Title: Portfolio management

Abstract: Portfolio Management/asset management refers to the professional management of securities and other assets, it will basically manage an individuals and company's securities such as stocks. The management is executed in accordance with specific investment profile and takes into consideration the level of risk. The solution would look like it takes data of various securities and applying machine learning algorithm to analyze the common pattern among them and then applying sentiment analysis and extracting relation among clusters, suggest an optimized portfolio suggestion. I will be using 2 to 3 algorithms as my models depending on time. I will show visually representation of the dataset before and after balancing as a part of data exploration.

Related domain of study (application domain): This is basically aligned to financial domain, it would benefit the individuals or companies who want to invest money in securities and create a portfolio with low risk and high return.

Algorithms:

- Why these algorithms are important and how they are used.

  I am using algorithms such as K-means, linear regression, neural network.

  Currently I am working on K- means and linear regression because I want to implement unsupervised learning and find out patterns in my dataset so for that K- means algorithm is important and then I would like to find correlation among the stocks so for that I am using linear regression.

  After finding pattern, I would like to apply neural network on the dataset to predict the future stock price using sentiment analysis.

- How does the data have to be manipulated or conform to use these algorithms (e.g. remove nulls, split data into training vs test sets...)?

  I will remove the null values in the data if there is any, then I need to either sort the stock data by date and take uniform date range for all individual stocks. For stock prices prediction I will split the data into train and testing.

- How would they be used in real application (e.g. batch processing, data streaming, offline training, online comparison)
  In real application after creating cluster and implementing unsupervised learning for behavioral stock prediction, it would use real time data streaming from application like twitter to know the polarity or sentiment of the stock in the market.

- How are results expected to be interpreted or used?

We can use these results to suggest the companies for the stock creation and to decide whether someone wants to invest in the stock or not.

- **Data sources**:
  https://finance.yahoo.com/
  https://www.kaggle.com/dgawlik/nyse

- Show samples of data source (up to 10 lines of data)

Symbol	Name	Price	Price/Earnings	Earnings/Share	Market Cap
MMM	3M Company	222.89	24.31	7.92	1.390000e+11
AOS	A.O. Smith Corp	60.24	27.76	1.70	1.078342e+10
ABT	Abbott Laboratories	56.27	22.51	0.26	1.020000e+11
ABBV	AbbVie Inc.	108.48	19.41	3.29	1.810000e+11
ACN	Accenture plc	150.51	25.47	5.44	9.876586e+10
ATVI	Activision Blizzard	65.83	31.80	1.28	5.251867e+10
AYI	Acuity Brands Inc	145.41	18.22	7.43	6.242378e+09
ADBE	Adobe Systems Inc	185.16	52.31	3.39	9.455021e+10
AAP	Advance Auto Parts	109.63	19.54	6.19	8.123612e+09
AMD	Advanced Micro Devices Inc	11.22	187.00	0.03	1.119166e+10

In the above table we can see that it is the data of 10 S&P500 listed companies with their symbol, name, their market cap, price of share, price/earnings and earnings/share.
For the analysis purpose we will only use price/earnings, earning/share and market cap.

- Show or describe the steps you will take to transform the sample and make it ready for the analysis

Step 1- I will explore the data both statistically and analytically
Step 2- I will look for any missing values and make data sort by date and take data for certain time span for all the companies
Step 3- I will remove outliers or any unwanted variables to improve my accuracy

Graphics:
What type of graphs, charts, tables are you planning to use to describe the end results
Histogram, Pie chart etc.

What are your current challenges:
I need to scrap data for the sentiment analysis and prepare it to predict and implement neural network.

At least 5 references URLs with phrases or paragraphs you are planning to cite.

https://www.quantinsti.com/blog/sentiment-analysis-news-python
https://devpost.com/software/twitter-sentiment-analysis-for-stock-prediction
http://stanford.edu/~cpiech/cs221/handouts/kmeans.html
http://www.aviarampatzis.com/publications/PCI2016.pdf
https://wesclock777.github.io/
https://medium.com/@chengweizhang2012/simple-stock-sentiment-analysis-with-news-data-in-keras-1478b96dd693
https://www.kaggle.com/dgawlik/nyse