Running head: PREDICTION OF S&P 500 USING MACHINE LEARNING

**Prediction of S&P 500 using technical indicators in conjunction with random forest classification**

Jimmy Ng

CUNY School of Professional Studies

Data 698

Spring 2022

Final Paper

**Abstract**

This study[[1]](#footnote-1) investigated the effectiveness of various technical indicators in conjunction with machine learning model to predict stock price movement from the S&P 500. We employed the Moving Average Convergence Divergence (MACD), Relative Strength Index (RSI), and Commodity Channel Index (CCI) as our core trading rules. In addition, we applied feature engineering to come up with a set of confirmation logic for price movement. Trading strategies were executed only if the signals generated by the technical indicators were confirmed by the logic. For instance, we calculated two Exponential Moving Averages (EMA) – 10-day versus 30-day EMA. The crossover between the two EMAs confirmed an up- or downtrend signal. Furthermore, we passed all these features into a Random Forest (RF) classification algorithm for identifying market trend. We compared the actual return of each trading strategy among the 458 stocks across 11 industry sectors from the S&P 500 between January 2018 and March 2022. While none of these trading strategies could outperform the market, the random forest classification consistently, significantly outperformed the traditional technical indicators. Overall, technical analysis in conjunction with machine learning can substantially enhance profitability in the stock market.

**Introduction**

Market volatility is unsettling, but not unusual. Profitable trading relies on making as many correct predictions as quickly as possible to maximizing return. That means, experienced traders tend to quickly discern signal from noises while making the least number of false positive in the long run. However, there are overwhelmingly strong evidence drawn from empiric research that oppose the idea of making profits from prediction of stock market prices. The rationale is that asset prices already reflect all available information, and therefore no trading strategy can reliably outperform the market. The Efficient Market Hypothesis (EMH) argues that markets are efficient, and therefore there is no room left to generate excess profits by investing since everything is already fairly and accurately priced (Fama, 1970). In other words, it is impossible to “beat the market” consistently on a risk-adjusted basis since market prices should only react to new information. Thus, the idea of generating profits reliably from prediction based on existing trading data is considered futile. It is arguable that one can make extreme profit in the market only when it is overpriced (aka when market is in bubble).

If we were to assume that the markets are efficient, then no model or trading strategy can ever “beat the market” to provide excess return to investors. However, investors realize that excess return can still be reasonably generated by various prediction methods. Technical analysis generates trading rules and signals that guide investors to make informed decision. Many technical analyses have been practiced and applied effectively in a variety of asset markets in many decades. It is sometimes seen as more art than science. There is a strong assumption that prices move in patterns that can be detected and taken advantages of by investors, and that the durations of these patterns sometimes can last long enough to compensate for any transactional costs and losses that could be incurred due to false positives. This assumption reinforces the innovative use of technical analysis across different financial markets.

Obviously, this assumption of technical analysis is at odds with the widely accepted EMH proposed by Fama and others in the academic circle. Many studies have shown that trading strategies generated by technical analysis have not made any significant or acceptable level of profitability, not to mention that considerable portfolio turnover generated by noises that resulted in additional transaction cost that could offset any profit accumulated during the investing period. In short, researchers were reasonably skeptical about the effectiveness and application of these trading strategies generated by technical analysis (Fama & Blume, 1966; Jensen & Benington, 1970). Technical analysis has had its ups and downs over the past few decades, depending on the extent of the prevalence of EMH in academic circles. This represents a challenge for behavioral science both in terms of the choice of investment strategy as well as the theoretical basis of its application.

**Literature Review**

Park and Irwin (2007) conducted a comprehensive literature review of the performance of technical trading strategies. Their detailed analysis included a total of 137 studies which covered the stock markets, foreign exchange markets and future markets during the period between 1960 and 2004. All the studied material was divided into two groups: early studies (1960-1987) and modern studies (1988-2004). The early studies analyzed application of only several simple trading rules and in most cases the trading strategies were not implemented and tested in appropriate way. In general, this group of studies disproved the efficacy and application of technical trading strategies. At least, the early studies indicated that technical analysis only seemed to be profitable in foreign exchange and futures markets, but not in stock markets in general.

On the other hand, modern studies demonstrated significantly more mature and flexible use of trading strategies adopted across various technical indicators. This group of studies mostly applied appropriate way of back testing and they were able to offer suggestion about enhancing accuracy of trading signals, strategies in general. In most cases (about 60%) of the modern studies, Park and Irwin were able to summarize and confirm the profitability of various trading strategies based on technical analysis. The rest of modern studies showed either mixed or negative results. By and large, the modern studies indicated the effectiveness and consistency of generating profitable returns in a variety of speculative markets at least until the early 1990s. Despite some positive evidence on the profitability of technical analysis, Park and Irwin concluded that majority of these empirical studies lacked appropriate way of estimating and controlling risk and transaction cost in their trading strategies. As a result, researchers remained very skeptical about the use of technical indicators.

More recent studies in the past decade, however, have neither strongly supported nor opposed the effectiveness and application of trading strategies based on technical indicators. Instead, success of technical trading is now considered to be adaptive and responsive to the market (or environment) – in accordance with the Adaptive Market Hypothesis (AMH) proposed by Andrew Lo in the early 2000s. Under this new paradigm, traditional models of modern financial economics can now coexist with behavioral models (Todea, Ulici & Silaghi, 2009; Todea, Zoicas-Ienciu & Filip, 2009). The AMH argues that investors are capable of optimal dynamic allocation and many behaviors previously seen as irrational and impractical can now be explained as adaptive and consistent with an evolutionary perspective, i.e., individuals are striven to adapt and respond to environment (market) using simple heuristics for enhancing, maximizing the chance of survival (profitability). Economic irrationality is therefore seen as basic instinct driven by evolutionary forces. For example, market participants drive stock prices way above their value in relation to some system of stock valuation – aka creating a stock market bubble. This seemingly irrational thinking and behavior are completely understandable under AMH. The stock market bubble is attributed to cognitive biases that lead to groupthink and herd behavior in response to market volatility and the “fear of missing out” (FOMO). The AMH is treated as an alternative to EMH, and the AMH is characterized as the missing piece between EMH and behavioral finance. The dynamic of the financial market, the chance for arbitrage and the irrationality of investors all point to an unstable relationship between risk and return, proposed by the AMH. Therefore, Lo (2004) stated that investment strategies "wax and wane, performing well in certain environments and performing poorly in other environments". Investors can still “beat the market”, but there is no set of trading rules that can guarantee to work every time in every market.

Opposing and ambivalent attitude on the performance and application of technical analyses is now considered relative, and mostly depending on the time and market that is being studied. Nowadays, many investors would agree that success in technical trading rules largely depends on the conditions on the financial market, primarily (non)liquidity, and to a lesser extent macroeconomic (in)stability, including the ability to short-sell stocks (Taylor, 2014). Although the existing literature on the application of technical analysis is still quite controversial, technical analysis is widely accepted and communicated in academic and professional circles. Many empirical studies have been conducted to find optimization of trading strategies (Gehrig & Menkhoff, 2006; Menkhoff, 1997). For example, Moving Average Convergence Divergence (MACD) and Relative Strength Index (RSI) are two of the most popular technical indicators for developing trading strategy. They are found consistently effective in optimizing investors’ portfolio in the financial markets (Rosillo, de la Fuente & Brugos, 2013; Stankovic et al., 2015; Wu & Diao, 2015). Commodity Channel Index (CCI) is another tool commonly employed by aggressive option-traders. The CCI significantly helps in assessing momentum and gives hint about future volatility in prices (Maitah, Prochazka, Cermak & Šrédl 2016). In addition, trading strategies based on crossover signal generated by moving averages significantly helps in reducing investment risk and confirming trading signal (Anghel, 2013).

More recent studies suggest that, if the information obtained from stock prices is carefully pre-processed and then run by complex machine learning algorithms, the trend or stock price index movement is highly predictable (Patel, Shah, Thakkar & Kotecha, 2015). The advance of machine learning and more cost efficient of big data technology have made significant contribution to investment strategy, e.g., from quantitative to sentiment analysis using Natural Language Processing (NLP). Machine learning is gaining tremendous attention in the financial market. Concepts such as support vector machines, genetic algorithms, artificial neural networks, fuzzy logic, and chaos theory have become new buzzwords in Wall Street. Many studies have demonstrated the application of various machine learning algorithms in the successful prediction of stock price movement and thus significantly contributed to the increase in profitability and reducing the risk involved in trading. Some of the most sophisticated algorithms in the market include but not limited to Artificial Neural Networks (ANNs) (Boyacioglu, Kara & Baykan, 2009), linear and multi-linear regression (LR, MLR) (Atsalakis & Valavanis, 2009), genetic algorithm (GA) (Atsalakis & Valavanis, 2009), and Support Vector Machine (SVM) (Stankovic, Markovic & Stojanovic, 2015). The methods most widely used with high success rate for predicting stock market trend are the approaches based on Support Vector Machine (SVM) and Random Forest (RF) (Chen, Chen & Liu, 2020; Lohrmann & Luukka, 2019).

On the one hand, technical analysis seems to have become obsolete and replaceable in the age of machine learning, in which sometimes it works like a black box to investors. However, some have found evidence that machine learning cannot always reliably outperform or work as well as simple technical indicators. For example, Jian and Jakubowicz (2017) applied four classic classification algorithms: random forest, gradient boosted trees, artificial neural network, and logistic regression in predicting 463 stocks of the S&P 500. They carried several experiments thoroughly to study the predictability of these stocks using the mentioned algorithms. Furthermore, they validated each prediction algorithm by applying standard cross validation, sequential validation, and single validation. Surprisingly, they were not able to predict stocks future prices from their past using any of these algorithms. Although the financial sector was relatively easier to predict, none of these algorithms was able to predict, generate meaningful signals that could be profitable. Similarly, Qian and Rasheed (2007) investigated the predictability of the Dow Jones Industrial Average Index using artificial neural network, decision tree, and k-nearest neighbor. They concluded that none of these algorithms could achieve more than sixty-five percent accuracy. Their work seemingly supported the argument proposed by the Efficient Market Hypothesis, i.e., stock prices should follow a random walk pattern; trading strategies based on prediction of stock price movement would not generate any considerable profit in the long run.

On the other hand, there are still overwhelmingly strong evidence in favoring the use of machine learning in predicting stock price movement. Chen and colleagues (2020) explored the application of machine learning in prediction of S&P 500 stock price movement between 2014 and 2018. They compared three machine learning models: Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Random Forest (RF). They concluded that all three models outperformed benchmark market index, and that RF generated the best performance by risk-adjusted measures, followed by SVM and ANN. Others found similar and positive evidence in support of the application of machine learning in stock price prediction, Random Forest in particular (Lohrmann & Luukka, 2019; Thakur & Kumar, 2018).

Recently, researchers have become interested in exploring emerging markets, since these markets are now being recognized as important alternative of investment opportunities; however, studies on these markets showed mixed results (Stankovic et al., 2015). Unlike these emerging markets, many studies have demonstrated the ease of access to the historical data and consistent profitability generated by technical indicators in the case of developed markets. As a result, the present study focused on the application of technical indicators and machine learning algorithm to evaluate actual return from a mature market, i.e., S&P 500.

The S&P 500 is widely considered a significant representation for the overall US economy as it includes some of the biggest corporations from the NYSE and the NASDAQ stock exchanges. Basically, the list of S&P 500 includes some of the largest (in terms of capitalization) and most influential global companies across eleven industry sectors. Any efficient prediction of the S&P 500 is of great importance to investors and policy makers in general, since their decisions may trigger large scale buying or selling opportunities that can bring significant impact to the global market and overall economy. There is already compelling evidence in favor of the use and experiment of trading strategies using historical data from NASDAQ, Dow Jones or S&P 500 for model tuning and back testing purposes (Chen, et al., 2020; Jian & Jakubowicz, 2017; Lohrmann & Luukka, 2019; Thakur & Kumar, 2018). Studies in EU or other emerging markets showed that predictive power of technical analysis or machine learning model was only effective in small and medium sized capitalization markets (Metghalchi, Marcucci & Chang, 2012). Others found mixed evidence in the world’s emerging markets (Fifield, Power & Donald, 2005; McKenzie, 2007). Overall, there is still a lack of consensus about the profitability of technical trading strategies in European frontier markets (Stankovic et al., 2015).

As a result, we focused only on the S&P 500 and applied a specific machine learning algorithm, i.e., Random Forest (RF), to investigate the effectiveness of trading strategies based on historical pricing data between October 2014 and March 2022. This study built a set of simple trading strategies based on signals generated from Moving Average Convergence Divergence (MACD), Relative Strength Index (RSI), and Commodity Channel Index (CCI). Signals extracted from these indicators along with other features (such as the crossover strategy by two Exponential Moving Averages) were used as inputs to build a classification model that would signal market trend for each security in the S&P 500. Result of this classification model would form an alternative trading strategy in comparing against the traditional technical trading rules. We measured the success of these trading strategies by comparing their actual returns at the end of evaluation period.

**Data and Methods**

Technical analysis applies various qualitative and quantitative methods to evaluate asset price trends. The simplest qualitative method is based on charting asset prices along with trading volume to discern any pattern that signals optimal instant for entry or exit. Technical indicators are quantitative methods in which they represent simple mathematical expression of price and volume changes. Investors usually use a combination of both methods, in conjunction with fundamental and sentiment analyses, to provide a more precise overview of market trends.

In this study, we selected 458 stocks from the current S&P 500 and analyzed their prices between October 2014 and March 2022. We employed some of the most common technical indicators in the field, i.e., Moving Average Convergence Divergence (MACD), Relative Strength Index (RSI), and Commodity Channel Index (CCI) to capture buy and sell signals. Trading strategies were based on signals generated from these indicators in conjunction with other features described in the following section.

*Moving Average Convergence Divergence (MACD)*

The Moving Average Convergence-Divergence (MACD) is a commonly used indicator for trend monitoring. It is a popular momentum indicator that can visually display the relationship between two EMAs of a financial asset price. The MACD is simply calculated by subtracting the 26-period EMA from the 12-period EMA. The result of that calculation is the "MACD line". Subsequently, we calculate a nine-day EMA of the MACD line and that is referred to the "signal line". A buy signal is triggered when the MACD line crosses above its signal line; conversely, a sell signal is triggered when the signal line crosses above the MACD line. MACD is a trend-following indicator that can allow investors to anticipate the optimal time of buying and selling a security. In addition, investors pay attention to the speed of crossover as it can reveal whether the security is overbought or oversold. MACD helps investors to understand and anticipate whether the bullish or bearish movement in the price is strengthening or diminishing.

In this study, the buy signal was only triggered when the MACD instantaneously crossed above the signal line (meaning the MACD line from the day prior must still be below the signal line, whereas the present day the MACD line just barely crossed it to trigger a buy signal); likewise, the same condition applied to the sell signal (meaning the signal line from the day prior was still above the MACD line and then it crossed below it). That was to reduce false positives.

*Relative Strength Index (RSI)*

Another popular indicator is the Relative Strength Index (RSI), which is also a momentum indicator that measures the magnitude of recent price changes to evaluate overbought or oversold conditions of an asset. The RSI is sometimes referred to as an oscillator as it can be visually displayed as a line between two extremes of 0 and 100. The interpretation is quite simple as the RSI goes above the value of 70 would indicate an overbought condition; on the other hand, if the value goes below 30, it would signal an oversold condition. The overbought/oversold condition alarmed by the RSI value would indicate a trend reversal or corrective pullback of an asset price in near future.

Like the above MACD trigger, the RSI from this study would only trigger a buy signal when the RSI from the day prior was still above the threshold (30), and then crossed below it in the present day to trigger an oversold condition; similar condition applied to the overbought condition, i.e., RSI must be below the threshold (<70) the day prior to the current day until it went above it in order to trigger a sell signal.

*Commodity Channel Index (CCI)*

The Commodity Channel Index (CCI) is like the RSI in which the CCI is also a momentum-based oscillator that helps in detecting whether an asset is being overbought or oversold. The CCI assesses trend direction and strength, primarily used for spotting any new trend in development, watching overbought or oversold condition, and signaling weakness in trends when the indicator diverges with price. When the CCI moves from below 0 to above 100, it signals that the price is moving in an uptrend direction; on the other hand, if it goes from above 0 to below -100, it signals a downtrend momentum.

In this study, the buy/sell signal was generated only when the price entered the overbought (+100) or oversold (-100) zone. When the CCI continued to rise in two consecutive days, i.e., today > yesterday and yesterday > the prior day, it signaled a trend reversal when the CCI was in oversold condition, or it signaled a strong momentum in the overbought condition that flagged a buying opportunity. Thus, it indicated a buy signal. Likewise, similar logic applied for sell signal when the CCI dropped in two consecutive days in either the overbought or oversold zone.

*Other Features*

The above MACD, RSI and CCI indicators had to be “confirmed” by at least two of the following features in the same trading day.

a) Moving Average (MA): a running moving average or simply Moving Average (MA) represents the average of the price of a financial asset over a certain period. This is a frequently used indicator, and usually the first step in time series analysis when investors look at the trend of a financial asset. The purpose of calculating MA is to smooth the trend of asset prices by removing noise in the time series. In this study, when the current closing price was higher than its 5-day simple moving average, we flagged a positive signal; otherwise, a negative signal.

b) Exponential Moving Average (EMA): traders tend to apply the Exponential Moving Average (EMA) for generating insight (and confirmation) of market trend. The EMA is still a type of MA, but it places a greater weight, e.g., a smoothing constant, on the most recent data points when calculating an average over a certain period. MA or EMA is considered a lagging indicator as it displays signal of a trend that is already in place. Thus, a trading strategy solely based on the signals triggered by MA or EMA tend to miss the optimal buying or selling window, but we should use it as confirmation of buying/selling opportunity. In this study, we adapted a trading system that involved the calculation of two EMAs. That was, we calculated the EMA for the past 10-day versus the past 30-day. When there was a crossover between the two EMAs, that would trigger a positive signal of market uptrend (10-day EMA > 30-day EMA) or negative signal for downtrend (10-day EMA < 30-day EMA).

c) On-Balance-Volume (OBV): OBV is a lesser-known technical indicator that uses volume flow to predict changes in prices. OBV provides a running, cumulative total of volume (positive and negative) of an asset based on the change in closing price day-over-day. The OBV is implemented based on three simple logics, i.e.,

1. If today's closing price is higher than yesterday's closing price, then: Current OBV = Previous OBV + today's volume

2. If today's closing price is lower than yesterday's closing price, then: Current OBV = Previous OBV - today's volume

3. If today's closing price equals yesterday's closing price, then: Current OBV = Previous OBV

Consider that if today closing price is higher than yesterday closing price, the market is more likely to be in an uptrend (also reflected by volume change in the OBV formula); the opposite is true for downtrend movement. Applying the OBV alone is not enough for generating any meaningful buy/sell signal; however, we treated it like the MA/EMA as a confirmation strategy in this study. If the OBV of the current day was higher than the day prior to it, then we flagged it as a positive signal; otherwise, a negative signal.

d) Overnight-flag: it was a custom logic that we applied to the “open” and “close” prices. We generated a positive signal when the opening price was higher than the closing price the day prior to it, and simultaneously the closing price of the day must be higher than its opening. It indicated an uptrend momentum. Likewise, a negative signal was triggered when the current opening price was lower than the closing price the day prior to it, and simultaneously the closing price of the day must be lower than its opening. It indicated a downtrend momentum.

None of the above features were strong enough to generate an impactful buy/sell signal alone. Instead, they served as confirmation signal for the MACD, RSI, and CCI strategy. In this study, the “buy” or “sell” signal from each of the three technical indicators must be met by at least 2 of these confirmations in the same trading day.

For extra measure, we took a step further to collect an additional sample on top of the 458 stocks. The additional sample was the symbol “SPY” – aka the SPDR S&P 500 trust and that is an exchange-traded fund which trades on the NYSE Arca. It is designed to track the S&P 500 stock market index, and therefore, it serves as a proxy of the market performance. First, the “buy” signal of the MACD, RSI or CCI strategy must be “confirmed”, and then it had to align with the market, i.e., the SPY must also flag a “buy” signal in the same trading day. We used this consensus to minimize risk and reduce noise from the indicators. On the other hand, we did not apply the same logic for “sell” because it was safer to have more false positive for sell than buy signal. We tried to be extra cautious before entering the market; in contrast, we wanted to quickly exit a trading position as soon as we detected any anomaly to minimize loss.

*Random Forest (RF)*

Random Forest (RF) is a very popular classification algorithm that is made up by numerous decision trees. It consists of many individual decision trees that operate as an ensemble. The algorithm applies randomness from a set of independent variables to build each individual tree in random to promote uncorrelated forest. Subsequently, each tree in the random forest spits out a class prediction and then the class with the most votes becomes the model's prediction, aka majority vote.

In this study, we passed all the above signals, features into the random forest algorithm to generate buy/sell signal. These included the MA, EMA, OBV, Overnight-flag, MACD, RSI, and CCI. We defined our dependent variable as a flag for either “up”, “down”, or “flat” to signal near-future price movement. This dependent variable was based on the cross-over of closing price and moving average of the proceeding 5 trading days. For instance, if 80% or more (4 out of 5) of the following trading days had the closing price above its corresponding 5-day MA, it signaled an uptrend condition (buy); conversely, if 80% or more of the time, the closing price fell below the corresponding 5-day MA, it signaled a downtrend condition (sell); other than these, it returned a flat condition. This dependent variable served as a proxy of whether the price was in an uptrend or downtrend momentum. In this study, we split the data into two for 1) model building, and 2) back testing. We built the model using data between October 2014 and December 2017. We generated signal for each security between January 2018 and March 2022. We did not care so much about accuracy, F1 score or sensitivity of our model, since making the most accurate classification was never our intention. Instead, our goal was to signal the optimal entry/exit point to maximize return.

We executed a trading strategy (based on buy/sell signal) in the following trading day using the “open” price. We used the period between January 2018 and March 2022 for back testing. We measured the actual return at the end of March 2022 for each strategy, including the Buy-and-Hold strategy. We applied the Analysis of Variance (ANOVA) to measure any statistically significant difference in means across these strategies and industry sectors.

**Research Question**

Can machine learning model trained on a variety of traditional technical indicators be used to reliably outperform the market? Can machine learning model reliably outperform the traditional technical analysis?

**Results**

There were 458 samples (stocks) across 11 industry sectors collected between October 2014 and March 2022. We used the time between October 2014 and December 2017 as training period. Subsequently, we used 2018 and beyond as evaluation period for back testing of our strategies. We measured the actual return of each sample (stock) at the end of evaluation period. The market was extremely volatile and displayed a significant rebound, uptrend in the year of 2020 during the early phrase of global pandemic, and then entered the bear market since November 2021. Overall, the market saw a spectacular return among all sectors with an average return of 78% (*SD* = 0.96). The actual return varied across sectors, i.e., Information Technology led the way by 145% (*SD* = 1.63), whereas Energy trailed the most behind by up only 30% (*SD* = 0.5).

All 4 strategies in addition to the Buy-and-Hold strategy generated positive return. A one-way subject analysis of variance (ANOVA) was conducted to compare the return among strategies. It revealed a significant main effect, *F*(4, 2285) = 82.95, *p* < .00. Post hoc comparisons using the Tukey HSD test indicated that the market outperformed all the other strategies, i.e., Buy-and-Hold strategy generated 78% return (*SD* = 0.97), far above the others. The mean of return by Random Forest (*M* = 0.41, *SD* = 0.63) was statistically, significantly higher than the rest. While MACD (*M* = 0.28, *SD* = 0.45) significantly outperformed CCI (*M* = 0.18, *SD* = 0.39) and RSI (*M* = 0.15, *SD* = 0.29), the latter two were not significantly different from each other. Table 1 presented the actual return of the various strategies. Figure 1 displayed the boxplot of actual return among the strategies.

The average return from all strategies were significantly different from 0 (as indicated by the lower and upper bound from Table 1); however, none of the technical or machine learning strategies performed any better than the market in any sector in terms of average return, as confirmed by another set of ANOVA done on individual sector. Table 2 presented the actual return by strategies across the 11 sectors. Figure 2 displayed the boxplot of actual return among the strategies by sector.

**Discussion**

All trading strategies generated significantly positive return. Machine learning, i.e., random forest in our case, significantly outperformed the technical indicators. While MACD outperformed the RSI and CCI, the latter two were not considerably different from each other. However, none of these strategies could outperform the market in this study between 2018 and March 2022. Although these strategies generated positive return on average, none of them could “beat the market” in any industry sector during the extremely volatile period of global pandemic.

There were several reasons to explain the downside of these strategies. First, the goal of technical indicators was different from the machine learning algorithm. While they all signaled buying or selling opportunity, the technical indicators were programed for optimal entry and exit at the best possible timing while being extremely cautious (required layers of confirmation). For example, the RSI and CCI flagged very few buying or selling signals, and these signals had to be confirmed by other features, also agreed by SPY in the same trading day. It turned out that the RSI had the least number of buy and sell signals, followed by the CCI in this study. In fact, these two strategies, stood on the sideline most of the time during the evaluation period. They often entered a position late and exited too soon.

On the other hand, the classification model was built to classify or predict market trend. We used all the available information, features to identify market trend. The performance of each model (for each security) was not impressive at all, i.e., rarely achieved more than 60% accuracy. However, the machine learning strategy generated the most buy and sell signals and therefore it was trading frequently. The goal of this machine learning algorithm was not to enter or exit the market at the optimal point; instead, it tried to predict whether the price was going to go up, down or stay flat in near future, and then acted accordingly. As a result, this machine learning strategy always exited too soon whenever it sensed that there was more than 33% chance that the closing price was going to move in a negative direction in coming days. In other words, the model was hypersensitive to negative, downtrend movement. In addition, the model was memoryless, i.e., it did not account for historical accumulative influence, and it did not factor in long term growth, trend either.

The technical indicators entered or exited the market infrequently (only few times a year for RSI and CCI), and the machine learning strategy tended to enter and exit too frequently before waiting for the price to become stabilized to maximizing profit. Although these strategies were programmed to be conservative, they were still profitable during the extremely volatile time in the market. They were not designed to optimize an investment portfolio. Instead, this study was meant to compare the technical indicators with machine learning model. In addition, this study investigated the possibility of outperforming the market consistently using the employed trading strategies.

While the machine learning strategy worked significantly better than traditional technical indicators, none could consistently outperform the market. There were several ways to improve these strategies and this study. First, we could calculate the return multiple times using different time periods instead of relying on just one. For example, like the concept of moving average, we could evaluate, calculate the return for every past 100 days. We could average the list of returns by strategy and then come up with a fairer comparison that would be less affected by noise toward the end of evaluation period. Second, we should go back and test the threshold of these indicators. For example, what if we lowered the threshold for oversold and overbought for RSI and CCI? What if we changed the number of confirmation signal from two to one? Would it just trigger more false positives? In addition, we could extend our data to different machine learning algorithms, such as support vector machine, naïve bayes, linear discriminant analysis, etc. We should fine tune our model for better accuracy and sensitivity. We should also revisit the validity of our dependent variable to ensure that its definition really connected to what we intended to measure, i.e., market trend movement. Finally, this study did not account for risk or expected return for each security. We could extend this study to design for portfolio optimization that includes risk and expected return. In such case, we should extend the data set to cover for sentiment and fundamental analyses to get a more comprehensive view of investment strategy that goes beyond historical pricing data.

In conclusion, prediction of stock market is rewarding but extremely challenging due to its highly volatile nature. Stock market can be influenced by any change and impact from foreign commodities like emotional behavior of investors, geopolitical, psychological, and various economic factors. Investors sell out at the wrong time and often fail to gain or maintain any profit. Successful investing is therefore based on timing, i.e., whether it is time to buy, hold, or sell an asset. Although it is clearly not possible to predict stock market movement with full accuracy, losses from selling stocks at wrong time can still be avoided to greater extent using machine learning in conjunction with technical, sentiment, and fundamental analyses. Machine learning and traditional technical indicators effectively supplement each other to come up with simpler but more efficient trading strategies that can produce less noisy signals. The predictability of stock market and profitability of model-based trading strategies rely on the maturity of the market, sophistication of algorithm(s) employed, and the technical feasibility for generating accurate signal in real-time trading. We should move away from debating whether the financial markets are generally efficient, or which specific technical indicator(s) or machine learning algorithm(s) are the most effective; instead, researchers and practitioners should focus on the wide application of technical indicators in conjunction with machine learning in mature market to maximizing return.

**References**

Anghel, G.D.I. (2013). How reliable is the moving average crossover rule for an investor on the Romanian stock market? *The Review of Finance and Banking, 5*(2), 89-115.

Atsalakis, G. S. & Valavanis, K. P. (2009). Forecasting stock market short-term trends using a neuro-fuzzy based methodology. *Expert Systems with Applications, 36*(3), 10696-10707.

Boyacioglu, M. A., Kara, Y., & Baykan, O. K. (2009). Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: A comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks in Turkey. *Expert Systems with Applications, 36*(2), 3355-3366.

Chen, C. C., Chen, C. H., & Liu, T. Y. (2020). Investment performance of machine learning: Analysis of S&P 500 index. *International Journal of Economics and Financial Issues, 10*(1), 59-66.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance, 25*(2), 383-417.

Fama, E. F., & Blume, M. E. (1966). Filter rules and stock-market trading. *Journal of Business, 39*(1), 226-241.

Fifield, S. G., Power, D. M., & Donald, S. C. (2005). An analysis of trading strategies in eleven European stock markets. *European Journal of Finance, 11*(6), 531-548.

Gehrig, T., & Menkhoff, L. (2006). Extended evidence on the use of technical analysis in foreign exchange. *International Journal of Finance & Economics, 11*(4), 327-338.

Huang, W., Nakamori, Y., & Wang, S. Y. (2005). Forecasting stock market movement direction with support vector machine. *Computers and Operations Research, 32*(10), 2513-2522.

Jensen, M. C., & Benington, G. A. (1970). Random walks and technical theories: Some additional Evidence. *The Journal of Finance, 25*(2), 469-482.

Jiao, Y., & Jakubowicz, J. (2017). Predicting stock movement direction with machine learning: An extensive study on S&P 500 stocks. *IEEE International Conference on Big Data (Big Data),* 4705-4713.

Lo, A. W. (2004). The adaptive markets hypothesis: Market efficiency from an evolutionary perspective. *Journal of Portfolio Management, 30*, 15-29.

Lohrmann, C., & Luukka, P. (2019). Classification of intraday S&P500 returns with a random forest. *International Journal of Forecasting, 35*(1), 390-407.

Maitah, M., Prochazka, P., Cermak, M., & Šrédl, K. (2016). An empirical study on options trading strategy using 'Commodity Channel Index' for NSE’s Nifty options in India. *International Journal of Economics and Financial Issues, 6*(1), 176-178.

McKenzie, M. D. (2007). Technical trading rules in emerging markets and the 1997 Asian currency crises. *Emerging Markets Finance and Trade, 43*(4), 46-73.

Menkhoff, L. (1997). Examining the use of technical currency analysis. *International Journal of Finance & Economics, 2*(4), 307-318.

Metghalchi, M., Marcucci, J., & Chang, Y. H. (2012). Are moving average trading rules profitable? Evidence from the European stock markets. *Applied Economics, 44*(12), 1539-1559.

Park, C. H., & Irwin, S. H. (2007). What do we know about the profitability of technical analysis? *Journal of Economic Surveys, 21*(4), 786- 826.

Patel, J., Shah, S., Thakkar, P., Kotecha, K. (2015), Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications, 42*(1), 259-268.

Qian, B., & Rasheed, K. (2007). Stock market prediction with multiple classifiers. *Applied Intelligence, 26*, 25–33.

Rosillo, R., de la Fuente, D., Brugos, J. A. L. (2013). Technical analysis and the Spanish stock exchange: testing the RSI, MACD, momentum and stochastic rules using Spanish market companies. *Applied Economics, 45*(12), 1541-1550

Taylor, N. (2014). The rise and fall of technical trading rule success. *Journal of Banking & Finance, 40*, 286-302.

Thakur, M., & Kumar, D. (2018). A hybrid financial trading support system using multi-category classifiers and random forest. *Applied Soft Computing, 67*, 337-349.

Todea, A., Ulici, M., & Silaghi, S. (2009). Adaptive market hypothesis: Evidence from Asia-Pacific financial markets. *The Review of Finance and Banking, 1*(1), 7-13.

Todea, A., Zoicas-Ienciu, A., & Filip, A. (2009). Profitability of the moving average strategy and the episodic dependencies: Empirical evidence from European stock markets. *European Research Studies, 11*(1), 63-72.

Wang, Y., & Choi, I.C. (2013). Market index and stock price direction predicting using machine learning techniques: An empirical study on the K0SPI and HSI. *ARXIV, 1*, 1-13.

Wu, M. & Diao, X. (2015). Technical analysis of three stock oscillators testing MACD, RSI and KDJ rules in SH & SZ stock markets. International Conference on Computer Science and Network Technology (ICCSNT), 4, 320-323.

Table 1

*Percent of actual return among strategies*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **strategy** | **n** | **mean** | **sd** | **se** | **lower bound** | **upper bound** | **minimum** | **maximum** |
| **CCI** | 458 | 0.18 | 0.39 | 0.02 | 0.14 | 0.22 | -0.80 | 1.98 |
| **MACD** | 458 | 0.28 | 0.45 | 0.02 | 0.24 | 0.32 | -0.73 | 3.01 |
| **RSI** | 458 | 0.15 | 0.29 | 0.01 | 0.13 | 0.18 | -0.71 | 1.55 |
| **Random Forest** | 458 | 0.41 | 0.63 | 0.03 | 0.35 | 0.47 | -0.71 | 4.19 |
| **Buy-and-Hold** | 458 | 0.78 | 0.97 | 0.05 | 0.69 | 0.87 | -0.68 | 9.86 |

Table 2

*Percent of actual return among strategies across the 11 sectors*

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **strategy** | **sector** | **n** | **mean** | **sd** | **se** | **lower bound** | **upper bound** | **minimum** | **maximum** |
| Buy-and-Hold | Communication Services | 21 | 0.5 | 0.53 | 0.12 | 0.28 | 0.73 | -0.36 | 1.76 |
| Buy-and-Hold | Consumer Discretionary | 57 | 0.55 | 0.88 | 0.12 | 0.32 | 0.78 | -0.68 | 4.37 |
| Buy-and-Hold | Consumer Staples | 30 | 0.49 | 0.6 | 0.11 | 0.28 | 0.71 | -0.5 | 2.26 |
| Buy-and-Hold | Energy | 21 | 0.3 | 0.5 | 0.11 | 0.09 | 0.52 | -0.65 | 1.44 |
| Buy-and-Hold | Financials | 61 | 0.66 | 0.61 | 0.08 | 0.51 | 0.81 | -0.32 | 3.16 |
| Buy-and-Hold | Health Care | 59 | 1.04 | 1.2 | 0.16 | 0.74 | 1.35 | -0.54 | 7.95 |
| Buy-and-Hold | Industrials | 67 | 0.68 | 0.68 | 0.08 | 0.52 | 0.84 | -0.65 | 2.53 |
| Buy-and-Hold | Information Technology | 62 | 1.45 | 1.63 | 0.21 | 1.05 | 1.86 | -0.49 | 9.86 |
| Buy-and-Hold | Materials | 23 | 0.72 | 0.59 | 0.12 | 0.48 | 0.96 | -0.44 | 1.67 |
| Buy-and-Hold | Real Estate | 30 | 0.7 | 0.53 | 0.1 | 0.51 | 0.89 | -0.25 | 1.85 |
| Buy-and-Hold | Utilities | 27 | 0.65 | 0.35 | 0.07 | 0.52 | 0.79 | 0.09 | 1.68 |
| CCI | Communication Services | 21 | 0.22 | 0.33 | 0.07 | 0.08 | 0.36 | -0.29 | 0.8 |
| CCI | Consumer Discretionary | 57 | 0.13 | 0.49 | 0.06 | 0 | 0.25 | -0.64 | 1.98 |
| CCI | Consumer Staples | 30 | 0.22 | 0.4 | 0.07 | 0.08 | 0.36 | -0.25 | 1.92 |
| CCI | Energy | 21 | -0.09 | 0.5 | 0.11 | -0.31 | 0.12 | -0.8 | 1.02 |
| CCI | Financials | 61 | 0.15 | 0.29 | 0.04 | 0.08 | 0.22 | -0.29 | 0.8 |
| CCI | Health Care | 59 | 0.23 | 0.35 | 0.05 | 0.14 | 0.31 | -0.38 | 1.44 |
| CCI | Industrials | 67 | 0.07 | 0.29 | 0.04 | 0 | 0.14 | -0.48 | 1.38 |
| CCI | Information Technology | 62 | 0.22 | 0.4 | 0.05 | 0.12 | 0.32 | -0.49 | 1.36 |
| CCI | Materials | 23 | 0.13 | 0.46 | 0.09 | -0.05 | 0.32 | -0.39 | 1.43 |
| CCI | Real Estate | 30 | 0.51 | 0.37 | 0.07 | 0.38 | 0.65 | -0.23 | 1.26 |
| CCI | Utilities | 27 | 0.24 | 0.28 | 0.05 | 0.14 | 0.35 | -0.18 | 1.13 |
| MACD | Communication Services | 21 | 0.2 | 0.55 | 0.12 | -0.04 | 0.43 | -0.46 | 1.73 |
| MACD | Consumer Discretionary | 57 | 0.3 | 0.42 | 0.06 | 0.19 | 0.41 | -0.73 | 1.52 |
| MACD | Consumer Staples | 30 | 0.05 | 0.48 | 0.09 | -0.12 | 0.23 | -0.5 | 2.27 |
| MACD | Energy | 21 | 0.75 | 0.83 | 0.18 | 0.39 | 1.1 | -0.18 | 2.97 |
| MACD | Financials | 61 | 0.28 | 0.3 | 0.04 | 0.21 | 0.35 | -0.38 | 1.12 |
| MACD | Health Care | 59 | 0.21 | 0.51 | 0.07 | 0.08 | 0.34 | -0.46 | 3.01 |
| MACD | Industrials | 67 | 0.31 | 0.36 | 0.04 | 0.22 | 0.4 | -0.33 | 1.46 |
| MACD | Information Technology | 62 | 0.32 | 0.43 | 0.05 | 0.22 | 0.43 | -0.53 | 1.61 |
| MACD | Materials | 23 | 0.27 | 0.35 | 0.07 | 0.13 | 0.41 | -0.31 | 1.15 |
| MACD | Real Estate | 30 | 0.2 | 0.35 | 0.06 | 0.08 | 0.33 | -0.33 | 1.22 |
| MACD | Utilities | 27 | 0.28 | 0.28 | 0.05 | 0.17 | 0.38 | -0.26 | 0.85 |
| RSI | Communication Services | 21 | 0.08 | 0.24 | 0.05 | -0.02 | 0.18 | -0.22 | 0.82 |
| RSI | Consumer Discretionary | 57 | 0.19 | 0.4 | 0.05 | 0.09 | 0.3 | -0.57 | 1.55 |
| RSI | Consumer Staples | 30 | 0.21 | 0.31 | 0.06 | 0.1 | 0.32 | -0.71 | 0.8 |
| RSI | Energy | 21 | -0.15 | 0.24 | 0.05 | -0.25 | -0.05 | -0.62 | 0.38 |
| RSI | Financials | 61 | 0.1 | 0.21 | 0.03 | 0.05 | 0.15 | -0.34 | 0.79 |
| RSI | Health Care | 59 | 0.23 | 0.29 | 0.04 | 0.15 | 0.3 | -0.41 | 1.43 |
| RSI | Industrials | 67 | 0.09 | 0.25 | 0.03 | 0.03 | 0.15 | -0.43 | 0.88 |
| RSI | Information Technology | 62 | 0.2 | 0.29 | 0.04 | 0.13 | 0.27 | -0.22 | 1.5 |
| RSI | Materials | 23 | 0.24 | 0.38 | 0.08 | 0.08 | 0.39 | -0.17 | 1.27 |
| RSI | Real Estate | 30 | 0.16 | 0.22 | 0.04 | 0.08 | 0.23 | -0.38 | 0.79 |
| RSI | Utilities | 27 | 0.26 | 0.22 | 0.04 | 0.18 | 0.34 | -0.03 | 0.68 |
| Random Forest | Communication Services | 21 | 0.3 | 0.69 | 0.15 | 0.01 | 0.59 | -0.43 | 2.22 |
| Random Forest | Consumer Discretionary | 57 | 0.44 | 0.67 | 0.09 | 0.27 | 0.62 | -0.71 | 2.38 |
| Random Forest | Consumer Staples | 30 | 0.22 | 0.36 | 0.07 | 0.09 | 0.35 | -0.38 | 1.26 |
| Random Forest | Energy | 21 | 0.58 | 0.89 | 0.19 | 0.2 | 0.96 | -0.42 | 2.86 |
| Random Forest | Financials | 61 | 0.38 | 0.49 | 0.06 | 0.25 | 0.5 | -0.65 | 1.35 |
| Random Forest | Health Care | 59 | 0.58 | 0.84 | 0.11 | 0.37 | 0.8 | -0.38 | 4.19 |
| Random Forest | Industrials | 67 | 0.45 | 0.56 | 0.07 | 0.32 | 0.59 | -0.37 | 2.81 |
| Random Forest | Information Technology | 62 | 0.44 | 0.75 | 0.09 | 0.26 | 0.63 | -0.48 | 3.25 |
| Random Forest | Materials | 23 | 0.44 | 0.5 | 0.1 | 0.23 | 0.64 | -0.16 | 2.15 |
| Random Forest | Real Estate | 30 | 0.28 | 0.39 | 0.07 | 0.14 | 0.42 | -0.37 | 1.37 |
| Random Forest | Utilities | 27 | 0.17 | 0.32 | 0.06 | 0.05 | 0.29 | -0.34 | 1.07 |

Figure 1

*Percent of actual return by strategy*

Chart, box and whisker chart

Description automatically generated

Figure 2

*Percent of actual return by sector, strategy*

Chart

Description automatically generated

1. The entire study was conducted using open-source data (i.e., quantmod and yahoo finance) and coded in R language. All materials can be found here: https://github.com/myvioletrose/school\_of\_professional\_studies/tree/master/698.%20Capstone/algorithmic\_trading\_code [↑](#footnote-ref-1)