Predicting Buy Sell Patterns of Stock Market Experts

Authors: Nishant Jain
Nitesh Jaswal

Motivation behind this:

- We always see models trying to predict the stock market effectively which includes prediction of the stock prices, securities, bonds and other financial instruments to base the buy sell decisions
- What we wanted to do is a bit different, we wanted to predict the buy sell patterns of the so called 'stock market expert investors', e.g. Warren Buffet, Bill Gates

Challenges:

- Getting the dataset, as it is with the every problem in the world which we try to solve through models
- Expertise knowledge of finance domain
- Vast scope of modelling techniques

Data:

- We tried to find data in clean format using various resources but none worked out because it is hard to get financial data for free
- So we turned to web scraping to get the data, the website we used:
  - https://www.gurufocus.com/

Data Format:

- The data contained the buy, sell, new buy and sell out decisions by investors at different point in time. We picked 6 famous investors and used historical data till the point it was available (from Year 2010 – 2018) with quarterly frequency. This was to avoid the discrepancies due to financial crisis in 2008-09.

Model

Variables:

- Ticker (Stock), Price and growth(3 months and average), High, Low, S&P500 Index and growth (3 months), Investor ID, Current Stocks

Datasets:

- First data containing multiple stocks for multiple investors
- Second data containing only Apple stock for multiple investors
- Split both into train and test data, used most recent data (about 10%) as the test data due to this being a time series data, rest as the training data

Models:

- Used Keras framework in Python which utilizes tensor flow in the backend
- Fitted the data using multilayer neural network, and random forest
- **Results for 1st Data:**

Training samples: 9613 Unique Stocks: 1870 Unique Investors: 6
Testing samples: 1000 Unique Stocks: 461 Unique Investors: 6

Method 1 Neural Network:
Train Accuracy: 78.29% Test Accuracy: 72.8%

Method 2 Random Forest:
Train Accuracy: 77.69% Test Accuracy: 72.6%

- **Results for 2nd Data (Apple):**

Training samples: 479 Testing samples: 50

Method 1 Neural Network:
Train Accuracy: 55.11% Test Accuracy: 68%

Method 2 Random Forest:
Train Accuracy: 90.8% Test Accuracy: 90%

Next Steps:

- Use Q learning models to make better predictions, and train a model to do trading
- Incorporate more variables and more data
- Analyze the effect, applicability and practicality of "Impact on portfolio" variable, it was giving some unusual results like accuracy more than 90% using random forest for the 1st Data

References:

1. https://medium.com/mlreview/a-simple-deep-learning-model-for-stock-price-prediction-using-tensorflow-30505541d877
2. http://www.empirical.net/wp-content/uploads/2014/12/Barber-Buys-Holds-and-Sells.pdf
3. http://cs229.stanford.edu/proj2016spr/report/049.pdf
3. https://ropensci.org/tutorials/rselenium_tutorial/ - Selenium in R for Web Scraping
4. https://keras.io/ - ML models implementation in Python using Keras