**An Examination of AI and Machine Learning Techniques for Stock Market Predictions**

[A graph on a screen

AI-generated content may be incorrect.](https://github.com/im360wayne/AI-ML-StockMarketPredictions/blob/main/images/stock-banner.jpg)

**Author: Wayne H. Lee**  
Email: [wayneone@gmail.com](mailto:wayneone@gmail.com)

**Executive summary**

**OVERVIEW**

The goal of this project is to determine whether a model can be developed to predict the days on which a stock is expected to rise—using historical data—with better results than buying on random days. Ideally, if a trader buys only on days indicated by the model, they would expect a net gain over losses.

**Expected Results**

It is expected that some momentum exists in the market based on fundamentals, so some correlation should be observable in at least certain time frames for some stocks.

**Rationale**

Being able to predict stock price movement can help inform investment decisions. If a stock is predicted to go up within a particular timeframe, it can be bought, held for that period, and then sold for a gain.

**Research Question**

Can stock price movement be predicted using historic stock market data?

**Data Sources**

Historical stock price data from Yahoo! Finance stocks including historic prices, volume, dividends and other stock metrics such as a volatility index.

**Methodology**

The industry standard model of CRISP-DM will be used to guide the analysis.

**Processing Techniques**

The investigation will be framed as a classification problem: Will the historical inputs result in a classification of the price as "Up" or "Down"?

The raw historical stock data—including price, volume, and dividends—will be used. The following engineered features, calculated from existing data, will be employed:

* Historic Prices: Price of the stock x-days back
* Historic Deltas: Difference between today's opening price and that from x-days ago.
* Horizons: Ratio of current price to an x-day average
* Number of Days of Stock Increases: Count of days the stock increased.

For the training dataset, an "Up" or "Down" result is calculated based on whether the stock increased or decreased in value from the previous day's opening price.

Decision Trees, K-Nearest Neighbors (KNN), and Support Vector Machines (SVM) models—with various hyper-parameters tuned using GridSearch—will be used to perform the classification.

**Results**

Preliminary analysis shows that a test precision score of up to 68% can be reached indicating that such a model is possible and would beat out purchasing stocks on a random selection of days.

**Additional Observations**

* Engineered features, such as a horizon ration, and historical time shifted deltas were the most important in predictions.
* As part of exploratory data analysis (EDA), it was discovered that the maximum correlation values were associated with the ratio of the two-day average price change and the previous day’s price, respectively. This finding suggests some market momentum, where price increases on previous days can help predict continued upward movement.
* While trading volume follows a roughly normal distribution, it is very skewed towards large trades. This could be due to a high number of institutional investors that react upon certain market changes with big trades.
* Changes in the daily price (delta) appear to form an approximately bell-shaped curve. This indicates that small corrections in the stock value occur more frequently, whereas large price changes—either positive or negative—are relatively rare. Future research could leverage this distribution to estimate the potential magnitude of price changes offering additional insights for traders.

**Next steps**

* Continue developing and tuning the model.
* Complete full model development and tuning.
* Identify comprehensive next steps.

**Outline of project**

* [StockMarketPredictions.ipynb](https://github.com/im360wayne/AI-ML-StockMarketPredictions/blob/main/StockMarketPredictions.ipynb)