UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ARTES, CIÊNCIAS E HUMANIDADES
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM DE SISTEMAS COMPLEXOS

FERNANDO DANILO DE MELO

Otimização de portfólio: uma análise através de técnicas de *Reinforcement Learning* e *Autoencoders*

São Paulo
2022

FERNANDO DANILO DE MELO

Otimização de portfólio: uma análise através de técnicas de *Reinforcement Learning* e *Autoencoders*

Dissertação apresentada à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo para obtenção do título de Mestre em Ciências pelo Programa de Modelagem de Sistemas Complexos.

Área de concentração: Aprendizado de Máquina e Economia

Versão corrigida contendo as alterações solicitadas pela comissão julgadora em 19 de Setembro de 2022. A versão original encontra-se em acervo reservado na Biblioteca da EACH-USP e na Biblioteca Digital de Teses e Dissertações da USP (BDTD), de acordo com a Resolução CoPGr 6018, de 13 de outubro de 2011.

Orientador: Prof. Dr. Camilo Rodrigues Neto

São Paulo

2022

Ficha catalográfica elaborada pela Biblioteca da Escola de Artes, Ciências e Humanidades,
com os dados inseridos pelo(a) autor(a)
Brenda Fontes Malheiros de Castro CRB 8-7012; Sandra Tokarevicz CRB 8-4936

de Melo, Fernando Danilo
Otimização de portfólio: uma análise através de técnicas de Reinforcement Learning e Autoencoders / Fernando Danilo de Melo; orientador, Camilo Rodrigues Neto. -- São Paulo, 2022.
67 p: il.

Dissertacao (Mestrado em Ciencias) - Programa de Pós-Graduação em Modelagem de Sistemas Complexos, Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, 2022.
Versão corrigida

1. Aprendizado de Máquina. 2. Reinforcement Learning. 3. Otimização de Portfólio. 4. Mercado de Ações. 5. Criptomoedas. 6. Autoencoders. I. Rodrigues Neto, Camilo, orient. II. Título.

Dissertação de autoria de Fernando Danilo de Melo, sob o título **"Otimização de portfólio: uma análise através de técnicas de *Reinforcement Learning* e *Autoencoders*"**, apresentada à Escola de Artes, Ciências e Humanidades da Universidade de São Paulo, para obtenção do título de Mestre em Ciências pelo Programa de Pós-graduação em Modelagem de Sistemas Complexos, na área de concentração Sistemas Complexos, aprovada em ____ de _____ de _____ pela comissão julgadora constituída pelos doutores:

Prof. Dr.
Instituição
Presidente

Prof. Dr.
Instituição

Prof. Dr.
Instituição

Prof. Dr.
Instituição

Prof. Dr.
Instituição

Dedico esse trabalho a Fabrícia, Léia, meu irmão, meus pais, família e amigos, que sempre estão ao meu lado.

Agradecimentos

Primeiramente gostaria de agradecer a minha esposa pelo suporte, paciência e apoio incondicional agora e nos últimos onze anos, me incentivando todos momentos que parecia difícil seguir em frente.

Agradecer aos meus pais, irmão e família pelo apoio e incentivo a minha educação, a todo esforço que todos fizeram para que minhas realizações se tornassem possível.

Um agradecimento ao meu Orientador Prof. Dr. Camilo Rodrigues Neto, que desde a graduação me apoia no desafio de estudar o mercado de investimentos por meio de uma visão quantitativa e de modelagem.

Agradeço aos amigos Francisco Caio Lima Paiva, Eric Muszalska, Diego Bezerra Lira, entre outros, que me deram sugestões valiosas, que me ajudaram a enriquecer o trabalho com diversas discussões sobre o tema e tornaram possível várias abordagens que estão presentes nesse trabalho

Finalmente agradeço aos meus amigos, professores, meus colegas de trabalho e de mestrado, que me ouviram e me apoiaram durante todo esse período, tornando a experiência o mais enriquecedora possível.

Resumo

MELO, Fernando Danilo de. **Otimização de portfólio**: uma análise através de técnicas de *Reinforcement Learning* e *Autoencoders*. 2022. 67 f. Dissertação (Mestrado em Ciências) – Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, 2022.

Com o desenvolvimento dos algoritmos de *Reinforcement Learning* nos últimos anos, houve um aumento no número de estudos relacionados à negociação de ativos e otimização de portfólio. Embora trabalhos com dados de análise técnica e fundamentalista ganharam notoriedade nos últimos anos, poucos incluem ambos. Outro tema pouco explorado é o impacto do uso de *Autoencoders* para extrair variáveis e conexões entre os dados. Buscando explorar esses pontos e entender o impacto da introdução dessas variáveis, propomos um sistema inteligente para otimizar um portfólio por meio de análises de dados técnicos e fundamentalistas, bem como as variáveis geradas utilizando *Autoencoders*. Avaliamos o modelo em dois mercados distintos (o mercado Norte Americano de Ações e o de Criptoativos) em mais de 10 ativos, buscando avaliar o desempenho do agente em relação a modelos tradicionais. Posteriormente, esta avaliação permitiu-nos entender o impacto dos dados dos ativos em seu desempenho e como o agente se comporta em um mercado tradicional, como o de ações, e em mercados menos regulamentados, como o de criptomoedas.

Palavras-chaves: Aprendizado de Máquina. *Reinforcement Learning*. Otimização de Portfólio. Mercado de Ações. Criptomoedas. *Autoencoders*. Dados Técnicas. Dados Fundamentalistas.

Abstract

MELO, Fernando Danilo de. **Portfolio optimization**: an analysis through Reinforcement Learning techniques and Autoencoders. 2022. 67 p. Dissertation (Master of Science) – School of Arts, Sciences and Humanities, University of São Paulo, São Paulo, 2022.

With the development of Reinforcement Learning algorithms in recent years, there has been an increase in the number of studies related to trading and portfolio optimization. Although works with technical and fundamental analysis data have gained notoriety recently, few include both of them. Another little explored subject is the impact of using Autoencoders to extract variables and connections among data. Seeking to explore these points and understand the impact of introducing these variables, we propose an intelligent system for optimizing a portfolio via analyses of technical and fundamental data as well as the variables generated through Autoencoder. We evaluated the model in ten markets (U.S. Stocks and Crypto assets) and more than 10 assets with hourly data, seeking to assess the agent's performance about baselines. Subsequently, this evaluation allowed us to evaluate the impact of asset data on its performance and how the agent behaves in a more traditional market, such as stocks, and in less regulated markets, such as cryptocurrencies.

Keywords: Machine Learning. Reinforcement Learning. Portfolio Optimization. Stock Market. Cryptocurrencies. Autoencoders. Technical Data. Fundamental Data.

Lista de figuras

Lista de algoritmos

Lista de tabelas

Lista de abreviaturas e siglas

MPT	Modern Portfolio Theory
MV	Média-Variância
AE	Autoencoder
DRL	Deep Reinforcement Learning
MLP	Multilayer Perceptron
CNN	Convolutional Neural Network
HME	Hipótese do Mercado Eficiente
TA	Análise Técnica
FA	Análise Fundamentalista
OHLC	Preço de Abertura, Máxima, Mínima e Fechamento
TMC	Teoria do Mercado de Capitais
ATP	Arbitrage pricing theory
CAPM	Capital Asset Pricing Model
PMPT	Post-Modern Portfolio Theory
PPO	Proximal Policy Optimization
DDPG	Deep Deterministic Policy Gradient
PG	Policy Gradient
USD	Dólar Americano

Sumário

1 Introdução

Construir um portfólio é uma tarefa importante para qualquer um que queira investir, pois ele quer comprar ativos, como por exemplo, ações, moedas, criptomoedas e outros, de forma que o portfólio se torne lucrativo mesmo em mercados em baixa ou, pelo menos, perder o menor valor possível. O processo de alocação de ativos é constante, e a realocação pode ocorrer sempre que o gestor da carteira entender que há uma necessidade latente de mudança. Neste caso, ela pode ser a alteração de posição de ativos já adquiridos ou a venda e compra de novos ativos (REILLY; BROWN, 2011) .

Uma parte importante desse processo envolve a análise de dados históricos, técnicos e fundamentalistas buscando identificar padrões que podem ajudar a determinar tendências e o que comprar ou vender. No entanto, do ponto de vista humano, não é fácil analisar a enorme quantidade de dados relativos a cada ativo e determinar como os recursos devem ser alocados. Nesse contexto, podemos aplicar técnicas como *Deep Reinforcement Learning* (DRL) para entender os dados de mercado e encontrar relacionamentos que não estavam visíveis inicialmente, ajudando o processo de tomada de decisão.

Trabalhos recentes (ABOUSSALAH; LEE, 2020a; YE *et al.*, 2020; ABOUSSALAH; LEE, 2020b; BETANCOURT; CHEN, 2021) discutiram como as técnicas de *Reinforcement Learning* são aplicadas à otimização de portfólio, mostrando resultados promissores ao aplicá-las a processos contínuos de tomada de decisão para encontrar a melhor estratégia que maximize o valor do portfólio. Além disso, como parte dos esforços para melhorar a representação dos ativos e mercado, indo além dos preços dos ativos exclusivamente, alguns estudos exploraram indicadores técnicos ou fundamentalistas, ou uma combinação deles (ZHANG; MARINGER, 2016; YE *et al.*, 2020), mostrando que eles agregam em uma descrição mais efetiva do estado do mercado.

Este trabalho agrega diferentes tipos de dados de mercado, tornando as informações do agente mais completas. Nossas contribuições podem ser resumidas da seguinte forma:

- Propomos indicadores fundamentalistas e técnicos, agregando-os para ajudar a uma representação mais precisa do mercado.
- Tentando entender a relação subjacente entre os recursos e gerar novos recursos, usando *Autoencoders*.

- Examinamos nossa abordagem em um mercado de ações tradicional e criptomoedas, buscando entender como nosso agente se comporta em diferentes mercados e sua capacidade de generalização.

Organizamos nosso artigo da seguinte forma: Na seção 1.1, revisamos trabalhos relacionados a *Deep Reinforcement Learning* aplicado à otimização de portfólio. No capítulo 2, nós realizamos uma revisão dos fundamentos teóricos de Mercado Financeiro e otimização de portfólio. Na seção 3, definimos conceitos de *Machine Learning* e *Reinforcement Learning*. Nossa metodologia e resultados experimentais são apresentados na seção 4. Por fim, na seção 5, apresentamos nossas conclusões e apontamos para estudos futuros.

1.1 Trabalhos Relacionados

Moore *et al.* (1965), diz que o número de transistores em um chip iria dobrar a cada dois anos, mantendo o custo de produção, essa observação é conhecida como Lei de Moore, (BROCK; MOORE, 2006) e com esse aumento vertiginoso do poder computacional nos últimos 50 anos, podemos lidar e agregar muito mais informações em nossos modelos.

No caso do mercado financeiro, existem dados diários ou até na granularidade de milissegundos, informações técnicas da posição da ação e comportamento gráfico, dados fundamentalistas e até dados de notícia e análise de sentimento em tempo real. Outro ponto que essa evolução proporcionou é a utilização de técnicas mais avançadas de *Machine* e *Deep Learning*, que requerem muito poder computacional e dados para execução (MIIKKULAINEN *et al.*, 2019).

Estudos mostram que a utilização dessas técnicas, como Redes Neurais, podem ser promissoras na identificação padrões e na previsão do comportamento de um ativo, (VAN; ROBERT, 1997), (CHENG; WAGNER; LIN, 1996), (TAY; CAO, 2001), esses resultados se dão porque segundo Hornik, Stinchcombe e White (1989) as Redes Neurais podem ser consideradas aproximadores universais, podendo mapear qualquer função não linear. Além disso, elas lidam bem com o ruído nos dados, caudas longas e são mais flexíveis que os métodos tradicionais.

Porém, por mais que esses algoritmos tenham se mostrados eficientes em entender e prever o comportamento de ativos, ainda existem ações e decisões a serem tomadas, como a compra e venda, quando e qual ativo a ser utilizado. Logo, é necessária outra

camada para negociação que leve em conta também os riscos da operação que não estão relacionados ao valor do ativo. Entre eles, os custos de negociação, o volume de negociação do ativo, adaptabilidade as condições futuras do mercado, etc, (SHAO; KIM; IMRAN, 2016).

Para essa camada adicional de tomada de ação e decisão, alguns estudos como o de Jiang, Xu e Liang (2017) têm mostrado resultados promissores em lidar com esses pontos utilizando técnicas de *Reinforcement Learning*. Esses métodos, diferentemente de *Machine/Deep Learning*, onde ele busca classificar ou prever um valor, exploram o aprendizado e execução de ações que busquem maximizar o retorno para cada situação. Estudos recentes utilizando Redes Neurais dentro de *Reinforcement Learning* para ajudar a estimativa de parâmetros mostram que essa combinação de técnicas tem se mostrado promissora na Otimização de Portfólios e tem sido superior aos modelos tradicionais (FILOS, 2019).

Em termos de problemas de otimização de portfólio, as técnicas de *Reinforcement Learning* parecem ser uma solução viável, pois mapeiam observações de cada ativo e ambiente para uma estratégia que busca ser lucrativa. A primeira aplicação de *Reinforcement Learning* para otimização de portfólio feita por Neuneier (1995), Neuneier (1997) utiliza o método *Q-learning*, no qual busca encontrar a função ótima $Q^*(S_t, A_t)$.

Outros trabalhos utilizam Redes Neurais para aproximar a política (DING *et al.*, 2018; DENG *et al.*, 2016; ABOUSSALAH; LEE, 2020a) a ser aprendida tanto para aquisição e venda de ativos, quanto para a otimização de sistemas de negociação e portfólios. Em resposta à influência desses trabalhos e ao surgimento de técnicas destinadas a aprimorar os métodos *Actor-Crtic*, vários trabalhos foram lançados explorando essas técnicas (YE *et al.*, 2020; LI; ZHENG; ZHENG, 2019; YU *et al.*, 2019; LIANG *et al.*, 2018). Este artigo realiza uma revisão de trabalhos que utilizam métodos de *Policy Gradient*, como *Deep Deterministic Policy Gradient (DDPG)* (LILLICRAP *et al.*, 2015) e *Proximal Policy Optimization (PPO)* (SCHULMAN *et al.*, 2017).

- **Indicadores Técnicos e Fundamentalistas**: (ALIMORADI; KASHAN, 2018; DANTAS; SILVA, 2018; EILERS *et al.*, 2014) como característica que nos ajudam a encontrar uma melhor definição dos ativos e do mercado, fornecendo uma visão mais ampla das ações que serão tomadas permitindo que o agente seja mais assertivo em cada passo.

- **Criptomoedas e Ações**: buscando testar as capacidades de generalização dos agentes, vamos explorar a robustez do modelo analisando dois mercados diferentes, sendo eles, o Mercado de ações com as dez maiores empresas do SP500 (ABOUSSALAH; LEE, 2020a) e criptomoedas (JIANG; LIANG, 2017; WENG *et al.*, 2020; YE *et al.*, 2020). Vale ressaltar que o processo de seleção de ativos é explorado na seção 4.4.
- ***Autoencoders*** (LI; ZHENG; ZHENG, 2019; PARK; SIM; CHOI, 2020): onde iremos explorar a capacidade dos *Autoencoders* de amenizar o ruído nos dados e sua capacidade de reconhecer a relação subjacente entre eles, gerando novos recursos.
- ***Proximal Policy Optimization*** (SCHULMAN *et al.*, 2017): também conhecido como PPO, é um dos algoritmos mais recentes lançado. Uma de suas principais características é limitar o tamanho da atualização da política. Isso ocorre tomando a proporção entre as políticas novas e antigas e limitando essa proporção por meio de um hiperparâmetro ϵ.

  Destaca-se que o modelo é abordado com maiores detalhes na seção 3.3.3.

Embora essas técnicas já tenham sido utilizadas em outros trabalhos, ainda não foram exploradas juntamente na otimização do portfólio. Por isso, buscamos contribuir com a área explorando o impacto da combinação de indicadores técnicos, fundamentalistas e os dados gerados pelo *Autoencoder*, com uma técnica de ponta como o PPO.

2 Fundamentação Teórica

No capítulo anterior fizemos uma revisão histórica passando pelos trabalhos mais relevantes. Agora iremos nos aprofundar na definição e formulação dos conceitos abordados pelos modelos clássicos.

2.1 Sistemas Complexos

Podemos definir um sistema como complexo a partir de caraterísticas que eles apresentam. Segundo Boccara (2010), a principal delas é **emergência**, ou seja, um fenômeno que emerge por meio das interações entre os agentes, que não era claro a partir do comportamento de suas partes individualmente. Porém temos outras características que surgem em sistemas complexos, como:

- **Adaptação**: o agente muda seu comportamento de maneira a se adaptar ao ambiente.
- **Não-estacionariedade**: não se pode assumir que a dinâmica e características passadas, serão as mesmas para o futuro.
- **Feedback**: a presença dessa característica implica que o sistema pode responder a uma ação passada e manter o padrão no futuro de maneira não trivial.
- Um grande número de **agentes** interagindo entre si.
- **Auto-Organização**: nesse caso, as ações individuais dos agentes geram ordem em escala global.

O mercado financeiro detém várias dessas características. Com isso podemos considerá-lo um Sistema Complexo, onde a dinâmica se dá pela interação dos agentes com os ativos (LADYMAN; LAMBERT; WIESNER, 2013).

2.2 A imprevisibilidade do mercado

Em 1900, Bachelier publica seu trabalho *Théorie de la spéculation*, onde ele propõe que sendo $P(t)$ o preço de um ativo no período t, as diferenças do preço ativo com um momento posterior $P(t+T) - P(t)$ é independente (MANDELBROT, 1997).

Esse comportamento observado por Bachelier, é descrito como uma *Random Walk* (termo cunhado por Pearson (1905)), podendo ser representado por uma curva Gaussiana ou Normal.

2.2.1 Teoria do Mercado Eficiente

Kendall e Hill (1953) examinam o comportamento semanal dos preços da *British Industrial*, onde segundo o autor:

> "once a week the Demon of Chance drew a random number from a symetrical population of fixed dispersion and added it to the current price to determine the next week's price".

Por mais que as evidências suportassem uma *Random Walk* para os retornos, não existia uma formalização do significado econômico do porquê o mercado se comportava dessa maneira.

Em 1965 e 1966, Delcey (2019) e Mandelbrot (1966), respectivamente, apresentam paralelamente trabalhos que descreveram a aleatoriedade como uma martingale, ou seja, o momento $T + 1$ só depende do momento T e nenhum momento anterior. Ele define que, sendo o preço futuro do ativo Z_{t+T}, correspondendo t o valor no momento de avaliação e T o momento futuro, como podemos observar a seguir

$$E(Z_{t+T}|Z_t) = Z_t, \tag{1}$$

podemos observar que o retorno esperado depende somente do preço no momento t.

Porém, até 1970, não se tinha o entendimento desse fenômeno. Isso muda com Fama (1970), que apresenta a Hipótese do Mercado Eficiente (HME), onde formaliza que o preço de mercado das ações é uma avaliação racional do valor da empresa dadas as informações disponíveis. De acordo com essa teoria, o mercado é eficiente em termos de informação, ou seja, o valor de uma ação reflete toda informação disponível sobre a companhia.

Fama (1970) divide o mercado em três tipos:

- ***Weak form***: não se pode tirar proveito de informações ou indicadores técnicos passados para auferir melhores retornos no futuro, porém, informações fundamentalistas podem ajudar a melhorar os retornos futuros em relação ao retorno médio do mercado.

- ***Semi-Strong form***: o valor do ativo reflete todas informações públicas disponíveis se adaptando rapidamente a novas informações, porém, informações que ainda não foram disponibilizadas ao público podem ajudar a melhorar os retornos.
- ***Strong form***: o valor do ativo reflete todas informações públicas e privadas disponíveis.

2.3 Risco, Incerteza, Prêmio de Risco e Ativos Livres de Risco

Frank (1921) diferencia o risco de incerteza de maneira que o risco pode ser observado, suas probabilidades calculadas, se comportando de maneira estável. Já incerteza é aquilo que não pode ser previsto e que suas probabilidades não podem ser calculadas.

Podemos definir que investimento é o retorno do capital comprometido em algum ativo, de maneira a compensar: (a) o tempo do investimento, (b) a inflação do período e (c) o risco envolvido, (REILLY; BROWN, 2011).

Quando falamos de utilizar capital em busca de um retorno ao longo do tempo, temos que selecionar ativos que façam sentido para o perfil do investidor. Nesse caso, podemos escolher ativos com retorno variável e maior risco, como ações, ou os ativos conhecidos como Livre de Risco. Tais ativos são conceituais, onde se conhece o retorno com precisão e sem influência de risco, ou seja, o retorno esperado é igual ao real (DAMODARAN, 1999).

Porém, nem todos os investimentos são livres de risco e em alguns casos o investidor pode aceitar um risco maior, esperando um maior retorno, chamado de Prêmio de Risco.

Alguém que invista R$ 100 hoje, e espera um retorno livre de risco de 5% em um ano, teria no final do período R$ 105, porém, nesse caso, não consideramos a inflação. Caso tenhamos uma inflação de 3% e o investidor espera os mesmos 5% de retorno, a taxa nominal do investimento deveria ser de 8%, ou seja, R$ 108 se considerarmos o exemplo descrito acima.

Existem diversos tipos de investimento e vários tipos de riscos atrelados a eles, segundo (REILLY; BROWN, 2011), podemos defini-los, como:

- **Risco do Negócio**: que esta relacionado da natureza do negócio. Nesse caso temos negócios com uma operação e fluxo de caixa mais estável durante o ano, como um banco. Em contrapartida. temos empresas que sofrem de sazonalidades durante o ano, como as do ramo hoteleiro.

- **Risco Financeiro**: podemos relacionar esse risco a maneira que a empresa financia seus próprios investimentos, se ela faz operações de crédito para fomentar crescimento, ela adquiriu uma obrigação, que se sobrepõe as obrigações de retorno com investidores.
- **Risco na Liquidez**: quando um investidor compra um ativo de um mercado secundário, ele espera poder vender esse ativo quando julgar interessante (nesse caso, evitando perdas, ou realizando ganhos), porém devido às características dessas operações, estamos expostos ao risco de não conseguir realizar a operação no momento que optamos por ela. Sendo assim, podemos ter um preço de venda/compra diferente e também receber/comprar em uma janela de tempo diferente da que queremos.
- **Risco Cambial**: é o risco que o investidor tem quando compra um ativo em uma moeda diferente da sua, estando sujeito a efeitos de valorização ou desvalorização da moeda. Com isso, alterando a taxa de retorno do ativo.
- **Risco Política**: é o risco relacionado as alterações que políticas podem trazer para um país.

Logo, podemos entender que o processo de investimento está diretamente ligado ao risco assumido. No caso do retorno esperado, quando assumimos um risco maior, esperamos também um Prêmio de Risco maior e quando montamos um portfólio otimizado, buscamos o maior retorno dentro do risco que o investidor aceita assumir.

2.4 As Teorias Clássicas de Alocação de Capital

A literatura de otimização de portfólio é ampla e data do começo da década de 50. Nesse capítulo iremos realizar uma revisão histórica passando por alguns dos trabalhos mais relevantes no que diz respeito a formulação teórica da alocação de portfólio.

2.4.1 Modern Portfolio Theory

Markowitz (1952) apresenta o trabalho *Portfolio Selection*, baseado no trabalho de Williams (1938), que propõe que o investidor maximiza o valor descontado dos retornos futuros, porém segundo Markowitz essa abordagem é incorreta, pois se considerado somente

o retorno futuro, ignorando o fator de risco, não existiria motivo para diversificação de portfólio.

Dessa maneira, Markowitz assume duas regras:

- O objetivo do investidor é maximizar o retorno para determinado nível de risco.
- O risco pode ser minimizado com por meio da diversificação do portfólio entre ativos não correlacionados.

Portanto, ele assume que o investidor busca maximizar o retorno esperado e diminuir a variância, ou seja, o risco, o fator indesejável. Dessa maneira, carteiras devem buscar a menor variância por meio da diversificação dos ativos.

O autor supracitado sustenta que um modelo otimizado de portfólio deve ser construído considerando que o risco da carteira não é somente a média dos ativos, mas a correlação entre eles. Também deve se levar em consideração que o desempenho analisado não é o individual, mas da carteira como um todo.

Além desses conceitos, Markowitz (1952) introduz a variância do retorno do portfólio, como critério para seleção do mesmo. Com a apresentação desse trabalho, se deu início ao que é chamado de *Modern Portfolio Theory - MPT* (Teoria Moderna do portfólio, em tradução literal).

Inicialmente o retorno esperado do portfólio é definido como:

$$E(R_{port}) = \sum_{i=1}^{n} w_i E(R_i) \tag{2}$$

sendo, w_i o peso do ativo i, e $E(R_i)$ é o retorno esperado para o ativo i.

Com o retorno definido, podemos calcular a variância e o desvio padrão de cada ativo e do portfólio, a variância de cada ativo separadamente:

$$Variância = \sigma^2 = [R_i - E(R_i)]^2 P_i. \tag{3}$$

O Desvio padrão (Dp):

$$Dp = \sigma = \sqrt[2]{[R_i - E(R_i)]^2 P_i}, \tag{4}$$

onde, R_i representa uma possível taxa de retorno, $E(R_i)$ o retorno esperado e P_i a probabilidade de retorno do ativo i.

Com a introdução do MPT, a variância de um portfólio passou a considerar a relação entre os ativos. Com isso temos que entender dois novos conceitos, *covariância* e *correlação*.

Segundo Walpole (2009), *covariância* descreve a relação linear entre duas variáveis, a covariância dos retornos de dois ativos, em termos matemáticos:

$$Cov_{ij} = E[R_i - E(R_i)][R_j - E(R_j)]. \tag{5}$$

Porém a covariância é dependente de escala, sendo de difícil interpretação e comparação do resultado. Buscando padronizar, temos o *coeficiente de correlação*, que nada mais é que a covariância dividida pelo desvio padrão das variáveis. Para dois ativos podemos calcular da seguinte maneira:

$$Corr_{ij} = \frac{Cov_{ij}}{\sigma_i^2 \sigma_j^2}. \tag{6}$$

Com esses dois conceitos apresentados, Markowitz define a variância do portfólio como:

$$\sigma_p^2 = \sum_{i=1}^{n} w_i^2 \sigma_i^2 + \sum_{i=1}^{n} \sum_{j \neq i}^{n} w_i w_j Cov_{ij}. \tag{7}$$

E o desvio Padrão:

$$\sigma_p = \sqrt[2]{\sum_{i=1}^{n} w_i^2 \sigma_i^2 + \sum_{i=1}^{n} \sum_{j \neq i}^{n} w_i w_j Cov_{ij}}, \tag{8}$$

sendo, σ_p^2 é a variância do portfólio, σ_p o desvio padrão, w_i o peso de cada ativo, sendo que $\sum w = 1$ e σ_i^2 a variância do ativo i , e Cov_{ij} a covariância entre os ativos. Podemos observar que a equação é dividida em duas partes, a primeira, $\sum_{i=1}^{n} w_i^2 \sigma_i^2$, é o fator responsável por cada ativo, na segunda parte, $\sum_{i=1}^{n} \sum_{j \neq i}^{n} w_i w_j Cov_{ij}$, é o fator de relação entre os ativos.

Substituindo a Equação 6 em 8, temos:

$$\sigma_p^2 = \sum_{i=1}^{n} w_i^2 \sigma_i^2 + \sum_{i=1}^{n} \sum_{j \neq i}^{n} w_i w_j \sigma_i^2 \sigma_j^2 Corr_{ij}. \tag{9}$$

Se mantivermos os demais fatores, a variância da carteira é maior quanto maior ou mais positiva a correlação entre os ativos, e menor, quanto menor ou mais negativa. Sendo assim, é preferível ativos com correlação baixa ou negativa.

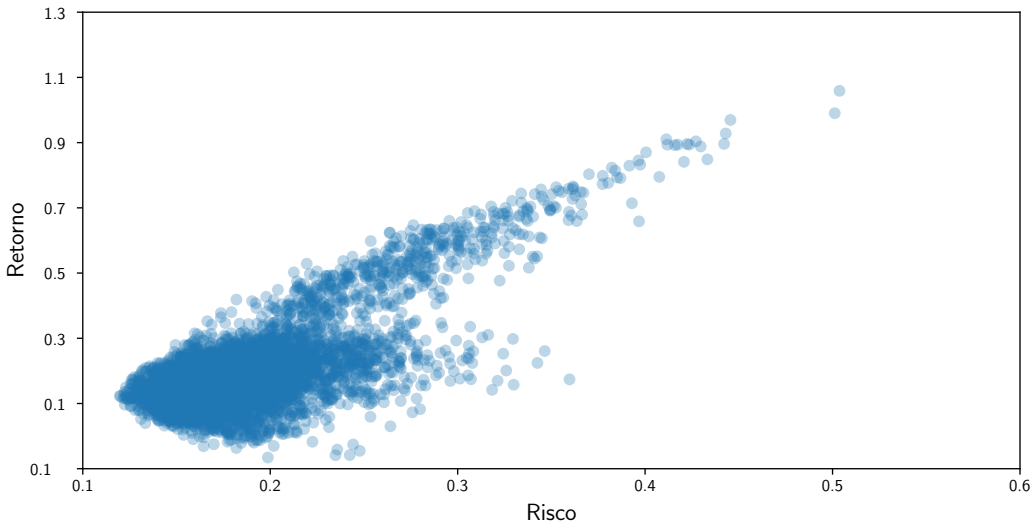
Observando a Equação 2, o retorno dos ativos não depende da relação entre eles, logo buscamos ativos para o portfólio que busquem manter nível de retorno, porém minimizando

o risco do portfólio. Dessa maneira uma carteira não perfeitamente correlacionada oferece um risco-retorno melhor do que cada componente por si só (MARCUS; BODIE; KANE, 2013).

Na figura 1 são apresentados o risco-retorno de diversos portfólios construídos buscando ilustrar os conceitos discutidos anteriormente. Randomizando os pesos em cada carteira, podemos observar que o risco e o retorno podem variar bastante. Dessa forma, todas as carteiras com melhor retorno para determinado nível de risco formam o que é conhecido fronteira eficiente.

Nessa curva estão todos portfólios ótimos na relação risco-retorno, cabendo ao investidor a escolha da carteira que melhor supre suas necessidades.

Figura 1 – Relação do risco-retorno de diversas carteiras construídas com o MPT.



Fonte – Fernando Melo, 2021

2.5 Teoria do Mercado de Capitais

Segundo Jensen (1972), a Teoria do Mercado de Capitais (TMC) busca entender quais as implicações da introdução de um título livre de risco em uma carteira de Markowitz, ela parte de algumas premissas:

- Todos investidores são eficientes em relação ao modelo de Markowitz.
- Investidores podem comprar e emprestar dinheiro a uma taxa livre de risco.
- Todos investidores tem expectativas homogêneas em relação ao retorno.

- Todos investimentos são infinitamente divisíveis, ou seja, podemos comprar e vender frações dos ativos.
- Não são consideradas taxas, custos transacionais, alterações na inflação, o mercado está em equilíbrio.

2.5.1 Linha do Mercado de Capitais

No final da década de 50, Tobin (1958) introduz um ativo livre de risco (iremos utilizar rf para facilitar a notação) no MPT. Esse ativo tem retorno conhecido, logo a variância σ_{rf}^2, é 0, se considerarmos a equação 5, a covariância de qualquer ativo com um ativo livre de risco também é 0. Todavia, o que acontece se combinarmos um ativo livre de risco com um portfólio? Utilizando a equação 7, o termo $w_{rf}w_pCov_{rfp}$ é 0, devido a $\sigma_{rf}^2 = 0$, sendo o peso do portfólio com um ativo de livre risco, $w_p = 1 - w_{rf}$, dessa maneira podemos definir a variância do novo portfólio n, com a adição de um ativo livre de risco como,

$$\sigma_n^2 = (1 - w_{rf})^2\sigma_p^2 + w_{rf}^2\sigma_{rf}^2, \tag{10}$$

de modo que $\sigma_{rf}^2 = 0$,

$$\sigma_n^2 = (1 - w_{rf})^2\sigma_p^2. \tag{11}$$

O desvio padrão:

$$\begin{aligned} \sigma_n &= \sqrt[2]{(1 - w_{rf})^2\sigma_p^2} \\ &= (1 - w_{rf})\sigma_p. \end{aligned} \tag{12}$$

Utilizando as equações acima em conjunto com a 2, podemos chegar que o retorno do portfólio junto com um ativo livre de risco é:

$$E(R_p) = R_{rf} + \sigma_p \left[\frac{E(R_n) - R_{rf}}{\sigma_n} \right]. \tag{13}$$

Analisando a equação 13, observamos que o retorno de um investidor que distribui os recursos entre ativos livres de risco e de risco, é o retorno do primeiro mais o retorno do segundo por unidade de risco assumida, que como definido anteriormente, é o Prêmio de Risco.

Essa equação é chamada de Linha de Mercado de Capitais. Com ela podemos construir uma linha com carteiras compostas por ativos de risco e livres de risco, que relaciona o retorno e o risco.

2.5.2 Portfólio de Mercado

Portfólio de mercado é um modelo conceitual que contém a todos ativos de risco do mercado, onde os pesos são proporcionais a sua relevância no mercado.

Logo todo o risco não sistemático de cada ativo foi completamente diversificado, restando somente o risco sistemático ou não diversificável. O risco sistemático pode ser calculado a partir da equação 8.

2.5.3 Capital Asset Pricing Model

Novos modelos surgiram para simplificar o cálculo da covariância, como Modelo de Índice Único, por Sharpe (1963), onde ele decompõe o risco entre sistemático e específico da empresa, com isso o número de cálculos a serem feitos diminui e o portfólio se torna mais assertivo.

Com a diversificação, o MPT ajuda a eliminar o risco específico de cada empresa, porém não resolve o problema do risco sistemático do mercado. Em meados da década de 60, Sharpe (1964), Lintner (1965) e Mossin (1966) apresentam quase que ao mesmo tempo o Modelo de Precificação de Ativos Financeiros (*Capital Asset Pricing Model - (CAPM)*). Eles utilizam o modelo de Markowitz como base. O método ajuda a calcular o retorno de um ativo em relação ao risco sistemático.

No MPT e na Teoria do Mercado de Capitais, o risco é tratado como um todo. No CAPM consideramos somente o risco não diversificável, que é chamado de **coeficiente beta** (β). Se observarmos a equação 13, poderíamos substituir a variância σ_p pela variância σ_i do ativo de risco i, multiplicado pela correlação desse ativo com o portfólio de mercado M, dessa maneira

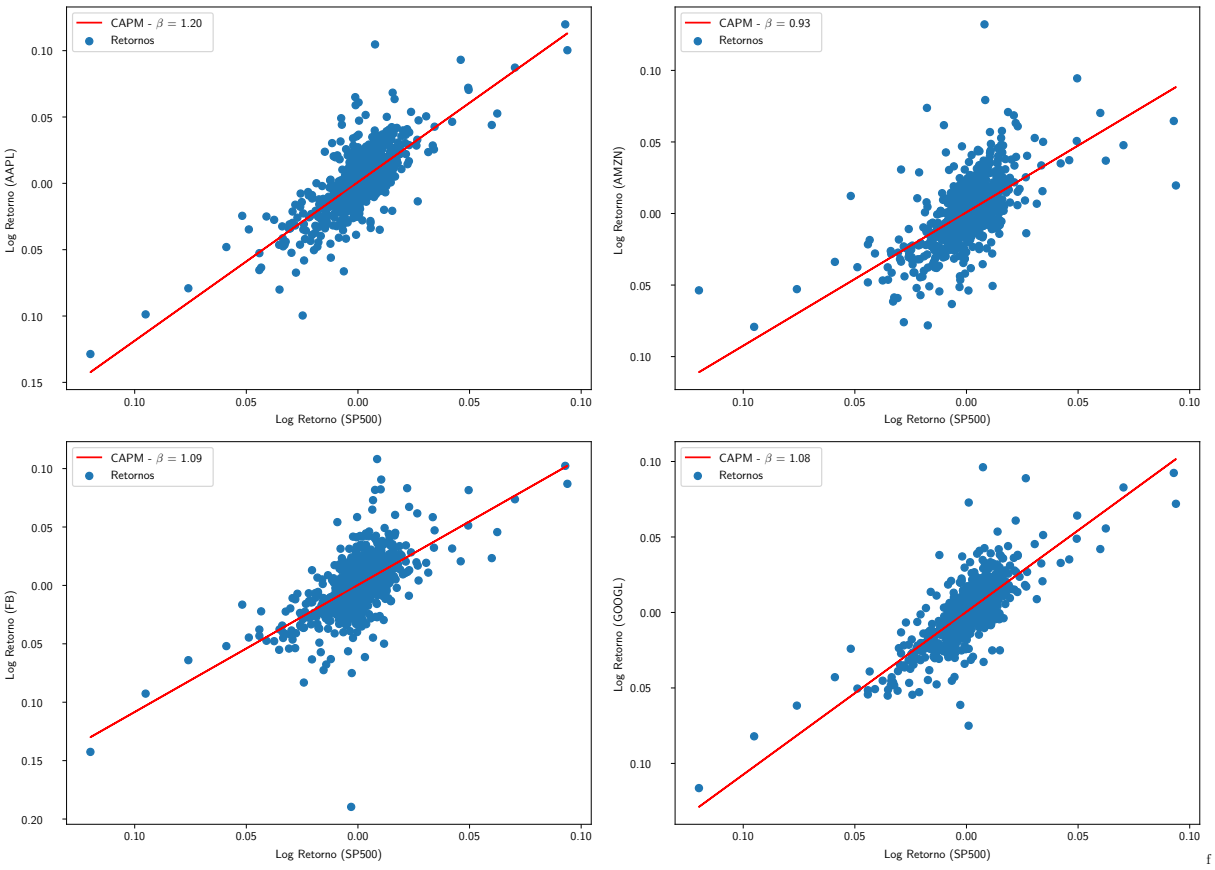
$$E(R_i) = R_{rf} + \sigma_i Corr_{iM} \left[\frac{E(R_M) - R_{rf}}{\sigma_M} \right]. \tag{14}$$

Ajustando:

$$\begin{aligned} E(R_i) &= R_{rf} + \left(\frac{\sigma_i Corr_{iM}}{\sigma_M}\right)[E(R_M) - R_{rf}] \\ &= R_{rf} + \beta_i\,[E(R_M) - R_{rf}]\,. \end{aligned} \tag{15}$$

A equação 15 implica uma relação linear entre o retorno esperado $E(R_i)$ e o retorno esperado do portfólio de mercado $E(R_M)$. Utilizando os dados empíricos, podemos através de técnicas estatísticas calcular o β_i. O β_i captura a parte do risco que é não-diversificável em relação ao mercado como um todo, por definição o risco do mercado é 1, exemplificando:

Figura 2 – CAPM calculado para alguns ativos - Apple (AAPL), Amazon (AMZN), Facebook (FB), Google (GOOGL) em relação ao portfólio de mercado SP500.



Fonte – Fernando Melo, 2021

Utilizando os ativos, Apple (AAPL), Amazon (AMZN), Facebook (FB), Google (GOOGL) e o utilizando o SP500 como exemplo de portfólio de mercado, aplicamos a equação 15 com o log nos retornos.

Um $\beta > 1$ (Apple, Facebook, Google), implica que o ativo é mais volátil que o Portfólio de mercado, logo um $\beta < 1$ (Amazon), tem risco menor que o Portfólio de mercado.

2.6 Arbitrage pricing theory

O MPT e o CAPM trouxeram grande evolução no campo da gestão de investimento, porém eles consideram um único fator de risco. Banz (1981), sugere que esses modelos, não conseguem descrever o risco adequadamente para algumas empresas, como empresas menores que tem retorno maior em determinado nível de risco que empresas maiores. Logo, uma abordagem mais específica na descrição do risco pode ajudar a construir portfólios mais lucrativos.

Buscando criar um modelo multifatorial na descrição do risco, Ross (1976) apresenta o *Arbitrage pricing theory - APT*, em tradução literal, Teoria de Precificação por arbitragem, relacionando a sensibilidade do ativo a um componente de risco específico e o Prêmio de Risco daquele daquele mesmo fator.

Podemos calcular o retorno esperado como

$$E(R_i) = \lambda_0 + \lambda_1 b_{i1} + \lambda_2 b_{i2} + ... + \lambda_k b_{ik}, \tag{16}$$

onde, λ_0 é o retorno esperado de um ativo livre de risco, λ_i o Prêmio de Risco relacionado ao fator de risco k, e b_{ik} é a responsividade do ativo i para o fator k. Diferentemente do CAPM, onde podemos utilizar técnicas estatísticas para definição do β, no APT não temos definido quais são os fatores, variando de ativo pra ativo, entre diferentes países e economias. Dessa maneira a natureza dos riscos é empírica, cabendo uma análise singular dos riscos que afetam cada ativo.

Mesmo os riscos sendo individualizados, os ativos sofrem de fatores macroeconômicos. Alguns estudos tentaram definir fatores comuns que impactam ativos na bolsa, como mostrado por Chen, Roll e Ross (1986). Entre eles, temos a inflação, produção industrial, mudanças na curva de rendimento, etc.

Os modelos citados anteriormente são modelos de *Média-Variância - MV*, ou seja, baseados no retorno médio e variância como medida de risco. Uma das maiores críticas a esse tipo de modelo é que eles assumem que o retorno dos ativos tem uma distribuição

normal. Porém, alguns estudos sugerem de maneira empírica (FAMA, 1965), que os retornos seguem uma distribuição de Pareto.

Além de assumir a normalidade da distribuição, os modelos de MV também assumem a variância da média dos retornos como medida de risco, sem considerar o objetivo dos investidores. Buscando contornar essas limitações, Rom e Ferguson (1993) apresentam a *Post-Modern Portfolio Theory - PMPT*.

No modelo clássico, todo risco é tratado da mesma maneira: tanto o risco de subida, quanto o de descida. Entretanto, o PMPT só considera o risco abaixo do objetivo do investidor, e tudo acima disso é considerado oportunidade sem risco.

Esse risco é chamado de *Downside Risk*, e o retorno chamamos é Retorno mínimo aceitável. Enquanto no MPT necessita de uma curva normal, no PMPT aceita distribuições assimétricas, o que se aproxima mais da realidade.

2.7 Medindo a performance de um portfólio

Até o presente momento, descreveu-se maneiras de avaliar risco, retorno e construir portfólios. Com isso surge a pergunta, entre todos os portfólios construídos, qual apresenta a melhor relação risco-retorno? Com essa questão em mente, precisamos de métricas que nos permitam comparar a performance dos portfólios construídos, podemos listar, nos títulos a seguir, quatro métricas mais relevantes que são amplamente utilizadas na literatura e no mercado.

Na avaliação do nosso modelos iremos utilizar o Índice Sharpe que é o mais comumente utilizado, porém a seguir iremos expandir para outras métricas, buscando uma contextualização mais formal dos métodos que podemos utilizar para medir a performance.

2.7.1 Índice Sharpe

Introduzido por Sharpe (1966), o Índice Sharpe é o retorno médio de um portfólio/ativo em relação a um ativo de livre de risco por unidade de risco. Dessa maneira, podemos examinar a performance ajustada pelo risco.

$$I_{Sharpe} = \frac{R_p - R_f}{\sigma_p}, \tag{17}$$

sendo, R_p é o retorno do portfólio, R_f o retorno de um ativo livre de risco e σ_p o desvio padrão do portfólio. O ativo livre de risco pode ser considerado uma constante, logo, quando calculamos o índice quanto maior a razão entre retorno e risco, mais eficiente é o portfólio em relação a um ativo livre de risco, e o portfólio com maior índice é o portfólio ótimo entre todos os possíveis.

2.7.2 Índice de Treynor

Desenvolvido por Treynor e Mazuy (1966), essa métrica é uma extensão do Índice Sharpe, pois ele utiliza o risco sistemático (β) ao invés do risco total. Podemos definir o Índice de Treynor como:

$$I_{Treynor} = \frac{R_p - R_f}{\beta_p}, \tag{18}$$

onde, R_p é o retorno do portfólio, R_f o retorno de um ativo livre de risco e β_p que é o risco do portfólio em relação ao mercado como um todo.

Como no índice Sharpe, quanto melhor a razão entre o retorno do portfólio e o risco sistemático, mais eficiente se torna o portfólio. Ambos os índices conseguem nos dar um valor, que torna possível classificar os portfólios de maneira ranqueada, porém nenhum dos dois consegue nos dizer quão melhor nosso portfólio foi em relação a carteira de mercado.

2.7.3 Alfa de Jensen

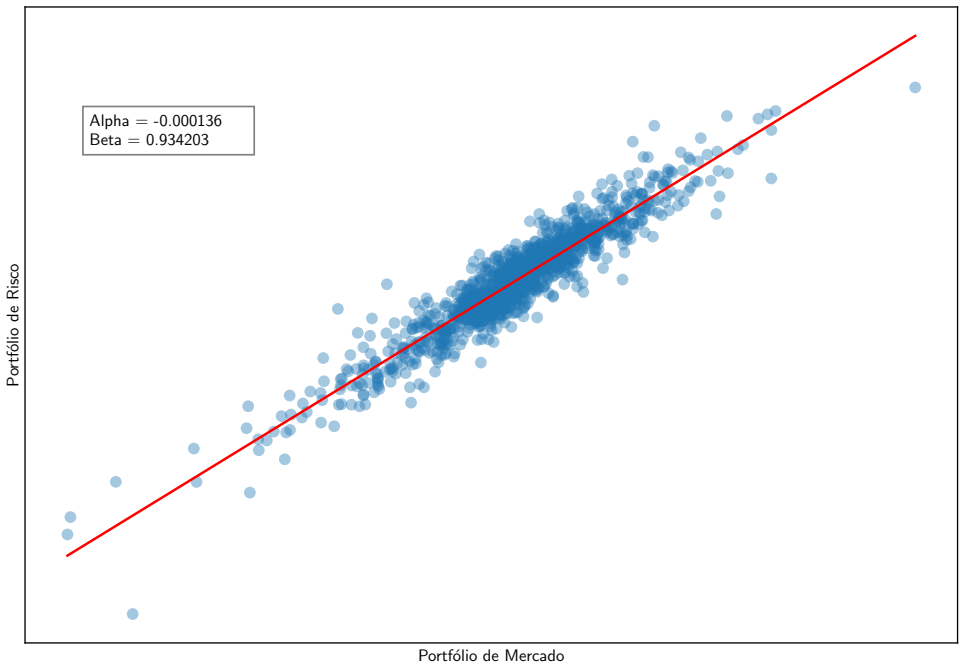
Apresentado por Jensen (1968), o Alfa (α) de Jensen é uma métrica que mede o retorno adicional de um ativo em relação ao previsto pelo CAPM. Podemos defini-la como:

$$(R_i - R_{rf}) = \alpha_p + \beta_i \left[E(R_M) - R_{rf}\right]. \tag{19}$$

Desse modo podemos calcular o α e β através de métodos estatísticos, como por exemplo uma regressão.

No gráfico a seguir, podemos observar o retorno adicional dos portfólios de risco em relação ao Portfólio de Mercado.

Figura 3 – Exemplo de cálculo de Alfa e Beta.



f

Fonte – Fernando Melo, 2021

Podemos interpretar, que o Portfólio de Risco, cresceu 0.0136% menos que o Portfólio de Mercado, porém o β mostra que o portfólio é quase 7% menos volátil que o índice de mercado.

2.7.4 Information Ratio

Essa métrica é definida como a relação entre os retornos de um ativo em relação a outro ativo/portfólio que utilizamos como parâmetro dividido pelo *Tracking Error (TE)* (GOODWIN, 1998).

Muitos portfólios são avaliados através da comparação com algum outro portfólio de referência, como um índice, a diferença entre o retorno do portfólio e da referência, é o que chamamos de *Tracking Error* ou risco ativo, ou seja, é o desvio do retorno atingido pelo

portfólio através do gerenciamento em relação ao nosso comparativo. Matematicamente é definido como:

$$TE = \sigma_{\triangle} = \sqrt{\frac{\sum_{t=1}^{T}(\triangle_t - \overline{\triangle})^2}{T-1}}, \tag{20}$$

onde $\triangle_t$ é diferença percentual entre o retorno dos ativos em questão, $\overline{\triangle}$ a média da diferença dos retornos no período e T o número de observações utilizadas. Com a compreensão do *Tracking Error*, podemos definir o *Information Ratio*:

$$IR_j = \frac{R_p - R_b}{\sigma_{\triangle}}, \tag{21}$$

sendo, R_p o retorno do portfólio p, R_b o retorno do ativo/portfólio que utilizamos como parâmetro e $\sigma_{\triangle}$ o *Tracking Error*. Podemos interpretar o *Information Ratio* como a capacidade do gestor do portfólio em gerencia-lo de maneira a ser mais lucrativo que o portfólio que utilizamos como referência.

3 Machine Learning

Segundo Samuel (1959), *Machine Learning* é a ciência onde buscamos que o computador execute atividades das quais eles não foram diretamente programados, ou seja, a partir de um conjunto de dados ou ações, o computador se torne capaz de realizar tarefas das quais ele não tinha sido ensinado.

Essas tarefas podem ser uma classificação ou regressão, onde ele busca estimar o valor de y através do conjunto de variáveis X, ou até mesmo aprender ações a serem tomadas a partir de tentativa e erro.

O processo de aprendizado dos algoritmos de *Machine Learning* acontece através da otimização de parâmetros utilizando o conjunto de dados para treinamento. Com esses parâmetros treinados, podemos utilizar o modelo para inferir em cima do problema a qual ele foi concebido.

Na próxima seção serão apresentados os principais conceitos para desenvolvimento dos modelos que serão utilizados nesse trabalho.

3.1 Redes Neurais

As Redes Neurais datam da década de 40, onde McCulloch e Pitts (1943) apresentam um modelo conceitual que busca mimetizar o funcionamento do cérebro humano através de um modelo matemático, porém não existia nenhuma maneira de treinar uma rede.

Na década de 80, Rumelhart, Hinton e Williams (1986) apresentam o o algoritmo de *Backpropagation*, permitindo assim que a rede seja treinada.

Com maior poder computacional, se tornou possível construir redes mais complexas que apresentam excelentes resultados em tarefas, como reconhecimento de imagem, reconhecimento de texto, entre outros (LECUN; BENGIO; HINTON, 2015).

Iremos descrevê-las a seguir a partir da definição de Goodfellow *et al.* (2016).

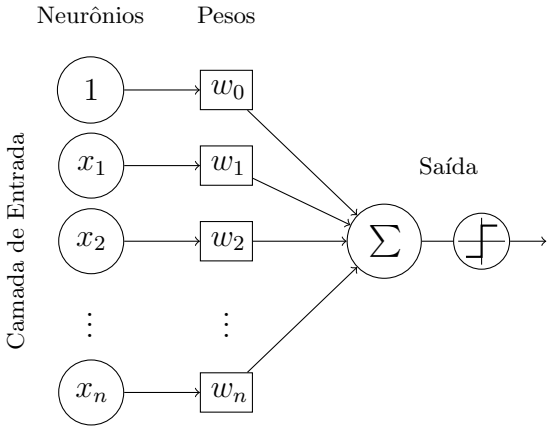
As estruturas base para uma Rede Neural são:

- **Neurônio (em inglês, *Neuron*)**: é a unidade mais elementar de uma rede neural, ele recebe um *input* (entrada) de dados e envia uma saída. De maneira geral, os neurônios são espaços onde ocorrem transformações matemáticas (como por exemplo as funções de ativação). São essas unidades que formam as camadas citadas abaixo.

- **Camada de entrada, (em inglês, *Input Layer*)**: é a camada de entrada dos dados, para um processamento subsequente da rede.
- **Pesos, (em inglês, *Weights*)**: normalmente representados com a letra w são as conexões entre as camadas. Temos pesos atrelados a cada uma delas. Eles podem ser inicializados aleatoriamente ou através de um inicializador pré-definido. São essas estruturas que treinamos e aprendemos no modelo.
- **Camada(s) Oculta(s), (em inglês, *Hidden Layers*)**: um modelo pode ter várias camadas ocultas, o número delas é um dos fatores que define a arquitetura da Rede, essa divisão é composta por neurônios que recebem os dados da camada anterior, aplica os pesos e as envia através de uma função de ativação. Ou seja, ela realiza transformações não lineares nos dados recebidos.
- **Camada de saída, (em inglês, *Output Layer*)**: é onde geramos a resposta da rede.

Podemos então construir uma Rede Neural da seguinte maneira:

Figura 4 – Ilustração de uma possível arquitetura de uma Rede Neural.



Fonte – Fernando Melo, 2021

Onde $x_1, x_2, ..., x_n$ é o conjunto de características que utilizamos para treinar o modelo, $w_0, w_1, w_2, ..., w_n$ são os pesos, nesse caso, eles são os parâmetros que treinamos buscando aprimorar o modelo através do conjunto de dados. A relação entre as características e pesos se dá através da equação:

$$A = \sum_{i=1}^{n} x_i w_i. \tag{22}$$

Ou seja, é a soma ponderada, que será utilizada como saída de cada camada, ou como entrada da próxima camada (Camada(s) oculta(s)).

3.1.1 Funções de Ativação

Quando observamos na equação 22, podemos notar que é uma função linear, e se combinarmos várias delas, mantemos a linearidade. Buscando introduzir não-linearidade ao modelo aplicamos funções de ativação, essas funções ajudam o modelo a aprender padrões/funções mais complexas. As funções de ativação são utilizadas nas camadas ocultas e na camada de saída. De maneira matemática:

$$A = g(Z^i), \tag{23}$$

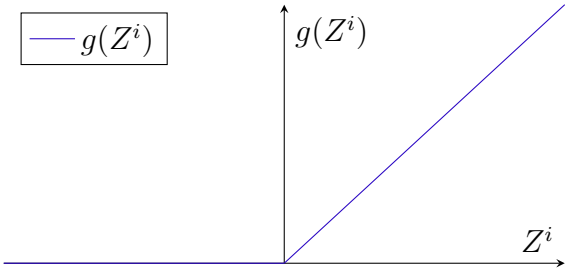
onde, $g(Z^i)$ é a função de ativação aplicada ao resultado da soma ponderada. Existem diversão funções de ativação, algumas delas são:

- ***Rectified Linear Unit, ou ReLU***: é uma das funções mais utilizadas, principalmente em reconhecimento de imagens, ela transforma a saída de 22, se menor que 0, o valor assume 0, se maior que 0, o valor mantém Z. Matematicamente:

$$g(Z^i) = \max(0, Z^i). \tag{24}$$

  Graficamente, pode ser representado como:

Figura 5 – Função de Ativação ReLU.



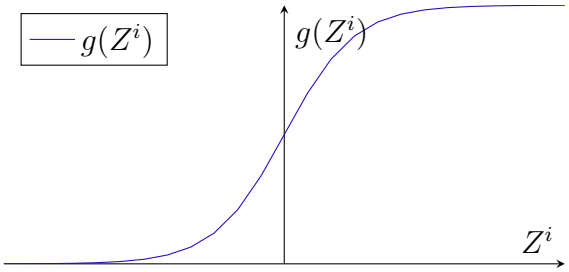
Fonte – Fernando Melo, 2021

- **Função Sigmoide ou Logística**: é outra função muito utilizada, ela é muito aplicada na última camada, devido a sua saída estar sempre entre 0 e 1, esse valor é interpretado estatisticamente como a chance do valor ser classificado como 0 ou 1, como por exemplo classificar uma imagem entre ser um cachorro ou gato. Matematicamente, definimos como:

$$g(Z^i) = \frac{1}{1 + e^{-Z^i}}. \tag{25}$$

Graficamente:

Figura 6 – Função de Ativação Sigmoide.



Fonte – Fernando Melo, 2021

- ***Softmax***: é muito parecido com a Sigmoide. Sua saída também está entre 0 e 1, porém ele é capaz de prever qual a probabilidade entre várias classes, como classificar uma imagem entre diversos animais possíveis. Matematicamente, definimos como:

$$g(Z^i) = \frac{e^{Z^i}}{\sum_{i=1}^{n} e^{Z^n}}. \tag{26}$$

  Onde n, é o número de classes do modelo.

3.2 Treinamento

De maneira macro, o processo de treinamento de uma Rede Neural acontece em duas fases e a primeira é o processo que descrevemos anteriormente, a *Feedforward*. Nela, os dados fluem através da rede criada, passando por transformações lineares e não lineares. A segunda parte, nós medimos o erro e executamos um processo chamado de *Backpropagation*, onde os pesos são atualizados e o processo de treino acontece.

3.2.1 Épocas

Épocas (em inglês, *Epoch*) é o número de vezes que passamos os dados através do modelo para atualizar os pesos. Caso o conjunto de dados seja muito grande, podemos definir conjuntos menores de dados (normalmente chamamos de lotes, ou em inglês, *Batch*), e atualizamos os pesos a cada passagem.

3.2.2 Erro

Com a saída do modelo é necessário medir quão distante está ela está do valor real. Para isso são utilizadas Funções de Custo (J). Existem diversas funções que podem exercer esse papel, listando duas que sãos mais utilizadas:

- ***Cross Entropy***, normalmente em uma classificação podem ser utilizadas uma Função Sigmoid ou Softmax na última camada como função de ativação. Podemos usar a *Cross Entropy* para medir a performance do nosso classificador. Para um classificador binário, pode-se expressar matematicamente como:

$$J = -\frac{1}{N}\sum_{j=1}^{N}(y_j \log(p_j) + (1 - y_j)\log(1 - p_j)), \tag{27}$$

  onde, N é o número de amostras, y_j é o valor real da amostra j e p_j a probabilidade para da amostra j ser 1.
- **Erro Quadrático Médio** é a função de custo padrão quando uma rede neural é utilizada para regressão. Ela mede a diferença média entre o valor predito e o real. Matematicamente, podemos expressar como:

$$J = -\frac{1}{N}\sum_{j=1}^{N}(\hat{y}_j - y_j)^2, \tag{28}$$

  sendo, $\hat{y}$ é o valor predito para amostra j e y o valor real da amostra j. Podemos notar que por utilizar o fator quadrático ela acaba penalizando mais erros maiores

3.2.3 *Backpropagation*

Até nesse ponto os dados foram processados pela rede e o erro foi medido, porém não ocorreu uma atualização dos pesos. O treinamento ocorre em uma processo chamado de *Backpropagation*. Esse algoritmo ficou popular a partir do trabalho de Rumelhart, Hinton e Williams (1986), onde é definido como:

> *Repeatedly adjusts the weights of the connections in the network so as to minimize a measure of the difference between the actual output vector of the net and the desired output vector.*

Portanto, utilizando a função gradiente em conjunto com as funções de custo, 28 ou 27, podemos ajustar os pesos da rede, indo do fim para o começo. Matematicamente podemos definir como:

$$w_{ij}^{l} = w_{ij}^{l} - \alpha \left(\frac{\partial(J)}{\partial w_{ij}^{l}} \right), \tag{29}$$

onde, w_{ij}^{l}, esquerdo, é o valor atualizado dos pesos entre as camadas l e $l-1$, sendo w_{ij}^{l}, o peso atualizado entre o node j na camada $l-1$ e o node i na camada l, w_{ij}^{l}, direito, o peso atual, $\frac{\partial(J)}{\partial w_{ij}^{l}}$ a derivada parcial da função custo em relação a w_{ij}^{l}.

Com esse processo, a cada época, podemos atualizar os pesos de cada camada, convergindo-os ao valor ótimo, de maneira a melhorar a performance do modelo.

3.3 Reinforcement Learning

Diferente das técnicas descritas anteriormente, onde aprendizado ocorre com as informações que foram disponibilizadas, em *Reinforcement Learning* o aprendizado ocorre pela interação do agente com o ambiente, onde ele realiza ações recebendo uma recompensa, positiva ou negativa, com o objetivo de aprender quais são as melhores ações a serem tomadas em cada instante de tempo em determinado estado s, de maneira a maximizar a recompensa.

Nesse trabalho, iremos adotar o formalismo estabelecido por Sutton e Barto (2018) onde a busca pela solução ótima é um Processo de Decisão de Markov (MDP ou $\mathcal{M}$), sendo representado como:

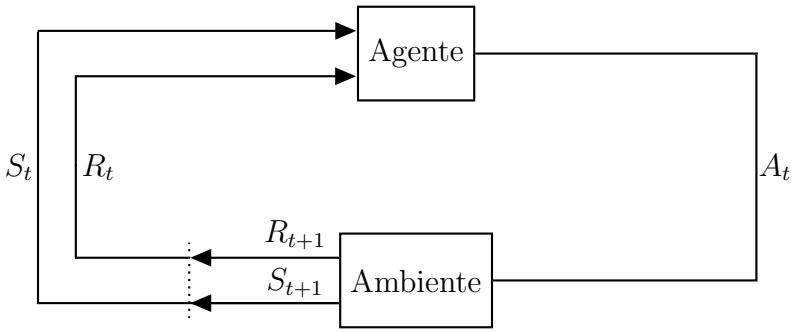
$$\mathcal{M} = (\mathcal{S}, \mathcal{A}, \mathcal{R}, \mathcal{P}), \tag{30}$$

Onde essa tupla representa todos os estados possíveis $\mathcal{S}$, todas ações que o agente pode executar $\mathcal{A}$, as recompensas recebidas nas ações tomadas em cada determinado estado, $\mathcal{R}$ **:** $\mathcal{S}\times\mathrm{A}\mapsto \mathcal{R}$, e $\mathcal{P}$ a distribuição de probabilidades de transições de estados dada determinada ação, $\mathcal{P}$ **:** $\mathcal{S}\times\mathrm{A}\mapsto \mathcal{P}$.

O agente tem o papel de aprender e tomar decisões, com o processo de aprendizado acontecendo através da interação contínua com o Ambiente, em cada interação no instante t, temos o estado s_t e recebemos uma recompensa r_t pela ação tomada, em um novo instante novas situações são apresentadas gerando um novo estado s_{t+1} e o agente recebe

uma nova recompensa r_{t+1} dependendo da ação tomada, podemos desenhar o diagrama a seguir:

Figura 7 – Simplificação de um processo MDP.



Fonte – Sutton, 2018

Matematicamente, podemos definir essa dinâmica, como:

$$p(s', r|s, a) \doteq Pr\{S_t = s', R_t = r|S_{t-1} = s, A_{t-1} = a\}, \tag{31}$$

sendo, p é a distribuição de probabilidade de cada estado e ação. Podemos observar, que o estado seguinte depende unicamente no estado atual, nesse caso, consideramos que o estado, traz toda informação necessária da interação passada do agente com o ambiente, essa característica é conhecida como Propriedade de Markov.

O objetivo principal do *Reinforcement Learning*, é maximizar a recompensa acumulada no final do processo. Para processos finitos, podemos definir formalmente como:

$$G_t \doteq R_{t+1} + R_{t+2} + R_{t+3} + ... + R_{t+n}. \tag{32}$$

Porém, temos casos contínuos, como melhoria contínua de um sistema, nesse tipo de aplicação, a recompensa pode tender a infinito, podemos definir a equação do retorno de maneira mais geral, englobando os dois casos:

$$G_t \doteq R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \gamma^3 R_{t+4} + ... \tag{33}$$

Onde γ é um parâmetro contido entre 0 e 1, que recebe o nome de taxa de desconto. Se igual a 1 temos a equação 32, caso esteja mais perto de 0, a recompensa é menor no futuro, fazendo o agente olhar mais para as recompensas seguintes, se mais próximo de 1, as recompensas futuras passam a ser mais relevantes para o agente.

Em cada estado s, podemos tomar diferentes ações que podem impactar diretamente a recompensa daquela ação e o próximo estado. Esse mapeamento de quais ações podemos tomar em cada estado, é o que chamamos de Política (π).

Se seguimos uma política (π), estando no estado s' em um momento t, podemos ter vantagem ou não em relação a nossa busca pela recompensa. Essa medida de como estamos na posição em relação ao nosso objetivo final, é o que chamamos de Função de Valor.

A Função de Valor de um estado s, em respeito a política π pode ser escrito como:

$$v_\pi(s) \doteq \mathbb{E}_\pi[G_t \,|\, S_t = s], \tag{34}$$

onde, $\mathbb{E}_\pi$ é o valor esperado considerando o estado s, tomando ações de acordo com a política (π).

Podemos também realizar a ação a, essa ação pode ser vantajosa ou não, essa medida da ação que tomada no estado s, seguindo a política π é chamada de Função de Ação-Valor. A Função de Ação-Valor tomando uma ação a, em um estado s, em respeito a política π pode ser escrito como:

$$q_\pi(s, a) \doteq \mathbb{E}_\pi[G_t \,|\, S_t = s, A_t = a], \tag{35}$$

sendo, $\mathbb{E}_\pi$ é o valor esperado considerando que o agente tomou ações de acordo com a política (π).

3.3.1 Policy Gradient

Policy Gradient (PG), são métodos onde buscamos otimizar os parâmetros θ de uma política $\pi(a|s, \theta) = Pr\{a_t = a|\, S_t = s, \theta_t = \theta\}$, através de uma métrica de performance $J(\theta)$, sem necessariamente aprender uma função de valor, ou seja, buscamos atualizar a chance de cada ação ser tomada, baseado na recompensa que ela trouxe, sem o apoio da função de valor (SUTTON *et al.*, 1999) (SILVER *et al.*, 2014). Podemos representar matematicamente como:

$$\theta_{t+1} = \theta_t + \alpha\widehat{\nabla J(\theta_t)}, \tag{36}$$

sendo, $\widehat{\nabla J(\theta_t)}$ o gradiente ascendente da métrica de performance $J(\theta_t)$ e α a taxa de aprendizado.

3.3.2 Algoritmos Actor-Critic

Uma das arquiteturas mais utilizadas com *Policy Gradient*, é a *Actor-Critic*, que consiste em duas estruturas, o *actor* busca aprender a política e atualizar os θ da política $\pi_\theta(s)$, e o *critic* visa aprender a função de valor $Q^w(s, a)$ avaliando a ação tomada pelo *actor*.

Nos últimos 10 anos, diversos trabalhos surgiram utilizando uma Rede Neural para aproximar o *actor* e o *critic* (BARTO; SUTTON; ANDERSON, 1983), (SCHULMAN *et al.*, 2015a), (MNIH *et al.*, 2016), (LILLICRAP *et al.*, 2015), essa utilização de Redes Neurais em algoritmos de *Reinforcement Learning*, é o que chamamos de *Deep Reinforcement Learning*.

Nesse modelo, os parâmetros são atualizados a cada episódio. No instante $t = 0$, o agente não tem conhecimento prévio, consequentemente ele toma uma ação aleatória (*Actor*), para avaliar essa ação treinamos uma rede neural que estima o valor da retorno (*Critic*). Com esse *feedback* atualizamos a rede neural da política, que vai melhorando as ações que são tomadas a cada instante t.

3.3.3 Proximal Policy Optimization

Nos métodos citados anteriormente, nós atualizamos nossa função custo $J(\theta)$ atualizando os parâmetros θ através do gradiente ascendente, porém se damos um passo muito grande temos variabilidade no treino e um passo muito pequeno, o treinamento se torna muito lento.

Buscando contornar esse problema Schulman *et al.* (2017) introduz o *Proximal Policy Optimization (PPO)* com a ideia de limitar o tamanho da atualização da política em cada passo, melhorando a estabilidade.

Esse processo se da através da introdução uma nova função objetiva chamada *"Clipped surrogate objective function"*, que vai limitar o tamanho da atualização da política para um determinado range.

Uma maneira de entender o impacto das ações, é através da função a seguir:

$$R_{t\theta} = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta old}(a_t|s_t)}, \tag{37}$$

onde, $\pi_\theta(a_t|s_t)$ é a probabilidade ação na atual política e $\pi_{\theta old}(a_t|s_t)$ é a probabilidade ação na política anterior. Se $R_t(\theta) > 1$ a ação é mais provável no estado atual que o anterior, se o valor estiver entre 0 e 1, a ação é menos provável no estado atual do que no anterior.

Dessa maneira nossa nova função objetiva é:

$$L^{CPI}(\theta) = \hat{\mathbb{E}}_t \left[\frac{\pi_\theta(a_t|s_t)}{\pi_{\theta old}(a_t|s_t)} \hat{A}_t \right] = \hat{\mathbb{E}}_t[r_t(\theta)\hat{A}_t], \tag{38}$$

sendo, $\hat{A}_t$, a *Advantage* que pode ser entendido como a diferença do valor de q da função 35 e a média das ações que poderiam ter sido tomadas naquele estado, ou seja, a recompensa extra que poderia ser obtida pelo agente ao realizar uma ação específica (SCHULMAN *et al.*, 2015b).

Porém na equação 38 , sem limitações, a atualização da política ainda pode ser muito grande, com isso se torna necessário colocar limitações no tamanho da atualização, que no artigo original, pode variar entre 0.8 e 1.2.

Podemos então limitar o tamanho da atualização direto na função objetiva, como:

$$L^{CLIP}(\theta) = \hat{\mathbb{E}}_t[min(r_t(\theta)\hat{A}_t, clip(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t)]. \tag{39}$$

Nesse caso temos dois valores, um limitado entre $1-\epsilon$ e $1+\epsilon$ e outra sem limitação, e selecionamos o menor valor entre os dois, evitando grande atualizações da política π_θ.

Se utilizarmos uma rede neural que compartilha os parâmetros entre a função de valor e a política, devemos acrescentar um termo de erro para função de valor e adicionar um parâmetro de entropia para garantir um bom nível de exploração, dessa maneira:

$$L^{CLIP+VF+S}(\theta) = \hat{\mathbb{E}}_t[L^{CLIP+VF+S} - c_1 L_t^{VF}(\theta) + c_2 S[\pi_\theta](s_t)], \tag{40}$$

sendo, c_1 e c_2 coeficientes, S o fator de entropia e L_t^{VF} é o erro quadrado.

Implementando o algoritmo para N *actors* em paralelo, em T interações por episódio e depois atualizamos as redes.

Algoritmo 1 Algoritmo PPO.

for $interação = 1, 2, ...$ **do**
  for $actor = 1, 2, ...$ **do**
    Executa a política $\pi_{\theta old}$ no ambiente for T interações.
    Calcula as vantagens estimadas $\hat{A}_1, ..., \hat{A}_t$.
  Otimiza L em relação a θ, em K épocas e mini batchs de tamanho $M \leq NT$
  $\theta_{old} \leftarrow \theta$

4 Metodologia e Resultados

Nessa seção iremos nos aprofundar na metodologia utilizada para os modelos e na apresentação e discussão dos resultados obtidos.

4.1 Problema de Pesquisa

Considerando técnicas de aprendizado de máquina em um conjunto de dados amplo, conseguimos criar um algoritmo completamente autônomo na construção, otimização e manutenção de um portfólio? Ele tem potencial para ter um retorno melhor que modelos clássicos?

4.2 Hipótese

O Mercado financeiro pode ser considerado um Sistema Complexo devido as características que o definem. Esse viés multifacetado abre a possibilidade para estudos com diversas abordagens. Uma das maneiras mais comuns de abordá-lo é através de técnicas estatísticas que tragam previsibilidade no movimento dos ativos buscando vantagem competitiva, de maneira a auferir lucros mediante a sua negociação.

Os modelos clássicos partem do cálculo do retorno esperado em uma janela de tempo considerando dados históricos. Entretanto Fama (1970), apresenta a Hipótese do Mercado Eficiente, onde defende que o preço de mercado das ações é uma avaliação racional do valor da empresa dadas as informações disponíveis. De acordo com essa teoria, o mercado é eficiente em termos de informação, ou seja, o valor de uma ação reflete toda informação disponível sobre a companhia.

Segundo, Marcus, Bodie e Kane (2013), isso implica que o preço do ativo é suscetível as notícias, que por sua vez não podem ser previstas, tendo assim um comportamento randômico. Considerando essa hipótese, a única coisa que pode mudar o valor de uma ação são as informações que alteram a percepção do valor da empresa para o mercado.

Nos últimos anos, temos gerado dia após dia uma grande quantidade de dados estruturados e não estruturados, um fenômeno que chamamos de "Big Data". Com isso o maior desafio deixa de ser obter o dado, mas a transformação do dado bruto em informação. Em paralelo a isso, temos uma maior acessibilidade de poder computacional e a evolução de

ambientes *Cloud*, onde temos Infraestrutura com precificação sobre demanda se tornando mais acessível, proporcionando escalabilidade quase que infinita. É importante salientar o surgimento de arquiteturas que possibilitam armazenar, processar e extrair essa grande quantidade de dados, como Apache Hadoop[1] e Apache Spark[2], (DAVENPORT; BARTH; BEAN, 2012).

Podemos destacar, alguns fundos, que de maneira constante conseguem a mais de 30 anos manter um retorno médio de 30% ao ano[3] através de métodos quantitativos.

Com isso levantamos a hipótese, que o grande número de dados gerados, em conjunto com técnicas de modelagem estatística e poder computacional, se torna possível o desenvolvimento de estratégias, com atuação autônoma, que consigam gerar retornos maiores que os modelos clássicos e acima da média do mercado, contrariando a Teoria do Mercado Eficiente.

4.3 Objetivo

O objetivo desse trabalho, é propor uma abordagem automatizada através da criação de um agente utilizando técnicas de *Reinforcement Learning*, de maneira que ele seja capaz, sem intervenção humana de escolher os ativos, distribuir os pesos, e executar operações de compra e venda que resultem em portfólios melhores que os modelos clássicos.

Um dos pontos importantes desse trabalho, é agregar dados de análise de técnica e fundamentalista na série histórica em conjuntos com os *Autoencoders*, nos permitindo descrever de maneira mais assertiva o estado que o agente se encontra, melhorando o aprendizado e as ações a serem tomadas.

4.4 Dados

Para construção de nossos portfólios, consideramos 2 mercados distintos, o Americano e o de Criptoativos, filtrando as maiores empresas por capitalização de mercado e pela disponibilidade dos dados. Para uma abordagem mais ampla e buscando entender a capacidade do agente em se adaptar e agir em uma gama maior de ativos com características funcionais diferentes consideramos o mercado de Criptomoedas, todos ativos

[1] https://hadoop.apache.org/
[2] https://spark.apache.org/
[3] https://www.wsj.com/articles/the-making-of-the-worlds-greatest-investor-11572667202

Tabela 1 – Lista de Criptomoedas.

Ativo	Operação	Pontos de Treino	Período	Fonte
ETH	Ininterrupta	36096	08-05-2018 a 20-06-2022	Binance
BTC	Ininterrupta	36096	08-05-2018 a 20-06-2022	Binance
BNB	Ininterrupta	36096	08-05-2018 a 20-06-2022	Binance
XRP	Ininterrupta	36096	08-05-2018 a 20-06-2022	Binance
ADA	Ininterrupta	36096	08-05-2018 a 20-06-2022	Binance

Fonte – Fernando Melo, 2022

Tabela 2 – Nome das empresas utilizadas.

Ativo	Nome da Empresa
AAPL	Apple Inc.
FB	Meta Platforms, Inc.
PEP	PepsiCo, Inc.
GOOGL	Alphabet Inc.
TSLA	Tesla, Inc.
AMD	Advanced Micro Devices, Inc.
HON	Honeywell International Inc.
NVDA	NVIDIA Corporation
PYPL	PayPal Holdings, Inc.
TXN	Texas Instruments Incorporated

Fonte – Fernando Melo, 2022

estão emparelhados com o dólar Americano e com granularidade de hora. A fonte das informações é o site Dukascopy [4,5] para os ativos da Bolsa de valores Americana e a API da Binance para extração dos dados de Criptomoedas [6,7]. No mercado de Criptomoedas temos operação sem pausas, logo temos mais pontos que o mercado de Ações como podemos observar na tabelas 1 e 1.

Nosso processo de seleção de dados leva em consideração fatores como valor de mercado, volume, setor da empresa e disponibilidade de dados. Para as criptomoedas, selecionamos cinco ativos com base em sua capitalização de mercado, com intervalo de uma hora e excluindo *stablecoins*. *Stablecoins* têm uma forte correlação com ativos fiduciários, como o dólar Americano, mas no caso do nosso agente, estamos testando seu desempenho contra ativos mais voláteis. Podemos verificar a lista de criptomoedas na Tabela 1.

Quanto às ações, coletamos dados sobre ações Americanas, com frequência de 1 hora, atendendo aos critérios mencionados, o nome e a matriz de dados dos ativos podem ser observados nas tabelas 3 e 2.

[4] https://www.dukascopy.com/
[5] https://github.com/Leo4815162342/dukascopy-node
[6] https://www.binance.com/pt-BR
[7] https://python-binance.readthedocs.io/en/latest/

Tabela 3 – Lista de Ativos da Bolsa Americana.

Ativo	Operação	Pontos de Treino	Período	Fonte
AAPL	9:30 a 16:00	9486	26-01-2017 a 27-05-2022	Dukascopy
FB	9:30 a 16:00	9486	26-01-2017 a 27-05-2022	Dukascopy
PEP	9:30 a 16:00	9486	26-01-2017 a 27-05-2022	Dukascopy
GOOGL	9:30 a 16:00	9486	26-01-2017 a 27-05-2022	Dukascopy
TSLA	9:30 a 16:00	9486	26-01-2017 a 27-05-2022	Dukascopy
AMD	9:30 a 16:00	9486	26-01-2017 a 27-05-2022	Dukascopy
HON	9:30 a 16:00	9486	26-01-2017 a 27-05-2022	Dukascopy
NVDA	9:30 a 16:00	9486	26-01-2017 a 27-05-2022	Dukascopy
PYPL	9:30 a 16:00	9486	26-01-2017 a 27-05-2022	Dukascopy
TXNU	9:30 a 16:00	9486	26-01-2017 a 27-05-2022	Dukascopy

Fonte – Fernando Melo, 2022

4.4.1 Dados Técnicos, Fundamentalistas e do *Autoencoder*

Os dados da Análise Técnica que usamos são indicadores como Média Móvel Simples, Bandas de Bollinger, Parabólica e Reversa (PSAR), Média Móvel Convergência Divergência (MACD) e Índice de Força Relativa (RSI).

Para Ações, também foram utilizados dados de análise fundamentalista, na qual extraímos dados relacionados ao valor da empresa, fluxo de caixa, balanço patrimonial, crescimento, dividendos e *ratings*.

Nosso *Autoencoder* será uma Rede Neural Convolucional 1-D, que codifica e decodifica os recursos, ajudando a eliminar o ruído dos dados e capturar a relação subjacente no vetor de características.

Para geração dos dados de análise técnica, foi utilizado o pacote *ta* [8].

Para as ações do mercado Americano, também foram utilizados os dados de análise fundamentalista, através do pacote *fundamentalanalysis* [9].

4.5 Implementação

Iremos implementar o algoritmo PPO, como foi descrito na seção 3.3.3, utilizando como saída uma função *Softmax* como na equação 26 com dimensão igual ao número de ativos no portfólio, sendo que o valor de saída, é o peso de cada ativo no portfólio.

[8] https://technical-analysis-library-in-python.readthedocs.io/en/latest/
[9] https://pypi.org/project/fundamentalanalysis/

4.5.1 A Tarefa

Otimização de portfólio é o processo de reorganização dos ativos para atingir os objetivos definidos pelo investidor. Considere que temos N ativos e que queremos otimizar os pesos em cada etapa do tempo t. Seja $v_m(t) \in \mathbb{R}^N$ onde v são os preços de fechamento de cada ativo m no momento t. Nosso vetor relativo de preço para o período t de negociação, y_t, é definido como:

$$y_t := \left(1, \frac{v_{1,t}}{v_{1,t-1}}, \frac{v_{2,t}}{v_{2,t-1}}, ..., \frac{v_{m,t}}{v_{m,t-1}}\right)^T, \tag{41}$$

onde cada elemento de y_t é a razão do fechamento do ativo no período t em relação ao período anterior $t-1$.

Os pesos do portfólio podem ser definidos como $w(t) \in \mathbb{R}^N$, especialmente considerando que w é o peso do portfólio no momento t, que é atualizado a cada passo.

$$w = (w_1, w_2, ..., w_m)^T, \tag{42}$$

sendo, w é o vetor, que contém a contribuição de todos os ativos até m para o portfólio p.

Seja $p(t) \in \mathbb{R}$ o valor do portfólio em cada passo, podemos definir nosso valor relativo do portfólio como a razão entre o valor do portfólio no final do momento t e o final do momento $t-1$, reorganizando,

$$p_t = p_{t-1} y_t \cdot w_{t-1}, \tag{43}$$

onde w_{t-1} representa os pesos da carteira no início do momento t.

A taxa de retorno na forma logarítmica, é definida como:

$$r_t = \ln \frac{p_t}{p_{t-1}}. \tag{44}$$

Para o instante final t_f:

$$p_f = p_0 \prod_{t=1}^{t_{f+1}} y_t \cdot w_{t-1}. \tag{45}$$

Podemos observar que o valor do portfólio final p_f é o valor inicial p_0 multiplicado pelo produto dos valores de cada ativo em cada momento t pelo peso no momento anterior $t-1$, até o momento $t_f + 1$ onde liquidamos todas as posições.

4.5.2 Agente e o estado

Nosso **Espaço de Ações** $\mathcal{A}$ corresponde ao vetor de peso da carteira em cada passo de tempo t. Para isso, na última camada da rede neural, utilizamos um *Softmax* e interpretamos a saída como os pesos de cada ativo do portfólio.

Logo podemos descrever a ação do agente como:

$$a_t = w_t. \tag{46}$$

Como não temos nenhum controle sobre o preço dos ativos, então devemos nos concentrar em tentar encontrar o valor ótimo dos pesos $w*_{(t-1)}$ para o passo seguinte t, e essa é otimização do portfólio, em termos matemáticos, pode ser definido como:

$$w^*(t) := \arg\max w_t y_t \cdot w_{t-1}. \tag{47}$$

O **Espaço de Estados** $\mathcal{S}$ é o conjunto de características que ajuda a descrever o mercado no instante t, para ações Norte Americanas usaremos os dados de abertura, máxima, mínima e fechamento (em inglês, OHLC) que pode ser escrito como (X_t^{OHLC}), indicadores técnicos (X_t^{TA}), indicadores fundamentalistas (X_t^{FA}) e os dados gerados pelo *Autoencoder* (X_t^{AE}). Usaremos os mesmos recursos para Criptomoedas, exceto para dados fundamentalistas. Assim, o estado no início do passo de tempo t é

$$s(t) = (X_t^{OHLC}, X_t^{TA}, X_t^{FA}, X_t^{AE}). \tag{48}$$

4.5.3 Recompensa

Se considerarmos a equação 45, entendemos que o objetivo do nosso agente é maximizar o valor do portfólio a cada passo t, logo, podemos considerar nossa recompensa como a razão logarítmica do valor do portfólio em t em relação a $t-1$, podendo ser descrito como:

$$r = \ln\left(\frac{p_{f,t}}{p_{f,t-1}}\right). \tag{49}$$

4.5.4 Autoencoders

Autoencoders são conhecidos por suas capacidades de compactar os dados através da parte de codificação (LI; ZHENG; ZHENG, 2019; PARK; SIM; CHOI, 2020). É possível referir-se a esse conjunto de recursos originais como representação de espaço latente, o que significa que esses dados de dimensão inferior tentam representar os dados originais e reconstruí-los para a forma original por meio da parte de um decodificador. Entre os principais recursos dos *Autoencoders* estão a capacidade de diminuir o ruído dos dados e sua capacidade de reconhecer a relação subjacente entre os recursos, gerando novos recursos a partir deles.

Um *Autoencoder* é composto de duas partes, o *decoder* e o *encoder*, onde o *encoder*, cria representação de espaço latente, e o *decoder* reconstrói os dados para a forma inicial.

O *encoder* pode ser representado como:

$$\phi : \mathcal{X} \to \mathcal{F}. \tag{50}$$

E o *decoder*:

$$\psi : \mathcal{F} \to \mathcal{X}. \tag{51}$$

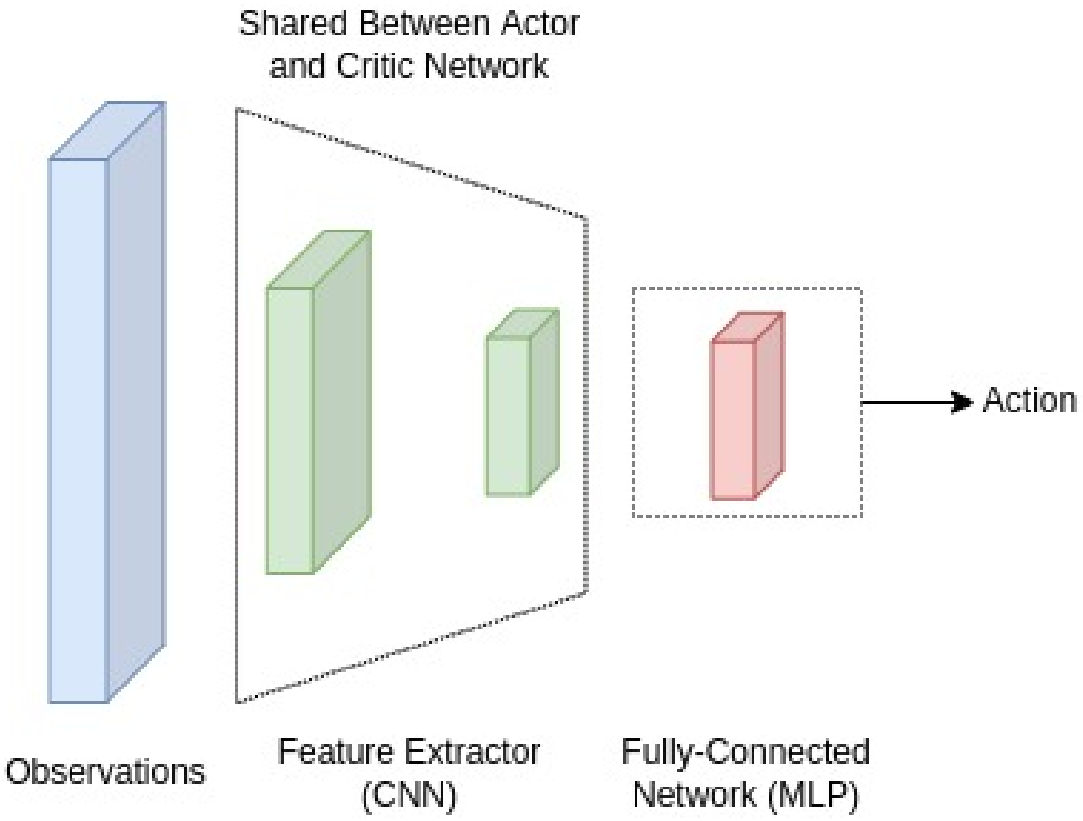
Onde $x \in \mathbb{R}^d = \mathcal{X}$ é o conjunto inicial de dados, e $h \in \mathbb{R}^p = \mathcal{F}$ nossas variáveis latentes.

4.5.5 Actor-Critic Networks

Será construída uma rede Neural que busca representar nosso *actor* e *critic*. A rede pode ser construída de diversas maneiras. Iremos utilizar 2 arquiteturas distintas, CNN (*Convolutional Neural Network*) que faz o papel extrair características do vetor de entrada X_t. Como pode ser observado na Figura 8, ela é composta por 64 filtros na primeira camada e 32 na segunda, a Tangente Hiperbólica como função de ativação e o tamanho do *kernel* com a mesma dimensão do nosso espaço de observação e uma MLP (*Multi Layer Perceptron*) com duas camadas de 64 neurônios, em todas as redes o conjunto de dados será o vetor de característica X_t.

Nas 2 topologias citadas, iremos utilizar o conceito de camadas compartilhadas, ou seja, a camada de entrada e as camadas ocultas são compartilhadas entre o *actor* e *critic*,

Figura 8 – Arquitetura do nosso extrator de características.



Fonte – Fernando Melo, 2022

só alterando a camada de saída, onde no *actor* iremos utilizar uma função *Softmax*, que gera o vetor de pesos w_t e no *critic* não iremos utilizar nenhuma função de ativação para saída.

4.5.6 Treino e avaliação

Dividimos nosso conjunto de dados em dois períodos. Um deles para treinamento, abrangendo 27.600 pontos por ativo. No entanto, quando utilizamos o *Autoencoder*, precisamos utilizar parte do conjunto de dados para treinar e validar o *Autoencoder*. Utilizamos os primeiros 6.800 pontos, reduzindo nossos pontos de treinamento do agente para 12.070 e 8.730 pontos abrangendo 11 meses para avaliação. Para ações do mercado Americano, temos 9.486 pontos, 3.000 dos quais serão usados para treinar e validar nosso *Autoencoder*, 5.613 dos quais serão usados para treinamento e 873 distribuídos para avaliação.

Para determinar quais hiperparâmetros são os mais adequados para o conjunto de dados, realizamos uma seleção em um pequeno conjunto de dados. No entanto, devido ao

Figura 9 – Esquema genérico de como os dados são separados entre o *Autoencoder*, treino e teste.

Fonte – Fernando Melo, 2022

grande número de testes a serem realizados, testar mais parâmetros tornou-se proibitivo em termos de tempo.

Em nossa análise para testar os hiperparâmetros e nosso comportamento de recompensa ao longo do treinamento, determinamos nosso ϵ como 0,3 e descobrimos que, após 200 episódios, não há mudanças significativas na recompensa e na Função de Perda do *Actor-critic*, o que foi posteriormente confirmado no conjunto de dados de teste.

Nosso Agente será avaliado principalmente por duas métricas: o retorno total em um período de 11 meses e o Índice de Sharpe da carteira, que será comparado com o Modelo de Média-Variância (MV) e *Buy and Hold* (BH).

4.6 Resultados

Esta seção revisará nossos resultados e examinará os pontos fortes e fracos do nosso agente DRL.

4.6.1 Análise de Desempenho

Na Tabela 4, podemos ver que MV tem uma grande margem de ganho sobre todos os Agentes DRL, mais de 10%, mas se olharmos mais de perto para a Figura 12 e Figura 13, podemos entender melhor a situação.

Para Criptomoedas, o mercado teve uma queda significativa nos últimos dois meses, com os ativos caindo mais de 40% em todos os modelos, exceto MV, que caiu 32%. Com relação ao *Bull Market*, nos primeiros quatro meses, nosso modelo MLP com apenas recursos gerados pelo *Autoencoder* foi capaz de superar todos os outros modelos, mesmo o MV por uma margem de 10% e atingir o maior índice de Sharpe. O comportamento ao longo do tempo entre todos os modelos pode ser observado na Figura 14.

Os *Autoencoders* desempenharam um papel crucial na melhoria das capacidades de nossos agentes no *Bull Market*, resultando no maior retorno de 74,83%, mas não generalizaram bem no *Bear Market*. Com o agente CNN obtendo o maior retorno com os dados de TA e FA e sem os recursos gerados pelo *Autoencoder*, podemos ver uma comparação do comportamento ao longo do tempo entre o nosso melhor portfólio construído com o uso de DRL e o desenvolvido por meio do MV na Figura 15.

As ações seguem um padrão semelhante. Nosso agente MLP apenas com *Autoencoder* teve um bom desempenho no mercado de tendência de alta nos primeiros cinco meses de nossa série, com um retorno de 17,57%, superando o MV em mais de 15%, mas caindo em um mercado de tendência de baixa, em quase 20%.

Entre nossos Agentes, o agente MLP com características Técnicas e Fundamentalistas alcança os melhores resultados, tanto no *Bear Market* quanto no *Bull Market*.

Nosso *Autoencoder* se saiu muito mal no mercado de tendência de baixa. Uma possível explicação é o fato de que o período que usamos para treina-los não foi capaz de capturar todas as informações necessárias. Então para trabalhos futuros podemos treinar *Autoencoder* em uma maior diversidade de períodos.

Quando estamos em um *Bull Market*, o Agente DRL tem um desempenho muito bom, pois pode capturar sinais que foram capazes de melhorar as posições ao longo do tempo, mas não é capaz de melhorar os resultados em um *Bear Market*. Isso pode ser explicado pelo fato de que no período de treinamento temos um *Bull Market* dominante, tanto para ações quanto para criptomoedas.

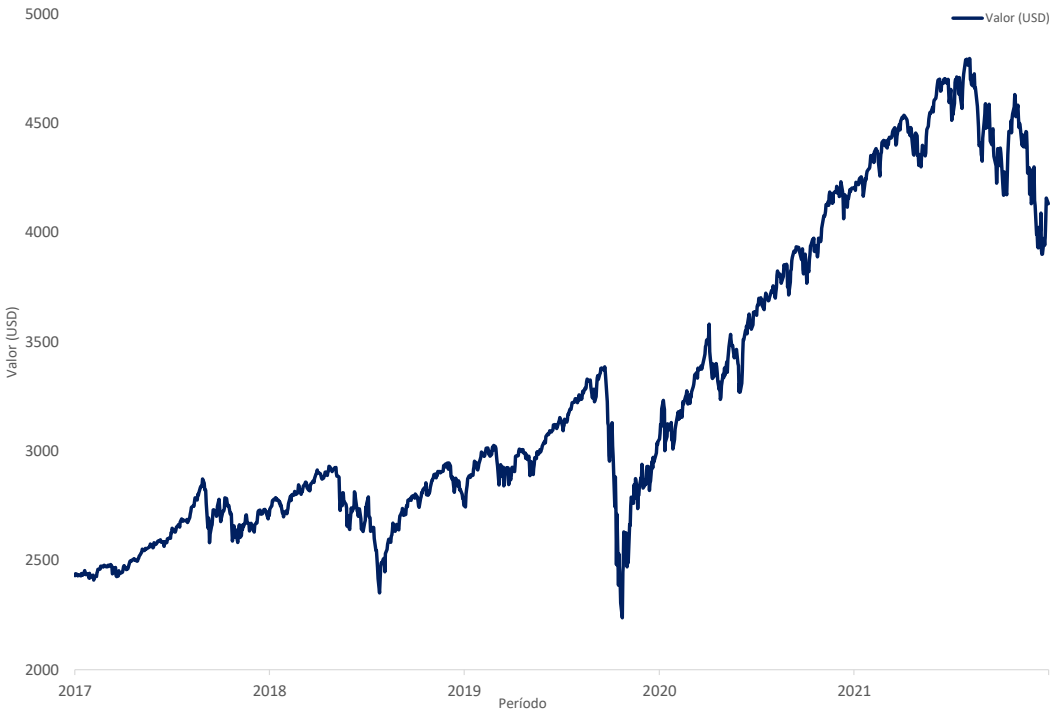
Podemos entender melhor o comportamento do mercado nos últimos anos se examinarmos o *S&P 500 Index* na Figura 10[10] e o *S&P Cryptocurrency Broad Digital*

[10] https://finance.yahoo.com/quote/%5EGSPC/history?p=%5EGSPC

Market Index na Figura 11[11]. Temos visto uma tendência de alta predominante em ambos os mercados ao longo dos anos, com apenas alguns períodos de baixa.

Consequentemente, no processo de aprendizagem, o agente lida principalmente com tendências de alta, sem a capacidade de agir adequadamente em um mercado com tendência de baixa perdendo assim sua capacidade de generalização. Embora a Carteira MV seja parametrizada para risco mínimo, ela não tem um desempenho tão bom quanto nosso agente DRL em mercados de alta, mas é capaz de manter um desempenho aceitável em um mercado de tendência de baixa.

Figura 10 – *S&P 500 Index* nos últimos cinco anos.

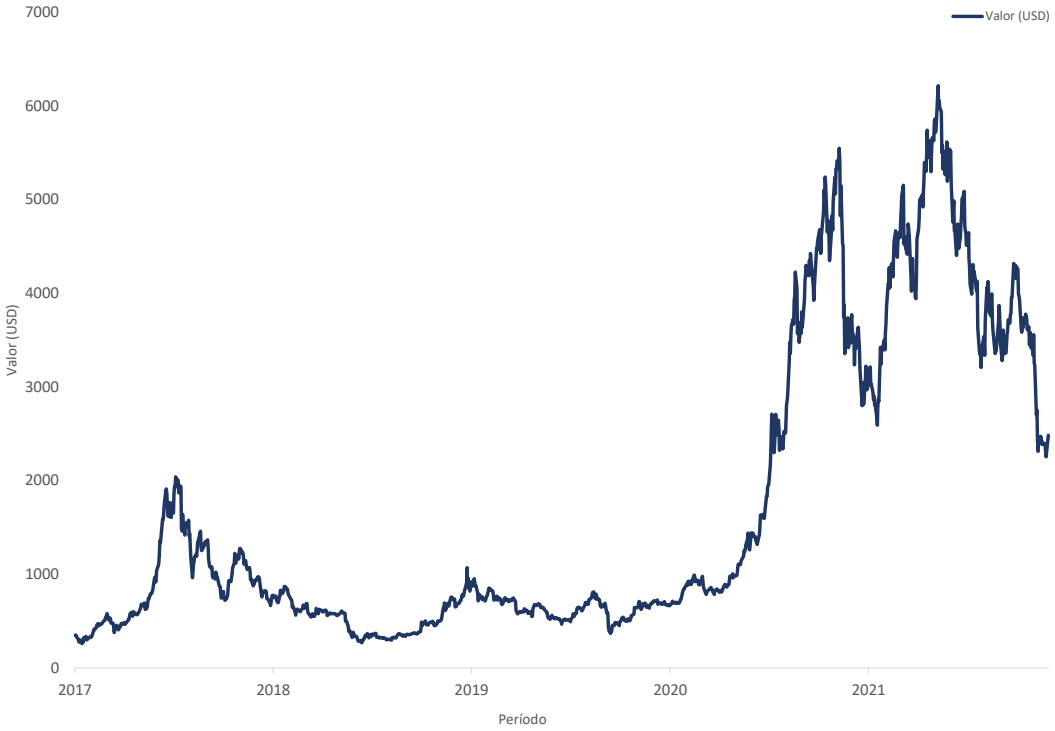


Fonte – Fernando Melo, 2022

Em relação ao Índice Sharpe para Criptomoedas, todos os modelos tiveram resultados ruins próximos de 0. Tabela 4 ilustra que, baseado apenas em mercados de tendência de alta, nosso modelo MLP com apenas *Autoencoders* tem o maior índice de Sharpe de 0,49. Em Ações, o Índice de Sharpe é negativo para todos os modelos, exceto para MV,

[11] https://www.spglobal.com/spdji/pt/indices/digital-assets/sp-cryptocurrency-broad-digital-market-index/overview

Figura 11 – *S&P Cryptocurrency Broad Digital Market Index* nos últimos cinco anos.



Fonte – Fernando Melo, 2022

Tabela 4 – Desempenho de cada portfólio para Criptomoedas.

Modelo	Retorno	Índice Sharpe
CNN (TA + AE)	-20,67%	-0,01
CNN (TA)	-14,94%	0,01
CNN (AE)	-22,76%	-0,02
MLP (TA + AE)	-21,58%	-0,02
MLP (TA)	-22,35%	-0,01
MLP (AE)	-19,72%	-0,01
MV	-3,04%	0,04
BH	-19,40%	-0,01

Fonte – Fernando Melo, 2022

mas ainda é baixo como pode ser visto na Tabela 5. Em mercados de tendência de alta, nosso melhor modelo é o Agente MLP com apenas *Autoencoder*, com um Índice de Sharpe de 0,97 comparado a 0,35 para MV.

Considerando o modelo de BH com os ativos de maiores volume como um aproximador do comportamento do mercado, quando olhamos para a Hipótese do Mercado Eficiente,

Tabela 5 – Desempenho de cada carteira de ações.

Modelo	Retorno	Índice Sharpe
CNN (TA + FA + AE)	-14.59%	-0.18
CNN (TA + FA)	-14.00%	-0.17
CNN (AE)	-14.57%	-0.18
MLP (TA + FA + AE)	-14.84%	-0.19
MLP (TA + FA)	-12.886	-0.14
MLP (AE)	-15.22%	-0.19
MV	4.63%	0.14
BH	-11.94%	-0.14

Fonte – Fernando Melo, 2022

Figura 12 – Mapa de calor dos retornos da carteira de ações dos Americanas, onde as cores verdes significam retorno positivo no período e a paleta de cores amarela/vermelha significa retorno negativo. No Eixo vertical temos os modelos testados e no Eixo Horizontal temos os períodos por mês.

	2021-07	2021-08	2021-09	2021-10	2021-11	2021-12	2022-01	2022-02	2022-03	2022-04	2022-05
CNN (TA + FA + AE)	3,89%	4,98%	-5,70%	9,20%	4,34%	-0,42%	-8,06%	-8,87%	5,86%	-14,69%	-3,26%
CNN (TA + FA)	3,71%	4,74%	-5,54%	13,37%	4,05%	-0,21%	-7,82%	-9,16%	5,61%	-14,32%	-3,12%
CNN (AE)	3,89%	5,01%	-5,62%	9,32%	4,41%	-0,35%	-7,97%	-8,87%	5,72%	-14,48%	-3,15%
MLP (TA + FA + AE)	4,07%	4,49%	-5,41%	7,90%	4,04%	-0,19%	-7,71%	-8,46%	5,28%	-13,85%	-3,55%
MLP (TA + FA)	3,49%	5,57%	-5,41%	11,04%	5,41%	-1,06%	-8,52%	-8,39%	6,56%	-15,45%	-3,59%
MLP (AE)	3,51%	5,35%	-5,80%	9,28%	4,75%	-0,44%	-8,14%	-9,52%	5,79%	-14,74%	-3,40%
MV	4,64%	-0,61%	-6,55%	10,22%	-2,42%	6,84%	-2,04%	-5,42%	4,04%	-2,49%	-0,34%
BH	3,84%	4,87%	-5,62%	8,74%	3,83%	0,15%	-7,82%	-9,17%	5,60%	-14,37%	0,38%

Fonte – Fernando Melo, 2022

nosso algoritmo foi capaz de supera-lo por uma margem de quase 5%, contrariando a *weak form*, isso está em linha com outros trabalhos que mostram a ineficiência da HME na sua forma mais fraca para Criptoativos como foi muito bem explorado e revisado por Kyriazis (2019).

Porém no mercado de ações, nosso agente DRL tem performance melhor que o BH em mercado de alta, porém 2% atrás quando consideramos o mercado todo.

Figura 13 – Mapa de calor dos retornos dos Portfólios de Criptomoedas, onde as cores verdes significam retorno positivo no período e a paleta de cores amarelo/vermelho significa retorno negativo. No eixo vertical temos os modelos testados e no eixo horizontal temos os períodos por mês.

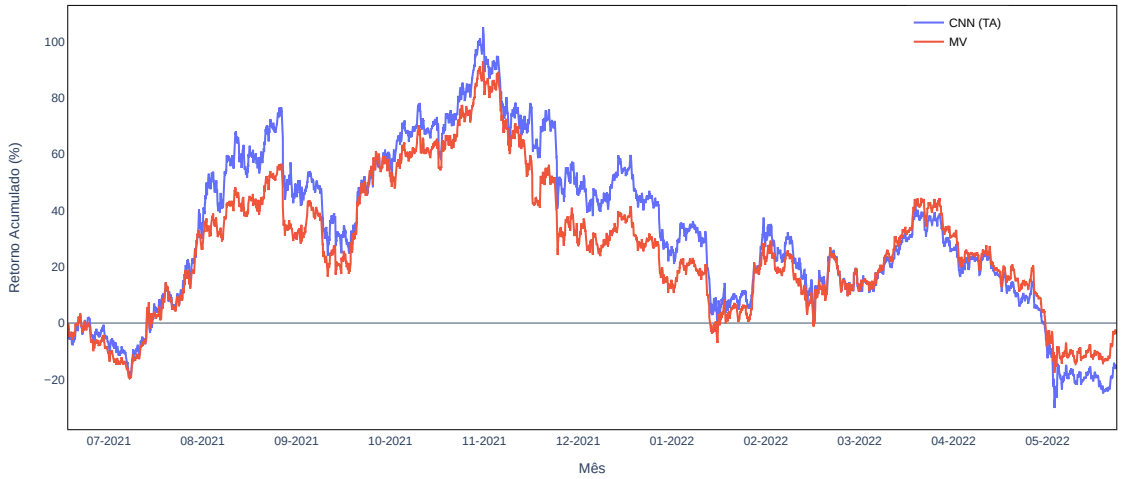
	2021-07	2021-08	2021-09	2021-10	2021-11	2021-12	2022-01	2022-02	2022-03	2022-04	2022-05
CNN (TA + AE)	9,16%	50,11%	-15,69%	23,98%	-3,48%	-17,30%	-22,57%	8,14%	11,02%	-21,09%	-20,90%
CNN (TA)	11,26%	42,53%	-14,99%	28,48%	-0,77%	-17,28%	-23,07%	12,00%	8,71%	-19,11%	-21,03%
CNN (AE)	8,54%	52,48%	-16,17%	23,65%	-2,50%	-17,35%	-23,18%	7,61%	11,54%	-21,07%	-23,24%
MLP (TA + AE)	7,06%	57,86%	-17,35%	20,10%	-2,21%	-16,99%	-22,90%	4,57%	12,36%	-21,38%	-19,15%
MLP (TA)	11,20%	43,28%	-14,73%	26,84%	-4,63%	-17,19%	-22,40%	13,30%	8,60%	-20,81%	-22,04%
MLP (AE)	8,31%	44,80%	-13,63%	29,07%	-3,65%	-17,65%	-22,98%	10,77%	9,85%	-23,35%	-22,08%
MV	9,61%	30,57%	-8,08%	24,82%	-7,40%	-17,08%	-17,80%	14,69%	15,50%	-18,08%	-13,98%
BH	9,45%	49,06%	-15,65%	25,20%	-2,31%	-17,37%	-23,02%	8,92%	10,73%	-20,57%	-24,42%

Fonte – Fernando Melo, 2022

Figura 14 – Retorno acumulado do portfólio de Criptoativos de cada modelo ao longo do nosso período de teste.

Fonte – Fernando Melo, 2022

Figura 15 – Retorno acumulado do portfólio de Criptoativos somente do melhor Agente DRL e o modelo de Média-Variância.



Fonte – Fernando Melo, 2022

Figura 16 – Retorno acumulado do portfólio de Ações de cada modelo ao longo do nosso período de teste.

Fonte – Fernando Melo, 2022

Figura 17 – Retorno acumulado do portfólio de Ações somente do melhor Agente DRL e o modelo de Média-Variância.

Fonte – Fernando Melo, 2022

5 Conclusões

Esse trabalho apresenta um agente PPO que foi testado empiricamente contra ações do Mercado Americano e Criptomoedas usando diferentes tipos de vetores de características e diferentes arquiteturas para a Rede Neural. De acordo com nossos experimentos, os Agentes DRL têm um bom desempenho em mercados de tendência de alta, superando MV e BH, mas uma performance ruim em mercados de tendência de baixa, perdendo por uma grande margem para MV. Como resultado de nossa análise, pudemos demonstrar como é difícil superar consistentemente o MV, especialmente quando o Momentum do mercado muda rapidamente, mesmo com diferentes tipos de informações sendo agregadas.

No *Bull Market*, o *Autoencoder* teve um bom desempenho, dando melhores resultados do que outros conjuntos de variáveis, mas resultados mais fracos em geral. Uma investigação mais profunda é necessária para melhorar a arquitetura e os parâmetros dos *Autoencoders*, aprimorando suas capacidades em diferentes tipos de tendências de mercado.

Quando olhamos com o viés da HME, nosso algoritmo teve uma performance melhor que o BH em Criptoativos e pior em Ações como foi explorado na seção 4.6.1, essa divergência de comportamento entro os dois mercados pode ocorrer devido a alguns fatores como, (a) a generalização ruim do nosso agente no mercado de baixa, que se comporta pior em ações do que no mercado de Criptoativos, (b) a maior variância dos retornos no mercado de Criptomoedas em relação ao mercado de Ações ou (c) a inaptidão do Agente em superar a HME, dado que as ações tem uma simetria maior em relação a informações.

Uma possível explicação para o ponto levantado na seção 4.2, é a capacidade que essas empresas têm de adquirir fontes de dados diversas, chegando em alguns casos há mais 10000 fontes [1], esse grande número de informações ajuda a enriquecer os modelos, proporcionando inferência mais assertivas e aumentando o lucro das operações ao longo do tempo.

Em trabalhos futuros, um período mais extenso de dados, abrangendo uma gama maior de cenários de mercado, pode ajudar a criar um agente DRL capaz de ter um bom desempenho em um mercado com tendência de queda ou adicionar recursos que permitam enviar mais sinais precisos durante momentos de turbulência ou ponto de virada no momento do mercado, além disso, uma maior quantidade de dados acrescenta novos períodos de teste, que tornam a avaliação de performance mais abrangente.

[1] https://www.twosigma.com/approach/

Referências

ABOUSSALAH, A. M.; LEE, C.-G. Continuous control with stacked deep dynamic recurrent reinforcement learning for portfolio optimization. *Expert Systems with Applications*, Elsevier, v. 140, p. 112891, 2020. Citado 3 vezes nas páginas 14, 16 e 17.

ABOUSSALAH, A. M.; LEE, C.-G. Continuous control with Stacked Deep Dynamic Recurrent Reinforcement Learning for portfolio optimization. *Expert Systems with Applications*, Elsevier Ltd, v. 140, p. 112891, feb 2020. ISSN 09574174. Citado na página 14.

ALIMORADI, M. R.; KASHAN, A. H. A league championship algorithm equipped with network structure and backward q-learning for extracting stock trading rules. *Applied soft computing*, Elsevier, v. 68, p. 478–493, 2018. Citado na página 16.

BANZ, R. W. The relationship between return and market value of common stocks. *Journal of Financial Economics*, v. 9, n. 1, p. 3–18, 1981. ISSN 0304-405X. Disponível em: ⟨https://www.sciencedirect.com/science/article/pii/0304405X81900180⟩. Citado na página 28.

BARTO, A. G.; SUTTON, R. S.; ANDERSON, C. W. Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE transactions on systems, man, and cybernetics*, IEEE, n. 5, p. 834–846, 1983. Citado na página 41.

BETANCOURT, C.; CHEN, W.-H. Deep reinforcement learning for portfolio management of markets with a dynamic number of assets. *Expert Systems with Applications*, Elsevier, v. 164, p. 114002, 2021. Citado na página 14.

BOCCARA, N. *Modeling complex systems*. [S.l.]: Springer Science & Business Media, 2010. Citado na página 18.

BROCK, D. C.; MOORE, G. E. *Understanding Moore's law: four decades of innovation*. [S.l.]: Chemical Heritage Foundation, 2006. Citado na página 15.

CHEN, N.-F.; ROLL, R.; ROSS, S. A. Economic forces and the stock market. *The Journal of Business*, University of Chicago Press, v. 59, n. 3, p. 383–403, 1986. ISSN 00219398, 15375374. Disponível em: ⟨http://www.jstor.org/stable/2352710⟩. Citado na página 28.

CHENG, W.; WAGNER, W.; LIN, C.-H. Forecasting the 30-year us treasury bond with a system of neural networks. *NeuroVe $ t Journal*, v. 1, n. 2, 1996. Citado na página 15.

DAMODARAN, A. Estimating risk free rates. *WP, Stern School of Business, New York*, 1999. Citado na página 20.

DANTAS, S. G.; SILVA, D. G. Equity trading at the brazilian stock market using a q-learning based system. In: IEEE. *2018 7th Brazilian Conference on Intelligent Systems (BRACIS)*. [S.l.], 2018. p. 133–138. Citado na página 16.

DAVENPORT, T. H.; BARTH, P.; BEAN, R. How'big data'is different. MIT Sloan Management Review, 2012. Citado na página 45.

DELCEY, T. Samuelson vs fama on the efficient market hypothesis: The point of view of expertise. *Œconomia. History, Methodology, Philosophy*, Association Œconomia, n. 9-1, p. 37–58, 2019. Citado na página 19.

DENG, Y.; BAO, F.; KONG, Y.; REN, Z.; DAI, Q. Deep direct reinforcement learning for financial signal representation and trading. *IEEE transactions on neural networks and learning systems*, IEEE, v. 28, n. 3, p. 653–664, 2016. Citado na página 16.

DING, Y.; LIU, W.; BIAN, J.; ZHANG, D.; LIU, T.-Y. Investor-imitator: A framework for trading knowledge extraction. In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. [S.l.: s.n.], 2018. p. 1310–1319. Citado na página 16.

EILERS, D.; DUNIS, C. L.; METTENHEIM, H.-J. von; BREITNER, M. H. Intelligent trading of seasonal effects: A decision support algorithm based on reinforcement learning. *Decision support systems*, Elsevier, v. 64, p. 100–108, 2014. Citado na página 16.

FAMA, E. F. The behavior of stock-market prices. *The Journal of Business*, University of Chicago Press, v. 38, n. 1, p. 34–105, 1965. ISSN 00219398, 15375374. Disponível em: ⟨http://www.jstor.org/stable/2350752⟩. Citado na página 29.

FAMA, E. F. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, [American Finance Association, Wiley], v. 25, n. 2, p. 383–417, 1970. ISSN 00221082, 15406261. Disponível em: ⟨http://www.jstor.org/stable/2325486⟩. Citado 2 vezes nas páginas 19 e 44.

FILOS, A. Reinforcement learning for portfolio management. *arXiv preprint arXiv:1909.09571*, 2019. Citado na página 16.

FRANK, H. K. Risk, uncertainty and profit. ,): Hart, Schaffner & Marx, 1921. Citado na página 20.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A.; BENGIO, Y. *Deep learning*. [S.l.]: MIT press Cambridge, 2016. v. 1. Citado na página 33.

GOODWIN, T. H. The information ratio. *Financial Analysts Journal*, Routledge, v. 54, n. 4, p. 34–43, 1998. Disponível em: ⟨https://doi.org/10.2469/faj.v54.n4.2196⟩. Citado na página 31.

HORNIK, K.; STINCHCOMBE, M.; WHITE, H. Multilayer feedforward networks are universal approximators. *Neural networks*, Elsevier, v. 2, n. 5, p. 359–366, 1989. Citado na página 15.

JENSEN, M. C. The performance of mutual funds in the period 1945-1964. *The Journal of finance*, JSTOR, v. 23, n. 2, p. 389–416, 1968. Citado na página 30.

JENSEN, M. C. The foundations and current state of capital market theory. Praeger Publishers, 1972. Citado na página 24.

JIANG, Z.; LIANG, J. Cryptocurrency portfolio management with deep reinforcement learning. In: IEEE. *2017 Intelligent Systems Conference (IntelliSys)*. [S.l.], 2017. p. 905–913. Citado na página 17.

JIANG, Z.; XU, D.; LIANG, J. A deep reinforcement learning framework for the financial portfolio management problem. *arXiv preprint arXiv:1706.10059*, 2017. Citado na página 16.

KENDALL, M. G.; HILL, A. B. The analysis of economic time-series-part i: Prices. *Journal of the Royal Statistical Society. Series A (General)*, [Royal Statistical Society, Wiley], v. 116, n. 1, p. 11–34, 1953. ISSN 00359238. Disponível em: ⟨http://www.jstor.org/stable/2980947⟩. Citado na página 19.

KYRIAZIS, N. A. A survey on efficiency and profitable trading opportunities in cryptocurrency markets. *Journal of Risk and Financial Management*, MDPI, v. 12, n. 2, p. 67, 2019. Citado na página 56.

LADYMAN, J.; LAMBERT, J.; WIESNER, K. What is a complex system? *European Journal for Philosophy of Science*, Springer, v. 3, n. 1, p. 33–67, 2013. Citado na página 18.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *nature*, Nature Publishing Group, v. 521, n. 7553, p. 436–444, 2015. Citado na página 33.

LI, Y.; ZHENG, W.; ZHENG, Z. Deep robust reinforcement learning for practical algorithmic trading. *IEEE Access*, IEEE, v. 7, p. 108014–108022, 2019. Citado 3 vezes nas páginas 16, 17 e 50.

LIANG, Z.; CHEN, H.; ZHU, J.; JIANG, K.; LI, Y. Adversarial deep reinforcement learning in portfolio management. *arXiv preprint arXiv:1808.09940*, 2018. Citado na página 16.

LILLICRAP, T. P.; HUNT, J. J.; PRITZEL, A.; HEESS, N.; EREZ, T.; TASSA, Y.; SILVER, D.; WIERSTRA, D. Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*, 2015. Citado 2 vezes nas páginas 16 e 41.

LINTNER, J. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics*, The MIT Press, v. 47, n. 1, p. 13–37, 1965. ISSN 00346535, 15309142. Disponível em: ⟨http://www.jstor.org/stable/1924119⟩. Citado na página 26.

MANDELBROT, B. Forecasts of future prices, unbiased markets, and "martingale"models. *The Journal of Business*, University of Chicago Press, v. 39, n. 1, p. 242–255, 1966. ISSN 00219398, 15375374. Disponível em: ⟨http://www.jstor.org/stable/2351745⟩. Citado na página 19.

MANDELBROT, B. B. The variation of certain speculative prices. In: *Fractals and scaling in finance*. [S.l.]: Springer, 1997. p. 371–418. Citado na página 18.

MARCUS, P. A. J.; BODIE, P. Z.; KANE, A. *Investments*. McGraw-Hill Education, 2013. ISBN 9780077861674. Disponível em: ⟨https://books.google.com.br/books?id=eJ71mgEACAAJ⟩. Citado 2 vezes nas páginas 24 e 44.

MARKOWITZ, H. Portfolio selection. *Journal of Finance*, v. 7, n. 1, p. 77–91, 1952. Disponível em: ⟨https://EconPapers.repec.org/RePEc:bla:jfinan:v:7:y:1952:i:1:p:77-91⟩. Citado 2 vezes nas páginas 21 e 22.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, Springer, v. 5, n. 4, p. 115–133, 1943. Citado na página 33.

MIIKKULAINEN, R.; LIANG, J.; MEYERSON, E.; RAWAL, A.; FINK, D.; FRANCON, O.; RAJU, B.; SHAHRZAD, H.; NAVRUZYAN, A.; DUFFY, N.; HODJAT, B. Chapter 15 - evolving deep neural networks. In: KOZMA, R.; ALIPPI, C.; CHOE, Y.; MORABITO, F. C. (Ed.). *Artificial Intelligence in the Age of Neural Networks and Brain Computing*. Academic Press, 2019. p. 293–312. ISBN 978-0-12-815480-9. Disponível em: ⟨https://www.sciencedirect.com/science/article/pii/B9780128154809000153⟩. Citado na página 15.

MNIH, V.; BADIA, A. P.; MIRZA, M.; GRAVES, A.; LILLICRAP, T.; HARLEY, T.; SILVER, D.; KAVUKCUOGLU, K. Asynchronous methods for deep reinforcement learning. In: PMLR. *International conference on machine learning*. [S.l.], 2016. p. 1928–1937. Citado na página 41.

MOORE, G. E. *et al. Cramming more components onto integrated circuits*. [S.l.]: McGraw-Hill New York, 1965. Citado na página 15.

MOSSIN, J. Equilibrium in a capital asset market. *Econometrica*, [Wiley, Econometric Society], v. 34, n. 4, p. 768–783, 1966. ISSN 00129682, 14680262. Disponível em: ⟨http://www.jstor.org/stable/1910098⟩. Citado na página 26.

NEUNEIER, R. Optimal asset allocation using adaptive dynamic programming. *Advances in Neural Information Processing Systems*, v. 8, 1995. Citado na página 16.

NEUNEIER, R. Enhancing q-learning for optimal asset allocation. *Advances in neural information processing systems*, v. 10, 1997. Citado na página 16.

PARK, H.; SIM, M. K.; CHOI, D. G. An intelligent financial portfolio trading strategy using deep q-learning. *Expert Systems with Applications*, Elsevier, v. 158, p. 113573, 2020. Citado 2 vezes nas páginas 17 e 50.

PEARSON, K. The problem of the random walk. *Nature*, Nature Publishing Group, v. 72, n. 1867, p. 342–342, 1905. Citado na página 19.

REILLY, F.; BROWN, K. *Investment Analysis and Portfolio Management*. Cengage Learning, 2011. ISBN 9780538482387. Disponível em: ⟨https://books.google.com.br/books?id=CfB-qTXqRWEC⟩. Citado 2 vezes nas páginas 14 e 20.

ROM, B. M.; FERGUSON, K. W. Post-modern portfolio theory comes of age. *The Journal of Investing*, Institutional Investor Journals Umbrella, v. 2, n. 4, p. 27–33, 1993. ISSN 1068-0896. Disponível em: ⟨https://joi.pm-research.com/content/2/4/27⟩. Citado na página 29.

ROSS, S. A. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, v. 13, n. 3, p. 341–360, 1976. ISSN 0022-0531. Disponível em: ⟨https://www.sciencedirect.com/science/article/pii/0022053176900466⟩. Citado na página 28.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. *nature*, Nature Publishing Group, v. 323, n. 6088, p. 533–536, 1986. Citado 2 vezes nas páginas 33 e 37.

SAMUEL, A. L. Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, IBM, v. 3, n. 3, p. 210–229, 1959. Citado na página 33.

SCHULMAN, J.; LEVINE, S.; ABBEEL, P.; JORDAN, M.; MORITZ, P. Trust region policy optimization. In: PMLR. *International conference on machine learning*. [S.l.], 2015. p. 1889–1897. Citado na página 41.

SCHULMAN, J.; MORITZ, P.; LEVINE, S.; JORDAN, M.; ABBEEL, P. High-dimensional continuous control using generalized advantage estimation. *arXiv preprint arXiv:1506.02438*, 2015. Citado na página 42.

SCHULMAN, J.; WOLSKI, F.; DHARIWAL, P.; RADFORD, A.; KLIMOV, O. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017. Citado 3 vezes nas páginas 16, 17 e 41.

SHAO, G. N.; KIM, H.; IMRAN, S. https://www. sciencedirect. com/science/article/abs/pii/s092633731500346x. Wiley, 2016. Citado na página 16.

SHARPE, W. A simplified model for portfolio analysis. *Management Science*, v. 9, n. 2, p. 277–293, 1963. Disponível em: ⟨https://EconPapers.repec.org/RePEc:inm:ormnsc:v:9:y:1963:i:2:p:277-293⟩. Citado na página 26.

SHARPE, W. F. Capital asset prices: A theory of market equilibrium under conditions of risk*. *The Journal of Finance*, v. 19, n. 3, p. 425–442, 1964. Disponível em: ⟨https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1964.tb02865.x⟩. Citado na página 26.

SHARPE, W. F. Mutual fund performance. *The Journal of business*, JSTOR, v. 39, n. 1, p. 119–138, 1966. Citado na página 29.

SILVER, D.; LEVER, G.; HEESS, N.; DEGRIS, T.; WIERSTRA, D.; RIEDMILLER, M. Deterministic policy gradient algorithms. In: PMLR. *International conference on machine learning*. [S.l.], 2014. p. 387–395. Citado na página 40.

SUTTON, R. S.; BARTO, A. G. *Reinforcement learning: An introduction*. [S.l.]: MIT press, 2018. Citado na página 38.

SUTTON, R. S.; MCALLESTER, D. A.; SINGH, S. P.; MANSOUR, Y. *et al.* Policy gradient methods for reinforcement learning with function approximation. In: CITESEER. *NIPs*. [S.l.], 1999. v. 99, p. 1057–1063. Citado na página 40.

TAY, F. E.; CAO, L. Application of support vector machines in financial time series forecasting. *omega*, Elsevier, v. 29, n. 4, p. 309–317, 2001. Citado na página 15.

TOBIN, J. Liquidity Preference as Behavior Towards Risk1. *The Review of Economic Studies*, v. 25, n. 2, p. 65–86, 02 1958. ISSN 0034-6527. Disponível em: ⟨https://doi.org/10.2307/2296205⟩. Citado na página 25.

TREYNOR, J.; MAZUY, K. Can mutual funds outguess the market. *Harvard business review*, v. 44, n. 4, p. 131–136, 1966. Citado na página 30.

VAN, E.; ROBERT, J. The application of neural networks in the forecasting of share prices. *Haymarket, VA, USA: Finance & Technology Publishing*, 1997. Citado na página 15.

WALPOLE, R. *Probabilidade & estatística para engenharia e ciências*. Pearson Prentice Hall, 2009. ISBN 9788576051992. Disponível em: ⟨https://books.google.com.br/books?id=3_OTPgAACAAJ⟩. Citado na página 23.

WENG, L.; SUN, X.; XIA, M.; LIU, J.; XU, Y. Portfolio trading system of digital currencies: A deep reinforcement learning with multidimensional attention gating mechanism. *Neurocomputing*, Elsevier, v. 402, p. 171–182, 2020. Citado na página 17.

WILLIAMS, J. *The Theory of Investment Value*. Harvard University Press, 1938. (Investment value). ISBN 9780678080504. Disponível em: ⟨https://books.google.com.br/books?id=cIhCAAAAIAAJ⟩. Citado na página 21.

YE, Y.; PEI, H.; WANG, B.; CHEN, P.-Y.; ZHU, Y.; XIAO, J.; LI, B. Reinforcement-learning based portfolio management with augmented asset movement prediction states. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. [S.l.: s.n.], 2020. v. 34, n. 01, p. 1112–1119. Citado 3 vezes nas páginas 14, 16 e 17.

YU, P.; LEE, J. S.; KULYATIN, I.; SHI, Z.; DASGUPTA, S. Model-based deep reinforcement learning for dynamic portfolio optimization. *arXiv preprint arXiv:1901.08740*, 2019. Citado na página 16.

ZHANG, J.; MARINGER, D. Using a genetic algorithm to improve recurrent reinforcement learning for equity trading. *Computational Economics*, Springer, v. 47, n. 4, p. 551–567, 2016. Citado na página 14.

Glossário

Apache Hadoop é um framework de código aberto para armazenamento e processamento para grandes volumes de dados.

Apache Spark é um framework de código aberto de processamento para grandes volumes de dados com foco em performance.

Bear Market é quando o mercado está em um tendência de baixa por um período prolongado de tempo.

Bull Market é quando o mercado está em um tendência de alta por um período prolongado de tempo.

Buy and Hold é uma estratégia tradicional de comprar um ativo e mantê-lo na carteira por um longo período independente das flutuações de curto e médio prazo do mercado.

Prêmio de Risco é o retorno adicional em relação a um ativo livre de risco, de maneira a compensar o investidor pelo risco adicional tomado.

SP500 é um índice composto pelas 500 maiores empresas pertencentes a bolsa Nasdaq ou NYSE.