INSTITUTO
UNIVERSITÁRIO
DE LISBOA

O Impacto da Aprendizagem Híbrida por Reforço e Supervisionada no *Trading* - Ativo do Setor Tecnológico

Diego Júnior de Souza

Mestrado em Ciência de Dados

Orientadoras:
Doutora Diana Elisabeta Aldea Mendes, Professora Associada,
ISCTE - Instituto Universitário de Lisboa

Doutora Anabela Ribeiro Dias da Costa, Professora Auxiliar,
ISCTE - Instituto Universitário de Lisboa

Setembro, 2025

BUSINESS SCHOOL

TECNOLOGIAS E ARQUITETURA

Departamento de Métodos Quantitativos para Gestão e Economia

Departamento de Ciências e Tecnologia da Informação

O Impacto da Aprendizagem Híbrida por Reforço e Supervisionada no *Trading* - Ativo do Setor Tecnológico

Diego Júnior de Souza

Mestrado em Ciência de Dados

Orientadoras:
Doutora Diana Elisabeta Aldea Mendes, Professora Associada, ISCTE - Instituto Universitário de Lisboa

Doutora Anabela Ribeiro Dias da Costa, Professora Auxiliar, ISCTE - Instituto Universitário de Lisboa

Setembro, 2025

Agradecimentos

Em primeiro lugar, expresso a minha mais profunda gratidão às minhas orientadoras, Professora Diana Mendes e Professora Anabela Costa, pelo acompanhamento atento e pelo rigor científico que orientaram cada etapa deste percurso. Agradeço a disponibilidade constante, a partilha de conhecimento e a dedicação demonstrada, que foram muito além de um simples papel académico. O vosso apoio foi determinante não só para o desenvolvimento desta tese, mas também para o meu crescimento pessoal e académico.

À minha família, em especial à minha mãe e às minhas irmãs, expresso o meu profundo reconhecimento, pela compreensão diante das minhas ausências e pela flexibilidade em ajustar as dinâmicas familiares ao longo desta jornada. Mais do que ajuda prática, foram uma verdadeira rede de suporte, garantindo equilíbrio emocional, cuidado e incentivo. A vossa preocupação constante e a vossa presença permitiram-me manter a consciência da importância deste esforço, ajudando-me a ser resiliente perante os desafios que enfrentei.

Aos meus amigos, que mostraram compreensão e paciência quando precisei de me ausentar ou reduzir o convívio social, agradeço igualmente o apoio e o encorajamento.

Este trabalho é também fruto dessa rede de suporte, não apenas em aspetos práticos, mas sobretudo no apoio emocional e na motivação que recebi. A todos os que contribuíram direta ou indiretamente para que eu pudesse manter a resiliência e a determinação, deixo a minha sincera gratidão.

Dedico este estudo à minha avó, Francisca Eleutério, cuja sabedoria, transmitida ao longo das gerações, ensinou o valor da ética, a importância de adquirir conhecimento e o dever de aplicá-lo de forma sábia.

Resumo

Esta dissertação tem como objetivo investigar o impacto da aprendizagem híbrida por reforço e supervisionada aplicada ao *trading*, com destaque para ativos do setor tecnológico. Num contexto caracterizado por volatilidade elevada e incerteza significativa, os modelos híbridos combinando aprendizagem supervisionada (*XGBoost*) com aprendizagem por reforço (*Q-learning* e *Deep Q-Network*) são explorados para prever movimentos financeiros. Os modelos híbridos desenvolvidos incluem versões com variáveis exógenas e mecanismos específicos de controlo de risco, nomeadamente as estratégias de *stop-loss* e *trailing stop*. Estes modelos híbridos são comparados com um modelo base que utiliza exclusivamente aprendizagem por reforço. A comparação é realizada através de métricas de desempenho que avaliam a rentabilidade, a precisão das previsões e a eficácia na gestão do risco. Conclui-se que a aprendizagem híbrida por reforço constitui uma abordagem eficaz e promissora para otimizar estratégias de investimento em mercados altamente complexos e voláteis.

Palavras-chave: *Trading*, Modelos Híbridos; Aprendizagem por Reforço; *Q-learning*; *XGBoost* e *Deep Q-Network*.

Abstract

This dissertation aims to investigate the impact of hybrid reinforcement and supervised learning applied to trading, with a particular focus on assets from the technology sector. In a context marked by high volatility and considerable uncertainty, hybrid models that combine supervised learning (XGBoost) with reinforcement learning (Q-learning and Deep Q-Network) are explored to predict financial movements. The hybrid models developed include versions that incorporate exogenous variables as well as specific risk management mechanisms, namely stop-loss and trailing stop strategies. These hybrid approaches are compared with a baseline model relying exclusively on reinforcement learning. The comparison is conducted through performance metrics that assess profitability, predictive accuracy, and effectiveness in risk management. The findings suggest that hybrid reinforcement learning constitutes an effective and promising approach to optimizing investment strategies in highly complex and volatile markets.

Keywords: Trading; Hybrid models; Reinforcement learning; Q-learning; XGBoost; Deep Q-Network.

Índice

Agradecimentos i
Resumo iii
Abstract v
Índice de Tabelas ix
Índice de Figuras xi
Abreviaturas xiii
1. Introdução 1
1.1. Contexto Geral 1
1.2. Motivação e Objetivos 2
1.3. Estrutura 3
2. Revisão da Literatura 5
2.1. Evolução na Modelação Preditiva em *Trading* 5
2.2. Métodos na classificação da tendência 6
2.3. Métodos na previsão de preço 7
2.4. Métodos Híbridos e *Ensemble* 8
2.5. Análise comparativa 10
3. Metodologia 13
3.1. Contexto Geral 13
3.2 Análise de Dados 13
3.3 Fundamentos dos modelos utilizados 16
3.3.1 Aprendizagem por Reforço 17
3.3.1.1 *Q-learning* 20
3.3.1.2 *Deep Q-Network* 21
3.3.2 *XGBoost* 23
3.4 Modelos Híbridos 25
3.4.1 *Q-learning* com *XGBoost* 26
3.4.2 *Q-learning* com *XGBoost* e Variáveis Exógenas 27

3.4.3 *Q-learning* com *XGBoost* e Controlo de Risco 28
3.4.4 *DQN* com *XGBoost* 29
3.5 Métricas de Desempenho 31
4. Análise de Resultados..... 33
5. Conclusões 39
Bibliografia..... 41
Anexo 47

Índice de Tabelas

Tabela 1 - Estatísticas descritivas dos preços de fecho 14

Tabela 2 – Resultados comparativos dos modelos no intervalo de 22/06/2022 a 30/09/2024 33

Tabela 3 - Resultados comparativos dos modelos no intervalo de 01/01/2022 a 30/06/2023 36

Tabela 4 – Resultados dos Métodos na classificação da tendência 47

Tabela 5 – Resultados dos métodos na previsão de preço 48

Tabela 6 - Resultados dos Métodos Híbridos e *Ensemble* 49

Tabela 7 - Parâmetros e métricas dos modelos no intervalo de 22 de junho de 2022 a 30 de setembro de 2024 50

Tabela 8 - Parâmetros e métricas dos modelos no intervalo de 1 de janeiro de 2022 a 30 de junho de 2023 50

Tabela 9 - Parâmetros e métricas dos modelos *DQN+XGBoost* 51

Índice de Figuras

Figura 1 - Evolução histórica dos preços diários de fecho da NVIDIA (2014–2024) 15
Figura 2 - Retornos Diários da NVIDIA (2014–2024)..... 15
Figura 3 - Detecção de outliers no Volume negociado NVIDIA (2014–2024) 16
Figura 4 - Ilustração da interação do agente com o sistema..... 17
Figura 5 - Diagrama do modelo híbrido proposto combinando *XGBoost* e *Q-Learning* 26
Figura 8 - Impacto da gestão de risco na trajetória do retorno acumulado 34
Figura 7 - Comparação dos modelos em termos de *Sharpe Ratio* e *CAGR* 35
Figura 6 - Gráfico do retorno acumulado dos modelos comparados com *Buy & Hold* 36

Abreviaturas

AI / IA – Artificial Intelligence (Inteligência Artificial)

ARIMA – *Autoregressive Integrated Moving Average*

CNN – *Convolutional Neural Network*

CVaR – *Conditional Value at Risk*

DJIA – *Dow Jones Industrial Average*

DQL – *Deep Q-Learning*

DQN – *Deep Q-Network*

DRL – *Deep Reinforcement Learning*

ETF – *Exchange-Traded Fund*

GA – *Genetic Algorithm*

GC=F – *Gold Futures* (Commodity Exchange)

GRU – *Gated Recurrent Unit*

HDRL-Trader – *Hybrid Deep Reinforcement Learning Trader*

KNN – *K-Nearest Neighbors*

LSTM – *Long Short-Term Memory*

MAPE – *Mean Absolute Percentage Error*

MAE – *Mean Absolute Error*

MAR – *Mean Absolute Relative Error*

MSE – *Mean Squared Error*

ML – *Machine Learning*

MLP – *Multilayer Perceptron*

NVDA – ticker do NVIDIA Corporation

P-DDQN – *Pairs Trading Double Deep Q-Network*

PER – *Prioritized Experience Replay*

PPO – *Proximal Policy Optimization*

PT-DQN – *Pairs Trading Deep Q-Network*

Q-Learning (QL) – *Reinforcement Learning Q-Learning*

QML – *Quantum Machine Learning*

QA3C – *Quantum Asynchronous Advantage Actor-Critic*

QLSTM – *Quantum Long Short-Term Memory*

RNN – *Recurrent Neural Network*

RMSE – *Root Mean Squared Error*

RSI – *Relative Strength Index*

S&P 500 – *Standard & Poor's 500 Index*

SVM – *Support Vector Machine*

VaR – *Value at Risk*

VIX – *Volatility Index*

WTO – *West Texas Intermediate (Petróleo Bruto)*

XGBoost – *Extreme Gradient Boosting*

1. Introdução

1.1. Contexto Geral

Nas últimas décadas, os mercados de ações consolidaram-se como uma componente fundamental da economia global. As suas oscilações afetam não apenas o desempenho financeiro de empresas e investidores, mas também a estabilidade económica de países. Para além de financiar o crescimento empresarial, o mercado acionista representa uma via de investimento amplamente utilizada, especialmente devido à possibilidade de obtenção de retornos significativos. No entanto, este retorno está intrinsecamente associado a um grau de risco considerável, resultante da elevada volatilidade, da complexidade dos fatores intervenientes e da imprevisibilidade do comportamento dos investidores (Ananthi et al., 2021).

Neste contexto, torna-se evidente a importância do desenvolvimento de modelos preditivos capazes de antecipar os movimentos do mercado, contribuindo para uma tomada de decisão mais informada e estratégica (Fama, 1995). A previsão da tendência do mercado acionista constitui, no entanto, um desafio exigente, atendendo à natureza não estacionária dos dados, à presença de ruído, à elevada incerteza e às relações complexas entre variáveis (Zhong et al., 2017).

A capacidade de antecipar movimentos de mercado com maior precisão representa uma vantagem competitiva significativa, permitindo otimizar decisões de investimento em contextos financeiros marcados por elevada volatilidade e complexidade.

Nos últimos anos, os modelos utilizados na previsão de estratégias de *trading* evoluíram de forma expressiva, acompanhando os progressos tecnológicos e metodológicos no campo da ciência de dados. A aprendizagem supervisionada tem-se destacado pela sua capacidade de lidar com grandes volumes de dados e identificar padrões complexos, enquanto a aprendizagem por reforço tem demonstrado potencial para lidar com problemas sequenciais de decisão sob incerteza. Esta última revelou-se especialmente promissora no domínio do *trading* automatizado, onde a tomada de decisão sequencial e a adaptação a contextos dinâmicos são essenciais (Sutton & Barto, 2018).

Combinando estas duas abordagens, os modelos híbridos surgem como uma solução promissora, ao integrar a capacidade preditiva dos modelos supervisionados com a adaptabilidade dos métodos por reforço. Esta integração visa explorar as vantagens complementares de ambas as abordagens. No caso específico desta investigação, os modelos foram aplicados ao *trading* do ativo financeiro NVIDIA, permitindo avaliar o seu desempenho num mercado altamente volátil e competitivo.

Assim, esta dissertação tem como objetivo principal desenvolver e avaliar modelos híbridos baseados em aprendizagem por reforço e supervisionada, aplicados ao *trading* com especial enfoque em ativos do setor tecnológico. A investigação propõe-se analisar diferentes algoritmos utilizados nesta problemática, procurando identificar as

abordagens que têm demonstrado melhor desempenho em termos de previsão de tendências, gestão de risco e rentabilidade.

1.2. Motivação e Objetivos

A presente investigação surge da necessidade de desenvolver abordagens inovadoras para enfrentar os desafios inerentes à previsão financeira e o desenvolvimento de estratégias de *trading* eficazes, particularmente no contexto altamente dinâmico dos ativos tecnológicos.

Estes ativos têm-se destacado por apresentar elevados níveis de volatilidade, forte crescimento e sensibilidade significativa de fatores económicos e geopolíticos. Tais características tornam evidente a necessidade de modelos preditivos mais sofisticados e adaptativos, capazes de captar padrões complexos e responder adequadamente às rápidas mudanças de contexto.

Neste enquadramento, os modelos híbridos assumem um papel central. A combinação de aprendizagem supervisionada e aprendizagem por reforço oferece benefícios que dificilmente são alcançáveis quando estas metodologias são utilizadas isoladamente. A aprendizagem supervisionada, ao fornecer previsões explícitas sobre retornos ou tendências, contribui para melhorar a capacidade de antecipação do agente, enquanto a aprendizagem por reforço permite otimizar decisões sequenciais considerando o impacto cumulativo das ações no tempo. Esta combinação torna os modelos híbridos particularmente úteis em ambientes financeiros caracterizados por elevada incerteza, regimes de mercado variáveis e relações não lineares entre variáveis.

Adicionalmente, pretende-se analisar o impacto da incorporação de variáveis exógenas - como o índice do dólar (DXY), o índice de volatilidade (VIX), bem como os preços dos contratos futuros de ouro negociados na *Commodity Exchange* (GC=F) e do petróleo bruto WTI (*West Texas Intermediate*) negociados na *New York Mercantile Exchange* (CL=F) - na performance preditiva e na gestão do risco. Será igualmente considerado um modelo que integra uma estratégia de controlo de risco baseada em *stop-loss* e *trailing stop* com o objetivo de aferir a eficácia de mecanismos simples de proteção em ambientes voláteis.

Assim, os modelos híbridos não só combinam diferentes fontes de informação, como permitem aproveitar as vantagens específicas do *XGBoost* na previsão e da aprendizagem por reforço na otimização sequencial das decisões. Através da comparação entre um modelo base de aprendizagem por reforço e as versões híbridas desenvolvidas, será possível avaliar de forma rigorosa os benefícios destas estratégias em termos de rentabilidade, estabilidade e capacidade de adaptação ao mercado.

Através da comparação entre um modelo base de aprendizagem por reforço e os modelos híbridos, será possível avaliar os benefícios das diferentes estratégias em termos de rentabilidade, estabilidade e capacidade de adaptação ao mercado. Com esta investigação, pretende-se contribuir para a área do *trading* quantitativo através do

desenvolvimento de um modelo robusto, facilmente replicável e adequado às exigências dos mercados financeiros atuais.

1.3. Estrutura

A presente dissertação está organizada em cinco capítulos, que se articulam de forma a proporcionar um percurso coerente desde o enquadramento teórico até à implementação prática e análise dos resultados.

O **Capítulo 1 - Introdução** apresenta o contexto geral da investigação, a motivação subjacente, os objetivos do estudo e a estrutura do documento.

O **Capítulo 2 - Revisão da Literatura** examina a evolução dos métodos de modelação preditiva aplicados ao *trading*, com especial enfoque nos métodos de classificação de tendência, previsão de preços, abordagens híbridas e *ensemble*. Inclui ainda uma análise comparativa dos trabalhos mais relevantes, oferecendo o enquadramento teórico necessário ao desenvolvimento da investigação.

O **Capítulo 3 - Metodologia** descreve o enquadramento técnico da investigação, detalhando os dados utilizados, os fundamentos dos modelos adotados, incluindo o *Q-learning*, o *Deep Q-Network* e o *XGBoost*, bem como a conceção dos modelos híbridos. Apresenta ainda as diferentes variantes propostas, as variáveis exógenas, assim como mecanismos de controlo de risco.

O **Capítulo 4 – Análise de Resultados** apresenta as métricas de desempenho utilizadas, os resultados obtidos com cada modelo e uma análise comparativa aprofundada. São discutidas as vantagens e limitações de cada abordagem, bem como as suas implicações práticas no contexto do *trading*.

Por fim, o **Capítulo 5 – Conclusão** sintetiza os principais achados do estudo, identifica as contribuições científicas e práticas da investigação e propõe possíveis direções para trabalhos futuros.

Esta organização visa assegurar um percurso lógico, fundamentado e progressivo, permitindo uma compreensão clara do problema investigado, das soluções propostas e dos contributos alcançados no domínio do *trading* quantitativo com modelos híbridos de aprendizagem por reforço.

2. Revisão da Literatura

2.1. Evolução na Modelação Preditiva em *Trading*

Nas últimas décadas, a modelação preditiva aplicada ao *trading* registou uma evolução significativa, impulsionada pelos avanços na inteligência artificial, no poder computacional e na disponibilidade de dados financeiros. Inicialmente centradas em abordagens estatísticas tradicionais, como regressões lineares e modelos *Autoregressive Integrated Moving Average* (ARIMA), as estratégias de previsão foram gradualmente substituídas ou complementadas por técnicas mais sofisticadas, capazes de captar padrões não lineares e complexas interações entre variáveis. Desde os modelos estatísticos tradicionais até as redes neurais avançadas, como as *Long Short-Term Memory* (LSTM), bem como os modelos *ensemble* e híbridos, tem-se observado um aumento substancial no desempenho dessas técnicas para prever valores e tendências de mercado (Yinget al., 2024).

Estudos demonstram que modelos de aprendizagem automática, como as *Recurrent Neural Network* (RNN) e LSTM, têm mostrado elevado potencial em aplicações de *trading* devido à sua capacidade de lidar com dados sequenciais e dinâmicos (Selvin, 2017). Chen et al. (2016) apresentaram o *XGBoost* como uma evolução mais avançada e escalável do *gradient boosting*, destacando melhorias como o suporte a dados esparsos, processamento paralelo e a inclusão de regularização (Zhang et al., 2018). Na área de *trading*, o *XGBoost* tem sido amplamente utilizado, graças ao seu desempenho superior em comparação com modelos tradicionais e menor tempo de execução comparado com o LSTM.

Com o surgimento da aprendizagem por reforço e dos modelos híbridos, inicialmente projetados para tarefas como o processamento de linguagem natural (Kim et al., 2022), a previsão em *trading* também experimentou avanços notáveis. O sucesso de modelos *ensemble* e híbridos em áreas distintas despertou grande interesse na comunidade de finanças e *trading*, incentivando pesquisadores a adaptarem essas arquiteturas para lidar com os desafios do mercado financeiro (Stempień & Ślepaczuk, 2025; Darwish, 2025).

Neste contexto, esta revisão estrutura-se em três secções principais: a primeira aborda os métodos utilizados na **classificação da tendência dos preços**, destacando os algoritmos de *machine learning* (ML) que categorizam ações em termos de decisão de compra, manutenção ou venda; a segunda foca-se nos **modelos voltados à previsão de preços e retornos**, com ênfase nos avanços em redes neurais recorrentes, nomeadamente LSTM e variantes; por fim, a terceira secção analisa os **métodos híbridos e *ensemble***, que combinam diferentes abordagens com o intuito de melhorar o desempenho preditivo e a robustez das estratégias de *trading*.

2.2. Métodos na classificação da tendência

No estudo realizado por Caparrini et al. (2024), os autores exploram a eficácia de diferentes algoritmos de *machine learning* na seleção de ações do índice S&P 500, recorrendo a dados diários do período entre 2005 e 2019. O objetivo principal da pesquisa é comparar o desempenho de três modelos distintos: *Decision Trees*, *Random Forest* e *XGBoost*.

A análise comparativa revelou que o *XGBoost* superou os modelos *Decision Tree* e *Random Forest* em várias métricas de desempenho. O *XGBoost* alcançou um retorno total de 502,16% e um retorno anualizado de 11,26%, superando o *Random Forest* e o *Decision Tree*. Embora apresente uma volatilidade anual ligeiramente superior em comparação aos outros modelos, o *XGBoost* compensa com um *Sharpe Ratio* mais elevado (0,29), indicando uma melhor relação retorno-risco. Além disso, o *XGBoost* demonstrou uma precisão média de 40,4%, superior ao *Random Forest* (29,3%) e ao *Decision Tree* (29,2%), evidenciando sua eficácia na identificação de ações com potencial de valorização no S&P 500.

O modelo *XGBoost* tem evidenciado ser um dos melhores algoritmos para problemas de classificação encontrados na literatura, sendo que supera os outros existentes e é capaz de prever a longo prazo (Dey et al., 2016).

Wang et al. (2023) propuseram um modelo preditivo para a classificação de tendências no mercado acionista, integrando dados financeiros e informações de sentimento extraídas de notícias e redes sociais. O estudo utilizou dados de 2014 a 2018 referentes a empresas como Google, Amazon, Apple, e ao índice DJIA, recorrendo a fontes como o *Yahoo Finance* e a API do New York Times. O modelo desenvolvido teve como objetivo prever as tendências do mercado, classificando as ações nas categorias "comprar", "manter" ou "vender".

Na comparação entre modelos preditivos de *machine learning* e arquiteturas de redes neuronais de aprendizagem profunda, os resultados indicaram que não há uma vantagem significativa associada aos algoritmos de aprendizagem profunda. Os modelos tradicionais de *machine learning* demonstraram ser capazes de alcançar desempenhos competitivos, mesmo em tarefas consideradas relativamente simples. Dentre os modelos analisados, a *Rede Neural Convolucional* (CNN) destacou-se como o mais eficaz na previsão de tendências de ações, seguido de perto pelo modelo *XGBoost*, que apresentou uma leve diminuição de 0,72 pontos percentuais na acurácia em relação ao CNN.

No estudo de Ampomah et al. (2020), os autores avaliaram a eficácia de seis modelos de aprendizagem de máquina para prever a direção dos preços de ações, utilizando dados de 2005 a 2019. Os modelos analisados foram *AdaBoost*, *Random Forest*, *XGBoost*, *Bagging Classifier*, *Extra Trees* e *Voting Classifier*. As ações incluídas na análise foram *Bank of America* (BAC), *Exxon Mobil* (XOM), *Microsoft* (MSFT), *CarMax* (KMX), *Tata Steel* (TATASTEEL), *HCL Technologies* (HCLTECH), além dos índices S&P 500 e *Dow Jones* (DJIA).

O *AdaBoost* apresentou a melhor acurácia média no conjunto de treino, destacando-se em diversos ativos, como BAC, DJIA e TATASTEEL. O *XGBoost* revelou desempenho consistente, com acurácia entre 81,22% e 89,91%, e precisão entre 82,42% e 88,22%. Já o *Extra Trees* registou a melhor acurácia média em testes, particularmente nos dados do DJIA e HCLTECH.

A análise foi complementada com métricas como precisão, *recall*, *F1-score*, especificidade e AUC-ROC, evidenciando que, embora o *AdaBoost* se destaque na acurácia, o *Extra Trees* apresenta melhor desempenho noutras métricas. Os autores sublinham a importância de uma avaliação multidimensional dos modelos e sugerem, para estudos futuros, a inclusão de abordagens mais avançadas, como processos Gaussianos, técnicas de regularização e métodos baseados em *kernel*. Para resultados adicionais e detalhes comparativos, ver a *Tabela 4* nos Anexos.

2.3. Métodos na previsão de preço

Diversos estudos recentes têm demonstrado a eficácia dos modelos baseados em redes LSTM na previsão de preços de ações e dos respetivos retornos. No entanto, o seu desempenho depende fortemente da adequada configuração dos hiperparâmetros, como evidenciado por Hutter et al. (2014).

Bhandari et al. (2022) analisaram o desempenho de LSTM com uma e múltiplas camadas na previsão do preço de fecho diário do índice S&P 500, utilizando dados de 2006 a 2020. O modelo de camada única obteve melhores resultados, especialmente quando complementado com variáveis macroeconómicas. O melhor desempenho foi registado com o otimizador *Adam*, apresentando uma Raiz do Erro Quadrático Médio (RMSE) de 28,65 para uma configuração com 10 neurónios, taxa de aprendizagem de 0,001 e *batch size* de 8.

Zou et al. (2022) compararam variantes de LSTM — incluindo *Stacked* LSTM e *Attention-Based* LSTM — com o modelo ARIMA tradicional, usando dados financeiros de empresas do S&P 500. As variáveis de entrada incluíram indicadores como volume de negociação, retorno sobre o património e margem de lucro. O modelo *Attention*-LSTM obteve os melhores resultados, evidenciando maior capacidade de maximizar retornos em estratégias de *trading*, superando consistentemente os demais modelos avaliados.

Sethia et al. (2019), desenvolveram um modelo inovador para a previsão do preço das ações com o objetivo de possibilitar uma estratégia diária de compra e venda para o Índice S&P 500. O estudo focou na previsão do preço das ações para o quinto dia seguinte, utilizando um conjunto abrangente de dados que se estendeu por mais de 12 anos, entre 2000 e 2017. Durante esse período, foram coletados e processados dados financeiros, incluindo preços de abertura, máxima, mínima, fechamento e volume, além de 50 indicadores técnicos adicionais que ajudam a filtrar ruídos momentâneos nos dados.

Os modelos utilizados na pesquisa incluíram *Long Short Term Memory* (LSTM), *Gated Recurrent Unit* (GRU), *Support Vector Machines* (SVM) e *Multilayer Perceptron*

(MLP). A pesquisa revelou que o modelo LSTM apresentou um desempenho superior em comparação com os demais, alcançando um erro médio quadrático (RMSE) de 0.000428.

Ao contrário de Sethia et al. (2019), que adotaram abordagens mais tradicionais, Awoke et al. (2020) compararam modelos de aprendizagem profunda, nomeadamente LSTM e GRU, na previsão do preço do Bitcoin. Utilizando dados entre janeiro de 2014 e fevereiro de 2018, o estudo incorporou sete variáveis, como preço de abertura, máximo, mínimo e fecho. A avaliação dos modelos foi realizada com base no Erro Percentual Absoluto Médio (MAPE) e no RMSE, para previsões de 1, 3, 5, 7 e 15 dias, com diferentes tamanhos de janela. Os resultados indicaram que o GRU apresentou melhor desempenho em contextos de elevada volatilidade, revelando-se mais adequado para séries temporais com forte variação.

Num estudo mais recente, Gupta et al. (2023) investigaram a previsão de preços de ações com recurso a Regressão Linear Multicamada, Redes Neuronais Convolucionais (CNN) e LSTM. A análise foi conduzida sobre dados diários de preços históricos, entre 2012 e 2022, referentes a cinco empresas de diferentes setores. O modelo LSTM obteve os melhores resultados na previsão do preço de fecho, demonstrando superioridade face aos restantes métodos, independentemente do comprimento da sequência utilizada. Para informações complementares e resultados detalhados, consultar a *Tabela 5* nos Anexos.

2.4. Métodos Híbridos e *Ensemble*

Vários estudos têm comparado o desempenho de modelos híbridos e *ensemble* face a modelos individuais, evidenciando ganhos significativos em termos de precisão preditiva quando técnicas combinadas são utilizadas (Chou et al., 2013). Enquanto os modelos *ensemble* combinam previsões de múltiplos algoritmos de forma independente, os modelos híbridos integram diferentes métodos num único processo preditivo.

Tong et al. (2021) testaram diferentes abordagens na previsão do movimento de preços de ações de seis empresas de setores distintos, incluindo Apple, Johnson & Johnson e JPMorgan, com dados entre 2016 e 2023. Foram comparados os modelos ARIMA, LSTM, *XGBoost* e um modelo híbrido composto por uma rede neural com mecanismo de atenção e um *XGBoost*. Embora *ARIMA* e *LSTM* tenham apresentado desempenhos semelhantes em termos de erro quadrático médio (MSE) e erro absoluto médio relativo (*MAR*), o modelo híbrido superou consistentemente os restantes. Os resultados demonstraram que a combinação de um componente de aprendizagem profunda com atenção e a robustez do *XGBoost* proporcionou melhor desempenho preditivo em todas as empresas analisadas.

Li et al. (2020) propuseram uma abordagem híbrida que combina redes neurais convolucionais (CNN) com redes LSTM para prever o preço do *Bitcoin*. O modelo utiliza dados históricos da criptomoeda, variáveis macroeconómicas e indicadores de atenção do investidor. A CNN é responsável pela extração de características relevantes, enquanto a LSTM realiza a previsão. Os resultados, avaliados com base no Erro Absoluto Médio (MAE), RMSE e MAPE, demonstraram que o modelo CNN-LSTM superou os modelos

individuais, revelando-se mais adequado para prever a dinâmica do mercado de criptomoedas.

Yun et al. (2021) desenvolveram um modelo *ensemble* combinando Algoritmos Genéticos (GA) para seleção de características com *XGBoost*, aplicado ao índice de ações sul-coreano entre 2000 e 2016. O modelo previu a direção do mercado no dia seguinte e foi comparado com diversos algoritmos, como *Decision Tree*, *Extra Trees*, *Random Forest*, SVM e *K-Nearest Neighbors* (KNN). O GA-*XGBoost* apresentou o melhor desempenho, com uma taxa de precisão de 93,28%.

Kim et al. (2022) analisaram 20 ações do S&P 500 entre 2010 e 2020, comparando cinco estratégias de *trading*, incluindo modelos baseados em aprendizagem por reforço, *Pairs Trading Deep Q-Network (PT-DQN)* e *Pairs Trading Double Deep Q-Network (P-DDQN),* uma estratégia passiva (*Buy and Hold*), e dois modelos híbridos: *P-Trader* e *HDRL-Trader*. Os modelos híbridos *P-Trader* e *HDRL-Trader* diferem principalmente pelo *HDRL-Trader* incorporar técnicas adicionais de redução de dimensionalidade, *clustering*, aprendizagem de representação de estados e um mecanismo de *stop-loss*, o que lhe confere maior robustez e desempenho superior na gestão de risco e retorno. O *HDRL-Trader* alcançou a maior taxa média de retorno (82,4%), superando em 25,7% a segunda melhor abordagem, e demonstrou robustez na gestão de risco.

Esta combinação torna o *HDRL-Trader* particularmente robusto e eficaz em diferentes condições de mercado, proporcionando um desempenho superior tanto em termos de retorno como de controlo de risco. Trata-se do único modelo que integra de forma simultânea técnicas avançadas, como estrutura de *gate*, modelo de regressão, clonagem de comportamento, *Prioritized Experience Replay* (PER), atraso dinâmico e aprendizagem por reforço híbrida.

Mais recentemente, Aweda et al. (2025) apresentaram um modelo híbrido direcionado ao mercado cambial (*Foreign Exchange – Forex*), no qual integraram uma arquitetura *Convolutional Neural Network – Long Short-Term Memory (CNN–LSTM)* para a previsão de preços e um algoritmo de *Deep Q-Learning (DQL)* para a definição de estratégias de *trading*. O estudo baseou-se em dados históricos do par Euro/Dólar Americano (EUR/USD) entre 2014 e 2024, complementados por indicadores técnicos, como o *Relative Strength Index (RSI)*, o *Moving Average Convergence Divergence (MACD)* e médias móveis.

Os resultados evidenciaram que a componente CNN–LSTM apresentou um desempenho superior face a modelos tradicionais, como o *Autoregressive Integrated Moving Average (ARIMA)* e o LSTM isolado, registando valores inferiores de Raiz do Erro Quadrático Médio (*Root Mean Squared Error – RMSE*) (0,0025) e de Erro Absoluto Médio (*Mean Absolute Error – MAE*) (0,0017). Paralelamente, o agente DQL obteve um retorno acumulado de 49,2% e um *Sharpe Ratio* de 2,87, superando metodologias baseadas em regras, como as médias móveis ou a estratégia *Buy-and-Hold*.

Em síntese, os resultados demonstraram o potencial das abordagens híbridas que combinam Aprendizagem Profunda e Aprendizagem por Reforço (DL–RL) em contextos de elevada volatilidade e não linearidade, como é o caso do Forex, proporcionando simultaneamente maior precisão preditiva e estratégias de *trading* mais adaptativas.

Por sua vez, Chen et al. (2025) introduziram uma abordagem inovadora ao explorarem a aplicação de *Quantum Machine Learning (QML)* em modelos híbridos de *trading*. Os autores desenvolveram um agente que articula uma rede *Quantum Long Short-Term Memory (QLSTM)*, destinada à previsão de tendências de curto prazo, com o modelo *Quantum Asynchronous Advantage Actor-Critic (QA3C)*, responsável pela tomada de decisão. Este agente foi treinado com dados diários do par Dólar Americano/Novo Dólar Taiwanês (USD/TWD) no período compreendido entre 2000 e 2025, obtendo, na fase de teste (2020–2025), um retorno total de 11,87% e um *drawdown* máximo de apenas 0,92%, superando, ou pelo menos igualando, o desempenho de diversos *Exchange-Traded Funds (ETF)* cambiais.

O QLSTM desempenhou um papel relevante na extração de características quânticas, permitindo uma representação mais eficiente dos estados e reduzindo a complexidade amostral exigida pelo agente de reforço. O QA3C, por sua vez, beneficiou da utilização de camadas variacionais quânticas, o que lhe conferiu maior eficiência em termos de parâmetros face ao modelo clássico A3C. Estes resultados reforçam a relevância das abordagens híbridas quântico-clássicas, que, mesmo em ambiente de simulação, já evidenciam ganhos significativos no *trading* algorítmico, antecipando o impacto que a futura implementação em *hardware* quântico poderá proporcionar (Chen et al., 2025). Para uma síntese comparativa dos modelos híbridos e ensemble analisados, ver a *Tabela 6* nos Anexos.

2.5. Análise comparativa

No domínio da classificação de tendências, modelos como o *XGBoost*, *AdaBoost* e *Extra Trees* se destacaram em diversas métricas de desempenho, especialmente no retorno total, acurácia e relação risco-retorno. A análise comparativa dos algoritmos, como evidenciado pelos estudos de Caparrini et al. (2024) e Ampomah et al. (2020), demonstrou que, embora os modelos baseados em árvores como o *Random Forest* e o *Decision Tree* tenham mostrado um desempenho competitivo, o *XGBoost* obteve resultados superiores, principalmente devido ao seu *Sharpe Ratio* mais elevado e maior precisão na classificação das ações do S&P 500.

Por outro lado, na previsão de preços de ações, os modelos LSTM e suas variantes, como o *Stacked* LSTM e o *Attention*-LSTM, mostraram-se extremamente eficazes, superando modelos tradicionais, como o ARIMA, em termos de precisão e capacidade de captura de padrões temporais complexos. Estudos como o de Bhandari et al. (2022) e Zou et al. (2022) indicam que o LSTM, especialmente quando ajustado corretamente em termos de hiperparâmetros, pode fornecer previsões precisas, mesmo quando comparado com modelos mais simples.

No entanto, a pesquisa de Awoke et al. (2020) e Gupta et al. (2023) sugeriu que modelos híbridos, como CNN-LSTM e *Attention*-LSTM, podem melhorar ainda mais a acurácia ao combinar diferentes técnicas de aprendizado profundo, mostrando uma tendência crescente no uso de abordagens híbridas para lidar com a complexidade dos dados financeiros.

Os métodos híbridos e *ensemble*, que combinam múltiplos modelos, têm demonstrado vantagens significativas em comparação com modelos isolados. Tong et al. (2021) evidenciaram isso ao usar *XGBoost* em conjunto com mecanismos de atenção, enquanto Li et al. (2020) aplicaram uma combinação de CNN e LSTM para prever o preço do Bitcoin. Além disso, o modelo *GA-XGBoost* de Yun et al. (2021) obteve resultados substancialmente superiores em termos de métricas de erro e precisão em relação aos modelos individuais.

Além disso, a aplicação de métodos de aprendizagem por reforço híbrida, como exemplificado no artigo de Kim et al. (2022) com o *HDRL-Trader*, trouxe avanços significativos ao integrar técnicas como *clustering* e aprendizagem de representação de estado. Essa abordagem, ao lidar com a complexidade do mercado financeiro, não só demonstrou maior robustez em diferentes condições de mercado, como também garantiu maior rentabilidade, mesmo com custos elevados de transação. O *HDRL-Trader* também se mostrou eficaz em resolver desafios críticos, como a dependência entre fronteiras de *stop-loss* e *trading*, evidenciando seu impacto na otimização de estratégias de *trading.*

Os estudos mais recentes ampliam esta tendência ao explorar novos horizontes metodológicos. Aweda et al. (2025) evidenciaram que a integração de *Convolutional Neural Network – Long Short-Term Memory (CNN–LSTM)* com *Deep Q-Learning (DQL)* melhora a precisão preditiva e o desempenho em *trading* no mercado cambial, destacando-se em métricas de erro, retorno e risco. Já Chen et al. (2025) introduziram a dimensão da *Quantum Machine Learning (QML)*, propondo um modelo híbrido que combina *Quantum LSTM (QLSTM)* e *Quantum Asynchronous Advantage Actor-Critic (QA3C)*, capaz de alcançar retornos estáveis e baixos níveis de *drawdown*, antecipando o potencial dos modelos quântico-clássicos para o *trading* algorítmico.

Em síntese, verifica-se uma dupla evolução: de um lado, o refinamento das arquiteturas híbridas baseadas em Aprendizagem Profunda e Aprendizagem por Reforço; de outro, a emergência de abordagens quânticas, que prometem ganhos adicionais em eficiência e robustez, consolidando a ideia de que o futuro do *trading* automatizado passará pela combinação de paradigmas distintos.

3. Metodologia

3.1. Contexto Geral

A NVIDIA *Corporation*, fundada em 1993 e com sede em Santa Clara, Califórnia, afirma-se como uma das empresas mais influentes do sector tecnológico a nível global. Reconhecida como pioneira no desenvolvimento de unidades de processamento gráfico (GPU), a empresa lidera soluções inovadoras aplicadas a domínios como os videojogos, a inteligência artificial (IA), os veículos autónomos e os centros de dados.

Em termos financeiros, no segundo trimestre de 2023, a empresa registou receitas de 13,51 mil milhões de dólares, representando um crescimento de 101 % face ao período homólogo, impulsionado principalmente pelo segmento de centros de dados, que gerou 10,32 mil milhões de dólares (Chen, 2024).

A NVIDIA apresenta uma rentabilidade sustentada e uma posição de liderança consolidada no mercado. Segundo a *Jon Peddie Research*, entre 2022 e 2024, a empresa deteve uma quota de 88 % no segmento de GPUs discretas, reforçando a sua posição quase monopolista face a concorrentes como a AMD e a Intel.

A escolha da NVIDIA como objecto de análise justifica-se pela sua posição estratégica na vanguarda da inovação tecnológica, pelo desempenho financeiro consistente e pela sua influência em áreas emergentes como a inteligência artificial generativa e a computação acelerada. Estes factores tornam-na um caso representativo para a análise do sector tecnológico atual.

3.2 Análise de Dados

Nesta secção são descritos os procedimentos utilizados na obtenção, tratamento e exploração inicial dos dados referentes à ação da NVIDIA, que constituem a base dos modelos desenvolvidos no âmbito da presente dissertação.

Os dados históricos da ação da NVIDIA foram extraídos da plataforma *Yahoo Finance*, abrangendo o período de 1 de janeiro de 2014 a 30 de setembro de 2024. Os dados correspondem a cotações diárias referentes a dias úteis de negociação (*business days*). A seleção deste intervalo temporal justifica-se pela relevância dos acontecimentos económicos e tecnológicos ocorridos ao longo deste período, permitindo uma análise abrangente do comportamento da ação em diferentes contextos do mercado. Destacam-se, particularmente, os seguintes eventos:

- A crise do petróleo entre meados de 2014 e o início de 2016, caracterizada por uma queda acentuada nos preços das matérias-primas, com impacto significativo em diversos setores económicos.

- O período de recuperação económica entre 2017 e 2019, assinalado por um crescimento sustentado e uma valorização expressiva das ações do setor tecnológico.
- A pandemia de COVID-19, a partir de março de 2020, que inicialmente provocou uma forte queda no mercado acionista, seguida de uma recuperação impulsionada pela procura de soluções tecnológicas associadas ao teletrabalho e ao entretenimento digital.
- A crise inflacionária entre 2022 e 2023, acompanhada pelo aumento das taxas de juro por parte dos principais bancos centrais, especialmente pela Reserva Federal dos Estados Unidos, resultou numa correção significativa nos preços das ações tecnológicas.
- O crescimento acelerado da inteligência artificial ao longo de 2023 e no início de 2024, que impulsionou uma valorização notável da ação da NVIDIA, dada a sua posição dominante no mercado de processadores gráficos utilizados em aplicações de IA.

Os dados descarregados apresentam uma estrutura composta pelas seguintes variáveis principais:

- **Open:** preço de abertura da ação no dia.
- **High:** preço máximo alcançado durante o dia.
- **Low:** preço mínimo alcançado durante o dia.
- **Close:** preço de fecho da ação no dia.
- **Volume:** volume total de ações negociadas no dia.

A base de dados recolhida contém mais de 2.700 observações. Após uma análise preliminar, confirmou-se a inexistência de valores omissos ou inconsistentes, dispensando qualquer procedimento adicional de pré-processamento nesse domínio.

Com o intuito de caracterizar a volatilidade associada à ação, foram calculadas estatísticas descritivas dos preços de fecho, sintetizadas na tabela seguinte:

Tabela 1 - Estatísticas descritivas dos preços de fecho

Estatística	Valor
Média	16,08 USD
Desvio Padrão	25,68 USD
Valor mínimo	0,36 USD
1º Quartil (25%)	1,54 USD
Mediana (50%)	5,76 USD
3º Quartil (75%)	17,90 USD
Valor máximo	135,54 USD

Os indicadores acima revelam uma elevada dispersão dos preços, com diferenças acentuadas entre os valores mínimos e máximos, indicando uma elevada volatilidade. A diferença substancial entre a média e a mediana sugere uma distribuição enviesada à direita, reflexo da valorização expressiva da ação nos últimos anos.

Complementarmente, foi realizada uma análise gráfica dos preços de fecho e dos retornos diários. A Figura 1 ilustra a evolução temporal do preço de fecho da ação NVIDIA ao longo do período em estudo. Observa-se um claro impacto evidente dos eventos anteriormente destacados na volatilidade e nas tendências de longo prazo.

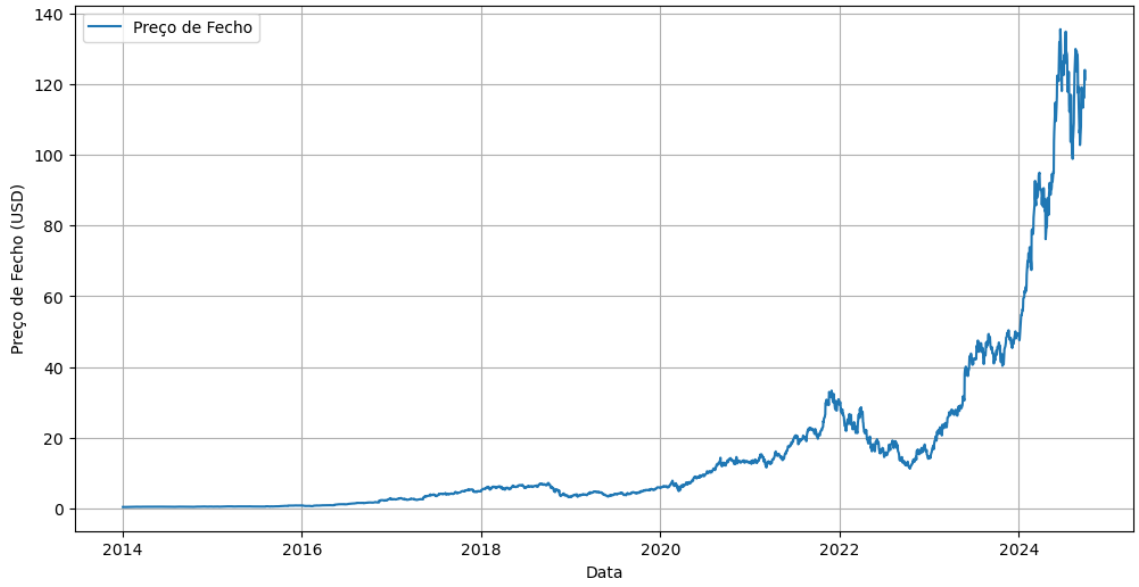


Figura 1 - Evolução histórica dos preços diários de fecho da NVIDIA (2014–2024)

A Figura 2 mostra os retornos diários, permitindo identificar períodos de maior dispersão, sobretudo durante crises globais e eventos macroeconómicos relevantes, como a pandemia COVID-19 e tensões comerciais internacionais.

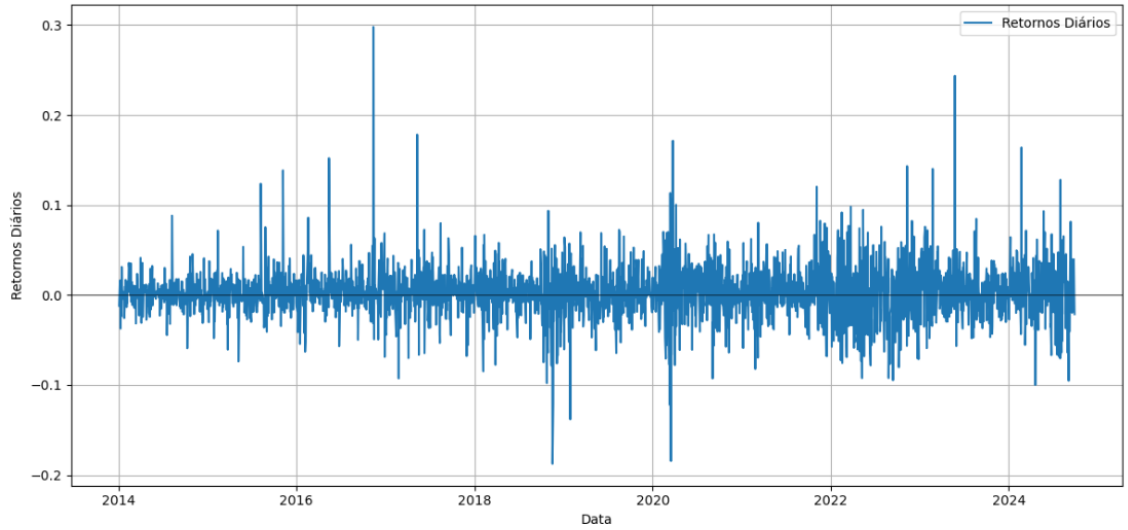


Figura 2 - Retornos Diários da NVIDIA (2014–2024)

Através de *boxplots*, identificaram-se *outliers* significativos, sobretudo na variável *Volume*, fenómeno habitual em séries temporais de mercados financeiros, em particular durante períodos de elevada atividade, como anúncios de resultados ou eventos macroeconómicos relevantes. Devido à natureza específica deste domínio, tais valores não são considerados anómalos no sentido tradicional, pelo que não foram aplicados procedimentos de remoção ou correção de *outliers*, como é comum noutras áreas.

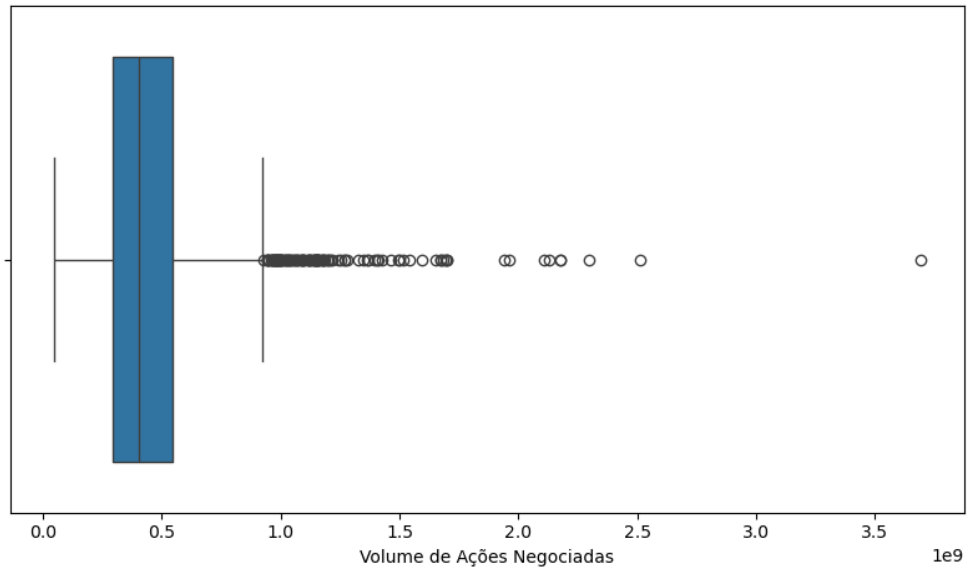


Figura 3 - Detecção de outliers no Volume negociado NVIDIA (2014–2024)

Por fim, procedeu-se à divisão dos dados em dois conjuntos fundamentais: treino e teste. O conjunto de treino compreende o período entre 1 de janeiro de 2014 e 21 de junho de 2022, enquanto o conjunto de teste abrange o intervalo de 22 de junho de 2022 a 30 de setembro de 2024. Esta divisão possibilita uma análise mais clara do desempenho preditivo e da estabilidade dos modelos apresentados nos capítulos seguintes.

3.3 Fundamentos dos modelos utilizados

A presente secção descreve os princípios teóricos que sustentam os modelos aplicados neste trabalho, cobrindo tanto abordagens de aprendizagem por reforço como de aprendizagem supervisionada. Dentro deste paradigma, serão explorados o *Q-Learning*, na sua forma tabular, e a sua extensão baseada em redes neuronais profundas, o *Deep Q-Network* (DQN), ambos capazes de aprender políticas ótimas através da interação com o ambiente. Em paralelo, recorre-se ao *XGBoost*, um modelo supervisionado reconhecido pela sua eficiência e capacidade de capturar relações não lineares complexas. A integração destas técnicas em modelos híbridos visa explorar sinergias entre previsão supervisionada e otimização sequencial de decisões, criando estratégias mais robustas e adaptáveis às condições do mercado.

3.3.1 Aprendizagem por Reforço

Um processo de decisão de *Markov* (*Markov Decision Process* - MDP) constitui uma formalização matemática de problemas de decisão sequenciais, nos quais as ações tomadas influenciam não apenas as recompensas imediatas, mas também a trajetória do sistema. Assim, a tomada de decisão deve considerar não só o retorno imediato, mas igualmente o impacto a médio e longo prazo (Sutton e Barto, 2008).

Um MDP finito com horizonte infinito pode ser especificado pelo conjunto ($\mathcal{S}$, $\mathcal{A}$, $\mathcal{R}$, $\mathcal{P}$, γ), onde t = 0, 1,2,…, e onde:

- $\mathcal{S}$ é o conjunto finito de estados do sistema;
- $\mathcal{A}$ é o conjunto finito de ações possíveis;
- $\mathcal{R}$ é o conjunto finito de recompensas;
- $\mathcal{P}$: $\mathcal{S}$ x $\mathcal{R}$ x $\mathcal{S}$ x $\mathcal{A} \rightarrow [0, 1]$ é a função que define a probabilidade conjunta de transição para um determinado estado e de obtenção de uma dada recompensa, condicionada pelo estado anterior do sistema e pela ação executada.

$$\mathcal{P}(s_{t+1}, r_{t+1} | s_t, a_t) = \mathbb{P}(S_{t+1}, \mathrm{R}_{t+1} \mid \mathrm{S}_t, \mathrm{A}_t)$$

- $\gamma \in [0, 1[$ é o fator de desconto.

Num processo de decisão de *Markov*, considera-se a existência de uma agente, que, em instantes discretos, interage com o ambiente segundo um ciclo sequencial. Em cada instante t, o agente obtém informação sobre o estado atual do sistema, representado por $s_t \in \mathcal{S}$, e, de acordo com um determinado critério ou política, seleciona uma ação $a_t \in \mathcal{A}$ a ser executada. Após essa decisão, o tempo avança uma unidade, o sistema transita para um novo estado $s_{t+1} \in \mathcal{S}$ e o agente recebe uma recompensa $r_{t+1} \in \mathcal{R}$, cuja ocorrência está associada a uma probabilidade $\mathcal{P}(s_{t+1}, r_{t+1} | s_t, a_t)$ (Sutton e Barto, 2008). A Figura 4 apresenta uma representação deste processo de interação.

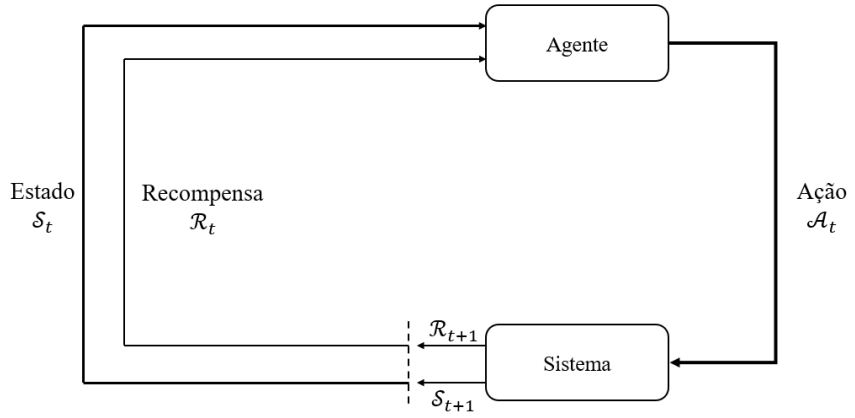


Figura 4 - Ilustração da interação do agente com o sistema

A função $\mathcal{P}$ descreve integralmente a dinâmica do sistema, uma vez que, conhecendo-se o estado presente e a ação executada, a probabilidade de transição para qualquer estado subsequente, assim como a obtenção de determinada recompensa, não depende da sequência de estados e ações anteriores. Esta característica corresponde à denominada propriedade de *Markov*.

O critério que define a forma como o agente seleciona as suas ações denomina-se *política* (Sutton & Barto, 2018). Esta pode ser estacionária ou não estacionária. Considera-se não estacionária quando a decisão depende simultaneamente do estado e do instante temporal em que o sistema se encontra; e estacionária quando é independente do tempo (Puterman, 2014).

As políticas classificam-se também como determinísticas ou estocásticas (Sutton & Barto, 2018). Numa política estacionária determinística, uma função associa a cada estado uma única ação. Por outro lado, uma política estacionária estocástica define, para cada par (s, a) a probabilidade de o agente escolher a ação $\boldsymbol{a}$ quando se encontra no estado $\boldsymbol{s}$. No presente trabalho, assume-se que as políticas adotadas são estacionárias e determinísticas, sendo o conjunto de todas elas representado por Π (Puterman, 2014).

Definido o conceito de política, a resolução de um processo de decisão de *Markov* corresponde à determinação da política ótima (Sutton & Barto, 2018). Ao interagir com o sistema, o agente acumula recompensas que representam o valor do percurso realizado. Assim, a ação ideal num dado instante é aquela que maximiza o valor esperado das recompensas futuras, com o objetivo de aumentar a quantidade e a qualidade das recompensas obtidas. Formalmente, este objetivo traduz-se na maximização da soma descontada das recompensas futuras, G_t, ao selecionar uma ação no instante t (Bertsekas, 2012), onde:

$$G_t := R_{t+1} + \gamma \times R_{t+2} + \gamma^2 \times R_{t+3} + \cdots = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}$$

O *fator de descon*to (γ) representa o peso atribuído às recompensas futuras no processo de decisão (Sutton & Barto, 2018). Quando $\gamma = 0$, apenas a recompensa imediata é considerada relevante, à medida que γ se aproxima de 1, aumenta a importância conferida às recompensas de longo prazo. Dado que $\gamma \in [0, 1[$, recompensas mais distantes no tempo têm, inevitavelmente, menor influência no valor total acumulado (Puterman, 2014).

A política π, enquanto critério que orienta a escolha das ações, influencia diretamente as recompensas obtidas. Assim, existe uma relação intrínseca entre cada política e o valor esperado das recompensas futuras que podem ser obtidas a partir de um determinado estado. Essa relação é expressa pelas funções de valor, definidas a seguir (Sutton & Barto, 2018).

Definição: A função valor *V* para a política π é a função que a cada estado $s \in \mathcal{S}$ atribui o valor $\mathrm{E}_\pi[G_t|S_t = \mathrm{s}]$ e escreve-se $v_\pi(\mathrm{s})$. A função Q para a política π é a função que a cada par $(s, a) \in \mathcal{S} \times \mathcal{A}$, atribui o valor $\mathrm{E}_\pi[G_t|S_t = \mathrm{s},\ A_t = \mathrm{a}]$ e denota-se por $q_\pi(\mathrm{s}, \mathrm{a})$.

Sobre esta definição impõe-se fazer as seguintes considerações:

- A notação $\mathrm{E}_\pi[.]$ representa o valor esperado considerando que as ações são escolhidas seguindo a política π;
- As funções *V* e Q não dependem explicitamente do instante t, evitando assim qualquer ambiguidade temporal.

Quando a política π está definida, a função *V* fornece uma medida de qualidade associada a cada estado, correspondendo ao valor esperado de seguir π a partir desse estado no futuro. A função Q fornece a medida de qualidade associada a executar uma ação a num estado s e, posteriormente, continuar a seguir π.

Uma propriedade fundamental da função Q é que esta obedece à equação de *Bellman* (Bertsekas, 2012). Usando a igualdade $G_t = R_{t+1} + \gamma \times G_{t+1}$, vem que:

$$
\begin{aligned}
q_\pi(s,a) &= \mathrm{E}_\pi[G_t|S_t = \mathrm{s},\ A_t = a] \\
&= \mathrm{E}_\pi[\, R_{t+1} + \gamma \times G_{t+1} \mid S_t = \mathrm{s},\ A_t = a] \\
&= \sum_{s_{t+1}, r_{t+1}} \mathcal{P}(s_{t+1}, r_{t+1} |\, s_t, a_t) \times [\, r + \gamma \times \mathrm{E}_\pi[G_{t+1}|S_{t+1}] \,] \\
&= \sum_{s_{t+1}, r_{t+1}} \mathcal{P}(s_{t+1}, r_{t+1} |\, s_t, a_t) \times [r + \gamma \times v_\pi(s_{t+1})] \\
&= \sum_{s_{t+1}, r_{t+1}} \mathcal{P}(s_{t+1}, r_{t+1} |\, s_t, a_t) \times [r + \gamma \times q_\pi(s_{t+1}, \pi(s_{t+1}))] \\
&= \mathrm{E}_\pi[\, R_{t+1} + \gamma \times q_\pi(S_{t+1}, \pi(S_{t+1})) \mid S_t = \mathrm{s}, A_t = a]
\end{aligned}
$$

Assim, para uma política π, a função Q pode ser expressa da seguinte forma:

$$q_\pi(s,a) = \mathrm{E}_\pi\big[\, R_{t+1} + \gamma \times q_\pi\big(S_{t+1}, \pi(S_{t+1})\big) \,\big|\, S_t = \mathrm{s}, A_t = a\big]$$

Como já foi referido, a função *V* mede o valor de cada estado segundo uma determinada política. Desta forma, usando estas funções, consegue-se comparar as diversas políticas e, consequentemente, definir o conceito de política ótima (Sutton & Barto, 2018).

Definição. Diz-se que a política π_1 é melhor que a política π_2 se $v_{\pi_1}(s) \geq v_{\pi_2}(s)$ para todo o $s \in \mathcal{S}$. Diz-se que uma política é uma política ótima se for melhor que todas as outras.

Definição. À função que a cada s ∈ $\mathcal{S}$ atribui o valor $\max_{\pi \in \Pi} v_\pi\ (s)$ dá-se o nome de função *V* ótima e escreve-se $v_*(.)$. À função que a cada par $(s,\ a) \in \mathcal{S} \times \mathcal{A}$ atribui o valor $\max_{\pi \in \Pi} q_\pi\ (s, a)$ dá-se o nome de função *Q* ótima e escreve-se $q_*(.,.)$.

Num MDP finito existe sempre uma política ótima e, caso exista mais do que uma, elas partilham as mesmas funções *V* e *Q*. Para além disso, quando se trata de uma política ótima, a função *V* é v_* e a função *Q* é q_* (Sutton & Barto, 2018). Assim, para as funções *V* e *Q* associadas a uma política ótima usa-se a notação v_* e q_*, respetivamente.

Além do mais, para uma política ótima π^*, mostra-se que a equação de *Bellman* para a função *Q* pode ser escrita de uma forma ligeiramente diferente:

$$q_*(s,a) = \mathrm{E}_{\pi^*}\left[R_{t+1} + \gamma \times \max_{a' \in \mathcal{A}} q_*(S_{t+1}, a') \middle|\ S_t = \mathrm{s}, A_t = \ \ a\right]$$

Designa-se, neste caso, por *equação de Bellman ótima*. O conhecimento da função *Q* ótima permite então definir a política ótima através da seguinte expressão:

$$\pi(s) := arg \max_{a \in \mathcal{A}} q_*(s,a) \quad \forall_s \in \mathcal{S}$$

No entanto, em problemas de grande dimensão ou elevada complexidade, esta abordagem torna-se computacionalmente dispendiosa e, por vezes, impraticável (Sutton & Barto, 2018). Acresce que a sua aplicação requer o conhecimento exato da dinâmica do sistema, representada pela função $\mathcal{P}$. Quando não se tem conhecimento da função $\mathcal{P}$ pode-se optar por métodos que visam aproximar a função *Q* ótima através da equação de *Bellman* ótima. No presente trabalho, adota-se o algoritmo de aprendizagem por reforço *Q-Learning* que será detalhado na secção seguinte.

3.3.1.1 *Q-learning*

O *Q-Learning* é um algoritmo de aprendizagem por reforço cujo objetivo consiste em obter, por aproximações sucessivas baseadas na equação de *Bellman* ótima, uma estimativa da função *Q* ótima (Watkins & Dayan, 1992). Para tal, o agente interage interactivamente com o ambiente, observando as consequências das suas ações sob a forma de recompensas numéricas.

Na sua formulação clássica, o *Q-Learning* utiliza uma tabela de *Q*-valores (*Q-table*) para as estimativas da função *Q*. Esta tabela é organizada de forma que cada linha corresponde a um estado $s \in \mathcal{S}$ e cada coluna a uma ação $a \in \mathcal{A}$, existindo, portanto, uma entrada distinta para cada par $(s,\ a)$. O valor associado a esse par, frequentemente designado $h(s,\ a)$, é inicialmente definido como zero em toda a tabela (Busoniu et al., 2010).

Supondo que, no instante t, o agente observa o estado s_t e escolhe, de alguma forma a definir, a ação a_t. O sistema transita para um novo estado s_{t+1} e o agente recebe uma recompensa r_{t+1}, formando assim a experiência $(s_t,\ a_t,\ s_{t+1},\ r_{t+1})$. Antes de

escolher uma nova ação a_{t+1} e formar uma nova experiência, o valor de $h(s_t,\ a_t)$ é atualizado segundo forma:

$$h^{novo}(s_t, a_t) \leftarrow (1 - \alpha) \times h(s_t, a_t) + \alpha \times (r_{t+1} + \gamma \times \max_{a' \in \mathcal{A}} h(s_{t+1}, a'))$$

onde $\alpha \in [0,1]$ é um valor fixado à priori a que se dá o nome de taxa de aprendizagem e $\gamma \in [0,1]$ o fator de desconto. Este processo repete-se ao longo de múltiplos episódios de interação entre o agente e o ambiente, até que os valores da função Q estabilizem (convergência) ou se atinja um número máximo de iterações definido pelo investigador, constituindo assim o treino do agente (Bertsekas, 2012).

De forma resumida, a estrutura do algoritmo *Q-Learning* segue os seguintes passos (Sutton & Barto, 2018):

1. Inicializa-se a tabela
2. A cada momento do tempo:
    - Observa-se o estado
    - Escolhe-se e executa-se a ação
    - Observa-se o novo estado e a recompensa
    - Atualiza-se a tabela

A convergência deste método está garantida sob determinadas condições (Watkins & Dayan, 1992; Jaakkola et al., 1994), nomeadamente execução repetida e suficiente de todas as ações em todos os estados. No entanto, este requisito não é eficiente se o número de estados e de ações não for suficientemente pequeno. Para ultrapassar este problema, pode-se usar uma função paramétrica Y para estimar os valores da função Q ótima em detrimento do método que se acabou de abordar.

3.3.1.2 *Deep Q-Network*

O *Deep Q-Network* (DQN) constitui uma extensão do *Q-learning* tradicional, concebida para superar as limitações deste último quando aplicado a ambientes com espaços de estados de grande dimensão ou contínuos, nos quais a representação explícita de todos os valores $q\ (s, a)$ numa tabela é impraticável (Mnih et al., 2015; Sutton & Barto, 2018). Nesta abordagem, recorre-se a uma rede neuronal profunda como função de aproximação, parametrizada por θ, para estimar os valores ótimos da função Q.

Ao contrário do *Q-learning* tradicional, que armazena os valores Q numa tabela, o DQN aproxima a função $q\ (s, a)$ através de uma rede neuronal parametrizada por θ. Esta rede recebe como entrada o estado s e devolve uma estimativa dos valores Q para todas as ações possíveis:

$$q\ (s, a, \theta)\ \approx\ q^*\ (s, a)$$

O funcionamento do DQN segue a mesma lógica de interação do *Q-Learning*: em cada passo temporal *t*, o agente observa o estado s_t, seleciona uma ação a_t, recebe uma recompensa r_{t+1} e observa o estado seguinte s_{t+1}. Estes elementos formam uma transição $(s_t, a_t, r_{t+1}, s_{t+1})$ que é utilizada para atualizar os pesos da rede.

Para atualizar a rede, utiliza-se uma função de perda baseada na equação de *Bellman* ótima (Bellman, 1957). O valor-alvo y_t para a atualização é dado por:

$$y_t = r_{t+1} + \gamma \times \max_{a' \in \mathcal{A}} q(s_{t+1}, a';\, \theta^-)$$

onde γ é o fator de desconto e θ− são os parâmetros de uma rede alvo.

O objetivo é minimizar o erro quadrático entre y_t e $q\,(s_t, a_t; \theta)$:

$$\mathcal{L}(\theta) = \left(y_t - q\,(s_t, a_t; \theta)\right)^2$$

Para assegurar a estabilidade do treino, o DQN introduz dois mecanismos fundamentais:

- *Replay de Experiência (Experience Replay):* as transições vividas pelo agente são armazenadas numa memória de experiências. Durante o treino, em vez de usar apenas a última transição, seleciona-se aleatoriamente um conjunto de transições (*mini-batch*), o que reduz a correlação entre as amostras e melhora a eficiência da aprendizagem (Lin, 1992).
- *Rede Alvo (Target Network):* para calcular o valor-alvo y_t, utiliza-se uma cópia da rede principal, utilizada para calcular os valores-alvo na atualização da função *Q*. Os seus parâmetros θ^- permanecem fixos durante um intervalo de tempo e são atualizados periodicamente através da cópia dos parâmetros da rede principal (θ^- ← θ), contribuindo para estabilizar o treino e reduzir oscilações numéricas (Mnih et al., 2015).

A estrutura geral do algoritmo DQN pode ser sintetizada da seguinte forma:

1. Inicializa-se a rede $q\,(s, a, \theta)$ com pesos aleatórios.
2. Inicializa-se a rede alvo $q\,(s, a, \theta^-) = q\,(s, a, \theta)$.
3. A cada iteração:
   - Observa-se o estado s_t;
   - Seleciona-se uma ação a_t (por exemplo, usando uma política *ε-greedy*);
   - Executa-se a_t, observando r_t e s_{t+1};
   - Armazena-se a transição na memória de experiência;
   - Seleciona-se um *mini-batch* de transições aleatórias;

- Calcula-se o valor-alvo y_t;
- Atualiza-se a rede θ minimizando $\mathcal{L}(\theta)$,
- Atualiza-se a rede alvo periodicamente: $\theta^- \longleftarrow \theta$

O DQN, inicialmente proposto por Mnih et al. (2015), demonstrou resultados expressivos ao atingir desempenho humano em diversos jogos do *Atari 2600*. Ao conjugar a robustez do *Q-Learning* com a capacidade de generalização das redes neuronais profundas, este método permite aprender políticas eficazes em ambientes de elevada dimensionalidade sem a necessidade de modelar explicitamente a dinâmica do ambiente (Hasselt et al., 2016; Arulkumaran et al., 2017).

3.3.2 *XGBoost*

O *Extreme Gradient Boosting* (XGBoost), desenvolvido por Chen e Guestrin (2016), é uma extensão do *gradient boosting* que alia eficiência computacional a boa capacidade de generalização, sendo eficaz na análise de dados de elevada dimensionalidade e relações não lineares.

A arquitetura do *XGBoost* segue o princípio aditivo do *gradient boosting*, construindo iterativamente um conjunto de modelos fracos — geralmente árvores de decisão —, onde cada novo modelo é ajustado de forma a corrigir os erros residuais dos modelos anteriores (Friedman, 2001; Chen & Guestrin, 2016).

A principal inovação do *XGBoost* reside na introdução de regularização explícita na função objetivo e na utilização de uma aproximação de segunda ordem (*Newton boosting*), que considera não apenas o gradiente de primeira ordem, mas também a informação de curvatura da função de perda. Esta abordagem permite otimizações mais precisas e estabilidade acrescida durante o treino (Chen & Guestrin, 2016; Natekin & Knoll, 2013).

Considere-se um conjunto de dados com *N* elementos, cada um com *x* variáveis explicativas. O algoritmo *XGBoost* irá construir uma coleção de *K* árvores, de f_1 até f_k. Cada árvore atribui um valor de folha, *w*, às variáveis fornecidas. Em outras palavras, $f(x) = w$. Naturalmente, a estrutura das árvores varia, pelo que diferentes árvores podem produzir valores de saída distintos para as mesmas variáveis.

Formalmente, o modelo é composto por um conjunto de *K* árvores de regressão, sendo a previsão dada por:

$$\hat{y} = \sum_{k=1}^{K} f_k(x), \qquad f_k \in \mathrm{F}$$

Onde F representa o espaço das árvores de decisão. Cada árvore associa um valor de saída (*score*) a cada folha, que é determinado de forma a minimizar a função objetivo regularizada:

$$L^{(t)} = \sum_{i=1}^{N} l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t)$$

Com o termo de regularização, $\Omega(f)$, definido por:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda\|\omega\|^2$$

onde T representa o número de folhas, γ controla a penalização pela complexidade da árvore (*pruning*) e λ regula a magnitude dos valores atribuídos às folhas (*ridge regularization*). Aqui, o termo $\frac{1}{2}\lambda\|\omega\|^2$ corresponde à regularização L2, também conhecida como *ridge regression penalty*, que penaliza o quadrado dos valores das folhas, promovendo um modelo mais conservador e reduzindo o risco de *overfitting* (Chen & Guestrin, 2016).

A função objetivo é otimizada através de uma expansão de Taylor de segunda ordem, permitindo determinar o valor ótimo de cada folha:

$$w^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda}$$

sendo g_i e h_i os gradientes de primeira e segunda ordem, respetivamente. Esta formulação conduz ao chamado *Structure Score*, métrica que orienta a divisão dos nós da árvore, maximizando a capacidade preditiva e controlando simultaneamente o risco de *overfitting*.

Para além da regularização L2 (λ) e do parâmetro γ para *pruning*, o *XGBoost* implementa técnicas adicionais de controlo de complexidade, como o *shrinkage* (ou taxa de aprendizagem η), que reduz o impacto de cada nova árvore na previsão final, e o *early stopping*, que interrompe o treino quando não se verificam melhorias consistentes no erro de validação.

A sua elevada eficiência deve-se, ainda, à utilização de estruturas de dados otimizadas para partições em memória (*block structure*), paralelização no treino das árvores e suporte para *sparse data*. Estas características tornam o *XGBoost* particularmente atrativo para problemas de larga escala, mantendo tempos de treino reduzidos e um desempenho competitivo face a métodos baseados em redes neuronais (Ke et al., 2017; Prokhorenkova et al., 2018).

A combinação de um processo de otimização de segunda ordem, regularização explícita e múltiplos mecanismos de prevenção de *overfitting* contribuiu para que o *XGBoost* se tornasse uma das ferramentas mais utilizadas em competições de ciência de dados e em aplicações industriais, evidenciando-se em tarefas como previsão de séries temporais, classificação, regressão e deteção de anomalias.

3.4 Modelos Híbridos

Este subcapítulo descreve os modelos desenvolvidos no âmbito deste trabalho, com base nos fundamentos teóricos apresentados na secção anterior. A construção dos modelos segue uma abordagem progressiva, partindo de uma configuração base exclusivamente assente em aprendizagem por reforço, para versões híbridas que integram elementos supervisionados, variáveis exógenas e mecanismos de controlo de risco.

O modelo base considerado na presente dissertação consiste num *Q-Learning* tabular de reduzida complexidade, com dois estados e duas ações. O espaço de estados representa a condição de exposição do agente ao mercado, podendo corresponder à ausência de posição ou à manutenção de uma posição ativa. O espaço de ações compreende duas possibilidades: transacionar ou não transacionar, correspondendo a uma compra quando o agente se encontra fora do mercado e a uma venda quando está no mercado. Este modelo, pela sua simplicidade estrutural e operacional, é adotado como linha de base para aferir o contributo incremental das variantes híbridas desenvolvidas, permitindo isolar e avaliar de forma rigorosa o impacto das melhorias introduzidas.

Na aplicação prática ao ativo NVIDIA, o agente observa, a cada instante temporal, o estado de exposição e seleciona a ação de acordo com uma política ε-*greedy* com ε fixo, promovendo o equilíbrio entre exploração e aproveitamento do conhecimento acumulado. A atualização da tabela *Q* segue a forma clássica baseada na equação ótima de *Bellman*, aproximando a política que maximiza a recompensa acumulada ao longo da série temporal. A eficácia do modelo é avaliada num período fora da amostra, recorrendo a métricas de desempenho adequadas, o que permite aferir a capacidade de generalização da política aprendida.

Nos subcapítulos seguintes detalham-se quatro modelos híbridos, todos assentes na integração de aprendizagem por reforço com sinais supervisionados e, em alguns casos, com informação macro e regras de gestão de risco: (i) *Q-Learning* + *XGBoost*, onde um classificador *XGBoost* fornece um sinal trinário (comprar, vender, manter) baseado em retornos defasados e esse sinal é incorporado no estado do agente tabular, orientando a escolha entre *buy/sell/hold*; (ii) *Q-Learning* + *XGBoost* + *variáveis exógenas*, que estende o anterior ao incluir no *XGBoost* e no estado do agente indicadores de regime macro/volatilidade, permitindo ajustar a política a condições de mercado; (iii) *Q-Learning* + *XGBoost* + controlo de risco, que mantém a arquitetura preditiva com sinais supervisionados e variáveis exógenas, mas introduz, na fase de execução, regras explícitas de *stop loss* fixo e *trailing stop*, visando limitar perdas e proteger ganhos sem alterar o processo de aprendizagem; e (iv) *DQN* + *XGBoost,* no qual o agente tabular é substituído por uma rede DQN capaz de operar num espaço de estados contínuo e multivariado, integrando diretamente o sinal do *XGBoost* e variáveis técnicas/exógenas, com *replay buffer* e *target network* para maior estabilidade.

Esta estrutura incremental permite avaliar, de forma sistemática, o contributo de cada componente — poder preditivo híbrido, sensibilidade a variantes macro e mitigação

de risco, bem como capacidade de generalização em espaços de maior dimensão — para a qualidade das decisões em *trading*.

3.4.1 *Q-learning* com *XGBoost*

A estratégia híbrida *Q-Learning* com *XGBoost* integra aprendizagem por reforço e aprendizagem supervisionada com o objetivo de melhorar a tomada de decisão face a um modelo exclusivamente de reforço ou a estratégias simples.

O algoritmo *XGBoost* estima a direção de curto prazo do preço, classificando-a em subida, descida ou neutra a partir de características técnicas. Esse sinal preditivo é incorporado no estado observado pelo agente de *Q-Learning*, que decide entre comprar, vender ou manter e atualiza a função valor segundo a equação de *Bellman* ao longo dos episódios. Deste modo, a política aprendida deixa de depender apenas das recompensas históricas e passa a beneficiar de informação preditiva adicional, procurando elevar o desempenho, em particular o retorno ajustado ao risco, relativamente ao modelo base e a referências como a estratégia de comprar e manter.

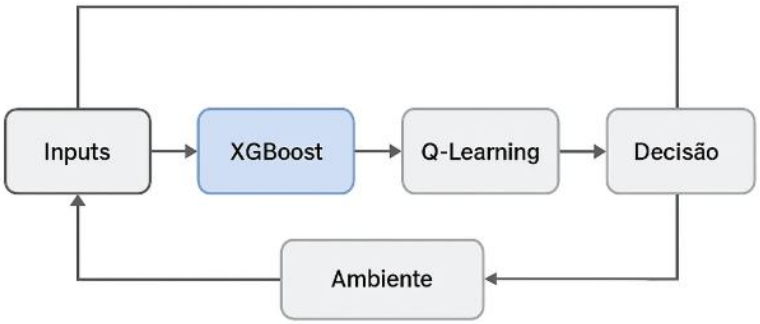


Figura 5 - Diagrama do modelo híbrido proposto combinando XGBoost e Q-Learning

O modelo *XGBoost* gera um sinal de previsão (comprar, vender ou manter) a partir de dados de mercado, nomeadamente retornos defasados, e esse sinal é incorporado no estado observado pelo agente de *Q-Learning*, em conjunto com informação sobre padrões recentes de retornos e a posição corrente do agente, para suportar a decisão de negociação. A ação selecionada é executada no ambiente de mercado e a recompensa observada (por exemplo, o retorno financeiro do período seguinte) é utilizada para atualizar a função valor do agente segundo o algoritmo de *Q-Learning*, completando o ciclo de aprendizagem.

Neste modelo híbrido, treina-se, em primeiro lugar, um classificador *XGBoost* para prever o movimento diário do ativo com base em características técnicas do preço. As variáveis preditoras são os retornos dos 1, 2 e 3 dias anteriores e o conjunto de treino abrange o histórico de 2014 até meados de 2022. O alvo é definido de forma trinária: +1 (subida) quando o retorno esperado excede um limiar positivo (por exemplo, +0,1%), −1 (queda) quando fica abaixo de um limiar negativo (por exemplo, −0,1%) e 0 (neutro) nos restantes casos. Esta codificação permite filtrar variações pouco relevantes e concentrar o sinal em situações com magnitude previsional suficiente.

Na implementação considerada, o *XGBoost* é ajustado com 50 árvores e profundidade máxima igual a 3, parâmetros determinados empiricamente. O modelo fornece, para cada dia, uma previsão categórica (comprar, vender ou manter) que é integrada no componente de reforço. Importa notar que o *XGBoost* atua como um assessor estático durante a simulação: não é atualizado em tempo real e permanece fixo após o treino com dados históricos até ao final do período de calibração.

O agente de *Q-Learning* representa um decisor automatizado que observa, em cada instante, um estado composto pelo padrão dos últimos três retornos diários (assinalados como positivos ou negativos), pela previsão do *XGBoost* e pela posição atual (comprado ou fora do mercado). Esta representação origina 48 combinações possíveis de estados e permite que a política aprendida seja sensível tanto à dinâmica recente do preço como ao sinal supervisionado. A seleção de ações segue uma política *ε-greedy*: com probabilidade de 10% procede-se à exploração por escolha aleatória e, caso contrário, adota-se a ação que maximiza o valor estimado da função Q para o estado observado.

A função de recompensa corresponde ao retorno diário quando o agente se encontra posicionado; caso contrário, a recompensa é nula. O processo de aprendizagem decorre sobre o período de 2014 até meados de 2022, com atualização iterativa da tabela Q em função dos ganhos e perdas observados, estabilizando após cerca de 100 épocas. Na fase de teste, entre junho de 2022 e setembro de 2024, o agente aplica a política aprendida sem exploração adicional, permitindo avaliar o desempenho fora da amostra.

3.4.2 *Q-learning* com *XGBoost* e Variáveis Exógenas

Este modelo híbrido constitui uma extensão direta do apresentado na secção anterior, mantendo inalterada a arquitetura base que combina aprendizagem supervisionada via *XGBoost* com aprendizagem por reforço via *Q-Learning* e introduzindo variáveis exógenas com o propósito de aprimorar a qualidade das decisões.

Em particular, são incorporadas quatro variáveis exógenas de reconhecida relevância macrofinanceira. O índice do dólar norte-americano (DXY) traduz a evolução relativa do valor do dólar face a uma cesta de moedas de referência, refletindo condições monetárias globais. O índice de volatilidade implícita (VIX), frequentemente designado por "índice do medo", sintetiza as expectativas de volatilidade futura, sendo que níveis elevados tendem a sinalizar maior aversão ao risco e potenciais correções nos mercados (Whaley, 2000). Os contratos futuros de ouro, negociados na *Commodity Exchange* (GC=F), assumem historicamente o papel de ativo de refúgio em períodos de instabilidade financeira. Por fim, os contratos futuros de petróleo bruto WTI (*West Texas Intermediate*), transacionados na *New York Mercantile Exchange* (CL=F), constituem uma das principais referências internacionais para a formação de preços no setor energético.

Embora a NVIDIA, enquanto empresa do setor tecnológico, não apresente uma relação direta com os mercados de ouro e petróleo, estes ativos exercem influência indireta através de mecanismos macrofinanceiros. O preço do petróleo, ao impactar

custos de produção, inflação e crescimento económico, repercute-se nos fluxos de capital e no apetite por risco nos mercados acionistas em geral (Sadorsky, 1999; Broadstock & Filis, 2014). Já o ouro, tradicionalmente considerado um ativo seguro, tende a registar movimentos inversos aos das ações em períodos de instabilidade, funcionando como indicador da aversão ao risco dos investidores (Baur & Lucey, 2010; Baur & McDermott, 2010).

Neste modelo, as variáveis exógenas são incorporadas como *features* adicionais no XGBoost, complementando as variáveis originais de retornos defasados da ação da NVIDIA (NVDA). Desta forma, as previsões do modelo supervisionado refletem não apenas padrões técnicos próprios do ativo, mas também o enquadramento macroeconómico e as condições gerais do mercado. A introdução destas variáveis procura capturar efeitos indiretos e correlações intermercado que dificilmente seriam identificados com base apenas na análise dos retornos históricos.

A estrutura do estado do agente foi modificada para incorporar explicitamente estas variáveis exógenas, aumentando o espaço de estados possíveis de 48 para 96 combinações distintas. Cada estado passa a considerar:

- Padrões dos últimos três retornos diários (positivo ou negativo);
- Previsão do *XGBoost* (comprar, vender ou manter);
- Posição atual do agente (comprado ou fora do mercado);
- Sinal diário da variável exógena DXY, VIX, ouro (GC=F) e petróleo (CL=F) (positivos ou negativos).

A função de recompensa permanece idêntica à do modelo de referência, correspondendo ao retorno diário obtido quando o agente se encontra posicionado e sendo nula quando está fora do mercado. A atualização da tabela *Q* segue o procedimento padrão do *Q-Learning*, utilizando as recompensas observadas para aproximar a política que maximiza o retorno acumulado, agora informada por sinais supervisionados enriquecidos com variáveis exógenas.

3.4.3 *Q-learning* com *XGBoost* e Controlo de Risco

Esta versão híbrida constitui um incremento direto ao modelo anterior — o *Q-Learning* com *XGBoost* e variáveis exógenas — através da incorporação explícita de estratégias de controlo de risco, nomeadamente o *Stop Loss* fixo e o *Trailing Stop* dinâmico. O núcleo da abordagem permanece idêntico: um agente de aprendizagem por reforço, orientado por sinais preditivos do *XGBoost* e por variáveis macroeconómicas através das variáveis exógenas. A inovação aqui reside na forma como o agente gere o risco de mercado, ajustando a sua política de decisão não apenas com base no retorno esperado, mas também em regras automáticas de mitigação de perdas.

O *Stop Loss* é uma técnica largamente adotada em finanças quantitativas e *trading* algorítmico para limitar perdas potenciais. Conforme definido por Chan (2009), trata-se

de uma ordem de venda que é acionada automaticamente quando o preço do ativo atinge um determinado nível inferior ao preço de entrada. Neste modelo, a posição comprada é encerrada se o preço do ativo descer mais de 5% em relação ao preço de entrada, uma decisão programada de forma transparente no código.

Por sua vez, o *Trailing Stop* é um mecanismo mais adaptativo que visa proteger ganhos acumulados. Esta técnica ajusta o nível de *stop* dinamicamente à medida que o preço do ativo sobe, de modo a preservar parte do lucro obtido. Neste trabalho, a posição é encerrada se o preço atual cair mais de 3% em relação ao ponto mais alto atingido desde a entrada. Segundo Prado (2018), esta técnica permite capturar movimentos favoráveis de preços, enquanto protege o capital do agente quando ocorrem reversões.

A integração destas regras de controlo de risco é feita diretamente na lógica da simulação de mercado, sem alterar o processo de treino do agente. Assim, o agente *Q-Learning* continua a aprender a sua política com base nas recompensas históricas, mas, durante a execução da política no período de teste, está sujeito às restrições impostas pelos mecanismos de *stop*. Esta separação entre treino e execução permite observar com clareza os efeitos das regras de controlo de risco sobre a *performance*.

Em termos operacionais, sempre que o agente entra numa posição comprada, o sistema regista automaticamente o preço de entrada e monitoriza tanto esse valor como o pico máximo atingido durante a posição aberta. Se o preço atual quebrar o limite de perda (*stop fixo*) ou o de recuo a partir do máximo (*trailing stop*), a posição é encerrada imediatamente e o capital é realizado. A tabela *Q* permanece intacta, permitindo à política aprendida ser testada sob diferentes regimes de execução.

Esta abordagem permite reforçar a robustez do sistema de *trading*, uma vez que responde a dois riscos típicos enfrentados em mercados voláteis: queda abrupta de preços e reversão de tendência após subida. Estudos prévios apontam para a eficácia da combinação entre técnicas de *machine learning* e controlos de risco explícitos (Zhang et al., 2020; Liang et al., 2022), especialmente em ambientes não estacionários como os mercados financeiros.

A implementação destas regras torna o modelo mais realista do ponto de vista de uma aplicação prática, aproximando-o de sistemas de *trading* algorítmico utilizados por *traders* institucionais. Além disso, proporciona uma linha de base comparativa importante para avaliar se os ganhos gerados pelo *Q-Learning* e pelo *XGBoost* são sustentáveis sob restrições realistas de perda aceitável, algo essencial do ponto de vista da gestão de risco.

3.4.4 *DQN* com *XGBoost*

Neste modelo híbrido avançado, o agente tabular de *Q-Learning* com *XGBoost* e variáveis exógenas é substituído por um agente *Deep Q-Network* (DQN), mantendo-se o uso de sinais supervisionados e contextuais. O propósito desta transição é capitalizar a capacidade das redes neurais de lidar com espaços de estado contínuos e multivariados, melhorando a robustez e a adaptabilidade da política de negociação.

O processo inicia-se com o treinamento de um classificador *XGBoost*, que produz um sinal binário de previsão (subida ou descida) com base em variáveis técnicas (como retorno, médias móveis e volatilidade) e exógenas. Este sinal é então incorporado como um componente adicional na observação do agente.

A arquitetura do ambiente de simulação é implementada através de um *Gym* personalizado, um *framework* de código aberto amplamente utilizado em *Reinforcement Learning* (Brockman et al., 2016). A sua importância decorre do facto de fornecer uma interface padronizada para a interação entre agentes e ambientes, o que facilita a implementação, reprodutibilidade e comparação de resultados entre diferentes estudos.

No presente caso, o *Gym* foi adaptado para integrar variáveis financeiras e sinais preditivos, permitindo que o vetor de observação inclua as variáveis selecionadas e o sinal previsto. As decisões do agente são tomadas mediante uma política de rede neural (*MlpPolicy*), treinada por *Deep Reinforcement Learning* segundo o paradigma DQN, com recurso a *replay buffer* e atualização de *target network*, conforme originalmente proposto por Mnih et al. (2015).

A seleção dos hiperparâmetros mais adequados, como taxa de aprendizagem, tamanho do *buffer*, parâmetro de desconto (*gamma*) e frequência de treino, é realizada de forma automática através da biblioteca *Optuna*. Esta biblioteca de otimização de hiperparâmetros recorre a métodos de pesquisa sequencial baseados em otimização bayesiana, suportando estratégias como *define-by-run* e *pruning* de experimentos. Tal abordagem permite interromper ensaios pouco promissores e concentrar os recursos computacionais nas configurações mais promissoras, assegurando uma afinação eficiente dos parâmetros que contribui para melhorar o desempenho do agente no ambiente de *trading*.

A substituição do agente tabular por *DQN* elimina a necessidade de discretização dos estados, permitindo explorar diretamente o espaço contínuo de múltiplas variáveis. Além disso, o facto de integrar o sinal *XGBoost* no processo de decisão neural cria uma sinergia entre o poder heurístico supervisionado e a capacidade adaptativa da aprendizagem profunda — configuração essa que tem sido apontada como promissora em aplicações financeiras recentes.

Modelos similares, que combinam sinais de *machine learning* supervisionado com agentes DQN, têm demonstrado ganhos significativos em mercados com comportamento estocástico, como o *trading* de ativos digitais. Contudo, este modelo destaca-se por integrar simultaneamente três fluxos de informação — técnico, preditivo e macroeconómico — dentro de um ambiente profundo otimizado.

3.5 Métricas de Desempenho

As estratégias de *trading* são avaliadas não apenas pelos retornos absolutos que geram, mas sobretudo pela sua eficiência ajustada ao risco e pela forma como se comportam em condições adversas. Para esse efeito, o conjunto de métricas utilizado — Retorno Total, *Compound Annual Growth Rat (CAGR)*, *Sharpe Ratio*, Variância, *CVaR* (5 %), *Alpha* e *Beta* — permite avaliar o desempenho das estratégias de forma completa e fiável. Tal diversidade de indicadores tranquiliza que o objetivo central de maximizar o retorno ajustado ao risco seja alcançado de forma sustentada (Sharpe, 1966; CAPM, Investopedia).

O *Sharpe Ratio*, proposto por *William Sharpe* (1966), é reconhecido como a métrica de referência para quantificar o retorno excecional por unidade de risco total (volatilidade). Valores superiores a 1 são geralmente considerados satisfatórios, enquanto valores abaixo de 0 indicam que o ativo ou a estratégia apresenta pior desempenho do que um ativo livre de risco. No entanto, diversos autores alertam para a sua limitação em distribuições de retornos não normais, onde pode subestimar o risco associado a eventos extremos (Lhabitant, 2004; Cont, 2001).

A *Compound Annual Growth Rate* (CAGR) quantifica a taxa média anual composta de crescimento do capital investido. Esta métrica é crucial para comparar o desempenho de diferentes estratégias num horizonte temporal homogéneo, dado que elimina distorções causadas por variações pontuais nos retornos (Pinto et al., 2019).

A Variância, por sua vez, é um dos indicadores estatísticos mais clássicos de risco, representando a dispersão dos retornos em torno da média. É a base conceptual para várias métricas modernas, incluindo o *Sharpe Ratio*, e traduz de forma simples a volatilidade associada a uma estratégia (Markowitz, 1952).

No campo do risco extremo, destaca-se o *Conditional Value at Risk* (*CVaR*), também conhecido como *Expected Shortfall*. Esta métrica representa o valor médio das piores perdas possíveis num determinado quantil de probabilidade, como 5%. Ao contrário do *Value at Risk* (VaR), o *CVaR* é considerado uma medida de risco coerente (Artzner et al., 1999) e fornece uma visão mais precisa sobre o impacto potencial de eventos raros, mas severos. A sua importância no contexto de *trading* reside na capacidade de avaliar a resiliência de uma estratégia perante choques de mercado.

Para avaliar a performance relativa ao mercado de referência, recorrem-se a *Alpha* **e** *Beta*, métricas derivadas do *Capital Asset Pricing Model* (CAPM) (Sharpe, 1964; Lintner, 1965). O *Alpha* mede o retorno adicional obtido por uma estratégia face ao retorno esperado pelo risco sistemático assumido. Um *Alpha* positivo indica criação de valor acima do mercado, refletindo a competência da estratégia em gerar ganhos adicionais. O *Beta*, por outro lado, mede a sensibilidade de uma estratégia às variações do mercado de referência: valores superiores a 1 revelam elevada exposição sistémica, enquanto valores próximos de 0 indicam uma estratégia menos dependente da direção geral do mercado (Bodie et al., 2021).

Em conjunto, estas métricas garantem uma avaliação abrangente, equilibrando retorno absoluto com risco (*Sharpe*, *Variância*), risco extremo (*CVaR*) e *benchmarking* face ao mercado (*Alpha*/*Beta*). Esta abordagem permite não apenas comparar estratégias de *trading* de forma mais rigorosa, mas também compreender a sua robustez em cenários adversos, o que é fundamental para a análise crítica e estatística desenvolvida no capítulo seguinte.

4. Análise de Resultados

Neste capítulo apresenta-se a análise comparativa do desempenho de estratégias híbridas de *trading*, resultantes da combinação de aprendizagem por reforço e supervisionada, considerando dois intervalos temporais distintos. A avaliação baseou-se em métricas financeiras como *Sharpe Ratio*, retorno total, *CAGR*, variância e *CVaR* 5 %, permitindo analisar a eficácia e robustez das estratégias face a diferentes condições de mercado, incluindo o efeito de variáveis exógenas e mecanismos de gestão de risco.

A avaliação comparativa dos modelos, realizada com base nas métricas calculadas sobre o conjunto de teste entre 22 de junho de 2022 e 30 de setembro de 2024, evidencia diferenças significativas na eficácia de cada estratégia. A *Tabela 2* sintetiza, de forma clara e objetiva, os resultados obtidos pelos melhores modelos.

Tabela 2 – Resultados comparativos dos modelos no intervalo de 22/06/2022 a 30/09/2024

Modelo	Sharpe	Retorno Total (%)	CAGR (%)	Variância	CVaR 5 % (%)
M1 - Q-Learning isolado	1.66	233.93	70.26	0.000510	–4.88
M2 - Híbrido (XGBoost)	2.03	469.08	116.29	0.000720	-5.50
M3 - Híbrido + exógenas	1.77	317.64	88.76	0.000650	-5.12
M4 - Híbrido + gestão de risco + exógenas	1.66	274.95	79.93	0.000648	-5.23
M5 - DQN + XGBoost + exógenas	**14.90**	**521.34**	**124.25**	**0.000012**	**-0.17**

No *Q-Learning* foram testadas doze combinações de hiperparâmetros, variando a taxa de aprendizagem (α) entre 0.1 e 0.2, o fator de desconto (γ) entre 0.9 e 0.95, a taxa de exploração inicial (ε) entre 0.2 e 0.3 com decréscimo progressivo até 0.01, e o número de episódios (N) entre 5 000, 10 000 e 20 000. A configuração ótima foi α=0.1, γ=0.9, ε=0.2 e N=5 000, com tempo de execução de cerca de 446 segundos. Este modelo alcançou um *Sharpe Ratio* de 1.66, um retorno total de 233.93% e um CAGR de 70.26%, confirmando viabilidade.

O modelo híbrido combina o classificador supervisionado *XGBoost* com o agente de *Q-Learning* tabular. O *XGBoost*, alimentado com *lags* de retornos diários (1, 2 e 3 dias) e um limiar de 0.001, gera previsões discretas de três classes — comprar, vender ou manter — que são incorporadas no espaço de estados do *Q-Learning*, juntamente com a posição do agente e a discretização binária dos sinais. Foram definidos cinquenta estimadores, profundidade de três e métrica de avaliação *logarithmic loss* (*logloss)* usada durante o treino. No *Q-Learning* exploraram-se doze combinações com α=0.1 e 0.2, γ=0.9 e 0.95, ε=0.2 e 0.3 (com decréscimo até 0.01) e N=1 000, 2 000 e 5 000. A configuração ótima correspondeu a α=0.1, γ=0.95, ε=0.2 e N=5 000, com execução de 446 segundos.

Os resultados superaram o *Q-Learning* isolado, alcançando *Sharpe Ratio* de 2.03, retorno de 469.08% e CAGR de 116.29%.

Na versão híbrida com variáveis exógenas, foram integrados fatores macroeconómicos e de mercado, nomeadamente DXY, VIX, ouro e petróleo. O espaço de estados manteve-se idêntico, mas passou a incorporar previsões mais informadas. O *XGBoost* manteve cinquenta estimadores, profundidade de três e métrica *logloss*. No *Q-Learning* foram testadas dezoito combinações: α=0.1 e 0.2, γ=0.9 e 0.95, ε=0.2 e 0.3 (até 0.01) e N=500, 1 000 e 2 000. A solução ótima foi α=0.1, γ=0.95, ε=0.2 e N=500, com tempo de 1 894 segundos. O modelo registou *Sharpe Ratio* de 1.77, retorno de 317.64% e CAGR de 88.76%, com melhoria no CVaR (–5.12%) e menor variância, indicando maior equilíbrio risco-retorno.

Na versão híbrida com gestão de risco, introduziram-se mecanismos automáticos como *stop-loss* fixo de 5% e *trailing stop* de 3%. Testaram-se dezoito combinações no *Q-Learning*, sendo a melhor α=0.1, γ=0.95, ε=0.2 e N=1 000, com tempo de 4 269 segundos. O modelo alcançou *Sharpe Ratio* de 1.66, retorno de 274.95% e CAGR de 79.93%. Apesar da redução da rentabilidade face ao híbrido simples, os mecanismos de controlo mitigaram perdas extremas e suavizaram a variância.

A introdução da gestão de risco, como *stop-loss* fixo e *trailing stop*, contribuiu para suavizar perdas severas. A *Figura 8* mostra o impacto da gestão de risco na trajetória do retorno acumulado, evidenciando a redução das perdas temporárias em relação aos picos anteriores em comparação com a versão híbrida sem *stops*, embora à custa de menor rentabilidade final.

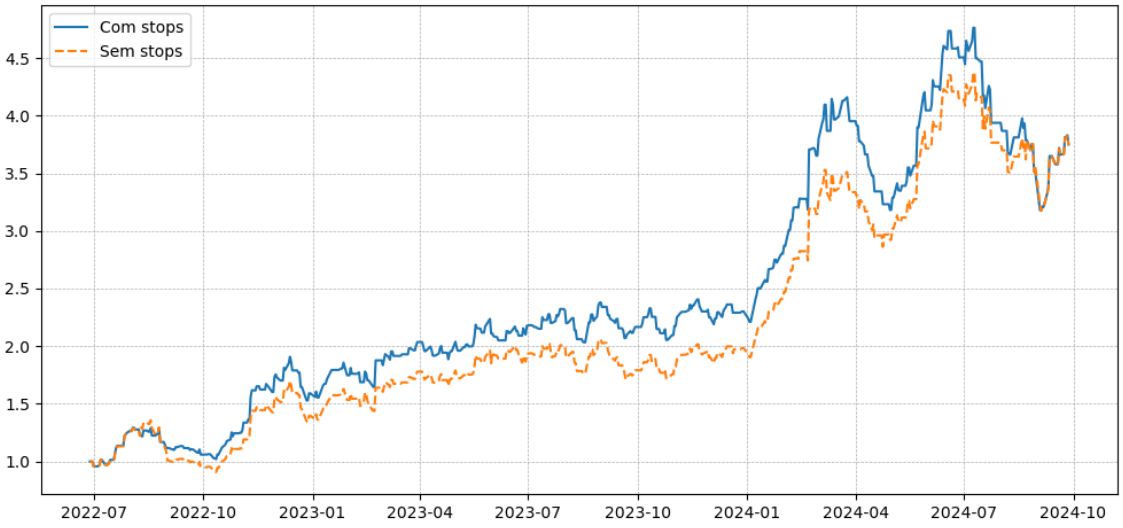


Figura 6 - Impacto da gestão de risco na trajetória do retorno acumulado

A gestão de risco, através de mecanismos como *stop-loss* e *trailing stop*, é fundamental para limitar perdas e proteger o capital. Estes mecanismos previnem que oscilações abruptas do mercado se transformem em perdas significativas, garantindo que o investidor mantém parte do capital e pode continuar a operar de forma sustentável.

Em 2025, por exemplo, o *crash* das criptomoedas em outubro provocou desvalorizações superiores a 70 % em várias *altcoins* e mais de 50 % na *Bitcoin*, enquanto mercados de ações globais registaram quedas de 5–8 % em apenas um dia devido a tensões comerciais. Sem gestão de risco, estas perdas poderiam ter comprometido totalmente o capital investido, evidenciando a importância de mecanismos de controlo sistemático (Zhang & Watts, 2025).

O modelo mais avançado combina *Deep Q-Networks* (*DQN*) com *XGBoost*, integrando variáveis exógenas (*DXY* e *VIX*). A inclusão de ouro e petróleo deteriorou a performance, pela introdução de ruído e colinearidade. A configuração ótima envolveu noventa estimadores, profundidade máxima de oito e taxa de aprendizagem ≈0.023 no *XGBoost*, enquanto no *DQN* foram otimizados $\alpha \approx 0.0027$, *buffer* de 20 000, *batch* de 64, $\gamma = 0.943$, $\varepsilon = 0.253$, atualização da rede-alvo a cada 500 passos e frequência de treino de 8. O modelo atingiu *Sharpe Ratio* de 14.90, retorno de 521.34% e CAGR de 124.25%, com variância quase nula (0.000012) e CVaR de –0.17%. O treino exigiu cerca de 2 horas para 100 000 *timesteps*.

O modelo *DQN + XGBoost* destacou-se pela sua trajetória praticamente isenta de perdas acentuadas, revelando superioridade clara em termos de crescimento sustentado. Complementarmente, a *Figura 7* sintetiza a performance em termos de *Sharpe Ratio* e *CAGR*, métricas que permitem avaliar, respetivamente, a eficiência risco-retorno e a taxa de crescimento anual composta. Os resultados confirmam que os modelos híbridos superaram o *Q-Learning* isolado, com destaque para o *DQN + XGBoost*, que alcançou valores substancialmente superiores em ambas as métricas.

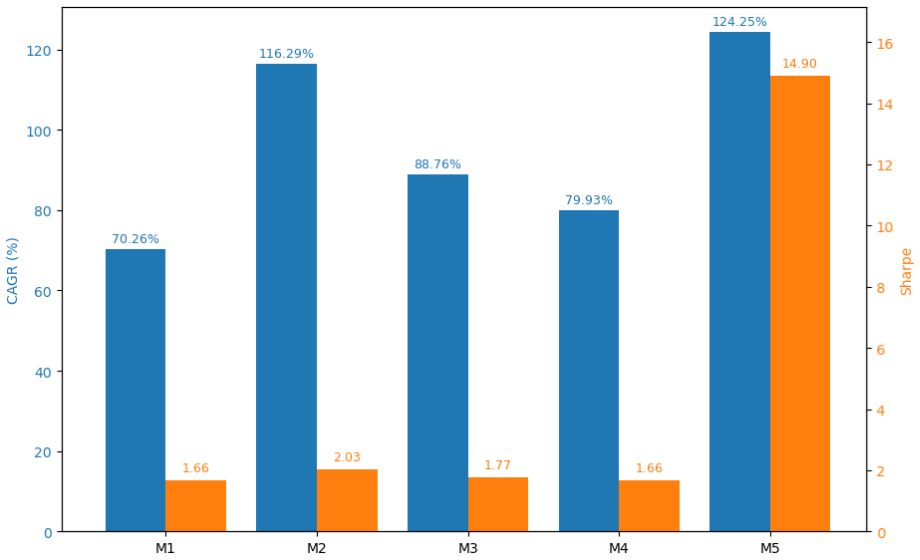


Figura 7 - Comparação dos modelos em termos de Sharpe Ratio e CAGR

A *Figura 6* apresenta as curvas do retorno acumulado de cada modelo em comparação com a abordagem *Buy & Hold*, ou seja, a estratégia passiva de comprar e manter o ativo ao longo do tempo.

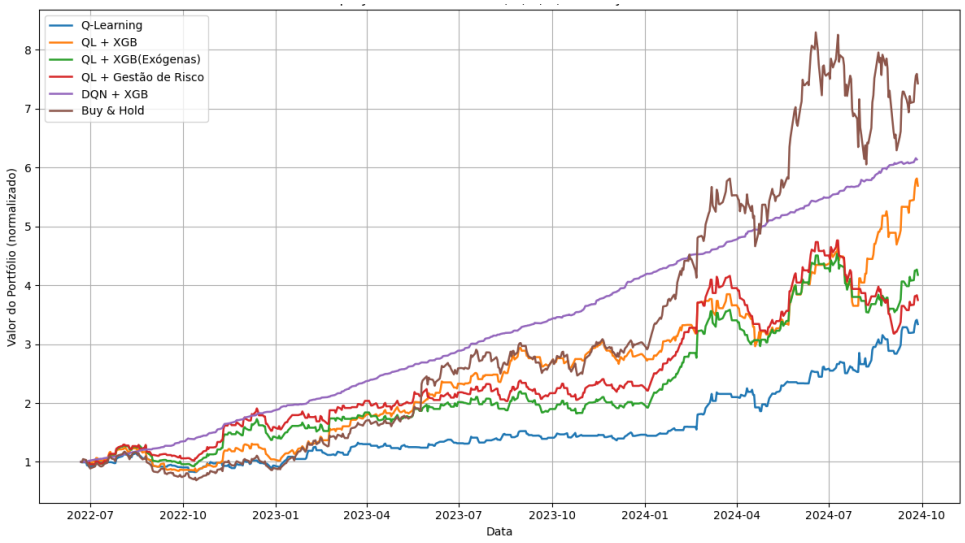


Figura 8 - Gráfico do retorno acumulado dos modelos comparados com Buy & Hold

De forma geral, os modelos híbridos superaram o *Q-Learning* isolado, validando a integração de métodos supervisionados. O modelo híbrido *Q-learning* com *XGBoost* destacou-se em retorno e *Sharpe Ratio*, enquanto a introdução de variáveis exógenas reduziu a rentabilidade, mas aumentou o equilíbrio risco-retorno. A gestão de risco automática suavizou perdas severas, embora com menor retorno. O modelo híbrido DQN com *XGBoost* revelou-se a estratégia mais eficaz e consistente, maximizando o retorno ajustado ao risco e demonstrando robustez em diferentes condições de mercado.

Na segunda fase de testes, compreendida entre 1 de janeiro de 2022 e 30 de junho de 2023, período caracterizado por elevada volatilidade e ausência de tendência, foram avaliados apenas os modelos que apresentaram melhor desempenho no intervalo mais alargado anterior. O objetivo foi analisar a sua robustez em condições de mercado adversas. Os resultados obtidos encontram-se apresentados na *Tabela 3*.

Tabela 3 - Resultados comparativos dos modelos no intervalo de 01/01/2022 a 30/06/2023

Modelo	**Sharpe**	**Retorno Total (%)**	**CAGR (%)**	**Variância**	**CVaR 5 % (%)**
Q-Learning isolado	1.08	70.08	43.02	0.000662	–6.02
Híbrido (XGBoost)	1.52	103.78	62.18	0.000521	-4.84
Híbrido + exógenas	0.66	35.56	23.03	0.001014	-6.82
Híbrido + gestão de risco + exógenas	1.11	83.29	51.09	0.000880	-6.28
DQN + XGBoost + exógenas	**16.95**	**346.82**	**174.94**	**0.000014**	**-0.16**

Esta distinção temporal nos períodos de teste explica diferenças relevantes nos resultados obtidos, tanto ao nível quantitativo como no comportamento das estratégias face às condições de mercado. No período de tendência ascendente (2022–2024), todos os modelos registaram desempenhos expressivos, destacando-se as variantes híbridas, que superaram claramente o *Q-Learning* isolado.

Já no período mais curto (2022–2023), marcado por ausência de tendência definida, as diferenças de comportamento tornaram-se mais evidentes, reforçando a importância da análise em múltiplos cenários. O *Q-Learning* isolado registou desempenho modesto (*Sharpe Ratio* de 1.08; retorno de 70.08%), enquanto o híbrido simples manteve vantagem relativa (*Sharpe Ratio* de 1.52; retorno de 103.78%), confirmando a utilidade da integração de métodos supervisionados. Contudo, a inclusão de variáveis exógenas deteriorou a performance (*Sharpe Ratio* de 0.66; retorno de 35.56%), sugerindo que, em contextos laterais ou de elevada incerteza, os sinais macroeconómicos podem introduzir ruído adicional. O híbrido com gestão de risco apresentou resultados intermédios, mitigando perdas, mas com menor rentabilidade. Uma vez mais, a estratégia *DQN* com *XGBoost* destacou-se de forma robusta (*Sharpe Ratio* de 16.95; retorno de 346.82%), demonstrando capacidade de adaptação em cenários instáveis, em contraste com os modelos tabulares, mais sensíveis ao regime de mercado.

Estudos anteriores têm mostrado que, em regimes de mercado sem tendência bem definida, os modelos baseados em *machine learning* enfrentam dificuldades acrescidas para captar padrões úteis, devido ao aumento do ruído informacional, instabilidade nas correlações entre variáveis e rápidas mudanças de regime (Suárez-Cetrulo, Quintana & Cervantes, 2023; Letteri, 2023). Quando o mercado está lateral, sem tendência, os movimentos de preços tendem a ser aleatórios, dificultando a distinção entre sinais reais e flutuações momentâneas. Além disso, relações que eram estáveis no passado podem deixar de ser confiáveis, tornando menos eficazes os modelos treinados com dados históricos. Por estas razões, abordagens puramente tabulares ou rígidas apresentam limitações, enquanto modelos híbridos e de aprendizagem profunda conseguem lidar melhor com a incerteza, captando padrões relevantes e mantendo maior consistência e robustez nas estratégias de *trading*.

Assim, no período de tendência claramente crescente, todos os modelos exibiram desempenhos positivos, com destaque para o *DQN* com *XGBoost*. No entanto, em ambiente volátil e sem direção definida, apenas esta última abordagem manteve resultados sólidos, enquanto as demais revelaram maior vulnerabilidade. Importa sublinhar que, embora o *DQN* com *XGBoost* preserve uma superioridade clara face às alternativas, a sua robustez mostrou-se menos expressiva em fases de instabilidade.

Comparando os dois períodos, conclui-se que as tendências gerais permanecem válidas: os modelos híbridos superam o *Q-Learning* isolado; a gestão de risco contribui para suavizar perdas, ainda que à custa de menor rentabilidade; e o *DQN* com *XGBoost* confirma-se como a solução mais eficaz e consistente. O contraste entre cenários de tendência e de lateralidade evidencia que, para além dos resultados absolutos, a resiliência em diferentes regimes de mercado constitui critério fundamental de avaliação, confirmando que a combinação de aprendizagem profunda com métodos supervisionados confere uma vantagem estrutural em termos de robustez e consistência. Para detalhes complementares sobre os parâmetros e métricas testados, consultar as *Tabelas 7*, *8* e *9* nos Anexos.

5. Conclusões

Este trabalho teve como principal objetivo investigar, comparar e avaliar a eficácia de abordagens de aprendizagem por reforço, simples e híbridas, no contexto de *trading* sobre a ação da NVIDIA. A análise empírica demonstrou que o *Q-Learning* isolado apresenta resultados sólidos, com retorno total de 233,93 % e *Sharpe Ratio* de 1,66, confirmando a viabilidade do modelo base sem previsão supervisionada nem mecanismos explícitos de controlo de risco. Em contraste, o modelo híbrido *Q-Learning* com *XGBoost* elevou substancialmente a qualidade do desempenho — retorno total de 469,08 % e *Sharpe Ratio* de 2,03 — evidenciando a vantagem de integrar previsão supervisionada na política sequencial de decisão.

A extensão para o modelo híbrido *Q-Learning* com *XGBoost* e variáveis exógenas (DXY, VIX, ouro e petróleo) introduziu sensibilidade macroeconómica, alcançando retorno total de 317,64 %, *Sharpe Ratio* de 1,77 e *CVaR* de –5,12 %. Embora menos lucrativo do que o híbrido simples, este modelo exibiu menor variância e perdas extremas mais contidas, sugerindo melhor equilíbrio risco-retorno em determinados regimes de mercado. Por sua vez, a incorporação de regras operacionais no modelo híbrido *Q-Learning* com *XGBoost* e controlo de risco (*stop loss* fixo e *trailing stop*) resultou em retorno total de 274,95 %, *Sharpe Ratio* de 1,66 e CVaR de –5,23 %, confirmando o papel das restrições de execução na mitigação de perdas severas, ainda que com algum custo em termos de retorno absoluto.

O destaque pertenceu ao modelo híbrido *DQN* com *XGBoost*, que combinou aprendizagem profunda com sinal supervisionado, atingindo retorno total de 517,75 %, *Sharpe Ratio* de 14,79 e CVaR de –0,19 %, praticamente eliminando perdas extremas. Estes resultados alinham-se com a literatura recente em *Deep Reinforcement Learning* aplicado a finanças, que aponta para arquiteturas profundas sensíveis ao risco como particularmente eficazes em ambientes voláteis (Pricope, 2021; Bai et al., 2024).

Na segunda fase de avaliação, compreendida entre 1 de janeiro de 2022 e 30 de junho de 2023, período marcado por elevada volatilidade e ausência de tendência definida, foram reavaliados apenas os modelos híbridos de melhor desempenho, juntamente com o modelo base. Os resultados evidenciaram que tanto o *Q-Learning* isolado como os modelos híbridos *Q-Learning* com *XGBoost* registaram desempenhos mais modestos ou mesmo deteriorados. Em contraste, o modelo híbrido *DQN* com *XGBoost* manteve uma performance robusta (*Sharpe Ratio* de 16,95; retorno total de 346,82 %), demonstrando maior resiliência a mudanças de regime e à presença de ruído informacional.

Em termos comparativos, no período de tendência ascendente entre 22 de junho de 2022 e 30 de setembro de 2024 todos os modelos alcançaram desempenhos expressivos, destacando-se a superioridade dos modelos híbridos e, em particular, do modelo híbrido *DQN* com *XGBoost*. Já no período compreendido entre 1 de janeiro de

2022 e 30 de junho de 2023, caracterizado por elevada volatilidade e ausência de tendência definida (cenário lateral), o modelo híbrido *DQN* com *XGBoost* manteve resultados consistentes, enquanto as variantes dos modelos híbridos baseados em *Q-Learning* e *XGBoost* evidenciaram maior vulnerabilidade à ausência de tendência. Assim, mais do que os retornos absolutos obtidos em fases de clara tendência ascendente, evidencia-se a importância da capacidade de sustentar desempenho tanto em contextos direcionais como em fases laterais de mercado, confirmando o modelo híbrido *DQN* com *XGBoost* como a solução estruturalmente mais robusta.

Do ponto de vista prático, os resultados sustentam a utilização de modelos híbridos em ambientes reais de *trading* algorítmico: a integração de previsão supervisionada (*XGBoost*) com modelos de aprendizagem por reforço (*Q-Learning*/*DQN*), eventualmente enriquecida com variáveis exógenas e regras de execução, concilia retorno elevado com perfil de risco controlado e maior resiliência a alterações de regime.

Importa salientar que esta investigação apresenta algumas limitações: foi aplicada exclusivamente a um único ativo, não considerou custos operacionais e restringiu-se a posições *long* (compradas), o que limita a possibilidade de generalização para mercados mais complexos. Para trabalhos futuros, recomenda-se a extensão da análise a múltiplos ativos, a incorporação de custos de transação, a inclusão de posições *short* (vendidas) e a exploração de outras arquiteturas de *Deep Reinforcement Learning* (DRL), como *Proximal Policy Optimization* (PPO) ou *Soft Actor-Critic* (SAC).

Em conclusão, a evidência obtida confirma que os modelos híbridos, em particular o modelo híbrido *DQN* com *XGBoost*, constituem uma via promissora para estratégias automatizadas de *trading*. Destacam-se não apenas pelos níveis de retorno alcançados, mas sobretudo pela robustez e consistência demonstradas em diferentes contextos de mercado, contribuindo de forma relevante para o avanço do estado da arte em finanças computacionais.

Bibliografia

Ampomah, G., Nasiru, I., & Nyarko-Boateng, O. (2020). *Ensemble Learning for Stock Market Direction Prediction Using Machine Learning Algorithms*. Journal of Big Data, 7(1), 1–20.

Ananthi, M. & Vijayakumar, K. (2021). Stock market analysis using candlestick regression and market trend prediction. Journal of Ambient Intelligence and Humanized Computing, vol. 12, no. 5, pp. 4819–4826, 2021.

Artzner, P., Delbaen, F., Eber, J. M., & Heath, D. (1999). *Coherent measures of risk. Mathematical Finance*, 9(3), 203–228.

Aweda, O. T., Woods, N. C., & Ojo, A. K. (2025). Development of Hybrid Deep Learning and Reinforcement Learning for Intelligent Trading in Forex Markets. *Journal of Advances in Mathematics and Computer Science, 40*(7), 149–158. https://doi.org/10.9734/jamcs/2025/v40i72028

Awoke, A. A., & Awoke, T. (2020). *Comparative Study of LSTM and GRU in Forecasting Volatile Bitcoin Prices*. Procedia Computer Science, 175, 180–185.

Bai, Y., Gao, L., Han, D., & Jiang, Z. (2024). *A Review of Reinforcement Learning in Financial Applications*. arXiv preprint arXiv:2411.12746. Disponível em: https://arxiv.org/abs/2411.12746

Baur, D. G., & Lucey, B. M. (2010*). Is gold a hedge or a safe haven? An analysis of stocks, bonds and gold.* Financial Review, 45(2), 217–229. https://doi.org/10.1111/j.1540-6288.2010.00244.x

Baur, D. G., & McDermott, T. K. (2010). *Is gold a safe haven? International evidence. Journal of Banking & Finance*, 34(8), 1886–1898. https://doi.org/10.1016/j.jbankfin.2009.12.008

Bellman, R. (1957*). Dynamic Programming*. Princeton University Press.

Bertsekas, D. P. (2012*). Dynamic Programming and Optimal Control* (Vol. 2). Athena Scientific. ISBN 978-1886529264.

Bhandari, D., Gautam, P., & Dev, P. (2022). *Comparative Analysis of Deep Learning Models in Predicting S&P 500 Stock Index*. Journal of Finance and Data Science, 8(2), 123–134.

Bodie, Z., Kane, A., & Marcus, A. J. (2021). *Investments* (12th ed.). McGraw-Hill Education.

Broadstock, D. C., & Filis, G. (2014). *Oil price shocks and stock market returns: New evidence from the United States and China*. Journal of International Financial Markets, Institutions and Money, 33, 417–433. https://doi.org/10.1016/j.intfin.2014.09.007

Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J., & Zaremba, W. (2016). *OpenAI Gym*. arXiv preprint arXiv:1606.01540.

Busoniu, L., Babuska, R., De Schutter, B., & Ernst, D. (2010). *Reinforcement Learning and Dynamic Programming Using Function Approxima*tors. CRC Press.

Caparrini, S., Torres, M. A., & Almeida, L. (2024). *Machine Learning Models for Stock Selection in the S&P 500 Index: A Comparative Approach*. Applied Artificial Intelligence, 38(1), 45–62.

Carr, P. (2017). *Why is VIX a fear gauge? Risk and Decision Analysis*, 1–7.

Chen, S. (2024). Research on Nvidia Investment Strategies and Analysis. In Highlights in Business, Economics and Management – EMFRM 2023 (Vol. 24, pp. 2234–2240). Department of Economics, State University of New York at Buffalo.

Chen, T., & Guestrin, C. (2016). *XGBoost: A scalable tree boosting system. Proceedings of the 22nd* ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794.

Chen, X., Li, Y., Wang, H., & Zhang, Z. (2025). *Quantum-enhanced reinforcement learning for forex trading: QLSTM-QA3C hybrid approach*. arXiv preprint. Disponível em: https://arxiv.org/abs/2509.09176v1

Chou, J. S., Chong, W. T., & Tzen, C. H. (2013). *Hybrid machine learning models for stock price forecasting*. Applied Soft Computing, 13(4), 2171–2180.

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

Darwish, A. (2025). *Stock market forecasting: From traditional predictive models to hybrid and ensemble learning approac*hes. Computational Economics. Springer. https://doi.org/10.1007/s10614-025-11024-w

Dey, S., Roy, R., & Das, A. (2016). *Forecasting Stock Market Using Hybrid Model*. Procedia Computer Science, 85, 392–400.

Elton, E. J., Gruber, M. J., Brown, S. J., & Goetzmann, W. N. (2014). *Modern Portfolio Theory and Investment Analysis* (9th ed.). Wiley.

Fama, Eugene F. 1995. Random walks in stock market prices. Financial Analysts Journal 51: 75–80.

Friedman, J. H. (2001). *Greedy function approximation: A gradient boosting machine. Annals of Statistics*, 29(5), 1189–1232.

Gupta, R., Chawla, V., & Bansal, A. (2023). *Forecasting Stock Prices Using Deep Learning Models: A Comparative Study Across Sectors*. International Journal of Forecasting, 39(1), 99–114.

Hambly, B., Xu, R., & Yang, H. (2021). *Recent Advances in Reinforcement Learning in Finance*. arXiv preprint arXiv:2112.04553

Hasselt, H. van, Guez, A., & Silver, D. (2016). *Deep Reinforcement Learning with Double Q-Learning. Proceedings of the AAAI Conference on Artificial Intelligence*, 30(1).

Hutter, F., Kotthoff, L., & Vanschoren, J. (2014). *Automated Machine Learning: Methods, Systems, Challenges*. Springer.

Jaakkola, T., Jordan, M. I., & Singh, S. P. (1994). On the convergence of stochastic iterative dynamic programming algorithms. Neural Computation, 6(6), 1185–1201. https://doi.org/10.1162/neco.1994.6.6.1185

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). LightGBM: *A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems*, 30.

Kim, H., Lee, D., & Kim, J. (2022). *HDRL-Trader: A Hybrid Deep Reinforcement Learning Approach for Robust Stock Trading*. Expert Systems with Applications, 198, 116915.

Letteri, I. (2023). VolTS: *A Volatility-based Trading System to forecast Stock Markets Trend using Statistics and Machine Learning*. arXiv:2307.13422. https://doi.org/10.48550/arXiv.2307.13422

Lhabitant, F.-S. (2004). *Hedge Funds: Quantitative Insights*. Wiley.

Li, X., Xie, H., Chen, L., Wang, J., & Deng, X. (2020). *A CNN-LSTM Quantitative Trading Strategy for Cryptocurrency Prediction*. Financial Innovation, 6(1), 1–12.

Lin, L.-J. (1992). *Self-improving reactive agents based on reinforcement learning, planning and teaching*. Machine Learning, 8(3–4), 293–321. https://doi.org/10.1007/BF00992699

Lintner, J. (1965). *The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. The Review of Economics and Statistics*, 47(1), 13–37.

Markowitz, H. (1952). *Portfolio selection. The Journal of Finance*, 7(1), 77–91.

Mnih, V., et al. (2015). *Human-level control through deep reinforcement learning.* Nature, 518(7540), 529–533. https://doi.org/10.1038/nature14236

Natekin, A., & Knoll, A. (2013). *Gradient boosting machines, a tutorial. Frontiers in Neurorobotics*, 7, 21.

Pinto, J. E., Henry, E., Robinson, T. R., & Stowe, J. D. (2019). *Equity Asset Valuation* (4th ed.). CFA Institute Investment Series.

Preis, T., Moat, H. S., & Stanley, H. E. (2013). *Quantifying Trading Behavior in Financial Markets Using Google Trends*. Scientific Reports, 3, 1684.

Pricope, T. (2021). *Deep Reinforcement Learning in Quantitative Algorithmic Trading*. arXiv preprint arXiv:2106.00123. Disponível em: https://arxiv.org/abs/2106.00123

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). *CatBoost: unbiased boosting with categorical features. Advances in Neural Information Processing Systems*, 31.

Puterman, M. L. (2014). *Markov Decision Processes: Discrete Stochastic Dynamic Programming.* John Wiley & Sons. ISBN 978-1118625873.

Riedmiller, M. (2005). *Neural fitted Q iteration – First experiences with a data efficient neural reinforcement learning method*. In Proceedings of the 16th European Conference on Machine Learning (pp. 317–328). Springer.

Sadorsky, P. (1999). *Oil price shocks and stock market activity. Energy Economics*, 21(5), 449–469. https://doi.org/10.1016/S0140-9883(99)00020-1

Selvin, S., Vinayakumar, R., Gopalakrishnan, E. A., Menon, V. K., & Soman, K. P. (2017). *Stock price prediction using LSTM, RNN and CNN-sliding window model*. In 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI) (pp. 1643–1647).

Sethia, A., Mehta, S., & Srinivasan, K. (2019). *Deep Learning for Predicting Stock Market Trends*. International Journal of Engineering and Advanced Technology, 8(6), 5326–5331.

Sharpe, W. F. (1964). *Capital asset prices: A theory of market equilibrium under conditions of risk.* The Journal of Finance, 19(3), 425–442.

Sharpe, W. F. (1966). *Mutual fund performance*. The Journal of Business, 39(1), 119–138.

Stempień, J., & Ślepaczuk, R. (2025). *Hybrid models for financial forecasting: Combining econometric, machine learning, and deep learning models*. arXiv preprint arXiv:2505.19617. https://arxiv.org/abs/2505.19617

Suárez-Cetrulo, A. L., Quintana, D., & Cervantes, A. (2023). *Machine Learning for Financial Prediction Under Regime Change Using Technical Analysis*: A Systematic Review. *International Journal of Interactive Multimedia and Artificial Intelligence, 10*(3), 6–15. https://doi.org/10.9781/ijimai.2023.06.003

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

Théate, T., & Ernst, D. (2020). *An Application of Deep Reinforcement Learning to Algorithmic Trading*. arXiv preprint arXiv:2004.06627.

Tong, X., Zhang, Y., & Li, H. (2021). *A Hybrid Attention-XGBoost Model for Stock Price Movement Prediction*. Expert Systems with Applications, 168, 114481.

Van Hasselt, H., Guez, A., & Silver, D. (2016). *Deep Reinforcement Learning with Double Q-Learning*. AAAI Conference on Artificial Intelligence, 30(1).

Wang, J., Chen, H., & Sun, Y. (2023). *Stock Trend Prediction by Integrating Market Sentiment and Technical Indicators Using Deep Learning*. Journal of Computational Finance and Economics, 12(1), 55–74.

Watkins, C. J. C. H., & Dayan, P. (1992). *Q-learning. Machine Learning*, 8(3–4), 279–292. https://doi.org/10.1007/BF00992698

Ying, X., Zhao, Z., & Lin, Y. (2024). *Advances in Time Series Forecasting Using Hybrid Machine Learning Models: A Financial Market Perspective*. Applied Intelligence, 54(3), 3112–3127.

Yun, Y., Park, J., & Cho, S. (2021). *A Hybrid Genetic Algorithm and XGBoost Model for Stock Market Forecasting*. Expert Systems with Applications, 176, 114811.

Zhang, S., Zheng, Y., & Wang, P. (2018). *Improving Stock Market Prediction with XGBoost and Feature Engineering*. Procedia Computer Science, 131, 173–179.

Zhang, W., & Watts, A. (2025). *HODL Strategy or Fantasy? 480 Million Crypto Market Simulations and the Macro-Sentiment Effect*. arXiv preprint arXiv:2512.02029. https://arxiv.org/abs/2512.02029

Zhong, Xiao, and David Enke. 2017. *Forecasting daily stock market return using dimensionality reduction*. Expert Systems with Applications 67: 126–39.

Anexo

Autores	Modelos	Métricas	Desempenho
Caparrini et al. (2024)	Decision Tree (**DT**) Random Forest (**RF**) XGBoost	Retorno Total (RT) Retorno Anualizado (RA) Sharpe Ratio (SR) Volatilidade Anual (VA)	**DT** RT 289.63% RA 7.29% SR 0.22 VA 33.54% **RF** RT 329.63% RA 8.18% SR 0.21 VA 38.7% **XGBoost** RT 502.16% RA 11.26% SR 0.29 VA 39.42%
Wang et al. (2023)	MLP, DT, NB, RF, LR, XGBoost, LSTM CNN	Acurácia F1-score	Com Análise Técnica (TA) e de Sentimento (SA) Prinicpais resultados (DJIA): **Acurácia**: MLP 64.67 DT 64.08 NB 72.29 RF 72.69 LR 72.69 XGBoost 73.09 LSTM 72.69 CNN 73.41 ***F1***: MLP 64.67 DT 64.08 NB 73.67 RF 63.80 LR 63.11 XGBoost 64.08 LSTM 62.06 CNN 62.89
Ampomah et al (2020)	AdaBoost Random Forest XGBoost, Bagging Classifier Extra Trees Voting Classifier	Acurácia Precisão Recall F1-score Especificidade AUC-ROC	Média de Acurácia no Treinamento: AdaBoost: 0.9103 Random Forest (RF): 0.9098 XGBoost (XG): 0.9150 Bagging Classifier (BC): 0.8909 Extra Trees (ET): 0.9261 Voting Classifier (VC): 0.9198 Média de Acurácia no Teste: AdaBoost: 0.9103 Random Forest (RF): 0.9098 XGBoost (XG): 0.9150 Bagging Classifier (BC): 0.8909 Extra Trees (ET): 0.9261 Voting Classifier (VC): 0.9198

Tabela 4 – Resultados dos Métodos na classificação da tendência

Autores	Modelos	Métricas	Desempenho
Bhandari et al. (2022)	LSTM	RMSE MAPE R	Melhor RMSE com Adam: 28.65 Melhor RMSE com multilayer LSTM (150, 100): 49.8362 MAPE e R mostraram desempenho superior para modelos de uma camada em comparação com multilayer LSTM. Neurônios em um modelo de uma camada (150): Melhor desempenho em RMSE, MAPE e R.
Zou et al. (2022)	ARIMA RNN com LSTM RNN com Stacked-LSTM RNN com LSTM + Attention	MSE	Valores médio para as 10 ações. **ARIMA**: MSE = 0.0546 **LSTM**: MSE = 0,0429 **Stacked-LSTM**: MSE = 0,0525 **Attention-LSTM**: MSE = 0,029 Retorno por estratégia **LSTM**: long-only 173% e long-short 99% **Attention-LSTM**: long-only 266% e long-short 263%
Sethia et al. (2019)	LSTM GRU MLP SVM	R^2 RMS error índice de otimismo (IO) índice de pessimismo (IP) Taxa de Retorno (R)	**LSTM**: RSME 0.00043, R^2 0.9486, R 4.308, IO 0.3104, IP 0.2038 **GRU**: RSME 0.00051, R^2 0.9387, R 5.7222, IO 0.4479, IP 0.1209 **SVM**: RSME 0.00054, R^2 0.9350, R -1.858, IO 0.0801, IP 0.6313 **MLP**: RSME 0.00105, R^2 0.8740, R 2.4787, IO 0.6890, IP 0.1154
Awoke et al. (2020)	LSTM GRU	RSME MAPE	Para **LSTM**: 1 dia, RMSE: 0.092, MAPE: 0.068 3 dias, RMSE: 0.079, MAPE: 0.057 5 dias, RMSE: 0.081, MAPE: 0.060 7 dias, RMSE: 0.045, MAPE: 0.030 15 dias, RMSE: 0.067, MAPE: 0.048 Para **GRU**: 1 dia, RMSE: 0.075, MAPE: 0.065 3 dias, RMSE: 0.065, MAPE: 0.046 5 dias, RMSE: 0.087, MAPE: 0.062 7 dias, RMSE: 0.051, MAPE: 0.035 15 dias, RMSE: 0.067, MAPE: 0.058
Gupta et al. (2023)	Regressão Linear Multicamadas (RLM) LSTM CNN	MSE MAE RMSE	Valores médios: **MSE:** Reg. Linear: 0.972, LSTM: 0.065 CNN: 0.097 **MAE:** Reg. Linear: 0.629, LSTM: 0.114 CNN: 0.123 **RMSE:** Reg. Linear: 0.869, LSTM: 0.134 CNN: 0.146

Tabela 5 – Resultados dos métodos na previsão de preço

Autores	Modelos	Métricas	Desempenho
Tong et al. (2023)	ARIMA LSTM XGBoost Híbrido	MSE MAR	Os valores médios das empresas: **MSE / MAR** ARIMA: 5,84 / 1,66 LSTM: 5,97 / 1,70 XGBoost: 4,23 / 1,39 Híbrido: 3,12 / 1,19
Li et al. (2020)	BP CNN LSTM CNN-LSTM	MAE RMSE MAPE Precisão Recall F1-score	**BP** MAE: 416.09 RMSE: 515.35 MAPE: 2.85% Precisão: 0.59 Recall: 0.58 F1: 0.55 **CNN** MAE: 215.98 RMSE: 261.90 MAPE: 2.03% Precisão: 0.64 Recall: 0.63 F1: 0.60 **LSTM** MAE: 229.78 RMSE: 297.97 MAPE: 2.75% Precisão: 0.58 Recall: 0.73 F1: 0.63 **CNN-LSTM** MAE: 209.89 RMSE: 258.31 MAPE: 2.35% Precisão: 0.64 Recall: 0.81 F1: 0.69
Yun et al. (2021)	Extra Trees (ET) Decision Tree (DT) Random Forest (RF) SVM com Kernel K-Nearest Neighbour (KNN) GA-XGBoost	Acurácia Precisão Recall F1 Score AUC	Acurácia média dos modelos. ET 93.71% DT 90.78% RF 92.42% SVM 85.46% KNN 80.24% GA-XGBoost 93.82%
Kim et al. (2022)	B&H P-DDQN PT-DQN P-Trader HDRL-Trader	Taxa de Retorno (R) Sharpe Ratio (SR) Máximo Drawdown (MDD)	**B&H** R: 109.3% SR: 1.81 MDD: 0.58 **P-DDQN** R: 88.8% SR: 1.51 MDD: 0.63 **PT-DQN** R: 109.6% SR: 1.74 MDD: 0.00 **P-Trader** R: 99.8% SR: 1.72 MDD: 0.00 **HDRL-Trader** R: 125.2% SR: 2.12 MDD: 0.36

Tabela 6 - Resultados dos Métodos Híbridos e Ensemble

Modelos	alpha	gamma	epsilon	episodes/ epochs	Return Total (%)	Sharpe Ratio	CAGR (%)	Variância	CVaR (%)	Alpha	Beta	Tempo (s)
Q-learning	0.1	0.9	0.2	5000	233.93	1.66	70.26	0.000510	-4.88	0.1431	0.4407	398.45
Modelo Híbrido (Q-learning+XGBoost)	0.1	0.95	0.2	5000	469.08	2.03	116.29	0.00072	-5.5	0.2252	0.6251	446.45
Modelo Híbrido com variáveis Exógenas												
DXY e *VIX*	0.1	0.95	0.2	1000	117.89	1.05	41.36	0.000662	-5.57	0.1711	0.5743	331.17
DXY, VIX, ouro e petróleo	0.1	0.95	0.2	500	317.64	1.77	88.76	0.000650	- 5.12	0.1278	0.5640	1894.10
Modelo Híbrido e Gestão de Risco												
Sem variáveis exógenas	0.1	0.95	0.2	500	142.19	1.22	48.06	0.000582	-5.47	-4.73%	0.5032	2224.47
DXY e *VIX*	0.1	0.95	0.2	500	42.97	0.62	17.22	0.000534	-5.55	-0.2593	0.4640	2411.05
DXY, VIX, ouro e petróleo	0.1	0.95	0.2	1000	274.95	1.66	79.93	0.000648	- 5.23	0.0816	0.5620	4268.82

Tabela 7 - Parâmetros e métricas dos modelos no intervalo de 22 de junho de 2022 a 30 de setembro de 2024

Modelos	alpha	gamma	epsilon	episodes/ epochs	Returno Total (%)	Sharpe Ratio	CAGR (%)	Variância	CVaR (%)	Alpha	Beta	Tempo (s)
Q-learning	0.1	0.95	0.23	5000	70.08	1.08	43.02	0.000662	-6.02	0.27	0.4348	377.55
Modelo Híbrido (Q-learning+XGBoost)	0.1	0.9	0.2	1000	103.78	1.52	62.18	0.000521	-4.84	39.88%	0.3424	77.55
Modelo Híbrido com variáveis Exógenas												
DXY e *VIX*	0.1	0.95	0.2	5000	24.01	0.54	15.78	0.000995	-6.63	-0.0336	0.6523	163.01
DXY, VIX, ouro e petróleo	0.1	0.95	0.2	1000	35.56	0.66	23.03	0.001014	-6.82	0.0209	0.6816	4138
Modelo Híbrido e Gestão de Risco												
Sem variáveis exógenas	0.1	0.95	0.2	500	93.76	1.29	56.72	0.000690	-6.20	0.3374	0.4535	2262.18
DXY e *VIX*	0.1	0.95	0.2	500	47.04	0.79	30.03	0.000892	-6.28	0.1010	0.5848	2313.34
DXY, VIX, ouro e petróleo	0.1	0.95	0.2	500	83.29	1.11	51.09	0.000880	-6.28	0.2553	0.5776	2258.43

Tabela 8 - Parâmetros e métricas dos modelos no intervalo de 1 de janeiro de 2022 a 30 de junho de 2023

Parâmetro	22/06/2022 – 30/09/2024		01/01/2022 – 30/06/2023	
	2 Exógenas	**4 Exógenas**	**2 Exógenas**	**4 Exógenas**
xgb_n	213	157	118	263
xgb_d	8	5	5	5
xgb_lr	0.0145	0.0080	0.0275	0.0292
dqn_lr	0.00333	0.00325	0.00156	0.00185
buf	25	15	35	35
bs	64	64	32	128
gamma	0.9206	0.9729	0.9020	0.9494
ef	0.0593	0.4293	0.3065	0.1224
tui	700	700	300	400
tf	8	6	7	1
Variáveis Exógenas	2	4	2	4
Sharpe Ratio	14.79	1.02	16.95	3.36
Retorno Total	517.75%	35.74%	346.82%	68.77%
CAGR	123.68%	8.54%	174.94%	34.69%
Variância	0.000012	0.000001	0.000014	0.000003
CVaR(5%)	-0.19%	-0.01%	-0.16%	-0.03%
Alpha	80.88%	5.58%	101.98%	20.22%
Beta	-0.0296	-0.0020	-0.0389	-0.0077

Tabela 9 - Parâmetros e métricas dos modelos DQN+XGBoost

Legenda dos Hiperparâmetros
xgb_n – Número de árvores (estimadores) do modelo *XGBoost*.
xgb_d – Profundidade máxima das árvores no *XGBoost*.
xgb_lr – Taxa de aprendizagem (*learning rate*) do *XGBoost*.
dqn_lr – Taxa de aprendizagem da rede neural no *DQN*.
buf – Tamanho do *replay buffer*, ou seja, quantas transições o agente armazena para treino.
bs – Batch size, número de amostras retiradas do buffer para cada atualização.
gamma – Fator de desconto (γ*gammaγ*) usado no Q-*Learning*/*DQN* para valorizar recompensas futuras.
ef – *Epsilon* final, valor mínimo da taxa de exploração *epsilon* no esquema de exploração-exploração.
tui – *Target update interval*, frequência (em passos) de atualização da rede-alvo no *DQN*.
tf – Número de camadas (ou *layers*) na rede neural do DQN.

Legenda das Métricas
Sharpe Ratio – Índice de *Sharpe*, mede retorno ajustado ao risco.
Total Retorno – Retorno acumulado total do período de teste.
CAGR – *Compound Annual Growth Rate*, taxa de crescimento anual composta.
Variância – Variância dos retornos, indicador de risco.
CVaR(5%) – *Conditional Value at Risk* a 5%, risco esperado em cenários adversos.
Alpha – Excesso de retorno em relação ao *benchmark*, ajustado ao risco.
Beta – Sensibilidade do portfólio em relação ao mercado (menor que zero indica descorrelação).