**Selected Stock Portfolio Optimization**

*Kehinde Afolabi, Sonny Desai, Jorge Lambis, Philip Melanchthon*

November 2020

**Abstract**

This project deals with an analysis of stock market projections, and how they can best be altered to increase performance results. The problem we are trying to solve is how to best increase yielded returns from a stock market portfolio, especially in comparison to other funds. Thus, the question we are asking is: What is the correct combination of stocks in a portfolio to achieve the optimal return on investment? Investors want to determine the performance of their portfolios from several stock investments, mainly through Quarter-Over-Quarter, and an evaluation of average ROI Year-Over-Year (YOY) for each investment. We seek to provide investors with risk evaluation for multiple stock purchase combinations by advising investment diversification options while controlling volatility. We then compared the results of these tools against real-world ROI on specific markets. Our methodology was to solve the aforementioned problem by maximizing performance of our Investments. We gauge success by viewing the dollar amount to Invest in each stock validated by upward trend in performance. To do this programmatically we used the Markowitz mean variance model. Our results from this model were a 3% return on investment, which beat the S&P 500 and Vanguard 2040 targets dated retirement funds. After further analysis though, we concluded these results might be an unrealistic comparison because the aforementioned targeted funds have bonds and other financial products to lower volatility, so their returns are understandably less.

**Key Words**

Portfolio, Performance, Optimization, Returns, Market.

**1 Introduction**

One of the most important issues and questions constantly facing the financial industry is how to create the best mix of stocks in a portfolio to yield the best returns. The financial industry, like most other industries, has faced disruption through technology and innovation, with the usage of data analysis and machine learning often been attributed to yield these better returns for investors. Stocks and other financial products have become more accessible to the lay user, with apps like Robinhood and Stash bringing financial products directly to many consumers, including those who had never thought about financial portfolios in the past.

As part of these offerings and disruption, deep analysis of stock trends, performance, and predictions are often used as tools to help entice potential investors and to increase the value of certain funds and portfolios. Numerous new predictive techniques have been created to get the best return on a portfolio or mix of stocks. Financial companies seek to pair the best mix of stocks, with a certain amount of risk measured with reward, to sell to their consumers. Part of this process is using data to help predict future returns. With so many investors having been preyed upon by bad advice, predatory tactics and Ponzi schemes, the use of statistical process with data in predictive modeling offers a viable alternative with data backed results.

Our motivation for this project was to continue that trend of using machine learning and data to aid in disruption and to see if we could bring better results from certain portfolios. Our problem we ware trying to answer is: What is the correct combination of stocks in a portfolio to achieve the optimal return on investment? Our goal in answering this question in regard to predictive analytics, is to assess whether certain predictive analytics techniques can help us solve for the optimal solution of mix of stocks which will lead to a higher yield of results, especially in comparison to other popular funds. Our hypothesis is that we can use the predictive analytics method of the Markowitz mean variance model to solve for the optimization of stocks in a portfolio which will lead to the highest yield of return to be compared against other financial funds, notably S&P 500 and Vanguard 2040.

1. **Literature Review**

There are numerous other predictive analytics papers on financial portfolios which is not surprising because on the size of the financial industry worldwide, the amount of tech disruption that has continuously been a part of the industry, and the rising popularity of applying predictive modeling and machine learning to financial products. The size of the financial industry, along with the wealth contributed, affords a plethora of attention opportunities for innovation. This paper, and the review of others like it, focuses on how predictive analytics through statistical methods use data to predict future behavior and returns of stocks – with the end goal being how to solve the optimization problem of providing the best mix of stocks in a portfolio which lead the highest returns. The benefit of solving this problem is obvious, not only will new methods of innovation be brought to a competitive industry, but the prospect of yielding better results for one’s consumers would bring new business and a great opportunity to increase wealth. This is one of the pervasive themes that bind together the numerous articles we reviewed that were similar to our problem and methodology.

One of the main aspects of this predictive model, or any predictive analytics problem, is determining the model which can best be used to achieve optimal results for the problem the model is trying to predict. The articles we reviewed show a wide variety of usage of models to try to solve this problem and various reasons for why they were chosen as the best fit. One study looked to see how tweets and other social media posts including message boards lead to a predictor of stock market prices from the Standard & Poor’s 500 index (S&P 500). This paper uses two methods, the first being a sentiment analysis on the social media posts which have been accumulated from various sites around the internet. The researchers accumulated these messages and them attributed them to individual stocks. Not all stocks were attributed the same number of messages, so the researchers adjusted the sentiment analysis based upon average number of messages. Finally, the team created a sentiment predictor which compared the amount of positive and negative messages to the 30-day simple moving average (SMA 30) to predict future stock price changes for the days after the messages had been posted. Another team had a similar approach in that they viewed how marketing affected future stock predictions. These researchers created a CLV model which estimates the timing and value of future cash flows but does not account for the inherent risks associated with the future stream of cash flows from customers. “The corresponding lifts in stock price for both companies was predicted and then compared with actual stock prices over the nine-month duration of the field experiment. The implication for marketing is to account for cash flow risks when estimating the future cash flow streams from customers”. The last team we viewed team used two models to try to predict future potential of the three biggest stocks in the Indian stock market. The researchers used the statistical tools of IBM’s SPSS and Neural Works Predict to build their models and get their results.

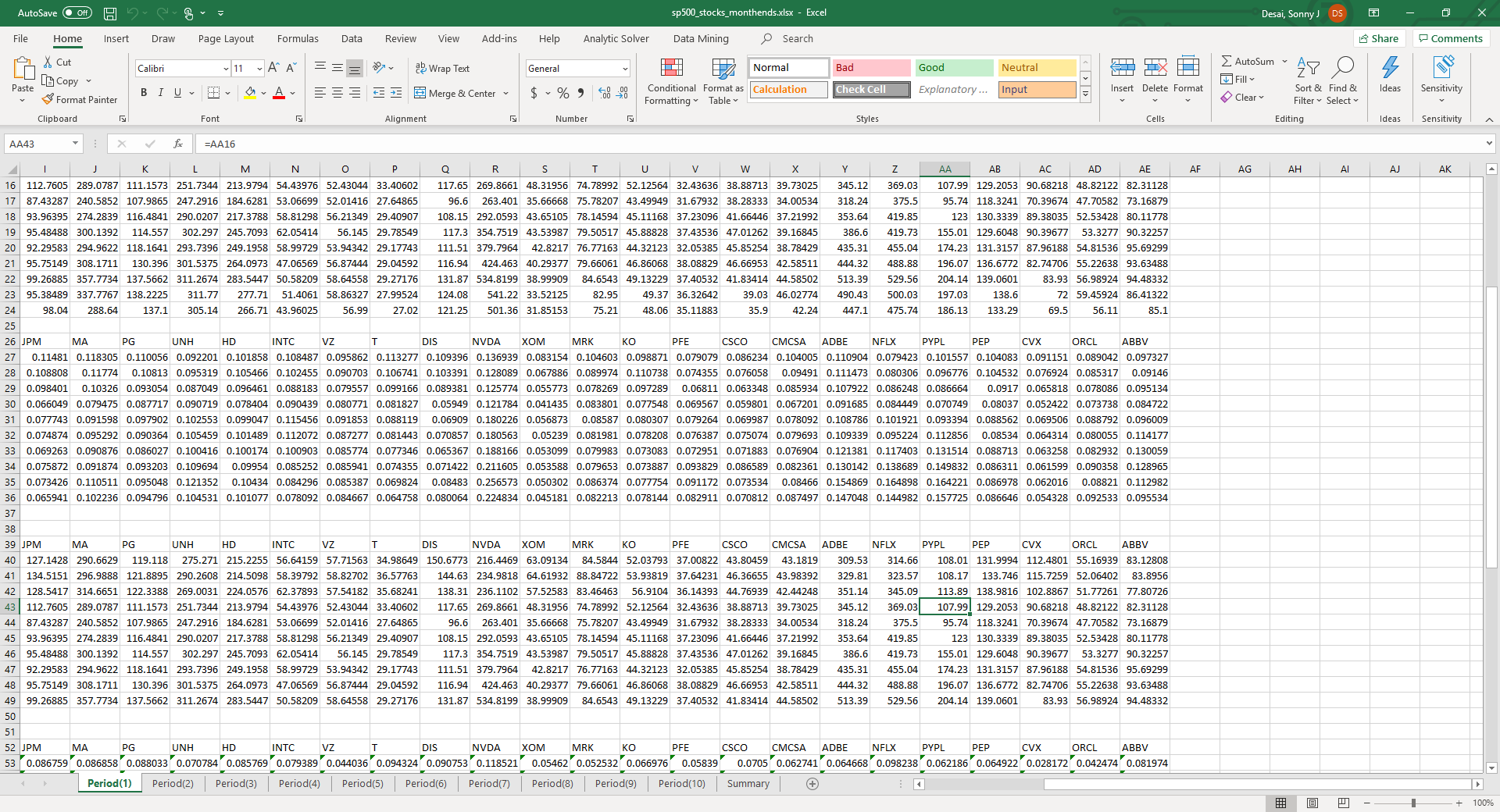
These three other teams of researchers that we viewed had similar models to ours in that they were all viewing the affects of predictive models on predicting future stock behavior. Two of the models used external factors to view how this affect the predictions of their models, but they all show how predictive analytics are sued to make these predictions. The first team found that social media posts did have an affect on future predictions of stock returns. Specifically, they found that 60 percent of all stocks which sentiment had significant prediction ratios in the past delivered correct predictions in the periods considered. The advantages here was the use of an external factor in predicting stock returns. This allowed the team to focus in on the sentiment analysis as a tool to help them predict what the stocks will do. The drawback is the assumptions sued for the sentiment analysis. Although it seems that most people would be privy to comments about stocks they are seeking to trade on, it is uncertain how they are co-related. By using different comments from different sources, bias and other factors might not be properly attributed to the outcomes of their predictions. Having a proper understanding of bias would have led to better results. In addition, they returned a 60 percent correlation rate, which is good but still relatively low in making stock predictions.

The second team also found an external factor was a factor in future predictions of stock performance. This team found that expenditure on marketing had a positive impact on stock predictions and performance, especially in comparison between their competitors. They “compared the predicted values with the actual stock prices and found that during the nine-month observation period, the marketing strategy outcomes (as indicated by the increase in CLV) corresponded to the actual stock price movement for both companies within an error range of 12% to 13%. An important implication of this result was that the marketing department was actually able to quantify the increase in stock price based on the performance outcomes of marketing strategies’. Again, as was the case with the first team, the advantage Is having a specific factor to base stock predictions on. This really helps narrow the scope for the predictive model and allow for more concise analysis. The drawback with this team was again with assumptions. How did other factors affect the eventual results that the team received? Could other economic factors have been instrumental in affecting the prediction of the stock price? These themes were not properly examined by the research team and would have to be accounted for in the overall analysis of the team.

The last team had the most analogous comparison of models and problem set between their research and ours, however their market research was conducted upon Indian stocks in the Indian stock market. This team, however, found that their theory was not accepted following the examination of their results. “The analysis clearly indicates that the stock prices of all the three sample stocks (RIL, TCS and HDFC Bank), varied widely, in tune with price variants, namely opening price, high price, low price and closing price during intra-day transactions, during the study period, on all parameters of descriptive statistics, used in this study. Hence the hypothesis H1 (There is no stochastic movement between the stock prices of Reliance”. Industries Limited, Tata Consultancy Services Limited and HDFC Bank Limited during the study period), is not accepted”. This team, unlike the first two, took various other external factors into account in analyzing the results of their model. Their results were varied, and thus more data and testing would need to be done on the model, but the main drawback in comparison are the difference between the American and Indian stock markets. Although functionally similar, the inherent culture and socioeconomic difference between the two countries would have to be further examined in order to make a like-for-like comparison.

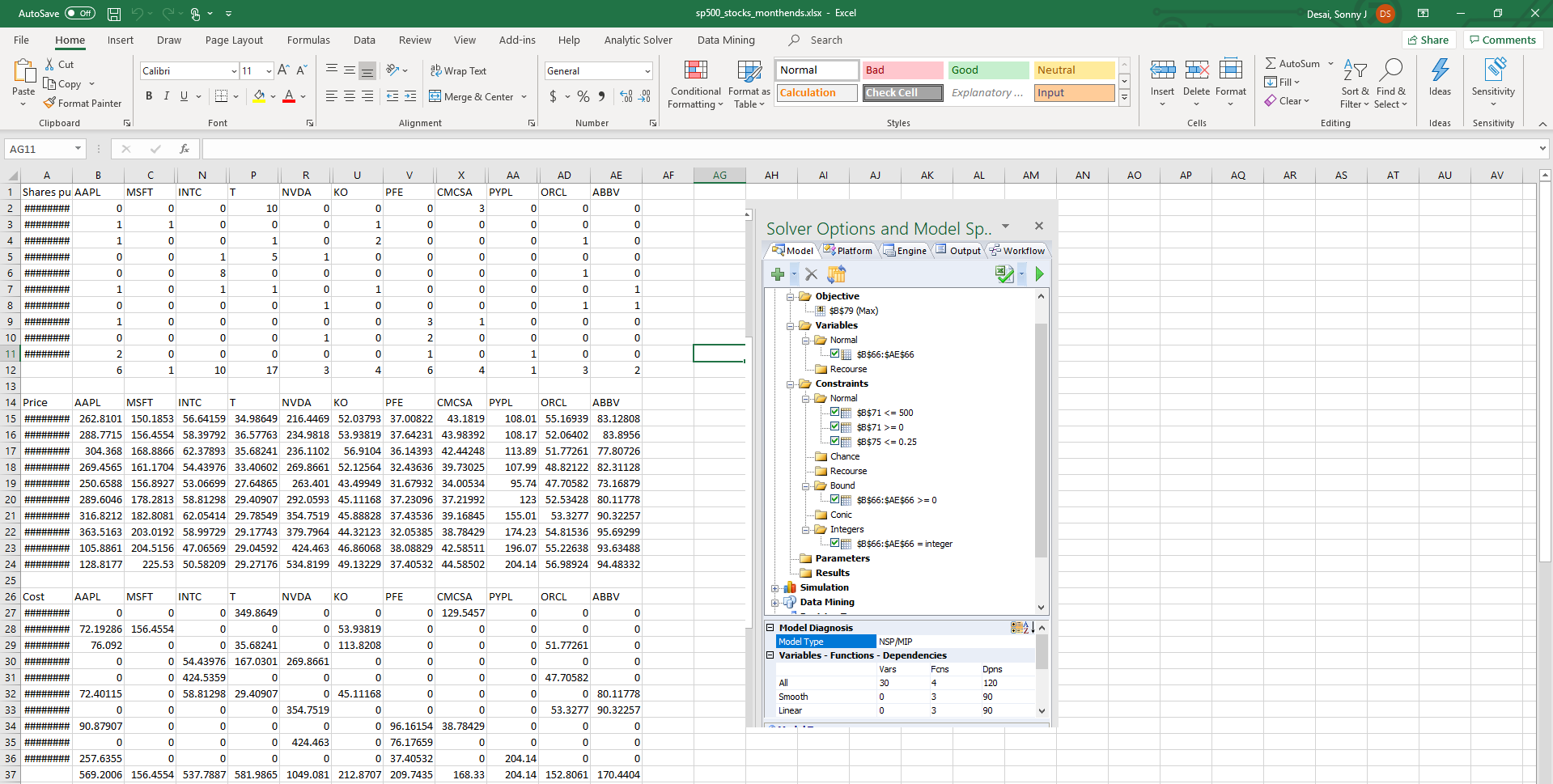
1. **Methodology**

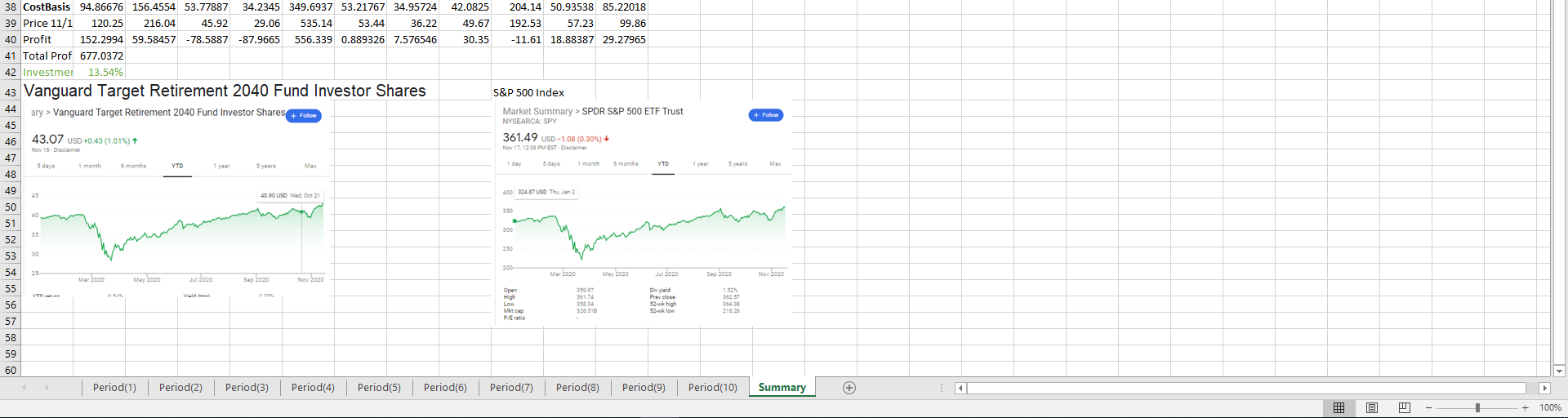
Our team decided to build a model using the Markowitz mean variance model. The Markowitz mean-variance variance model models the rate of returns on assets as random variables. The goal here is to choose the portfolio weighting factors optimally. In the context of the Markowitz theory an optimal set of weights is one in which the portfolio achieves an acceptable baseline expected rate of return with minimal volatility. Our statement of objectives is that by using the aforementioned Markowitz model, we will be able to predict future return and make the best mix of stocks by maximizing return, while controlling volatility. For the preparation of our data we gathered data for 12 months prior the current data and over a period of 2 years. It is a rolling 12 months of historical returns in which we viewed 30 Stocks performance data over 2 years. We also included associated volatility data for these stocks. The other data we researched and added to our report was the number of stocks purchased per month, the monthly budget for stock purchases; the month end closing cost of stocks; and finally, the previous day closing price of stocks. Our data parameters were built around Roth IRA contributions over a calendar year, so we have a max dollar limit over the course of a year ($6,000). Each month was calculated with $500 in savings added to the fund, starting with January with $500, February the sum being $1,000. Stock selections were made from studying their performance in 2020 Alone. We also had the desire to keep volatility low and with a focus on investors who are only Interested in long term gains. Our decision variable was that the dollar amount to invest in each stock was validated by an upward trend in performance. Our assumptions were that a year's worth of historical performance is good enough to make a selection each month. The second is that an investor will want to keep the volatility low because it's a retirement account (measured by variance). The constraints we had are: the stocks included in this portfolio are ones with prices low enough to be Acquired for $500, there is only one portfolio, the portfolio return should not be lower than a certain value, we only selected the top 30 Stocks in the Market, the initial month starts with a zero balance, and that succeeding months roll over the balance of the previous month. To construct and execute this model, we loaded the data from the 30 stocks into Microsoft Excel and split the data into 10 periods. We then ran the Markowitz mean model using Excel’s Analytic Solver with the aforementioned constraints factored in.



1. **Computational Experiment and Results**

Our team successfully ran the Markowitz mean model and found that the stock mix we found had an optimal return of 3% Return on Investment which beat both the S&P 500 and Vanguard 2040 targets dated retirement fund. Our findings were broken down into returns for each individual stock by showing the amount of stocks purchased, their price, their cost, and cost basis.



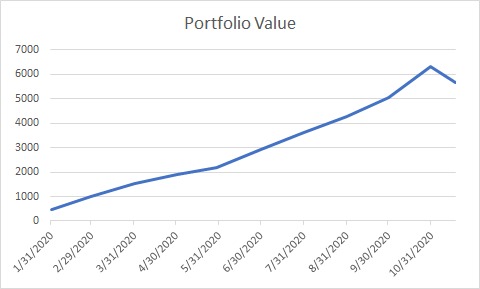


Graphical user interface, table

Description automatically generated

1. **Discussion and Conclusions**

As mentioned previously, the results to our model and problem set was a finding of 3% return on investment for the 30 stocks we viewed for the period of time that we had in our rolling 12 month range. Further analysis shows that although a nice result our project would be subject to these factors: stock split, inflation/deflation, trade wards, the pandemic outbreak and economic crash. In addition, we adjusted AAPL's price within the sheets for it's August stock split but if there were others we might have missed them. W only included stocks from the S&P 500 with a price low enough that you could conceivably purchase from 500 a month savings starting with $0 balance so some major stocks were not bought, with Amazon being the primary example. WE further concluded that our findings comparing our results to the S & P 500 and Vanguard Fund, might be an unrealistic comparison because the targeted funds have bonds and other products to lower volatility, so their returns are understandably less.



Chart, line chart

Description automatically generated

Our results proved a successful optimization of stock picks to lead to a 3% return, however future renditions of this project will likely add some improvements. We noted that we needed to research more data sources, have a more granular study on data for over 2 years, we can apply machine learning techniques for preparing data and usage in data analysis and lastly have a larger sample of available stocks.