EE4080 FINAL YEAR PROJECT

PROGRESS REPORT

PART A 3

PART B 4

1. INTRODUCTION 4
2. BACKGROUND STUDY AND LITERATURE REVIEW 4
   1. FUNDAMENTALS OF INVESTMENTS 4
      1. Asset 4
      2. Stocks 4
      3. Portfolio 4
   2. FINANCIAL TIME-SERIES CONCEPTS 5
      1. Prices 5
      2. Returns 5
      3. Volatility or Risk 6
   3. FUNDAMENTALS OF PORTFOLIO THEORIES 6
      1. Modern Portfolio Theory 6
      2. Portfolio Expected Returns 6
      3. Efficient Frontier 6
      4. Sharpe Ratio 6
   4. MACHINE LEARNING APPROACHES 7
      1. Clustering Methods for Diversification 7
3. METHODOLOGY AND IMPLEMENTATION 7
   1. DATA COLLECTION 7
      1. Standard & Poor List of Companies 7
      2. Stock Data 7
   2. STOCK SELECTION AND CLUSTERING 8
      1. SECTOR BASED STOCK SELECTION 8
      2. STOCK CLUSTERING 9
      3. STOCK CORRELATION 10
      4. EFFICIENT FRONTIER AND SHARPE RATIO 11
4. SUMMARY OF OBJECTIVES 12
5. REFRENCES 13

# PART A

# PART B

### INTRODUCTION

In the world of finance, portfolio management and optimisation is a problem which is solved everyday by asset managers in the field. Portfolio management and optimisation is the process of allocating capital across different financial assets such as equities, bonds and mutual funds to build investment solutions. While building a portfolio, the investor must have an investment objective and must address various factors such as how much risk they are prepared to take, the duration of their investment and many other factors which will be discussed in this paper.

Over the past decade, there has a been exponential growth and research in the field of artificial intelligence and machine learning. Advanced algorithms are being developed and tested in multiple areas and finance is one of the front runners. In this project, I will tackle portfolio optimisation with help of reinforcement learning techniques and different machine learning models to optimally allocate weights to various equity assets and build portfolios which have returns for the risk appetite of the investor.

### BACKGROUND STUDY AND LITERATURE REVIEW

#### FUNDAMENTALS OF INVESTMENTS

#### Asset

An asset is essentially anything which possesses monetary value. Some example of assets include cash, checking and savings accounts, treasury bills, stocks, bonds etc [1]. In this project I will be focusing on stocks.

#### Stocks

A stock or equity is an asset or a form of security which indicates a proportionate ownership in a company by the holder of the stock. Stocks are generally issued by corporations to raise funds in order to operate their businesses. Stocks are bought mostly on stock exchanges. Some common and famous stock exchanges are the New York Stock Exchange (NYSE), the London Stock Exchange (LSE), NASDAQ to name a few. Stocks are usually the foundation of almost every portfolio built [2].

#### Portfolio

A portfolio is a set or group of publicly traded assets such as equities, bonds, mutual funds, commodities etc. Sometimes portfolios can also include non-publicly traded assets such as real estate and art. Investors construct their portfolios based on their risk tolerance their investment objectives and these portfolios are generally managed by the investors themselves or wealth managers.

* + 1. Portfolio Construction and Optimisation

Portfolio Construction is all about selecting the right assets based on the investors objectives and risk tolerance. The most important aspect of constructing a portfolio is the asset allocation [3]. Breaking down the portfolio optimisation process, there are four important steps one must consider:

* + - 1. Risk Profiling - High Risk or Low Risk based on a scale determined by the investor.
      2. Asset Allocation - The right combination of assets for the portfolio.
      3. Fine-Tuning the Portfolio - Based on the risk profile and the existing assets, add or remove the underlying assets most suitable to the investor, potentially reducing the risk and increasing the returns.
      4. Rebalancing - The investor must review the portfolio from time to time and rebalance the weight allocated to the underlying asset based on the market.

#### FINANCIAL TIME-SERIES CONCEPTS

#### Prices

Let *pt* be the price of a stock at a discrete time index of *t.* The time series sequence can then be shown as *p1, p2, p3, …. pT-1, pT .* For the convenience of portfolio analysis, we define the price vector *pt* such that [4],

pt = [p1,t, p2,t, p3,t,….., pM,t ] ∈ ℝ

The data being used for the project contains three types of prices for each security. The three types are as follows:

* + 1. **Open Price** - This represents the securities price at the start of the day’s business
    2. **Close Price** - This represents the securities price at the end of the day’s business
    3. **Adjusted Close Price** - This factor in anything that might affect the stock price after the market closes and amends the close price.

#### Returns

Returns are in the simplest terms the change in the price of an asset or a period of time which can be represented in terms of price change or percentage. In a portfolio, positive returns represent a profit in the investment made and negative returns indicate a loss [4].

There are two general methods of calculating returns:

* + - 1. **Simple Returns**

The change in percentage of asset prices from *(t-1)* to *t* is called simple return. It can be calculated by using the following formula:

* + - 1. **Log Returns**

Logarithmic returns provide a symmetric method to calculate future returns as opposed to simple returns which is asymmetric. Log returns are additive hence making it easier to process time series data. It can be calculated using the following formula:

#### Volatility or Risk

In investments, risk or volatility generally defines the possibility or chance that an outcome of an investment or the investment gain actually differs from the expected outcome or the returns. It refers to the possibility of losing either some or all of the original investment amount. With regards to equities and the historic prices, standard deviation is a common measure used to quantify risk. It is used to provide the volatility of an equity, comparing it to its historic average [5].

#### FUNDAMENTALS OF PORTFOLIO THEORIES

#### Modern Portfolio Theory

In 1952Harry Markowitz published a paper called “Portfolio Selection” where he introduced the Modern Portfolio Theory (MPT). MPT theorizes on how a risk-averse investor can build and optimize their portfolios to achieve maximizing expected returns for a given level of market risk [6].

#### Portfolio Expected Returns

In a portfolio of *N* assets, the expected returns are calculated as the weighted average of the individual returns of all *N* assets. For example, if a portfolio contains 5 equally weighted assets, each with individual returns of 6%, 7%, 12%, 4% and 15% respectively, then the expected returns of the portfolio will be calculated by:

*E(R) = (0.06 x 0.20) + (0.07 x 0.20) + (0.12 x 0.2) + (0.04 x 0.20) + (0.15 x 0.20) = 8.8%*

#### Efficient Frontier

Every possible weighted combination of assets in a portfolio can plotted onto a graph showcasing the portfolio risk on the X-axis and the portfolio expected return on the Y-axis. Based on this graph an investor can choose the most desirable combination of assets. The upward sloping hyperbola which connects all the most efficient portfolio asset combinations is called the efficient frontier. For an investor following the Modern Portfolio Theory, investing in a portfolio not on the efficient frontier is not desirable [6].

#### Sharpe Ratio

Nobel laureate William Sharpe introduced a new method to understand a portfolio’s performance. The Sharpe Ratio is used by investors to understand the return on investment (ROI) compared to the amount of risk undertaken. The Sharpe ratio can be calculated by using the formula [6]:

Where:

#### MACHINE LEARNING APPROACHES

#### Clustering Methods for Diversification

The process of Diversification is used by investors to choose investments which help in reducing the exposure to any one particular asset. This can be done by investing in a mixed variety of assets. Correlation plays a big role in helping to build diversified portfolios. When the asset prices are not correlated i.e., they do not move together, then a diversified portfolio of those assets will have a lower volatility.

Cluster Analysis is the task of grouping a set of objects that shares similar characteristics to each other than other sets of groups. Clustering is generally used for statistical analytics, but investors can use clustering to build diversified portfolios. Clustering can be divided into two basic algorithms, hierarchical and partition clustering. For this project, I will be using partitional clustering algorithms which directly partitions them data into a set of disjoint clusters.

**K-Means Clustering - Partitional Clustering**

K-Means clustering is one of the most common clustering algorithms used for unsupervised machine learning tasks. K-Means is a centroid-based or distance-based heuristic algorithm which calculates the distance between a point and the centroid and then assigns the point to a cluster. One of the main objectives of using K-Means is to minimize the sum of distances between respective points and their cluster centroids [7].

### METHODOLOGY AND IMPLEMENTATION

#### DATA COLLECTION

#### Standard & Poor List of Companies

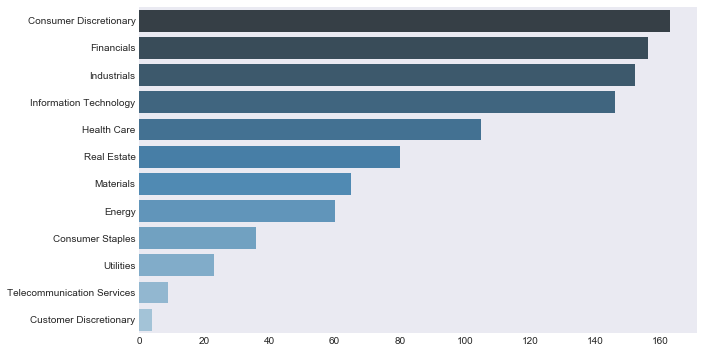
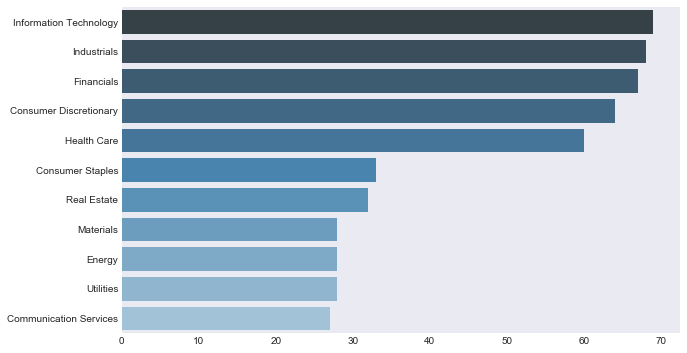
Currently for this project I will be using the S&P list of 500 large market capital companies, S&P list of 700 mid-market capital companies and the S&P list of 300 small market capital companies. Using the inbuilt library function for retrieving HTML data provided by the pandas library, I retrieve all the thousand five hundred company ticker information along with the Global Industry Classification Standard Sector information.

#### Stock Data

The stock price time-series data is being collected using the Yahoo Finance API which also inbuilt into the pandas data-reader library. For the purpose of the portfolio construction, we only need the Close market-price data. We also retrieve the latest market capital value for each of the thousand five hundred companies. Currently the time series used for the project is from ’01/01/2017’ to the current date.

#### STOCK SELECTION AND CLUSTERING

#### SECTOR BASED STOCK SELECTION

Diversification is one of the most important aspects of portfolio construction as discussed in the background study. Selecting stocks from different sectors help in diversifying the portfolio. Since we have over thousand five hundred stocks to choose from, it becomes hard to select a group. Approaching a systematic method, to reduce the number of stocks in the portfolio universe, I use the market capital as a feature to select the group of stocks.

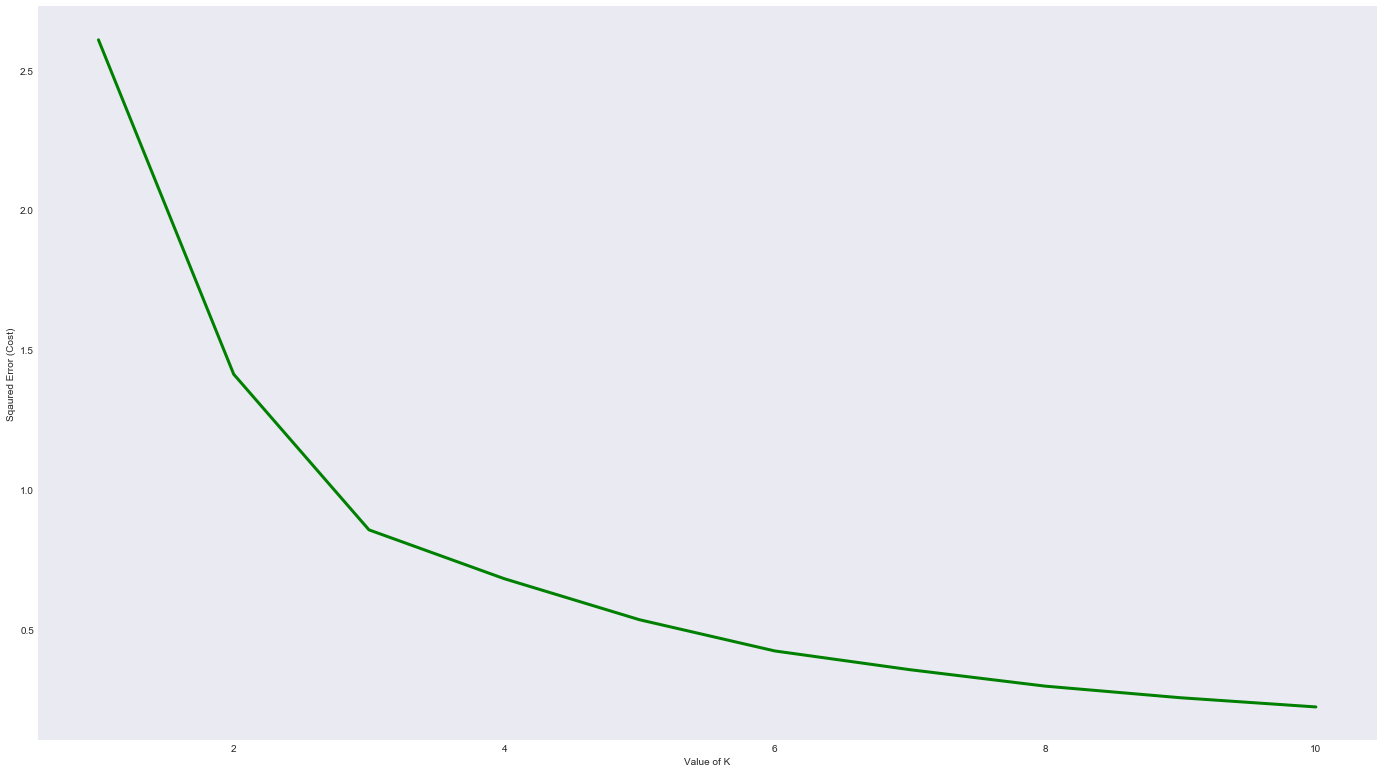
It is assumed that stocks with higher market-capitalization are easier to trade as they have higher liquidity as compared to the stocks with lower market-capitalization. Lower market- capitalization stocks are considered to be illiquid with smaller trading volumes. Using them for the portfolio would not be efficient as they increase the volatility.

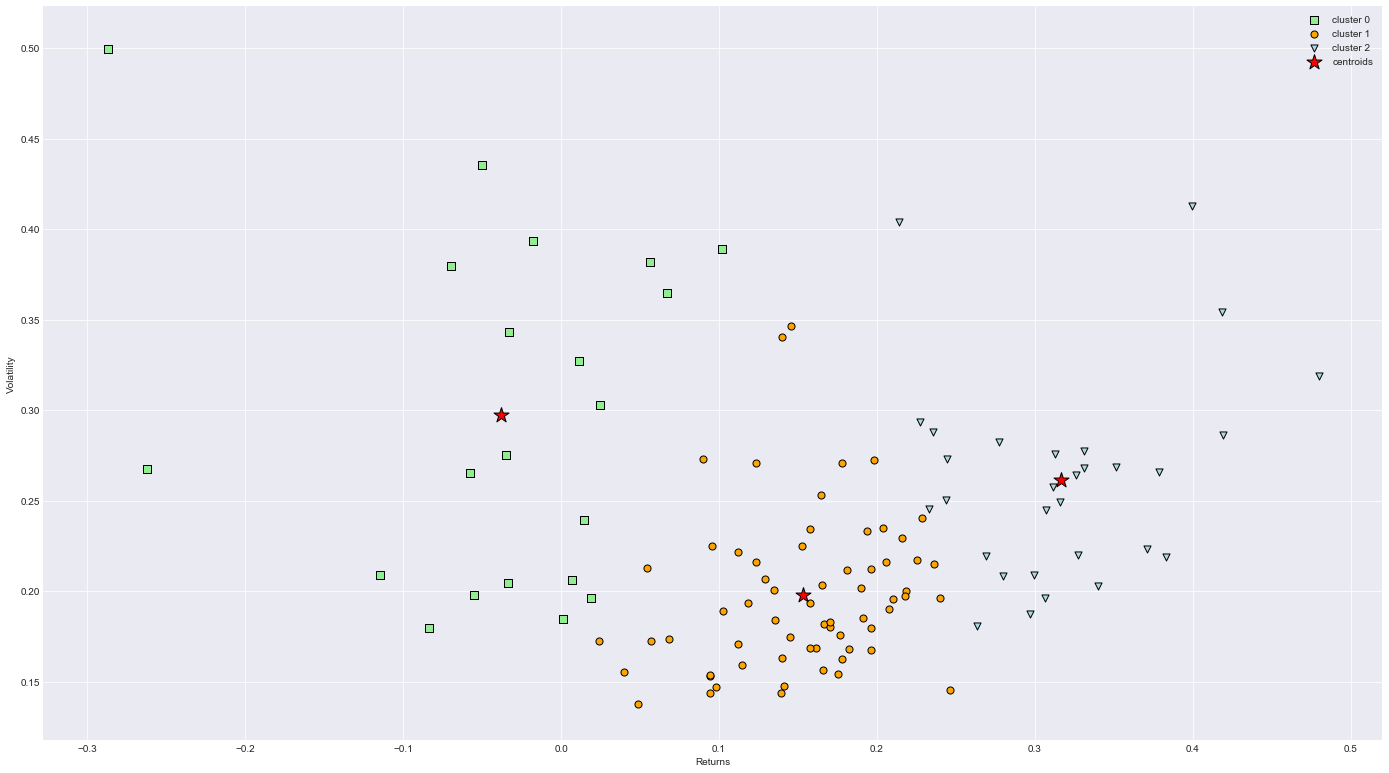
Using the market capitalization information retrieved from yahoo finance, I am able to group-by the sectors and select the top five stocks with the higher market-capitalization for large-cap, mid-cap and small-cap stocks.

#### STOCK CLUSTERING

After the sector-based stock selection, we have about a hundred stocks for clustering. Using the yahoo finance API, I retrieve the time-series close price information for each of the hundred stocks. I then calculate the annual return and volatility for each stock. These two numerical features are used as the input features for the K-Means clustering algorithms. Iterating the Means algorithm for *i=1000* where *i = number of clusters.*  Plotting the Cost (Squared Error) against the number of clusters, I can determine the right number of clusters by find the elbow of the graph.

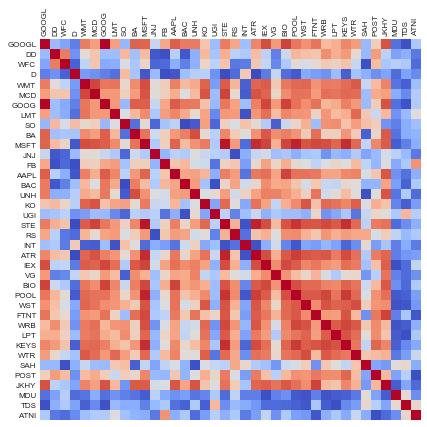
The hundred stocks are then clustered into three groups with each stock assigned to a group based on the distance to the centroid. I then selected fifty percent of stocks from each cluster to reduce the correlation. The stocks are selected based on the distance to their respective centroids.



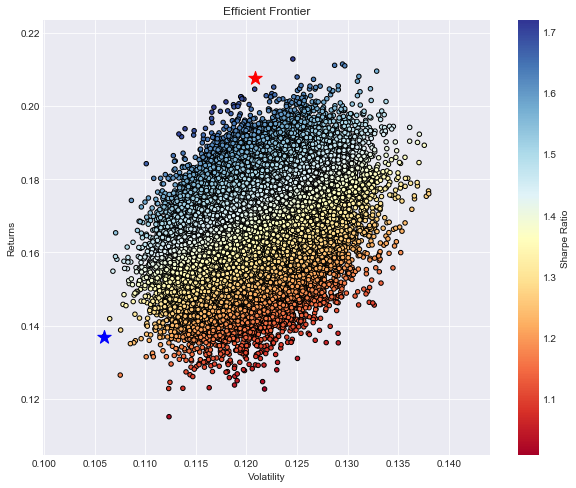


#### STOCK CORRELATION

Although I select stocks from different clusters for the portfolio for diversification, some stocks can still be very highly correlated. In order to determine which highly correlated stock-pairs to remove, I plot the correlation matrix of all the stocks in the portfolio. Based on the plot and correlation coefficient I drop the stock-pairs which are highly correlated.



#### EFFICIENT FRONTIER AND SHARPE RATIO

 After completing all the previous steps, we now have the group of stocks for the portfolio and can calculate the weights using mean-variance optimization. The efficient frontier graph is a helpful visualization of all the possible weight combinations which is set to fifty thousand iterations. The Sharpe Ratio and the Minimum Variance portfolios are also plotted on the efficient frontier graph.

For this portfolio of 31 equity assets, optimising the weights using mean-variance method, we get the Sharpe portfolio with the following results:

### SUMMARY OF OBJECTIVES

|  |  |  |
| --- | --- | --- |
| Objectives | Addressed  *(please tick)* | Percentage achieved  *(please estimate)* |
| Project Outline and Goals |  | 60% |
| Literature Study |  | 60% |
| Introduction to Investments |  | 90% |
| Introduction to Portfolio Construction and Analysis on Python |  | 90% |
| Developing Python Algorithms for Portfolio Construction |  | 100% |
| Introduction to Portfolio Optimisation Techniques |  | 60% |
| Developing Optimisation Algorithms on Python |  | 60% |
| Introduction to Clustering Algorithms for Stocks |  | 90% |
| Interpretation of Results |  | 100% |



Semester A Project Timeline

### REFRENCES

[1] Investopedia. (2015) Assets.

[2] Investopedia. (2015) Assets.

[3] Investopedia. (2015) Assets.

[4] Investopedia. (2015) Assets.

[5] Investopedia. (2015) Assets.