

Modelagem e Análise de Dados Obtidos a Partir de Arbitragem em Sites de Apostas

Pedro Rios Bustamante

Departamento de Ciências da Computação

Universidade Federal de Minas Gerais

Belo Horizonte, Brasil

pedro.rios.bust@gmail.com

Crisitano Arbex Valle

Departamento de Ciências da Computação

Universidade Federal de Minas Gerais

Belo Horizonte, Brasil

arbex@dcc.ufmg.br

Abstract—Este trabalho investiga a viabilidade e a rentabilidade das apostas de arbitragem, utilizando ferramentas de web scraping e modelagem quantitativa. Com o crescimento do mercado de apostas on-line no Brasil, especialmente em eSports, identificamos oportunidades significativas para apostas de arbitragem. Utilizando dados reais coletados de grandes casas de apostas, realizamos simulações para avaliar os lucros potenciais com e sem a inclusão de riscos. As simulações demonstraram que os lucros mensais podem ser extremamente significativos, evidenciando o potencial desse método. No entanto, riscos associados, como erros nas odds após a aposta, foram considerados e modelados.

Além disso, exploramos estratégias alternativas, como apostar em uma única casa com odds favoráveis, e analisamos os desafios e limitações das abordagens, incluindo a necessidade de uma banca maior e o aumento da volatilidade. As conclusões indicam que, apesar dos desafios, a arbitragem oferece um campo promissor e pouco explorado, com muito potencial para futuras pesquisas. Sugerimos a exploração de diversas casas de apostas, a identificação de padrões de erro nas odds e a automatização do processo para maximizar os lucros. Também discutimos a importância de estratégias para as casas de apostas evitarem situações de arbitragem de forma eficaz e atraente para os jogadores. Concluimos que o estudo de apostas de arbitragem é uma área emergente e promissora, com vastas oportunidades de descoberta e otimização.

I. INTRODUCTION

Nos últimos anos, o mercado de apostas on-line tem crescido exponencialmente, impulsionado por uma série de fatores, incluindo avanços tecnológicos, maior acesso à internet e um aumento no interesse por esportes e jogos de azar. No Brasil, eventos recentes, como a regulamentação das apostas esportivas, têm contribuído significativamente para a expansão desse mercado. Esse crescimento apresenta uma série de oportunidades e desafios, tanto para as casas de apostas quanto para os apostadores.

A motivação para este estudo surge do interesse em compreender um fenômeno específico dentro do mercado de apostas on-line: as apostas de arbitragem. Este tipo de aposta, embora lucrativo para os apostadores, é considerado prejudicial pelas casas de apostas, pois permite que os apostadores garantam lucros independentemente do resultado do evento. A arbitragem de apostas ocorre quando discrepâncias nas odds oferecidas por diferentes casas de apostas são exploradas para garantir um retorno positivo, sem risco.

O objetivo deste trabalho é modelar as apostas de arbitragem através de técnicas de finanças quantitativas, interpretando-as como investimentos. Essa abordagem permite a realização de paralelos entre o mercado de apostas e o mercado financeiro, proporcionando uma análise numérica robusta das apostas de arbitragem. A meta final é identificar o potencial de exploração dessas apostas, facilitando uma futura análise comportamental tanto das casas de apostas quanto dos apostadores.

Este artigo está estruturado da seguinte maneira: na Seção II, apresentamos a revisão da literatura relevante, destacando os principais estudos sobre apostas de arbitragem e finanças quantitativas. Na Seção III, detalhamos a metodologia utilizada para modelar as apostas de arbitragem e discutimos os resultados obtidos a partir da modelagem e análise dos dados. Por fim, na Seção IV, concluímos o estudo com considerações finais e sugestões para pesquisas futuras.

Com esta estrutura, esperamos fornecer uma compreensão aprofundada sobre as apostas de arbitragem e suas implicações para o mercado de apostas on-line, bem como abrir caminho para estudos subsequentes que possam contribuir para uma melhor regulação e gestão desse mercado em crescimento.

Nesta seção, apresentamos uma revisão da literatura sobre casas de apostas, com foco em estudos que comprovam a capacidade dessas casas de fornecer odds melhores nos dias atuais. Além disso, discutimos a exploração de apostas em jogos de BlackJack e traçamos paralelos com apostas de arbitragem. Por fim, revisamos a literatura sobre finanças quantitativas, abordando conceitos de modelagem e previsão de risco em investimentos, o conceito de banca disponível e o que é considerado um retorno desejável em um investimento.

II. REVISÃO DA LITERATURA

Nesta seção, apresentamos uma revisão da literatura sobre casas de apostas, com foco em estudos que comprovam a capacidade dessas casas de fornecer odds melhores nos dias atuais. Além disso, discutimos a exploração de apostas em jogos de BlackJack e traçamos paralelos com apostas de arbitragem. Por fim, revisamos a literatura sobre finanças quantitativas, abordando conceitos de modelagem e previsão de risco em investimentos, o conceito de banca disponível e o que é considerado um retorno desejável em um investimento.

A. Estudos sobre Casas de Apostas

Vários estudos têm investigado a eficiência das casas de apostas em fornecer odds competitivas. Segundo [1], a competição entre as casas de apostas tem levado a uma melhoria significativa nas odds oferecidas aos apostadores. Isso se deve em parte ao aumento do número de operadores no mercado e ao uso de tecnologias avançadas de análise de dados para ajustar as odds em tempo real. [2] demonstra que as casas de apostas modernas utilizam algoritmos sofisticados para ajustar as odds, o que lhes permite oferecer odds mais atrativas em comparação aos métodos tradicionais.

B. Exploração de Apostas em Jogos de BlackJack

O BlackJack é frequentemente citado na literatura como um exemplo de jogo onde os apostadores podem obter uma vantagem sobre a casa. Estudos como [3] mostram que, com estratégias de contagem de cartas e gestão de banca eficaz, os apostadores podem reduzir a vantagem da casa a ponto de obter um retorno positivo a longo prazo. Este cenário é semelhante às apostas de arbitragem, onde os apostadores procuram explorar discrepâncias nas odds oferecidas por diferentes casas de apostas. Ao contrário do BlackJack, onde a aposta é feita em uma única opção, a arbitragem envolve apostar em múltiplas odds para garantir um lucro independente do resultado do evento, no entanto é possível traçar um paralelo na aposta feita com uma única opção com a aposta feita em somente uma das odds. Isso será trabalhado mais adiante.

C. Finanças Quantitativas e Modelagem de Investimentos

A aplicação de finanças quantitativas em apostas de arbitragem é fundamentada em conceitos de modelagem e previsão de risco, amplamente discutidos na literatura financeira. Segundo [4], a modelagem quantitativa de investimentos envolve o uso de técnicas matemáticas e estatísticas para prever retornos e avaliar riscos. Estudos como [5] destacam a importância de uma gestão de risco eficaz e a definição de uma banca disponível (o capital que o investidor está disposto a arriscar). A literatura também sugere que um retorno desejável em um investimento varia conforme o perfil do investidor, mas geralmente é considerado atrativo um retorno ajustado pelo risco que supere a taxa livre de risco, conforme discutido por [6].

D. Conceito de Banca Disponível

O conceito de banca disponível é crucial na gestão de apostas de arbitragem, similar à gestão de portfólios de investimentos. Estudos como [7] explicam que a banca deve ser gerida de forma a minimizar o risco de ruína e maximizar o crescimento a longo prazo. Técnicas como a estratégia de Kelly [8] são frequentemente mencionadas na literatura como métodos eficazes para determinar o tamanho das apostas em relação à banca disponível.

E. Retorno Desejável em Investimentos

A definição de um retorno desejável em investimentos é amplamente discutida na literatura de finanças. [9] sugerem que, para muitos investidores, um retorno anual que exceda a inflação e a taxa livre de risco é considerado desejável. Em apostas de arbitragem, o retorno desejável pode ser definido como um valor que compense adequadamente o risco envolvido e justifique o esforço e o tempo gastos na identificação e execução das apostas.

F. Ferramentas de Web Scraping para Arbitragem

Por último, mas não menos importante, um outro trabalho foi feito anterior a este, intitulado "Ferramenta de web scraping para arbitragem", que mostrou que é possível minimizar drasticamente o esforço e o tempo utilizado na identificação de apostas de arbitragem. Isso é imprescindível para que um estudo em larga escala faça sentido, pois reduz a carga operacional sobre os apostadores e aumenta a eficiência na identificação de oportunidades de arbitragem.

III. METODOLOGIA

Este trabalho inclui uma série de arquivos Python que serão anexados no apêndice. Esses arquivos contêm tanto uma cópia do código do trabalho anterior realizado para o web scraping, quanto as modelagens utilizadas para gerar as informações apresentadas nos resultados deste estudo.

É relevante mencionar que, embora os códigos Python sejam fornecidos para referência, este trabalho pretende ser explícito o suficiente para que a consulta aos códigos não seja necessária. Assim, mesmo aqueles que não possuem familiaridade com Python poderão acompanhar o processo de descrição das atividades realizadas sem impedimentos. A metodologia será descrita de forma clara e detalhada, permitindo uma compreensão completa das técnicas e abordagens utilizadas.

Para compreender a modelagem das apostas de arbitragem, é essencial primeiro entender o processo "manual" envolvido. Na prática, uma pessoa encontraria uma SureBet utilizando dois sites de apostas distintos. Uma SureBet ocorre quando há uma discrepância suficientemente grande nas odds oferecidas por diferentes casas de apostas, permitindo ao apostador garantir um lucro independente do resultado do evento.

O cálculo de como apostar em cada uma das odds de uma aposta é feito utilizando a seguinte fórmula:

$$A_1 = \frac{S}{O_1} \quad \text{e} \quad A_2 = \frac{S}{O_2}$$

onde: - A_1 é o valor a ser apostado na primeira casa de apostas, - A_2 é o valor a ser apostado na segunda casa de apostas, - S é o valor total a ser apostado (soma de A_1 e A_2), - O_1 e O_2 são as odds oferecidas pelas duas casas de apostas.

Para calcular o lucro garantido com essa SureBet, utilizamos a seguinte fórmula:

$$L = \frac{S}{\left(\frac{1}{O_1} + \frac{1}{O_2}\right)} - S$$

onde: - L é o lucro garantido, - S é o valor total apostado, - O_1 e O_2 são as odds oferecidas pelas duas casas de apostas.

A trivialidade de associar esse processo a um investimento reside no fato de que alguém que realiza uma SureBet obtém um retorno que é proporcional ao valor investido (apostado). Isso significa que, assim como em investimentos tradicionais, o retorno financeiro é diretamente relacionado ao capital aplicado, permitindo uma analogia direta entre apostas de arbitragem e estratégias de investimento.

A. Riscos e Limitações do Processo de Arbitragem

Embora as apostas de arbitragem possam garantir um lucro seguro em teoria, existem vários riscos e limitações associados ao processo que precisam ser explicitados.

Um dos riscos contornáveis envolve casas de apostas que não informam a quantidade máxima de valor que pode ser apostado. Isso é prejudicial porque, para garantir o lucro, precisamos adequar a quantidade da segunda aposta à quantidade da primeira. Esse risco pode ser contornado de duas formas:

1. **Apostando valores baixos:** Esta abordagem, embora reduza o risco, não é ideal para aqueles que buscam lucros expressivos. 2. **Utilizando casas de apostas que informam o limite máximo permitido:** Esta é a abordagem mais ideal, pois neutraliza virtualmente o risco, permitindo uma melhor gestão das apostas.

Todas as casas de apostas possuem um limite máximo para cada aposta, e esses limites tendem a ser menores em casos de arbitragem. Isso se deve ao fato de que apostas com valores mais altos são mais monitoradas. Para nossas simulações, adotaremos uma abordagem conservadora e assumiremos que estamos apostando um total de 500 reais, distribuídos entre as duas apostas, um valor que geralmente está abaixo dos limites da maioria das casas de apostas.

Outra limitação notável é a alocação de banca. Considerando que um indivíduo tem uma quantidade de dinheiro que deve ser dividida entre as duas casas de apostas e retida até que a aposta se encerre, há um tempo envolvido na movimentação dos fundos. Para maximizar as oportunidades de apostas nas simulações, assumiremos que a banca é de 5000 reais (10 vezes maior que o valor apostado). Esta estimativa fornece uma margem generosa para garantir que nenhuma oportunidade de aposta seja perdida devido à falta de fundos.

Um risco significativo é a possibilidade de que, após uma aposta ser feita, as odds da aposta no outro site mudem, reduzindo o lucro ou até mesmo resultando em prejuízo. Esse risco pode ser minimizado evitando apostas muito voláteis (como apostas ao vivo) e sincronizando as apostas para que ocorram simultaneamente. Nos testes realizados (que serão discutidos em breve), esse risco não se concretizou. No entanto, para uma modelagem completa, consideraremos uma estimativa pessimista: assumiremos que há uma chance de 5% de ocorrer uma perda, e que quando essa perda ocorre, o prejuízo será de 5% do valor investido.

B. Base de Dados Utilizada para as Análises

Para a geração das análises, utilizamos uma tabela com os dados obtidos pela ferramenta de web scraping. Esses dados

foram coletados em uma única sessão de apostas de dois sites diferentes: Betano e Rivalry, duas grandes casas de apostas no Brasil. A sessão de apostas foi focada em um jogo de eSports (Valorant).

A quantidade relativamente limitada de dados utilizada para gerar estimativas maiores se deve ao fato de que essas apostas foram de fato realizadas e confirmadas, gerando empiricamente um lucro e provando que o método funciona na prática. A informação não foi apenas coletada, mas sim validada por meio de apostas reais. É interessante notar que nenhuma perda ocorreu, mas mesmo assim, um risco associado será adicionado em alguns cálculos devido ao tamanho da amostra base.

A Tabela I apresenta as informações detalhadas das apostas feitas.

TABLE I
INFORMAÇÕES DAS APOSTAS FEITAS

OddA	OddB	Dias do Retorno
1.51	3.00	1
1.61	2.75	2
1.95	2.10	5
1.51	3.00	5
1.55	3.00	7
1.62	2.75	8
1.41	3.75	8
1.32	4.33	7
1.65	2.62	8
1.38	4.00	6

A tabela acima mostra as odds oferecidas pelas duas casas de apostas (OddA e OddB) e os dias do retorno (considerando o início da semana como dia 0) para cada aposta. Esses dados serão utilizados para gerar as análises subsequentes e para demonstrar o funcionamento do método de arbitragem.

A partir desses dados, um programa consegue traçar uma distribuição normal que nos ajude a gerar, através dos devidos cálculos de média e desvio padrão, tanto o tempo médio para que uma aposta de arbitragem surja quanto um valor aleatoriamente gerado (dentro da normal) de seu lucro. Os valores de lucro negativos podem ser desconsiderados, pois o programa não os considerará como arbitragem, e assim, serão considerados apenas como nenhuma aposta encontrada.

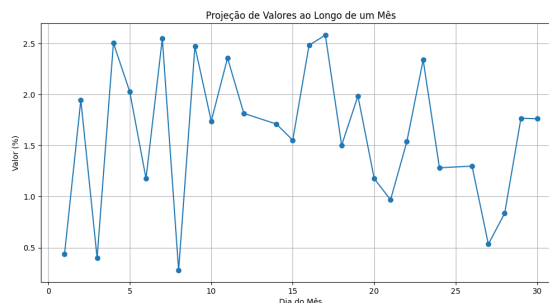


Fig. 1. Lucro por Aposta por Dia na Simulação de um Mês

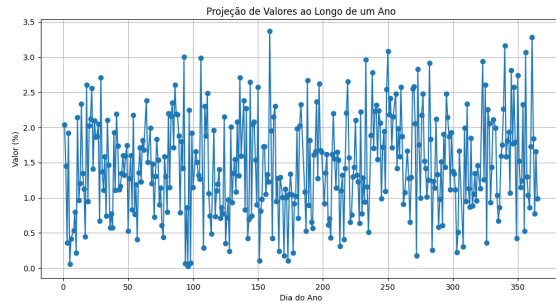


Fig. 2. Lucro por Aposta por Dia na Simulação de um Ano

C. Projeções e Simulações de Apostas

Nesta subseção, colocamos os dados à prova e coletamos informações sobre projeções, adicionando elementos relevantes aos cálculos. Primeiro, consideramos que não há elemento de risco (risco 0) e fazemos a projeção de uma sequência de 1000 simulações mensais para obter uma distribuição normal. Em seguida, adicionamos o risco mencionado anteriormente (5% de risco de uma perda de 5% acontecer) e repetimos a simulação para obter outra distribuição normal.

Para todas as apostas, estamos considerando que um total de 500 reais é aportado, e nossa banca total é de 5000 reais. O lucro em porcentagem é calculado com base no valor total da banca.

1) *Simulação Sem Risco*: Na primeira etapa, realizamos 1000 simulações mensais sem considerar o risco. Cada simulação utiliza os dados obtidos e calcula o lucro esperado para cada mês. Os resultados dessas simulações nos permitem obter uma distribuição normal do lucro mensal esperado sem a presença de risco.

A média do lucro percentual sem risco foi de 4.11%, com um desvio padrão de 0.43%.

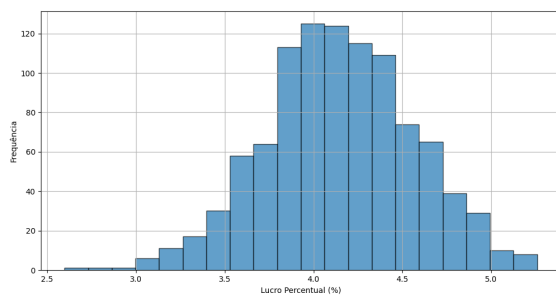


Fig. 3. Distribuição Normal do Lucro Mensal Sem Risco

2) *Simulação Com Risco*: Na segunda etapa, repetimos o processo de 1000 simulações mensais, desta vez incluindo o risco de uma perda de 5%. Em cada simulação, há uma probabilidade de 5% de que uma perda de 5% ocorra em uma das apostas. Os resultados dessas simulações nos permitem

obter uma distribuição normal do lucro mensal esperado com a presença de risco.

A média do lucro percentual com risco foi de 3.23%, com um desvio padrão de 0.85%.

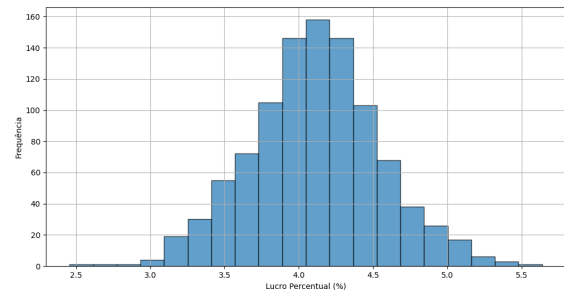


Fig. 4. Distribuição Normal do Lucro Mensal Com Risco

Estas simulações são cruciais para entender como o risco impacta os lucros das apostas de arbitragem. Ao comparar as distribuições normais das simulações com e sem risco, podemos avaliar a robustez do método de arbitragem e a importância da gestão de risco nas apostas.

D. Análise dos Resultados e Considerações Adicionais

Lucros mensais de 4% em um investimento são extremamente significativos, como discutido por [6]. Considerando o esforço mínimo dado à ferramenta de scraping, esses resultados são ainda mais impressionantes. No entanto, é importante considerar que, por mais atraentes que os números sejam, existem possibilidades fora da modelagem que não estão sendo consideradas.

Por exemplo, pode haver um erro em um dos sites depois que a aposta seja feita no outro, resultando em uma potencial perda significativa de dinheiro. Além disso, uma possível restrição nas apostas pode ocorrer se houver troca de dados entre os sites, permitindo que identifiquem as surebets e limitem as apostas.

Ao esclarecer o paralelo entre o BlackJack e as arbitragens, observamos que o comportamento das odds, após a arbitragem ser identificada empiricamente, tende a se balancear. Em quase todos os casos, as duas odds nas quais estaríamos apostando abaixam em seus respectivos sites, indicando que o site reavalia a aposta e a considera "boa demais" para o apostador. Em vez de tentar garantir o lucro correndo os riscos mencionados acima, podemos aproveitar essas apostas "boas" da mesma forma que um jogador de BlackJack aproveita quando a vantagem não está mais com a casa.

Para isso, vamos modificar um pouco a modelagem, que será apresentada a seguir.

E. A Casa Sempre Perde

Como mencionado anteriormente, se uma aposta de arbitragem aparece, o mais provável é que o cálculo de probabilidades das duas casas esteja defasado (inclusive, defasado o suficiente para ultrapassar a margem de lucro das casas). Nesse

caso, a menos que o erro de cálculo das probabilidades seja sempre maior na mesma casa, podemos assumir que os erros, por mais que sejam ocasionalmente diferentes, convergirão para o mesmo valor nas duas casas.

Para explicar isso com um exemplo, vamos considerar o segundo exemplo da Tabela I, onde temos $OddA = 1.61$ e $OddB = 2.75$.

Calculamos as probabilidades implícitas para essas odds da seguinte maneira:

$$P(A) = \frac{1}{O_1} = \frac{1}{1.61} \approx 0.6211$$

$$P(B) = \frac{1}{O_2} = \frac{1}{2.75} \approx 0.3636$$

A soma das probabilidades implícitas é:

$$P(A) + P(B) = 0.6211 + 0.3636 \approx 0.9847$$

Para calcular os valores a serem apostados em cada odd, usamos a fórmula discutida anteriormente:

$$A_1 = \frac{S}{O_1} = \frac{500}{1.61} \approx 310.56 \text{ reais}$$

$$A_2 = \frac{S}{O_2} = \frac{500}{2.75} \approx 181.82 \text{ reais}$$

Portanto, para uma aposta total de 500 reais, deve-se apostar aproximadamente 310.56 reais na OddA e 181.82 reais na OddB.

Este cálculo demonstra que, quando uma aposta de arbitragem é identificada, as probabilidades implícitas indicam que a soma das probabilidades não chega a 1 (ou 100%), o que sugere que existe uma margem de lucro para o apostador. Em teoria, a menos que ocorra um erro sistemático na mesma casa de apostas, os erros nas odds convergirão, eventualmente ajustando-se para refletir o verdadeiro equilíbrio das probabilidades.

Assumindo probabilidades igualmente distribuídas, o ganho esperado dessa aposta de exemplo seria o seguinte: se apostamos 63.07% (porcentagem de OddA no valor total apostado), temos 63.07% de chance de receber 101.55 reais e 36.93% (o restante) de chances de perder. Com um bom controle de montante e uma diversificação nas apostas, é possível tornar esse padrão de apostas muito rentável, apostando em somente uma casa.

No entanto, quando lidamos dessa forma com as apostas, apesar da margem de lucro continuar semelhante e de probabilisticamente nossas chances de ganhar serem maiores que as de perder, o risco aumenta muito, e com ele a necessidade de uma banca maior, para uma garantia menor de lucros. Ao simularmos 1000 situações onde, durante um mês, regularmente apostamos de forma coesa em uma mesma casa, geramos um gráfico (ver Figura 5) que nos mostra que, apesar da média do lucro ser coerente com o esperado (20.11 reais de lucro em um mês, com apostas de em média 50 reais), o desvio padrão do lucro é muito alto (243.35 reais), o que significa

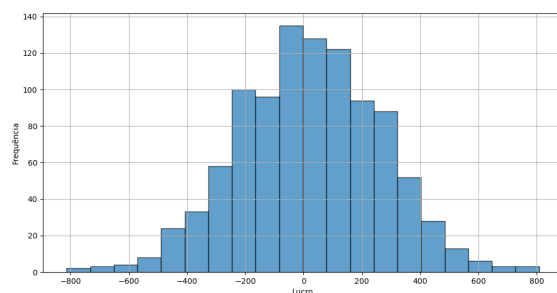


Fig. 5. Simulação de 1000 Situações de Apostas em Uma Única Casa por Mês

que a chance de perder (ou ganhar) muito dinheiro é sempre muito alta.

Isso é o esperado, considerando a volatilidade do mercado de apostas, mas a intuição diz que um refinamento maior da decisão de em qual das casas apostar pode tornar essa opção muito mais atraente, o que instiga pesquisas futuras.

IV. CONSIDERAÇÕES FINAIS E SUGESTÕES PARA PESQUISAS FUTURAS

Esta pesquisa evidenciou que os lucros com arbitragem podem ser extremamente significativos, proporcionando uma base sólida para que estudos mais extensos sejam realizados no futuro. Através da análise e modelagem dos dados, foi possível demonstrar a viabilidade e a rentabilidade das apostas de arbitragem, especialmente quando suportadas por ferramentas de web scraping e técnicas de modelagem quantitativa.

Uma das principais conclusões é a possibilidade de maximizar ainda mais esses lucros explorando diversas abordagens. Por exemplo, o uso de várias casas de apostas diferentes pode aumentar as oportunidades de arbitragem. Além disso, entender melhor quais casas de apostas "erram" mais frequentemente pode permitir a elaboração de estratégias mais eficazes para os apostadores. A implementação de automatizações que auxiliem na identificação e execução de apostas de arbitragem também pode ser uma área de pesquisa frutífera, permitindo retornos ainda mais altos para os investidores.

Conseqüentemente, há também a necessidade de pesquisas voltadas para as casas de apostas, para que possam desenvolver melhores estratégias para evitar situações de arbitragem que não são interessantes para elas. Estratégias que envolvem diminuir muito suas odds ou limitar excessivamente os apostadores são indesejáveis, pois tornam a casa menos atraente para os jogadores em geral. Portanto, estratégias mais complexas que identifiquem e evitem arbitragens em colaboração com outras casas de apostas são desejáveis e merecem uma investigação mais profunda.

Outra área promissora é a combinação das informações obtidas nesta pesquisa com outros conhecimentos técnicos de maximização de lucro em casas de apostas. A integração dessas técnicas pode levar a descobertas ainda mais valiosas e à otimização dos retornos financeiros.

Concluimos, finalmente, que a área de arbitragem em apostas é muito promissora e com muito a ser descoberto. A popularização recente dessa prática indica que há um vasto campo de pesquisa a ser explorado, com potencial para proporcionar insights significativos tanto para apostadores quanto para as próprias casas de apostas.

REFERENCES

- [1] D. Forrest, I. McHale, and J. McAuley, "Using Statistics to Detect Match Fixing in Sport," 2008.
- [2] D. Forrest and I. McHale, "Gambling and problem gambling among young adolescents in Great Britain," *Journal of Gambling Studies*, 2012.
- [3] E. O. Thorp, "Beat the Dealer: A Winning Strategy for the Game of Twenty-One," 1966.
- [4] J. C. Hull, "Options, Futures, and Other Derivatives," 2017.
- [5] H. Markowitz, "Portfolio Selection: Efficient Diversification of Investments," 1952.
- [6] W. F. Sharpe, "The Sharpe Ratio," *Journal of Portfolio Management*, 1994.
- [7] E. O. Thorp, "Portfolio Choice and the Kelly Criterion," *Journal of Portfolio Management*, 2012.
- [8] J. L. Kelly, "A new interpretation of information rate," *Bell System Technical Journal*, 1956.
- [9] E. F. Fama and K. R. French, "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, 1993.