Janda, Karel; Petit, Mathieu

Working Paper

Analyzing decision-making in deep-Q reinforcement learning for trading: A case study on Tesla company and its supply chain

IES Working Paper, No. 40/2024

Provided in Cooperation with:
Charles University, Institute of Economic Studies (IES)

Suggested Citation: Janda, Karel; Petit, Mathieu (2024) : Analyzing decision-making in deep-Q reinforcement learning for trading: A case study on Tesla company and its supply chain, IES Working Paper, No. 40/2024, Charles University in Prague, Institute of Economic Studies (IES), Prague

This Version is available at:
https://hdl.handle.net/10419/311563

INSTITUTE
OF ECONOMIC STUDIES
Faculty of Social Sciences
Charles University

ANALYZING DECISION-MAKING IN DEEP-Q REINFORCEMENT LEARNING FOR TRADING: A CASE STUDY ON TESLA COMPANY AND ITS SUPPLY CHAIN

Karel Janda
Mathieu Petit

IES Working Paper 40/2024

Institute of Economic Studies,
Faculty of Social Sciences,
Charles University in Prague

[UK FSV – IES]

Opletalova 26
CZ-110 00, Prague
E-mail : ies@fsv.cuni.cz
http://ies.fsv.cuni.cz

Institut ekonomických studií
Fakulta sociálních věd
Univerzita Karlova v Praze

Opletalova 26
110 00 Praha 1

E-mail : ies@fsv.cuni.cz
http://ies.fsv.cuni.cz

Bibliographic information:
Janda K., Petit M. (2024): " Analyzing Decision-Making in Deep-Q Reinforcement Learning for Trading: A Case Study on Tesla Company and its Supply Chain " IES Working Papers 40/2024. IES FSV. Charles University.

Analyzing Decision-Making in Deep-Q Reinforcement Learning for Trading: A Case Study on Tesla Company and its Supply Chain

Karel Janda[a,b]
Mathieu Petit[a,*]

[a]Institute of Economic Studies, Faculty of Social Sciences, Charles University, Prague, Czech Republic
[b]Department of Banking and Insurance, Faculty of Finance and Accounting, Prague University of Economics and Business, Czech Republic
*corresponding author

November 2024

Abstract:
This study addresses the economic rationale behind algorithmic trading in the Electric Vehicle (EV) sector, enhancing the interpretability of Q-learning agents. By integrating EV-specific data, such as Tesla's stock fundamentals and key supply chain players such as Albemarle and Panasonic Holdings Corporation, this paper uses a Q-Reinforcement Learning (Q-RL) framework to generate a profitable trading agent. The agent's decisions are analyzed and interpreted using a decision tree to reveal the influence of supply chain dynamics. Tested on a holdout period, the agent achieves monthly profitability above a 2% threshold. The agent shows sensitivity to supply chain instability and identifies potential disruptions impacting Tesla by treating supplier stock movements as proxies for broader economic and market conditions. Indirectly, this approach improves understanding and trust in Q-RL-based algorithmic trading within the EV market.

JEL: G17, Q42, C45, Q55
Keywords: Electric Vehicle Supply Chain, Algorithmic Trading, Machine Learning, Q-Reinforcement Learning, Interpretability

1. Introduction

The current state of knowledge in the supply chain of the Electric Vehicle (EV) sector presents an interesting opportunity for the development of informed and profitable algorithmic trading strategies. This paper therefore advances the fields of financial and energy economics by addressing critical gaps in the interpretability and transparency of algorithmic trading agents within the electric vehicle (EV) sector. It focuses on interpreting the economic rationale underlying the actions of profitable algorithmic trading agents, providing insights into the intricate factors driving EV market movements for researchers and investors. Indirectly and in addition, this paper enhances transparency by identifying key features driving algorithmic trading Q-Reinforcement Learning (Q-RL) agents decisions. This transparency improves investor's trust. Rather than striving to create the most profitable agent, this study aims to produce an agent whose decisions can be understood by the financial analyst familiar with Q-RL and EV market fundamentals.

Reinforcement Learning (RL) as a formalized concept has roots that date back to the 1950s and 1960s (Bellman and Kalaba, 1957; Howard, 1960; Minsky, 1961; Sutton and Barto, 2018). The foundational ideas can be traced to the works of researchers such as Richard Sutton and Andrew G. Barto, who popularized RL in the 1980s and 1990s (Barto et al., 1983; Sutton, 1984; Sutton and Barto, 2018). While Reinforcement Learning (RL) estimators, particularly Q-Reinforcement Learning (Q-RL) and its Deep Learning counterparts, have demonstrated significant success in generating trading signals, their inherent lack of interpretability remains a critical limitation (Mnih et al., 2015; Fischer and Krauss, 2018). Interpreting the economic rationale behind the Q-RL trading agent's actions would provide insights into the EV market. In addition, transparency is crucial in financial markets, particularly in high-stakes trading situations where stakeholders need to trust and comprehend the systems guiding their trading decisions.

The literature on Explainable AI (XAI) and Deep Reinforcement Learning

(DRL) in finance reflects a rising interest in interpretability to address challenges in algorithmic trading, especially in high-stakes financial environments such as EV stocks. While recent reviews (Weber et al., 2024) highlight a trend toward post-hoc explainability, other studies note significant trust issues with current Explainable Reinforcement Learning (XRL) methods, which remain a barrier to widespread adoption (Puiutta and Veith, 2020; Hickling et al., 2023). Interpretability frameworks, such as the one proposed by Zhang et al. (2021), provide taxonomies for neural network interpretability. Although progress has been made with rule-based algorithms such as SIRUS (Bénard et al., 2021), the field still lacks transparent, interpretable frameworks that combine profitability with clear decision-making logic in trading contexts.

Most studies exploring XRL in financial trading apply general machine learning techniques rather than domain-specific approaches, as seen in work on fuzzy reinforcement learning and associative classifiers that boost profitability (Bekiros, 2010; Attanasio et al., 2020). Notably, DRL-based systems are recognized for their adaptability in uncertain market conditions (Mosavi et al., 2020; Sahu et al., 2023), but refining these systems to enhance interpretability without sacrificing performance remains an open issue. Despite efforts to enhance profitability by optimizing policy functions and action space in DRL (Corazza, 2021), interpretability remains a secondary consideration in these studies, often leaving stakeholders in the dark about decision rationales.

Research into the EV and lithium markets emphasizes unique drivers, such as market and tech factors, as well as supply chain dynamics. While studies such as Plante (2023) identify co-movement drivers in EV stocks, and others highlight links between EV demand and lithium price dynamics (Sun et al., 2022; Mo and Jeon, 2018), few integrate these insights into trading algorithms. This gap underlines the need for frameworks that leverage EV-specific fundamentals to enhance trading algorithms' interpretability, a focus of the present study. By tailoring Q-RL with domain-specific data, this research aims to bridge gaps in interpretability and economic rationale, advancing the use of explainable frameworks in EV stock trading.

The remainder of this paper is organized as follows. The second section offers a brief review of XAI in Finance and the Electric Vehicle supply chain. The third section outlines the methodology employed in this research. In the fourth section, the data sources and results are presented. The last section concludes.

2. Interpretable Reinforcement Learning in Finance and Market Dynamics of the EV Supply Chain

2.1. XAI and DRL in Finance: Interpretability

Weber et al. (2024) did a systematic review of XAI in Finance, highlighting a recent preference for post-hoc explainability methods. Puiutta and Veith (2020), Alharin et al. (2020), Glanois et al. (2024), Wells and Bednarz (2021), and Hickling et al. (2023) show that most XRL methods still faces challenges such as the lack of trust. Trust can be enhanced through the application of interpretability methods, which can be approached from various perspectives. However, the diversity of interpretability approaches can lead to confusion and a lack of clarity, creating challenges in navigating this field. Zhang et al. (2021) propose a novel taxonomy for neural network interpretability, organizing research into three dimensions—engagement type, explanation type, and interpretability focus. Various studies aimed at developing methodologies to improve transparency of such estimators. Sequeira and Gervasio (2020) found that using visual summaries of an agent's behavior improved humans' assessment of the DRL agent's strengths and limitations. Bénard et al. (2021) introduce SIRUS, an interpretable rule-based algorithm, demonstrating comparable accuracy to competitors with a higher stability.

Among studies focusing on the adoption of Interpretable Machine Learning methods to develop trading algorithms, Bekiros (2010), Attanasio et al. (2020), and Wang et al. (2020) demonstrate that integrating fuzzy reinforcement learning or associative classifiers, enhances stock market profitability and interpretability. Although momentum-based models work satisfactorily for specific stocks (Nguyen et al., 2021), several notable studies have focused on the use

of DRL for profitable algorithmic trading. Mosavi et al. (2020) and Sahu et al. (2023) highlight that DRL excels in performance under market uncertainty.

Numerous studies including Corazza (2021), Yang et al. (2021), Kong and So (2023), Zhang et al. (2020), Duan et al. (2022) have aimed to enhance the profitability on a wide range of financial products of DRL agents by refining policy functions and the action space search, or adjusting the estimator structure. Other studies, such as Liu et al. (2022), focused on improving profitability by synthetic Data Augmentation.

2.2. Explaining Agent Actions in Financial Trading: A Comparison of XRL Methods

Similar to our research objectives, Kumar et al. (2022) introduces an XRL method using SHAP on a Deep Q Network (DQN) to explain agent actions in financial stock trading, tested on SENSEX and DJIA datasets. According to the interpretability taxonomy proposed by Zhang et al. (2021), this approach is passive, offering explanations through examples to achieve Global Interpretability.

According to this taxonomy our proposed approach is also passive for Global Interpretability but it provides explanations by rules. The focus has been more on refining techniques than input data, but our approach tailors EV market-specific data for the Q-RL agent to optimize trading decisions.

2.3. Market Dynamics and Drivers in the Electric Vehicle and Lithium Industries

Recent overviews of the Electric Vehicle (EV) and EV batteries markets are provided by Mohammadi and Saif (2023) and Rapson and Muehlegger (2023). Plante (2023) decomposed EV and battery supply chain stock returns, identifying three main drivers of EV stock co-movement: the market factor, the tech factor, and a risk factor from latent factors on S&P 500 and Nasdaq 100 stocks. Other studies, such as Mu et al. (2023), identified hidden risks in the EV Lithium-Ion Battery (LIB) supply chain, dominated by manufacturers in China, Japan, and South Korea.

Research also focused on the relationship between the EV and lithium markets. Sun et al. (2022) noted that optimistic EV sales forecasts triggered overproduction across the supply chain, contributing to the 2022 lithium price spike. Mo and Jeon (2018) linked EV demand to short-term lithium price dynamics, while Burney and Killins (2023) found no robust evidence for significant effects of battery material prices on automobile manufacturers' equity prices. Additionally, Baur and Todorova (2018) highlighted that Tesla's equity exhibits a positive sensitivity to oil prices, contrasting with negative oil price sensitivity observed for traditional automobile manufacturers, likely reflecting substitution effects between electric and combustion-engine vehicles. Baur and Gan (2018) identified an "EV-demand effect" for Chinese manufacturers and a "production-cost effect" for a German manufacturer. Alekseev et al. (2024) provided evidence of return spillovers between the EV and lithium markets. These studies suggest that stock market indexes, the tech sector, battery manufacturers, and lithium producers may be key drivers of the EV market.

3. Methodology

3.1. Overview

In this paper, our primary objective is to develop a Q-learning reinforcement learning (Q-RL) trading agent for Tesla's stock that achieves a baseline level of profitability and to provide interpretable insights into its trading decisions. Rather than striving to create the most profitable agent, we aim to produce an agent whose decisions can be understood using one of the most interpretable machine learning estimators. To achieve this, we utilize a decision tree to analyze the agent's actions, with a focus on the role of Tesla's supply chain in its price movements. As this study centers on interpretability rather than maximizing profitability, we do not employ a walk-forward validation approach, which is commonly used in trading agent research to better estimate performance over time. Instead, we use a single holdout period to test the agent, training it on past data and evaluating it on a future period. Given the substantial computational cost (24 hours per training session on a standard CPU), a walk-forward approach

or extensive hyperparameter tuning would be prohibitive. Instead, we rely on hyperparameters that have shown promising performance in related literature (Mnih, 2013; Heaton et al., 2018; Li et al., 2018).

The Q-learning framework used here is a model-free reinforcement learning algorithm designed for decision-making in environments with delayed rewards. We implement an exploration-exploitation strategy through an epsilon-greedy policy, which enables the agent to alternate between exploring new actions and exploiting known profitable strategies. Specifically, the agent's objective is to optimize cumulative rewards, represented as profits in this trading context, over the test period. Given this, Q-RL updates its action-value function $Q(s, a)$, where each Q-value represents the expected future reward of taking action a in state s. This function is updated iteratively based on the Bellman equation: $Q(s, a) = Q(s, a) + \alpha\left(r + \gamma \max_{a'} Q(s', a') - Q(s, a)\right)$, where α is the learning rate, r the reward, and γ the discount factor for future rewards. In our case, the agent is trained to leverage price action, Tesla's stock fundamentals, and key supply chain data to take long, hold, or short actions.

After training, the agent's performance is evaluated with profitability defined as achieving a return above 2% for the test month. If this threshold is met, the agent is deemed sufficiently profitable for further interpretation. A decision tree is then used to analyze the agent's actions, providing an interpretable estimator to trace how supply chain-related factors may influence Tesla's price movements. This approach allows us to examine the economic rationale driving its decisions. Following the interpretability taxonomy proposed by Zhang et al. (2021), our approach to interpretability is considered "passive" because it provides insights based on analyzing already-generated examples from the system rather than intervening in or actively probing the model's decision-making processes. Furthermore, our methodology aims for Global Interpretability, focusing on offering a broad, overarching understanding of how the system operates and makes decisions, as opposed to justifying specific, individual predictions.

3.2. Training a profitable Q-RL agent

3.2.1. Trading Environment

We choose the *gym-trading-env* Python package for its simplicity of usage. The *gym-trading-env* package offers a customizable reinforcement learning environment tailored for training agents on financial trading tasks. Key features include flexibility in defining positions, importing data, and setting up market parameters such as trading fees and borrow interest rates, all of which enhance its use for training a Double Deep Q-Network (DDQN) or other trading algorithms.

The environment is designed around a position-based action space, where the agent's actions determine its allocation in Tesla's stock. We rule that there are only three possible actions, namely long, hold, and short. Although actions could technically range from 1 (full investment) to -1 (full short), including fractional positions that allow mixed allocations (e.g., 0.5 for 50% invested), we opted for a simpler action space. By restricting our action space to just long, hold, and short, we aim to streamline decision-making and focus on fundamental trading strategies, rather than risk and portfolio management.

The environment requires a dataset that includes at least the close price and optionally open, high, low, and volume columns. Custom features, such as price differentials or volume indicators, can be created to enrich observations and tailor them to the agent's strategy needs. These features are static or dynamic and can include rolling calculations that provide context over time.

The environment calculates rewards based on the performance of the agent's portfolio relative to market returns. Typically, the reward r at each time step corresponds to the change in portfolio value pv due to the agent's position, calculated as

$$r_t = ln\left(\frac{pv_t}{pv_{t-1}}\right).$$

The portfolio valuation includes penalties for trading costs, such as fees and borrow interest for short positions, which add realism and incentivize the agent to consider trade-offs between returns and costs. We refer the reader to the

gym-trading-env GitHub repository[1] for further details.

3.2.2. Agent Architecture

The Tesla trading agent is built using a DDQN architecture that approximates the Q-values of the three trading actions, i.e., long, hold, and short. This agent consists of two identical neural networks: the Online Network, used to select actions based on Q-value estimates, and the Target Network, providing more stable Q-value targets during training. Each network has two hidden layers penalized with L_2-penalization, each with $n_1 = n_2 = 256$ neurons, with Rectified Linear Unit (ReLU) activation, a dropout layer at rate $p = 0.1$, and a final output layer for Q-values, where each neuron represents the expected return of taking a specific action. Figure 1 represents the estimator architecture of the DDQN agent.

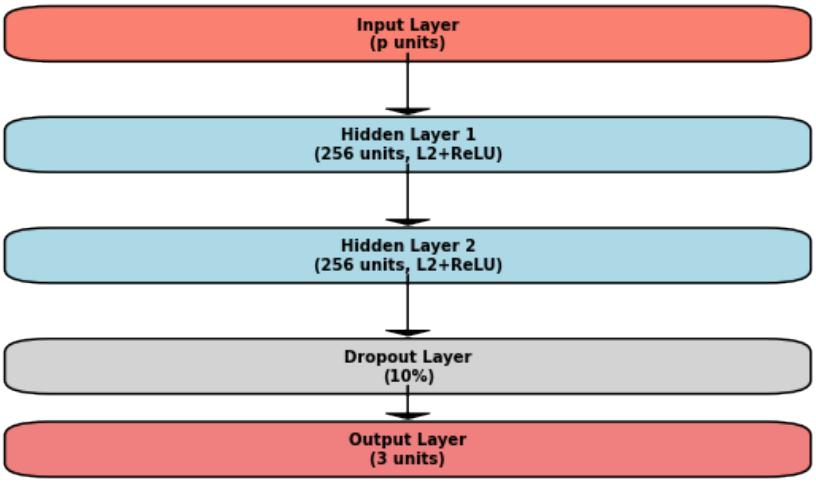

Figure 1: DDQN Architecture for Tesla Trading Agent.

3.2.3. Q-Learning Objective

The core of the DDQN estimator is to approximate the Q-values, which represent the expected return of taking a specific action a from a state s. The

[1] Available here.

Q-value update equation for a DDQN agent is

$$Q(s,a) = r + \gamma Q_{target}(s', \arg\max_{a'} Q_{online}(s', a')), \tag{1}$$

where r is the reward received for taking action a in state s, s' is the next state after taking action a, γ is the discount factor, and Q_{online} and Q_{target} represent the Q-values from the online and target networks, respectively.

3.2.4. Experience Replay Mechanism

To stabilize learning, the agent uses an experience replay buffer, storing experiences in the form (s, a, r', not_done). At each step, the agent samples a random minibatch from this buffer and trains the online network on past experiences. The Experience Replay Loss function is

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}(y_i - Q(s_i, a_i; \theta))^2 + \lambda \sum_{l=1}^{L}\sum_{k=1}^{n_l}\sum_{j=1}^{n_{l-1}}(w_{jk}^l)^2, \tag{2}$$

where N is the batch size, $y_i = r_i + \gamma Q_{target}(s', \arg\max_{a'} Q_{online}(s', a'))$, θ are the parameters of the online network, λ is the L_2-regularization parameter, L is the number of hidden layers, n_l is the number of units in hidden layer $l \geq 1$ (with this notation, the input layer index is 0), and w_{jk}^l is the weight of unit u_j^{l-1} in unit u_k^l of layer l. (We denote by u_k^l the k^{th} unit of layer l.)

3.2.5. Target Network Update

The target network's weights θ_{target} are periodically updated to match the online network's weights θ. This mechanism helps avoid instability in Q-value updates. We introduce an integer τ that controls the rate of updating the target network towards the online network. We define τ as the number of trading days between two updates of the target network's weights.

3.2.6. Epsilon-Greedy Action Selection Policy

To balance exploration and exploitation, the agent follows an ϵ-greedy policy, where the probability of choosing a random action (exploration), ϵ, decays over time. ϵ is updated as follows. For the first $\epsilon_{decaysteps}$ episodes,

$$\epsilon \leftarrow \epsilon - \epsilon_{decay},$$

and for the remaining episodes,

$$\epsilon \leftarrow \epsilon \times \epsilon_{expdecay},$$

where

$$\epsilon_{decay} = \frac{\epsilon_{start} - \epsilon_{end}}{\epsilon_{decaysteps}},$$

ϵ_{start} and ϵ_{end} are the initial and final values of epsilon, and $\epsilon_{expdecay}$ is a constant between 0 and 1.

Figure 2 shows epsilon decays over the episodes, using the constants we choose to train the DDQN, i.e., $\epsilon_{start} = 1.0$, $\epsilon_{end} = 0.01$, $\epsilon_{decaysteps} = 250$, and $\epsilon_{expdecay} = 0.99$.

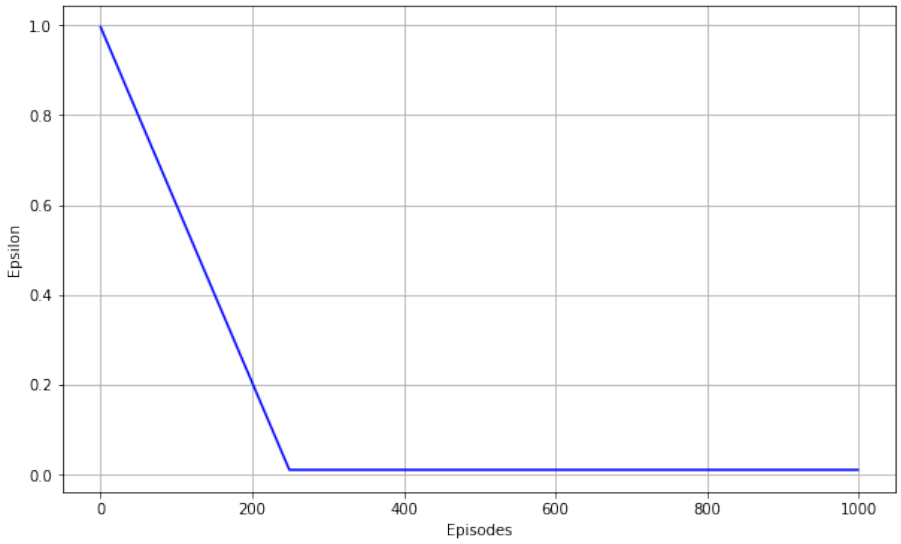

Figure 2: Epsilon Decay Over Episodes.

3.2.7. Training Process Flow

Algorithm 1 outlines the structured steps for training the DDQN, including parameter initialization, action selection via an epsilon-greedy policy, reward calculation, experience replay for Q-value updates, and periodic target network synchronization.

Algorithm 1 Q-Learning Training Process

Initialize Parameters: Set initial Q-values, ϵ, and experience buffer.
for each episode **do**
  Environment Reset: Start from a random point in the training data.
  while episode not done **do**
    Action Selection: Use ϵ-greedy policy to choose action a_t for state s_t.
    Reward Calculation: Observe reward r_t and next state s_{t+1} based on action taken.
    Store Transition: Save (s_t, a_t, r_t, s_{t+1}) in experience buffer.
    Experience Replay: Sample mini-batch from experience buffer.
    Q-value Update: Update Q-values based on target network estimates:

$$Q(s_t, a_t) \leftarrow r_t + \gamma \max_{a'} Q_{\text{target}}(s_{t+1}, a') \tag{3}$$

    if episodes $\equiv \tau = 0$ **then**
      Target Network Update: Sync target network with online network.
    end if
  end while
end for

3.2.8. Parameters and DDQN's hyperparameters

We specify the number of trading days for each episode at 252, reflecting the typical number of trading days in a year. We set $\gamma = 0.99$ which determines the importance of future rewards, indicating that future rewards are nearly as significant as immediate rewards. ϵ is evolving using $\epsilon_{start} = 1.0$, $\epsilon_{end} = 0.01$, $\epsilon_{decaysteps} = 250$, and $\epsilon_{exponentialdecay} = 0.99$. We set $\tau = 100$ which controls the frequency of updates to the target network, ensuring stable learning.

The following hyperparameters were configured for training the DDQN, aligning with established literature that suggests these settings contribute to the development of profitable Q-RL trading agents using the DDQN architecture. In the order of their importance, we set the learning rate constant at 10^{-4}, the momentum term $\beta = 0.99$ to smooth the gradient descent, the mini-batch size at 4096, the number of units in hidden layers at 256, and the number of layers at 2. We are penalizing large weights in the network by applying exclusively to the hidden layers a L_2-regularization and we include a dropout layer at rate $p = 0.1$ between the last hidden layer and the output layer to reduce

overfitting and improve generalization. Due to the limited scope of this paper, we direct the reader to Goodfellow et al. (2016) for additional insights regarding these hyperparameters and their intended purpose.

3.3. Interpreting its decisions

3.3.1. Overview

Input data from the testing period was used to generate the agent's actions, resulting in a dataset that compiles both the input features and corresponding actions taken by the agent for the next trading day. Using this dataset, we trained a decision tree with a maximum depth of 3 to capture simplified rules that reveal the economic rationale behind the agent's decision-making process. This approach allows us to distill the factors driving the agent's long, hold, or short decisions in a transparent and interpretable manner.

3.3.2. Decision Tree Classifier Overview

Leo Breiman, along with Jerome Friedman, Richard Olshen, and Charles Stone, formally introduced decision trees as a systematic method for classification and regression in their 1984 book Classification and Regression Trees (often referred to as "CART") (Breiman et al., 1984). This seminal work established the foundational algorithms and concepts for decision tree estimators, including recursive binary splitting, pruning, and handling both categorical and continuous data. The CART framework has since become a cornerstone in machine learning, inspiring further developments such as bootstrap aggregating (in short, "bagging"), introduced by Breiman (1996) to reduce variance, and random forests, introduced by Breiman (2001), which extend bagging by incorporating random feature selection.

In a decision tree classifier, decisions are modeled in a hierarchical structure that branches based on specific conditions on the input features. For a trading agent with three potential actions (long, hold, short), each path through the tree represents a logical sequence that leads to one of these actions.

A decision tree begins at the root node, which contains the entire dataset. Each node splits based on a feature threshold, dividing the data into two subsets

that lead to subsequent branches. This splitting continues, creating a tree structure until reaching a "leaf" node, where each path represents a class prediction (long, hold, short).

The splits are chosen to maximize the "purity" of each node, measured using the Gini impurity,

$$G(t) = 1 - \sum_{i=1}^{n} p_i^2, \tag{4}$$

where p_i is the proportion of samples belonging to class i (long, hold, short) in node t.

Limiting the decision tree's maximum depth is a common technique for reducing overfitting and enhancing generalization. In our study, we set the maximum depth to 3, prioritizing interpretability. This choice strikes a balance between achieving simple, clear decision rules and retaining sufficient depth to meaningfully capture the agent's trading logic.

To train the classifier, the input data fed to the trading agent (e.g., stock prices, supply chain variables) is used alongside the agent's actions for each time point. The classifier learns decision rules that map specific patterns in the input data to actions such as long, hold, or short, aiming to approximate the agent's decision-making.

Figure 3 is an example of what a decision tree might look like when interpreting trading agent actions. Each split indicates a condition on a feature, leading to a final class prediction (long, hold, short) at each leaf.

Each branch in the decision tree represents a rule that contributes to the trading agent's decision. For example, the far-left branch of the decision tree indicates that when Feature 2 has a value below 0.051 and Feature 1 is less than 0.803, the agent is likely to take a short position for the following trading day. By tracing paths that lead to specific actions, one can derive economic insights into how variables —such as stock prices or supply chain factors— impact the agent's behavior.

Interpreting a Q-learning agent using a decision tree provides a transparent

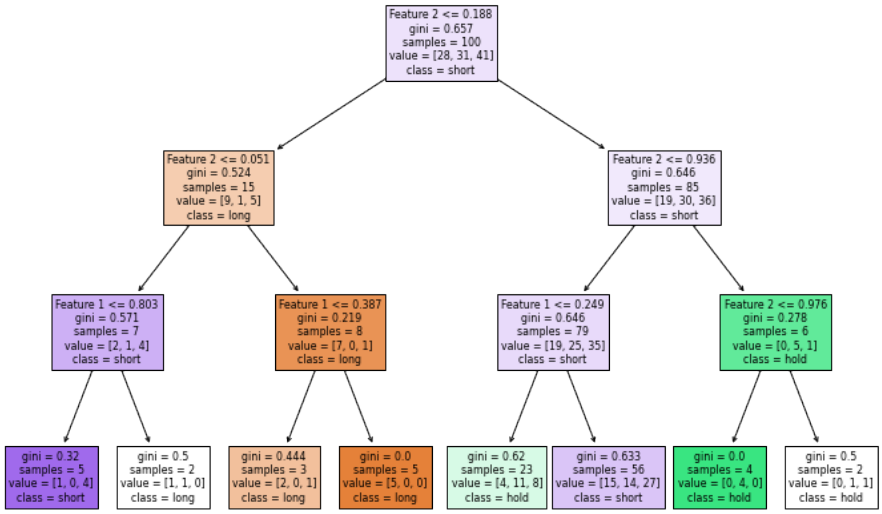

Figure 3: Example decision tree representation using random input data.

framework that improves stakeholder trust. By translating complex decision logic into simple rules, one can better understand how market and supply chain dynamics impact the EV stock trading strategy, bridging the gap between interpretability and profitable decision-making in high-stakes environments.

4. Data and Empirical Results

4.1. Available Data and Transformations

First, we hypothesize a significant influence of Tesla' stock (NASDAQ: TSLA) fundamentals on its close price. Therefore, we utilize a dataset that includes stock fundamentals for Tesla as control variables. To investigate this, we utilize a detailed dataset that includes key stock fundamentals for Tesla as control variables. Specifically, we focus on Tesla's market capitalization and price-to-earnings ratio. We also explored additional stock fundamentals—debt-to-equity ratio, market beta, dividend yield, earnings before interest and taxes, and free cash flow—but their limited data availability led to their exclusion. We provide

	D/E	Beta	EBIT	FCF	Market Cap.	P/E
count	56	144	61	46	3638	1101
mean	152.2	2.3	1.9×10^{09}	3.0×10^{08}	2.4×10^{11}	322.7
std	134.0	0.02	4.3×10^{09}	1.2×10^{09}	3.3×10^{11}	408.1
min	4.6	2.27	-2.2×10^{09}	-2.5×10^{09}	1.5×10^{09}	33.0
25%	33.8	2.28	-3.8×10^{08}	-5.4×10^{08}	2.2×10^{10}	61.1
50%	136.9	2.3	-1.3×10^{08}	1.1×10^{08}	4.5×10^{10}	89.0
75%	243.2	2.31	2.0×10^{09}	9.6×10^{08}	5.6×10^{11}	396.6
max	705.6	2.36	1.4×10^{10}	3.3×10^{09}	1.2×10^{12}	1722.5

Table 1: The descriptive statistics of Tesla's stock daily fundamentals. EBIT, FCF, and Market Cap. are reported in US Dollar (USD). Note: D/E = Debt-to-Equity Ratio; Beta = Beta Coefficient; EBIT = Earnings Before Interest and Taxes; FCF = Free Cash Flow; Market Cap. = Market Capitalization; P/E = Price-to-Earnings Ratio.

descriptive statistics for each of these stock fundamental variables in Table 1 to support our decision.

In addition, we include Tesla's daily open, high, low, and close prices in our dataset. The close price is obviously essential as it is used to compute the agent's returns and is included among the features that guide the agent's trading decisions for the following day. The inclusion of open, high, and low prices enables us to construct informative price action features, enhancing the agent's decision-making framework. Figure 4 provides a comprehensive overview of these price points.

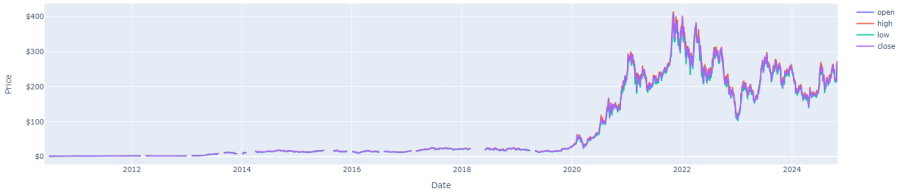

Figure 4: Time series of Tesla's daily stock prices, illustrating the open, high, low, and close prices. The chart highlights the availability of price data.

As a central aspect of this study, we expand our dataset to include the broader electric vehicle supply chain by incorporating the close stock prices of two major lithium producers, namely Albemarle Corporation (NYSE: ALB)

	ALB	SQM	CATL	Panasonic	Samsung SDI	LG
count	7708	7426	1560	9997	10678	685
mean	58.22	24.69	166.61	1582.92	129352	454972
std	61.07	23.17	94.07	528.66	166322	76602
min	6.19	1.49	20.11	385.0	4301	321000
25%	12.31	3.86	63.53	1209.52	23832	398500
50%	40.87	18.48	184.94	1485.0	66400	436500
75%	73.04	40.94	231.51	1926.0	147500	525000
max	325.38	113.52	382.22	3230.0	817000	624000

Table 2: The descriptive statistics for the close prices of the major lithium and lithium-ion battery manufacturers. ALB and SQM are reported in US Dollar (USD), CATL and Panasonic are reported in Japenese Yen (JPY), and Samsung SDI and LG are reported in South Korean Won (KRW). Note: ALB = Albemarle Corporation; SQM = Sociedad Química y Minera de Chile; CATL = Contemporary Amperex Technology Co Ltd; Panasonic = Panasonic Holdings Corp; Samsung SDI = Samsung SDI Co Ltd; LG = LG Energy Solution Ltd.

and Sociedad Química y Minera de Chile (NYSE: SQM), alongside leading battery manufacturers, specifically Contemporary Amperex Technology Co Ltd (CATL) (Shenzhen SE: 300750), Panasonic Holdings Corp (Tokyo SE: 6752), and Samsung SDI Co Ltd (Korea SE: 006400). Due to limited data availability, we exclude LG Energy Solution Ltd. (Korea SE: 373220), another prominent lithium-ion battery manufacturer. We provide descriptive statistics for the close prices of these stocks in Table 2 to support our decision.

The dataset, sourced through the Refinitiv Eikon API for Python, spans from July 29, 2020, to November 13, 2024. The training period lies from July 29, 2020 to September 30, 2024. The testing period lies from October 1, 2024 to November 13, 2024. Table 3 presents descriptive statistics of the training and test sets combined. The dataset is the intersection of the fundamentals, the Tesla's stock prices, and the close prices of the previously introduced stocks of the electric vehicle supply chain. We introduced three additional intraday-specific features: the Open-to-Close (O/C), High-to-Close (H/C), and Low-to-Close (L/C) ratios, which provide some insights into Tesla stock's intraday price action. For instance, a high H/C ratio on trading day d indicates that the highest trade price recorded on that day exceeded the closing price at the session's end. This indicates that Tesla's stock price peaked significantly above

the closing price during the trading session, suggesting high intraday volatility. This could reflect heightened investor enthusiasm or speculation, leading to temporary surges in the price. However, the fact that the price closed lower than the high may indicate that, despite the strong buying interest earlier in the day, sellers gained control by the end of the session, causing the price to fall back. In practice, this pattern could signal potential price resistance or profit-taking.

	P/E	Market Cap.	ALB	SQM	Samsung SDI	CATL	Panasonic	TSLA	O/C	H/C	L/C
count	1076	1076	1076	1076	1076	1076	1076	1076	1076	1076	1076
mean	325.5	7.1×10^{11}	179.65	60.46	573079	223.67	1283.84	229.49	1.0	1.02	0.98
std	409.4	1.8×10^{11}	60.87	20.32	135185	57.84	186.35	58.74	0.03	0.02	0.02
min	33.0	2.6×10^{11}	72.85	27.88	255500	102.4	862.9	91.63	0.87	1.0	0.87
25%	60.3	5.8×10^{11}	125.34	46.02	446375	185.38	1153.5	188.87	0.98	1.01	0.97
50%	91.6	7.0×10^{11}	181.98	53.2	591000	222.3	1273.5	227.06	1.0	1.02	0.98
75%	405.0	8.1×10^{11}	228.0	76.31	690000	257.38	1395.12	261.23	1.02	1.03	0.99
max	1722.5	1.2×10^{12}	325.38	113.52	817000	382.22	1794.5	409.97	1.15	1.15	1.0

Table 3: The descriptive statistics of the training and test sets combined. Market Cap., ALB, SQM, and TSLA are reported in US Dollar (USD), Samsung SDI is reported in South Korean Won (KRW), and CATL and Panasonic are reported in Japenese Yen (JPY). Note: P/E = Price-to-Earnings Ratio; Market Cap. = Market Capitalization; ALB = Albemarle Corporation; SQM = Sociedad Química y Minera de Chile; Samsung SDI = Samsung SDI Co Ltd; CATL = Contemporary Amperex Technology Co Ltd; Panasonic = Panasonic Holdings Corp; TSLA = Tesla Inc; O/C = Open-to-Close ratio; H/C = High-to-Close ratio; L/C = Low-to-Close ratio.

To facilitate convergence towards a local optimum of the trained parameters in the Multi-Layer Perceptron, we apply two standard data transformations, namely differentiation followed by a power transformation. The former centers data around zero and improves stationarity and the later promotes a near-normal distribution of features. Additionally, we introduce four lagged features of Tesla's stock daily transformed close price at lags 2, 5, 10, and 21. It offers the trading agent a broader context on historical price movements, which can improve its ability to predict the stock's next-day performance. Descriptive statistics, unit root and normality tests of the transformed data are provided in Tables 4 and 5.

	Market Cap.	P/E	ALB	SQM	SDI	CATL	Panasonic
Observations	1054	1054	1054	1054	1054	1054	1054
Mean	0.01	-0.0	0.0	-0.0	-0.01	0.0	0.01
Std.dev	1.02	0.99	0.99	0.99	0.99	1.05	1.01
Minimum	-4.86	-10.49	-6.96	-6.14	-3.63	-7.66	-5.86
Median	-0.01	-0.05	-0.04	-0.03	0.04	0.02	-0.03
Maximum	5.73	8.77	3.94	4.6	3.6	6.45	4.77
Skewness	2.45	2.45	2.45	2.45	2.45	2.45	2.45
Kurtosis	7.0	7.0	7.0	7.0	7.0	7.0	7.0
Jarque-Bera	566.19^{***}	46852.24^{***}	544.46^{***}	712.56^{***}	77.02^{***}	1260.83^{***}	627.3^{***}
Shapiro-Wilk	0.96^{***}	0.55^{***}	0.97^{***}	0.95^{***}	0.98^{***}	0.94^{***}	0.95^{***}
ADF	-9.54^{***}	-7.28^{***}	-31.08^{***}	-31.9^{***}	-33.03^{***}	-20.28^{***}	-21.66^{***}
PP	-32.81^{***}	-34.12^{***}	-31.04^{***}	-31.99^{***}	-33.03^{***}	-34.25^{***}	-32.34^{***}
KPSS	0.08	0.66^{**}	0.31	0.25	0.27	0.13	0.06

Table 4: The descriptive statistics of the transformed training and test sets combined with unit root and stationary tests. Note 1: P/E = Price-to-Earnings Ratio; Market Cap. = Market Capitalization; ALB = Albemarle Corporation; SQM = Sociedad Química y Minera de Chile; SDI = Samsung SDI Co Ltd; CATL = Contemporary Amperex Technology Co Ltd; Panasonic = Panasonic Holdings Corp. Note 2: ***, **, * indicate statistical significance at 1%, 5% and 10%.

	C	O/C	H/C	L/C	C (d-2)	C (d-5)	C (d-10)	C (d-21)
Observations	1054	1054	1054	1054	1054	1054	1054	1054
Mean	0.01	0.0	0.0	0.0	0.01	0.0	0.0	-0.0
Std.dev	1.03	0.99	0.99	0.99	1.02	1.02	1.01	0.99
Minimum	-4.95	-4.2	-4.44	-4.27	-4.95	-4.95	-4.95	-4.95
Median	0.0	-0.02	0.0	0.02	0.0	0.0	0.0	0.0
Maximum	5.63	3.71	4.55	4.01	5.63	5.63	5.63	5.63
Skewness	2.45	2.45	2.45	2.45	2.45	2.45	2.45	2.45
Kurtosis	7.0	7.0	7.0	7.0	7.0	7.0	7.0	7.0
Jarque-Bera	561.0^{***}	66.51^{***}	177.86^{***}	135.96^{***}	558.87^{***}	584.68^{***}	610.13^{***}	421.99^{***}
Shapiro-Wilk	0.96^{***}	0.99^{***}	0.98^{***}	0.98^{***}	0.96^{***}	0.96^{***}	0.96^{***}	0.97^{***}
ADF	-9.57^{***}	-12.83^{***}	-12.36^{***}	-13.61^{***}	-9.46^{***}	-9.45^{***}	-9.66^{***}	-9.63^{***}
PP	-33.15^{***}	-176.88^{***}	-161.97^{***}	-166.61^{***}	-33.13^{***}	-33.16^{***}	-33.11^{***}	-33.6^{***}
KPSS	0.08	0.06	0.04	0.07	0.08	0.08	0.09	0.13

Table 5: The descriptive statistics of the transformed training and test sets combined with unit root and stationary tests. Note 1: C = Tesla's Close price; O/C = Open-to-Close ratio; H/C = High-to-Close ratio; L/C = Low-to-Close ratio; C (d-x) = Lagged Tesla's close price at d-x. Note 2: ***, **, * indicate statistical significance at 1%, 5% and 10%.

4.2. Empirical Results

4.2.1. Agent's returns

Figure 5 shows the monthly returns of a DDQN trained on the period from July 29, 2020, to September 30, 2024. In this plot, light-colored bars represent monthly returns during the training phase, while darker-colored bars indicate performance in the testing period. Throughout the training period, the agent achieved significant positive returns, with all months reaching over 10%, indicating successful learning of profitable trading strategies during the training period. However, the returns show variability, with some months yielding low returns and others reaching over 80%, suggesting that market conditions during certain periods posed more challenges for the agent. The authors hypothesize that this variability is largely driven by Tesla's stock price volatility, which significantly impacts the agent's performance in both positive and negative directions.

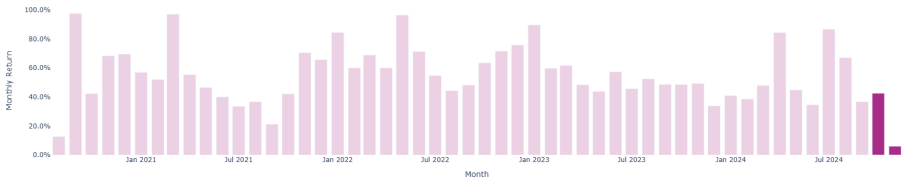

Figure 5: Monthly Returns of the DDQN: Light-colored bars represent returns during the training period (July 2020 - September 2024), while darker bars indicate returns in the testing period, highlighting system performance in both familiar and unseen market conditions.

In the testing period, represented by darker bars toward the end of the timeline, the agent's returns for October fall within the typical range of returns observed during the training period, indicating that the system performed as expected in the absence of significant market disruptions. However, the November monthly return was notably impacted by the U.S. presidential election. The election results were announced by the end of November 5th, leading to a significant market response. Between November 4th and 11th, Tesla's stock price surged by 44%, a sharp increase that the agent could not anticipate, as it was not trained to recognize the impact of such political events. This unexpected

event is one element of explanation of the agent's low performance for November compared to other months, as it lacked information on the election's influence on the market.

4.2.2. Interpretation of the agent's actions

Figure 6 presents a trained classifier which is used to interpret the actions of the DDQN. The target variable is the agent's actions (long, hold, or short) on trading day d for trading day $d+1$, and the features are identical to those used in training the agent, except without Power transformation to enhance interpretability. The features include: *feature_open*, defined as the difference between the Tesla's open/close (O/C) ratio for trading day d and that of day $d-1$; *feature_low*, defined as the difference between the Tesla's low/close (L/C) ratio for trading day d and that of day $d-1$; *feature_close*, the difference in Tesla's close prices between days d and $d-1$; *feature_close_5*, the difference in Tesla's close prices between days $d-5$ and $d-6$; *feature_low*, the difference between the Tesla's low/close (L/C) ratio for day d and that of day $d-1$; *feature_006400_close*, the difference in Samsung SDI's stock close prices between days d and $d-1$; *feature_300750_close*, the difference in CATL's stock close prices between days d and $d-1$; and *feature_market_capitalization*, the difference in Tesla's market capitalization between days d and $d-1$. To prioritize interpretability, the tree's maximum depth is set to 3, highlighting the most significant decision rules. Each node shows a feature threshold used to split the data, along with the Gini index and sample distribution, revealing the primary factors influencing the agent's trading decisions. This structure provides insight into the agent's strategy, showing how specific market and company features drive its actions.

The authors arbitrarily focus on the interpretation of two scenarios: (1) described by the total left branch and (2) described by the shortest branch of the fitted decision tree. The agent's trading strategy, as described in the two scenarios, reflects a nuanced approach to interpreting price movements of key players in the EV supply chain, specifically Samsung SDI and CATL, and

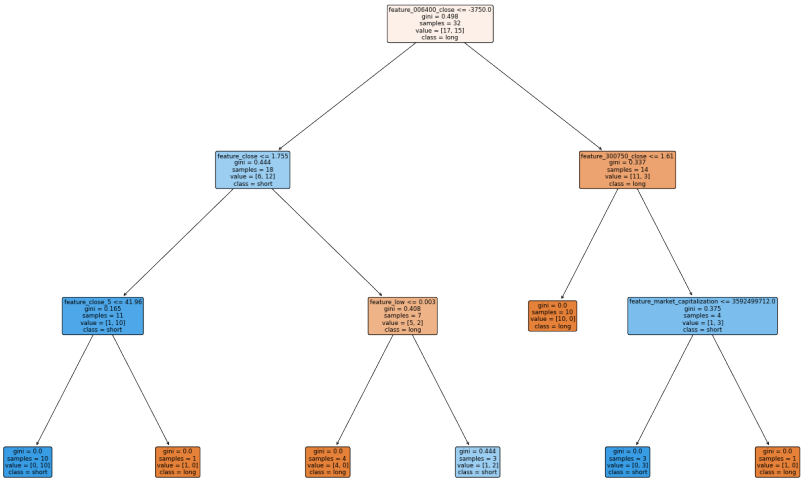

Figure 6: Decision tree illustrating the agent's actions, trained on the same features as the agent (without Power transformation) with a maximum depth of 3 for interpretability. Each node represents a decision rule, showing key features influencing the agent's 'long' or 'short' trading decisions.

using them as indicators for Tesla's stock movements. These decisions highlight the agent's attempt to capture supply chain dynamics, market sentiment, and potential future demand for Tesla's products.

This paragraph focuses on the interpretation of scenario (1). Samsung SDI, as a major lithium-battery manufacturer, is a key supplier in Tesla's supply chain. A decline in Samsung SDI's stock price might suggest a perceived short-term weakness in the EV battery supply chain, either due to supply constraints, decreased demand, or external pressures affecting the sector. The agent's initial check of Samsung SDI's price reflects a reliance on upstream suppliers as an early indicator of potential downstream impacts on Tesla. After observing a decline in Samsung SDI's price, the agent looks at Tesla's own price movement. If Tesla's stock is also declining, this may reinforce the indication of broader negative sentiment or challenges facing Tesla, perhaps due to sector-wide issues or anticipated lower demand. The agent uses this as a filter to validate whether

the perceived weakness in the supply chain might also be impacting Tesla. If, however, the agent observes that Tesla had an extreme price increase five days ago, it takes a long position for tomorrow. This suggests that the agent views such a large recent price increase as an indicator of underlying strength or demand resilience that might counteract the current short-term weakness. By going long, the agent expresses confidence that the price dip may be temporary and that the stock could rebound, driven by prior momentum or demand fundamentals. Conversely, if Tesla did not exhibit an extreme increase five days ago, the agent takes a short position. This conservative approach suggests that the agent interprets the lack of recent strong momentum as an indicator that Tesla might lack the resilience to withstand the current downward pressures, perhaps due to concerns about demand or profitability. In this case, the agent views Tesla's near-term outlook as negative, given the combined signals of weakness from both the supplier (Samsung SDI) and Tesla's current price trend. This decision rule reflects a nuanced view of the EV market's interconnectedness. The agent considers upstream supply chain signals (Samsung SDI) as initial indicators of possible future impacts on Tesla, then seeks confirmation through Tesla's price trend and momentum. The need for an "extreme increase" five days ago for a long position implies the agent's reliance on strong historical price momentum as a counterweight to recent supply chain weakness. This shows the agent's preference for conservative, momentum-driven entries and exits, aiming to minimize exposure to Tesla during periods of perceived sector-wide instability unless strong prior gains indicate possible resilience.

This paragraph focuses on the interpretation of scenario (2). When Samsung SDI's stock price increases, the agent sees this as an initial positive signal for the EV supply chain. An increase in Samsung SDI's stock suggests investor confidence in the battery supplier, possibly due to positive market conditions, demand forecasts, or operational strength. This could signal improved supply chain conditions or optimism about future EV demand, indirectly benefiting Tesla. Despite Samsung SDI's positive movement, the agent checks CATL's price movement for a counter-signal. If CATL's price is declining, the agent

interprets this as a potential market anomaly, where two major suppliers are moving in opposite directions. This divergence might indicate that CATL's decline is temporary or specific to CATL, rather than an indicator of broader EV market weakness. The agent's decision to go long on Tesla when Samsung SDI is up and CATL is down suggests that it perceives the overall supply chain environment as favorable or stabilizing. The decline in CATL, juxtaposed with Samsung SDI's increase, could be seen as an opportunity, where CATL's price decline does not necessarily impact Tesla directly. The agent's long position in this scenario shows confidence that Tesla may benefit from improved supply conditions or sector sentiment, with CATL's decline being viewed as a non-critical or temporary setback within the supply chain. The agent's trading rule in this scenario reflects a contrarian view, where it capitalizes on perceived misalignments within the EV supply chain. By going long on Tesla when Samsung SDI is up and CATL is down, the agent demonstrates an understanding of the heterogeneity among suppliers and a willingness to interpret one supplier's weakness (CATL) as a potential opportunity rather than a risk factor. This suggests the agent's belief that diverging movements among suppliers might provide favorable entry points for Tesla, as it sees the broader market sentiment (reflected in Samsung SDI's increase) as more influential for Tesla's outlook than isolated declines among specific suppliers.

In the first scenario, the agent uses supply chain weakness (Samsung SDI's decline) as a cautionary signal but is willing to take a long position if there was recent positive momentum in Tesla's stock. This suggests that the agent interprets supply chain signals as potentially temporary, using recent price increases to indicate resilience. Without this momentum, it takes a cautious stance and shorts Tesla, indicating sensitivity to supply chain instability. In the second scenario, the agent takes a contrarian approach, interpreting divergence between Samsung SDI's increase and CATL's decline as a possible opportunity to go long on Tesla. This reflects the agent's perception that not all supply chain signals are equally influential and that a mixed signal may still offer a favorable outlook for Tesla, especially if broader supply chain sentiment is positive

(Samsung SDI up). Both decision rules reveal the agent's sophisticated use of supply chain stock prices to gauge market sentiment and future demand. By treating supplier stock movements as proxies for broader economic and market conditions, the agent aims to identify potential demand or supply disruptions impacting Tesla. This highlights the agent's understanding of the interdependency between Tesla and its suppliers and the use of this knowledge to make informed trading decisions.

5. Conclusion

This study addresses gaps in the economic rationale of algorithmic trading agents' decisions within the EV sector. RL, especially Q-RL and DRL, has shown promise for trading, but often lacks transparency, limiting trust among stakeholders. XAI in finance emphasizes the need for interpretability, yet many methods fail to provide a clear economic rationale for trading decisions. Previous research has examined EV market dynamics, such as co-movement drivers and the relationship between EV and lithium stock prices, but few studies integrate these insights into trading algorithms. By customizing Q-RL with EV-specific data, this research aims to enhance both interpretability and economic rationale, providing investors and researchers with insights into the factors driving trading decisions and fostering greater transparency in the EV market.

Our study uses a dataset focusing on Tesla's stock fundamentals, including market capitalization and price-to-earnings ratio, alongside daily open, high, low, and close prices. We extended the dataset to encompass the close prices of key players in the EV supply chain, such as Albemarle, SQM, CATL, Panasonic, and Samsung SDI. Additionally, we created intraday price action features (O/C, H/C, L/C ratios) to enhance the agent's decision-making framework. Data transformations (differentiation and power transformation) and four lagged Tesla close prices provide context on historical price movements, improving predictability for the trading agent.

Our study develops a Q-learning agent for Tesla's stock, prioritizing interpretability over profitability. Using a decision tree, we analyze the agent's

actions to reveal supply chain influences. The system is tested on a holdout period, optimizing cumulative rewards, and evaluated for monthly profitability above a 2% threshold.

The agent's approach shows a keen awareness of EV supply chain dynamics and an ability to distinguish between short-term volatility and fundamental shifts in market sentiment. By balancing caution with contrarian strategies, it seeks to capture upside opportunities in Tesla while hedging against supply chain risks, using nuanced price patterns from both Tesla and its key suppliers to guide its decisions.

To enhance this study, future research could explore several improvements: adjusting the training and testing split to 50%/50% rather than allocating most of the historical data for training, training at least 30 distinct DDQRL agents to increase robustness in profitability estimation and interpretability, expanding features to include market sentiment, macroeconomic indicators such as interest rates, and market news, and broadening the analysis to encompass all EV manufacturers, beyond just Tesla.

CRediT authorship contribution statement

Karel Janda: Conceptualization, Methodology, Writing - Review & Editing, Supervision. **Mathieu Petit:** Data Curation, Software, Validation, Formal analysis, Writing - Original Draft.

Acknowledgements

This paper is part of a project GEOCEP that has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 870245. Karel Janda acknowledges financial support from the Czech Science Foundation (grant no. 24-10008S). The views expressed here are those of the authors and not necessarily those of our institutions. All remaining errors are solely our responsibility.

References

Alekseev, O., Janda, K., Petit, M., and Zilberman, D. (2024). Return and volatility spillovers between the raw material and electric vehicles markets. Energy Economics, 137(C).

Alharin, A., Doan, T.-N., and Sartipi, M. (2020). Reinforcement learning interpretation methods: A survey. IEEE Access, 8:171058–171077.

Attanasio, G., Cagliero, L., and Baralis, E. (2020). Leveraging the explainability of associative classifiers to support quantitative stock trading. In Proceedings of the Sixth International Workshop on Data Science for Macro-Modeling, DSMM '20, New York, NY, USA. Association for Computing Machinery.

Barto, A. G., Sutton, R. S., and Anderson, C. W. (1983). Neuronlike adaptive elements that can solve difficult learning control problems. IEEE transactions on systems, man, and cybernetics, SMC-13(5):834–846.

Baur, D. G. and Gan, D. (2018). Electric vehicle production and the price of lithium. SSRN Electronic Journal.

Baur, D. G. and Todorova, N. (2018). Automobile manufacturers, electric vehicles and the price of oil. Energy Economics, 74.

Bekiros, S. (2010). Heterogeneous trading strategies with adaptive fuzzy actor-critic reinforcement learning: A behavioral approach. Journal of Economic Dynamics and Control, 34(6):1153–1170.

Bellman, R. and Kalaba, R. (1957). Dynamic programming and statistical communication theory. Proceedings of the National Academy of Sciences, 43(8):749–751.

Bénard, C., Biau, G., da Veiga, S., and Scornet, E. (2021). Interpretable random forests via rule extraction. In Banerjee, A. and Fukumizu, K., editors, Proceedings of The 24th International Conference on Artificial Intelligence

and Statistics, volume 130 of Proceedings of Machine Learning Research, pages 937–945. PMLR.

Breiman, L. (1996). Bagging predictors. Machine Learning, 24(2):123–140.

Breiman, L. (2001). Random forests. Machine Learning, 45:5–32.

Breiman, L., Friedman, J., Olshen, R., and Stone, C. (1984). Classification and Regression Trees. Chapman and Hall/CRC, 1st edition.

Burney, R. B. and Killins, R. N. (2023). Sustainability and electrification of the automobile industry: Battery metals and equity returns. The Journal of Investing, 32:63–78.

Corazza, M. (2021). Q-learning-based financial trading: Some results and comparisons. In Esposito, A., Faundez-Zanuy, M., Morabito, F. C., and Pasero, E., editors, Progresses in Artificial Intelligence and Neural Systems, pages 343–355. Springer Singapore, Singapore.

Duan, Z., Chen, C., Cheng, D., Liang, Y., and Qian, W. (2022). Optimal action space search: An effective deep reinforcement learning method for algorithmic trading. In Proceedings of the 31st ACM International Conference on Information & Knowledge Management, CIKM '22, page 406–415, New York, NY, USA. Association for Computing Machinery.

Fischer, T. and Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270(2):654–669.

Glanois, C., Weng, P., Zimmer, M., Li, D., Yang, T., Hao, J., and Liu, W. (2024). A survey on interpretable reinforcement learning. Machine Learning, 113(8):5847–5890.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). Deep Learning. MIT Press. http://www.deeplearningbook.org.

Heaton, J. B., Polson, N. G., and Witte, J. H. (2018). Deep learning in finance. Preprint.

Hickling, T., Zenati, A., Aouf, N., and Spencer, P. (2023). Explainability in deep reinforcement learning: A review into current methods and applications. ACM Computing Surveys, 56(5):1–35.

Howard, R. A. (1960). Dynamic Programming and Markov Processes, pages 39–47. MIT Press.

Kong, M. and So, J. (2023). Empirical analysis of automated stock trading using deep reinforcement learning. Applied Sciences, 13(1).

Kumar, S., Vishal, M., and Ravi, V. (2022). Explainable reinforcement learning on financial stock trading using shap. Preprint.

Li, L., Jamieson, K., DeSalvo, G., Rostamizadeh, A., and Talwalkar, A. (2018). Hyperband: A novel bandit-based approach to hyperparameter optimization. Journal of Machine Learning Research, 18(185):1–52.

Liu, C., Ventre, C., and Polukarov, M. (2022). Synthetic data augmentation for deep reinforcement learning in financial trading. In Proceedings of the Third ACM International Conference on AI in Finance, ICAIF '22, page 343–351, New York, NY, USA. Association for Computing Machinery.

Minsky, M. (1961). Steps toward artificial intelligence. In Proceedings of the IRE, volume 49, pages 8–30.

Mnih, V. (2013). Playing atari with deep reinforcement learning. Preprint.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., et al. (2015). Human-level control through deep reinforcement learning. Nature, 518(7540):529–533.

Mo, J. Y. and Jeon, W. (2018). The impact of electric vehicle demand and battery recycling on price dynamics of lithium-ion battery cathode materials: A vector error correction model (VECM) analysis. Sustainability, 10(8).

Mohammadi, F. and Saif, M. (2023). A comprehensive overview of electric vehicle batteries market. e-Prime - Advances in Electrical Engineering, Electronics and Energy, 3:100127.

Mosavi, A., Faghan, Y., Ghamisi, P., Duan, P., Ardabili, S. F., Salwana, E., and Band, S. S. (2020). Comprehensive review of deep reinforcement learning methods and applications in economics. Mathematics, 8(10).

Mu, D., Ren, H., Wang, C., Yue, X., Du, J., and Ghadimi, P. (2023). Structural characteristics and disruption ripple effect in a meso-level electric vehicle lithium-ion battery supply chain network. Resources Policy, 80:103225.

Nguyen, D. K., Sensoy, A., Vo, D.-T., and von Mettenheim, H.-J. (2021). Does short-term technical trading exist in the Vietnamese stock market? Borsa Istanbul Review, 21(1):23–35.

Plante, M. D. (2023). Investing in the batteries and vehicles of the future: A view through the stock market. Working Paper No. 2314, Federal Reserve Bank of Dallas. Available at SSRN or http://dx.doi.org/10.24149/wp2314r1.

Puiutta, E. and Veith, E. M. S. P. (2020). Explainable reinforcement learning: A survey. In Holzinger, A., Kieseberg, P., Tjoa, A. M., and Weippl, E., editors, Machine Learning and Knowledge Extraction, pages 77–95, Cham. Springer International Publishing.

Rapson, D. S. and Muehlegger, E. (2023). The economics of electric vehicles. Review of Environmental Economics and Policy, 17(2):274–294.

Sahu, S. K., Mokhade, A., and Bokde, N. D. (2023). An overview of machine learning, deep learning, and reinforcement learning-based techniques in quantitative finance: Recent progress and challenges. Applied Sciences, 13(3):1956.

Sequeira, P. and Gervasio, M. (2020). Interestingness elements for explainable reinforcement learning: Understanding agents' capabilities and limitations. Artificial Intelligence, 288:103367.

Sun, X., Ouyang, M., and Hao, H. (2022). Surging lithium price will not impede the electric vehicle boom. Joule, 6(8):1738–1742.

Sutton, R. S. (1984). Temporal Credit Assignment in Reinforcement Learning. University of Massachusetts Amherst.

Sutton, R. S. and Barto, A. G. (2018). Reinforcement Learning: An Introduction. MIT Press, Cambridge, MA, 2nd edition.

Wang, X., Zhang, Y., and Chen, Y. (2020). A novel lasso regression model for sector rotation trading strategies with "economy-policy" cycles. In Proceedings of the 2020 IEEE International Conference on Big Data (Big Data), pages 5473–5479.

Weber, P., Carl, K. V., and Hinz, O. (2024). Applications of explainable artificial intelligence in finance—a systematic review of finance, information systems, and computer science literature. Management Review Quarterly, 74(2):867–907.

Wells, L. and Bednarz, T. (2021). Explainable AI and reinforcement learning—a systematic review of current approaches and trends. Frontiers in Artificial Intelligence, 4.

Yang, H., Liu, X.-Y., Zhong, S., and Walid, A. (2021). Deep reinforcement learning for automated stock trading: an ensemble strategy. In Proceedings of the First ACM International Conference on AI in Finance, ICAIF '20, New York, NY, USA. Association for Computing Machinery.

Zhang, Y., Tiňo, P., Leonardis, A., and Tang, K. (2021). A survey on neural network interpretability. IEEE Transactions on Emerging Topics in Computational Intelligence, 5(5):726–742.

Zhang, Z., Zohren, S., and Roberts, S. (2020). Deep reinforcement learning for trading. The Journal of Financial Data Science, 2:25–40.

IES Working Paper Series

2024

1. Nino Buliskeria, Jaromir Baxa, Tomáš Šestořád: *Uncertain Trends in Economic Policy Uncertainty*
2. Martina Lušková: *The Effect of Face Masks on Covid Transmission: A Meta-Analysis*
3. Jaromir Baxa, Tomáš Šestořád: *How Different are the Alternative Economic Policy Uncertainty Indices? The Case of European Countries.*
4. Sophie Ghvanidze, Soo K. Kang, Milan Ščasný, Jon Henrich Hanf: *Profiling Cannabis Consumption Motivation and Situations as Casual Leisure*
5. Lorena Skufi, Meri Papavangjeli, Adam Gersl: *Migration, Remittances, and Wage-Inflation Spillovers: The Case of Albania*
6. Katarina Gomoryova: *Female Leadership and Financial Performance: A Meta-Analysis*
7. Fisnik Bajrami: *Macroprudential Policies and Dollarisation: Implications for the Financial System and a Cross-Exchange Rate Regime Analysis*
8. Josef Simpart: *Military Expenditure and Economic Growth: A Meta-Analysis*
9. Anna Alberini, Milan Ščasný: *Climate Change, Large Risks, Small Risks, and the Value per Statistical Life*
10. Josef Bajzík: *Does Shareholder Activism Have a Long-Lasting Impact on Company Value? A Meta-Analysis*
11. Martin Gregor, Beatrice Michaeli: *Board Bias, Information, and Investment Efficiency*
12. Martin Gregor, Beatrice Michaeli: *Board Compensation and Investment Efficiency*
13. Lenka Šlegerová: *The Accessibility of Primary Care and Paediatric Hospitalisations for Ambulatory Care Sensitive Conditions in Czechia*
14. Kseniya Bortnikova, Tomas Havranek, Zuzana Irsova: *Beauty and Professional Success: A Meta-Analysis*
15. Fan Yang, Tomas Havranek, Zuzana Irsova, Jiri Novak: *Where Have All the Alphas Gone? A Meta-Analysis of Hedge Fund Performance*
16. Martina Lušková, Kseniya Bortnikova: *Cost-Effectiveness of Women's Vaccination Against HPV: Results for the Czech Republic*
17. Tersoo David Iorngurum: *Interest Rate Pass-Through Asymmetry: A Meta-Analytical Approach*
18. Inaki Veruete Villegas, Milan Ščasný*: Input-Output Modeling Amidst Crisis: Tracing Natural Gas Pathways in the Czech Republic During the War-Induced Energy Turmoil*
19. Theodor Petřík: *Distribution Strategy Planning: A Comprehensive Probabilistic Approach for Unpredictable Environment*
20. Meri Papavangjeli, Adam Geršl*: Monetary Policy, Macro-Financial Vulnerabilities, and Macroeconomic Outcomes*

21. Attila Sarkany, Lukáš Janásek, Jozef Baruník: *Quantile Preferences in Portfolio Choice: A Q-DRL Approach to Dynamic Diversification*
22. Jiri Kukacka, Erik Zila: *Unraveling Timing Uncertainty of Event-driven Connectedness among Oil-Based Energy Commodities*
23. Samuel Fiifi Eshun, Evžen Kočenda: *Money Talks, Green Walks: Does Financial Inclusion Promote Green Sustainability in Africa?*
24. Mathieu Petit, Karel Janda: *The Optimal Investment Size in the Electricity Sector in EU Countries*
25. Alessandro Chiari: *Do Tax Havens Affect Financial Management? The Case of U.S. Multinational Companies*
26. Lenka Nechvátalová: *Autoencoder Asset Pricing Models and Economic Restrictions – International Evidence*
27. Markéta Malá: *Exploring Foreign Direct Investments and Engagements of Socialist Multinational Enterprises: A Case Study of Skoda Works in the 1970s and 1980s*
28. Veronika Plachá: *Does Childbirth Change the Gender Gap in Well-Being between Partners?*
29. Jan Žalman: *The Effect of Financial Transparency on Aid Diversion*
30. Aleksandra Jandrić, Adam Geršl: *Exploring Institutional Determinants of Private Equity and Venture Capital Activity in Europe*
31. Tomáš Boukal: *Where Do Multinationals Locate Profits: Evidence from Country-by-Country Reporting*
32. Karel Janda, Vendula Letovska, Jan Sila, David Zilberman: *Impact of Ethanol Blending Policies on U.S. Gasoline Prices*
33. Anton Grui: *Wartime Interest Rate Pass-Through in Ukraine: The Role of Prudential Indicators*
34. Jaromír Baxa, Tomáš Šestořád: *Economic Policy Uncertainty in Europe: Spillovers and Common Shocks*
35. Daniel Kolář: *Poverty in the Czech Republic: Unemployment, Pensions, and Regional Differences*
36. Tomáš Šestořád, Natálie Dvořáková: *Origins of Post-COVID-19 Inflation in Central European Countries*
37. Bathusi Gabanatlhong: *Stock Market Reaction to Increased Transparency: An Analysis of Country-By-Country Reporting in Developing Countries*
38. Lukas Petraseka, Jiri Kukacka: *US Equity Announcement Risk Premia*
39. Tomáš Boukal, Petr Janský, Miroslav Palanský: *Global Minimum Tax and Profit Shifting*
40. Karel Janda, Mathieu Petit: *Analyzing Decision-Making in Deep-Q Reinforcement Learning for Trading: A Case Study on Tesla Company and its Supply Chain*

All papers can be downloaded at: http://ies.fsv.cuni.cz

Univerzita Karlova v Praze, Fakulta sociálních věd
Institut ekonomických studií [UK FSV – IES] Praha 1, Opletalova 26
E-mail : ies@fsv.cuni.cz http://ies.fsv.cuni.cz