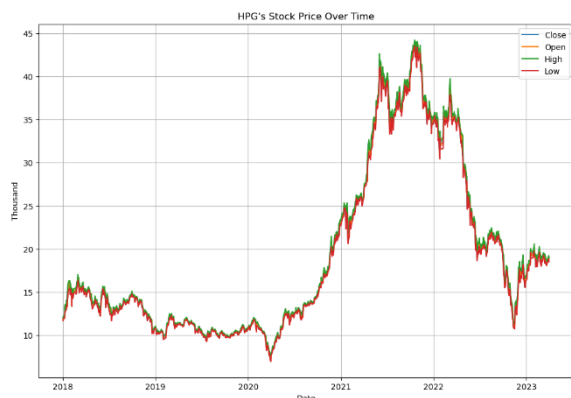


PREDICT THE PRICE OF THE STOCK HPG

- Astrophile -

1. Exploring Data Analysis



From Figure 1, it is clearly seen that the four factors including Closing price, Opening price, the Highest Price and the Lowest witnessed the same trend during the period tested.

More importantly, we can see that HPG stock price had a close relationship with economic and social fluctuations. After the emergence of COVID19 pandemic, the stock market had witnessed a remarkable drop by March of 2020 before it boomed. The low-interest-rate environment, the urgency to find new investment channels amidst a period of social isolation, and the expectation of a high rate in return all contribute to stocks lead to a record high increase in the stock market as well as in HPG stock. However, Vietnam stock

market in general and HPG stock in particular witnessed the sharpest decrease after the first quarter of 2022 due to several reasons. The geological tension between Russia and Ukraine causing disruption in the global production and supply chain, high inflation and rising interest rate after more than a year of loose monetary policy, especially, the prosecution and arrest of former chairman of FLC Group for manipulating the stock market severely affected investor sentiment, leading to a wave of capital withdrawal.

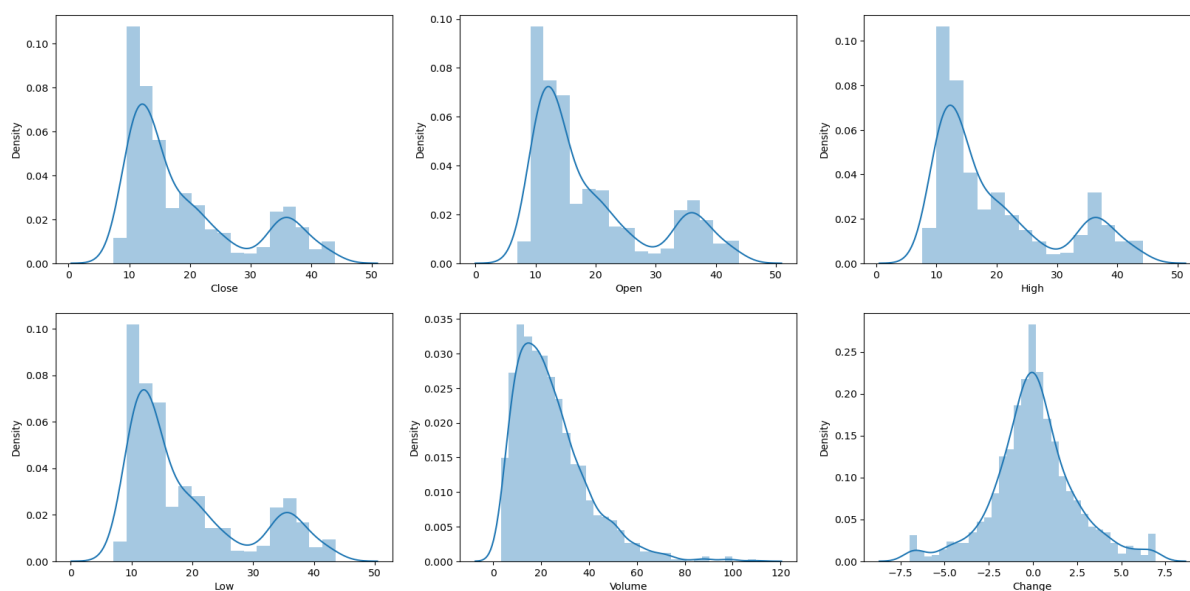


Figure 2. Distribution Histogram

Figure 2 displays six histograms with overlaid density plots. Each plot represents the distribution of different financial data attributes. We can see that:

- Most closing prices are clustered around 10 to 20, with fewer occurrences as the price increases.
- The density plot shows that the majority of opening prices fall between 10 and 20, with a noticeable peak. There are fewer instances of higher opening prices, with some dispersion above 30.
- Similar to the closing and opening prices, the high prices are mostly concentrated around 10 to 20. There are occasional peaks at higher price levels, indicating some higher price points.
- The density plot reveals that low prices predominantly range from 10 to 20. As with other price attributes, there's a lesser occurrence of higher low prices.
- The density plot shows that trading volume is highest at the lower end, particularly between 0 and 20. There's a long tail towards higher volumes, indicating some days with significantly higher trading activity.
- The density plot indicates that most price changes are around 0, showing a normal distribution. This suggests that the HPG stock price typically does not change drastically day-to-day, with changes mostly clustering around small positive or negative values.

2. Modeling

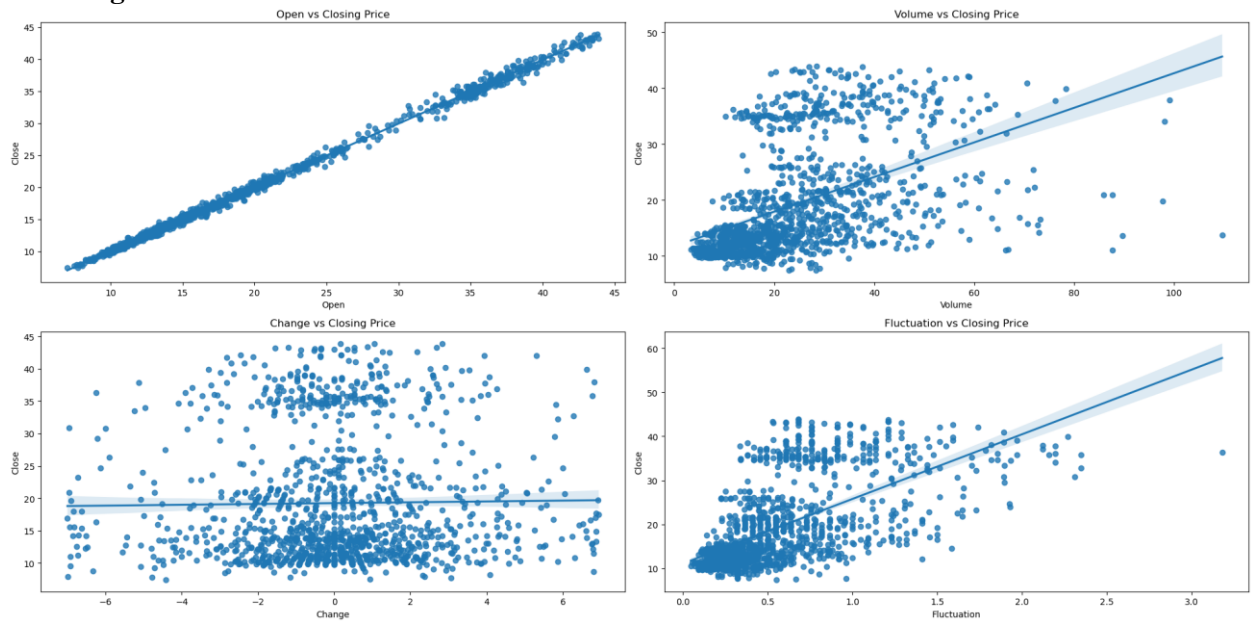


Figure 3. Scatter plots histogram

Figure 3 displays scatter plots with regression lines, representing relationships between different variables. There is a positive correlation between Opening Price and Closing Price. Higher Opening Prices are generally associated with higher Closing prices.

There is an insignificant correlation between trading volume and Closing Price. While trading volumes follow an upward trend, Closing Price follows a non-linear distribution.

There is no significant correlation between price changes and Closing Price. The price changes do not strongly influence the Closing Price.

The correlation between Fluctuation (differences between the highest prices and the lowest prices) and Closing Price is similar to the correlation between trading Volume and Closing Price.

Table 1. ARIMAX model and LSTM model comparison

Model	ARIMAX(p,d,q)	LSTM
Process	1. Data Preparation: <ul style="list-style-type: none"> Use first-order differencing to make the series stationary. Conduct a stationarity test using the Dickey-Fuller test. Split the data into training and testing sets. 2. Model Building: <ul style="list-style-type: none"> Determine the best p, d, and q values using ACF and PACF plots. Train the ARIMAX model using the determined parameters. Use external features for the ARIMAX model. 3. Forecasting and Evaluation: <ul style="list-style-type: none"> Predict the stock prices for the testing period. Evaluate the model using R square, MAPE, and RMSE metrics. 	1. Data Preparation: <ul style="list-style-type: none"> Use Min-Max scaling and standardization for data preprocessing. Split the data into training and testing sets. Create windows for LSTM input. 2. Model Building: <ul style="list-style-type: none"> Use hyper-parameter tuning techniques (Optuna, RandomizedSearchCV) to find the best parameters for the LSTM model. 3. Training and Evaluation: <ul style="list-style-type: none"> Train the models on the training data (January 2nd, 2018 to February 28th, 2022). Predict the stock prices for the testing period (March 1st, 2022 to March 31st, 2023). Evaluate models using R square, MAPE, and RMSE metrics.
The best model	$p = 2, d = 1, q = 2$	

parameters		
Metrics		
R-Squared	0.996112	0.921211
MAPE	0.012974	0.069128
RMSE	0.397135	1.787683

3. Result



Figure 4. Forecasted results using LSTM model

Toward Figure 4, we can see that the forecasting model captures the general trends of the actual closing prices but struggles with high volatility periods. During stable periods, the forecast aligns closely with the actual prices. However, there are noticeable discrepancies, particularly during periods of high volatility. The forecasted prices show more frequent fluctuations compared to the actual prices.



Figure 5. Forecasted results using ARIMAX model

Looking at the Figure showing results using the ARIMAX model, we can see that the forecasting model captures the general trends of the actual closing prices pretty well. The forecast prices align closely with the actual prices, even during periods of high volatility. Clearly, the ARIMAX model can be considered to make accuracy improvement during volatile periods and reduce the sensitivity to short-term fluctuations in comparison to the LSTM model.