CIS 550 Project Proposal

https://github.com/ChengChen-Steven/CIS550_Final_Project/blob/main/document/Project_Proposal.md

Table of Contents

1. Team Member

- Wenting Zhao, Email: wentingz@seas.upenn.edu, GitHub: Wentinggg
- Ruikang Liu, Email: liurk@seas.upenn.edu, GitHub: ruikangpenn
- Puran Zhang, Email: puran@seas.upenn.edu, GitHub: puran-debugger
- Cheng Chen, Email: cheng24@seas.upenn.edu, GitHub: ChengChen-Steven

2. Application Description

The application we are building is a multifunctional web application that allows users to quickly retrieve information of stock price, financials, and other key stats for the companies of their choices. The app will also provide service for users, like searching similar stocks to follow or pushing recommendations of stocks/news to users based on their preference.

3. Dataset Overview

3.1 Data Source

We will use a dataset from financial market sites API, like Yahoo Finance/AlphaVantage, and a dataset from financial news sites API, like Twitter/Bloomberg, to filter out the data to actively inform users with price alerts and news/tweets of the companies they follow.

We will fetch the data from a variety of data sources, including but not limited to the following:

Yahoo Finance API: * Description: API that contains stock price information and company information. * Link: https://github.com/ranaroussi/yfinance

Alpha Vantage API: * Description: API that contains stock price information and company information. * Link: https://www.alphavantage.co/support/#api-key

Twitter Developer API: * Description: API developed by Twitter where we can search for the related tweets and news about a stock on Twitter, which can be further used for sentiment calculation and news feed information retrieval. * Link: https://developer.twitter.com/en/products/twitter-api/enterprise

Google Trends API: * Description: API developed by Google to provide some information on the popularity of the companies. * Link: https://trends.google.com/trends/?geo=US

3.2 Database Tables

We will create a few tables in our own database. * Stock/Company basic information (MySQL or MongoDB). * Stock historical trading information (MySQL). * Stock social feed information and news feed information (MySQL or MongoDB). * (TBD) User-level information, login credentials, login history, favorite stocks, portfolio management. (MySQL or MongoDB).

3.3 Outlook

Our data collection process is quite challenging given the complexity in data source unification, data quality sanity check, API reliability, job scheduling and failure retry. Considering different APIs might have constrains in usage, we will change our data injection accordingly based on the requirements so that we can retain some resilience should error happens.

In our application, we currently won't focus on intra-day minute level data availability given the resource restriction, we will first fetch the data after the market closes on each trading day. Thus, we will schedule a cronjob to fetch trading data from the data sources on a daily basis, and first inject the data into our database. Our application will utilize trading data directly from our own database.

We will likely start from a subset of US stocks, for example S&P 500 and later on ramp to all US stocks. In the future, we will consider the expansion to other asset categories, like international stocks, Gold/Silver, Futures, Crypto, and OTC. We will also consider integrating with intra-day data and possibly enabling trading API to place asset orders. In the meantime, we will be exploring some open-source information that we can collect from different sources.

4. Example Queries

Query #1:
Search for specific stock/company basic information/updates based on user's condition filter.

Query #2:
Inner join to get year-to-date cumulating return-on-investment of each stock.

Query #3:
Select top/bottom 20 performing stocks of the latest trading day to display on homepage.

Query #4:
Select top/bottom 3 performing stocks in each industry of the latest trading day to display on homepage.

Query #5:
Provide recommendations of similar stocks/companies based on similarity calculation (we will measure and quantify each stock from different aspects).