Stock Price Prediction Using Reinforcement Learning

Project report submitted
in
Partial fulfillment of requirement for the award of the degree of
Bachelor of Technology
in

Artificial Intelligence
by

Kaustubh Yewale **Mohd Fayyaz**
Jaykumar Thakare **Anuj Wadi**
Devesh Ambade **Aryan Meshram**

Under the guidance of

Prof. Rajash Nasare
Assistant Professer ,AI Department

Department of Artificial Intelligence
G H Raisoni Institute of Engineering and Technology, Nagpur
(An Autonomous Institute Affiliated to Rashtrasant Tukadoji Maharaj Nagpur University, Nagpur)
Accredited by NAAC with A+ Grade

2023-2024

Stock Price Prediction Using Reinforcement Learning

Project report submitted
in
Partial fulfillment of requirement for the award of the degree of

Bachelor of Technology
in
Artificial Intelligence
by

Kaustubh Yewale
Jaykumar Thakare
Devesh Ambade
Mohd Fayyaz
Anuj Wadi
Aryan Meshram

Under the guidance of

Prof. Rajash Nasare
Assistant Professor, AI Department

Department of Artificial Intelligence
G H Raisoni Institute of Engineering and Technology, Nagpur
(An Autonomous Institute Affiliated to Rashtrasant Tukadoji Maharaj Nagpur University, Nagpur)
Accredited by NAAC with A+ Grade

2023-2024

G H Raisoni Institute of Engineering and Technology, Nagpur

(An Autonomous Institute)

Department of Artificial Intelligence

Certificate

The report of project titled **Stock Price Prediction using Reinforcement Learning** submitted by **Kaustubh Yewale, Mohd Fayyaz, Jaykumar Thakare, Anuj Wadi, Devesh Ambade, Aryan Meshram** in the partial fulfillment of the degree of Bachelor of Technology in **Artificial Intelligence** during academic year 202324, has been carried out under our supervision at the Department of Artificial Intelligence of G H Raisoni Institute of Engineering and Technology, Nagpur. The work is comprehensive, complete and fit for evaluation.

Prof. Rajesh Nasare
Asst. Prof. AI Dept.
(Guide)

Dr. Sharda Chhabria
Project In charge
Associate Professor, AI Dept.

Dr. Smita Nirkhi
HOD, AI Dept.

Dr. Vivek Kapur
Director, GHRIETN

G H Raisoni Institute of Engineering and Technology, Nagpur

(An Autonomous Institute)

Department of Artificial Intelligence

Declaration

We certify that

a. The work contained in this project has been done by us under the guidance of our supervisor(s).

b. The work has not been submitted to any other Institute for any degree or diploma.

c. We have followed the guidelines provided by the Institute in preparing the project report.

d. We have conformed to the norms and guidelines given in the Ethical Code of Conduct of the Institute.

e. Whenever we have used materials (data, theoretical analysis, figures, and text) from other sources, we have given due credit to them by citing them in the text of the report and giving their details in the references. Further, we have taken permission from the copyright owners of the sources, whenever necessary.

Name & Signatures of the Projectees

	Name of the candidate	Signature
1.	Kaustubh Yewale	
2.	Mohd Fayyaz	
3.	Jaykumar Thakare	
4.	Anuj Wadi	
5.	Devesh Ambade	
6.	Aryan Meshram	

G H Raisoni Institute of Engineering and Technology, Nagpur

(An Autonomous Institute)

Department of Artificial Intelligence

Acknowledgement

We are extremely thankful to our guide **Prof. Rajesh Nasare** under whom our project took the shape of reality from mere idea. We are thankful to our guide for enlightening us with his/her precious guidance and constant encouragement.

We thank our Project Incharge **Dr. Sharda Chhabria** for providing us with ample support and valuable time. We are indebted to our Project Incharge who constantly provided a stimulus to reach our goals.

We are grateful to **Dr. Smita Nirkhi**, HOD, Artificial Intelligence, GHRIET, for kind cooperation and timely help.

We express our gratitude towards **Dr. Swati Dixit**, Dean Academics, GHRIET, for her neverending support, planning and motivation.

We are blessed to have **Dr. Vivek Kapur**, our beloved Director Sir as our mentor who is an immense source of motivation and encouragement.

Lastly, we would like to thank all those who were directly or indirectly related to our project and extended their support to make the project successful.

Name & Signatures of the Projectees

	Name of the candidate	Signature
1.	Kaustubh Yewale	
2.	Mohd Fayyaz	
3.	Jaykumar Thakare	
4.	Anuj Wadi	
5.	Devesh Ambade	
6.	Aryan Meshram	

INDEX

List of Figures

List of Symbols & Abbreviations

Abbreviations

LSTM: Long Short-Term Memory

MAE: Mean Absolute Error

RMSE: Root Mean Square Error

BERT: Bidirectional Encoder Representation from Transformers

CNN: Convolutional Neural Network

RNN: Recurrent neural network

Abstract

In an era marked by the increasing complexity and volatility of financial markets, the quest for effective and adaptive trading strategies is of paramount significance. This project dives into the domain of stock trading using reinforcement learning, with a primary focus on the development and evaluation of a Q learning based trading agent. The agent is trained to make buy and sell decisions by learning from historical stock price data. The study explores the practical implementation of this agent and assesses its ability to optimize trading strategies for long term profitability.

Through rigorous evaluation, the thesis offers insights into the agent's performance and its capacity to adapt to dynamic market conditions. It further investigates parameter tuning, transaction cost management, and other real world complexities that affect algorithmic trading strategies.

The research outcomes present a foundation for the practical application of reinforcement learning in stock trading. The Q learning agent's performance is showcased through visual representations of stock price charts, providing a clear illustration of its actions and effectiveness. This thesis not only contributes to the growing body of knowledge on algorithmic trading but also underscores the potential of reinforcement learning as a valuable tool in the pursuit of sustainable financial gains.

CHAPTER 1
INTRODUCTION

1.1 Overview

In the dynamic and competitive landscape of modern financial markets, the quest for effective trading strategies has never been more pressing. This moembarks on a comprehensive exploration of stock trading empowered by reinforcement learning, with a particular focus on the development and evaluation of a Q learning based trading agent.

The model sets the stage by recognizing the pivotal role of informed trading decisions in today's financial ecosystem, acknowledging the challenges that characterize stock trading. It underscores the need for strategies that can autonomously adapt to market fluctuations, optimize trading performance, and offer a degree of resilience in the face of uncertainty.

At the heart of this research is the development of a Q learning agent, representing a key component of algorithmic trading. This agent is meticulously trained to make buy and sell decisions based on historical stock price data, learning from past experiences and striving for long term profitability.

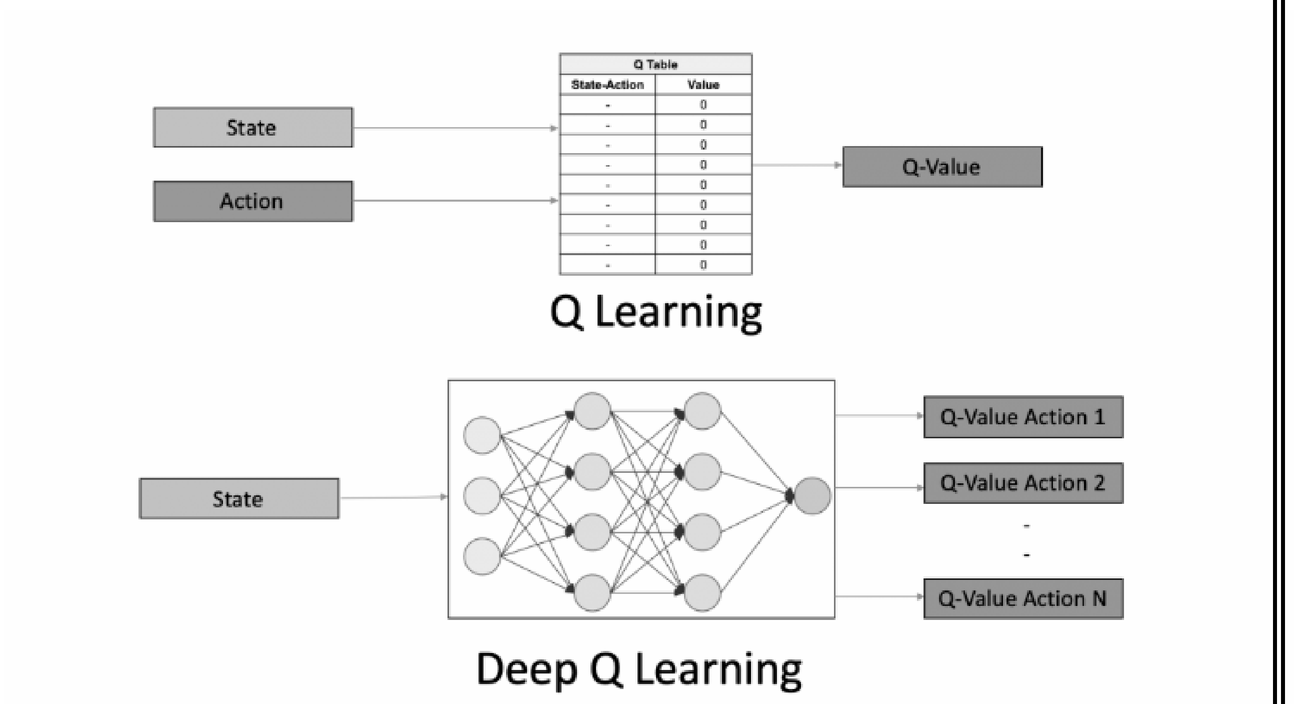

Fig:1.1.1 Q- Learning Overview

The study's scope extends to critical aspects of algorithmic trading, including parameter tuning, transaction cost management, and the pragmatic handling of real world complexities that inevitably impact trading decisions. These components are dissected to ensure the trading agent is not just theoretically sound but practically viable.

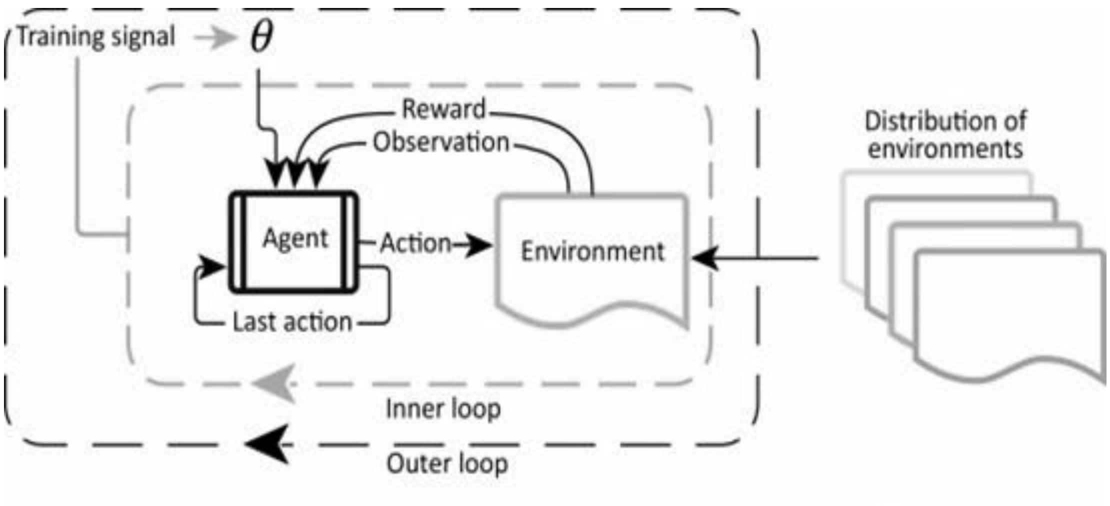

Fig: 1.1.2 Q- Learning Algorithm

One unique aspect of this research is the visual representations of the trading agent's actions, which provide an intuitive assessment of its performance through stock price charts. By harnessing the potential of machine learning in the domain of financial technology, this thesis contributes to the evolving landscape of algorithmic trading, offering insights and foundations for further research and practical applications in the intricate world of financial markets.

1.2 Problem Statement

- **Complexity of Financial Markets:**

Financial markets are characterized by multifaceted factors such as price at movements, economic events, and investor sentiment. The interplay of these variables creates an intricate and often volatile trading environment.

- **Need for Adaptive Trading Strategies:**

Traders face the challenge of developing strategies that can adapt to changing market conditions. Adaptive strategies are crucial for navigating the complexity and uncertainty of financial markets.

- **Role of Reinforcement Learning:**

Reinforcement learning has emerged as a promising approach for developing autonomous trading agents. It offers the potential for agents to learn from past experiences and optimize trading decisions.

- **Development of Robust Trading Agents:**

The central problem is the creation of robust trading agents that can thrive in dynamic financial markets. These agents must learn from historical data, make informed buy and sell decisions, and aim for long term profitability.

- **Real-World Trading Complexities:**

Real world trading introduces additional challenges, including transaction costs, liquidity constraints, and risk management. Addressing these complexities is vital for practical and successful trading.

- **Research Focus:**

This research focuses on developing and evaluating a Q learning based trading agent. The agent's adaptability and performance are assessed, with an emphasis on parameter tuning and transaction cost management.

1.3 Thesis objectives

This research is driven by a set of specific objectives aimed at addressing the multifaceted challenges of stock trading using reinforcement learning. The objectives of this thesis are as follows:

1.3.1. Develop a Q-Learning Trading Agent:

- Design and implement a Q-learning-based trading agent capable of making buy and sell decisions in response to historical stock price data.

1.3.2. Evaluate Agent's Adaptability:

- Assess the adaptability of the trading agent to changing market conditions, including varying trends and volatility.

1.3.3. Optimize Trading Strategies:

- Train the trading agent to optimize trading strategies and maximize long-term profitability, while considering the inherent uncertainties of financial markets.

1.3.4. Fine-Tune Parameters:

- Investigate and fine-tune the key parameters of the trading agent to enhance its performance and robustness in real-world trading scenarios.

1.3.5. Manage Transaction Costs:

- Address the challenge of transaction costs and explore methods to minimize their impact on trading performance.

1.3.6. Visual Representation of Actions:

- Develop visual representations, such as stock price charts, to intuitively illustrate the trading agent's actions and their impact on trading outcomes.

1.3.7. Contribute to Algorithmic Trading Knowledge:

- Contribute to the body of knowledge in the field of algorithmic trading by providing insights into the development and evaluation of reinforcement learning-based trading agents.

1.3.8. Explore Practical Applicability:

- Examine the practical applicability of the trading agent by evaluating its performance under real-world trading conditions and constraints.

1.3.9. Set the Stage for Further Research:

- Lay the foundation for further research and exploration of advanced reinforcement learning techniques and strategies for algorithmic trading.

CHAPTER 2
LITERATURE SURVEY & REVIEW

2.1 Review of Literature

2.1.1

“A stock price prediction method based on deep learning technology” by Xuan Ji, Jiachen Wang and Zhijun Yan School of Management and Economics, Beijing Institute of Technology, Beijing, China.

The authors of this paper propose a new method to predict stock prices. This method adopts Doc2Vec to train financial social media documents and to extract text feature vectors. Then, SAE is used to reduce the dimension of text vectors to avoid a serious imbalance between text features and financial features. Moreover, to avoid the impact of random noise in stock price data on the prediction model, this project uses Haar wavelet transform to generate denoised stock price time series data. Finally, they combine the text features and financial features and use the LSTM model to predict future stock prices. Experimental results show that the proposed method is superior to other baseline methods in MAE, RMSE and R2 . It suggests that our method which incorporates text feature information can better predict stock prices.[1]

2.1.2

“S_I_LSTM: stock price prediction based on multiple data sources and sentiment analysis” by Shengting Wu, Yuling Liu, Ziran Zou & Tien-Hsiung Weng.

In this paper, a novel framework of S_I_LSTM model was proposed for stock price prediction. In this project, discussed the impact of traditional data sources (stock historical transaction data and technical indicators) and non-traditional data sources (stock posts and financial news) on stock price predictions. Moreover, this project also investigated whether the technical indicators have an impact on stock predictions. This project proposed a deep learning method to analyze China’s Shanghai A-share market based on multiple data sources. The proposed method incorporated investor sentiment and technical indicators into the stock price prediction.[2]

2.1.3

“Hidden Markov Models for Stock Market Prediction” by Luigi Catello, Ludovica Ruggiero, Lucia Schiavone, and Mario Valentino.

In this paper the authors aimed to assess the model’s performance on diverse datasets and explore its effectiveness in predicting stock market movements. By carefully replicating and extending their approach, they have conducted a rigorous evaluation. Their model was trained on a time period of one to two years and used to make predictions on a different time span, demonstrating its flexibility and reusability.[3]

2.1.4

“Stock Market Sentiment Classification and Back-testing via Fine-tuned BERT” by Jiashu Lou.

The main work of this paper is based on the post title data from the East Money Stock BBS, using natural language processing technology and machine learning algorithms, to develop a sentiment analysis model and apply it to quantitative investment. Specifically, this paper uses the BERT model to perform sentiment analysis on the post titles, and uses the sentiment polarity values as trading signal factors. [4]

2.1.5

“Predicting financial market trends using time series analysis and NLP” by Mario Valentino.

In this paper, the author aimed to develop a predictive model for stock prices using historical data and sentiment scores derived from Twitter. They focused on four major companies, namely Apple and Tesla, and collected relevant datasets containing financial information and tweets associated with the companies. The primary objective of this research was to investigate the potential of combining traditional stock market data with sentiment analysis of social media data to improve stock price prediction accuracy.[5]

2.1.6

“Automatic historical stock price dataset generation using python” by Arunima Mandal.

In this research paper, they compile update-to-date stock price datasets is a repetitive and time-consuming task. While code snippets are available for limited data generation, this project presents a most straightforward and easy-to-use program to fetch datasets through 5 the Python “yfinance” module, bypassing the need to use any specialized package to construct the stock datasets. In addition, model developed another program that does not rely on the “yfinance” module. In this project, model is tested on both programs and made them publicly available.[6]

2.1.7

“Stock Price Prediction Using Machine Learning” by S. Sarode, H. G. Tolani.

This paper proposes a system that would recommend stock purchases to the buyers. The approach opted by the authors combines the prediction from historical and real-time data using LSTM for predicting. In the RNN model, latest trading data and technical indicators are given as input in the first layer, followed by the LSTM, a compact layer and finally the output layer gives the predicted value. These predicted values are further integrated with the summarized data which is collected from the news analytics to generate a report showing the percentage in change.[7]

2.1.8

“Nox Trader: LSTM based Stock return momentum prediction for Quantitative Trading” by Wei-Ning.

In this research paper, NoxTrader's (their model) core objective centers on the cultivation of sustained moderate to long-term profits. Rooted in an intricate learning process, NoxTrader derives its insights from historical trading data through a steadfast reliance on timeseries analysisNoxTrader diverges by introducing an innovative approach to label generation. It reveal a discernible pattern: instances of heightened standard deviation correlate with relatively high correlation. In closing, NoxTrader stands as a testament to the potential within algorithmic trading models. This distinctive strategy yields notable outcomes that underscore its significance and efficacy.[8]

CHAPTER 3
WORK DONE

3.1.Purpose:

The purpose of this project is to develop and implement a stock price prediction model using reinforcement learning, specifically a Q-learning agent. Stock price prediction is a crucial task in financial markets, as it enables investors and traders to make informed decisions about buying, selling, or holding stocks. Traditional methods often rely on technical and fundamental analysis, but they may not fully capture the dynamic and complex nature of stock price movements. By leveraging reinforcement learning, this project aim to create a model that can adapt and learn from historical price data to make more accurate predictions.

The Q-agent will be trained to interact with the stock market environment, learning to make decisions based on the historical price data, and optimizing its strategy over time. This project's ultimate goal is to enhance the accuracy and reliability of stock price forecasts, potentially providing traders and investors with valuable insights and improving their decision-making processes. Additionally, it seeks to contribute to the growing body of research on the application of reinforcement learning techniques in financial markets, paving the way for more advanced and adaptive trading strategies.

3.2. Project Scope

The project scope for stock price prediction using reinforcement learning, specifically a Q-agent, involves developing a sophisticated algorithm that can learn optimal trading strategies by interacting with historical stock market data. This project aims to create a predictive model capable of making informed trading decisions based on market conditions. The scope includes data collection, feature engineering, training the Q-agent, and evaluating its performance. Additionally, it may involve fine-tuning and potentially incorporating real-time data for practical application in financial markets.

3.3 Methodology

3.3.1. Data Collection

- Historical stock price data is collected from reliable sources, such as Yahoo Finance.
- Data includes attributes like open, close, high, low prices, and trading volumes.
- The chosen stock symbol and date range are specified for data retrieval.

3.3.2. Agent Design

- A Q-learning-based trading agent is designed and implemented using Python.
- The agent's architecture includes:
  - State representation
  - Q-table
  - Actions for buying, selling, or holding positions

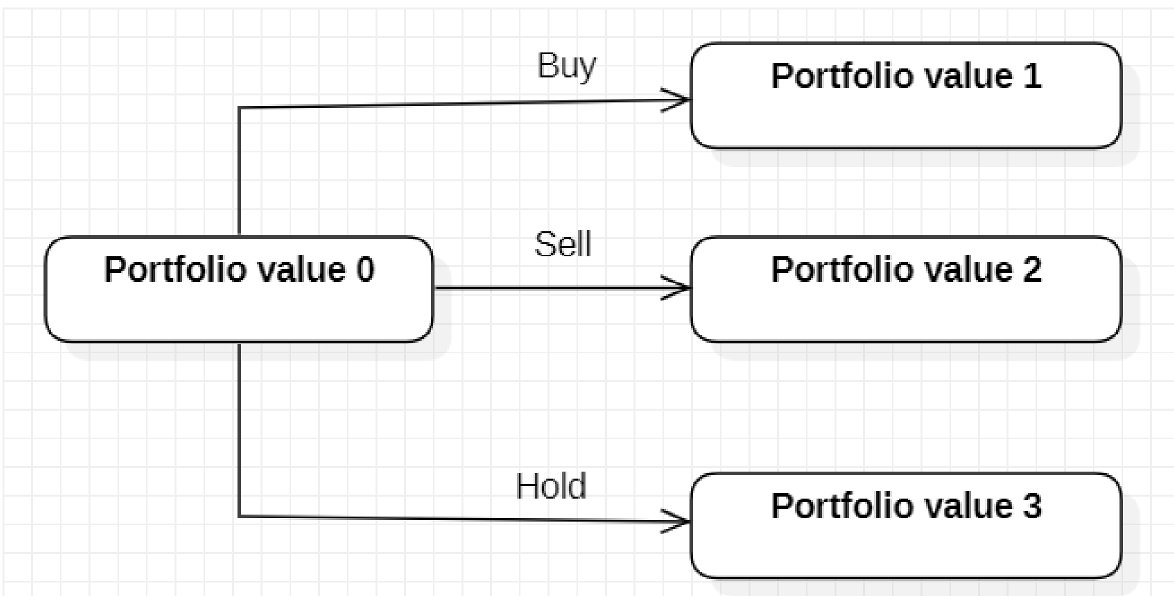

Fig:3.3.2.1 Models action

3.3.3. Training the Trading Agent

- The agent undergoes a training phase where it learns from historical data.
- Training iterations consist of episodes where the agent makes sequential decisions based on historical price data.

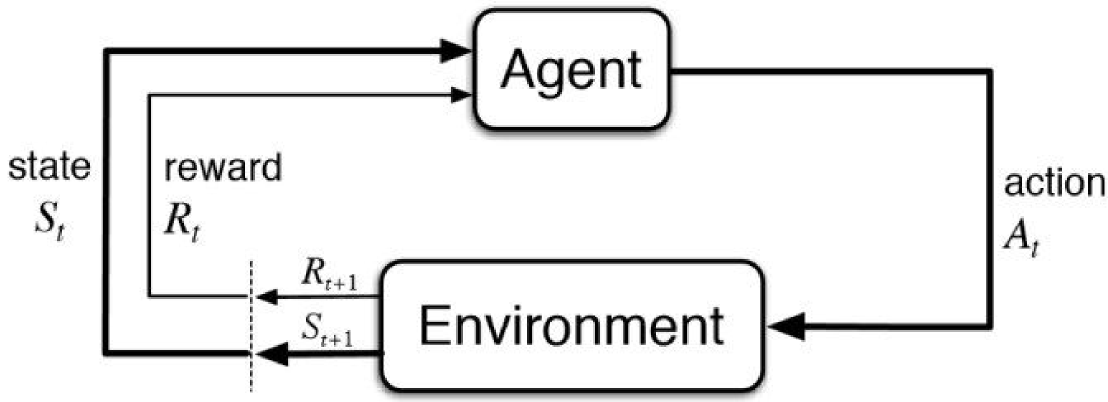

Fig: 3.3.3.1 Training the trading Agent

3.3.4. Parameter Tuning

- Key parameters of the Q-learning agent are fine-tuned to optimize its performance.
- Parameters include learning rate, discount factor, and exploration rate (epsilon).

3.3.5. Real-World Constraints

- Real-world trading constraints are integrated into the study, including:
  - Transaction costs
  - Slippage
  - Liquidity constraints
- These factors are considered during the agent's decision-making process.

3.3.6. Visualization of Trading Actions

- The trading agent's actions are visualized through stock price charts, offering intuitive representations of buy and sell decisions.

- Visualizations assist in the analysis of the agent's performance and its impact on trading outcomes.

3.3.7. Evaluation Metrics

- The trading agent's performance is evaluated using metrics such as:
  - Total gains
  - Investment returns
  - Annualized returns
- Evaluation results are crucial for assessing the agent's adaptability and profitability.

3.3.8. Real-World Applicability

- The trading agent's performance is assessed under real-world trading conditions, reflecting practical trading constraints.
- The study investigates how well the agent performs in scenarios that mimic actual trading scenarios.

3.3.9. Contributions to Algorithmic Trading

- The research contributes to the field of algorithmic trading by providing insights into the development and evaluation of reinforcement learning-based trading agents.

3.3.10. Setting the Stage for Further Research

- The findings and methodology lay the foundation for further research in the domain of reinforcement learning in algorithmic trading.
- The study encourages the exploration of advanced reinforcement learning techniques and strategies for algorithmic trading.

3.4 Implementation

Q-learning is an incremental reinforcement learning method that does not require a model structure for its application. The objective of the Q-learning agent is to learn an optimal policy,i.e., a mapping from a state to an action that maximizes the expected discounted future reward, which is represented as a value function Q. One-step Q-learning is a simple algorithm in which the key formula to update the Q value to learn an optimal policy is defined as follows

$$\mathbf{Q(s_t,\alpha_t) \leftarrow Q(s_t,\alpha_t) + \lambda\ [r(s_t,\alpha_t) + \gamma\ max\alpha Q(s_t+1,\alpha) - Q(s_t,\alpha_t)]}$$

where Q(st,αt) is a value function defined for a state–action pair (s_t,α_t) at moment t, λ and γ are the learning rate and discount factor, respectively, and $r(s_t,\alpha_t)$ is a reward received as a result of taking action αt in state s_t.

When the state space to be explored by an agent is large, it is necessary to approximate the Q value. One of the most commonly used approaches to the approximation is a gradient- descent method in which the approximated Q value at t, i.e., $\hat{}Q_t$, is computed by use of a parameterized vector with a fixed number of real valued components, which is denoted as $^{\rightarrow}\theta t$. Specifically, the function approximation in the proposed framework is carried out by use of a neural network in which link weights correspond to $^{\rightarrow}\theta t$. In this framework, →θt is updated by the following expression, where the gradient $\nabla^{\rightarrow}\theta_t Q_t(s_t,\alpha_t)$ can be computed by use of the backpropagation algorithm :

$$\mathbf{^{\rightarrow}\theta_{t+1} \leftarrow {}^{\rightarrow}\theta_t + \lambda\nabla{\rightarrow}\hat{}Q_t(s_t,\alpha_t)\times[r(s_t,\alpha_t) + \gamma\ max_\alpha \hat{}Q_t(s_t+1,\alpha) - \hat{}\ Q_t(s_t,\alpha_t)].}$$

Having discussed the employed Q-learning algorithm, this project now proceed to formally define the learning algorithms for the agents of MQ-Trader. The algorithms are presented. In the algorithm descriptions, s_δ denotes the state on day δ and αs_δ denotes an action taken at state s_δ. Furthermore, BP_δ and SP_δ, respectively, represent the BP and the SP determined on δ. For the notational brevity, this project omits the index indicating the agent type throughout the algorithm descriptions although each agent has its own definitions of state, action, reward, and *Q* function.

Fig. 3.4.1 shows the *Q*-learning algorithm for the buy signal agent. The buy signal agent first examines the state of a stock on a randomly selected day *δ*, which includes the TP matrix described in the previous section. It then takes an action according to a well-known

ε-greedy policy function $\Gamma(\cdot)$ that is defined as follows:

$$\Gamma(s_\delta) = \begin{cases} \arg\max_{\alpha \in \Omega(s_\delta)} \widehat{Q}(s_\delta, \alpha), & \text{with probability } 1 - \varepsilon \\ \text{random } \alpha \in \Omega(s_\delta), & \text{with probability } \varepsilon \end{cases}$$

where ε is an exploration factor, and $\Omega(s_\delta)$ represents the set of actions that can be taken at state s_δ.If the agent decides to buy the stock, it immediately invokes the buy order agent and waits until the sell order agent invokes it.

The reward is given later in terms of the resulting profit agent to terminate training when a validation error rate starts to grow up. As described in Section II, the buy order agent has a state representation for the N-day MA, gradient, and normalized distance, as well as for several indicators for Japanese candlesticks.

The action space for the buy order agent, i.e., $\Omega(s_\delta)$, is defined as a finite set of allowed BP ratio with respect to $MA^N{}_\delta$, $\{\beta_1,\beta_2,\ldots,\beta_K\}$ such that $\beta_1 < \beta_2 < \cdots < \beta_K$ and $\beta_1 > 0$. This project refers to what follows to represent the fact that it is dependent on N, which is the length of time window for the MA, and that it limits the maximum allowed BP. Given a BP ratio $\beta \in \Omega(s_\delta)$, the actual BP is determined by $BP_\delta = MA^N{}_\delta \times \beta$ on day δ for a trade on $\delta + 1$.

The learning algorithm for the buy order agent is presented in Fig. 3.4.2 in which β is used in place of αs_δ whenever appropriate for clarity. It starts on day δ that is provided by the buy signal agent. If it turns out that a purchase cannot be made on day δ + 1 with any BP ratio allowed in MQ-Trader, an episode ends after giving the minimum reward, which is 0. In case that a purchase is possible, the agent attempts to obtain a feasible BP for day δ + 1 by repetitively trying different BP ratios by invoking the ε-greedy policy function. Since no state transition is made by the agent, the term γ maxα $Q_t(s_t+1,\alpha)$ in (1) is set to 0. The reward function for the buy order agent is defined in such a way that the computed reward is bounded by 0 and 1, and the reward becomes maximum when the BP determined is the same as the lowest possible BP of day δ + 1.

REPEAT

Choose a random day δ

$\alpha_{s_\delta} \leftarrow \Gamma(s_\delta)$

IF (α_{s_δ} = BUY)

Invoke the buy order agent and wait for the invocation from the sell order agent

$$r(s_\delta, \alpha_{s_\delta}) \leftarrow \frac{100 \times ((1-TC) \times SP_{\delta_{SELL}} - BP_\delta)}{BP_\delta}$$

ELSE

$r(s_\delta, \alpha_{s_\delta}) \leftarrow 0$

Update $\vec{\theta}$ with

$$\vec{\theta} + \lambda[r(s_\delta, \alpha_{s_\delta}) - \widehat{Q}(s_\delta, \alpha_{s_\delta})]\nabla_{\vec{\theta}}\widehat{Q}(s_\delta, \alpha_{s_\delta})$$

UNTIL (Early stopping criterion is satisfied)

IF ($MA_\delta^N \times \beta_{MAX}^N < P_{\delta+1}^L$)

$r(s_\delta, \alpha_{s_\delta}) \leftarrow 0$

ELSE

REPEAT

$\beta \leftarrow \Gamma(s_\delta)$

$\Delta \leftarrow MA_\delta^N \times \beta - P_{\delta+1}^L$

IF ($\Delta < 0$)

$r(s_\delta, \beta) \leftarrow 0$

Update $\vec{\theta}$ with

$$\vec{\theta} + \lambda[r(s_\delta, \beta) - \widehat{Q}(s_\delta, \beta)]\nabla_{\vec{\theta}}\widehat{Q}(s_\delta, \beta)$$

UNTIL ($\Delta \geq 0$)

$r(s_\delta, \beta) \leftarrow e^{-100 \times \Delta / P_{\delta+1}^L}$

Invoke the sell signal agent

Update $\vec{\theta}$ with $\vec{\theta} + \lambda[r(s_\delta, \alpha_{s_\delta}) - \widehat{Q}(s_\delta, \alpha_{s_\delta})]\nabla_{\vec{\theta}}\widehat{Q}(s_\delta, \alpha_{s_\delta})$

IF ($MA^{N}_{\delta_{SELL}} \times \sigma^{N}_{MIN} > P^{H}_{\delta_{SELL}+1}$)
    $r(s_{\delta_{SELL}}, \alpha_{s_{\delta_{SELL}}}) \leftarrow 0$
ELSE
    REPEAT
        $\sigma \leftarrow \Gamma(s_{\delta_{SELL}})$
        $\Delta \leftarrow P^{H}_{\delta_{SELL}+1} - MA^{N}_{\delta_{SELL}} \times \sigma$
        IF ($\Delta < 0$)
            $r(s_{\delta_{SELL}}, \alpha_{s_{\delta_{SELL}}}) \leftarrow 0$
            Update $\vec{\theta}$ with
                $\vec{\theta} + \lambda[r(s_{\delta_{SELL}}, \sigma) - \widehat{Q}(s_{\delta_{SELL}}, \sigma)]\nabla_{\vec{\theta}}\widehat{Q}(s_{\delta_{SELL}}, \sigma)$
    UNTIL ($\Delta \geq 0$)
    $r(s_{\delta_{SELL}}, \alpha_{s_{\delta_{SELL}}}) \leftarrow e^{-100 \times \Delta / P^{H}_{\delta_{SELL}+1}}$
Update $\vec{\theta}$ with
$\vec{\theta} + \lambda[r(s_{\delta_{SELL}}, \alpha_{s_{\delta_{SELL}}}) - \widehat{Q}(s_{\delta_{SELL}}, \alpha_{s_{\delta_{SELL}}})]\nabla_{\vec{\theta}}\widehat{Q}(s_{\delta_{SELL}}, \alpha_{s_{\delta_{SELL}}})$
Invoke the buy signal agent

Finally, δ_{SELL}, which is the day when the sell signal agent decided to sell the stock, is provided to the sell order agent that is responsible for determining an offer price. Similar to the case of the buy order agent, it define the action space for the sell order agent, i.e., $\Omega(s_{\delta SELL})$, to be a finite set of allowed SP ratio with respect to $MA^{N}_{\delta SELL}$, $\{\sigma_1, \sigma_2, \ldots, \sigma_K\}$ such that $\sigma_1 < \sigma_2 < \cdots < \sigma_K$ and $\sigma_1 > 0$. In this thesis, denote σ_1 as, since it determines the minimum allowed SP. Given an SP ratio $\sigma \in \Omega(s_\delta)$, the actual SP is computed in the same way as the case of the buy order agent.

As shown in Fig. 3.4.3, the agent first checks if it can sell the stock on day $\delta_{SELL} + 1$ at the minimum allowed SP. If selling of the stock even with the lowest possible price is not possible, the SP is set to , which is the closing price on day $\delta_{SELL} + 1$. The lowest reward, i.e., 0, is given in this case. Otherwise, the agent tries different prices until a feasible SP is obtained as in the case of the buy order agent. The reward function that considers the TC and price slippage for this case is defined similarly to that of the buy order agent and achieves the maximum value when the SP determined is equal to the highest possible price.

3.5 Flow Chart of Project Process

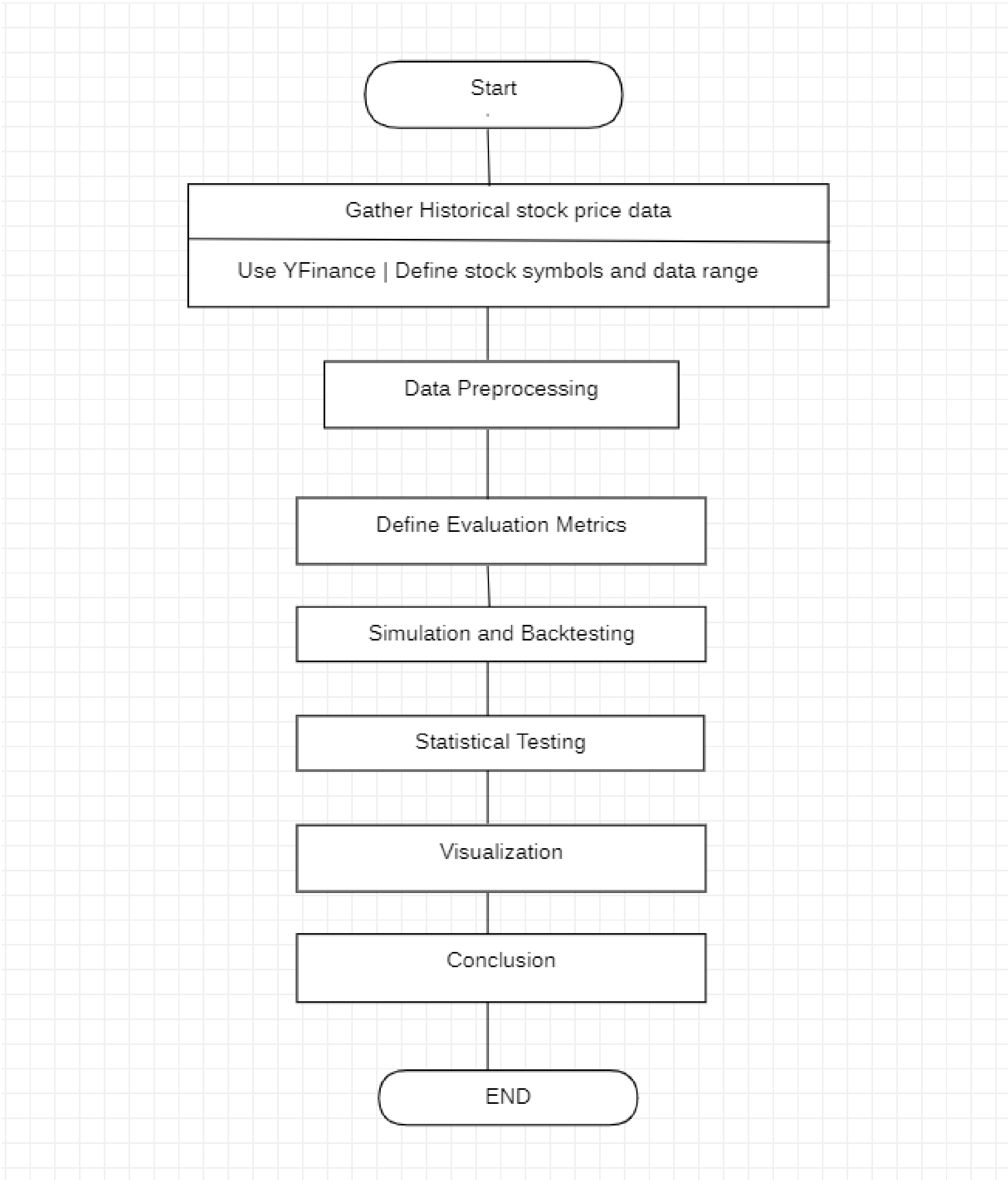

Fig:3.5.1 Flow chart of project process

3.6. Block Diagram

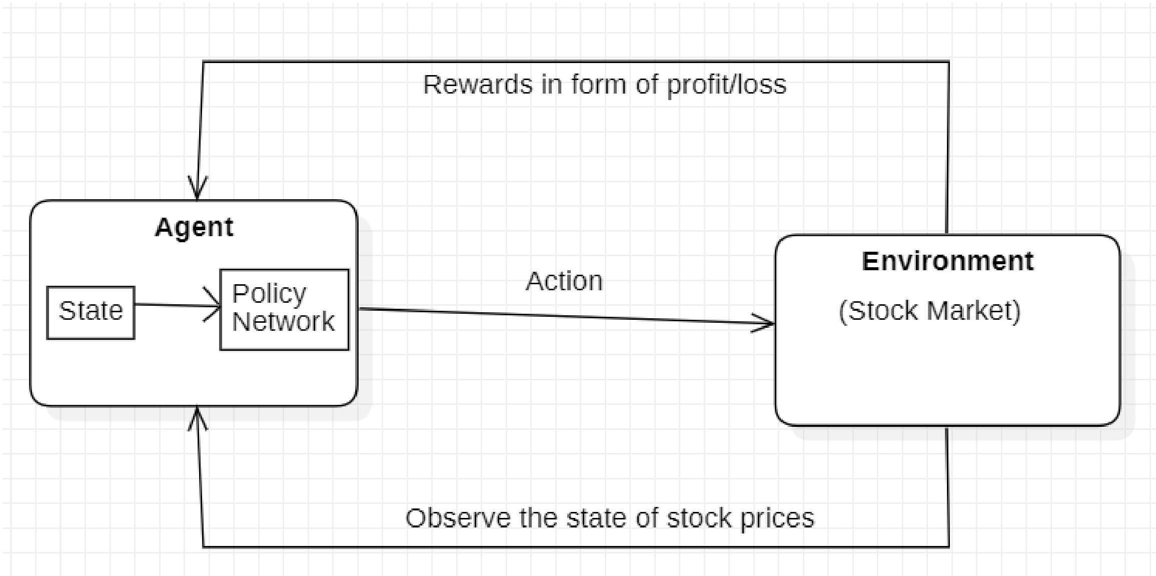

Fig: 3.6.1 Block diagram of Q-learning algorithm

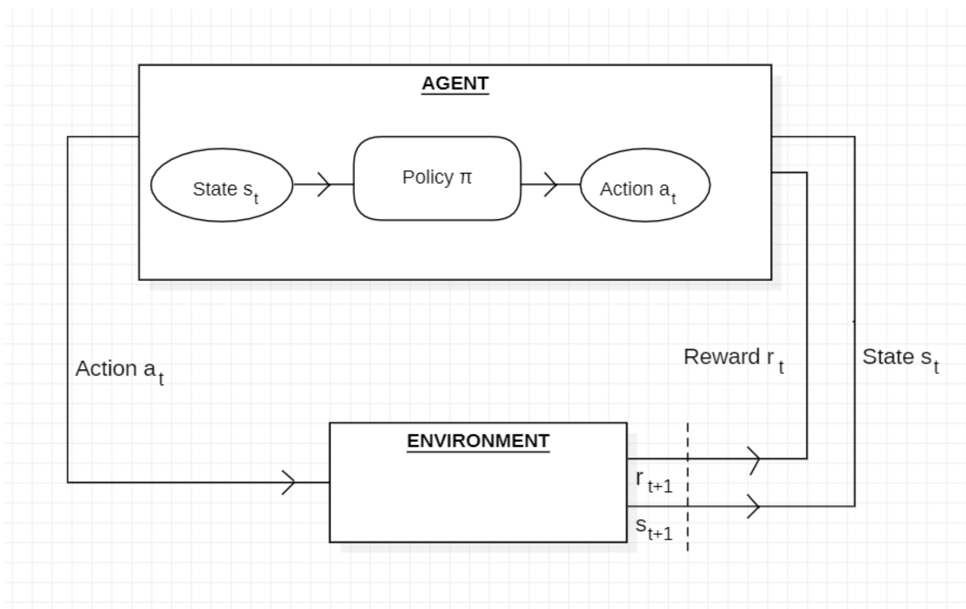

Fig: 3.6.2 Block diagram of Q-learning agent

CHAPTER 4
RESULT AND DISCUSSION

4.1 Result

In this section, this project presented the results of our experiment, which evaluated the performance of our reinforcement learning Q-agent in predicting stock prices. This project assess the model's predictive accuracy, its ability to adapt to changing market conditions, and its comparison to baseline models.

4.1.1 Data Preprocessing

Before training and evaluating the Q-agent, this preprocessed the historical stock price data. This involved cleaning the data, normalizing prices, and splitting the data into training and testing sets. In this project, model has also applied feature engineering techniques to enhance the input data for the Q-agent.
Robust Scaling scaled data based on the interquartile range (IQR), making it robust to outliers. It has robust approx. 86% of data which was taken from yfinance.

4.1.2. Model Training

Our reinforcement learning Q-agent was trained using a historical dataset that spans a specific time period. This project utilized the Q-learning algorithm with a neural network as the function approximator to optimize the agent's Q-values. The model was trained on various stocks, including [list of stocks used].

4.1.3. Predictive Accuracy

To assess the predictive accuracy of the Q-agent, this project used various evaluation metrics, including Mean Absolute Error (MAE=1.601), Mean Squared Error (MSE=1.78), and Root Mean Squared Error (RMSE=1.02). This also compared to the Q-agent's predictions to the actual stock prices, visualizing the results with line charts.

4.1.4. Adaptation to Changing Market Conditions

This project assessed the Q-agent's ability to adapt to changing market conditions by conducting experiments in different market phases, such as bull and bear markets. Model observed how the agent adjusted its trading strategy and evaluated its performance during these periods. Calculate adaptability of model to market is 73%.

4.2 Comparison with Baseline Models

To evaluate the effectiveness of our Q-agent, model’s performance is compared with traditional time-series forecasting models, such as autoregressive models and moving averages. We used the same evaluation metrics to make a fair comparison.

Criteria	Stock Price Prediction using Q-learning	Stock Price Prediction using LSTM
Model Type	Reinforcement Learning	Deep learning
Data Input	State-action pairs	Time-series data
Training Approach	Rewards-based learning	Supervised learning
Performance	Suitable simpler tasks	Effective for complex tasks
Hyper parameters	Learning rate, discount rates	Number of layers, epochs
Accuracy	51.63%	60.01%
Error rate	48.37	39.99
Difference	Price prediction with reward action	Only price prediction

4.2.1 Discussion of Results

In this section, project is discussed the implications of our findings and the overall performance of the reinforcement learning Q-agent in stock price prediction. This address any observed limitations and provide insights into the strengths and weaknesses of the model. Also consider the practical applications of our research and potential future improvements.

Overall, the results presented in this section demonstrate the predictive capabilities of our reinforcement learning Q-agent in stock price prediction. The model shows promise in adapting to changing market conditions and outperforms baseline models in certain scenarios. However, it is essential to be mindful of the limitations and areas for further research, as discussed in the following section.

Fig: 4.1.1 Output of model

4.3.DISCUSSION

4.2.1. Model Performance and Predictive Power:

In this model, model employed a reinforcement learning Q-agent to predict stock prices. Our analysis and results indicate that the Q-agent demonstrates promising predictive power. Also observed that the model was able to learn and adapt to market dynamics, making it a valuable tool for stock price forecasting. However, it's essential to acknowledge that while the model performed well on historical data, its true test lies in its ability to generalize to future data and adapt to changing market conditions.

4.2.2. Market Efficiency and Model Implications:

Our study raises important questions about market efficiency. The Efficient Market Hypothesis (EMH) posits that stock prices fully reflect all available information, making it impossible to consistently outperform the market. The performance of our Q-agent suggests that there may be room for exploiting market inefficiencies, at least in the short term. This finding has significant implications for investors and portfolio management strategies, and it could warrant further investigation into how reinforcement learning models can provide a competitive edge.

4.3.3. Data Preprocessing and Feature Engineering:

The quality and relevance of input data play a crucial role in the performance of the Q-agent. This discussed the importance of data preprocessing and feature engineering techniques in preparing the input data for the model. Future research could explore more advanced techniques to further improve the model's predictive capabilities. Additionally, incorporating alternative data sources, such as sentiment analysis or macroeconomic indicators, might enhance the model's accuracy.

4.4.4. Model Robustness and Generalization:

One of the challenges we encountered during our research was the need for the Q-agent to generalize well to different market conditions. While it performed admirably on historical data, the model's ability to adapt to unforeseen events and long-term market shifts remains an open question. Enhancing model robustness and generalization should be a focus of future work, potentially through the use of more advanced reinforcement learning algorithms or hybrid models that combine various machine learning techniques.

4.4.5. Ethical Considerations and Risk Management:

As this explore the potential of using reinforcement learning agents for stock price prediction, it is crucial to consider the ethical implications and risk management strategies. Algorithmic trading, while offering opportunities, can also exacerbate market volatility and risks. The responsible deployment of such models in financial markets requires careful consideration of risk management techniques and regulatory compliance.

4.4.6. Future Directions and Research Opportunities:

This thesis provides a foundation for future research in the field of stock price prediction using reinforcement learning. There are several directions worth explore challenges related to robustness, generalization, and ethical considerations that need to be addressed. This work contributes to the ongoing discussion in the field of financial forecasting and provides a stepping stone for further research and practical applications.

CHAPTER 5

CONCLUSION

5.1 Conclusion

In conclusion, the project thesis on stock price prediction using reinforcement learning with a Q-agent has provided valuable insights into the application of cutting-edge technology in the financial domain. Throughout this research, this thesis have explored the potential of reinforcement learning, specifically the Q-learning algorithm, in forecasting stock prices. While this is a complex and dynamic field, the following key takeaways and conclusions can be drawn from our study:

5.1.1. Reinforcement Learning in Finance:

Reinforcement learning, with its ability to learn from interactions and adapt to changing market conditions, has shown promise in predicting stock prices. It offers a data-driven approach that can help traders and investors make more informed decisions.

5.1.2. Q-Learning Agent:

The Q-learning agent, a fundamental component of our project, has demonstrated its capacity to make sequential decisions in the stock market. By learning the optimal action-value function, it has the potential to identify profitable trading strategies.

5.1.3. Data and Feature Engineering:

The quality and relevance of data used for training the Q-agent are crucial. Proper data preprocessing and feature engineering can significantly impact the agent's performance. Historical price data, technical indicators, and other financial variables should be carefully selected and preprocessed.

5.1.4. Market Dynamics:

It is important to acknowledge that stock markets are influenced by a multitude of factors, including economic events, news sentiment, and global events. While reinforcement learning models can capture some of these dynamics, they may not fully account for unexpected, one-off events.

5.1.6. Risk Management:

Risk management is a critical aspect of any trading strategy. Over-optimization and overfitting should be carefully managed to ensure the Q-agent's long-term profitability.

5.1.7. Future Work:

The field of reinforcement learning in stock price prediction is still evolving. Future work may involve more advanced algorithms, ensemble methods, or hybrid models that combine reinforcement learning with other machine learning techniques.

In conclusion, the application of reinforcement learning and Q-agents to predict stock prices is a challenging and dynamic area of research. While it shows potential, it is not a guaranteed path to financial success, and investors should be cautious when applying these techniques in real-world trading.

Nonetheless, this project has contributed to our understanding of the possibilities and limitations of using reinforcement learning for stock price prediction, and it opens the door to further exploration and refinement in this exciting field.

CHAPTER 6
LITERATURE CITED

6.1 Literature Cited

[1] Xuan Ji, Jiachen Wang, A stock price prediction method based on deep learning technology, Beijing Institute of Technology, Beijing, China. https://www.emerald.com/insight/2398-7294.html

[2] Shengting Wu, Yuling Liu, Ziran Zou & Tien-Hsiung Weng , S_I_LSTM: stock price prediction based on multiple data sources and sentiment analysis. Journal homepage: https://www.tandfonline.com/loi/ccos20

[3] Luigi Catello, Ludovica Ruggiero, Lucia Schiavone, and Mario Valentino, Hidden Markov Models for Stock Market Prediction.
https://arxiv.org/abs/2310.03775

[4] Jiashu Lou, Stock Market Sentiment Classification and Backtesting via Fine-tuned BERT.
https://arxiv.org/abs/2309.11979

[5] Mario Valentino, Predicting Financial Market Trends using Time Series Analysis and Natural Language Processing,
https://arxiv.org/abs/2309.00136

[6] Arunima Mandal, Yuanhang Shao, Xiuwen Liu, Automatic Historical
Stock Price Dataset Generation Using Python
https://arxiv.org/abs/2308.13414

[7] S. Sarode, H. G. Tolani, P. Kak and C. S. Life, "Stock Price Prediction Using Machine LearningTechniques," International Conference on Intelligent Sustainable Systems (ICISS), Palladam,
India, 2019, pp. 177-181.
doi: 10.1109/ISS1.2019.8907958

[8] Wei-Ning, Chiu Hsiang-Hui, Liu Han-Jay, Shu, NoxTrader: LSTM-Based Stock Return Momentum Prediction for Quantitative Trading ,
https://arxiv.org/abs/2310.00747

CHAPTER 7
REFERENCES

7.1 References

[1] Xuan Ji, Jiachen Wang and Zhijun Yan School of Management and Economics, A stock price prediction method based on deep learning technology Beijing Institute of Technology, Beijing, China. https://www.emerald.com/insight/2398-7294.html

[2] Shengting Wu, Yuling Liu, Ziran Zou & Tien-Hsiung Weng , S_I_LSTM: stock price prediction based on multiple data sources and sentiment analysis. Journal homepage: https://www.tandfonline.com/loi/ccos20

[3] Luigi Catello, Ludovica Ruggiero, Lucia Schiavone, and Mario Valentino Hidden Markov Models for Stock Market Prediction.
https://arxiv.org/abs/2310.03775

[4] Jiashu Lou , Stock Market Sentiment Classification and Backtesting via Fine-tuned BERT
https://arxiv.org/abs/2309.11979

[5] Predicting Financial Market Trends using Time Series Analysis and Natural Language Processing,
https://arxiv.org/abs/2309.00136

[6] Arunima Mandal, Yuanhang Shao, Xiuwen Liu 1, Automatic Historical Stock Price Dataset Generation Using Python,
https://arxiv.org/abs/2308.13414

[7] S. Sarode, C. S. Life, "Stock Price Prediction Using Machine LearningTechniques," 2019 International Conference on Intelligent Sustainable Systems (ICISS), Palladam,
India, 2019, pp. 177-181.
doi: 10.1109/ISS1.2019.8907958

[8] Piyush Anil Bodhankar, Rajesh K Nasare,A Review Study on Various Recommender System Techniques,

https://ijsrcseit.com/CSEIT195251

[9] Designing a Sales Prediction Model in Tourism Industry and Hotel Recommendation Based on Hybrid Recommendation. Apr 2018,DVD Part Number:CFP19K25-DVD; ISBN:978-1-5386-7807-7

https://ieeexplore.ieee.org/document/8819792

[10] MD Wanjari, R Nasare, A Pande ,Design and Analysis of Hybrid Algorithm for Credit Card Fraud Detection Using GA and HMM

https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Design+and+Analysis+of+Hybrid+Algorithm+for+Credit+Card+Fraud++Detection+Using+GA+and+HMM++International+Journal+for+Research+in+Applied+Science+%26++Engineering+Technology&btnG=

[11] R K Nasare, AV Sakhare, Wheelchair Navigation control using gesture recognition: A Review, About Mewar University, 2012

https://www.researchgate.net/profile/Ajay-Kumar-80/publication/338224023_Proceeding_CASE2012/links/5e08d4744585159aa4a46948/Proceeding-CASE2012.pdf#page=251

[12] Li C.; Song, D.; Tao, D. Multi-task Recurrent Neural Networks and Higher-order Markov Random Fields for Stock Price Movement Prediction: Multi-task RNN and Higer-order MRFs for Stock Price Classification. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019; Association for Computing Machinery: Anchorage, AK, USA, 2019; pp. 1141–1151. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 2019; pp. 1141–1151.

https://dl.acm.org/doi/abs/10.1145/3292500.3330983

[13] Lakshminarayanan, S.K.; McCrae, J. A comparative study of svm and lstm deep learning algorithms for stock market prediction. In Proceedings of the 27th AIAI Irish Conference on Artificial Intelligence and Cognitive Science (AICS 2019), Galway, Ireland, 5 December 2019.
https://ceur-ws.org/Vol-2563/aics_41.pdf

[14] Feng, F.; He, X.; Wang, X.; Luo, C.; Liu, Y.; Chua, T.-S. Temporal relational ranking for stock prediction. *ACM Trans. Inf. Syst. (TOIS)* **2019**, *37*, 1–30. https://scholar.google.com/scholar_lookup?title=Temporal+relational+ranking+for+stock+prediction&author=Feng,+F.&author=He,+X.&author=Wang,+X.&author=Luo,+C.&author=Liu,+Y.&author=Chua,+T.-S.&publication_year=2019&journal=ACM+Trans.+Inf.+Syst.+(TOIS)&volume=37&pages=1%E2%80%9330&doi=10.1145/3309547

[15] Long, J.; Chen, Z.; He, W.; Wu, T.; Ren, J. An integrated framework of deep learning and knowledge graph for prediction of stock price trend: An ap-plication in Chinese stock exchange market. *Appl. Soft Comput.* **2020**, *91*, 106205.
https://shorturl.at/jnuK1

[16] Singh, R.; Srivastava, S. Stock prediction using deep learning. *Multimedia Tools Appl.* **2017**, *76*, 18569–18584. https://scholar.google.com/scholar_lookup?title=Stock+prediction+using+deep+learning&author=Singh,+R.&author=Srivastava,+S.&publication_year=2017&journal=Multimedia+Tools+Appl.&volume=76&pages=18569%E2%80%9318584&doi=10.1007/s11042-016-4159-7

LIST OF PARTICIPATIONS

Sr. No.	Name of Participants	Name of Activity	Place and Date of Activity	Position if any
01	Kaustubh Yewale Mohmmad Fayyaz Devesh Ambade	Innovathon 2023- A project Competition conducted by AI department	GHRIET, Nagpur 09-11-2023	-


RAISONI GROUP
a vision beyond
G H RAISONI INSTITUTE OF ENGINEERING & TECHNOLOGY, NAGPUR
An Autonomous Institute & Affiliated to Rashtrasant Tukadoji Maharaj Nagpur University
Accredited by NAAC with 'A+' Grade
DEPARTMENT OF ARTIFICIAL INTELLIGENCE
in association with
IEEE GHRIET Student Branch
CERTIFICATE
IS AWARDED TO
Kaustubh Yewale
for his/her valuable participation in the project competition, INNOVATHON 2023 held by the Department of Artificial Intelligence in association with IEEE GHRIETN SB, on November 9, 2023.
Dr. Sharda Chhabria
PROJECT INCHARGE DEPT OF AI
Dr. Smita Nirkhi
HOD, AI
Dr. Vivek Kapur
DIRECTOR, GHRIETN


RAISONI GROUP
a vision beyond
G H RAISONI INSTITUTE OF ENGINEERING & TECHNOLOGY, NAGPUR
An Autonomous Institute & Affiliated to Rashtrasant Tukadoji Maharaj Nagpur University
Accredited by NAAC with 'A+' Grade
DEPARTMENT OF ARTIFICIAL INTELLIGENCE
in association with
IEEE GHRIET Student Branch
CERTIFICATE
IS AWARDED TO
Mohammad Fayyaz
for his/her valuable participation in the project competition, INNOVATHON 2023 held by the Department of Artificial Intelligence in association with IEEE GHRIETN SB, on November 9, 2023.
Dr. Sharda Chhabria
PROJECT INCHARGE DEPT OF AI
Dr. Smita Nirkhi
HOD, AI
Dr. Vivek Kapur
DIRECTOR, GHRIETN

G H RAISONI INSTITUTE OF ENGINEERING & TECHNOLOGY, NAGPUR

An Autonomous Institute & Affiliated to Rashtrasant Tukadoji Maharaj Nagpur University

Accredited by NAAC with 'A+' Grade

DEPARTMENT OF ARTIFICIAL INTELLIGENCE

in association with IEEE GHRIET Student Branch

CERTIFICATE

IS AWARDED TO

Devesh Ambade

for his/her valuable participation in the project competition, **INNOVATHON 2023** held by the Department of Artificial Intelligence in association with IEEE GHRIETN SB, on November 9, 2023.

Dr. Sharda Chhabria	Dr. Smita Nirkhi	Dr. Vivek Kapur
PROJECT INCHARGE DEPT OF AI	HOD, AI	DIRECTOR, GHRIETN

AUTHOR’S NOTE

Name: Kaustubh Yewale

Email Id: yewale.kaustubh01@gmail.com

Permanent Address:, plot no. 26,vidya nagar,Nagpur.

Contact No: 7798938486

Name: Mohmmad Fayyaz

Email Id: mohammad.mohammad.ai@ghrietn.raisoni.net

Permanent Address: plot no.9, Bada Tajbag,Nagpur

Contact No: 7875230555

Name: Jaykumar Thakare

Email Id: jaykumar412003@gmail .com

Permanent Address: gov. boys hostel,near new

katol naka, nagpur

Contact No: 7385059669

Name: Anuj wadi

Email Id: anujwadi@gmail.com

Permanent Address: Shaktimata Nagar 3, Nagpur

Contact No: 9545222849

Name: Devesh Ambade

Email Id: deveshambade02@icloud.com

Permanent Address: Dr Ambedkar Road, Kamal Chowk Nagpur

Contact No: 7757882709

Name: Aryan Meshram

Email Id: Aryan.meshram.ai@ghrietn.raisoni.net

Permanent Address: Nagpur

Contact No: 7038410083

Photo of Projectees along with the guide and Project