|  |
| --- |
|  |
|  |
| January 14  Birmingham city university  Student id - 23178059  Authored by: Namagiri Mukul |

# Machine learning Report

# CMP7228 S1 2023/4

Index

Domain description --------------------------------------------------------------------------- 3-5

Problem definition --------------------------------------------------------------------------- 5

Dataset description -------------------------------------------------------------------------- 5-7

Dataset exploration -------------------------------------------------------------------------- 8-

Experiments -------------------------------------------------------------------------- 10

Analysis and Results ------------------------------------------------------------------------- 12

Conclusions ------------------------------------------------------------------------- 14

References ------------------------------------------------------------------------- 16

Appendixes ------------------------------------------------------------------------- 18

Domain Description

A portfolio is a collection of investments owned by a person/individual or organization. These can range from securities, stocks, individual businesses, bonds, and mutual funds. The objective of a portfolio manager is to evaluate, manage, and optimize the composition of assets dynamically (Johnson,2014). With the advent of machine learning strategies and artificial intelligence algorithms, there is a drastic shift in the way portfolios are managed nowadays. These technologies help in processing large data sets, building models, and predicting market trends. Discretionary portfolio managers or systematic portfolio managers often tend to build theories and rationale with a heavy reliance on judgment and intuition rather than logic and fact-based, empirical evidence (Lopez, de Prado, 2018). Portfolio analysis has several aspects that include Security analysis, Stock selection, Risk management, portfolio construction, Performance Construction, Performance monitoring, Rebalancing, Dividend analysis, Tax considerations, Market and economic analysis. The flow chart below will represent the decision flow of a typical structured portfolio management.

Exploring different aspects of the individual steps presented

1. Security analysis refers to the process of evaluating financial instruments such as stocks to determine investment potential it has two main subcategories namely fundamental analysis and technical analysis.

* Fundamental analysis – examining the intrinsic value of security like management quality, industry conditions, economic indicators, etc.
* Technical analysis – analyzing the historical prices and volume data to predict future price movements.

1. Stock selection involves adding stocks based on growth potential, valuation, and risk factors. It also involves diversifying holdings across different industries.
2. Risk management involves assessing the individual stocks, including factors like volatility, beta, and correlation with other assets also includes setting up stopping criteria to hedge against adverse price movements.
3. Portfolio construction involves combining individual stocks with the investor’s goals in mind like risk tolerance, and time horizon.

A diagram of a risk management

Description automatically generated

Figure 1

1. Performance monitoring involves regularly reviewing individual companies and assessing the impact of market conditions on stock performance.
2. Rebalancing involves active assessment of assets with investor's objectives and changing market conditions.
3. Dividend analysis involves catering to income-oriented investors making sure that dividend yield and payout ratio are optimal while changing stock allotment.
4. Tax considerations should be optimized considering capital gains tax and dividends.
5. Market and economic analysis involves staying informed about market trends, economic indicators, and geopolitical issues that may impact stocks and also incorporating macroeconomic analysis into portfolio management and stock selection (Johnson and Parente,2013).

Problem definition

The goal of this project is to solve three problems associated with portfolio management using data analysis and machine learning techniques, the primary question of the 3 problems are presented below

Regression problem

Using regression-based machine learning models to find the future values of the opening, closing, high, and low values of the selected stocks. (Ravi Kumar & Saraf, 2020)

Classification problem

Using binary classification techniques to classify if the stock is going to go up or down to decide to buy or sell stock.

Clustering problem

Clustering companies based on their risk profile i.e., to identify classes of companies that have the same risk profile we want to diversify our portfolio based on the investor's needs so identifying and clustering them lets us decide which stocks to pick.

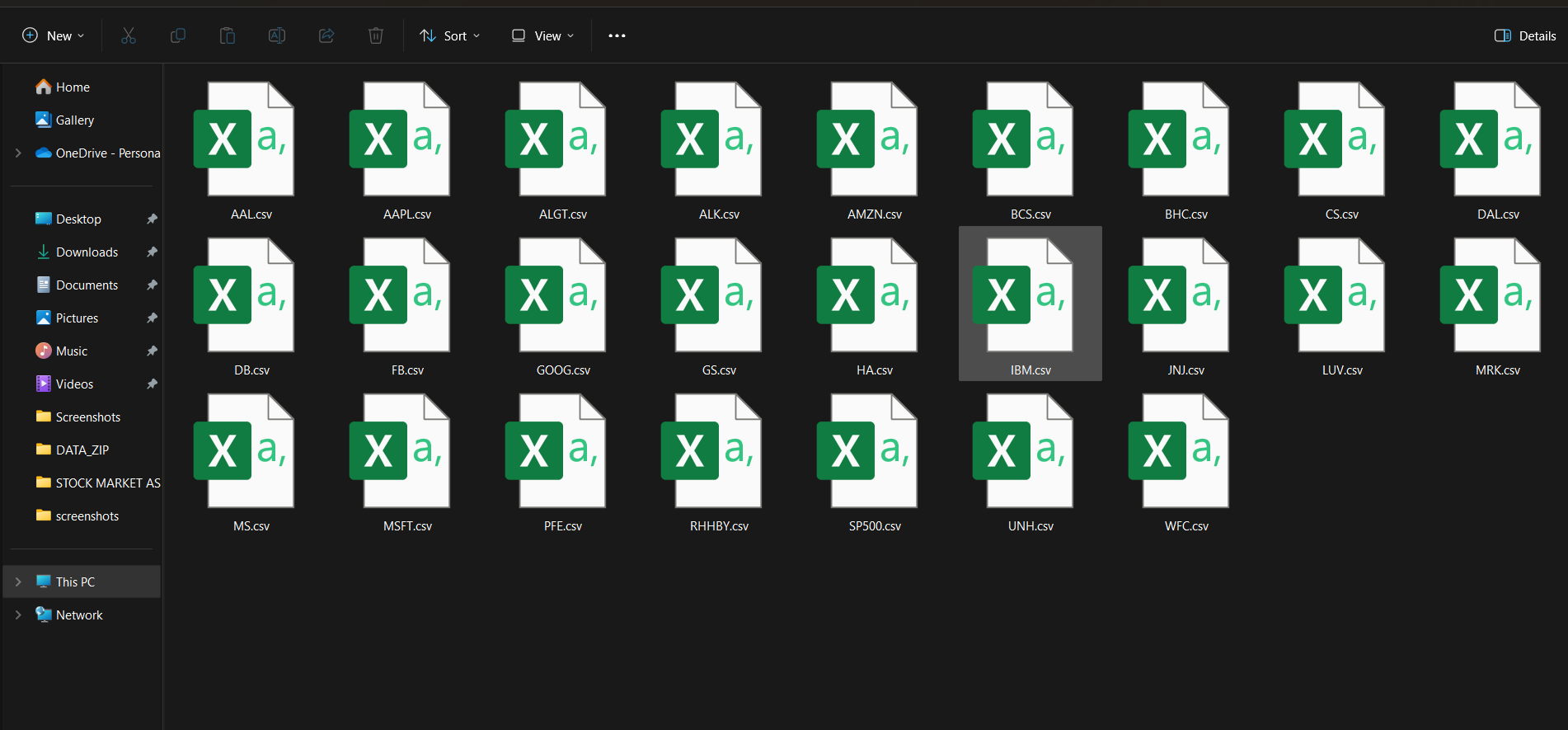
Dataset description

This project uses data from the Kaggle data set (Stock portfolio – financial risk analytics) with a usability rating of 4.71 – tags (Investing)

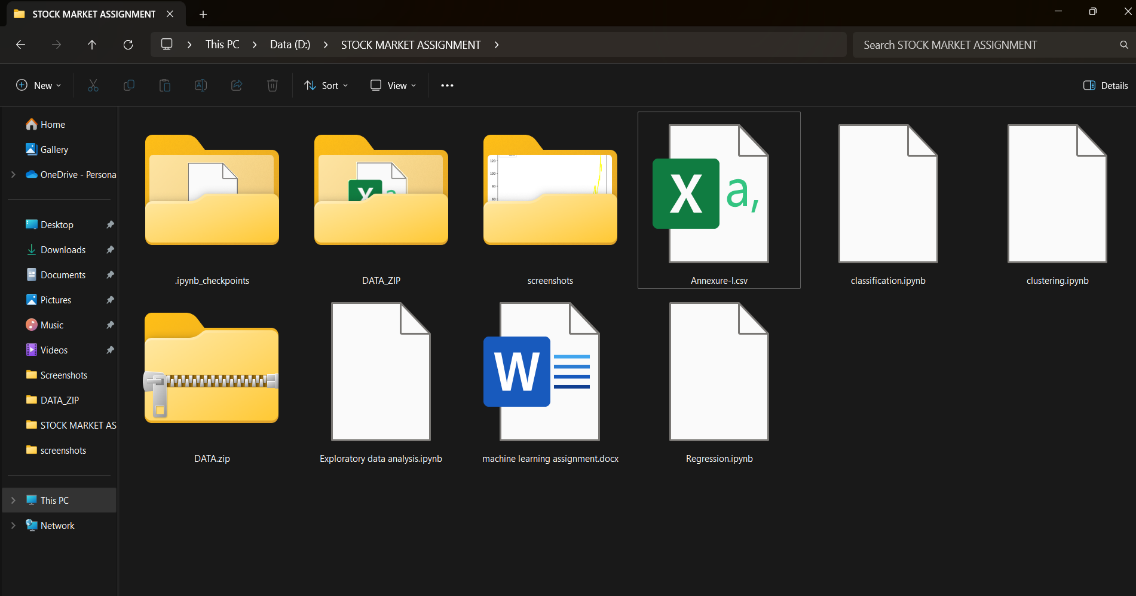
Problem statement presented in Kaggle – how to identify the right investment opportunity and recommend a portfolio as per the client's exact needs.

From a broad perspective, there are two types of investors conservative investors, and aggressive investors according to traditional investment norms conservative investors tend to be more cautious and want less risky alternatives even if the returns are not very high but aggressive investors tend to be more risk-taking for better returns.

The file directory is named DATA\_ZIP it contains 24 different company stocks and 1 CSV file for S&P 500 index.



One of the CSV file is Annexure – 1 is the index file for the whole data set containing the ticker symbol for the stock being traded, the industry it belongs to, and the name of the company name.



A close-up of a table

Description automatically generated

In this analysis we are only going to look at technology, finance and aviation companies

Dataset download link 🡪 <https://www.kaggle.com/datasets/ankurnapa/stock-portfolio-financial-risk-analytics>

Data set description link 🡪 <https://www.kaggle.com/datasets/ankurnapa/stock-portfolio-financial-risk-analytics?select=Annexure-I.csv>

Dataset Exploration

Describing each of the columns present in the data set

Date:

Date represents the trading day of the company at which the data is recorded.

Open:

Open represents the opening price of the stock for the trading day.

Close:

Close represents the closing price of the stock for the trading day.

High:

High represents the highest price recorded for the trading day.

Low:

Low represents the lowest price recorded for the trading day.

Adjusted closing price:

Adjusted closing price of the stock on a trading day is used to indicate the closing price taking dividends, stock splits, and new stock offerings into account it presents a more accurate value of the actual stock value.

Volume:

Volume refers to the total number of shares of a particular stock traded during a trading day.

The table below represents the column data for the index fund S&P 500

|  |  |  |
| --- | --- | --- |
| column | Data type | initial value |
| Date | object | 01-10-2010 |
| Open | Float64 | 1143.48 |
| Close | Float64 | 1146.23 |
| High | Float64 | 1150.30 |
| low | Float64 | 1139.42 |
| volume | Float64 | 4296910000 |
| Adjusted closing price | Float64 | 1146.239990(since there are no stock splits or dividends for INDEX FUND) |

Since manually opening each of the files is not possible using the OS module to read all the files and making a data frame that contains the first row of the CSV files and making a new column named company to identify which stock it belongs to

A table of numbers and letters

Description automatically generated

As evident from the table there seems to be inconsistent date formatting and varying start date for different companies of different domains. We want to check the overall general trend of the companies based on the domain of interest.

Analysis of technology companies:

A graph of a stock market

Description automatically generated with medium confidence

Line plot of different technology companies

As seen from the plots above the technology companies have seen a meteoric rise in recent years with Apple, Amazon, Google, and Microsoft showing very drastic increases whereas IBM shows a more stable plot since IBM has been in the market for a much longer period as expected.

A group of red and white bars

Description automatically generated

Volume traded per year for different technology companies.

The above plot shows the volume that has been traded across different years apple, google, Microsoft and Facebook have shown a slight downward trend in the late 2010s whereas IBM and Amazon have shown a relatively stable volume of the stock that is being traded.

The peak volume traded for

Apple is around 2011 and 2012

Amazon is in 2011 and 2018

Microsoft in 2011

IBM in 2011 and 2018

Google 2011 and 2012

Analysis of finance companies:

A graph of blue lines

Description automatically generated with medium confidence

Line plot of different finance companies

As seen from the plots above, finance companies have not had much of a dramatic rise like technology companies and by the late 2010s have shown a dramatic decrease the only companies that did manage to have a relatively stable are Morgan Stanley and Goldman Sachs.

Barclays, Credit Suisse, and Deutsche Bank have shown a relatively downward trend

While Wells Fargo has reached all time low by the end of 2019.

A group of blue and white bars

Description automatically generated

Volume traded per year for different finance companies

The volume traded for Goldman Sachs and Morgan Stanley has relatively decreased low trading volume with increased prices signifying a tendency to hold onto the assets.

The peak volume traded for

Barclays is in 2016

Credit Suisse is in 2016

Morgan Stanley in 2011 and 2012

Wells Fargo in the 2010s and 2020s (with a decreased price indicating loss of value)

Goldman Sachs is around the early 2010s.

Analysis of aviation companies:

A group of green lines

Description automatically generated

Line plot of different aviation companies

As seen from the plots above, the aviation industry has been relatively stable without any meteoric increase or decreases but the growth seemed to have plateaued by the late 2020s.

Alaska Airlines Group, Hawaiian Holdings Inc., and Allegiant Travel Company have shown a similar trend.

While Delta Airlines and Southwest Airlines seemed to do well until early 2019 by decreased by 2020 like other aviation companies

A group of green bars

Description automatically generated

Volume traded per year for different aviation companies.

The peak volume traded for

American airline group in 2020 [with a drastic decrease in price indicating severe losses.]

Followed by Delta Airlines and Southwest Airlines around 2020 [with a similar trend across each year.]

Alaskan air group and allegiant travel company in 2020.

2020 has been a bad year for the aviation industry with several big companies trading at less price with high volume traded

Correlation analysis of different companies in each sector

Correlation is one of the important aspects of picking good stocks if the picked companies are highly correlated this might lead to concentration risk and could lead to huge potential losses to avoid this we should analyze the correlation of different companies across each individual sector to find the ideal stock.

Correlation analysis of technology companies

A graph of numbers and a number

Description automatically generated with medium confidenceA colorful squares with numbers

Description automatically generated

Heatmap of closing prices correlation Heatmap of volume traded correlation

As it can be inferred from the correlation matrix Amazon and Apple are highly correlated also, Microsoft and Apple are highly correlated with respect to prices of the trading period, and Google and Apple and Microsoft and Apple are highly correlated in terms of volume.

IBM seems to be negatively correlated with several other technology companies indicating a good opportunity to diversify portfolio.

Correlation analysis of finance companies

A screenshot of a color chart

Description automatically generatedA screenshot of a color chart

Description automatically generated

Heatmap of closing prices correlation Heatmap of volume traded correlation

The correlation analysis of closing prices of finance companies shows both positive and negative correlation, unlike most of the technology companies.

As it can be inferred from the above charts credit Suisse is highly correlated with Deutsche Bank and Morgan Stanley is highly correlated with Goldman Sachs.

But there seems to be a very high negative correlation between traditional banks and investment banks as expected which presents us with a chance to categorize the finance companies into two subsects and a potential opportunity to diversify the assets.

Correlation analysis of Aviation companies

Heatmap of closing prices correlation Heatmap of volume traded correlation

# 

# S. Ravikumar and P. Saraf, "Prediction of Stock Prices using Machine Learning (Regression, Classification) Algorithms," 2020 International Conference for Emerging Technology (INCET), Belgaum, India, 2020, pp. 1-5, doi: 10.1109/INCET49848.2020.9154061.

# Johnson, William H. A., and Diane H. Parente. *Project Strategy and Strategic Portfolio Management : A Primer*, Business Expert Press, 2013.*ProQuest Ebook Central*,

# http://ebookcentral.proquest.com/lib/bcu/detail.action?docID=1097921. Created from bcu on 2024-01-10 11:11:28.

# Johnson, R. Stafford. *Equity Markets and Portfolio Analysis*, John Wiley & Sons, Incorporated, 2014.*ProQuest Ebook Central*, http://ebookcentral.proquest.com/lib/bcu/detail.action?docID=827137. Created from bcu on 2024-01-08 10:28:06.

# Lopez, de Prado, Marcos. *Advances in Financial Machine Learning*, John Wiley & Sons, Incorporated, 2018.*ProQuest Ebook Central*, http://ebookcentral.proquest.com/lib/bcu/detail.action?docID=5240570. Created from bcu on 2024-01-08 11:23:44.