Does the use of technical & fundamental analysis improve stock choice? : A data mining approach applied to the Australian stock market

Conference Paper · September 2012

DOI: 10.1109/ICSSBE.2012.6396537

Carol Hargreaves
National University of Singapore

Does the use of Technical & Fundamental Analysis improve Stock Choice?

: A Data Mining Approach applied to the Australian Stock Market

Carol Hargreaves
Institute of Systems Science
National University of Singapore
Singapore
carol.hargreaves@nus.edu.sg

Yi Hao
School of Business Economics
Wilfrid Laurier University
Waterloo, Canada
haox5200@mylaurier.ca

***Abstract*—With the easy access to share information and data, many investors worldwide are interested in predicting stock prices. The prediction of stock prices using data mining techniques applied to technical variables has been widely researched but not much research to date has been done in applying data mining techniques to both technical and fundamental information. This paper is based on a personal approach to stock selection, using both technical and fundamental information. In this paper we construct a framework that enables us to make class predictions about industrial stock companies' financial performances. In order to have a systemized approach for the selection of stocks and a high likelihood of the performance of the stock price increasing, a Data Mining Approach is applied. A trading strategy is also designed and the performance of the stocks evaluated. Our two goals are to validate our stock selection methodology and to determine whether our trading strategy allows us to outperform the Australian market. Simulation results show that our selected stock portfolios outperform the Australian All-Ordinaries Index. Our findings justify the use of data mining techniques for classification and prediction purposes. Further, in conclusion, we can safely say that our stock selection and trading strategy outperformed the Australian Ordinary index.**

Keywords- stock price prediction; stock selection; stock market; data mining; decision trees; neural networks (NN); trading strategy

I. INTRODUCTION

A stock market is a private or public market for the trading of company stock and derivatives of company stock at an agreed price. Many financial companies such as stock markets produce large datasets and are looking to find efficient ways to discover useful information about stocks and the market for investment decisions. Further, with the easy access to stock information and data, many private investors worldwide are interested in predicting stock prices and hope to maximize on the opportunities in the market and become rich. The problem is:

- There are so many stocks in the market,
- Large amounts of stock information easily available on the internet, through newspapers, magazines, radio and television.

How can investors go about selecting stocks in a systematic way so that they select winning stocks and make a profit with minimum risk?

A. Aim of the Present Study

The focus of this paper is to investigate whether stocks selected by

- the application of a personal systematic approach or with a data mining approach will outperform the Australian stock market
- when monitored and managed using a trading strategy will outperform the Australian stock market.

B. Hypothesis

The following hypotheses will be tested in this study:

- H1: The stocks selected through a personal systematic approach outperforms the Australian stock market
- H2: The stocks selected through a data mining approach outperforms the Australian stock market

In the remainder of the paper, we explain the methodology for our study, by discussing our stock selection system, followed by our trading strategy. We conclude our paper with the results and the value of our study with possible further research ideas.

II. METHODOLOGY

Data mining can be described as "making sense of data". Over the last few decades, increasingly huge amounts of past data have been stored electronically and this volume is expected to grow considerably in the future. According to the

Gartner Group, "Data mining is the process of discovering meaningful new correlations, patterns and trends by sifting through large amounts of data stored in repositories, using pattern recognition technologies as well as statistical and mathematical techniques". Data mining is one way of predicting if future stock prices will increase or decrease. We treated our problem as an unsupervised task and therefore we created the class variable at the beginning and only after the class variable was constructed, we trained the classifier. We used decision trees and neural networks to classify our stocks as to whether they will increase or decrease in price. These stock predictions assisted us in our decision of whether to buy a stock or not.

A. Decision Trees

Decision trees are a form of multiple variable analyses and are powerful and popular tools for classification and prediction. The attractiveness of decision trees lies in their ease of interpretation, relative power, robustness with a variety of data and levels of measurement, and ease of use. Decision trees attempt to find a strong relationship between input values and target values in a group of observations that form a data set [3]. In contrast to neural networks, decision trees represent rules and rules can readily be expressed so that humans can understand them or even directly used in database access language like SQL so that records falling into particular category may be retrieved [6].

In our study, the decision trees generated some rules of how the performance of different industrial company stocks were predicted as data became available. For a detailed technical explanation of the two decision trees applied (CHAID and C5.0) in this study, the reader is referred to [7] for C5.0 and [3] for CHAID.

B. Neural Networks

Why use Neural Networks for forecasting stock price? There are numerous reasons why Neural Networks offer an advantage in the quest to forecast stock prices. Firstly, there is no need for any assumptions to be made used by Efficient Market Hypothesis (EMH) and no need for Normality assumptions. The EMH [4, 5] assumes that the price of a stock reflects all of the information available and that everyone has same degree of access to information and that when ever a change in financial outlook occurs, the market will instantly adjust the stock price to reflect the new information.
Many studies [1] have established that non-linearity exists in financial data and Neural Networks can be successfully used to model the relationship among this data. So the applications of Neural Networks to financial forecasting have become very popular over the last few years.

- The main advantage of Neural Networks is that they can approximate any nonlinear function to a degree of accuracy with a suitable number of hidden units
- Neural Networks can cope with "fuzzy patterns" – patterns that are difficult to reduce into precise rules.
- Neural Networks can be retrained and thus can adapt to changing market behaviour.
- Neural Networks can play a crucial role in deciding which technical variables to follow when analyzing past prices.

Based on the above information, we ran a Neural Network with our Australian industrial stock data using a multi-layered perceptron Neural Network model, trained with a back propagation algorithm, using the Hyperbolic Tangent Solution.

III. THE EXPERIMENT - THE STOCK SELECTION PROCESS

Our method is more from a top-down perspective. Figure 1 below, illustrates the path taken to arrive at our 6 stocks for selection consideration. Since there are more than 2000 stocks available in the Australian stock market, our first step is to narrow down the universe of stocks we want to choose from.

To do this, we first identify the best performing sector by comparing all sector price trends against the ASX (Australian Ordinary) index for the latest 3 months. In our experiment, the industrial sector turned out to have the best sector performance when compared to the Australian Ordinary Index.

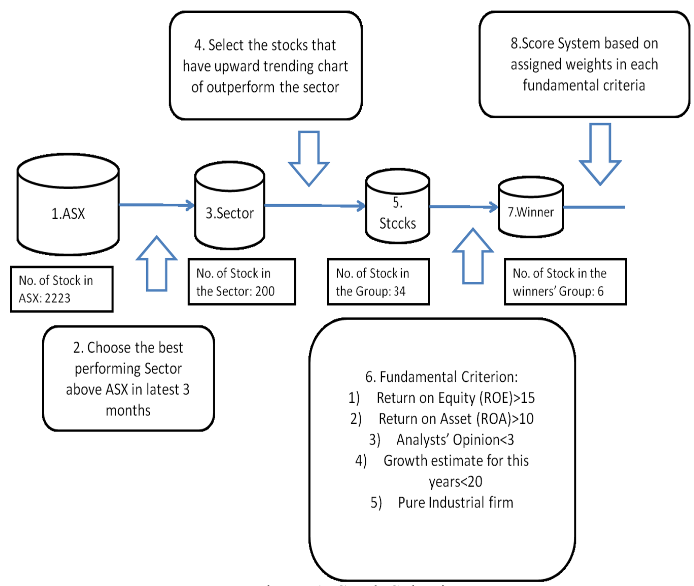

Figure 1: Stock Selection Process.

So, we instantly reduced our universe of stocks from 2223 to 200 as there are about 200 stocks in the Industrial sector. Data on the 200 industrial stocks was then collected from Yahoo Finance.

The four stock selection strategies were applied in this study. One personal trading strategy, two based on decision trees (CHAID & C5.0) and one based on Neural Networks. Each selection strategy selected six stocks and roughly ten thousand dollars was invested in the market for each selected stock. To keep the input variables consistent, five variables were used for all of the strategies:

- Return on Equity
- Return on Assets
- Analyst Opinion
- Growth this year
- Price

In the next few sections we will describe each of the four stock selection processes. For the stock selection processes using a data mining approach, 6 winning stocks were selected from 69 industrial stocks whose financial information was published.

A. Personal Trading Strategy

We then selected all those industrial stocks that had an upward trend in the market and were outperforming the Australian Ordinary Index. There were 34 of these stocks. We then selected the best 6 stocks based on the following criteria:

- Return on Equity (ROE) >15%,
- Return on Assets (ROA)>10%
- Analysts' opinion < 3
- Growth estimate for the current year>20%
- Pure Industrial Firm

As a result, 6 stocks were selected for our personal strategy (Option 1). Our scoring system assigned weights to stocks based on their fundamental ratios, price, analysts' opinion and growth estimates, where the sum of all weights is one. We next ranked our stocks from 1 to 6 with 1 as the highest preferred score to 6 as the lowest preferred score (for example, the higher the ROE the lower the rank). The final score is the sum product of the stock ranking and assigned weights. The assumption made is that the smaller the score, the more likely the stock will make profit.

B. Price Trading Strategy

We ran a C5.0 decision tree with the five input metrics on the 55 industrial stocks. The 'price' metric was the most important. More specifically the rule generated was 'price >1.05'. So the six stocks that had the highest probability of increasing their price and meeting this criterion the closest were selected for this portfolio, where the lower priced stocks were preferred with as they had a higher probability of increasing price.

C. Growth Trading Strategy

We ran a CHAID decision tree with the five input metrics and the 55 industrial stocks. The only metric that was important was 'growth this year', where the rule generated was 'growth this year >13.9%'. So the six stocks with the highest probability of increasing their price and meeting this criterion the closest were selected for this portfolio, where the highest 'growth this year' stocks were preferred.

D. Growth & Value Trading Strategy

We ran a Neural Network with the five input metrics and the 55 industrial stocks. Three metrics were important 'growth this year', 'analyst opinion' and 'return on assets'. So the six stocks with the highest probability of increasing their price, was selected for this portfolio.

IV. Trading Strategy

The goal of our trading strategy is to reduce the universe of stocks to a manageable few. We use technical analysis (charts) and fundamental analysis (financial ratios) to assist us in our selection of stocks with good indicators for growth & value respectively. After the stock selections process in Figure 1, we set the exit strategy as either a stock gain of 10% or more, or a loss of 5% or more (see Figure 2). To evaluate whether our trading strategy performs well, we compare it with the Australian Ordinary Index.

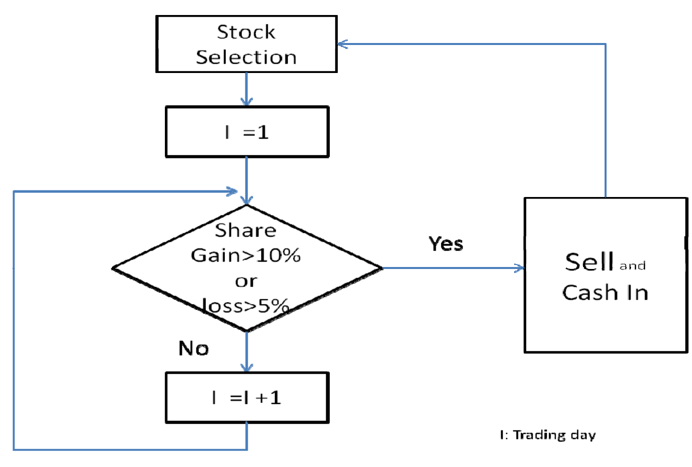

Figure 2: Trading Strategy.

V. Results

Accuracy can be compared in many ways. In the simplest way, a decision tree is grown and the predicted classification is tested against the data set used to train the form of the tree. This is called a re-substitution test [3]. The performance of CHAID is determined by a test of significance of 0.05. For our CHAID, C5.0 and Neural network models, we obtained the following results (See Table 1, 2, 3):

Table 1: C5.0 Accuracy Table

	Actual		
Predicted	***Increase Stock Price***	***Decrease Stock Price***	***Sum***
Decrease Stock price	1	23	24
Increase Stock Price	42	3	45
Sum	43	26	69

Table 2: CHAID Accuracy Table

	Actual		
Predicted	***Increase Stock Price***	***Decrease Stock Price***	***Sum***
Decrease Stock price	6	12	18
Increase Stock Price	37	14	51
Sum	43	26	69

Both Table 1 and Table 2 show the C5.0 and CHAID decision tree classification techniques to achieve high accuracy rates.

Table 3: Neural Network Accuracy Table

	Actual		
Predicted	***Increase Stock Price***	***Decrease Stock Price***	***Sum***
Decrease Stock price	3	8	11
Increase Stock Price	40	18	58
Sum	43	26	69

According to [3], sensitivity and specificity may be computed using equations (1) and (2).

$$\text{Sensitivity} = \frac{[\text{Number of True Positive}]}{[\text{Number of True Positives} + \text{Number of False Negatives}]} \quad (1)$$

$$\text{Specificity} = \frac{[\text{Number of True Positive}]}{[\text{Number of True Positives} + \text{Number of False Negatives}]} \quad (2)$$

The accuracy rates for the data mining models are presented in Table 4 below.

Table 4: Summary Accuracy Table

Accuracy Measure	**C5.0**	**CHAID**	**Neural Network**	**Overall**
Sensitivity	98%	86%	93%	92%
Specificity	88%	46%	31%	55%

Overall the models performed very well, with average sensitivity of 92% and specificity of 55%. The C5.0 decision tree had the best overall performance. Sensitivity is the more important measure for us as we are more interested in correctly classifying stocks with increasing trends. So, the overall accuracy level (sensitivity of 92%), provides support that our data mining models are reliable and have high accuracy. els

A. Portfolio Results

The personal strategy, option 1 portfolio yielded a higher return, than the Australian Ordinary Index (AOI) (7.9% versus 1.7%) during a 20 day trading period. This yield is more than 4.5 times higher than the AOI return (See Figure 3).

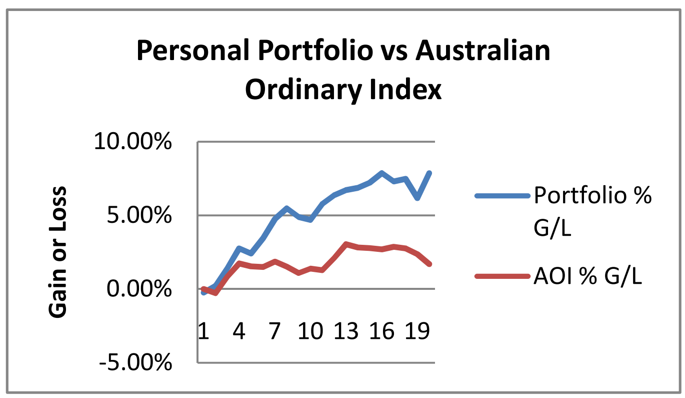

Figure 3: Personal Strategy Portfolio versus Australian Ordinary Index.

The growth value strategy (using a C5.0 decision tree), option 2 portfolio also yielded a higher return, than the Australian Ordinary Index (AOI) (19.9% versus 1.7%) during a 20 day trading period. This yield is more than 11.7 times the AOI return (See Figure 4). One thing to take note with the low price strategy is the volatility. The volatility is higher than the personal strategy. It confirms the notion that the higher the volatility, the higher the possible returns. Stocks with lower prices are more volatile than others but may produce higher returns.

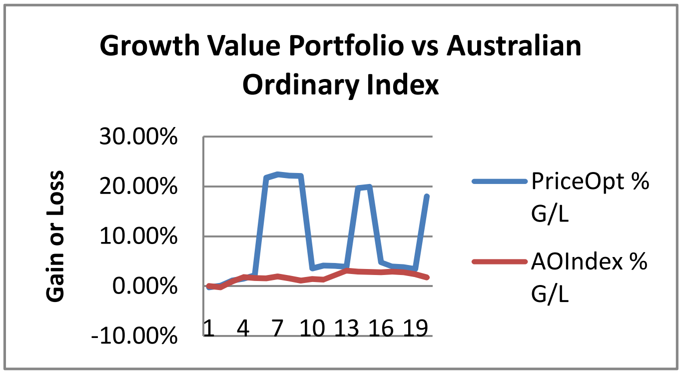

Figure 4: Growth Value (C5.0) Portfolio versus Australian Ordinary Index.

Similarly, the price strategy (using a CHAID decision tree), option 3 portfolio also yielded a higher return, than the Australian Ordinary Index (AOI) (20.8% versus 1.7%) during a 20 day trading period. This yield is twelve times the AOI return (See Figure 5).

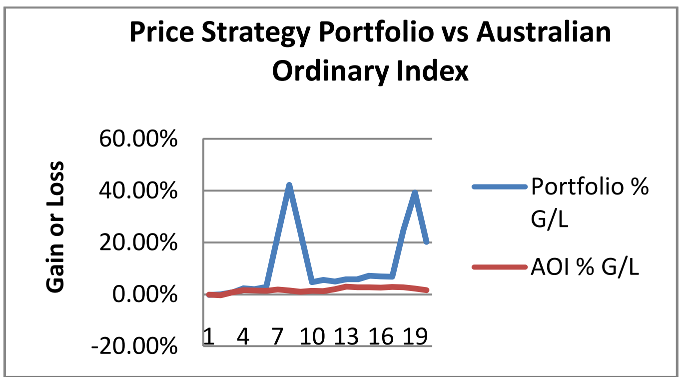

Figure 5: Price Strategy Portfolio versus Australian Ordinary Index.

The growth & value strategy, option 4, also yielded a higher return than the Australian Ordinary Index (AOI) (23.2% versus 1.7%). This yield is almost 14 times more than the AOI return (See Figure 6).

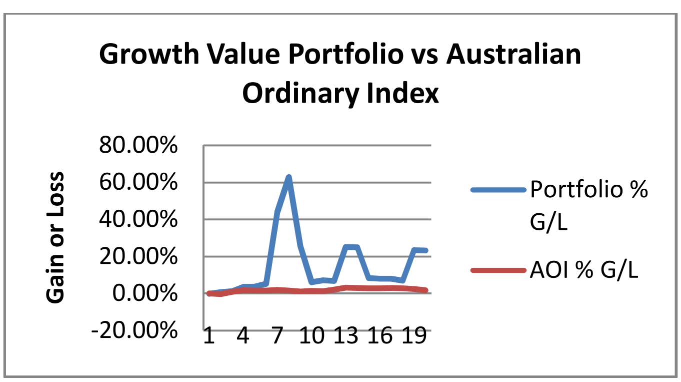

Figure 6: Growth Value Portfolio versus Australian Ordinary Index.

It is quite clear from the above figures, that our hypothesis:

H1: The stocks selected through a personal systematic approach outperforms the Australian stock market significantly

H2: The stocks selected through a data mining approach outperforms the Australian stock market significantly

Further, a statistical test was performed at the 5% level and it confirmed that the returns of all 4 portfolios were significantly different to the returns of the Australian stock market.

B. *Limitations*

The data set size is smaller than usual and missing data was imputed using the median value. The specific form of a decision tree, particularly at lower levels, cannot be exactly reproduced when applied to new data. One of the main problems of a single decision tree model is that small changes in the data set can produce substantial changes in the model. Thus, small changes can easily change the size and shape of the decision tree. One solution is to grow multiple decision trees using a randomisation approach and then combine the information from the multiple decision trees into one summary representation. Another limitation is that there are many other learning algorithms in the Neural Network to be explored. Secondly, we focused on the industrial sector so our stock selection criteria and trading strategy is currently limited to this sector.

VI. Conclusion

A. *Findings*

In this paper we showed how Data Mining techniques can be used to build classification models with stock data from the Industrial Sector of the Australian Stock Market. In conclusion, we can safely say that our stock selection and trading strategy outperformed the Australian Ordinary index. If we annualize our portfolio under the assumptions that we could consistently trade in this way each month, then expected returns for each portfolio would be 94.3% for our personal strategy, 216% for the low price strategy, 243.5% for the growth strategy and 263% for the growth value strategy.

Further, we have obtained high accuracy rates for the three data mining models under study, the C5.0 decision tree, the CHAID tree and the Neural Network model. The decision tree models performed better in classifying the stocks than the Neural Network model, probably because the data set size was smaller than usual and the Neural Network model will perform better for larger sample sizes.

Our results are good with regards to the validation criteria. We have confirmed that the metrics, ROE, ROA, growth this year, price, and analyst opinion, are good predictors for classifying stocks into two groups, namely, stocks that are highly likely to increase in price and stocks they are not likely to increase in price.

Our findings justify the use of data mining techniques for classification and prediction purposes. Lastly, our trading strategy consistently produced results that outperformed the Australian Ordinary Index, whether our personal strategy or the low price strategy or the growth strategy or the growth value strategy was chosen, because our trading strategy was employed we consistently outperformed the Australian market.

B. Further Research

Future researchers may include more methods for finding the best model for predicting stock prices. We used stocks from the Industrial Sector however it would be interesting to expand our study to see whether our stock selection and trading strategy will work in other sectors. Another interesting idea we have in mind, is to build an application that applies our stock selection and trading strategy. Finally, our trading strategy has a 10% upper limitation on the portfolio and an exit strategy on a loss of 5% or more. Other trading strategies may be investigated.

REFERENCES

[1] R.Apostolos-Paul, "Neural Networks in the Capital Markets," 1995.

[2] A. Costea, "new Approached and Perspectives in eth use of quantitative methods". Journal of Applied Quantitative Methods. Vol.1 No.2 Winter, 2006.

[3] B. de Ville. "decision trees for Business Intelligence and Data mining:Using SAS Enterprise Miner. SAS Institute Inc., Cary, NC, USA.ISBN-13:978-1-59047-567-6, 2006.

[4] E. Fama. "The behaviour of stock market prices." Journal of Business 38, 34-105, 1965.

[5] E. Fama. "Random Walks in stock market prices". Financial Analysts Journal 21, 55-9, 1965.

[6] E. Hajizadeh; H.D. Ardakani; J. Shahrabi. "Application of data mining techniques in stock markets: A Survey.", Journal of economics and International Finance. Vol. 2 (7), pp. 109-118, 2010.

[7] J.R. Quinlan. "C4.5: Programs for Machine Learning", Morgan Kaufmann Publishers, 1993.

[8] The Gartner Group, www.garner.com