**Predicting the Probability of Next Days Candlestick Pattern of HDFC Bank**

By: - Abhijeet Chavan

**Introduction:**

The stock market is a complex and dynamic environment where investors and traders aim to make informed decisions to maximize their returns. The ability to forecast future stock price movements is a valuable asset in this endeavor. In this project, we focus on analyzing the historical data of HDFC Bank stock and developing a predictive model to forecast the candle colors (green or red) for future days. By leveraging various features and employing machine learning techniques, we aim to gain insights into the potential movement of HDFC Bank stock and make informed trading decisions.

**Data Preprocessing:**

To begin the analysis, historical data of HDFC Bank stock is obtained and loaded into the system from an Excel file. The dataset includes attributes such as Open, Low, Close, and High prices, as well as the trading volume. However, before proceeding with the analysis, it is essential to preprocess the data and ensure it is in a suitable format.

One crucial aspect of data preprocessing is the calculation of the Relative Strength Index (RSI). RSI is a momentum oscillator that measures the strength and speed of price movements. By calculating RSI using the historical data, we obtain a valuable indicator that provides insights into the underlying trends and potential reversals in the stock's price movement.

In addition to RSI, we also calculate the 14-day average volume and 14-day average price. These features offer valuable information about the trading activity and price trends over a specific period. By incorporating these calculated features, we enhance the dataset and provide additional context for the predictive model.

**Model Training and Evaluation:**

The next step in the analysis is model training and evaluation. We employ the HistGradientBoostingClassifier, a powerful gradient boosting classifier algorithm, to train our predictive model. To ensure the reliability of the model's predictions, we split the dataset into training and testing sets, with 80% of the data used for training and 20% reserved for evaluating the model's performance.

During the training phase, the model learns from the historical data and identifies patterns and relationships between the features and the corresponding candle colors. Once the model is trained, it is evaluated on the testing set to assess its ability to generalize to unseen data.

The accuracy score is used as the evaluation metric to measure the performance of the model. The accuracy on the training set provides an indication of how well the model captures the underlying patterns in the data. In our analysis, the accuracy on the training set is found to be 0.9897, which suggests that the model performs exceptionally well in capturing historical trends.

Furthermore, the accuracy on the testing set is calculated to evaluate the model's ability to generalize to new, unseen data. The accuracy on the testing set is 0.7992, indicating that the model performs reasonably well in predicting the candle colors on unseen data. These results highlight the model's robustness and its potential to make accurate predictions.

**Predicting Next Day's Candle Probability:**

Having trained and evaluated the model, we proceed to predict the probabilities of the next day's candle color using the entire dataset. This step allows us to anticipate the potential movement of HDFC Bank stock and make informed trading decisions.

To predict the next day's candle color, we extract the most recent information from the dataset, including the features such as Open, Low, Close, High prices, RSI, and the calculated 14-day average volume and 14-day average price. We input this information into the trained model, which leverages its learned patterns and relationships to predict the probabilities of the candle being green or red.

The predicted probabilities offer valuable insights into the potential movement of HDFC Bank stock for the next day. For example, the predicted probability of a green candle is 0.9881, while the predicted probability of a red candle is 0.0119 for the next day (23/06/2023). These probabilities provide market participants with an understanding of the potential outcomes and can guide their trading decisions accordingly.

**Conclusion:**

In conclusion, this project has focused on the analysis and prediction of HDFC Bank stock candle colors using historical data and machine learning techniques. By incorporating features such as Open, Low, Close, and High prices, along with derived indicators like RSI, 14-day average volume, and 14-day average price, we have developed a predictive model capable of forecasting future candle colors.

The model demonstrated high accuracy on the training set (0.9897), indicating its ability to capture historical patterns effectively. Moreover, the model performed reasonably well on the testing set (0.7992), showcasing its potential to generalize to unseen data and make accurate predictions.

By leveraging the predictive model, we were able to anticipate the probabilities of the next day's candle colors. The predicted probabilities provided valuable insights into the potential movement of HDFC Bank stock and assisted market participants in making informed trading decisions.

This project highlights the significance of data analysis and machine learning techniques in the financial domain. The ability to analyze historical data, extract meaningful features, and develop predictive models opens up new avenues for traders and investors to enhance their decision-making process and potentially maximize their returns in the stock market.