International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 2 Feb. 2023, pp: 254-259 www.ijaem.net ISSN: 2395-5252

# Methodology to identify and trade high potentialstocks using Machine Learning and AlgorithmicTrading

Sidharth Sunilkumar

Date of Submission: 01-02-2023 Date of Acceptance: 10-02-2023

**ABSTRACT**—This paper explores a methodology to identify and trade high potential stocks using various combinations and variations of algorithmic trading methods. Various algorithmic trading methods and machine learning algorithms are analyzed and best performing ones are used to create a stock trading methodology that maximizes profit and minimizes risk.

## I. INTRODUCTION

Algorithmic trading is a process where computer programs are used for executing trade orders instead of orders placed manually. Generally, some strategies are used to decide the specifics of the trade. It is usually implemented considering various factors of a company like a company's assets, share price, revenue, etc. So one might assume the stock prices of a company can be predicted by looking at various financial aspects of the company. If this was true, then creating a foolproof trading strategy would be easy and everyone would make money. But therein lies the problem. Stock Markets are inefficient i.e. share prices do not accurately reflect a company's performance. So algorithmic trading is not easy to maintain and execute.

In this paper, we explore a methodology to identify and trade high potential stocks using machine learning and various algorithmic trading strategies. Trading has two important steps choosing the stocks to trade and then deciding at what price to trade the stock. For the first step, machine learning was used to used to predict the expected growth of stocks. Based on the results, some stocks can be considered for trading. Then those stocks can be traded with various algorithmic trading methods. To decide which trading method is best, performances of vari- ations and combinations of various algorithms were evaluated.

## II. RELATED WORK

There are two important steps in trading. First is to choose the stock to trade and second is to trade those chosen stocks. For choosing stocks, various machine learning algorithms were used to predict the growth of stocks.

Paper [Joseph D. Piotroski 2002] shows how to calculate the Piotroski F-score of a firm. This is a discrete score between 0-9 that reflects 9 criteria used to determine the strength of a firm's financial position. This score can be used to separate high potential stocks from lower potential ones. These 9 factors were first considered as the inputs for the Machine Learning models in order to predict the price growth of stocks.

For trading, 9 different trading strategies were implemented and tested to pick out the best performing strategies. These strategies are discussed below.

Dual moving crossover trading [2] based on a technique called moving average [3] which is a simple technical analysis tool that smooths out price data by creating a constantly updated average price. The average is taken over a specific period of time the trader chooses. It can be used to determine the moving trend of stock price of a company. There are many types of moving average strategies. One of which is called Crossover strategy [4]. It signals a potential change in trend when the price of a stock crosses the moving average line. The dual moving average crossover trading is a variation of the above strategy where two moving averages are used, one short and one long. Then trades are executed when the short moving average line crosses the long moving average line. When the short term moving average crosses the long term one by going lower, it is recommended to buy stocks and sell it once it crosses the long term line by going higher. A determining factor of the success of this strategy is the time periods taken for the short and long moving averages. Another factor

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 2 Feb. 2023, pp: 254-259 www.ijaem.net ISSN: 2395-5252

is the type of moving average used. For example, simple moving average, exponential moving average, weightedmoving average, etc.

Three moving average crossover strategy [9] is a variation of the above strategy. The difference being that is uses three different moving averages [3] rather than two. All three moving averages should be of different time spans- short moving average (SMA), medium moving average (MMA) and a long moving average (LMA). Apart from the three moving averages, it also uses two flags, a long flag and a short flag.In the strategy, a stock is bought when either of the following two conditions are true. The MMA is lesser than the LMAand the SMA is lesser than the MMA. In this case, the short flag is set to true. The other condition is when the MMA is greater than the LMA and the SMA is greater than the MMA. In this case, the long flag is set to true. Then the stock is soldif the SMA is greater than the MMA and the short flag is true, or the SMA is lesser than the MMA and the long flag is set to true. After the stock is sold, all flags are reset to false and the cycle continues.

Another strategy is the Moving Average Convergence Di- vergence (MACD) [10]. It is another indicator that shows the price momentum of a stock. In this method, two exponential moving averages, one fast moving average and one slow mov-ing average, are calculated. Then their difference is plotted. The difference is called a Signal line. The slope of the signal line determines the momentum of the price.

On-balance volume (OBV) [6] is a technical trading mo- mentum indicator that uses volume flow to predict changes in stock price. Volume (in trading) shows crowd sentiment as price bars carve out patterns that predict a bullish or bearish outcome. On-balance volume provides a running total of an asset's trading volume and indicates whether this volume is flowing in or out of a given security or currency pair. It generates a smooth indicator line that carves out highs, lows, and trend lines similar to price bars. OBV indicator line is calculated using 3 rules. If today's closing price is higher than yesterday's closing price, then current OBV is the sum of previous OBV and today's volume. If today's closing price is lower than yesterday's closing price, then the current OBV is the difference of previous OBV and today's volume. If today's closing price equals yesterday's closing price, then the current OBV is the same as previous OBV. Comparing the OBV and the exponential moving average of OBV signals the price trendof a stock.

Bollinger Bands [7] is an indicator to determine if a stockis overbought or oversold at a particular time. When a stock is overbought that usually means the price on the stock is bound to decrease. The vice versa is also true. Bollinger Bands are composed of three lines. The middle line is calculated using a simple moving average. The upper band is calculated by taking the middle band and adding twice the daily standard deviation to that amount. The lower band is calculated by taking the middle band and subtracting two times the daily standard deviation. When the price of the stock breaks below the lower band of the Bollinger Bands, it signifies that the stock is oversold. On the other hand, when the price breaks above the upper band, the market is perhaps overbought.

The Money Flow Index (MFI) [8] is a technical oscillator that uses price and volume data for identifying overboughtor oversold signals for an asset. It can also be used to spot divergences which warn of a trend change in price. The oscillator moves between 0 and 100. When this oscillatorgoes above a certain threshold, it signifies that a stock isoverbought and will likely decrease soon. On the other hand,if the oscillator falls falls below a certain threshold, it signifiesthat the stock is oversold.

## III. METHODOLOGY

### *A.* Stock selection process

For the stock selection process, stocks are selected by its expected price growth for the next quarter. The expected price growth is calculated by taking various financial data ofa company that conveys the its financial performance. This data can be taken from the financial statements published by the company at the end of every quarter. The expected price growth is calculated as follows-

$$Expected\ Price\ Growth = \left(\frac{P_{q+1} - P_q}{P_q}\right) * 100$$

where $P_n$ is the average stock price of a stock in a particularquarter 'n'.

6 different Machine Learning models were tested to pre-dict the expected price growth of stocks. The models tested were Mulivariate Linear Regression, Support Vector Machine (Linear kernel), Support Vector Machine (Polynomial kernel),Support Vector Machine (RBF kernel), Support Vector Ma- chine (Sigmoid kernel) and Random Forest Regressor.

The 9 factors used in Piotroski F-score method are Return on Assets, Operating Cash Flow, Change in Return of As- sets, Accruals,

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 2 Feb. 2023, pp: 254-259 www.ijaem.net ISSN: 2395-5252

Change in Leverage, Change in Current ratio, Change in the number of shares, Change in Gross Marginand Change in Asset Turnover ratio. These factors were taken as inputs for the Machine Learning models to predict the expected price growth of stocks. To improve the accuracy of the predictions, the inputs were divided into 2 sets - one with the 9 factors and other with percentage change of the factors from the previous quarter. Then all combinations of inputs, each input picked either from set 1 or 2, were tested to getthe most accurate result. To try and improve the accuracy, Earnings per share and Diluted earnings per share were also added to the set of inputs.

### *B.* Stock trading process

For the stock trading process, 9 trading strategies were first implemented. The 9 strategies are discussed above in chapter Related Works. These strategies were tested using historical data of stocks. Its performance was measured using 2 criteria average profit and average risk. Average risk was calculated as follows -

$$Avg.\ risk = \left(\frac{L_n}{L_n + P_n}\right) 100$$

where $L_n$ is the number of transactions that resulted in loss and $P_n$ is the number of transactions that resulted in profit.

During testing, it was observed that some strategies yielded more profits than others whereas some strategies had lesser risk. To potentially improve the strategies, 3 strategies were created by combining other strategies - Bollinger bands + three moving average, On-Balance moving average + exponential moving average, and Money Flow index + exponential moving average. All 12 strategies were tried with various parameters and used to trade.

## IV. OBSERVATIONS AND INFERENCES

The stocks of the S&P 500 companies were used for testing. All their historical financial data were obtained using FinancialModeling Prep API.

### *A.* Stock selection process

For the stock selection process, quarterly results of the S&P 500 stocks since its Initial Public Offering were used for testing.

In the first test, the 9 factors of Piotroski F-score were taken as the inputs for 6 Machine Learning models to predict the expected price growth of the stock. The R-squared scores for the test are shown in Table I under column 'T1'. For the second test, the some of the input parameters were modified to its percent change from previous quarters. All combinations of inputs were tested and the best one was taken. The R-squared score of the models are shown in Table I under column 'T2'. The algorithm was then combined with input factors suchas Earnings per share and Diluted earnings per share which resulted in further increase of performance of the models. The R-squared score of the models are shown in Table I under column 'T3'.

| R-squared Score Table | | | |
|---|---|---|---|
| **Model Name** | **T1** | **T2** | **T3** |
| Multivariate Linear Regression | -0.08 | 0.13 | 0.19 |
| SVM - Linear | -0.02 | 0.21 | 0.26 |
| SVM - Polynomial | -0.18 | 0.06 | 0.09 |
| SVM- RBF | 0.12 | 0.23 | 0.27 |
| SVM - Sigmoid | -0.21 | 0.14 | 0.20 |
| Random Forest Regressor | 0.06 | 0.32 | 0.36 |

TABLE I
R-SQUARED SCORE OF THE MACHINE LEARNING MODELS

For predicting the price growth of a stock, the model with the highest R-squared score for each stock is selected to predict the price growth of the stock using the latest quarterly results available. Stocks are then shortlisted according to their expected price growth.

### *B.* Stock trading process

For the stock trading process, 5 years of historical data (01- 01-2016 to 31-12-2020) of the S&P 500 stocks were used for testing.

First, 9 strategies discussed in the chapter Related Works section were implemented and tested. It was observed that some strategies gave more profits whereas other strategies hadlower risk. To potentially improve the strategies, 3 strategies were created by combining other strategies - Bollinger bands + three moving average, On-Balance moving average + expo- nential moving average, and

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 2 Feb. 2023, pp: 254-259 www.ijaem.net ISSN: 2395-5252

Money Flow index + exponential moving average. The 3 strategies were also implemented and tested. All 12 strategies were tested with various sets of parameters and given 100,000$ to trade. Considering each parameter set as a different method, the total number of algorithms tested was 399. Each algorithm's average risk and average profit were calculated. Fig. 1 shows the scatter plot for average risk vs average profit of the algorithms.

The next step was to shortlist the algorithms which max- imized average profit and minimized average risk. To do this, firstly, the dominated algorithms [9] were removed. This shortlisted 19 methods from the 399 algorithms. The short- listed non-dominated algorithms included different variations of Bollinger Bands, Money Flow Index, three moving average,Bollinger Bands + three moving average and Money Flow Index + exponential moving average. Fig. 2 shows the scatter plot for average risk vs average profit of the all 399 algorithms as blue dots and the non-dominated algorithms as red stars.

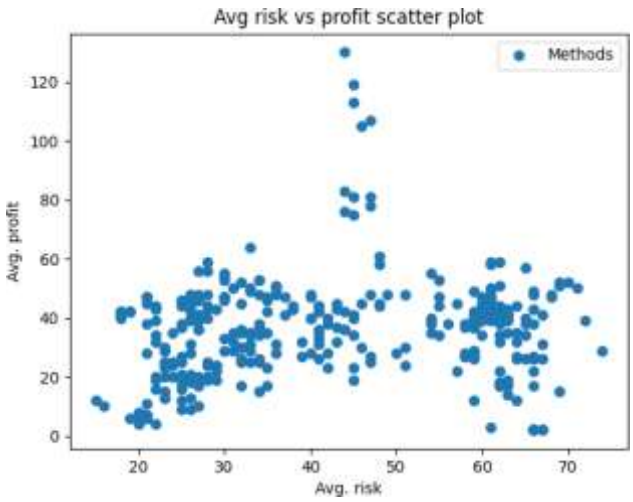


Fig. 1. Average risk vs average profit scatter plot of algorithms

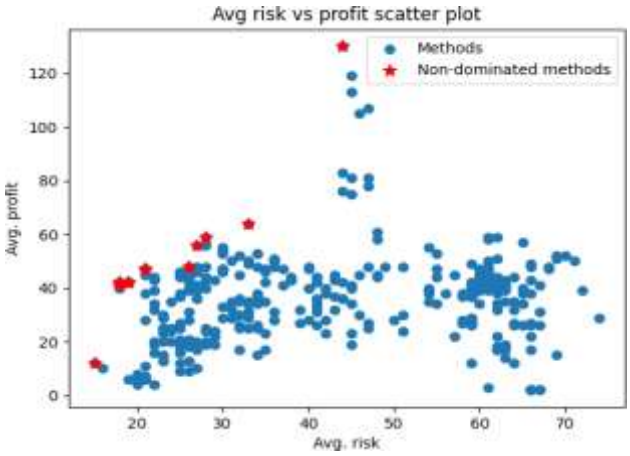


Fig. 2. Average risk vs average profit scatter plot of dominated and non-dominated algorithms

To further shortlist the algorithms, the crowding distance [9] of each algorithm was considered and the top 5 were selected. Fig. 3 shows the scatter plot for average risk vs average profit of the all 399 algorithms as blue dots and the shortlistedalgorithms as red stars. The 5 shortlisted algorithms are shown in Table II.

### *C.* Combination of Stock selection process and Stock trading process

Table II shows the average risk and profit

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 2 Feb. 2023, pp: 254-259 www.ijaem.net ISSN: 2395-5252

from all the S&P 500 companies. In theory, the average risk should decrease and average profit should increase when the stocks shortlisted by the Stock selection engine is used. To test this, 20 quarters of historical data (01-01-2016 to 31-12-2020) was used. For each quarter, the top 10 shortlisted stocks were given 100,000$ and traded with the shortlisted algorithms. The results are showsin Table III.

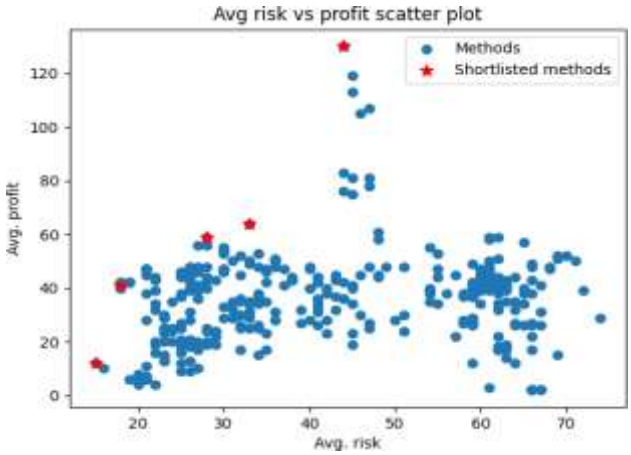


Fig. 3. average risk vs average profit scatter plot

| **Top 5 shortlisted algorithms** | | | |
|---|---|---|---|
| **Algorithm name** | **Parameter set** | **Average risk** | **Average profit($)** |
| Three moving average | 10,20,30 | 0.44 | 13013 |
| Money Flow Index | 5,(60,40) | 0.33 | 64270 |
| Money Flow Index | 12,(70,30) | 0.28 | 59070 |
| Bollinger Bands + Three mov-ing average | 100,5,20,50 | 0.18 | 41500 |
| Money Flow Index + Expo-nential moving average | 100,15,30 | 0.15 | 12426 |

TABLE II
AVERAGE RISK AND PROFIT TABLE FOR TOP 5 SHORTLISTED METHODS

## V. RESULTS

1) Using factors of Piotroski F-Score algorithm as inputs for Machine Learning models and then modifying them improved the accuracy of the models when predicting future prices of stocks.
2) Variations and combinations of different algorithmic trading methods yielded better profits and at the same time minimized risk.
3) Combining the Stock selection process with the Stock trading process improved the performance of the algo-rithmic trading methods.

## REFERENCES

[1] Joseph D. Piotroski. 2002. ”Value Investing: The Use of Historical Financial Statement Information to Separate Winners from Losers” URL: **https://www.chicagobooth.edu/ /media/FE874EE65F624AAEBD0166B1974FD74D.pdf** [Accessed: Mar 01,2021]

[2] Wisdomtrading ”DUAL MOVING AVERAGE CROSSOVER SYS- TEM EXPLAINED” URL: **https://www.wisdomtrading.com/public-trading-systems/dual-moving-average-crossover-trading-system** [Ac- cessed: Mar 01, 2021]

[3] Investopedia ”How to Use a Moving Average to Buy Stocks” URL: **https://www.investopedia.com/articles/active-trading/052014/how- use-moving-average-buy-stocks.asp** [Accessed: Mar

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 2 Feb. 2023, pp: 254-259 www.ijaem.net ISSN: 2395-5252

01, 2021]

[4] Investopedia "Trading Strategies—Crossovers" URL: **https://www.investopedia.com/articles/active-trading/052014/how- use-moving-average-buy-stocks.asp** [Accessed: Mar 01, 2021]

[5] Financial Modeling Prep "Financial Modeling Prep API" URL: **https://financialmodelingprep.com/developer/docs** [Accessed: Mar01, 2021]

[6] Investopedia "On-Balance Volume" URL: **https://www.investopedia.com/terms/o/onbalancevolume.asp** [Accessed: Mar 01, 2021]

[7] Investopedia "Using Bollinger Bands to Gauge Trends" URL: **https://www.investopedia.com/trading/using-bollinger-bands-to- gauge-trends/** [Accessed: Mar 01, 2021]

[8] Investopedia "Money Flow Index - MFI Definition and Uses" URL: **https://www.investopedia.com/terms/m/mfi.asp** [Accessed: Mar 01,2021]

[9] NewTraderU "3 Moving Average Crossover Strategy" URL: **https://www.newtraderu.com/2020/11/29/3-moving-average- crossover-strategy/** [Accessed: Mar 01, 2021]

[10] Babypips "How to Use the MACD Indicator" URL: **https://www.babypips.com/learn/forex/macd** [Accessed: Mar 01, 2021]

[11] NPTEL "Multi-objective Optimization" URL: **https://www.youtube.com/watch?v=Hm2LK4vJzRwab**channel = IITKharagpurJuly2018[Accessed : Mar01, 2021]

| | All | | Shortlisted | | Change | |
|---|---|---|---|---|---|---|
| **Method name** | **Avg. Risk** | **Avg. Profit ($** | **)Avg. Risk** | **Avg. Profit ($** | **)Risk change** | **Profit change** |
| Three moving aver-age | 0.44 | 3307 | 0.39 | 6348 | -11.36% | 91.92% |
| Money Flow Index (5, (60, 40)) | 0.34 | 2545 | 0.29 | 4314 | -14.70% | 69.45% |
| Money Flow Index (12, (70, 30)) | 0.37 | 1506 | 0.31 | 1742 | -14.86% | 15.65% |
| Bollinger Bands + Three moving average | 0.34 | 924 | 0.14 | 1364 | -59.41% | 47.62% |
| Money Flow Index + Exponential mov-ing average | 0.47 | 179 | 0.10 | 268 | -78.68% | 49.40% |

TABLE III
AVERAGE RISK AND PROFIT TABLE FOR TOP 10 SHORTLISTED STOCKS VS ALL S&P 500 STOCKS