XAI in Financial Time Series Forecasting

Nguyen Quoc Bao, Nguyen Duc Huynh, Nguyen Thanh Long,
Pham Vu Nhat Minh
FPT University, Ho Chi Minh Campus, Vietnam
{baonqse181700, huynhldse180086, longntse181506, minhpvnse180448}@fpt.edu.vn

Abstract

This project applies Explainable Artificial Intelligence (XAI) techniques to forecast financial time series data, focusing on daily stock prices of Google (GOOGL) from 2020 to 2025. Traditional models (ARIMA, SARIMAX) and deep learning models (LSTM, Transformer) are implemented and evaluated, with SARIMAX and LSTM demonstrating superior accuracy (RMSE of 3.44 and 4.62, respectively). XAI methods, including SHAP, LIME, and Attention Mechanisms, are utilized to enhance model interpretability, revealing the critical role of recent price data (e.g., High, Close, Low) in predictions. An interactive Streamlit application is developed to visualize forecasts and XAI insights, supporting investment decision-making with improved transparency and usability.

Index Terms

XAI, Time Series Forecasting, ARIMA, SARIMA, Deep learning, LSTM, Time Series Transformer, SHAP, LIME, Attention

I. Introduction

Financial time series forecasting plays a crucial role in investment decision-making, yet it remains challenging due to the inherent volatility, seasonality, and complexity of stock price data. This project addresses these challenges by forecasting the daily stock prices of Google (GOOGL) from 2020 to 2025, comparing traditional models (ARIMA, SARIMAX) with deep learning approaches (LSTM, Transformer). To enhance transparency and interpretability, Explainable Artificial Intelligence (XAI) techniques, including SHAP, LIME, and Attention Mechanisms, are employed to uncover the key drivers behind predictions, such as recent price trends and market sentiment. The project also develops an interactive Streamlit application, enabling investors to visualize forecasts and XAI insights, thereby supporting more informed and transparent investment strategies.

II. Exploratory Data Analysis (EDA)

A. Data

This dataset retrieved from kaggle [1] and included the daily historical stock prices for Google (GOOGL) spanning from 2020 to 2025. It features essential financial metrics such as opening and closing prices, daily highs and lows, adjusted close prices, and trading volumes.

- **Date:** Date of the stock data (needs cleaning as the first two rows are headers).
- **Adj Close:** Adjusted closing price, accounting for events like dividends and splits.

- **Close:** Closing price of the stock at the end of the trading day.
- **High:** Highest price of the stock during the trading day.
- **Low:** Lowest price of the stock during the trading day.
- **Open:** Opening price of the stock at the start of the trading day.
- **Volume:** Number of shares traded during the day.

This project will use stock data for the purposes of work follow in the figure 1

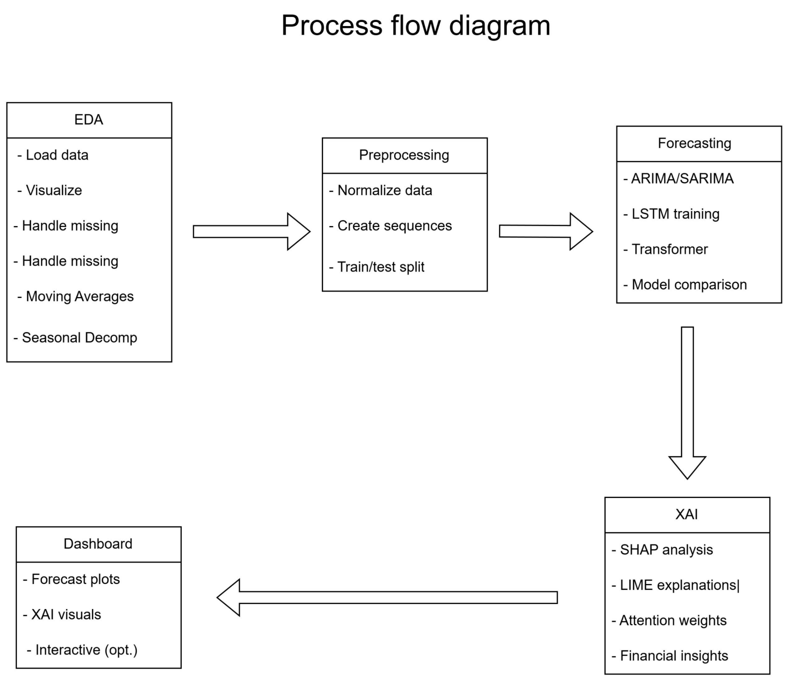


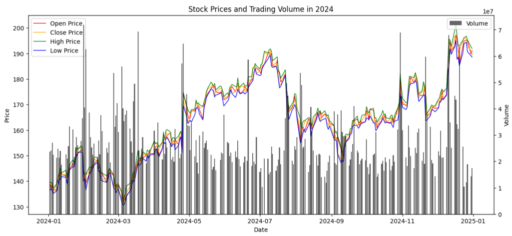
Fig. 1: Overview of the forecasting process

B. Data Visualization

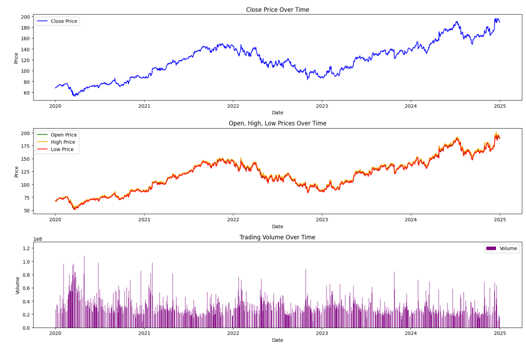
Stock price data (Close, High, Low, Open, Volume) visualized over time from start of 2020 to end of 2024 in Figure 2. Overall, stock prices tend to grow over the long term with fluctuations in daily trading volume. Specifically, in 2024 when prices increase, trading volume tends to decrease and vice versa. Stock price data is not missing but the trading volume has some days of unusual spikes in the box plot and those points are identified as outliers. Moving averages of a stock price show its average price over a certain period of time to analyze the short-term (one week, two weeks, etc.) and long-term (one month, one quarter, etc.) trends of the stock. Volatility measures the degree of fluctuation of the stock price and reflects the risk and opportunity.

C. Seasonality and Trends

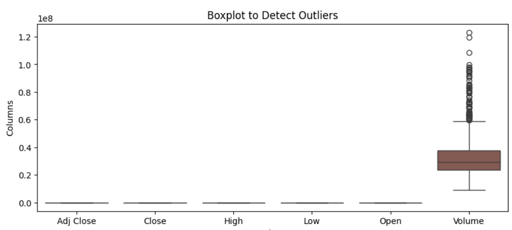
Seasonal decomposition reveals trends and seasonal patterns in the data, as shown in 3.



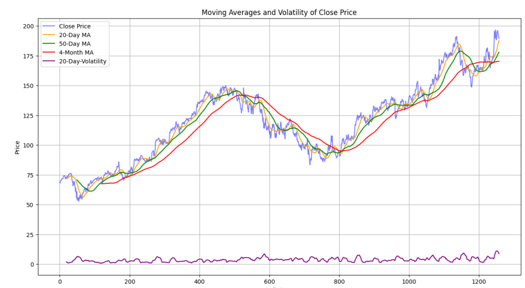
(a) Stock Prices and Trading Volume in 2024



(b) Stock Price Trends



(c) Boxplot to Detect Outliers



(d) Moving Averages and Volatility of Close Price

Fig. 2: Comprehensive Analysis of Stock Price Data

- Observed Data (Blue):
  - The stock price shows an upward trend over time, but there are some periods of decline (notably in 2023).
  - A strong recovery is observed in late 2023, continuing to rise into 2024–2025.
- Long-term Trend (Red):
  - From 2020 to 2022, the stock price increased significantly before experiencing a slight decline in mid-2022.
  - After that, the price begins to recover and grow again from late 2022 to 2025.
- Seasonal Component (Green):
  - There is a repeating cyclic fluctuation, indicating that the stock price follows a clear seasonal pattern (repeating in cycles).
  - The price tends to rise at specific times each year, suggesting potential influences from quarterly financial reports, shopping seasons, or economic factors.
- Residual Component (Blue):

– This represents fluctuations that cannot be explained by trend or seasonality.
– Volatility appears to increase in certain periods, particularly from 2021 onwards, possibly due to other influencing factors.

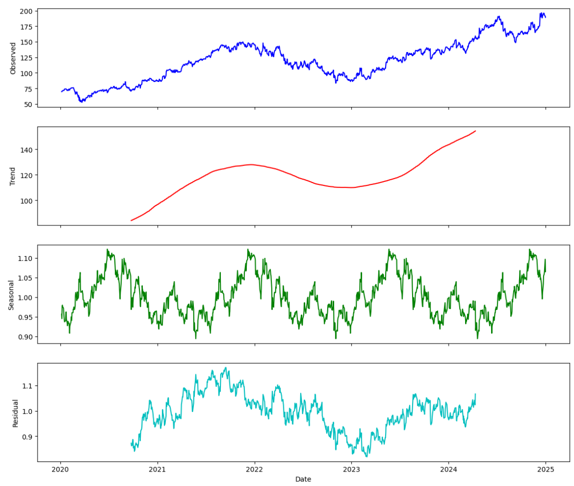


Fig. 3: Seasonal Decomposition of Close Price

D. Correlation Analysis

A correlation heatmap highlights relationships of stock prices in figure 4.

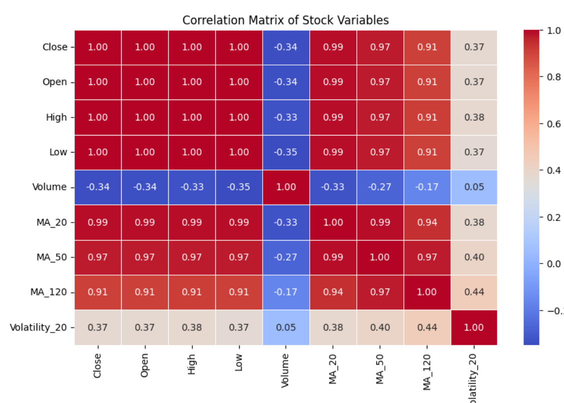


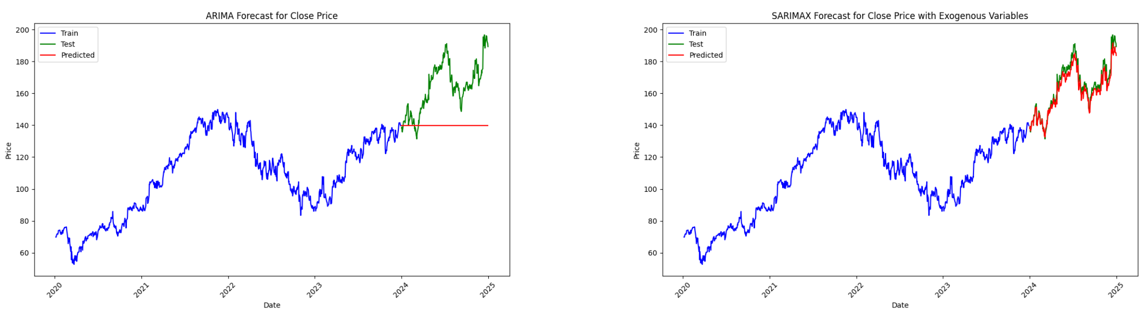
Fig. 4: Correlation Matrix of Stock Variables

- Strong relationship between Open, Close, High, and Low prices have a very high correlation (1.00). This is expected since these price points tend to move together within a trading day.
- Moving Averages strongly correlate with stock prices with MA_20 (0.99), MA_50 (0.97), and MA_120 (0.91).

- Trading Volume has a negative correlation with price with correlation ranges from -0.34 to -0.35.
  - When Volume increases, stock prices tend to decrease slightly.
  - This may indicate selling pressure or strong trading activity during market corrections.
- Volatility (Volatility_20) has a moderate correlation with prices with correlation ranges from 0.37 to 0.38.
  - When volatility increases, stock prices also tend to experience stronger fluctuations.
  - However, the correlation is not too strong, meaning volatility is influenced by multiple factors.

III. Time Series Forecasting

A. *Traditional Methods*



(a) ARIMA Forecast for Close Price

(b) SARIMAX Forecast for Close Price

Fig. 5: Forecast Comparison for Close Price

The input data comprises the closing price (Close) as the target variable and supplementary features (High, Low, Open, Volume). The dataset is divided into a training set (80%) and a test set (20%) for model training and evaluation. For the SARIMAX model, the Close price serves as the endogenous variable, while High, Low, Open, and Volume act as exogenous variables. Optimal parameters (p, d, q) and seasonal parameters (P, D, Q, $m = 12$) are automatically determined using the `auto_arima` function with annual seasonality ($m = 12$), after which the SARIMAX model is trained on the training data. For the ARIMA model, only the Close price is utilized, with parameters (p, d, q) also automatically identified via `auto_arima` using the same seasonal configuration. The ARIMA model is then trained and employed to forecast the Close price on the test set, with performance evaluated using metrics such as RMSE and MAE.

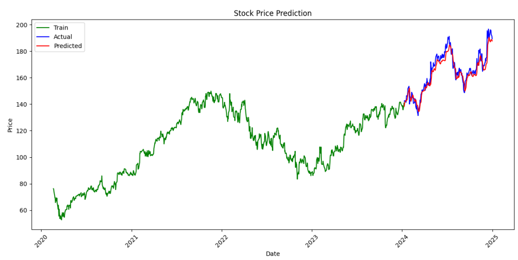
The performance of the two models in figure 5 is as follows:

- ARIMA: RMSE = 28.59, MAE = 24.55.
- SARIMAX: RMSE = 3.44, MAE = 2.93.

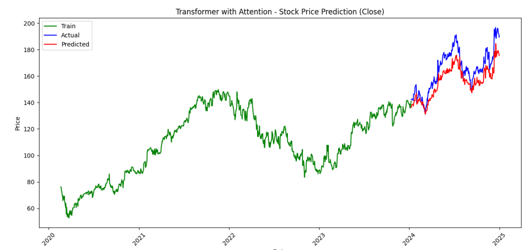
The SARIMAX model outperforms the ARIMA model in terms of accuracy, as indicated by its lower RMSE and MAE values. This suggests that SARIMAX handles trends and seasonality more effectively than ARIMA.

B. Deep Learning Models

The input data consists of stock price features (Close, High, Low, Open) and trading volume (Volume), extracted from a DataFrame and converted into an array of values. To normalize the data, the MinMaxScaler method is applied, scaling the values to the range [0, 1] to ensure consistency across features. Subsequently, the data is segmented into sequences with a window size of 30 days, where each sequence comprises data from 30 consecutive days and the Close price of the following dayas the forecasting target. Finally, the dataset is split into a training set (80%) and a test set (20%) to facilitate model training and evaluation. The performance of two



(a) LSTM model for stock price prediction



(b) Time Series Transformer for stock price prediction

Fig. 6: Deep Learning Models

model in figure 6 is as show::

- LSTM: RMSE = 4.62, MAE = 3.60.
- Transformer: RMSE = 9.90, MAE = 8.82.

With low MAE and RMSE, the LSTM model predicts stock price trends better than the Time Series Transformer model in both performance and capturing the overall trend.

C. Model Comparison

Model	RMSE	MAE
ARIMA	28.59	24.55
LSTM	4.62	3.60
Transformer	9.90	8.82

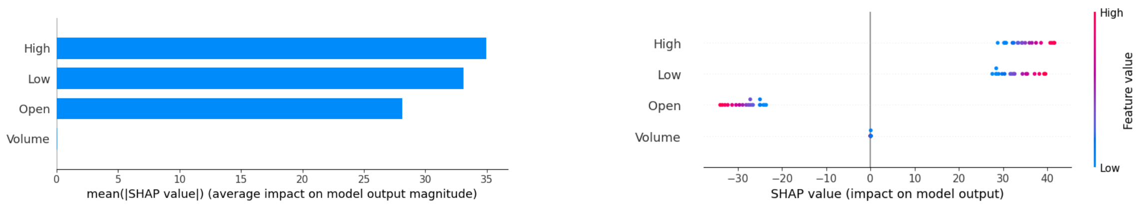
TABLE I: Comparison of Forecasting Models

The predicted results in figure 5,6 and performance in table I show that transformer outperforms ARIMA in accuracy, while LSTM balances accuracy and interpretability.

IV. Explainable Artificial Intelligence (XAI)

A. Feature Importance Analysis (SHAP)

SHAP values identify key factors influencing stock price predictions, as shown in Figure 7. High price is the most important factor affecting stock price with the highest average SHAP value of about 35%. High high price usually increases the prediction of Close price and other prices (low, Open) also have an influence but smaller while Volume has almost no influence.



(a) Bar Plot (b) Scatter Plot

Fig. 7: SHAP Feature Importance Analysis

B. *Local Interpretability (LIME)*

LIME explains individual predictions from the LSTM model. The LSTM model with Attention

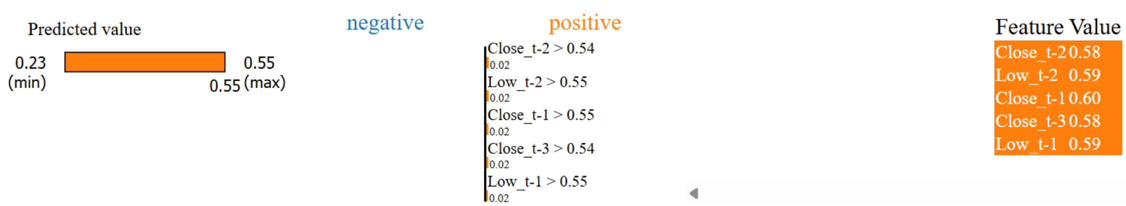


Fig. 8: LIME Explanation for a Specific Prediction

mechanism predicted a normalized `Close` price of 0.23 shown in figure 8 , which is low within the normalized range of [0.23, 0.55]. Using LIME for interpretation, the key features influencing this prediction were identified as follows: Close_t-1 (value: 0.60, weight: 0.02), Close_t-2 (value: 0.58, weight: 0.02), Close_t-3 (value: 0.57, weight: 0.02), Low_t-1 (value: 0.59, weight: 0.02), and Low_t-2 (value: 0.59, weight: 0.02). All of these features, with high values, contributed positively to the prediction and especially they were all prices on the last day of the series, showing that the model focuses heavily on recent data to make predictions. However, the influence level of each feature has a quite small value (0.02), so their total influence is not large enough to push the prediction up.

C. *Attention Mechanism*

Attention weights highlight critical time periods in the LSTM model. The results in figure 9 shows the attention weight for the first 30-day sample corresponding to the timestep from the day before the predicted date t-0 increasing gradually to the furthest days t-29 in the test set of the LSTM model. The results from the graph show that the time step at t-0 has the highest weight of about 0.33 and the time steps near t-0 are t-1, t-2, t-3, t-4, t-5 have high weights from 0.10 to 0.25, indicating that the model focuses on the data of the few days closest to the forecast

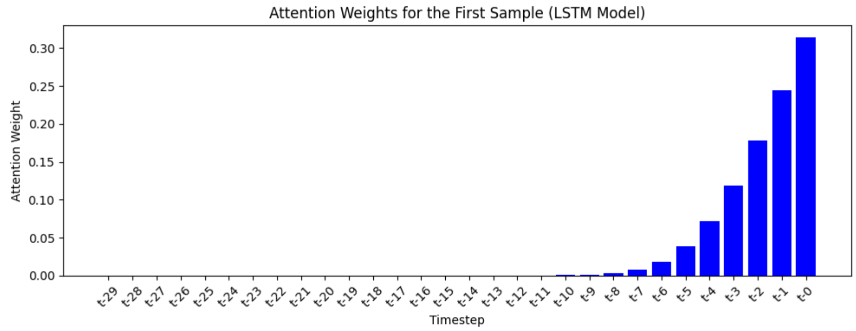


Fig. 9: Attention weight for the first sample of LSSTM model

Original Prediction: 158.81

Counterfactual Prediction: 155.00

Counterfactual Data (last 5 timesteps):

	High	Low	Open	SMA20	SMA50	% Change	RSI	MACD
-5	165.1179	161.1532	162.8914	164.029	185.6786	1.5828	43.2772	-5.4054
-4	167.4289	164.3311	164.7988	163.7799	185.2218	1.0553	44.3494	-4.6738
-3	166.3775	160.8354	164.8183	163.3817	184.6674	-3.887	39.1136	-4.4929
-2	162.15	157.9588	160.4792	163.0502	184.0533	-2.3603	34.4941	-4.5289
-1	158.4962	149.4245	156.4107	162.2341	183.1651	-5.5163	34.2907	-5.1451

Fig. 10: ICFTS

date. The distant time steps from t-29 to t-10 have very low weights (0.0), indicating that the model almost ignores the data from distant days.

D. Inverse Counterfactual Theory Selection (ICFTS)

ICFTS is a counterfactual selection method that helps find counterfactual but plausible explanations, helping users gain deeper understanding of the model. Figure 10 shows the difference between the original and counterfactual predictions of stock prices:

- Original predictions: 158.81
- counterfactual predictions: 155.00

The prediction slightly decreases as the input data changes. This shows that the input factors have an influence on the predictive model. Counterfactual data:

- A falling RSI value indicates a decrease in buying pressure, which could contribute to the expected price drop.
- SMA20 and SMA50 are also slightly decreasing, indicating that the average price trend is showing signs of weakening.
- %Change has more negative fluctuations, indicating that the price is tending to decrease.
- MACD continues to be negative, indicating that the downtrend is dominant.

So this means that if the input features change in the direction of the reference data, the predicted value will decrease.

V. INANCIAL INSIGHTS

According to Shap 7, the High is the highest price of the day, reflecting the maximum price that investors are willing to pay, showing strong buying pressure and positive market sentiment. When the High is high, it often pushes the Close price up, as traders tend to close near the high of the day if market sentiment is positive. However, a high High can also signal risk if the market is at a top, as there may be selling pressure (profit-taking). For LIME 8 and Attention 9, the Close Price is the final price of the day, reflecting the value accepted by the market after a trading day. High Close Prices on recent days indicate a short-term bullish trend, which is usually a positive signal for investors. However, in this case, the Close Price prediction is still low, possibly due to the pattern identifying an overall bearish trend or selling pressure from other factors. The high Low Price indicates that there is no strong selling pressure, as the price did not fall much during the day. This reinforces positive short-term market sentiment, but is not enough to push the Close Price prediction up.

In relation to financial reality:

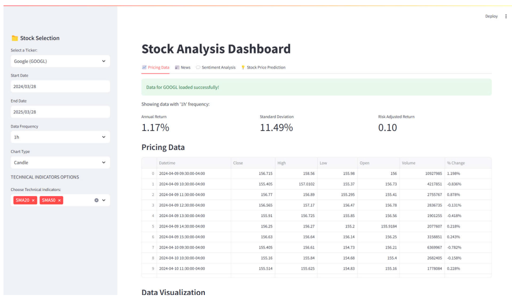
- High prices reflect strong buying pressure, which is a positive signal but also comes with risks if the market is at its peak.
- The high Close and Low prices over the past 5 days indicate a short-term upward trend; however, the low predicted price may signal selling pressure or a potential correction.

Applications in finance: Investors should focus on the High, Close, and Low prices of the last 5 days to assess short-term trends, and monitor trading volume to detect signs of correction or sell-off.

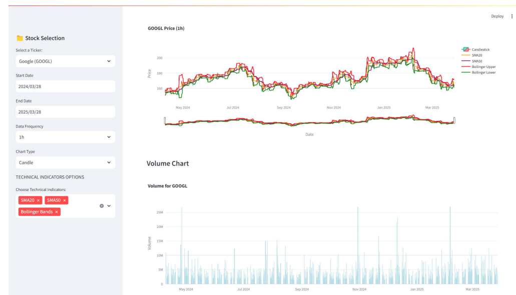
VI. STREAMLIT APPLICATION

An interactive Streamlit application is developed [2] with user interface shown in Figure 11 and allowing users to:

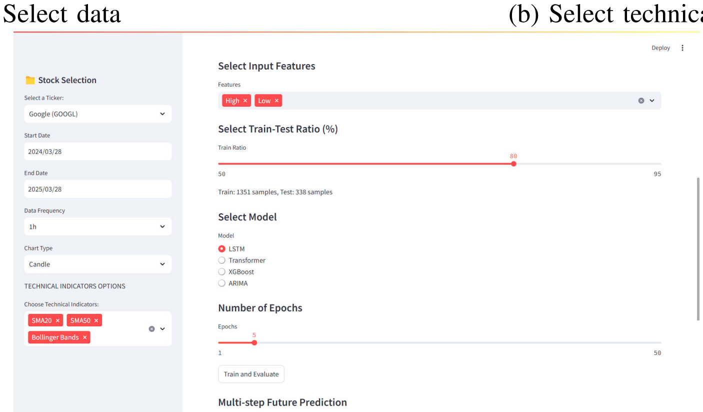
- Select datasets (Amazon, Alibaba, etc.) or upload custom CSV files .
- Allows users to select technical indicators (SMA20, SMA50, RSI, MACD,...) to display on the price chart, helping to determine trends, buy/sell signals, and volatility levels.
- Choose forecasting models (LSTM, Transformer, XGBoost, ARIMA).
- Adjust parameters (e.g., forecast horizon: 1, 5, 30 days, Select Train-Test Ratio).
- Use DAVOTS to visualize the results of weight distribution in forecasts (Attention).
- Implement Visual Explanations to clarify model decisions using XAI (SHAP and LIME charts)
- In addition, the app also adds a feature that allows users to view news related to the stock code and analyze whether the news is positive or negative, show in figure 12. This can help user evaluate the stock trend in the period near the predicted date.



(a) Select data

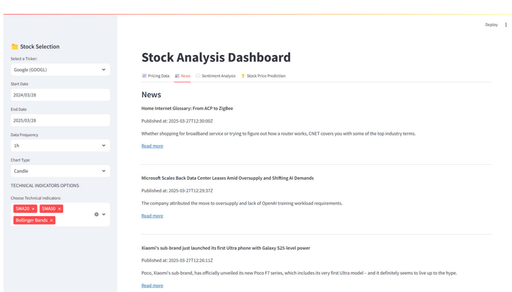


(b) Select technical indicators

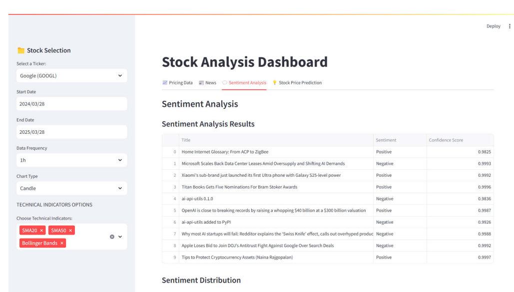


(c) Choose parameters and forecasting mode

Fig. 11: User Interface for Data Selection and Forecasting



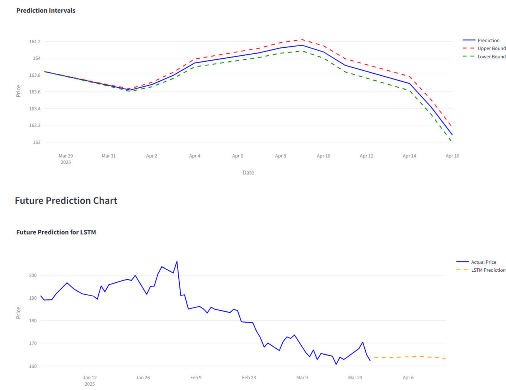
(a) News about stock ticker



(b) Sentiment of news

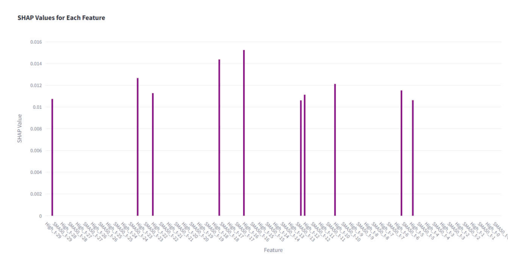
Fig. 12: News and Sentiment Analysis

(a) Prediction of LSTM

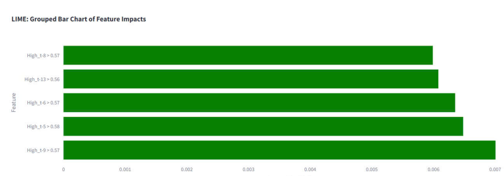


(b) Future Prediction of LSTM

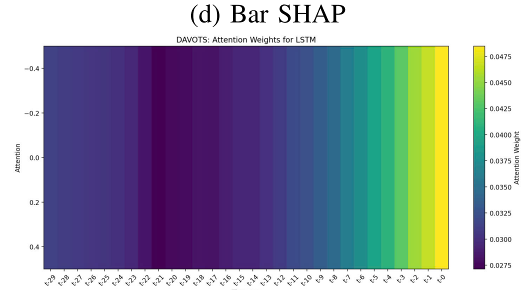
(c) Waterfall SHAP



(d) Bar SHAP



(e) LIME



(f) Attention weight

Original Prediction: 158.81

Counterfactual Prediction: 155.00

Counterfactual Data (last 5 timesteps):

	High	Low	Open	SMA20	SMA50	% Change	RSI	MACD
-5	165.1179	161.1532	162.8914	164.029	185.6786	1.5828	43.2772	-5.4054
-4	167.4289	164.3311	164.7988	163.7799	185.2218	1.0553	44.3494	-4.6738
-3	166.3775	160.8354	164.8183	163.3817	184.6674	-3.887	39.1136	-4.4929
-2	162.15	157.9588	160.4792	163.0502	184.0533	-2.3603	34.4941	-4.5289
-1	158.4962	149.4245	156.4107	162.2341	183.1651	-5.5163	34.2907	-5.1451

(g) ICFTS

Fig. 13: The results of predictions and explanations with XAI

Some of results from app, the GOOGLE stock prediction shown in figure **??** and explanations of the Google stock ticker shown in Figure 13. Through the app with analytical techniques, it can help investors analyze data about factors affecting stock prices. Then they can consider choosing suitable features to forecast prices more accurately, understand model decisions to give more reliable prediction results. Finally, they can make smart trading decisions, buy and sell stocks based on trends, signals and risks.

VII. Conclusion

This project demonstrates that SARIMAX and LSTM models outperform ARIMA and Transformer in forecasting Google (GOOGL) stock prices, achieving lower RMSE (3.44 and 4.62, respectively) and MAE (2.93 and 3.60, respectively). XAI techniques, including SHAP, LIME, and Attention Mechanisms, provide valuable insights into prediction drivers, highlighting the critical role of recent price data (e.g., High, Close, Low within the last 5 days) and the limited impact of trading volume. These findings align with financial reality, where short-term trends and market sentiment often drive stock price movements. The interactive Streamlit application enhances usability and transparency, enabling investors to visualize forecasts and XAI insights, with potential applications in investment decision-making and risk management. Future work could explore additional datasets (e.g., Amazon, NVIDIA) and incorporate external factors (e.g., macroeconomic indicators) to further improve forecasting accuracy and interpretability.

References

[1] M. Zohaib Zeeshan, "Google Stock Price Data 2020-2025 (GOOGL)," Kaggle, 2025. [Online]. Available: https://www.kaggle.com/datasets/mzohaibzeeshan/google-stock-price-data-2020-2025-googl.

[2] Streamlit Documentation, https://real-time-stock-analysis-fptuhcm.streamlit.app/.