APS 360 Project Proposal - Smart Equity Analysis During a Time of Financial Uncertainty

Samuel Gaskin - 1004985215
Aakash Budhera -1004974272
John Perry - 1003912947
Brandon Liu - 1005389857

Overview

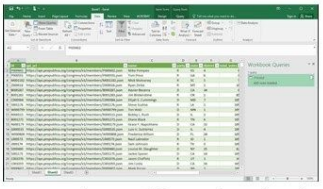
Our project is a smart equity analysis AI algorithm. The goal is to create a tool that can simulate the result of hours of research that bankers and analysts spend on valuation. This tool will predict the most common consensus “Buy/Hold/Sell” rating among analysts using a variety of financial stock data and we will compare our rating to the most common consensus to measure success.

This is a classic classification problem with multiple input variables. As such an artificial neural network is appropriate to solve this problem.

We will be using data From the first half of 2020. With the recent market volatility due to the coronavirus pandemic, this analytic model can help differentiate winning stocks from losing ones during a future economic crash.

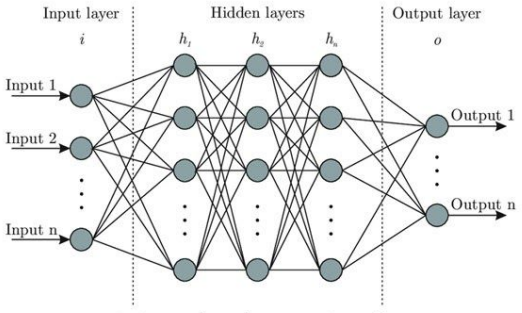
Illustration

Raw Economic Data



Preprocessing and Preparing the data

Outputs and Validation



Training and Implementation of ANN

Related Work

Financial prediction algorithms have a mixed history. There have been tremendous successes alongside massive failures. Two notable examples of financial projection models are:

1. **Long term capital management**
    a. $1.2 trillion Hedge fund propelled in part by complex computerized, or program trading, models [1].
    b. Several unexpected high sigma “black swan” events including the Asian financial crisis and the Russian government debt default of 1998 resulted in a -92% return after consistent ~40% returns since 1994 (average market returns were ~7%) [2].

2. **RBC Capital Markets “Aiden” Algorithm**
    a. reinforcement learning algorithm based product that can learn to perform volume-weighted average price (VWAP) stock execution
    b. Aiden has traded over 700 different stocks (Canadian, US, Europe), 50 billion dollars of capital since June 2018 and it is out-performing the existing VWAP execution algorithm [3].

Data Processing

The majority of our data will be taken from the bloomberg terminal. One of our team members has a bloomberg membership allowing us to use the bloomberg API in excel. We are able to pull historical and current data of our choosing into a csv file and then manipulate it how we want.

We will use ratios such as earnings per share ratio, price to earnings ratio, book value of equity per share, beta, and weighted average cost of capital. Ratios are useful because they are normalized such that they can be used to compare companies. We will also use measures of spread such as bid/ask spread, average change in price from 3:40 to MOC (Market on Close). We will normalize all dollars to bps (divide $ output value by a price at a constant starting time i.e +300 bps is an increase in 3%) so that we can compare companies with a huge discrepancy in value. Another important data piece we will be using from Bloomberg is the security sector (i.e technology, mining, retail ect.) as different sectors respond differently to different global affairs affecting market price.

An additional source of data will be from google trends. This tracks the volume of google searches of imputed key words over a specified time range.

Below is a visual representation of how the Bloomberg Terminal looks and how we would organize our data in excel. We can either type in specific stocks into excel or use built in functions to pull in groups of stocks from bloomberg. Then we can pull in specific data we want using the Bloomberg API and manipulate/normalize it in excel or python before we train our ANN.

Figure 1: Security Overview Page of Google in the Bloomberg Terminal

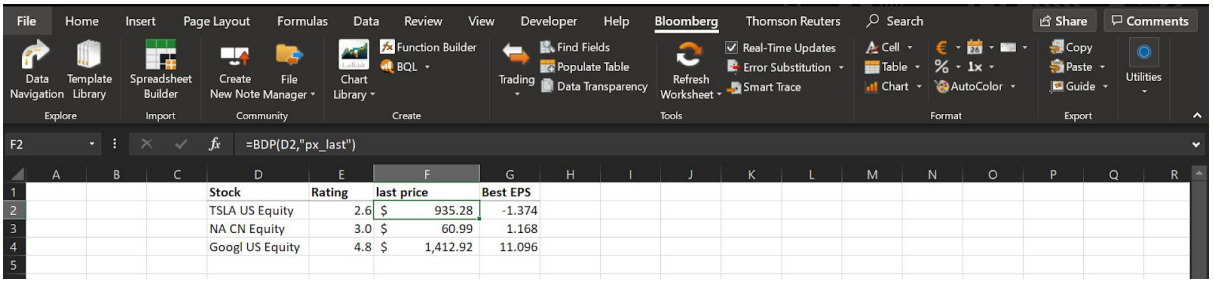


Figure 2: Representation of How the Bloomberg API Works in Excel

Architecture

Given that the issue we're trying to solve is a classification problem, an ANN will be the most suitable model for our project. Since our model will assign a stock a rating out of 5 (excluding 0), our program will have 5 outputs representing strong buy to strong sell ratings. The ANN will also have to contain multiple hidden layers, as it will be a nonlinear problem.

Baseline Model

Given that we will be comparing the results of our model to professional analyst ratings, a machine learning base model is not required. We can use hard coded inputs and outputs based on the ratings. Therefore, a simple python script will suffice to test our model's accuracy.

Ethical Considerations

Given the nature of how we're collecting the data, the data collection will only be possible for people or corporations who have access to very expensive financial tools. This means that a vast majority of people won't be able to complete the data collection at the

same level as us. This could be seen as an unfair advantage, as typically only wealthier individuals would be able to use this tool at its maximum effectiveness.

Project Plan

Our main source of communication is over facebook messenger. This is for day to day communication and updates. As many messages can be sent/lost in a group chat, we agreed that if you write an important message, we specify to “thumbs up” the message to show the sender it has been read.

We also use google teams for a more organized form of communication . We are able to schedule meetings through teams that go into our Teams and UofT mail calendars. The use of teams is to keep our meeting schedule organized, post updates on code progress, and organize larger deadlines we set as a group.

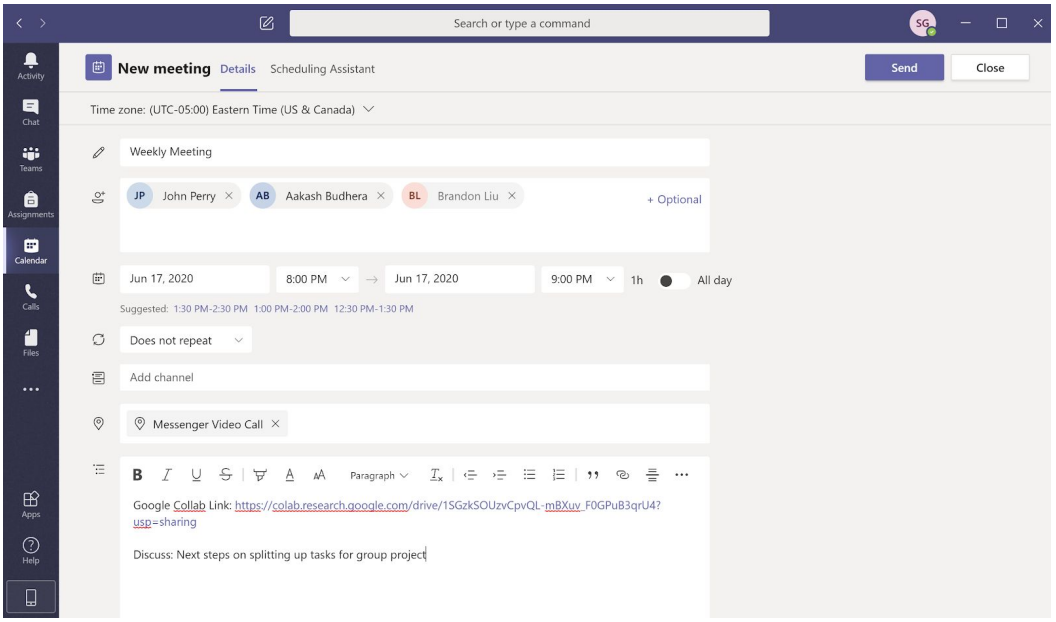


Figure 3: Samuel Gaskin Sending a Meeting Request to Team Members over Microsoft Teams and Outlook UofT Mail

As a group, we meet every wednesday night after our tutorial sesion. We will also plan extra meetings through teams when deadlines approach.

In terms of work distribution, Sam and Aakash are taking the lead on financial theory, and data collection/formatting, however, John and Brandon are also helping do research on data selection and formatting. John and Brandon are taking the lead on overall architecture for the ANN but Sam and Aakash will also be writing code and will help create this architecture. This way we have leadership roles within different aspects of the project but are all on the same page within every aspect.

Risk Register

One risk to consider in regard to our project are human factors influencing economic data. Often, humans are irrational resulting in behaviour against what is seemingly backed by evidence. Additionally, factors outside of just economic data can dramatically impact the performance of stocks such as international events, eg: corona virus and corona beer. Subsequently, these factors must be accounted for in some way or form prior to data processing to avoid inconsistencies in the data introduced to the neural network. We can account for this issue by analyzing data outside of financials including google searches.

Another potential risk when considering our project is either the overfitting or underfitting of our network when it comes to training sets due to the high variation across different stocks even within a common sector. For example, within any sector, smaller corporations are definitely much more subject to external market forces compared to the largest corporations. This could be addressed by implementing constraints on the input data in our data sets such as market cap and volume traded to mitigate these discrepancies.

A final risk that needs to be addressed would be a very low accuracy of our neural network even when the hyperparameters have been optimized. This would suggest that our current model is unable to provide adequate results with the data provided. Consequently, we could possibly narrow our evaluation to a specific sector or securities within a larger ETF to reduce variation in stock trends.

Link to Github or Colab Notebook

https://colab.research.google.com/drive/1SGzkSOUzvCpvQL-mBXuv_F0GPuB3qrU4?usp=sharing

References

[1] Traders Magazine Editorial Staff, "The Age of the Algorithm," 31 August 2004. [Online]. Available: https://www.tradersmagazine.com/news/the-age-of-the-algorithm/.

[2] W. W. Slome, "Investing in a World of Black Swans," 4 August 2009. [Online]. Available: https://investingcaffeine.com/tag/long-term-capital-management/.

[3] "David Shi," 14 June 2020. [Online]. Available: https://www.linkedin.com/in/david-shi-8526075b/?originalSubdomain=ca.