CS526 Stock Market Correlation and News

Amit Patil
Rutgers University
Piscataway, NJ, USA
amit.patil@rutgers.edu

Gaurav Ahlawat
Rutgers University
Piscataway, NJ, USA
gaurav.ahlawat@rutgers.edu

Sharvani Pratinidhi
Rutgers University
Piscataway, NJ, USA
sharvani.pratinidhi@rutgers.edu

***Abstract*— You don't want to lose all your money next time a housing market crash happens or Facebook is caught selling your information to Cambridge Analytica. This is where our project and our tool helps you out. Make educated decision next time you go investing, diversify your portfolio.**

I. Project Description

Diversification is a risk management strategy used by all the big investors and people around the world, which mixes a wide variety of investments, which can be different stocks or commodities or Government bonds, within a portfolio, which can be the 401K you are saving for your retirement. In this project, we created a tool which helps you diversify your stock portfolio, so that next time you are able to make a educated investment and are able to better handle the shocks during a financial downturn. According to much research in portfolio construction, the best portfolios should contain between 25 and 30 stocks,and with thousands of stocks out there, which stock you should buy, which particular combination you should go for, with billions of billions of possible combination- this is where you can use our tool, where you can compare and contrast different stocks in different industries and sectors and see which one of them are least correlated and put your money there.

The project has four stages: Gathering, Design, Infrastructure Implementation, and User Interface.

A. Stage1 - The Requirement Gathering Stage.

The objective of diversification is to mitigate the risk of loss by having a variety of investments in a single portfolio. For example, consider an investment that consists of only stocks issued by a single company. If this stock suffers a serious downturn, the portfolio will sustain the full brunt of the decline.But, by splitting the investment between the stocks from different companies, we can reduce the potential risk to the portfolio.In effect, by spreading the risk among different investments, a higher reward can be expected. Our system helps in diversification by providing the correlation between the companies along with the relevantly ranked articles from reliable sources like The Guardian, The New York Times,etc. so that the users can make an informed decision.
The different types of users for this system are:

- Stock Traders : A stock trader interested in investing his/her money in the stock market in an informed and data driven manner will find our system useful as it will provide him/her will the historical stock prices of the company, along with the relevant and ranked articles. The stock trader can also compare the prices of two companies over time, to see how they perform with respect to one another. The correlation table will also provide the stock trader with the information about how other companies are related to the selected one, so that he/she can consider these companies to diversify the investment.
  - Scenario 1: Stock trader Joe wants buy stocks of Facebook.
  - System Data Input for Scenario 1:
    In the 'Company Correlation' tab, the ticker for Facebook 'FB' is entered in the input box for company names.
  - System Data Output for Scenario 1:
    The System outputs a graph of the Stock Prices of Facebook over time. This graph is in terms of the 'Closing Prices' and can also be changed to 'Volume Traded' and 'Opening Prices'. The articles relevant to Facebook are also displayed by their titles. Clicking on the titles will cause the link to the original article to open in a new tab of the browser.The Correlation Table will display the different companies with their correlation values to Facebook.

  - Scenario 2: Stock trader Joe wants see the relation between stocks of Facebook and Apple.
  - System Data Input for Scenario 2:
    In the 'Company Correlation' tab, the ticker for Facebook 'FB' and the ticker for Apple 'AAPL' are entered in the input boxes for company names.
  - System Data Output for Scenario 2:
    The System outputs a graph of the Stock Prices of Facebook and Apple over time. This graph is in terms of the 'Closing Prices' and can also be changed to 'Volume Traded' and 'Opening Prices'. The articles relevant to Facebook and Apple are also displayed by their titles. Clicking on the titles will cause the link to the original article to open in a new tab of the browser.

- Portfolio Manager : A portfolio manager interested in creating a portfolio for a client in the stock market in an informed and data driven manner will find our

system useful as it will provide him/her will the historical stock prices of the companies, industries and sectors. The portfolio manager can also compare the prices of two companies/industries/sectors over time, to see how they perform with respect to one another. The correlation table will also provide the portfolio manager with the information about how other companies/industries/sectors are related to the selected one, so that he/she can consider these companies/industries/sectors to diversify the investment.

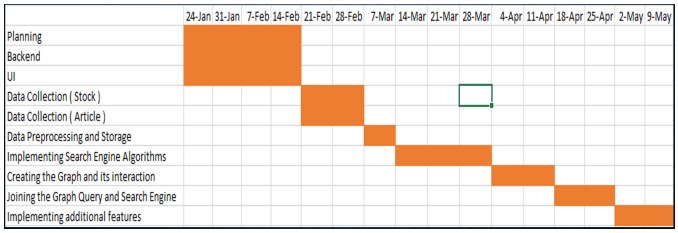
- Scenario 1: Portfolio Manager Mary wants the client to invest in the industries of Banks.
- System Data Input for Scenario 1:
  In the 'Industry Correlation' tab, the industry name 'Banks' is entered in the input box for industry names.
- System Data Output for Scenario 1:
  The System outputs a graph the overall performance of the industry in terms of stock prices of Banks over time.The Correlation Table will display the different industries with their correlation values to Banks.

- Scenario 2: Portfolio Manager Mary wants to check the movement of the sectors for 'Energy' and 'Finance' over time.
- System Data Input for Scenario 2:
  In the 'Sector Correlation' tab, the sector names 'Energy' and 'Finance' are entered in the input boxes for sector names.
- System Data Output for Scenario 2:
  The System outputs a graph the overall performance of the sector in terms of stock prices of 'Energy' and 'Finance' over time.The Correlation Table will display the different sectors.

- Investors : An investor interested in investing his/her money in the a company in an informed and data driven manner will find our system useful as it will provide him/her will the historical stock prices of the company, along with the relevant and ranked articles. The investor can also compare the prices of two companies over time, to see how they perform with respect to one another. The correlation table will also provide the investor with the information about how other companies are related to the selected one, so that he/she can consider these companies to diversify the investment.
  - Scenario 1: Investor Jenna wants to invest in Microsoft.
  - System Data Input for Scenario 1:
    In the 'Company Correlation' tab, the ticker for Microsoft 'MSFT' is entered in the input box for company names.
  - System Data Output for Scenario 1:
    The System outputs a graph of the Stock Prices of Microsoft over time. This graph is in terms of the 'Closing Prices' and can also be changed to 'Volume Traded' and 'Opening Prices'. The articles relevant to Microsoft are also displayed by their titles. Clicking on the titles will cause the link to the original article to open in a new tab of the browser.The Correlation Table will display the different companies with their correlation values to Microsoft. So now Jenna is equipped with all the information about Microsoft and can make an informed decision.

  - Scenario 2: Investor Jenna wants to invest in Microsoft or Apple but is confused about which one she should choose.
  - System Data Input for Scenario 2:
    In the 'Company Correlation' tab, the ticker for Microsoft 'MSFT' and the ticker for Apple 'AAPL' are entered in the input boxes for company names.
  - System Data Output for Scenario 2:
    The System outputs a graph of the Stock Prices of Microsoft and Apple over time. This graph is in terms of the 'Closing Prices' and can also be changed to 'Volume Traded' and 'Opening Prices'. The articles relevant to Microsoft and Apple are also displayed by their titles. Clicking on the titles will cause the link to the original article to open in a new tab of the browser. So now Jenna is equipped with all the information about Microsoft and Apple and can make an informed decision.

- Project Time line and Divison of Labor.

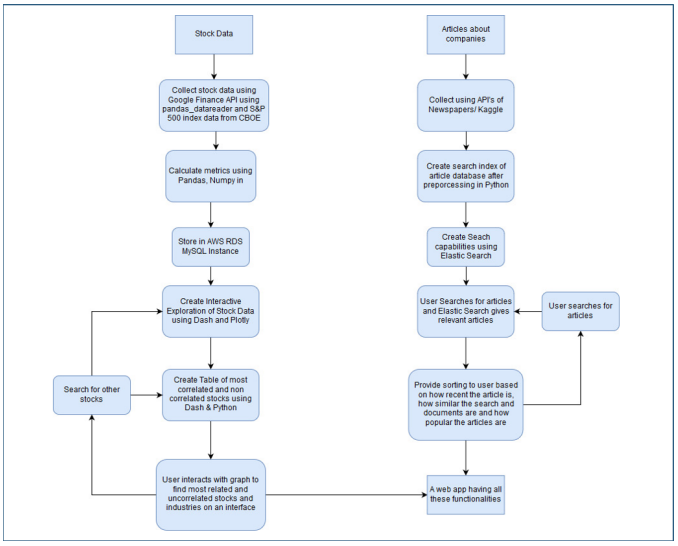


B. Stage2 - The Design Stage.

- Short Textual Project Description. We are mainly performing two tasks in this project. The first one is to find the correlation between the companies, industries or sectors. And the second one is to display the news articles that are related to the companies than you are comparing. For stock data, we have calculated the correlation coefficient for each company with each company using Pandas and numpy libraries in python. Then we stored this data on to the AWS server. Here we have used Client-Server architecture, client queries the AWS server for the required data. Then we have created the interactive application using Dash and Plotly in python where user can type in the names of companies/sectors/industries and

can compare them. For articles, we are preprocessing them using python and then storing into the database. For searching the articles for particular company, we have used elastic search. When user compares companies for stock prices, he will get related articles for that companies also.

- Flow Diagrams:



- High Level Pseudo Code System Description:
  - Company Correlation(Company1, Company2):
    {
    plot_stock_pricing(Company1,Company2);
    get_articles(Company1,Company2);
    Correlation_table(Company1,Company2);
    }
- Algorithms and Data Structures.
  - For finding the correlation between the companies we have used Pearson's correlation coefficient..
  - For searching the related articles in the database we have used the elastic search algorithm. It ranks the articles based on the relevance score. For calculating the relevance score it used combination of Vector space model and TF-IDF.
- Flow Diagram Major Constraints.
  - Integrity Constraint. User must spell the company name correctly. If the user spells the company wrong, he/she will not get the results.

C. Stage3 - The Implementation Stage.

The implementation was done in two stages: one part comparing the daily stock price data we have for the companies and the second part in getting the news articles where these companies were mentioned and their sentiment around them. We primarily used Python as our programming languages and created a Client-Server Architecture, where all our processed data was residing on the AWS servers and the client was making calls to the server for getting back the data and building the visualizations and helping provide information to the user about different stocks from different industries and sectors. The different libraries and frameworks we worked with are follows:

- Flask
- Dash
- Pandas
- Numpy
- Elastic Search

The deliverables for this stage include the following items:

- Sample small data snippet.
- Sample small output
- Working code
- Demo and sample findings
  - Data size: We had time series data for more than 4000 companies which led to about a millions records which were stored on AWS storage. All data was on the disk on AWS.
  - The tool we created could be used by professionals as well as individual people who make investments in the stock market. A normal worker who doesn't have time to read the news and follow the stock everyday, can come to this tool and see how the stock they are interested in has been doing over the past year or so and then compare and contrast to their current portfolio and see if this reduces their overall risk and get better results. This will help you not lose all your money on one stock investment you made on your whim or a relative's advice and are able to make educated and data backed decisions.

D. Stage4 - User Interface.

We have provided 3 tabs to the user:

1. Company Correlation
2. Industry Correlation
3. Sector Correlation

The user can click on any of the desired tabs. After clicking on the particular tab: The user will see the two text box where user can type the company names or Industry names or sector names based on which tab user has clicked. After giving the names of the companies user will see the plot on the left side of the window showing the stock values of both the entities. We have used orange color to represent one entity and blue color to represent the other entity so the user can easily compere the stock prices. On the plot user can zoom-in zoom-out, select the time window over which user wants to compare the prices. We have also provided one drop-down from which user can compare the plot based on opening value or the closing value. On the right side of the window user can see the articles related to the input companies he/she has given. At the bottom we have displayed one table which lists the top 25 most correlated or least correlated based on which option user choose from the drop-down.

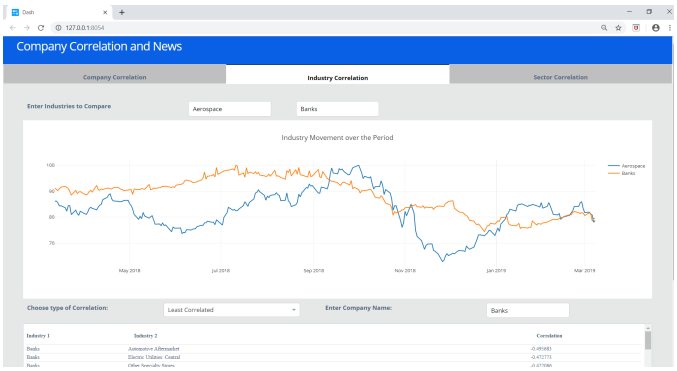


Fig. 1. Industry Correlation Tab

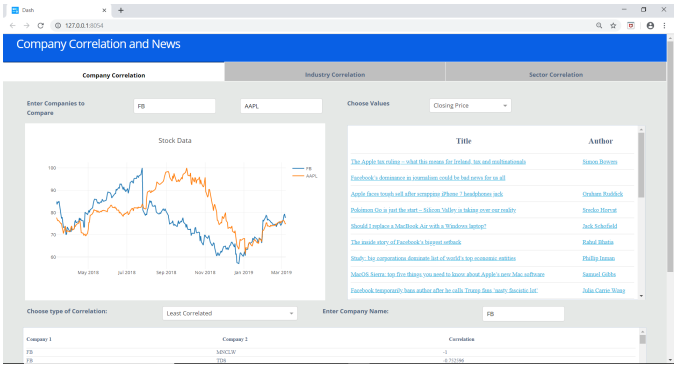


Fig. 2. Company Correlation Tab

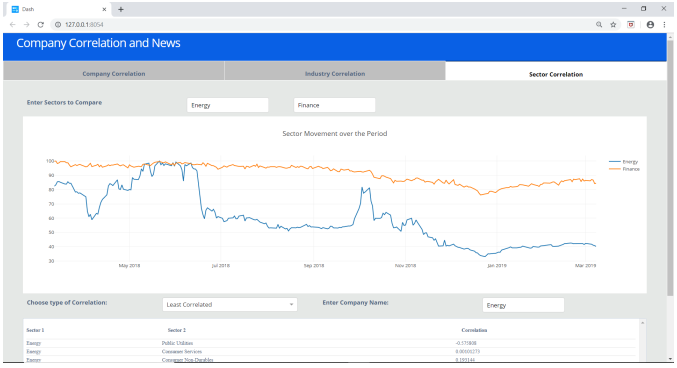


Fig. 3. Sector Correlation Tab

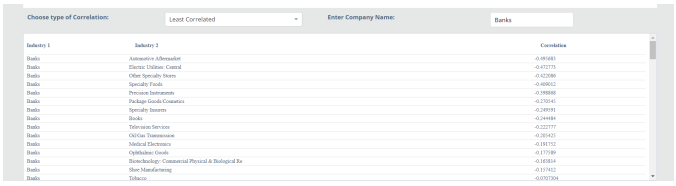


Fig. 4. Correlation Table