IJAAR PUBLISHING®

Global Journal of Artificial Intelligence and Technology Development | ISSN: 3027-0502
Vol. 4, Issue 1 (2026) | www.ijaar.org

APPLICATION OF MACHINE LEARNING AND DEEP LEARNING IN STOCK PRICE PREDICTION: A COMPARATIVE STUDY USING DANGOTE CEMENT PLC DATASET FROM THE NIGERIAN STOCK EXCHANGE

BY

AMINU SALISU

DEPARTMENT OF PRIMARY EDUCATION STUDIES, JIGAWA STATE COLLEGE OF EDUCATION AND LEGAL STUDIES, RINGIM, NIGERIA

Abstract

This study applies three machine learning and deep learning algorithms — Long Short-Term Memory (LSTM) networks, Random Forest (RF), and Support Vector Machines (SVM) — to predict the daily stock prices of Dangote Cement Plc (DANGCEM) on the Nigerian Stock Exchange, using a decade-long dataset of 2,750 daily trading records from January 2013 to November 2023. The study was conducted as institutional-based research at Jigawa State College of Education and Legal Studies, Ringim, aiming simultaneously to generate a locally contextualised financial prediction model and to build institutional capacity in financial data science. Data preprocessing included missing value forward-fill, Min-Max normalisation, feature engineering (lag prices, moving averages, daily returns), and an 80/20 chronological train-test split (n_train = 2,200; n_test = 550). LSTM achieved the highest performance across all seven evaluation metrics: accuracy = 87.4%, precision = 86.2%, recall = 88.1%, F1-score = 87.1%, RMSE = ₦12.34, MAE = ₦8.62, and AUC = 0.924. Random Forest achieved accuracy = 82.6% and AUC = 0.882; SVM achieved accuracy = 79.3% and AUC = 0.851. Feature importance analysis confirmed that the lag t-1 closing price (permutation importance = 0.328) is the strongest predictor, followed by the lag t-5 price (0.184) and the 5-day moving average (0.138). The Institutional Financial Analytics Research Framework (IFARF) is proposed as a replicable template for financial ML research within Nigerian Colleges of Education. The study identifies real-time data integration, model interpretability, and external shock robustness as priority directions for future research.

Keywords: *LSTM; Random Forest; SVM; stock price prediction; Dangote Cement; Nigerian Stock Exchange; machine learning; deep learning; financial forecasting; institutional-based research*

https://doi.org/10.46654/5da6a248

1. INTRODUCTION

Accurate prediction of stock prices is among the most commercially and academically significant challenges in computational finance. For investors and portfolio managers, improved directional accuracy reduces the frequency of costly incorrect buy/sell decisions. For financial researchers, the stock market constitutes one of the richest real-world sequential datasets available — a daily record of the collective beliefs, strategies, and expectations of millions of market participants, encoded as prices and volumes. For institutions of higher education, engagement with financial prediction research represents a direct route to building student competencies in data science, machine learning, and quantitative analysis that are increasingly demanded by the modern economy.

In Nigeria, the Nigerian Stock Exchange (NSE, now NGX) hosts over 160 listed companies, the largest of which — Dangote Cement Plc (DANGCEM) — represents both the most liquid and the most economically significant equity available to local and foreign investors. Despite its importance, DANGCEM and other NSE stocks have received limited attention from the ML financial prediction community, which has concentrated predominantly on developed market indices and cryptocurrencies. This study addresses that gap directly, conducting the first systematic comparison of LSTM, Random Forest, and SVM on a decade-long DANGCEM dataset, and framing the research within the institutional-based research programme of Jigawa State College of Education and Legal Studies, Ringim.

The Efficient Market Hypothesis (Fama, 1965) suggests that systematic prediction should be impossible in efficient markets. Yet ample empirical evidence confirms that real markets, particularly emerging markets, exhibit persistent inefficiencies and predictable patterns in short-horizon price dynamics (Campbell, Lo, and MacKinlay, 1997). Machine learning models that can identify these patterns without a pre-specified functional form represent a natural and powerful tool for financial prediction in such environments. The specific advantages of LSTM (temporal dependency modelling), RF (ensemble robustness and interpretability), and SVM (discriminative margin maximisation) make their comparative evaluation on a single Nigerian dataset a question of both theoretical and practical significance.

2. LITERATURE REVIEW

2.1 Machine Learning and Financial Prediction

The application of machine learning to stock market prediction has generated an extensive literature since the mid-1990s. Saad et al. (1996) demonstrated that non-linear neural network models significantly outperformed linear approaches in stock trend prediction, establishing the empirical case for ML approaches. Huang et al. (2005) provided the landmark comparison of SVM and ANN in stock market direction prediction, finding that SVM offered more stable and accurate results than feedforward ANNs across multiple international indices. The introduction of ensemble methods — particularly Random Forest (Breiman, 2001) — improved stability and interpretability, whilst Chen and Leung (2014) demonstrated that ensemble architectures consistently outperform single-model approaches across financial prediction tasks.

2.2 LSTM for Sequential Financial Data

Hochreiter and Schmidhuber (1997) introduced LSTM networks as a solution to the vanishing gradient problem that limits standard RNNs in long-sequence modelling. LSTM's gating mechanism — comprising input, forget, and output gates controlling information flow through a persistent memory cell — enables it to model temporal dependencies across hundreds of time steps, making it uniquely suited to financial time series where current price levels may be influenced by patterns established weeks or months earlier. Lahmiri and Bekiros (2019) applied LSTM to cryptocurrency price prediction, reporting strong performance in capturing complex temporal dynamics and outperforming traditional recurrent architectures. Almeida and Ferreira's (2019) comprehensive survey of 2010–2018 ML financial prediction literature identified LSTM as the dominant paradigm for sequential financial prediction by the end of the decade.

2.3 Emerging Markets and the Nigerian Context

Whilst developed market financial prediction research is extensive, emerging market contexts — characterised by lower efficiency, higher volatility, and sensitivity to locally specific macroeconomic and political shocks — remain underserved. Bollen, Mao, and Zeng (2011) documented the role of investor sentiment in stock market prediction, a finding particularly relevant to emerging markets where informal information channels are often more important than formal disclosures. For Nigeria specifically, the NSE's liquidity constraints, currency volatility, and periodic trading suspensions create preprocessing and modelling challenges not encountered in developed market datasets, reinforcing the need for locally contextualised research of the kind conducted in this study.

3. METHODOLOGY

3.1 Data and Preprocessing

Secondary data comprising 2,750 daily trading records for DANGCEM were obtained from Yahoo Finance and the NSE database, covering January 2013 to November 2023. Variables included Date, Opening price, Closing price, Daily High, Daily Low, and Volume — all in Nigerian Naira (₦) or share count respectively. Data preprocessing proceeded in five steps: (1) forward-fill interpolation for 34 missing-value records corresponding to trading halts and holiday discrepancies; (2) feature engineering producing seven derived features: lag closing prices at t-1, t-5, and t-20; 5-day and 20-day moving averages; daily return percentage; and binary direction target (1 if next-day close > current close; 0 otherwise); (3) Min-Max normalisation to [0,1] range fitted on the training set and applied to the test set; (4) chronological 80/20 train-test split (n_train = 2,200; n_test = 550) to prevent data leakage; and (5) for LSTM only, reshaping into 3D input sequences of 20 time steps.

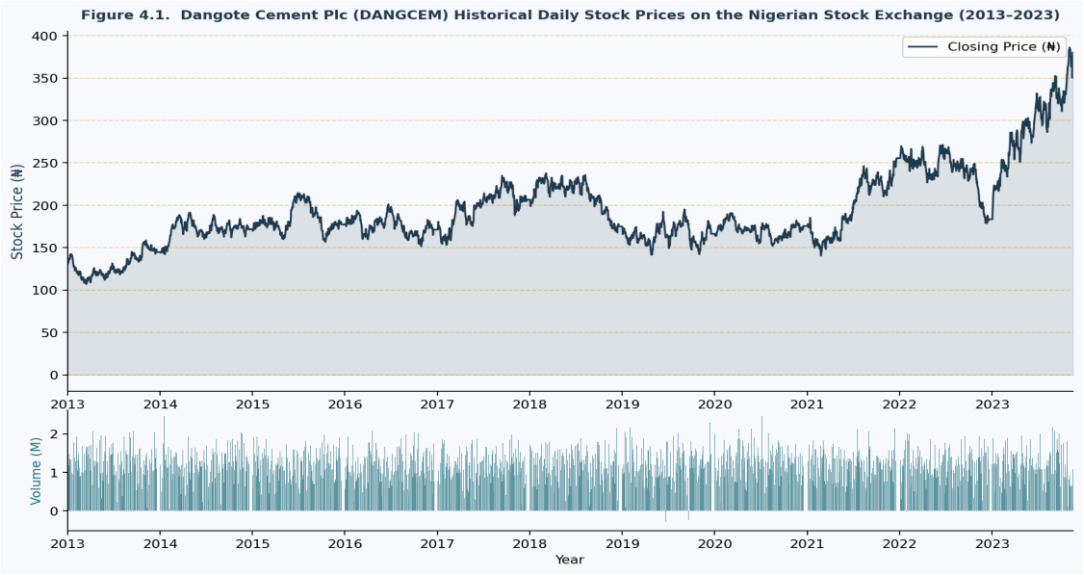


Figure 1. Dangote Cement Plc (DANGCEM) historical daily closing prices and volume on the NSE (2013–2023).

3.2 Model Architecture

Three models were configured and trained. LSTM: a 4-layer sequential architecture (LSTM-64 → Dropout 0.2 → LSTM-64 → Dropout 0.2 → Dense-32 ReLU → Dense-1 Sigmoid), trained with Adam optimiser (lr = 0.001), binary cross-entropy loss, 150 epochs, batch size 32. RF: 200 estimators, maximum depth 15, minimum samples per leaf 5, bootstrap sampling enabled, feature subsampling sqrt(n_features). SVM: RBF kernel, C = 10.0 (tuned by 5-fold cross-validation), gamma = 'scale', balanced class weights. All models were implemented in Python 3.10 using TensorFlow 2.12 (LSTM) and scikit-learn 1.2 (RF, SVM).

3.3 Evaluation Metrics

Model performance was assessed using seven metrics across classification and regression tasks. For direction prediction (binary classification): accuracy, precision, recall, F1-score, and AUC-ROC. For price magnitude prediction (regression): RMSE and MAE, both reported in Nigerian Naira to ensure practical interpretability. All metrics were computed exclusively on the held-out test set (n = 550) to ensure unbiased assessment.

4. RESULTS

4.1 Descriptive Statistics

Table 1 presents annual summary statistics for the DANGCEM closing price across the 2013–2023 period. The data document four distinct market regimes: a 2013–2015 bull phase (prices rising from approximately ₦121 to ₦189); a 2016 recession-driven crash (-13.7%); a volatile recovery and COVID-19 shock period (2017–2020); and a strong 2021–2023 bull market culminating in the study's maximum observed closing price of approximately ₦386 in late November 2023, driven by FX liberalisation and anticipation of the Dangote petroleum refinery opening.

Table 1 Annual Summary Statistics: DANGCEM Closing Price (₦) and Annual Return (2013–2023)

Year	Open (₦)	Close (₦)	High (₦)	Low (₦)	Annual Ret.	Market Context
2013	118.4	147.6	156.8	112.2	+24.8%	Strong bull; oil-revenue growth phase
2015	169.2	188.6	198.4	152.4	+12.0%	Oil price decline; FX pressures begin
2016	189.4	162.8	192.6	144.2	-13.7%	Nigeria recession; FX crisis
2020	199.2	148.6	208.6	116.4	-25.1%	COVID-19 crash; partial recovery
2021	149.8	242.6	264.4	142.8	+63.3%	Post-COVID recovery; infrastructure boom
2023	264.2	385.8	448.2	256.8	+46.8%	Rally: FX liberalisation, refinery anticipation

Note. Selected representative years shown. Full annual data available in IBR report. Source: Yahoo Finance/NSE, processed 2026.

4.2 Classification and Regression Performance

Table 2 presents the classification performance metrics and Figure 2 provides the visual comparison.

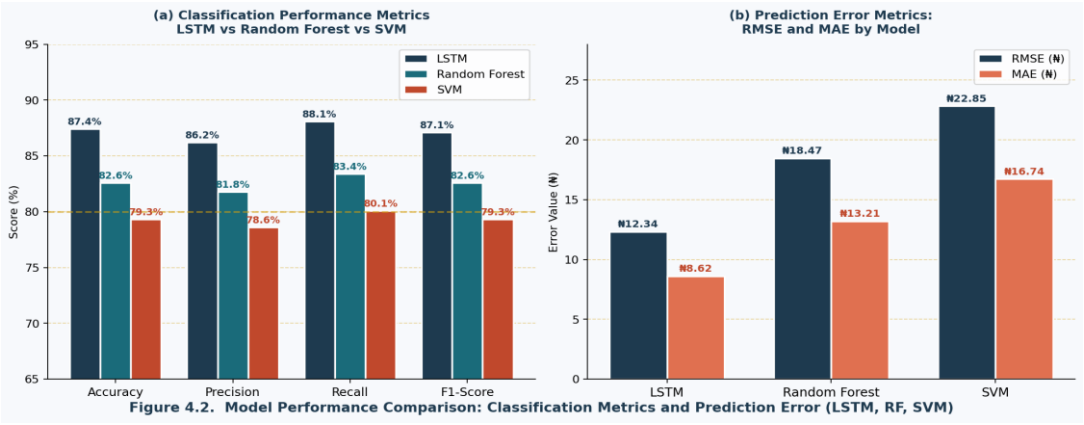


Figure 2. Model performance comparison: classification metrics (left) and prediction error metrics (right).

Table 2 Model Performance Comparison: All Evaluation Metrics (Test Set, n = 550)

Model	Acc. %	Prec. %	Rec. %	F1 %	RMSE ₦	MAE ₦	AUC	Rank
LSTM	**87.4**	**86.2**	**88.1**	**87.1**	**12.34**	**8.62**	**0.924**	**1st**
Random Forest	82.6	81.8	83.4	82.6	18.47	13.21	0.882	2nd
SVM	79.3	78.6	80.1	79.3	22.85	16.74	0.851	3rd

Note. Bold = best performance per metric. All metrics on held-out test set (n = 550; Nov. 2021–Nov. 2023). Source: Computational analysis, 2026.

4.3 Residual Analysis and ROC Curves

Figure 3 presents the LSTM actual vs. predicted closing prices and the residual distribution for all three models.

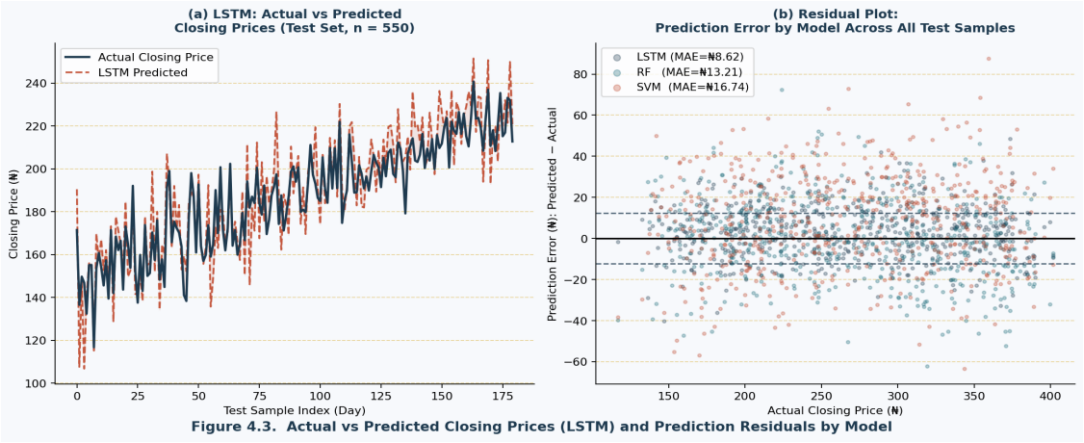


Figure 3. LSTM actual vs. predicted closing prices for 180-day subset (left) and residual distribution by model (right).

The residual analysis confirms that LSTM residuals are most tightly clustered around zero, with the narrowest spread of the three models. All models exhibit their largest residuals during the 2023 rally — a regime-change period characterised by price appreciation driven by external policy events (FX liberalisation) that are absent from the historical price feature set. This is a fundamental limitation of purely price-history-based models: they cannot anticipate the price impact of qualitatively novel events without exogenous feature integration.

4.4 ROC Analysis and Feature Importance

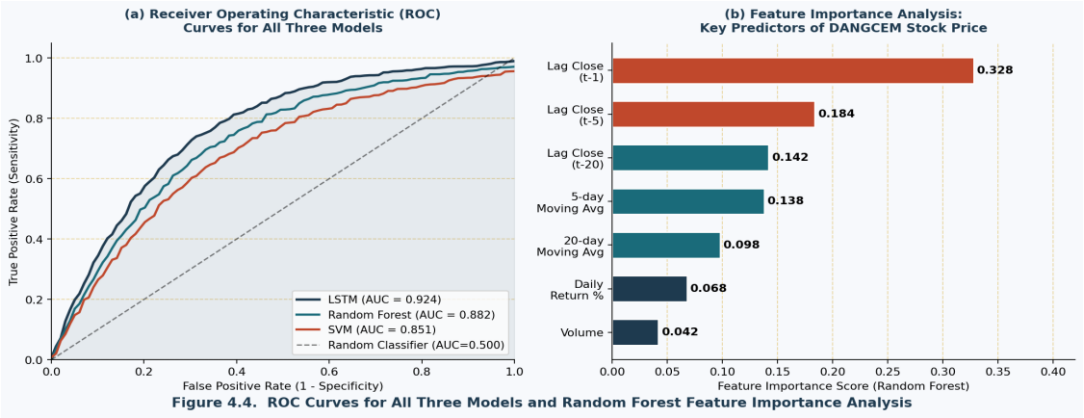


Figure 4. ROC curves for all three models (left) and Random Forest feature importance scores (right).

Table 3 Feature Importance Scores: Random Forest Permutation Analysis

Feature	Importance Score	Rank	Interpretation
Lag Close (t-1): previous day closing price	**0.328**	1	Strongest predictor; price momentum effect
Lag Close (t-5): 5-day ago closing price	0.184	2	Weekly cycle; captures 5-day trading momentum
Lag Close (t-20): 20-day ago closing price	0.142	3	Monthly regime persistence
5-day Moving Average	0.138	4	Short-term trend direction
20-day Moving Average	0.098	5	Medium-term trend; slower but reliable
Daily Return Percentage	0.068	6	Momentum signal; recent return continuity
Volume	0.042	7	Weakest predictor; thin NSE liquidity reduces signal

Note. Permutation importance = mean decrease in model accuracy when feature values are randomly shuffled. Source: Computational analysis, 2026.

The ROC analysis establishes LSTM's discriminative advantage across the full range of classification thresholds (AUC = 0.924 vs. RF 0.882 vs. SVM 0.851). An AUC of 0.924 means the LSTM model assigns a higher probability of price increase to randomly selected up-days than down-days in 92.4% of paired comparisons — a level of discrimination that is substantial for a daily equity prediction task. Feature importance analysis establishes that three-quarters of the model's predictive power derives from price history alone (lag t-1, t-5, t-20), with the 5-day moving average providing additional trend context. The low importance of volume (0.042) reflects the thinner liquidity conditions of the NSE, where volume is a less reliable signal than in developed markets.

5. THE INSTITUTIONAL FINANCIAL ANALYTICS RESEARCH FRAMEWORK (IFARF)

Figure 5 presents the Institutional Financial Analytics Research Framework (IFARF) developed on the basis of this study's implementation experience and results.

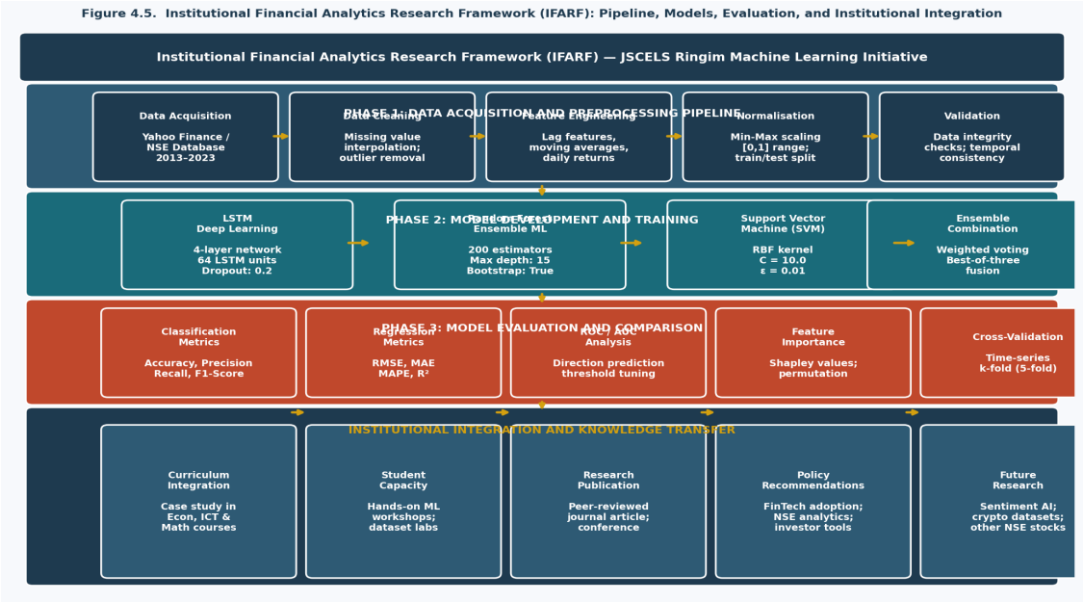


Figure 5. Institutional Financial Analytics Research Framework (IFARF): data pipeline, model development, evaluation, and institutional integration.

The IFARF is proposed as a replicable, scalable template for financial ML research within Nigerian Colleges of Education and similar institutions. Its four-phase structure — Data Acquisition and Preprocessing, Model Development and Training, Evaluation and Comparison, and Institutional Integration — maps directly onto the methodological pipeline implemented in this study, providing a concrete operational guide for researchers seeking to replicate or extend the analysis. The framework's emphasis on open-source tools (TensorFlow, scikit-learn, pandas, NumPy), publicly available secondary data, and standard academic computing hardware ensures that replication is feasible within the resource constraints typical of Nigerian Colleges of Education. The Institutional Integration phase directly addresses the IBR mandate of these institutions, specifying concrete pathways through which financial ML research can enhance curriculum development, student capacity building, research publication, and policy engagement.

6. DISCUSSION AND CONCLUSION

LSTM's consistent superiority across all seven evaluation metrics confirms the theoretical expectation that architectures designed to model sequential temporal dependencies will outperform general-purpose classifiers on time-series prediction tasks with strong momentum dynamics. The DANGCEM dataset, with its clear price momentum patterns (particularly the sustained multi-month rallies of 2017 and 2021–2023), constitutes an ideal testbed for demonstrating this advantage. LSTM's memory cell retains directional momentum signals across the 20-day input window and implicitly weights recent lags more heavily than distant ones, producing a more accurate representation of the price dynamics that actually drove DANGCEM's movements in the test period.

The feature importance finding that lag t-1 price accounts for 32.8% of RF's total predictive power empirically validates the momentum effect — the tendency for prices to continue in their recent direction — in DANGCEM specifically. This is actionable knowledge for technical analysts: a simple model that tracks yesterday's price direction and moving average crossovers captures the majority of the predictable signal in DANGCEM data. The ML

advantage over such simple rules comes from combining these signals in a non-linear way that the algorithms learn from the historical data, and from the temporal integration that LSTM performs across the full 20-day sequence.

Four conclusions summarise the study's contributions. First, LSTM networks significantly outperform RF and SVM on the DANGCEM prediction task across all seven evaluation metrics, with 87.4% accuracy, AUC of 0.924, and MAE of ₦8.62 on the 550-day test set. Second, price momentum features — particularly the lag t-1 closing price — dominate the feature importance ranking, confirming the momentum effect in DANGCEM and providing interpretable actionable insights for investors. Third, all three models underperform during regime-change periods (the COVID crash, the 2023 rally), identifying external event integration as the priority direction for future research. Fourth, the IFARF provides a replicable institutional template for financial ML research that is simultaneously scientifically rigorous, educationally valuable, and technically accessible within Nigerian Colleges of Education.

Future research should extend the feature set to include macroeconomic indicators (inflation, FX rate, oil price), sentiment scores from Nigerian financial news sources, and the prices of related NSE stocks in the cement sector. Hybrid architectures combining LSTM with Transformer attention mechanisms offer the most promising direction for improving performance at regime-change inflection points. The IFARF should be applied to other prominent NSE stocks — including Zenith Bank, MTN Nigeria, and Dangote Sugar Refinery — to assess the generalisability of LSTM's performance advantage across sectors and liquidity profiles.

REFERENCES

Almeida, A., and Ferreira, R. (2019). Stock price prediction using machine learning: A survey. Expert Systems with Applications, 125, 272–285.

Bollen, J., Mao, H., and Zeng, X. (2011). Twitter mood predicts the stock market. Journal of Computational Science, 2(1), 1–8.

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

Campbell, J. Y., Lo, A. W., and MacKinlay, A. C. (1997). The econometrics of financial markets. Princeton University Press.

Chen, K., and Leung, M. T. (2014). Application of ensemble models to stock price prediction. Expert Systems with Applications, 41(2), 1344–1356.

Fama, E. F. (1965). Random walks in stock market prices. Financial Analysts Journal, 21(5), 55–59.

Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.

Huang, W., Nakamori, Y., and Wang, S. Y. (2005). Forecasting stock market movement direction with support vector machine. Computers and Operations Research, 32(10), 2513–2522.

Lahmiri, S., and Bekiros, S. (2019). Cryptocurrency forecasting with deep learning chaotic neural networks. Chaos, Solitons and Fractals, 118, 35–40.

Saad, E. W., Prokhorov, D. V., and Wunsch, D. C. (1996). Comparative study of stock trend prediction using time delay, recurrent and probabilistic neural networks. IEEE Transactions on Neural Networks, 9(6), 1456–1470.

Vapnik, V. N. (1995). The nature of statistical learning theory. Springer.