CS 464 Introduction to Machine Learning 2021-2022 Fall Progress Report - Group 10

Göktuğ Çağıran (21803574), Kübra Okumuş (21600980), Cem Bingöl (21502436), Ulaş Anıl Eren (21501761), Mehmet Ali Altunsoy (21702531)

Project Title: Financial Data Estimator

1. Introduction and Background Information

As we progressed, we decided to make an update on our goal. As we started to work on the project, we noticed examining the behavior difference between time intervals in the stock market requires financial knowledge out of our scope. So instead of the time intervals, we decided to focus on future value prediction of the assets we choose. We believe our algorithm will be suitable to train with countless of asset data but for the sake of convenience and simplicity we'll name 3 of them officially. These are: bitcoin, ethereum and gold.

Bitcoin is the first cryptocurrency that is focused on decentralized and anonymous use. It was born in a forum as a proposal to modern economic assets by some user named Satoshi Nakamoto. After this people kept updating and improving the application of bitcoin. "Bitcoin was the first cryptocurrency to use blockchain and has been the market leader since the first bitcoin was mined in 2009" [1]. By the time bitcoin, has gained a lot of popularity and undergone rapid growth. Today there are numerous businesses that accept bitcoin as payment.

Ethereum, proposed as an improved version of bitcoin, is usually called the 2nd generation of bitcoin. It is an open source blockchain with smart contract functionality. It also has the 2nd biggest market cap coming just after bitcoin.Ethereum has other applications such as realizing creation and exchange of NFTs, namely digital works of art. "Ethereum is a type of cryptocurrency payment platform where transactions are completed by clients, with machines executing the requested operations. More broadly, ether is a crucial incentive that ensures that developers write and maintain quality throughout transactions" [2]. We will use this asset as well because it will help us to see the difference between cryptocurrencies with different features.

Gold was a worldwide asset that has been accepted since ancient times, before it was standardised as the basis for the international monetary system from the 1870s to the early 1920s. Gold has been a preferred form of money due to its rarity, durability, divisibility, fungibility and ease of identification [3]. Gold is not dependent on a currency because it's standard worldwide and stays unaffected by the fluctuations of specific currencies. Because of these characteristic attributes of gold, we decided to use it as one of our assets in this project.

For this project we will use the data of these assets from the "Alpha Vantage" database. We will evaluate the dataset we have by using the feature extraction method and the end product will be fear and greed index. Our goal is to implement the codes we are working on according to these

indexes and the nature of our dataset in the following weeks. We will make our tests using a cross validation method at the end.

Artificial Neural Network

Neural networks are suitable tools for performing financial forecasts. Some studies say that neural networks are more effective than other conventional models in most cases. Neural networks usually use historical data. Based on the previous data, future predictions are produced. The most basic principle when making predictions is to find a rule that will map past and future data appropriately. [4]

A multilayer perceptron (MLP) is one of the feedforward ANN. It is capable of learning non-linear models. MLP consists of at least three layers. Input, output, and nonlinear hidden layers. Each layer after input layer transforms the data to the next layer with a weighted linear summation. After the last hidden layer, the output layer receives the values and transforms them to output values. [5]

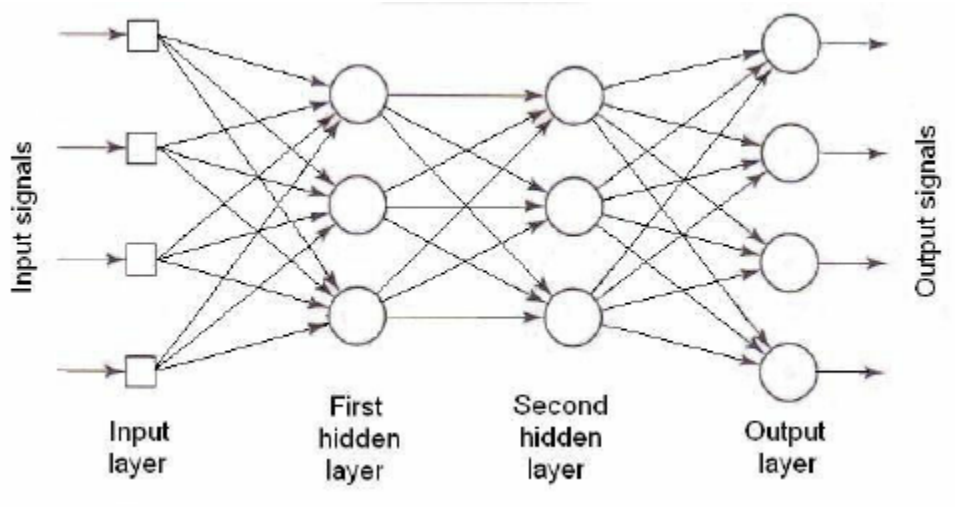

Figure: Example of MLP layers.

Usable Libraries:

- Scikit-learn: Used for setting and training svm and multi layer perceptron algorithms for regression.
- CSV: Used for handling .csv files.
- Pandas: Used for organising data.
- NumPy: Used for organising data.
- H5PY: Used for manipulating the data from NumPy.
- SciPy: Used for scientific computations.
- Matplot: Used for visualizing results to get a better look for efficiency and accuracy.

2. What is Done

We concluded our research about finance, stocks, commodities, crypto currencies and machine learning algorithms and applications being used in these areas. We also found open-source programs to examine and learn from. Based on these we tried to determine features which can forecast future values of commodities. Based on our research we decided to implement a Multi-layer Neural Network model.

We gathered information about crypto currencies and determined some features to build our model on. We decided to use data which had daily intervals. These features are:

- Opening Value: Value indicating the price at which a commodity traded with on the first exchange of the day.
- Closing Value: Value indicating the price at which a commodity traded with on the first exchange of the day.
- Daily High: Maximum price a commodity was traded in a day.
- Daily Low: Lowest price a commodity was traded in a day.
- Date: The date which the data points belong to
- Market Cap: Total value of all shares of a stock or a crypto currency.
- Dominance (Specific to cryptocurrency): The ratio between the market cap of Bitcoin to the rest of the crypto currencies.
- RSI (Relative Strength Index): Momentum indicator that indicates overbought or oversold situations of an asset or cryptocurrency.
- Fear and Greed Index: Index is calculated by examining several different factors to estimate how much fear and greed there is in the market. On this index 0 (zero) means “extreme fear” and 100 (one hundred) means “extreme greed”. For crypto currencies Fear and Greed Index is calculated from:
    - Volatility
    - Market Volume
    - Social Media: Twitter is analysed based on hashtags and how fast these hashtags get interaction.
    - Dominance
    - Google Trends data is analyzed for crypto currencies. (e.g. increase in “bitcoin price manipulation” lowers the index)
- New Addresses (Specific to crypto currencies): Number of new crypto wallets opened.
- Entities In Profit (Specific to crypto currencies): Number of wallets holding crypto currencies which are currently in profit.

We started to implement the code. First step was making research for libraries that we are going to use. After that we properly manage the data we found to simplify our work for further improvements.

3. What remains to be done

1. Research different activation functions currently being used in financial forecasting.
2. Implement algorithms using different activation functions.
3. Train ANN with data gathered and measure performance metrics.
4. Find the best activation function and implementation for financial forecasting
5. Report the findings.

4. Description of the Division of Work

We divide the work to sub work packages and contribute to different work packages.

WP#	Description of Work Package	Contributors
WP1	Background research.	Mehmet Ali Altunsoy, Göktuğ Çağıran, Ulaş Anıl Eren, Cem Bingöl, Kübra Okumuş
WP2	Development of the theoretical solution using mathematics.	Göktuğ Çağıran, Kübra Okumuş
WP3	Implementation of the algorithms with different activation functions.	Ulaş Anıl Eren, Mehmet Ali Altunsoy, Göktuğ Çağıran, Kübra Okumuş, Cem Bingöl
WP4	Training ANN with the data.	Göktuğ Çağıran, Mehmet Ali Altunsoy
WP5	Testing the model using the data reserved for testing and calculating performance metrics.	Göktuğ Çağıran, Mehmet Ali Altunsoy, Ulaş Anıl Eren
WP6	Report findings.	Ulaş Anıl Eren, Mehmet Ali Altunsoy, Göktuğ Çağıran, Kübra Okumuş, Cem Bingöl
WP7	Google CoLab integration and code trials.	Mehmet Ali Altunsoy, Cem Bingöl

References

[1] D. L. E. E. Kuo Chuen, L. Guo, and Y. Wang, “Cryptocurrency: A new investment opportunity?,” *The Journal of Alternative Investments*, vol. 20, no. 3, pp. 16–40, 2017.

[2] P. M., A. Sharma, V. V., V. Bhardwaj, A. P. Sharma, R. Iqbal, and R. Kumar, “Prediction of the price of Ethereum blockchain cryptocurrency in an industrial finance system,” *Computers & Electrical Engineering*, 03-Dec-2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0045790618331343?casa_token=tFXpwTzwGnkAAAAA%3Al_vdIl5yg-VGSMQ5Ax38dNAC_OrT6VyaIl9_Zz0mnFrhuJ6QPqUU3ibyBbZZt1-iamW0fgBlxg. [Accessed: 27-Nov-2021].

[3] “Gold standard,” *Wikipedia*, 27-Nov-2021. [Online]. Available: https://en.wikipedia.org/wiki/Gold_standard#Theory. [Accessed: 28-Nov-2021].

[4] J. T. Yao, “Guidelines for financial forecasting with neural networks,” Uregina.ca. [Online]. Available: http://www2.cs.uregina.ca/~jtyao/Papers/guide_iconip01.pdf. [Accessed: 28-Nov-2021].

[5] “1.17. Neural network models (supervised),” Scikit-learn.org. [Online]. Available: https://scikit-learn.org/stable/modules/neural_networks_supervised.html. [Accessed: 28-Nov-2021].