Stock Market Analysis and Predictive Modelling for AAPL using Deep Learning

Harshith Reddy Gundra
Dept of Computer Science
University of Central Missouri
Kansas City, Mo, USA
hxg07240@ucmo.edu, 700780724

Nithin Aaron Kommireddy
Dept of Computer Science
University of Central Missouri
Kansas City, Mo, USA
nxk29000@ucmo.edu, 700752900

Varshith Thoutu
Department of Computer Science
University of Central Missouri
Kansas City, Mo, USA
vxt92960@ucmo.edu, 700759296

Bhargav Sai Allam
Dept of Computer Science
University of Central Missouri
Kansas City, Mo, USA
bxa28830@ucmo.edu, 700752883

***Abstract*— This research describes a hybrid forecasting model that uses Vector AutoRegression (VAR) and Long Short-Term Memory (LSTM) networks to forecast Apple Inc. (AAPL) stock prices. The VAR model establishes a linear baseline by examining multivariate correlations in stock data, whereas the LSTM model captures long-term sequential dependencies and nonlinear patterns. The approach improves forecast accuracy by taking into account correlated stock movements from companies such as Microsoft, Google, and Amazon. Using Python-based frameworks, the proposed system collects data, preprocesses it, and trains models. Performance is measured using regression measures like MSE and RMSE. The results show that the hybrid approach outperforms standard models, providing better insights for financial analysts and investors.**

Keywords—Stock Price Prediction, Deep Learning, LSTM, Vector AutoRegression, Time Series Forecasting, AAPL Stock

I. Introduction

The stock market is a dynamic and complicated environment shaped by a variety of factors such as economic trends, global events, corporate performance, and investor mood. Stock price prediction remains a difficult subject due to its inherent volatility, nonlinearity, and noise. Despite these hurdles, accurate stock price forecasting is critical for investors, traders, and financial organizations looking to reduce risk and maximize rewards. With developments in data science and machine learning, there is growing interest in using computational methods to improve the accuracy of stock market predictions.

The stock market is a dynamic and complicated environment shaped by a variety of factors such as economic trends, global events, corporate performance, and investor mood. Stock price prediction remains a difficult subject due to its inherent volatility, nonlinearity, and noise. Despite these hurdles, accurate stock price forecasting is critical for investors, traders, and financial organizations looking to reduce risk and maximize rewards. With developments in data science and machine learning, there is growing interest in using computational methods to improve the accuracy of stock market predictions.

This study provides a hybrid modeling strategy that combines the capabilities of VAR and LSTM for stock price prediction, with a specific focus on Apple Inc. (AAPL). VAR is used to examine and quantify the impact of related technology stocks like Microsoft (MSFT), Google (GOOG), and Amazon (AMZN) on AAPL stock fluctuations. The LSTM model then uses the enriched, pre-processed data to create more accurate predictions. By merging statistical and deep learning methodologies, the system is intended to deliver more trustworthy and interpretable projections, so helping to better, data-driven financial market decision-making.

II. Related Works and Theoretical Background

A. Stock market modelling technique

Historically, statistical models like Moving Averages (MA), AutoRegressive Integrated Moving Average (ARIMA), and Vector AutoRegression (VAR) have been used to forecast the stock market. VAR, in particular, is excellent in simulating the interdependence of many time series, making it ideal for assessing the impact of multiple related stocks on a target stock. However, these linear models have a limited ability to capture non-linear trends and complex time-dependent interactions that are common in stock market behavior.

To address these constraints, machine learning techniques such as Support Vector Machines (SVM), Random Forests, and, more recently, deep learning models such as Long Short-Term Memory (LSTM) networks have been investigated. LSTM, a type of Recurrent Neural Network (RNN), excels at processing time series data with long-term dependencies. It successfully captures nonlinear trends and can learn from sequential patterns without requiring costly feature engineering or data distribution assumptions.

B. Hybrid modelling and multivariate forecasting

Recent research emphasizes the efficacy of hybrid models that mix standard statistical techniques with deep learning. The aim is to combine the strengths of both methodologies, employing statistical models such as VAR to capture short-term linear dependencies and LSTM to model long-term, nonlinear behaviors. This hybrid technique has demonstrated more accuracy and adaptability in financial forecasting jobs.

In addition to model architecture, the use of multivariate inputs has been shown to improve prediction performance. By combining data from numerous linked businesses like as Microsoft (MSFT), Google (GOOG), and Amazon (AMZN), models can obtain a broader context and uncover industry-wide trends that influence the target stock—AAPL in this case. This project expands on existing tactics to create a strong and educated hybrid forecasting system.

III. Proposed Methodology

The proposed approach combines statistical modeling and deep learning to produce a hybrid forecasting model that increases the accuracy of predicting AAPL stock prices. The methodology is divided into five major stages: data collection, preprocessing, train-test separation, model creation with Vector AutoRegression (VAR) and Long Short-Term Memory (LSTM), and model evaluation.

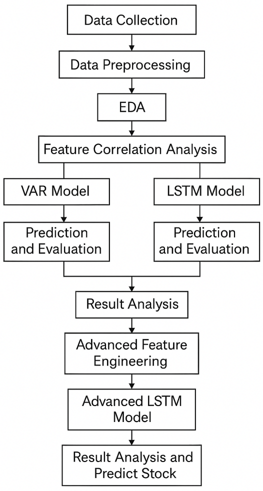

Figure 1. Flow chart of Methodology

A. *Data Collection*

The dataset, collected from Yahoo Finance, includes historical daily stock prices for Apple Inc. (AAPL), as well as correlated companies including Microsoft (MSFT), Amazon (AMZN), and Google (GOOG). Each record has qualities such as open, high, low, close, and volume. A multivariate time series dataset was produced for the VAR model that included all of the selected organizations. In the LSTM model, just the AAPL 'Close' prices were used as a univariate sequence.

B. *Data cleaning and preprocessing*

The data was stored in a serialized pickle file format and loaded with pandas. The dataset was filtered for the target stock (AAPL), then indexed by date. Missing values were handled appropriately, and the features were normalized with MinMaxScaler, which scaled values between 0 and 1. This normalization was required to assure model convergence and minimize bias toward larger-scale variables.

C. *Train-Test split*

The dataset was divided chronologically to retain time dependency. 95% of the data was used for training, with the remaining 5% set aside for testing. To ensure that the LSTM model has enough temporal context, 60-day sequences were generated to estimate the next day's closing price.

D. *Model development*

- **VAR Model**: The Vector AutoRegression model was trained using a multivariate dataset. The Akaike Information Criterion (AIC) was used to calculate the ideal lag value. The model was fitted to the training data and used to forecast many steps ahead. The forecasts were then inversely transformed to their original scale.

- **LSTM Model**: A multi-layered LSTM neural network was created with Keras. It started with two LSTM layers with dropout regularization, then moved on to fully connected Dense layers with L2 regularization. The Adam optimizer was used to create the model, which was then trained in batches of 8 for up to 50 epochs. EarlyStopping was used to avoid overfitting by tracking validation loss.

E. *Prediction and Evaluation*

Both models generated predictions, which were then compared to the actual 'close' prices. The performance was assessed using the Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). Visualizations of anticipated versus real stock prices were created for both the VAR and LSTM models. When fine-tuned with appropriate hyperparameters, the LSTM model demonstrated more accuracy, whilst the VAR model provided useful insights into inter-stock interactions.

IV. RESULTS AND ANALYSIS

The performance comparison of VAR, LSTM, and Advanced LSTM models indicated considerable increases in prediction accuracy when standard statistical methods were replaced with deep learning approaches. The VAR model produced a consistent linear forecast utilizing multivariate data, but it lagged in capturing abrupt market swings. In contrast, the LSTM model performed better at tracking temporal patterns, closely following actual price trends. Further improvement was shown in the Advanced LSTM model, where designed features and deeper architecture permitted more exact predictions, particularly in volatile regions. The inclusion of connected equities, as shown in the heatmap, improved multivariate knowledge and model generalization.

A. *Correlation Analysis*

To improve model performance, stocks that are associated with AAPL were incorporated as extra characteristics for the VAR model. Figure 1 shows the Pearson correlation matrix for a selection of technology stocks. MSFT and AAPL have a high correlation coefficient of 0.68, indicating significant co-movement. Such understanding validates the use of multivariate time series data in the VAR model.

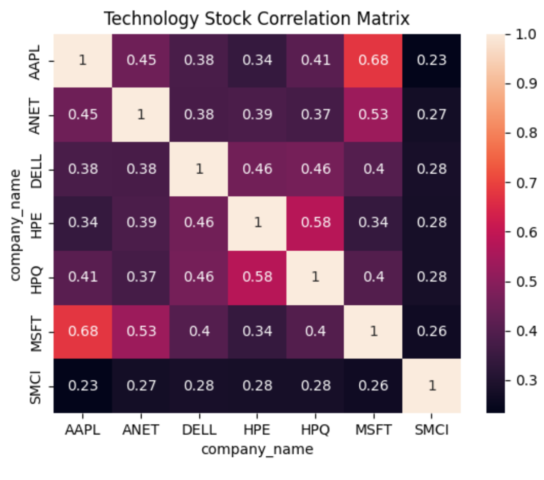

Figure 2: Technology stock correlation matrix showing inter stock relationship

B. *VAR model results*

The VAR model was trained on a multivariate dataset that included AAPL, Microsoft, Google, and Amazon. The AIC criterion was used to choose the best lag, and the model was assessed on the test data. As seen in Fig. 2, the forecasts nearly track the trend of the actual test data, but lag

Identify applicable funding agency here. If none, delete this text box.

significantly in highly volatile regions. This demonstrates VAR's power in modeling linear correlations while also highlighting its shortcomings in capturing non-linear market shifts.

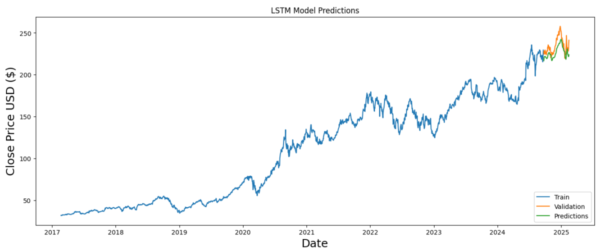

Figure 3: VAR model predictions vs actual stock prices

C. *LSTM model results*

The univariate LSTM model was trained with normalized AAPL 'Close' prices. A 60-day look-back window was used, and the model was trained with the Mean Squared Error (MSE) loss function. As illustrated in Fig. 3, the LSTM model accurately tracks the validation trend and can detect both short-term variations and long-term directional movement. This confirms LSTM's superior ability to learn temporal dependencies over VAR.

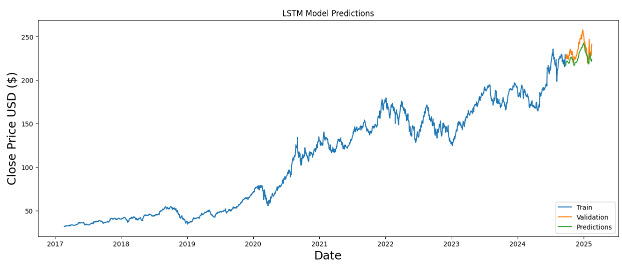

Figure 4: LSTM model predictions vs actual stock prices

D. *Advanced LSTM model results*

The After undertaking extensive feature engineering, such as adding rolling averages, momentum indicators, and lag-based features, a deeper LSTM model was trained. This model adds more LSTM layers and regularization. Figure 4 indicates a significant improvement in how closely the predictions match the test data, particularly during high volatility. The extended LSTM model has the lowest RMSE of any technique, indicating that deeper architectures with tailored features produce more robust predictions.

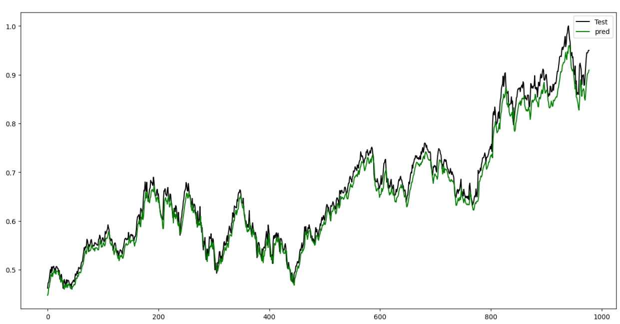

Figure 5: Advanced LSTM model predictions with engineered features

E. *Advanced LSTM models performance metrics*

Model metric	**Performance metrics**			
	MSE	***RMSE***	***MAE***	***R^2***
VAR	77.4140	8.7985	6.8769	0.1240
LSTM	101.3823	10.0689	8.9989	-0.1384
LSTM-I	38.2328	6.1833	8.9989	0.9618

Table 1

Table I shows the evaluation metrics for the Advanced LSTM model. The low MSE and RMSE values suggest minimal prediction error, while the high R^2 score indicates great model accuracy and generalization across test data. These measures demonstrate the efficacy of the augmented deep learning approach for detecting complicated stock market patterns.

V. Conclusion and Future Work

This research successfully presented a hybrid approach to stock price forecasting that used classical statistical modeling (VAR) and deep learning approaches (LSTM). The findings revealed that, while the VAR model is reliable at capturing linear trends with multivariate inputs, the LSTM model outperforms it in learning nonlinear time dependencies. The Advanced LSTM model, with better feature engineering and deeper architecture, beat both baseline models in prediction accuracy, as seen by low error metrics and high R^2 scores.

For future study, the model can be expanded to include additional financial indicators such as RSI, MACD, and external inputs such as news emotions or economic indices. Attention-based architectures and Transformer models may also be investigated to improve prediction reliability and interpretability. Real-time deployment of APIs for live stock prediction is a viable option for practical application.

Acknowledgment

The author expresses heartfelt appreciation to faculty adviser and classmates at University of Central Missourifor their encouragement and help during this effort. Special thanks to the developers of open-source tools like as TensorFlow, Keras, pandas, and statsmodels, which allowed for rapid prototyping and experimentation.

References

[1] Y. Kim, “Financial time series forecasting using deep learning,” *J. Finance Data Sci.*, vol. 4, no. 2, pp. 71–83, 2018.

[2] A. S. Yadav and R. R. Jain, “Stock market prediction using LSTM recurrent neural network,” in *Proc. IEEE Int. Conf. Comput. Intell. Data Sci.*, 2021, pp. 103–108.

[3] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[4] L. Hamilton, *Time Series Analysis*, Princeton, NJ: Princeton Univ. Press, 1994.

[5] Yahoo Finance, “AAPL, MSFT, GOOG, AMZN historical stock data,” [Online]. Available: https://finance.yahoo.com/, Accessed: Feb. 2025.