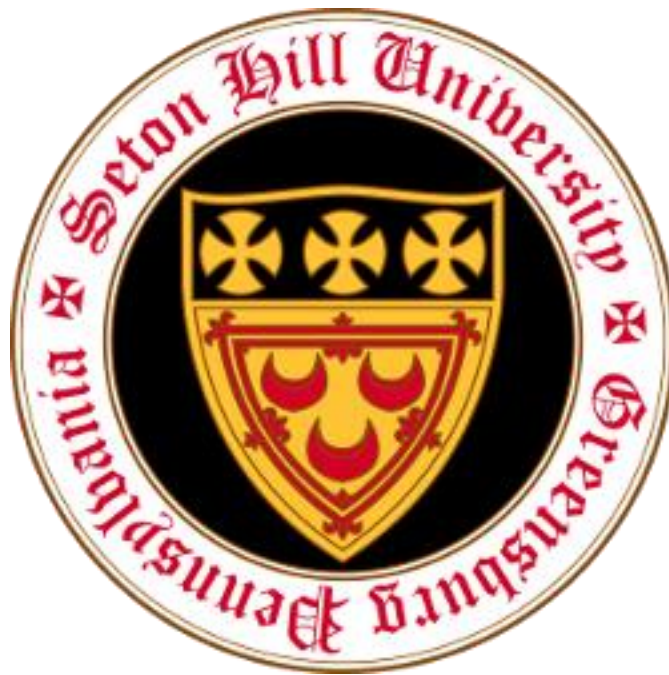


Senior Capstone Project

Evaluating Success of Technical Analysis Investment Strategies



Name: Benjamin Nicholson
School: Seton Hill University
Advisor: Dr Jared Burns

Abstract

The Senior Capstone Project is an integral part the Data Science Major at Seton Hill University. Under the guidance of Dr. Jared Burns I have completed a comprehensive investigation into how data science is currently transforming the financial sector. ‘Enhancing Technical Analysis (TA) with Data Science’ looks to incorporate data science techniques such as stochastic modeling, machine learning and mathematical modeling to understand expected returns in the US stock market. The evaluation includes backtesting across different macroeconomic cycles while utilizing sector Exchanged Traded Funds (ETFs) against passive buy and hold strategy. It was found that TA struggled to outperform baseline buy & hold strategies in periods of strong economic growth. However it was clear that during weak economic periods, TA outperformed buy & hold strategies, upwards of 10%. The findings from the investigation lay the foundation for advanced data science techniques such as reinforcement learning and neural networks which will look to uncover underlying patterns in the data, leading to more powerful predictive models. The accumulation of market knowledge and analysis of existing TA strategies will allow for a meaningful contribution to understand how data science can continue to contribute towards a changing financial industry.

Introduction

The advancements in modern computational power, alongside advancements in artificial intelligence/machine learning has seen drastic changes across the financial industry. Financial investors such as investment banks and financial tech companies are in constant battle for a competitive edge over one another. The traditional approach look to maximize returns through market knowledge, however the recent advancements in statistical data science has shifted how financial investors compete. The typical fund manager of the past has become machine intelligence which can see patterns far beyond humans understanding, paving the

way for an evolutionary change to the finance industry. An example of this is the CEO of Blackrock who said ‘the days of relying solely on intuition or market savvy are behind us,’ pointing out the need to shift operations to continue to gain a competitive edge. This emerging field gives me the opportunity to be a pioneer which excites me greatly to complete a project with such a concentration.

Purpose

Utilizing sector ETFs across a range of macroeconomic cycles, what data science optimized TA technique yields the highest return across the US stock market.

Financial Background

The financial data from the US market is free and widely available to those who desire. The nature of the market is meant to be free and widely available so that those interested can make informed decisions. As a result companies release financial data such as financial statements, income statements and other essential information that investors can use to make informed decision. It is also easy to access the price of a stock through the NASDAQ or US Stock Market which has live updates based on demand and supply.

While financial data is essential for all investors, it is specifically used for investment strategies that incorporate algorithmic trading. Some of these include fundamental analysis, technical analysis and High Frequency Trading (HFT).

Fundamental analysis looks to identify inefficient prices in equity prices such as a stock or an ETF in the market. An inefficient price is one which either over values or under values a company which can allow for investment opportunities. To understand what the price of an equity should be, an analysis into a companies financial data can help inform predictive

models to find what the 'true' price of an equity should be. Data science techniques could be multiple linear regression, clustering or neural networks.

Another type of investment that has become increasingly used is called High Frequency Data. This is where the time scale of investment can be as small as a few seconds, thereby thousands or even millions of actions can be made on a single day. HFT has a massive emphasis on advanced statistics and mathematical modeling. There is little to no concentration on the type of investment being made, rather it is simply about identifying which price movements are most likely to move upwards or downward. HFT requires live environments for testing, allowing for data science techniques such as reinforcement learning, these give real time feedback to the machine, rewarding it for a return on actions made.

While fundamental analysis looks heavily on domain knowledge and HFT is more based on mathematical patterns, Technical Analysis finds a balance between the two. TA is the use of historical price and volume data to make predictions about the movement of the market. TA was first established in the United States in the late 19th century by Charles Dow but has continued to evolve since. With the availability of modern computational power increasing in the 1960's and 1970's, there was massive improvements made in. There was now the opportunity to backtest models on historical data you could simulate the performance of different strategies before implementing them into the market. In the modern day, it is easier than ever with companies such as Robinhood offer no brokerage fee which allows for trading cadence to increase without the risk of losing profits due to brokerage costs.

The investment strategies discussed assume that misprices / inefficient prices are possible in the financial market. Remember the financial system is supposed to allow for all financial data to be freely available, which would lead the *Efficient Market Hypothesis* (EMH) to be true. It essentially states that an equities price is always perfect and any new information will be represented in an appropriate price change. Suggesting that it is impossible to beat the market, deeming TA futile. This project looks to demonstrate TA should allow one to beat the market, proving that inefficient pricing is possible.

Sector ETFs

ETFs or Exchanged Traded Funds are a type of investment that looks to represent the accumulation of a specific index or economic subsection, for example a sector ETF. By taking a basket of stocks within a certain sector you reduce the emphasis on a singular company and concentrate on movement of the entire sector. A sector is a component of the economy which contributes to the overall market, such as healthcare, real estate or materials. The sector is usually determined by one of GICS, the Global Industry Classification Standard which is an industry wide identification of sectors within the economy. The following GICS industries and ETFs representing them are below:

Sector ETFs:

XLB - Materials

XLI - Industrials

XLF - Financials

XLK - Information Technology

XLY – Consumer Discretionary

XLP - Consumer Staples

XLE - Energy

XLV - Healthcare

VOX - Communication Services

XLU - Utilities

IYR – Real Estate

Macroeconomic Cycles

The economy goes through periods of highs and lows, and understanding the dynamics between these shifts is crucial. Different investment strategies yield higher returns depending on the state of the market. These are defined by an expansion, peak, contraction and a trough which have different characters as seen below. Different macroeconomic cycles have different properties and by backtesting across a whole variety of environments can make for a better understanding of which time period is best for TA. To achieve this goal the following macroeconomic periods have been created from the past 20 years of US stock data:

Economic Period	Start Date	End Date	Level of economic activity
Expansion	2012/01/01	2015/01/01	Increasing economic activity, consumer spending and confidence rises
Peak	2019/06/01	2020/02/01	Maximum economic activity, interest rates rise
Contraction	2007/12/01	2008/10/01	Decreasing economic activity
Trough	2008/10/01	2009/06/01	Lowest economic activity, low interest rates, low consumer confidence
All Data	2005/01/01	2024/06/01	Multiple market movements captured

Technical Analysis Techniques

TA is a term to describe investment strategies that use historical prices to signal buy, sell or hold signals. TA is a genre of investing covering hundreds of strategies so I am going to use some of the most popular techniques to evaluate the performance as a whole.

1. Bollinger Bands

The Bollinger Bands was created by John Bollinger in the 1980's. The goal is to measure market volatility and utilize moving averages to understand whether a stock is overbought (price is high) or oversold (price is low), signalling buy or sell signals.

Bollinger Band technique creates the following parameters:

- Middle band: The 20 day moving average
- Upper band: The middle band plus 2 standard deviations of the current moving average
- Lower band: The middle band minus 2 standard deviations of the current moving average

If the price moves above the upper band then a sell signal is made suggesting the price is too high, while if the price goes below the lower band then a buy signal is made suggesting the stock is too low.

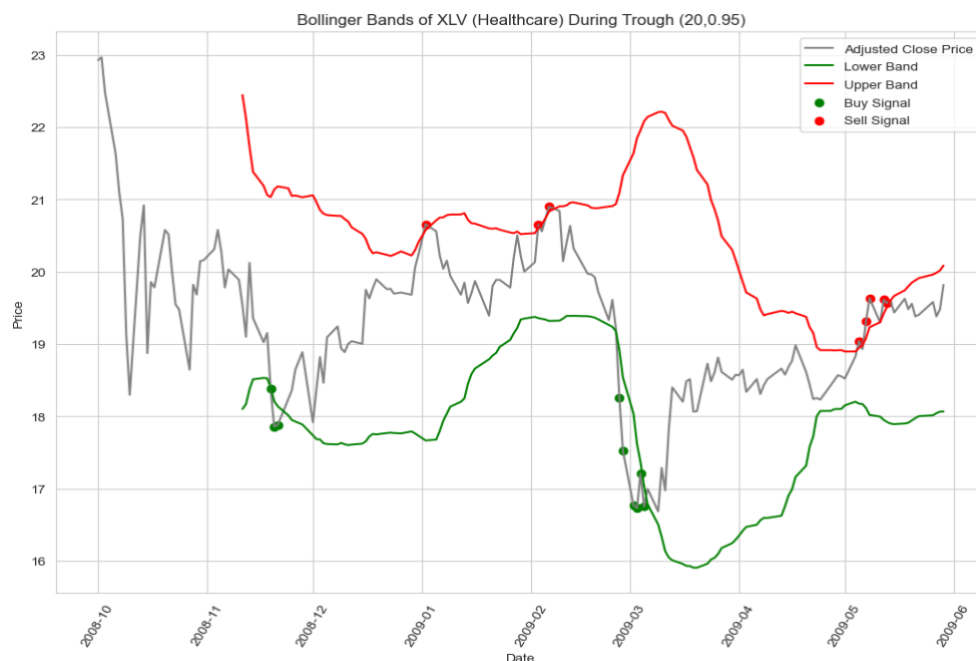


Image 1: Bollinger Bands of Healthcare Sector during Trough (20, 0.95)

2. Relative Strength Index (RSI)

RSI identifies the momentum of a price movement, similar to Bollinger Bands as it looks to identify when equities have been overbought or oversold. The RSI is a moving oscillator and falls between a value of 0 and 100. It is typically plotted below the line of an equity's price to get an overview of the movement of the stock. An asset is overbought when the value is greater than 70 which implies a sell signal and an asset is oversold when the value is less than 30, triggering a buy signal.

RSI parameters are created by the following:

1. Difference the data, by comparing the price at day n to day $n + 1$
2. Track average gain/loss where gain is the positive movement from day to day and loss is the negative movement
3. Create the Relative Strength (RS) via the average gain/average loss
4. Create the RSI value by $100 - (100/(1+RS))$



Image 2: RSI for Healthcare Sector during Trough

3. Moving Average Coverage Divergence (MACD)

The MACD identifies trends and changes in momentum by analyzing the relationship between two parameters. The first parameter/line, is referred to as the MACD line which is found by looking at longer term EMA (typically 26) / shorter term EMA (typically 12). The second line is referred to as the signal line which is typically a 9 day EMA of the MACD line. It can suggest the following:

- Bullish Signal: When the MACD Line crosses above the Signal Line, it can indicate upward momentum, a possible buy signal.
- Bearish Signal: When the MACD Line crosses below the Signal Line, it can indicate downward momentum, a possible sell signal.

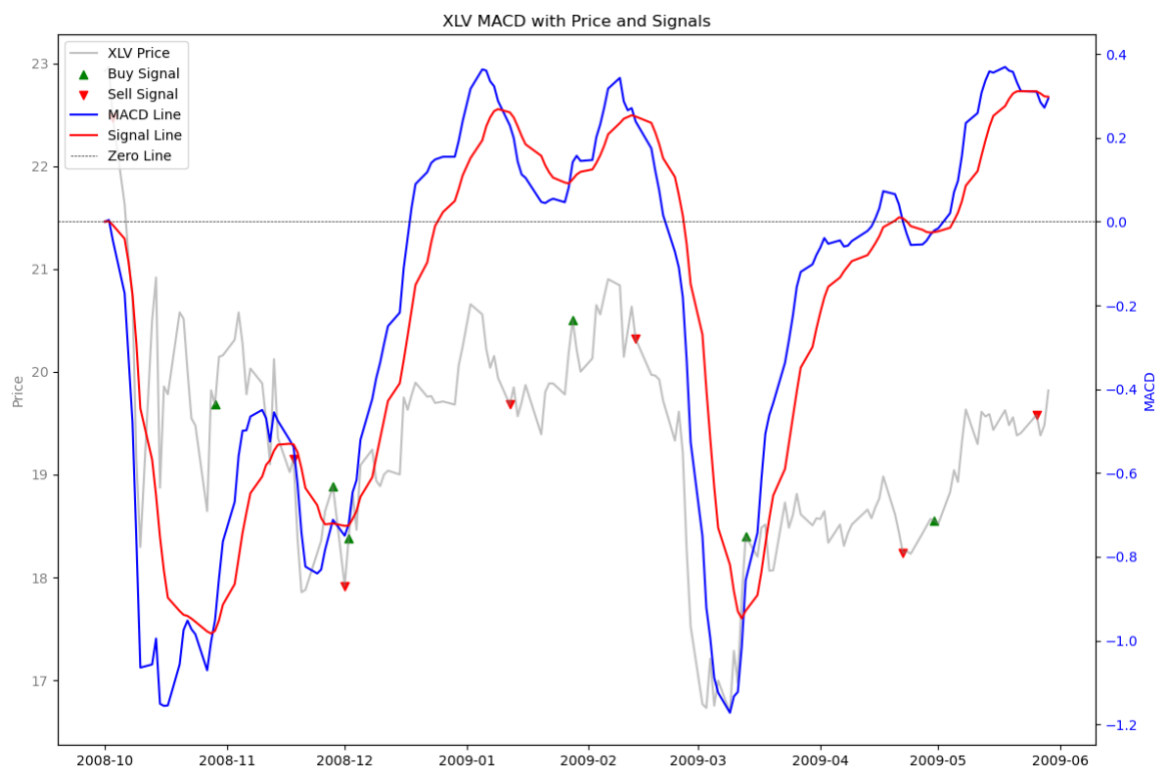


Image 3: MACD of Healthcare Sector during Trough

Backtesting methodology

To evaluate the success of TA there needs to be an extensive pipeline to take historical data, test TA method and then return an objective result.

Step 1. Pull historical data

The data that is required for this project is the adjusted close prices of stocks. While the volume traded usually belongs to fundamental analysis and open, low, close and high data might belong more to HFT. Therefore, historical data needs to be collected from the internet through an API which can bring the data locally to be analyzed. The YFinance API can be accessed locally through a Python IDE such as Visual Studio Code. YFinance provides real time and historical data on all essential financial information. By utilizing the YFinance API it allows for intensive backtesting on a local machine.

Step 2. Create Investment Period

The backtesting strategy for the project will use stochastic modeling to evaluate performance. By randomly choosing a 90 day investment period it creates a distribution of returns rather than a singular return. This is extremely important for stock investments as the time at which you enter or leave the market will likely have a high correlation with returns. To prevent this from causing inaccurate returns in the evaluation of TA strategies, using random sampling through stochastic modeling will create expected returns through the mean of the distribution.

Step 3. Create Parameters

The process of backtesting allows you to create the parameters needed to create the actions buy, sell or hold. The parameters are identified through analyzing historical data and creating

data based on the demands of the TA strategy, such as a rolling average for the Bollinger Band.

Step 4. Signal Creation

Utilizing the parameters, compare the adjusted closed price to the requirements of the TA strategy to signal a buy, hold or sell.

Step 5. Run simulation

Begin with the first day of the 90 day investment period and iterate through the signals to decide whether to buy, sell or hold. By comparing how the balance of cash and stock balance is at the end of the investment period with the start, it can return profit margins. This is repeated for 1000 90 day investment periods.

Step 6. Repeat steps 2-5 for each sector ETF across each macroeconomic cycle

By repeating the steps 2-5 for each sector ETF across each macroeconomic cycle it will allow for 2 major findings.

Firstly it will allow for an understanding of different sectors success using the different models. By looking at multiple movements through different sectors it is possible to understand what dynamics of an industry may encourage or diminish returns of TA.

Secondly by comparing how the performance different time periods it will aid in inferring what market dynamics reward TA.

Baseline test

To evaluate TA strategies a baseline method is going to be used to understand an industry standard of success. This will be a passive investment strategy. These strategies consist of purchasing a stock and holding on with no interference, allowing the movement of the economy to drive returns. Over the past 20 years this has been very successful as the growth in the US due to the tech market boom alongside other industries has been the driving force behind one of the strongest economy in the world. The passive investment strategy is very popular amongst most investors as it is lower risk, especially in a diversified portfolio, it is less stress as it does not require many actions to be made. The results of the baseline model should lay the foundation of comparison for the TA evaluations.

The methodology described prior creates a distribution of returns. By using randomness in the investment period it creates a list of 1000 different returns which can be crafted into a histogram.

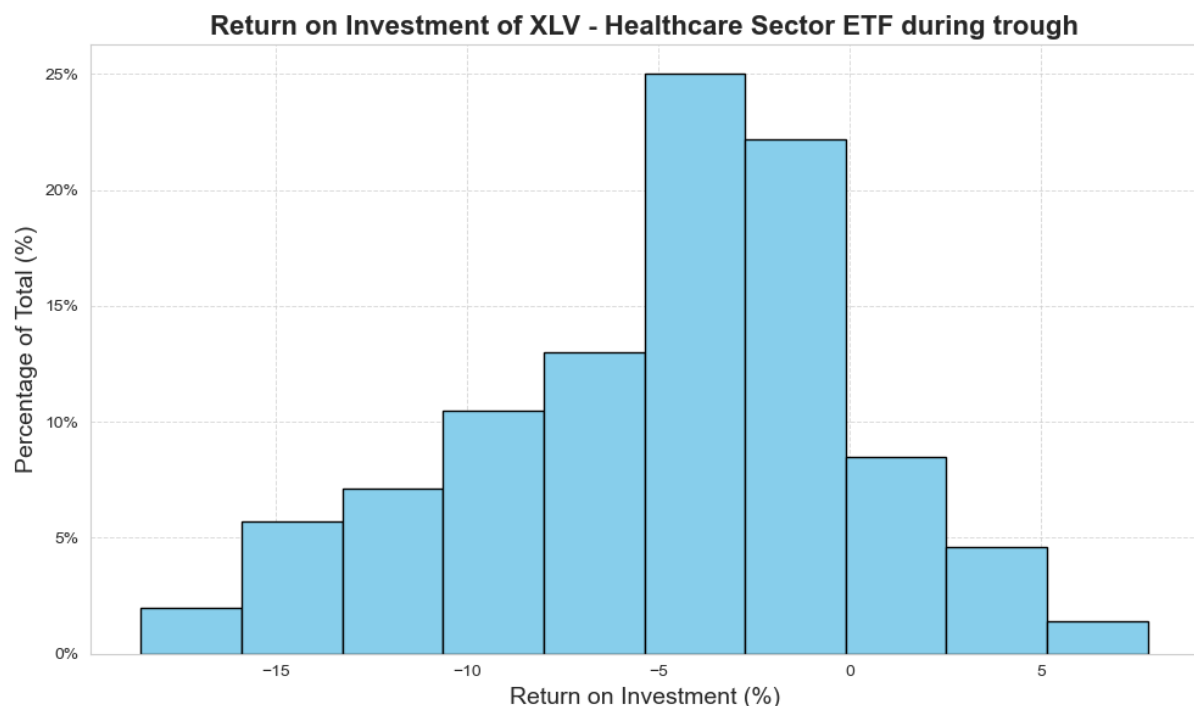


Image 4: Buy & Hold Healthcare Distribution of Returns

Image 4 shows the distribution of returns that have come from the buy and hold strategy. Observe that the distribution looks to be normally distributed around a mean of -2% return. This is expected of the healthcare sector as during the trough of the Global Financial Crisis most industries were in a deficit. It is interesting to observe that the expected return is normally distributed. This suggests that the returns for any given investment period could be as equally higher than or lower than the mean. Having the normal distribution can incorporate standard deviation to find a standard error on expected returns. The higher the tails are either side of the mean the lower the higher the risk. The normal distribution was found to be the case on almost all expected returns of the sector ETFs.

	trough	expansion	peak	contraction	all_data
XLB	-0.7610973	5.10224772	4.03005582	-0.1936342	3.79728184
XLI	-13.100474	6.55504437	6.81267623	-4.4371167	4.24395392
XLF	-21.387772	7.26148523	9.52855914	-15.437852	2.94366462
XLK	3.4761016	5.55051828	12.5958567	-2.6188525	5.78023523
XLY	-1.0104247	6.99860838	3.41100044	-4.1901145	4.5357148
XLP	-8.0759947	5.88652626	5.15358747	0.54166119	3.50875703
XLE	-4.8479924	3.89725745	-0.4590522	3.63243556	3.84319739
XLV	-4.7374016	9.12532985	8.43989242	-3.6049969	4.09325026
VOX	4.15607028	5.63322553	6.11301833	-5.1872052	3.42745425
XLU	-8.9184888	5.19492371	6.30886641	-2.4845293	3.23273705
IYR	-17.15656	4.44728482	4.30989818	-0.1692817	3.40310599
Average	-6.5785485	5.96840469	6.02221445	-3.1044987	3.89175931

Table 1: Buy & Hold Mean Returns

Table 1 has many implications. Firstly it shows that the expansion and peak time periods do correspond with far greater returns in Buy & Hold strategies in comparison with the trough and contraction periods. It is important to see that all data (time from 2005-2024) has a mean return of 3.9% with the greatest returns of 5.7% return being in the technology sector. This coincides with the background knowledge that the technology sector within the US has been the driving factor for the strong economic progress in the US financial market.

Another insight is done through comparing expected returns across different sectors, an economical understanding of the different sectors can be made. The financial, real estate and industrial sectors perform far below the average of -6.6% returns with the financial sector standing out with a 21% decrease on average in a 90 day investment period. This is expedited by the GFC which was taking place however it does show that finances follow the state of the economy with large oscillations as the expansion and peak show returns of 7.3% and 9.5% return. Where as some sectors such as materials and energy have far lower fluctuations in return. These are usually what are referred to as defensive industries as they are quite countercyclical to the movement of the economy. There is a full paper discussing a sector analysis in the US stock exchange in *source 1*.

Baseline Bollinger Band

The baseline Bollinger Band is the standard parameters described by John Bollinger. With a rolling average of 20 days and a confidence interval of 0.95 (1.96 standard deviation), it is the standard way of representing how Bollinger Bands yield returns.

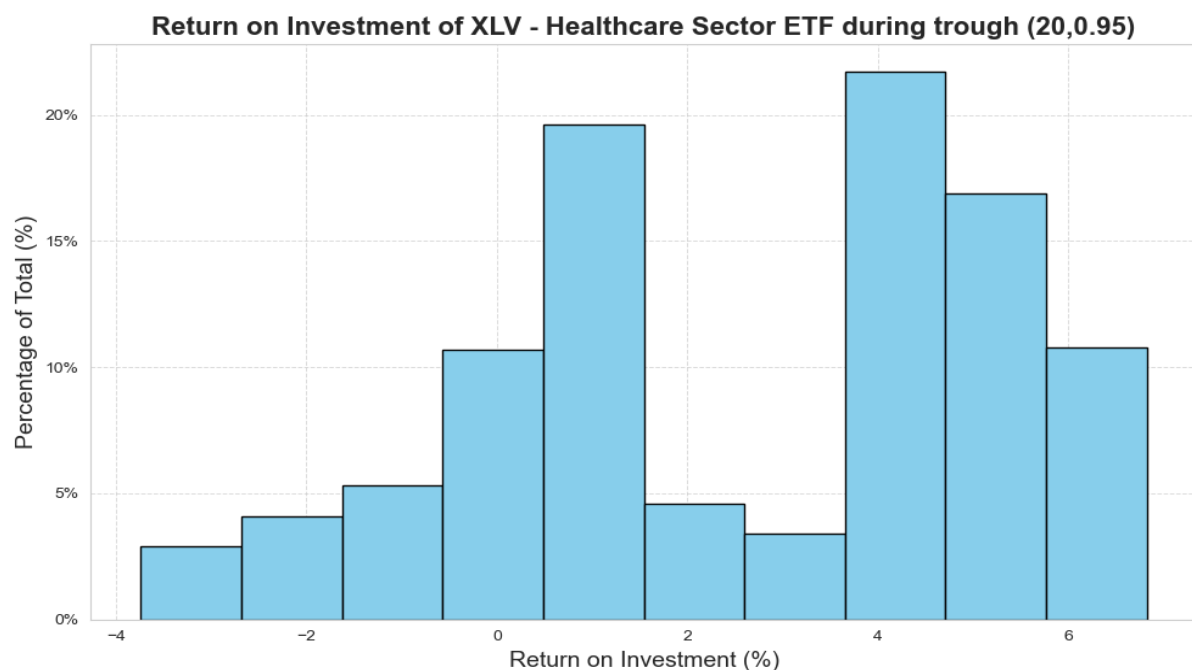


Image 5: Bollinger Band Healthcare Distribution of Returns

Image 5 shows the distribution of returns using Bollinger Bands is that there is less of normal distribution of data, this reduces the importance of standard deviation. There was an equal chance of having a 1% return as there is a 4% return. This is why it is so important to have distributions across multiple sectors during backtesting to best understand what the expected return on investments would be.

	trough	expansion	peak	contraction	all_data
XLB	6.70250076	1.91301905	3.66521073	1.15086682	1.03517115
XLI	5.32642927	2.49634469	3.45356946	-0.3405587	1.22637602
XLF	2.3645048	2.40095609	5.51544535	-0.8636046	1.22129376
XLK	3.48685532	2.07504506	2.75049257	-0.9310741	1.59740503
XLY	7.65697496	2.22784353	2.32907876	1.66687245	1.31951245
XLP	1.92654157	2.11013333	1.02385448	1.10170488	1.34505633
XLE	4.90929081	1.26168955	1.9993428	1.70919226	1.14100527
XLV	2.68524093	2.50851207	2.15208228	-0.211904	1.43179713
VOX	8.36608632	1.37719734	2.64492401	-1.328792	1.06021061
XLU	-0.7201294	1.04778276	1.16160587	-1.357912	1.09241303
IYR	1.51401037	0.94002783	0.92712627	2.87211268	1.22745363
Average	4.01984598	1.85077739	2.51115751	0.31517307	1.24524495

Table 2: Baseline Bollinger Band Returns (20,0.95)

There is a 4% return on investment during a trough macroeconomic period, meanwhile there are lower returns during the expansionary and peak time periods. This is an inverse relationship to what was observed with the buy and hold strategy. The buy & hold strategy suggests that as economic times improve, the higher expected returns. During times of low economic activity such as in a trough, there is higher volatility which allows for a better performance of Bollinger Bands, this is a strong basis to explore how a combination of Bollinger Bands and a Buy & Hold investment may improve success of investment strategies.

Optimized Bollinger Band

Optimized Bollinger Bands was the use of brute force grid search techniques to get the highest performing hyper parameters, this was an increase in rolling average of 20 days to 30 days. The impact of this increase is going to smooth out the upper and lower band allows for a create clustering of signals when overbought and oversold time periods are identified.

	trough	expansion	peak	contraction	all_data
XLB	10.8718564	1.95594828	2.00341474	1.76509138	1.42140014
XLI	8.4291905	2.88584042	2.30628814	-0.414606	1.31011951
XLF	4.2037909	2.82931131	2.96308561	0.56527831	1.11393186
XLK	7.22171225	1.65404835	0.83871082	-0.5488647	1.56951395
XLY	9.98765935	2.5659513	1.81490127	1.46092571	1.41851045
XLP	1.83497525	2.12394621	0.60523913	1.43002911	1.35607927
XLE	5.73326403	0.71549232	1.25889409	-0.8816989	0.71975396
XLV	5.57196179	2.26961798	3.37689304	0.67435635	1.45426417
VOX	10.4903588	1.60659262	1.0798114	-1.6101261	0.95391791
XLU	-0.4364928	1.11785008	1.38650072	-1.7295394	1.27025324
IYR	6.86989647	1.21066713	0.91895498	2.43522836	1.00314985
Average	6.43437936	1.903206	1.68660854	0.28600674	1.23553585

Table 3: Optimized Bollinger Band Return (30,0.95)

Table 3 shows the performance of Bollinger Bands to be optimized in the trough period as seen by the 6.4% return, with some stocks seeing upwards of 10% return in a 90 day period.

The 6.4% return during a trough is upwards of 10% improvement on the Buy & Hold strategy, the relationship between reliance on Bollinger Bands relative to a Buy & Hold strategy will be explored in *Graph 1*.

It is clear that there are some which performed better in the trough investment period such as XLB, XLY and VOX where as XLU and XLP show worse returns. Further studies will explore what the reasons that contribute towards a better investment for one stock compared to the rest, even during the same macroeconomic cycle.

Baseline Relative Strength Index

The relative strength index is made up of one variable, the number n days which will be used to compare average gain compared to average loss. If the RSI value is above 70 a signal is triggered to sell meanwhile an RSI value falls below 30 then a trigger to buy is signalled. The baseline parameter of the RSI will be 14 days.

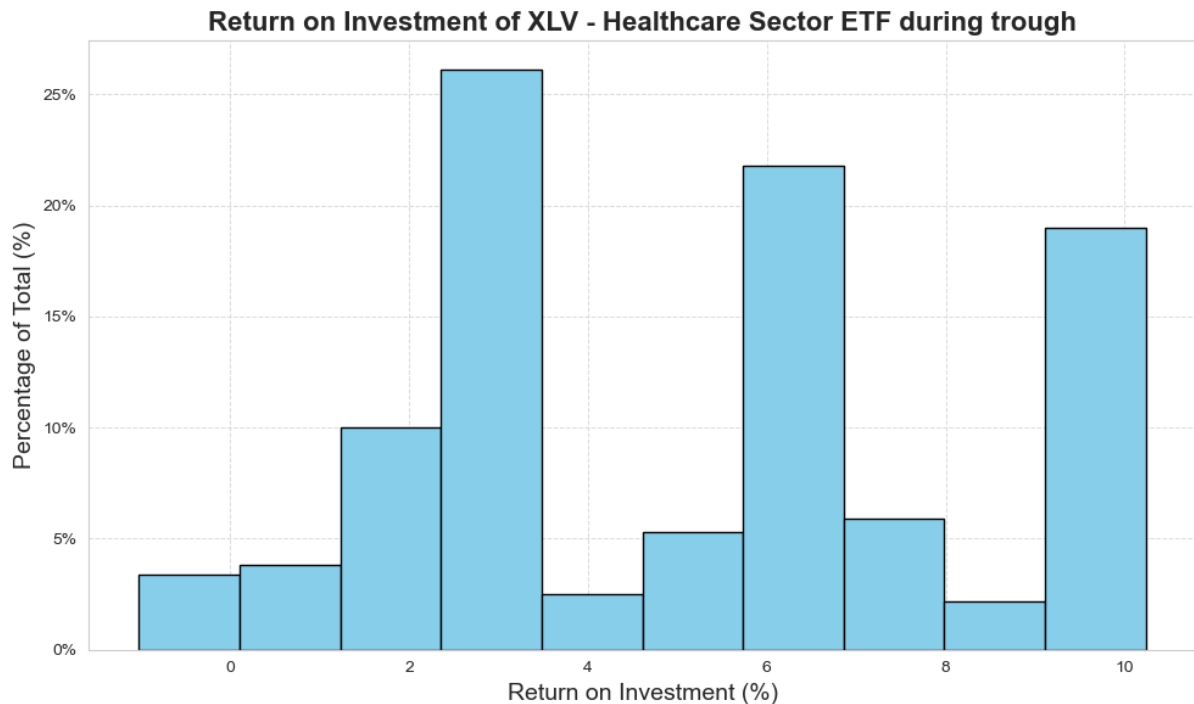


Image 6: RSI Distribution of Returns during Trough

Image 6 shows that there are 3 highest expected returns at 3%, 6% and 10% returns. The impact of this type of distribution is a clear understanding that it is not normally distributed. There are clusters of investment periods which yield the higher returns if identified by the backtesting methodology.

	trough	expansion	peak	contraction	all_data
XLB	10.0280347	1.89934591	2.61114671	0.36596075	1.6242131
XLI	2.28401724	1.84470503	5.55876708	-0.7445368	1.10398103
XLF	5.4681827	1.76275871	4.40612081	-5.1575661	1.2099588
XLK	8.72764253	0.90788374	0.86277411	-0.9462175	1.02781305
XLY	6.58547654	1.78101558	2.60166669	1.44159322	1.25023648
XLP	0.47176152	1.58162061	0	1.73098442	1.03363268
XLE	6.49321033	0.60028198	5.13970574	0.52217164	1.7632764
XLV	5.11468554	1.63681449	1.23341035	-0.1273942	1.11373394
VOX	5.74632353	2.08804363	2.79660858	-1.3509825	0.95748084
XLU	-1.2980332	1.34464534	0.98836969	-0.6552193	1.45848383
IYR	9.85462774	0.99039829	1.03328624	3.4197798	0.97733869
Average	5.40690266	1.49431939	2.47562327	-0.1364933	1.22910444

Table 4: Relative Strength Index Return (14)

The RSI return tells a similar story to Bollinger Bands with a 5.4% return during the trough time period with lower returns in expansionary and peak which suggests that weaker economic periods will result in higher returns. Both XLU and XLP struggled in the RSI and Bollinger Bands suggesting that there is some commonality amongst the characteristics of the price movement of these equities. The contraction period continues to struggle to show performance success. The XLP during the peak showed a 0 return which is due to the lack of creation a buy signal. If there is no time where a the RSI falls below a certain value then there will be no opportunity to purchase the sector ETF.

Optimized Relative Strength Index

The optimized RSI signal was found by the brute force testing of different lengths of average gains and average losses. The tested ranges were from 5-17, with the n of 17 yielding the highest return during the trough time period.

	trough	expansion	peak	contraction	all_data
XLB	10.6985197	1.32761488	1.11928194	0.29538627	1.41139892
XLI	4.16915852	1.25014715	3.23718832	0.43290843	0.87686154
XLF	11.2581819	1.32581048	1.25489488	-5.3917196	0.7281452
XLK	7.78176551	0.7053951	0	-0.4645496	0.57792728
XLY	9.88644464	1.19339837	1.22494519	0.81716973	0.94069363
XLP	0.93415641	1.40488402	0	1.47766024	0.75661846
XLE	8.4725507	0.46074012	5.17057853	1.67023291	1.62101804
XLV	5.24572786	1.00451928	0.37400426	-1.758107	0.69707774
VOX	5.53420276	1.70213356	1.99266296	-0.9667656	0.76601882
XLU	0.1081007	1.23378277	0	-0.5522076	0.99770999
IYR	11.026073	0.95476257	0.57171794	3.19750446	0.74694532
Average	6.82862561	1.14210803	1.35866127	-0.1129534	0.92003772

Table 5: Optimized Relative Strength Index Return (17)

A 6.8% return during a trough time period of 90 days is the best seen across all strategies, however this is at the detriment of the average return across all data which showed a 0.92% return. This continues to show that as the emphasis on the TA increases in the trough the higher the return, in comparison to the counter for stronger economic periods.

MACD

The Moving Average Convergence Divergence TA method is the most advanced. It incorporates multiple lines based on averaging. There are 3 different parameters.

The first is the short exponential moving average which is going to be 12 for the baseline. A long exponential moving at moving average of 26 and a moving average length of contrasting the relationship between the short and long exponential moving average, which is 9.

	trough	expansion	peak	contraction	all_data
XLB	0.33888727	0.50916778	0.84745988	-0.2997783	0.63040917
XLI	-0.5948356	1.20273771	1.54089549	-1.0352083	0.75075974
XLF	1.40854032	1.27836139	1.91176922	-1.5586875	0.64239218
XLK	1.40060212	0.94023988	2.43428763	0.22144233	1.05426119
XLY	1.59922671	1.20010237	0.77320227	-0.5165435	0.59297843
XLP	-0.6045926	0.98223306	0.87594835	0.03898548	0.61181236
XLE	-1.4224086	0.87648832	0.81120532	0.02825247	0.79385191
XLV	-0.0638768	1.62661955	2.10481356	-0.7343039	0.64455837
VOX	0.24019243	1.02868428	1.39775899	-0.1944954	0.51877939
XLU	-1.3510099	1.08805095	1.28131915	-0.7245577	0.60230434
IYR	-2.9740897	0.89960266	0.62934056	-0.3219264	0.50366518
Average	-0.1839422	1.05748072	1.32800004	-0.4633473	0.66779748

Table 6: Moving Average Convergence Divergence Return (12,26,9)

This TA method shows to perform quite badly. There does not seem to be any improvements in any time period or any specific stocks when comparing with other strategies. A baseline Buy & Hold would perform far better than the MACD.

Baseline multiple signals

As explored in isolated TA evaluation there are some strategies which perform better than others. By giving different weights to the different investment strategies and combining their buy and hold signals it should allow for more signals to give a strength of investment with the number of buy signals instead of isolated TA evaluation. The baseline multiple signal is going to have Bollinger Bands with 0.4, RSI with 0.4 and MACD with 0.2. Meaning that each Bollinger Band and RSI signal will have a 0.4 proportion of a buy/sell signal each. While MACD has 0.2. This allows for a maximim investment to take place when all 3 suggest a buy meanwhile a weak signal would be when only one suggests a buy or a sell. A buy is a positive weight meanwhile a sell is a negative weight.

The Bollinger Band and RSI are using their optimized strategies for the multiple signals in the returns below.

	trough	expansion	peak	contraction	all_data
XLB	10.4350801	1.54203759	1.3867833	1.64262052	1.55403876
XLI	6.60318169	1.80516005	2.49529473	0.58405471	1.13356306
XLF	10.0002424	2.03261107	1.90770334	-2.2383928	0.69889826
XLK	6.7592151	1.08446215	0.32795518	-0.1425347	0.62063233
XLY	9.59044894	1.72209888	1.26430298	1.30508608	0.83829018
XLP	1.45815802	1.51979673	0.24128726	1.22912727	0.76945984
XLE	6.21966919	0.75804259	2.5549507	1.38730636	1.04920516
XLV	5.20286935	1.51187621	1.28878074	-0.5319845	0.56205477
VOX	7.31739842	1.47304105	1.27537885	-0.6949579	1.09773597
XLU	0.17809612	1.04943731	0.72643849	-0.6971309	0.75254863
IYR	8.06699551	1.01593324	0.64916226	2.58989298	0.93653905
Average	6.53012317	1.41040881	1.28345798	0.40300792	0.91026964

Table 7: Baseline Multiple Signals Return (BB: 0.4, RSI: 0.4, MACD: 0.2)

Multiple signals show similar returns to the Bollinger Band and RSI, they used the optimal models so that would make sense to see a 6.5% return during the trough. It makes sense that RSI and BB would show similar returns as they have similar buy, sell or hold signals.

Optimized multiple signals and Buy & Hold Strategy

The final model is going to incorporate the Buy & Hold Strategy which showed success during times of strong economic periods as well as TA strategies. Using a brute force grid search technique to find the optimal signals it was found that giving RSI a 0.8 strength and Bollinger Bands a 0.2 strength it would have the highest returns. This would again struggle across strong economic periods. By looking at *image 7* an understanding of the relationship between cash balance compared to stock balance was made. This would lead the model to use a 40% of cash balance already invested in the stock.

	trough	expansion	peak	contraction	all_data
XLB	5.37924198	2.39324259	3.21527175	0.41622424	1.73497734
XLI	0.10566289	2.40894134	5.53338588	-1.1351328	1.36772728
XLF	-2.1406435	2.47474422	4.91945484	-6.3425129	1.20720602
XLK	6.36739948	1.80790752	2.75093581	-1.1402131	1.53674021
XLY	4.75117408	2.51022739	2.96737513	0.46065818	1.50967256
XLP	-0.7590929	1.88204504	1.08911194	1.33793624	1.28275333
XLE	3.51137225	1.11606398	4.48964506	0.62738403	1.95587167
XLV	3.67730763	2.44271333	1.74445962	-1.2094913	1.39196173
VOX	5.65599536	2.43632465	3.3073076	-2.0782727	1.16247867
XLU	-2.1944762	1.68989449	1.85561804	-1.526706	1.67048802
IYR	1.08148112	1.4291231	1.86013013	2.80858605	1.16603676
Average	2.31231111	2.05374797	3.06660871	-0.7074127	1.45326487

Table 8: Optimized Multiple Signals with Buy & Hold

The returns for trough are reduced to 2.3% but for peak and expansion they improve from the TA strategies. However, contraction still manages to struggle with a return of -0.7 and all data has some improvement to 1.45% but still not the same level of the Buy & Hold Strategy.

Notice that depending on the type of sector some trough returns will be worst than others, for example the XLF which showed terrible performance with the Buy & Hold Strategy showed a -2% return on average.

Discussion

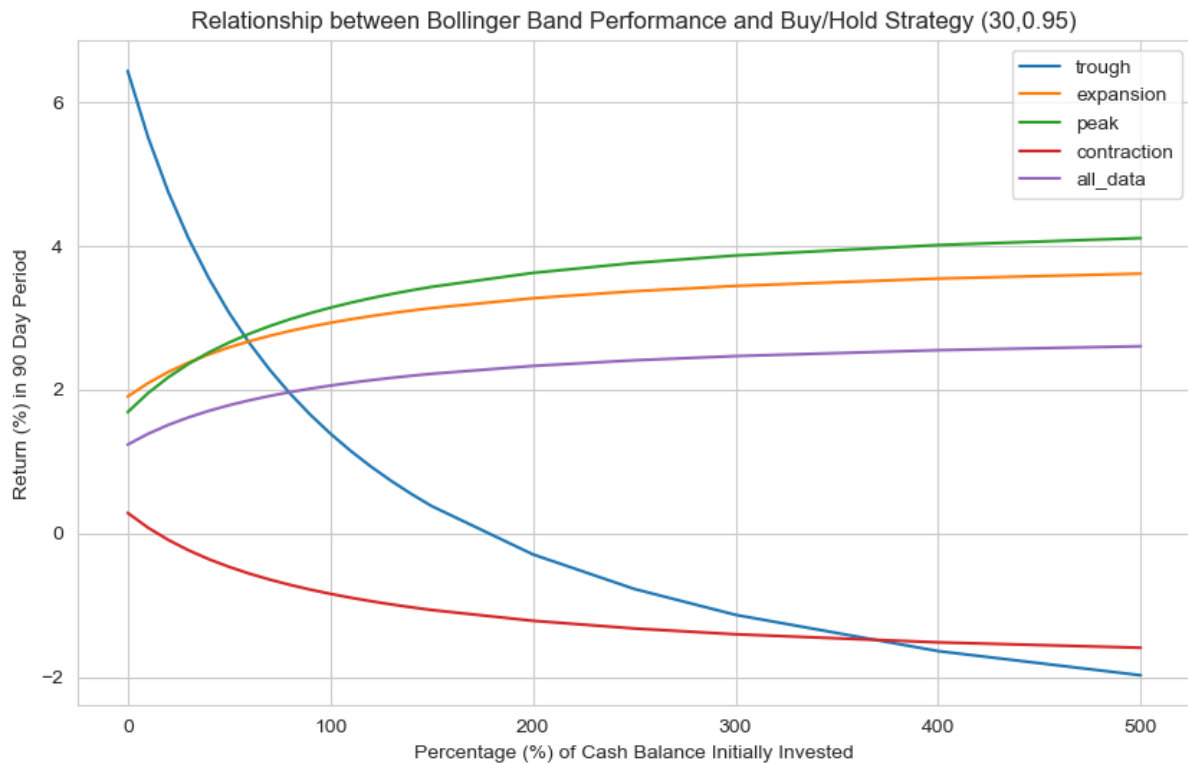


Image 7: Relationship between Optimized Bollinger Band and Buy & Hold Strategy

Image 7 is an analysis of how differing the amount in cash relative to initial stock investment yields different returns. If you observe the trough macroeconomic period, there is a clear inverse relationship as you increase the amount invested in cash. When there is no initial investment in stock you can see a 6% return on investment during a trough, however as you increase the amount of initial stock investment the average returns diminish. In comparison to a peak which shows the opposite effect. With a 2% return during with no initial investment and a 4% return with a 500% initial investment. As you move along you can see that there is a intersection of expansion, peak and trough all around the 40% of stock invested relative to cash balance. This should be used as a good representation of how much to incorporate a Buy & Hold strategy with TA. As you can see that as you increase the amount invested initially it reduces the emphasis of the TA and follows the market movement rather than TA.

Conclusion

The financial industry continues to be shaped by the innovation in applied mathematics, statistical inference and machine learning/intelligence. For these improvements yield higher returns than traditional approaches to investment. 'Enhancing Technical Analysis with Data Science' explored the success of Technical Analysis across multiple macroeconomic periods using sector ETFs. The greatest success was identified in utilizing TA during the trough time period in the 2008 GFC. This economic crises known for its financial sector impact was countered by techniques such as Bollinger Bands and RSI which showed upward of 10% returns in some sectors. Suggesting that when the market is in crises there are more opportunities to identify overbought and oversold equities. However, the passive Buy & Hold Strategy showed strong signs of success during economic periods of growth. With expansions and peaks seeing 4% returns. An inverse relationship between economic growth and expected returns using TA was identified, where as the direct relationship was observed with the Buy & Hold. Allowing for the incorporation of both to create a balanced model, which yielded 2.3% return during troughs which is far better than a -4% return with a Buy & Hold strategy yet upholding a 3% return during the peak.

Ethical Considerations

The financial markets are emersed in ethical dilemmas. Every time you make a profit you are making money off the ignorance or bad decisions of others, after all the price of a stock is based on the level of supply and demand. Using data science in the financial investment industry is an extension of this, with an even greater magnitude. You are using computers to automate your decision making, not your own financial knowledge. For some this may be considered unethical but utilizing your own skills for your own prosperity is not. As long as you are not involved with insider trading that is not possible to gain from public information,

you are not doing anything that someone else could not. The patterns are there, you just have to have the time and skillset to decipher what is useful and what is not.

Further Studies

To further improve upon traditional TA strategies, creating a reward strategy based on the success of a buy/sell/hold signal can be used to engineer my own model. This project was also tested across very safe sector ETFs which are not known for volatility, so studying how TA performs for stocks specifically known for higher volatility could yield higher returns. The combination of fundamental, technical and sentimental analysis could be used to make an all encompassing predictive model of buy/sell/hold signals. The most important next step is understanding what makes a good signal. This can be done through advanced Data Science techniques such as logistic regression, neural networks and reinforcement learning which will advance the pattern identification. The importance is in the direction of the stock after a buy is up and the direction after a sell is down, rewarding machines that identify this can be used to make a dynamic and evolving model capable of live deployment that adapts and evolves to its changing environment.

References

Source 1: Bartolini, M., & Dong, A. (n.d.). **Sector business cycle analysis**. SPDR Americas Research. Retrieved from [<https://www.ssga.com/library-content/products/fund-docs/etfs/us/insights-investment-ideas/sector-business-cycle-analysis.pdf>]

Source 2: S. Pagar, A. Jaiswal, A. Auti, V. Purohit and V. Mishra, "**Technical Analysis of Stock Market using Data Science and its tools**," 2023 7th International Conference On

Computing, Communication, Control And Automation (ICCUBEA), Pune, India, 2023, pp. 1-5, doi: 10.1109/ICCUBEA58933.2023.10392184