Stock Market Analysis

Shreya Maheshwari
Department of CS&E
PES University
Bangalore, India

Vishal Sathyanarayana
Department of CS&E
PES University
Bangalore, India

Rithvik Chandan
Department of CS&E
PES University
Bangalore, India

***Abstract*—Our project aims at finding the stocks which are the most optimal for investment at a given point of time, based on their historical data. It will help pick out stocks for investment in such a way that one can minimize risks and maximize profits – by engendering a profitable stock trading decision through technical analysis of the historical data.**

I. INTRODUCTION

In today's day and age, trading in the stock market is an extremely popular channel of financial investment. There are various investment opportunities such as trading bonds, shares, foreign investments and precious metals to name a few. Investors try to create wealth by buying and selling their investments at a proper time. There are various advantages of investing in the stock market – investment gains, dividend income, diversification and ownership. One of the primary reasons for investing in the stock market is the chance to multiply your money. Over time, companies tend to grow in value, thereby increasing the stock prices by a proportionate amount. Investment in stable companies make profit for investors. Some stocks also provide income in the form of dividends. Dividends represent income on top of the profit that comes from selling the stock. An investor need not put all his money into one stock – he/she can diversify. This helps in building wealth by leveraging different sectors of the economy. Buying shares of a company means taking ownership stake in the company. This implies that investing in the stock market also brings benefits as those which are enjoyed by being one of the business' owners.

The key of a good investment is to invest at a suitable time with minimum risk. The aforementioned is extremely hard to determine due to the dynamic and highly volatile nature of the stock market. Technical indicators, as calculated using historical data, are extremely helpful in assisting an investor to execute buy and sell decisions. One cannot predict the exact behavior of a stock, as it does not solely depend on the technical indicators – long term trends, cyclical variations, seasonal variations and irregular movements also play a role. However, technical analysis reduces the uncertainty involved with investment decisions. Even when faced with certain unforeseen circumstances, fundamentally good stocks, ones which exhibit strong technical indicators, don't tend to fall as much as the rest of the other stocks in the market. Exploiting this can potentially make you high profits with a "low" risk. Even in case of a loss scenario the amount of money lost will be minimized.

Our objective is to develop a model which can indicate the stability of a stock and provide every stock with a certain score using the calculated technical indicators. The stocks will then be listed in non-increasing order of scores, and based on how many stocks you would like to invest in, we expect our model to output an investment scheme which would give information as to what percentage of your current money needs to be invested into which stock to try to maximize profit with a lower risk factor.

II. WHAT HAVE OTHERS DONE TO SOLVE THE PROBLEM

Reference [1], Kimoto, Takashi et al base their buy and sell timing prediction on Modular Neural Networks. The output of the Network indicated whether a stock needs to be sold or bought. They use a moving simulation method, where the model learns for M months and predicts next L months and the window is continually shifted. In their model, an output value > 0.5 indicated 'Buy' and < 0.5 indicated 'Sell'. However, this overlooks a third, more practical option in "holding" a particular stock.

In [2], Birgul Egeli et al implement an ANN with the traditional Back Propagation (BP) algorithm used to acquire weights of connection to predict the Istanbul Stock Exchange Market Index Values to find out if it does better than the models based on Moving Averages. Although it did do better, there are some limitations like vanishing gradient as they used the Sigmoid function as the Activation function for a neuron in the network. This could lead to slow convergence and increase the chances of getting stuck at a local minimum.

In [3], Lamartine Almeida Teixeira and Adriano Lorena Inácio de Oliveira combine technical analysis and nearest neighbor classification. They use combinations of technical indicators to such as RSI filter, stop loss, stop gain etc. to come up with 22 features to be used as inputs to a k-NN classifier. This however results in a requirement for large computational power as a k-NN classifier does all of its processing during test time as the "distances" need to be calculated w.r.t 22 features along with cross-validation to get an ideal value for k and might still result in lower accuracy as compared to say a Support Vector Machine.

A commonly seen problem with ANN's is that it ignores noise and various non-stationary characteristics in the data. Training of a Back-Propagation algorithm is difficult due to the noise as it is hard to incorporate market variables into the model without making certain assumptions as a result, the network might end up always predicting the most common output.

Unlike an ANN, the Autoregressive Integrated Moving Average (ARIMA) is a model which is used explicitly for time series data and has a certain structure to it. In [4], Ayodele A. Adebiyi. Et al build an ARIMA model to provide investors successfully with a short-term prediction to aid them with their decision.

This model however does not take into account any technical indicators as it is completely dependent only on

previous target value data. This could lead to inaccurate results in the long run.

Due to this, researchers tend to move towards hybrid solutions by synergistically combining both, the ANN model and the ARIMA model. Studies have been conducted to see if a hybrid model yields better results than individual models [5]. Similarly, Ping-Feng Pai Et al [6] showed that promising results can be obtained from a hybrid model of ARIMA and SVM [6].

III. PROPOSED PROBLEM STATEMENT

To develop a model based on historical data of the stocks listed on the National Stock Exchange in order to find sustainable stocks for safe investments.

A. *Collection of Data*

Historical data of the stocks listed on NSE was collected from nseindia.com using a web scraping software called GetBhavCopy. One text file was generated for each day that trading was carried out. The text file consists of 7 columns as follows: the name of the company, the date (in yyyymmdd format), open price, highest price on that day, lowest price on that day, close price and volume. We have collected data for the last 25 years i.e. from 1995 to 2019.

Information about splits in stocks was collected by web scraping from the website www.moneycontrol.com.

B. *Cleaning of Data*

The data that was collected from the software contained equities as well as indexes. Our primary focus is equity stocks, so the records containing indexes were dropped. Furthermore, some records contained "-"and some records had 8 columns which we have dropped.

While plotting basic graphs for closing price, we noticed that there were certain steep drops in the same. This is attributed to splits in company stocks. We handled this by finding the date after which a stock had been split and multiplied the closing price with the appropriate split factor. We accounted for a stock splitting multiple times by multiplying by the latest split factor after the corresponding date. All records of companies whose data was missing in between have been dropped.

C. *Exploratory Data Analysis*

Line graphs of various stocks were plotted to see the general trend that the companies follow. Figure 1 shows a sudden drop in the closing price of the stock over a very short period of time. This was the case with many such stocks. On further research, we came across the term "split" in regard with stocks which was causing this. A stock split is a decision taken by the company's board of directors to increase the number of shares that are outstanding by issuing more shares to the current shareholders. As a result of this, the stock price is also affected – which was unadjusted in the raw data set that we had obtained from the software. Thus, on calculation of CAGR and number of years of positive returns, we were getting wrong values. In order to rectify this, we scraped data about stock splits which have taken place over the years, which we will apply to our data for further analysis.

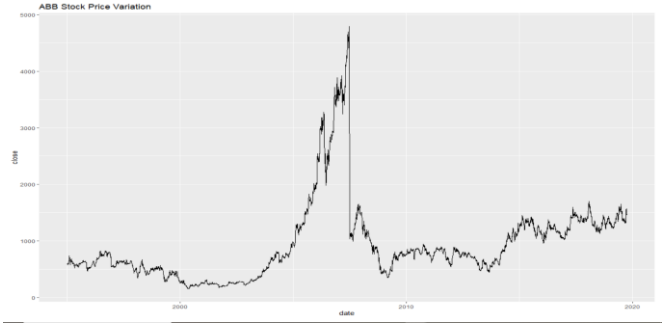

Figure 1: General Trend for ABB stock between 1995 and 2020.

We carried out feature engineering by finding the CAGR (Compound Annual Growth Rate) and number of years of positive and number of years of negative returns.

$$CAGR = (EB/BB) \wedge (1/n) - 1 \quad (1)$$

CAGR: Compound Annual Growth Rate

EB: Ending Balance

BB: Starting Balance

n: Number of years

Higher the value of CAGR, more the company has grown over the years of its existence.

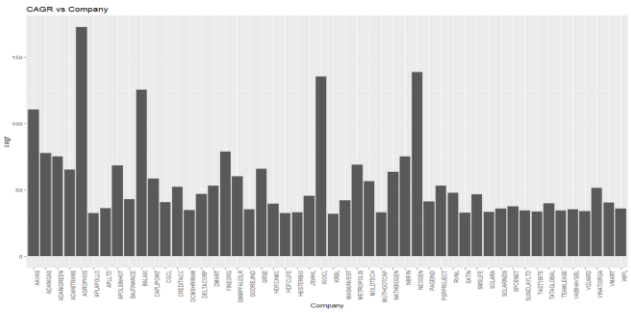

Figure 2: Bar graph for CAGR of top 50 companies.

We also calculated the ratio of number of years of positive returns to total number of years. The number of positive years was found as the closing price on the last trading day of the year and the closing price on the first trading day of the year. If their difference was positive, one was added to the number of years of positive returns. If not, one was added to the number of years of negative returns. Total was the sum of these two values.

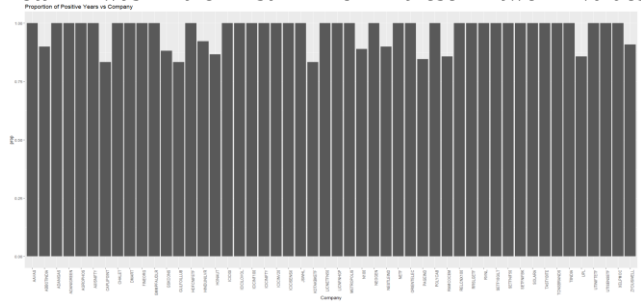

Figure 3: Bar graph for Proportion of positive returns of top 50 companies

A combination of these factors would help us determine which companies are stable – those with high value of

CAGR and proportionately higher number of years of positive returns as compared to the total number of years.

REFERENCES

[1] Kimoto, Takashi, et al. "Stock market prediction system with modular neural networks." *1990 IJCNN international joint conference on neural networks*. IEEE, 1990.

[2] Birgul Egeli, Asst. "Stock market prediction using artificial neural networks." *Decision Support Systems* 22 (2003): 171-185.

[3] Teixeira, Lamartine Almeida, and Adriano Lorena Inacio De Oliveira. "A method for automatic stock trading combining technical analysis and nearest neighbor classification." *Expert systems with applications* 37.10 (2010): 6885-6890.

[4] Ariyo, Adebiyi A., Adewumi O. Adewumi, and Charles K. Ayo. "Stock price prediction using the ARIMA model." *2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*. IEEE, 2014.

[5] Merh, Nitin, Vinod P. Saxena, and Kamal Raj Pardasani. "A comparison between hybrid approaches of ANN and ARIMA for Indian stock trend forecasting." *Business Intelligence Journal* 3.2 (2010): 23-43.

[6] Pai, Ping-Feng, and Chih-Sheng Lin. "A hybrid ARIMA and support vector machines model in stock price forecasting." *Omega* 33.6 (2005): 497-505.