Thesis

Portfolio Optimization Using Financial Ratios and Machine Learning Applications

Abstract:

Stock Market is one of the less volatile and highly anticipatory trading options for most investors of the present world. The primary reason behind investment would always succumb to the most important needs of well-being, secondary earning source and understanding the market insights would drive people to make better investment decisions. The various opportunities like forex, crypto-currencies, options, commodities and various other options which are much larger money market opportunities the risk sentiment still remains higher in these options due to factors like time-based expiry of these options and lack of important information related to trade positions. According to Investopedia reports its seen that S&P 500 stocks have made a 10% of yearly returns since last 100 years which is way better in comparison to other investment opportunities ranging both in higher and lower return sections. Country financial suggested that among various high yield returns, stock is still preferably the best options for investments in terms of long payouts because of no expiry option.

The overall process of stock judgment in our paper would rely on various machine learning techniques starting from K-Means Clustering which would primarily distinguish a stock based on their price and volume criteria’s. Our process would extract the top stocks from all the various cluster levels using these two criteria’s. We would also use PCA for Dimensionality reduction because there are various financial and statistical ratios which all are needed for effective stock analysis decision and their overall value would represent how good or bad the stock is. By applying PCA we will convert all the ratios into one denominating number that would represent the stock. Finally, we will use the clustering process again to determine the best possible stocks from different clusters with low risk and high return criteria. After determining the specific group of stocks, the model would preferably start forecasting the future prices of these best stocks with variant methods such as SVR, LSTM and RNN for stock price forecasting method and test their effectiveness in relation to accuracy with previous data’s. The model will then determine an effective portfolio that ranges with the best possible stocks with highest return and lowest risk values and gets compared with portfolio stock data determined through apriori algorithm for the existent best possible portfolio and compare with its risk and return. All the above methods would be evaluated using previous significance in various related scholarly articles.

Content

1. List of Figures
2. List of Tables
3. Introduction
4. Appendices
5. Insights: Investment Needs, Stocks & Methodologies
6. Why invest & what is the benefit?
7. Financial Calculation as proof for Need of Investment – Literature reviews as needs to invest
8. Why Stocks?
9. Literature Reviews for Suggesting Stocks
10. Analytical Proof for Stock Investment
11. Important Notions of Stocks
12. What is Stock?

* How Stock Market is Beneficial as an Investment Alternative?
* How is Stock market the best mix of Volatility and gain measurement?
* How to determine the investment capital for each smaller time segment?
* Retirement fund accumulation process

1. Financial Ratios for Stock Analysis

* What is Financial Ratio Analysis?
* Which Financial Ratios are important – for Stock Analysis
* Weighted Importance of various Financial Ratios for Stock Analysis
* Literature Reviews suggesting importance of Financial Ratios in Stock Analysis

1. Process: Defining various data sources, Importance of Data for Stock Analysis, Adapted Methodologies, Significance Analysis of Applied Methods, Evaluation of Applied Methods, Hypothesis Development & Suggestion
2. K-Means Clustering:
3. What is K-Means Clustering and what is its significance
4. Why should K-Means Clustering be used in Stock Classification

* Classifying Stocks using Stock Financial Ratios
* Determining best stocks for various clusters

1. Applying various Stock Prediction techniques in order to predict the most efficient method Stock price calculation.
2. Considering one of the best stocks from previous data samples.
3. Using Linear Regression for Stock Price Prediction
4. Using SVR for Stock Price Prediction
5. What is SVR and how is it significant for Stock price analysis in accordance with Literature Reviews
6. Technical Aspects of SVR
7. Using SVR to predict Stock Price
8. Testing & evaluating the result
9. ARIMA for Stock Price Prediction
10. What is ARIMA?
11. Importance of ARIMA for Stock Price Prediction
12. Accuracy Result Using ARIMA
13. Using LSTM for Stock Price Prediction
14. What is LSTM and its significance with Stock Price analysis in accordance with Literature reviews
15. Technical aspects of LSTM
16. Using LSTM to predict Stock Price
17. Testing & Evaluating the result
18. Comparing the SVR & LSTM results for Stock Price Prediction – checking the effectiveness of either methods
19. Determining the best mix of Stocks for Portfolio and evaluating their performance – with a given contribution amount
20. Apply Portfolio Optimization on the most popular high returning assets
21. What is Portfolio? How Portfolio Optimization has been used in
22. Technical Aspects of Portfolio Optimization
23. Evaluate risk and return for existing best performing portfolio
24. Evaluate risk and return of suggested best performing stocks by the previously accumulated results of Clustering and LSTM
25. Compare and Evaluate the risk & return of the new portfolio with the best performing portfolios
26. Discussion & Recommendations of the Result
27. Suggestive Measures of improvement
28. Discuss where the model lacks in confidence
29. Finally draw a Conclusion
30. Acknowledgements & References

**Introduction:**

Evaluating Mahmoud Rajablu in “Value Investing: review of Warren Buffets investment philosophy and practice” its determined that the best investment decisions are made from careful observations based on issues such as economic reality not accounting reality as lots of intangible assets are never portrayed in accounting cycles and risk-free investments being the best investment alternative. Buffet considered that portfolio diversification would reduce risk and increase return with performance measurement being entangled with intrinsic value rather than book value. The justified reasoning towards making long term investments with good return options is viable with stocks rather than other high returning options like futures or forwards, forex and options, commodities all are short term highly risk-oriented investments. Hence through following the best investors options, stocks are justified as the best investment alternative as the simple illustration of long-term investment is viable with stocks with solid return potential.

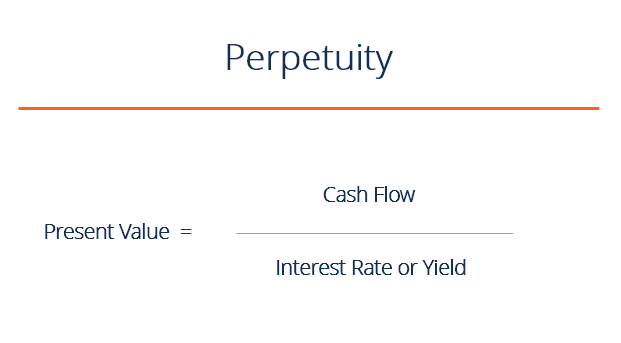
Another reason to invest in Stocks is pertained with the size of investment. Generally, stocks and bonds both are investment opportunities with Bonds being less risky but more confirmed return on investment but generally for a bigger deposit in investments. Bonds are essentially loans provided to any company which are supposed to be paid off after its due date with contractual terms of coupon payments. Whereas Stocks are ownership of a company with the risk of both the company getting undervalued or overvalued than the purchase price of stock but with dividend payments and long-term association even an undervalued stock becomes a better investment than bond. Generally, a bond with shorter maturity will have very low earning in respect to interest rate growths as per Morningstar research at 5-6% yearly rate. Whereas Stocks with proper market incentives and understanding will certainly make a greater return but in case of an undervalued stock the immediate disposal of a stock with lowest possible difference in purchase price and selling price would provide the money in hand to be invested in a better stock (will be explained quantitative applications in further proceedings)

According to SEC’s Saving and Investing “A RoadMap to Your Financial Security Through Saving and Investing” implies that if an account holds a solid 1 Million USD worth of cash what would be that money worth at some far future point of time. Its seen that there is yearly average growth of inflation rate at 2.5 to 3 % which would actually reduce the entire amount on the basis of compounding principle to a solid US$ of 443144 at a time period of 30 years and if there were any products which was worth $1, 30 years prior will be worth $2.25 nearly 30 years later. So, in reality the total amount of value worth of 1 Million USD will be 1 million items 30 years priorly and close to 197000, 30 years later with the effect of inflation for the same product. The overall effect didn’t even associate the related factors of taxation affect in investment.

According to Inspired Investor’s RBC Direct Investing, Stocks are historically better investors than low risk investment alternative than bonds as long run investment alternative. Kimberly Amadeo suggested in “Benefits of Investing in Stocks Vs Disadvantages” that usually stock investments make an average yearly return of 10% which is way higher than the yearly inflation rate of 2.9% (a solid 7.1% positive impact on growth), so stock investments remain inflation proof.

The hypothetical figure of $1 Million in savings has created a havoc in the investment world since a long time and there are various reasons behind it. In accordance to Money CNN a person who retires at 65 with a savings of $1 Million would earn $5,660 throughout their entire lifetime. The perpetuity-based investment could be explained in table 1.1

Perpetuity in the financial system is a situation where a stream of cash flow payments continues indefinitely or is an annuity that has no end. In valuation analysis, perpetuities are used to find the present value of a company’s future projected cash flow stream and the company’s terminal value. Essentially, a perpetuity is a series of cash flows that keep paying out forever.



Here present value is referred as the total savings on the basis of which the continuous cash flows would be generated. Interest rate or yield rate refers to various stable yield rates the investing funds can generate. With $1 Million in savings and if the return rate varies between 8% to any other higher terms the monthly money generations could be illustrated in the table with an inflation adjusted rate of 2.9% for various funds.

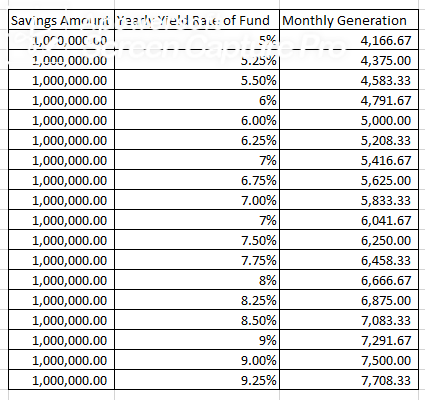


Table 1.1

According to the regulations and requirements of Securities Act of 1933 the minimum requirement for investment reasons to $1 Million in personal assets and understanding of Finance. The various investment options for monthly earning of $5660 or more is explained in the sources of US Money and Statista 2017 report of annualized return from various source.

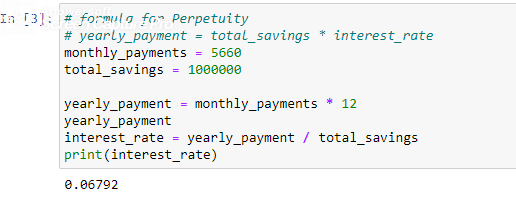


Figure 1.2

As there are various sources of investment options are present the return capability to all these funds varies on the basis of various operations set. The ETF’s, Mutual Funds, Hedge Funds are various options to invest but in the process of investment the ROI is completely dependent on those fund’s performance and decision process of investment. The choice process of these internal funds is highly determined by their internal management process whereas in this paper we would focus on how minimum investment could result towards subsequent increment in personal asset and worth sustainably.

**Description of Various Funds:**

An exchange-traded fund (ETF) is an investment fund traded on stock exchanges, much like stocks. An ETF holds assets such as stocks, commodities, or bonds and generally operates with an arbitrage mechanism designed to keep it trading close to its net asset value, although deviations can occasionally occur. As most ETFs track an index, such as a stock market index or bond market index hence the ROI will be relative to various market following index’s in accordance to Mazumder et al discussing in “Investing in Exchange Traded Funds” but that process wouldn’t follow reasonable ROI analysis towards investments except the process to follow would just bring sound return on investment by simply following the same method that good performers in market do. Similarly, various other issues such as tracking error (which defines return of ETF’s with return of reference index), liquidity risks, counterparty risks, regulatory risks and various other risks involved

A mutual fund is a type of financial vehicle made up of a pool of money collected from many investors to invest in securities like stocks, bonds, money market instruments, and other assets. Mutual funds are operated by professional money managers, who allocate the fund's assets and attempt to produce capital gains or income for the fund's investors. A mutual fund's portfolio is structured and maintained to match the investment objectives stated in its prospectus.

A well performing Mutual Fund is very similar to a portfolio which has best possible investment understandings and would require a huge cash balance in their account to make proper investments decisions that would accrue proper ROI. Moreover, the investment account will have fluctuating returns which at time of requirement might be negative standing. Mutual Funds generally practice a high fees, commissions and other expenses which brings down the overall ROI. Generally, a ROI of positive return is predicted through most Mutual Funds but a risk is still associated with negative return with high maintenance fees and inflation rate.

Hedge funds are alternative investments using pooled funds that employ different strategies to earn active return, or alpha, for their investors. Hedge funds may be aggressively managed or make use of derivatives and leverage in both domestic and international markets with the goal of generating high returns (either in an absolute sense or over a specified market benchmark). It is important to note that hedge funds are generally only accessible to accredited investors as they require less SEC regulations than other funds. One aspect that has set the hedge fund industry apart is the fact that hedge funds face less regulation than mutual funds and other investment vehicles.

Risks involved with investment options in Hedge funds are unregistered investments as most other funds follow lots of SEC registration and disclosure requirements, the funds generated in hedge funds are sourced from various unregistered or non-disclosed sources, the opportunities to engage in fraud for hedge fund managers are higher. The investment options in hedge funds are much riskier than other funds as they follow much aggressive process of investment to generate more profits. The process of controlled earning and sustainable growth generally don’t work for hedge fund earning process. Hedge funds employs options, forwards and futures, commodities, forex and various other much riskier investment options and follow much riskier process to generate more profits hence the risk involved with funds are generally higher. Hedge funds, both the unregistered and registered variety, are illiquid investments and are subject to restrictions on transferability and resale. Registered hedge fund units may not be redeemable at the investor's option and there is probably no secondary market for the sale of the hedge fund units. Unlike mutual funds, there are no specific rules on hedge fund pricing, which means that during the time of need the moneys could not be liquidated as easily as possible.

Below various funds with their annualized yield percentage are described below:

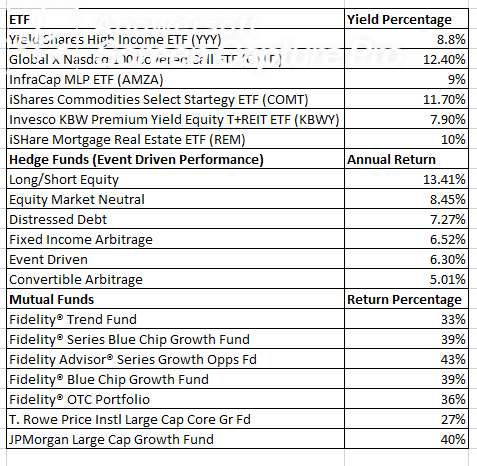


Table 1.3

In accordance to various updates provided in Table 1.2 it could easily be seen that there are various hedge funds which certainly makes way better investment return than the minimum amount of 6.7% yearly return. Although there are various hedge funds which participates with a minimum investment of $100,000 as the minimum investment amount but the yearly return comparatively becomes way smaller in amount.

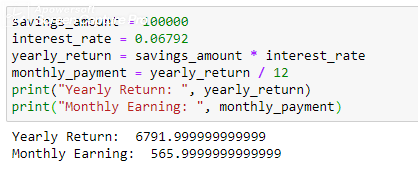


Figure 1.4

Its seen that with smaller investments the overall monthly and yearly earning both remains pretty low, hence for most investors it’s better to consider a larger investment which would generate a bigger yearly or monthly return and $1 million is considered as a solid step in achieving that target.

In this research paper we would suggest ways how we can reduce down the overall working age period of people and how people would live comfortably by contributing their savings of $1 Million in highly capable funds with solid monthly earnings after their retirement with comfortable monthly investments in a highly reliable stock portfolio-based fund.

**Scenario & Situational Analysis:**

Most of the employees are faced with high work & peer competition pressure, extreme stress and undervalued compensation throughout their entire beginning 10 years of life as 41% of employees listed job security as the main reason for staying in job (Willis Tower Watson, January, 2020).

The average yearly inflation in North America as a whole is 2.12% (dataworldbank.org, 2020), whereas the average starting salary for a college graduate is varying from $1042 a week making an average earning of $50000 a year. The average salary increases rates due to good economic condition is also forecasted at 3.1%, the best since 2008(Mercer 2015/2016 US Compensation Planning Survey). In accordance to National center for Education Statistics said that the average age of college graduation is 22, which gives most working individuals an average working lifespan of about 45 years. But due to various issues like graduate school, losing job, career change and other issues have given an average working time period of 41 years and 3 months (Bureau of Economics, Employment Survey 2017) for average college graduates, it seems that the overall earning period shortens for most employees. It’s also seen that the average monthly retirement earnings are $1470 from Social Services for USA and $1200 CAD for employment pension in Canada. More than half of US citizens aged 65 or older makes $24,244 or less in a year. In Canada despite being a sound retirement benefit plan, it still makes the mark of requirement management barely. For fully professional retirees the social security benefit plan stays at $3011/month at full retirement age. According to Investopedia, only 7% of US and Canadian retired households actually earn a yearly amount of more than $48,400 a year, which in reference to present days cost structure is very tight cost management budget.

Our sample study will happen with people who are minimum college graduates as for our continuous investment to be successful we need people with solid earning potential or people who will get back to professional jobs after a short while. The reason for such analysis is because in USA and Canada its seen that most people change their jobs 12 times in their entire lifetime (according to Bureau of Labor Statistics, 2019). According to US Bureau of Labor Statistics, 2019 report its seen that a college graduate and high school graduate diploma has an average difference of earning limits of $750 and $1280 respectively. In case of non-graduates, it seems that after paying the requisite amount of taxes and all expenses, the left-out portion will be less than $2700 for a whole month, which makes it very stringent for risk associated investment option. The overall earning period for college graduates and the non-graduates also seem highly different as the job security of non-college graduates are nearly 42% lower than college graduates hence their overall earning lifetime also falls down severely (average period of 40 years with job loss and 35.5 years without it, Carnevale et al, The College Payoff, cew.georgetown.edu).

Our analysis below is generated from the basic earning level denoted by US Bureau of Labor Statistics, 2019 report for average monthly salary, average expenditure and average monthly savings. Although S&P 500 shows an affirmative 12.25% positive return yearly but in an average any mixed portfolio will have 7% yearly return.

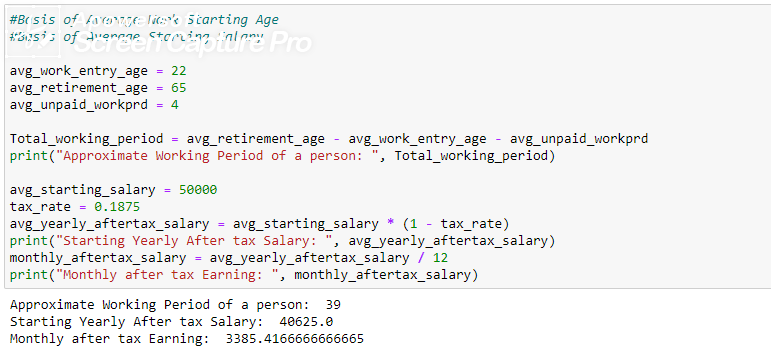


Figure 1.5

Its seen that throughout the entire lifetime of 39 years and through the change of 12 professional jobs most of the people are looking towards various factors for settlement in a job starting from compensation amount (which according to Towers Watson Survey 2017 nearly 59% employees feel they are underpaid), work environment and peer pressure, various other issues. According to Mercers 2015/16 US Compensation Planning Survey it’s seen that the expected raise an employee can expect is 3.0% in average with a five years consistency. The previous reports have suggested that the overall yearly growth rate remained at 2.6% to 2.8% considerably with various economic events driving the scenarios. The brightest employees will see a raise between 4.6% to 5.2% for least growing to highly growing job fields. The underperformers could expect a 1% yearly raise.

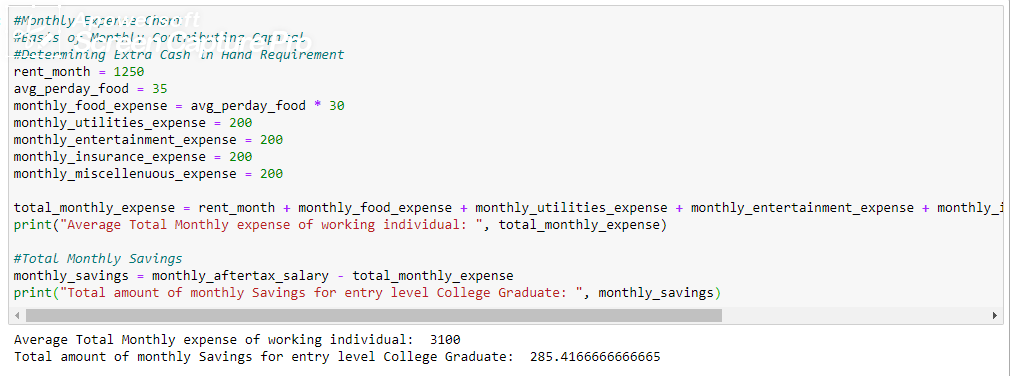


Figure 1.6

An average person who starts with a after tax starting salary would expect to see the above-mentioned expenses as their primary expense which would include rent, food expense, utilities expense, insurance payment, entertainment expenses and miscellaneous expenses as their major monthly expenses. In accordance to SEC’s Savings and Investing if a person wishes to have secured financial future an investment in both the savings account which would work as a cash basis and investment in money market which would adequately grow the money creates a positive impact in their savings. In accordance to Mercers report for 2015/16 with 3% average annual increase in salary the portrayed salary chart could be

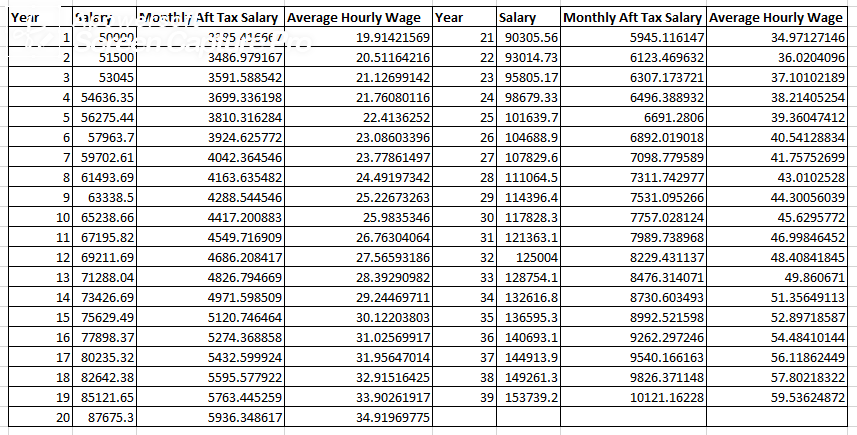


Table 1.4

If we analyze CRS report for “Real Wage Trends. 1978 to 2018” we can see the chart matches perfectly well with the after-tax hourly wage rates which seems our analysis matches with yearly growth trends. Similarly, the 2016-17 pay-scale college salary report also says that average yearly salary for mid-career ranges at $150k.

**Determining Contribution Amount:**

Its seen that at early career stage most working individuals with college degree or more than $3500 earning each month are capable of making very little amount of contribution to their savings. As the savings contribution ranges from $200 to $400 at maximum and in accordance to SEC Savings & Investing guideline people should contribute half their monthly savings towards savings account and half towards investment which makes the window very narrow for investment option. In reality with such options most investors cannot consider large cap stocks or riskless high return-based bonds which requires subsequently larger investment capability (as most highly reputed stocks are AAPL: $1217, MSFT: $1302, AMZN: $1119, FB: $523, BABA: $561) at the beginning stage of investment with smaller investment contribution capability and high-risk orientation with large cap stocks and lower earning potential. In the analysis process we can see that if we purchase a stock worth less than $10 and at the same time, we purchase one of the large cap stocks for $500 than even though the possibility of large cap stock with good reputation in the market making a loss is way less than the $10 stock but the large cap stock itself is carrying 50 times ($500 / $10 = 50) more risk with downfall of price. Yang et al (2016) have established that in general stocks have high degree of relativity and changing capability in reference to news and social media sentiments with people’s choice and perception. In that case the price volatility for large cap stocks with more popularity are more complimented to people’s perception than various other notions like statistical and financial ratio factors. Switzer et al suggested that large cap stocks perform better over peaks. Jonathon Burton suggested in WSJ the for longer term investment performance it’s better to invest in large cap stocks which are generally good performers for longer periods (historically its seen that every large cap stocks actually makes a rise in an operating cycle) that results to make them a highly priced stocks with time. But with small investment capability and much less contribution margin at the beginning of stock investment period it’s never an easy decision which stock to make an investment and for what price.

Prior to making a choice regarding the investment decisions, the important questions would be how much an individual coming from a certain social status and earning capability can contribute to investing each month. Its seen that according to  **U.S. Bureau of Labor Statistics (BLS) Consumer Expenditure Survey of 2018** the average household size of 2.6 people (ArcGIS data 2019) have an average monthly expenditure of $5,102 each month. Our analysis is aligned with various of the same expenditure aspects with increase of expense in few sectors as their expenses are increased for single person in some issues considerably. Our analysis in Figure 1.6 explains how for a single person the spending would increase considerably. Although in accordance to average household size and expenditure singular monthly contribution should be limited to $1962 (5102 / 2.6) but as the single person needs to pay off various extra payments on their own hence their spending margin increases.

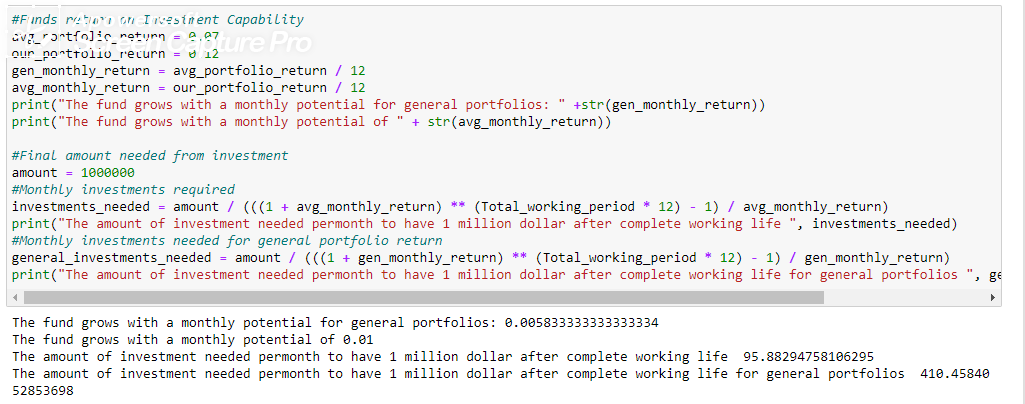


Figure 1.7

In the above figure we see how we can use the small contributing amount in investment and can make certain profit from it. As S & P 500’s best possible investment suggests that it makes considerable return of 10% in a year. According to table 1.3 it’s also seen that various Mutual Funds and ETF’s are present which makes considerable return on investment but all comes with various risk considerations. According to Financial-Samurai the historical evidence since 1926 on 100% stock-based portfolio’s the return is 10.2% which is very similar to S & P 500’s yearly ROI. The average return on investment falls to 7% effective growth with effects from inflation of 2.9% in average.

The simple target for the portfolio is clearly determined in the above figure using the principle of annuity. Considering the FV of Annuity in investments we have found that if any investment is capable of making a 1% positive ROI then that investment is capable of making 1 Million throughout their entire working lifetime with an affordable amount of payment for entire lifetime. The consideration of working lifetime differs from person to person and their capability to invest on the basis of their own belief system. Hence our portfolio-based return analysis will try to accommodate various options of investments on the basis of various complete work life.

The future value of an annuity is a way of calculating how much money a series of payments will be worth at a certain point in the future.

The formula for the future value of an ordinary annuity is as follows:

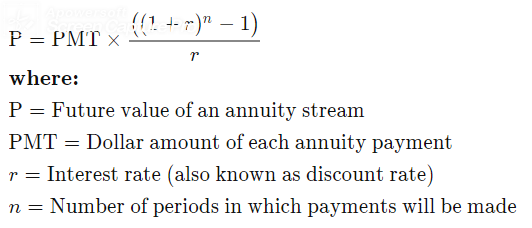


Figure 1.8

Considering People’s various choice pattern of investments and people’s varying work starting age there are various issues on the basis of which a person can have different contributing margin. Let’s consider a person have started job at a lucrative firm for well over 200k yearly salary but at the age of 35 years (complementing advanced degree and certifications needed for high performance) he/she can easily contribute to $685 a month in a portfolio, whereas in other hand a person who starts working at the age of 22 but have a low salary growth potential but a very long working time period can easily attain the 1 million savings benchmark with a very low investment of around $70 a month (equivalent to monthly phone plan). It’s considered that most individuals will have at-least 3 years no contribution in their entire lifetime on the basis of various issues like marriage, graduate studies, savings for home down-payment and other factors. Our calculation has left out more than three years noncontributing time period (as from 22 to 67 its 45 years of working lifetime). There are various issues like dividend payments which will only accelerate the early retirement option or extra earning potential through investments but we will only discuss about it in later chapters.

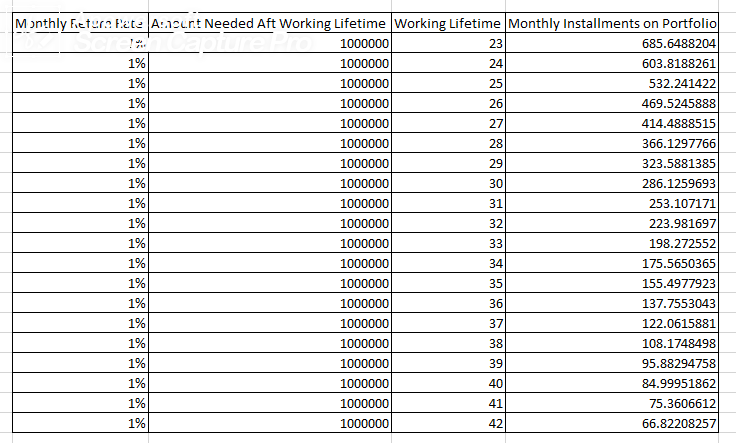


Table 1.5

Chapter 2

Portfolio Analysis

**What is Portfolio?**

It’s a mixture of various securities instruments like annuities, futures, options, stocks, bonds and various other investment classes. A portfolio is a collection of financial investments like stocks, bonds, commodities, cash, and cash equivalents, including closed-end funds and exchange-traded funds (ETFs).

**What is Portfolio Analysis?**

Portfolio analysis is an examination of the components included in a mix of products with the purpose of making decisions that are expected to improve overall return. Generally, a portfolio analyst or manager makes the application of the term to the process that allows a manager to recognize better ways to allocate resources with the goal of increasing profits.

**Managing investments into Portfolio’s:**

Generally, it’s found that from our previous analysis that every month a certain amount of investment would ensure a guaranteed monthly earning after their retirement in accordance with the calculations performed on the basis of basic perpetuity and annuity calculations. But in reality, the stock market faces various other important aspects like sensitivity to important and relative news (Joshi et al June 2016 and Yang et al February 2016), financial and statistical ratio indicators (Arkan 2016), overall aspects of Buffets indicators for positive performance indicators (Rajablu 2011) as the important contributors for making successful choices for investment analysis.

Its seen from our above analysis that a person would have to invest less with the increasing work life period which seems logical enough as there should be more accumulated investments which can gain relative timely profits to accumulate into $1 Million. But in reality, the margin for successful life savings and monthly contribution amount is changing with time. In 1980’s the simple concept of $5000 a month in relation to CPI and PPI was a very considerably lucrative amount which may not be so at this moment under the same parameters due to inflation and various other social factors. Also, there is an important question for attaining a specific monthly goal of 1% remaining with influencing factors contributing towards risk and return gain (Sharpe ratio indicator) if the investment is capable of earning a risk free 1% monthly. Also, our further analysis will show how we can increase our ROI and IRR of the portfolio from 1% to 2% or 3% monthly positive rates such as various Mutual Funds ROI (source: mutualfunds.com). Our analysis will also provide various statistical and financial ratio-based insights that will justify how the investments are best possible in accordance to various situational and conditional analytics. In order to make our primary investment process less risky our portfolio would remain focused only on Stocks and no other riskier investment alternatives like options, forwards or futures, commodities or forex etc. We would also not get involved in less risky investment alternative like bonds which would not make the monthly investment requirement benchmark limitations that stock investment can withstand and the same time the bond ROI remains extremely low to attain considerable amount of $1 Million for smaller investment section.

**What is a Stock?**

A stock (also known as equity) is a security that represents the ownership of a fraction of a corporation. This entitles the owner of the stock to a proportion of the corporation's assets and profits equal to how much stock they own. Units of stock are called "shares."

**What is a Stock Portfolio?**

Investing in a stock portfolio is a way of trying to profit from increases in the share price of a range of companies rather than simply betting on one.

By owning shares in a range of companies – carefully chosen to represent a balance of different sectors and company types, the portfolio owner will have more control over the risk associated with equity investment. The idea is that if one of the invested stocks experiences falls in price, the portfolio may get better compensations by increases in the price of other stocks in present in portfolio.

**Method Analysis: For Portfolio Construction:**

As there are various categories of stocks present in the market like large cap stocks, mid cap stocks, small cap stocks in reference to market capitalization. There are various other ways to classify stocks in relation to industry type, in relation to beta response and various other issues. Our analysis process would focus only on the best performers among these stocks in relation to investment capability and slowly shifting investments from riskier to less risky but more favorable investment alternatives.

**Large Cap Stocks:**

Large-cap stock also known as big caps are shares that trade for corporations with a market capitalization of $10 billion or more. Large-cap stocks tend to be less volatile during rough markets as investors fly to quality and stability and become more risk-averse. Switzer’s suggestion was that in reference to small and mid-cap stocks, large cap stocks are seen to provide long term and steady return, more regularized dividends and high liquidity standards in comparison to small and mid-cap stocks in reference to lower volatility. But in general, its seen that most large cap companies have a high beta which means their prices are reflected highly on the basis of news or social media responses and a lot of the price variations are not dependent on the factors of actual performance qualification of the company (Public News Impact on Prices, Ormos, February 2011). Hence in case of large cap stocks the best ideology for investment would be to make intrinsic analysis before making guided judgements for short term investments. In our analysis our primary focus would be to make short term investment decisions to gain profits.

**Mid Cap Stocks:**

Mid-cap is an approximate term that encapsulates companies and stocks which fall in between large-cap and small-cap category. The classification of respective company’s stocks depends on its market capitalization. Such classification is variable and can change with the change in a company’s market valuation.

Mid-cap is the term given to companies with a market capitalization (value) between $2 and $10 billion. As the name implies, a mid-cap company falls in the middle of the pack between large-cap (or big-cap) and small-cap companies.

Benefits of Mid Cap Stocks:

1. When interest rates are low and capital is cheap, corporate growth is stable. Mid-cap companies can get credit they need in order to grow, and they do well during the expansion part of the business cycle.
2. Mid-caps are not as risky as small-cap companies, which means they tend to do relatively well financially during times of economic turbulence.
3. Many mid-caps are well known, are often focused on one specific business and have been around long enough to make a niche in their target market.
4. And finally, because they are riskier than large caps, they may have a higher return, which could be more appealing to a less risk averse investor's bottom line.

**Small Cap Stocks:**

Small cap is a term used to classify companies with relatively small market capitalization. A company's market capitalization is the market value of its outstanding shares. The definition of small cap can vary among brokerages, but it is generally a company with a market capitalization of between US$300 million and $2 billion.

**Advantage of Investing – Small Cap Stocks:**

One of the most significant benefits of investing in small-cap stocks is the opportunity to beat institutional investors. Because mutual funds have restrictions that limit them from buying large portions of any one issuer's outstanding shares, some mutual funds would not be able to give the small-cap a meaningful position in the fund. The fund would usually have to file with the SEC to overcome these limitations. When a fund does this, it means tipping its hand and inflating the previously attractive price.

**Methodology:**

In accordance to Mark Hulbert’s article on WSJ regarding “The Big Reason Behind Small Cap’s Struggle” it is priorly understood that most Mutual funds and ETF’s do not plan to invest in small cap stocks. But a careful judgement of situation and better understanding regarding their financial status and future prospects could provide important insights regarding issues that suggest a good ROI from the investment in Small Cap Stocks. Hence, our methodology will not stop at the boundaries of general market principles but would be very certainly contained for investment decisions for general market guidelines.

Our Principle of stock analysis would follow through the below mentioned analysis process.

1. We will primarily select the best stocks present in the market. The stocks would be included in the best possible market groups of according to suggestive patterns provided as news outlines suggestions, online forum and portal suggestions, variant financial investment-based website suggestions and various well-known stock market index and various famous stock selections by Mutual Funds and ETF’s.
2. After selecting the big group of stocks, we would classify the stocks into various classes in terms of variant features – using K-Means Clustering.

**K – Means Clustering**

1. What is K-Means Clustering?

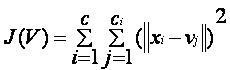
In accordance to Andrey Bu the objective of K-means is simple: group similar data points together and discover underlying patterns. To achieve this objective, K-means looks for a fixed number (*k*) of clusters in a dataset. (James MacQueen 1967)

Cluster: A cluster refers to a collection of data points aggregated together because of certain similarities.

Working Process of K-Means Algorithm:

1. Specify number of clusters *K*.
2. Initialize centroids by first shuffling the dataset and then randomly selecting *K*data points for the centroids without replacement.
3. Keep iterating until there is no change to the centroids. i.e assignment of data points to clusters isn’t changing.

* Compute the sum of the squared distance between data points and all centroids.
* Assign each data point to the closest cluster (centroid).
* Compute the centroids for the clusters by taking the average of the all data points that belong to each cluster.



where,

*‘||xi - vj||’* is the Euclidean distance between *xi* and *vj.*

*‘ci’* is the number of data points in *ith* cluster.

*‘c’* is the number of cluster centers.

Algorithmic Steps of K-Means Clustering:

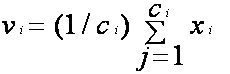
Let X = {x1,x2,x3,……..,xn} be the set of data points and V = {v1,v2,…….,vc} be the set of centers.

1) Randomly select *‘c’* cluster centers.

2) Calculate the distance between each data point and cluster centers.

3) Assign the data point to the cluster center whose distance from the cluster center is minimum of all the cluster centers.

4) Recalculate the new cluster center using: where, *‘ci’* represents the number of data points in *ith* cluster.



5) Recalculate the distance between each data point and new obtained cluster centers.

6) If no data point was reassigned then stop, otherwise repeat from step 3.

Advantages of K-Means Clustering:

1) Fast, robust and easier to understand.

2) Relatively efficient: O(tknd), where n is # objects, k is # clusters, d is # dimension of each object, and t is # iterations. Normally, k, t, d << n.

3) Gives best result when data set are distinct or well separated from each other.

Relativity with Cost based Stock Clusters:

The cluster of stocks (if determined) with smaller investment requirement would allow people with lower earning to be able to participate in the investment process. Our analysis previously has shown that there is variant investment amount based on the time period through which the investment could be made. Its seen that the amount of money contributed is inversely proportional to the total time requirement to reach certain investment amount.

1. Price of Stock

What is the cost of the stock? – The lower the cost of the stock the more favorable would be the stock for investment for shorter term gain.

The reasonable expectation behind such stock classification lies within the logical grounds of Dakota, September 2002 with “Growth Stocks outperform Value Stocks Over Long Term” which clearly indicates it’s better to make investments in growth stocks although which are highly priced as they have a tendency to perform well in the long run. Our Stock purchase strategy will rely heavily on the factor of short-term gains should reasonably deduct losses in the proportion to generate a reliable amount of positive ROI for a specific time period. But as the investors are incapable of providing higher investment thus the portfolio investment would get riskier with all value investments more. Hence even though the above analysis in the research paper suggests to make minimum investment to generate a solid positive 1% monthly ROI but in reality, to reduce the overall risk factor for stock investment it’s better to make investments in a larger amount to generate the same 1% return on investments.

1. Risk Assessment of the Stock – The least the risk associated with the price variability the better the stock behavior could be predicted. In reality its seen that the stock beta and the Sharpe ratio associated with beta is moreover an important aspect for risk analysis associated with the stock. Beta basically determines the stock’s value appreciation or depreciation in reference to market movement. Hence the diverge the beta value of stock from 1 the more stock price fluctuations will be observed in reference to market movement. Whereas if a stock exhibits a greater Sharpe ratio then it will initiate more risk adjusted return, as Sharpe ratio is the measure of an investments return compared to its risk.
2. Volumes of Stocks – currently being traded in the market and the 3 months average

The growing number and average of volumes of stock would suggest the stock is gaining more popularity and the investment in the stock is increasing. Chandrapala’s explanation regarding “The relationship Between Trading Volume and Stock Returns” suggests that there is positive relationship for stock return in reference to increase in Trading volume. Also, Moatemri et al (2017) have suggested that there is issue of overpricing that results to increased volume in shares in the market but the correct identification would always result for overpriced and underpriced stock groups and exactly priced stock groups and should be highly reliable with financial and statistical ratios regarding the stocks or market suggestions.

Relative Analysis:

Our primary Analysis would make classes of Stock using K-Means Clustering Algorithm in order to find stocks of different classes using their price factor. This method would differentiate highly priced stocks from low/cheap stocks on the basis of prices. In this way the stocks would get clustered into various groups of different price ranges. It’s important as our investors have different ranges of capability in reflection of their earning capability and its seen that the longer period the investments are required to achieve its original goals the smaller are the investment capabilities as discussed and shown in Table 1.5.

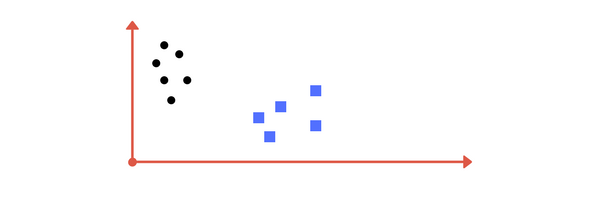
After making our primary clusters we will consider each group of stocks and start making clustering based on risk assessment of stocks which is dependent on Sharpe ratio and Stock beta. This analysis would result towards best performing stocks with highest return based on risk and their respective movement with market performance. Each stock from the price groups will have a list of best performing and worst performing stocks based on risk analysis and volatility with market movement.

The best groups of stocks will then be tested with another factor of LSTM process to figure out if the stocks trading volume would rise in the upcoming 30 days in the future. As LSTM is successfully capable of denoting the trait then at least success presumed would be the upward and downward stock volume. Its seen that through Chandrapala’s and Moatemi et al suggested that stock return is based on trading volume hence an approximate growth and decline could give insights regarding future stock ROI. The process of Stock price prediction needs to be tested through various market standard processes which are already existent like SVR (Support Vector Regressor) proposed by Henrique et al Sept 2018, HMM (Hidden Markov Models) proposed by Nguyen, March 2016, and finally LSTM (Long Short-Term Memory). Our analysis would recall also a further extension of analysis which follows the efficiency of each models based on test data relevancy with real data in order to show which models among them would be the most efficient.

The analysis of such process would effectively define the best stock price prediction method using efficiency of the data testing with predicted value and real value with the closest possible value prediction as the best measure for prediction.

**SVR (Support Vector Regression) - Using SVM (Support Vector Machine):**

SVM: A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplane. In simpler words, given labeled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples.



In the above figure as the two different classes are distinctively different from each other, the dataset gets differentiated into two distinct classes using SVM algorithm. Similarly, there are some more efficient ways to effectively differentiate various classes of data using kernel operations and nonlinear classification techniques all of which could be applied through SVM. SVM is capable of predicting the movement of direction of stocks in reference to increase or decrease in price through the usage of non-linear classification method that would differentiate lower price range from the higher price range.

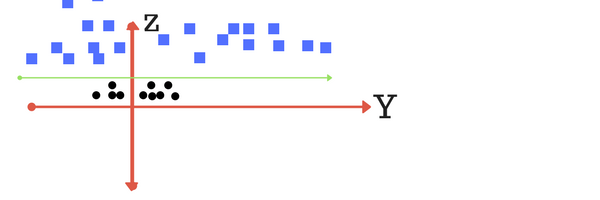
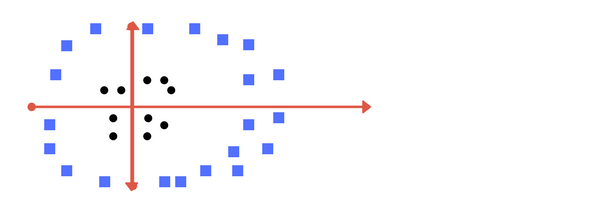


Figure: SVM

**Objective of SVM:**

The exact objective of an SVM is to find a hyperplane in an N-dimensional space (N – the number of features) that distinctly classifies the data points.

In order to separate between various classes of data points there should be many possible hyperplanes which could be chosen. Our objective is to find a plane that would clearly distinguish between different classes. There could be many hyperplanes that would satisfyingly distinguish between various classes. True objective of SVM will be to find an exact plane that will be able to define the maximum margin which represents the maximum distance between the data points for two or more classes

The positive reasoning behind the maximizing distance between classes would suffice that it provides enough reinforcement between the data points of the classes so that they could be easily distinguished with more precision for future data points.



Fig: SVM with Margin

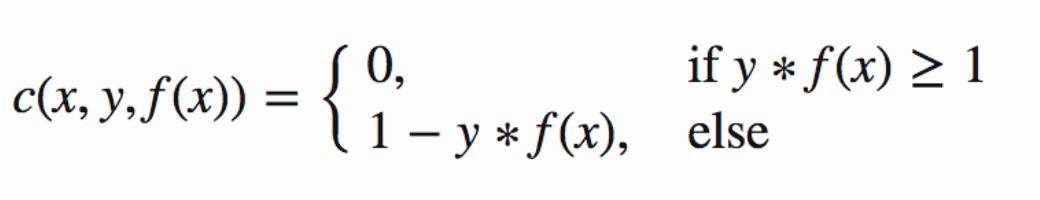
Our target for SVM will be to create hyperplanes that would be similar to the right image which has more margin in between the classes as that would clearly differentiate the classes with more effectiveness rather than the left image.

Generally, SVM gives us the continuous improvement option by making developments in the margin increments by reducing the loss associated with hyperplane difference. The support vectorization helps to maximize margin of the classifier.

**Large Margin Setup:**

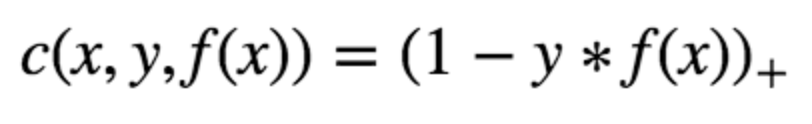
Similar to logistic regression when the output values are ranged within [0,1] with higher than 0.5 is assigned to 1 and others to 0. The SVM segregates the output values in different ranges from [1,-1] with values greater than 1 to 1 and values less than -1 to -1 thus making a larger difference in margin between the classes.

Loss Function:

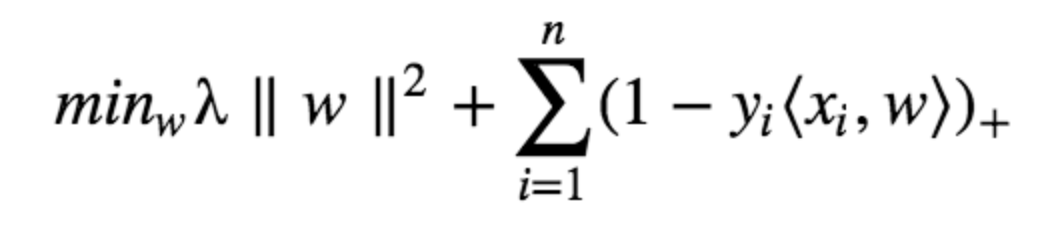


The loss function helps maximize the margin is considered as the hinge loss. The cost is 0 if the predicted value and actual value are in the same class. When it’s not then the regularization parameter is considered as the cost function. The objective of regularization parameter is to balance the margin maximization and loss.

Regularization Parameter:

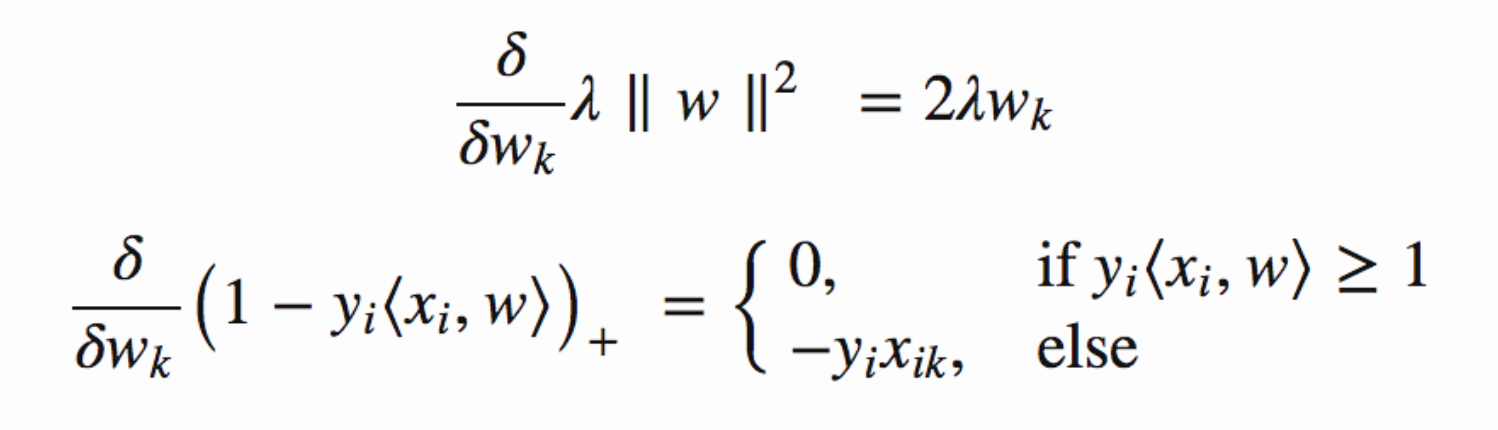


Adjusted Cost Function (With Regularization Parameter):



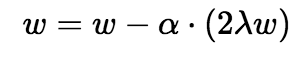
Generally, the adjusted cost function is adjusted with loss function with respect to the weights to find gradients. The partial derivatives with respect to weights find the gradients. Our weights are updated using the gradients.

After the completion of partial derivatives, the updated gradients are exhibited:

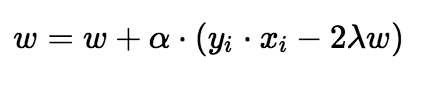


**Steps to Follow for Weight Update:**

* 1. When the cost function makes no mistake in classification, our model has correctly predicted the data points, in that case the gradient from the regularization parameter needs to be updated.



* 1. When there is a mistake in terms of making classification the loss along with the regularization parameter to perform gradient update is mentioned.



**ARIMA (Auto Regressive Integrated Moving Average)**

Auto regressive Integrated Moving Average is actually a class of models that explains a time series based on its past values which is integrated with its own lags and lagged forecast errors. An ARIMA model is characterized by three terms: p, d and q

P is the order for AR term.

Q is the order for MA term.

D is the number of differencing required to make the time series stationary.

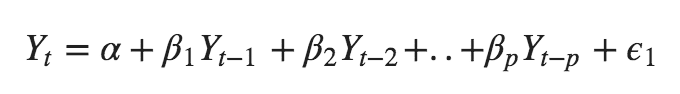
Generally, a time series with seasonality pattern is coined to be termed as SARIMA which is Seasonal ARIMA.

ARIMA is modified Linear Regression Model that uses its own lags as predictors. The best functionality of Linear regression model is achieved when predictors are not correlated and are generally independent of each other.

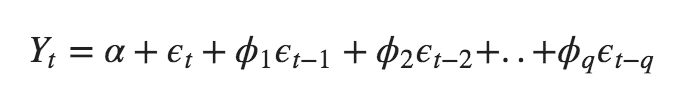
The process of making the time series stationary is appropriated through the subtraction of previous value with current value. In relation to the complexity of the series the differencing is needed more than once. The value of d thus becomes the minimum number of differencing required to make the time series stationary. If the time series is already stationary then d = 0.

In general, a ARIMA forecast never correlates completely with either AR or MA but its mostly based on the combination of both.

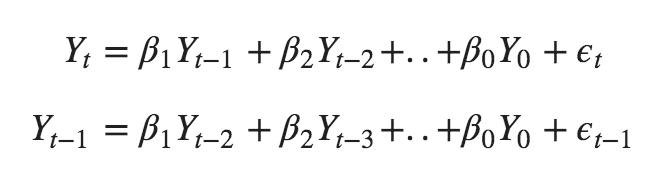
A pure AR (auto Regressive) Model is completely dependent on its lags for producing a predictable output. It is generally termed as the function for ‘lags of y(t)’.



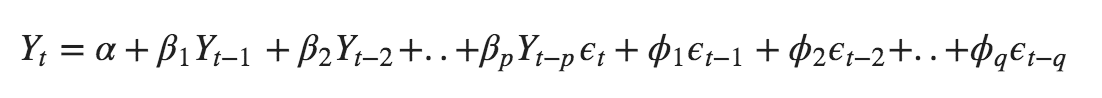
Here each Y(t-1) to Y(t – p) is the lag1 of the series. The beta1 to beta(p) is the coefficient of lag1 that the model estimates and alpha is the intercept term estimated by the model.

A pure MA (Moving Average) model is the one where y(t) depends only on the lagged forecast errors, where the error terms are the errors of the autoregressive models of the respective lags. 

In both the situations for forecasting y(t) and y(t-1) the estimated errors are represented by the following equations in reference to E(t) and E(t-1).



As ARIMA is the combination of both AR & MA then the only demand for perfect forecasting is the combination of both the processes while making them stationary.



The actual realization of ARIMA in words:

**Predicted Yt = Constant + Linear combination Lags of Y (upto p lags) + Linear Combination of Lagged forecast errors (upto q lags)**

**Order of Differencing in ARIMA (Determining d):**

**The purpose of** differencing is to make the Time Series Stationary. A carefully prepared forecasted model will not have higher degree of inefficiencies in relation to over differences because that will affect the model parameters.

The right order of differencing lies within the boundary of near -stationary series which roams around a defined mean and it relies to be performed by two processes.

ACF (Auto Correlation Plot)

PACF (Partial Auto Correlation Plot)

A fairly treated differencing by ACF reaches to zero fairly quick.

It’s generally a prediction that if the autocorrelations are positive for many numbers of lags (usually 10 or more), then the series needs further differencing. On the other process, if the lag 1 autocorrelation itself is too negative, then the series is probably over differenced.

First process will rely on the method of testing the series through Augmented Dickey Fuller Test, If the series easily predicts that d = 0 then no differencing is required.

In general, the result is tested through hypothesis testing processes that defines if p < 0.05 then the time series is indeed stationary. If P > 0.05 then the series is not stationary hence requires differencing.

**Order of AR (Auto Regressive) term:**

Partial autocorrelation can be imagined using correlation between the series and its lags. The process is performed after excluding the contributions from the intermediate lags. So, PACF conveys the pure correlation between a lag and series.

Thus, Partial Autocorrelation of Lag in the series is the coefficient of that lag in the autoregression equation of Y.

**Y(t) = Alpha(0) + alpha1 \* y(t-1) + alpha2 \* Y(t -2) + alpha3 \* y(t -3) + ………… + alpha n \* Y(t – n)**

Any autocorrelation in a stationed series can be rectified by adding enough AR terms. The order of AR term is equal to as many lags that crosses the significance limit of PACF plot.

**Order of MA (Moving Average) term:**

An MA term is technically the error of the lagged forecast. The ACF tells us how many MA terms are required to remove any autocorrelation in the standardized series.

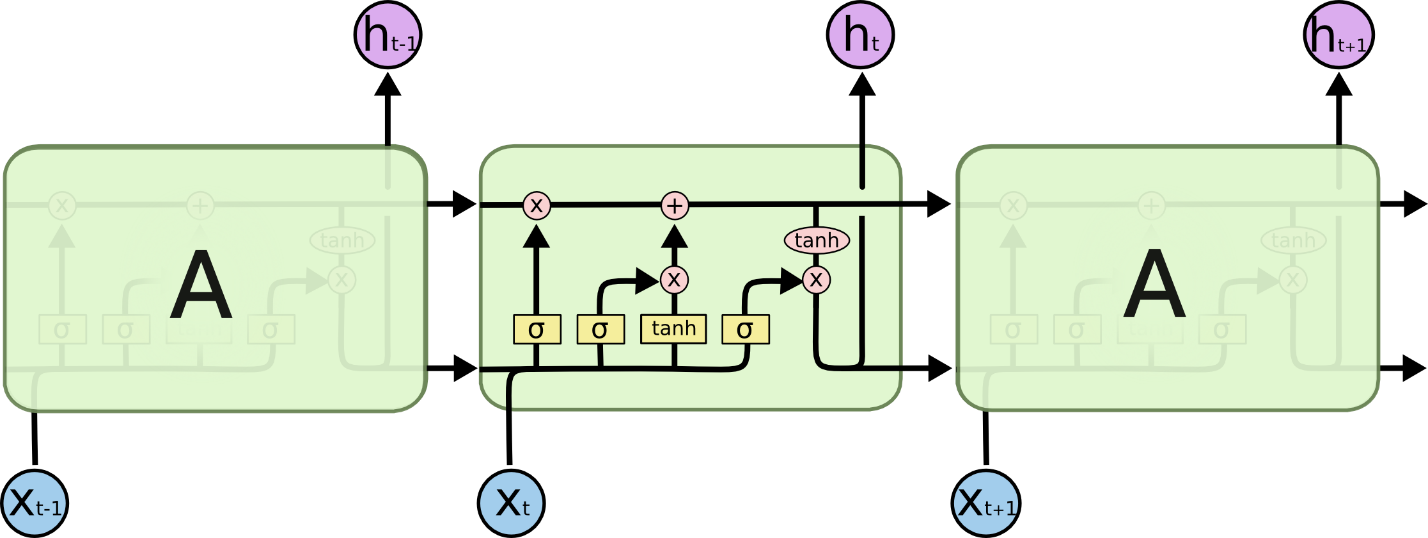
**Accuracy Metrics for Time Series Forecast:**

The commonly used accuracy metrics to judge forecasts are:

1. Mean Absolute Error
2. Mean Error
3. Mean Absolute Error
4. Mean Percentage Error
5. Root Mean Squared Error
6. Lag 1 Autocorrelation of Error
7. Correlation Between the Actual & the forecast
8. Min-Max Error

**LSTM (Long Short-Term Memory):**

A recurrent neural network can be thought of as multiple copies of the same network, each passing a message to a successor. An LSTM is a kind of RNN that are capable of learning capable of learning long-term dependencies (Hochreiter & Schmidhuber (1997). All the recurrent neural networks have the form of a chain of repeating modules of neural network. In standard RNNs, this repeating module will have a very simple structure, such as a single tanh layer. LSTMs also have this chain like structure, but the repeating module has a different structure. In case LSTM there are four layers instead of one interacting in a specific way.

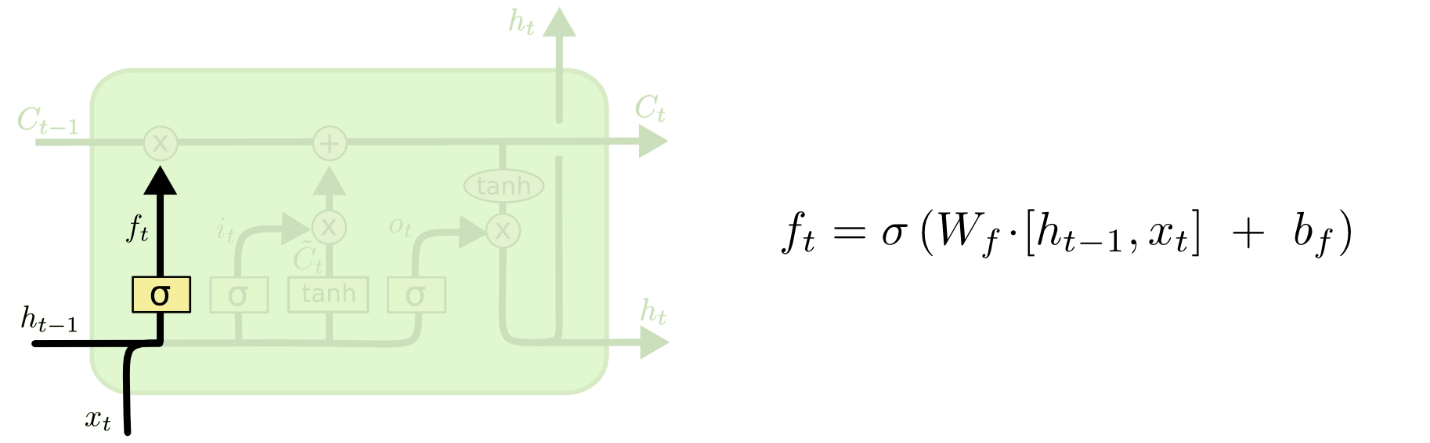


The core idea behind LSTM is the cell state, the horizontal line running through the top of the diagram. The cell state is kind of like a conveyor belt. It runs straight down the entire chain, with only some minor linear interactions. It’s very easy for information to just flow along it unchanged. The LSTM does have the ability to remove or add information to the cell state, carefully regulated by structures called gates. Gates are a way to optionally let information through. They are composed out of a sigmoid neural net layer and a pointwise multiplication operation.

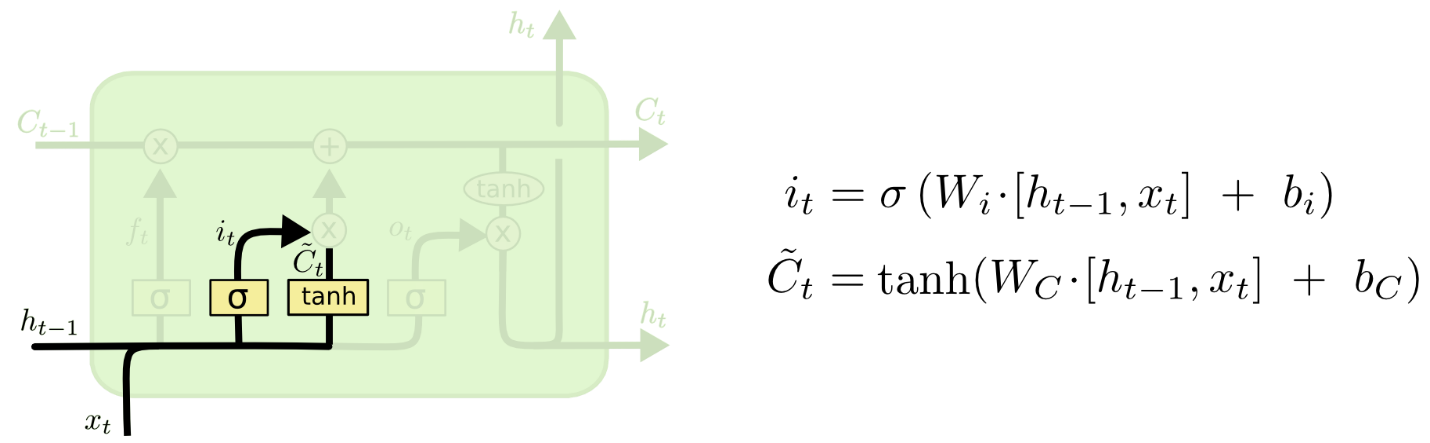


**Step by Step Walkthrough Process of LSTM:**

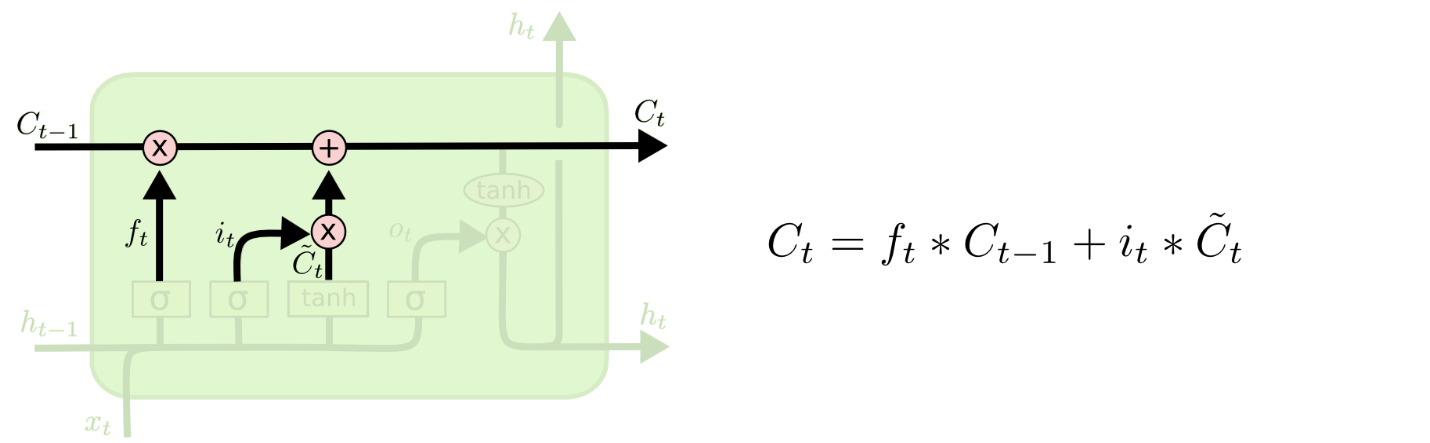
LSTM needs to decide within the context what information they need to keep and what information they will throw away in the next step in each cell state. The decision is made by sigmoid layer called the “forget gate layer”. The h(t -1) and x(t) are both looked at with 0 and 1 for each cell number c(t -1). The decision remains simple as 1 represents to remember the information whereas 0 represents to forget the information.



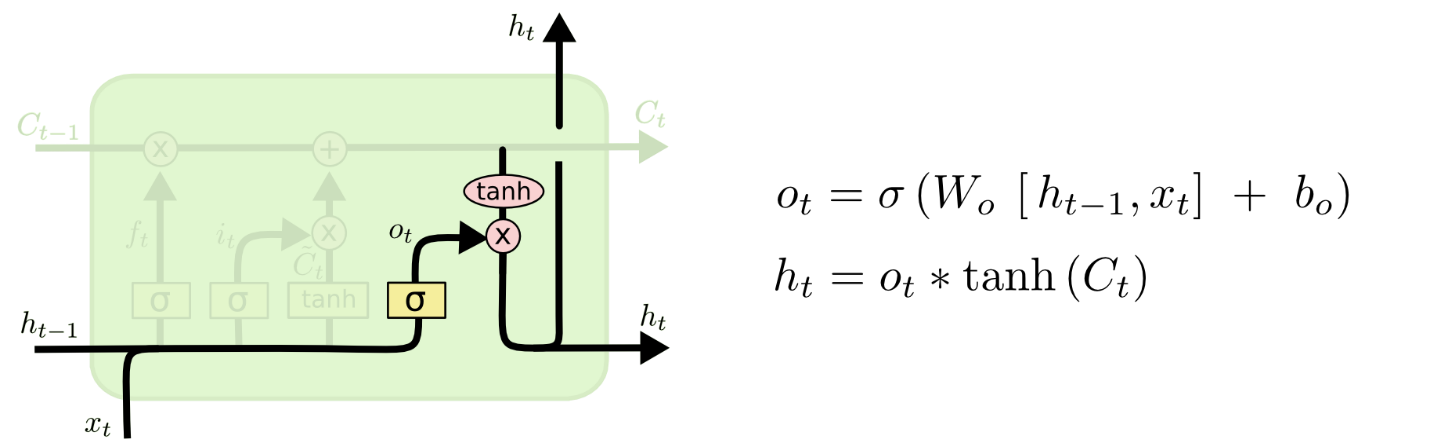
The second step decides what information we would require to store in the cell state. The “Sigmoid Layer” decides which values to update. A tanh layer creates a vector for the new candidate values C(t) in the new state.



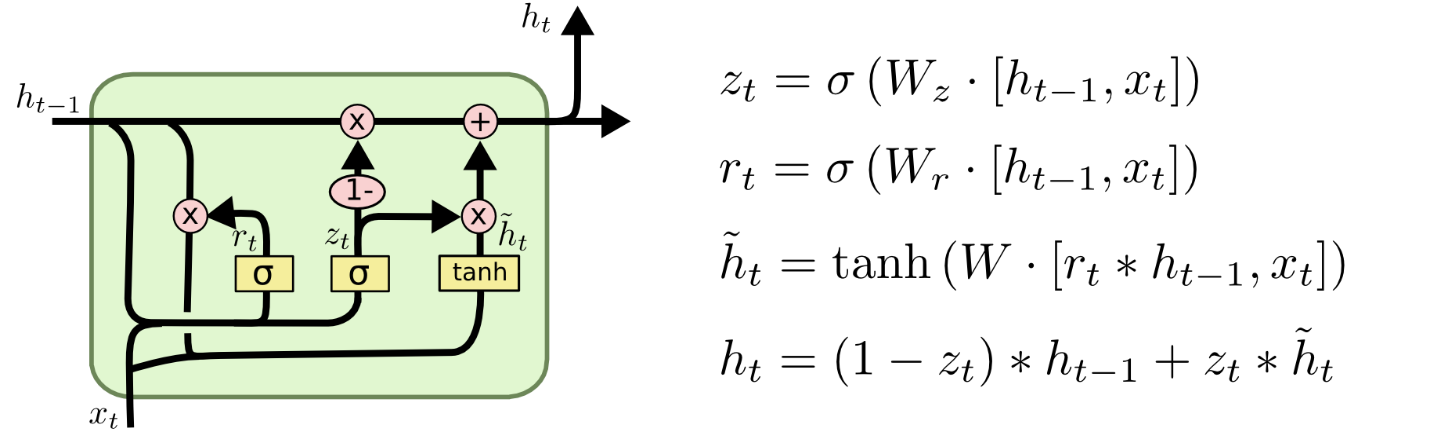
As the old cell state C(t -1) requires to perform the operation to provide new cell state of C(t) we just actually need to do it. By multiplying the old state with f(t) and then adding the i(t) with C(t) which we refer as the candidate value to update each state value.



In the position of deciding upon the update process, the decision is done based on the cell state. The decision also relies upon the factor of what we are going to update. The sigmoid layer decides what parts of the cell states its going to update hence output. Then the cell state is pushed through tanh (in order to push values between -1 and 1) and multiply it with the output of sigmoid gate so we can definitively output that we need to forecast.



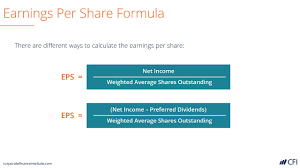
The complete internal state could be represented as such below:



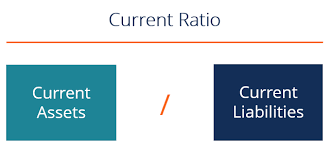
We will then formulate a portfolio based on the stocks which will have a growth from the perspective growth expectation depending on LSTM analysis process with least risk and most reward-based scenario.

After conducting the initial research on this method, we would go for further analysis to figure out the best possible stocks from the stocks of price clusters to evaluate on the basis of financial and statistical ratio analysis to check the best selection of stocks.

1. EPS Ratio: Earnings per share is one most important indicator that speaks within the context of development of any company. It says how much each share is earning in respect of any situational constraint. So, under general rule, higher value of EPS the better the company is performing and lower EPS shows a lower performance.



1. Current ratio: It basically shows how much current liabilities a company own in comparison to how much current assets the company possess under any given quarter. The higher the ratio of current assets divided by current liabilities the better the performance of the assets.



In our stock portfolio optimization process we will initiate stock clustering primarily based on Financial Ratio Analysis. In this Analysis process we will separate out stocks with positive values for both EPS and Current Ratio’s as benchmarking barriers for considering the stocks to be in portfolio. All the Stocks which will be considered for portfolio analysis would have a positive gain for the year of 2020 but our research paper would focus on reasonable analysis and suggestion for a group of stocks that will be considered for such processes.

|  |  |  |  |
| --- | --- | --- | --- |
| Testing Ratio’s | Accounting Formula for Ratios | Meaning of Result | Selection |
| EPS (Earnings Per Share) | Total Earnings / Outstanding Shares | Higher Earning indicates better financial condition for company | Higher Cluster |
| Current Ratio | Current Assets / Current Liabilities | Higher Current ratio Indicates better asset liabilities situation for the company. | Higher Cluster |

1. After selecting if we will go through two important aspects of stock.
2. We will collect the important details of 100 sample stocks which were winners for the period of 2020.
3. After making the collection we will perform K-Means Clustering on those two stocks based on the two available ratios. It’s true that due to serious problems of data unavailability we were unable to find data of other financial ratios for the stocks. (Christensen & Nikolaev, Springer 2013)
4. Only the positive financial ratios will be considered for good indicators as high EPS and better than 1 current ratio is a good indicator for a Stock.
5. Due to whole year pattern, we will initiate a two monthly stock portfolio optimization tests between negative ratio stocks and positive ratio stocks.
6. Also, the pricing methods would be tested through 6 step processes:
7. SVR for finding out future Prices.
8. ARIMA for finding out Future Prices for Stocks.
9. Linear Regression for Future Prices for Stocks.
10. LSTM by Single Variable Method for Future Prices.
11. LSTM by Multivariable Method for Future Prices.

Successive positive forecasts for all these steps would allow the stock to be a member of this portfolio and then we will make a mix of these stocks for consideration towards maximum Sharpe ratio value.

1. After finding out the positive values from the majority of the previously decided forecasted steps with higher accuracy tests, we will select assets for portfolio creation.

Steps in Portfolio Optimization Method:

1. Selection of assets using their positive indicators of financial ratios using K-Means Clustering for selection of assets for portfolio
2. Using the previous prices of investable assets, forecasted future prices for shorter intervals would be predicted both in terms of their prices.
3. The prices for stocks would be taken into consideration for a particularly longer time for training dataset.
4. The testing dataset will be used for each month as testing time frame for the year 2020.
5. The predicted values will then be tested with each actual values.
6. The highest accuracy among the selected assets will then be considered for predictions for January and February 2020 then undergone through same price prediction process. Only the gaining Assets would be then paired into the portfolio thus expecting a positive return
7. The Overall predicted values will then be tested through January & February asset prices in order to check if the stocks actually have made positive gain and if the Sharpe ratio of the portfolio is greater than 1.

Result Analysis:

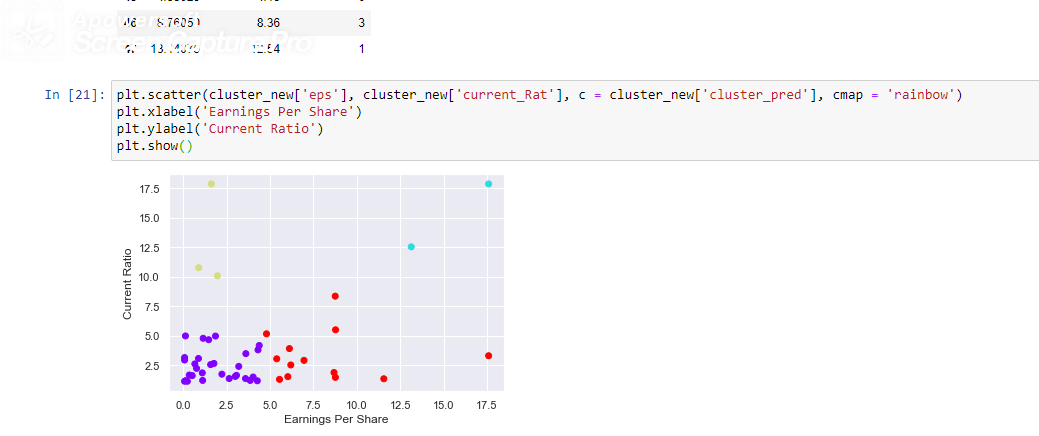
Our Results are completely based on the approaches modern Machine Learning applications have on the progresses in Analytical Finance. The primary approaches stared with asset classification based on financial ratio analysis. Due to unavailability of financial ratio, we have consolidated our ratio analysis process based on EPS and current ratio only. Our results have derived the following results.

Among the best performing stocks of 2020, there are variant stocks which do not possess both positive EPS and current ratio, higher than 2. We have used K Means Clustering method to significantly cluster the stocks in various different groups. Its known that if current ratios being higher than 2 then it means the companies may not be using their current assets or short-term financial facilities efficiently. In general, the higher the EPS, the better companies’ financial conditions are. On the basis of these constraints, we are formulating our clustering processes where the best clusters would suggest the highest EPS with medium current ratios.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Ticker | Beta | EPS | Current Ratio | Ticker | Beta | EPS | Current Ratio |
| ALRM | 1.15 | 1.47 | 4.67 | HD | 1.03 | 11.56 | 1.36 |
| JD | 0.92 | 0.15 | 1.2 | SNPS | 1.09 | 4.27 | 1.19 |
| LRCX | 1.34 | 17.6 | 3.31 | CELH | 2.17 | 0.08 | 3.16 |
| NTAP | 1.33 | 3.07 | 1.65 | NLS | 1.74 | 1.1 | 1.86 |
| NVDA | 1.45 | 6.12 | 3.92 | WYY | 1.4 | 0.24 | 1.15 |
| CRM | 1.17 | 3.86 | 1.22 | SHOP | 1.6 | 1.62 | 17.87 |
| SWKS | 1.21 | 4.8 | 5.17 | SPSC | 0.89 | 1.15 | 4.79 |
| SYNA | 1.17 | 3.2 | 2.41 | CTLT | 1.5 | 1.58 | 2.56 |
| TSM | 0.91 | 2.23 | 1.75 | CORT | 1.03 | 0.89 | 10.77 |
| MSFT | 0.83 | 6.2 | 2.53 | ALXN | 1.34 | 4.31 | 3.81 |
| FB | 1.18 | 8.78 | 5.51 | HZNP | 1.14 | 3.61 | 3.49 |
| TSLA | 2.19 | 0.52 | 1.63 | ACMR | 0.89 | 0.67 | 2.62 |
| ETSY | 1.65 | 1.86 | 4.98 | VNET | 0.44 | 0.77 | 2.24 |
| PYPL | 1.12 | 2.65 | 1.38 | KNSL | 0.82 | 3 | 1.56 |
| IPGP | 1.48 | 1.97 | 10.09 | BR | 0.84 | 4.03 | 1.51 |
| DE | 0.97 | 8.69 | 1.89 | OSTK | 4.41 | 0.35 | 1.68 |
| ROK | 1.37 | 8.77 | 1.48 | OIIM | 0.8 | 0.13 | 4.99 |
| ALB | 1.6 | 3.59 | 1.38 | LPTH | 1.45 | 0.079 | 2.95 |
| NKE | 0.84 | 1.77 | 2.66 | BRKS | 1.97 | 0.88 | 3.08 |
| CAT | 0.96 | 6.03 | 1.54 | PTC | 1.33 | 1.12 | 1.22 |
| ERIE | 0.42 | 5.55 | 1.31 | IRBT | 1.47 | 5.39 | 3.05 |
| ASML | 0.88 | 6.96 | 2.92 |  |  |  |  |

The best performing top 42 stocks which had a higher return during the period of 2020 with positive EPS and higher than 1 current ratio are considered being the primary dataset for data analysis.

Using K-Means Clustering Process we have figured out the best segment of performing stocks which could be used as suggestions for portfolio optimization processes used in Risk minimizing with return maximization option.



Using our portfolio analysis processes, its determined that for a specific period very big group of stocks could be good performer. In light of the scenarios for 2020, COVID 19 pandemic the financial markets have seen a serious turmoil and there were losses in opportunities in the market.

Our analysis process has surfaced various time periods for analysis, including each two months period for year 2020 as the benchmarks for price predictions and comparing them with real portfolio suggestions and price directions and analysis combinations using various Price prediction techniques.

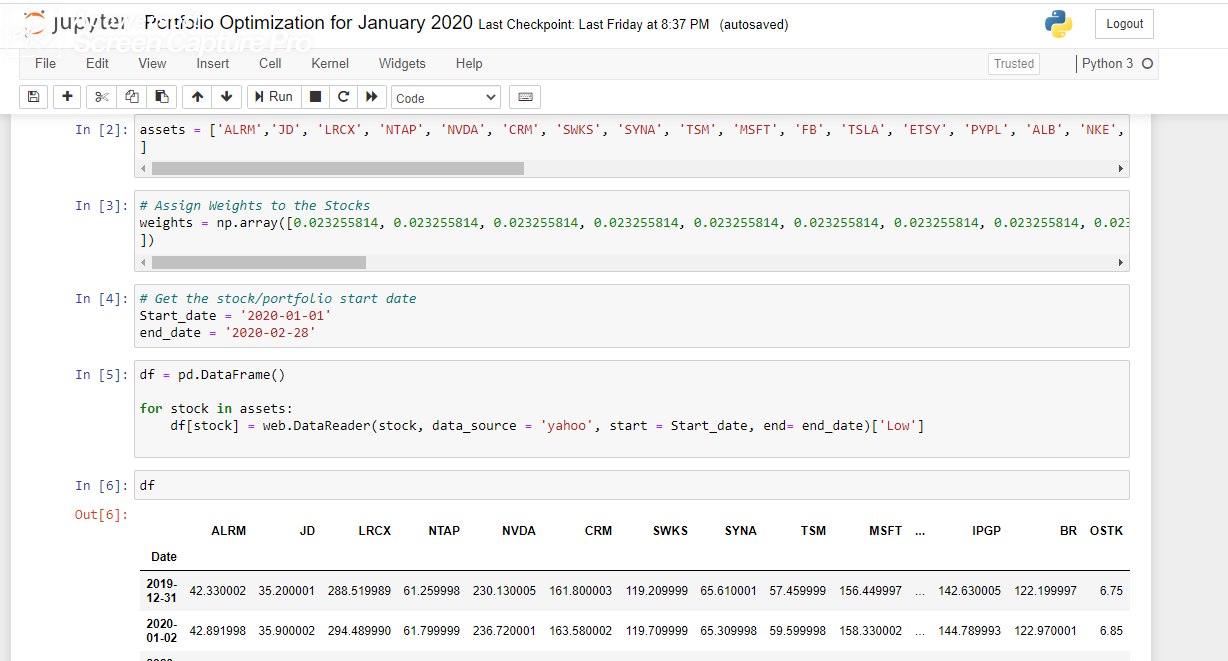
The consideration for such study has a very positive impact as 2020 is a benchmark year for understanding financial market in relation to time series analysis processes.

**Portfolio Analysis for Months of January & February 2020**

Due to severe financial pressure and scarcity in the worlds changing processes of management in supply chain efficiencies, and inefficiency arise due to COVID 19 pandemic. The hardest blow in the market was seen during the period of January and February 2020. In light of all these processes the market was capable of making a positive return with positive beta but perform well amidst financial turmoil.

In our analysis process we have considered all the best stocks with financial suitability in terms of positive EPS and current ratio as the selective measures for their performances.

In portfolio analysis process we have taken their prices for the period of beginning of January 2020 till end of February 2020 from the famous source of yahoo finance. By assigning equal weights for these stocks the portfolio analysis process is giving equal importance to all of these stocks which becomes beneficial to understand their effectiveness in relation to their contribution in a successful portfolio.



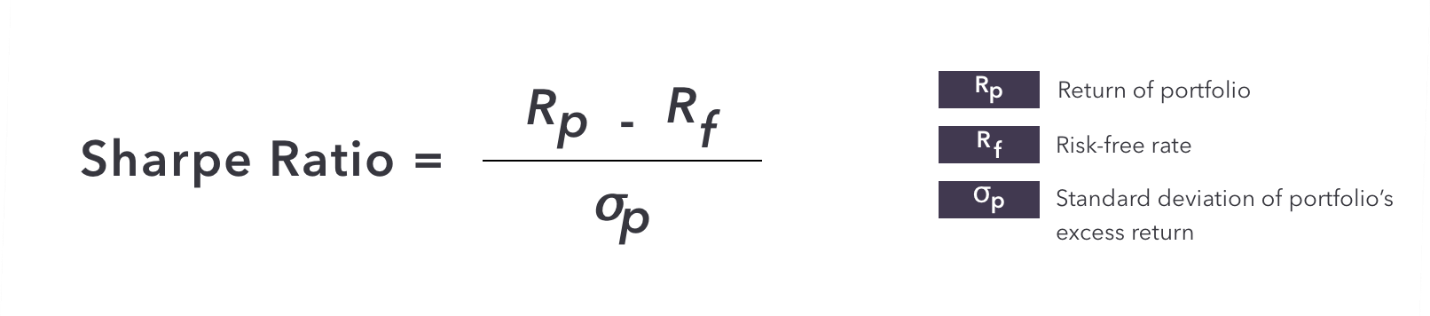
Using the measure of portfolio optimization as primary benchmark using efficient frontier with Sharpe ratio and portfolio performance, we have maximized portfolio performances.

**Analysis for Portfolio Optimization Ratios and their values under Measurement:**

**Sharpe Ratio:**

The Sharpe Ratio is a financial metric often used by investors when assessing the performance of investment management products and professionals. It consists of taking the excess return of the portfolio, relative to the risk-free rate, and dividing it by the standard deviation of the portfolio’s excess returns.

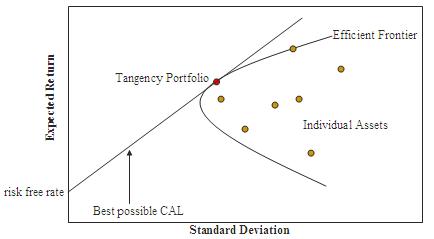
A good Sharpe ratio will have values greater than or equal to 1.00 which is generally considered “good”, as this would suggest that the portfolio is offering excess returns relative to its volatility. Investors will often compare the Sharpe Ratio of a portfolio relative to its peers. A portfolio with a Sharpe Ratio of 1.00 might be considered inadequate if the competitors in its peer group have an average Sharpe Ratio above 1.00.



**Efficient Frontier**

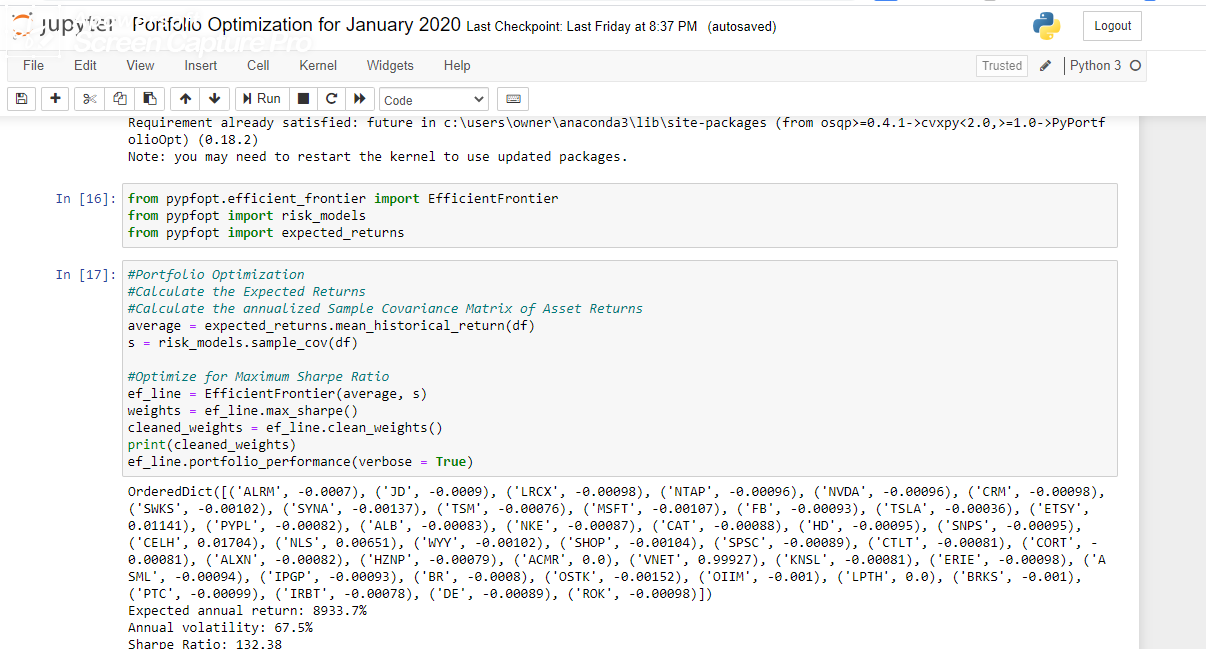
The efficient frontier is the set of optimal portfolios that offer the highest expected return for a defined level of risk or the lowest risk for a given level of expected return. Portfolios that lie below the efficient frontier are sub-optimal because they do not provide enough return for the level of risk. Portfolios that cluster to the right of the efficient frontier are sub-optimal because they have a higher level of risk for the defined rate of return.

Efficient frontier comprises investment portfolios that offer the highest expected return for a specific level of risk. Returns are dependent on the investment combinations that make up the portfolio. The standard deviation of a security is synonymous with risk. Lower covariance between portfolio securities results in lower portfolio standard deviation. Successful optimization of the return versus risk paradigm should place a portfolio along the efficient frontier line. Optimal portfolios that comprise the efficient frontier tend to have a higher degree of diversification.



**Result Analysis for Months of January & February 2020**

Portfolio Optimization method have suggested which stocks are the best performers for the period of first two months of the year 2020. Its seen that due to serious turmoil of finance industry very small number of stocks have performed extensively well under drastic financial conditions. There were various other stocks which have actually plummeted their values and became negative inhibitors for the portfolio optimization method.

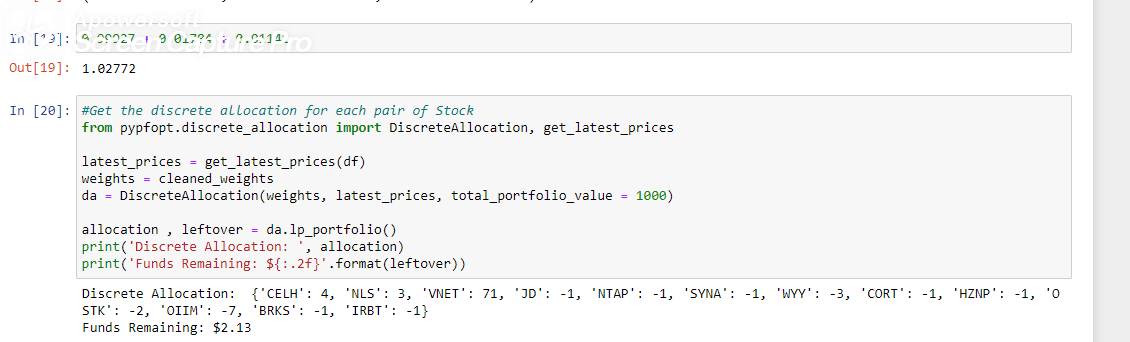


**Creating Suggestive Portfolio: Based on Successful Stock Growth Prediction**

In accordance to the analysis drawn from the above portfolio optimization process its seen that most stocks due to financial market and financial conditions have underperformed and made negative contributions to the portfolio. It makes the portfolio inefficient, hence its highly suggestive for the investors to know which stocks they require to make an investment during this period that could make suggestive positive ROI and would result towards positive growth.

There have been various time series prediction methods which were used till now by stock markets like ARIMA (Pai et al, 2005 and Ariyo et al, 2014), SVR (Guo et al 2018, Tsai et al 2009, Zhang et al 2019) and LSTM (Jiang et al 2018, Cao et Li 2019, Kim et Won 2018).

Primarily the portfolio optimization processes have figured out the high performing assets during a specific time frame but at the same time the underperforming and negatively impacting assets are portrayed through the portfolio optimization processes as well.

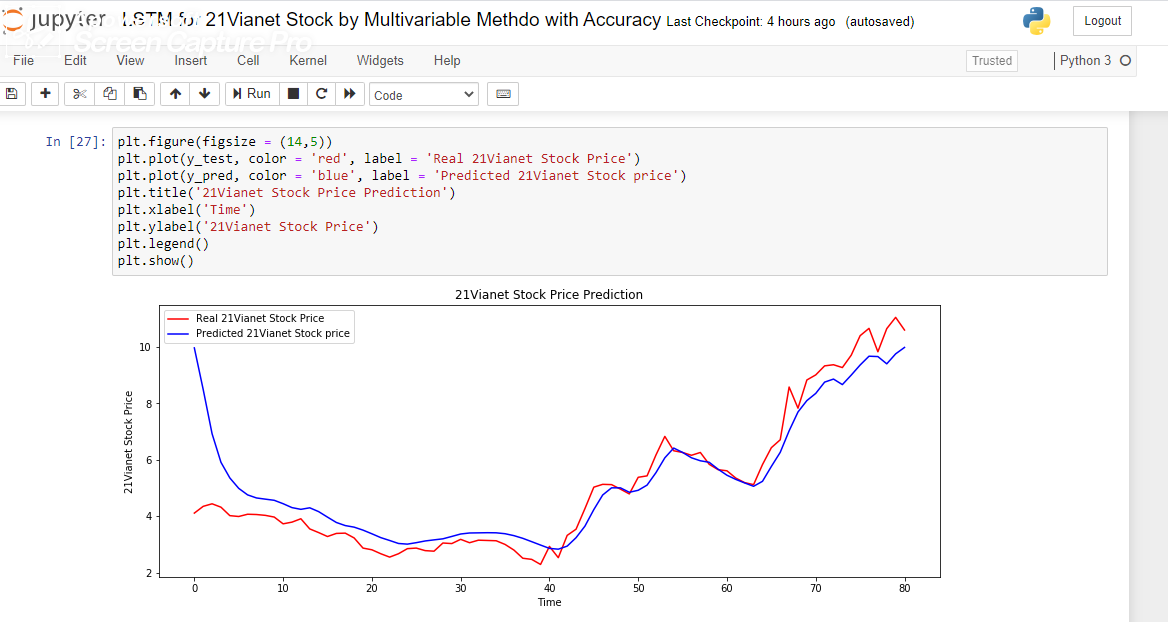


In accordance to the above results for the months of January and February 2020 we can observe that three different stocks were making positive returns using the Stock ticker as CELH, NLS and VNET. These three stocks made affirmative gains with while having positive beta but performed better than the market.

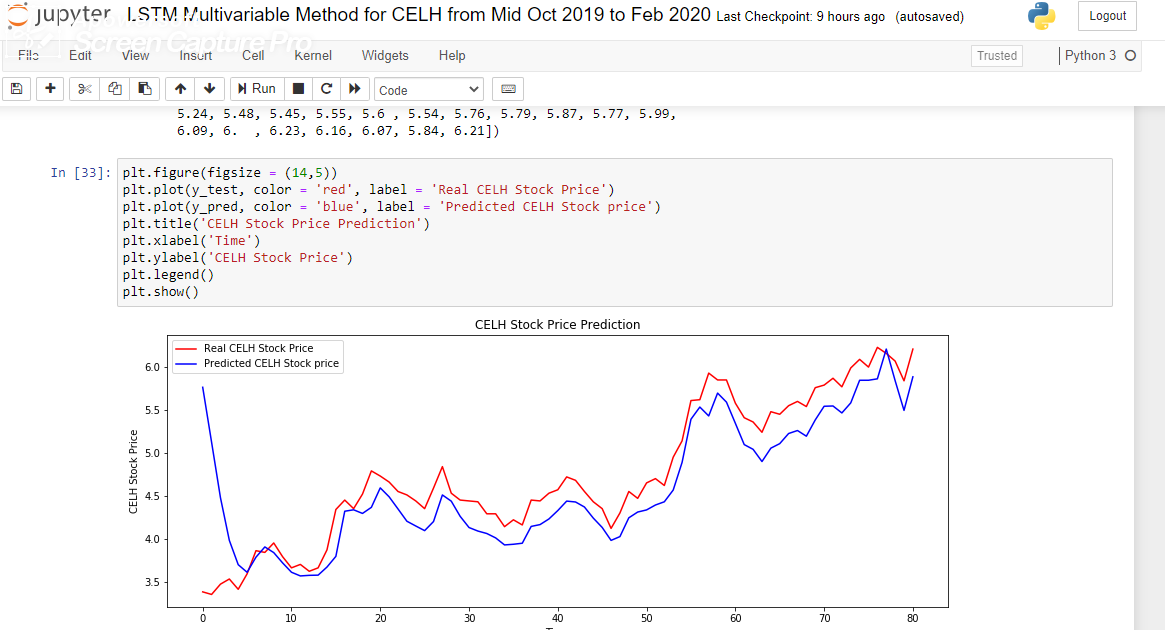
As the portfolio suggestions has been carried out it said that from the beginning till the end of period for October 2019 to February 2020, there has been a positive gain in the stock.

Through LSTM multivariable process of price prediction, the price has actually risen for this stock. The same has happened with the other two stocks.

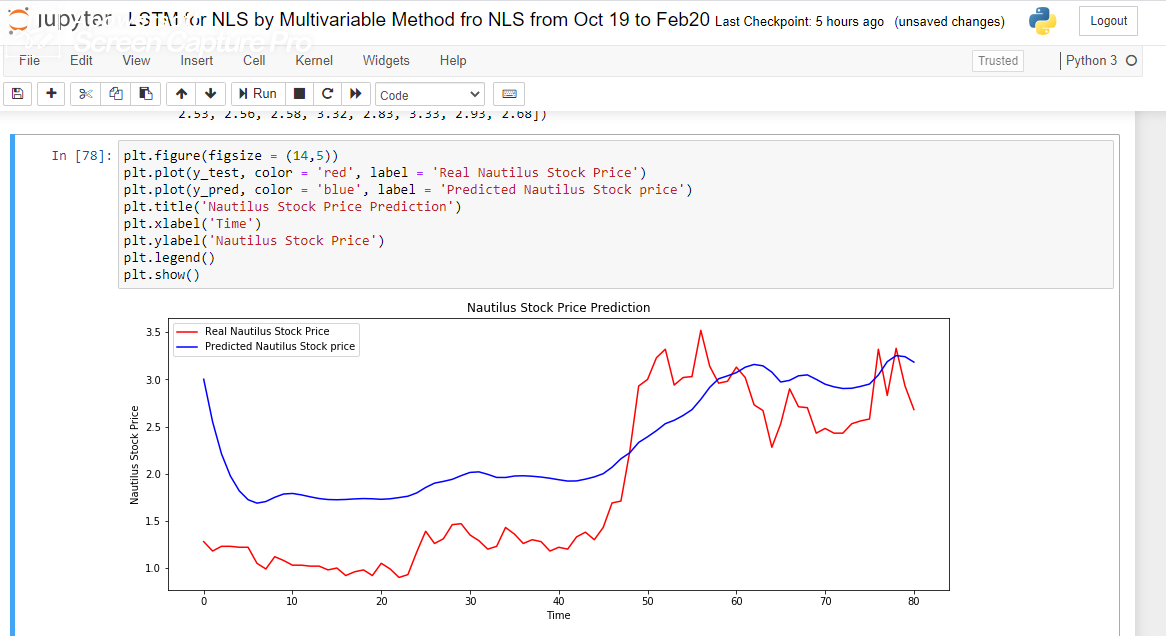
The postulated suggestion could be if there are favorable financial ratios for any stock then the stock should have better future gain in terms of price growth. Among all the 42 selected stocks all of them made subsequent recovery but these three stocks were completely COVID 19 pandemic proof in relation to market behavior.



|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **21Vianet** | | | | | | | | |
| **Predicted Price** | **Actual price** | **Accuracy** | **Predicted Price** | **Actual Price** | **Accuracy** | **Predicted Price** | **Actual Price** | **Accuracy** |
| 3.3770058 | 2.81 | 83.20981 |  |  |  |  |  |  |
| 3.2826242 | 3.03 | 92.3042 | 4.8483887 | 4.96 | 97.74977 | 8.099062 | 7.83 | 96.67786 |
| 3.3693428 | 3.18 | 94.38042 | 4.9190803 | 4.79 | 97.37593 | 8.353848 | 8.83 | 94.60757 |
| 3.405203 | 3.06 | 89.86248 | 5.102983 | 5.38 | 94.85099 | 8.754055 | 9.01 | 97.15932 |
| 3.4100487 | 3.15 | 92.37405 | 5.543754 | 5.43 | 97.94807 | 8.862083 | 9.33 | 94.98481 |
| 3.415363 | 3.14 | 91.93752 | 6.0698366 | 6.17 | 98.37661 | 8.666072 | 9.37 | 92.48743 |
| 3.4116347 | 3.13 | 91.74488 | 6.4161916 | 6.83 | 93.94131 | 9.003292 | 9.27 | 97.12289 |
| 3.379453 | 3 | 88.77176 | 6.261215 | 6.32 | 99.06986 | 9.360525 | 9.71 | 96.40088 |
| 3.311489 | 3.11 | 93.91546 | 6.0714707 | 6.26 | 96.98835 | 9.67264 | 10.4 | 93.00615 |
| 3.216187 | 2.8 | 87.05961 | 5.9651318 | 6.16 | 96.83656 | 9.659756 | 10.66 | 90.61685 |
| 3.097655 | 2.51 | 81.02904 | 5.91167 | 6.26 | 94.43562 | 9.401974 | 9.83 | 95.64572 |
| 2.977463 | 2.47 | 82.95653 | 5.6668825 | 5.84 | 97.03566 | 9.755708 | 10.65 | 91.60289 |
| 2.859509 | 2.29 | 80.08368 | 5.452518 | 5.65 | 96.50474 | 9.983758 | 11.05 | 90.35075 |
| 2.8308342 | 2.93 | 96.6155 | 5.304506 | 5.61 | 94.55447 | 5.304506 | 5.61 | 94.55447 |
| 2.9396086 | 2.53 | 86.06588 | 6.2647142 | 6.43 | 97.42946 | 5.1787224 | 5.35 | 96.79855 |
| 3.2383237 | 3.32 | 97.53987 | 7.0255404 | 6.71 | 95.50867 | 5.0605206 | 5.19 | 97.50521 |
| 3.6480381 | 3.54 | 97.03846 | 7.695 | 8.58 | 89.68531 | 5.236874 | 5.12 | 97.76825 |
| 4.2341413 | 4.27 | 99.16022 | 5.0061917 | 5.13 | 97.58658 | 5.764114 | 5.83 | 98.86988 |
| 4.752312 | 5.03 | 94.47936 | 5.00575 | 5.12 | 97.76855 | **Average Accuracy** |  | **94.0967** |

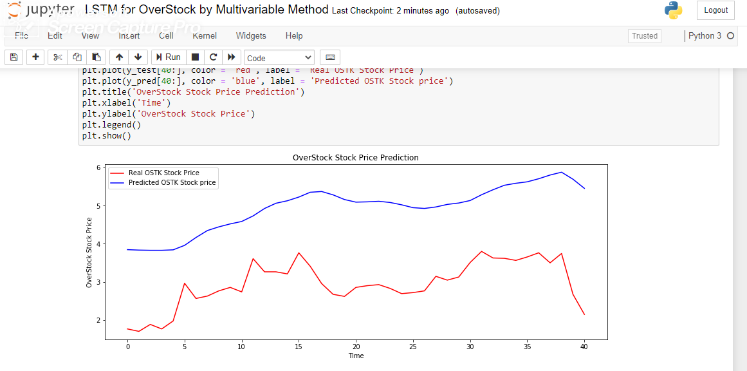


|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **CELH** | | | | | | | | |
| **Predicted Price** | **Actual Price** | **Accuracy** | **Predicted Price** | **Actual Price** | **Accuracy** | **Predicted Price** | **Actual Price** | **Accuracy** |
|  |  |  |  |  |  |  |  |  |
| 4.1299167 | 4.44 | 93.01614 | 4.24327 | 4.55 | 93.25868 | 5.260523 | 5.6 | 93.93791 |
| 4.0888295 | 4.43 | 92.29863 | 4.310508 | 4.47 | 96.43195 | 5.1943836 | 5.54 | 93.76144 |
| 4.060378 | 4.29 | 94.64751 | 4.335206 | 4.65 | 93.23024 | 5.383974 | 5.76 | 93.47177 |
| 4.0090623 | 4.29 | 93.45134 | 4.3924823 | 4.7 | 93.45707 | 5.54338 | 5.79 | 95.74059 |
| 3.9271548 | 4.14 | 94.85881 | 4.4299818 | 4.62 | 95.88705 | 5.5466967 | 5.87 | 94.49228 |
| 3.9349082 | 4.22 | 93.24427 | 4.5667663 | 4.95 | 92.25791 | 5.4654946 | 5.77 | 94.72261 |
| 3.9469433 | 4.16 | 94.87844 | 4.8891325 | 5.14 | 95.11931 | 5.5826225 | 5.99 | 93.19904 |
| 4.1428976 | 4.45 | 93.09882 | 5.392091 | 5.61 | 96.1157 | 5.8463435 | 6.09 | 95.99907 |
| 4.164159 | 4.44 | 93.78736 | 5.534269 | 5.62 | 98.47454 | 5.8465533 | 6 | 97.44256 |
| 4.2298565 | 4.53 | 93.37432 | 5.4305477 | 5.93 | 91.57753 | 5.863022 | 6.23 | 94.1095 |
| 4.3278255 | 4.57 | 94.70078 | 5.6969285 | 5.85 | 97.38339 | 6.2100725 | 6.16 | 99.19369 |
| 4.43798 | 4.72 | 94.025 | 5.591559 | 5.85 | 95.58221 | 5.8417387 | 6.07 | 96.23952 |
| 4.427628 | 4.68 | 94.60744 | 5.345384 | 5.58 | 95.79541 | 5.495076 | 5.84 | 94.09377 |
| 4.3692636 | 4.55 | 96.02777 | 5.09658 | 5.41 | 94.20665 | 5.8867307 | 6.21 | 94.79438 |
| 4.239617 | 4.43 | 95.70242 | 5.041468 | 5.36 | 94.05724 | 5.1077967 | 5.45 | 93.72104 |
| 4.1308975 | 4.35 | 94.96316 | 4.8990464 | 5.24 | 93.49325 | 5.2257795 | 5.55 | 94.15819 |
| 3.9799826 | 4.12 | 96.60152 | 5.0550885 | 5.48 | 92.24614 | 4.0243025 | 4.3 | 93.58843 |
| **Average Accuracy** | **94.31080747** | | | | | | | |

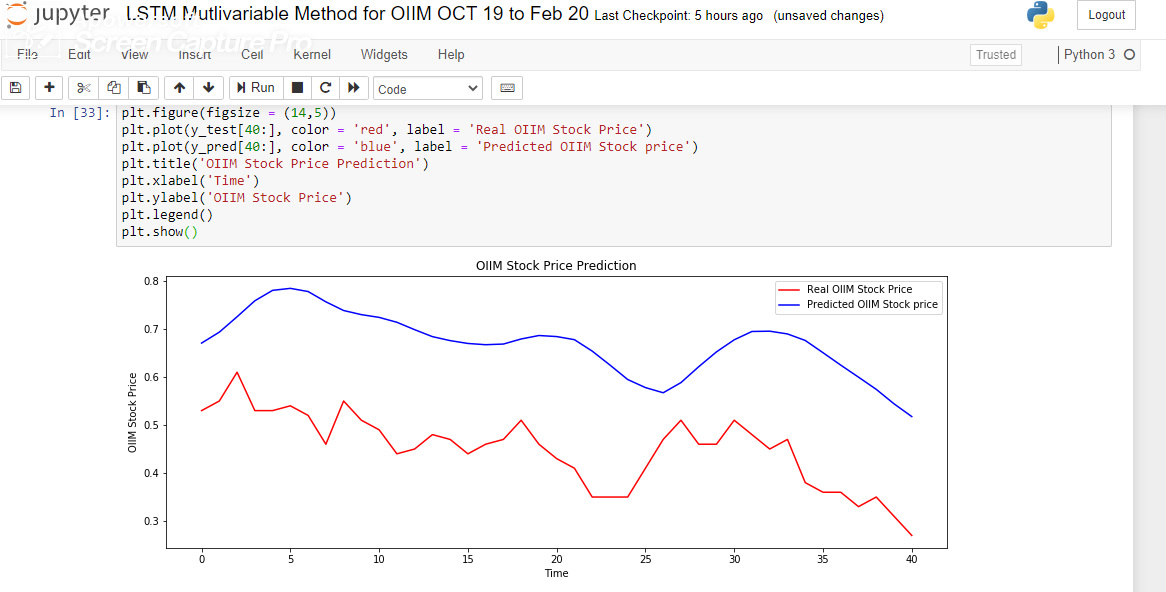


|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Nautilus Stock Price Predicting Using LSTM Multivariable Method** | | | | | | | | |
| **Predicted Price** | **Actual Price** | **Accuracy** | **Predicted Price** | **Actual Price** | **Accuracy** | **Predicted Price** | **Actual Price** | **Accuracy** |
|  |  |  |  |  |  |  |  |  |
| 1.9360833 | 1.22 | 63.01382 | 2.9156814 | 3.14 | 92.8561 | 2.9257305 | 2.56 | 87.49952 |
| 1.9220078 | 1.2 | 62.43471 | 3.006527 | 2.96 | 98.45247 | 2.9507096 | 2.58 | 87.43659 |
| 1.9232605 | 1.33 | 69.1534 | 3.0374634 | 2.98 | 98.10818 | 3.0468724 | 3.32 | 91.77327 |
| 1.9422415 | 1.38 | 71.05193 | 3.071845 | 3.13 | 98.14201 | 3.1879768 | 2.83 | 88.77103 |
| 1.9660524 | 1.3 | 66.12235 | 3.129417 | 3.02 | 96.5036 | 3.2521913 | 3.33 | 97.6634 |
| 2.000238 | 1.43 | 71.49149 | 3.1585543 | 2.73 | 86.43195 | 3.240787 | 2.93 | 90.41014 |
| 2.0687673 | 1.69 | 81.69116 | 3.1436436 | 2.67 | 84.93329 | 3.1832132 | 2.68 | 84.19166 |
| 2.1586077 | 1.71 | 79.21773 | 3.0768619 | 2.28 | 74.10147 | 3.0005505 | 2.43 | 80.98514 |
| 2.2228365 | 2.24 | 99.23377 | 2.972117 | 2.53 | 85.12451 | 2.9490466 | 2.48 | 84.09497 |
| 2.3317323 | 2.93 | 79.58131 | 2.9898777 | 2.9 | 96.99393 | 2.9209886 | 2.43 | 83.19101 |
| 2.3913903 | 3 | 79.71301 | 3.0379608 | 2.71 | 89.20457 | 2.9033823 | 2.43 | 83.69549 |
| 2.456459 | 3.23 | 76.05136 | 3.047829 | 2.7 | 88.58765 | 2.9059503 | 2.53 | 87.06274 |
| 2.5304282 | 3.32 | 76.21772 | 2.617141 | 3.02 | 86.6603 | 2.7891064 | 3.52 | 79.23598 |
| 2.566548 | 2.94 | 87.29755 | 2.6791675 | 3.03 | 88.42137 | **Average Accuracy** |  | **84.21472** |

After making the analysis for all the winning stocks during the period for 2020 and with getting the best accuracy rates we would further move into the scenario for stocks that made sharp decline for the period of January and February 2020. Its seen in the portfolio analysis that the stocks with ticker named as OSTK, WYY and OIIM made the sharpest possible decrease during this period.

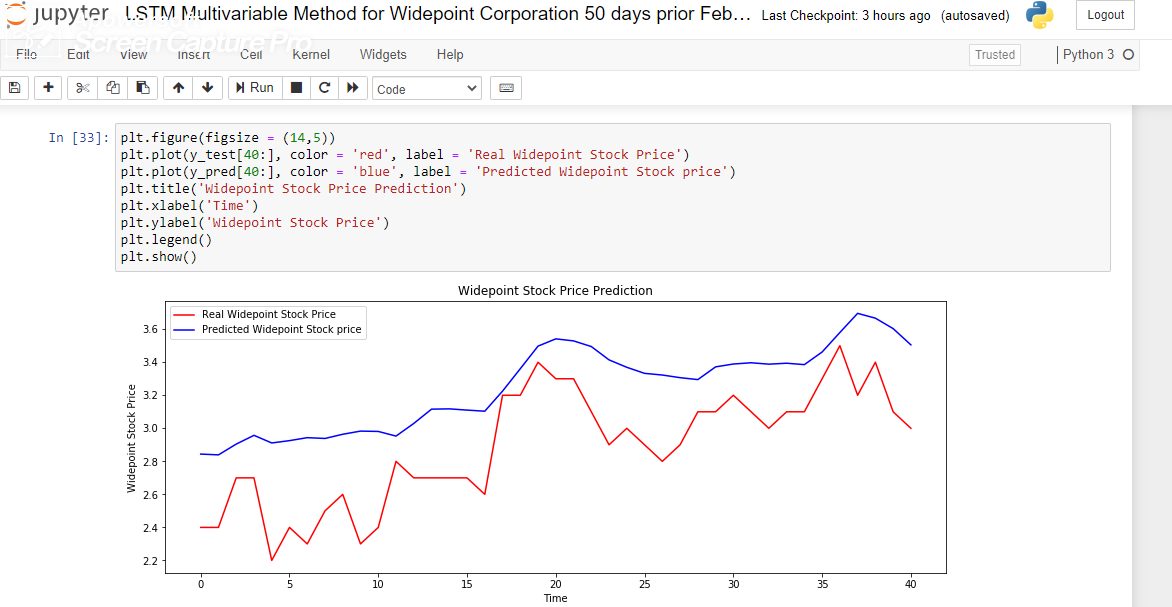


|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Overstock** | | | | | |
| **Predicted Price** | **Actual Price** | **Accuracy** | **Predicted Price** | **Actual Price** | **Accuracy** |
| 3.8449616 | 1.76322528 | 45.85807 | 5.092633 | 2.85388009 | 56.03938 |
| 3.833347 | 1.69960375 | 44.33733 | 5.1001973 | 2.89932404 | 56.84729 |
| 3.8276768 | 1.88137955 | 49.152 | 5.1147633 | 2.92659041 | 57.21849 |
| 3.8275974 | 1.76322528 | 46.06611 | 5.083392 | 2.82661372 | 55.60487 |
| 3.8401976 | 1.97226745 | 51.35849 | 5.022731 | 2.69028187 | 53.56213 |
| 3.957552 | 2.96294557 | 74.86814 | 4.947522 | 2.71754824 | 54.92746 |
| 4.163969 | 2.56303881 | 61.55278 | 4.9271436 | 2.76299219 | 56.07696 |
| 4.3472285 | 2.62666034 | 60.42149 | 4.9659004 | 3.14472137 | 63.32631 |
| 4.443476 | 2.76299219 | 62.18087 | 5.0330677 | 3.04474468 | 60.49481 |
| 4.5213203 | 2.85388009 | 63.1205 | 5.0692053 | 3.12654379 | 61.6772 |
| 4.584711 | 2.73572582 | 59.67063 | 5.1350365 | 3.50827297 | 68.32031 |
| 4.7308054 | 3.60824967 | 76.27136 | 5.286693 | 3.79911426 | 71.86183 |
| 4.925879 | 3.26287564 | 66.23946 | 5.417931 | 3.62642725 | 66.9338 |
| 5.0633535 | 3.26287564 | 64.441 | 5.535875 | 3.61733846 | 65.34357 |
| 5.1288524 | 3.2083429 | 62.55479 | 5.5876274 | 3.56280571 | 63.76241 |
| 5.2256794 | 3.7627591 | 72.00517 | 5.626335 | 3.65369362 | 64.93914 |
| 5.3520737 | 3.40829628 | 63.68179 | 5.7071 | 3.7627591 | 65.93119 |
| 5.3719826 | 2.95385678 | 54.98634 | 5.804484 | 3.49918418 | 60.28416 |
| 5.2836432 | 2.67210429 | 50.57314 | 5.8785872 | 3.74458152 | 63.69866 |
| 5.1611395 | 2.61757155 | 50.71693 | 5.6950064 | 2.67210429 | 46.92013 |
| 5.454542 | 2.14495446 | 39.32419 | **Average Accuracy** |  | **59.34514** |



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **OIIM** | | | | | |
| **Predicted Accuracy** | **ActuaL Price** | **Accuracy** | **Predicted Price** | **Actual Price** | **Accuracy** |
| 0.67064035 | 0.53 | 79.02895 | 0.6791988 | 0.51 | 75.08847 |
| 0.69346064 | 0.55 | 79.31236 | 0.6864075 | 0.46 | 67.01558 |
| 0.72551394 | 0.61 | 84.07833 | 0.68421656 | 0.43 | 62.8456 |
| 0.75875527 | 0.53 | 69.85124 | 0.6777497 | 0.41 | 60.49431 |
| 0.7804857 | 0.53 | 67.90643 | 0.6539447 | 0.35 | 53.52135 |
| 0.7846321 | 0.54 | 68.82206 | 0.6249379 | 0.35 | 56.00556 |
| 0.777999 | 0.52 | 66.83813 | 0.59475964 | 0.35 | 58.8473 |
| 0.75661385 | 0.46 | 60.7972 | 0.577944 | 0.41 | 70.94113 |
| 0.73848325 | 0.55 | 74.47698 | 0.5671843 | 0.47 | 82.86548 |
| 0.7298988 | 0.51 | 69.8727 | 0.5883321 | 0.51 | 86.68573 |
| 0.72415787 | 0.49 | 67.6648 | 0.62142277 | 0.46 | 74.02368 |
| 0.71399695 | 0.44 | 61.62491 | 0.65231115 | 0.46 | 70.51849 |
| 0.69857496 | 0.45 | 64.41685 | 0.6774806 | 0.51 | 75.27891 |
| 0.684086 | 0.48 | 70.16662 | 0.6948192 | 0.48 | 69.08272 |
| 0.67579925 | 0.47 | 69.54728 | 0.6954122 | 0.45 | 64.70982 |
| 0.6698471 | 0.44 | 65.68663 | 0.6897018 | 0.47 | 68.14539 |
| 0.6672578 | 0.46 | 68.93887 | 0.67611134 | 0.38 | 56.20376 |
| 0.6686415 | 0.47 | 70.29178 | 0.65064245 | 0.36 | 55.32993 |
| 0.624751 | 0.36 | 57.62296 | 0.54419655 | 0.31 | 56.96471 |
| 0.5996928 | 0.33 | 55.02817 | 0.51757133 | 0.27 | 52.16672 |
| 0.57426834 | 0.35 | 60.94712 | **Average Accuracy** |  | **67.06476** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Widepoint Corporation** | | | | | |
| **Predicted Price** | **Actual Price** | **Accuracy** | **Predicted Price** | **Actual Price** | **Accuracy** |
| 2.8438988 | 2.4 | 84.39119 | 3.529086 | 3.3 | 93.50863 |
| 2.8391263 | 2.4 | 84.53305 | 3.494904 | 3.1 | 88.70058 |
| 2.9043179 | 2.7 | 92.96503 | 3.4141414 | 2.9 | 84.94083 |
| 2.9574573 | 2.7 | 91.29464 | 3.3692527 | 3 | 89.04052 |
| 2.9111953 | 2.2 | 75.57033 | 3.33261 | 2.9 | 87.01888 |
| 2.9252968 | 2.4 | 82.04296 | 3.322507 | 2.8 | 84.27371 |
| 2.943291 | 2.3 | 78.14382 | 3.306518 | 2.9 | 87.70556 |
| 2.9384644 | 2.5 | 85.07845 | 3.2948472 | 3.1 | 94.08631 |
| 2.9638891 | 2.6 | 87.72258 | 3.3718665 | 3.1 | 91.93721 |
| 2.9829998 | 2.3 | 77.10359 | 3.3891406 | 3.2 | 94.41922 |
| 2.981001 | 2.4 | 80.50987 | 3.396749 | 3.1 | 91.26373 |
| 2.9528744 | 2.8 | 94.82286 | 3.3882606 | 3 | 88.54101 |
| 3.0293639 | 2.7 | 89.12762 | 3.3939714 | 3.1 | 91.33842 |
| 3.115881 | 2.7 | 86.65286 | 3.3854265 | 3.1 | 91.56896 |
| 3.11765 | 2.7 | 86.60369 | 3.4618912 | 3.3 | 95.32362 |
| 3.1102133 | 2.7 | 86.81077 | 3.5798528 | 3.5 | 97.76938 |
| 3.1034334 | 2.6 | 83.77818 | 3.695852 | 3.2 | 86.58355 |
| 3.2249498 | 3.2 | 99.22635 | 3.6666822 | 3.4 | 92.72688 |
| 3.360772 | 3.2 | 95.21622 | 3.6035721 | 3.1 | 86.02575 |
| 3.4976144 | 3.4 | 97.20911 | 3.5050693 | 3 | 85.59032 |
| 3.541482 | 3.3 | 93.18133 | **Average Accuracy** |  | **88.64262** |



Its seen that in terms predictions the accuracy wasn’t very accurate as the results didn’t make a very accurate predictions like it did for the winning stocks. But the Propositional concept that if a stock would go up or not is very clearly determined in here.

**Reasons Behind the choice for LSTM Multivariable Method:**

Time series Analysis process is a wide range of prediction method that has existed since long time in the name of Linear regression and has got much more widespread application with the development of better computing methods. SVR for Stock Price Prediction (Henrique et al, 2018 and Trafalis & Ince, 2008) has shown how SVM has been used as an important measure to figure out Stock Prices for a shorter interval and for one specific period.

ARIMA (Auto Regressive Integrated Moving Average) (Pai et al, 2005 and Ariyo et al 2014) has been a used as an important measure for predicting Stock Prices as a specialized Linear Regression model for Stock Price.

In order to determine the best possible method for determining the very method that needs to be used for Stock Price prediction and on the basis of which we will actually perform our portfolio optimization function, we have made a sample test process.

The sample Stock that has been taken in account in NVDA (NYSE) ticker for the very famous company. The stock is also selected as one the best stocks among the 42 other stocks and have huge degrees of market capitalization and growth opportunity.

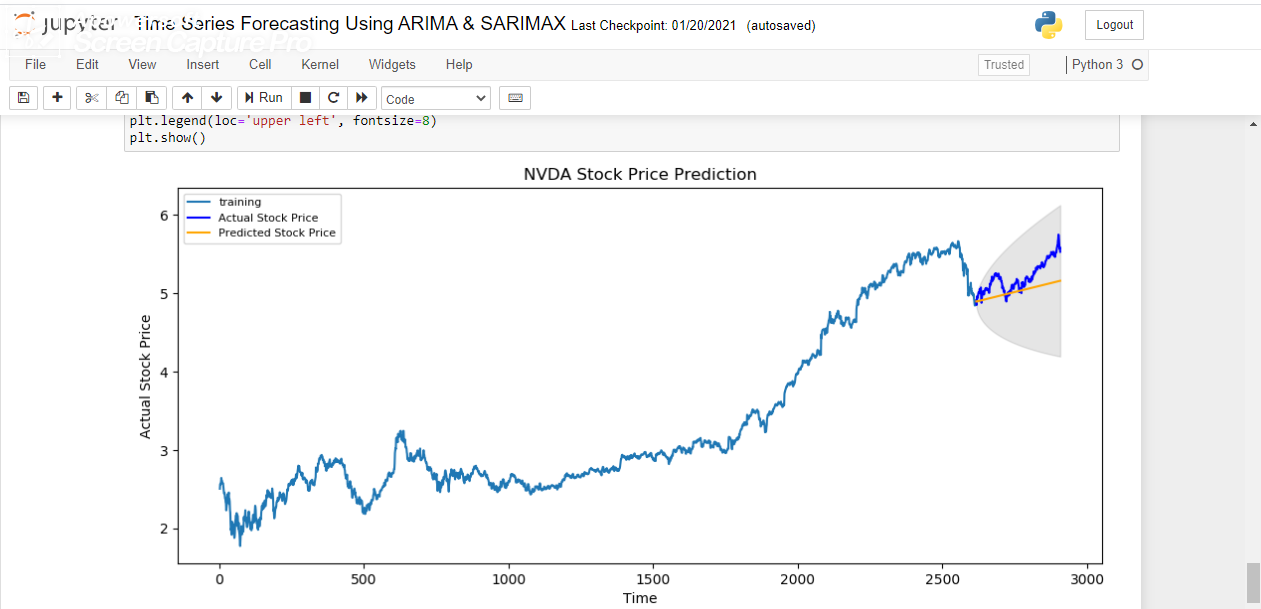
Price prediction tests in relation to accuracy has been performed on NVDA. The results are analyzed in the following:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **SVM** |  |  | **Linear** |  |  |
| **Predicted** | **Actual Price** | **Difference** | **Predicted** | **Actual Price** | **Difference** |
| 189.2019 | 230.899994 | 58.301932 | 206.0595 | 230.899994 | 75.1595449 |
| 206.7333 | 238.75 | 67.983316 | 214.1511 | 238.75 | 75.4011363 |
| 234.9479 | 235.100006 | 99.847944 | 209.8603 | 235.100006 | 74.7602791 |
| 195.6016 | 232.320007 | 63.281601 | 213.0509 | 232.320007 | 80.7309109 |
| 205.6176 | 238.199997 | 67.417583 | 212.0507 | 238.199997 | 73.8507232 |
| 196.1086 | 239.759995 | 56.348615 | 212.7609 | 239.759995 | 73.0008765 |
| 248.0138 | 243.850006 | 95.836213 | 223.0829 | 243.850006 | 79.2329205 |
| 223.5588 | 247.330002 | 76.228835 | 218.8721 | 247.330002 | 71.5420762 |
| 235.1051 | 247.660004 | 87.445073 | 220.2734 | 247.660004 | 72.6133964 |
| 223.1944 | 250.259995 | 72.934431 | 218.772 | 250.259995 | 68.5120003 |
| 223.6901 | 247.789993 | 75.90008 | 211.275 | 247.789993 | 63.4849658 |
| 232.5484 | 247.720001 | 84.828364 | 209.6835 | 247.720001 | 61.9634888 |
| 219.682 | 249.699997 | 69.982012 | 211.4451 | 249.699997 | 61.745121 |
| 233.4333 | 247.809998 | 85.623272 | 210.7645 | 247.809998 | 62.9544853 |
| 207.3817 | 249.759995 | 57.621675 | 214.1977 | 249.759995 | 64.4377109 |
| 207.3817 | 251.720001 | 55.661669 | 214.1977 | 251.720001 | 62.4777049 |
| 219.2235 | 257.5 | 61.723497 | 216.0094 | 257.5 | 58.5094056 |
| 226.1677 | 238.240005 | 87.927717 | 219.3726 | 238.240005 | 81.1325459 |
| 253.4288 | 242.850006 | 89.421157 | 226.1089 | 242.850006 | 83.2588695 |
| 253.0556 | 247.389999 | 94.334386 | 226.0288 | 247.389999 | 78.6387833 |
| 254.0071 | 241.610001 | 87.602938 | 227.1899 | 241.610001 | 85.579854 |
| 234.7044 | 243.970001 | 90.734367 | 230.3328 | 243.970001 | 86.3627846 |
| 235.6733 | 235.699997 | 99.973335 | 231.5539 | 235.699997 | 95.8539346 |
| 253.1275 | 246.029999 | 92.902451 | 237.5095 | 246.029999 | 91.4795255 |
| 225.1013 | 251.990005 | 73.111318 | 241.4232 | 251.990005 | 89.43318 |
| 224.4577 | 252.139999 | 72.317674 | 240.8727 | 252.139999 | 88.7326868 |
| 225.098 | 252.639999 | 72.457956 | 240.6725 | 252.639999 | 88.0324909 |
| 224.5259 | 252.5 | 72.025938 | 241.243 | 252.5 | 88.7430239 |
| 247.2661 | 267.320007 | 79.946102 | 238.9208 | 267.320007 | 71.6008265 |
| 238.3195 | 270.01 | 68.309509 | 234.3666 | 270.01 | 64.3565778 |
| **Prediction Accuracy** |  | **77.267699** | **Prediction Accuracy** |  | **75.786061** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **LSTM Single Variable** |  | **LSTM Multivariable** |  |  |
| **Actual** | **LSTM (SINGLE)** | **Difference** | **Predicted Value** | **Actual Value** | **Accuracy** |
| 230.9 | 237.6551065 | 93.2448875 | 192.805 | 203.6999967 | 89.10496 |
| 238.75 | 238.4649316 | 99.7149316 | 189.6125 | 197.8899987 | 91.72253 |
| 235.1 | 239.6492301 | 95.4507759 | 188.5863 | 205.0299987 | 83.55626 |
| 232.32 | 240.8173169 | 91.5026901 | 189.0961 | 201.0700067 | 88.02604 |
| 238.2 | 241.9467321 | 96.2532649 | 190.4094 | 204.9999997 | 85.40938 |
| 239.76 | 243.0255026 | 96.7344924 | 192.131 | 205.0899957 | 87.04104 |
| 243.85 | 244.0482414 | 99.8017647 | 193.7587 | 210.0399927 | 83.71866 |
| 247.33 | 245.0115204 | 97.6815184 | 196.185 | 214.4999997 | 81.68504 |
| 247.66 | 245.9128403 | 98.2528363 | 198.5875 | 212.6699977 | 85.91754 |
| 250.26 | 246.7504617 | 96.4904667 | 200.7008 | 210.9999997 | 89.70076 |
| 247.79 | 247.524182 | 99.734189 | 202.0344 | 210.4600067 | 91.57434 |
| 247.72 | 248.2354535 | 99.4845475 | 200.6749 | 195.8699947 | 95.19511 |
| 249.7 | 248.8872652 | 99.1872682 | 196.5917 | 205.0200037 | 91.57172 |
| 247.81 | 249.4837544 | 98.3262436 | 194.3753 | 204.3000027 | 90.07526 |
| 249.76 | 250.029751 | 99.730244 | 193.7959 | 205.5200037 | 88.27585 |
| 251.72 | 250.5304055 | 98.8104045 | 194.5704 | 205.2499997 | 89.3204 |
| 257.5 | 250.9906827 | 93.4906827 | 195.7857 | 207.3800047 | 88.40572 |
| 238.24 | 251.4154968 | 86.8245082 | 197.1876 | 208.3999937 | 88.7876 |
| 242.85 | 251.8088501 | 91.0411559 | 198.8566 | 210.1799927 | 88.67664 |
| 247.39 | 252.1746435 | 95.2153555 | 200.8991 | 218.1900017 | 82.7091 |
| 241.61 | 252.5158153 | 89.0941857 | 203.5716 | 219.8999937 | 83.67159 |
| 243.97 | 252.8349663 | 91.1350347 | 206.2606 | 219.8999937 | 86.36065 |
| 235.7 | 253.1344099 | 82.5655871 | 208.6404 | 223.6900017 | 84.9504 |
| 246.03 | 253.4157505 | 92.6142485 | 210.9407 | 224.8999937 | 86.04068 |
| 251.99 | 253.680339 | 98.309666 | 213.2931 | 232.1300047 | 81.16312 |
| 252.14 | 253.929425 | 98.210574 | 216.2147 | 234.5200037 | 81.69465 |
| 252.64 | 254.1638361 | 98.4761629 | 219.0638 | 232.1999967 | 86.8638 |
| 252.5 | 254.384501 | 98.1154991 | 221.2519 | 232.7599947 | 88.49195 |
| 267.32 | 254.5919432 | 87.2719362 | 222.8841 | 233.8000027 | 89.08405 |
| 270.01 | 254.7870409 | 84.7770409 | 223.2063 | 229.9900047 | 93.21625 |
| **Prediction Accuracy** |  | **94.9180721** | **Prediction Accuracy** |  | **87.40037** |

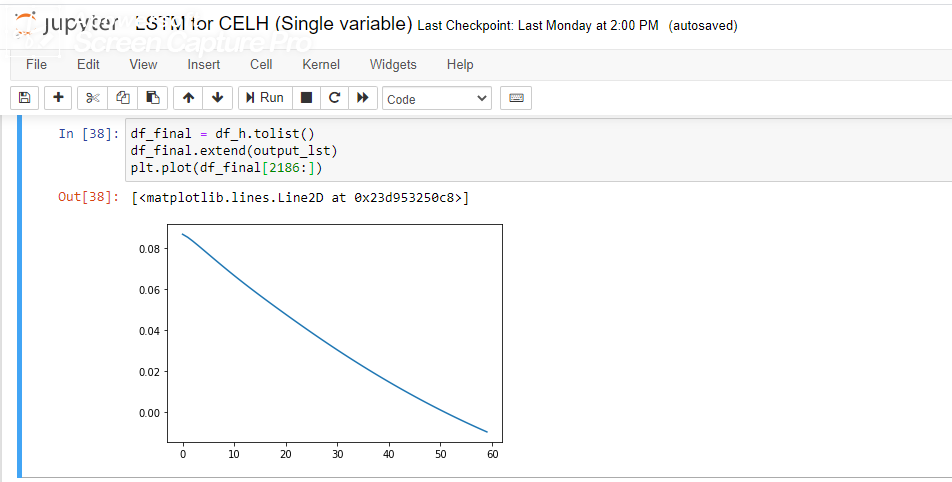
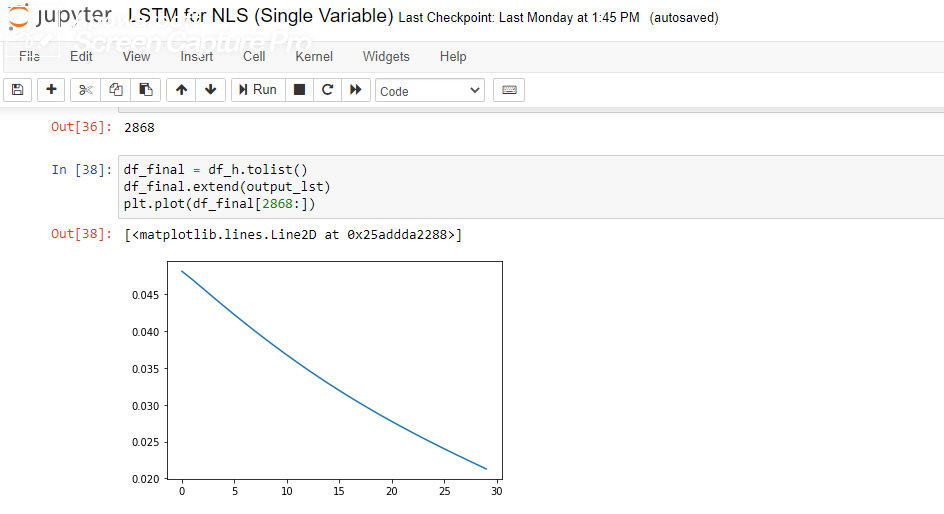
In the above tables for Stock price prediction modelling its clearly seen that LSTM has one of the best possible methods for Stock price prediction in relation other methods such as SVR and Linear Prediction modelling. An ARIMA has also lower accuracy than LSTM but the directional prediction remained accurate.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ARIMA for NVDA** | | | | | |
| **Predicted Price** | **Actual Price** | **Accuracy** | **Prediction Price** | **Actual Price** | **Accuracy** |
| 166.916209 | 231.937271 | 71.9661 | 170.013257 | 245.135498 | 69.35481 |
| 167.06972 | 234.912384 | 71.12001 | 170.169601 | 245.40506 | 69.34234 |
| 167.223373 | 239.514771 | 69.81756 | 170.326087 | 236.040497 | 72.15969 |
| 167.377165 | 235.681122 | 71.01849 | 170.482716 | 239.934097 | 71.05398 |
| 167.531098 | 236.669479 | 70.78695 | 170.639488 | 246.722885 | 69.16241 |
| 167.685172 | 239.534729 | 70.00453 | 170.796404 | 250.346893 | 68.2239 |
| 167.839387 | 239.984009 | 69.93774 | 170.953463 | 253.831161 | 67.34928 |
| 167.993743 | 242.619659 | 69.2416 | 171.110666 | 251.175537 | 68.12394 |
| 168.14824 | 243.917511 | 68.93652 | 171.268012 | 262.536804 | 65.23581 |
| 168.302878 | 251.56488 | 66.90237 | 171.425503 | 267.448669 | 64.0966 |
| 168.457658 | 246.872635 | 68.23667 | 171.583137 | 272.091034 | 63.06093 |
| 168.612579 | 245.165436 | 68.77502 | 171.740916 | 270.333923 | 63.52918 |
| 168.767642 | 248.519913 | 67.9091 | 171.898838 | 289.312592 | 59.4163 |
| 168.922846 | 248.869354 | 67.87611 | 172.056905 | 296.081451 | 58.11134 |
| 169.078192 | 247.531555 | 68.30571 | 172.215117 | 314.18161 | 54.81388 |
| 169.233681 | 249.697983 | 67.77535 | 172.373473 | 308.191467 | 55.93064 |
| 169.389311 | 252.443451 | 67.0999 | 172.531974 | 293.585571 | 58.76718 |
| 169.545084 | 250.067352 | 67.79977 | 172.69062 | 272.829834 | 63.29609 |
| 169.700999 | 239.804306 | 70.76645 | 172.849411 | 261.618317 | 66.06931 |
| 169.857057 | 247.561493 | 68.61207 | 173.008347 | 267.209076 | 64.74643 |
| 173.167428 | 252.334732 | 68.62608 | **Average Accuracy** |  | **66.91117** |



As its seen by the above tests that LSTM makes the best possible assumption in reference to price predictions hence its sufficient enough evidence that LSTM should be chosen for price prediction purposes.

As the single variable LSTM prediction method was the best option for NVDA in accordance to accuracy hence we started planning calculating accuracies in terms of single variable method in terms of the suggestions those were postulated through portfolio optimization method. In case of accuracy prediction, its seen that for both NLS and VNET it showed a downward growth whereas in reference to actual pricing the stocks made positive growth and were the very stocks that made gains during the COVID 19 pandemic.



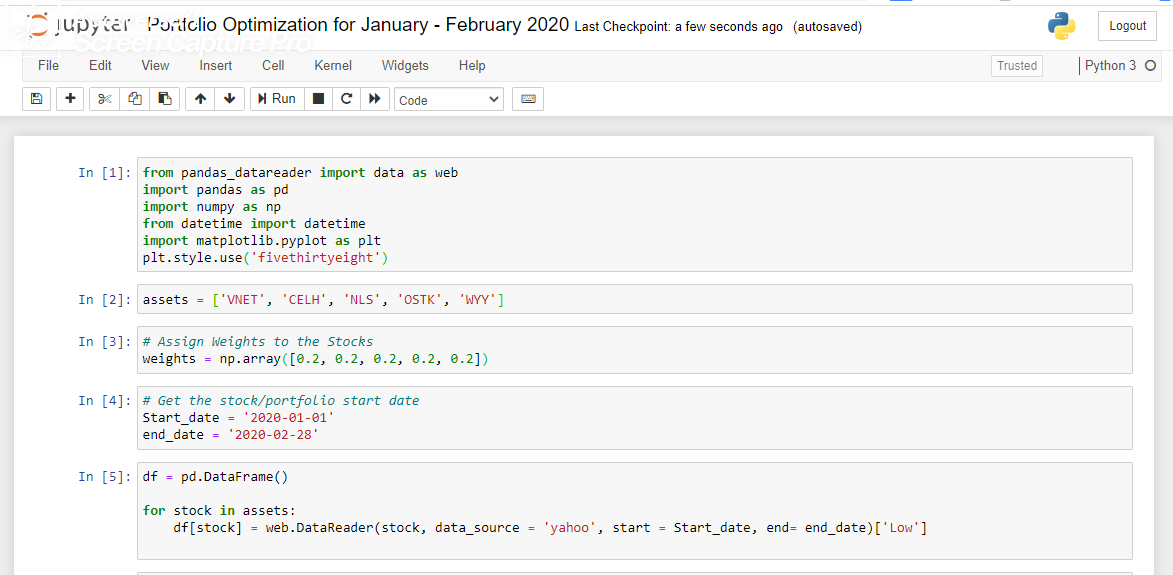
Hence for price prediction method we shifted towards the second-best method which is LSTM by multivariable method for price forecasting. Its seen that in terms of accuracy the LSTM multivariable method follows with the following levels of accuracy.

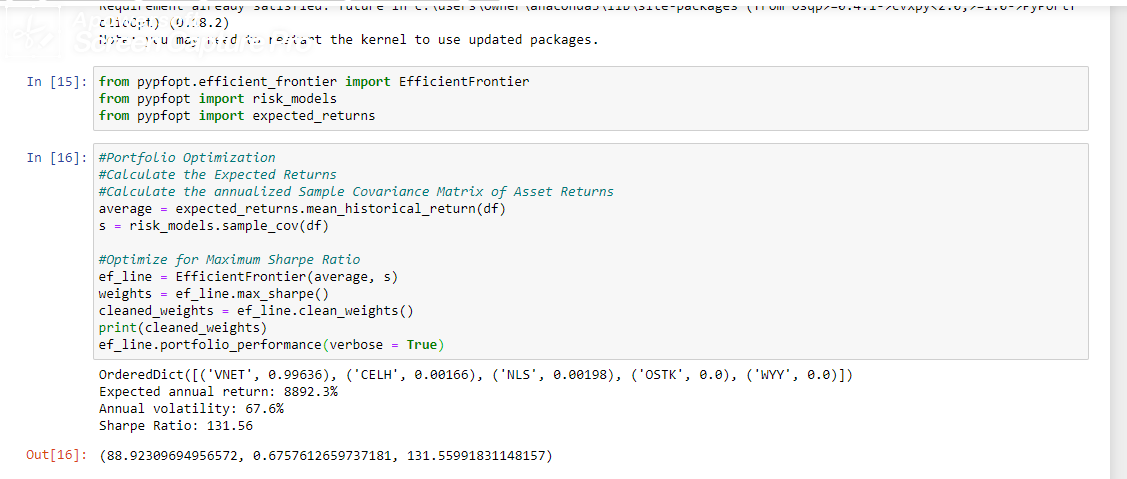
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Portfolio** | | | | | | |
| **Selected Stocks** | **Accuracy** | **Direction** | **Actual Price (Start)** | **Actual Price (Ending)** | **Predicted Price (Start)** | **Predicted Price (End)** |
| 21Vianet | 94.0967 | Upward | 4.96 | 5.83 | 3.28 | 5.76 |
| CELH | 94.3108 | Upward | 4.44 | 5.48 | 4.13 | 5.22 |
| Nautilus | 84.21472 | Upward | 1.22 | 2.53 | 1.93 | 2.78 |
| Overstock | 59.34514 | Upward | 1.67 | 2.67 | 3.84 | 5.695 |
| Wide point | 88.64262 | Upward | 2.4 | 3 | 2.84 | 3.5 |
| OIIM | 67.06476 | Downward | 0.53 | 0.27 | 0.67 | 0.51 |

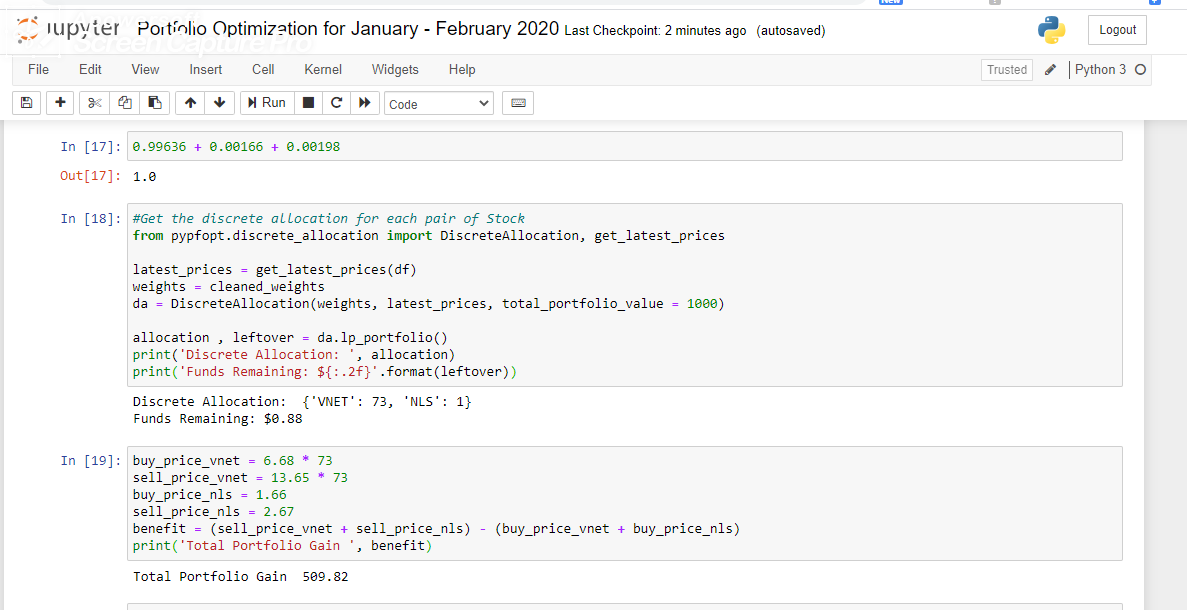
Hence with the prices in actual periods its easily seen that beginning price and ending price easily represented which stocks should actually make the best contributions in terms of both the growth aptitude and predicted price measures. In reality as this paper primarily focuses on the factors of testing the data samples on the basis of price forecasting techniques based on time series prediction techniques. In reality most of the portfolio optimizations gets carried by the practices of following processes:

1. Selecting the correct asset class from the various assets on the basis of financial ratios
2. Perform the time series analysis process on the best selected class.
3. After performing the analysis for future price forecasts selecting those which intends to make positive contributions in portfolio optimization processes
4. Selecting those assets and make a predicted optimized portfolio on the basis of selected assets and their forecasted prices.
5. Investing in each of these assets on the basis of allocated contributions
6. Finally, after the period testing it out with previous forecasted portfolio and newly formed prices in terms of its accuracy.

The final target would also remain same in accordance to the portfolio optimization and it would focus on collecting such assets that makes the best benefits in the portfolio optimization processes in terms of profit maximization and the best time series forecasting method would be the primary stepping stone for portfolio optimization.



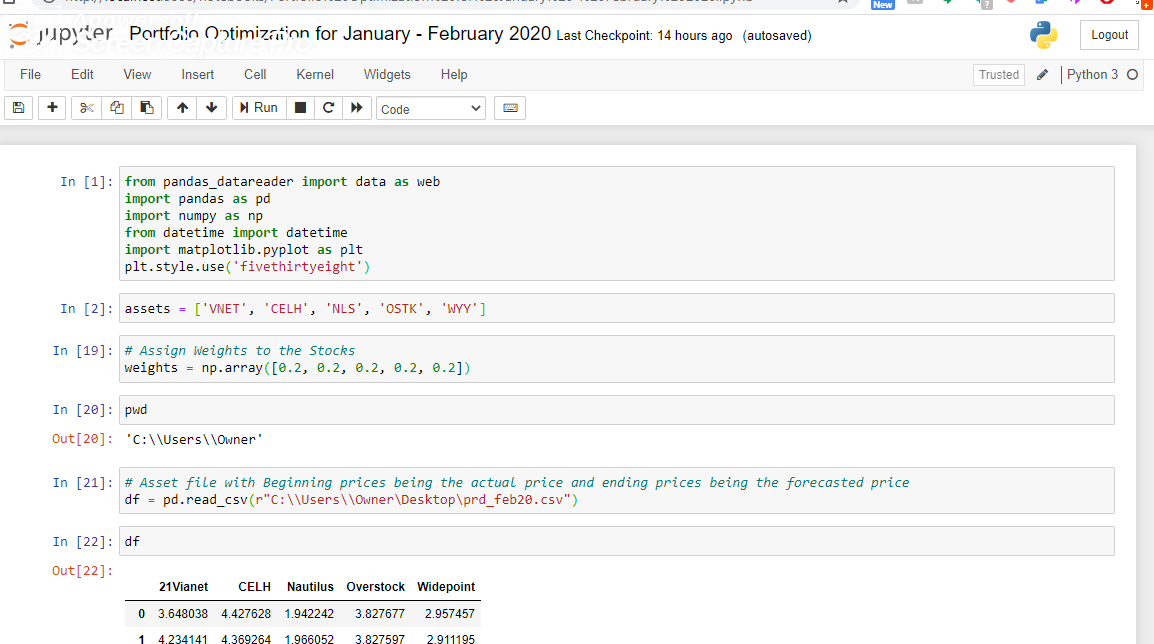




This is the approximating portfolio gain we predict to allocate with the best suggestive assets in the portfolio with highest possible gain ranges. In reality, the possibility of figuring out the exact prices for any future dates are not possible, hence we plan on creating our portfolio optimization based on predicted prices that we have predicted for the best possible gaining assets. And allocate them in portfolio optimization plan in order to check with possible gain allocation. The finally created portfolio gain could then be compared with actual portfolio gains in order to understand the accuracy of forecasted portfolio.

In accordance to the analysis for the real portfolio its seen that the real portfolio only allocates two stocks for profit maximization and risk minimization whereas the predicted portfolio uses for asset prices into portfolio allocation process. In reality the portfolio with actual prices have seen better portfolio returns than the predicted ones as we have seen that in our prediction models were successful in predicting a pessimistic price view (in most upward trending prices the actual prices remained higher than the predicted prices).

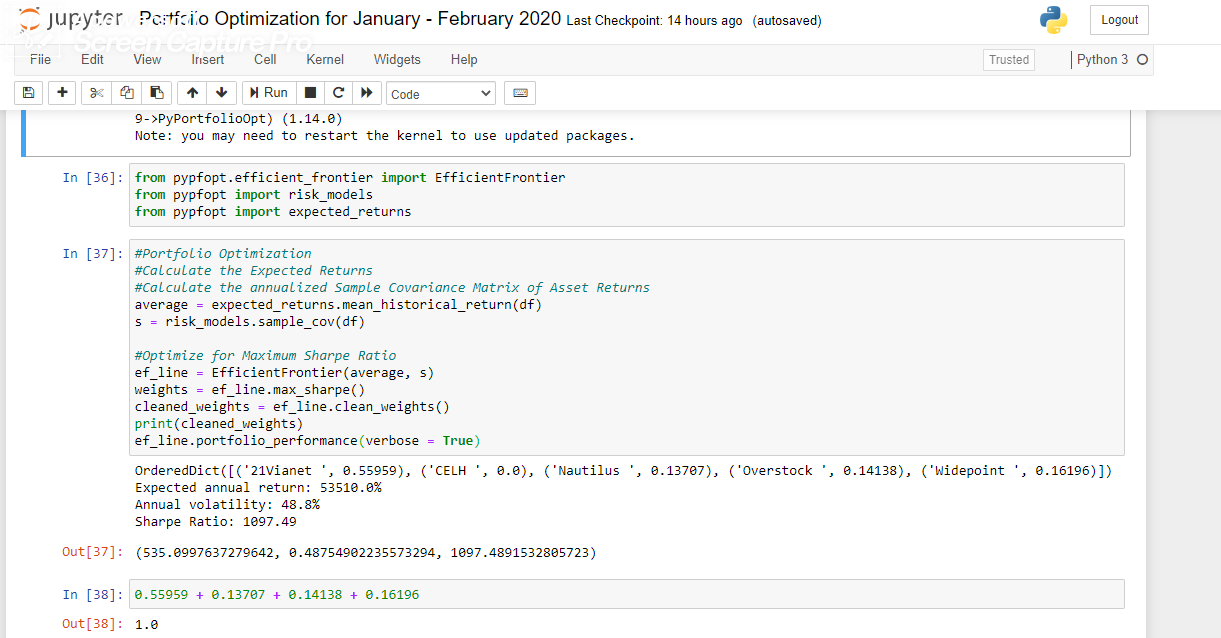
The forecasted prices drawn using LSTM multivariable method is discussed below as price associations for portfolio optimization.



The portfolio variance, volatility and the simple two months return are showcased below:

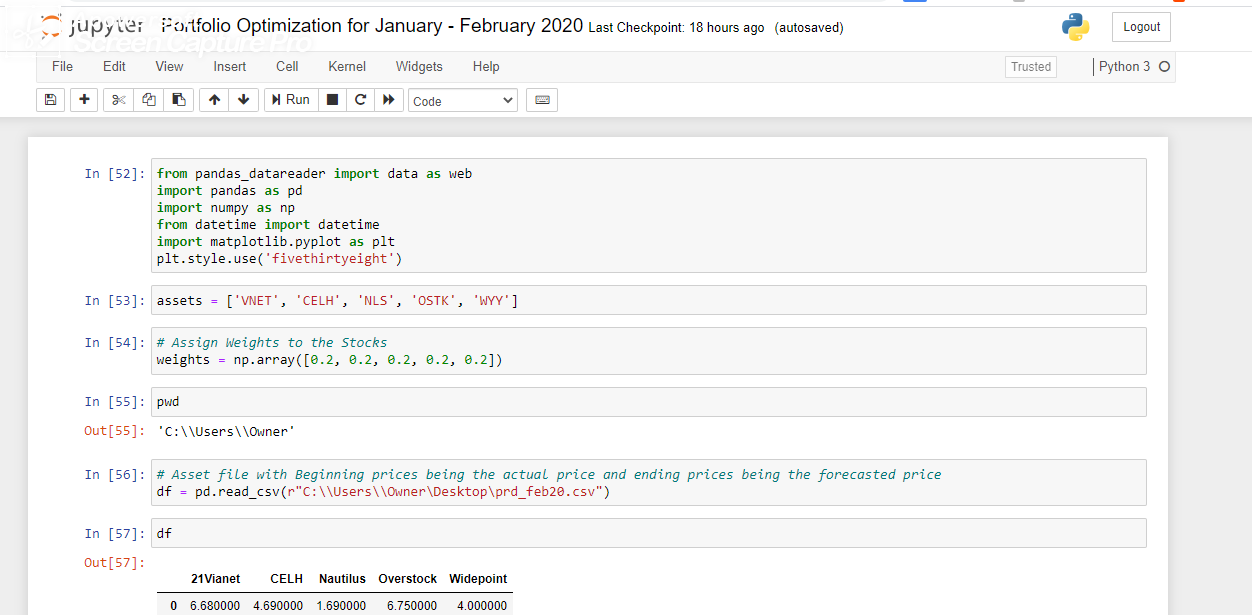


Using pypPorfolioOpt from Python as a library actually denotes the best possible combinations for the allocated assets in terms of profit maximizations and risk minimizations.



So, the forecasted portfolio will allocate its weights on the basis of best possible allocations of resources. The portfolio will allocate only the gaining assets for the portfolio optimization functions. As the beginning date for portfolio stock purchase will be January first 2020, hence the prices for that period needs to be recorded instead of the forecasted price for that day.

After that, we will initiate the portfolio optimization using the PypPortfolioOpt library associated with Python and then allocate the assets accordingly. The newly formed forecasted prices with one actual beginning price will create a new allocation to assets in accordance to the future prices. Finally, the new assets price at the date of ending will be tested in accordance to real prices with same number of allocated stock purchase for each asset and checked with final gain through stock purchasing. The portfolio gains are measured in two variant methods, one through actual prices at the end and other through forecasted prices at the end date.

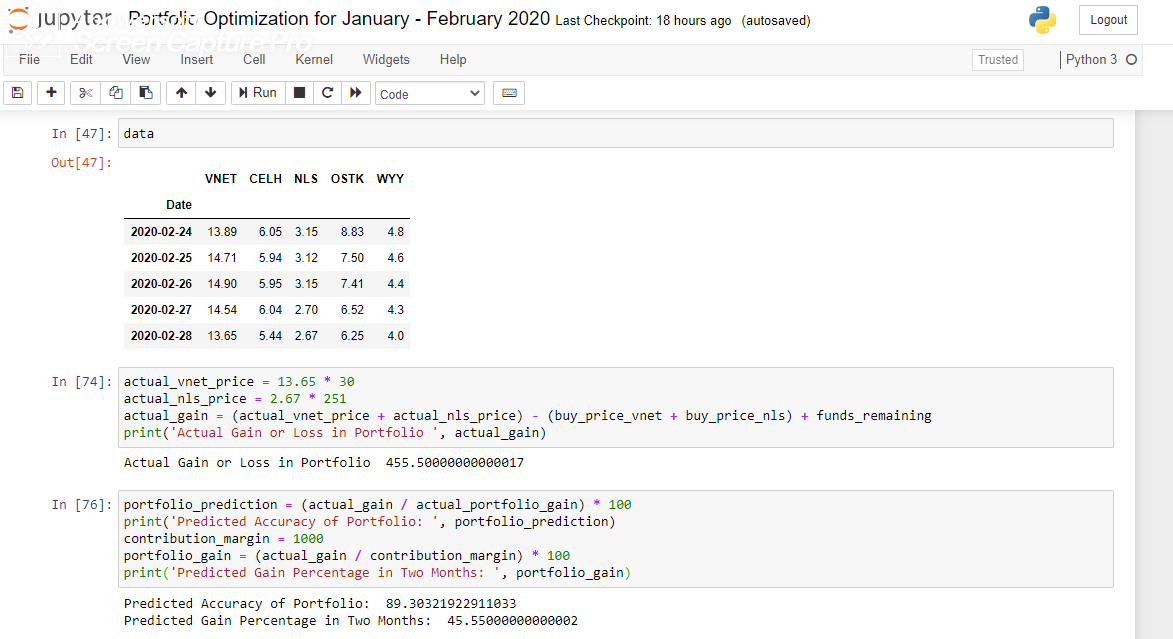


After checking with predicted future prices its seen that the future gains are lower in comparison to real prices.





The predicted makes a gain of 37.391% in two months. The predicted prices are more modest in comparison to the actual prices.



The real portfolio gain is way more than the predicted gain as it showing a gain of 45.55% in two months using the similar distribution of assets in terms of allocation numbers. The predicted portfolio accuracy has been compared with actual gain for the same capital contribution modelling which shows that in real scenario the contribution should had been done on VNET more than NLS whereas our predicted portfolio made more contributions on NLS rather than VNET hence the benefit margin has dropped down but our predicted price based portfolio and actual portfolio both suggested the same two stocks as comprehensive measure and the overall gain from the predicting processes also is pretty high (around 89.3%).

Recommendations & Suggestions:

Although our Stock price prediction method has been widely focused on the usage of Deep Learning as important measures of suggestions but there have been serious requirements to make this time series forecasting method appropriate in association with various suggestion undergone through MAE and RMSE suggestions. Its seen that our method mostly was highly capable of determining if the stocks were basically growing or not for a particular time frame rather than i.e. it was highly capable of determining the direction of growth or fall in reference to time. The method also saw high degree of error in reference to actual price predictions but was successful enough at the end in reference to gaining or losing model due to proper time series analysis method taken for consideration. The single variable based LSTM method well performed with one very well-known stock for testing purpose but in reality, couldn’t have performed well with the required stocks hence needs to make adjustments for single method processes as well. The ARIMA and SVR also made effective directional suggestions and how we can make such processes much more efficient could be a measurement of development towards financial time series analysis. The average accuracy for the six stocks were 81.23% in terms of price prediction through LSTM which also is not the best accuracy hence needs to be further evaluated. The K Means clustering process was initiated under the recognition of availability of very few financial ratios and the problem is associated with less data being available for all the stocks in open sources. The problem with these data sources is that they have serious implications in reference to changing scenarios, market noise and news, weather information’s and volatile political and social norms.

References: