**Introduction**

Stock market prediction remains a complex and dynamic challenge at the intersection of finance, economics, and data science. With market movements driven by a mix of structured numerical data and unstructured factors like investor sentiment and macroeconomic trends, traditional forecasting models often fall short of capturing these multidimensional influences. Advances in statistical modeling, machine learning (ML) and deep learning (DL) have introduced promising tools for identifying hidden patterns and dependencies in historical stock data, offering a more sophisticated approach to stock trend forecasting.

This research presents a two-phase system for stock market prediction. The first phase employs regression and time-series analysis to uncover trends during historically stable periods. The second phase incorporates broader context—such as geopolitical events, financial news, and social sentiment—through hybrid ML and DL models. The ultimate goal is to improve predictive accuracy while enhancing model interpretability and practical relevance in real-time decision-making environments.

**Literature Review**

Recent research highlights the value of hybrid approaches that combine traditional statistical models with advanced deep learning techniques for stock prediction. (Liu, Zhang, Gao, Xu, & Wang, 2025)introduced a Vector Error Correction Model (VECM) combined with a Bidirectional Long Short-Term Memory (BiLSTM) model to forecast China's energy stock price index. Their hybrid framework effectively integrates spillover effects from global economic indicators with deep learning’s sequential modeling capabilities. Results showed improved performance over classical models like ARIMA and Support Vector Regression (SVR), particularly in terms of Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE).

Lin et al. (2022) presented a Spatial-Temporal Attention-based Convolutional Network (STACN) that combines numerical and textual data for stock price prediction. Their architecture uses Convolutional Neural Networks (CNNs) to extract local features, Long Short-Term Memory (LSTM) networks to handle temporal dependencies, and attention mechanisms to highlight critical input elements. The model demonstrated significant improvements in predictive accuracy over CNN-only and LSTM-only baselines, especially when incorporating financial news and sentiment.

In line with these advancements, the present study uses a phased approach to model stock behavior. The first phase involves logistic and linear regression to identify trends in calm market periods, while the second phase explores DL models and external information integration, as outlined in the project design document. This structure enables both performance benchmarking and contextual model refinement.

**Methodology**

This study follows a two-phase methodology designed to isolate the predictive value of both structured market data and external contextual factors.

**Phase 1: Stable Market Trend Detection**

* **Data Source:** Historical stock data from a stable market period (free of major economic disruptions).
* **Techniques:** Linear regression and logistic regression.
* **Goal:** Identify baseline stock movement trends using open, close, high, and low prices.
* **Evaluation Metrics:** Accuracy, R² score, and confusion matrix for classification tasks.

**Phase 2: Context-Enriched Prediction**

* **Data Source:** Time series stock data combined with sentiment analysis of financial news and social media, macroeconomic indicators (e.g., interest rates, inflation), and significant global events.
* **Techniques:** Hybrid deep learning models including BiLSTM, CNN, and attention-based architectures.
* **Tools:** Python, TensorFlow/Keras, Scikit-learn, and financial APIs (e.g., Alpha Vantage, Twitter API).
* **Goal:** Enhance model robustness by incorporating external drivers of volatility.
* **Evaluation Metrics:** RMSE, MAPE, Precision, Recall, and F1-Score.

The dual-phase design allows for controlled evaluation: first measuring predictive accuracy in clean environments, then quantifying performance improvements when complexity and real-world variables are introduced.

**References (APA Style)**

Liu, X., Chen, C., & Yang, Z. (2025). *China’s energy stock price index prediction based on VECM–BiLSTM model*. Journal of Computational and Cognitive Engineering. https://doi.org/10.1007/s11831-025-00123

Lin, Y., Zhang, Z., & Zhang, W. (2022). Spatial-temporal attention-based convolutional network with text and numerical information for stock price prediction. *Applied Intelligence*, 52, 8570–8587. https://doi.org/10.1007/s10489-021-02653-z