

Advanced Portfolio Management
Using
Markowitz Portfolio Theory

by

Souvik Kundu
19141015

Mustaqim Khan
19141013

Faisal Rahman
19141014

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
Brac University
December 2019

© 2019. Brac University
All rights reserved.

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Faisal Rahman
19141014

Souvik Kundu
19141015

Mustaqim Khan
19141013

Approval

The thesis/project titled “Advanced Portfolio Management Using Markowitz Portfolio Theory” submitted by

1. Souvik Kundu (19141015)
2. Mustaqim Khan (19141013)
3. Faisal Rahman (19141014)

Of Fall, 2019 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on December 24, 2019.

Examining Committee:

Supervisor:
(Member)

Mahbubul Alam Majumdar, PhD
Professor and Chairperson(CSE), Interim Dean, School of Sciences
Brac University

Program Coordinator:
(Member)

Md. Golam Rabiul Alam, PhD
Associate Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Mahbubul Alam Majumdar, PhD
Professor and Chairperson(CSE), Interim Dean, School of Sciences
Brac University

Abstract

A portfolio is a collection of stocks made from different companies, the number of stocks can range from 10 to 30 depending on expected return by the investors. Portfolio management is finding the right group of stocks to invest in with detailed risk and return assessment. Finding the right combination is easier said than done, we opted to work with S&P 500 data set. We filter the data set picking the top 10 stocks from each sector based on different criteria, using Markowitz portfolio theory, we generate random portfolios and compare between them on the basis of Volatility and Sharpe ratio, a ratio generated from return and risk. When plotting all the portfolios a curve is generated called the efficient frontier from which we can select an optimum portfolio based on volatility and return. We then compare our generated portfolio which is dynamic based on the requirements by looking at the most recent stock market data and determine the accuracy for future prediction.

Keywords: Portfolio management; S&P 500; Markowitz model; Sharpe ratio; Efficient frontier; Volatility; Expected return.

Acknowledgement

Firstly, all praise to the Great Allah for whom our thesis have been completed without any major interruption.

Secondly, to our advisor Mr. Mahbubul Alam Majumder sir for his kind support and advice in our work. He helped us whenever we needed his help.

And finally to our parents without their throughout support it may not be possible. With their kind support and prayer we are now on the verge of our graduation.

Table of Contents

Declaration	i
Approval	ii
Abstract	iii
Acknowledgment	iv
Table of Contents	v
List of Figures	vii
List of Tables	viii
Nomenclature	ix
1 Introduction	1
1.1 What is portfolio management?	1
1.2 Key Elements	1
1.3 Why Is Portfolio Management Important?	2
1.4 Benefits of Portfolio Management	3
2 Related Work	4
3 Dataset	5
3.1 Selecting dataset	5
3.2 How S&P 500 works	5
3.2.1 Market Capitalization	5
3.2.2 Blue Chip Stock	6
3.3 Sectors	7
3.4 Understanding the Data	8
3.5 Data Collection	9
3.5.1 First Stage	9
3.5.2 Second Stage	9
4 Markowitz model	10
4.1 Markowitz Model	10
4.2 Assumption of Markowitz theory	11
4.3 Data Training	12
4.3.1 Expected Return	12

4.3.2	Market Risk	12
4.3.3	Standard deviation	13
4.3.4	The Efficient Frontier	14
4.3.5	Sharpe Ratio	14
5	Proposed model	16
5.1	Data Segmentation	16
5.2	Clustering Stocks	17
5.2.1	Annual Return	17
5.2.2	Sharpe Ratio	18
5.2.3	Market Cap	19
6	Implementation	21
6.1	Clustering	21
6.2	Markowitz Model	21
7	Future Plans	24
7.1	Alternate Datasets	24
7.2	Incorporate Blue Chip	24
7.3	Implementation of LSTM for prediction	24
8	Analysis and comparison	26
8.1	Annual Return Based	26
8.2	Individual Sharpe Ratio	26
8.3	Market Capitalization	27
9	Hybrid Model	28
10	Validation	29
	Bibliography	30

List of Figures

3.1	Sectors of the stock market Source:[7]	7
3.2	Comparing sectors over the years Source: [4]	8
4.1	Markowitz model	11
4.2	Standard deviation Source: [10]	13
4.3	Efficient frontier Source: [8]	14
5.1	Industry Cluster	16
5.2	Annual Returns	18
5.3	Sharpe Ratio	19
5.4	Market Capitalization	20
6.1	Annual Returns	22
6.2	Two optimized clusters based on Annual Returns	23
8.1	Highest Sharpe Ratio Based on Annual Returns	26
8.2	Minimum Volatility Based On Annual Returns	26
8.3	Highest Sharpe Ratio Based On Sharpe Ratio	26
8.4	Minimum Volatility Based On Sharpe Ratio	27
8.5	Highest Sharpe Ratio Based On Market Cap	27
8.6	Minimum Volatility Based On Market Cap	27
9.1	Hybrid A	28
9.2	Hybrid B	28
10.1	Updated Value	29

List of Tables

5.1 Industry Cluster	17
--------------------------------	----

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

CAGR compound annual growth

Chapter 1

Introduction

1.1 What is portfolio management?

In general sense, portfolio management means the process of making decisions about investment diversity and policies, contenting investments to objectives and create equilibrium between risk and return. The whole idea is to find out all strengths, weaknesses, opportunities, and threats in case of making choice for certain expectations. Portfolio management is done on behalf of the demands of the investors and it's a dynamic and independent process. The epitome of this optimization is to maximize the client's expected return considering the optimum level of risk exposure.

Portfolio management basically can be classified into either Passive way or Active way. Active portfolio management opts to 'beat the market' by determining undervalued assets, often through short-term trades and market timing. Passive (indexed) portfolio management seeks to copy the broader market while it maintains costs and risks to a minimum level.

Moreover, on the basis of discretionary powers portfolio management system is of two types. A discretionary portfolio management system means the portfolio manager has the right to take financial decisions based on the requirements of the investor and monitor the entire process so that all expectations are met. On the other hand, non-discretionary portfolio management refers that the portfolio manager only works as an advisor for better return generation and in the end, all the decisions are made themselves by the investors.

1.2 Key Elements

1. **Asset Allocation** The main success factor to an effective portfolio management system depends on the long-term mixture of assets. The characteristics of assets vary significantly from each other. Some do not change in concert and some have more volatility compared to others. This process mainly tries to create an optimized cluster that will satisfy the risk/return expectation of an investor. Whenever these allocations are made the correlation between the assets is also considered so that no cluster exists with only highly correlated companies. Investors with highly aggressive profiles tend more to weight their portfolio with volatile investments expecting a higher return in the end. On the contrary, conventional investors mostly give focus on stable investments.

The main goal of this optimization is to remain prepared to deal with every scenario that can come from the investors. The model returns an optimized solution based on dynamic allocations of assets.

2. **Diversification** In case of investing the only constant factor is that it is not possible to assume or predict the winners and losers simultaneously and constantly. As a solution to this, the efficient approach is to develop clusters that provide a comparatively broader view inside a given asset class. It is very hard to determine which subset is going to perform better than others; diversification tends to gather information about returns of all of the sectors over a given period of time but with less volatility. The impact of diversification can be noticed in sectors of geography, economy, and sociology as well as in nature. So while creating an optimized portfolio for investors diversification is a crucial factor to consider.
3. **Rebalancing** This segment ensures the flexibility of a portfolio. The whole idea of making an efficient portfolio depends on adaptability. The basic constant in this field of study is that the market is never fully predictable. So, all models must be immune to as much as possible to market fluctuations. Rebalancing ensures this adaptability and flexibility. This method is used to return a portfolio back to the originally expected output at annual intervals. Prices of invested companies may vary from time to time. So, the portfolio needs to be updated according to the changes in the stock price so that the investor always gets the expected return given a certain amount of risk. If the portfolio gets stagnant there is a higher possibility that even though a portfolio may work properly now but due to market changes it will expose the investor to a much higher risk in the upcoming fiscal years. So, the annual calculation of rebalancing gives the opportunity for investors to capture higher gains and broaden the opportunity for growth in better sectors with higher potential while maintaining the risk/return ratio.

1.3 Why Is Portfolio Management Important?

There are many factors such as age, income, risk-taking capability, amount of available resources, etc. that work when we consider the financial market. The portfolio management system considers all these factors into consideration while proposing any possible cluster to an investor. It helps to monitor the risk factor as the entire mechanism works keeping “Risk Minimization” as the focus. Moreover, investors have a variety of options to reflect their preferences. The entire model is dynamic and everything is based on the requirements that are provided by the investors. So, customization and diversity are some of the main pillars of this management system. In case of, preference share or debenture where an investor might be exposed to interest rate risk or price risk of security; this system can help the investor to immunize the portfolio from the risk factors. These are the key reasons why a portfolio management system is so important and plays a vital role in the field of financial predictions.

1.4 Benefits of Portfolio Management

One of the key factors of the portfolio management system is that it helps an investor to avoid any disastrous incident. The portfolio management system not only gives emphasis on large companies but also considers all other possible factors that contribute to stock investment. For example, it's a basic tendency to invest in big companies that have huge resource values. Some investors get a sense of security investing only in big companies while not considering other factors. The financial market is very volatile. There's no way to predict the sudden changes and fluctuations. If an investor invests everything on a big company that might have high return value now but there is no way to predict that things will remain constant in the future as well. If the price of that big company drops drastically then there is no way left for the investor to overcome this because all the investments were under only one company. These types of unpredictable situations are not rare here in this business. So, diversity is very important. The portfolio management system gives huge priority on diversification so that investors remain immune to sudden market changes.

Besides diversification, reducing risk is also very important. According to the risk-return trade off theory, the amount of potential return increases with a rise in risk. The relationship between risk and return is proportional. This does not necessarily mean that one should just look at the return value and invest there. There is a threshold point for maximum risk. If this is not considered then investors are exposed to huge risk threats without any protection that can neutralize the sudden market fluctuations. The portfolio management system creates a balance between these two factors. This system creates a cluster keeping in mind the minimum risk involved and simultaneously assuring the highest return that is possible in the given circumstances.

Moreover, an investor having limited resources and expecting the maximum return from it is the biggest dilemma that is addressed by the portfolio management system. Buying stocks at random without any probable reason can never be appreciated. It never assures that an investor will get the expected output. However, the portfolio management system provides a pure structure and performs many financial calculations to create a model and help investors to decide how much return to expect and the proper layout to achieve that. This systematic approach is the most convenient way to allocate limited funds dynamically and ensure the most return. This system returns weights showing how much to invest on which company and for how long to get the desired return value while keeping the risk factor to a minimum.

Chapter 2

Related Work

Advancement in portfolio optimization has been a big issue all over these years and investors are rapidly looking for an optimal solution. Starting from selecting the right stocks to distribute resources among them are the most important issue all over the time. On the verge of fixing this issue, different sorts of research took place in a very similar manner. Such use of Generative Adversarial Network in portfolio selection which identifies market uncertainty as a prime factor. This helps to identify risk on different stocks and helps to construct an efficient portfolio minimizing risk as much as possible and increases return. [9] [1]

Another important concept which had a strong effect on our work is not having a profile for a longer period of time. Along with it, volatility is not a good measure risk and it does not strongly incorporated with market changes. Rather computing probability of failing can increase efficiency in the model. This can help in proper risk-free asset allocation and the return rate does have a better effect.[3] [2]

One remarkable progress of work has been using Bayesian Learning theory combining with Markowitz Portfolio theory where the decisions for optimizing weights comes from the Bayesian probability distribution. This data get produced from a large chunk of historical asset pricing data. An investor can utilize this portion of knowledge and updates the information by using Dynamic Programming for better optimization. Afterwards assessing those values to incorporate Sharpe Ratio for the portfolio which helps to validate more strongly.[6]

The daily return for stocks are a good way to analyse market trend also it does a greater help to pick individual stock judging their performances. Calculating risks are a prime concern for MPT. One research has used an index called ERI(Extreme Risk Index) for finding out higher risks for large scale portfolio losses and minimize the possibility for this. This has been compared with minimum variance and equally weighted portfolio. This comparison reflects how ERI selects a better combination of portfolio than traditional methods for portfolio management.[5]

Chapter 3

Dataset

3.1 Selecting dataset

Since we are dealing with stocks the first thing, we needed to do is decide on a stock market, we went with the US Stock Exchange since it's the leading stock exchange in the world. There are about 2800 companies listed trading approximately 1.46 billion shares each day.

The financial market never stops growing, which means last year's data is not good enough to build an efficient model. To collect data, we use Stock market data API's which offer real time or historical data on financial assets that are currently being traded in the markets. We use the Yahoo Finance API which has been the gold standard for some time now, it provides access to more than 5 years of daily Open, High, Low and Close price data.

As stated earlier there are about 2800 companies listed on the US Stock Exchange which is a very high number, so instead we decided to rely on S&P 500 index. S&P stands for Standard and Poor which are the founding companies is a stock market index which calculates the performance of the top 500 companies in the US Stock Exchange instead, and this is regarded by most as the best representation of the stock market.

3.2 How S&P 500 works

S&P 500 index tracks the stocks of the top 500 companies, these are also referred to as large-cap companies. Before we discuss about large-cap companies we need to understand what market capitalization is.

3.2.1 Market Capitalization

It is typically called market cap for short. It is the total value of the company measured by the product of stock price with the number of shares issued to the public. So, for example if a company has 2 million shares and each share is being sold for 10 then the market cap for the company would be 20 million. This also means you can acquire the company if you had 20 million to spend, however it still depends on whether the current stockholders are willing to sell.

The stock market is divided into 3 segments based on the market cap of companies.

- **Small cap:** Companies with a market cap of less than \$1 billion. These are smaller companies or recent startups which have started growing. They are riskier investments but they have the potential to grow and become very profitable.
- **Mid cap:** These companies have a market cap between \$1 billion to \$5 billion. Mid-cap companies are less risky however the room for growth is less as well.
- **Large cap:** The market cap for these companies are \$5 billion or more. They are the market leaders and among the biggest companies which means the opportunity for growth is minimum. They have the least risk, their returns might not be as high as other small or mid cap stocks, but they are more likely to reward stockholders with dividends.

Realistically speaking, new companies are always likely to come and join the elite Large Cap segment of the stock market. For example, Facebook did not even exist 20 years ago but now it is on of the top 10 companies with a market cap of \$500 billion plus. Whereas there are a few which have been among the top 500 companies to decades now.

So, investing in both mid-cap and large-cap companies are usually the safe options, although keeping in mind that mid-cap companies are riskier to invest compared to large-cap companies.

Here is why we should invest in Large-Cap stocks

- **Stable:** Since these companies are very large, their size makes it very unlikely that they will go out of business therefore they are a safer option compared to small-cap companies.
- **Safer in a downturn:** It would not be accurate to say that Large-cap companies are immune to recession, however they are more likely to withstand a dip in stocks without going out of business.
- **They pay dividends:** For investors these stocks become a source of income. Due to stagnant stock price they need to compensate investors and they do so with dividends from their own profits, which is feasible for such a large company.

3.2.2 Blue Chip Stock

Another terminology widely used during stock market analysis are blue-chip large-cap stocks. The “blue chip” represents the top tier companies that are among the SP 500. To clarify, being among the top does not necessarily mean they are among the top 10, the “blue chip” stocks are high performing stocks, they have a long history of stable earnings, little debt and most importantly they represent diversified business. Which means if one of their businesses has bad year, it won’t affect their stock as another business of theirs will have a good year which will make up for it. To sum up, owning a blue-chip stock give and instant diversification and reduces risk as well.

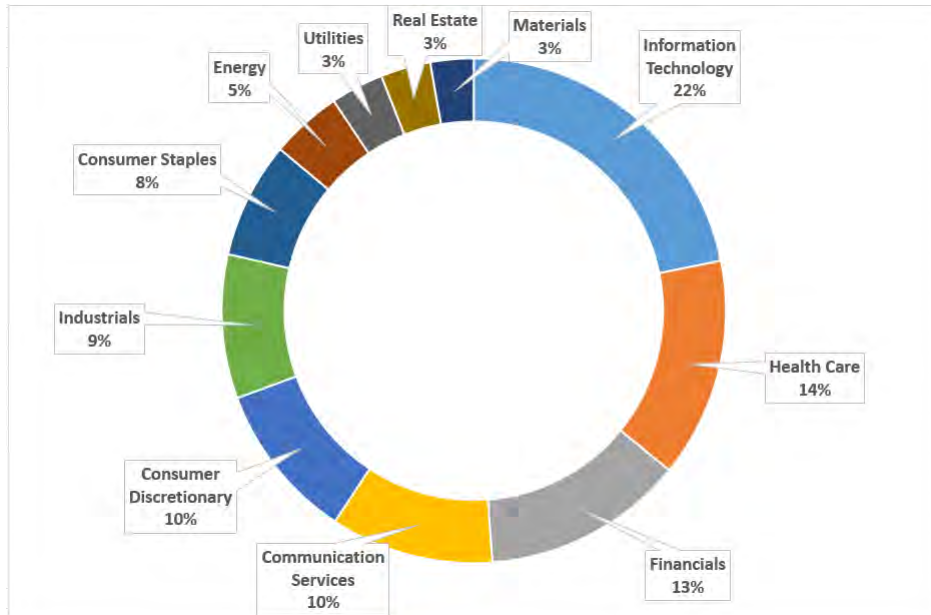


Figure 3.1: Sectors of the stock market
Source:[7]

3.3 Sectors

If we were to divide the stock market by industries then we would find 11 different sectors.

- **Information Technology**
- **Healthcare**
- **Financial**
- **Communications services**
- **Consumer Discretionary**
- **Industrial**
- **Consumer Staples**
- **Energy**
- **Utilities**
- **Real Estate**
- **Materials**

As we can see from Figure 3.1, the most dominant field is Information Technology while the least dominant would be utilities, real estate and materials. However, this was not always the case. If we look at Figure 3.2 below, we can see that during the 1990's the scenario was quite different. The tech industry only consisted of 6.3%

Historical Sector Weightings of the S&P 500: 1990 - Current																	
Sector	1990	1993	1996	1998	1999	2000	2001	2002	2004	2006	2007	2008	3/09	2009	2010	2011	Current
Tech	6.3	5.9	12.4	17.7	29.2	21.2	17.6	14.3	16.1	15.1	16.7	15.3	17.6	19.9	18.7	19.0	19.0
Financials	7.5	11.2	15.0	15.4	13.0	17.3	17.8	20.5	20.6	22.3	17.6	13.3	8.9	14.4	16.1	13.4	14.1
Energy	13.4	10.0	9.2	6.3	5.6	6.6	6.3	6.0	7.2	9.8	12.9	13.3	14.3	11.5	12.0	12.3	12.0
H. Care	10.4	8.2	10.4	12.3	9.3	14.4	14.4	14.9	12.7	12.0	12.0	14.8	16.1	12.6	10.9	11.9	11.8
Cons Stap	14.0	12.5	12.7	11.1	7.2	8.1	8.2	9.5	10.5	9.3	10.2	12.9	13.8	11.4	10.6	11.5	11.1
Industrials	13.6	13.9	12.7	10.1	9.9	10.6	11.3	11.5	11.8	10.8	11.5	11.1	9.5	10.3	11.0	10.7	11.0
Cons Disc	12.8	16.4	11.7	12.5	12.7	10.3	13.1	13.4	11.9	10.6	8.5	8.4	8.3	9.6	10.6	10.7	10.8
Materials	7.2	7.1	5.8	3.1	3.0	2.3	2.6	2.8	3.1	3.0	3.3	2.9	3.2	3.6	3.7	3.5	3.7
Utilities	6.2	5.6	3.7	3.0	2.2	3.8	3.1	2.9	2.9	3.6	3.6	4.2	4.4	3.7	3.3	3.9	3.7
Telecom	8.7	9.1	6.5	8.4	7.9	5.5	5.5	4.2	3.3	3.5	3.6	3.8	4.0	3.2	3.1	3.2	2.8

Figure 3.2: Comparing sectors over the years

Source: [4]

of the whole market whereas consumer stable dominated the stock market with 14%.

From this what we can understand is the stock market is ever changing. Industries which are dominating the market today might be overtaken by a new industry altogether in 20 or 30 years. This unpredictability of the stock market is what makes it so challenging to invest and profit from the stock market.

3.4 Understanding the Data

There are a few terminologies in stock trading that we need to be familiar with before moving forward. Whenever we fetch data for a particular company, each row denotes a particular day. There are 5 columns which are Open, Low, High, Close and Adjusted Close.

- **Open:** The price at which a particular stock opened on that day.
- **High:** The highest price of a particular stock on that day.
- **Low:** The lowest price of a stock on that day.
- **Close:** The price at which the stock closed, i.e. the last price of the stock on that particular day.
- **Adjusted Close:** Adj close is calculated after all distributions such as dividends and corporate actions such as split/bonus share are deducted.

For data analysis we use the adjusted closing price as it is filtered. If we were to use only closing prices and not take into account dividends or splits then the data would be highly inaccurate. If we were to look at a real example it would be much clearer, Money Chimp a S&P 500 company shows returns return from Jan 1980 to Dec 2012 as +3264% yet, the index only rose from 107.94 to 1426.19 or +1221%, the rest was deducted due to dividends paid and splits/bonuses.

3.5 Data Collection

3.5.1 First Stage

To request data for a particular company using the API we have to pass a ticker, a ticker is the symbol under which the stocks of a company is collected. For example, the ticker for google is GOOGL.

1. So first we need to collect the tickers of all the 500 companies listen under the SP index.
2. To do this we use a library called BeautifulSoup, it is a python library which can be used to parse or pull data from HTML or XML files.
3. We go on the Wikipedia page for SP 500 and using Beautiful soup we pull the list of tickers from the “Symbol’ column and store it in an array.

3.5.2 Second Stage

Now we have all the names of the 505 companies in SP 500 in an array.

1. We create a data frame to store the data that we will fetch from yahoo finance.
2. We implement pandas datareader package where we pass the data source as “yahoo finance”, we also mention the time frame of data we want to pull.
3. We filter the fetched data and only select the columns that we intend to work with.

Chapter 4

Markowitz model

4.1 Markowitz Model

Harry Markowitz is recognized as the founder of Modern Portfolio Theory (MTP). This theory is based on some assumptions that are derived from real-world scenarios. This model considers all possible variables while considering making a portfolio that is optimized and satisfies the criteria of investors. We have decided to go with this model because this is globally renowned and suits the purpose of our research the most.

In general, investors are regarded as risk averts. They expect to maximize the amount of profit given a minimum level of risk. Markowitz's model shines in this area. The model portrays that the risk and return characteristics don't need to separate for diverse investment options whereas we need to detect how these factors jointly contribute to the performance of a portfolio. The basic assumption of this theory is that investors only tend to take high risk only when high return is involved. In short – “High risk, high reward”.

This model basically tends to find possible clusters for investors to create an efficient set of portfolios that will ensure the expected return and will also keep in mind the minimum risk factor. This process is called creating an efficient frontier. The amount of risk an investor can take determines the type of the frontier. If an investor wants a very low-risk factor this model will create a cluster that will always ensure that the investor gets the expected return with the minimum risk. For sure no Utopian value can be achieved with this model. However, this model always ensures the highest possible value of return with minimum volatility that possible in the given market and immunizes the investor to sudden market fluctuations. The entire model is dynamic and totally justifiable according to the requirements of investors. Diversity and optimization are some of the key features of this model.

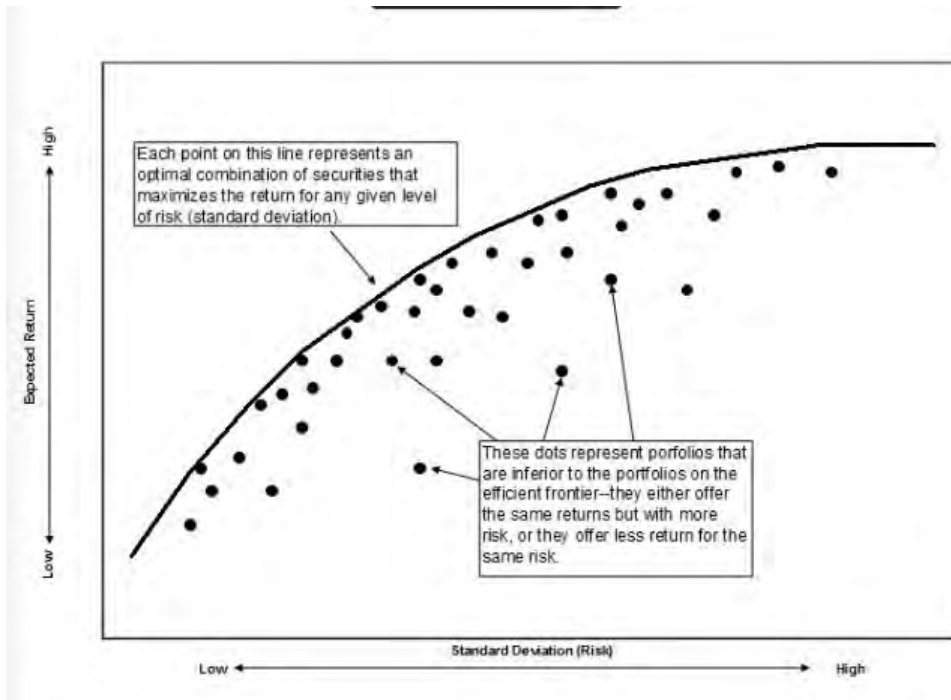


Figure 4.1: Markowitz model

4.2 Assumption of Markowitz theory

There are some basic assumptions that play a vital in the case of this model. Keeping these assumptions into consideration this model returns-optimized and dynamic clusters to the investors. Some basic assumptions are listed below:

1. It is assumed that all investors are rational and their manner of behaving reflects that they expect to maximize their utility within a given level of resource.
2. Free access to fair and correct information regarding exact return and risk are available to all investors.
3. Given any market, it is assumed that the market is efficient and absorbs the information fast and perfectly simultaneously.
4. The basic nature of investors is that they want to adverse risk while maximizing the return keeping the risk level to a minimum.
5. The decisions of investors are based on the expected returns and on the standard deviation of these returns from the value of the mean.
6. On any given day investors will pick higher returns to lower return for a certain amount of risk.

So, these are the basic assumptions that are considered under this model. These assumptions are relevant to the real-world scenario so that optimized portfolios can be created.

4.3 Data Training

Under this model, the aim of reducing risk is done by using the concept of diversification, which reduces the value of variability or standard deviation. However, this is not the only importance of diversification for this model. Diversification also helps to reduce the co-variance or interactive risk among the securities in a portfolio. Co-variance means how strongly correlated companies are in a portfolio. Highly correlated companies are never encouraged to be together in one cluster as this significantly increases the risk factor. The model opts to get the value of the standard deviation closest and if possible to zero and the coefficient of correlation to -1(negative) so that the overall risk of a given portfolio is negligible. However, achieving these values is practically impossible. As a result, the model does several permutation and combinations so that we can achieve optimal clusters that satisfy the most.

4.3.1 Expected Return

The basic idea of expected return is the amount of profit or loss that an investor can expect from a particular investment. This value is calculated by multiplying potential outcomes that involve both profit or loss and adding them together. For example, investment has a 60% chance of gaining 20% and a 40% chance of losing 15%. Then the expected return for this given scenario is $(60\% \times 20\% + 40\% \times -15\% = 6\%)$. The main purpose of calculating this value is to find out whether an investment has a positive or negative net outcome. The value is dependent on historic data and it does not provide full assurance and thus other factors are also considered along with this value for finding out the optimized cluster for an investor. Moreover, there can be scenarios where the value of the expected return can be the same for several clusters. So, it is never wise to get to any conclusion only based on expected return. We need to consider other factors like standard deviation, co variance, risk percentage, etc. to decide the end result. However, initially, this value helps to sort out the possible companies to invest in.

4.3.2 Market Risk

Market risk is the chance that an investor might experience a loss due to the factors that affect the overall performance of the financial market in which the investor is involved. This kind of risk is also regarded as systematic risk and this risk cannot be eliminated by diversification or other methods and this is very hard to predict how much this will affect a market in a given amount of time. However, there are ways that can be used in order to hedge this. Common examples of market risk are recessions, natural disasters, political turmoil, the sudden change in interest rates, etc. These incidents are independent and there is no way to prevent these to happen. However, the Markowitz model also keeps this into consideration. This model gives the chance to investors to use hedging strategies like put options and index options to protect against the volatility and market risk.

4.3.3 Standard deviation

Standard Deviation is one of the most important terms in the area of statistics. For any financial modeling and prediction understanding the underlying theory of standard deviation is very important. In general senses, this is a measurement of the amount of the variation or dispersion among the set of values. The less this value is means that the value is closer to the actual mean of data entries given. Standard deviation can also be described by variance. The square root value of variance of a data set is called standard deviation. So, standard deviation and variance term are correlated. This value is traditionally expressed using the Greek letter Sigma (σ). In order to calculate first of all we need to find the mean value. After that we find the difference of each data entry and the mean value and square the value. This process is repeated for every single data in the set and the summation of all these squared different values is divided by the size of the data set. This gives us the value of variance. If this variance value is square rooted then we get the desired value of Standard Deviation (SD). For the implementation of Markowitz Model, we need this concept. For, generating the efficient frontier we need to plot standard deviation against expected return so that we can generate a portfolio than ensures a certain amount of return within the minimum risk expectancy.

$$\sigma = \sqrt{\frac{1}{N}[(x_2 - \mu)^2 + \dots + (x_N - \mu)^2]} \quad (4.1)$$

Where:

$$\mu = \frac{1}{N}(x_1 + \dots + x_N)$$

$$S = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (4.2)$$

$$S^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 \quad (4.3)$$

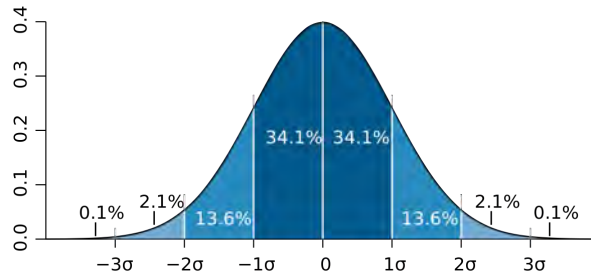


Figure 4.2: Standard deviation

Source: [10]

4.3.4 The Efficient Frontier

It is the graphical representation of portfolios that maximize the return values given a certain amount of risk. The common way of plotting is to represent the annual return on the y-axis and risk on the x-axis. The standard deviation value is regarded as synonymous with the risk factor. The main goal of an investor to populate the portfolio to ensure maximum return and combined less standard deviation value compared to individual standard deviation values. The less co variance among the securities inside a portfolio then the less standard deviation. If this process of combining risk versus returns functions properly then the portfolio should align along the efficient frontier line. The efficient frontier line shows the optimal portfolios that investors can choose to fulfill their requirements. This line also gives us the portfolio that ensures minimum risk with optimal return as well as the highest possible return cluster with more risk. So, the investor has a wide variety of options to pick upon the volatility and risk value. This line helps to make the entire model more dynamic. From this line, we can pick any cluster and assure the investor about the end output. However, this line is drawn assuming the fact that asset returns always follow the normal distribution, which is not always the case in reality. Asset returns sometime follow the leptokurtic or heavy-tailed distribution. So, this is a drawback of this graphical representation. As a result, even though it is a critical part of this model we consider other factors while choosing the optimum cluster.

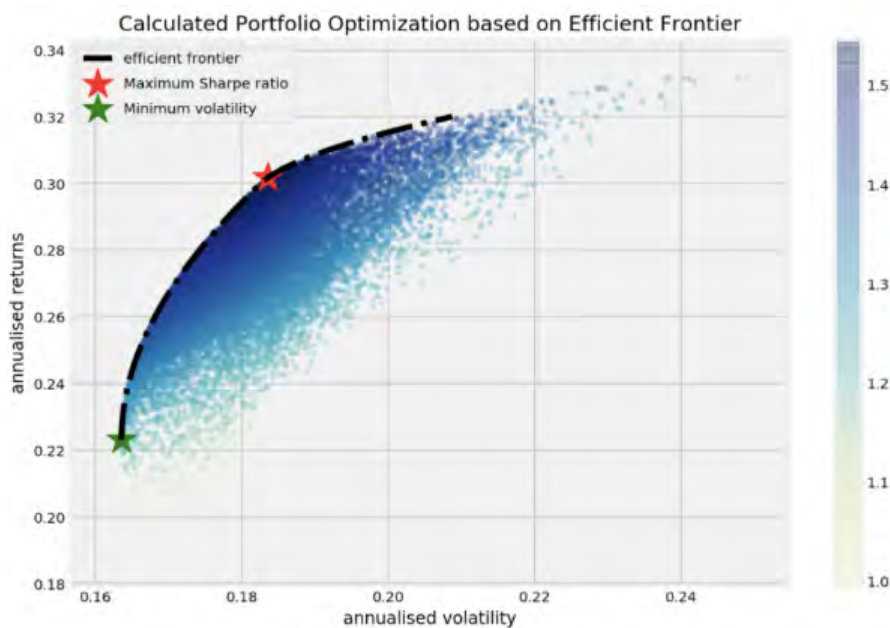


Figure 4.3: Efficient frontier

Source: [8]

4.3.5 Sharpe Ratio

The idea of the sharp ratio is related to the concept of risk-adjusted return. Risk-adjusted return refines measures how much risk is engaged in the process of generating a return and this is expressed as a number or rating. This unique idea of ratio was given my William Sharp who won Nobel Memorial Prize in Economic Sciences.

This ratio shows how much excess return can be generated for the extra volatility that one endures for holding a riskier asset. In general, the ratio represents the average return that one can achieve more than the risk-free rate per unit of volatility. This risk-free rate varies from market to market and it is determined under a certain period of time. The formula for calculating Sharpe Ratio:

$$\text{Sharperatio} = \frac{R_p - R_f}{\sigma_p} \quad (4.4)$$

Where:

R_p = Return of portfolio

R_f = Risk free rate

σ_p = standard deviation of the portfolio

The Sharpe ratio can be used to interpret the past performance of a portfolio where actual returns are used in the above-mentioned formula. In general, a better sharp ratio represents a better risk-adjusted return. If the sharp ratio is negative then it either means the risk-free rate is greater than the portfolio's actual return or somehow the return of the portfolio is expected to be negative. Neither of these scenarios is expected. So, in most cases, we opt for the better Sharpe ratio value considering a certain volatility level.

Chapter 5

Proposed model

5.1 Data Segmentation

In the very first place from the Wikipedia we collected the updated SNP 500 list that consists of 505 companies. However, all these companies fall under some unique industry. In order to, create an optimized portfolio it is never wise to consider all 505 companies. So, initial sort is necessary. Under our model sorting all these companies according to their respected industries did the first filtering. After this we got 11 unique industries under which all 505 companies fall. The industries are: Industrial, Health Care, Information Technology, Communication Services, Financial, Utilities, Consumer Discretionary, Real Estate, Consumer Staples, Energy and Materials. Each of these industries has several numbers of companies and this is the first segmentation of our data set. Apart from sorting, another idea behind this segmentation is that we want to have equal amount of companies from all industry sectors before feeding the data to the Markowitz Model. The main basic ideology of our model is that it needs to be dynamic so all industries should be given equal priority. The below table shows how many companies fall under each industry.

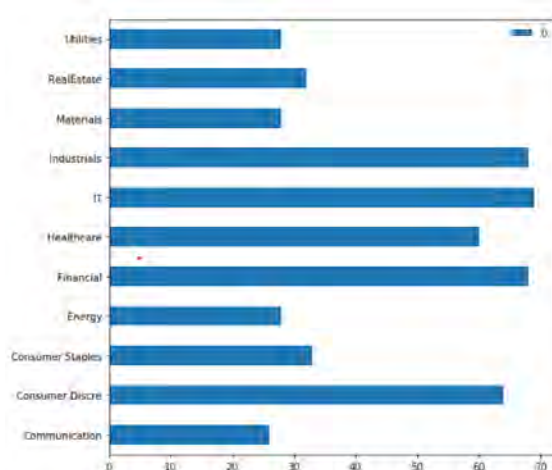


Figure 5.1: Industry Cluster

Industry Name	Number of companies
Industrial	68
Healthcare	60
Information Technology	69
Communication Services	26
Financial	68
Utilities	28
Consumer Discretionary	64
Real State	32
Consumer Staples	33
Energy	28
Materials	28

Table 5.1: Industry Cluster

5.2 Clustering Stocks

Clustering Stocks: Clustering of financial stocks are a bit different compared to other traditional clustering methods. There are many variables method so it is tough to decide one single clustering feature. So, for developing our model we decided to cluster under three criteria and perform comparison to portray the best option in a given scenario. The criteria are:

5.2.1 Annual Return

Annual return means the return an investor expects over a period of time and this value is showed as a time-weighted annual percentage. This represents a geometric mean rather than just arithmetic mean. This value is very helpful when we want to see how an investment has performed over time and help to compare two or more investments.

Example Annual Return Calculation:

$$CAGR = \frac{EndingValue^{\frac{1}{year}}}{Beginningvalue} \quad (5.1)$$

Where:

CAGR= compound annual growth
 Years= holding period, in years

Calculation Process:

1. Percentage change of each stock during a given period of time
2. Calculating mean of the percentage change
3. Regarding 250 as working days for a fiscal year
4. Multiply the mean value percentage by 250
5. Sorting the companies based on annual return for each industry

After completing this sort 5 companies were selected from each industry. So on total we received 55 companies based on the highest annual return value and created a separate cluster just based on this so that we can run Markowitz model and find the most suitable cluster. However, choosing 5 companies from each industry is not some constant value. It totally depends on the investor. If the investor wants to invest on more companies then this value is updated dynamically and based on that a new whole cluster is created. For simpler understanding we have decided to go with 5 companies from each unique industry.

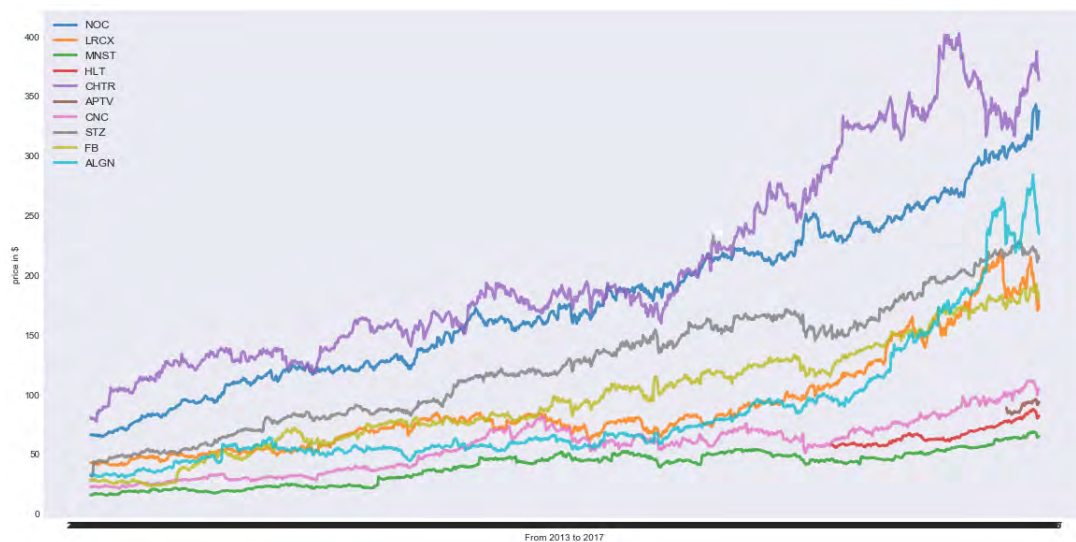


Figure 5.2: Annual Returns

The above diagram shows companies that have highest annual return value and the change of price over time.

5.2.2 Sharpe Ratio

This ratio shows how much excess return can be generated for the extra volatility that one endures for holding a riskier asset. In general, the ratio represents the average return that one can achieve more than the risk-free rate per unit of volatility. This risk-free rate varies from market to market and it is determined under a certain period of time. It is one of key factors from which we can determine how much return we can expect from one investment and how much risk factor is involved in this process.

Calculation Process:

1. Percentage change of each stock during a given period of time
2. Calculating annual return of a stock
3. Consider risk-free factor which is constant for a given market
4. Subtract the risk-free factor from return of a particular stock
5. Divide the subtracted value by a stock's individual standard deviation value.

After calculating individual Sharpe ratio values we sort the cluster based on this and pick 55 companies jointly from all the industries and do the same process and feed this cluster to Markowitz model.

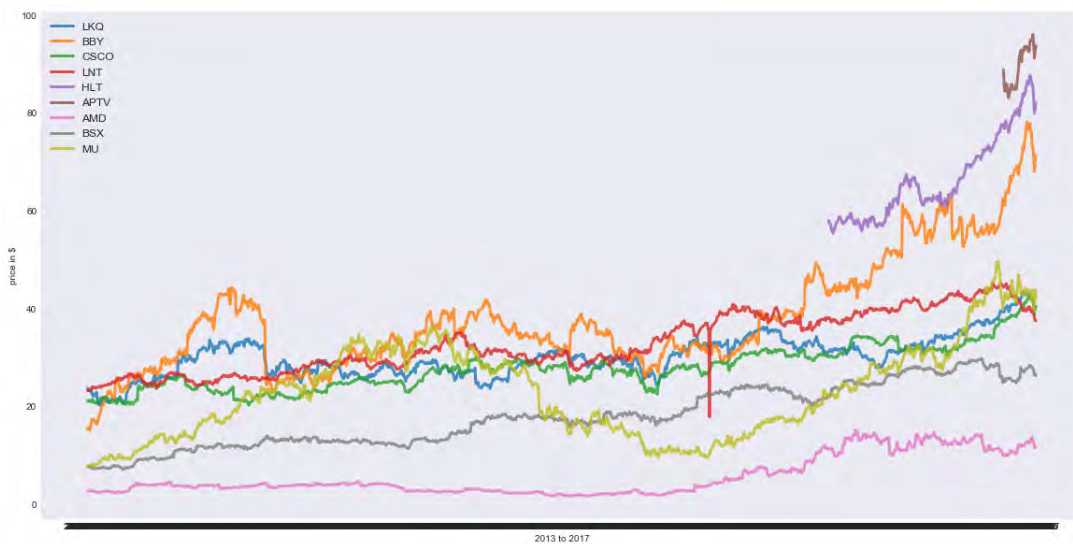


Figure 5.3: Sharpe Ratio

5.2.3 Market Cap

Market Cap is defined as the total market value of all outstanding shares of a company and this value represents the worth of a company in a particular market. In order to calculate this we multiply the value of one share with total number of outstanding shares. There are three classifications based on market cap. They are: Large-cap (\$10 billion or more), Mid-cap (\$2 billion to \$10 billion), and Small-cap (\$300 million to \$2 billion). Since for this model we are working with SNP500 dataset; it mostly covers the large-cap companies. Even though market cap is mostly used to describe a company; it does not necessarily measure the equity.

Calculation Process:

1. Determine price of one share of a particular company
2. Find out numeric value of all outstanding shares
3. Multiply price of one share by number of total outstanding shares
4. Sort companies based on market cap

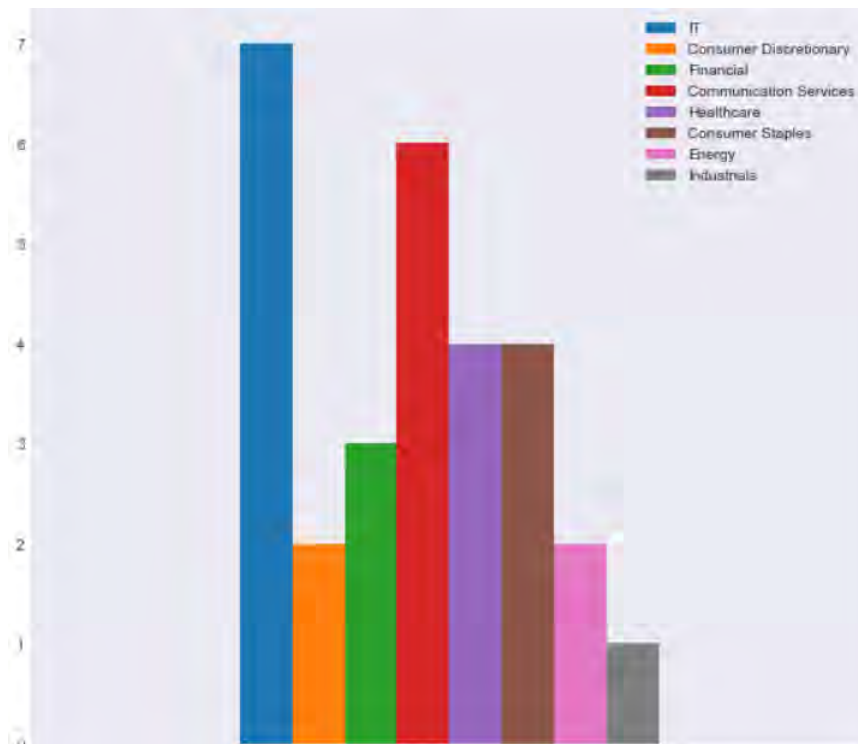


Figure 5.4: Market Capitalization

Chapter 6

Implementation

6.1 Clustering

In S&P 500 stock market, for ensuring equal ratio of stocks from all industries we clustered the stocks based on the three proposed models described before. Those are, i) Annual Return; ii) Individual Sharpe Ratio; iii) Market Capitalization. Using each method we have created a cluster of total 55 stocks which get computed by Markowitz Model to calculate the efficient frontier. Among these 55 stocks, we again filter the data based on their performance over a certain period of time. Then we can select how many stocks a single portfolio will be consisted on and this totally depends on the investor on how much capital he is willing or capable to invest.

6.2 Markowitz Model

The revolutionary model by Harry Markowitz has an amazing way of calculating. Markowitz model takes prices for each stock over a certain period of time as an input. The input should have the stock tickers as column values and each row represents adjusted closing prices of each stock. Then the model has certain steps which calculates the weights each stock should be assigned and also returns the return rate with Sharpe ratio and volatility. All of the values come as percentage. The model does the calculation by:

1. It generates random weights
2. The sum of the weights are always one
3. Calculates return = `np.dot(weights , annual_return)`
4. Calculates volatility = `np.sqrt(np.dot(weights.T , np.dot(annual_covariance , weights)))`
5. Then it calculates the Sharpe ratio for the whole portfolio, Sharpe = returns / volatility
6. This whole loop is iterated over 50k times so we can get as much as possible portfolios

7. From all these we can find the best portfolio with highest Sharpe ratio and minimum variance portfolio

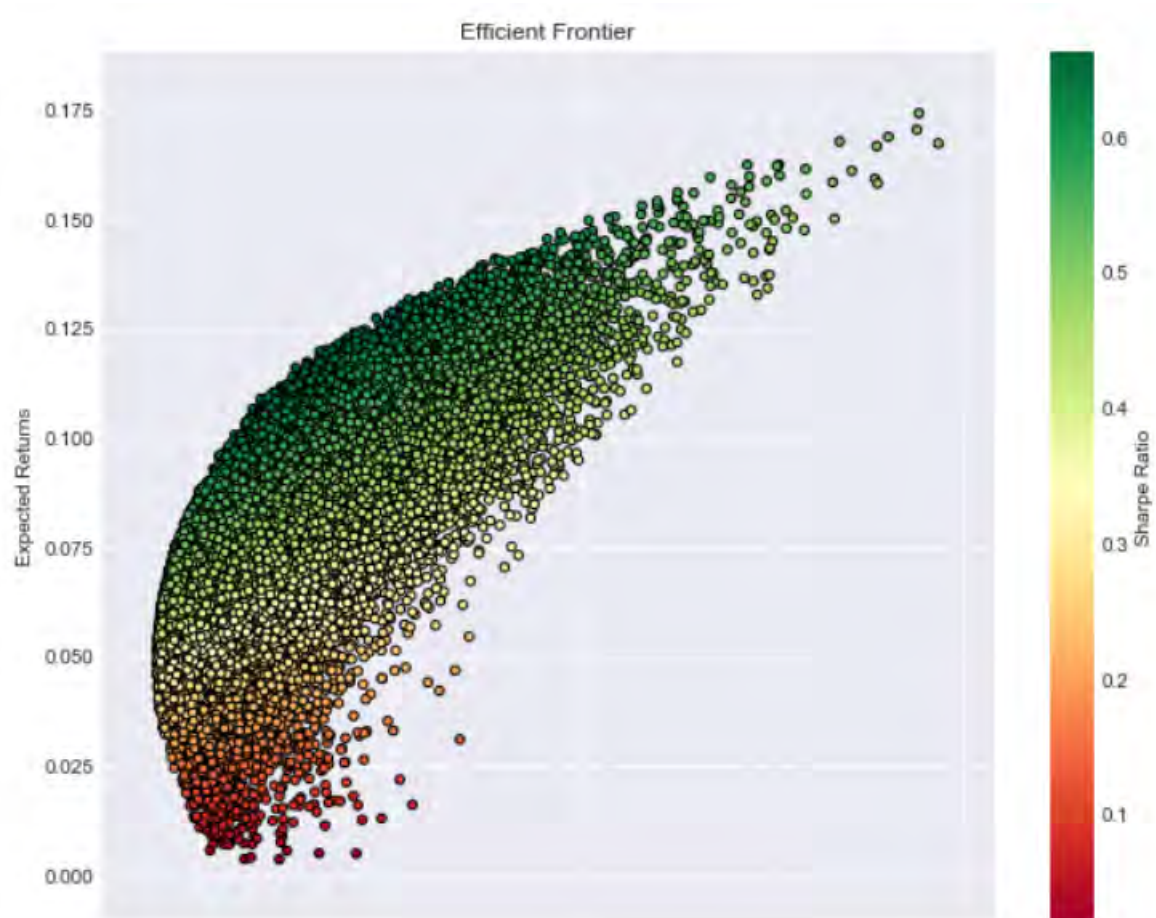


Figure 6.1: Annual Returns

The whole concept is to generate a large number of combinations where each of the scatter points are denoted as a single portfolio. Every single portfolio is generated by computing according to Markowitz portfolio theory. Afterwards they are plotted by expected returns and the and risk. As a result, as an example of an ideal portfolio it came as a bell shaped curve where the more it increases towards y-axis means the returns are increasing and the x-axis shows how risky the portfolio is. According to the concept the higher the risk means higher the return but that is not always the case. That's why the whole plot is colored according to their characteristics. For the bell shaped feature the red ones shows those portfolios are risky as well as have bad returns. Then the portfolios turned into yellow colour and then into green which shows the portfolios have higher returns and than previous ones. The more

upper we move the higher returns become. The darker the green colour gets, the higher the Sharpe ratio becomes. Afterwards the more we move to the right the return does not increase by a large quantity but the risk increases. As a result, we have to find the best combination among those.

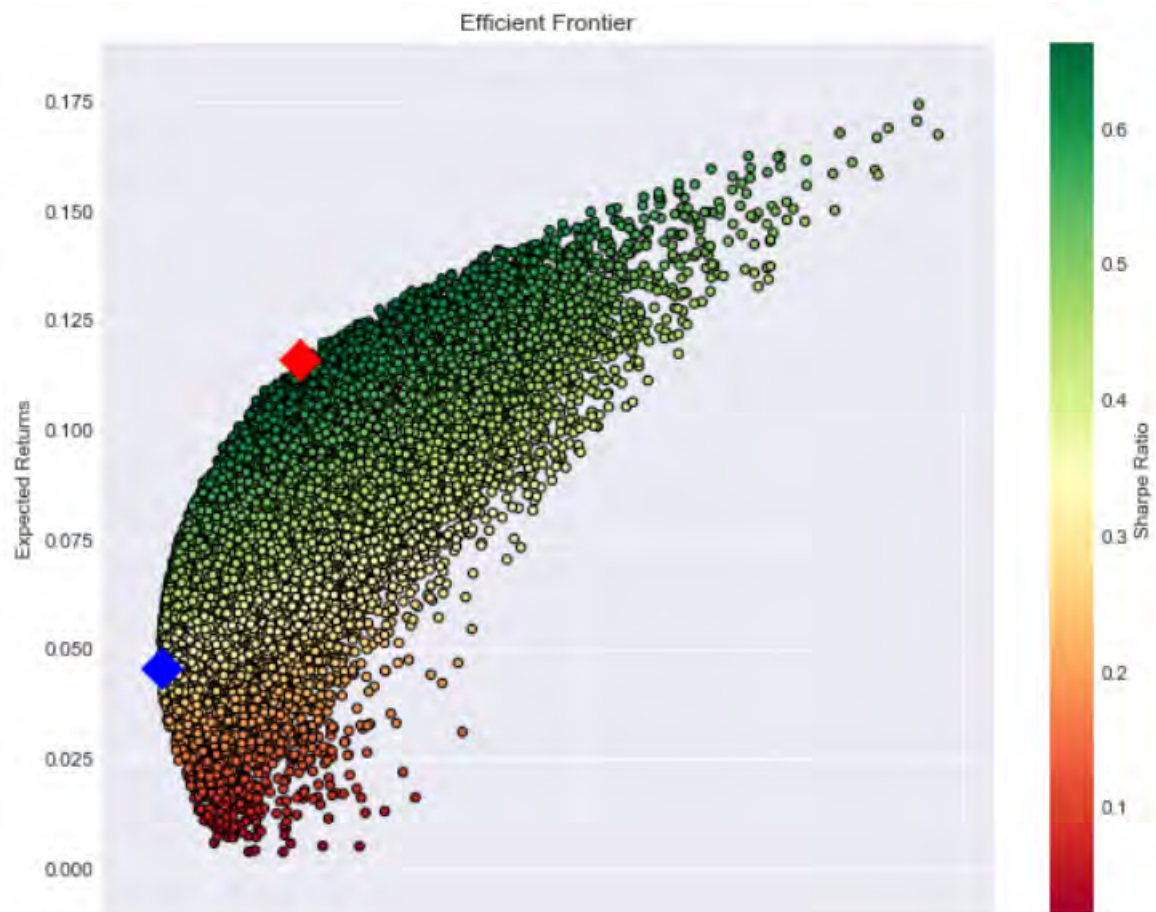


Figure 6.2: Two optimized clusters based on Annual Returns

Looking more closely to the plots, we can see two coloured points. The red one shows the portfolio with the Sharpe portfolio and red one denotes as min variance portfolio. The highest Sharpe ratio means that portfolio has the best returns in terms of riskiness of their stocks. On the other hand the min variance means the portfolio with the highest return considering least risk.

Chapter 7

Future Plans

7.1 Alternate Datasets

In the first place, here we have trained and fine-tuned our model based on the SNP 500 dataset. This dataset is very renowned for financial predictions but we hope to work with more datasets that will increase our exposure. More exposure means we can incorporate more features and generate an optimum portfolio for all kinds of investors. Particularly we look forward to working with NYSE and Nasdaq for better understanding.

7.2 Incorporate Blue Chip

Secondly, we would like to involve the idea of blue-chip companies. The characteristics of these blue-chip companies are very different compared to traditional companies. These are usually very big multinational companies having a wide variety of businesses. It's comparatively safer to invest in blue-chip companies but they have a high stock prices in the first place and sometimes they prone the investor to more risk because of high correlation values. However, some companies provide dividends to neutralize the risk factor. So, the entire model needs to be tweaked just to implement these companies and we opt to do this in future research.

7.3 Implementation of LSTM for prediction

Finally, for validating our model we used historical data. In the cross-validation segment we predicted the weights for certain stocks based on data from 2015-1-1 to 2018-11-31. Later on, we used the weights that were predicted from this timeline in 2019 to actually see how efficient our model is in real life. However, for this justification, we always need actual data for the ongoing fiscal year. This scenario is not going to be possible in every case. So, to address this issue in the future we plan to use LSTM to predict future stock value. LSTM is widely used for this purpose and the average accuracy using the base LSTM theory is quite satisfactory. We plan to implement this in our model. So, using the historical data firstly we will find an optimum cluster and get the value of weights for each stock. Secondly, we will use the same historical data to feed in LSTM in order to predict the future stock value. Afterward, the weights will be used in the LSTM given stock value to calculate how

much return we can generate from the cluster given a certain amount of risk. This will full proof our entire idea. With the implementation of this our hybrid model is going to be immune to any market in any given timeline.

Chapter 8

Analysis and comparison

8.1 Annual Return Based

Selecting the cluster based on annual return for individual stock, we put that as an input in our markowitz model to find the number of possible portfolios. In this section, the best portfolio combination came with highest sharpe ratio of 2.37 and volatility of 0.164 with 39% return rate.

	Returns	Volatility	Sharpe Ratio
13476	0.387604	0.163568	2.369685

Figure 8.1: Highest Sharpe Ratio Based on Annual Returns

	Returns	Volatility	Sharpe Ratio
17857	0.332897	0.151587	2.196075

Figure 8.2: Minimum Volatility Based On Annual Returns

8.2 Individual Sharpe Ratio

While clustering based on individual Sharpe ratio gives a totally different result which has a good portfolio Sharpe ratio but not as good as annual return based but the volatility rate decrease with high margin which is a very good indicator of risk-free allocation of stock with minimal return.

	Returns	Volatility	Sharpe Ratio
11984	0.274932	0.146292	1.879343

Figure 8.3: Highest Sharpe Ratio Based On Sharpe Ratio

	Returns	Volatility	Sharpe Ratio
19552	0.194085	0.118605	1.636402

Figure 8.4: Minimum Volatility Based On Sharpe Ratio

8.3 Market Capitalization

Calculation of market capitalization always shows how large are the companies. For this method, we chose the top 30 highest market cap companies and computed a portfolio distribution and surprisingly performance is very poor for this case. Best portfolio combination came with Sharpe ratio of 1.37 and min variance came with 1.063 Sharpe ratio and return also decreased.

	Returns	Volatility	Sharpe Ratio
10590	0.173073	0.12664	1.366649

Figure 8.5: Highest Sharpe Ratio Based On Market Cap

	Returns	Volatility	Sharpe Ratio
978	0.120878	0.113613	1.063944

Figure 8.6: Minimum Volatility Based On Market Cap

Chapter 9

Hybrid Model

After computing for multiple times on different clusters, we have eventually computed portfolio selection for the most productive stocks from companies across all industries. After going through this research process we had a very good intuition over the whole stock market and which particular stocks perform well in course of time and other factors. So we came up with a hybrid model in which we will include the best performing stocks from annual return and the top ones from individual Sharpe ratio and put those into one stack and compute efficient frontier to allocate suitable weights distribution. The ratio for selecting annual return would be higher than Sharpe ratio as our prime focus is increasing the return and for minimizing risk we will keep a certain amount of stocks from individual Sharpe ratio. With this model, we achieve a very good Sharpe ratio and more importantly we were able to decrease volatility. In the best portfolio combination the Sharpe ratio is above 2 which is very good and volatility is less than previous models whereas the minimum variance has a very less volatility with a good amount of Sharpe ratio.

	Returns	Volatility	Sharpe Ratio
10241	0.361409	0.161045	2.24415

Figure 9.1: Hybrid A

	Returns	Volatility	Sharpe Ratio
29304	0.325203	0.148399	2.191414

Figure 9.2: Hybrid B

Chapter 10

Validation

Using our proposed hybrid model and running it multiple times we eventually found some calculated weights and possible portfolios. In this segment, we have computed the Markowitz model on stock data from the year 2015 to 2018 and with that, we have calculated all possible outcomes. More importantly, we have collected the weights assigned to each stock of the portfolio. After that, for the purpose of cross-validation, we used those weights and took data from 2019 January to 2019 November. With this data, we fed it into our model with collected weights from the previous experiment. As a result, we have got a single portfolio as an output from our model with that optimized weights which gave us a very satisfying result. The previous result for portfolio selection, the Sharpe portfolio has a return of around 36% and 2.24 Sharpe ratio with a 16% volatility rate. With this collected weight, running it into 2019 data has also given a very good result of 37% return and 16% volatility and Sharpe ratio of 2.34. Now the stocks we have selected for our portfolio and looking into the data from 2018 and 2019 and comparing those, we can see the selected stocks have an up rise of their values where some of the stocks eventually dropped. This proves the correctness of the volatility rate along with the return rate and Sharpe ratio. According to this observation, we can strongly state that our proposed model has optimized the portfolio selection and correctly distributed the weights taking consideration of market changes.

```
In [18]: sharpe_portfolio
```

```
Out[18]:
```

	Returns	Volatility	Sharpe Ratio	MKTX Weight	PGR Weight	GRMN Weight	FMC Weight	MSCI Weight	BBY Weight	CAT Weight	...	MU Weight	WCG Weight	TTWO Weight	ANET Weight	LW Weight
0	0.377512	0.161182	2.342146	0.046281	0.065867	0.028585	0.015779	0.050864	0.015664	0.006809	...	0.010904	0.069886	0.001121	0.037029	0.084123

1 rows x 33 columns

```
In [19]: min_variance_port
```

```
Out[19]:
```

	Returns	Volatility	Sharpe Ratio	MKTX Weight	PGR Weight	GRMN Weight	FMC Weight	MSCI Weight	BBY Weight	CAT Weight	...	MU Weight	WCG Weight	TTWO Weight	ANET Weight	LW Weight
0	0.377512	0.161182	2.342146	0.046281	0.065867	0.028585	0.015779	0.050864	0.015664	0.006809	...	0.010904	0.069886	0.001121	0.037029	0.084123

1 rows x 33 columns

Figure 10.1: Updated Value

Bibliography

- [1] K. K. Hung, C. C. Cheung, and L. Xu, “New sharpe-ratio-related methods for portfolio selection”, in *Proceedings of the IEEE/IAFE/INFORMS 2000 Conference on Computational Intelligence for Financial Engineering (CIFEr)* (Cat. No. 00TH8520), IEEE, 2000, pp. 34–37.
- [2] X. Yu and L. Xu, “Adaptive improved portfolio sharpe ratio maximization with diversification”, in *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium*, IEEE, vol. 4, 2000, pp. 472–476.
- [3] M. H. Cohen and V. D. Natoli, “Risk and utility in portfolio optimization”, *Physica A: Statistical Mechanics and its Applications*, vol. 324, no. 1-2, pp. 81–88, 2003.
- [4] B. I. Group, *Sp 500 historical sector weightings*, Jan. 2012. [Online]. Available: <https://seekingalpha.com/article/320168-s-and-p-500-historical-sector-weightings>.
- [5] G. Mainik, G. Mitov, and L. Rüschemdorf, “Portfolio optimization for heavy-tailed assets: Extreme risk index vs. markowitz”, *Journal of Empirical Finance*, vol. 32, pp. 115–134, 2015.
- [6] C. De Franco, J. Nicolle, and H. Pham, “Bayesian learning for the markowitz portfolio selection problem”, *arXiv preprint arXiv:1811.06893*, 2018.
- [7] E. is a co-portfolio manager in a global systematic multi-strategy hedge fund. He has been in the fund, finance industry since 2006, is the owner of the blog Investment Cache., R. Posts, and N. (required), *Hedge funds is nothing like sp 500*, Sep. 2019. [Online]. Available: <https://investmentcache.com/hedge-funds-is-nothing-like-sp-500/>.
- [8] R. Kim, *Efficient frontier portfolio optimisation in python*, Jan. 2019. [Online]. Available: <https://towardsdatascience.com/efficient-frontier-portfolio-optimisation-in-python-e7844051e7f>.
- [9] G. Mariani, Y. Zhu, J. Li, F. Scheidegger, R. Istrate, C. Bekas, and A. C. I. Malossi, “Pagan: Portfolio analysis with generative adversarial networks”, *arXiv preprint arXiv:1909.10578*, 2019.
- [10] *Standard deviation*, Dec. 2019. [Online]. Available: https://en.wikipedia.org/wiki/Standard_deviation?fbclid=IwAR0Z4CCiWkAtPSHmEHRPI440o4M8c2ie_AwhLV7zpZgOyYKKnHbR.JbvIE9E.