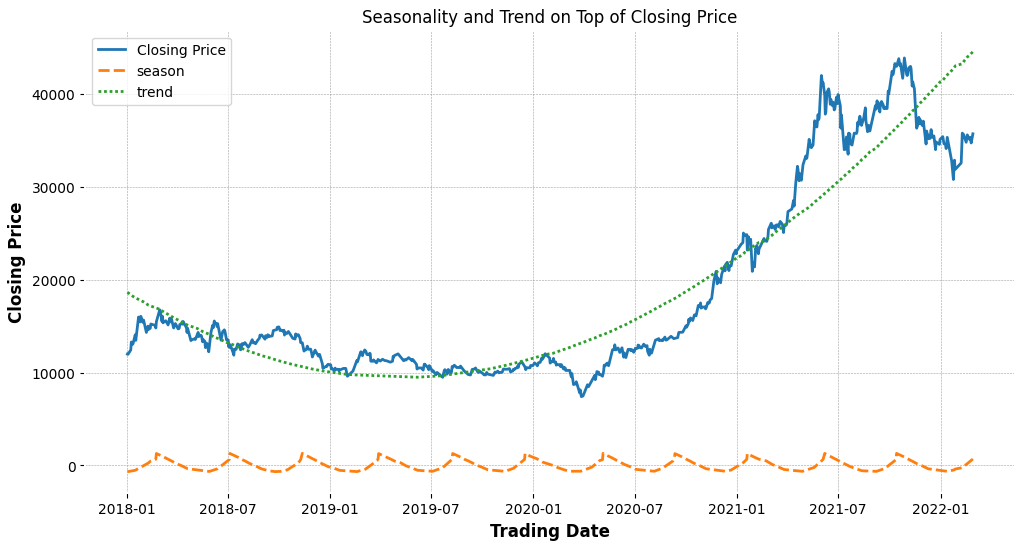
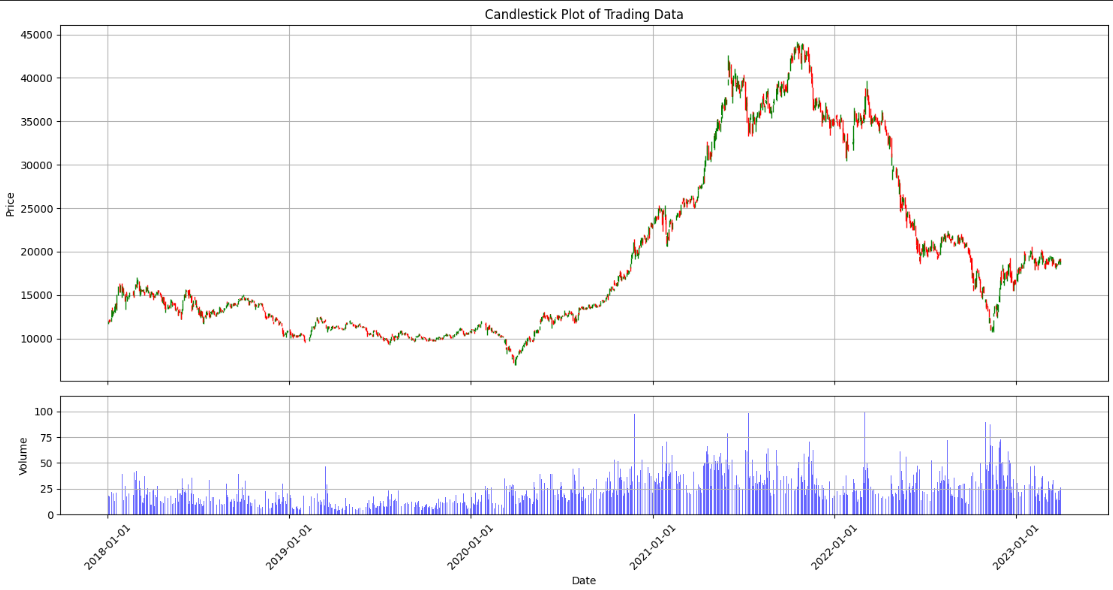
**DATA ANALYSIS AND PREDICTIVE MODELING FOR HPG STOCK PRICE (2018-2023)**

1. ***Exploratory Data Analysis (EDA)***

Time series plots of closing prices for training and test periods showed overall stock price trends, revealing phases of growth and decline. The 50-day and 200-day moving averages highlighted these trends, indicating periods of bullish and bearish behavior. Additionally, daily returns showed variability with high and low volatility periods. The 50-day rolling volatility plot provided insights into return consistency, indicating periods of higher risk and fluctuation.

A scatter plot of trading volume versus daily returns suggested that higher trading volumes might be linked to larger price changes. A corresponding correlation analysis indicates a potential connection between trading activity and price movements.

Overall, HPG's stock price showed a strong upward trend, reflecting robust financial performance and strategic expansions.



In 2019, HPG's stock price increased to VND 31,000, driven by a revenue of VND 64,678 billion and a net profit of VND 7,578 billion, according to the annual financial report. One of the most important causes is the effective functioning of the Dung Quat Steel Complex, which boosted capacity, improved competitiveness, and led to stock price growth.

Considering the various worldwide obstacles faced by the COVID-19 pandemic, HPG's stock price remained resilient in 2020. Furthermore, despite mid-year turbulence, the stock rallied sharply. According to the annual report, sales increased to VND 91,279 billion, with a net profit of VND 13,506 billion. Strategic expansions and government policies that promote economic growth helped to retain investor confidence and demand for steel.

HPG’s stock price peaked in 2021, reaching VND 58,000 mid-year. This remarkable performance was driven by exceptional financial results, with the annual report showing revenue at VND 149,680 billion and net profit at VND 34,521 billion. The launch of the Dung Quat 2 Steel Complex and the acquisition of the Roper Valley Iron Ore Mine Project boosted production and secured raw material supply, positively influencing investor sentiment.

A remarkable decline in total sales as well as a 30% drop in net profit from the previous year resulted in a 20% loss in stock price in 2022. Rising raw material costs and worldwide disruptions in supply chains affected profitability, while rising competition pushed margins down, causing investor sentiment to decrease. Missed earnings expectations further contributed to the negative trend, with the company’s EPS and revenues not meeting analyst projections.

HPG showed some recovery in the first quarter of 2023. The EPS stood at VND 896, and despite the reduced profit margin of 8.1%, it still suggested probable stability. According to market experts, major investments in new technologies and expansions at the Dung Quat 3 Steel Complex are projected to boost investor confidence. Furthermore, excellent results in the third quarter of 2023, with consolidated sales and a net profit after tax of VND 2,000 billion, supported the recovery trajectory greatly.

1. ***Predictive Analysis and Results***

| **Model** | **Preprocessing** | **Model Configuration** | **Parameters** | **Technology** |
| --- | --- | --- | --- | --- |
| **KNN-DTW** | MinMaxScaler | Using GridSearchCV, we systematically explored variations in the number of neighbors (n\_neighbors) and different distance metrics, including specialized methods like Dynamic Time Warping (DTW) for time series data. This approach allowed us to identify the most effective configuration for our regression task through cross-validation. | metric: DTW  n\_neighbors: 3 |  |
| **XGBoost** | MinMaxScaler | Through a rigorous grid search process using GridSearchCV, we optimized hyperparameters such as the number of trees (n\_estimators), maximum depth of each tree (max\_depth), learning rate (learning\_rate), and other parameters. The final XGBoost model was selected based on its performance metrics derived from cross-validation. | colsample\_bytree: 1.0  gamma: 0 learning\_rate: 0.1 max\_depth: 4 min\_child\_weight: 1  n\_estimators: 100  subsample: 0.8 |  |
| **ARIMA** | Log transformation,  First - order differencing | Using ARIMA on train data to predict. After each loop, the train data will be renewed by adding real data from the test data to predict afterward. | p = 2  d = 1  q = 2 |  |
| **LSTM** | MinMaxScaler | The LSTM model includes three LSTM layers (256, 128, 64 units) with tanh activation and ‘return\_sequences=True’ for the first two, each followed by Dropout (0.3). Two Dense layers (64, 32 units) with ‘relu’ activation and L2 regularization are added, ending with a Dense layer of one unit for prediction. | learning\_rate = 0.001; batch\_size = 32; epochs=200; time\_step=20; patience=10 | Learning Rate Scheduler; Adam Optimizer; Early Stopping |

| **MODEL** | **METRICS** | | |
| --- | --- | --- | --- |
| **RMSE** | **MAPE** | **R-Squared** |
| KNN-DTW | 407.06 | 1.22% | 1.00 |
| XGBoost | 351.59 | 1.22% | 1.00 |
| ARIMA | 608.14 | 2.23% | 0.99 |
| LSTM | 1032.55 | 3.99% | 0.97 |

Overall, all models show fairly good performance. Among them, KNN-DTW and XGBoost perform exceptionally well with a perfect R-squared value of 1.00 and very low RMSE, MAPE, indicating high accuracy and prediction capability. Additionally, ARIMA performs moderately well with a high R-squared of 0.99, but higher RMSE and MAPE compared to KNN-DTW and XGBoost, suitable for time series data with clear trends. Finally, LSTM, despite its potential for long-term time series data, shows the lowest performance with the highest RMSE and MAPE, and a lower R-squared of 0.97, indicating the need for further tuning. In conclusion, XGBoost emerges as the best-performing model for this task.

[You can find the code here](https://drive.google.com/drive/folders/1IEVT1ZJ-f6XC1TH_HNunmvQxeoCfJw_G?usp=sharing)