HIGH LEVEL DESIGN DOCUMENT

DEEP RL BASED STOCK PORTFOLIO PREDICTION USING CANDLESTICK CHARTS

UE17CS490A – Capstone Project Phase – 1

Submitted by:

Rishabh Jain **<PES1201700099>**

Phalguni Kamani **<PES1201700759>**

Vidhaan Mishra **<PES1201700055>**

Under the guidance of
Prof. Ramamoorthy Srinath

PES University

August - December 2020

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
FACULTY OF ENGINEERING
PES UNIVERSITY
(Established under Karnataka Act No. 16 of 2013)
100ft Ring Road, Bengaluru – 560 085, Karnataka, India

TABLE OF CONTENTS

<u>Note:</u>

Section – 1 & Section 2	**Common for Product Based and Research Projects**
Section 3 to Section 11	**High-Level Design for Product Based Projects.**
Section 12	**High-Level Design for Research Projects.**
Appendix	**Provide details appropriately**

1. Introduction

The High Level Design Document highlights the concept and necessary detail for implementing a Deep RL based stock portfolio. Candlestick charts are essential when it comes to trading and it provides a lot of insights into the market. Infact, Candlestick is representative of the entire stock movement in that given time period and can be very insightful when closely examined. We believe that a candlestick chart is a more "complete" representation of the stock movement. Hence, we plan on exploring research opportunities in this particular topic.

2. Current System [if applicable]

Unpredictability and high risks of the stock market have encouraged numerous researchers to explore ways to better understand this phenomena. There have been multiple works that have gone into predicting "the best" stock portfolios for the users, using the historical time series data that are readily available. When it comes to the stock market, there are a lot of factors such as recent company policies, public sentiment, etc. that also play a role in determining the direction of the market. As we all know, there is no "best" way to determine the ideal stock portfolio and hence we want to research another methodology and observe the results.

3. Design Considerations

3.1. Design Goals

- We believe that a candlestick chart is a more "complete" representation of the stock movement. It traces the stock movement over a period of time which incorporates the market sentiments, and public perception of the company. Hence, we feel that building an AI that can read and understand these charts would perform better than just reading numbers.
- Candlestick charts are an essential when it comes to comprehend the market and make some trading decisions by the traders.
- We want to imitate the human understanding of the charts and based on comprehending the visual, and then make intelligent trading decisions.
- Hence, we have a CNN model that serves the purpose of the "visual analysis" and the RL agent which can be analogous to the "intelligent decision-maker".
- Our aim is to optimise the return within the given user time constraint. The user provides us with an expected ROI and a time frame within which our

model will arrive at an optimal portfolio that maximises the returns for the user.

3.2. Architecture Choices

CNN

Pros

- CNN's are extremely effective for image recognition. We are going to feed in chart images into the CNN and hence we choose CNN as our first architecture.
- CNN can update its convolution kernel by backward propagation and train the appropriate weights to extract excellent image features.

Cons

- Computationally expensive during the training time.
- With candlestick charts, there is always a chance of over/underfitting of the model.

RL

Pros

- Reinforcement Learning can very quickly learn based on the actions it takes and make very smart decisions. They have also shown to outperform humans in many thinking-intesnive "games".
- By defining the reward function as the change of the portfolio value, Deep Reinforcement Learning maximizes the portfolio value over time.
- The stock market provides sequential feedback. RL coupled with a CNN (Deep RL) can sequentially increase the model performance during the training process.

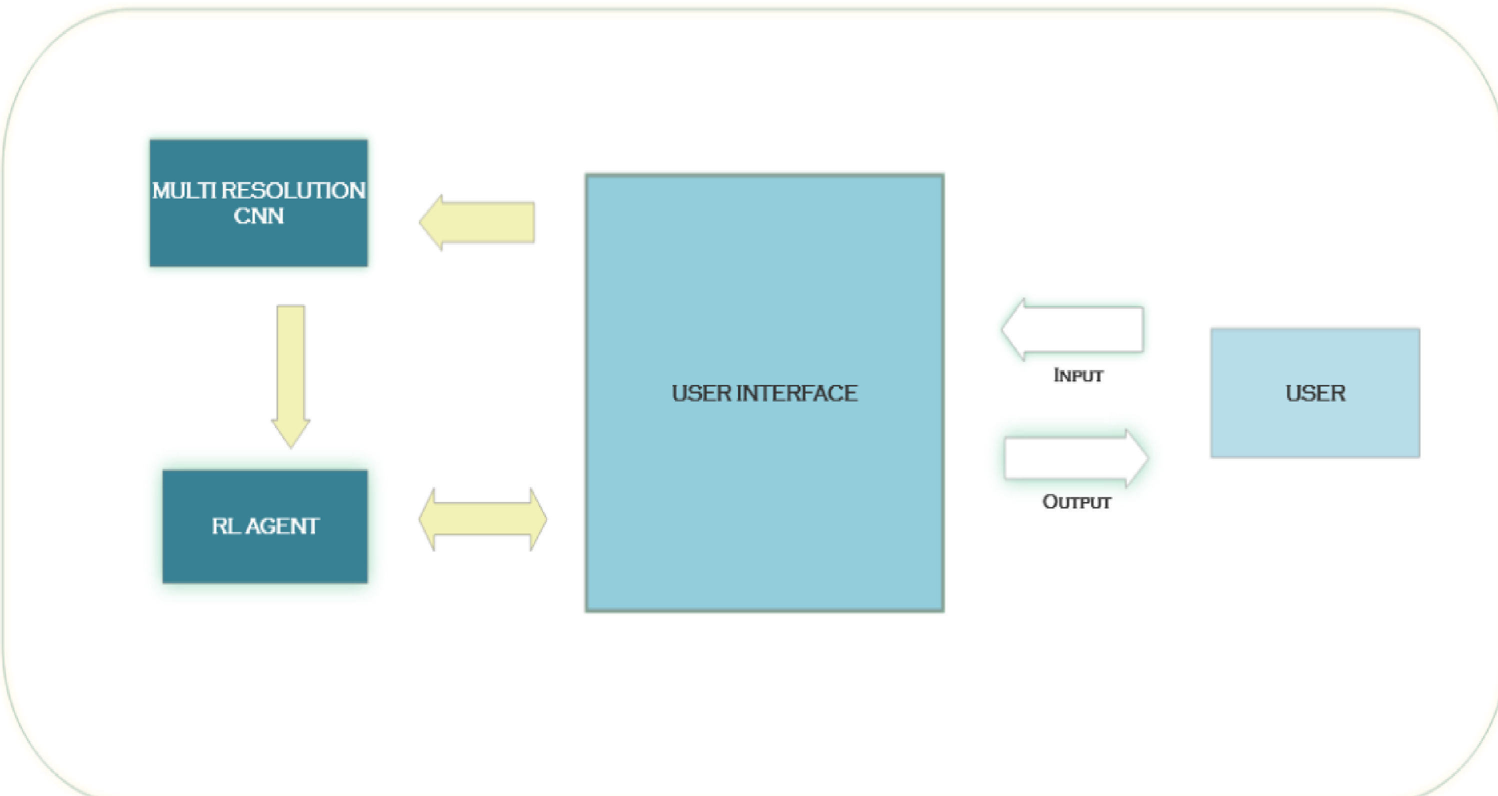
Cons

- Reinforcement Learning algorithms such as Q-Learning can be computationally expensive over a larger action space, however, after being coupled with a CNN it performs significantly better.
- For the task we want to carry out, there seems to be no feasible alternative to the DRL that we plan on using.

3.3. Constraints, Assumptions and Dependencies

- Data repository and distribution requirements
  - Expect stock to have hourly and yearly datasets available
- Discuss the performance related issues as relevant.
  - As size of data increases, and more operations are performed parallelly the number of GPU's and CPU's will need to be increased.
- Availability of Resources
  - Need large amounts of data for the particular stock to perform analysis
- Hardware Requirements
  - Dedicated GPUs to accelerate the training time and performance of Neural Networks.
- Software Requirements
  - Keras - 2.3.1
  - TensorFlow -2.0.0
  - MatplotLib -3.1.1
  - Python - 3.7.7
  - Google Colab
- Assumptions
  - All stocks generally follow a particular pattern which can be identified and hence used in predicting future prices of the stock.
  - A threshold on the loss is kept. This will force the bot to make a sale before the stock value dips further and results in a massive loss to our users.

4. High Level System Design



Logical User groups and characteristics :

- **Individual Trading Enthusiasts**

  Individual traders can be characterised as individuals who have an interest in
  trading. People not knowing the nitty-gritties of stock trading can make use of our product to make profits without getting too involved in the process.

- **Stock Brokers**

  A stock broker is a professional trader that buys or shares trades on behalf of the clients. Stock brokers can compare their trading strategies to that of the bot and see if their decisions can be made better or not.

- **Trading Bots**

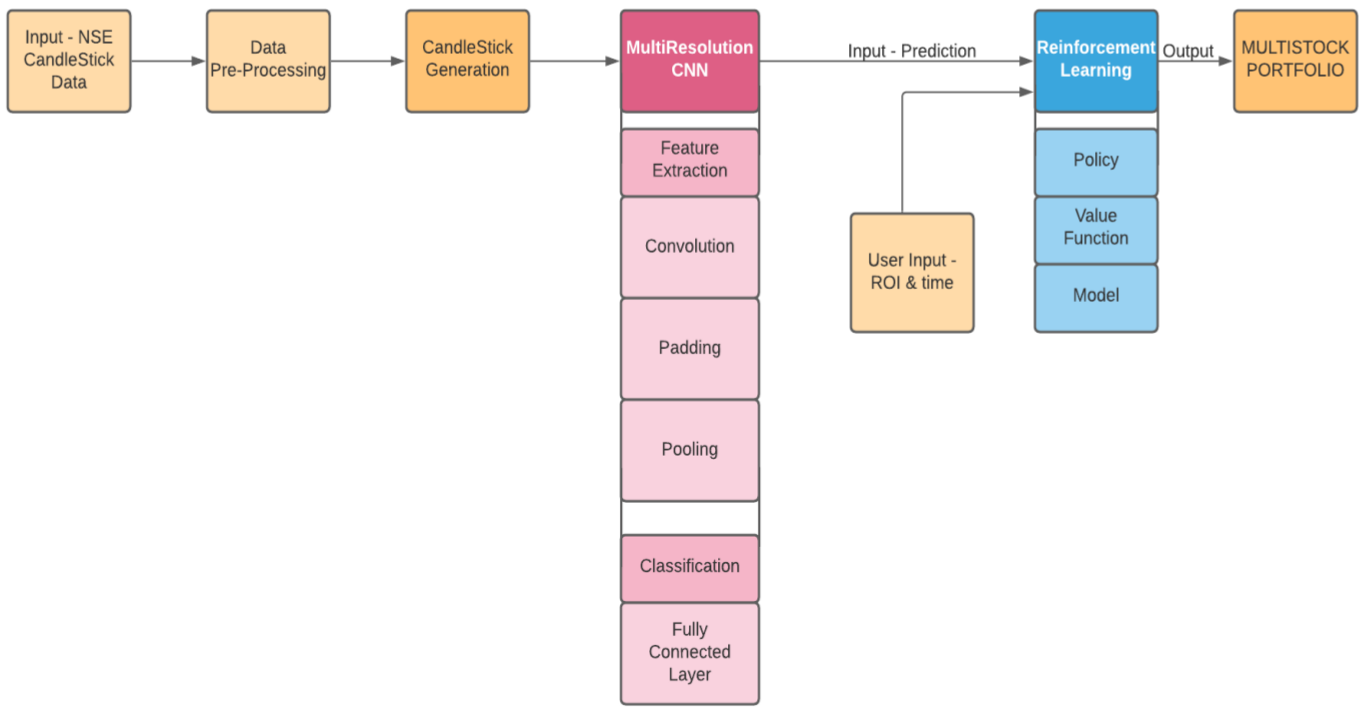
  Trading bots can also be compared with each other in terms of their overall performance, strategies, etc.

Data components

- Nifty-Fifty DataSet procured from Kaggle(Daily) and YFinance(Hourly)
- The fields it constitutes of are :
  - Date
  - Open
  - High
  - Low
  - Close
  - Volume

5. Design Description

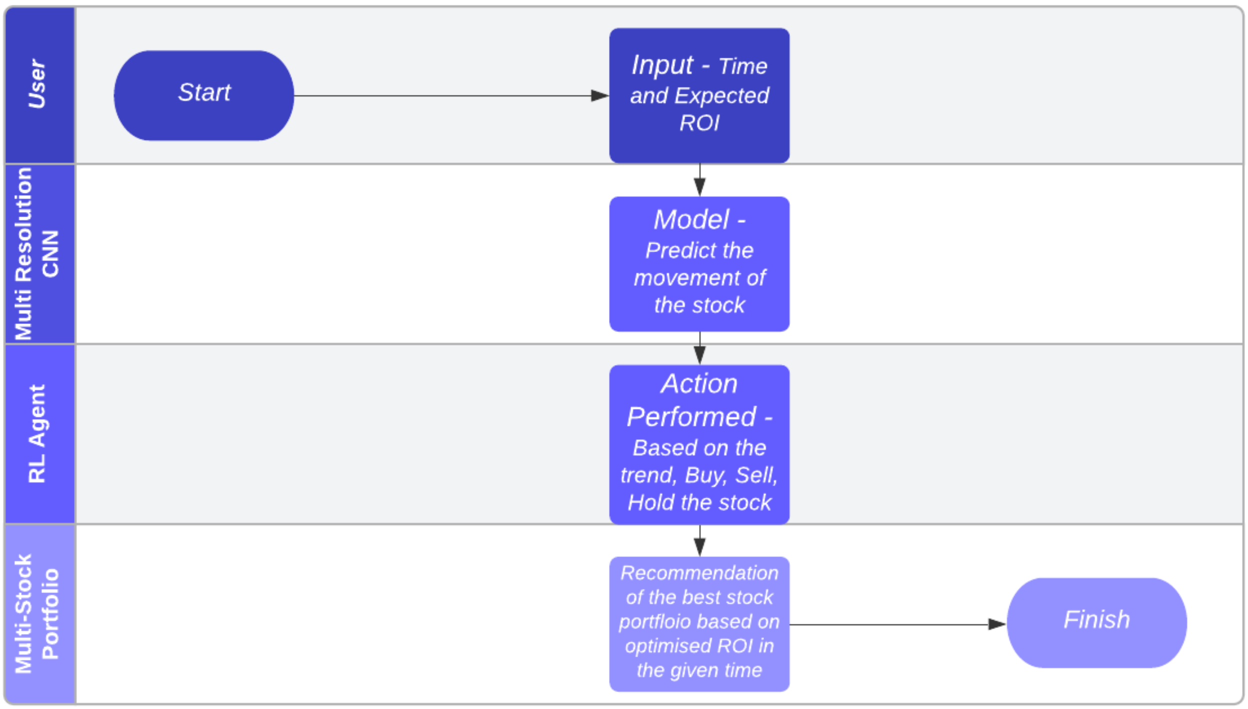
5.1. Master Class Diagram



5.2. Reusability Considerations

- Yahoo Finance Data gets refreshed daily. The code written for the model to download hourly stock data can be run repetitively to refresh our dataset with the next day's data by providing a new Start date and End date.
- Our neural network and RL agent can be re-used for any stock in addition to Nifty 50.

6. Swimlane Diagram



7. User Interface Diagrams

The user-friendly interface provides :

- A login page to gain access to their account
- View Stock data (Browse the different CandleStick charts of their choice)
- Input the Time and ROI
- The future predictions for the stocks

8. Packaging and Deployment Diagram

- The entire code and data files will be zipped and shipped. The requirements.txt file present in it can be installed first after which the model can be run.

ALTERNATIVELY

- It can be procured from Google Drive

9. Help

- A User Manual Page will be provided on the Website for giving a brief idea to the users on how to go about using the application and preventing them from stalling at any point.
- As the interface will be the only point of interaction, a technical manual will not be required.

10. Design Details

1.1. Novelty -
CNN architecture is combined with RL agent.

1.2. Innovativeness -
Different types of datasets will be run on our model to decide which works best for accurately predicting stock movement, not just changing parameters of the CNN and RL agent.

1.3. Interoperability -
At present, the model focuses on 50 different Indian Stocks, part of Nifty 50.

1.4. Performance -
Dedicated GPUs to accelerate the training time and performance of Neural Networks.

1.5. Security -
Login Page implemented so that the user has to verify his/her identity to access the account.

1.6. Reliability -
Highly accurate results of above or equal to 80% to predict stock movement.

1.7. Portability -
Model can be deployed on any hardware/software as long as the requirements have been downloaded.

Appendix A: Definitions, Acronyms and Abbreviations

CNN (Convolutional Neural Network) - A type of Artificial Neural Networks used in image recognition and processing that is specifically designed to process pixel data.

UI (User Interface) - Refers to the visual elements involved in the interaction between the user and the system.

RL (Reinforcement Learning) - It is an area of machine learning concerned with how software agents ought to take actions in an environment in order to maximize the notion of cumulative reward.

DL (Deep Learning) - A broader family of machine learning methods based on artificial Neural Networks with representation of learning.

DRL (Deep Reinforcement Learning)

NSE (National Stock Exchange)

Appendix B: References

[1] Chen, J., Tsai, Y. *Encoding candlesticks as images for pattern classification using convolutional neural networks*. Financ Innov, (2020)

[2] Azhikodan, Akhil & Bhat, Anvitha & Jadhav, Mamatha. (2019). Stock Trading Bot Using Deep Reinforcement Learning. 10.1007/978-981-10-8201-6_5.

[3] P. Bhandia, V. Yamnur, K. Rakesh and S. S. Shylaja, "Augmenting Deep Learned Representation Based Portfolio Selection With Predictive Shallow Networks," 2019 Fifteenth International Conference on Information Processing (ICINPRO), Bengaluru, India, 2019

[4] Rosdyana Mangir Irawan Kusuma, Trang-Thi Ho, Wei-Chun Kao, Yu-Yen Ou and Kai-Lung Hua. *Using Deep Learning Neural Networks and Candlestick Chart Representation to Predict Stock Market*. (2019)

[5] Hu, Guosheng, et al. ”Deep Stock Representation Learning: From Candlestick Charts to Investment Decisions.” 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018.

[6] Marc Velay and Fabrice Daniel. *Stock Chart Pattern Recognition with Deep*

Learning. (2018)

[7] Lee, Sang Il, and Seong Joon Yoo. ”A new method for portfolio construction using a deep predictive model.” Proceedings of the 7th International Conference on Emerging Databases. Springer, Singapore, 2018.

[8] Fischer, Thomas, and Christopher Krauss. ”Deep learning with long short-term memory networks for financial market predictions.” European Journal of Operational Research 270.2 (2018): 654-669.

[9] Saad Albawi, Tareq Abed Mohammed, and Saad Al-Zawi. ***Understanding of a Convolutional Neural Network***. (2017)

[10] C. Olah. ***Understanding LSTM Networks***. Available : http://colah.github.io/, (2015)

[11] C.-F. Tsai and Z.-Y. Quan. ***Stock prediction by searching for similarities in candlestick charts***. ACM Transactions on Management Information Systems (TMIS), (2014)

[12] A. V. Devadoss, T. A. A. Ligori, "**Forecasting of stock prices using multi layer perceptron**", Int J Comput Algorithm, vol. 2, pp. 440–449, 2013.

[13] H.A.doPrado, E.Ferneda, L.C.Morais, A.J.Luiz, and E.Matsura. On the effectiveness of candlestick chart analysis for the brazilian stock market. Procedia Computer Science, (2013)

[14] Krizhevsky, Alex & Sutskever, Ilya & Hinton, Geoffrey. ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. (2012).

[15] G. L. Morris. Candlestick Charting Explained: Timeless Techniques for Trading Stocks and Futures: Timeless Techniques for Trading stocks and Sutures. McGraw Hill Professional, (2006)

[16] Sutton, R.S., Barto : A.G., Reinforcement Learning: An Introduction in Advances in Neural Information Processing Systems, MIT Press (1998)

[17] Sharpe, William F. ”The sharpe ratio.” Journal of portfolio management 21.1 (1994): 49-58.

[18] E. Schöneburg. Stock price prediction using neural networks: A project report. Neuro- computing, (1990)

[19] Markowitz, Harry M. ”The early history of portfolio theory: 16001960.” Financial Analysts Journal 55.4 (1999): 5-16.

[20] Harry Markowitz, “Portfolio selection,” The journal of finance, vol. 7, no. 1, pp. 77–91, 1952.