

Henrique Junqueira de Brito

Eficiência do mercado de apostas de jogos eletrônicos

Belo Horizonte, Minas Gerais

2023

Henrique Junqueira de Brito

Eficiência do mercado de apostas de jogos eletrônicos

Relatório Final - POC II

Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Ciência da Computação

Orientador: Vinícius dos Santos

Belo Horizonte, Minas Gerais
2023

Sumário

1	INTRODUÇÃO	4
1.1	Objetivos Gerais	5
1.2	Objetivos Específicos	5
2	REFERENCIAL TEÓRICO	6
3	ARBITRAGEM	7
4	DESENVOLVIMENTO E RESULTADOS	9
4.1	Tecnologias utilizadas	9
4.1.1	<i>Python</i>	9
4.1.2	<i>PostgreSQL</i>	9
4.1.3	<i>Selenium</i>	9
4.1.4	<i>Beautiful Soup</i>	9
4.2	Websites Relevantes	10
4.3	Estrutura do Banco de Dados	10
4.4	Web Scraper	11
4.5	Coleta dos dados	12
4.6	Processamento dos dados	12
5	RESULTADOS	13
5.1	Possibilidade de Arbitragem	13
5.2	Discussão	14
5.3	Trabalhos Futuros	14
	REFERÊNCIAS	16

Resumo

Nesse trabalho é construído um conjunto de dados com informações sobre precificações de apostas de partidas de *Counter Strike: Global Offensive*, a fim de evidenciar a possibilidade de estratégias de aposta com retorno positivo garantido. O conjunto de dados foi coletado fazendo o *scrapping* de 4 *websites* de apostas esportivas.

1 Introdução

A relação de esportes com apostas é bastante antiga: existem evidências de apostas esportivas sendo feitas desde à Grécia antiga e, em 2022, o tamanho mercado de apostas esportivas foi projetado em 231 bilhões de dólares, pela a Statista. Esse mercado está presente em esportes tradicionais (futebol, beisebol, futebol americano, etc) e esportes criados em função de apostas, como a corrida de cavalos ou cachorros. Com a expansão da internet e a adoção de computadores pessoais pela maioria da população, jogos eletrônicos competitivos foram criados, gerando um novo segmento de esportes conhecido como esportes eletrônicos, que por sua vez desenvolveu um mercado de apostas análogo aos esportes tradicionais.

Nesse mercado, uma aposta é um contrato entre a casa de apostas e o apostador que coloca um certo dinheiro na previsão de um resultado, a um certo preço (*odd*, ou inverso da probabilidade). Se a previsão estiver certa, o apostador recebe o dinheiro que apostou mais o lucro dessa aposta que é calculado de acordo com a *odd* à qual apostou.

Existem diversas casas de apostas que atuam no mesmo mercado, com modelos de precificação bastante distintos. Com isso, é possível que ocorram diferenças significativas de *odds* em um dado jogo, a ponto de existirem oportunidades de arbitragem, isto é, a possibilidade de uma pessoa apostar em resultados opostos em 2 casas de apostas distintas de modo que o retorno seja positivo, independente do resultado da partida em questão. Além disso, é possível que, mesmo sem oportunidade de arbitragem, exista alguma estratégia que obtenha retorno positivo com alta probabilidade.

Dentre os diversos jogos eletrônicos existentes, esse trabalho foca somente no jogo *Counter Strike*, que é um jogo de tiro em primeira pessoa baseado em rodadas nas quais equipes de contraterroristas e terroristas combatem-se até a eliminação completa de um dos times, e tem como objetivo principal plantar e desarmar bombas. O foco nesse jogo se justifica por 4 fatores:

- Frequência alta de partidas contempladas nas casas de aposta
- Simplificação da extração dos dados, por se tratar de 1 jogo apenas
- Estrutura dinâmica do jogo possibilita vários tipos de apostas
- Existe um site que divulga os resultados das partidas, enriquecendo as amostras coletadas

1.1 Objetivos Gerais

O principal objetivo desse trabalho é investigar a diferença de precificação de várias casas de apostas em torno de campeonatos de jogos eletrônicos, mais especificamente de *Counter Strike*, e, a partir dessa análise, averiguar se existe alguma estratégia que pode ser adotada para garantir um retorno positivo, ou então uma estratégia que obtenha um retorno positivo com alta probabilidade. Assim, o trabalho visa responder a seguinte questão de pesquisa (QP):

- Existe alguma estratégia de aposta que envolva o uso simultâneo de 2 ou mais casas de apostas para obter um retorno positivo com alta probabilidade?

1.2 Objetivos Específicos

Implementou-se um *web scraper* para 4 sites de aposta a fim de criar um banco de dados com a precificação das apostas do jogo *Counter Strike*.

Para isso foi feito:

- Escolha das casas de aposta que vão ter seus dados extraídos.
- Implementação de um *web scraper* para cada um dos sites.
- Estrutura de coleta e armazenamento dos dados.
- Processamento e análise dos dados.
- Discussão sobre os resultados.

2 Referencial Teórico

Como apostas esportivas no geral é um fenômeno bastante antigo existem vários trabalhos que abordam sobre a eficiência do mercado de apostas antes mesmo da popularização das casas de aposta na internet, entre eles, destaca-se o [Hausch e Ziemba 1990] que identificou oportunidades de arbitragem em apostas de corrida de cavalos na época em que casas de apostas diferentes começaram a atuar nas mesmas corridas. O trabalho [Gray e Gray 1997] analisou *odds* da *NFL* e criou um modelo baseado em *probit* a fim de identificar apostas com uma alta probabilidade de sucesso, além disso mostrou que ineficiências já publicadas no mercado eram ultimamente corrigidas pelas casas de aposta. Um pouco mais recente, [Vlastakis, Dotsis e Markellos 2009] modelou a quantidade de gols feitos em uma partida de futebol usando distribuição de *Poisson* para criar uma estratégia de aposta bem sucedida nos campeonatos de futebol europeu. O trabalho [Franck, Verbeek e Nüesch 2013], também baseado no futebol Europeu, mostrou que 19.2% das partidas possuíam algum oportunidade de arbitragem. Com a exposição de dados estruturados sobre as partidas, facilita-se o uso de técnicas de aprendizado de máquina para tentar prever resultados, o que foi feito em [Knoll e Stübinger 2020], em que o modelo criado conseguiu gerar lucros de forma consistente.

No geral, quase nenhum trabalho aborda de estratégias de apostas específicas para jogos eletrônicos, possivelmente por ser recente e acompanhado por menos pessoas. Portanto, é possível que exista uma ineficiência de mercado que possa ser explorada para obter retornos positivos com alta probabilidade.

3 Arbitragem

Uma aposta é um contrato entre a casa de apostas e o apostador que coloca um certo dinheiro na previsão de um resultado, a um certo preço (*odd*, ou inverso da probabilidade). Se a previsão estiver certa, o apostador recebe o dinheiro que apostou mais o lucro dessa aposta que é calculado de acordo com a *odd* à qual apostou.

Em um exemplo simples, temos o lançamento de uma moeda não viciada, em que a probabilidade do resultado ser cara é 50% e coroa 50%. Em um cenário justo, uma casa de aposta precificaria as *odds* de cara e coroa em 2, que equivalem às probabilidades reais. No entanto, as casas de aposta adicionam um *spread*, que pode ser interpretado como uma comissão que a casa de aposta cobra pela aposta. Com isso, um cenário realista seria *odds* de 1.9 para cara e coroa.

Como nos jogos reais as probabilidades reais não são conhecidas previamente e as casas de aposta usam modelos distintos para a previsão de resultados, é plausível a existência desse cenário:

Casa de aposta 1:

- *odd* para Coroa: 2.1
- *odd* para Cara: 1.8

Casa de aposta 2:

- *odd* para Coroa: 1.8
- *odd* para Cara: 2.1

Com isso, temos a seguinte estratégia que garante um retorno positivo:

- Apostar 1 real em Coroa na casa de aposta 1
- Apostar 1 real em Cara na casa de aposta 2

Nesse caso, teremos um retorno de 5% no investimento, independente do resultado do lançamento da moeda.

Pode-se generalizar esse exemplo para casos em que existem mais de 2 resultados possíveis, como um jogo de futebol pode ter vitória, empate e derrota, e para casos em que há *odds* de mais de 2 casas de aposta. De acordo com [Franck, Verbeek e Nüesch 2013], temos:

$$R = \left(\frac{1}{\sum_e 1/\bar{o}_e} \right) - 1 \quad (3.1)$$

Em que \bar{o}_e é a maior *odd* entre todas as casas de aposta de um resultado “e” (vitória, empate, etc). R pode ser interpretado como o retorno máximo garantido, isto é, o valor que multiplicado pelo valor apostado retorna o saldo obtido, que pode ser negativo.

Com isso, pode-se inferir que uma possibilidade de arbitragem existe quando:

$$\sum_e \frac{1}{\bar{o}_e} < 1 \quad (3.2)$$

Além disso, para obter o retorno garantido R da equação (3.1), basta o apostar na *odd* \bar{o}_e um valor inversamente proporcional à *odd*. Ou seja:

$$v_e = M \frac{1}{\bar{o}_e} \left(\frac{1}{\sum_f 1/\bar{o}_f} \right) \quad (3.3)$$

Sendo, v_e o valor a ser apostado no resultado “e” e M o montante total a ser apostado.

Nesse trabalho, foram considerados formatos de arbitragem com apenas 2 resultados possíveis, com isso foi utilizado uma fórmula equivalente mais simples para calcular R :

$$R = \frac{odd_1 odd_2}{odd_1 + odd_2} - 1 \quad (3.4)$$

Com as equações descritas acima, fica evidente que determinar a possibilidade de arbitragem dado um conjunto de *odds* é uma tarefa simples.

4 Desenvolvimento e resultados

4.1 Tecnologias utilizadas

Nesta seção serão abordados aspectos relacionados às tecnologias empregadas neste trabalho, como a linguagem de programação, softwares e bibliotecas.

4.1.1 *Python*

Python é uma linguagem de programação de propósito geral, interpretada e de alto nível. Como Python é uma linguagem de código aberto, dinamicamente tipada, com bom suporte para o processamento de dados e possui uma vasta gama de bibliotecas que auxiliam no desenvolvimento de projetos de pequena escala, ela foi a escolha para a implementação do *web scraper* proposto.

4.1.2 *PostgreSQL*

O Sistema de Gerenciamento de Banco de Dados (SGBD) escolhido atualmente para o projeto é o PostgreSQL, também conhecido como *Postgres*. PostgreSQL é um SGBD para dados relacionais, de código aberto e que respeita os princípios ACID (Atomicidade, Consistência, Isolamento e Durabilidade), sendo projetado para suportar desde pequenos projetos locais a *Data Warehouses* e *webservices* com diversas consultas em concorrência. Dessa forma, Postgres demonstrou ser capaz de atender os requisitos desejados para este trabalho.

4.1.3 *Selenium*

Selenium é biblioteca de código aberto destinada a oferecer suporte à automação de tarefas do navegador. Ela é essencial para navegar automaticamente pelas *URLs* fornecidas a fim de coletar todos as páginas *HTMLs* que possuem dados de precificações de uma partida.

4.1.4 *Beautiful Soup*

Beautiful Soup é biblioteca de código aberto destinada a processar arquivos *HTML*, a fim poder extrair os dados de forma mais estruturada. Ela é fundamental para extrair os dados das páginas *HTMLs* encontradas usando o *Selenium*.

4.2 Websites Relevantes

Nesse trabalho, foram consideradas 4 casas de apostas para os quais o *Web Scraper* foi desenvolvido.

- GG.BET - [https://gg.bet/en/?sportIds\[\]=esports_counter_strike](https://gg.bet/en/?sportIds[]=esports_counter_strike)
- Pinnacle - <https://www.pinnacle.com/en/esports/games/csgo/matchups>
- 1xBet - <https://br.1x001.com/en/cyber/live/shooters/counter-strike>
- Betway - <https://betway.com/pt/sports/sct/esports/cs-go>

4.3 Estrutura do Banco de Dados

Nessa seção, será explicitado os tipos de dados coletados pelo *Web Scraper* e armazenados no banco de dados *SQL*. Cada entrada na tabela consiste em um par de *odd* coletada.

- *timestamp_id* - Consiste em um identificador do momento da coleta dos dados.
- *team_1* - Nome do time 1
- *team_2* - Nome do time 2
- *match_id* - Consiste em um identificador da partida, construído com: <Nome do Time 1> + \$ + <Nome do Time 2> + \$ + <Nome do Evento>
- *bookie_id* - Nome da casa de aposta
- *Odd1* - Valor da precificação 1
- *Odd2* - Valor da precificação 2
- *odd_type* - Consiste em um identificador do tipo de *odd* coletado. Tais tipos podem ser:
 - *Money_Line* - Ganhador da partida. *Odd1* se refere à vitória do primeiro time e *Odd2* ao segundo
 - *Handicap_1* - *Odd1* se refere à vitória do primeiro time sem perder mapas e *Odd2* se refere à derrota do primeiro time sem ganhar mapas.
 - *Handicap_2* - *Odd1* se refere à vitória do segundo time sem perder mapas e *Odd2* se refere à derrota do segundo time sem ganhar mapas.
 - *Total_Map* - Número total de mapas jogados na série “melhor de 3”.

4.4 Web Scraper

O *Web Scraper* consiste em várias etapas. Para cada uma das 4 *URLs* citadas acima, é executado o seguinte:

Primeiramente, com a biblioteca *Selenium*, é inicializado um *web driver*, que é uma interface do código *Python* com o navegador (nesse caso o *Chrome*) que tem o objetivo de coletar as páginas *HTML* dado as suas respectivas *URLs*. Com isso, é coletada a página *HTML* que lista todos os jogos de *Counter Strike* cobertos pela casa de aposta. Para cada jogo listado, a página *HTML* correspondente é coletada e armazenada em memória.

Em segundo lugar, é utilizado a biblioteca *Beautiful Soup* para realizar o processamento das páginas *HTML* armazenadas a fim de extrair as tabelas de *odds* contidas nelas. Essa foi a etapa mais trabalhosa do projeto inteiro, visto que páginas *HTML* podem estar mal formatadas ou bastante “suja” com informações irrelevantes à coleta. Foi gasto bastante tempo tentando identificar padrões que identifiquem as tabelas com as informações desejadas. Uma forma de fazer isso é utilizando a funcionalidade de inspecionar páginas *HTML* em navegadores modernos, ilustrado pela Figura 1, uma vez que eles fornecem uma forma dinâmica de explorar o que cada componente do *HTML* corresponde na página vista no navegador normal. No fim, diversas regras de processamento foram elaboradas para a detecção e extração das tabelas de *odds* corretamente.

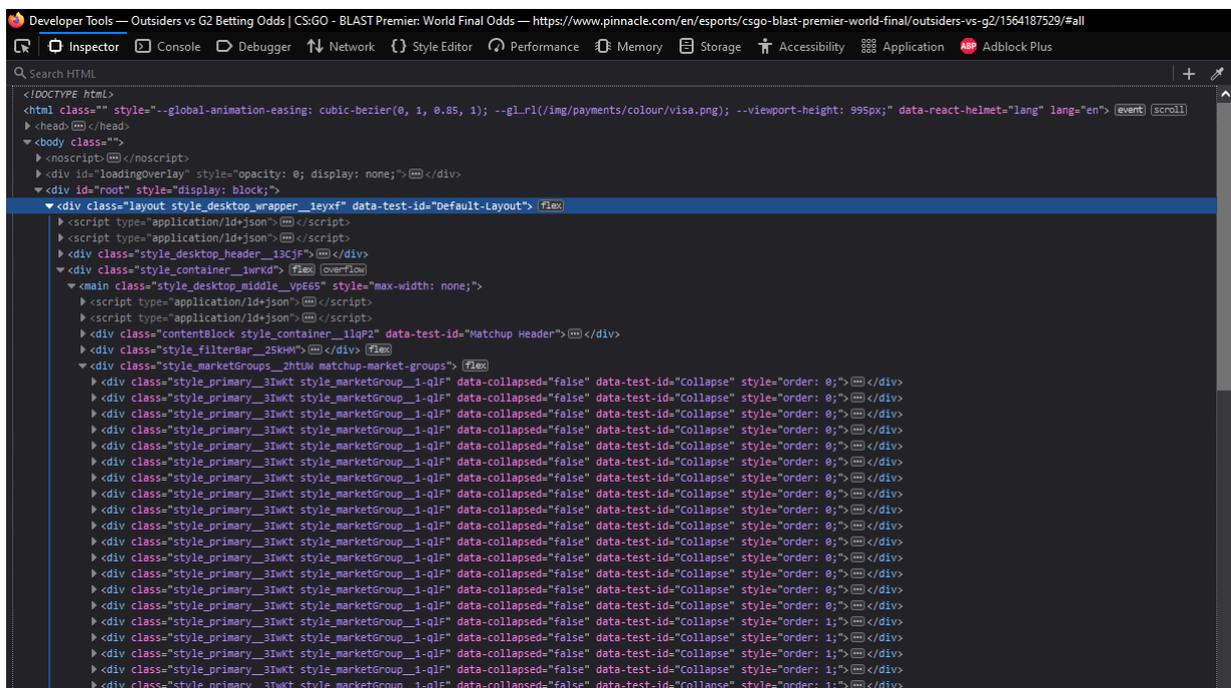


Figura 1 – Visualização interativa de uma página *HTML* do site da Pinnacle

Por último, cada tabela de *odds* é processada e formatada para o armazenamento na tabela *SQL*. Nessa etapa, algumas informações são padronizadas para facilitar a análise posterior, como exemplo, os 2 times da partida são ordenados em ordem lexicográfica. Além

disso, é computado o `match_id`, dado os nomes dos times e do evento e o `timestamp_id`, de acordo com o data e o horário da coleta (dia ou noite).

O programa é projetado para ser executado idealmente de 12 em 12 horas, uma vez que as casas de aposta normalmente disponibilizam precificações 24 horas antes do início da partida. Desse modo, temos uma alta probabilidade de capturar as precificações de uma dada partida, mesmo em caso de falha esporádica na execução do programa.

4.5 Coleta dos dados

A coleta dos dados foi feita utilizando o a ferramenta `cron` em uma máquina local, executando o *Web Scraper* duas vezes ao dia, uma de manhã e outra de noite, de 17 de Maio até 20 de Junho. Com isso, foram cobertos 42 eventos do jogo *Counter Strike: Global Offensive*, com 446 jogos registrados e 1784 mercados cobertos, sendo cada um deles com 2 a 4 casas de aposta. No total, foram feitas 57 coletas bem sucedidas.

4.6 Processamento dos dados

Essa etapa consiste em basicamente realizar o pareamento de cada par de *odd* extraída com as *odds* equivalentes das outras casas de aposta, ou seja, agrupar os conjuntos de *odds* em relação à partida e ao tipo de aposta. Cada conjunto desse tipo constituiu um mercado, para o qual vai ser calculado a possibilidade de arbitragem. Esse processo de pareamento pode parecer trivial, uma vez que bastaria agrupar cada par de *odd* com base em `match_id`, `timestamp_id` e `odd_type`, no entanto, diferenças de nomenclatura de eventos e times em casos pontuais dificultaram o pareamento direto em diversos casos. Como exemplo, um time Norte Americano bastante conhecido com o nome de *Evil Geniuses* é frequentemente referido como *EG* nas casas de aposta. Para isso, alguns nomes foram padronizados manualmente antes da execução do processamento, mas é provável que alguns casos tenham passado despercebidos. Esse problema pode ser um gargalo notável para a escalabilidade de um processo mais robusto de detecção de arbitragem, visto que existe uma quantidade bem mais elevada de times de esportes eletrônicos quando comparado aos esportes tradicionais.

5 Resultados

5.1 Possibilidade de Arbitragem

Com esses dados, é possível calcular a possibilidade de arbitragem em cada mercado coberto. Ao agregar tudo, temos:

- 41 mercados cobertos possuem possibilidade de arbitragem. O que é aproximadamente 2.3% do total de mercados.
- Retorno médio em cada arbitragem: 3.6%
- Desvio padrão: 2.12%
- Maior retorno encontrado: 12.012%

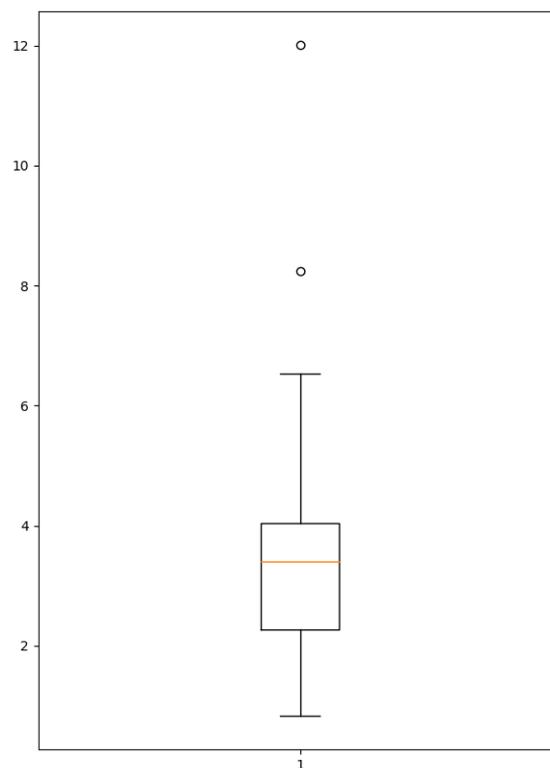


Figura 2 – *Boxplot* da distribuição de retorno nas arbitragens coletadas

Ao analisar a figura acima, é pode-se inferir que os retornos encontrados são valores baixos, por exemplo, 75% dos retornos é menor ou igual a 4%, contudo existem 2 *outliers* notáveis, de 12.012% e 8.236%, o que eleva a média no geral.

Com isso, é possível dizer que a possibilidade de arbitragem é possível, mesmo que em poucos casos com um retorno garantido pequeno. Como o banco de dados reflete um conjunto pequeno de casas de aposta e com uma quantidade relativamente pequena de amostras (de 12 em 12 horas) é provável que um trabalho mais robusto consiga extrair arbitragens com maior frequência. Além disso, precificações durante o jogo possuem uma volatilidade muito maior, de modo que uma coleta que consiga extrair dados durante o jogo deve encontrar arbitragens mais acentuadas e frequentes.

5.2 Discussão

O resultado demonstrado acima desperta imediatamente questionamentos sobre a viabilidade de usufruir dessas arbitragens, seja de forma automatizada ou manual. Contudo, existe uma série de problemas que dificultam uma operação desse tipo, algumas delas são:

1. Não concretização de uma aposta

Uma arbitragem depende de pelo menos 2 apostas em *websites* distintos, por isso é possível que uma aposta seja concretizada em um site enquanto outra seja cancelada ou impedida de ser executada, de modo que o apostador fique exposto ao risco da primeira aposta. Considerando o valor médio do retorno de arbitragem extraído, um erro desse tipo pode corresponder a 40 ou mais arbitragens bem sucedidas, diminuindo substancialmente a possibilidade de lucro total do apostador

2. Banimento de contas

Como a industria de aposta é pouco regulamentada, muitas casas de aposta se dão o direito de banir contas de apostadores por qualquer motivo. Alguns relatos em fóruns sobre aposta na internet afirmam muitas contas são impedidas de realizar qualquer aposta após o *website* detectar que o apostador está conseguindo um retorno médio positivo de forma consistente. Com isso, mesmo com o apostador podendo recuperar o saldo existente no *website*, o *pool* de casas de apostas é reduzido, já que existem medidas de identificação pessoal em cada conta (RG, CPF) que impossibilitam a criação de uma outra conta. Com isso, as oportunidades vão se reduzindo ao longo do tempo.

5.3 Trabalhos Futuros

Esse trabalho apresenta uma tentativa modesta de coleta de dados de precificações de apostas esportivas, uma vez que um trabalho robusto demanda um considerável esforço

de tratamento de *corner cases* e de limpeza de dados. Dentre os desafios que podem ser encontrados, destaca-se:

- Processamento correto de páginas *HTML*.
- Pareamento de partidas entre casas de aposta.
- Mudança no *layout* de um *website*, demandando a reformulação do código.
- Possível *rate limit* por parte do servidor, caso a coleta seja feita com muita frequência.

No entanto, caso um banco de dados maior pudesse ser coletado, seria interessante poder obter *insights* sobre as condições em que as arbitragens mais acontecem, seja por tipo de evento, times participantes, horário da partida, notoriedade da partida, tipos de mercados, entre outros. Além disso, existe a oportunidade de extensão da coleta para outros tipos de aposta ou mesmo de outros esportes, eletrônicos ou tradicionais, possibilitando a comparação entre os esportes.

Referências

- [Franck, Verbeek e Nüesch 2013]FRANCK, E.; VERBEEK, E.; NÜESCH, S. Inter-market arbitrage in betting. *Economica*, Wiley Online Library, v. 80, n. 318, p. 300–325, 2013.
- [Gray e Gray 1997]GRAY, P. K.; GRAY, S. F. Testing market efficiency: Evidence from the nfl sports betting market. *The Journal of Finance*, Wiley Online Library, v. 52, n. 4, p. 1725–1737, 1997.
- [Hausch e Ziemba 1990]HAUSCH, D. B.; ZIEMBA, W. T. Arbitrage strategies for cross-track betting on major horse races. *Journal of Business*, JSTOR, p. 61–78, 1990.
- [Knoll e Stübinger 2020]KNOLL, J.; STÜBINGER, J. Machine-learning-based statistical arbitrage football betting. *KI-Künstliche Intelligenz*, Springer, v. 34, n. 1, p. 69–80, 2020.
- [Vlastakis, Dotsis e Markellos 2009]VLASTAKIS, N.; DOTSI, G.; MARKELLOS, R. N. How efficient is the european football betting market? evidence from arbitrage and trading strategies. *Journal of Forecasting*, Wiley Online Library, v. 28, n. 5, p. 426–444, 2009.