Portfolio Robo-Advisor

PAPER BY

ANWESHA KERKETTA
HARMANDEEP KAUR
MANI KASTURI
SUSHANT BELAPURKAR

(MSBA-2019)

TABLE OF CONTENTS

BACKGROUND

Currently, the Capital market is one of the segments having tremendous data. Still, this segment is not so advanced in Data Science. As data is very complex, it is crucial to understand the data with the domain. But once we have the fundamental domain awareness, we can use the data in a great way to understand its working and its impact on the market. Furthermore, with new high technology, Machine Learning (ML), and Artificial intelligence (AI), it is also challenging and exciting to understand different patterns of the historical data to know the market situation.

With this domain knowledge and enormous data, many businesses developed and developing different tools to research the stock market data and creating automation tools to invest in the stock market. Based on this automation and AI, many companies have developed the Robo -advisors. These robots invest in the stock market without any help from the traditional financial advisors, which saves the investors cost. But if we look at the performance of these Robo-advisors, they are struggling to provide substantial investment possibilities to the investors. Few of the main reasons are[1]:

- These Robo-advisors are dependent on the other funds to manage the investments.

- Most of the robots look for historical prices and the fund's performance.

---

[1] https://www.liberatedstocktrader.com/do-robo-advisors-beat-the-market/

- Robots try to search for the stock market trend and end up losing the money as the stock market trend is a random walk.

- Most importantly, these robots' programs focused primarily on profit-making than concentrating on stable & reliable investments, which could benefit investors in the long run.

OUR APPROCH

The stock market refers to the trading of shares of publicly held companies. But surprisingly, while trading the stocks- emotions, political activities, and the world economy, inflation plays a significant role in the stock price movement more than company performance itself. As far as company performance is concerned, stock price shows a considerable movement with more or less change in quarterly data.[2]

Also, if we look at the successful investors, they always suggest finding a company with strong financials than depending on the stock market movement to buy/sell the company shares.

Keeping this in mind, we decided to create a Robo-advisor, which solely depends on the company's financial strength. We are focusing mainly on investment in companies having a solid foundation that will provide investors to create a wealth in the long run than making short term profits. To know how to study the financial statements in a required manner, we took help from the concepts explained in the book 'The intelligent investor' written by Mr. Benjamin Graham.

Below paper will give information about the source of the data, ideas of the portfolios & implementation of the same using Python and MySQL.

---

[2] https://www.investopedia.com/articles/basics/04/100804.asp

Our programmed Robo-advisor will build three different sorts of portfolios for various kind of risk-taker investors. These portfolios are as follows:

1. **Value Portfolio:** For investors having a low-risk appetite

2. **Growth Portfolio:** For moderate risk-takers

3. **AI-driven Portfolio:** Portfolio based on the machine learning mathematical model for investors who are willing to take a high risk.

   We will discuss the details of each in proceeding sections.

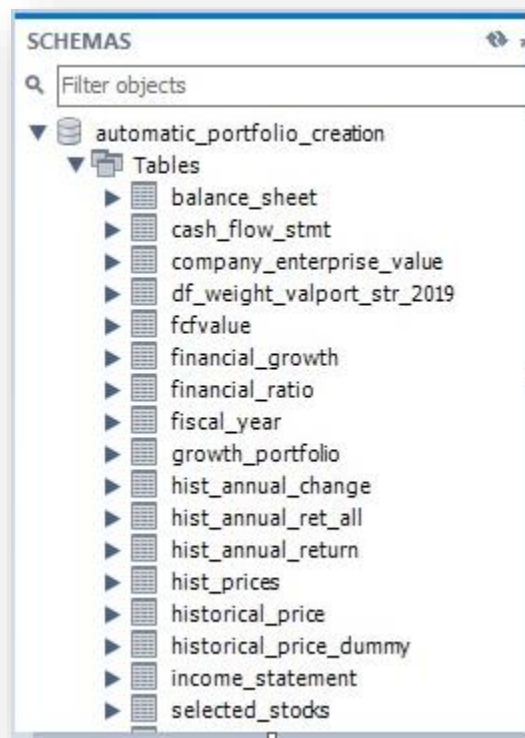## DATA AND DATABASE DETAILS

Our data consisted of three financial statements: Balance sheet, cash flow statement and income statement. The balance sheet reports a company's assets and liabilities and is often used to calculate rates of return. Cash Flow statement summarizes the amount of cash entering and exiting a company. Income Statement primarily focuses on the company's revenues and expenses during a particular period in time. These three financial statements were extracted from a financial modeling API[3]. Each financial statement has ten years of historical data, starting from 2009 to 2019. Federal reserve economic data[4] was also used to find the recession probability and treasury rate of history. Since the financial data was extracted from an API, it required merging different file formats, cleansing the data and creating Python to MySQL interface as our data in stored in MySQL. Our MySQL consists of around 10 tables, where in we are storing the data from the three main financial statements for each company along with other tables that house information like

---

[3] https://github.com/antoinevulcain/Financial-Modeling-Prep-API
[4] https://fred.stlouisfed.org/
  https://www.sec.gov/

historical prices, company profile, value stocks, growth stocks to name a few. We also have created customized views to fetch the data as needed for our analysis. Overall, we have around 4,000 stocks financial data which consists of over 1 million rows of data. Overall, we have around 4,000 stocks financial data which consists of over 1 million rows of data.

Our database structure looks like as:

```
SCHEMAS
Q  Filter objects
▼ 🗄 automatic_portfolio_creation
  ▼ 🗂 Tables
    ▶ 🔲 balance_sheet
    ▶ 🔲 cash_flow_stmt
    ▶ 🔲 company_enterprise_value
    ▶ 🔲 df_weight_valport_str_2019
    ▶ 🔲 fcfvalue
    ▶ 🔲 financial_growth
    ▶ 🔲 financial_ratio
    ▶ 🔲 fiscal_year
    ▶ 🔲 growth_portfolio
    ▶ 🔲 hist_annual_change
    ▶ 🔲 hist_annual_ret_all
    ▶ 🔲 hist_annual_return
    ▶ 🔲 hist_prices
    ▶ 🔲 historical_price
    ▶ 🔲 historical_price_dummy
    ▶ 🔲 income_statement
    ▶ 🔲 selected_stocks
```

INTRINSIC VALUE OR FIRM VALUATION

Once we have all the required data loaded, we began filtering out our stocks and the first step towards that process was finding out the intrinsic value of a firm. While potential gain is the main motive for an investor to invest in a stock, the onus of actual money making, falls on the company which accepts the investment. Any company is profitable if its able to create value, the value

creation in other words is to make net earnings from selling a product, excluding all the expenses. The money earned could be used to expand the company or give it back to investors in the form of dividends. Either ways a company is successful only if it creates value or generates wealth in other words its turning profits in business. Financial analysts look at trailing profits of a company to determine its financial health and thereby decide whether to invest in a company or not. Value investing is another approach where investors analyze intrinsic value of stock using its financial data. There is more than one approach to determine the intrinsic value of stock using financial data by employing different valuation methods. Some of the approaches that are used by some of the top accounting companies to valuate a company are Discounted Cash Flow method, Economic value method, residual earnings method. It requires accurate financial data for at-least 5-7 years to feed into the evaluation model to arrive at valuation of the company. Once we find the intrinsic value of a company, we can derive the intrinsic share value by dividing the value by number of shares. Every company that is traded in prominent stock exchanges file for annual statements or 10K statements to SEC every financial year. This information is publicly available.

By collecting this data and using valuation methodology we can determine if a stock at current price is undervalued or overvalued.

Discounted Cash Flow Method is our choice to valuate companies as value investing being the pinnacle of our portfolio and discounted cash flow method is best suited for stable companies with good financial health. Discounted cash flow method utilizes two components, firstly free cash flow of a company and secondly the weighted average cost of capital, in other words the expenses on capital to earn this cash flow. Using past data such as sales for earnings and cost of goods sold, operational expenses, interest expense, depreciation and takes we calculate the net amount (or cash flow CF) that will be earned by the company each year. Assuming that sales grow over time (as

7

percentage) we project the sales as well as average expenses for next 5 years. This way we are finding the free cash flow for future and this is discounted by the weighted average of capital.

Weighted average cost of capital is nothing but a percentage that determines the discount rate (dr) at which free cash flow is available for the firm for the next five years. It uses following components as well.

Long term debt

Equity

rm – Market return rate

rf – Risk free rate / Treasury bond rate

rd– Long term debt interest rate

β – Beta for the stock or industry

You can refer to the above formula on how the discounted cash flow is determined. In the formula n represents max number of years for which the projection is calculation, which is 5 in our case. All other parameters used to calculate the discounted rate can vary depending on when the valuation is being calculated, in other words parameter can be manipulated to simulate the valuation. This is extremely useful if you want to base your valuation on varying interest rates for your long-term debt or if the Federal bank adjusts their interest rates. Financial analysts can use this method as a leading indicator the determine the trading or support levels of the stock price.

In summary, valuation methods are the best tools to be used to convert the book value to share price. Many times, we come across news articles which say that "Apple" is a trillion-dollar

company and its share price is something, we can verify the share value by using 10K statements files by Apple to SEC to find its real worth. There are many pros and cons in using different valuation methods, but discounted cash flow is more robust than other two other two methods are more forecast biased and doesn't include risk factor. Another common flaw for valuation methods is that they only consider the tangible assets which is in the financial data. They don't include intangible assets such as brand value or goodwill for a company.

Our company valuation data frame looks like:

Calc_year	Closing_price	STK_TKR	Share_price
2019	34.05	CMCSA	52.2746254327189
2019	46.93	INTC	46.5031693828478
2019	31.73	MU	59.3051671120497
2019	157.74	AAPL	116.625529551222
2019	101.57	MSFT	17.3680176193811
2019	5.71	SIRI	3.6505474403151
2019	20.46	HPQ	274.548912046999
2019	7.65	F	301.856624728206
2019	18.46	AMD	-0.321028691433878
2019	28.54	T	59.2710246572354
2019	13.19	VALE	19.3660359140153
2019	2.25	AKS	0.148036903035.1

SELECTION OF VALUE AND GROWTH PORTFOLIO

Right from the inception, we have divided our investor base into 2 main categories,

1. The conservative or risk averse investor and

2. The enterprise or the risk taker

The conservative investors are those who prefers stocks that are steady, comparatively less risky and stocks that are likely to yield good return over a period. This idea of investing fits well with what has been conventionally termed as the 'Value Investing Strategy'. Value investing strategy works mainly in two steps, first the analyst looks at the company financial statements to estimate a fair or intrinsic value indicating what the actual worth or potential of the company is. Through this process they try to identify companies that have strong financial foundation, when identified, these companies may or may not be performing to their full potential, but they have the potential to yield consistent and better results over the time. Going by the company's financial documents, the intrinsic value that we have calculated in the previous step (firm valuation) gives us an estimate of what could be a fair value for the stock currently. The main idea behind value investing is to pick stocks by only looking at their merit and quantitative data and not to give in to the sentiments or emotions of the market. Value investors believe that market can sometimes run high on emotions which may result in buying/selling of stocks which may not be backed by their numbers. By taking out emotion from the equation, value investors in fact see the situation as a good opportunity to buy stocks that have the potential and are undervalued.

What does it mean when the analyst says that a stock is undervalued? This question leads us to the second part of the value investing strategy; once the present value or the intrinsic value of a company is calculated, the next step is to check what is the current market price of this stock? is it trading over the intrinsic value or under intrinsic value?
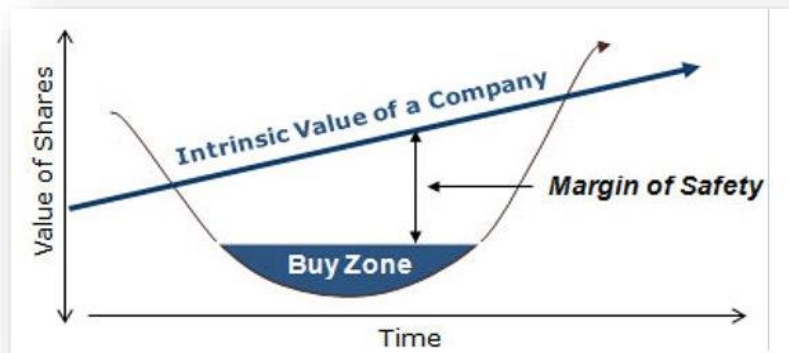
Fig-1[5]

For the companies that we have calculated to have a positive intrinsic value which indicates strong financials (excluding the extraneous conditions that can impact the price), if this intrinsic value is greater than the current market price(meaning the stock is undervalued,) then it would be ideal to buy a stock at this price because it gives us a cushion against market fluctuations.

This difference between the price is also known as margin of safety. On the other hand, when the market price is over the intrinsic value of the stock is said to be overvalued. Mr. Benjamin Graham in his book 'The Intelligent Investor' mentions that the current market price of a stock should be at least one third below its current market price. With this as a guideline, we also used ratios like P/B, P/E, current ratio etc. to further narrow down stocks selected that have the positive intrinsic value.

---

[5] https://httpss://investingpr.com/value-investing/

Value investing though fundamentally strong needs a lot of diligence and patience. As one of the most famous value investors Mr. Warren Buffet said, "Only buy something that you be perfectly happy to hold if the market shut down for 10 years". This confidence comes from using accurate data to calculate the financial ratios from the company statements, using the right formulae and thresholds as per the industry, diversifying the portfolio and most importantly not buying an overvalued stock.

After applying above parameters into logic our value stock portfolio (listing few stocks) looks as below:

STOCK_TIKR	SECTOR	BETA	FAIR_VALUE	Calc_year	ADDED_ON	CLOSE_PRICE	BUY_NOW
ACCO	Industrials	1.818291	19.0	2019	2019-12-18	9.585	Yes
ADM	Consumer Defensive	0.955928	45.0	2019	2019-12-18	46.235	Yes
AIR	Industrials	1.009082	28.0	2019	2019-12-18	45.415	No
ALG	Industrials	0.850361	39.0	2019	2019-12-18	117.990	No
AP	Industrials	2.154915	13.0	2019	2019-12-18	2.990	Yes

GROWTH PORTFOLIO -- ENTERPRISE CUSTOMER

For our second classification of customers, the enterprise investors who have a little more appetite for risk, we offer a portfolio of growth stock which comprises of stocks that have above average return in their sectors. These are usually young or small companies that are coming up with innovative ideas/technologies which are expected to do well and result in some profits or capital appreciation. Rather than giving dividends to customers typically growth companies tend to reinvest their profit which in turn results in stock price appreciation.

These stocks are more volatile in nature. The financial ratios that determine if a particular stock has the potential to yield profits in the future are much different when compared to the value investing. Typically, the ratios that we consider are Company's ROE, EPS, Strong earnings growth etc.

For our growth portfolio, we have considered stocks whose price is within 20 percent of the intrinsic value of the firm and tried to apply some formulas on top involving ratios like ROE, EPS etc. which helped us choose stocks for our growth portfolio.

After applying above parameters into logic our value stock portfolio (listing few stocks) looks as below:

| STOCK_TIKR	SECTOR	BETA	GrowthEPS	Calc_year	FAIR_VALUE	ADDED_ON	CLOSE_PRICE	BUY_NOW
TEL	Technology	1.047484	Yes	2019	71.0	2019-12-18	95.450	No
TRIP	Consumer Cyclical	1.299399	Yes	2019	12.0	2019-12-18	29.195	No
UAA	Consumer Cyclical	0.419624	Yes	2019	3.0	2019-12-18	20.575	No
VLO	Energy	1.184196	Yes	2019	123.0	2019-12-18	95.670	Yes
WBA	Healthcare	0.807362	Yes	2019	74.0	2019-12-18	57.135	Yes

PORTFOLIO OPTIMIZATION

Economist Harry Markowitz introduced Modern Portfolio theory in 1952 for which he was awarded a Nobel Prize in Economics. This theory explains a mathematical structure to gather a list of stocks, assets etc. within the portfolio such that the expected rate of return is maximized for a given level of risk.

13

Considering the Modern Portfolio theory, we have calculated the risk and expected return for stocks after creating the Value and Growth portfolio to optimize our automated portfolio. Optimization is the selection of best component from some set of available alternatives called constraints, it's a set of restrictions or limitations.

The constraints taken into consideration here is a) the sum of the weights distributed among the selected stocks based on past 5 years stock performance is equal to one. Secondly b) the weights are not less than 0 which negates the stocks which are short. Using the python library cvxopt and using all the above constraints we ran the solver to get the optimal solution which is maximum return and minimum risk for the stock trend for the next 5 years. However, the trend is also calculated based on the volatility of the stocks which is calculated using Sharpe ratio. Sharpe ratio measures the performance of an investment compared to a risk-free rate after adjusting the risk. Risk free rate is the rate of return of an investment over a given period of time. The time taken into consideration is annual risk-free rate from the treasury.gov of 1.59.

- We are giving the range of returns between 5 to 10 percent. For each of this return, our function will calculate the weights of all the stocks.
- For each return, we are also calculating Sharpe ratio.
- The return having the highest Sharpe ratio, we will take those weights into consideration for our portfolio stocks.

From the below graph Efficient Frontier for the Growth stock we can apprehend that the expected return is maximized while the expected volatility is minimized after optimizing the portfolio.

Efficient Frontier - Growth Stock
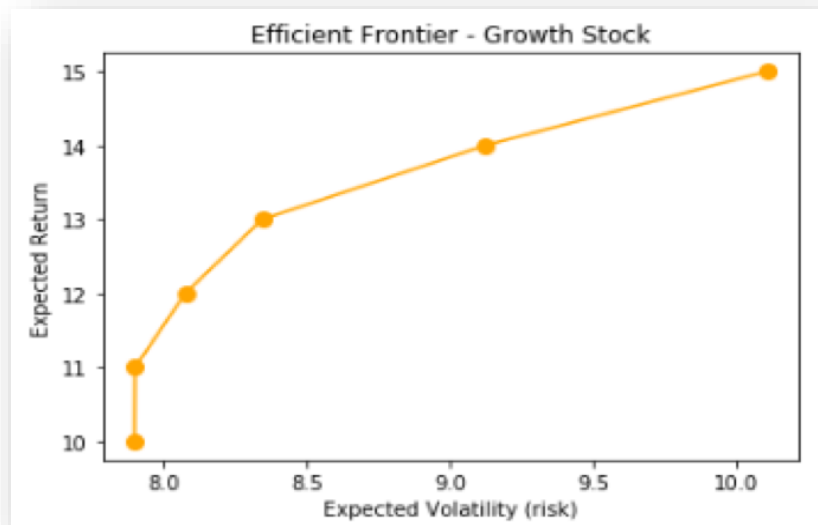
FINALIZING PORTFOLIO AND PERFORMANCE AGAINST S&P

After doing our analysis and assigning weights to various stocks, we bucked our analysis in different time frames. Exhibit shows results of 2010 to 2014 growth analysis. Our Growth analysis consisted of forty-three potential stocks and sixteen undervalued stocks that were added to our portfolio. Our portfolio significantly surpassed S&P 500, the approximate rate of our portfolio is 13 percent, while the S&P 500 return rate is roughly 6 percent. In addition to our portfolio having a higher rate of return, our portfolio is well diverse incorporating companies from different sectors such as technology, natural energy and more.

Stock selection period	Potential	Undervalued
2010-2014	43	16

S&P vs Growth Port. 2015-18
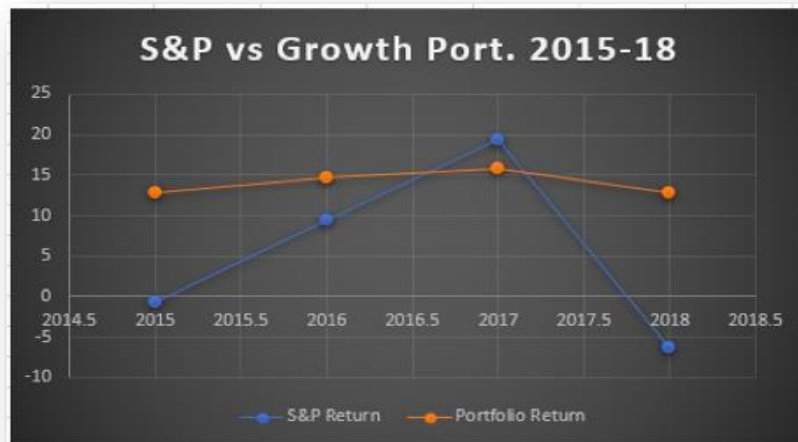
S&P500 Returns	Growth portfolio Return
5.50%	12.94%

The same analysis process was followed for value analysis. We bucked our analysis in different time frames. Exhibit __ shows results of 2010 to 2014 value analysis. Our value analysis had fifty-one potential stocks and eighteen undervalued stocks that were included in our value portfolio. Our portfolio performed remarkably better than S&P 500. The rate of return for our value portfolio is about 12 percent, in comparison to S&P 500 return rate being 6 percent. As the image be shows the stocks in our value portfolio is across various sectors.

Stock selection period	Potential	Undervalued
2010-2014	46	31

S&P vs Value Port. 2015-2018
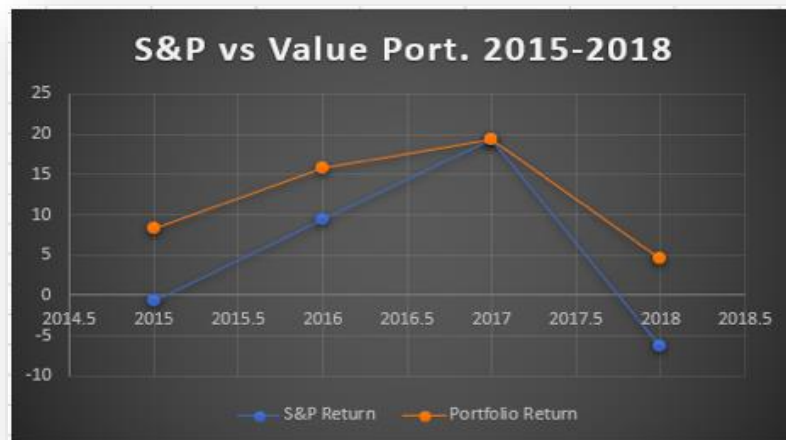
S&P500 Returns	Growth portfolio Return
5.50%	11.99%

PORTFOLIO REBALANCING

Rebalancing is the act of periodically buying or selling assets to restore your portfolio to its original target allocation range. In our system, annually we load our three main tables (Balance Sheet, Cash flow and Income statement). Once we have the new data in tables, we re-run our stock picker code and see if there are any new stocks that would satisfy our criteria for value and growth portfolio,

If yes, the new entrants would then be routed through our optimization engine to determine what weights would be assigned to the new array of stocks in our portfolio. Every year when we re-run the stock picker code, we keep track to see if every stock in our portfolio has been consistently meeting out criteria or not. If a given stock A, has made it through our stock picker code, satisfying

all the criteria we defined, two out of three times then at the end of three years we would replace

stock A with a stock that has been meeting our criteria all the 3 times or rather with a stock that

has been performing more consistently.
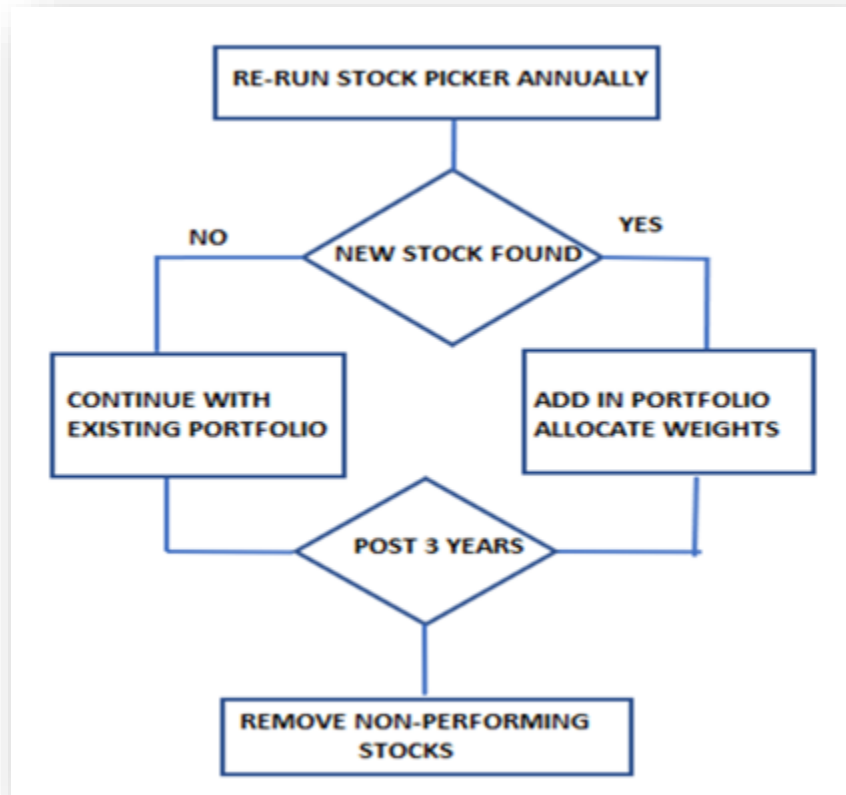
AI DRIVEN PORTFOLIO - BACKGROUND

For a few years, many companies started using machine learning (ML) and artificial neural network (AI) with stock market data. Stock market data is a typical example for which ML can work more nicely. Algorithms of ML and neural networks are meant to handle and process such a large amount of data. If the ML algorithms train suitably, it is exciting to find meaningful information with stock market data.

When we research ML in the stock market, except few, most of the models are trying to predict the stock/index price based on the historical price data[6]. But as we stated earlier, share's price movement is dependent on many factors which technology cannot handle as of now, such as sentiments and political factors. As a result, existing price prediction ML models are not performing up to the mark as expected[7].

Considering this, we admitted using ML in a way that models should not impact by non-technical features.

As we are building our programmed Robo-advisor based on financial statements to select the stocks, we finalized to use the same data, not to predict the stock price but to predict the trend of the company shares.

Still, the stock trend over a short period depends on many non-technical/non-arithmetic parameters. Thus, to monitor the direction of a stock over the historical time frame, we examined annual stock returns based on companies' annual financial results, considering the yearly

---

[6]
https://www.google.com/search?biw=1366&bih=576&ei=L5T4XdX1BZC60PEPtMy78AI&q=research+in+stock+market+machine+learning+&oq=research+in+stock+market+machine+learning+&gs_l=psy-ab.3..33i22i29i30l9.120105.134275..135059...0.5..0.136.3499.30j9......0....1..gws-wiz.......0i71j0i273j0j0i131j0i67j0i70i255j0i22i30.ESeG2BX8ajA&ved=0ahUKEwiVye6YpLzmAhUQHTQIHTTmDi4Q4dUDCAs&uact=5

[7] https://blog.usejournal.com/why-stock-predicting-ai-will-never-take-over-the-world-b1b411decc21

performance of the company has more influence of financial data than the other forces of stock market movement.

With annual financial results, we are also taking historical treasury rate and recession data into consideration, thinking it may influence the long-term performance of a company.

With this, we are considering the annual returns of the S&P 500 for our model, to select the stocks not only having a positive trend but also have better performance than the S&P 500.

In a nutshell, we will try to select the stocks which may perform better than S&P next year, considering the company's past annual results with historical treasury rate and recession data.

MODEL SELECTION, PERFORMANCE

As discussed above, we consolidated the historical ten years (2009-2019) of data for the balance sheet, income & cash flow statement, recession probability and treasury rate for all the stocks.

Based on this data and with S&P500 and stock returns, we marked each stock performance as either a UP trend or Downtrend for all the ten years. This dataset resulted in the binary classification with around eighty features in it. We further clean and converted the data to categorical values as required for the machine learning models. With total of around 32000 rows, our dataset contains balanced output, or we can say that our dataset is the balanced one as it has almost equal amount of up and downtrend stock movements over the past ten years.

Dividing the dataset into training and testing, we have run around 9 classification algorithms. The accuracy of each of the models is as shown below.

| Acc_Score	Model
100.00	Random Forest
100.00	Decision Tree
80.64	XGBooster
77.98	KNN
55.66	Support Vector Machines
53.73	Perceptron
53.29	Logistic Regression
52.60	Stochastic Gradient Decent
48.55	Naive Bayes

From above we can see that Random Forrest gives highest accuracy. But this 100% accuracy seems impractical. So, we ran ten-fold cross-validation method on Random forest to come up with the better validated accuracy. And this accuracy comes around 72% for random forest. We also got precision and recall, F1 as follows:

	precision	recall	f1-score	support
0	0.74	0.74	0.74	4819
1	0.73	0.72	0.72	4532
avg / total	0.73	0.73	0.73	9351

With this we also decided to crosscheck the data accuracy and performance using deep neural network (DNN). With Relu activation function and R2 regularization, we build DNN model with 5 hidden layers and 300 epochs. As we have around eighty features, this DNN model resulted into overfitting with accuracy around 65% as shown below.
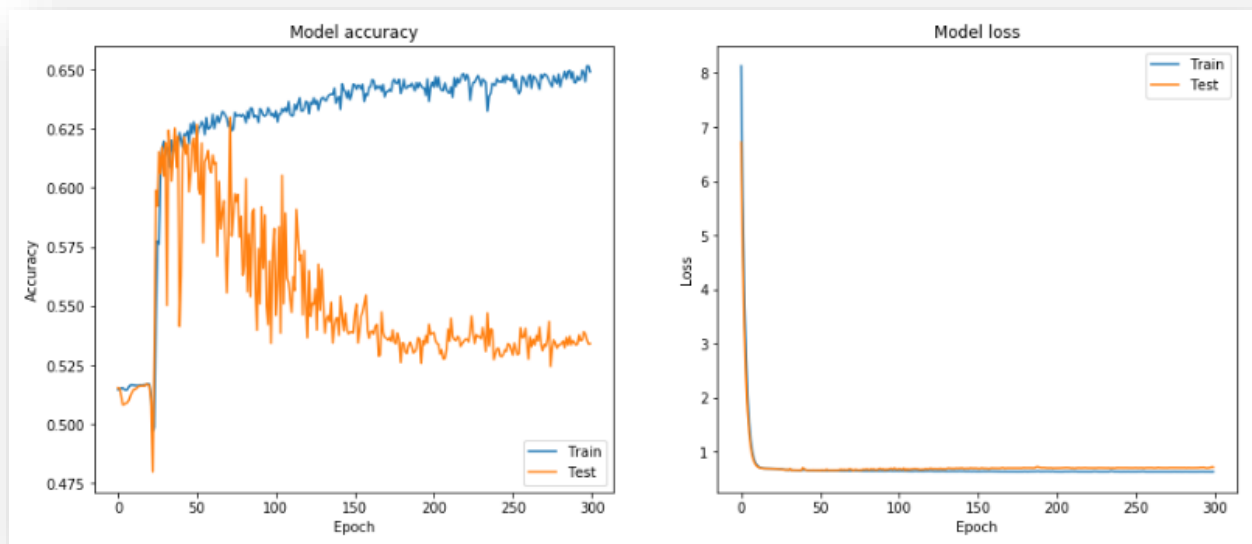
To overcome this overfitting and to achieve better accuracy, we used PCA analysis. PCA suggested that 18 components of the dataset explain the 95% of variance of the dataset. To decide which 18 components to be used, we perform the feature importance using random forest feature importance method. Below is the chart of the feature importance data.
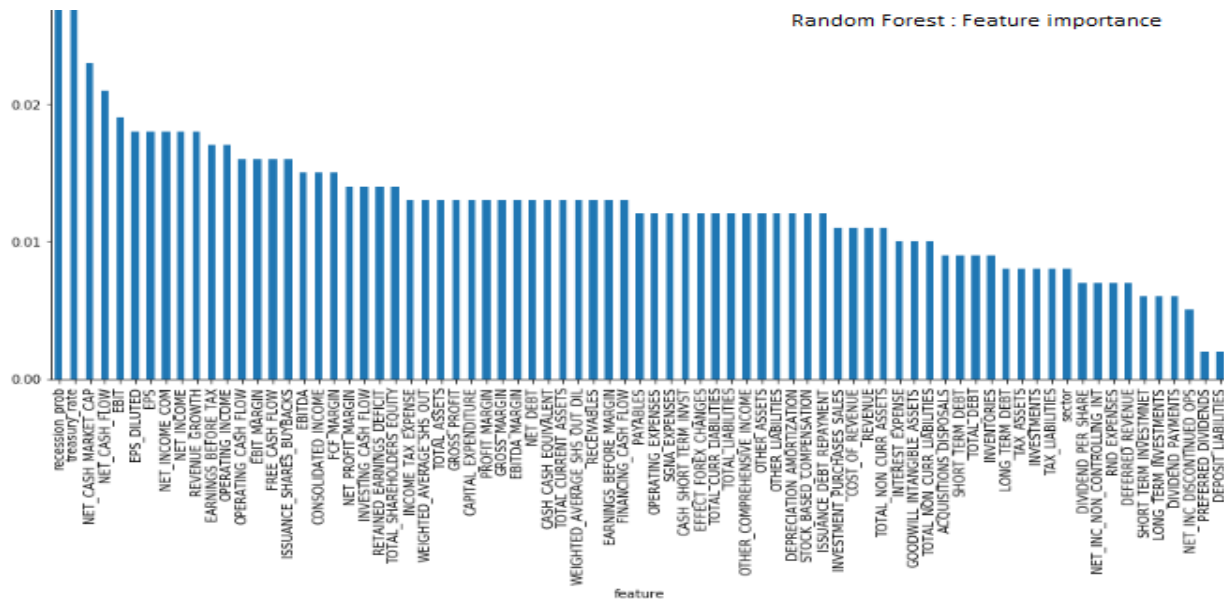
We ran the random forest model and DNN model, with cross validation with 5 folds and with top 18 features of above chart.

With less features, we end up with random forest accuracy, with 74%. But at the same time DNN performance increase a lot without any overfitting as shown below.
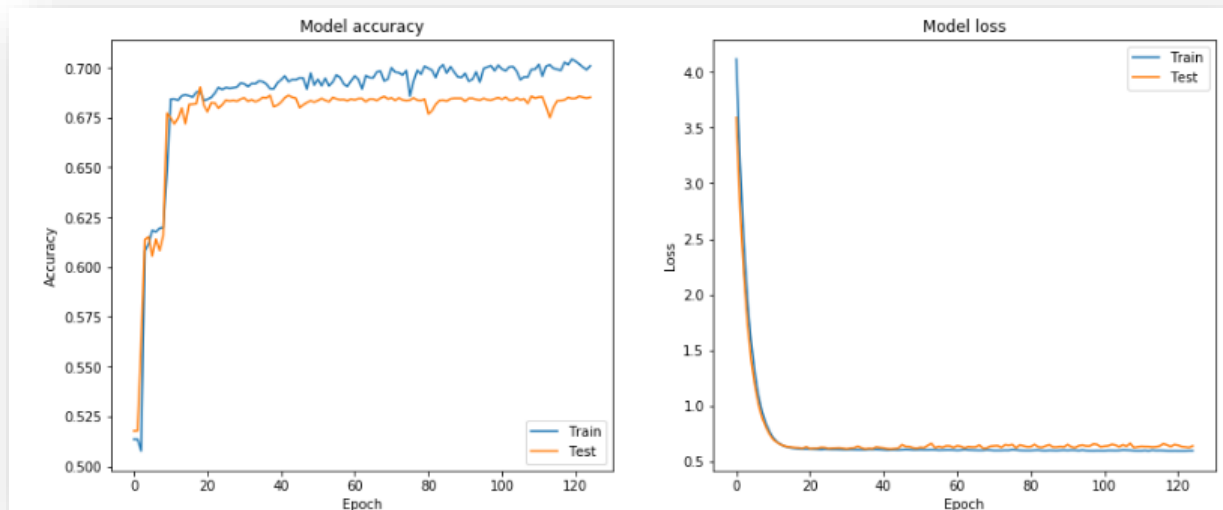
With this improved model, we predict our data and found 15 stocks that may perform better than S&P500 in next year, few of which we have listed below:

STOCK_TIKR	COUNT	Next_Yr_beat_SnP
ALNY	7	Yes
AMD	6	Yes
CXO	6	Yes
DQ	6	Yes
ERJ	6	Yes

CONCLUSION

Companies' shares are trading on the exchange for decades. But still, it is considered as one of the risky investment options. As discussed in the paper, the average investor interested in quick money and depends on external factors than the company performance, which can be analyzed using company financial statements.

We applied the techniques invented by successful investors and Nobel laureates to build an automated system that will pick such solid stocks for the portfolio.

The most crucial purposes for selecting these stocks are:

1. Select companies that will sustain in the weak economic conditions.

2. Build portfolios suitable for all kinds of risk-takers.

3. Instead of offering short term profits, create a long-lasting wealth creation portfolio.

4. Maintain the best risk-return balance in the portfolio.

We cannot give guaranty that our portfolio will give the best returns all the time. However, we have tested our portfolio for different time frames from 2009-2018 with the data for the same time period. We examined that our portfolio did well and gave returns near or better than S&P500 majority of the timeframes.

We have used advanced technical tools and algorithms to work in a better way on all arithmetic data, which gives us confidence in the results.

There is a lot of room to improve our performance and to build a more productive system. We are working on the same.

Last but not least, we sincerely thank Prof. Navid Shabbaghi for his excellent suggestions & guidance which resulted in the improvement of our Robo-Advisor.

Stock/Shares Investment is subject to market risk. But as Warren Buffet said, 'Risk comes from not knowing what you are doing'. We tried to mitigate this risk at all levels while building our portfolio so that common investors can confidently invest in our Robo-Advisor.