Financial Analysis Using Data Mining Techniques and Modelling

Shantanu Jha, Rajeshwar Sisodia, Yash Khandelwal, Nidhi Dube, Anita Gunjal
Computer Engineering Department
MITCOE
Pune, India

***Abstract*—The financial markets see a fair amount of usage of predictive technology and automated computer programs that use mining. Data mining utilizes the principle that historic data is a reliable basis for estimating performance in the future. The technology is designed to help investors make better investment decisions by extracting hidden patterns from the available historic data. From the obtained result, we can build various models (using neural networks) to value the company. A company's valuation can help other companies to weigh up possible mergers and acquisitions. Stock market analysis is widely regarded as a challenging problem in financial time series prediction. This paper discusses the various techniques used for modelling the vast financial data available electronically and the portfolio management methodologies required to help investors and institutions make better investment decisions. Also, it investigates various geo-political events that place a challenge on successful stock market prediction.**

Keywords—neural networks; data mining; mergers; acquisitions; portfolio management

I. INTRODUCTION

Knowledge discovery extracts useful information and hidden patterns from the increasingly unmanageable amounts of digital data that we humans face in these times. Over the last few years an abundance of financial data has been stored electronically and this volume is going to grow exponentially in the coming future. Despite this wealth of data, financial institutions and fund managers have been unable to fully utilize them as they are not able to uncover the implicit hidden patterns and information buried in this expansive data set which is very difficult to discern. Automated programs utilize data mining techniques to predict the future performance of companies and give buy/sell signals to investors.

Market forecasting includes finding trends, formulating investment strategies and identifying the best time to purchase and sell the stocks. Like in other computational methodologies, numerous data mining technique and method has been used in financial modelling. The efficient market theory denies the discovery of stable long-term trading regularities. It is based on the principle that if such trends did exist, then it would be discovered by all the players in the market thereby beating the purpose of finding them in the first place. Data mining neither accepts nor rejects the efficient market hypothesis. It only strives to create tools that uncover subtle and short-term trends and conditional patterns in a wide range of markets.

In finance, prediction tasks are in one of the two forms: (1) Direct prediction of the numeric market characteristics like the stock return, or (2) the prediction if that market characteristic will rise or fall to decide when and which stock to buy/sell. In the second case, we need to determine a threshold above which the market characteristic needs to increase or decrease as we need to take into account the significance of trading returns and trading costs. We use five methods in this approach. These were Typical Price, Chaikin Money Flow Indicator, Moving Average, Relative Strength Index and Bollinger Bands. For the direct stock return prediction, we use Neural Networks that help institutions and investors in making informed decisions.

II. NEURAL NETWORKS

Neural Networks are modelled after the biological nervous systems and are an information processing paradigm that renders data like the brain processes information. It consists of many connecting nodes working together to solve a specific problem. A Neural Network is designed for a particular matching by utilizing a learning process. These mechanisms can be used for portfolio management of both investors and financial institutions by forecasting stock returns [1]. Firstly, we collect the historical data for stock S1 from a data warehouse, say for 15-20 years. Then, we construct a neural network NN1 that predicts the stock return value for S1. We repeat this procedure for every stock Si which is under the investor's radar. For each stock, a neural network is generated. We compute NNi(Si(t))=Si(t+k) for each stock i, k days ahead for forecasting the stock return. Subsequently, select the n highest Si(t+k) values of predicted return and compute the sum of all the selected returns T. Then the investors are advised to invest in stocks proportional to Si(t+k)/T. The obtained share return has a twofold use. Firstly, it can be used by private investors to make better and informed investment decisions. Secondly, it can be used by companies to monitor their potential mergers and acquisitions (M&A) targets. They can do valuation of their target firm which can also be backed by other modeling techniques like the three statement models, which can be derived from the data pulled from the three financial statements of a company: income statement, balance sheet and cash-flow statement. Other models that can be used for support are the Discounted Cash Flow (DCF) model and regression analysis.

In the Discounted Cash Flow model, the discounted values of the future free cash flows of a company are used to arrive at an estimated present value. The goal of the DCF analysis is to estimate the return on investment for an investor with the time value of money taken into account. The time value of money is the idea that any money available at the present day is worth more than an equal amount in the future due to the opportunity cost that needs to be taken into account. The free cash flows that are predicted in this model represent the cash amount that the company generates after spending the money required to expand or maintain its asset base.

While discounting the values of the future cash flows, it is common to use the weighted average cost of capital (WACC) as the discount rate. The weighted average cost of capital is the calculation of a firm's cost of capital with each type of capital weighted proportionally. The various categories of capital include bonds, long-term debt, common stocks and preferred stocks. The traditional method of the DCF modelling employs art to predict the near-term growth rate of future cash flows with the use of a financial analyst's expertise. It is not based on concrete science which leads to erroneous or inconclusive valuation of companies. We propose the use of the backpropagation algorithm for training artificial neural networks to estimate the future cash flows based on their historic value. The backpropagation algorithm uses two phases: the propagation phase and the weight update phase. The propagation phase takes in an input vector and propagates it through each layer of the neural network till it reaches the output layer. Then, the output of the neural network is compared to an expected output and the error between them is presented back to the system to modify the internal weightings of the historic input states. Over multiple progressions, input weightings get adjusted to form an accurate network for predicting the future output.

Applying the backpropagation algorithm to the historic cash flows of a company can be used to estimate its near-term future cash flows. But to find the final valuation of a company, we need to find its terminal value which is all the future cash flows in the model. To calculate the terminal value, a long-term growth rate of the cash flows is used after a certain number of years of near term future cash flows are estimated using the network. The long-term growth rate is assumed with some qualitative analysis of the company and should not exceed the growth rate of the overall economy by much.

The DCF model using artificial neural networks works best for established companies with a concrete track record as there is a much larger data set for training the network. Also, such companies are the ones which have reached the stage where they generate reliable cash flows as they already have a solid asset base and hence do not need to shift the flow of funds from their operating profit to capital expenditure like in the case of a startup.

III. METHODOLOGIES

The prediction of stocks is done by combining and analyzing different methodologies, It is done to predict whether the following day's closing price would increase or decrease. To predict if the stock price will increase or decrease, all the methods need to be in agreement. The methods are Typical Price (TP), Moving Average (MA), Chaikin Money Flow Indicator (CMI), Stochastic Momentum Indicator (SMI), Relative Strength Index (RSI), Bollinger Bands (BB) and Bollinger Signal.

A. Typical Price

Here, the inputs for the algorithm are the daily high, low and closing values of the share. The typical price is calculated by adding high, low and closing price together and then dividing the sum by three. This average acts as an indicator and is compared with the benchmark to decide whether to buy or sell stocks.

B. Moving Average

Moving Average shows the average price over a given time period. It is the most popular indicator. For a 20-day moving average, you should add the closing prices of the days under consideration and then divide the sum by 20. 20, 30, 50, 100 and 200 days are the most common test samples for taking averages. Longer time span average is less affected by daily price fluctuations. On a graph of price changes, moving average is plotted as a line. When the price falls below the moving average they have a tendency to continue falling. Similarly, when the price fall above the moving average they have a tendency to continue rising.

C. Stochastic Momentum Index

The Stochastic Momentum Index (SMI) is derived from the Stochastic Oscillator. The difference is that the SMI finds the position of close relative to the midpoint of the high/low range, while the Stochastic Oscillator finds where the close is relative to the entire high/low range. The range of values for the SMI is from +100 to -100 [2]. The SMI is above zero if the close is greater than the midpoint. On the other hand, if the close is less than the midpoint, the SMI will sub-zero. The SMI and the Stochastic Oscillator are interpreted in the same way. Unusually high/low values of the SMI point to overbought/oversold conditions. A rise above -50 of the value of the SMI generates a buy signal whereas as fall below +50 indicates a sell signal.

D. Chaikin Money Flow Indicator

Chaikin money flow indicator is derived from Chaikin accumulation/distribution. In accumulation/distribution, midpoint is an important factor. Midpoint is calculated as the average of the high and the low for the day. Accumulation is done when the stock price closes above it midpoint that day. Similarly, if the stock price closes below its midpoint then distribution is done that day. By summing the accumulation/distribution of 13 periods and then dividing it by 13, chaikin money flow is calculated.

A bullish stock will have a relatively high close price within its daily range and have increasing volume. This is the main assumption made in CMI [2]. However, an indicator of a weak security is a stock consistently closing with a relatively low close price within its daily range with high volume. When a stock closes is in the upper half of a period's range, there is a buying pressure and when a stock closes in the lower half of the period's trading range, there is a selling pressure. According to the sensitivity sought and the time horizon of individual investor the exact number of periods for the indicator should be varied. When Chaikin Money Flow is less than zero it is known as an obvious bearish signal. A security is under selling pressure or experiencing distribution when a reading of less than zero is indicated.

An obvious bearish signal is when Chaikin Money Flow is less than zero. A reading of less than zero indicates that a security is under selling pressure or experiencing distribution. The length of time that Chaikin Money Flow has remained less than zero is a second potential bearish signal. The greater the evidence of sustained selling pressure or distribution, the longer it remains negative. Bearish sentiment towards the underlying security and downward pressure on the price can be indicated by extended periods below zero. The degree of selling pressure is the third potential bearish signal. This can be determined by the oscillator's absolute level [2]. Readings on either side of the zero line or plus or minus 0.10 are usually not considered strong enough to warrant either a bullish or bearish signal. Degree selling pressure begins to warrant a bearish signal when the indicator moves below -0.10. Similarly, to warrant a bullish signal the indicator should move above +0.10. A reading below -0.25 is considered to be indicative of strong selling pressure. Conversely, a reading above +0.25 is considered to be indicative of strong buying pressure.

The assumption on which Chaikin Money Flow is based is that a bullish stock will have a relatively high close price within its daily range and have increasing volume. This condition would be indicative of a strong security. However, a weak security is indicated if it consistently closed with a relatively low close price within its daily range and high volume.

IV. GLOBAL EVENTS

The biggest threat to data mining for financial prediction is the impact of unpredictable events arising due to various geo-political changes around the globe. These events have a high risk associated with them and can render forecasting through any modeling wrong [3]. The best way to deal with such a situation is to find the sensitivity of each stock with volatility in the markets and to take it into consideration while developing financial models for prediction. Some stocks are highly sensitive to the macro-economic conditions and they are impacted the heaviest in the wake of such global geo-political events.

While it is very difficult to quantify the effects of such events beforehand, it is very important to remember that many times the volatility in the markets due to their occurrence is short-lived especially if it is purely based on public consensus. In such a case, the result due to the impact evens itself out within a short span of time [4]. But, in the case when the event directly affects the industry or the market concerning the stock under consideration, all the models built need to be recalibrated for the future.

V. CONCLUSION

The algorithms and methodologies discussed in the paper might not making anyone rich, but they are very useful for trading and valuation analysis using historic data. They can perhaps be used as a confidence booster for a trader's prediction of stock performance. For a data mining project to be successful, it should drive usage needs and the results obtained should be tested as quickly as possible. The uniqueness of financial applications is that they can be tested quickly. The stock price could be forecasted using some algorithms and their prediction could be tested the very next trading day. The process can be repeated continuously for quality estimates.

We expect that data mining in finance will evolve as a distinguished field in the coming years, one that blends knowledge from data mining and the financial domain. It could reach the same maturity as bioinformatics where the field specifics are closely integrated. Over time, it should be able to replace the art side of financial modelling with scientific certainty as much as possible for improved results for investors.

ACKNOWLEDGMENT

The authors would like to thank Prof. R.K Bedi, Dr. Ramachandra Pujeri and Dr. Siddhivinayak Kulkarni for their continued support during the work.

REFERENCES

[1] Hossein Nikooa, Mahdi Azarpeikanb, Mohammad Reza Yousefib,c, Reza Ebrahimpourb,c, and Abolfazl Shahrabadia,d (2007). "Using A Trainable Neural Network Ensemble for Trend Prediction of Tehran Stock Exchange" IJCSNS International Journal of Computer Science and Network Security, VOL.7 No.12,

[2] K. Senthamarai Kannan, P. Sailapathi Sekar, M.Mohamed Sathik and P. Arumugam "Financial Stock Forecast Using Data mining techniques".

[3] Justin Wolfers and Eric Zitzewitz (2006). "Prediction Markets in Theory and Practice"NBER Working Paper No.12083.

[4] Hol, E.M.J.H. (2003): ING Group Credit Risk Management, Amsterdam, "The Netherlands Empirical studies on Volatility in International Stock Markets