**Algorithmic Trading and Market Trends: A Study of Statistical and Machine Learning Models**

**Introduction**

Both algorithmic and manual trading focus on capitalizing on market inefficiencies. Algorithmic trading involves using automated computer models to analyse market data and execute trades. It is ideal for addressing persistent inefficiencies, where trading opportunities occur repeatedly in a similar pattern.

For example, if gold prices are consistently lower on a Japanese exchange compared to a US exchange (after factoring in the USDJPY currency rate), one can continuously buy on the cheaper exchange and sell on the higher-priced one whenever the price gap appears.

Manual trading is more suited for capturing one-off inefficiencies. An example of this is hedge fund manager John Paulson, who personally profited $3.7 billion by betting against the US housing market in 2007, an exceptional, one-time event.

This study investigates the efficacy of various statistical and machine learning models in algorithmic trading, focusing on AAPL stock across two distinct periods: 2010–2018 and 2018–2024. The primary goal is to evaluate and compare the performance of statistical models like the Simple Moving Average (SMA) strategy, Random Walk Hypothesis (RWH), and frequency-based strategies with machine learning models, including Support Vector Machine (SVM), Gaussian Naïve Bayes (Gaussian NB), Logistic Regression, and Natural Language Processing (NLP) for sentiment analysis.

The adoption of big data analytics in finance has provided unprecedented opportunities to uncover complex relationships between market variables that traditional methods often overlook. By incorporating machine learning and advanced statistical techniques, this research seeks to highlight innovative trading strategies and their implications for modern financial markets.

**What is a Strategy?**

A strategy in algorithmic trading is a predefined set of rules or algorithms designed to identify trading opportunities, execute trades, and manage risk, typically based on market data, patterns, or signals.

For E.g:

* Asset A typically moves ahead of Asset B. We buy or short Asset B when we observe a movement in Asset A.
* We execute a trade swiftly (using automated systems) before the market has a chance to react to a news event.
* Stocks from the same country and industry often move in tandem. If one stock behaves unusually, we buy or short it, expecting its behaviour to revert to the norm.

A trading strategy is made up of two key components: 1) **Inefficiency Discovery** and 2) **Inefficiency Exploitation**.

**Inefficiency Discovery** refers to the process of identifying opportunities or inefficiencies within the market, where prices deviate from their expected value or behaviour due to various factors, such as mispricing, market imbalances, or information gaps.

**Inefficiency Exploitation** is the process of designing and executing trades to take advantage of these discovered inefficiencies. It involves determining the most effective way to act on the inefficiencies, such as choosing the right entry and exit points, position sizing, and managing risks to maximize returns.

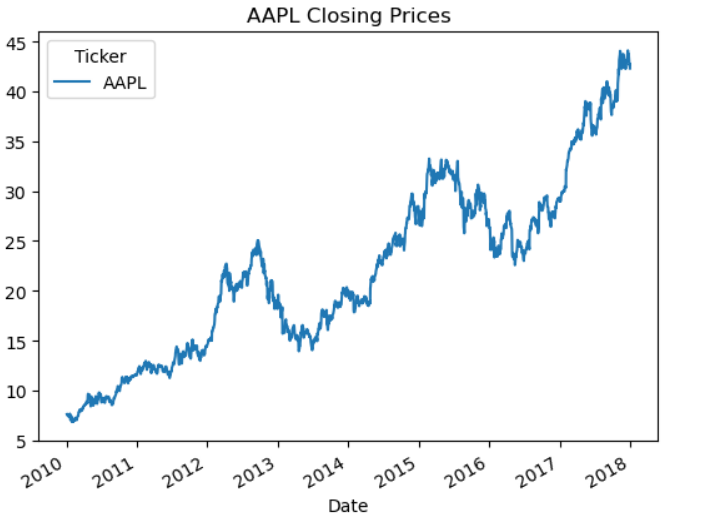
In a trading strategy, we may use algorithmic methods for discovering inefficiencies and then employ manual methods to exploit them, or alternatively, we could apply an algorithmic approach for both discovery and execution. This combination allows for greater flexibility in responding to market opportunities.

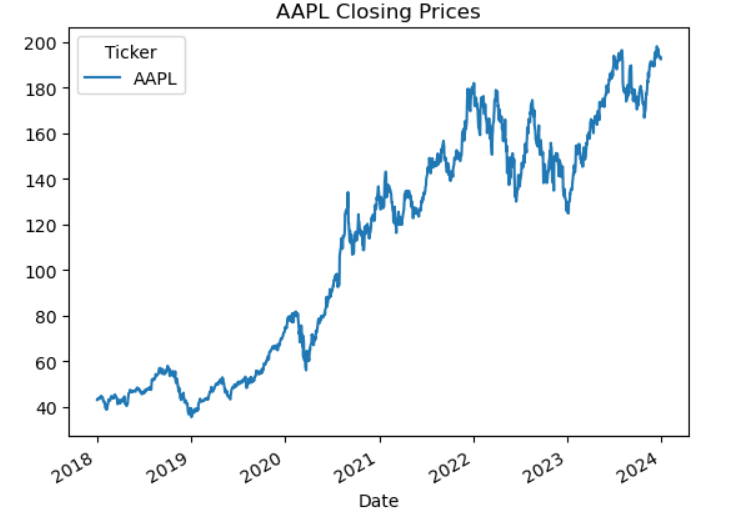
**Data Collection**

The dataset comprises daily AAPL stock prices, trading volumes, and related financial metrics from 2010 to 2024. Sentiment data was collected from Twitter and financial news platforms. The analysis is divided into two periods: 2010–2018 (pre-ML adoption) and 2018–2024 (post-ML adoption).

Between 2010 and 2018, Apple's stock (AAPL) experienced significant growth, driven by the company's expansion into new markets and the introduction of innovative products. In 2010, the stock price was approximately $30 per share, and by the end of 2018, it had risen to around $40 per share. This period saw Apple's market capitalization surpass $1 trillion in August 2018, marking a historic milestone. (Kerin,2024)

2018-2024: This period was marked by increased market volatility, driven by events such as the COVID-19 pandemic, inflation concerns, and swings in the tech sector. The market experienced more frequent short-term fluctuations compared to the previous period. In such volatile environments, lagged returns (historical returns) may be more effective in capturing market patterns, such as mean-reversion (when prices return to their long-term average after a deviation) or momentum effects (when an asset's price continues in the same direction after a trend). These factors made the stock price movements less predictable and more reactive to external shocks and broader economic trends. (Sun,2021)





**Traditional Financial Strategies**

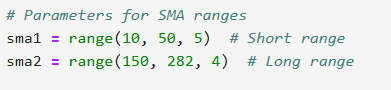
**Simple Moving Average (SMA)**

The SMA is a widely used technical analysis tool that identifies market trends by calculating the average price over a specific period. (Fama, 1970) introduced the Efficient Market Hypothesis (EMH), which challenges the predictive validity of SMA by asserting that market prices already reflect all available information. However, (Brock et al.,1992) demonstrated that SMA-based strategies can yield profitable outcomes in less efficient markets.

The **SMA (Simple Moving Average)** strategy refers to the use of a moving average as a technical indicator to smooth out price data over a specified period, typically to identify trends and potential trading signals.

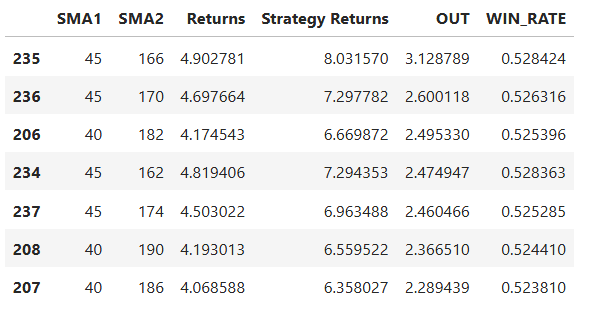
* **Buy Signal**: When a shorter-term SMA (e.g., 30-day) crosses above a longer-term SMA (e.g., 100-day), indicating a potential upward trend.
* **Sell Signal**: When the shorter-term SMA crosses below the longer-term SMA, signalling a potential downward trend.

In the code, we created multiple SMA windows and ran vectorized back testing on it, to find the maximum profit yielding strategy



The same strategy was applied to two different data sets belonging to 2010-2018 and 2018-2024 AAPL stock data set.

As expected, The **SMA strategy** which is a trend-following approach that works well in markets with sustained growth, performed better than Market Returns for Apple’s stock from 2010-2018.



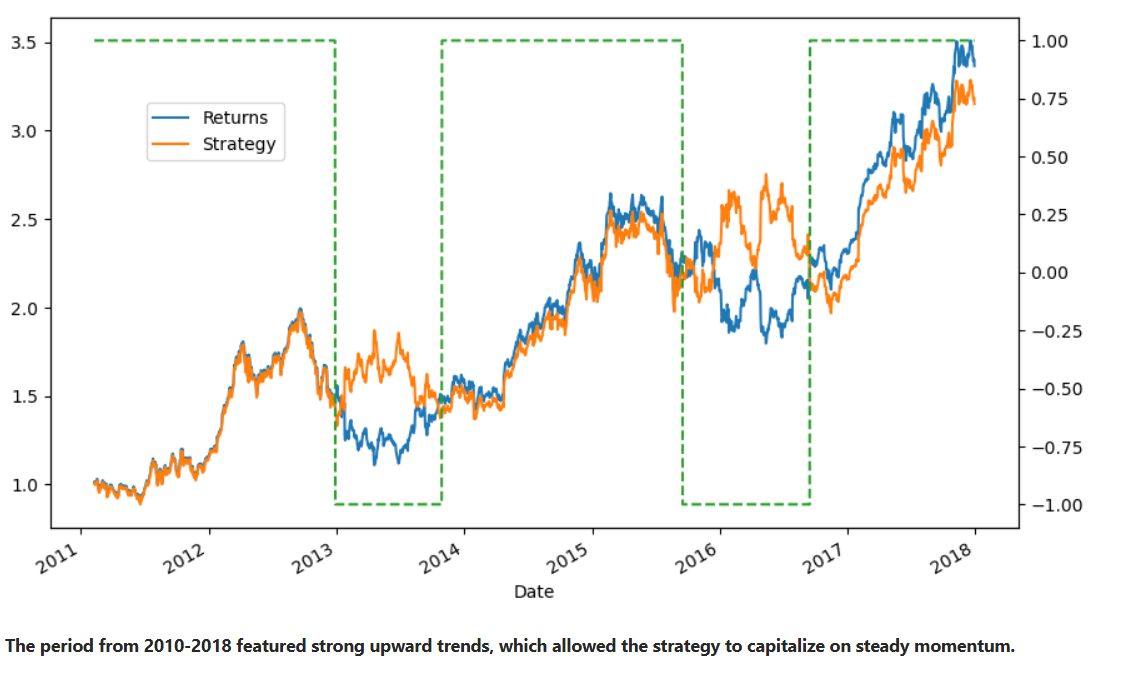
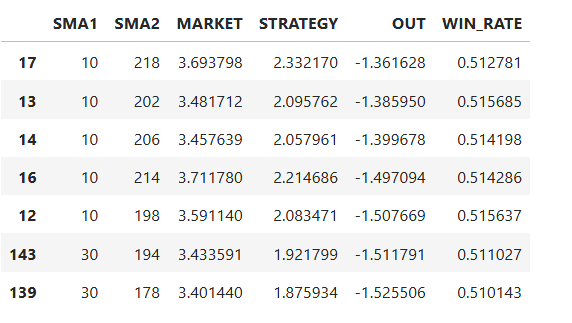


Figure shows that for the short-term Rolling window =45 and long-term Rolling window = 166, the SMA strategy performance significantly outperforms the Market Returns. This is simply because he **SMA strategy** is a trend-following approach that works well in markets with sustained growth, like Apple’s stock from 2010-2018. During the period from 2010-2018, Apple experienced several extended uptrends, particularly driven by major product releases (iPhone, iPad, etc.) and its growing market share. The strategy works well in these trends by entering when a short-term trend (SMA1) crosses above a long-term trend (SMA2), indicating the start of an uptrend, and exiting when the short-term trend crosses below the long-term trend, signalling a potential reversal or slowdown.

However, for the 2018-2024 time frame, the SMA strategy significantly underperforms as compared to the market returns.





The **SMA strategy** underperformed Apple stock from 2018-2024 due to increased market volatility, including events like the COVID-19 pandemic, which led to false buy and sell signals. The strategy struggled in a market characterized by frequent short-term reversals and choppy movements, unlike the sustained trends from 2010-2018. As a lagging indicator, the SMA was slow to react to rapid shifts, missing early gains or exiting too late. Additionally, Apple’s stock experienced momentum and sharp reversals that the SMA strategy failed to capture effectively, while external factors like product innovation and macroeconomic changes influenced growth in ways that didn’t align with the SMA’s signals.

Critically, SMA's reliance on historical data renders it less effective in rapidly changing market conditions. Its lagging nature results in delayed buy and sell signals, often missing critical market reversals. Moreover, SMA's predictive power diminishes in highly efficient markets dominated by algorithmic trading, as evidenced by recent studies highlighting its declining profitability.

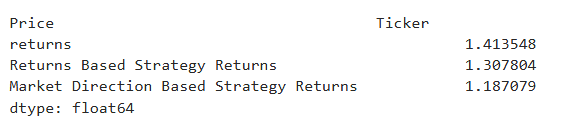
**Random Walk Hypothesis (RWH)**

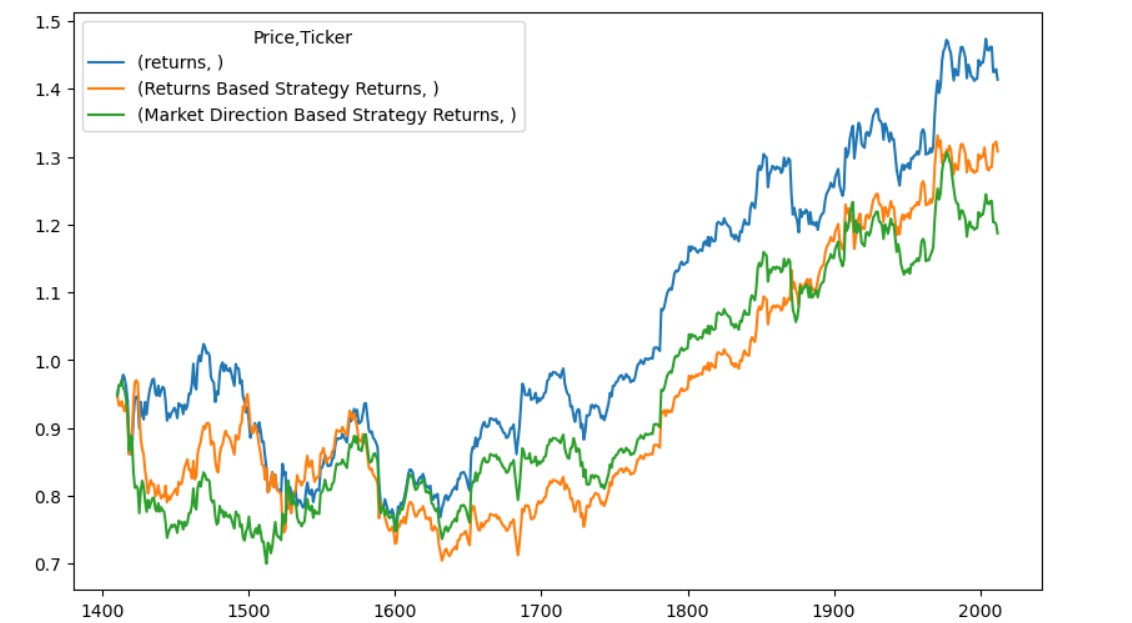
The RWH posits that stock prices follow a stochastic process, making future movements inherently unpredictable (Malkiel, 1973). This concept serves as a cornerstone for assessing market efficiency. Despite its theoretical importance, empirical evidence reveals significant limitations. For instance, anomalies like momentum and mean reversion challenge the RWH, suggesting that markets are not entirely random.

The Random Walk Hypothesis (RWH) and the Efficient Market Hypothesis (EMH) are consistent in suggesting that stock prices fully reflect all available information. RWH asserts that stock price changes are random and unpredictable, meaning future movements cannot be forecasted from past trends. Similarly, EMH claims that markets are efficient, and asset prices incorporate all relevant information at any given time, making it impossible for investors to consistently outperform the market. Both theories emphasize that market prices adjust quickly to new information, rendering it impossible to achieve superior returns based on public or historical data.

The second strategy applies the **Random Walk Hypothesis (RWH)** to market price prediction. In this approach, a financial time series of historical market prices is utilized to generate several lagged versions of the data. These lagged prices represent the historical values from previous days. Then, **Ordinary Least Squares (OLS) regression** is employed to model the relationship between the current market price and these lagged values. The fundamental premise of this strategy is that the market prices from a few days prior such as 4 days before can provide useful information to predict the price of the market today, leveraging past price movements to forecast future ones. Two strategies are created, Return based and Market Direction Based.

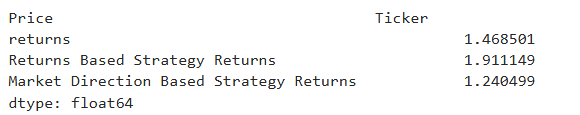
Interestingly, for the 2010-2018 period, both strategies underperform as compared to the market returns, when used with a 5-day lag.





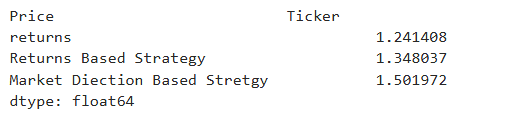
In calmer periods, like 2010-2018, where market volatility is lower, stock returns tend to exhibit less correlation across different lags. This can indeed limit the utility of lagged features in models like OLS regression for several reasons like, lagged returns might not offer predictive power because past performance doesn’t strongly influence future price movements. During calmer periods, markets might not display strong momentum or mean-reversion tendencies that are often captured through lagged returns. Without these trends, using past returns as predictors becomes less effective, as the relationship between lagged returns and future returns weakens. (Yieldstreet,2023)

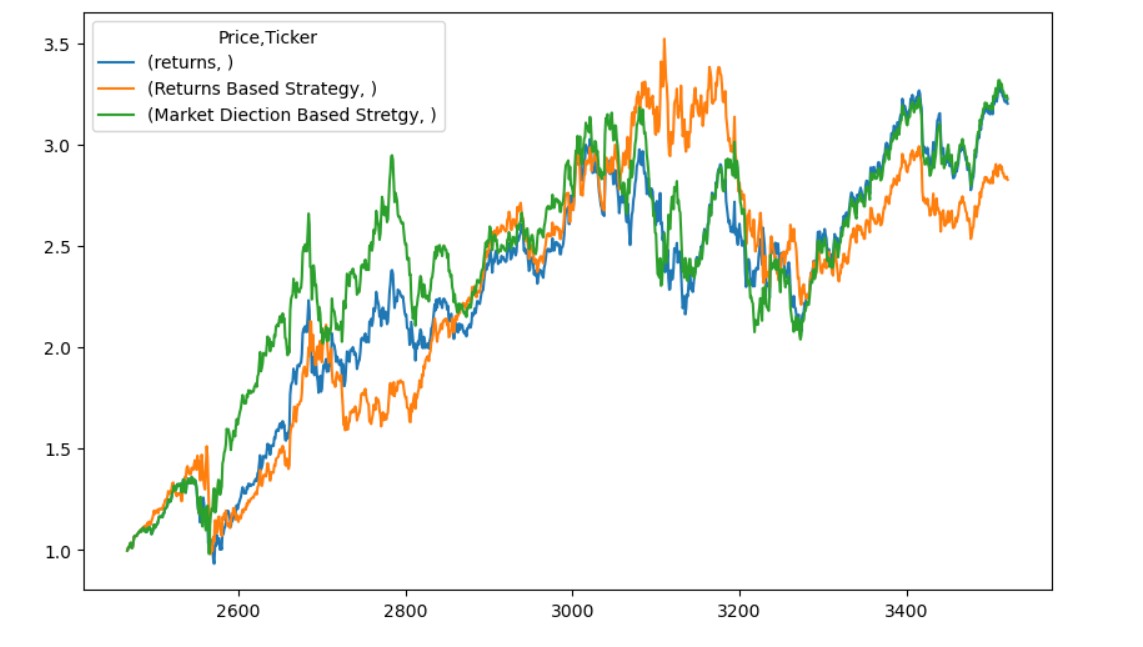
However, when a 10-day lag is used instead of 5-days, the Returns Based OLS regression strategy improves in performance



This happens because A longer lag (10 days) gives the model more time to capture trends and patterns that might not be visible in the short term. This helps in incorporating a broader range of information (such as market cycles, momentum, and past performance) that could improve prediction accuracy. While the direction-based strategy underperforms due to its simplicity and inability to capture the full range of market information. Direction is too limited as a predictor compared to the richer data provided by actual returns, making it harder for a direction-based strategy to outperform the market.

The 2018-2024 window yields better results for RWH based, OLS regression strategies.





The period from 2018 to 2024 includes events like the COVID-19 pandemic and heightened geopolitical tensions, which increased market volatility. Higher volatility often leads to stronger trends and patterns, making lagged features (both returns and direction) more predictive. During volatile or trend-driven periods, stocks like Apple may exhibit more pronounced momentum (e.g., persistent upward trends) or mean reversion (e.g., corrections following extreme movements). These behaviours can be captured more effectively by regression models using lagged returns or directions even with a small lag window of 5 days. (Roszyk & Ślepaczuk,2024)

Additionally, the RWH oversimplifies market dynamics by ignoring external factors such as macroeconomic events and behavioural biases. Critics argue that while the hypothesis provides a useful benchmark, it fails to account for the complex, nonlinear relationships that often drive market behaviour.

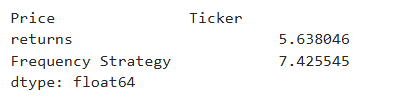
**Frequency-Based Strategies**

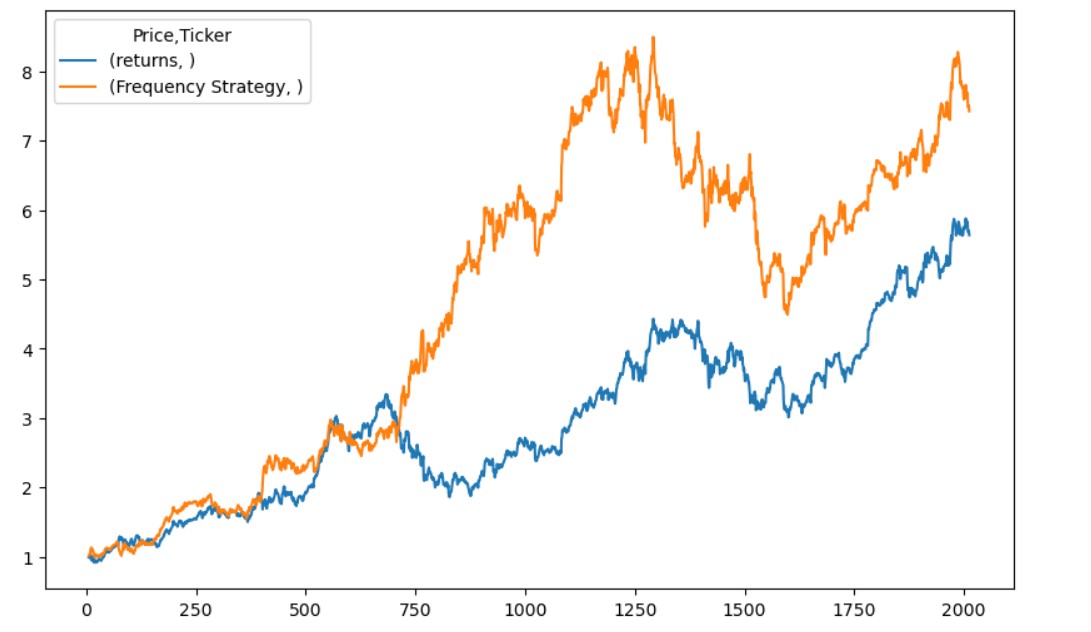
The frequency approach to market prediction involves converting continuous market data into binary outcomes, such as 0 or 1, to represent different market scenarios. By analysing historical data, this method calculates the probability of upward or downward movements for each scenario, offering a straightforward probabilistic framework for predicting market direction without relying on complex models.

This approach is related to techniques used in binary options trading, where traders make directional bets on market movements. Binary options are financial instruments that allow traders to speculate on the price movement of an underlying asset, with a fixed payout if the prediction is correct and a loss if it is not. Strategies in binary options trading often involve analysing market trends and patterns to make informed predictions. (Liss, 2024)

In the Third Strategy, the frequency approach streamlines market predictions by transforming continuous data into binary outcomes (0 or 1). Each binary pair (e.g., (0,0), (0,1)) corresponds to a distinct market scenario. Using historical data, it calculates the probabilities of upward or downward movements for each scenario, providing a straightforward probabilistic framework for forecasting market direction without relying on complicated models. (Hilpisch, 2019)

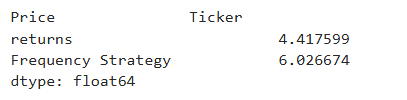
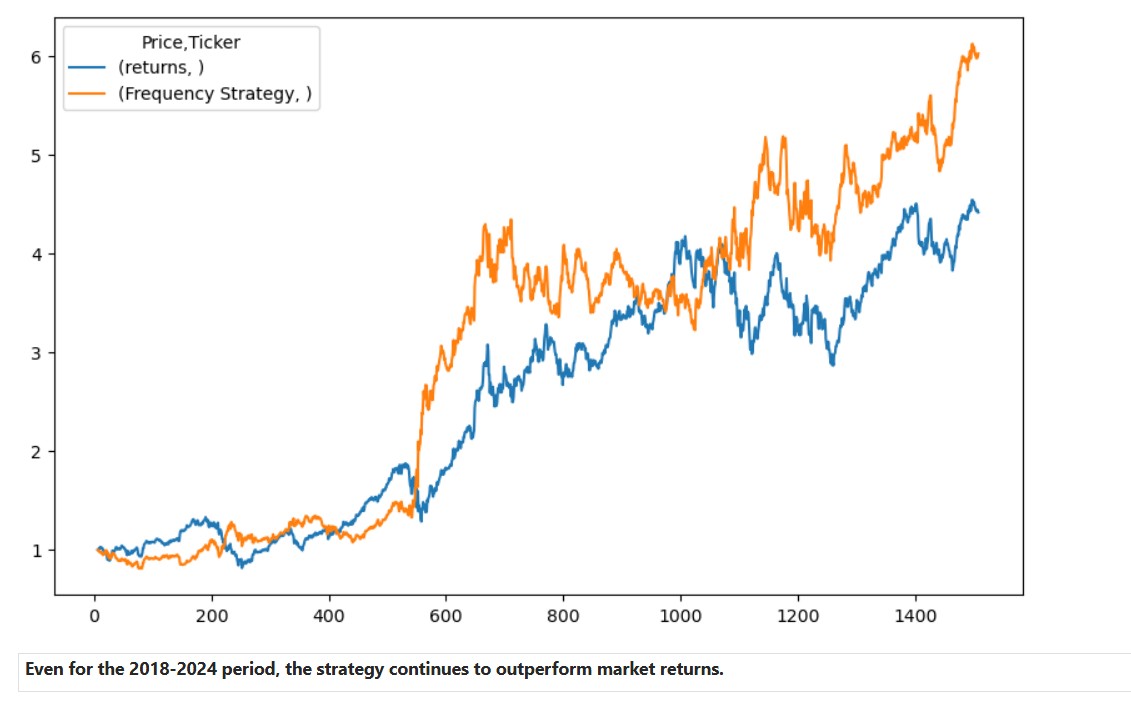
The Strategy significantly outperforms the Market Returns, for the 2010-2018 window.





Apple's stock during this period showed strong upward trends driven by consistent innovation, product launches, and financial performance. The approach could capture these patterns effectively by focusing on historical probabilities. Moreover, the binary conversion and scenario-based analysis might excel in exploiting short-term price volatility, identifying profitable opportunities that traditional models overlook. By reducing continuous data to binary outcomes, the method filters out minor fluctuations, focusing on meaningful market movements. This can result in more precise and actionable predictions.

Even for the 2018-2024 period, the strategy continues to outperform market returns.

****

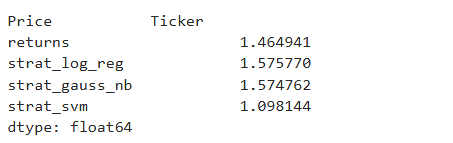
This is Because Apple maintained a strong growth with the expansion of its services ecosystem (e.g., Apple Music, iCloud, App Store), product diversification (e.g., wearables like Apple Watch), and flagship device upgrades. These consistent patterns make historical probabilities effective predictors of market direction. Apple's status as a market leader likely resulted in predictable investor confidence and stable upward trends, which the frequency approach could exploit by identifying recurring patterns in stock movements. While the Volatility increased, the binary nature of this strategy enabled it to quickly adapt to these fluctuations and identify profitable opportunities. (Sazuka, 2018).

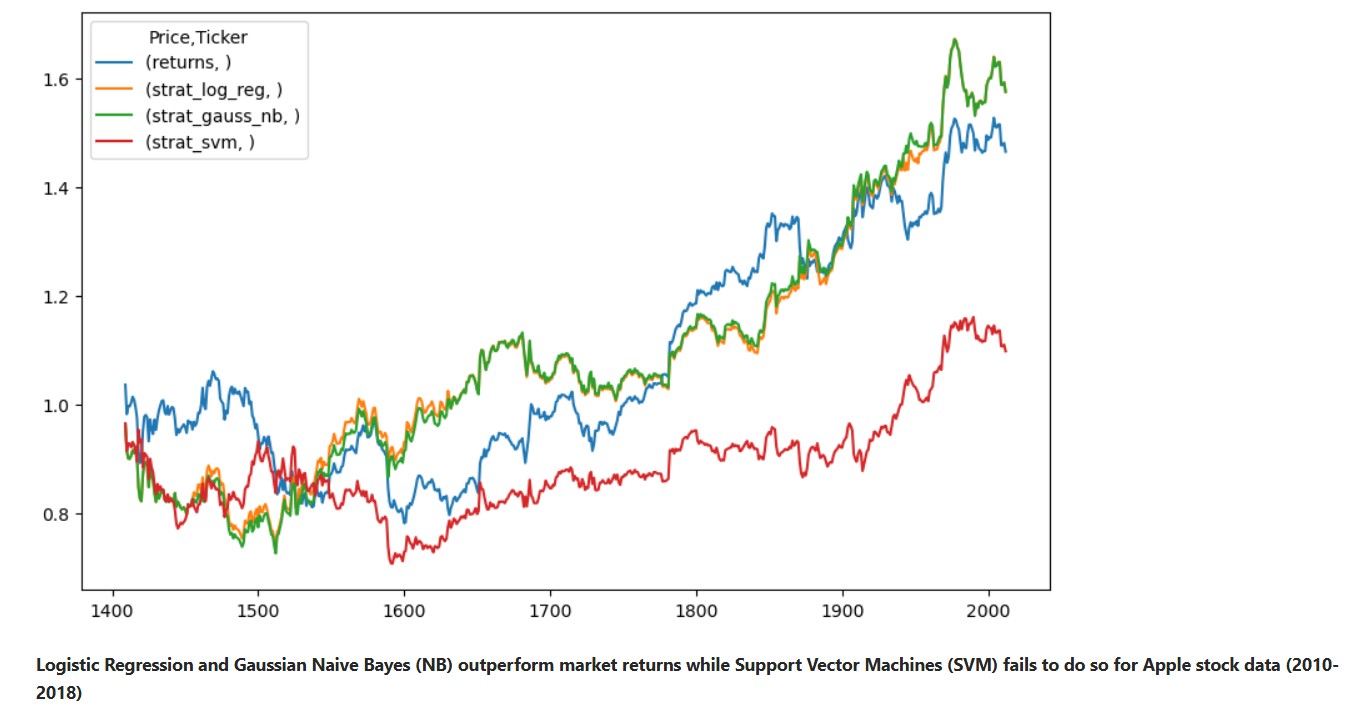
Despite the positive performance results, the widespread adoption of frequency-based strategies has raised concerns about market stability. High-frequency trading has been linked to increased volatility and flash crashes, undermining investor confidence. Furthermore, the intense competition among algorithmic traders has led to diminishing marginal returns, challenging the sustainability of such strategies.

**Big Data and Machine Learning Approaches**

Machine learning has greatly improved predictive modelling and strategy optimization in algorithmic trading. Support Vector Machines (SVM) and Logistic Regression are effective at classifying market trends based on historical data, learning complex patterns to predict future movements (Chen et al., 2023). Gaussian Naive Bayes (NB), though simpler, provides a probabilistic approach to trend prediction, assuming feature independence and offering insights into market behaviour (Zhang, 2004). These techniques help optimize trading strategies and enhance decision-making by adapting to market changes.

For the 2010-2018 time period we noticed a curious outcome, Logistic Regression and Gaussian Naive Bayes (NB) outperform market returns while Support Vector Machines (SVM) fails to do so for Apple stock data (2010-2018)



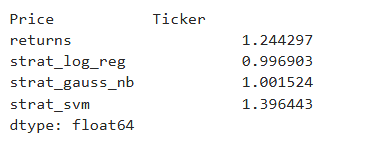


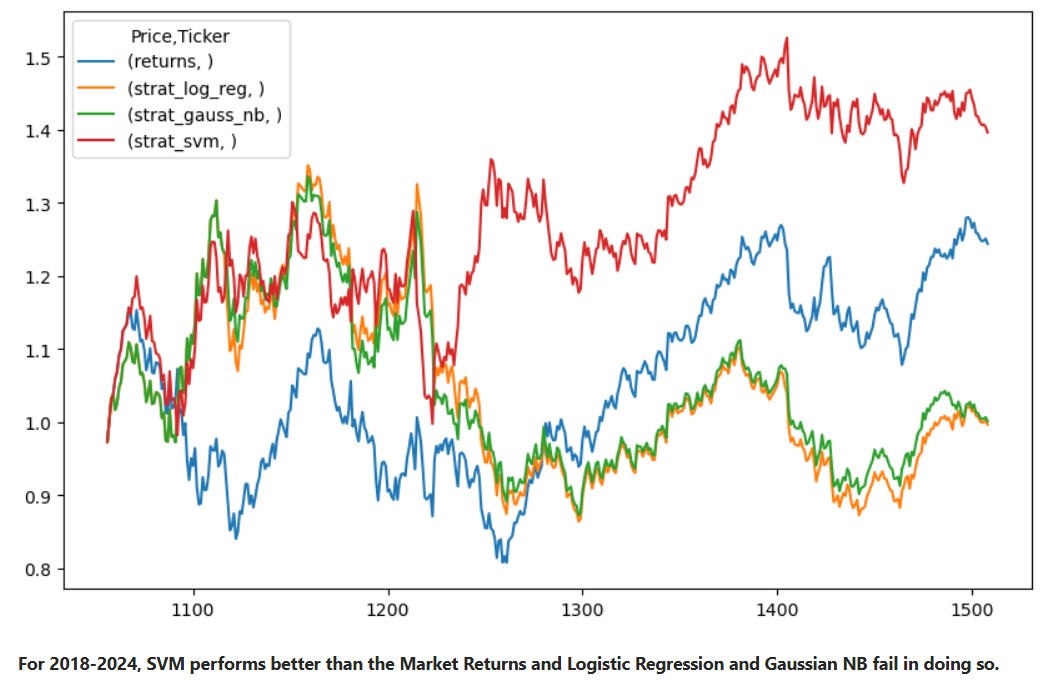
**Logistic Regression** and **Gaussian NB** are simpler models that make specific assumptions about the data. Logistic Regression assumes a linear relationship between the features and the target, while Gaussian NB assumes that the features are independent and follow a Gaussian distribution. These assumptions align well with the structure of the stock data, especially when market trends follow relatively predictable, linear patterns.

**SVM**, on the other hand, is more complex and involves finding a hyperplane to separate different classes, which may not always work well with financial data that is noisy and non-linear. If the stock data is not linearly separable or if it has too much noise, SVM's performance degrades. It is also prone to overfitting, especially with small or noisy datasets like stock market data.

Essentially Apple stock data from 2010-2018 exhibited more consistent patterns or linear trends, which would favour models like **Logistic Regression** and **Gaussian NB** that perform well under such conditions, while **SVM** is a powerful algorithm, its complexity and sensitivity to hyperparameters and noise explains why it struggled to outperform market returns on Apple stock data. (Ullah et al., 2023)

For 2018-2024 data, a complete inverse trend in noticed, SVM performs better than the Market Returns and Logistic Regression and Gaussian NB fail in doing so.





As has been seen before, 2018-2024 was a time of high volatility, such irregularities often involve data patterns that are non-linear, noisy, and difficult to capture using simpler models like **Logistic Regression** or **Gaussian NB**. **SVM**'s ability to handle such volatility and find more complex decision boundaries helps it outperform the simpler models, which may have struggled with the changing dynamics of the stock market.

Despite their advantages, traditional machine learning models face limitations in handling high-dimensional financial data. The curse of dimensionality and susceptibility to overfitting often hinder their predictive performance. Additionally, these models struggle to capture the complex, nonlinear dependencies characteristic of financial markets, necessitating the adoption of more advanced techniques.

**Conclusion**

This study compared traditional statistical models and machine learning techniques for algorithmic trading using AAPL stock data from 2010-2018 and 2018-2024. The performance of each model varied with market conditions. In the stable 2010-2018 period, traditional strategies like Simple Moving Average (SMA) and machine learning models such as Logistic Regression and Gaussian Naive Bayes (NB) performed well, as the market followed consistent trends. However, in the volatile 2018-2024 period, SMA underperformed due to frequent market reversals, while Support Vector Machines (SVM) outperformed simpler models due to its ability to handle complex, non-linear data. The underlying idea for all these strategies is simple, to find predictive value, analyse it and to derive some sort of useful information to strengthen the investors decisions.

To achieve more robust and adaptable trading strategies, a hybrid approach is recommended. By combining the strengths of various models such as using trend-following strategies like SMA during stable growth periods, and leveraging complex models like SVM during volatile conditions a more comprehensive and flexible algorithmic trading system can be developed. Additionally, integrating external data sources like sentiment analysis and enhancing risk management practices would further improve the predictive power and profitability of these strategies. This hybrid approach would allow for more effective discovery and exploitation of market inefficiencies across different market environments, ultimately optimizing trading outcomes.

Bibliography

BROCK, W., LAKONISHOK, J., & LeBARON, B. (1992). Simple Technical Trading Rules and the Stochastic Properties of Stock Returns.

Chen, J., Wen, Y., Nanehkaran, Y., Suzauddola, M., Chen, W., & Zhang, D. (2023). *Machine learning techniques for stock price prediction and graphic signal recognition.*

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work.

Hilpisch, Y. (2019). Algorithmic Trading. In Y. Hilpisch, *Python For Finance.*

KERIN, J. (2024). *https://www.investopedia.com/articles/stocks/12/history-apple-stock-increases.asp?* Retrieved from https://www.investopedia.com.

Liss, R. (2024). *https://www.investopedia.com/articles/active-trading/052014/binary-options-strategies.asp?* Retrieved from https://www.investopedia.com.

Malkiel, B. G. (1973). A Random Walk Down Wall Street.

Roszyk, N., & Ślepaczuk, R. (2024). *The Hybrid Forecast of S&P 500 Volatility ensembled from VIX, GARCH and LSTM models.*

Sazuka, N. (2018). *Analysis of Binarized High Frequency Financial Data.*

Sun, W. (2021). The Changes of Apple’s Stock Price During the.

Ullah, A. K., Ihsan, M. U., Imtiaz, F., Alam, M. R., & Majumdar, M. (2023). *Combining Machine Learning Classifiers for Stock Trading.*

YieldStreet. (2023). *https://www.yieldstreet.com/blog/article/past-performance-not-indicative-of-future-results/*. Retrieved from https://www.yieldstreet.com.

Zhang, H. (2004). *The Optimality of Naive Bayes.*