Journal of
Intelligent Financial
Management Perspectives

ISSN: 3116-0964

Designing an Optimized Algorithmic Trading Strategy for the Gold Futures Market Using Deep Reinforcement Learning

Pedram Salehi [1]*, Helia Mirzaei [2], Keyvan Ashrafi[3]

[1] *PhD in Financial Management, University of Shiraz, Shiraz, Iran (Coresponding autour), Email: p.salehi@shirazu.ac.ir*
[2] *PhD in Financial Management, University of Shiraz, Shiraz, Iran*
[3] *MSc in Financial Engineering, University of Shiraz, Shiraz, Iran*

ARTICLE INFO

Article history:
Received:04/01/2026
Received in revised form:15/01/2026
Accepted:09/02/2026
Available online:01/03/2026

Keywords:

Algorithmic Trading
Deep Reinforcement Learning
Gold Futures
Portfolio Optimization
DQN

ABSTRACT

In recent years, the expansion of financial markets and the increasing complexity of price behavior have made the use of intelligent methods in designing trading strategies a necessity. The gold futures market, as one of the highly volatile markets sensitive to macroeconomic variables, provides a suitable environment for developing and evaluating advanced algorithmic trading systems. In this study, a novel framework based on Deep Reinforcement Learning is proposed for designing an optimal algorithmic trading system in the gold futures market.

The proposed model combines deep neural networks with reinforcement learning algorithms such as Deep Q-Network (DQN) and Proximal Policy Optimization (PPO) to learn an optimal trading policy through continuous interaction with the market environment. In this framework, the market state includes variables such as price, trading volume, and technical indicators, while the reward function is designed to consider not only profitability but also risk in the decision-making process.

Preliminary simulation results show that the proposed model outperforms traditional technical analysis strategies such as Moving Average and MACD in terms of risk-adjusted return (Sharpe Ratio) and drawdown control. The findings indicate that deep reinforcement learning can serve as an effective tool for designing intelligent trading systems in complex and nonlinear financial markets.

Article Type: Research Paper

Journal of Intelligent Financial Management, 2025, Vol. 1, No.4, pp. 38- 55

Publish by:
Tolou-e Binish-e Ayandeh Scientific Institute

https://doi.org/ 10.25843/JIFM.2025.8563.22558

©Authors

Cite: Salehi,P , Mirzaei,H and Ashrafi,K . (2025). Designing an Optimized Algorithmic Trading Strategy for the Gold Futures Market Using Deep Reinforcement Learning. *Journal of Intelligent Financial Management, 1(4), 38-55*.

نگرش های
مدیریت مالی هوشمند
Journal of Intelligent Financial Management Perspectives
ISSN: 3116-0964

طراحی الگوریتم معاملات الگوریتمی بهینه در بازار آتی سکه با رویکرد یادگیری تقویتی عمیق

پدرام صالحی [۱و*]، هلیا میرزایی [۲]،کیوان اشرفی [۳]

۱ و *- دکتری مدیریت مالی، دانشگاه شیراز، شیراز، ایران (نویسنده مسئول)، ایمیل نویسنده مسئول:p.salehi@shirazu.ac.ir

۲- دکتری مدیریت مالی، دانشگاه شیراز، شیراز، ایران

۳- کارشناسی ارشد مهندسی مالی، دانشگاه شیراز، شیراز، ایران

اطلاعات مقاله

تاریخچه مقاله:

تاریخ دریافت:۱۴۰۴/۱۰/۱۴

تاریخ بازنگری:۱۴۰۴/۱۰/۲۵

تاریخ پذیرش:۱۴۰۴/۱۱/۲۰

تاریخ انتشار:۱۴۰۴/۱۲/۱۰

کلیدواژه‌ها:

معاملات الگوریتمی

یادگیری تقویتی عمیق

قراردادهای آتی طلا

بهینه‌سازی سبد سرمایه‌گذاری

شبکه کیو عمیق

چکیده

در سال‌های اخیر، گسترش بازارهای مالی و افزایش پیچیدگی رفتار قیمت‌ها، استفاده از روش‌های هوشمند در طراحی راهبردهای معاملاتی را به یک ضرورت تبدیل کرده است. بازار آتی سکه به‌عنوان یکی از بازارهای پرنوسان و حساس به متغیرهای کلان اقتصادی، محیطی مناسب برای توسعه و ارزیابی الگوریتم‌های معاملاتی پیشرفته محسوب می‌شود. در این پژوهش، یک چارچوب نوین مبتنی بر یادگیری تقویتی عمیق برای طراحی سامانه معاملات الگوریتمی بهینه در بازار آتی سکه ارائه می‌شود. مدل پیشنهادی با ترکیب شبکه‌های عصبی عمیق و الگوریتم‌های یادگیری تقویتی، اقدام به یادگیری سیاست معاملاتی بهینه از طریق تعامل مستمر با محیط بازار می‌کند.در این چارچوب، وضعیت بازار شامل متغیرهایی مانند قیمت، حجم معاملات و شاخص‌های فنی بوده و تابع پاداش به‌گونه‌ای طراحی شده است که علاوه بر بازدهی، ریسک نیز در فرایند تصمیم‌گیری لحاظ شود. نتایج اولیه حاصل از شبیه‌سازی نشان می‌دهد که مدل پیشنهادی نسبت به راهبردهای سنتی مبتنی بر تحلیل فنی، عملکرد بهتری از نظر بازده تعدیل‌شده بر اساس ریسک و کنترل افت سرمایه دارد. یافته‌های این پژوهش بیانگر آن است که یادگیری تقویتی عمیق می‌تواند ابزاری مؤثر برای طراحی سامانه‌های معاملاتی هوشمند در بازارهای مالی پیچیده و غیرخطی باشد.همچنین تحلیل حساسیت نتایج نشان می‌دهد که عملکرد مدل در شرایط مختلف بازار، از جمله روندهای صعودی، نزولی و خنثی، نسبتاً پایدار بوده و نوسانات عملکردی آن در مقایسه با روش‌های کلاسیک کمتر است. این موضوع بیانگر توانایی مدل در تطبیق با تغییر رژیم‌های بازار و یادگیری الگوهای رفتاری متغیر در طول زمان است. افزون بر این، لحاظ کردن هزینه‌های معاملاتی در فرایند شبیه‌سازی باعث شده است که نتایج به شرایط واقعی بازار نزدیک‌تر باشد و از بیش‌برآورد عملکرد جلوگیری شود. از منظر مدیریت ریسک نیز مدل پیشنهادی توانسته است با کنترل هم‌زمان بازده و افت سرمایه، تعادل مناسبی میان سودآوری و پایداری سرمایه ایجاد کند.در نهایت، این پژوهش نشان می‌دهد که ترکیب یادگیری تقویتی عمیق با داده‌های مالی می‌تواند مسیر مناسبی برای توسعه نسل جدید سامانه‌های هوشمند معاملات الگوریتمی فراهم سازد.

نشریه مدیریت مالی هوشمند، ۱۴۰۴، دوره ۱، شماره ۴، صفحه ۳۸-۵۵.

ناشر: موسسه علمی طلوع بینش آینده

https://doi.org/ 10.25843/JIFM.2025.8563. 22558

نوع مقاله: پژوهشی

© نویسندگان

استناد: صالحی,پدرام , میرزایی,هلیا و اشرفی,کیوان . (۱۴۰۴). طراحی الگوریتم معاملات الگوریتمی بهینه در بازار آتی سکه با رویکرد یادگیری تقویتی عمیق .*مدیریت مالی هوشمند* ، ۱(۴)، ۳۸-۵۵.

ISSN: 3116-0964

۱-مقدمه

در دهه‌های اخیر، ساختار بازارهای مالی در سطح جهانی دستخوش تغییرات بنیادین شده است. این تغییرات عمدتاً ناشی از پیشرفت‌های سریع در حوزه فناوری اطلاعات، افزایش قدرت پردازش داده‌ها، و ورود گسترده الگوریتم‌های هوشمند به فرآیندهای تصمیم‌گیری مالی بوده است. در چنین شرایطی، نقش انسان در تصمیم‌گیری‌های معاملاتی به‌تدریج کاهش یافته و جای خود را به سیستم‌های خودکار و الگوریتم‌های مبتنی بر داده داده است. این تحول ساختاری موجب شکل‌گیری حوزه‌ای تحت عنوان معاملات الگوریتمی شده است که امروزه یکی از مهم‌ترین شاخه‌های مهندسی مالی و علوم داده محسوب می‌شود(Aldridge, 2013) .معاملات الگوریتمی به فرآیند اجرای خودکار معاملات در بازارهای مالی اشاره دارد که در آن تصمیم‌های خرید و فروش بر اساس مدل‌های ریاضی، آماری و محاسباتی اتخاذ می‌شوند. این نوع معاملات معمولاً با هدف افزایش سرعت اجرا، کاهش خطای انسانی، حذف احساسات از تصمیم‌گیری و بهینه‌سازی بازدهی تعدیل‌شده بر اساس ریسک طراحی می‌شوند. در بازارهای مالی توسعه‌یافته مانند بازار سهام ایالات متحده، استفاده از این الگوریتم‌ها به بخش عمده‌ای از حجم معاملات روزانه تبدیل شده است(Chan, 2017) .

با وجود این پیشرفت‌ها، چالش اصلی در طراحی سیستم‌های معاملاتی هوشمند، پیچیدگی ذاتی رفتار بازارهای مالی است. داده‌های مالی معمولاً دارای ویژگی‌هایی همچون نویز بالا، غیرایستایی، وابستگی زمانی پیچیده و رفتارهای غیرخطی هستند. این ویژگی‌ها باعث می‌شود که مدل‌های کلاسیک اقتصادسنجی مانند ARIMA یاGARCH ، اگرچه در برخی شرایط مفید هستند، اما در بسیاری از موارد توانایی کافی برای مدل‌سازی دقیق رفتار بازار را نداشته باشند(Cont, 2001) .در این میان، بازارهای نوظهور مانند بازار آتی سکه در ایران، به دلیل ویژگی‌های خاص خود، پیچیدگی بیشتری نسبت به بازارهای توسعه‌یافته دارند. بازار آتی سکه تحت تأثیر مجموعه‌ای از عوامل کلان اقتصادی از جمله نرخ ارز، قیمت جهانی طلا، سیاست‌های پولی بانک مرکزی، انتظارات تورمی و حتی شوک‌های سیاسی قرار دارد. این عوامل باعث ایجاد نوسانات شدید و غیرقابل پیش‌بینی در قیمت‌ها می‌شوند که مدل‌سازی آن‌ها را با روش‌های سنتی بسیار دشوار می‌سازد.از منظر مالی، بازار آتی سکه یکی از ابزارهای مشتقه مهم برای پوشش ریسک و سفته‌بازی محسوب می‌شود. قراردادهای آتی به سرمایه‌گذاران این امکان را می‌دهند که دارایی پایه (در اینجا سکه طلا) را در آینده با قیمت مشخص معامله کنند. همین ویژگی باعث شده است که این بازار هم برای معامله‌گران حرفه‌ای و هم برای سرمایه‌گذاران ریسک‌پذیر جذاب باشد. با این حال، ماهیت اهرمی این بازار باعث افزایش ریسک نیز می‌شود و مدیریت ریسک در آن اهمیت ویژه‌ای دارد.

در ادبیات مالی کلاسیک، تلاش‌های زیادی برای پیش‌بینی قیمت دارایی‌ها صورت گرفته است. مدل‌های سری زمانی خطی، روش‌های اقتصادسنجی و تحلیل تکنیکال از جمله ابزارهای رایج در این حوزه بوده‌اند. تحلیل تکنیکال بر این فرض استوار است که الگوهای تاریخی قیمت می‌توانند در پیش‌بینی رفتار آینده بازار مفید باشند. ابزارهایی مانند میانگین متحرک، شاخص قدرت نسبی (RSI) و MACD از جمله روش‌های پرکاربرد در این حوزه هستند. با این حال، این روش‌ها اغلب فاقد توانایی لازم برای درک روابط پیچیده و غیرخطی موجود در داده‌های مالی هستند.با ظهور یادگیری ماشین، رویکردهای جدیدی برای تحلیل بازارهای مالی معرفی شد. الگوریتم‌هایی مانند ماشین بردار پشتیبان ، جنگل تصادفی و شبکه‌های عصبی مصنوعی توانستند عملکرد بهتری نسبت به مدل‌های سنتی ارائه دهند. این مدل‌ها قادرند الگوهای غیرخطی را از داده‌ها استخراج کنند و در بسیاری از موارد دقت پیش‌بینی را افزایش دهند. با این وجود، این روش‌ها عمدتاً بر مسئله پیش‌بینی قیمت تمرکز دارند و نه بر تصمیم‌گیری مستقیم معاملاتی(Sezer, Gudelek & Ozbayoglu, 2019) .

در واقع، مسئله اصلی در معاملات مالی نه صرفاً پیش‌بینی قیمت آینده، بلکه انتخاب بهترین اقدام در هر لحظه زمانی است. معامله‌گر باید تصمیم بگیرد که در هر وضعیت بازار، اقدام به خرید، فروش یا نگهداری دارایی کند. این مسئله ذاتاً یک مسئله تصمیم‌گیری ترتیبی تحت عدم قطعیت است که در چارچوب فرآیندهای تصمیم‌گیری مارکوف قابل مدل‌سازی است.در چنین چارچوبی، یادگیری تقویتی به‌عنوان یکی از شاخه‌های مهم هوش مصنوعی، ابزار مناسبی برای حل این نوع مسائل ارائه می‌دهد. در یادگیری تقویتی، یک عامل هوشمند با محیط تعامل می‌کند، در هر وضعیت یک اقدام انجام می‌دهد و در ازای آن پاداش یا جریمه دریافت می‌کند. هدف عامل، یادگیری سیاستی است که مجموع پاداش‌های بلندمدت را بیشینه کند(Sutton & Barto, 2018) .

مزیت مهم یادگیری تقویتی نسبت به روش‌های پیش‌بینی محور این است که به جای تمرکز بر تخمین مقدار آینده، مستقیماً بر بهینه‌سازی تصمیم تمرکز دارد. این ویژگی باعث می‌شود که این روش برای کاربردهای مالی، که در آن هدف نهایی کسب سود و مدیریت ریسک است، بسیار مناسب باشد.

با این حال، یادگیری تقویتی کلاسیک نیز محدودیت‌هایی دارد. یکی از مهم‌ترین محدودیت‌ها، مشکل «فضای حالت بزرگ» است. در بازارهای مالی، تعداد متغیرهای تأثیرگذار بسیار زیاد است و مدل‌های ساده RL قادر به مدیریت این پیچیدگی نیستند. این مسئله با ظهور یادگیری تقویتی عمیق تا حد زیادی حل شده است.یادگیری تقویتی عمیق با ترکیب شبکه‌های عصبی عمیق و الگوریتم‌های تقویتی، امکان یادگیری از داده‌های پیچیده و با ابعاد بالا را فراهم می‌کند. الگوریتم‌هایی مانند Deep Q-Network (DQN) توانستند نشان دهند که می‌توانند مستقیماً از داده‌های خام، سیاست‌های تصمیم‌گیری بهینه استخراج کنند. (Mnih et al., 2015) همچنین الگوریتم Proximal Policy Optimization (PPO) به‌عنوان یکی از پایدارترین روش‌هایpolicy gradient ، در بسیاری از مسائل پیچیده عملکرد بسیار خوبی از خود نشان داده است (Schulman et al., 2017).

در حوزه مالی نیز استفاده از یادگیری تقویتی عمیق در سال‌های اخیر رشد قابل توجهی داشته است. پژوهش‌ها نشان داده‌اند که این مدل‌ها قادرند روابط پیچیده بین متغیرهای مالی را شناسایی کرده و استراتژی‌های معاملاتی سودآور تولید کنند. به عنوان مثال، استفاده از DQN در طراحی سیستم‌های معاملاتی نشان داده است که این روش می‌تواند عملکرد بهتری نسبت به مدل‌های سنتی تحلیل تکنیکال داشته باشد (Deng et al., 2016).

با وجود این پیشرفت‌ها، چالش‌های متعددی در به‌کارگیری یادگیری تقویتی عمیق در بازارهای مالی وجود دارد. یکی از مهم‌ترین چالش‌ها، طراحی تابع پاداش مناسب است. در بسیاری از موارد، استفاده صرف از سود به‌عنوان پاداش می‌تواند منجر به رفتارهای پرریسک و غیرواقعی شود. بنابراین، لازم است معیارهایی مانند ریسک، افت سرمایه و نوسانات نیز در تابع پاداش لحاظ شوند.چالش دیگر، غیرایستا بودن بازارهای مالی است. برخلاف محیط‌های کنترل‌شده در علوم مهندسی، بازارهای مالی دائماً در حال تغییر هستند و الگوهای گذشته لزوماً در آینده تکرار نمی‌شوند. این ویژگی باعث می‌شود که مدل‌های یادگیری ماشین در معرض خطر بیش‌برازش قرار گیرند و عملکرد آن‌ها در دوره‌های زمانی مختلف ناپایدار باشد.در بازار آتی سکه ایران، این چالش‌ها حتی شدیدتر هستند. این بازار نه تنها تحت تأثیر عوامل جهانی مانند قیمت طلا قرار دارد، بلکه به شدت از شرایط اقتصادی داخلی نیز تأثیر می‌پذیرد. تغییرات نرخ ارز، سیاست‌های پولی و انتظارات تورمی نقش مهمی در تعیین قیمت‌ها دارند. بنابراین، طراحی یک سیستم معاملاتی هوشمند برای این بازار نیازمند مدلی است که بتواند با عدم قطعیت بالا و تغییرات سریع سازگار شود.

بررسی ادبیات پژوهشی نشان می‌دهد که بخش عمده‌ای از تحقیقات انجام‌شده در حوزه معاملات الگوریتمی بر بازارهای توسعه‌یافته مانند بازار سهام آمریکا، بازار فارکس و ارزهای دیجیتال متمرکز بوده است. در مقابل، بازارهای نوظهور مانند ایران کمتر مورد توجه قرار گرفته‌اند. این موضوع یک شکاف پژوهشی مهم ایجاد کرده است که نیازمند بررسی‌های بیشتر است.بر این اساس، هدف اصلی این پژوهش طراحی و توسعه یک چارچوب معاملاتی مبتنی بر یادگیری تقویتی عمیق برای بازار آتی سکه است. این چارچوب تلاش می‌کند با ترکیب الگوریتم‌های DQN و PPOو استفاده از داده‌های واقعی بازار، یک سیاست معاملاتی بهینه ارائه دهد که بتواند همزمان بازدهی بالا و ریسک کنترل‌شده ایجاد کند.

در نهایت، سؤال اساسی این پژوهش این است که آیا یادگیری تقویتی عمیق می‌تواند نسبت به روش‌های سنتی و مدل‌های یادگیری ماشین کلاسیک، عملکرد بهتری در بازار آتی سکه ارائه دهد یا خیر. پاسخ به این سؤال می‌تواند نقش مهمی در توسعه سیستم‌های هوشمند معاملاتی در بازارهای مالی نوظهور داشته باشد و مسیر جدیدی برای تحقیقات آینده در حوزه مهندسی مالی و هوش مصنوعی باز کند.

۲-مبانی نظری

۲-۱ رویکردهای کلاسیک در پیش‌بینی بازارهای مالی

مطالعات اولیه در حوزه تحلیل و پیش‌بینی بازارهای مالی عمدتاً بر پایه مدل‌های اقتصادسنجی و روش‌های آماری کلاسیک شکل گرفته‌اند. این رویکردها بر این فرض بنیادین استوار هستند که رفتار سری‌های زمانی مالی را می‌توان از طریق روابط خطی یا نیمه‌خطی میان متغیرهای گذشته

و آینده مدل‌سازی کرد. در این میان، مدل‌هایی نظیر ARIMA و GARCH از مهم‌ترین ابزارهای مورد استفاده در تحلیل داده‌های مالی بوده‌اند که به‌طور گسترده در پیش‌بینی قیمت دارایی‌ها، نرخ بازده و نوسانات بازار به کار رفته‌اند.

مدل ARIMA به‌عنوان یکی از رایج‌ترین روش‌های سری زمانی، بر اساس ترکیب اجزای خودرگرسیو، میانگین متحرک و تفاضل‌گیری طراحی شده است و تلاش می‌کند الگوهای خطی موجود در داده‌های تاریخی را استخراج کرده و برای پیش‌بینی مقادیر آینده به کار گیرد. در ادامه، مدل‌های خانواده GARCH با تمرکز بر مدل‌سازی واریانس شرطی، امکان تحلیل نوسانات خوشه‌ای در بازارهای مالی را فراهم کردند. این ویژگی به‌خصوص در بازارهایی که دارای دوره‌های آرام و پرنوسان متوالی هستند، اهمیت زیادی پیدا کرد.

با این حال، یکی از مفروضات اساسی این مدل‌ها، ایستایی نسبی ساختار داده‌ها در طول زمان است؛ به این معنا که فرض می‌شود روابط آماری گذشته می‌تواند در آینده نیز تا حد زیادی معتبر باقی بماند. این در حالی است که شواهد تجربی در بازارهای مالی واقعی، به‌ویژه در بازارهای نوظهور و کم‌عمق، نشان می‌دهد که رفتار قیمت‌ها اغلب تحت تأثیر عوامل غیرخطی، شوک‌های بیرونی، تغییرات ساختاری و رفتارهای هیجانی سرمایه‌گذاران قرار دارد. همین مسئله موجب می‌شود که کارایی مدل‌های کلاسیک در شرایط واقعی بازار محدود شود و در بسیاری از دوره‌ها نتوانند پیش‌بینی‌های پایداری ارائه دهند(Cont, 2001) .

علاوه بر این، مدل‌های اقتصادسنجی کلاسیک معمولاً در مواجهه با داده‌های با نویز بالا و روابط پیچیده میان متغیرهای مالی دچار افت عملکرد می‌شوند. بازارهای مالی واقعی نه‌تنها غیرایستا هستند، بلکه دارای ویژگی‌هایی مانند وابستگی‌های بلندمدت، جهش‌های ناگهانی قیمت و تغییر رژیم‌های رفتاری هستند که در چارچوب مدل‌های خطی به‌سختی قابل توضیح‌اند. از این رو، اگرچه این مدل‌ها از نظر نظری پایه‌گذار بسیاری از روش‌های تحلیل مالی مدرن محسوب می‌شوند، اما از منظر کاربردی در محیط‌های پیچیده مالی محدودیت‌های قابل توجهی دارند.

در کنار رویکردهای اقتصادسنجی، تحلیل تکنیکال نیز به‌عنوان یکی از روش‌های پرکاربرد در پیش‌بینی قیمت دارایی‌ها مطرح شده است. این رویکرد به جای تمرکز بر متغیرهای بنیادی، بر الگوهای قیمتی و حجم معاملات در گذشته تکیه دارد و تلاش می‌کند از طریق شناسایی الگوهای تکرارشونده، رفتار آینده بازار را پیش‌بینی کند. ابزارهایی مانند میانگین‌های متحرک، شاخص قدرت نسبی(RSI) ، مکدی (MACD) و سایر اسیلاتورها از جمله مهم‌ترین ابزارهای تحلیل تکنیکال هستند که در میان معامله‌گران فردی و نهادی به‌طور گسترده مورد استفاده قرار می‌گیرند.

با وجود محبوبیت بالای تحلیل تکنیکال، مطالعات تجربی متعدد نشان داده‌اند که این روش‌ها در بسیاری از موارد قادر به ایجاد بازدهی پایدار و قابل اتکا نیستند. یکی از دلایل اصلی این مسئله، وابستگی این روش‌ها به قواعد ثابت و از پیش تعریف‌شده است. در واقع، تحلیل تکنیکال فاقد یک چارچوب یادگیری تطبیقی است که بتواند خود را با تغییر شرایط بازار سازگار کند. به عبارت دیگر، این روش‌ها در ذات خود ایستا هستند و در مواجهه با تغییر رژیم‌های بازار، کارایی آن‌ها به‌طور قابل توجهی کاهش می‌یابد.

علاوه بر این، تحلیل تکنیکال معمولاً بر فرض تکرارپذیری الگوهای تاریخی استوار است، در حالی که در بازارهای مالی واقعی، رفتار سرمایه‌گذاران و ساختار بازار به‌طور مداوم در حال تغییر است. این موضوع باعث می‌شود که سیگنال‌های تولید شده توسط این روش‌ها در برخی دوره‌ها کارایی داشته باشند، اما در بسیاری از شرایط دیگر منجر به سیگنال‌های نادرست یا تأخیری شوند. به همین دلیل، علی‌رغم استفاده گسترده از این ابزارها در عمل، شواهد علمی قوی و پایداری در حمایت از کارایی بلندمدت آن‌ها وجود ندارد(Murphy, 1999) .

در مجموع، می‌توان بیان کرد که رویکردهای کلاسیک در پیش‌بینی بازارهای مالی، شامل مدل‌های اقتصادسنجی و تحلیل تکنیکال، اگرچه نقش مهمی در توسعه اولیه علم مالی و تحلیل داده‌های بازار داشته‌اند، اما به دلیل اتکا به فرضیات ساده‌سازی‌شده و عدم توانایی در مدل‌سازی پیچیدگی‌های واقعی بازار، در بسیاری از کاربردهای عملی با محدودیت مواجه هستند. همین محدودیت‌ها زمینه‌ساز حرکت پژوهش‌های جدید به سمت روش‌های پیشرفته‌تر مانند یادگیری ماشین و به‌ویژه یادگیری تقویتی شده است که در بخش‌های بعدی مورد بررسی قرار می‌گیرند.

۲-۲ ورود یادگیری ماشین به بازارهای مالی

با گسترش علوم داده و افزایش توان پردازشی سیستم‌های محاسباتی، استفاده از الگوریتم‌های یادگیری ماشین در تحلیل و پیش‌بینی بازارهای مالی به‌طور گسترده‌ای توسعه یافته است. برخلاف رویکردهای کلاسیک که عمدتاً مبتنی بر فرض‌های خطی بودن و ایستایی داده‌ها هستند، روش‌های یادگیری ماشین امکان مدل‌سازی روابط پیچیده، غیرخطی و پنهان میان متغیرهای مالی را فراهم کرده‌اند. همین ویژگی باعث شده است که این دسته از مدل‌ها در حوزه‌های مختلفی مانند پیش‌بینی بازده دارایی‌ها، تشخیص روند بازار و مدیریت ریسک مورد توجه جدی پژوهشگران قرار گیرند(Hastie, Tibshirani & Friedman, 2009; Murphy, 2012) .

در میان الگوریتم‌های مختلف، روش‌هایی مانند ماشین بردار پشتیبان، جنگل تصادفی و روش‌های مبتنی بر تقویت گرادیانی از پرکاربردترین ابزارها در ادبیات مالی محسوب می‌شوند. این مدل‌ها با استفاده از داده‌های تاریخی قادرند الگوهای پیچیده و غیرخطی را استخراج کرده و در بسیاری از موارد نسبت به مدل‌های اقتصادسنجی سنتی دقت پیش‌بینی بالاتری ارائه دهند (Krauss, Do & Huck, 2017; Fischer & Krauss, 2018). همچنین برخی مطالعات نشان داده‌اند که ترکیب ویژگی‌های مالی و تکنیک‌های یادگیری ماشین می‌تواند منجر به بهبود معنادار در پیش‌بینی جهت بازار شود(Gu, Kelly & Xiu, 2020) .

در بازارهای مالی داخلی نیز نتایج مشابهی گزارش شده است. برای مثال، پژوهش‌های انجام‌شده در بورس تهران نشان می‌دهد که مدل‌های مبتنی بر درخت تصمیم و روش‌های ترکیبی در مقایسه با مدل‌های آماری کلاسیک، توانسته‌اند دقت پیش‌بینی بازده سهام را بهبود دهند و در شرایط غیرخطی بازار عملکرد بهتری داشته باشند (کریمی و محمدی، ۱۳۹۹؛ رضایی و شریفی، ۱۴۰۰). همچنین برخی مطالعات داخلی تأکید کرده‌اند که استفاده از یادگیری ماشین در بازارهای کم‌عمق و پرنوسان می‌تواند به کاهش خطای پیش‌بینی و افزایش قابلیت تطبیق مدل کمک کند (احمدی و حسینی، ۱۳۹۸).با وجود این پیشرفت‌ها، یکی از محدودیت‌های اساسی روش‌های یادگیری ماشین در حوزه مالی این است که این مدل‌ها عمدتاً به‌صورت ایستا عمل می‌کنند و فاقد تعامل مستقیم با محیط بازار هستند. به عبارت دیگر، این الگوریتم‌ها معمولاً بر اساس داده‌های تاریخی آموزش داده می‌شوند و پس از آموزش تنها نقش پیش‌بینی یا طبقه‌بندی را ایفا می‌کنند، بدون آنکه بتوانند در طول زمان و با تغییر شرایط بازار، سیاست تصمیم‌گیری خود را به‌صورت پویا اصلاح کنند(Hastie et al., 2009; Murphy, 2012) .

این ویژگی موجب می‌شود که خروجی این مدل‌ها بیشتر به‌عنوان سیگنال‌های تحلیلی تلقی شود تا یک سیستم تصمیم‌گیر معاملاتی کامل. در واقع، فاصله میان «پیش‌بینی قیمت» و «تصمیم‌گیری معاملاتی بهینه» یکی از چالش‌های اساسی این رویکردها محسوب می‌شود. حتی در مواردی که دقت پیش‌بینی بالا است، تبدیل این پیش‌بینی‌ها به استراتژی سودآور نیازمند طراحی قواعد اضافی مدیریت سرمایه و کنترل ریسک است که خارج از ساختار اصلی مدل قرار دارد(Fischer & Krauss, 2018) .

از سوی دیگر، بازارهای مالی به‌طور مداوم در حال تغییر رژیم هستند و تحت تأثیر عوامل اقتصادی، سیاسی و رفتاری قرار دارند. این مسئله باعث می‌شود که مدل‌های ایستا در برخی دوره‌ها عملکرد مناسبی داشته باشند، اما در دوره‌های دیگر با افت شدید دقت مواجه شوند. به همین دلیل، نبود سازوکار یادگیری تطبیقی در بسیاری از مدل‌های یادگیری ماشین، یکی از نقاط ضعف مهم آن‌ها در کاربردهای مالی محسوب می‌شود (Cont, 2001; Gu et al., 2020).در مجموع، می‌توان بیان کرد که اگرچه ورود یادگیری ماشین به حوزه مالی موجب ارتقای قابل توجه دقت پیش‌بینی و توسعه روش‌های تحلیل داده شده است، اما ماهیت غیرتعاملی و ایستای این مدل‌ها همچنان محدودیت‌هایی در کاربرد عملی آن‌ها ایجاد می‌کند. همین محدودیت‌ها زمینه‌ساز توسعه رویکردهای پیشرفته‌تری مانند یادگیری عمیق و به‌ویژه یادگیری تقویتی شده‌اند که در آن‌ها تمرکز از صرف پیش‌بینی به سمت تصمیم‌گیری پویا و تعامل مستقیم با محیط بازار منتقل شده است.

۲-۳ یادگیری عمیق در تحلیل سری‌های زمانی مالی

در سال‌های اخیر، با پیشرفت زیرشاخه‌های یادگیری ماشین، رویکردهای مبتنی بر یادگیری عمیق به‌عنوان یکی از قدرتمندترین ابزارها در تحلیل داده‌های پیچیده و حجیم مطرح شده‌اند. این رویکردها به دلیل توانایی بالا در استخراج ویژگی‌های غیرخطی از داده‌های خام، به‌ویژه در حوزه تحلیل سری‌های زمانی مالی، توجه گسترده‌ای را به خود جلب کرده‌اند. در واقع، برخلاف مدل‌های سنتی که نیازمند طراحی دستی ویژگی‌ها هستند، شبکه‌های عصبی عمیق قادرند به‌صورت خودکار الگوهای پنهان در داده‌ها را استخراج کرده و ساختارهای پیچیده موجود در بازارهای مالی را مدل‌سازی کنند(Goodfellow, Bengio & Courville, 2016; Hastie et al., 2009) .در میان معماری‌های مختلف یادگیری عمیق، شبکه‌های عصبی بازگشتی و به‌ویژه مدل حافظه بلندمدت–کوتاه‌مدت (LSTM) نقش بسیار مهمی در تحلیل داده‌های سری زمانی ایفا کرده‌اند. این مدل‌ها به‌طور خاص برای حل مشکل وابستگی‌های زمانی طراحی شده‌اند و قادرند اطلاعات گذشته را در حافظه خود نگه‌داری کرده و از آن‌ها برای پیش‌بینی مقادیر آینده استفاده کنند. این ویژگی باعث شده است که LSTM در بسیاری از کاربردهای مالی مانند پیش‌بینی قیمت سهام، نرخ ارز و تحلیل روند بازار مورد استفاده قرار گیرد(Fischer & Krauss, 2018; Zhang, Aggarwal & Qi, 2020) .

یکی از مزیت‌های اصلی این مدل‌ها نسبت به روش‌های کلاسیک و حتی برخی الگوریتم‌های یادگیری ماشین، توانایی آن‌ها در شناسایی وابستگی‌های بلندمدت در داده‌ها است. در بازارهای مالی، رفتار قیمت‌ها معمولاً تحت تأثیر الگوهایی قرار دارد که در بازه‌های زمانی طولانی شکل می‌گیرند و مدل‌های ساده‌تر قادر به شناسایی این روابط نیستند. در مقابل، شبکه‌های LSTM می‌توانند این وابستگی‌ها را در ساختار خود

ذخیره کرده و در فرآیند پیش‌بینی مورد استفاده قرار دهند، که این امر منجر به افزایش دقت مدل در بسیاری از مطالعات شده است (Fischer & Krauss, 2018; Sezer, Gudelek & Ozbayoglu, 2020).

در ادبیات پژوهشی داخلی نیز نتایج مشابهی گزارش شده است. برخی مطالعات نشان داده‌اند که استفاده از شبکه‌های عصبی عمیق در پیش‌بینی قیمت دارایی‌ها در بازار سرمایه ایران می‌تواند دقت بالاتری نسبت به مدل‌های آماری و برخی روش‌های یادگیری ماشین سنتی ارائه دهد (کریمی و محمدی، ۱۳۹۹؛ رضایی و شریفی، ۱۴۰۰). همچنین نتایج برخی پژوهش‌ها نشان می‌دهد که ترکیب داده‌های تکنیکال با ساختارهای یادگیری عمیق می‌تواند به بهبود عملکرد پیش‌بینی در بازارهای پرنوسان کمک کند (احمدی و حسینی، ۱۳۹۸).با وجود این پیشرفت‌ها، مدل‌های یادگیری عمیق نیز با محدودیت‌های مهمی مواجه هستند. مهم‌ترین محدودیت این مدل‌ها در کاربردهای مالی این است که تمرکز اصلی آن‌ها بر پیش‌بینی مقدار آینده متغیرها است و نه تصمیم‌گیری بهینه. به عبارت دیگر، این مدل‌ها معمولاً خروجی‌هایی به شکل پیش‌بینی قیمت یا جهت بازار تولید می‌کنند، اما تبدیل این پیش‌بینی‌ها به یک استراتژی معاملاتی سودآور نیازمند طراحی جداگانه سیستم تصمیم‌گیری و مدیریت ریسک است(Goodfellow et al., 2016) .

از سوی دیگر، این مدل‌ها برای دستیابی به عملکرد مناسب نیازمند حجم بالایی از داده و فرآیند آموزش پیچیده هستند. علاوه بر این، در شرایطی که ساختار بازار تغییر می‌کند یا رژیم‌های جدید رفتاری شکل می‌گیرد، مدل‌های یادگیری عمیق ممکن است با افت عملکرد مواجه شوند، مگر آنکه به‌صورت مداوم بازآموزی شوند. این مسئله در بازارهای مالی غیرایستا مانند بازار آتی سکه اهمیت بیشتری پیدا می‌کند، زیرا تغییرات ناگهانی در شرایط اقتصادی یا سیاسی می‌تواند الگوهای تاریخی را بی‌اعتبار کند(Zhang et al., 2020; Cont, 2001) .

بنابراین، اگرچه یادگیری عمیق توانسته است نسبت به روش‌های کلاسیک و حتی برخی مدل‌های یادگیری ماشین بهبود قابل توجهی در دقت پیش‌بینی ایجاد کند، اما همچنان محدودیت اساسی آن در تمرکز بر پیش‌بینی و نه تصمیم‌گیری مستقیم باقی مانده است. همین مسئله زمینه‌ساز توسعه رویکردهای پیشرفته‌تری مانند یادگیری تقویتی عمیق شده است که در آن هدف نه تنها پیش‌بینی، بلکه یادگیری سیاست بهینه تصمیم‌گیری در تعامل با محیط بازار است.

٤-٢ یادگیری تقویتی در بازارهای مالی

یادگیری تقویتی به‌عنوان یکی از شاخه‌های مهم هوش مصنوعی، چارچوبی برای حل مسائل تصمیم‌گیری ترتیبی ارائه می‌دهد که در آن یک عامل هوشمند از طریق تعامل مداوم با محیط، رفتار خود را به‌گونه‌ای تنظیم می‌کند که بیشترین پاداش تجمعی را در طول زمان کسب کند. در این چارچوب، سیستم به‌جای یادگیری صرف از داده‌های ثابت، از تجربه‌های حاصل از تعامل با محیط یاد می‌گیرد و بر اساس بازخوردهای دریافتی، سیاست تصمیم‌گیری خود را به‌صورت تدریجی بهینه می‌سازد(Sutton & Barto, 2018) .این ویژگی موجب شده است که یادگیری تقویتی از منظر مفهومی، تطابق بسیار بالایی با مسائل مالی و به‌ویژه معاملات الگوریتمی داشته باشد. بازارهای مالی ذاتاً محیط‌هایی پویا، غیرایستا و مبتنی بر تصمیم‌گیری‌های متوالی هستند؛ به‌گونه‌ای که هر تصمیم معاملاتی نه‌تنها بر نتیجه لحظه‌ای، بلکه بر فرصت‌ها و ریسک‌های آینده نیز اثرگذار است. بنابراین، برخلاف مدل‌های صرفاً پیش‌بینی‌محور، یادگیری تقویتی می‌تواند فرآیند معامله را به‌عنوان یک مسئله بهینه‌سازی بلندمدت مدل‌سازی کند که در آن هدف، بیشینه‌سازی بازده تجمعی تعدیل‌شده بر اساس ریسک در طول زمان است (Moody & Saffell, 2001; Sutton & Barto, 2018).

در ادبیات اولیه این حوزه، الگوریتم‌های پایه‌ای مانند Q-learning به‌عنوان نخستین تلاش‌ها برای کاربرد یادگیری تقویتی در مسائل مالی مورد استفاده قرار گرفتند. این الگوریتم‌ها تلاش می‌کردند ارزش هر حالت–کنش را بر اساس تجربه‌های گذشته برآورد کرده و سیاست معاملاتی بهینه را استخراج کنند. نتایج اولیه نشان داد که این روش‌ها می‌توانند در طراحی استراتژی‌های ساده معاملاتی، به‌ویژه در محیط‌های با فضای حالت محدود، عملکرد قابل قبولی داشته باشند. (Kaelbling, Littman & Moore, 1996) با این حال، با افزایش پیچیدگی بازارهای مالی و بزرگ شدن فضای حالت و کنش، محدودیت‌های این روش‌ها به‌تدریج آشکار شد.یکی از مهم‌ترین چالش‌های الگوریتم‌های پایه یادگیری تقویتی، مشکل «بُعد بالا» در فضای حالت است. در بازارهای مالی واقعی، متغیرهای متعددی مانند قیمت، حجم، شاخص‌های تکنیکال، اخبار و عوامل کلان اقتصادی می‌توانند به‌عنوان ورودی مدل در نظر گرفته شوند. این تنوع و پیچیدگی باعث می‌شود که روش‌های سنتی یادگیری تقویتی نتوانند به‌صورت مؤثر کل فضای حالت را پوشش دهند و در نتیجه، عملکرد آن‌ها در محیط‌های واقعی محدود شود(Sutton & Barto, 2018) .

با این وجود، ترکیب یادگیری تقویتی با شبکه‌های عصبی عمیق منجر به شکل‌گیری رویکردهای پیشرفته‌تری شده است که تحت عنوان یادگیری تقویتی عمیق شناخته می‌شوند. این رویکردها با استفاده از شبکه‌های عصبی، امکان تقریب تابع ارزش یا سیاست را در فضاهای حالت بزرگ فراهم می‌کنند و به همین دلیل توانسته‌اند بسیاری از محدودیت‌های روش‌های کلاسیک را برطرف کنند. (Mnih et al., 2015) در این چارچوب، مدل قادر است به‌جای ذخیره‌سازی صریح مقادیر ارزش برای هر حالت، از یک تابع تقریب‌زننده استفاده کند که قابلیت تعمیم‌پذیری بالاتری دارد.

در حوزه مالی، این ویژگی اهمیت بسیار زیادی دارد، زیرا بازارها معمولاً دارای ابعاد بالا، نویز زیاد و ساختار غیرخطی هستند. یادگیری تقویتی عمیق این امکان را فراهم می‌کند که عامل معاملاتی بتواند از داده‌های پیچیده و چندبعدی برای یادگیری سیاست بهینه استفاده کند، بدون آنکه نیاز به تعریف صریح قواعد معاملاتی از پیش تعیین‌شده باشد. به همین دلیل، در سال‌های اخیر استفاده از این رویکرد در طراحی سیستم‌های معاملات الگوریتمی و مدیریت پرتفوی به‌طور قابل توجهی افزایش یافته است(Deng et al., 2016; Moody & Saffell, 2001) .

در مطالعات انجام‌شده نیز نشان داده شده است که یادگیری تقویتی می‌تواند در مقایسه با مدل‌های صرفاً پیش‌بینی‌محور، عملکرد بهتری در تولید استراتژی‌های معاملاتی پویا داشته باشد، زیرا به‌جای تمرکز بر پیش‌بینی قیمت، مستقیماً بر بهینه‌سازی تصمیم‌های معاملاتی تمرکز می‌کند. با این حال، موفقیت این رویکرد به عوامل متعددی مانند طراحی صحیح تابع پاداش، تعریف مناسب فضای حالت و در نظر گرفتن هزینه‌های معاملاتی وابسته است(Sutton & Barto, 2018; Mnih et al., 2015) .

در مجموع، می‌توان بیان کرد که یادگیری تقویتی به دلیل ماهیت تعاملی و تصمیم‌محور خود، یکی از مناسب‌ترین چارچوب‌ها برای مدل‌سازی مسائل مالی پویا محسوب می‌شود. با وجود محدودیت‌های اولیه در نسخه‌های ساده‌تر این الگوریتم‌ها، توسعه نسخه‌های عمیق آن توانسته است زمینه را برای کاربردهای پیشرفته‌تر در حوزه معاملات الگوریتمی فراهم کند و مسیر را برای طراحی سیستم‌های هوشمندتر و تطبیق‌پذیرتر در بازارهای مالی هموار سازد.

۲-۵یادگیری تقویتی عمیق

با پیشرفت همزمان دو حوزه یادگیری عمیق و یادگیری تقویتی، رویکردی ترکیبی تحت عنوان یادگیری تقویتی عمیق شکل گرفت که توانست بسیاری از محدودیت‌های هر دو حوزه را به‌صورت جداگانه کاهش دهد. در این چارچوب، شبکه‌های عصبی عمیق به‌عنوان تقریب‌زننده تابع ارزش یا سیاست مورد استفاده قرار می‌گیرند و به عامل اجازه می‌دهند در فضاهای حالت بسیار بزرگ و پیچیده، تصمیم‌گیری بهینه انجام دهد. این ویژگی باعث شده است که یادگیری تقویتی عمیق به یکی از قدرتمندترین ابزارها برای حل مسائل تصمیم‌گیری در محیط‌های غیرخطی و پویای امروزی تبدیل شود (Goodfellow, Bengio & Courville, 2016; Sutton & Barto, 2018).

یکی از نخستین و مهم‌ترین دستاوردهای این حوزه، الگوریتم شبکه Q عمیق است که توسط پژوهشگران شرکت دیپ‌مایند معرفی شد. این الگوریتم با ترکیب شبکه‌های عصبی کانولوشنی و چارچوب Q-learning توانست در محیط‌های پیچیده بازی‌های آتاری، بدون نیاز به مهندسی ویژگی دستی، عملکردی در سطح انسان یا حتی بالاتر ارائه دهد. (Mnih et al., 2015) اهمیت این دستاورد در آن بود که برای نخستین بار نشان داد یک عامل هوشمند می‌تواند تنها از طریق مشاهده داده‌های خام و دریافت پاداش، سیاست‌های تصمیم‌گیری پیچیده را یاد بگیرد.

پس از موفقیتDQN ، محدودیت‌های این الگوریتم مانند ناپایداری آموزش و همگرایی کند، زمینه‌ساز توسعه روش‌های پیشرفته‌تری در خانواده یادگیری تقویتی عمیق شد. یکی از مهم‌ترین این روش‌ها، الگوریتم‌های مبتنی بر گرادیان سیاست هستند که به‌جای یادگیری تابع ارزش، مستقیماً سیاست تصمیم‌گیری را بهینه می‌کنند. در این میان، الگوریتم بهینه‌سازی سیاست نزدیک (PPO) به دلیل سادگی نسبی، پایداری در آموزش و عملکرد مناسب در محیط‌های پیچیده، به یکی از پرکاربردترین روش‌ها در مسائل تصمیم‌گیری پیوسته تبدیل شده است (Schulman et al., 2017).

در مقایسه با روش‌های ارزش‌محور مانندDQN ، الگوریتم‌های سیاست‌محور معمولاً توانایی بیشتری در مواجهه با فضاهای کنش پیوسته و مسائل با ساختار پیچیده دارند. این ویژگی به‌ویژه در کاربردهای مالی اهمیت دارد، زیرا تصمیم‌گیری معاملاتی اغلب شامل انتخاب‌های پیوسته مانند تعیین حجم معامله یا تنظیم سطح ریسک است. علاوه بر این، نسخه‌های ترکیبی مانند Actor-Critic نیز توسعه یافته‌اند که تلاش می‌کنند مزایای هر دو رویکرد ارزش‌محور و سیاست‌محور را به‌صورت همزمان بهره‌برداری کنند و پایداری یادگیری را افزایش دهند (Sutton & Barto, 2018).

در حوزه مالی، یادگیری تقویتی عمیق به‌عنوان یکی از رویکردهای نوظهور و بسیار امیدوارکننده برای طراحی سیستم‌های معاملات الگوریتمی شناخته می‌شود. برخلاف مدل‌های سنتی و حتی بسیاری از روش‌های یادگیری ماشین که عمدتاً بر پیش‌بینی قیمت یا سیگنال‌های خرید و فروش تمرکز دارند، DRLمستقیماً فرآیند تصمیم‌گیری معاملاتی را هدف قرار می‌دهد. به عبارت دیگر، هدف این مدل‌ها نه صرفاً پیش‌بینی آینده بازار، بلکه یادگیری یک سیاست بهینه برای بیشینه‌سازی بازده تعدیل‌شده بر اساس ریسک در طول زمان است (Deng et al., 2016; Moody & Saffell, 2001).

یکی از مزیت‌های مهم این رویکرد در بازارهای مالی، توانایی آن در یادگیری رفتارهای تطبیقی در شرایط متغیر بازار است. بازارهای مالی معمولاً غیرایستا هستند و تحت تأثیر عوامل اقتصادی، سیاسی و رفتاری به‌طور مداوم تغییر می‌کنند. در چنین شرایطی، مدل‌های مبتنی بر یادگیری تقویتی عمیق می‌توانند با استفاده از بازخوردهای محیطی، سیاست معاملاتی خود را به‌روزرسانی کرده و با تغییر رژیم‌های بازار سازگار شوند. این ویژگی باعث شده است که DRL در سال‌های اخیر به‌طور گسترده در طراحی سیستم‌های معاملاتی خودکار، مدیریت پرتفوی و بهینه‌سازی استراتژی‌های سرمایه‌گذاری مورد توجه قرار گیرد(Fischer & Krauss, 2018; Zhang, Aggarwal & Qi, 2020) .

با این حال، علی‌رغم توانمندی‌های قابل توجه، کاربرد یادگیری تقویتی عمیق در حوزه مالی همچنان با چالش‌هایی همراه است. از جمله این چالش‌ها می‌توان به حساسیت بالا به طراحی تابع پاداش، نیاز به حجم بالای داده برای آموزش پایدار، و دشواری در تضمین تعمیم‌پذیری مدل در شرایط واقعی بازار اشاره کرد. علاوه بر این، هزینه‌های معاملاتی و محدودیت‌های اجرایی در بازار واقعی نیز می‌توانند عملکرد مدل را تحت تأثیر قرار دهند. بنابراین، طراحی یک سیستم DRL کارآمد در حوزه مالی نیازمند توجه دقیق به جنبه‌های مدل‌سازی محیط، تعریف مناسب فضای حالت و کنش، و همچنین تنظیم دقیق تابع پاداش است(Sutton & Barto, 2018; Mnih et al., 2015) .

در مجموع، یادگیری تقویتی عمیق را می‌توان نقطه اتصال میان تحلیل داده‌های مالی و تصمیم‌گیری هوشمند دانست که توانسته است مسیر جدیدی را در توسعه سیستم‌های معاملاتی الگوریتمی ایجاد کند. این رویکرد با ترکیب قدرت استخراج ویژگی‌های پیچیده از داده‌ها و توانایی یادگیری سیاست‌های تصمیم‌گیری، زمینه را برای طراحی مدل‌های پیشرفته‌تر در بازارهای مالی فراهم کرده است و به‌عنوان یکی از اصلی‌ترین پایه‌های پژوهش حاضر مورد استفاده قرار می‌گیرد.

٢-٦کاربرد یادگیری تقویتی عمیق در معاملات الگوریتمی

در سال‌های اخیر، با گسترش کاربردهای هوش مصنوعی در بازارهای مالی، یادگیری تقویتی عمیق به‌عنوان یکی از رویکردهای پیشرو در طراحی سیستم‌های معاملات الگوریتمی مورد توجه گسترده قرار گرفته است. دلیل اصلی این توجه، توانایی این روش در مدل‌سازی فرآیند تصمیم‌گیری پویا و ترتیبی در محیط‌های پیچیده و غیرایستا مانند بازارهای مالی است. برخلاف روش‌های سنتی که عمدتاً بر پیش‌بینی قیمت یا استخراج سیگنال‌های آماری تمرکز دارند، یادگیری تقویتی عمیق مستقیماً مسئله انتخاب اقدام بهینه در هر لحظه از زمان را هدف قرار می‌دهد و بدین ترتیب، ساختار مسئله معاملات را به‌صورت یک فرآیند بهینه‌سازی بلندمدت مدل‌سازی می‌کند (Sutton & Barto, 2018; Moody & Saffell, 2001).

مطالعات اخیر نشان داده‌اند که استفاده از این رویکرد می‌تواند در طراحی استراتژی‌های معاملاتی پویا و سازگار با شرایط متغیر بازار بسیار مؤثر باشد. برای مثال، در برخی پژوهش‌ها نشان داده شده است که الگوریتم‌های یادگیری تقویتی عمیق قادرند با یادگیری مستقیم از داده‌های بازار، سیاست‌هایی تولید کنند که منجر به بهبود بازده و کاهش ریسک نسبت به روش‌های سنتی شوند (Deng et al., 2016). این نتایج بیانگر آن است که DRL می‌تواند به‌عنوان یک چارچوب مناسب برای یادگیری رفتار معاملاتی بهینه در محیط‌های مالی مورد استفاده قرار گیرد.

یکی از مهم‌ترین مزیت‌های کاربرد یادگیری تقویتی عمیق در معاملات الگوریتمی، توانایی آن در ترکیب تحلیل ویژگی‌های پیچیده بازار با فرآیند تصمیم‌گیری است. در این رویکرد، شبکه‌های عصبی عمیق وظیفه استخراج ویژگی‌های پنهان از داده‌های خام بازار مانند قیمت، حجم معاملات و شاخص‌های تکنیکال را بر عهده دارند. سپس عامل یادگیرنده بر اساس این ویژگی‌ها و با توجه به بازخوردهای محیط، اقدام به تصمیم‌گیری در خصوص خرید، فروش یا نگهداری دارایی می‌کند. این ساختار باعث می‌شود که فرآیند تصمیم‌گیری به‌صورت یکپارچه و داده‌محور انجام شود، بدون آنکه نیاز به طراحی دستی قواعد معاملاتی پیچیده وجود داشته باشد (Fischer & Krauss, 2018) .

در برخی مطالعات پیشرفته‌تر نیز نشان داده شده است که ترکیب ویژگی‌های تکنیکال با چارچوب یادگیری تقویتی عمیق می‌تواند عملکرد سیستم‌های معاملاتی را به‌طور قابل توجهی بهبود دهد. در این حالت، شاخص‌های تکنیکال به‌عنوان ورودی‌های مکمل در کنار داده‌های خام

قیمت و حجم مورد استفاده قرار می‌گیرند و شبکه عصبی از طریق یادگیری عمیق قادر به استخراج روابط غیرخطی میان این متغیرها می‌شود. این موضوع موجب افزایش دقت تصمیم‌گیری و بهبود عملکرد استراتژی‌های معاملاتی در شرایط مختلف بازار شده است (Gu, Kelly & Xiu, 2020; Fischer & Krauss, 2018).

با این حال، علی‌رغم نتایج امیدوارکننده، استفاده از یادگیری تقویتی عمیق در معاملات الگوریتمی همچنان با چالش‌هایی مواجه است. از جمله مهم‌ترین این چالش‌ها می‌توان به حساسیت بالای مدل نسبت به طراحی تابع پاداش، نیاز به تنظیم دقیق پارامترها، و پیچیدگی بالای فرآیند آموزش اشاره کرد. همچنین در محیط‌های واقعی بازار، عواملی مانند هزینه‌های معاملاتی، محدودیت نقدشوندگی و لغزش قیمت می‌توانند به‌طور قابل توجهی بر عملکرد مدل تأثیرگذار باشند و در صورت عدم لحاظ مناسب در طراحی محیط، منجر به نتایج غیرواقعی شوند (Sutton & Barto, 2018; Moody & Saffell, 2001).

در مجموع، نتایج مطالعات انجام‌شده نشان می‌دهد که یادگیری تقویتی عمیق یکی از رویکردهای بسیار توانمند در توسعه سیستم‌های معاملات الگوریتمی هوشمند محسوب می‌شود. این روش با ترکیب قابلیت استخراج ویژگی‌های پیچیده از داده‌ها و توانایی یادگیری سیاست‌های تصمیم‌گیری بهینه، امکان طراحی استراتژی‌های معاملاتی تطبیق‌پذیر و پویا را فراهم می‌کند و به همین دلیل، به‌عنوان یکی از پایه‌های اصلی پژوهش‌های نوین در حوزه مالی الگوریتمی مورد استفاده قرار می‌گیرد.

۲-۷ چالش‌های موجود در یادگیری تقویتی عمیق در بازارهای مالی

با وجود موفقیت‌ها و پیشرفت‌های قابل توجه یادگیری تقویتی عمیق در حوزه معاملات الگوریتمی، کاربرد عملی این رویکرد در بازارهای مالی همچنان با مجموعه‌ای از چالش‌های اساسی مواجه است که می‌تواند عملکرد آن را در شرایط واقعی محدود کند. این چالش‌ها عمدتاً ناشی از ماهیت پیچیده، غیرایستا و نویزی بازارهای مالی هستند و در بازارهای نوظهور، شدت بیشتری نیز پیدا می‌کنند.

یکی از مهم‌ترین چالش‌ها، غیرایستا بودن ساختار بازارهای مالی است. در بسیاری از مدل‌های یادگیری تقویتی، فرض می‌شود که توزیع داده‌ها و قوانین حاکم بر محیط در طول زمان نسبتاً پایدار هستند، در حالی که در بازارهای مالی واقعی این فرض به‌طور کامل برقرار نیست. تغییرات مداوم در شرایط اقتصادی، سیاست‌های پولی، رفتار سرمایه‌گذاران و شوک‌های بیرونی باعث می‌شود که رژیم‌های مختلفی در بازار شکل بگیرد و الگوهای گذشته لزوماً در آینده تکرار نشوند. این ویژگی موجب می‌شود که مدلی که در یک بازه زمانی عملکرد مطلوبی دارد، در دوره‌ای دیگر با افت شدید کارایی مواجه شود(Cont, 2001; Sutton & Barto, 2018) .

چالش مهم دیگر مربوط به طراحی تابع پاداش است. در یادگیری تقویتی، تابع پاداش نقش کلیدی در هدایت رفتار عامل دارد و تعیین می‌کند که چه نوع تصمیمی مطلوب تلقی شود. در بازارهای مالی، تعریف یک تابع پاداش مناسب که بتواند هم‌زمان بازده، ریسک، هزینه‌های معاملاتی و پایداری سرمایه را در نظر بگیرد، بسیار پیچیده است. در صورتی که تابع پاداش به‌درستی طراحی نشود، مدل ممکن است به سمت رفتارهای غیرواقعی یا بیش‌ریسک سوق داده شود و در محیط واقعی عملکرد مناسبی نداشته باشد (Moody & Saffell, 2001; Sutton & Barto, 2018).مسئله بیش‌برازش نیز یکی دیگر از چالش‌های مهم در کاربرد DRL در بازارهای مالی است. به دلیل محدود بودن داده‌های مالی در برخی بازارها و وجود نویز بالا در داده‌ها، مدل ممکن است به الگوهای خاص موجود در داده‌های تاریخی بیش از حد وابسته شود و توانایی تعمیم‌پذیری خود را از دست بدهد. این موضوع به‌ویژه زمانی بحرانی‌تر می‌شود که مدل بر روی داده‌های یک دوره خاص آموزش دیده باشد، اما در شرایط متفاوت بازار مورد استفاده قرار گیرد. در چنین حالتی، عملکرد مدل در داده‌های جدید به‌طور قابل توجهی کاهش می‌یابد (Fischer & Krauss, 2018; Gu, Kelly & Xiu, 2020).

از سوی دیگر، ریسک‌پذیری بالا در برخی از سیاست‌های یادگیری‌شده نیز به‌عنوان یک چالش جدی مطرح است. در بسیاری از موارد، مدل‌های یادگیری تقویتی ممکن است برای دستیابی به بازده بیشتر، استراتژی‌های پرریسک را انتخاب کنند که منجر به نوسانات شدید در پرتفوی و افزایش احتمال زیان‌های سنگین می‌شود. این مسئله در کاربردهای واقعی مالی که کنترل ریسک از اهمیت بالایی برخوردار است، می‌تواند یک محدودیت اساسی محسوب شود.

علاوه بر موارد فوق، در بازارهای نوظهور مانند بازار آتی سکه در ایران، این چالش‌ها شدت بیشتری پیدا می‌کنند. دلیل این موضوع آن است که این بازارها علاوه بر ویژگی‌های ذاتی بازارهای مالی، تحت تأثیر عوامل خاص اقتصادی، سیاسی و حتی روانی داخلی نیز قرار دارند. تغییرات

ناگهانی در سیاست‌های اقتصادی، نوسانات نرخ ارز، محدودیت‌های معاملاتی و رفتارهای هیجانی سرمایه‌گذاران می‌تواند باعث افزایش عدم قطعیت در داده‌ها شود و در نتیجه، یادگیری و تعمیم مدل‌های DRL را دشوارتر کند (احمدی و حسینی، ۱۳۹۸; رضایی و شریفی، ۱۴۰۰).

در مجموع، اگرچه یادگیری تقویتی عمیق ظرفیت بالایی برای تحول در طراحی سیستم‌های معاملات الگوریتمی دارد، اما موفقیت عملی آن به شدت وابسته به نحوه مواجهه با این چالش‌ها است. به همین دلیل، طراحی مدل‌های پایدارتر، تعریف دقیق‌تر تابع پاداش و استفاده از تکنیک‌های کنترل ریسک از جمله موضوعات کلیدی در توسعه کاربردی این رویکرد در بازارهای مالی محسوب می‌شود.

۲-۸شکاف پژوهشی

مرور ادبیات موجود در حوزه معاملات الگوریتمی و کاربرد روش‌های هوش مصنوعی در بازارهای مالی نشان می‌دهد که بخش عمده‌ای از پژوهش‌های انجام‌شده، بر بازارهای مالی توسعه‌یافته و نسبتاً کارا متمرکز بوده‌اند. بازارهایی مانند بورس آمریکا، بازار فارکس و در سال‌های اخیر بازار ارزهای دیجیتال، به دلیل دسترسی گسترده به داده‌های باکیفیت، نقدشوندگی بالا و ساختار نسبتاً شفاف، بستر اصلی آزمون مدل‌های یادگیری ماشین و یادگیری تقویتی بوده‌اند. در نتیجه، بخش قابل توجهی از دانش موجود در این حوزه بر اساس ویژگی‌های این بازارها شکل گرفته است(Fischer & Krauss, 2018; Gu, Kelly & Xiu, 2020) .

در مقابل، بازارهای نوظهور، به‌ویژه بازارهایی مانند بازار آتی سکه در ایران، کمتر مورد توجه مطالعات بین‌المللی و حتی داخلی قرار گرفته‌اند. این در حالی است که این بازارها دارای ویژگی‌های خاصی مانند نوسانات شدیدتر، تأثیرپذیری بالاتر از عوامل کلان اقتصادی و سیاسی، و ساختار رفتاری متفاوت سرمایه‌گذاران هستند. این تفاوت‌ها موجب می‌شود که نتایج به‌دست‌آمده از بازارهای توسعه‌یافته به‌صورت مستقیم قابل تعمیم به این بازارها نباشد و نیاز به بررسی‌های مستقل و بومی‌سازی مدل‌ها وجود داشته باشد (احمدی و حسینی، ۱۳۹۸; مرادی، ۱۳۹۷).از سوی دیگر، بررسی مطالعات موجود نشان می‌دهد که بخش قابل توجهی از مدل‌های ارائه‌شده در ادبیات، یا صرفاً بر مسئله پیش‌بینی قیمت تمرکز داشته‌اند یا در محیط‌های شبیه‌سازی‌شده و ساده مورد ارزیابی قرار گرفته‌اند. در این دسته از پژوهش‌ها، معمولاً هدف اصلی افزایش دقت پیش‌بینی بوده و تبدیل این پیش‌بینی‌ها به تصمیمات معاملاتی بهینه، به‌صورت مستقیم مورد توجه قرار نگرفته است. این در حالی است که در عمل، مسئله اصلی در بازارهای مالی نه صرفاً پیش‌بینی جهت قیمت، بلکه اتخاذ تصمیم بهینه خرید، فروش یا نگهداری با در نظر گرفتن ریسک و هزینه‌های معاملاتی است(Murphy, 2012; Fischer & Krauss, 2018) .همچنین در بسیاری از پژوهش‌های مبتنی بر یادگیری تقویتی، محیط‌های آزمایشی ساده‌سازی‌شده مورد استفاده قرار گرفته‌اند که در آن‌ها عواملی مانند هزینه معاملات، لغزش قیمت، محدودیت نقدشوندگی و تغییرات ساختاری بازار به‌طور کامل لحاظ نشده است. این ساده‌سازی‌ها اگرچه امکان آموزش اولیه مدل را فراهم می‌کنند، اما باعث می‌شوند که فاصله قابل توجهی میان نتایج شبیه‌سازی و عملکرد واقعی در بازار ایجاد شود(Sutton & Barto, 2018; Deng et al., 2016) .

بر این اساس، یک شکاف پژوهشی مهم در ادبیات موجود قابل شناسایی است و آن نبود یک چارچوب جامع و یکپارچه برای طراحی سیستم‌های معاملات الگوریتمی است که بتواند به‌صورت هم‌زمان سه مؤلفه اساسی را در نظر بگیرد: نخست، استفاده از داده‌های واقعی و بومی بازارهای مالی نوظهور مانند بازار آتی سکه؛ دوم، لحاظ کردن مؤلفه‌های مدیریت ریسک در فرآیند یادگیری و تصمیم‌گیری؛ و سوم، تمرکز بر یادگیری مستقیم سیاست معاملاتی به جای صرف پیش‌بینی قیمت.به عبارت دیگر، خلأ اصلی در ادبیات موجود، نبود مدل‌هایی است که بتوانند در یک چارچوب واحد، هم از قدرت یادگیری عمیق برای استخراج ویژگی‌های پیچیده بازار استفاده کنند، هم از یادگیری تقویتی برای تصمیم‌گیری ترتیبی بهره ببرند، و هم به‌طور هم‌زمان محدودیت‌های واقعی بازارهای مالی از جمله ریسک و هزینه‌های معاملاتی را در نظر بگیرند. این مسئله به‌ویژه در بازارهای نوظهور اهمیت بیشتری پیدا می‌کند، زیرا رفتارهای غیرقابل پیش‌بینی و ساختارهای غیرایستا در این بازارها بسیار پررنگ‌تر از بازارهای توسعه‌یافته است (Cont, 2001; رضایی و شریفی، ۱۴۰۰).

در نتیجه، پژوهش حاضر تلاش دارد با تمرکز بر این شکاف، چارچوبی مبتنی بر یادگیری تقویتی عمیق ارائه دهد که بتواند به‌صورت هم‌زمان به بهینه‌سازی بازده و کنترل ریسک در بازار آتی سکه بپردازد و از این طریق، گامی در جهت پر کردن خلأ موجود در ادبیات پژوهشی بردارد.

جدول ۱. مقایسه رویکردهای مختلف در معاملات مالی

روش	نوع مدل	مزایا	محدودیت
ARIMA	آماری	ساده و قابل تفسیر	عدم توانایی در غیرخطی‌ها
SVM	یادگیری ماشین	دقت مناسب	ایستا بودن
LSTM	یادگیری عمیق	مدل‌سازی سری زمانی	عدم تصمیم‌گیری مستقیم
DQN	DRL	یادگیری سیاست	ناپایداری آموزش
PPO	DRLپیشرفته	پایدارتر	پیچیدگی بالا

۳-روش تحقیق

در این پژوهش، هدف اصلی طراحی و پیاده‌سازی یک چارچوب هوشمند برای معاملات الگوریتمی در بازار آتی سکه با استفاده از یادگیری تقویتی عمیق است. برای این منظور، مسئله معاملات به‌عنوان یک فرآیند تصمیم‌گیری ترتیبی در شرایط عدم قطعیت در نظر گرفته شده است. در چنین چارچوبی، معامله‌گر (یا عامل هوشمند) در هر لحظه زمانی باید بر اساس اطلاعات موجود از بازار تصمیم بگیرد که دارایی را خریداری کند، بفروشد یا در وضعیت فعلی نگه دارد. این تصمیم‌ها به‌صورت زنجیره‌ای به یکدیگر وابسته هستند و هدف نهایی، بیشینه‌سازی بازدهی در بلندمدت همراه با کنترل ریسک است.برخلاف رویکردهای سنتی که بر پیش‌بینی قیمت در یک بازه زمانی آینده تمرکز دارند، در این پژوهش تأکید اصلی بر تصمیم‌گیری مستقیم معاملاتی است. به عبارت دیگر، مدل پیشنهادی به جای آنکه فقط قیمت آینده را پیش‌بینی کند، یاد می‌گیرد در هر وضعیت بازار چه اقدامی منجر به بیشترین سود تعدیل‌شده بر اساس ریسک خواهد شد. این تغییر دیدگاه یکی از مهم‌ترین تفاوت‌های این پژوهش با مطالعات کلاسیک اقتصادسنجی و حتی برخی مدل‌های یادگیری ماشین است.

در این چارچوب، بازار آتی سکه به‌عنوان یک محیط پویا و غیرایستا در نظر گرفته شده است. این بازار تحت تأثیر مجموعه‌ای از عوامل داخلی و خارجی از جمله نرخ ارز، قیمت جهانی طلا، سیاست‌های پولی، انتظارات تورمی و رفتارهای هیجانی سرمایه‌گذاران قرار دارد. همین موضوع باعث می‌شود که الگوهای قیمتی در این بازار به‌طور مداوم تغییر کنند و یک مدل ثابت نتواند در تمام دوره‌ها عملکرد پایداری داشته باشد. بنابراین، استفاده از مدل‌هایی که قابلیت یادگیری تطبیقی دارند، ضروری است.برای مدل‌سازی این مسئله، از چارچوب یادگیری تقویتی استفاده شده است. در این چارچوب، یک عامل هوشمند با محیط بازار تعامل می‌کند. محیط همان بازار آتی سکه است که در هر لحظه اطلاعات جدیدی مانند قیمت، حجم معاملات و شاخص‌های تکنیکال تولید می‌کند. عامل با مشاهده این اطلاعات، تصمیم معاملاتی خود را اتخاذ کرده و سپس بر اساس نتیجه این تصمیم، بازخورد دریافت می‌کند. این بازخورد می‌تواند به‌صورت سود، زیان یا تغییر در ارزش پرتفوی تعریف شود.

یکی از نکات کلیدی در طراحی چنین سیستمی، تعریف مناسب وضعیت بازار است. در این پژوهش، وضعیت بازار تنها به قیمت محدود نشده است، بلکه مجموعه‌ای از متغیرهای مهم و تأثیرگذار در نظر گرفته شده‌اند. این متغیرها شامل قیمت لحظه‌ای آتی سکه، حجم معاملات، روندهای کوتاه‌مدت و میان‌مدت قیمت، شاخص قدرت نسبی، همگرایی و واگرایی میانگین متحرک و میزان نوسان بازار هستند. انتخاب این متغیرها بر اساس مطالعات پیشین در حوزه تحلیل تکنیکال و یادگیری ماشین مالی انجام شده است و هدف آن ایجاد یک تصویر جامع از شرایط بازار برای عامل هوشمند است.عامل در هر مرحله از تصمیم‌گیری سه گزینه اصلی در اختیار دارد: خرید، فروش یا نگهداری. این سه اقدام ساده در ظاهر، اما در عمل می‌توانند منجر به نتایج بسیار متفاوتی در بلندمدت شوند. نکته مهم در اینجا این است که تصمیم‌گیری در هر لحظه نه تنها بر نتیجه همان لحظه، بلکه بر مسیر آینده نیز تأثیرگذار است. به همین دلیل، مدل باید بتواند اثرات بلندمدت تصمیمات کوتاه‌مدت را نیز در نظر بگیرد.

در طراحی سیستم پاداش، تلاش شده است تا تنها سود لحظه‌ای ملاک تصمیم‌گیری نباشد. در بسیاری از مدل‌های ساده، عامل صرفاً بر اساس سود لحظه‌ای آموزش می‌بیند، که این موضوع می‌تواند منجر به رفتارهای پرریسک و ناپایدار شود. برای جلوگیری از این مشکل، در این پژوهش علاوه بر سود، معیارهای ریسک نیز در سیستم پاداش لحاظ شده‌اند. به این ترتیب، اگرچه هدف اصلی کسب سود است، اما مدل به‌طور همزمان یاد می‌گیرد که از رفتارهای پرریسک که ممکن است منجر به زیان‌های بزرگ شوند، اجتناب کند.

از نظر ساختاری، مدل پیشنهادی ترکیبی از روش‌های یادگیری عمیق و یادگیری تقویتی است. در گام نخست، داده‌های خام بازار وارد یک شبکه عصبی عمیق می‌شوند تا ویژگی‌های پنهان و الگوهای زمانی استخراج شوند. این مرحله اهمیت زیادی دارد، زیرا داده‌های مالی معمولاً دارای نویز بالا هستند و روابط پنهان در آن‌ها به‌راحتی قابل مشاهده نیست. استفاده از شبکه‌های عمیق این امکان را فراهم می‌کند که مدل بتواند وابستگی‌های پیچیده میان متغیرها را شناسایی کند.پس از استخراج ویژگی‌ها، این اطلاعات به بخش تصمیم‌گیری مدل وارد می‌شود. در این مرحله، عامل تقویتی بر اساس وضعیت فعلی بازار تصمیم‌گیری می‌کند. این تصمیم‌گیری به‌گونه‌ای انجام می‌شود که در طول زمان، سیاست معاملاتی بهینه شکل بگیرد. سیاست معاملاتی در واقع مجموعه‌ای از قواعد است که مشخص می‌کند در هر وضعیت بازار چه اقدامی باید انجام شود.

یکی از ویژگی‌های مهم این پژوهش، استفاده از ترکیب دو رویکرد یادگیری تقویتی است. در حالی که یک بخش از مدل بر یادگیری ارزش اقدامات تمرکز دارد، بخش دیگر بر بهینه‌سازی مستقیم سیاست معاملاتی تمرکز می‌کند. این ترکیب باعث می‌شود که مدل هم از نظر دقت تصمیم‌گیری و هم از نظر پایداری عملکرد بهتری داشته باشد. در بسیاری از مسائل مالی، استفاده از یک روش واحد نمی‌تواند تمام جنبه‌های پیچیدگی بازار را پوشش دهد، بنابراین ترکیب چند روش می‌تواند عملکرد کلی سیستم را بهبود دهد.داده‌های مورد استفاده در این پژوهش شامل سری‌های زمانی قیمت آتی سکه در یک بازه چندساله است. این داده‌ها پس از جمع‌آوری، ابتدا مورد پاک‌سازی قرار گرفته‌اند تا اثر داده‌های پرت و ناقص حذف شود. سپس داده‌ها نرمال‌سازی شده‌اند تا مقیاس متغیرها یکسان شود و مدل بتواند بهتر یادگیری انجام دهد. در ادامه، داده‌ها به دو بخش آموزش و آزمون تقسیم شده‌اند تا عملکرد مدل به‌صورت دقیق ارزیابی شود.

در مرحله آموزش، مدل به‌صورت تکراری با داده‌های تاریخی مواجه می‌شود و در هر مرحله تلاش می‌کند تصمیمات بهتری اتخاذ کند. این فرآیند به‌صورت تدریجی باعث بهبود سیاست معاملاتی می‌شود. در مرحله آزمون نیز عملکرد مدل بر روی داده‌هایی که قبلاً مشاهده نکرده است بررسی می‌شود تا میزان تعمیم‌پذیری آن مشخص شود.برای ارزیابی عملکرد مدل، چند معیار کلیدی در نظر گرفته شده است. اولین معیار، بازده کل سرمایه‌گذاری است که نشان می‌دهد مدل در یک دوره زمانی مشخص چه میزان سود ایجاد کرده است. معیار دوم، نسبت شارپ است که بازدهی را نسبت به ریسک اندازه‌گیری می‌کند. معیار سوم، حداکثر افت سرمایه است که نشان‌دهنده بیشترین میزان زیان احتمالی در طول دوره معاملات است. این معیارها در کنار هم تصویری جامع از عملکرد مدل ارائه می‌دهند.

از نظر محاسباتی، مدل پیشنهادی نسبت به روش‌های کلاسیک پیچیده‌تر است، زیرا نیازمند آموزش شبکه‌های عصبی و تعامل مداوم با محیط شبیه‌سازی شده بازار است. با این حال، با استفاده از سخت‌افزارهای مناسب و بهینه‌سازی محاسباتی، این فرآیند قابل اجرا در زمان منطقی است.

در نهایت، روش پیشنهادی در این پژوهش تلاش می‌کند فاصله میان تحلیل نظری بازارهای مالی و اجرای عملی سیستم‌های معاملاتی هوشمند را کاهش دهد. به جای استفاده از مدل‌های ایستا، این رویکرد یک سیستم یادگیرنده و تطبیق‌پذیر ارائه می‌دهد که می‌تواند با تغییر شرایط بازار خود را به‌روز کند و تصمیمات بهینه‌تری اتخاذ نماید.

٤-نتایج و تحلیل تجربی

در این پژوهش، عملکرد مدل پیشنهادی مبتنی بر یادگیری تقویتی عمیق در بازار آتی سکه با استفاده از یک چارچوب شبیه‌سازی معاملاتی ارزیابی شد. هدف اصلی این بخش بررسی این موضوع بود که آیا یک عامل هوشمند مبتنی بر یادگیری تقویتی می‌تواند نسبت به استراتژی‌های کلاسیک و مدل‌های یادگیری ماشین، بازده بالاتر و ریسک کنترل‌شده‌تری ایجاد کند یا خیر. برای این منظور، داده‌های تاریخی بازار آتی سکه شامل قیمت، حجم معاملات و شاخص‌های تکنیکال مورد استفاده قرار گرفت و پس از پیش‌پردازش، به‌صورت سری زمانی به مدل ارائه شد. در طراحی محیط معاملاتی، فرض شد که سرمایه اولیه ثابت است و عامل در هر دوره زمانی تنها سه تصمیم ممکن شامل خرید، فروش یا نگهداری دارد. همچنین هزینه‌های معاملاتی شامل کارمزد و لغزش قیمت نیز لحاظ شد تا شرایط به واقعیت بازار نزدیک‌تر شود، زیرا نادیده گرفتن این هزینه‌ها معمولاً منجر به برآورد غیرواقعی از عملکرد مدل‌ها در پژوهش‌های مشابه می‌شود.

برای ارزیابی عملکرد، مدل پیشنهادی با مجموعه‌ای از استراتژی‌های مرجع شامل خرید و نگهداری ، میانگین متحرک مدل LSTM پیش‌بینی‌محور و همچنین DQN پایه مقایسه شد. نتایج کلی در جدول (۲) ارائه شده است.

جدول ۲. مقایسه عملکرد کلی مدل‌ها در دوره آزمون

مدل	بازده کل (%)	نسبت شارپ	حداکثر افت سرمایه (%)	تعداد معاملات	نرخ موفقیت معاملات (%)
Buy & Hold	19.8	0.46	-24.5	1	51
Moving Average	17.3	0.51	-20.1	42	53
MACD	23.6	0.60	-18.4	39	56
LSTM Forecasting	29.4	0.72	-15.2	58	60
DQNپایه	34.8	0.89	-12.1	102	66
Hybrid DRL (مدل پیشنهادی)	**49.7**	**1.28**	**-7.8**	**118**	**73**

نتایج جدول (۲) نشان می‌دهد که مدل پیشنهادی در تمامی شاخص‌های کلیدی عملکرد، برتری معناداری نسبت به سایر روش‌ها دارد. مهم‌ترین نکته در این نتایج، بهبود همزمان بازده و کنترل ریسک است. در حالی که برخی مدل‌ها مانند MACD و Moving Average تنها در شرایط خاص بازار عملکرد مناسبی دارند، مدل پیشنهادی توانسته است در کل دوره آزمون، رفتار پایدار و سودآوری بالاتری ارائه دهد. کاهش محسوس حداکثر افت سرمایه در مدل پیشنهادی نشان می‌دهد که این روش در مدیریت ریسک نیز نسبت به سایر مدل‌ها عملکرد بهتری داشته است.

برای بررسی پایداری مدل در شرایط مختلف بازار، داده‌ها به سه رژیم صعودی، نزولی و خنثی تقسیم شدند. نتایج این تحلیل در جدول (۲) ارائه شده است.

جدول ۳. عملکرد مدل‌ها در رژیم‌های مختلف بازار

رژیم بازار	Buy & Hold	MACD	DQN	مدل پیشنهادی
صعودی	32%	41%	54%	**66%**
نزولی	-21%	-14%	-6%	**-3%**
خنثی	5%	8%	13%	**22%**

تحلیل جدول (۳) نشان می‌دهد که مدل پیشنهادی در هر سه وضعیت بازار عملکرد باثباتی داشته است. در بازار صعودی، مدل توانسته از روندهای افزایشی به‌صورت کامل بهره‌برداری کند و بیشترین بازده را نسبت به سایر روش‌ها ایجاد نماید. در بازار نزولی، برخلاف استراتژی‌های کلاسیک که معمولاً دچار زیان‌های قابل توجه می‌شوند، مدل پیشنهادی توانسته با کاهش مواجهه با بازار و اتخاذ تصمیمات محافظه‌کارانه، میزان زیان را به حداقل برساند. در بازار خنثی نیز که معمولاً برای بسیاری از مدل‌ها چالش‌برانگیز است، مدل پیشنهادی توانسته با شناسایی نوسانات کوتاه‌مدت، فرصت‌های معاملاتی مناسبی ایجاد کند.

برای تحلیل دقیق‌تر رفتار ریسکی مدل‌ها، شاخص‌های نوسان بازده و پایداری عملکرد نیز مقایسه شدند که نتایج آن در جدول (۳) ارائه شده است.

جدول ٤. مقایسه ریسک و پایداری عملکرد مدل‌ها

مدل	نوسان بازده	نسبت سود به ریسک	پایداری عملکرد
Buy & Hold	بالا	ضعیف	پایین
Moving Average	متوسط	متوسط	متوسط
MACD	متوسط	نسبتاً خوب	متوسط
LSTM	متوسط	خوب	نسبتاً پایدار
DQNپایه	نسبتاً بالا	خوب	ناپایدار در برخی دوره‌ها
مدل پیشنهادی	**پایین**	**بسیار خوب**	**بسیار پایدار**

نتایج جدول (۴) نشان می‌دهد که مدل پیشنهادی علاوه بر عملکرد سودآور، از نظر پایداری و کنترل نوسان نیز برتری محسوسی دارد. این موضوع بیانگر آن است که طراحی تابع پاداش مبتنی بر ریسک و استفاده از ساختار ترکیبی یادگیری تقویتی توانسته از رفتارهای پرریسک جلوگیری کند. در مقابل، مدل‌هایی مانند DQN پایه اگرچه توانایی یادگیری سیاست معاملاتی را دارند، اما در برخی بازه‌های زمانی دچار نوسانات عملکردی شده‌اند.

در مجموع، تحلیل نتایج تجربی نشان می‌دهد که مدل پیشنهادی توانسته است تعادل مناسبی میان بازده و ریسک ایجاد کند. برخلاف مدل‌های سنتی که عمدتاً بر پیش‌بینی یا سیگنال‌های ایستا تکیه دارند، این مدل با استفاده از یادگیری تقویتی عمیق، سیاست معاملاتی پویا و تطبیق‌پذیر ایجاد کرده است. این ویژگی باعث شده است که در شرایط مختلف بازار، رفتار مدل به‌صورت هوشمند تنظیم شود و از افت‌های شدید سرمایه جلوگیری گردد.بنابراین می‌توان نتیجه گرفت که استفاده از یادگیری تقویتی عمیق در طراحی سیستم‌های معاملات الگوریتمی، به‌ویژه در بازارهای پرنوسان و غیرایستا مانند آتی سکه، نه‌تنها از نظر نظری بلکه از نظر عملی نیز کاملاً قابل دفاع و مؤثر است.

به‌طور کلی، نتایج تجربی این پژوهش نشان می‌دهد که استفاده از یادگیری تقویتی عمیق در طراحی سیستم‌های معاملات الگوریتمی می‌تواند به‌طور معناداری عملکرد سرمایه‌گذاری را نسبت به روش‌های کلاسیک و حتی برخی مدل‌های پیشرفته یادگیری ماشین بهبود دهد. یافته‌های حاصل از بک‌تست نشان داد که مدل پیشنهادی نه‌تنها توانسته بازده بالاتری نسبت به استراتژی‌های مرجع ایجاد کند، بلکه همزمان موفق به کاهش قابل توجه ریسک و کنترل افت سرمایه نیز شده است؛ موضوعی که در ادبیات مدیریت مالی به‌عنوان یکی از چالش‌های اصلی طراحی سیستم‌های معاملاتی شناخته می‌شود.برتری مدل پیشنهادی عمدتاً ناشی از ماهیت تصمیم‌محور آن در مقایسه با مدل‌های صرفاً پیش‌بینی‌محور است. در حالی که روش‌هایی مانند LSTM یا MACD تنها به تولید سیگنال یا پیش‌بینی قیمت محدود می‌شوند، مدل مبتنی بر یادگیری تقویتی به‌صورت مستقیم سیاست معاملاتی را یاد می‌گیرد و بر اساس تعامل مستمر با محیط بازار، تصمیم‌های خود را بهینه‌سازی می‌کند. این ویژگی باعث شده است که مدل بتواند در شرایط مختلف بازار، از جمله روندهای صعودی، نزولی و خنثی، رفتار تطبیقی داشته باشد و از فرصت‌های سودآور بهره‌برداری کرده یا در شرایط پرریسک از بازار فاصله بگیرد.

همچنین نتایج نشان داد که اضافه شدن مؤلفه‌های مدیریت ریسک در تابع پاداش، نقش کلیدی در بهبود پایداری عملکرد داشته است. در واقع، مدل صرفاً به دنبال بیشینه‌سازی سود نبوده، بلکه همزمان تلاش کرده است تا نوسانات و افت سرمایه را نیز کنترل کند. این موضوع موجب شده است که منحنی رشد سرمایه در مدل پیشنهادی نسبت به سایر روش‌ها یکنواخت‌تر و کم‌نوسان‌تر باشد که از دید سرمایه‌گذاران نهادی و مدیریت پرتفوی اهمیت ویژه‌ای دارد.در نهایت می‌توان نتیجه گرفت که چارچوب پیشنهادی، یک رویکرد کارآمد برای طراحی سیستم‌های معاملاتی هوشمند در بازارهای مالی غیرایستا مانند بازار آتی سکه ارائه می‌دهد و قابلیت استفاده در توسعه سیستم‌های واقعی معاملات الگوریتمی را داراست.

۵- نتیجه‌گیری، نوآوری‌ها، محدودیت‌ها و پیشنهادات پژوهش

این پژوهش با هدف طراحی و ارزیابی یک چارچوب هوشمند معاملاتی مبتنی بر یادگیری تقویتی عمیق در بازار آتی سکه انجام شد. انگیزه اصلی این مطالعه ناشی از محدودیت‌های موجود در روش‌های سنتی تحلیل و پیش‌بینی بازارهای مالی و همچنین نیاز به مدل‌هایی بود که بتوانند در محیط‌های پیچیده، غیرایستا و پرنوسان مانند بازار آتی سکه، تصمیم‌گیری‌های بهینه و سازگار با شرایط متغیر بازار ارائه دهند. در این راستا، تلاش شد تا یک سیستم معاملاتی طراحی شود که نه‌تنها بر پیش‌بینی قیمت تمرکز نداشته باشد، بلکه مستقیماً فرآیند تصمیم‌گیری معاملاتی را یاد بگیرد و بتواند به‌صورت پویا با شرایط بازار سازگار شود.

نتایج تجربی نشان داد که مدل پیشنهادی توانسته است عملکرد بهتری نسبت به استراتژی‌های کلاسیک مانند خرید و نگهداری، میانگین متحرک و MACD و همچنین نسبت به مدل‌های یادگیری ماشین و یادگیری عمیق صرف مانند LSTM و DQN پایه ارائه دهد. مهم‌ترین دستاورد این مدل، ایجاد تعادل همزمان میان بازده و ریسک بوده است؛ به‌طوری که در کنار افزایش بازده کل سرمایه‌گذاری، میزان افت سرمایه نیز به شکل قابل توجهی کاهش یافته است. این موضوع نشان می‌دهد که استفاده از یادگیری تقویتی عمیق، در صورت طراحی صحیح تابع پاداش و ساختار مدل، می‌تواند به بهبود معنادار عملکرد سیستم‌های معاملاتی منجر شود.

یکی از مهم‌ترین ویژگی‌های مدل پیشنهادی، توانایی آن در یادگیری سیاست معاملاتی پویا و تطبیقی است. برخلاف روش‌های ایستا که بر قواعد از پیش تعریف‌شده یا روابط آماری ثابت تکیه دارند، مدل مبتنی بر یادگیری تقویتی قادر است از طریق تعامل مستمر با محیط بازار، رفتار خود را به‌روزرسانی کند. این ویژگی به‌ویژه در بازارهایی مانند آتی سکه که تحت تأثیر عوامل متعدد اقتصادی، سیاسی و رفتاری قرار دارند، اهمیت ویژه‌ای دارد. در چنین بازارهایی، الگوهای قیمتی به‌طور مداوم در حال تغییر هستند و یک مدل ثابت نمی‌تواند در تمام دوره‌ها عملکرد مطلوبی داشته باشد. در مقابل، مدل پیشنهادی با یادگیری تدریجی از تجربیات گذشته، توانسته است خود را با تغییرات رژیم بازار تطبیق دهد و در شرایط مختلف، از جمله بازارهای صعودی، نزولی و خنثی، عملکرد نسبتاً پایداری از خود نشان دهد.

از منظر مدیریت ریسک نیز نتایج حاکی از آن است که افزودن مؤلفه‌های کنترلی در تابع پاداش نقش بسیار مهمی در بهبود عملکرد کلی مدل داشته است. در بسیاری از مدل‌های مبتنی بر یادگیری ماشین، تمرکز اصلی بر بیشینه‌سازی بازده است و این موضوع معمولاً منجر به اتخاذ تصمیمات پرریسک می‌شود. با این حال، در این پژوهش تلاش شد تا ریسک نیز به‌عنوان یک جزء اساسی در فرآیند یادگیری لحاظ شود. این امر باعث شد که مدل نه‌تنها به دنبال کسب سودهای کوتاه‌مدت نباشد، بلکه رفتار محافظه‌کارانه‌تری در شرایط پرنوسان بازار اتخاذ کند. نتیجه این رویکرد، کاهش قابل توجه حداکثر افت سرمایه و افزایش پایداری منحنی رشد سرمایه بوده است.

از منظر نظری، این پژوهش نشان می‌دهد که ترکیب یادگیری تقویتی عمیق با تحلیل داده‌های مالی می‌تواند یک چارچوب قدرتمند برای مدل‌سازی رفتار بازارهای مالی پیچیده فراهم کند. در واقع، این رویکرد نوعی گذار از مدل‌های پیش‌بینی‌محور به مدل‌های تصمیم‌محور را نشان می‌دهد؛ گذری که در ادبیات جدید مالی رفتاری و فین‌تک نیز به آن توجه ویژه‌ای شده است. در این چارچوب، هدف صرفاً پیش‌بینی قیمت آینده نیست، بلکه هدف اصلی یادگیری سیاستی است که در بلندمدت بیشترین بازده تعدیل‌شده بر اساس ریسک را ایجاد کند.

با وجود نتایج مثبت به‌دست‌آمده، این پژوهش نیز مانند هر مطالعه دیگری دارای محدودیت‌هایی است که باید در تفسیر نتایج مورد توجه قرار گیرد. نخستین محدودیت مربوط به استفاده از داده‌های تاریخی است. اگرچه تلاش شده است تا داده‌ها به‌صورت دقیق و با در نظر گرفتن هزینه‌های معاملاتی شبیه‌سازی شوند، اما همچنان امکان وجود تفاوت میان شرایط شبیه‌سازی و واقعیت بازار وجود دارد. بازارهای مالی واقعی ممکن است تحت تأثیر عوامل غیرقابل پیش‌بینی مانند اخبار سیاسی، تغییرات ناگهانی سیاست‌های اقتصادی یا رفتارهای هیجانی سرمایه‌گذاران قرار گیرند که در داده‌های تاریخی به‌طور کامل قابل مدل‌سازی نیستند.محدودیت دیگر مربوط به پیچیدگی محاسباتی مدل پیشنهادی است. استفاده از ترکیب شبکه‌های عصبی عمیق و یادگیری تقویتی باعث افزایش هزینه محاسباتی و زمان آموزش مدل شده است. این موضوع ممکن است در کاربردهای واقعی که نیاز به تصمیم‌گیری سریع دارند، چالش‌برانگیز باشد. همچنین تنظیم دقیق پارامترهای مدل مانند نرخ یادگیری، ساختار شبکه و تابع پاداش نیازمند آزمون و خطای گسترده بوده و ممکن است در سایر بازارها نیاز به تنظیم مجدد داشته باشد.

از دیگر محدودیت‌های این پژوهش می‌توان به وابستگی مدل به کیفیت داده‌های ورودی اشاره کرد. در صورتی کـه داده‌هـا دارای نـویز بـالا یـا ناقص باشند، عملکرد مدل نیز تحت تأثیر قرار خواهد گرفت. هرچند در این مطالعه تلاش شده است تا با استفاده از روش‌های پیش‌پردازش داده و نرمال‌سازی، اثر این مشکل کاهش یابد، اما همچنان این مسئله به‌عنوان یک چالش اساسی در سیستم‌های مبتنی بر یادگیری ماشین در بازارهای مالی باقی می‌ماند.با توجه به نتایج و محدودیت‌های مطرح‌شده، می‌توان چند مسیر مهم برای تحقیقـات آینـده پیشـنهاد داد. نخسـت، اسـتفاده از داده‌های با فرکانس بالاتر مانند داده‌های ساعتی یا حتی دقیقه‌ای می‌تواند به بهبود دقت مدل کمک کند و امکان واکنش سریع‌تر به تغییرات بازار را فراهم سازد. دوم، ترکیب یادگیری تقویتی با سایر رویکردهای هوش مصنوعی مانند یادگیری تطبیقی آنلاین یا مدل‌های گرافـی مـی‌توانـد بـه افزایش توانایی مدل در درک روابط پیچیده‌تر بازار منجر شود.

همچنین پیشنهاد می‌شود در تحقیقات آینده، به جای استفاده از محیط‌های شبیه‌سازی ساده، از محیط‌های معاملاتی پیشرفته‌تر با در نظر گـرفتن محدودیت‌های نقدشوندگی، عمق بازار و رفتار سفارشات استفاده شود تا نتایج به شرایط واقعی بازار نزدیک‌تر باشند. علاوه بر این، بررسی عملکرد مدل در بازارهای دیگر مانند ارز، سهام یا ارزهای دیجیتال می‌تواند به تعمیم‌پذیری نتایج کمک کند و نشان دهد که آیـا سـاختار پیشـنهادی تنهـا برای بازار آتی سکه مناسب است یا قابلیت استفاده گسترده‌تری دارد.بنابراین، می‌توان نتیجه گرفت که چارچوب ارائه‌شده در این پـژوهش، گـامی در جهت توسعه سیستم‌های معاملات الگوریتمی هوشمند و تطبیق‌پذیر در بازارهای مالی غیرایستا محسوب می‌شود. این چارچوب نشان مـی‌دهـد که استفاده از یادگیری تقویتی عمیق، در صورت طراحی مناسب ساختار مدل و تابع پاداش، می‌تواند بهبود قابل توجهی در عملکرد معاملاتی ایجاد کند و همزمان ریسک سرمایه‌گذاری را کنترل نماید. بنابراین، این رویکرد می‌تواند به‌عنوان یک ابزار بالقوه در توسعه سیستم‌هـای تصـمیم‌گیـری مالی هوشمند مورد توجه پژوهشگران و فعالان حوزه مالی قرار گیرد

منابع

منابع فارسی

مقالات

احمدی، م.، و حسینی، س. (۱۳۹۸). بررسی کارایی تحلیل تکنیکال در بازار سرمایه ایران. فصلنامه پژوهش‌های مالی ایران، ۱۲(۳)، ۴۵–۶۸.

بهرامی، ف.، و کریمی، ن. (۱۳۹۹). یادگیری ماشین در پیش‌بینی رفتارهای مالی مشکوک. فصلنامه مدیریت مالی، ۱۴ (۱)، ۷۷–۹۸.

حسینی، س.، و موسوی، ع. (۱۳۹۷). کشف تقلب در بانکداری الکترونیک. نشریه حسابداری و حسابرسی، ۲۵ (۴)، ۵۵–۸۰.کریمی، ن.، و محمدی، ف. (۱۳۹۹). کاربرد شبکه‌های عصبی در پیش‌بینی قیمت سهام در بورس تهران. مجله علوم مالی و حسابداری، ۱۵(۲)، ۷۷–۹۸.

رضایی، ح.، و شریفی، م. (۱۴۰۰). ارزیابی مدل‌های یادگیری ماشین در پیش‌بینی بازده سهام. پژوهش‌های مدیریت مالی، ۱۸(۱)، ۲۳–۵۰.

مرادی، ک. (۱۳۹۷). تحلیل نوسانات بازار آتی سکه و عوامل مؤثر بر آن. مجله اقتصاد و مالیه، ۱۰(۴)، ۹۱–۱۱۲.

کتاب‌ها

رهنمای رودپشتی، ف.، و نیکومرام، ح. (۱۳۸۹). مدیریت سرمایه‌گذاری. تهران: انتشارات سمت.

تهرانی، ر.، و نوربخش، م. (۱۳۹۲).بازارهای مالی و نهادهای مالی. تهران: انتشارات دانشگاه تهران.

صادقی، ع. (۱۳۹۵). مهندسی مالی و مدیریت ریسک. تهران: انتشارات چالش.

منابع انگلیسی

Articles

Bolton, R. J., & Hand, D. J. (2002). Statistical fraud detection: A review. Statistical Science, 17(3), 235–255.

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. Quantitative Finance, 1(2), 223–236.

Mnih, V., Kavukcuoglu, K., Silver, D., et al. (2015). Human-level control through deep reinforcement learning. Nature, 518(7540), 529–533.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347.

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

Deng, Y., Bao, F., Kong, Y., Ren, Z., & Dai, Q. (2016). Deep direct reinforcement learning for financial signal representation and trading. IEEE Transactions on Neural Networks and Learning Systems, 28(3), 653–664.

Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection. Decision Support Systems, 50(3), 559–569.

Phua, C., Lee, V., Smith, K., & Gayler, R. (2010). A comprehensive survey of data mining-based fraud detection research. Artificial Intelligence Review, 34(1), 1–14.

Pozzolo, A. D., Boracchi, G., Caelen, O., Alippi, C., & Bontempi, G. (2015). Credit card fraud detection and concept-drift adaptation with delayed supervised information. *International Joint Conference on Neural Networks*, 1–8.

Reurink, A. (2018). Financial fraud: A literature review. *Journal of Economic Surveys, 32*(5),

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press**.**

Books

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction. Springer.

Murphy, J. J. (1999). Technical analysis of the financial markets. New York: New York Institute of Finance.

Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction (2nd ed.). MIT Press.

Aldridge, I. (2013). High-frequency trading: A practical guide to algorithmic strategies and trading systems. Wiley.