Otimização de um Portfólio de Algoritmos de Negociações Automatizadas utilizando Reinforcement Learning para o controle de risco

Ramon de Cerqueira Silva[1], **Carlos Alberto Rodrigues**[1]

[1]Universidade Estadual de Feira de Santana (UEFS)
Caixa Postal 44036-900 – Feira de Santana – BA – Brasil

ramondecerqueirasilva@gmail.com, carod@uefs.br

Abstract. *This work proposes an innovative method for optimizing Automated Trading Systems (ATS) portfolios using advanced Deep Reinforcement Learning (DRL) techniques. The algorithms A2C, DDPG, PPO, SAC, and TD3 are assessed for their ability to learn and adapt in volatile market conditions. The main goal is to enhance risk control and operational efficiency of ATS, using data from the Brazilian stock market. DRL models outperformed traditional benchmarks by offering superior risk management and better risk-adjusted returns. The findings demonstrate the potential of DRL algorithms in complex financial scenarios and lay the groundwork for future research on integrating machine learning in quantitative finance.*

Resumo. *Este trabalho apresenta uma abordagem inovadora para otimizar portfólios de sistemas de negociação automatizados (ATS) utilizando técnicas avançadas de Deep Reinforcement Learning (DRL). São analisados os algoritmos A2C, DDPG, PPO, SAC e TD3, visando avaliar suas performances em mercados voláteis. O principal objetivo é aprimorar o controle de risco e a eficiência operacional dos ATS com dados do mercado de ações brasileiro. Os modelos de DRL superaram os benchmarks, proporcionando melhor gestão de risco e retornos ajustados. Os resultados destacam o potencial dos algoritmos DRL em ambientes financeiros complexos e abrem caminhos para pesquisas futuras na integração do aprendizado de máquina em finanças quantitativas.*

1. Introdução

O *Reinforcement Learning* (RL) tem se destacado como uma poderosa ferramenta para enfrentar desafios em diversas áreas, incluindo a tomada de decisão em tempo real e previsões no mercado de ações. No RL, um agente aprende a maximizar recompensas interagindo com o ambiente, tornando-o promissor para aplicações financeiras, como a negociação automatizada [Chekhlov et al. 2005].

Os ATS (*Automated Trading Systems*) utilizam algoritmos para decisões de compra e venda, baseados em dados de mercado em tempo real. No entanto, mercados voláteis apresentam riscos, exigindo a constante otimização desses sistemas [Treleaven et al. 2013]. A aplicação de RL nesses modelos oferece uma abordagem adaptativa, permitindo maior flexibilidade e eficiência nas negociações.

Além disso, o RL se diferencia por eliminar a necessidade de previsões intermediárias, adaptando-se dinamicamente às mudanças do mercado. Estudos recentes

demonstram que o RL supera abordagens tradicionais em termos de rentabilidade e eficácia, mostrando-se uma técnica robusta em áreas como *High Frequency Trading* e gestão de portfólios [Buşoniu et al. 2018]. Estudos mostram que essas estratégias superaram abordagens tradicionais em termos de rentabilidade e eficácia [Chekhlov et al. 2005] [Framework 2013] [Parker and Fry 2020].

Este estudo busca explorar a otimização de modelos RL em um portfólio de ATS, comparando-os com a abordagem tradicional sem otimização e os índices do mercado, a fim de verificar as vantagens e limitações do RL no controle de risco.

2. Metodologia

2.1. Conjunto de dados

A aplicação dos algoritmos é realizada através do uso da biblioteca FinRL, especializada em aprendizado por reforço profundo (DRL) para negociação automatizada de ações [Liu et al. 2021].

Para a otimização, são feitos *backtests*, que consistem em uma simulações na base histórica de como um portfólio proposto teria se comportado se fosse implementado ao longo de um período passado. Com base nisso, os backtests são compostos por dados históricos de estratégias e retorno diários do Ibovespa.

Cada negociação envolve dois minicontratos do índice IBOVESPA (WIN) ou dólar (WDO), sendo que o histórico de dados abrange um total de 20.644 negociações. Ao todo, são consideradas 26 estratégias, sendo 18 aplicadas aos contratos WIN e 8 aos contratos WDO. Essas estratégias incluem tanto técnicas de tendência quanto osciladores e com exceção de uma estratégia, todas as negociações pertencem à categoria *day trade*. Essas estratégias de *day trade* utilizam *timeframes* de 15 e 20 minutos, permitindo uma análise granular e a execução rápida de operações ao longo do dia. O uso de *timeframes* curtos é fundamental para capturar movimentos de preço intradiários e aproveitar oportunidades de lucro em períodos de alta volatilidade [Day Trade Review 2023].

A biblioteca *yfinance* é utilizada para obter dados do principal índice de ações do mercado brasileiro, IBOVESPA através do Exchange Traded Fund (ETF) BOVA11, que tem como objetivo replicar o desempenho do Índice Bovespa, representando o outro conjunto de dados a ser utilizado como *backtest* do desempenho de negociação. Os dados são acessados no período que se estende de 12 de abril de 2018 até 11 de novembro de 2019.

A diversidade de métricas permitem a aplicação de técnicas de análise técnica e a simulação de *backtests* para avaliar o desempenho histórico das estratégias propostas.

2.2. Pré-Processamento de Dados

O conjunto inicial de dados históricos consiste em registros detalhados de operações, incluindo o tipo de operação (compra ou venda), datas e preços de entrada e saída, resultados em termos de lucro ou perda, volumes negociados e a identificação dos ATS envolvidos.

Após todo o processo o *dataset* gerado abrange a data, o nome da estratégia e o lucro (ou perda) das operações diárias, expressos em montantes financeiros. A coluna

data é renomeada para *date* a fim de padronizar os nomes das colunas. Em seguida, é feita uma transformação via *DataFrame* transformando seu formato de largo para longo, onde cada linha representa o lucro para uma data e um ATS específico, indicados pelas colunas *close* e *tic*, respectivamente.

Além disso, são calculados os indicadores técnicos para aprimorar as análises e auxiliar na tomada de decisões de negociação.

2.3. Otimização de portfólio utilizando RL

Uma abordagem para resolver o problema de otimização de portfólio é o uso de um agente RL. Nesta metodologia, o agente desenvolve uma política ao interagir diretamente com um ambiente. Em cada intervalo de tempo, o ambiente fornece observações que definem o estado do sistema. Com base neste estado, o agente decide qual ação realizar. Após a execução da ação, o ambiente retorna uma recompensa, permitindo ao agente avaliar a eficácia da ação escolhida. O objetivo do agente de RL é desenvolver uma política que maximize a soma esperada das recompensas ao longo do tempo [Sutton and Barto 2018].

No entanto, para ser eficaz em tarefas de otimização de portfólio, o agente de RL precisa lidar com espaços de estado complexos. Um portfólio é composto por múltiplos ativos, cada um com sua própria série de preços, o que resulta em um espaço de estado altamente dimensional. Utilizar aproximações de função, como redes neurais (NN), tem demonstrado resultados notáveis em diversas tarefas complexas [Sutton and Barto 2018]. Assim, os algoritmos de aprendizado por reforço profundo (DRL) são os mais adequados para essa finalidade.

2.4. Ambiente proposto

Assim, é necessário projetar uma solução de negociação automatizada para alocação de portfólio. O processo de negociação de ações é modelado como um problema de Markov Decision Process (MDP), envolvendo a observação de mudanças nos preços das ações, tomada de ações e cálculo de recompensas para ajustar a estratégia de negociação do agente [Sutton and Barto 2018]. Todo o pré-processamento é realizado para que os dados das negociações de ATS estivessem de acordo com este ambiente, onde o mesmo permite que o agente interaja e aprenda, considerando elementos cruciais como preços históricos das ações e indicadores técnicos [Liu et al. 2021].

Para treinar um agente de negociação com DRL, é implementado um ambiente que simula a negociação no mundo real usando o *OpenAI Gym*[Liu et al. 2020].

O ambiente inicializa carregando os dados de mercado para o dia atual e configurando o estado inicial, que inclui a matriz de covariância e os indicadores técnicos. A cada passo, o agente toma decisões de alocação, que são normalizadas e aplicadas para calcular o peso de cada ATS na carteira. O valor da carteira é atualizado de acordo com esse balanceamento, e a recompensa é definida como o novo valor da carteira. Se o episódio termina, gráficos de recompensas acumuladas e diárias são salvos, e estatísticas como o *Sharpe ratio* são calculadas e exibidas. O ambiente é então reinicializado para um novo episódio.

2.5. Treinamento do Agente DRL

A implementação dos algoritmos DRL a seguir é baseada em *Stable Baselines*. *Stable Baselines* é uma ramificação do *OpenAI Baselines*, com uma grande refatoração estrutural e

limpezas de código [Liu et al. 2020]. A partir desta biblioteca, A2C, DDPG, PPO, SAC e TD3, todos algoritmos de DRL, são implementados pelo fato de serem bastante utilizados na área de finanças [Liu et al. 2021] [Haarnoja et al. 2018] [Fujimoto et al. 2018]. A seleção destes algoritmos para um agente de RL é baseada em suas capacidades de lidar com espaços de ação contínuos, eficiência amostral, estabilidade de treinamento e desempenho robusto. Cada um desses algoritmos traz características únicas que podem ser exploradas para desenvolver estratégias de negociação eficazes e adaptativas, permitindo que os agentes aprendam e otimizem suas políticas de maneira eficiente e robusta [Buehler et al. 2019].

Com isso, a base de dados das negociações de ATS é separada por datas onde, o treinamento cobre o período do dia 1 de setembro de 2014 ao dia 11 abril de 2018. E para o teste, o período do dia 12 de abril de 2018 à 11 de novembro de 2019, totalizando 18486 negociações para treinamento e 10218 para o teste, com uma proporção aproximada de 65% e 35%, respectivamente.

Pra gerar os modelos DRL utilizando a biblioteca FinRL [Liu et al. 2020], é necessário importar e configurar os parâmetros de treinamento para todos os agentes. Diante da necessidade de melhorar o desempenho do agentes DRL, é utilizada a biblioteca Optuna [Akiba et al. 2019] para realizar a otimização dos hiperparâmetros.

Após o treinamento cada um dos modelos é usado para prever a performance do portfólio de ATS no ambiente definido utilizando o *dataset* de ATS, especificamente entre as datas de 6 de junho de 2018 a 11 de novembro de 2019, gerando dois conjuntos de resultados: os retornos diários e as ações tomadas pelos modelos. Esses resultados ajudaram a avaliar a eficácia de cada estratégia de modelo em termos de maximização de retorno e gestão de risco em um ambiente de otimização de portfólio.

2.6. Métricas de Avaliação

Para comparar diferentes abordagens de otimização de portfólio, além das métricas de desempenho previamente discutidas, é importante utilizar um *benchmark* confiável. Uma das abordagens mais comuns e reconhecidas para a otimização de portfólios é o modelo *Min-Variance* [Yang et al. 2018], que integra a Teoria Moderna do Portfólio, onde busca encontrar a combinação de ativos que minimiza a variância total do portfólio, proporcionando um equilíbrio eficiente entre risco e retorno.

2.6.1. *Min-Variance*

Este modelo busca a alocação de ativos que resulta na menor volatilidade possível, dado um nível esperado de retorno.

Neste trabalho, a implementação do modelo *Min-Variance* é feita utilizando a biblioteca *PyPortfolioOpt* [Martin 2021], que oferece ferramentas robustas para otimização de portfólios financeiros com base em teorias financeiras.

Utilizar o *Min-Variance* como *benchmark* é crucial porque ele estabelece uma referência de desempenho em termos de risco mínimo para um dado nível de retorno. Isso permite avaliar o quão eficazes são outras abordagens de otimização de portfólio em comparação a um modelo bem estabelecido. Se uma nova abordagem conseguir um

desempenho superior ao *Min-Variance* em termos de métricas como o *Sharpe ratio*, ela pode ser considerada mais eficiente.

3. Resultados

Para a análise dos resultados, são realizados três experimentos distintos. No primeiro experimento, a avaliação é focada exclusivamente nos contratos WIN. No segundo experimento, a análise é dedicada aos contratos WDO. Finalmente, no terceiro experimento, a base de dados é avaliada considerando ambos os contratos, WIN e WDO. Cada experimento tem como objetivo explorar a eficácia das estratégias aplicadas em diferentes contextos de negociação.

Em cada experimento é realizada uma avaliação de desempenho de cada estratégia de DRL utilizando as métricas de desempenho mencionadas anteriormente. Também são feitas comparações com um *baseline*, representado pelo portfólio sem otimização. A escolha da estratégia de DRL é baseada no maior valor de *Sharpe ratio*, pois este índice avalia a relação entre o retorno acumulado e a volatilidade dos retornos, oferecendo uma medida ajustada ao risco.

Posteriormente, a estratégia de DRL selecionada é comparada com *benchmarks* estabelecidos para contextualizar os resultados alcançados. No experimento, os *benchmarks* utilizados são o IBOVESPA e a mínima variância em todas as comparações.

3.1. Avaliação de Desempenho das Estratégias de DRL

Como já mencionado na Seção III, os cinco algoritmos DRL são treinados para encontrar a melhor configuração de parâmetros utilizando a técnica de otimização de hiperparâmetros com 50 tentativas. As tabelas apresentam os agentes com a melhor configuração obtida, para cada experimento, apresentando seu resultado acumulativo no período de 06/06/2018 até 11/11/2019.

3.1.1. Desempenho das Estratégias no Contrato WIN

Tabela 1. Comparação de Desempenho entre os DRL - Contrato WIN

Metrics Avaliation	**A2C**	**DDPG**	**PPO**	**SAC**	**TD3**	***Baseline***
Annual Return	17.4%	19.1%	17.7%	18.7%	17.2%	17.7%
Annual Volatily	4.6%	4.8%	4.1%	4.3%	4.3%	4.2%
Sharpe Ratio	3.47	3.67	3.92	3.96	3.73	3.93
Max Drawdown	1.4%	1.38%	1.26%	1.3%	1.33%	1.26%

Na Tabela 1, os retornos anuais variaram de 17,2% a 19,1%, indicando uma performance robusta em um período relativamente estável. O DDPG se destacou com o maior retorno anual de 19,1%, enquanto TD3 apresentou o menor, com 17,2%. A volatilidade anual dessas estratégias é consistentemente baixa, oscilando entre 4,1% e 4,8%, o que sugere uma considerável estabilidade nas operações diárias, sendo o PPO o método com a menor volatilidade, enquanto a *baseline* manteve uma volatilidade próxima à média do grupo, com 4,2%. Quanto ao índice de Sharpe, que mede a relação entre retorno e risco, todas as

estratégias apresentaram valores superiores a 3,47. A *baseline* teve um ótimo desempenho, com índice de Sharpe igual à 3,93, apenas ligeiramente abaixo do SAC, que obteve o maior valor de 3,96. Esses valores indicam uma ótima eficiência ajustada ao risco. O *drawdown* máximo, que indica a maior queda de valor do portfólio antes de uma nova alta, manteve-se abaixo de 1,4% para todas as estratégias e também para a *baseline*, com o PPO mostrando o menor *drawdown* máximo de apenas 1,26%. Isso demonstra uma resiliência notável dos modelos de DRL contra potenciais quedas do mercado.

Entre os modelos de DRL analisados, o SAC é selecionado como o mais eficiente devido ao seu índice de Sharpe mais alto, evidenciando sua superioridade no controle de risco.

3.1.2. Desempenho das Estratégias no Contrato WDO

Tabela 2. Comparação de Desempenho entre os DRL - Contrato WDO

Metrics Avaliation	**A2C**	**DDPG**	**PPO**	**SAC**	**TD3**	***Baseline***
Annual Return	12.2%	14.6%	11%	14.4%	11.3%	11.2%
Annual Volatily	7.1%	7.2%	6.7%	7.6%	6.9%	6.7%
Sharpe Ratio	1.66	1.93	1.59	1.81	1.59	1.60
Max Drawdown	3.7%	4.1%	5%	4.7%	5.1%	3.7%

Observando a Tabela 2, os retornos anuais variaram significativamente entre as estratégias, com o DDPG apresentando o maior retorno de 14,6% e o PPO o menor, com 11%. As volatilidades anuais dessas estratégias também mostraram variações, indo de 6,7% a 7,6%, com o SAC apresentando a maior volatilidade e com a *baseline* igualando a menor volatilidade observada, de 6,7%. Isso sugere que a *baseline* conseguiu manter uma estabilidade comparável à das estratégias mais complexas. Em termos de índice de Sharpe, que mede o retorno ajustado ao risco, os valores ficaram entre 1,59 e 1,93. O DDPG liderou com a maior índice de Sharpe, indicando uma eficiência superior na gestão de riscos relativos aos retornos obtidos. A *baseline*, com um índice de Sharpe de 1,60, ofereceu uma eficiência razoável, superando o PPO e o TD3 neste aspecto. A análise do *drawdown* máximo mostra uma perda máxima no valor do portfólio, variando entre 3,7% e 5,1%, com o PPO e o TD3 exibindo os maiores *drawdown*.

Das estratégias avaliadas, o DDPG é escolhido como o mais eficiente devido ao seu índice de Sharpe superior, que evidencia uma excelente capacidade de gestão de risco.

3.1.3. Desempenho das Estratégias nos Contratos WIN e WDO Combinados

A Tabela 3 mostra que o retorno anual das estratégias variou de 13,5% e 17,6%. Dentre elas, o SAC obteve o maior retorno anual, enquanto o A2C apresentou o menor retorno anual. A *baseline* obteve um retorno de 15,7%, superando o A2C e posicionando-se competitivamente entre as demais estratégias DRL. Quanto à volatilidade anual, os valores oscilaram entre 3,8% e 4,4%, com o SAC novamente apresentando o menor valor, onde a volatilidade menor implica em menor risco. O índice de Sharpe das estratégias variou

Tabela 3. Comparação de Desempenho entre os DRL - Contratos WIN e WDO

Metrics Avaliation	A2C	DDPG	PPO	SAC	TD3	*Baseline*
Annual Return	13.5%	15.5%	16%	17.6%	16.9%	15.7%
Annual Volatily	4.3%	4.4%	4%	3.8%	4.3%	4%
Sharpe Ratio	2.95	3.28	3.73	4.24	3.67	3.70
Max Drawdown	1.2%	1.4%	1.2%	1.2%	1.3%	1.3%

de 2,95 a 4,24, indicando a eficácia do SAC, que registrou o maior valor, em maximizar o retorno por unidade de risco. A *baseline* alcançou um índice de 3,70, mostrando uma performance robusta, apenas ligeiramente abaixo do PPO e do TD3. Em termos de *drawdown* máximo, todas as estratégias, incluindo a *baseline*, mostraram uma resiliência significativa com *drawdown* máximo entre 1,2% e 1,4%, indicando que são capazes de minimizar significativamente as perdas potenciais durante o período avaliado.

3.2. Comparação com *Benchmarks*

Nesta seção serão apresentados os resultados da aplicação dos métodos de DRL em comparação com seus respectivos benchmarks envolvendo os contratos de WIN, WDO e combinados.

3.2.1. Comparação no Experimento com Contratos WIN

Conforme apresentado na Tabela 4, o SAC registrou um retorno anual de 18,7%, posicionando-se abaixo do IBOVESPA, que teve um retorno de 28%, e do *Min-Variance*, com 22,8%. A volatilidade anual é consideravelmente menor para o SAC e o *Min-Variance*, comparado com o IBOVESPA que apresentou uma alta volatilidade de 21,2%. O índice de Sharpe é superior para o *Min-Variance*, alcançando um valor de 5,22, indicativo de uma gestão de risco excepcionalmente eficiente. O SAC também mostrou eficiência com um índice de 3,96, enquanto o IBOVESPA teve o menor valor de 1,27, refletindo maior risco relativo aos retornos gerados. O *drawdown* máximo é consideravelmente menor para o SAC e o *Min-Variance* comparado ao IBOVESPA, que experimentou um *drawdown* máximo significativo de 11,37%.

Tabela 4. O melhor Agente, IBOVESPA e o *Min-Variance* - Contrato WIN

(06/06/2018 até 11/11/2019)	SAC	IBOV	*Min-Variance*
Annual Return	18.7%	28%	22.8%
Annual Volatily	4.3%	21.2%	3.9%
Sharpe Ratio	3.96	1.27	5.22
Max Drawdown	1.3%	11.37%	1.1%

Em termos de retorno anual, expresso em porcentagem, o IBOVESPA liderou com 28%, seguido pelo *Min-Variance* com 22,8% e o SAC com 18,7%. Como mostrado na tabela 4, esses resultados destacam as diferenças nas estratégias de investimento, onde o IBOVESPA proporciona maiores retornos totais, porém com riscos consideravelmente mais elevados.

3.2.2. Comparação no Experimento com Contratos WDO

De acordo com os dados da Tabela 5, o DDPG apresentou um retorno anual de 14,6%, posicionando-se entre os valores mais baixos de retorno quando comparado com o IBOVESPA, que teve um retorno significativo de 28%, e acima do DJI e do *Min-Variance*, com retornos de 7% e 11,3% respectivamente. A volatilidade anual do DDPG é de 7,2%, demonstrando uma estabilidade maior em comparação com o DJI e o IBOVESPA, e comparável ao *Min-Variance* com 7,3%. Em termos de índice de Sharpe, o DDPG alcançou 1,93, superior ao DJI com 0,52 e ao IBOVESPA com 1,27, mas inferior ao *Min-Variance* com 1,5. A análise do *drawdown* máximo revelou que o DDPG teve um *drawdown* de 4,1%, significativamente menor que a do DJI e a do IBOVESPA, e comparável à do *Min-Variance* com 4,5%. Isso destaca a capacidade do DDPG e do *Min-Variance* de limitar perdas potenciais de forma mais eficaz do que os índices de mercado mais voláteis.

Em termos de retorno anual, expresso em porcentagem, o DDPG também se destacou, acumulando um aumento de 14,6%, indicando um desempenho sólido, mas não alcançando o alto valor de 28% do IBOVESPA. O *Min-Variance* e o DJI acumularam 11,3% e 7%, respectivamente, mostrando um desempenho mais conservador em termos de ganhos totais.

A análise dos resultados da Tabela 5 evidencia a eficiência do DDPG, mesmo comparado com o IBOVESPA, destacando sua capacidade de capitalizar sobre as oportunidades de mercado em comparação com os *benchmarks* tradicionais e de mínima variância.

Tabela 5. O melhor Agente, IBOVESPA e o *Min-Variance* - Contrato WDO

(06/06/2018 até 11/11/2019)	**DDPG**	**IBOV**	***Min-Variance***
Annual Return	14.6%	28%	11.3%
Annual Volatily	7.2%	21.2%	7.3%
Sharpe Ratio	1.93	1.27	1.5
Max Drawdown	4.1%	11.37%	4.5%

3.2.3. Comparação no Experimento com Contratos WIN e WDO Combinados

O SAC alcançou um retorno anual de 17,6%, menor do que o IBOVESPA que registrou 28% e ligeiramente inferior ao *Min-Variance* com 21,1%. Apesar do menor retorno anual, o SAC demonstrou uma volatilidade anual extremamente baixa de 3,8%, equivalente à do *Min-Variance* e muito abaixo dos 21,2% do IBOVESPA. Esta baixa volatilidade indica uma maior estabilidade do SAC e do *Min-Variance* em comparação com o mais volátil, IBOVESPA. O índice de Sharpe do SAC é de 4,24, refletindo uma alta eficiência no ajuste do retorno pelo risco assumido, embora o *Min-Variance* tenha apresentado um índice ainda superior de 5. Por outro lado, o IBOVESPA, com um índice de Sharpe de 1,27, mostrou menor eficiência sob a mesma métrica. O *drawdown* máximo, que mede a maior queda do valor do portfólio antes de uma nova alta, é de apenas 1,2% para o SAC e 1,1% para o *Min-Variance*, significativamente menor que os 11,37% do IBOVESPA. Este resultado enfatiza a robustez do SAC e do *Min-Variance* em termos de gestão de riscos e limitação de perdas.

Tabela 6. O melhor Agente, IBOVESPA e o *Min-Variance* - Contratos WDO e WIN

(06/06/2018 até 11/11/2019)	SAC	IBOV	*Min-Variance*
Annual Return	17.6%	28%	21.1%
Annual Volatily	3.8%	21.2%	3.8%
Sharpe Ratio	4.24	1.27	5
Max Drawdown	1.2%	11.37%	1.1%

De acordo com a Tabela 6, em termos de valor final acumulado do portfólio, o SAC teve um aumento de 26,56%, comparado com 41,78% do IBOVESPA e 31,1% do *Min-Variance*. Apesar do IBOVESPA ter oferecido um retorno total mais elevado, isso veio com riscos consideravelmente maiores. Notavelmente, o índice IBOVESPA apresenta uma recuperação significativa a partir de meados de 2019, ultrapassando as demais estratégias no último trimestre do período observado, destacando sua capacidade de recuperação após quedas.

Os resultados ilustram a diversidade de desempenho entre diferentes estratégias de investimento, especialmente em contextos voláteis. As estratégias baseadas em RL mostram potencial para superar *benchmarks* tradicionais como o IBOVESPA em certos períodos, embora com variações significativas entre elas em termos de retorno e risco. O *Min-Variance*, embora ofereça a menor volatilidade, também proporciona os menores retornos, confirmando sua adequação para investidores que priorizam a preservação de capital sobre o crescimento.

4. Conclusão

Este estudo apresentou uma abordagem inovadora para otimização de portfólios de sistemas automatizados de negociação (ATS) usando algoritmos de Deep Reinforcement Learning (DRL), com foco específico no controle de riscos em ambientes de mercado altamente voláteis. As técnicas de DRL, particularmente os algoritmos DDPG e SAC, demonstraram uma capacidade notável de aprender e adaptar estratégias de negociação em tempo real, otimizando os retornos enquanto gerenciam eficientemente os riscos associados, e superando a *baseline* em vários aspectos.

Os resultados obtidos indicam que o uso de DRL pode superar significativamente os métodos tradicionais de negociação, como os baseados em heurísticas ou mesmo outros modelos quantitativos que não incorporam aprendizado contínuo e adaptação. A capacidade de processar e reagir a condições de mercado em tempo real, aprendendo com as interações passadas sem a necessidade de previsões explícitas, coloca os sistemas baseados em DRL como ferramentas promissoras para a modernização das práticas de negociação financeira.

Este trabalho não apenas demonstra a eficácia dos modelos de DRL na redução de riscos e na otimização do desempenho de portfólios, mas também aponta o potencial de aplicar essas técnicas em outras áreas financeiras, indicando um campo promissor para futuras investigações. Pesquisas futuras poderiam explorar a integração de técnicas de RL com outras modalidades de dados, como sinais econômicos macro ou análises de sentimentos, para desenvolver sistemas ainda mais robustos e adaptativos.

Portanto, conclui-se que a aplicação de técnicas avançadas de RL, como o DRL,

no campo das finanças, representa uma direção promissora e inovadora, com implicações substanciais para a teoria e prática de gestão de investimentos e operações de mercado.

Referências

Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Buehler, H., Gonon, L., Teichmann, J., and Wood, B. (2019). Deep hedging. *Quantitative Finance*, 19(8):1271–1291.

Buşoniu, L., De Bruin, T., Tolić, D., Kober, J., and Palunko, I. (2018). Reinforcement learning for control: Performance, stability, and deep approximators. *Annual Reviews in Control*, 46:8–28.

Chekhlov, A., Uryasev, S., and Zabarankin, M. (2005). Drawdown measure in portfolio optimization. *International Journal of Theoretical and Applied Finance*, 8(01):13–58.

Day Trade Review (2023). Best time frame for day trading - when and how to trade. Accessed: 2024-07-24.

Framework, O. (2013). Review of business and economics studies. *Studies*, 1(1).

Fujimoto, S., Hoof, H., and Meger, D. (2018). Addressing function approximation error in actor-critic methods. In *International conference on machine learning*, pages 1587–1596. PMLR.

Haarnoja, T., Zhou, A., Abbeel, P., and Levine, S. (2018). Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. *arXiv preprint arXiv:1801.01290*.

Liu, X.-Y., Yang, H., Chen, Q., Zhang, R., Yang, L., Xiao, B., and Wang, C. D. (2020). Finrl: A deep reinforcement learning library for automated stock trading in quantitative finance. *arXiv preprint arXiv:2011.09607*.

Liu, X.-Y., Yang, H., Gao, J., and Wang, C. D. (2021). Finrl: Deep reinforcement learning framework to automate trading in quantitative finance. In *Proceedings of the second ACM international conference on AI in finance*, pages 1–9.

Martin, R. A. (2021). Pyportfolioopt: portfolio optimization in python. *Journal of Open Source Software*, 6(61):3066.

Parker, K. and Fry, R. (2020). More than half of us households have some investment in the stock market.

Sutton, R. S. and Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

Treleaven, P., Galas, M., and Lalchand, V. (2013). Algorithmic trading review. *Communications of the ACM*, 56(11):76–85.

Yang, H., Liu, X.-Y., and Wu, Q. (2018). A practical machine learning approach for dynamic stock recommendation. In *2018 17th IEEE international conference on trust, security and privacy in computing and communications/12th IEEE international conference on big data science and engineering (TrustCom/BigDataSE)*, pages 1693–1697. IEEE.