

Fundamental analysis of stock trading systems using classification techniques

¹Sahil Mane,Dhawal Shah, ²Ishika Diwan,Mr.Manoj Dhande

Student (Computer Engineering),Professor of Computer Engineering
Shah and Anchor Kutchhi Engineering College
,Mumbai,India

Abstract: There are numerous companies that are listed on the stock market and experts or portfolio managers usually pick them up using their expertise and fundamental analysis for picking long term stock holdings. But IT has made processing large datasets of companies easy and a layman can use it. Our paper gives an analysis of the company's stock on the basis of its fundamental indicators ie. by basically classifying stocks on the basis of fundamental analysis. It is important for investors to pick the right stock to buy as it will grow in the long term. This paper proposes a process to identify which company's stocks have strong fundamentals and those that have weak fundamentals. This paper uses a dataset that has been created using FinViz which contains many stocks from the S&P 500 and Indian Stock markets.

Index Terms Fundamental Analysis, Fundamental Indicators and Ratios, Random Forest Algorithm.

I. INTRODUCTION

In investing in equity, usually long term or midterm holders prefer doing fundamental analysis. Fundamental Value and Fundamental growth are the two procedures of the fundamental analysis. Fundamental growth means picking fundamentally strong companies that are small cap and have high growth potential in the future while fundamental value is picking up companies that are mid size to huge size by market capital but are undervalued i.e. below the intrinsic value and will grow in the long term. If the company has strong fundamentals an investor should consider adding the stocks to their portfolio. Strong fundamentals are indicators of a company's sustainability and a promising financial future.

II. DATA PRE-PROCESSING

The raw data which was obtained from the internet is then pre-processed. The data contains a lot of symbols related to its ratios, and some unavailable values named NaN. The not available values in the dataset are filled by using the mean, which is a known mathematical method to determine the missing values in a dataset. Useful features relevant to the project are extracted in order to work them on the model. The Signals of each Company Stock are fitted and transformed to numeric values.

III. DATASET

We are using a dataset that was created by following methods such as Web Scraping and extracting the data using Finviz. The dataset consists of over 600 companies listed in the United States market under S&P 500 index and their fundamental ratios such as Insider Own, EPS Q/Q, Change, Sales Q/Q, Gross Margin ,Sales past 5Y,52W High, EPS (ttm),Forward P/E and many such indicators. The dataset includes the intrinsic values of companies which determine the roots of a company.

IV. DATASET SPLITTING

The final pre-processed dataset is then split in a 70:30 proportion. The training data has a proportion of 70 and testing has a proportion of 30. The highlighted green portion in the figure is the training data. The highlighted red portion in the figure is testing data.

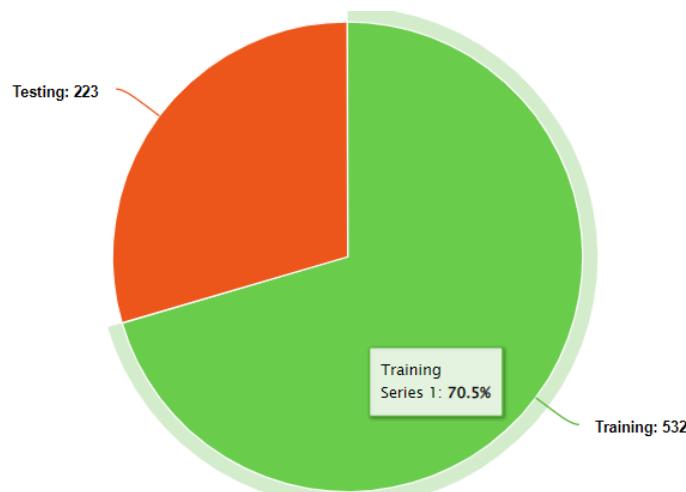


Figure 1: Train and Test data

V. TRAINING MODEL

The model training consists of five different classification algorithms. Classification methods such as K-Nearest Neighbor, Random Forest, Artificial Neural Network, Decision Tree and Logistic Regression are trained on the training dataset that was split. Each model has its own specific analysis and yields a different accuracy from one another. Random Forest and Decision Trees are best methods in order to solve a classification problem. Decision trees are individual sets of nodes that determine a decision based on a number of factors or values. Random Forest is a collection of these decision trees, which increases its accuracy as error minimizes. K-Nearest Neighbor works by finding the distance between a query and all the training data. The algorithm fails to perform as the size of data increases. Artificial Neural Network for classification involves passing the dataset through a number of layers. An activation function of ‘relu’ is used. Logistic Regression is a multinomial method, which means it will give results in more than two variables.

VI. TESTING MODEL

We pass on the testing dataset that was divided into a certain proportion. The model returns an array of numbers that consists of 0, 1 and 2 as an output. The models are compared and viewed based on their confusion matrix. Each model has its drawbacks. It is important for the training data to be accurate and should have minimum redundancy for the testing to go smoothly.

VII. EARNING PER SHARE

- EPS is earning per Share. It is calculated as follows;
- EPS = NET PROFIT/NO. OF SHARES IN CIRCULATION
- Basically, it is the net profit per unit of share. More the net profit per share the better it is hence, the more the value of EPS the better it is. EPS is never used alone and is used in conjunction with another financial ratio-the PE ratio.

VIII. PRICE TO EARNING RATIO

- PE ratio is price-earnings ratio. The P/E of a stock is calculated by dividing the current stock price by the Earning per Share (EPS). P/E indicates how expensive or cheap the stock is trading at. In other words it indicates if the stock is undervalued or overvalued.[2] If the PE ratio is low then the stock price is most probably below its intrinsic value and vice versa. When investing we need to compare the PE ratio of the company with the Industry PE ratio. If the Company PE ratio is less than Industry PE ratio, the company’s stock can be considered undervalued.

IX. FORWARD PE RATIO

- The forward PE ratio is calculated by taking into consideration the growth factor of the company. The growth factor helps find the estimated earnings for the next year.[2] These earnings are taken as denominators while calculating the forward PE ratio. It will usually be less than the PE ratio. The growth is calculated using the retention ratio and return on investment.

X. PRICE TO SALES RATIO

- In many cases, investors may use sales instead of earnings to value their investments. The earnings figure may not be true as some companies might be experiencing a cyclical low in their earning cycle[1]. Additionally, due to some accounting rules, a profitable company may seem to have no earnings at all, due to the huge write-offs applicable to that industry. So, investors would prefer to use this ratio. This ratio compares the stock price of the company with the company’s sales per share. The formula to calculate the P/S ratio is:
- Price to sales ratio = Current Share Price / Sales per Share
- A P/S ratio of ‘X’ times indicates that, for every Rs.1 of sales, the stock is valued Rs. ‘X’ times higher. Obviously, the higher the P/S ratio, the higher is the valuation.

XI. CURRENT RATIO

- Current ratio is a liquidity ratio and indicates the cash position of a firm at any given point of time.
- Current Ratio= Current Assets/Current Liabilities
- Higher the better.

XII. QUICK RATIO

- Quick Ratio only considers liquid assets.
- Quick Ratio=(Cash + Short term marketable securities + a/c receivables)/Current Liabilities
- It is a narrower approach. Quicker ratio is better for the company. It excludes inventories which are not liquid.

XIII. CASH PER SHARE

- Cash per share= Cash and cash equivalents/Total number of outstanding shares.
- This is also one of the measures for liquidity of a company. This cash per share indicates how much cash a company has per the number of outstanding shares. Higher the better. It is an indicator of how much liquidity a company has at any given point of time with which it can pay off its debt, put it into R&D etc. A very high or very low CPS both are undesirable. Its value should be close to the industry standard.

XIV. SHORT RATIO

- Short Ratio= Number of shares shorted/Average daily trading volume of shares
- It is also called the "days to cover" ratio, the short ratio is calculated by dividing the number of shares sold short by the average daily trading volume. For example, if a company has 20 million shares sold short and an average trading volume of 5 million shares, then its short ratio would be four days ($20 \text{ million} / 5 \text{ million} = 4 \text{ days}$). Knowing how many shares of a stock have been shorted is a good indication of how investors view that stock.[3] The short ratio tells investors approximately how many days it would take short sellers to cover their positions if the price of a given stock were to increase. The higher the short ratio, the longer it will take to buy back those borrowed shares.

XV. DEBT TO EQUITY RATIO

- This is a fairly straightforward ratio. Both the variables required for this computation can be found in the Balance Sheet. It measures the amount of the total debt capital with respect to the total equity capital. A value of 1 on this ratio indicates an equal amount of debt and equity capital. Higher debt to equity (more than 1) indicates higher leverage and hence one needs to be careful. Lower than 1 indicates a relatively bigger equity base with respect to the debt.
- The formula to calculate Debt to Equity ratio is: [Total Debt/Total Equity]

XVI. P/FCF

- P/fcf ratio= price of the share/free cash flow
- Free cash flow= operating cash flow- capital expenditure
- Lower P/fcf indicates that the company is undervalued. A very high ratio indicates the company being overvalued. It should be compared to its peers while analyzing [3].

XVII. BOOK VALUE PER SHARE

- BVPS= (Total equity-Preferred Equity)/ Total Number of Outstanding Shares
- Book value per share (BVPS) is the ratio of equity available to common shareholders divided by the number of outstanding shares. This figure represents the minimum value of a company's equity and measures the book value of a firm on a per-share basis.

XVIII. BETA

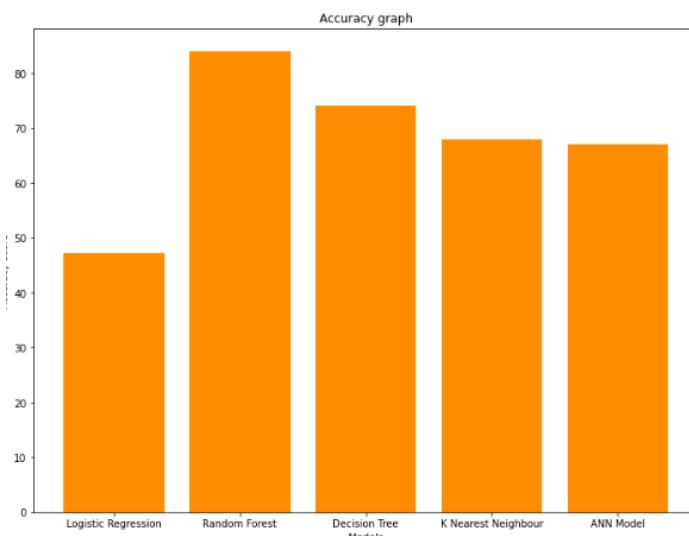
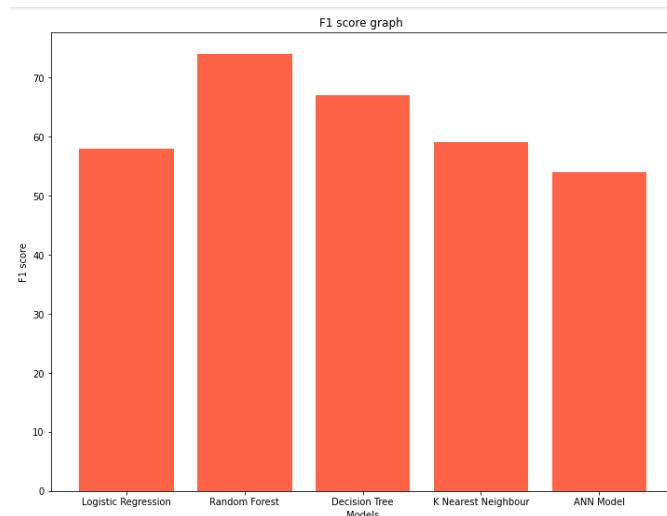
- Beta is the measure of the systematic risk of a firm. It indicates the correlation of the company with the market it is being traded in. If the Beta of the company is 2 it means for every 1% move in the market the stock price of the company moves by 2%.
- Beta = **Covariance (R_i, R_m) / Variance (R_m)**
- Covariance = Measure of a stock's return relative to that of the market
- Variance=Measure of how the market moves relative to its mean..

XIX. GROSS MARGIN

- Gross margin is net sales less the cost of goods sold (COGS). In other words, it's the amount of money a company retains after incurring the direct costs associated with producing the goods it sells and the services it provides. The higher the gross margin, the more capital a company retains, which it can then use to pay other costs or satisfy debt obligations. The net sales figure is gross revenue, less the returns, allowances, and discounts.
- Gross Margin= Net Sales- COGS

XX. RESULT

The results of each model are recorded in the form of a bar graph and the confusion matrix of each model is evaluated. Random Forest algorithm has obtained the highest accuracy of 84% and will be used to classify the fundamental values of a stock. The model is integrated in a Graphical User interface made on the Streamlit framework. The GUI for the model runs on LocalTunnel, a web service used to host the application on the internet. The final output recommends the user whether the fundamentals of a stock is strong or weak which will give the user a definite idea whether the company is worthy of investment or not.

**Figure 1.2: Accuracy Graph****Figure 1.3: F1 Score Graph**

A. ACKNOWLEDGMENT

In this paper we have presented a tool for Fundamental Analysis of a stock using Random Forest Algorithm. Several fundamental factors or indicators have been taken into consideration, which will help beginners to make investing decisions. Fundamental analysis of a stock is a time variable concept as the price of stock fluctuates the ratios from time to time. However, it is one of the most important strategies in the stock market that can be used to invest in a company.

REFERENCES

1. Moghar, Adil, and Mhamed Hamiche. "Stock market prediction using LSTM recurrent neural network." Procedia Computer Science.
2. Cheng, Ching-Hsue, and You-Shyang Chen. "Fundamental analysis of stock trading systems using classification techniques." 2007 International Conference on Machine Learning and Cybernetics. Vol. 3. IEEE, 2007.
3. Chen, Yuh-Jen, Yuh-Min Chen, and Chang Lin Lu. "Enhancement of stock market forecasting using an improved fundamental analysis-based approach." Soft Computing 21.13 (2017).
4. Huang, Yuxuan, Luiz Fernando Capretz, and Danny Ho. "Machine learning for stock prediction based on fundamental analysis." 2021 IEEE Symposium Series on Computational Intelligence (SSCI). IEEE, 2021.