


UM COMPARATIVO DOS MÉTODOS DE MARKOWITZ E DE APRENDIZADO POR REFORÇO PROFUNDO NA OTIMIZAÇÃO DE CARTEIRAS DE FUNDOS IMOBILIÁRIOS

A COMPARISON OF MARKOWITZ AND DEEP REINFORCEMENT LEARNING METHODS IN THE OPTIMIZING OF REAL ESTATE INVESTMENT FUND PORTFOLIOS

UNA COMPARACIÓN DE LOS MÉTODOS DE APRENDIZAJE DE REFUERZO PROFUNDO Y MARKOWITZ EN LA OPTIMIZACIÓN DE CARTERAS DE FONDOS DE INVERSIÓN INMOBILIARIA

 <https://doi.org/10.56238/rcsv16n3-005>

Data de submissão: 09/02/2026

Data de aprovação: 09/03/2026

Marcos Carvalho Corrêa Junior

Bacharel em Sistemas de Informação

Instituição: UPM Universidade Presbiteriana Mackenzie

E-mail: 10401844@mackenzista.com.br

Rogério de Oliveira

Doutor em Ciências Aplicadas

Instituição: Universidade Presb. Mackenzie (UPM), Inst. Mauá de Tecnologia (IMT)

E-mail: rogerio.oliveira@{mackenzie.br | maua.br}

Orcid: 0000-0001-5352-4420

Lattes: <https://lattes.cnpq.br/3067732992972770>

RESUMO

A otimização de carteiras tem por objetivo equilibrar riscos e retornos de diferentes ativos compondo uma carteira eficiente sendo essencial para gestão de investimentos. No mercado brasileiro de fundos imobiliários (FIIs), essa é uma tarefa complexa devido à heterogeneidade de liquidez e à influência de fatores macroeconômicos. Este trabalho compara e desenvolve a otimização de carteiras a partir de duas abordagens: o clássico modelo de Markowitz e o Aprendizado Profundo com Reforço (Deep Reinforcement Learning – DRL). São empregados dados históricos de FIIs listados na B3 entre 2020 e 2024, com liquidez mínima de R\$ 1 milhão/dia. Os resultados evidenciam a viabilidade da solução de Aprendizado por Reforço Profundo, embora soluções exatas ainda apresentem resultados melhores no curto prazo.

Palavras-chave: Otimização de Carteiras. Teoria Moderna do Portfólio. Fundos Imobiliários. Inteligência Artificial. Aprendizado por Reforço Profundo.

ABSTRACT

Portfolio optimization aims to balance the risks and returns of different assets by creating an efficient portfolio that is essential for investment management. The Brazilian real estate fund market (FIIs) is a complex task due to the heterogeneity of liquidity and the influence of macroeconomic factors. This work compares and develops portfolio optimization based on two approaches: the classic Markowitz model and Deep Reinforcement Learning (DRL). Historical data of FIIs listed in B3 between 2020 and 2024 are included, with minimum liquidity of R\$ 1 million/day. The results demonstrate the viability of the solution of Learned by Reforço Profundo, embora exatas solutions ainda presentem better results in the short term.

Keywords: Portfolio Optimization. Modern Portfolio Theory. Real Estate Investment Trusts. Artificial Intelligence. Deep Reinforcement Learning.

RESUMEN

La optimización de las carteras tem por objetivo equilibrar riesgos y retornos de diferentes activos compone una carteira eficiente sendo esencial para la gestión de inversiones. No hay mercado brasileño de fondos inmobiliarios (FII), es una tarefa compleja debida a la heterogeneidad de liquidez y a la influencia de factores macroeconómicos. Este trabajo compara e desarrolla una optimización de carteras a partir de dos abordajes: el modelo clásico de Markowitz y el Aprendizado Profundo con Reforço (Aprendizaje por refuerzo profundo – DRL). São empregados datos históricos de FIIs listados na B3 entre 2020 y 2024, con liquidez mínima de R\$ 1 millón/día. Los resultados evidencian la viabilidad de la solución de Aprendizado por Reforço Profundo, embora soluções exatas ainda apresentem resultados melhores no curto prazo.

Palabras clave: Optimización de Carteras. Teoría Moderna de Carteras. Fondos de Inversión Inmobiliaria. Inteligencia Artificial. Aprendizaje por Refuerzo Profundo.

1 INTRODUÇÃO

A gestão de investimentos é uma das áreas mais relevantes das finanças modernas, sendo responsável por definir estratégias de alocação de recursos que maximizem o retorno ajustado ao risco. Nesse contexto, a otimização de carteiras desempenha papel central, permitindo que investidores construam portfólios diversificados e alinhados ao seu perfil de risco. Com o avanço das técnicas quantitativas e o aumento da complexidade dos mercados, a busca por modelos capazes de oferecer decisões mais racionais e fundamentadas tornou-se um dos principais focos de pesquisa no campo das finanças computacionais.

A Teoria Moderna do Portfólio (TMP), desenvolvida por Markowitz (1952), estabeleceu as bases para essa abordagem quantitativa. O modelo introduziu o conceito de fronteira eficiente, demonstrando que é possível maximizar o retorno esperado de uma carteira para um determinado nível de risco por meio da diversificação e do cálculo das covariâncias entre os ativos.

Apesar de sua relevância teórica e ampla aplicação, a TMP apresenta várias limitações práticas indicadas na literatura. Bodie *et al.* (2014) apontam que a forte dependência de dados históricos para estimar retornos e covariâncias tende a gerar resultados instáveis quando aplicados a dados de mercado reais. Os pesos das carteiras obtidos a partir de estimativas amostrais de retorno e covariância tendem a ser sensíveis a erros de amostragem, e pequenas variações nos dados de entrada podem provocar mudanças expressivas na composição do portfólio, comprometendo sua estabilidade. No mercado brasileiro, especialmente no segmento de Fundos de Investimento Imobiliário (FIIs), essas limitações são agravadas pela menor liquidez média diária, pela concentração setorial e pela maior sensibilidade a fatores macroeconômicos, o que torna a construção de portfólios eficientes um desafio ainda maior.

Nesse cenário, o avanço das técnicas de Inteligência Artificial (IA) representa uma evolução significativa no processo de tomada de decisão em investimentos. Diversos estudos têm explorado o uso de algoritmos inteligentes para previsão de retornos, seleção de ativos e alocação de portfólios. Dentre essas técnicas, destacam-se as redes neurais artificiais, aplicadas na modelagem de séries temporais financeiras (ZHANG *et al.*, 1998), os algoritmos de aprendizado supervisionado, como árvores de decisão e *random forests*, empregados em classificação de tendências de mercado (GU *et al.*, 2020) e, mais recentemente, o Aprendizado por Reforço Profundo (*Deep Reinforcement Learning*, DRL), que combina redes neurais com aprendizado sequencial para otimização dinâmica de carteiras (DENG *et al.*, 2017).

Diferentemente dos modelos tradicionais, o DRL não depende de pressupostos estatísticos rígidos. O agente aprende por meio de interações sucessivas com o ambiente, ajustando as decisões de investimento com base nas recompensas acumuladas ao longo do tempo. Essa capacidade de aprendizado contínuo e adaptativo o torna especialmente promissor para contextos financeiros

marcados por incertezas, dinamicidade e não linearidades. Assim, o estudo tem como foco a aplicação prática das técnicas de otimização de carteiras baseadas no modelo clássico de Markowitz e no Aprendizado por Reforço Profundo (DRL), demonstrando como cada abordagem pode ser implementada computacionalmente e empregada no processo de alocação de recursos em fundos imobiliários.

2 REFERENCIAL TEÓRICO

O problema central da otimização de carteiras consiste em definir a distribuição ideal dos recursos entre diferentes ativos, de modo a equilibrar risco e retorno, considerando as incertezas inerentes aos mercados financeiros. Ao longo das últimas décadas, esse tema tem sido abordado sob duas grandes perspectivas teóricas: os métodos baseados na Teoria Moderna do Portfólio (TMP), que utilizam fundamentos estatísticos e o princípio da diversificação proposto por Markowitz (1952), e as técnicas de Inteligência Artificial (IA) aplicadas às finanças, que empregam algoritmos capazes de aprender padrões complexos nos dados e adaptar suas decisões às condições de mercado.

A Tabela 1 apresenta uma síntese de alguns trabalhos analisados, destacando as técnicas utilizadas, seus objetivos de otimização e as bases de dados empregadas.

Tabela 1 – Técnicas empregadas, objetivos de otimização e bases de dados de trabalhos correlatos.

Autor (Ano)	Técnica Utilizada	Objetivo da Otimização	Base de Dados
Vanelli & Neto (2024)	Algoritmo Genético	Sharpe Ratio	Ações Brasileiras
Melo (2022)	Deep Reinforcement Learning	Retorno Acumulado	Ações Americanas
Costa (2022)	Machine Learning	Risco-retorno ajustado	ETFs Internacionais
Carvalho (2021)	Markowitz	Diversificação e Retorno	Fundos Imobiliários
Tessari (2014)	Média e Mínima Variância	Desempenho da carteira	Ações Brasileiras
Souza (2017)	Markowitz (Solver/Excel)	Relação Risco-Retorno	Ações Brasileiras
Loke (2023)	Machine Learning	Relação Risco-Retorno	Mercados Simulados

Fonte: Elaborado pelos autores (2026).

Os trabalhos apresentados na Tabela 1 evidenciam o uso de diferentes abordagens para solucionar o problema de otimização de carteiras, incluindo técnicas de Inteligência Artificial, como redes neurais artificiais, aprendizado por reforço e algoritmos genéticos, além de métodos tradicionais baseados na Teoria Moderna do Portfólio. De forma geral, os estudos analisados demonstram a diversidade de metodologias aplicáveis e a busca por soluções que combinem fundamentos teóricos consolidados com ferramentas computacionais capazes de lidar com a complexidade dos mercados financeiros.

2.1 MÉTODO DE MARKOWITZ

A Teoria Moderna do Portfólio (TMP), proposta por Markowitz (1952), introduziu uma formulação quantitativa para a alocação eficiente de ativos, buscando o equilíbrio entre risco e retorno. Nessa estrutura, cada ativo do portfólio é caracterizado por seu retorno esperado, variância individual e covariância com os demais ativos. A relação entre essas variáveis permite estimar o risco total da carteira, não apenas pela soma dos riscos individuais, mas pela interação entre eles, princípio que fundamenta a diversificação eficiente.

No presente trabalho, o modelo de Markowitz foi implementado utilizando a biblioteca PyPortfolioOpt (MARTIN, 2025), que automatiza o cálculo dos retornos esperados e da matriz de covariância a partir dos dados históricos de preços. O processo de otimização consiste em encontrar o vetor de pesos que minimiza o risco total da carteira, sujeito a restrições definidas pelo usuário, como peso máximo por ativo e soma total dos pesos igual a 100%. A formulação empregada foi a de maximização da utilidade, equivalente à maximização do Índice de Sharpe (SHARPE, 1966), em vez da abordagem clássica de mínima variância. O Índice de Sharpe, mede a eficiência de uma carteira ao expressar quanto de retorno adicional é obtido para cada unidade de risco assumido. Quanto maior o valor do índice, mais eficiente é o portfólio, ou seja, ele oferece maior retorno ajustado ao risco.

A estimação dos pesos finais segue, portanto, um processo iterativo: os retornos médios são calculados a partir das séries de retornos diários dos fundos; a matriz de covariância é derivada diretamente dessas séries; e o algoritmo de otimização resolve o problema quadrático de maximização do Sharpe sob as restrições impostas. O resultado é uma carteira estática, cujos pesos permanecem fixos ao longo do período de análise, permitindo a comparação direta com o modelo dinâmico de Aprendizado por Reforço Profundo.

2.2 APRENDIZADO POR REFORÇO PROFUNDO

O Aprendizado por Reforço Profundo (*Deep Reinforcement Learning* – DRL) representa uma extensão do Aprendizado por Reforço tradicional (SUTTON & BARTO, 2018) combinando redes neurais profundas com políticas de decisão de reforço sequenciais. Em um contexto financeiro, o DRL busca aprender uma política de alocação ótima por meio de interação contínua com o ambiente de mercado. Diferentemente do modelo de Markowitz, que depende de estimativas estáticas de médias e covariâncias, o DRL ajusta suas decisões em tempo real, aprendendo pela experiência acumulada ao longo do treinamento.

O modelo desenvolvido foi implementado com a biblioteca Stable-Baselines3 (HILL, 2025), utilizando o algoritmo *Proximal Policy Optimization* (PPO). O ambiente de aprendizado foi configurado no framework Gymnasium (TOWERS *et al.*, 2025), onde o estado é composto por uma

janela deslizante de preços normalizados e pelos pesos correntes da carteira. A cada iteração, o agente executa alteração nos pesos dos ativos e recebe uma recompensa, definida como o retorno ajustado ao risco da carteira. Essa recompensa é positiva quando o portfólio apresenta bom desempenho e negativa quando o retorno é inferior ou a volatilidade se eleva acima do limite tolerado.

O processo de treinamento ocorre em ciclos: o agente observa o ambiente, executa uma ação, recebe a recompensa e ajusta sua política de decisão de forma a maximizar o retorno acumulado esperado. A função de otimização da rede neural busca maximizar a utilidade esperada do portfólio, incorporando as restrições impostas, como peso máximo por ativo, proibição de alocações negativas e taxa livre de risco. Dessa forma, o agente aprende não apenas a distribuir recursos entre os ativos, mas também a equilibrar o retorno e o risco de forma adaptativa.

Em termos práticos, o DRL substitui a formulação analítica da fronteira eficiente por um processo empírico e iterativo de aprendizado, em que o agente “descobre” políticas de investimento eficientes a partir da interação com dados históricos. Essa abordagem é especialmente útil em ambientes financeiros não lineares e dinâmicos, nos quais os parâmetros de risco e retorno variam ao longo do tempo, tornando o modelo clássico menos responsivo. Assim, o DRL oferece uma estrutura mais flexível, capaz de generalizar padrões de mercado e adaptar-se a novas condições sem necessidade de recalibração explícita.

3 METODOLOGIA

Os experimentos são conduzidos em três etapas: (i) coleta e preparação dos dados, (ii) aplicação de ambos os modelos de otimização, e (iii) análise dos resultados obtidos por meio de métricas de desempenho. A execução ocorre em uma interface interativa desenvolvida, permitindo selecionar os parâmetros de otimização, visualizar a composição das carteiras e comparar graficamente os resultados das duas abordagens.

3.1 COLETA E PREPARAÇÃO DOS DADOS

Foram utilizados dados diários de Fundos de Investimento Imobiliário (FIIs) listados na B3, abrangendo o período de janeiro de 2020 a agosto de 2025. Para o treinamento (construção dos modelos) foram empregados dados até dezembro de 2024, calculados os retornos diários e a estimação das volatilidades e covariâncias. Dados de janeiro de 2025 a agosto de 2025 são empregados para teste. A Tabela 2 apresenta a estrutura geral dos dados empregados nesta pesquisa.

Como o mercado de FIIs apresenta liquidez heterogênea, com volume de negociação concentrado em uma parcela reduzida dos fundos e spreads mais amplos nos ativos de menor liquidez, adotou-se um critério mínimo de elegibilidade, selecionando apenas fundos com média diária

negociada superior a R\$ 1.000.000,00. Após a aplicação desse filtro, 47 ativos foram mantidos na amostra, abrangendo diferentes segmentos do mercado imobiliário.

Tabela 2 – Estrutura Geral dos Dados. Preços diários ajustados dos ativos.

Data Pregao	ALZR11	BCFF11	BRCO11	BTLG11	...
2020-07-01	125,49	90,97	129,00	104,03	...
2020-07-02	122,80	90,74	127,90	104,20	...
2020-07-03	125,00	91,03	126,37	106,40	...

Fonte: Elaborado pelos autores.

3.2 IMPLEMENTAÇÃO DO MÉTODO CLÁSSICO DE MARKOWITZ

O método de Markowitz, é implementado em Python com a biblioteca PyPortfolioOpt (MARTIN, 2025). O modelo busca maximizar o retorno esperado para um dado nível de risco, ou, de forma equivalente, minimizar o risco total da carteira para um retorno-alvo predefinido, com base na relação média-variância.

O processo inicia-se com o cálculo dos retornos esperados, obtidos pela média histórica dos retornos diários dos ativos, e da matriz de covariância, que quantifica o grau de co-movimento entre os fundos. Em seguida, ocorre a seleção dos ativos que irão compor o portfólio, priorizando aqueles com maior Índice de Sharpe individual, de modo a favorecer combinações com melhor relação risco-retorno.

A otimização (classe EfficientFrontier da biblioteca PyPortfolioOpt), aplica os princípios da Fronteira Eficiente, considerando as restrições definidas pelo usuário, como peso máximo por ativo e retorno-alvo. O resultado é uma carteira cujos pesos atendem às restrições impostas e que maximiza o Índice de Sharpe, que indica o retorno excedente por unidade de risco total.

3.3 IMPLEMENTAÇÃO DO MÉTODO DE APRENDIZADO POR REFORÇO PROFUNDO

A implementação do modelo de Aprendizado por Reforço Profundo (DRL) emprega o framework Gymnasium (TOWERS *et al.*, 2025) que ajusta dinamicamente os pesos dos ativos a partir da interação entre o agente e o ambiente de investimento.

Nesse ambiente, o estado inclui uma janela deslizante de preços normalizados e os pesos correntes da carteira; as ações correspondem aos ajustes de alocação em cada ativo, normalizados por uma função softmax para garantir a soma total de 100% do capital; e a recompensa é definida como o retorno ajustado ao risco, baseado em uma proxy do Índice de Sharpe. O sistema penaliza automaticamente alocações que ultrapassem o peso máximo permitido por ativo, assegurando o cumprimento das restrições impostas.

O algoritmo *Proximal Policy Optimization* (PPO), disponível na biblioteca Stable-Baselines3 (HILL, 2025), é empregado pela sua robustez e estabilidade em ambientes contínuos de decisão. O agente é treinado sobre os retornos históricos (período de janeiro de 2020 a dezembro de 2024), utilizando o mesmo conjunto de ativos e parâmetros do modelo de Markowitz, o que permite uma comparação direta entre as abordagens. Ao final do treinamento, o modelo gera uma distribuição de pesos finais representando a política aprendida, que reflete o equilíbrio alcançado entre retorno e risco conforme os princípios do aprendizado por reforço.

3.4 PROCEDIMENTOS DE AVALIAÇÃO DOS RESULTADOS

A avaliação empírica foi conduzida por meio de *backtests* fora da amostra, abrangendo o período de janeiro de 2025 a agosto de 2025, com o objetivo de verificar a robustez dos modelos em dados não utilizados no treinamento. As carteiras geradas por cada método, Markowitz e Aprendizado por Reforço Profundo, foram testadas sob uma estratégia *buy-and-hold*, na qual os pesos são definidos no início do período e mantidos fixos até o final.

Foram realizados três experimentos variando a quantidade de ativos da carteira (6, 10 e 14), com o intuito de analisar o impacto da diversificação no desempenho e observar como cada método se comporta em contextos de maior ou menor restrição de alocação. Em cada cenário, os modelos foram avaliados com base em retorno anualizado, volatilidade, Índice de Sharpe (considerando a taxa livre de risco definida pelo usuário) e máximo *drawdown*, métrica que indica a maior perda acumulada no período.

4 RESULTADOS E DISCUSSÃO

Os resultados dos experimentos são apresentados na Tabela 3, que compara as métricas de desempenho: retorno anualizado, índice de Sharpe, máximo *drawdown* e volatilidade, obtidas pelos modelos de Markowitz e de Aprendizado por Reforço Profundo (DRL) em diferentes tamanhos de carteiras (6, 10 e 14 ativos). Em todos os casos, os parâmetros de entrada foram mantidos idênticos, permitindo uma avaliação direta da eficiência relativa entre as abordagens.

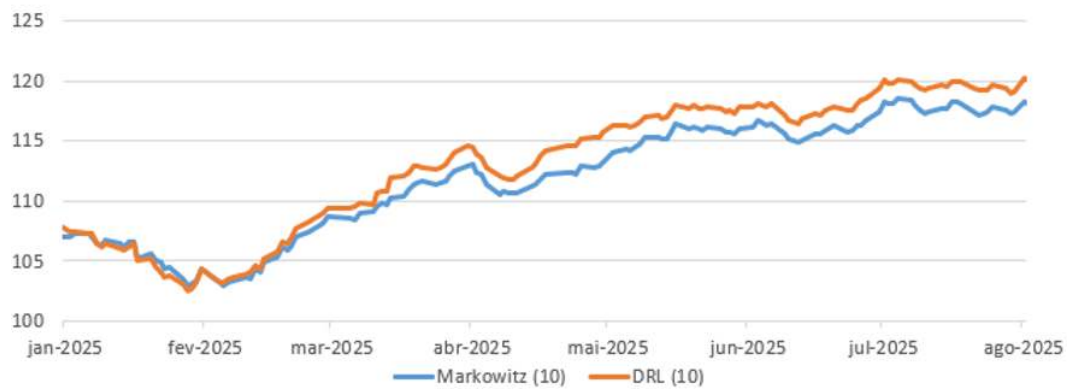
De modo geral, observa-se que ambos os métodos apresentam resultados consistentes e próximos, com diferenças marginais entre retorno e risco. O modelo de Markowitz demonstrou desempenho ligeiramente superior em termos de Índice de Sharpe, refletindo maior eficiência ajustada ao risco, enquanto o DRL apresentou ganhos discretos em retorno absoluto no cenário intermediário (Figura 1). Essa proximidade sugere que, para horizontes estáticos e carteiras com número moderado de ativos, o aprendizado por reforço tende a reproduzir padrões semelhantes ao modelo clássico, ainda que com maior flexibilidade na busca por soluções.

Tabela 3 – Métricas de desempenho para diferentes tamanhos de carteira para os modelos Markowitz e DRL.

Método - Quantidade Ativos	Retorno	Sharpe	Max Drawdown	Volatilidade
Markowitz - 6 Ativos	28,39%	2,37	-4,47%	7,53%
DRL - 6 Ativos	28,20%	2,36	-4,82%	7,48%
Markowitz - 10 Ativos	29,55%	2,49	-4,12%	7,53%
DRL - 10 Ativos	31,48%	2,54	-5,02%	7,52%
Markowitz - 14 Ativos	31,31%	2,53	-5,02%	8,21%
DRL - 14 Ativos	31,23%	2,39	-5,07%	8,69%

Fonte: Elaborado pelos autores.

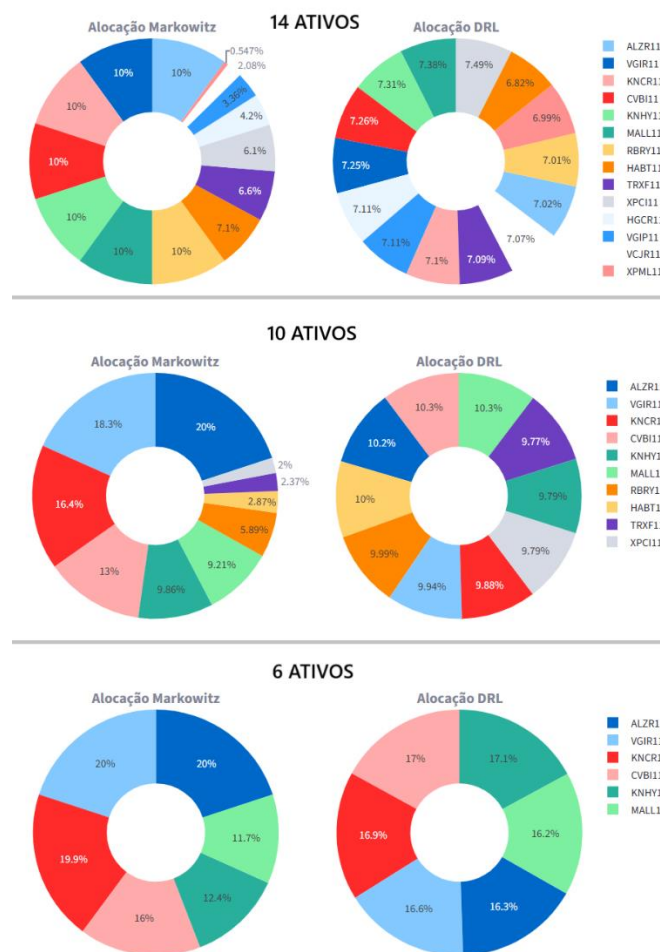
Figura 1 – Rentabilidade acumulada das carteiras otimizadas para o cenário de 10 ativos com ganhos discretos em retorno absoluto para a carteira DRL.



Fonte: Elaborado pelos autores.

A Figura 2 mostra a composição das carteiras otimizadas em cada experimento por cada modelo e permite comparar visualmente as alocações obtidas pelos dois métodos. O modelo de Markowitz tende a concentrar pesos em um número menor de fundos, refletindo sua estrutura analítica baseada na maximização do Índice de Sharpe e no posicionamento sobre a fronteira eficiente. Já o modelo de Aprendizado por Reforço Profundo (DRL) apresenta uma distribuição mais equilibrada e distribuída dos pesos, característica resultante da suavização das políticas aprendidas pelo agente durante o processo de treinamento. Essa diferença estrutural contribui para pequenas variações nos retornos obtidos, mas evidencia que ambos os métodos produzem carteiras diversificadas e coerentes com as restrições impostas.

Figura 2 – Comparação visual das alocações por ativos de ambos os modelos empregados.



Fonte: Elaborado pelos autores.

5 CONCLUSÃO

Esta pesquisa buscou avaliar o uso de técnicas baseadas em Inteligência Artificial na composição eficiente de carteiras de Fundos de Investimento Imobiliários no mercado brasileiro. Para isso, comparou os resultados do modelo clássico de Markowitz e o Aprendizado por Reforço Profundo (*Deep Reinforcement Learning* – DRL). A metodologia envolveu a construção de carteiras sob critérios de liquidez mínima, a definição de restrições de alocação e a condução de experimentos com diferentes tamanhos de portfólio. A avaliação empírica foi realizada por meio de *backtests* fora da amostra, calculando métricas como retorno anualizado, volatilidade, Índice de Sharpe e máximo *drawdown*.

Os resultados demonstraram que, em termos gerais, ambas as abordagens apresentaram desempenhos semelhantes, com o modelo de Markowitz exibindo leve vantagem na eficiência ajustada ao risco, enquanto o DRL apresentou ganhos marginais em retorno absoluto e maior suavização das alocações. Esses achados indicam que, para carteiras de tamanho moderado e horizontes estáticos, o modelo clássico permanece altamente competitivo. Por outro lado, a aplicação do DRL mostrou-se

bastante viável e promissora, evidenciando o potencial de aprendizado adaptativo em contextos financeiros mais dinâmicos ou com restrições não lineares.

Trabalhos futuros podem envolver o uso de estratégias de rebalanceamento periódico das carteiras, substituindo a abordagem estática de *buy-and-hold* por métodos que ajustem os pesos da carteira ao longo do tempo. A incorporação de custos de transação, restrições operacionais e simulações com janelas móveis possibilitaria também análises mais próximas da prática de mercado, parecendo promissor o uso de modelos de Aprendizado por Reforço para o desenvolvimento de modelos de aprendizado contínuo, capazes de atualizar políticas de alocação conforme novos dados chegam.

Os dados e códigos empregados neste estudo estão publicamente disponíveis no GitHub: <https://github.com/FII-MARKOTIWZ-REFORCO/RCS>.

REFERÊNCIAS

- BODIE, Z.; KANE, A.; MARCUS, A. J. *Investments*. McGraw-Hill Education, p. 242–322, 2014.
- DENG, Y. et al. Deep direct reinforcement learning for financial signal representation and trading. *IEEE Transactions on Neural Networks and Learning Systems*, v. 28, n. 3, p. 653–664, 2017.
- GU, S.; KELLY, B.; XIU, D. Empirical asset pricing via machine learning. *The Review of Financial Studies*, v. 33, n. 5, p. 2223–2273, 2020.
- HILL, A. *Stable-Baselines3: Reliable Reinforcement Learning Implementations*. Disponível em: <https://stable-baselines3.readthedocs.io/en/master>. Acesso em: 01 de outubro de 2025.
- MARKOWITZ, H. Portfolio selection. *The Journal of Finance* BV, p. 77–91, 1952.
- MARTIN, R. *PyPortfolioOpt: Portfolio Optimization in Python*. Disponível em: <https://pyportfolioopt.readthedocs.io/en/latest/UserGuide.html>. Acesso em: 01 de outubro de 2025.
- SHARPE, W. F. Mutual fund performance. *Journal of Business*, v. 39, n. 1, p. 119–138, 1966.
- SUTTON, R. S.; BARTO, A. G. *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press, 2018.
- TOWERS, M. et al. *Gymnasium: A Standard Interface for Reinforcement Learning Environments*. 2025. Disponível em: <https://arxiv.org/abs/2407.17032>. Acesso em: 02 de março de 2026.
- ZHANG, G.; PATUWO, B. E.; HU, M. Y. Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, v. 14, p. 35–62, 03 1998.