggregated into uniform 10-minute intervals, generating a continuous time series composed of standard OHLCV features (Open, High, Low, Close, Volume). This transformation, illustrated in Figure 1(b), enables consistent pattern recognition and modeling by capturing price behavior over regular time windows.

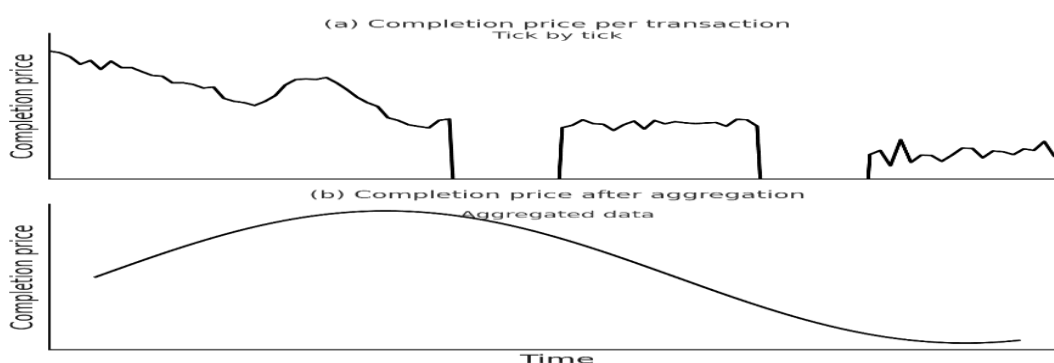


Figure 1: The Aggregation for raw data

Find Similar Patterns

To identify patterns that reflect similar price behavior, it is first necessary to segment the aggregated stock data into meaningful temporal patterns. In this study, we apply a sliding window method to generate overlapping patterns from the 10-minute interval OHLCV data, where each pattern spans a fixed length (e.g., a full trading day or 3-hour window). This approach results in a series of structured pattern instances that represent historical market behavior.

Figure 2 illustrates the process of generating these patterns from the continuous, aggregated time series. Each new pattern is offset by a small lag (e.g., 1 interval), ensuring that temporal trends are captured with high resolution. For instance, in a 3-hour window with 10-minute intervals, each pattern consists of 18 data points.

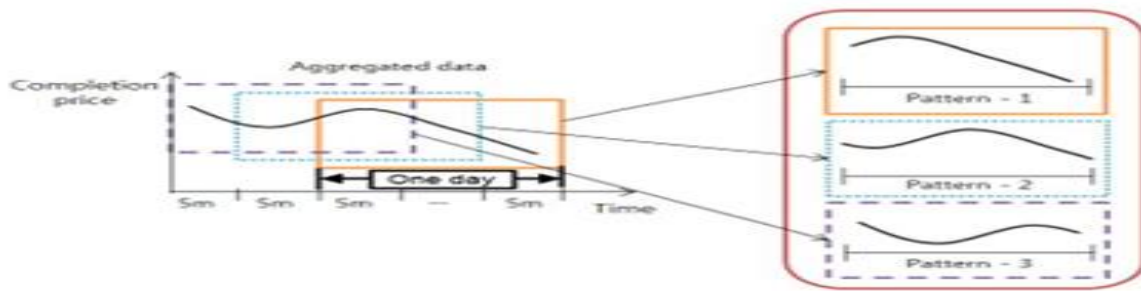


Figure 2 : Sliding window

To enable efficient pattern matching, we apply a hierarchical clustering algorithm to group structurally similar patterns based on their price movements. As shown in Figure 4, this method organizes patterns into a tree structure, where similar patterns are grouped as sibling or neighboring nodes. The current pattern (e.g., Pattern 1) is compared against historical patterns, and initial matches (e.g., Patterns 4 and 15) are identified from its local cluster neighborhood. If the number of similar patterns in the immediate cluster is insufficient for robust analysis, the search range can be extended within the clustering hierarchy. This dynamic adjustment allows the number of matched patterns to increase (e.g., from 2 to 12), improving the statistical reliability of the subsequent predictive modeling.

Figure 3 shows examples of similar patterns overlaid on real stock price trajectories, demonstrating how the clustering approach captures recurring shapes in price behavior. The use of hierarchical clustering not only accelerates the search process but also enhances the precision of pattern similarity analysis.

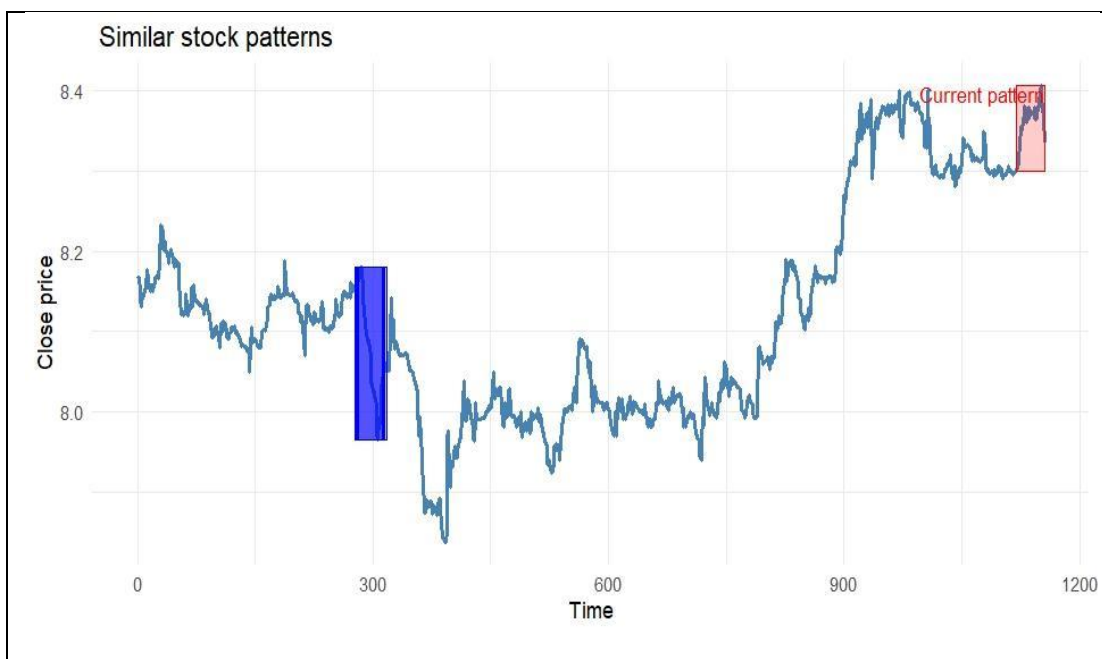


Figure 3: Similar stock patterns

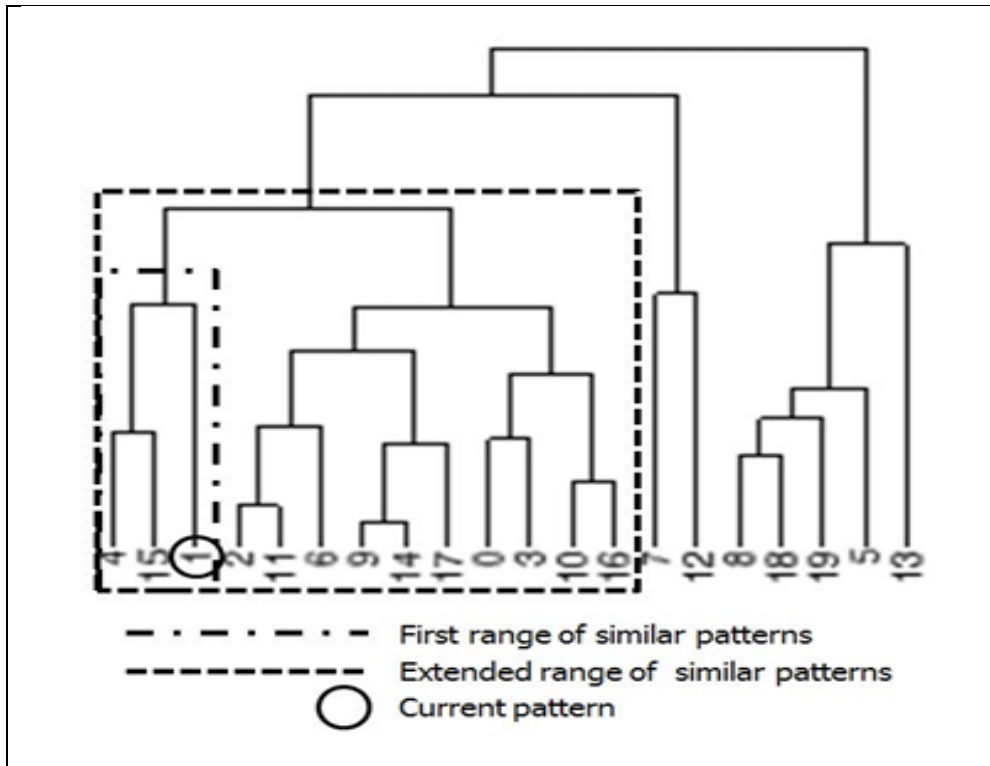


Figure 4 :Similar pattern through Hierarchical Cluster

Feature Selection using LASSO Regression

In the initial dataset, multiple variables may influence the stock price, but not all of them contribute meaningfully to the predictive model. Some features may introduce noise or redundancy, which can degrade model performance and interpretability. To address this, we apply for feature selection using the Least Absolute Shrinkage and Selection Operator (LASSO) regression technique.

Lasso performs both regularization and variable selection by imposing an L1 penalty on the regression coefficients. This constraint forces the coefficients of less important features to shrink toward zero, effectively removing them from the model. Unlike stepwise regression, which selects features incrementally through forward or backward elimination, LASSO considers all variables simultaneously and selects the optimal subset based on cross-validated prediction performance.

In this study, we consider the closing price as the dependent variable, and a set of predictor variables extracted from similar historical patterns (e.g., open, high, low, volume) as independent variables. The regression is conducted using the glmnet package in R. The model is fit with standardized variables, and the best regularization parameter (lambda) is determined using cross-validation. A total of five variables are removed after applying LASSO regression, as can be seen in Table 2.

Table 2 : Results of LASSO regression in Stock Data

Domestic purchase price Name	Choice	Domestic selling price Name	Choice	Foreign purchase price Name	Choice	Foreign selling price Name	Choice
Opening price	O	Opening price	O	Opening price	O	Opening price	O
High price	O	High price	O	High price	O	High price	O
Low price	X	Low price	X	Low price	X	Low price	X
Vol	0	Vol	X	Vol	0	Vol	0

Predicted Stock Data Generation using Artificial Neural Network (ANN)

Following the feature selection process, we employed an Artificial Neural Network (ANN) to forecast short-term stock prices. ANNs are widely recognized for their ability to model complex, nonlinear relationships in time series data (Kim & Han, 2000; Dunstall, 2005). The network is trained using a backpropagation algorithm, adjusting weights iteratively to minimize error between predicted and actual stock prices. The ANN structure includes one or more hidden layers to extract higher-level features from the input data. While deeper networks can capture more complexity, they also risk overfitting and learning instability (Li et al., 2023).

As shown in figure 5, to determine the optimal network architecture, models were trained with up to three hidden layers. The performance of each model was assessed using a set of widely accepted evaluation metrics. These included Mean Absolute Error (MAE), which quantifies the average magnitude of prediction errors, and Mean Squared Error (MSE), which penalizes larger deviations more heavily by squaring the residuals. Additionally, Root Mean Squared Error (RMSE) was employed as it represents the standard deviation of prediction errors, offering a more interpretable scale relative to the original data.

Beyond error-based metrics, the models were also evaluated using Precision, which measures the model's ability to correctly predict stock price movements within a defined error margin or threshold. This is particularly relevant in short-term financial forecasting, where accurate directional predictions are critical. The ANN model configuration that yielded the lowest MAE and RMSE values, along with high precision, was selected as the final model due to its superior performance in balancing accuracy and robustness.

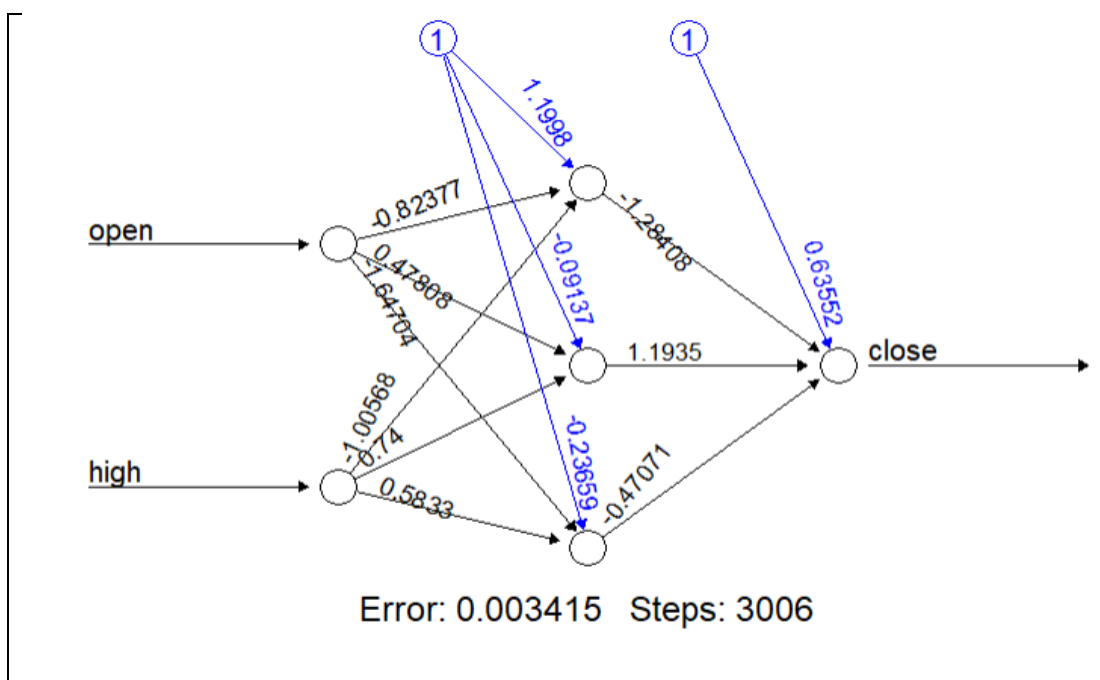


Figure 5 : ANN model with hidden three layer

Process:

1. Gather the predicted data output by the ANN and the actual data to perform a comparative analysis.
2. Choose appropriate accuracy measurement metrics. Common metrics include Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared, among others. Each metric provides insights into different aspects of prediction accuracy.
3. Compare the predicted values against the actual values using the selected metrics. This involves mathematical computations to quantify the discrepancies between predicted and real data.
4. Utilise graphical methods like plots or charts to visually represent the comparison, offering an intuitive understanding of the model's accuracy.

Real-Time Application and Continuous Improvement:

1. Deploy the model in real-time scenarios, monitor its predictions, and compare them with actual outcomes to gauge its real-time accuracy and reliability.
2. Establish a feedback loop where the model's predictions and their accuracies are continuously monitored. Utilise this data to constantly refine and enhance the model, ensuring it adapts to the evolving market dynamics and maintains optimal performance.

Evaluation

In this section, we describe the test dataset, which spans a three months period, and evaluate the prediction accuracy for each stock using Root Mean Squared Error (RMSE) as the primary performance metric.

Dataset and Test Results

To validate the effectiveness of the proposed model, we used a real historical stock dataset comprising multiple companies listed on BURSA Malaysia. The dataset spans a period from December 2024 to February 2025, and includes three representative stocks: CIMB, YTLPOWER, and TM. For model development, stock data from December 2024 to January 2025 was used as the training set, while data from January to February 2025 was reserved for testing. This split allows for a realistic short-term forecasting scenario that reflects actual market behavior. In the test scenario, the proposed forecasting framework—including hierarchical pattern clustering and Lasso-based feature selection—was applied to generate daily stock price predictions. The predicted values were then evaluated against the actual stock prices using Root Mean Squared Error (RMSE) to quantify prediction accuracy.

Evaluation Results

To evaluate the accuracy of the proposed forecasting method, we conducted experiments comparing actual stock data with predictions generated using the full model (including feature selection and machine learning) versus predictions made using feature selection alone. Figure 6 illustrates these comparisons for CIMB, YTLPOWER, and TM stocks on a selected test day. The x-axis represents time in 10-minute intervals, and the y-axis indicates the stock's closing price throughout the trading session. Figure 6(a) shows the comparison for CIMB. The predicted stock movement generated by the proposed model closely aligns with the actual price trends, capturing both the morning uptrend and afternoon decline. In contrast, the results from feature selection alone deviate more noticeably from the true price curve.

Follow by Figure 6(b) presents the case for YTLPOWER. While the overall trend follows the actual data, the predicted curve using the full model shows slight deviations, particularly in areas where price movements are less pronounced. Still, it outperforms the feature selection-only baseline. Then, Figure 6(c) illustrates the results for TM. Here, the proposed model maintains a close resemblance to the real stock price throughout the day, with minimal discrepancies in magnitude, demonstrating strong alignment.

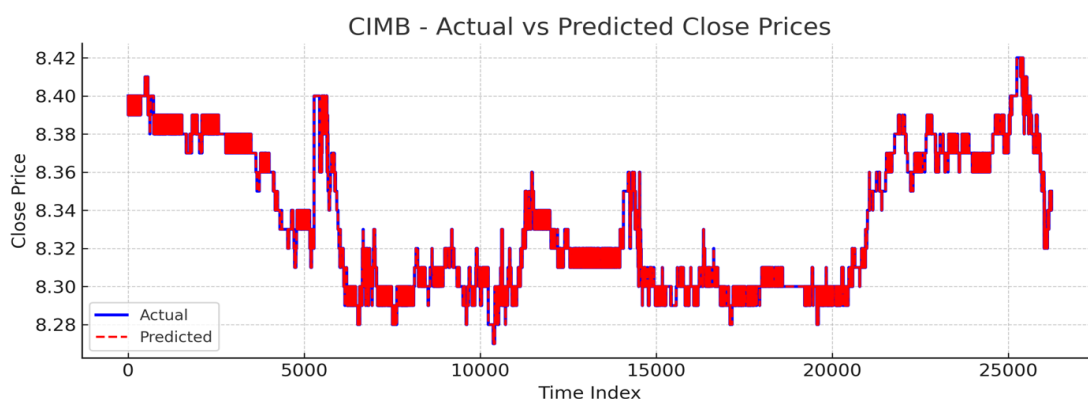


Figure 6(a) : Comparison Plot for CIMB

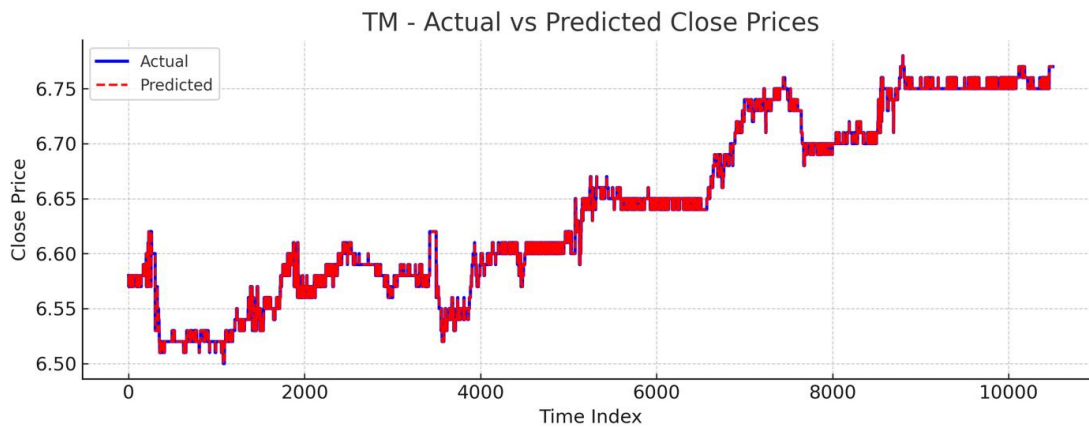


Figure 6 (b) : Comparison Plot for TM

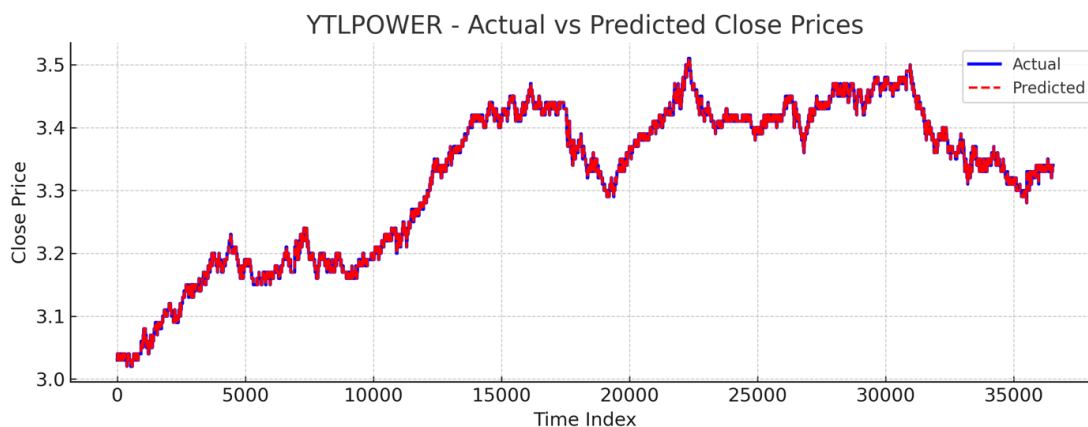


Figure 6 (c) : Comparison Plot for YTLPOWER

To quantify prediction performance, we used two widely adopted metrics in stock price forecasting: Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). These measures capture the average magnitude of prediction errors, with RMSE penalizing larger deviations more heavily.

As shown in Table 3, the proposed model achieved the highest accuracy for CIMB, with the lowest RMSE (0.00143) and MAE (0.00121), followed closely by TM, which also demonstrated low error values (RMSE: 0.00162, MAE: 0.00137). In contrast, YTLPOWER recorded higher errors (RMSE: 0.00530, MAE: 0.00470), suggesting either greater market volatility or reduced predictive power for this particular stock under the current model configuration. These results confirm that incorporating both pattern clustering and feature selection significantly improves prediction accuracy, particularly for more stable stocks, when compared to models relying solely on feature selection.

Table 3: Comparison of RMSE and MAE for difference stock data

Stock Name	RMSE	MAE
CIMB	0.00143	0.00121
YTLPOWER	0.0053	0.0047
TM	0.00162	0.00137

CONCLUSION

In this study, we proposed an integrated methodology for short-term stock price prediction that combines hierarchical pattern clustering, feature selection via Lasso regression, and Artificial Neural Networks (ANNs).

The primary motivation was to address the limitations of traditional time series models in capturing the complex, nonlinear, and high-frequency dynamic characteristic of modern financial markets.

Through extensive experimentation using intraday data from the FBMKLCI index, we demonstrated that historical stock price data contains sparse but meaningful repeating patterns. By segmenting the time series using a sliding window and applying hierarchical clustering, our model was able to identify patterns similar to the current trend. Feature selection using Lasso allowed us to isolate only the most relevant predictors—such as high and open prices—for each pattern, reducing model complexity and improving generalization.

We validated the effectiveness of this approach using real stock data from CIMB, TM, and YTLPOWER between December 2024 and February 2025. The performance was evaluated using RMSE, MAE, MSE, and precision, with CIMB and TM showing particularly high accuracy. The results clearly indicate that the combination of pattern-based clustering and data-driven feature selection leads to superior prediction performance compared to models using feature selection alone. Moreover, the entire pipeline was implemented using a big data processing environment based on R, allowing efficient handling of large-scale intraday datasets. This scalable architecture supports future expansion to more stocks and longer time periods.

RECOMMENDATION AND FUTURE WORKS

Based on the findings of this study, it is recommended that financial analysts and data scientists adopt a pattern-based predictive framework when dealing with high-frequency stock market data. The integration of hierarchical clustering and Lasso-based feature selection significantly improves model interpretability and forecasting accuracy by focusing on historically similar market behaviors and eliminating irrelevant variables. Organizations engaged in short-term trading or algorithmic investment strategies can benefit from implementing this hybrid approach using scalable data platforms such as Hadoop with R. Furthermore, it is advisable to prioritize model simplicity and generalization over complexity, as results show that selected features like opening and high prices often provide sufficient predictive power. Analysts should also regularly update the model with the latest patterns and re-evaluate feature relevance to ensure consistent performance in dynamic market environments. Lastly, regulatory compliance and ethical considerations in automated forecasting systems should be embedded in future implementations of such data-driven decision frameworks.

Future research should focus on expanding the proposed methodology to a broader set of stocks and longer time horizons. Incorporating multi-scale pattern detection could improve performance in volatile markets by identifying both short-term and long-term trends. Additionally, integrating external data sources, such as financial news, sentiment analysis, and macroeconomic indicators, may enhance the model's predictive capacity. Technically, the system could be strengthened by implementing a distributed parallel architecture to allow real-time predictions across the entire stock exchange. Finally, the use of advanced deep learning architectures like LSTM or Transformer models could further capture temporal dependencies in financial time series.

REFERENCES

1. Ampomah, E. K., Nyame, G., Qin, Z., Addo, P. C., Gyamfi, E. O., & Gyan, M. (2021). Stock market prediction with Gaussian naïve Bayes machine learning algorithm. *Informatica*, 45(2), 265-278
2. Awan, M. J., Rahim, M. S., Nobanee, H., Munawar, A., Yasin, A., & Zain Azlanmz, A. M. (2021). Social media and stock market prediction: A big data approach. *Computers, Materials & Continua*, 67(2).
3. Bansal, M., Goyal, A., & Choudhary, A. (2022). Stock market prediction with high accuracy using machine learning techniques. *Procedia Computer Science*, 215, 247–265.
4. Bhat, M. A., & Ahmad, T. (2021). Real-time stock prediction using LSTM and online learning: A dynamic approach. *International Journal of Advanced Computer Science and Applications*, 12(3), 89–95.
5. Chen, J. (2023). Analysis of bitcoin price prediction using machine learning. *Journal of Risk and Financial Management*, 16(1), 51. <https://doi.org/10.3390/jrfm16010051>

6. Dai, W., An, Y., & Long, W. (2022). Price change prediction of ultra high frequency financial data based on temporal convolutional network. *Procedia Computer Science*, 199, 1177–1183.
7. Demirel, U., Cam, H., & Ünlü, R. (2021). Predicting stock prices using machine learning methods and deep learning algorithms: The sample of the Istanbul Stock Exchange. *Gazi University Journal of Science*, 34(1), 63–82.
8. Dunstall, S., & Wirth, A. (2005). Heuristic methods for the identical parallel machine flowtime problem with set-up times. *Computers & Operations Research*, 32(9), 2479–2491.
9. Emioma, C. C., & Edeki, S. O. (2021). Stock price prediction using machine learning on least-squares linear regression basis. In *Journal of Physics: Conference Series* (Vol. 1734, No. 1, p. 012058). IOP Publishing.
10. Ghosh, A., Bose, S., Maji, G., Debnath, N., & Sen, S. (2019, September). Stock price prediction using LSTM on Indian share market. In *Proceedings of 32nd International Conference on* (Vol. 63, pp. 101–110).
11. Han, C., & Fu, X. (2023). Challenge and opportunity: Deep learning-based stock price prediction by using bi-directional LSTM model. *Frontiers in Business, Economics and Management*, 8(2), 51–54.
12. Jin, M., Wen, Q., Liang, Y., Zhang, C., Xue, S., Wang, X., ... & Xiong, H. (2023). Large models for time series and spatio-temporal data: A survey and outlook. *arXiv preprint arXiv:2310.10196*.
13. Jang, J., & Lee, K. (2020). Hybrid deep learning framework for stock market forecasting using technical indicators. *Applied Sciences*, 10(18), 6144. <https://doi.org/10.3390/app10186144>
14. Jeon, S., Hong, B., Lee, H. J., & Kim, J. (2016, April). Stock price prediction based on stock big data and pattern graph analysis. In *International conference on internet of things and big data* (Vol. 2, pp. 223–231). SCITEPRESS.
15. Kim, T., & Kim, H. Y. (2019). Forecasting stock prices with a feature fusion LSTM-CNN model using different representations of the same data. *PLOS ONE*, 14(2), e0212320. <https://doi.org/10.1371/journal.pone.0212320>
16. Kim, K. J., & Han, I. (2000). Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index. *Expert systems with Applications*, 19(2), 125–132.
17. Kumbure, M. M., Lohrmann, C., Luukka, P., & Porras, J. (2022). Machine learning techniques and data for stock market forecasting: A literature review. *Expert Systems with Applications*, 197, 116659.
18. Kumar, A., Alsadoon, A., Prasad, P. W. C., Abdullah, S., Rashid, T. A., Pham, D. T. H., & Nguyen, T. Q. V. (2022). Generative adversarial network (GAN) and enhanced root mean square error (ERMSE): Deep learning for stock price movement prediction. *Multimedia Tools and Applications*, 1–19, 81, 19337–19355. <https://doi.org/10.1007/s11042-021-11750-2>
19. Li, M., Zhu, Y., Shen, Y., & Angelova, M. (2023). Clustering-enhanced stock price prediction using deep learning. *World Wide Web*, 26(1), 207–232. <https://doi.org/10.1007/s11280-022-01074-w>
20. Lin, C. T., Wang, Y. K., Huang, P. L., Shi, Y., & Chang, Y. C. (2022). Spatial-temporal attention-based convolutional network with text and numerical information for stock price prediction. *Neural Computing and Applications*, 34(17), 14387–14395.
21. Li, J., Zhu, G., Hua, C., Feng, M., Bennamoun, B., Li, P., ... & Bennamoun, M. (2023). A systematic collection of medical image datasets for deep learning. *ACM Computing Surveys*, 56(5), 1–51.
22. Mehtab, S., & Sen, J. (2019). A robust predictive model for stock price prediction using deep learning and natural language processing. *arXiv preprint arXiv:1912.07700*.
23. Mehtab, S., & Sen, J. (2022). Analysis and forecasting of financial time series using CNN and LSTM-based deep learning models. In *Advances in Distributed Computing and Machine Learning: Proceedings of ICADCML 2021* (pp. 405–423). Springer Singapore.
24. Mittal, S., & Nagpal, C. K. (2022). Predicting a reliable stock for mid and long term investment. *Journal of King Saud University-Computer and Information Sciences*, 34(10), 8440–8448. <https://doi.org/10.1016/j.jksuci.2022.01.019>
25. Moghar, A., & Hamiche, M. (2020). Stock market prediction using LSTM recurrent neural network. *Procedia Computer Science*, 170, 1168–1173.
26. Moukalled, M. I. (2019). Automated stock price prediction using machine learning (Doctoral dissertation).
27. Muhammad, T., Aftab, A. B., Ibrahim, M., Ahsan, M. M., Muhu, M. M., Khan, S. I., & Alam, M. S. (2023). Transformer-based deep learning model for stock price prediction: A case study on Bangladesh

- stock market. *International Journal of Computational Intelligence and Applications*, 22(03), 2350013. <https://doi.org/10.1142/S1469026823500130>
28. Nikou, M., Mansourfar, G., & Bagherzadeh, J. (2019). Stock price prediction using deep learning algorithm and its comparison with machine learning algorithms. *Intelligent Systems in Accounting, Finance and Management*, 26(4), 164–174. <https://doi.org/10.1002/isaf.1469>
29. Olubusola, O., Mhlongo, N. Z., Daraojimba, D. O., Ajayi-Nifise, A. O., & Falaiye, T. (2024). Machine learning in financial forecasting: A US review: Exploring the advancements, challenges, and implications of AI-driven predictions in financial markets. *World Journal of Advanced Research and Reviews*, 21(2), 1969-1984.
30. Pahwa, N., Khalfay, N., Soni, V., & Vora, D. (2017). Stock prediction using machine learning: A review paper. *International Journal of Computer Applications*, 163(5), 36–43.
31. Pawaskar, S. (2022). Stock price prediction using machine learning algorithms. *International Journal of Research in Applied Science and Engineering Technology*, 10(1), 667–673.
32. Rouf, N., Malik, M. B., Arif, T., Sharma, S., Singh, S., Aich, S., & Kim, H. C. (2021). Stock market prediction using machine learning techniques: A decade survey on methodologies, recent developments, and future directions. *Electronics*, 10(21), 2717. <https://doi.org/10.3390/electronics10212717>
33. Shahi, T. B., Shrestha, A., Neupane, A., & Guo, W. (2020). Stock price forecasting with deep learning: A comparative study. *Mathematics*, 8(9), 1441. <https://doi.org/10.3390/math8091441>
34. Soni, P., Tewari, Y., & Krishnan, D. (2022). Machine learning approaches in stock price prediction: A systematic review. In *Journal of Physics: Conference Series* (Vol. 2161, No. 1, p. 012065). IOP Publishing. <https://doi.org/10.1088/1742-6596/2161/1/012065>
35. Schumaker, R. P., & Chen, H. (2009). Textual analysis of stock market prediction using breaking financial news: The AZFin text system. *ACM Transactions on Information Systems (TOIS)*, 27(2), 1–19. <https://doi.org/10.1145/1462198.1462204>
36. Vijh, M., Chandola, D., Tikkiwal, V. A., & Kumar, A. (2020). Stock closing price prediction using machine learning techniques. *Procedia Computer Science*, 167, 599–606.
37. Yang, H., Kim, H., & Kim, Y. (2020). Reinforcement learning-based stock trading strategy using risk-adjusted return. *Expert Systems with Applications*, 157, 113539. <https://doi.org/10.1016/j.eswa.2020.113539>
38. Zhang, J., Ye, L., & Lai, Y. (2023). Stock price prediction using CNN-BiLSTM-Attention model. *Mathematics*, 11(9), 1985. <https://doi.org/10.3390/math11091985>
39. Zou, J., Zhao, Q., Jiao, Y., Cao, H., Liu, Y., Yan, Q., ... & Shi, J. Q. (2022). Stock market prediction via deep learning techniques: A survey. *arXiv preprint arXiv:2212.12717*.
40. Zong, Z., & Guan, Y. (2024). AI-driven intelligent data analytics and predictive analysis in Industry 4.0: Transforming knowledge, innovation, and efficiency. *Journal of the Knowledge Economy*, 1-40.