CSE 546 Reinforcement Learning
Final Project Submission
Stock Market Prediction

Sruthikeerthi Rajagopalakrishnan
UB Person Number : 50359100
rajagop2@buffalo.edu

Sarveshwar Singhal
UB Person Number : 50418642
sarveshw@buffalo.edu

SaiKrishna Dirisala
UB Person Number : 50411928
sdirisal@buffalo.edu

Abstract

In this project, we have developed a stock market prediction system where the model learns the best course of actions which leads to the maximum profit. We used deep reinforcement learning algorithm, namely actor-critic model to learn this sequence. We get the total amount matured in the given course of days as the output.

1 Problem Statement

Stock market has a lot of nuances that has to be captured and interpreted while investing in the market. This project focuses on finding the most optimal way to earn a good amount of profit in the stock market. A reinforcement learning agent is a good option to learn the trends and implications to make a series of good decisions to earn the maximum or safe profit.

2 Literature Review

A stock prediction model using TD(0) models was presented in Jae Won Lee (1). In this work, Korean stock market data was considered as an infinite state model with different policies for each investor. R. Sathya et. al (2) describes a Partially Observable MDP model for predicting stock market with real-time and historic datasets. A. Moghar (3) describes an LSTM based RNN model to predict stock prices. They state the LSTM models hold memory of the previous stock nature to predict the current profit.

3 Project Pipeline

There are five basic steps involved in the implementation of this project. The flow diagram is given in Figure 1

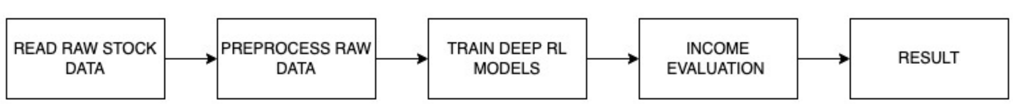
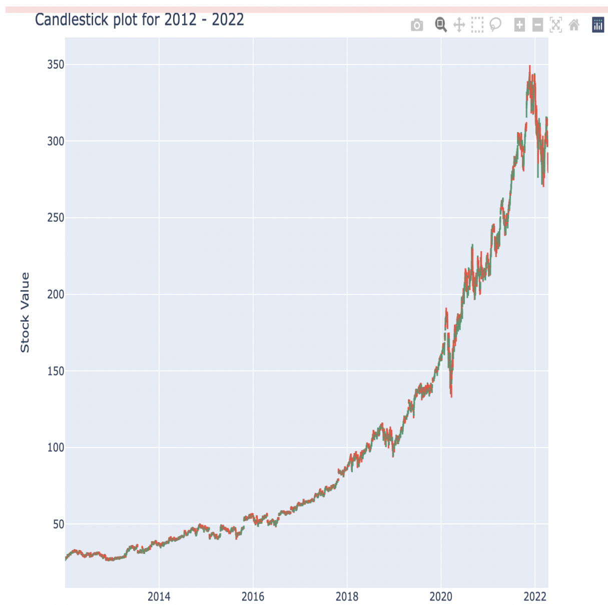


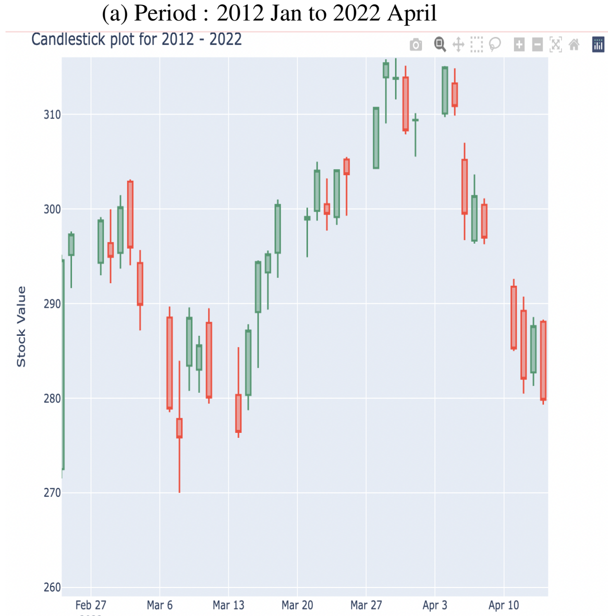
Figure 1: Data pipeline used in this project

3.1 Data Collection

The dataset was collected from Yahoo! Finance open source API. We chose Microsoft as our target company and collected stock market prices of the company since the inception of the company. The candle-wick plots for the data collected are given in Figure 2. We can see from Figure2a that the prices of the stock have a general trend of going up with upon inspecting a smaller section of data, as given in 2b, we can see the day-to-day tendencies of the stock requires regular monitoring and intense background study to predict the course of the price.



(a) Period : 2012 Jan to 2022 April



(b) Period : Feb 2022 to April 2022

Figure 2: Candle-wick plots for the collected MSFT stock data

3.2 Data Preprocessing

There are steps defined in preprocessing the stock prices dataset. They are

1. Finding the observations: In this step, we compute the rates of changes across the day's closing value w.r.t opening value, highest value w.r.t lowest value, lowest value w.r.t closing value and the opening values w.r.t the next day.

2. Normalizing the observation: In this step, we normalize the values to fit between zero and one.

3. Assigning the target value: Based on the observations computed with the rates of changes, the target value, i.e BUY, SELL or HOLD actions are assigned. If the rate of change of the closing value w.r.t opening value and the opening value w.r.t next day's opening value are both positive, then BUY action is assigned. If the rate of change of the closing value w.r.t opening value is positive but the opening value w.r.t next day's opening value is negative, then it's best to HOLD the stocks. If neither of these conditions are satisfied, then the SELL action is defined.

4. One hot encoding: The target values are one hot encoded.

3.3 Model Training

We have used Deep RL Actor Critic model to learn the good sequence to earn the best profit. The environment is design such that there is continuous state space where every state value consists of four components, opening value, closing value, highest value and lowest value of the stock in a day, a discrete action space with three actions and a corresponding reward metric.

1. BUY: The action BUY is given an intermediate reward of -7.

2. HOLD: The action HOLD is given an intermediate reward of 10.

3. SELL: The action SELL is given an intermediate reward of 5.

We used Keras module in Python to develop the neural network for the Actor and Critic. We introduce another neural network, ”Policy”, with similar architecture to the Actor network to learn the policy and predict the actions while computing the advantage for the Actor network. The architecture used for these three functions have one hidden layer of 32 units and an output layer with either one units of 3 units (i.e number of actions). The hidden and output layers in the Critic and ”Policy” networks have ReLu activation functions and for the Actor network has a softmax activation function in the output layer. The Actor and Critic networks are trained using the A2C algorithm with Adam and SGD optimizers over the entire dataset.

3.4 Income Evaluation

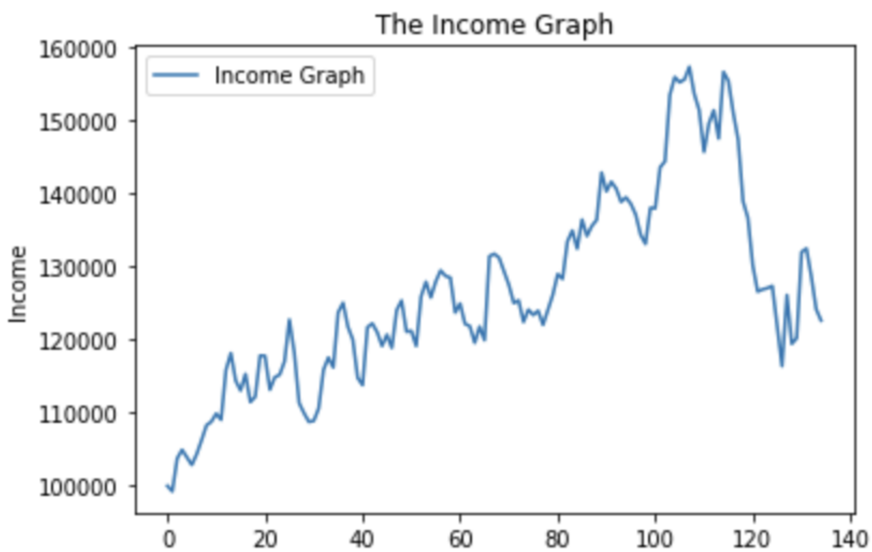
After obtaining the trained networks and rewards corresponding, the income was evaluated based on the number of shares bought, stock price when bought or sold and deducting taxes when sold.

4 Results

The model learns the policy and gives a testing accuracy of 70%-100% for different stocks. A test case scenario with a sample income value is given in Figure 3.

```
Enter the investment Capital:100000
Enter the number of days: 500
```

```
[205]: object_AC.Render_income()
```




```
[206]: object_AC.income

[206]: 122565.53411865234
```

Figure 3: Results of the Actor-Critic algorithm for a sample income of $100000 over a period of 500 days

We can see that the agent performs very well by earning a good profit till day 100 but drops a bit in profit after 100th day, but earns an overall profit in the period specified.

5 JIRA Board

We have 19 tasks in our JIRA board for this project. The status of the board after the checkpoint submission is given in Figure 4. The status after completing the project is given in Figure 5. A few tasks such executing double DQN and dueling DQN were left incomplete as the intermediate results were no favourable, hence we didnt pursue the model further. The paper mentioned solved the environment using $TD(\lambda)$ which was given unfavourable.

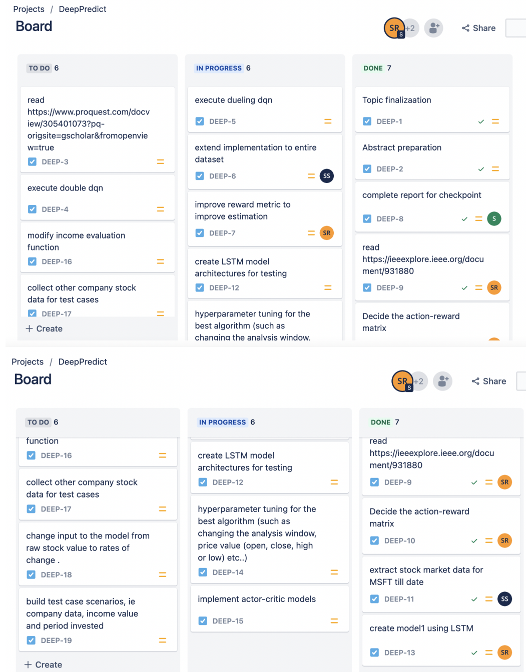


Figure 4: Jira Status after the checkpoint submission

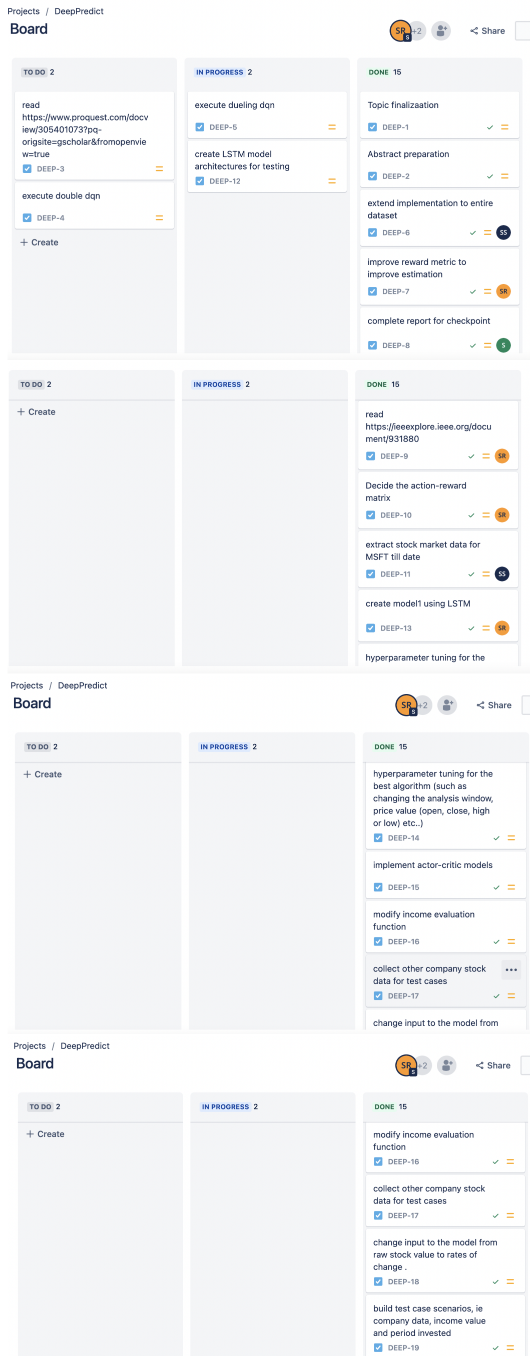


Figure 5: Jira Status after the completion

6 Group Contribution

Team Member	Contribution
Sruthikeerthi	33.33%
Sarveshwar	33.33%
SaiKrishna	33.33%

Table 1: Group Contribution

References

[1] Jae Won Lee, ”Stock price prediction using reinforcement learning,” ISIE 2001. 2001 IEEE International Symposium on Industrial Electronics Proceedings (Cat. No.01TH8570), 2001, pp. 690-695 vol.1, doi: 10.1109/ISIE.2001.931880.

[2] R. Sathya et al. “Stock Price Prediction using Reinforcement Learning and Feature Extraction” International Journal of Recent Technology and Engineering, vol. 8, no. 6. Blue Eyes Intelligence Engineering and Sciences Engineering and Sciences Publication - BEIESP, pp. 3324–3327, Mar. 30, 2020 [Online]. doi : http://dx.doi.org/10.35940/ijrte.F8606.038620.

[3] A. Moghar and M. Hamiche, “Stock Market Prediction Using LSTM Recurrent Neural Network,” Procedia Computer Science, vol. 170. Elsevier BV, pp. 1168–1173, 2020 [Online]. doi: http://dx.doi.org/10.1016/j.procs.2020.03.049.

[4] Mnih, Volodymyr, et al. ”Asynchronous methods for deep reinforcement learning.” International conference on machine learning. PMLR, 2016.