# Classification Model on Yahoo Finance

In this project, I set out to build a classification model using stock price data from Yahoo Finance to predict daily stock price movement. My first step was to introduce the problem and explain the importance of understanding short-term stock price fluctuations, as such predictions can be crucial for investors and financial analysts. I then explored the dataset, which included features like opening, high, low, and closing prices and trading volume. Given the goal of classifying price movement, I transformed these features and created a target variable representing whether the closing price was higher than the opening price.

To prepare the data, I handled missing values by dropping incomplete rows and converted the date column to a date-time format for future analysis. I calculated the daily returns and used them to define the target variable. Regarding feature engineering, I focused on using straightforward financial metrics rather than more complex indicators, which allowed the model to remain interpretable while utilizing the core financial signals in the data.

For modeling, I selected the XGBoost algorithm due to its ability to handle complex relationships and provide feature-importance insights. I trained the model on an 80-20 data split, optimizing for efficiency by setting a limited number of estimators and implementing early stopping to prevent overfitting. I then evaluated the model with metrics such as accuracy, precision, recall, and F1 score, which helped me understand the model's performance in identifying upward and downward movements. Additionally, I visualized feature importance and confusion matrix plots, which offered a better understanding of the model’s decision-making process.

Overall, this project emphasized the importance of selecting appropriate features and tuning model parameters and demonstrated how XGBoost can provide a reliable baseline for stock price classification tasks. Although the results were promising, further improvements could be achieved by integrating additional financial indicators or external market factors.