SCHOOL OF COMPUTING
Faculty of Engineering

Project Proposal Form MCST1043
Sem: 2 Session: 2024/25

SECTION A: Project Information.

Program Name:	**Masters of Science (Data Science)**
Subject Name:	**Project 1 (MCST1043)**
Student Name:	LEE HONG JIAN
Metric Number:	MCS241054
Student Email & Phone:	leehongjian@graduate.utm.my & +60127306995
Project Title:	Reinforcement learning for automated trading in stock markets
Supervisor 1:	
Supervisor 2 / Industry Advisor(if any):	

SECTION B: Project Proposal

Introduction:

Machine learning (ML) is the part of artificial intelligence (AI) which is ML mainly on creation of algorithms and models to let the agents learn from the data input and improve the performance over time without explicit programming (Vec et al. 2024). Generally, machine learning can divide into 3 parts which are supervised learning, unsupervised learning and reinforcement learning. Deep reinforcement learning (DRL) is where the agents are able to learn for making the decision by interphase with the real-time situation (environment) in the form of reward mechanism which is reward or penalty action will be based on the action taken by agent (Barto et al. 2025).

The aim is to create the policy that can maximize the cumulative reward over time. RL can be used in many different fields such as agriculture, trading, food, and, et al (Georg et al. 2024). The benefits of DRL in trading are adaptability, improved decision making, automation and optimization, and, et al. DRL was enhanced adaptability in trading by the dynamic's algorithms that can be adjusted based on the real-time environment (market condition). Based on the algorithms, agents able to develop the strategic that suitable used for the particular situation due to volatility of market and "new" trends (Huang et al. 2024).

The learning of market interaction allows agents to improve the decision making and maximize the return rate (Sangve et al. 2025). Based on the adaptive developed and trading of data-driven strategy, the effective of RL was performing well than the original and the potential of RL in decision making of forecasting. RL agents maximized reduce the resources used and the trading process will be more effective and efficient in presence of automation (Kabbani and Duman, 2022).

In trading field, RL was playing the importance roles in decision making because its will not distributed by emotional. An experienced trader may be affected by emotional by decision making however DRL can totally prevent the incident happen. The high potential of RL in trading especially due to the dynamics market floating (Kabbani and Duman, 2022). In trading field, RL was playing the importance roles in decision making because its will not distributed by emotional. An experienced trader may be affected by emotional by decision making however DRL can totally prevent the incident happen. The high potential of DRL in trading especially due to the dynamics market floating (Kabbani and Duman, 2022).

Problem Background:

In the reality of the financial market, the traditional trading strategy is based on the analysis (technical and fundamental) and statistical models to make decision, however, the limitation of traditional trading strategy also significant which when dueling with the highly dynamics of complex market. The traditional strategy relies on the assumption, rule-based, and predefined heuristics (sentimental) which will cause inflexibility and susceptibility toward market. In short, the traditional strategy often low capability in adaption of new data toward the dynamics market environment in the results the suboptimal performance while the high volatility of market (Chen et al., 2022).

In trading field, DRL was playing the importance roles in decision making because its will not distributed by emotional. An experienced trader may be affected by emotional by decision making however DRL can totally prevent the incident happen. The high potential of DRL in trading especially due to the dynamics market floating (Kabbani and Duman, 2022). The emotional biases occur often in the decision-making phase at the financial market which can affect the overall outcome significantly (Aziz et al. 2024). The human behaviors such as loss aversion, overconfidence, regret aversion, and herding behavior, leads to lose step investment decisions (Rafandito et al. 2024).

The traditional algorithmic trading strategy was not demanded in the growing complexity of financial markets in global. DRL have introduced into the trading market due to the innovation method to increase the outcome of trading with minimize the emotional biases. In automation decision making process, DRL able to learn from the interaction of market and make the improved decision based on the long-term rewards instead of the short-term reaction to market fluctuation. The ability of DRL is consistent improve the strategy without the influenced by emotional biases in trading. Overestimate, overconfidence, overreaction towards the news and fear of loss are the main human behavior that will affect the emotional of trader and leads to wrong decision making (misjudgment). The investors will loss the optimal strategy due to emotional biases in the ends the stability of market and the return rate will be affected (Huang et al., 2024).

DRL is one of the solutions for the emotional biases in financial markets. Compare to original/traditional systems, DRL algorithms able to learn directly from the actual market interactions and able to adjust the strategy to maximize the return rate based on the dynamics market. Without intervene of human, DRL agents able to learn and adapt to delimited the factor of emotional biases in trading. By minimize the human intervene, enhance the trading strategy thereby the emotional biases will be reduced. The performance of the DRL was better than human traders as the consistency and prevent the impetuous decision in the turbulence financial market (Sangve et al., 2025). The critical issues of trading decision are emotional biases in the trading markets. The performance of traders and stability of market will be affected. Automation decision making process will minimizes the emotional biases. By using of DRL, traders able to often improved financial strategy that increase the performance over time.

The famous traditional trading strategy called Golden Cross strategy which is the traders will buy in stock when the 50-day moving average cross over the 200-day moving average and sell out when reverse occurs. The traditional strategy is simple and direct in the stable market environment only because its low adaptive toward the dynamics market environment. However, DRL are able to working consistent and well in dynamics market environment such as the Two Sigma in DRL where the system will keep adapts the new various high amount of data and identify the profitable strategy that traditional trading strategy that unable to make it. The high profitability and adaptability without influenced by sentimental factor in the highly dynamics market environment (Huang et al. 2024).

Problem Statement:

Nowadays, the main problem of the trading is the human emotional biases which is affected in the decision making and decrease the trading strategy of effective. Those human behaviors lead to the inconsistent of the human traders, impetuous action and irrational. Developing the system with minimize the human intervene meanwhile automated learning and adaption with the complexity of global financial markets is the main challenges. The potential solution that can overcome the emotional biases is by the automation decision making and learning directly from the interaction global financial market (Huang et al., 2024). Exploration of the used of DRL to minimize human emotional biases in trading. DRL develop and implement based on the trading strategy that able to learn and adapt directly to the real-time global financial market to eliminate the emotional biases in decision making and replaced with the DRL the traditional system. The estimated outcome is to replace traditional trading strategy and increase the return rate over time instead of making decision based on the emotional biases and dynamics market (Sangve et al., 2025). The Self-rewarding deep reinforcement learning (SRDRL) is the Model that created to increase the profit and efficiency by grants challenges with the reward mechanism for agent to learning. The key features of SRDRL such as self-rewarding mechanism, improved efficiency of learning, algorithmic trading of application, exploration and exploitation and reward shaping. There are some model for the SRDRL such as deep Q-network (DQN), proximal policy optimization (PPO) and soft actor-critic (SAC). In the learning paradigm, DQN is based on the value which is mainly focus on the action value estimation however PPO and SAC are policy-based which is will optimize the policy consistently to achieve the high profit. For the action space capability, DQN created for discrete spaces only; SAC for continuous spaces only; PPO are able to handle both discrete and continuous spaces. In terms of exploration and exploitation, SAC mainly on exploration but DQN and PPO are balance in the exploration and exploitation. In summary, DQN benefits in trading which is capability of learning from the large pool of market data, adaption of the dynamics market, keep improve the strategy and eliminate the sentimental in decision making that leads to highest profit and consistent. SAC benefits in trading are achieve the better exploration, increase the efficiency, improve the decision making continuously and adaption of dynamics market environment. However, PPO benefits in trading are ensure the stability of learning by controlling the policy, balance the ratio of the exploration and exploitation and adaption toward discrete and continuous actions (Haarnoja et al., 2018)

Aim of the Project:

- Compare different model performance (DQN, PPO and SAC).
- Minimize the emotional biases
- Improved decision making; automation

Objectives of the Project:

- To obtain the policy that give optimized return.
- To train agent that will not be influenced by the sentimental with using SRDRL model.
- Develop a dashboard that visualize the return of the agent that trained on different mechanism.
- Compare the performance within model and discuss the strength and weakness between different model (DQN, PPO and SAC).

Scopes of the Project:

Computing tools (Python) will be used for the data collection and process. The range of time series data of the standard & poor 500 (S & P 500) from 01/01/2016 to 01/01/2024 will be collected from Yahoo Finance. After completion of data collection, the process of cleaning data will be conducted which is cleared the missing data and prepare work for the descriptive analysis. After that, the DQN, PPO and SAC will be used to do the decision making based on the policy and reward mechanism that can achieve the optimized return profit. The comparison of performance between model will using F1-score and optimized cumulative return rates. The strengths and weaknesses of the model also will be discussed based on the performance.

Expected Contribution of the Project:

In the project, the expected contribution is to train trader with the high efficiency and stable in sentimental. What kinds of the difference will be brought by using the different model in trading agents which is PPO, SAC and DQN. Indicate the strengths and weaknesses of the PPO, SAC and DQN in trading strategy. Helping stakeholder to harvest more profit without worry and sentimental impact. The trader will be automation decision making based on the policy and reward mechanism. The suitable policy and reward mechanism is needed to train the agent that can fully eliminated the sentimental and the decision-making will be more discipline and the profit will be optimized. Every model has their own strengths and weaknesses; by identify the advantages of model will let the stakeholder apply those models in more appropriately toward the real-time market environment. The model with the higher value in F1-scores and accuracy, and optimized cumulative return rates based on the back testing will be selected for the future trading strategy in decision making.

Project Requirements:

Software: Python and Power BI

Hardware: Desktop with XEON intel

Technology/Technique/ Methodology/Algorithm: deep Q-network (DQN), proximal policy optimization (PPO) and soft actor-critic (SAC)

Type of Project (Focusing on Data Science):

[ ] Data Preparation and Modeling

[ ] Data Analysis and Visualization

[ ] Business Intelligence and Analytics

[/] Machine Learning and Prediction

[ ] Data Science Application in Business Domain

Status of Project:

[/] New

[ ] Continued

If continued, what is the previous title?

SECTION C: Declaration

I declare that this project is proposed by:

[ ] Myself

[ ] Supervisor/Industry Advisor ( )

Student Name: LEE HONG JIAN

06/April/2025

Signature **Date**

SECTION D: Supervisor Acknowledgement

The Supervisor(s) shall complete this section.

I/We agree to become the supervisor(s) for this student under aforesaid proposed title.

Name of Supervisor 1:

Signature **Date**

Name of Supervisor 2 (if any):

Signature **Date**

SECTION E: Evaluation Panel Approval

The Evaluator(s) shall complete this section.

Result:

[ ] FULL APPROVAL [ ] CONDITIONAL APPROVAL (Major)*

[ ] CONDITIONAL APPROVAL (Minor) [ ] FAIL*

* Student has to submit new proposal form considering the evaluators' comments.

Comments:

Name of Evaluator 1:

Signature Date

Name of Evaluator 2:

Signature Date

REFERENCES

Prayudi, Rafandito & Purwanto, Eko. (2023). The Impact of Financial Literacy, Overconfidence Bias, Herding Bias and Loss Aversion Bias on Investment Decision. Indonesian Journal of Business Analytics. 3. 1873-1886. 10.55927/ijba.v3i5.5715.

Aziz, et al. (2024), Role of behavioral biases in the investment decisions of Pakistan Stock Exchange investors: Moderating role of investment experience. Investment Management & Financial Innovations; Sumy Vol. 21, Iss. 1, (2024): 146-156.

Sangve, et al. (2025), ProfitPulse: Reinforcement Learning-Driven Trading Strategy. Artificial Intelligence & Data Science, Vishwakarma Institute of Technology, Pune, IND.

Georg, et al. (2024), Current applications and potential future directions of reinforcement learning-based Digital Twins in agriculture, SBA Research gGmbH, Floragasse 7/5.OG, Vienna, 1040, Vienna, Austria.

Kabbani, T., & Duman, E. (2022). *Deep reinforcement learning approach for trading automation in the stock market*. Ithaca: doi:https://doi.org/10.1109/ACCESS.2022.3203697

Huang, Y., Zhou, C., Zhang, L., & Lu, X. (2024). A self-rewarding mechanism in deep reinforcement learning for trading strategy optimization. *Mathematics, 12*(24), 4020. doi:https://doi.org/10.3390/math12244020

Vec, V., Tomažič, S., Kos, A., & Umek, A. (2024). Trends in real-time artificial intelligence methods in sports: A systematic review. *Journal of Big Data, 11*(1), 148. doi:https://doi.org/10.1186/s40537-024-01026-0