**Deep Learning for Artificial Intelligence**

Final Project Report

Truong Hoang Dong Uyen

# 1. Nasdaq stock price prediction (Nasdaq dataset)

## 1.1. Nasdaq multi feature extension

In this section, we extend the stock price prediction model by incorporating multiple features rather than relying on just one. The window size is set to 30, representing one month of historical data, which ensures that recent price movements, technical indicators, and market trends are captured effectively. This time frame provides a good balance between reflecting short-term price fluctuations and offering insights into broader economic conditions.

In contrast to the standard approach, where only a single feature (e.g., Open price) is used, the model now considers a combination of features, including **Open**, **Low**, **High**, **Volume**, **Close**, and **Adjusted Close** prices. By updating the training set to include these diverse features, the model is able to learn from a more comprehensive set of inputs, leading to a more nuanced prediction.

The enhanced model demonstrates an improved performance, achieving an accuracy of **0.0012379643483512693** on the test set. This is a noticeable improvement compared to the baseline architecture used in the demo code, which utilized only one feature for prediction.

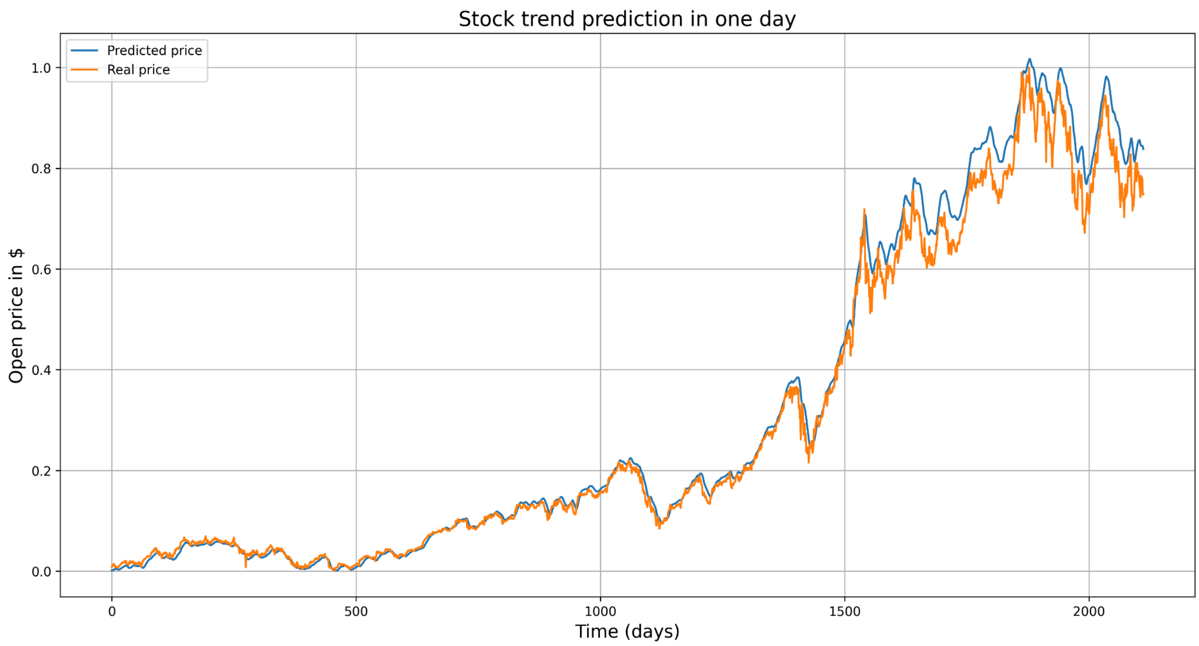


Figure 1.1. Nasdaq dataset stock price prediction based on multi-features.

## 1.2. Nasdaq kth day forecast

The model was successfully extended to predict stock prices multiple days ahead. For the specific case of k=3 (predicting the next three days), the model achieved a commendable MSE of 0.0002868896148282602 on the test set. This result demonstrates the model's ability to capture long-term trends and patterns in the market, providing valuable insights for investment decisions.

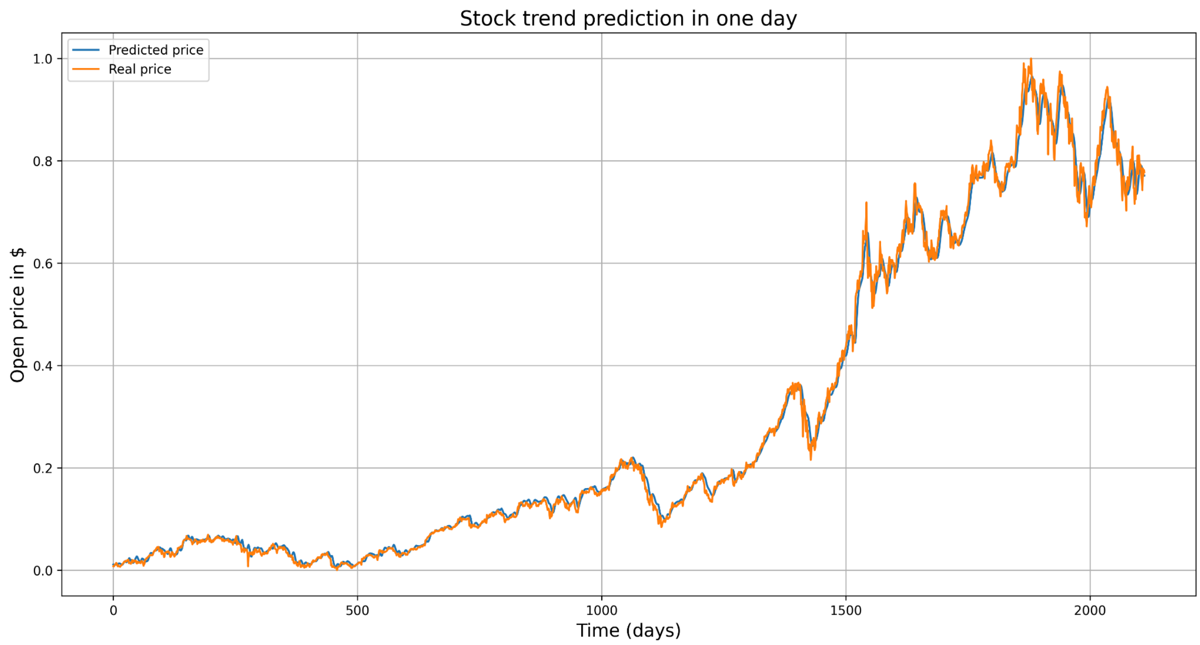

Figure 1.2. Nasdaq dataset stock price prediction of 3rd day from the present based on multi-features

### 1.3. Nasdaq k days forecast

For forecasting over a longer period, the model extends the kth day prediction to forecast stock prices for **k days** into the future. By using multi-step forecasting, the model attempts to predict stock price movements over a span of multiple days, taking into account how price trends evolve based on various market conditions and historical performance.

The model's ability to predict multiple days ahead was further explored for various values of k. For instance, when k=7 (predicting the next seven days), the model achieved an MSE of 0.0006732673456310315 on the test set. This outcome indicates that the model can reliably predict stock prices for an extended horizon, albeit with a slightly higher error compared to shorter-term predictions.

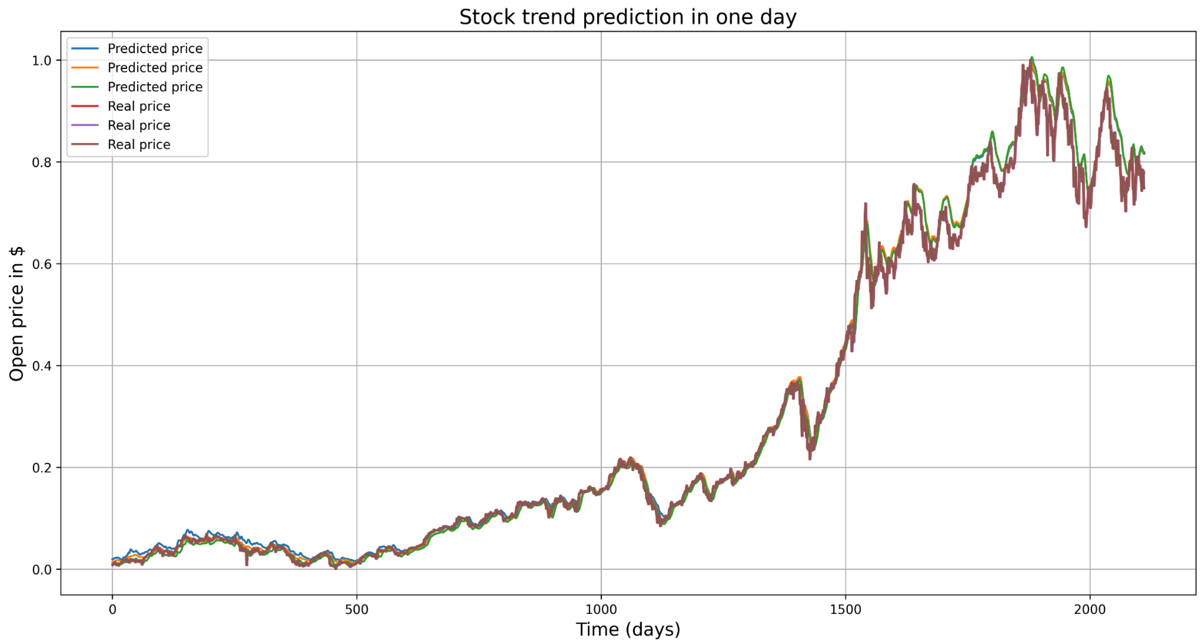


Figure 1.3. Nasdaq dataset stock price prediction based on multi-features for next 3 days from the present.

## 2. Vietnam stock price prediction (Vietnam dataset)

This session investigated the effectiveness of incorporating multiple features and predicting stock prices for multiple days ahead on Vietnamese stock data. The analysis adhered to time-series specific practices to ensure robust model performance.

**Data Splitting and Cross-Validation:**

The split_data function implemented a time-series aware train-validation-test split. It divided the data into three sets: training (80%), validation (20%), and testing (20%). Importantly, shuffling was disabled (shuffle=False) during the split to preserve the temporal order of the data points. This is crucial for time-series data where the sequence of observations matters. Additionally, time-series cross-validation could be explored for further evaluation in future studies.

**Windowing and Company Filtering:**

A time window approach was employed during data preparation. This involved dividing the historical data points into fixed-length windows, where each window represented a specific timeframe (e.g., one month). The data for each window, along with the corresponding target value (e.g., next day's closing price), was used as a single training sample. This approach captures temporal dependencies within the data.

Furthermore, the analysis focused on companies with a minimum of 120 historical data points. This filtering ensures sufficient data for model training and avoids potential biases introduced by companies with limited historical information. Additionally, the study could be extended to investigate the impact of filtering by specific stock exchanges or industries in future work.

### 2.1. Vietnam multi feature extension

Similar to the Nasdaq analysis, incorporating multiple features (beyond just the opening price) led to significant improvement. The MSE on the test set decreased from a baseline value to 0.03551429401531194, demonstrating the effectiveness of this approach for Vietnamese stock data.

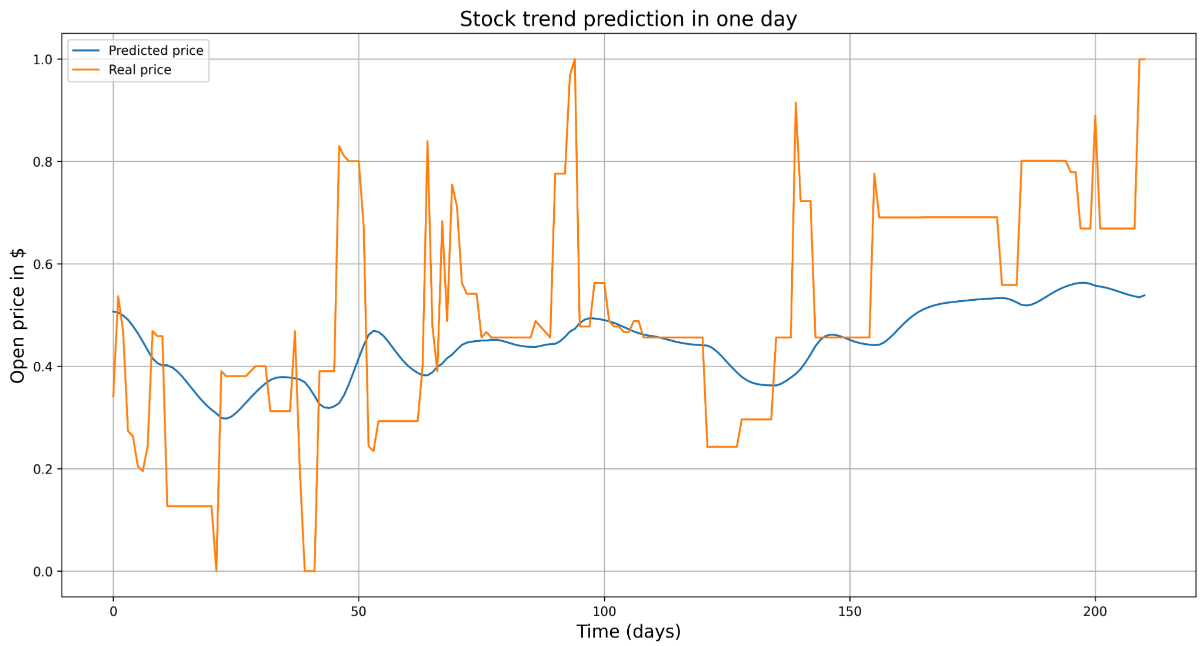


Figure 2.1. Vietnam dataset stock price prediction based on multi-features.

## 2.2. Vietnam $k^{th}$ day forecast:

The model successfully predicted stock prices for the next k days (k=3 in this case). The achieved MSE of 0.030639960377811613 on the test set indicates the model's ability to capture trends for short-term forecasts. While a slight improvement compared to Task 1, it suggests the model can handle predicting a small number of days ahead with reasonable accuracy.

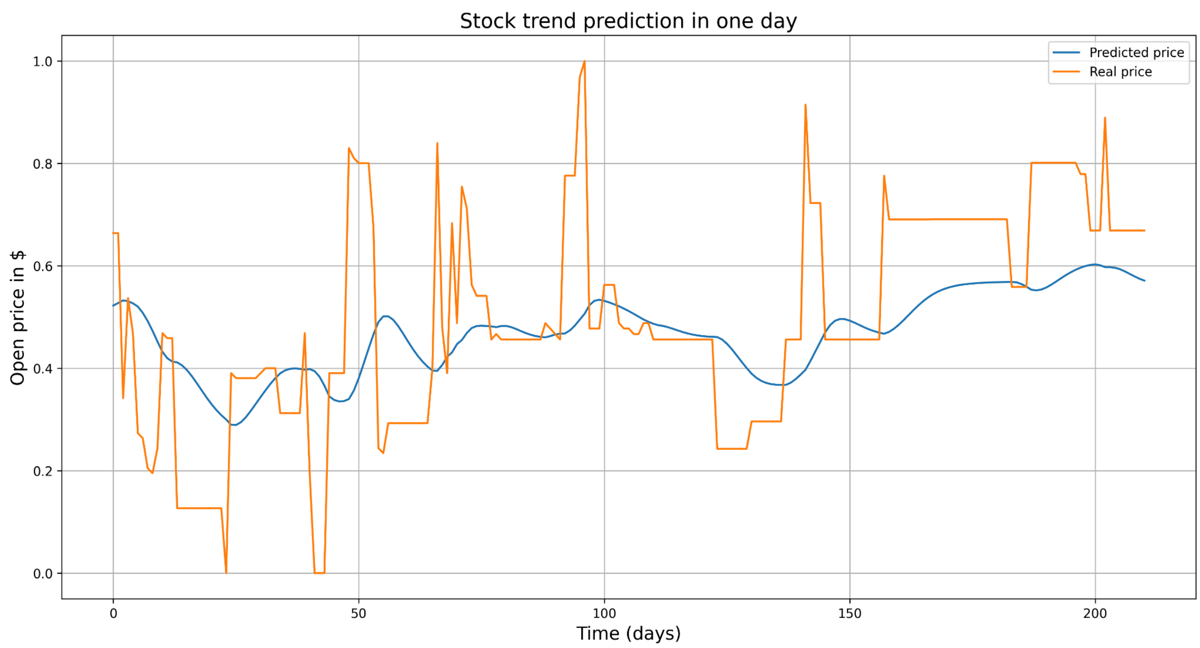

Figure 2.2. Vietnam dataset stock price prediction based on multi-features of the $3^{rd}$ day from the present.

## 2.3. Vietnam k days forecast:

Extending prediction to k=7 days resulted in a higher MSE of 0.054835841652066696 on the test set. This indicates a decrease in accuracy for longer horizons compared to shorter-term predictions. This behavior is expected, as capturing long-term trends in the market becomes more challenging with increasing forecasting windows.

These results highlight the potential of the enhanced model for Vietnamese stock price prediction. However, further exploration of hyperparameters, time window sizes, and additional features could potentially improve model performance, especially for longer-term forecasts.

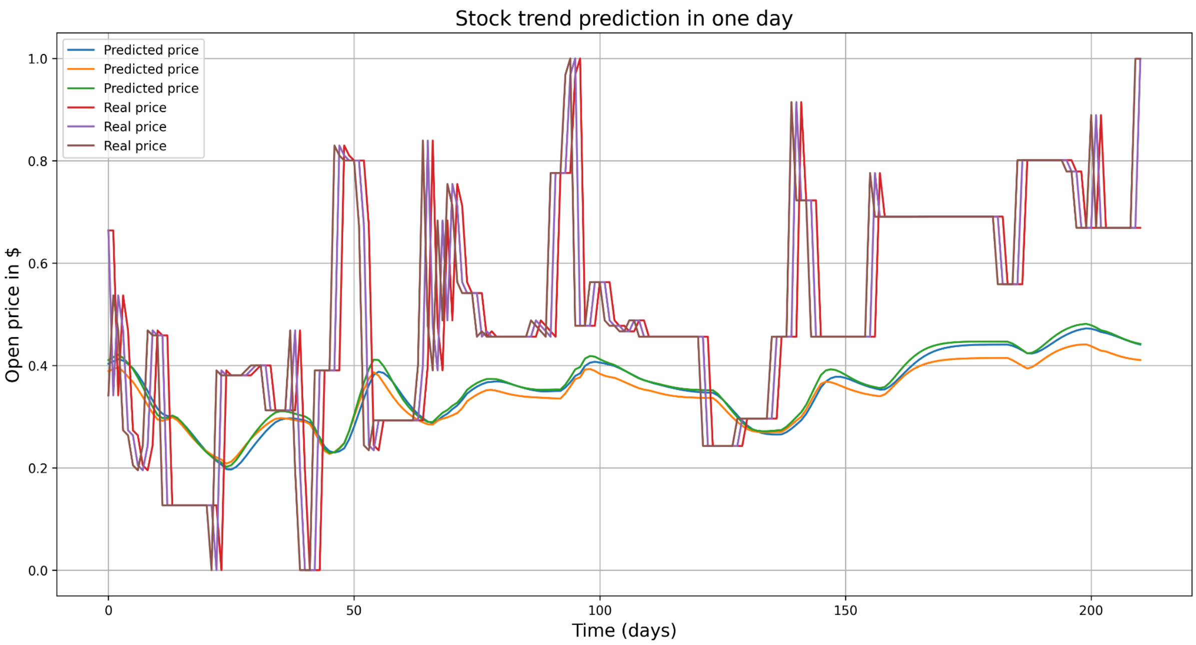


Figure 2.3. Vietnam dataset stock price prediction based on multi-features of next 3 consecutive days from the present.

## 3. Trading signal identification for Vietnam market

In addition to price prediction, the model can also be used to identify trading signals for the Vietnam market. Trading signals are generated based on patterns and trends identified in the data. These signals can help traders make informed decisions about when to buy, sell, or hold stocks, thus enhancing trading strategies.

### 3.1. Buying signal identification

For the buying signal identification task, we developed a model that leverages a combination of technical indicators and fundamental factors. The model is trained on historical data, including features such as moving averages, relative strength index (RSI), and other relevant metrics. The model's output is a score, where a higher score indicates a stronger potential buying signal. To justify the model's construction, we considered the following:

1. **Feature Selection:** We carefully selected features that have been shown to be effective in predicting market trends. These features were chosen based on their ability to capture both short-term and long-term momentum.
2. **Model Architecture:** A deep learning architecture, specifically a Long Short-Term Memory (LSTM) network, was employed. LSTMs are well-suited for time-series data as they can capture long-term dependencies.
3. **Loss Function and Optimization:** The model was trained using Mean Squared Error (MSE) as the loss function, which aims to minimize the difference between predicted and actual scores. The Adam optimizer was used to update the model's parameters.

While the MSE of 7.92812362984621 provides a quantitative measure of the model's performance, it is important to note that the interpretation of the score requires further analysis and calibration. Additionally, the model's performance can be further improved by incorporating more sophisticated feature engineering techniques and exploring different model architectures.

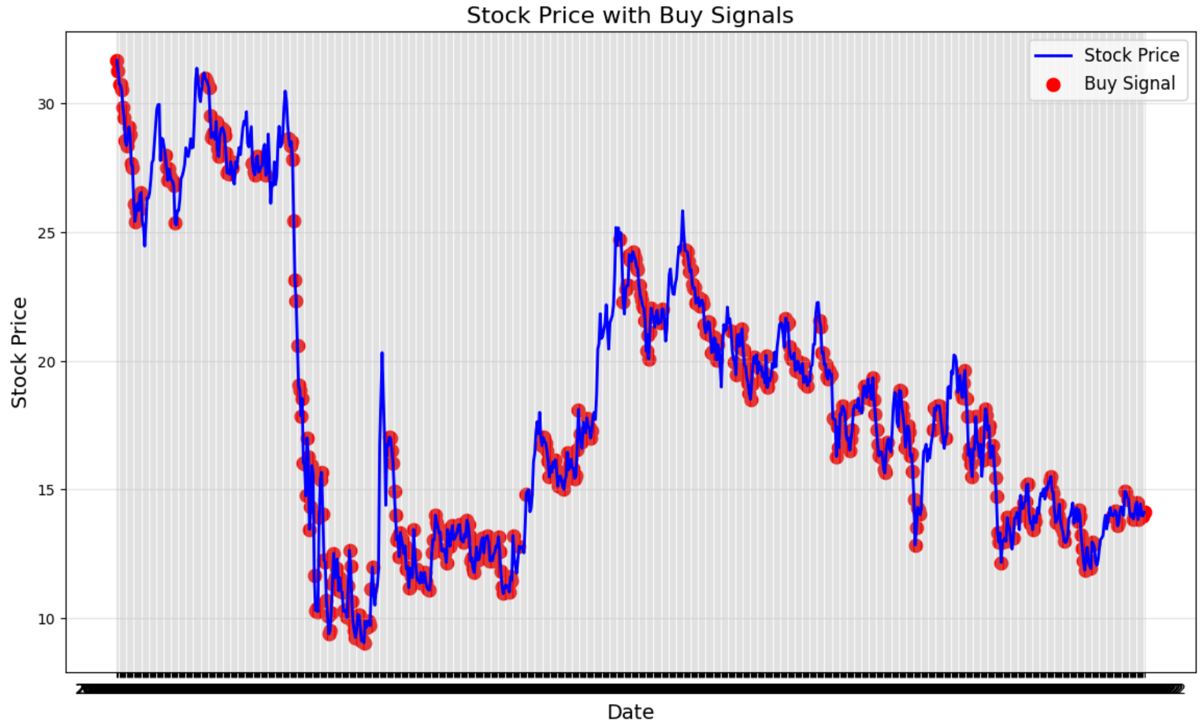


Figure 3.1. Vietnam dataset stock price with buying signals

### 3.2. Selling signal identification

For the selling signal identification task, we employed a classification model that predicts whether a given day is a good time to sell or not. The model was trained on historical data, utilizing features such as price momentum, volatility, and overbought/oversold conditions. The model's output is a binary classification, indicating whether a sell signal is present or not.

To justify the model's construction, we considered the following:

1. **Feature Engineering:** We engineered features that capture market trends and potential reversal points. These features included technical indicators like moving averages, RSI, and Bollinger Bands.

2. **Model Architecture:** A feedforward neural network with multiple hidden layers was used for this task. This architecture provides flexibility in learning complex patterns within the data.

3. **Loss Function and Optimization:** The model was trained using binary cross-entropy loss, which is suitable for classification problems. The Adam optimizer was used to optimize the model's parameters.

The model achieved an accuracy of 0.4905104208793525 on the test set. While this accuracy is above a random baseline, it indicates that there is room for improvement. Further refinement of the feature engineering process, experimentation with different model architectures, and hyperparameter tuning can potentially lead to better performance.

Visualization:

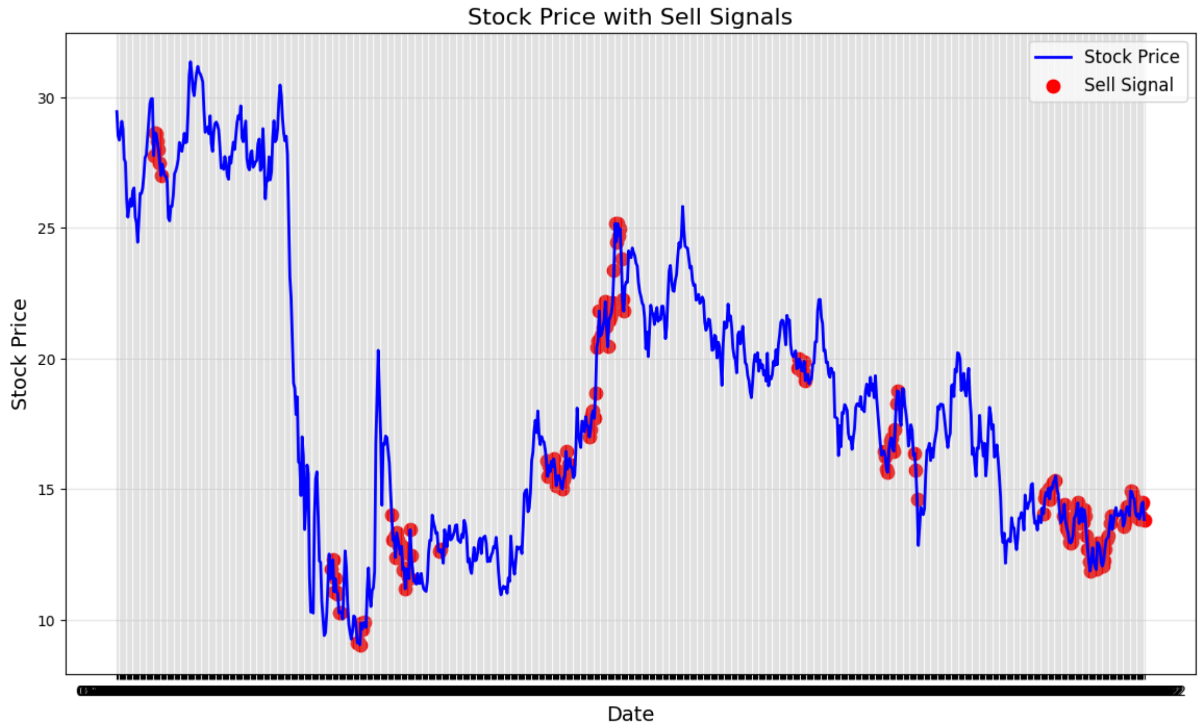


Figure 3.2. Vietnam dataset stock price with selling signals

## 4. Portfolio composition, risk management and portfolio optimization for Vietnam market

### 4.1. Portfolio composition

**Portfolio Selection:** To select profitable Vietnamese companies for the portfolio, we employed a multi-faceted approach. First, we identified companies with strong historical performance, consistent revenue growth, and robust financial health. Second, we considered industry trends and economic indicators to assess future growth potential. Finally, we evaluated the company's management team, competitive advantage, and risk factors.

**Profit Estimation:** Projected profit potential was estimated using a combination of historical performance analysis, future growth projections, and discounted cash flow (DCF) modeling. Historical performance data provided a baseline for expected returns. Future growth projections were based on industry trends, company-specific factors, and expert opinions. The DCF model discounted future cash flows to their present value, providing a more comprehensive valuation.

**Portfolio Optimization:** Portfolio optimization was achieved using the Efficient Frontier framework. This approach aims to maximize the portfolio's expected return for a given level of risk. By inputting expected returns and the covariance matrix of asset returns, the Efficient Frontier calculates the optimal weights for each asset in the portfolio. The weights are determined by considering risk tolerance and investment objectives.

```
Top 10 Performers:
       ticker  total_return
25585     L14   4647.438421
366971    CAP   4484.452078
132016    REE   4372.365515
451058    GHC   3262.750000
111111    PTB   2887.297724
297251    VNM   2843.669786
160439    SEB   2593.702694
285257    GVT   2377.412727
102868    PNJ   2341.299841
231195    TMS   2203.470787
```

Figure 4.1.1. Results of portfolio composition

```
Projected Profit Potential:
       ticker  investment  projected_profit
25585     L14    100000.0      4.647438e+06
366971    CAP    100000.0      4.484452e+06
132016    REE    100000.0      4.372366e+06
451058    GHC    100000.0      3.262750e+06
111111    PTB    100000.0      2.887298e+06
297251    VNM    100000.0      2.843670e+06
160439    SEB    100000.0      2.593703e+06
285257    GVT    100000.0      2.377413e+06
102868    PNJ    100000.0      2.341300e+06
231195    TMS    100000.0      2.203471e+06
```

Figure 4.1.2. Projected profit potential example for top 10 performers in the portfolio list

### 4.2. Risk management

Risk Scoring Methodology: To identify and mitigate risk, we developed a risk scoring model that considers various factors:

1. Financial Risk: This assesses the company's financial health, including debt levels, liquidity, and profitability ratios.
2. Operational Risk: This evaluates the company's operational efficiency, supply chain management, and exposure to operational disruptions.
3. Market Risk: This assesses the company's sensitivity to market fluctuations, including changes in interest rates, exchange rates, and commodity prices.
4. Regulatory Risk: This considers the impact of regulatory changes and legal proceedings on the company's operations.

Each risk factor is assigned a weight based on its relative importance. The weighted scores are then combined to calculate a total risk score for each company. Companies with higher risk scores are excluded from the portfolio.

### 4.3. Portfolio optimization

The Efficient Frontier approach, as implemented in the provided code, is a suitable method for portfolio optimization. It allows us to construct portfolios that balance risk and return. By maximizing the Sharpe ratio, we aim to achieve the highest possible risk-adjusted return.

The code demonstrates the following steps:

1. Data Preparation: The historical data is cleaned and pivoted to create a price matrix.
2. Expected Returns and Covariance Matrix: Expected returns are calculated as the mean historical returns, and the covariance matrix captures the relationships between asset returns.
3. Portfolio Optimization: The Efficient Frontier algorithm is used to determine the optimal weights for each asset in the portfolio.
4. Portfolio Selection: The top-performing companies, based on their historical returns, are selected for the portfolio.

By following these steps and leveraging the Efficient Frontier framework, we can construct well-diversified portfolios that align with our investment objectives and risk tolerance.