

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/301199235>

Stock Trading Using Analytics

Article · January 2016

CITATIONS

8

READS

6,741

2 authors:



Chandrika Mani

National University of Singapore

2 PUBLICATIONS 30 CITATIONS

[SEE PROFILE](#)



Carol Hargreaves

National University of Singapore

68 PUBLICATIONS 397 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Pairs Trading: An Application in the Australian Stock Market [View project](#)



Project on Generative Adversarial Networks [View project](#)

Stock Trading Using Analytics

Chandrika Kadirvel Mani*, Carol Anne Hargreaves

Business Analytics, Institute of Systems Science, National University of Singapore, Singapore, Singapore

Abstract

The primary challenges with stock trading are the identification of the profitable stocks and trading those stocks without human errors and interference of personal sentiments in order to reap better returns. In this paper, we propose the automated stock trading in a semi-automated manner. We combine the usage of both fundamental and technical variables in the prediction of profitable stocks using the machine learning algorithm SVM. These stocks are then traded using stop loss/gain criteria. We buy/hold/sell the stocks based on the rules configured in the self-developed automated trading application.

Keywords

Stocks, Auto Trading, Support Vector Machines, Trend Following, Machine Learning

Received: February 24, 2016 / Accepted: March 9, 2016 / Published online: March 18, 2016

@ 2016 The Authors. Published by American Institute of Science. This Open Access article is under the CC BY license.

<http://creativecommons.org/licenses/by/4.0/>

1. Introduction

Companies or corporations raise their capital through the issue and subscription of shares via exchanges or over-the-counter markets by giving investors a slice of ownership in the company. A trader seeks to profit from the price fluctuations of the shares held. Trading in the stock market can be very profitable or painfully unprofitable based on the stock-picking methodology.

In order to select profitable stocks, basically there are two different stock picking methodology namely fundamental and technical analysis. Fundamental analysis takes into account the overall economy, industry conditions, financial conditions and management of the organizations whereas in technical analysis instead of measuring the intrinsic value it mainly studies the stock charts to identify patterns and trend of the stock market. These two techniques are generally considered as two sides of a coin. It is always a debate which methodology is best [1].

With the advent of information technology, trading methods have considerably evolved from an open outcry manner on the stock exchange floor of the stock market to modern stock

trading conducted via electronic exchanges. In early 2000 the e-trading has taken a new dimension where trades are planned and executed by complex algorithms rather than humans. Auto trading executes trading systematically by eliminating the human emotional impacts and errors which results in increased profit for the trader and market liquidity. Another major advantage is that can trade continuously on all opened markets in the same time, applying the same algorithms over and over again, without any pause. Technical analysis is the foundation for Auto trading. Currently there are decision making algorithms such as Pairs Trading, Automated Gaming and Execution Algorithms such as Iceberg, Volume Weighted Average Price are implemented for algorithmic trading [2].

Algorithmic trading is widely used by investment banks, pension funds, mutual funds, and other buy-side (investor-driven) institutional traders, to divide large trades into several smaller trades to manage market impact and risk. Also algorithmic trading helps banks to trade foreign currency in order to reduce the risk encountered by the misconduct by human traders.

Motivation of automated trading strategy dates back to

* Corresponding author

E-mail address: chandrika.kadirvelmani@gmail.com (C. K. Mani)

1950's with portfolio optimization theory by Harry Markowitz in which he explains how an investor should allocate his wealth over risky securities so as to maximize his expected utility of total wealth. The second remarkable milestone is the development of Capital Asset Pricing model by Sharpe (1964), Lintner (1965) and Mossin (1966). This model determines theoretically appropriate required rate of return of an asset, if that asset is to be added to an already well-diversified portfolio, given that non-diversifiable risk of the asset. The third milestone is the linear multifactor risk model by Rosenberg (1974) in which individual stock returns were assumed to be related to a smaller number of common factors which was statistical and computational. A relatively recent innovation in financial trading is high frequency trading a form of automated trading that utilize the computational innovation and advancement in telecommunication. This trading accounts for 40 to 60% of all trading activity across the financial markets, including stocks, derivatives and liquid foreign currencies. [17].

The major benefits rendered by auto trading to the financial markets operating are liquidity, transaction costs and the efficiency of market prices.

The purpose of this paper is to come up with good stock picks using machine learning algorithm, trade those picked stocks by using trading rules in the semi-automated approach in the Australian Stock Market. The paper is structured into 7 sections. While Section 1 is the introduction, Section 2 gives a brief literature review, Section 3 the objectives of the study, Section 4 a brief overview of the methodology and statistical methods used, Section 5 the statistical analysis results, Section 6 paper trading, Section 7 the automated trading application results, after which Section 8 presents the conclusion.

2. Literature Review

In [3], the authors have proposed an algorithmic stock trading model that combines the signals from different technical indicators such as Moving Average Convergence-Divergence MCDM, Price Rate of Change and Stochastic Oscillator in order to provide more accurate trading signals.

Further, in [4], the authors have simulated an artificial market in which the different types of traders such as random agents, human agents, market makers agent and strategic agent trade based on certain rules.

In addition, in [5], the authors have studied the four algorithmic trading strategies. The first strategy is based on Static Order Book Imbalance in which the volume weighted averages of the prices are computed in the buy and sell order books. Then two differences are computed from each average

and the last price of the exchange. Depending upon the difference in the sell/buy side greater than the threshold buy/sell order is placed. The second strategy is based on the volume average weighted prices in which the weighted prices for the entire market is computed and compared with the average price for the first orders in the buy book. Based on the higher/lower than the threshold hold value buy/sell order placed. The third strategy is based on the trend following which depends only on the price movements. In this, two trend lines, with the help of regression is computed based on the 4 hour and 1 hour window. If the slope of the two trend line matches then a buy order is placed. The final strategy is reverse strategy in which buy order is placed when the price is falling and sell order is placed when the price is rising. It has been concluded that SOBI, reverse and TF are better.

Also in [6], the authors have combined the technical analysis with the nearest neighbour classification. The technical indicators such as moving averages, Relative strength index and Stochastics and Bollinger bands have been considered in this work. Apart from these indicators, the features extracted from the price history and trading volumes of the stocks has been used as input to K-NN algorithm. Based on the output from K-NN, stock is either bought/sold. Also stop/loss criteria has also been considered.

All of the above authors have all successfully demonstrated the algorithmic trading strategies of stocks. They have all used only the technical indicators based on past price and volume to predict the stock performance. It is a fact that only a good technical indicators do not always lead to better returns. Same holds good for fundamentals too. An organization which is having upper trend in the stock movement currently but if it does not have good fundamental say management effectiveness then the trend would not be a long standing one. In our research we intend to use both fundamental and technical indicator data in order to reap the benefits of both the methodologies. Also we have used the machine learning algorithm, support vector machine to predict the trend of the stocks.

Finally we hypothetically trade the top 5 stocks suggested by SVM using the auto trading application developed based on the trading rules such as stop/loss criteria.

3. Objective of the Study

Our research intends to demonstrate scientifically how analytical techniques can help to select good stocks and then such stocks traded using auto trading approach in a semi-automated process yield profit. Our research would help to identify the good stocks to invest by using machine learning algorithm. With a semi-automated process, selected stocks will be traded in a faster and smarter way without

interference of human sentiments ensuring a significance increase in profitability. Investment firms/brokers can use this approach to make money.

The Figure 1, illustrates the systematic approach to achieve the research objective. Using fundamental and technical

indicators we intend to predict the stocks with the upward trend movement by using machine learning algorithm. Infer rules and use them for making decisions such as buy, sell or hold the stocks. Automate the trading based on the algorithm developed using the rules and analyse the inferences.

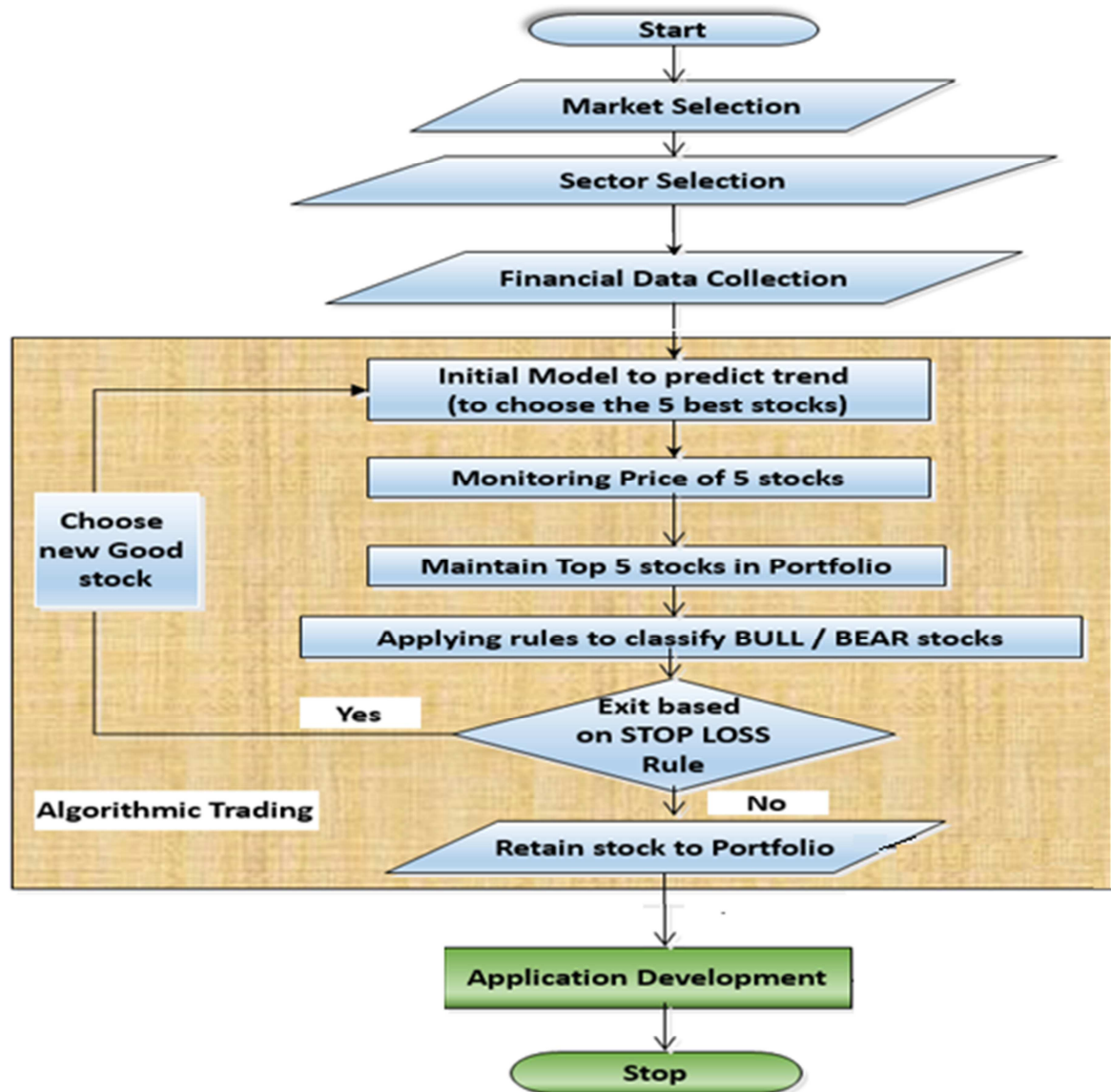


Figure 1. Systematic approach for Auto trading.

4. Methodology

4.1. Market Selection

Though there exists so many stock markets, we have chosen Australian stock market for our study as it is having good global economic position. There are around 2,100 stocks traded in Australian Exchange.

4.2. Sector Selection

The Australian stock market ASX is divided into the

following sectors namely Consumer Discretionary, Consumer Staples, Energy, Financials, Health Care, Industrials, Information Technology, Materials, Metals/Mining, Telecommunications Services, Utilities and Real Estate [10]. Out of these 12 sectors, in order to keep our research focused we decided to choose 1 good performing sector. We started our research study during the first quarter of 2015. So as per the sector trend analysis during the 1st quarter of 2015, the health care sector was performing well as depicted in the following figures 2.

HEALTH CARE (SECTOR)



Figure 2. Upward trend in healthcare sector.

ENERGY (SECTOR)



Figure 3. Downward trend in energy sector.

For Health Care sector the trend is upwards which means it is a good sign. The energy sector trend is depicted in figure 3 and the trend is downwards which means it is not a good sector to choose. There are totally 98 stocks in health care sector.

4.3. Data Collection

The fundamental indicators for the first quarter of 2015 for

each of the 98 health care stocks, is collected from yahoo finance [7] by web scraping using webscraper.io [8]. The figure 4 shows the highlighted fundamental variables collected from yahoo finance.

The data is collected for 3 different periods such as Jan, Feb and Mar 2015 for our research, in order to ensure repeatability and consistency of our modelling approach results.

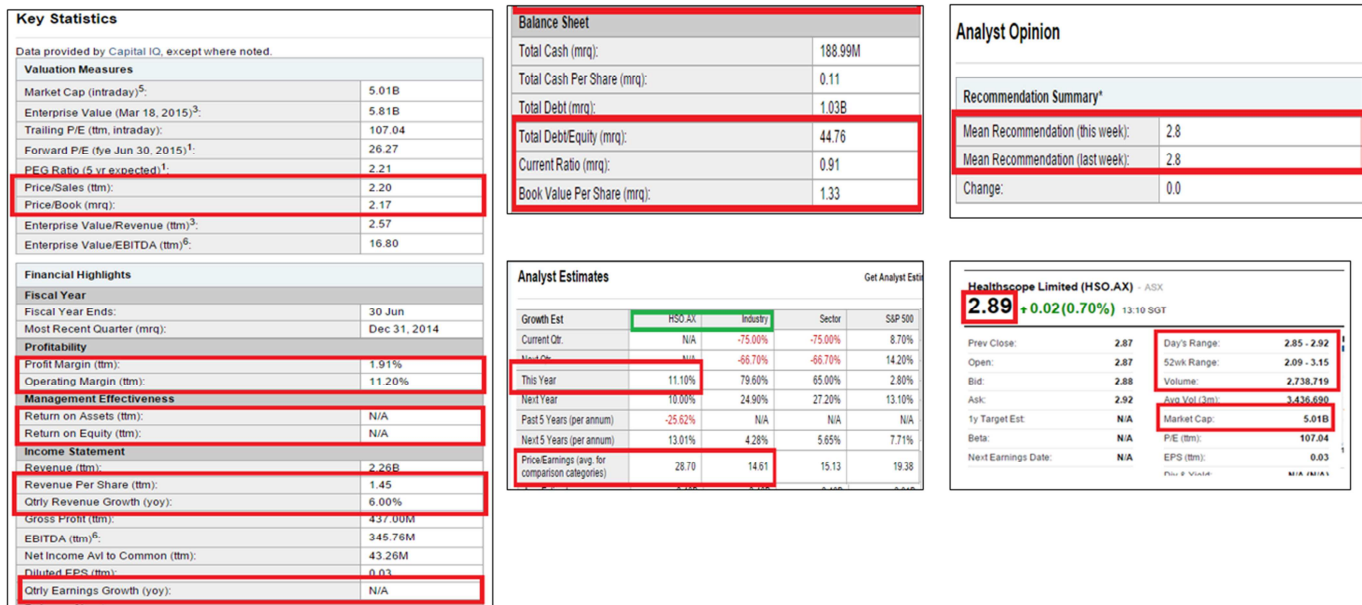


Figure 4. Variables collected from yahoo finance.

The Trend variable is collected from ASX. When the trend is upward as shown in figure 5, this indicates a good stock and hence the variable trend is coded as 1. If the trend is downwards as shown in figure 6 then it indicated not a good stock and hence the variable trend is coded as 0 for such stocks.

PRY, PRIMARY HEALTH CARE ORD

The chart of daily prices over 6 months for security PRY



Figure 5. Upward trend healthcare stock.

CUV, CLINUVEL PHARMACEUTICALS ORD

The chart of daily prices over 6 months for security **CUV**



Figure 6. Downward trend healthcare stock.

4.4. Data Cleaning

For the Health care sector, once the data is collected we formatted the data in a structured format so that we can further process using statistical tools. Then we checked for the presence of incorrect, incomplete and duplicate data. The missing data was imputed using Expectation maximization method. Also certain stocks which are not currently traded, stocks with volume of trade less than 10000 and also stocks having price less than 10 cents were removed from the analysis.

4.5. Trend Analysis

Trend analysis tries to predict the upward or downward trend for each of the stocks. There are various methods such as current market price calculation, moving averages and channel breakouts for trend following. Previous studies has compared the different trend following strategies such as momentum, channel, dual moving average crossover and time series momentum strategies [10] and confirmed that such trend following approach are indeed profitable in stock trading when followed. Also in another previous study which was conducted based on the inspiration of the Performance Probability Score (PPS) model, proved the possibility that trend following can yield profits [11].

In our study for each of the stocks, we tried to infer the

monthly trend of the stock by comparing the current day stock price with the previous day stock price. If the current day stock price is greater than the previous day stock price for about 40% of the trading days in a month then we conclude that the stock is having upward trend else it is considered to have downward trend for that particular month.

4.6. Support Vector Machines

Initially, we started to fit more simple models like logistic regression and decision trees and then even proceeded with ensemble models but they were unable to predict the upward trend stocks. The sensitivity was very low for most of these models. The SVM was giving relatively a very good accuracy, sensitivity and specificity. Hence we selected SVM.

Support Vector Machines (SVMs) are supervised learning models that analyse and recognize data for classification and regression analysis. A support vector machine constructs a hyper plane or set of hyper planes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks [12].

When there are numerous machine learning algorithms, the reason for choosing SVM is to address the unbalanced data problem more efficiently. The unbalanced classification problem occurs when there are many more samples of some classes than others. In our case we have 88% of stocks with

downward trend and 12% stocks with upward trend. The standard classifiers tend to be overwhelmed by the large-scale classes and ignore the small ones and tend to classify any example as belonging to the majority-class. Even though this is highly accurate under the standard measure of accuracy such a classifier is useless because it will be able to predict which stocks will tend to have downward trend and not the upward trend. We are more interested in the stock with upward price trend than downward price trend as we are interested to know the most profitable stock for investment.

Also in one of the previous research study, it was proved that SVM outperforms the other classification methods in the predictability of financial movement direction were they forecasted the weekly movement direction of NIKKEI 225 index [13].

SVMs belong to the class of algorithm kernel methods. A kernel method is an algorithm that depends on the data only through dot-products.

It is important to decide on which kernel to use and setting its parameters.

In our research we have a two class problem, the labels '1' and '0' denote the upward and downward trend respectively. Let us assume x denote a vector with components x_i . The notation x_i will denote the i^{th} vector in a dataset

$\{(x_i, y_i)\}_{i=1}^n$, where y_i is the label associated with x_i . The objects x_i are called patterns or examples.

A linear classifier is based on a linear discriminant function of the form

$$f(x) = w^T \phi(x) + b.$$

The vector w is known as the weight vector, and b is called the bias. Consider the case $b = 0$ first. The set of points x such that $w^T x = 0$ are all points that are perpendicular to w and go through the origin - a line in two dimensions, a plane in three dimensions, and more generally, a hyperplane. The bias b translates the hyperplane away from the origin. The hyperplane divides the space into two.

$$\{x: f(x) = w^T \phi(x) + b = 0\}$$

The boundary between regions classified as upward trend stocks and downward is called the decision boundary of the classifier. A classifier with a linear decision boundary is called a linear classifier. Conversely, when the decision boundary of a classifier depends on the data in a non-linear way the classifier is said to be non-linear. In our case it is a non-linear classifier as it provided better accuracy.

The naive way of making a non-linear classifier out of a linear classifier is to map our data from the input space X to a feature space F using a non-linear function $\phi: X \rightarrow F$. In the

space F the discriminant function is $f(x) = w^T \phi(x) + b$.

The approach of explicitly computing non-linear features does not scale well with the number of input features: when applying the mapping from the above example the dimensionality of the feature space F is quadratic in the dimensionality of the original space. This results in a quadratic increase in memory usage for storing the features and a quadratic increase in the time required to compute the discriminant function of the classifier. This quadratic complexity is feasible for low dimensional data; but when handling high dimensional data, quadratic complexity in the number of dimensions is not acceptable. Kernel methods solve this issue. In terms of the kernel function the discriminant function is

$$f(x) = \sum_{i=1}^n \alpha_i k(x, x_i) + b.$$

$i=1$

The distance from the decision surface to the closest data point determines the margin of the classifier. Maximizing the margin seems good because points near the decision surface represent very uncertain classification decisions: there is almost a 50% chance of the classifier deciding either way.

Training an SVM finds the large margin hyper plane, i.e. sets the parameters α_i and b . The SVM has another set of parameters called hyper parameters. The soft margin constant, C , and any parameters the kernel function may depend on kernel choice. In our case we have chosen polynomial of degree 2 kernel as it gave better accuracy when compared with linear or Gaussian or RBF kernel. The polynomial kernel of degree d is defined as follows

$$k(x, x') = (x^T x' + 1)^d.$$

5. Statistical Analysis Results

The results of trend following and performance of the machine learning algorithms are visualized using confusion matrix [10]. The upward and downward trends are denoted by 1 and 0 respectively. Sensitivity measures the proportion of upward trend stocks that are correctly identified. Specificity measures the proportion of downward trend stocks that are correctly identified. Accuracy is the proportion of both upwards and downward trend stock resulting among the total number of stocks examined.

5.1. Results of Trend Analysis

The trend analysis logic was implemented using R programming language. The predicted trend values are compared with the actual trend values collected as explained in the section 4.3. The figure 7 below illustrates the confusion matrix for the Trend analysis.

Today's Price > Yesterday's Price				
	Predicted			
		1	0	Total
	1	8	2	10
	0	19	56	75
Actual	Total	27	58	85

Inference from Confusion Matrix		Value
Sensitivity		80%
Specificity		74.66666667
Accuracy		75.29411765

Figure 7. Confusion Matrix for Trend Analysis.

5.2. Results of Support Vector Machines

The financial data collected from yahoo finance and the stock monthly trend data resulted from our trend analysis are used as the independent and dependent variable respectively for

SVM. It was implemented using the R programming language. The package used is caret. The figure 8 below illustrates the confusion matrix for the support vector machines.

SVM				
	Actual			
		1	0	Total
	1	22	1	23
	0	4	58	62
Predicted	Total	26	59	85

Inference from confusion Matrix		Value %
Sensitivity		84.61538462
Specificity		98.30508475
Accuracy		94.11764706

Figure 8. Confusion Matrix for SVM.

5.3. Best Stock Pick for Trading

For the health care sector, the best stocks which are predicted to have upward trend having probability greater than 0.8 for each of the month is shown in the figure 9. The below stocks are the ones chosen for hypothetical automatic trading with

high probability of moving upward trend. In case of the model suggesting equal probabilities then we have prioritized the stocks based on the fundamental indicators such as return on equity, return on assets, revenue per share, book value per share [9] with high values but price having low value.

Jan-15	Feb-15	Mar-15
COH.AX	ANN.AX	RMD
CSL.AX	PRY.AX	CAJ
RMD.AX	SHL.AX	ANN
RHC.AX	CSL.AX	SHL
SHL.AX	COH.AX	GLH
SIP.AX	SRX.AX	PRY
PRY.AX	CAJ.AX	SVA
REG.AX	CYP.AX	API
GXL.AX	MYX.AX	
UBI.AX	FPH.AX	

Figure 9. Health Care sector - good stock picks.

The highlighted top 5 stocks are suggested for stock trading. The stocks predicted using Jan, Feb and March 2015 data are used for the Feb, Mar and April automated trading respectively. Based on the stop loss rule once a stock goes out of portfolio next stock in the list will go into the portfolio.

6. Paper Trading

Based on the stocks predicted by SVM, hypothetically paper trading was done following the stop loss criteria and we were able to achieve around 16.64% revenue at the end of 3 months. The algorithm outperformed the ASX returns and this is shown in the below visualization.

Paper trading Simulation – SVM

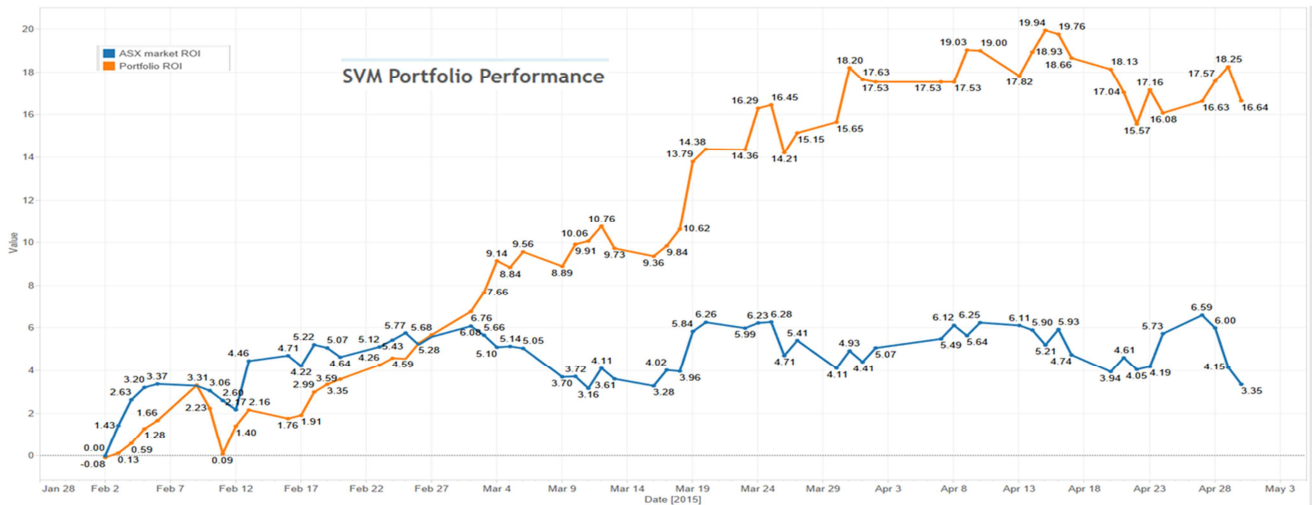


Figure 10. Visualization of returns from Paper Trading shown from Feb to April 2015.

7. Auto Trading Application

The methodology discussed in the previous sections for the selection of good stocks and trading rules followed for paper trading are implemented programmatically and semi-automatic trading application has been developed. The stocks predicted to have upward trend by SVM is given as input to the auto trading application. The application is programmed dynamically to pick the first 3 stocks from the list and invest equally the allocated fund. The total allocated fund is \$100K. Initially each of the three stocks are equally invested. The stop loss/gain rule needs to be configured in the application.

The rules configured are, sell the share when the daily returns from the share exceeds 10% in order to reap the gain or sell when the daily returns from the share goes below 3% in order to avoid further loss. The buying and selling charges are assumed to be 20\$ for each transaction. When one of the stock in the portfolio is sold due to profit or loss, the next stock in the list is automatically picked for trading.

The backend of the auto trading application is programmed using Microsoft EXCEL visual basic for applications. The front end is developed using R programming language using Shiny dashboard and rCharts packages giving interactive charts for the user to better understand their portfolio.

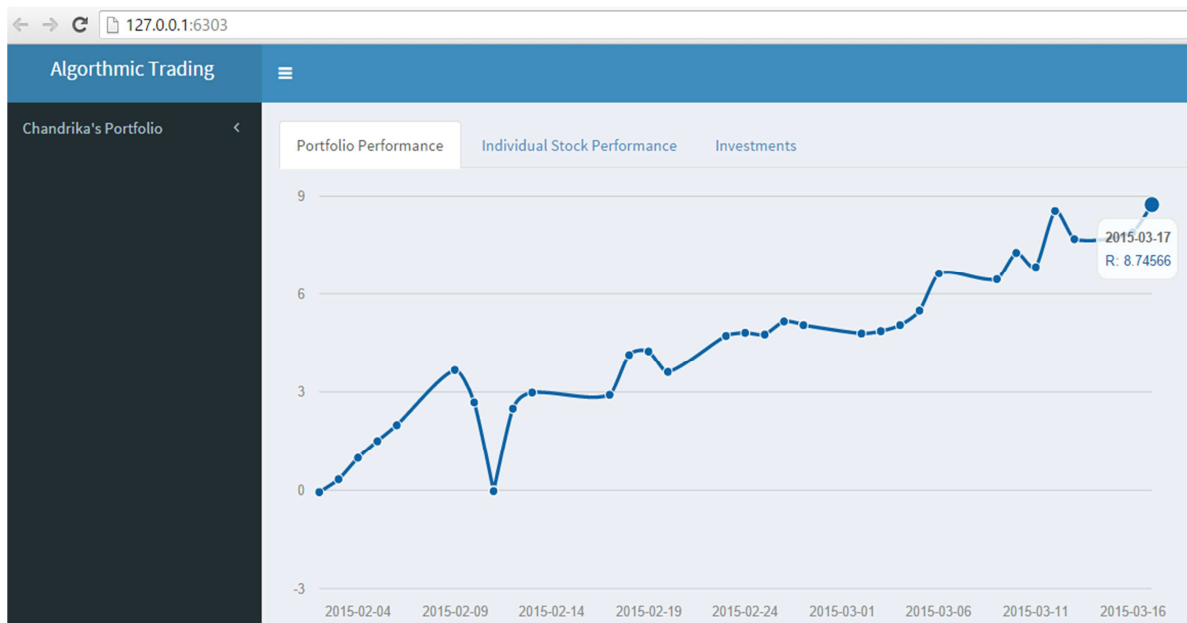


Figure 11. The below interactive chart provides visualization of the overall daily returns of the portfolio shown from Feb till mid of March 2015.

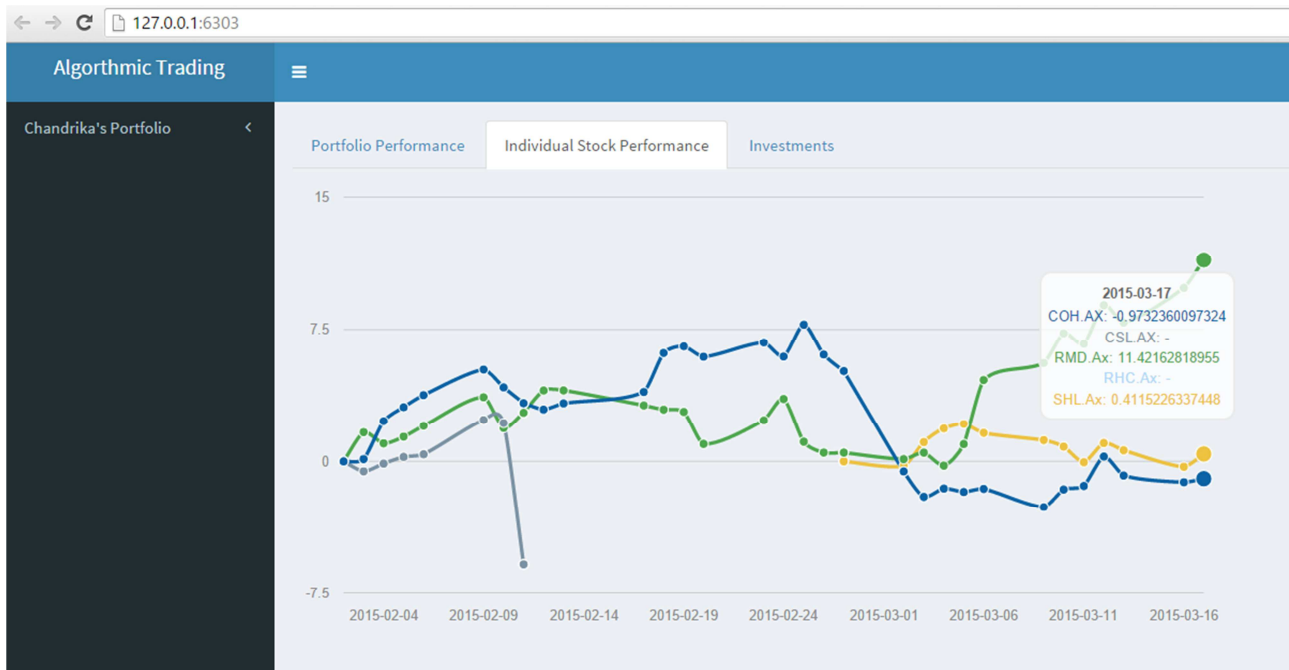


Figure 12. The below interactive chart provides visualization of the individual stock daily returns of the portfolio.

8. Conclusion

So far in our research we have achieved in screening the 2000+ stocks in Australian stock market based on the good performing sectors and come down to 98 stocks in health care sector. Out of these 98 stocks with the help of statistical modelling techniques we have further come down to 5 best stocks for each of the months namely Jan, Feb and March 2015. Thus we achieved to systematically pick few high return stocks (needles) from the hay of 2000+ stocks. These stocks were traded with the help of the auto trading application without human error and human sentiments interference to yield profit by configuring the trading rules in the application.

References

- [1] Fundamental and Technical Analysis. www.investopedia.com/ask/answers/131.asp
- [2] <https://www.sungard.com/~media/fs/capital-markets/resources/white-papers/Valdi-Algo-Trading-Complex-Map.ashx?sfcdCampaignId=701500000000Yyzo>.
- [3] Darie MOLDOVAN, Mircea MOCA, Ștefan NIȚCHI. "A Stock Trading Algorithm Model Proposal, based on Technical Indicators Signals". *Informatica Economică* vol. 15, no. 1/2011.
- [4] Daniel Paraschiv, Srinivas Raghavendra, and Laurentiu Vasiliu. "Algorithmic Trading on an Artificial Stock Market". C. Badica et al. (Eds.): *Intel. Distributed Comput., Systems & Appl.*, SCI 162, pp. 281–286, 2008. springerlink.com.
- [5] Gheorghe Cosmin Silaghi & Valentin Robu. "An Agent Strategy for Automated Stock Market Trading Combining Price and Order Book Information". 1-4244-0020-1/05/\$20.00 a2005 IEEE.
- [6] Lamartine Almeida Teixeir, Adriano Lorena Inácio de Oliveira. "A method for automatic stock trading combining technical analysis and nearest neighbor classification". *Expert Systems with Applications* 37 (2010) 6885–6890.
- [7] Fundamental data from Yahoo finance. <http://finance.yahoo.com/q/ks?s=HSO.AX+Key+Statistics>
- [8] Kadirvel Mani Chandrika (2015). "Web scraping in a simple way." <http://chandrikakadirvelmani.blogspot.sg/2015/04/web-scraping-in-simple-way.html>.
- [9] ASX charts for trend capturing. <http://hfgapps.hubb.com/asxtools/Charts.aspx>.
- [10] Andrew C. Szakmary, M. Carol Lancaster. "Trend-Following Trading Strategies in U. S. Stocks: A Revisit". *The Financial Review* 50 (2015) 221–255.
- [11] Simon Fong, Jackie Tai, Yain Whar Si, "Trend Following Algorithms for Technical Trading in Stock Market". *JOURNAL OF EMERGING TECHNOLOGIES IN WEB INTELLIGENCE* · MAY 2011.
- [12] Support Vector Machines. https://en.wikipedia.org/wiki/Support_vector_machine
- [13] Wei Huang; b, Yoshiteru Nakamoria, Shou-Yang Wang. "Forecasting stock market movement direction with support vector machine". *Computers & Operations Research* 32 (2005) 2513–2522.
- [14] Carol Anne Hargreaves, Chandrika Kadirvel Mani, "The Selection of Winning Stocks Using Principal Component Analysis". *American Journal of Marketing Research*, Vol. 1, No. 3, October 2015 Publish Date: Aug. 10, 2015 Pages: 183-188.
- [15] Confusion Matrix. https://en.wikipedia.org/wiki/Confusion_matrix

- [16] Asa Ben-Hur Department of Computer Science Colorado State University, Jason Weston NEC Labs America Princeton, NJ 08540 USA. "A User's Guide to Support Vector Machines".
- [17] Andrei A. Kirilenko and Andrew W. Lo, "Moore's Law versus Murphy's Law: Algorithmic Trading and Its Discontents". Journal of Economic Perspectives—Volume 27, Number 2—Spring 2013—Pages 51–72.