

To Study and Analyze To Foresee Market Using Data Mining Technique

¹Amit Khedkar, ²Prof. Rajendra Argiddi

^{1,2}Department of Computer Science & Engineering Walchand Institute of Technology, Solapur, India

ABSTRACT

In every field there is huge growth and demand in knowledge and information over the internet. The automation using data mining and predictive technologies are doing an advance amount of deals in the markets. Data mining is all based on the theory that the historic data holds the essential memory for predicting the future direction. This technology is designed to help shareholders to discover hidden patterns from the historic data that have probable predictive capability in their investment decisions. The prediction of stock markets is regarded as a challenging task of financial time series prediction. Data analysis is one way of predicting if future stocks prices will increase or decrease. There are some methods of analyzing stocks which were combined to predict if the day's closing price would increase or decrease. These methods include study of Price, Index, and Average. (For e.g. Typical Price (TP), Bands, Relative Strength Index (RSI), CMI and Moving Average (MA)).

KEYWORDS :Data mining, Stock prediction, Historical data

I. INTRODUCTION

Data mining means 'making better use of data'. Every human being is increasingly faced with unmanageable amounts of data; hence, data mining or knowledge discovery apparently affects all of us. It is therefore recognized as one of the key research areas. Ideally, we would like to develop techniques for "making better use of any kind of data for any purpose". However, we argue that this goal is too demanding yet. Over the last three decades, increasingly large amounts of historical data have been stored electronically and this volume is expected to continue to grow considerably in the future. Yet despite this wealth of data, many fund managers have been unable to fully capitalize on their value. This paper attempts to determine if it is possible to predict if the closing price of stocks will increase or decrease on the following day. The approach taken in this paper was to combine six methods of analyzing stocks and use them to automatically generate a prediction of whether or not stock prices will go up or go down. After the predictions were made they were tested with the following day's closing price. If the following day's closing price can be predicted to increase or decrease 70% of the time at the 0.07 confidence level, then this analysis would be an easy and useful aid in financial investing. Furthermore, the results would show that the results are better than random at a reasonable level of significance. Many fund management firms have invested heavily in information technology to help them manage their financial portfolios.

Over the last three decades, increasingly large amounts of historical data have been stored electronically and this volume is expected to continue to grow considerably in the future. Yet despite this wealth of data, many fund managers have been unable to fully capitalize on their value. This is because information that is implicit in the data for the purpose of investment is not easy to discern. For example, a fund manager may keep detailed information about each stock and its historic data but still it is difficult to pinpoint the subtle buying patterns until systematic explorative studies are conducted. The automated computer programs using data mining and predictive technologies do a fare amount of trades in the markets. Data mining is well founded on the theory that the historic data holds the essential memory for predicting the future direction. This technology is designed to help investors discover hidden patterns from the historic data that have probable predictive capability in their investment decisions. This is an attempt, made to maximize the prediction of financial stock markets using data mining techniques. Predictive patterns from quantitative time series analysis will be invented fortunately, a field known as data mining using quantitative analytical techniques is helping to discover previously undetected patterns present in the historic data to determine the buying and selling points of equities. When market beating strategies are discovered via data mining, there are a number of potential problems in making the leap from a back-tested strategy to successfully investing in future real world conditions.

The first problem is determining the probability that the relationships are not random at all market conditions. This is done using large historic market data to represent varying conditions and confirming that the time series patterns have statistically significant predictive power for high probability of profitable trades and high profitable.

II. PROBLEM STATEMENT

The aim is to propose a model which can calculate and tell the user about the investment in the stock market. Market analysis is the basis for the prediction.

III. SYSTEM DESIGN

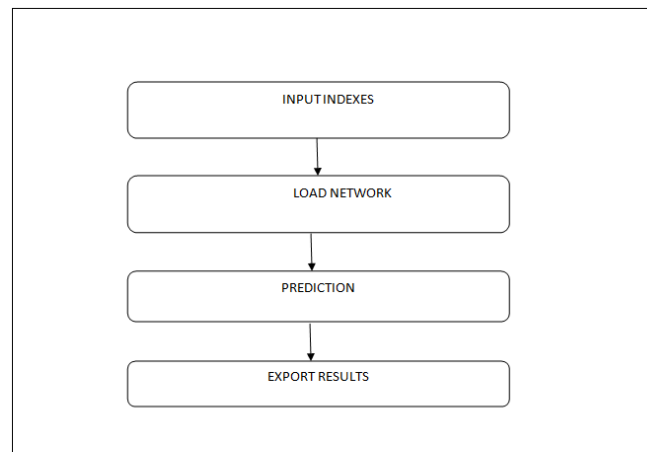


Fig. System Design

Indexes are the basic inputs for the network. Network is trained on provision of the input. It is possible to predict the future depending on the analysis. Design flow works on the same principal. At last whatever the results we got we save it for the prediction.

IV. METHODOLOGY

Five methods of analyzing stocks were combined to predict if the following day's closing price would increase or decrease. All five methods needed to be in agreement for the algorithm to predict a stock price increase or decrease. The five methods were Typical Price (TP), Chaikin Money Flow indicator (CMI), Stochastic Momentum Index (SMI), Relative Strength Index (RSI), Moving Average (MA) and Bollinger Signal.

Algorithm:

1. Give High, Low, Close values of the daily share as the input
2. Take an output array and add the values of H, L, and C
3. Divide the total by 3

$$(TP - [H + L + C]/3)$$

Where, H=High; L=Low; C=Close. Where the TP greater than the bench mark we have to sell or to buy.

Chaikin Money Flow Indicator

Chaikin's money flow is based on Chaikin's accumulation/distribution. Accumulation/distribution in turn, is based on the premise that if the stock closes above its midpoint $[(high+low)/2]$ for the day, then there was accumulation that day, and if it closes below its midpoint, then there was distribution that day. Chaikin's money flow is calculated by summing the values of accumulation/distribution for 13 periods and then dividing by the 13-period sum of the volume. It is based upon the assumption that a bullish stock will have a relatively high close price within its daily range and have increasing volume. However, if a stock consistently closed with a relatively low close price within its daily range with high volume, this would be indicative of a weak security. There is pressure to buy when a stock closes in the upper half of a period's range and there is selling pressure when a stock closes in the lower half of the period's trading range. Of course, the exact number of periods for the indicator should be varied according to the sensitivity sought and the time horizon of individual investor. 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. 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. A second potentially bearish signal is the length of time that Chaikin Money Flow has remained less than zero. The longer it remains negative, the greater the evidence of sustained selling pressure or distribution. Extended periods below zero can indicate bearish sentiment towards the underlying security and downward pressure on the price is likely. The third potentially bearish signal is the degree of selling pressure. This can be determined by the oscillator's absolute level. 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. Once the indicator moves below -0.10, the degree selling pressure begins to warrant a bearish signal. Likewise, a move above +0.10 would be significant enough to warrant a bullish signal. Marc Chaikin considers a reading below -0.25 to be indicative of strong selling pressure. Conversely, a reading above +0.25 is considered to be indicative of strong buying pressure. The Chaikin Money Flow is based upon the assumption 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, if it consistently closed with a relatively low close price within its daily range and high volume, this would be indicative of a weak security.

The Following formula was used to calculate CMI.

$$\text{CMI} = \left[\frac{\sum(\text{AD},n)}{\sum(\text{VOL},n)} \right];$$
$$\text{AD} = \text{VOL} \left[\frac{\text{CL} - \text{OP}}{\text{HL} - \text{LO}} \right]$$

Where, AD stands for Accumulation Distribution, n=Period, CL=today's close price; OP=today's open price, HI=High Value; LO=Low value

Stochastic Momentum Index

The Stochastic Momentum Index (SMI) is based on the Stochastic Oscillator. The difference is that the Stochastic Oscillator calculates where the close is relative to the high/low range, while the SMI calculates where the close is relative to the midpoint of the high/low range. The values of the SMI range from +100 to -100. When the close is greater than the midpoint, the SMI is above zero, when the close is less than the midpoint, the SMI is below zero. The SMI is interpreted the same way as the Stochastic Oscillator. Extreme high/low SMI values indicate overbought/oversold conditions. A buy signal is generated when the SMI rises above -50, or when it crosses above the signal line. A sell signal is generated when the SMI falls below +50, or when it crosses below the signal line. Also look for divergence with the price to signal the end of a trend or indicate a false trend.

Relative Strength Index

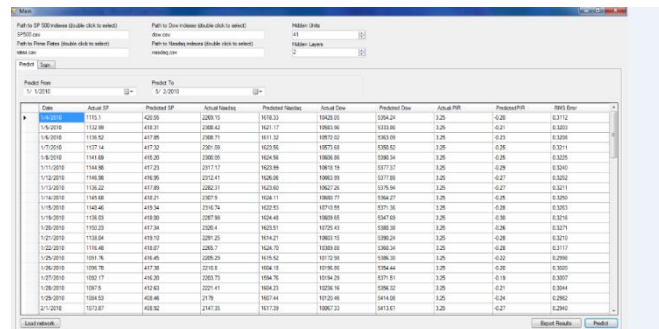
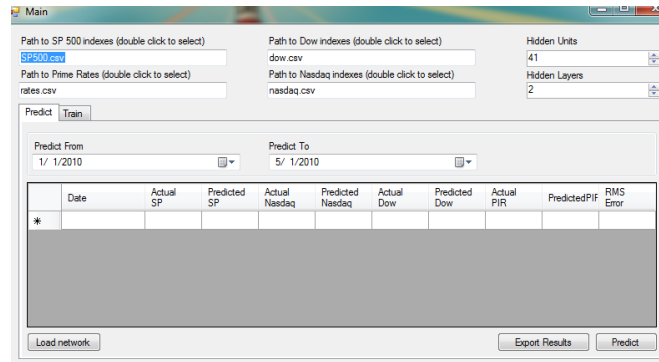
This indicator compares the number of days a stock finishes up with the number of days it finishes down. It is calculated for a certain time span usually between 9 and 15 days. The average number of up days is divided by the average number of down days. This number is added to one and the result is used to divide 100. This number is subtracted from 100. The RSI has a range between 0 and 100. A RSI of 70 or above can indicate a stock which is overbought and due for a fall in price. When the RSI falls below 30 the stock may be oversold and is a good they can vary depending on whether the market is bullish or bearish. RSI charted over longer periods tend to show less extremes of movement. Looking at historical charts over a period of a year or so can give a good indicator of how a stock price moves in relation to its RSI.

Moving Average

The most popular indicator is the moving average. This shows the average price over a period of time. For a 30 day moving average you add the closing prices for each of the 30 days and divide by 30. The most common averages are 20, 30, 50, 100, and 200 days. Longer time spans are less affected by daily price fluctuations. A moving average is plotted as a line on a graph of price changes. When prices fall below the moving average they have a tendency to keep on falling.

V. RESULTS AND ANALYSIS

The Output is best located on a web page given below in snapshots. The GUI for the application will look and seems much user friendly. The numerical data that will be retrieved are located in a systematic fashion.



Other information related to the results such as total no. of results retrieved and time to of retrieval is easy & shown in a simple but effective manner.

VI. CONCLUSION AND FINDINGS

The results show that this work is able to predict that the day's closing price would increase or decrease better than chance (50%) with some level of significance. Furthermore, this shows that there is some validity to technical analysis of stocks. These efforts may be useful for trading analysis.

REFERENCES

- [1] Financial Stock Market Forecast using Data Mining Techniques (K. Senthamarai Kannan, P. Sailapathi Sekar, M.Mohamed Sathik and P. Arumugam) (IMECS International Multiconference of Engineers and Computer Scientists, 2010, Hong Kong)
- [2] Hunter, (2007). Hedge Funds Do About 60% Of Bond Trading, Study Says, The Wall Street Journal.
- [3] Durbin Hunte, (2007). "Hedge Funds Do About 60% Of Bond Trading, Study Says", The Wall Street Journal.
- [4] Hossein Nikooa, Mahdi Azarpeikanb, Mohammad Reza Yousefiba,c, Reza Ebrahimpourb,c, and Abolfazl Shahrabadi,d (2007). "Using A Trainable Neural Ntwork Ensemble for Trend Prediction of Tehran Stock Exchange" IJCSNS International Journal of Computer Science and Network Security, VOL.7 No.12.
- [5] Jiarui Ni and Chengqi Zhang (2006). A Human-Friendly MAS for Mining Stock Data IEEE / WIC / ACM International onference on Web Intelligence and Intelligent Agent Technology.
- [6] Amaury Lendasse, Francesco Corona, Antti Sorjamaa, Elia Liitiäinen, Tuomas Kärnä, Yu Qi, Emil Eirola, Yoan Mich'e, Yongnang Ji, Olli Simula, (2006). "Time series prediction".
- [7] Juliana Yim , Heather Mitchell (2005). A comparison of corporate distress prediction models in Brazil, nova Economia_Belo Horizonte, 15 (1), 73-93.
- [8] Eamonn Keogh, Stefano Lonardi, (2004). Chotirat Ann Ratanamahatana Towards Parameter-Free Data Mining, KDD '04.
- [9] Garth Garner, Prediction of Closing Stock Prices, (2004). This work was completed as part of a course project for Engineering Data Analysis and Modeling at Portland State University.