

MEC-SETEC
INSTITUTO FEDERAL MINAS GERAIS - *Campus* Formiga
Curso de Ciência da Computação

**O USO DO NSGA-II PARA MAXIMIZAR O LUCRO E A
DISTRIBUIÇÃO DOS RECURSOS ENTRE APOSTAS ESPORTIVAS**

João Geraldo Borges Sales

Orientador: Everthon Valadão dos Santos

Formiga - MG

2023

JOÃO GERALDO BORGES SALES

**O USO DO NSGA-II PARA MAXIMIZAR O LUCRO E A
DISTRIBUIÇÃO DOS RECURSOS ENTRE APOSTAS ESPORTIVAS**

Monografia do trabalho de conclusão de curso apresentado ao Instituto Federal Minas Gerais - Campus Formiga, como requisito parcial para a obtenção do título de Bacharel em Ciência da Computação.

Orientador: Everthon Valadão dos Santos

Formiga - MG

2023

Sales, João Geraldo Borges

S163o O uso do NSGA-II para maximizar o lucro e a distribuição dos recursos entre apostas esportivas / João Geraldo Borges Sales – Formiga : IFMG, 2023.
70p. : il. color.

Orientador: Prof. Me. Everthon Valadão dos Santos

Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação)
Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais – *Campus*
Formiga.

1. Arbitragem de apostas. 2. Betting arbitrage. 3. Surebets. 4. Meta-heurística multiobjetivo. 5. Algoritmo NSGA-II. I. Santos, Everthon Valadão dos. III. Título.

CDD 004



MINISTÉRIO DA EDUCAÇÃO
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE MINAS GERAIS

Campus Formiga
Diretoria de Ensino
Docência Área Acadêmica de Computação
Rua São Luiz Gonzaga, s/n - Bairro São Luiz - CEP 35570-000 - Formiga - MG
- www.ifmg.edu.br

JOÃO GERALDO BORGES SALES

O uso do NSGA-II para maximizar o lucro e a distribuição dos recursos entre apostas esportivas

Trabalho de Conclusão de Curso apresentado ao Instituto Federal de Minas Gerais - Campus Formiga, como requisito parcial para obtenção do título de Bacharel em Ciência da Computação.

APROVADO em: 12 de dezembro de 2023.

BANCA EXAMINADORA

Prof.º Everthon Valadão dos Santos (orientador, IFMG)

Prof.º Mário Luiz Rodrigues Oliveira (IFMG)

Prof.º Marlon Jesus Lizarazo Urbina (IFMG)



Documento assinado eletronicamente por **Everthon Valadão dos Santos, Professor**, em 12/12/2023, às 12:14, conforme Decreto nº 10.543, de 13 de novembro de 2020.



Documento assinado eletronicamente por **Marlon Jesus Lizarazo Urbina, Professor Substituto**, em 12/12/2023, às 12:14, conforme Decreto nº 10.543, de 13 de novembro de 2020.



Documento assinado eletronicamente por **Mário Luiz Rodrigues Oliveira, Professor**, em 12/12/2023, às 12:19, conforme Decreto nº 10.543, de 13 de novembro de 2020.



A autenticidade do documento pode ser conferida no site <https://sei.ifmg.edu.br/consultadocs> informando o código verificador **1757258** e o código CRC **DF5700C8**.

Agradecimentos

Gostaria de agradecer a todos que estiveram presentes na minha caminhada durante o curso: família, colaboradores do campus, colegas e amigos.

Em especial, aos meus pais pelo apoio, incentivo e por dizerem que a educação é importante, mesmo eu não entendendo o significado disso antes. À minha namorada, que sempre esteve ao meu lado, dando todo o suporte e empurrão que eu precisava. Aos meus sogros, que são os melhores. Ao meu irmão, que ajudou isso a acontecer. E ao meu orientador Everthon, pela sua paciência e entusiasmo em ensinar sem medir esforços.

“Aprendemos com o fracasso, não com o sucesso!” (Bram Stoker)

Resumo

O mercado de apostas esportivas online está em ascensão, movimentando dezenas de bilhões de dólares. A abordagem convencional de aproveitar as oportunidades decorrentes de erros de precificação nas casas de apostas (*surebets*) pode não ser sustentável quando empregada sem moderação. Investimentos elevados, voltados para maximizar o lucro, facilitam a identificação dessa estratégia, como consequência possíveis restrições, podem ocorrer cancelamento de apostas ou até mesmo o bloqueio de suas contas por parte das casas de apostas. Este trabalho propõe o desenvolvimento de um protótipo de software para otimizar o lucro e a distribuição eficiente de recursos entre um conjunto de *surebets* disponíveis, aproveitando oportunidades decorrentes de erros das casas de apostas. A solução proposta adota o uso da meta-heurística *NSGA-II*, uma abordagem multi-objetivo, para gerar soluções que maximizem não só o lucro, mas também otimizem a distribuição do investimento entre as apostas com a intenção de evitar bloqueios e/ou limitações. Utilizando dados coletados de janeiro a outubro/2023 por meio de uma API e da construção de um algoritmo, montamos um *dataset* contendo mais de 2.400 jogos e centenas de *surebets* identificadas. Para validar o algoritmo foram considerados quatro cenários, explorando diferentes operadores de *crossover*. Com meta-heurística parametrizada, é gerado um conjunto de soluções que maximizam o lucro e a distribuição eficiente de recursos entre as *surebets*. Os resultados alcançados por este trabalho se traduzem em um protótipo de software que proporciona soluções adaptáveis a diversos perfis de apostadores, desde os mais conservadores, que preferem distribuir o investimento entre várias *surebets*, até os mais arrojados, que buscam maximizar o lucro e preferem investir na *surebet* que oferece o maior lucro.

Palavras-chave: Arbitragem de Apostas; *Betting Arbitrage*, *Surebets*; Meta-heurística Multiobjetivo; Algoritmo *NSGA-II*;

Abstract

The online sports betting market is on the rise, generating tens of billions of dollars. The conventional approach of capitalizing on opportunities arising from mispricing in bookmakers (surebets) may not be sustainable when employed without moderation. High investments aimed at maximizing profit make it easier to identify this strategy, resulting in potential restrictions, cancellation of bets, or even account blocking by bookmakers. This work proposes the development of a software prototype to optimize profit and efficiently distribute resources among a set of available surebets, leveraging opportunities arising from bookmakers' errors. The proposed solution adopts the use of the NSGA-II meta-heuristic, a multi-objective approach, to generate solutions that maximize not only profit but also optimize the distribution of investment among bets with the intention of avoiding blocks and/or limitations. Using data collected from January to October 2023 through an API and the development of an algorithm, we assembled a dataset containing over 2,400 games and hundreds of identified surebets. To validate the algorithm, four scenarios were considered, exploring different crossover operators. With the metaheuristic parameterized, a set of solutions is generated that maximizes profit and efficiently distributes resources among surebets. The results achieved by this work translate into a software prototype that provides adaptable solutions for various profiles of bettors, from the more conservative, who prefer to distribute investments among several surebets, to the more daring, who seek to maximize profit and prefer to invest in the surebet that offers the highest return.

Keywords: Betting Arbitrage; Surebets; Multi-objective meta-heuristic; NSGA-II algorithm;

Lista de ilustrações

Figura 1 – Fluxo iterativo do <i>NSGA-II</i>	28
Figura 2 – Representação híbrida do cromossomo para 3 <i>surebets</i>	37
Figura 3 – Estratégia de <i>Uniform Crossover</i> para geração de novos indivíduos na população.	38
Figura 4 – Mutação alterando valor binário do primeiro gene.	38
Figura 5 – Normalização sendo aplicada depois da mutação.	39
Figura 6 – Estabilização do <i>hypervolume</i> com os parâmetros do projeto fatorial.	48
Figura 7 – Estabilização do <i>hypervolume</i> para diferentes tamanhos de população.	49
Figura 8 – Gráfico com as <i>surebets</i> encontradas por mês.	49
Figura 9 – Gráfico com as <i>surebets</i> encontradas para o mês de setembro.	50
Figura 10 – Gráfico com as <i>surebets</i> encontradas para o mês de outubro.	51
Figura 11 – Estrutura dos dados coletados da API.	53
Figura 12 – Divisão das <i>surebets</i> encontradas em mês.	53
Figura 13 – <i>Surebets</i> encontradas por dia da semana, de janeiro até agosto/2023	54
Figura 14 – <i>Surebets</i> encontradas por dia da semana, em setembro e outubro/2023	54
Figura 15 – Divisão de <i>surebets</i> encontradas por casa de aposta	55
Figura 16 – Distribuição das <i>surebets</i> encontradas em intervalos de 1h	55
Figura 17 – <i>Surebets</i> encontradas a cada <i>slot</i> de 5 min., de 14h às 15h	56
Figura 18 – Evolução da fronteira de Pareto para o intervalo das 14:20h às 14:25h.	59
Figura 19 – Variação do lucro no intervalo das 14:00 às 15:00 para três finalidades	59
Figura 20 – Variação do lucro com os 3 objetivos sobrepostos.	60
Figura 21 – Lucro acumulado no intervalo de 14h–15h conforme cada objetivo.	61
Figura 22 – Lucro acumulado com os 3 objetivos sobrepostos.	62

Lista de tabelas

Tabela 1	–	Representação de <i>odds</i> americanas	20
Tabela 2	–	<i>Odds</i> para o lançamento de uma moeda	21
Tabela 3	–	<i>Odds</i> para o lançamento de uma moeda com comissão da casa.	22
Tabela 4	–	Diferentes probabilidades com seus respectivos retornos.	22
Tabela 5	–	Tabela com possibilidade de <i>surebet</i>	23
Tabela 6	–	Configurações dos computadores.	30
Tabela 7	–	Exemplo de Registros na Base de Dados	35
Tabela 8	–	Exemplo de três <i>surebets</i>	36
Tabela 9	–	Configuração inicial do primeiro projeto fatorial para todas as estratégias.	42
Tabela 10	–	Resultado do 1º projeto fatorial para o cruzamento <i>One Point Crossover</i>	43
Tabela 11	–	Resultado do 1º projeto fatorial para o cruzamento <i>Two Points Crossover</i>	44
Tabela 12	–	Resultado do 1º projeto fatorial para o cruzamento <i>Uniform Crossover</i>	44
Tabela 13	–	Configuração do 4º projeto fatorial para estratégia <i>One Point Crossover</i>	45
Tabela 14	–	Resultado do 4º projeto fatorial para estratégia <i>One Point Crossover</i>	45
Tabela 15	–	Configuração do 4º projeto fatorial para estratégia <i>Two Points Crossover</i>	46
Tabela 16	–	Resultado do 4º projeto fatorial para estratégia <i>Two Points Crossover</i>	46
Tabela 17	–	Configuração do 4º projeto fatorial para estratégia <i>Uniform Crossover</i>	47
Tabela 18	–	Resultado do 4º projeto fatorial para estratégia <i>Uniform Crossover</i>	47
Tabela 19	–	Lista de <i>surebets</i> para o intervalo de 14:20 às 14:25	56
Tabela 20	–	Comparação entre as 4 datas definidas como cenários de experimento.	63

Lista de quadros

Quadro 1 – Solução para maximização do lucro	57
Quadro 2 – Solução p/ maximização da quant. de <i>surebets</i> e alocação do recurso. .	58
Quadro 3 – Solução de equilíbrio entre os dois objetivos.	58

Lista de abreviaturas e siglas

API	<i>Aplication Programming Interface</i>
NSGA-II	<i>Non-dominated Sorting Genetic Algorithm</i>
RAM	<i>Random Access Memory</i>
MAT	<i>MATCHBOOK</i>
PIN	<i>PINNACLE</i>
UNI	<i>UNIBET_EU</i>
C.V	Coeficiente de Variação
H _v	<i>Hypervolume</i>

Sumário

1	INTRODUÇÃO	15
1.1	Justificativa	15
1.2	Objetivos	16
1.2.1	Objetivo Geral	16
1.2.2	Objetivos Específicos	16
2	FUNDAMENTAÇÃO TEÓRICA	18
2.1	Mercado de apostas esportivas <i>online</i>	18
2.2	Meta-Heurística Multiobjetivo	25
2.2.1	NSGA-II	25
2.2.2	Trabalhos relacionados	28
3	MATERIAIS E MÉTODO	30
3.1	Materiais	30
3.1.1	Hardware	30
3.1.2	Bibliotecas	30
3.2	Método	31
3.2.1	Coleta dos dados	31
3.2.2	Estudo da meta-heurística NSGA-II	31
3.2.2.1	Modelagem e codificação de um indivíduo como solução	31
3.2.2.2	Seleção dos pais	32
3.2.2.3	Implementação do operador de cruzamento	32
3.2.2.4	Implementação do operador de mutação	32
3.2.2.5	Implementação das funções objetivo	33
3.2.3	Calibração dos parâmetros do NSGA-II via Projeto Fatorial 2^k	33
3.2.4	Execução dos experimentos	33
4	DESENVOLVIMENTO	34
4.1	Coleta dos dados	34
4.2	Encontrando <i>surebets</i>	34
4.3	Implementação do NSGA-II	35
4.3.1	Representação do indivíduo	35
4.3.2	Cruzamento	37
4.3.3	Mutação	37
4.3.4	Função objetivo: Maximização do lucro	39
4.3.5	Função objetivo: Maximização da quantidade de <i>surebets</i> escolhidas	40

4.3.6	Escolha de uma solução dentre aquelas da fronteira de Pareto	41
4.4	Calibração dos parâmetros do NSGA-II via projeto fatorial 2^k	41
4.4.1	Resultado comparativo primeiro projeto fatorial 2^k	42
4.4.1.1	<i>One Point Crossover</i>	42
4.4.1.2	<i>Two Points Crossover</i>	43
4.4.1.3	<i>Uniform Crossover</i>	43
4.4.2	Resultado comparativo quarto (último) projeto fatorial 2^k	45
4.4.2.1	<i>One Point Crossover</i>	45
4.4.2.2	<i>Two Points Crossover</i>	46
4.4.2.3	<i>Uniform Crossover</i>	46
4.4.3	Definição dos parâmetros do NSGA-II	47
4.5	Definição dos cenários para realização dos experimentos	49
4.5.1	Setembro	50
4.5.2	Outubro	50
5	RESULTADOS E ANÁLISE	52
5.1	Caracterização do <i>dataset</i>	52
5.2	Experimento para data 02/09/2023	55
5.2.1	Maximização do lucro	56
5.2.2	Maximização de <i>surebets</i>	57
5.2.3	Equilíbrio entre os dois objetivos	57
5.2.4	Evolução da fronteira de Pareto	58
5.2.5	Variação do lucro no intervalo das 14:00h às 15:00h	59
5.2.6	Lucro acumulado no intervalo das 14:00h às 15:00h	60
5.3	Experimento comparativo	62
6	CONSIDERAÇÕES FINAIS	64
6.1	Principais contribuições	65
6.2	Trabalhos futuros	65
	REFERÊNCIAS	68

1 Introdução

Não é de hoje que o ser humano está envolvido com a atividade de realizar apostas. Seja jogos de dados, cartas, loterias, esportes ou cassino, todos esses segmentos também foram, ou são, utilizados com a prática de apostar (EADINGTON, 2008). Isso porque a atração por apostar está muitas vezes ligada ao entretenimento proporcionado por essa atividade. Seja qual for a modalidade escolhida, o apostador investe uma quantia em um possível resultado, esperando obter retorno financeiro caso sua aposta seja bem-sucedida.

As apostas esportivas *online*, funcionam de forma bem semelhante, mas possuem algumas particularidades. Nessa modalidade, o apostador utiliza uma plataforma *online* para realizar as suas apostas. As opções são amplas e variam desde a categoria esportiva, como o futebol, tênis, basquete, até os eventos esportivos dentro de cada categoria. Além disso, para cada evento esportivo existem variadas opções de mercados disponíveis para escolher no que apostar (essas opções podem ser apostar no time A como ganhador ou time B marcará 5 gols). O retorno a ser findado, dependerá da quantia investida pelo apostador, bem como os valores das *odds* (valor pago se um evento acontecer), e logicamente se o evento por ele apostado se concretizou.

Mundialmente falando, as apostas esportivas *online* tiveram um crescimento significativo. De acordo com Research (2023) em 2022, o mercado global atingiu 83,65 bilhões de dólares e a previsão é que vá atingir 182,12 bilhões de dólares de receita em 2030. Eles acreditam que esse crescimento se deu devido a fatores como legalização em setores de jogos de azar, acesso por dispositivos móveis e o desenvolvimento da infraestrutura digital.

Diante de um mercado expressivo como tal, é necessária uma estratégia que seja consistente, caso contrário os apostadores poderão não ter o sucesso desejado. Uma estratégia que questiona a eficiência do mercado é a arbitragem de apostas (ou em inglês *betting arbitrage*), basicamente um mercado é eficiente se ele consegue acompanhar as mudanças de um evento esportivo e projetar essas informações, rápido o suficiente, nas probabilidades do jogo/esporte. No entanto, ainda que exista essa estratégia, identificar a oportunidade pode não ser tão fácil devido à velocidade com que o mercado se movimenta, além de, quando identificado mais de uma, escolher o melhor conjunto que fará mais sentindo conforme as restrições, pode ser um desafio.

1.1 Justificativa

A cada ano que se passa, o mercado de apostas esportivas *online* cresce de forma notável e conforme as pesquisas de GrandViewResearch (2023) a estimativa é que o

mercado mundial vá atingir 130 bilhões de dólares em 2030. Entretanto, seja pelo uso de estratégias que não são assertivas, falta de agilidade para identificar oportunidades ou qualquer outro motivo, pessoas perdem dinheiro entrando nesse mercado. Para se ter uma ideia, ao passar apenas dois anos de liberação do mercado de apostas em no estado de Illinois (EUA), apostadores haviam perdido mais de 812 milhões de dólares e gerado 122 milhões em receitas para o estado ([TAX; ACCOUNTABILITY, 2022](#)).

Embora estratégias como a arbitragem de apostas, também conhecida como *betting arbitrage* ou *surebet*, se destaquem pela assertividade, empasses podem acompanhar esse método. Como a própria (capacidade de) identificação de tal oportunidade ou, quando já identificada e existindo mais de uma, quais devem ser escolhidas para atender o objetivo definido.

Existem diversos trabalhos na literatura abordando *surebets* (([FRANCK; VERBEEK; NÜESCH, 2013](#)), ([KAUNITZ; ZHONG; KREINER, 2017](#)), ([VLASTAKIS; DOTSI; MARKELLOS, 2009](#))), mas nenhum deles utilizam alguma estratégia para evitar que as casas de aposta apliquem limitações nos valores permitidos de apostar, ou que cancelem suas entradas, esses trabalhos estão focados em otimizar apenas o lucro. O nosso trabalho visa abordar o problema otimizando não apenas o lucro, mas também a distribuição eficiente do recurso entre as *surebets* escolhidas.

1.2 Objetivos

Nesta seção, serão apresentados os objetivos gerais e específicos que guiarão o desenvolvimento deste trabalho.

1.2.1 Objetivo Geral

O objetivo deste trabalho é desenvolver um protótipo de software capaz de identificar oportunidades para arbitragem em apostas esportivas e gerar um conjunto contendo a seleção das oportunidades encontradas. Esse conjunto terá o propósito de otimizar o lucro (em um ambiente simulado, mas com dados reais) e a distribuição eficiente dos recursos (dinheiro disponível) entre as *surebets* selecionadas.

1.2.2 Objetivos Específicos

1. Obter dados dos jogos de diversas casas de apostas *online*, de forma recorrente e periódica, persistindo tais informações para que com elas possam ser utilizadas nos treinamentos e validação do algoritmo.
2. Preparação dos dados coletados para estruturação/armazenamento em banco de dados, compondo o *dataset* utilizado no trabalho.

3. Fazer uma caracterização do *dataset* montado, para identificar tendências nas *surebets*.
4. Desenvolver um algoritmo para identificar a(s) possibilidade(s) de aplicação da estratégia de arbitragem de apostas (*surebet*) no *dataset* coletado.
5. Definir a heurística multi-objetivo a ser utilizada.
6. Implementar a heurística, modelando o problema na abstração (metáfora) dela, para atender ao objetivo definido.
7. Calibrar os parâmetros da heurística (via projeto fatorial) para atingir, no contexto do *dataset*, os melhores resultados com um menor tempo de processamento.
8. Realizar experimentos em ambiente simulado com a meta-heurística, consumindo o *dataset* construído para verificar a eficácia e eficiência da mesma.

2 Fundamentação Teórica

Nessa sessão serão abordados conceitos essenciais para compreender o problema e o desenvolvimento de uma solução eficiente. Estão relacionados a esses conceitos o mercado de apostas esportivas, coleta de dados, análise combinatorial e heurísticas.

2.1 Mercado de apostas esportivas *online*

Esse mercado está relacionado à atividade de realizar apostas em eventos esportivos por meio de plataformas *online*. Essas plataformas dão oportunidades aos jogadores de fazerem apostas em diversos esportes como o futebol, vôlei, basquete, corrida de cavalo, dentre outros. Nos próximos tópicos serão levantados itens relacionados ao mercado de aposta esportiva *online*.

- **Casa de aposta (ou *bookmaker*):** São as plataformas *online* que permitem realizar apostas.
- **Evento esportivo (ou *Event*):** São partidas/jogos que acontecem e os quais, os jogadores podem acompanhar para firmar uma aposta. Por exemplo, uma partida de futebol seria um evento em uma casa de aposta.
- **Mercado (ou *Market*):** Para cada evento esportivo nas casas de apostas, é disponibilizado ao jogador uma opção específica para apostar. Então, o mercado é uma forma do jogador investir em uma determinada perspectiva ou resultado do evento. Então em uma partida de futebol, pode-se ter os seguintes mercados (opções) para apostar:
 - Vencedor do jogo: o apostador tem a opção de dar um palpite sobre quem será o vencedor da partida.
 - Resultado correto: O jogador pode dizer, em sua opinião, qual seria o resultado da partida.
 - Marcador de gols: O jogador informa quais atletas marcarão os gols.

O termo evento, por vezes, também é utilizada como sendo para se referir a uma possibilidade de mercado, como mostrado nas opções anteriormente.

- **Cotação (ou *odd*):** No contexto de apostas esportivas *online*, as *odds* estão associadas ao valor que o apostador receberá caso a aposta seja bem sucedida. Algo interessante que podemos extrair em relação ao valor das *odds* são as intenções das

casas de apostas, ou melhor, o que elas acham que irá acontecer naquela partida, já que elas são obtidas com base em probabilidades. Além disso, podemos ver se existe um favorito para casa de aposta simplesmente olhando para o peso que ela coloca para cada opção de palpite disponível (ou evento).

Essas cotações são expressas em forma de números e podem ser apresentadas de várias configurações como, decimal, fracionário, formato americano ou outros.

- Decimal: Esse é o formato mais utilizado no Brasil, para saber o valor de retorno basta multiplicar diretamente. Suponhamos que o valor da *odd* para um palpite seja 2,50, então se for investido R\$ 100,00 e esse for o resultado correto, o retorno será:

$$\text{Retorno} = \text{odd} \times \text{investimento}$$

$$\text{Retorno} = 2.50 \times 100$$

$$\text{Retorno} = \text{R\$ } 250,00$$

para saber o valor da chance desse evento, ou palpite, acontecer em porcentagem fazemos:

$$\text{Porcentagem} = \frac{1}{2.5} \times 100 = 40\%$$

- Fracionário: Sendo o formato mais usado no Reino Unido, a sua representação é dada pelo formato $\frac{3}{2}$. Para saber o retorno que essa *odd* traria considerando que o investimento seja de R\$ 100,00, temos:

$$\text{Retorno} = \left(\frac{3}{2} + 1\right) \times \text{investimento}$$

$$\text{Retorno} = \left(\frac{3}{2} + 1\right) \times 100,00$$

$$\text{Retorno} = (1.5 + 1) \times 100,00$$

$$\text{Retorno} = \text{R\$ } 250,00$$

Para encontrar a chance desse evento ocorrer, em porcentagem, fazemos:

$$\text{Porcentagem} = \left(\frac{1}{\text{odd} + 1} \right) \times \text{investimento}$$

$$\text{Porcentagem} = \left(\frac{1}{\frac{3}{2} + 1} \right) \times 100,00$$

$$\text{Porcentagem} = \left(\frac{1}{2,5} \right) \times 100,00$$

$$\text{Porcentagem} = (0.4) \times 100,00$$

$$\text{Porcentagem} = 40\%$$

- Americano: Esse formato é um pouco mais complexo se comparado com os anteriores, podemos vê-lo sendo usado com frequência em ligas dos Estados Unidos. Nesse formato as *odds* são representadas por um número acompanhado de um sinal positivo(+) ou negativo (-) e possui como base o número 100.

Quando as *odds* são positivas, indicam o valor que é possível ganhar em uma aposta de 100 unidades. Por exemplo, se a *odd* valer +300, isto significa que caso o apostador ganhe, ele levará 300 unidades. Então se for apostado 100 unidades, e vencer é claro, o apostador terá como retorno 100 unidades (investidas) + 300 unidades ganhas.

Por outro lado, as apostas negativas, indicam o valor que é necessário desembolsar para receber 100 unidades de retorno. Então se a *odd* valer -300, deverá ser investido 300 unidades para ganhar 100. Dessa forma, se a aposta for vencida, terá como retorno 300 unidades (investimento) + 100 unidades ganhas.

Veja o resumo das *odds* americanas na tabela x.

Tabela 1 – Representação de *odds* americanas

<i>Odd</i>	Investimento	Retorno (se a aposta for vencida)
+300	R\$ 100,00	R\$ 300,00 (lucro) + R\$ 100,00 (investido)
-300	R\$ 300,00	R\$ 100,00 (lucro) + R\$ 300,00 (investido)

Fonte: Próprio autor.

- **Como as *odds* são criadas:** As casas de apostas possuem criadores de probabilidades e analistas para elaboração das *odds*. Para simplificar o entendimento de como são criadas essas probabilidades, considere um exemplo utilizando uma moeda:

No lançamento de uma moeda, podemos obter como resultado (face da moeda que ficará virada cima) CARA ou COROA. A probabilidade de sair cada um desses eventos é dado por:

$$P(A) = \frac{n(A)}{n(\Omega)}$$

onde,

$A \rightarrow$ Evento do espaço amostral

$\Omega \rightarrow$ Espaço amostral (Conjunto de todos os eventos)

$P(A) \rightarrow$ Chance do caso evento A se realizar.

$n(A) \rightarrow$ Número de casos favoráveis

$n(\Omega) \rightarrow$ Número total de possíveis casos.

Considerando que a moeda é justa, temos:

$$P(\text{CARA}) = \frac{1}{2} \rightarrow 0,5 \rightarrow 50\%$$

Como só existem duas faces e elas possuem as mesmas chances de saírem com a face virada para cima, chegamos a:

$$P(\text{CARA}) = 50\% \text{ e } P(\text{COROA}) = 50\%$$

Para obter o valor das *odds* justas, fazemos o seguinte cálculo:

$$\text{odds} = \frac{100}{\text{porcentagem}} \rightarrow \frac{100}{50} \rightarrow 2,0$$

Dessa forma, para cada evento (CARA ou COROA), a *odd* vale 2,0. Considerando esses eventos e a probabilidade 2,0, imagine o seguinte cenário onde dois apostadores (A e B) resolvem investir conforme a [Tabela 2](#):

Tabela 2 – *Odds* para o lançamento de uma moeda

Apostador (Evento escolhido)	Investimento	Odd	Retorno
A (CARA)	R\$ 100,00	2,0	R\$ 200,00
B (COROA)	R\$ 100,00	2,0	R\$ 200,00

Fonte: Próprio autor.

Perceba que um dos dois apostadores receberá o dinheiro do outro, isso porque $R\$ 100,00 * 2,0 = R\$ 200,00$, assim que funcionam as chamadas *odds* justas, entretanto se a realidade se fosse dessa forma as casas de apostas não lucrariam. Para ser possível obter lucro, os *bookmakers* cobram uma comissão descontada diretamente da probabilidade, veja a seguir na [Tabela 3](#), uma das possíveis *odds*:

Tabela 3 – Odds para o lançamento de uma moeda com comissão da casa.

Apostador (Evento escolhido)	Investimento	Odd	Retorno
A (CARA)	R\$ 100,00	1,90	R\$ 190,00
B (COROA)	R\$ 100,00	1,90	R\$ 190,00

Fonte: Próprio autor.

Com as *odds* valendo 1.90 para cada evento (CARA ou COROA), a casa de aposta lucra 5% em cada operação. Essa prática é conhecida como *Vigorish* e garante a longo prazo um lucro para as casas de apostas, independente do resultado. Outros nomes para essa prática são *Vig* ou *Fruit*. É dessa forma, então, que as casas de aposta esportivas *online* criam as suas *odds*. É claro que outros fatores podem contribuir, como estatísticas e acontecimentos externos.

A probabilidade e a cotação são inversamente proporcionais, então quanto maior uma, menor será a outra. Para confirmar essa propriedade, veja a seguir a probabilidade resultante aos termos diminuindo as *odds* de 2.0 para 1.90:

$$\begin{aligned} \text{probabilidade} &= \frac{100}{\text{odd}} + \frac{100}{\text{odd}} \\ \text{probabilidade} &= \frac{100}{1.90} + \frac{100}{1.90} \\ \text{probabilidade} &\approx 52,63 + 52,63 \\ \text{probabilidade} &\approx 105,26\% \end{aligned}$$

Para concluir, na [Tabela 4](#), podemos ver a relação entre as *odds* e as probabilidades. Quando a chance de todos os eventos bate 105% os *bookmakers* estão utilizando do *Vig* para tentar garantir um lucro a longo prazo, quando a probabilidade bate 100%, dizemos que temos as probabilidades justas, os *bookmakers* apenas repassam o dinheiro. Então se o apostador A ganhar considerando essas *odds* justas, ele receberá R\$ 100,00 que havia sido investido pelo apostador B. E por fim, quando a soma dos possíveis eventos fica abaixo de 100%, como a terceira linha com 95%, os *bookmakers* não só repassam o dinheiro, como também devem retirar do próprio bolso já que a proporção de investimento e pagamento está acima da probabilidade.

Tabela 4 – Diferentes probabilidades com seus respectivos retornos.

Probabilidade	Apostador	Investimento	Odd	Retorno
105%	A	R\$ 100,00	1,90	R\$ 190,00
	B	R\$ 100,00	1,90	R\$ 190
100%	A	R\$ 100,00	2,00	R\$ 200,00
	B	R\$ 100,00	2,00	R\$ 200,00
95%	A	R\$ 100,00	2,10	R\$ 210,00
	B	R\$ 100,00	2,10	R\$ 210,00

Fonte: Próprio autor.

- **Arbitragem de apostas (ou *surebet/betting arbitrage*):** É uma estratégia de aposta que visa aproveitar de *odds* com discrepância entre as casas de aposta, com essa estratégia é possível ter um lucro garantido independente do resultado da partida.

Para que isso seja possível, é necessário selecionar duas casas de apostas que estejam oferecendo apostas para o mesmo evento (ex.: uma partida de futebol) e disponibilizando a opção de apostar em um mercado comum (ex.: *Mais ou Menos 2,5 gols*). Além disso, o par de *odds* fornecidas por essas duas casas de apostas devem ser altas o suficiente, possibilitando a realização de apostas em todas as possibilidades e, assim, cobrir todos os resultados.

Só existem dois possíveis resultados para a partida apresentada na [Tabela 5](#), sair mais de 2,5 gols (isto é, 3 ou mais), ou então sair menos de 2,5 gols (que se traduz em 2 ou menos). Como dito anteriormente, para existir uma *surebet* deve ser possível cobrir todos os eventos e ainda ter um lucro.

Tabela 5 – Tabela com possibilidade de *surebet*.

Casa de Apostas	Mais de 2,5 Gols na partida	Menos de 2,5 Gols na partida
A	2,00	1,95
B	2,30	1,85

Fonte: Próprio autor.

Imagine que um jogador possua R\$ 100,00 e deseja apostar nessa falha, o valor para cada *odds* é dado por:

Identificando a chance dos eventos de ocorrerem:

$$Odd_{2,30} \rightarrow \frac{1}{2,30} \approx 0.435$$

$$Odd_{1,95} \rightarrow \frac{1}{1,95} \approx 0.513$$

Somando os dois temos:

$$0,435 + 0,513 \approx 0,948$$

Depois, devemos ver quanto deve ser investido em cada evento:

$$Odd_{2,30} \rightarrow \frac{0,435 * 100}{0,9480} \approx R\$ 45,80$$

$$Odd_{1,95} \rightarrow \frac{0,513 * 100}{0,9480} \approx R\$ 54,10$$

Agora, se multiplicarmos o valor encontrado para cada evento, temos:

$$Odd_{2,30} \rightarrow R\$ 45,80 * 2,30 \approx 105,34$$

$$Odd_{1,95} \rightarrow R\$ 54,10 * 1,95 \approx 105,49$$

no final, o resultado será positivo independente do resultado porque cobrimos todas as possibilidades, então se saírem 3 gols, o dinheiro iria todo para casa B, cobrindo a perda do A e um lucro de R\$ 5,34 seria alcançado.

É importante ressaltar que o lucro de R\$ 5,34, no exemplo apresentado, poderia ser ampliado caso a banca utilizada fosse maior. Como nos experimentos consideramos uma banca fictícia limitada ao valor de R\$ 100,00, podemos interpretar os resultados como um ganho percentual, equivalente a 5,34%. Assim, se a banca tivesse um valor de R\$10.000,00, o resultado seria um lucro de R\$ 534,00.

Alguns motivos pelos quais as *odds* ficam diferentes entre as casas de apostas são:

- Fórmulas: Cada *bookmaker* pode possuir diferentes probabilidades para cada evento esportivo, isso porque se baseiam em diferentes modelos e análises do mercado.
- Erro humano: Como dito anteriormente, os *bookmakers* utilizam de modelos e análises para criar as suas probabilidades e envolve pessoas trabalhando para criar essas probabilidades, erros humanos podem ocorrer, seja por colocar um valor que não deveria ou não alterar quando deveriam.
- Competição entre casas de apostas: pode existir competição entre as casas de apostas para serem mais atrativas e por consequência trazer mais apostadores para a sua plataforma, eles fazem isso aumentando o valor a ser pago em cada evento, o que no final contribui para uma discrepância entre as casas de aposta.
- Movimentação do mercado: por diversos motivos o mercado pode sofrer flutuações, como alguma notícia favorecendo ou desfavorecendo um time, grande volume de compra para um evento, lesões e assim por diante. De acordo com [Kaunitz, Zhong e Kreiner \(2017\)](#), até mesmo as casas de aposta fazem essa movimentação deliberadamente com o intuito de evitar grandes prejuízos, imagine que existam 3 possíveis resultados: time A vencer, Empate ou Time B vencer. Caso uma massa de apostas seja feita para o Time A vencer, por exemplo, a casa de aposta teria que arcar com o pagamento caso esse fosse o resultado. Dessa forma, a casa de aposta pode mudar aumentar as *odds* dos outros resultados (Empate e Time B para vencer) para atrair apostadores e assim criar um equilíbrio e não ter que bancar sozinha o resultado caso o Time A seja vencedor.

2.2 Meta-Heurística Multiobjetivo

Muitos desafios reais que demandem otimização, são de caráter multi objetivo, isto é, requerem que seja analisado e considerado mais de um critério/objetivo ao procurar uma solução satisfatória. Satisfatória porque é impossível encontrar uma solução ótima que atenda a todos os critérios ou objetivos.

Diferente dos problemas de cunho mono objetivo, que só possuem um critério a ser otimizado, o de caráter multi objetivo, possuem dois ou mais e ao melhorar um, os outros necessariamente sofrem algum grau de prejuízo, criando um dilema de *trade-off* (JONES; MIRRAZAVI; TAMIZ, 2002). E isso se agrava ainda mais quando os objetivos a serem otimizados são conflitantes entre si, para exemplificar podemos pensar no caso de uma montadora de veículos que deseja maximizar o seu lucro e minimizar a emissão de CO_2 , ao investir em menor emissão de CO_2 , essa montadora precisaria de uma tecnologia mais sofisticada, e isso atingiria o outro objetivo diretamente (de maximizar o lucro).

Por essa razão que não existe apenas uma solução quando se trata de problemas multi objetivo, nesse cenário temos um conjunto de soluções que sempre tenderá para um lado, isto é, temos uma solução boa no objetivo A e melhor no objetivo B, por outro lado, pode existir outra solução nesse conjunto ruim no objetivo A, mas é excelente no objetivo B. E para definir qual das respostas é a melhor, ou a que melhor atende, é utilizado um *decision-maker*, especialista ou pessoa interessada na resolução. Essa pessoa escolherá a solução que melhor lhe atende, conforme os seus critérios/experiências.

Para esses problemas multiobjetivos, estratégias conhecidas como Heurísticas ou Meta-Heurísticas multiobjetivo, são utilizadas. Esses métodos não garantirão uma solução ótima, visto que a combinação de múltiplos objetivos e seus *trade-off* o tornam impraticáveis, por outro lado, eles podem trazer um conjunto de soluções viáveis em um tempo razoável. Proposto por Deb et al. (2002), o NSGA-II é um exemplo de algoritmo genético que se destaca na resolução de problemas multiobjetivos. Ele utiliza a estratégia de seleção elitista, classificação não dominada e seu conjunto de soluções é denominado Pareto ótimo.

2.2.1 NSGA-II

Buscando na literatura algoritmos para resolver desafios que envolviam otimização de múltiplos objetivos, o NSGA-II foi o escolhido para tal função. A seguir, no pseudo-código 1, será apresentado as etapas e funcionamento do NSGA-II.

O NSGA-II, começa por gerando uma população inicial P de forma aleatória, como podemos ver na linha 2. Essa população gerada possui o tamanho N , definido previamente como parâmetro do algoritmo. No momento que isso é feito, é definido para cada indivíduo, a sua pontuação frente aos objetivos a serem otimizados no problema.

Algoritmo 1: NSGA-II (*Non-dominated Sorting Genetic Algorithm*)

```

1 begin
2    $P \leftarrow \text{INITIALIZE POPULATION}(N)$ 
3    $\text{EVALUATION}(P)$ 
4   for  $i$  from 1 to  $G$  do
5      $\text{STEP POPULATION} \leftarrow []$ 
6     while  $|\text{STEP POPULATION}| < N$  do
7        $I_0 \leftarrow \text{SELECT}(P)$ 
8        $I_1 \leftarrow \text{SELECT}(P)$ 
9        $\text{STEP POPULATION} \leftarrow \text{STEP POPULATION} \cup \text{EVOLVE}(I_0, I_1)$ 
10    end
11     $N \leftarrow N \cup \text{STEP POPULATION}[0 : N]$  ; /* Nova população com tamanho 2N */
12     $\text{EVALUATION}()$ 
13     $\text{SORT BY CROWDED COMPARISON}()$ 
14     $\text{SELECT SURVIVORS}()$ 
15  end
16 end
;                                     /* FONTE: Adaptado de (DEB et al., 2002) */

```

Na próxima etapa do algoritmo, no método $\text{EVALUATION}()$ da linha 3, ocorre um processo fundamental. Nesse ponto, a população P é submetida a uma avaliação que determina o grau de dominância que um indivíduo possui sobre o outro. Essa avaliação é indispensável para geração das fronteiras de Pareto. A dominância é um conceito importante e, é dito que uma solução x domina a solução y , se para todos os objetivos, x é melhor ou igual à solução y , além disso, deve existir pelo menos um objetivo onde a solução x é melhor que y (DEB et al., 2002).

Outro ponto importante nessa etapa do método $\text{EVALUATION}()$, é o cálculo de *Crowding Distance* (ou distância de aglomeração). Essa métrica, é calculada para todos os indivíduos de cada fronteira de modo a averiguar a dispersão das soluções no espaço de busca. A *Crowding Distance* é uma medida que será considerada em etapas posteriores do algoritmo. Ela também será usada para determinar quais indivíduos sobreviverão e avançarão para próxima geração. Em resumo, quanto maior o valor da *Crowding Distance*, maior é o espaçamento entre as soluções no espaço de busca. Esse espaçamento é desejável, para evitar que o algoritmo convirja prematuramente para uma solução local.

Entre as linhas 4 e 15, inicia-se o processo iterativo que se estenderá até que a quantidade de gerações definidas por G , seja alcançada. Primeiramente, uma lista vazia, denominada STEP POPULATION , é inicializada para conter as soluções resultantes do cruzamento da população P . Esse processo pode ser observado entre as linhas 5 e 10, e ele continua até que tenhamos uma população de tamanho N na STEP POPULATION .

Nas linhas 7 e 8, ocorre a seleção de dois indivíduos da população P para o processo de cruzamento. Na linha 9, o cruzamento é executado, gerando novos indivíduos. Nesse

momento, cada indivíduo pode sofrer uma mutação, a depender da taxa de mutação definida no NSGA-II e por fim, ele é adicionado à STEP POPULATION. Cada indivíduo resultante do cruzamento, passa por uma avaliação onde o seu desempenho em relação aos objetivos definidos, são analisados.

Após ter gerado uma população descendentes de P e tamanho N , é feito a união da população principal P e STEP POPULATION, isso pode ser visto na linha 11. Em um próximo passo (linha 12), é realizado a avaliação dessa nova população de tamanho $2N$, os mesmos processos que ocorreram na linha 3 se repetem aqui (classificação dos indivíduos em relação ao grau de dominância, montagem da fronteira de Pareto e calculado a distância de lotação para cada solução).

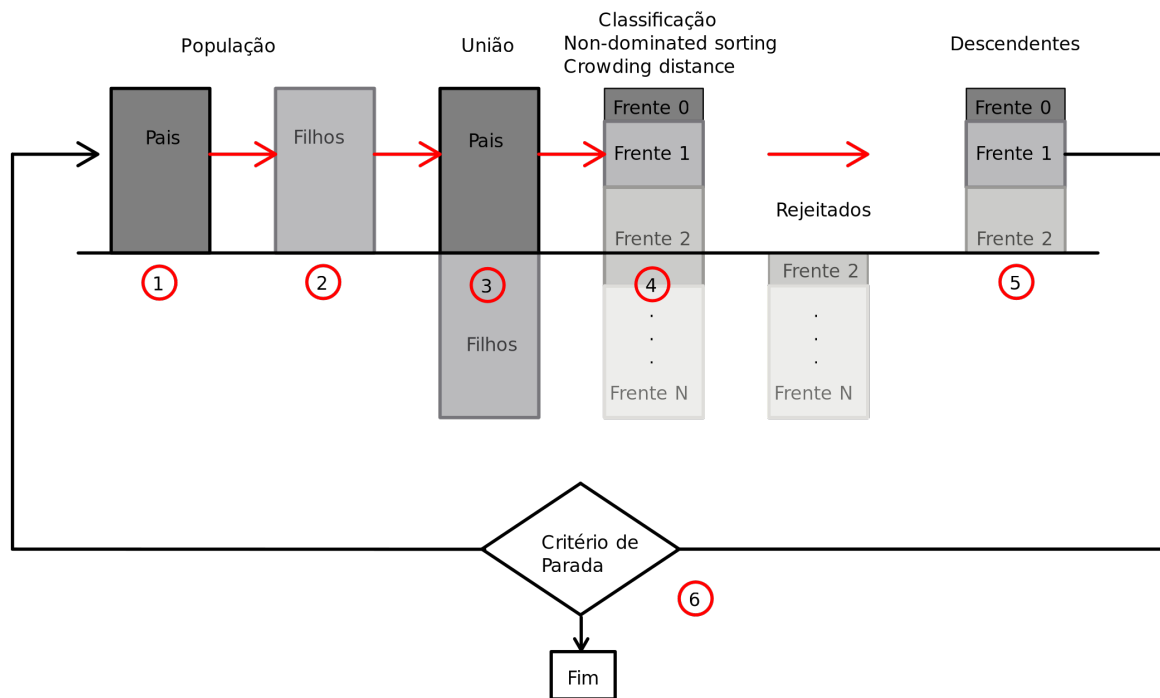
Na última parte do algoritmo (linhas 13 e 14), ocorre a ordenação das soluções de acordo com dois parâmetros, a fronteira que o indivíduo se encontra, crescentemente, e a distância de aglomeração decrescentemente. Essa ordenação será importante para definir quais indivíduos passarão para seguinte geração. E a escolha ocorre selecionando os N primeiros indivíduos da população P ordenada. Vale lembrar que em etapas anteriores ocorreu a união em P o que resultou em um tamanho $2N$. Ao fazer isso, a população que seguirá como base para próxima geração, conterá as melhores fronteiras e distância de aglomeração.

Depois disso, é dado sequência no processo iterativo, iniciando da linha 4 e indo até a 15, até que seja atingido o número de gerações definido como parâmetro do NSGA-II.

A Figura 1 apresenta de maneira clara o processo iterativo do funcionamento do NSGA-II. Na Etapa 1, ocorre a inicialização dos pais, representando N indivíduos na população. Em seguida, na Etapa 2, utilizando o cruzamento e aplicando a mutação, são gerados os descendentes. Logo após, na Etapa 3, as duas populações, de pais e filhos, são unidas, formando $2N$ indivíduos. Na Etapa 4, ocorre a organização dos indivíduos em fronteiras (*Non-dominated sorting*), e dentro de cada fronteira, há a ordenação decrescente do *Crowding distance*.

Uma parte da população é desconsiderada, sendo selecionada uma população de tamanho N na Etapa 5, onde ocorre o processo de elitismo do algoritmo. Isso significa serem selecionados os indivíduos que estão na melhor fronteira e possuem um maior espaçamento entre eles. Caso o critério de parada não tenha sido atingido na Etapa 6, é realizada outra iteração utilizando a população da Etapa 5 como os pais da próxima geração.

¹ Igor Carlos Pulini (Youtube/@PrOgRaMaNdO) <https://www.youtube.com/watch?v=L7Y_jD2BM2M&t=50s>

Figura 1 – Fluxo iterativo do *NSGA-II*

Fonte: Adaptado de Igor Carlos Pulini ¹

2.2.2 Trabalhos relacionados

Nessa seção, serão apresentados artigos e softwares do mercado relacionado ao trabalho desenvolvido.

O [oddsJam](#) é uma plataforma de assinatura com alguns recursos projetados para auxiliar os apostadores em suas atividades. Entre esses recursos, destacam-se na capacidade de detectar *surebets*, comparar *odds* entre casas de apostas e o seu banco de dados de estatística. Além disso, essa plataforma cobre uma extensa variedade de jogos esportivos. Apesar de possuir a detecção de *surebets*, o aproveitamento de todas essas oportunidades, em decorrência da precificação errada pelas casas de aposta, podem levar ao apostador ser descoberto e as casas de apostas bloquearem-lo/limitá-lo ou mesmo cancelar sua conta. A plataforma, não possui um módulo capaz de gerar um conjunto de arbitragens de acordo com alguns critérios. A única opção que existe é a lista de arbitragens disponíveis e os apostadores devem escolher cada uma, conforme o seu interesse. O trabalho desenvolvido consegue destacar e fornecer um conjunto de *surebets*, além de dividir o montante a ser apostado entre essas oportunidades, variando o lucro a depender do objetivo do apostador.

O autor [Levitt \(2004\)](#) estabelece uma comparação entre o mercado de apostas e o mercado financeiro, destacando que, apesar das semelhanças, eles são estruturados de maneira distinta. Pontos considerados como semelhança incluem: nos dois mercados, os

investidores possuem crenças e buscam lucrar em meio à incerteza; assim como no mercado de apostas, o mercado financeiro possui um *trader* de cada lado da transação; e ambos os mercados envolvem grandes quantias de dinheiro. Ele também aborda as estratégias das casas de apostas para obter lucro, enfatizando que elas possuem habilidades e informações superiores para definir os preços e aproveitar sistematicamente as tendências. Por exemplo, se souberem que os apostadores locais preferem times locais, podem ajustar os valores contra essas preferências para pagar menos. Sabemos que existem diversas maneiras das casas de apostas lucrarem, muitas vezes dominando o mercado. No âmbito deste trabalho, buscamos explorar as falhas de precificação por elas cometidas e mostrar, através de um ambiente simulado, mas com dados reais, que é possível lucrar em cima das casas de aposta aproveitando dessas falhas e possivelmente não ter suas operações bloqueadas ou limitadas.

Os autores [Kaunitz, Zhong e Kreiner \(2017\)](#) demonstraram, por meio de uma base histórica de 10 anos, que a estratégia *surebet* ou *betting arbitrage* é lucrativa. Durante um período de cinco meses fizeram apostas com dinheiro real nas casas de apostas e mostraram que o mercado de apostas *online* é ineficiente. Essa ineficiência indica que as probabilidades oferecidas pelas casas de apostas não refletem de maneira precisa as probabilidades reais dos eventos ocorrerem. No entanto, conforme os autores relatam, em poucas semanas as casas de aposta identificaram o uso da estratégia e colocaram limitações na quantidade a que poderia ser apostada. Em nosso trabalho, não visamos otimizar somente o lucro, a escolha e a distribuição eficiente dos recursos (banca para aposta) também é considerada.

3 Materiais e Método

Nesta seção, apresentamos os materiais e métodos que serão utilizados no desenvolvimento do trabalho.

3.1 Materiais

3.1.1 Hardware

Para a elaboração do projeto, foi utilizado um computador pessoal e *cluster* para se obter dados periodicamente, e as configurações se encontram na [Tabela 6](#).

Tabela 6 – Configurações dos computadores.

Item	Computador Pessoal	Cluster
Processador	Intel Core I5 9400	4x 1,2 GHz
Memória RAM	12 GB	12 GB
Armazenamento	256 GB SSD 1 TB HD	30 GB HD
Processador Gráfico	Nvi Geforce Gtx 1050 3 Gb	-
Sistema Operacional	Ubuntu 22.04.2 LTS	Ubuntu 22.04.2 LTS

3.1.2 Bibliotecas

A linguagem Python foi utilizada por oferecer uma vasta quantidade de bibliotecas, gratuitas, facilitando o desenvolvimento em diversos domínios. Além disso, se tratar de uma linguagem de alto nível e possui uma baixa curva de aprendizagem. As principais bibliotecas utilizadas foram:

- Requests: É uma biblioteca que permite fazer requisições HTTP de forma simples. Além disso, possui uma interface fácil e possibilita não só utilizar diversos métodos como GET, POST, como também manipular cookies e trabalhar com autenticação.
- Pydantic: Com essa biblioteca é possível fazer validações de objetos e definir tipos para modelos criados, ajudando a garantir uma consistência dos dados que está sendo trabalhado e se obter um *feedback* mais apurado em relação aos tipos.
- Sqlalchemy: Essa biblioteca faz o mapeamento de objetos para tabelas do banco de dados, também conhecido como mapeamento objeto-relacional ORM, ela oferece uma camada de abstração que torna forma mais fácil a interação com o banco de dados.

- Pandas: Bastante utilizada para manipular e realizar análise de dados, essa biblioteca possui um tipo de dado próprio, Dataframe, que permite trabalhar de forma tabular. Além disso, traz consigo vários métodos como filtragem, ordenação, agregação e outros.
- Pytz: É uma biblioteca que ajuda a trabalhar e manipular o fuso horário.
- Python-dotenv: permite carregar variáveis de ambiente a partir de um arquivo .env com a configuração CHAVE=VALOR.
- Injectable: Essa biblioteca fornece recursos para utilizar a injeção de dependência em Python.
- Pygmo: Biblioteca para gerar hypervolume da fronteira de Pareto.

3.2 Método

Abordaremos nessa seção, as etapas fundamentais para desenvolvimento do trabalho:

3.2.1 Coleta dos dados

Inicialmente, consideramos a possibilidade de realizar a coleta de dados por meio da técnica de raspagem de informações (*scrapy*) diretamente dos websites das casas de apostas. Entretanto, após uma análise dos termos de uso desses sites e consultas a professores de direito, optamos por abandonar essa abordagem, uma vez que os termos de uso proibiam essa prática.

Para contornar esse impedimento, a [API \(2023\)](#), foi utilizada, o que possibilitou a obtenção das *odds* e mercados esportivos de várias casas de aposta e jogos.

Devido maior familiaridade com os jogos de futebol, esse foi o evento esportivo escolhido para realizar a coleta de dados. Todos esses dados coletados podem ser visto no repositório do [GitHub](#). Veja a seção 4.1 para maiores detalhes sobre o desenvolvimento.

3.2.2 Estudo da meta-heurística NSGA-II

3.2.2.1 Modelagem e codificação de um indivíduo como solução

Um indivíduo é uma solução codificada do problema que desejamos otimizar. Essa codificação pode assumir diferentes formas, e a escolha da representação depende das características do problema em questão. A representação deve ser escolhida para refletir as características das soluções do problema de maneira eficiente. Quando decodificado, o

indivíduo fornece informações sobre os valores ou características da solução (COELLO; LAMONT; VELDHUIZEN, 2007).

Para esse trabalho, um indivíduo deveria carregar duas informações, sendo elas, dizer se uma *surebet* faz parte ou não da solução e o valor de investimento a ser alocado em cada *surebet*. Por isso, foi desenvolvido um indivíduo que trouxesse essas informações, além de ser possível realizar operações de mutação, cruzamento para evoluir e gerar novos indivíduos (população). Isso pode ser visto na seção 4.3.1.

3.2.2.2 Seleção dos pais

A seleção dos pais para o cruzamento adotou a estratégia de seleção por torneio. Nessa abordagem, dois indivíduos foram selecionados aleatoriamente da população em cada iteração, com o intuito de evitar qualquer tendência de seleção. A avaliação do vencedor desse torneio foi baseada em dois fatores:

- **Ranking:** Cada indivíduo foi classificado conforme a fronteira de Pareto à qual pertencia. Quanto menor o valor de ranking, melhor, uma vez que indicava que o indivíduo estava mais próximo da fronteira de Pareto ideal.
- **Distância de *Crowding*:** A distância de *crowding* é uma métrica que avalia a diversidade e a dispersão dos indivíduos. Quanto maior a distância de *crowding* de um indivíduo, melhor, pois indicava que a solução era mais espaçada e diversificada.

Essa estratégia de seleção por torneio foi adotada para garantir a escolha imparcial dos pais com base em critérios de ranking e diversidade, contribuindo para uma melhor exploração do espaço de soluções.

3.2.2.3 Implementação do operador de cruzamento

O cruzamento (também conhecido como recombinação) é um operador genético bastante frequente na geração de novos indivíduos e na evolução da população. Ele desempenha um papel importante ao combinar informações genéticas dos pais para criar descendentes que buscam soluções mais promissoras do que os próprios pais. Há diversas estratégias de cruzamento disponíveis, e a escolha do método mais apropriado depende das características específicas do problema. Neste trabalho, optamos por utilizar o cruzamento *Uniform Crossover*, conforme descrito na seção 4.3.2.

3.2.2.4 Implementação do operador de mutação

A mutação é outro operador genético que ajuda na evolução das soluções, mais especificamente esse operador introduz pequenas alterações na codificação do indivíduo

cujo objetivo é gerar uma diversidade na população, o que acaba ajudando a explorar outras regiões do espaço de busca. Mesmo não sendo tão frequente, possui um papel fundamental. Foi desenvolvido um operador de mutação que atua sobre a parte binária do gene, transformando de 1 para 0 e vice-versa, veja a seção 4.3.3.

3.2.2.5 Implementação das funções objetivo

Definir uma função objetivo que descreva bem a realidade não é tarefa fácil, mas é de suma importância, uma vez que são elas que darão um direcionamento para o algoritmo na busca de melhores soluções. Além disso, fica claro a distinção de uma solução boa de outra menos desejável, já que é possível quantificar, dar valor a cada solução.

Após ter desenvolvido a representação do indivíduo, operadores de cruzamento e mutação, foi elaborado as funções objetivo a serem otimizadas que se buscava maximizar, veja as seções 4.3.4 e 4.3.5.

3.2.3 Calibração dos parâmetros do NSGA-II via Projeto Fatorial 2^k

A otimização dos parâmetros da meta-heurística é outro ponto importante para garantir que o algoritmo trabalhe de maneira eficiente, evitando recursos desperdiçados e maximizando seu desempenho. Nesse sentido, utilizamos o projeto fatorial 2^k para calibrar os parâmetros que serão empregados nos experimentos descritos na seção 3.2.4.

Para esse processo de refinamento, realizamos quatro rodadas do projeto fatorial, abrangendo diferentes tipos de cruzamento. Mais informações sobre esse procedimento estão disponíveis na seção 4.4.

3.2.4 Execução dos experimentos

Para a condução dos experimentos, foram estabelecidos quatro cenários distintos, distribuídos entre os meses de setembro e outubro. Especificamente, as datas escolhidas foram 02/09/2023 e 24/09/2023 para setembro, e 02/10/2023 e 07/10/2023 para outubro. A escolha dessas datas considerou a presença de mais oportunidades de arbitragem identificadas, proporcionando uma quantidade significativa de dados para alimentar o algoritmo e avaliar seu desempenho.

Como era necessário definir uma quantia representativa para a banca, optou-se por realizar os experimentos com um montante de R\$ 100,00. No entanto, é importante destacar que esse valor não influencia diretamente nos resultados, uma vez que o lucro é proporcional à quantia alocada. Adicionalmente, é crucial ressaltar que os experimentos foram executados em um ambiente simulado, não envolvendo apostas reais no cenário real.

4 Desenvolvimento

No decorrer deste capítulo, serão detalhados os procedimentos e etapas fundamentais que deram forma a este trabalho. A seguir, exploraremos como ocorreu a coleta de dados que serviu como base, o desenvolvimento do NSGA-II, bem como a calibração de seus parâmetros para em seguida montar os cenários de estudo.

4.1 Coleta dos dados

A coleta de dados por meio da [API \(2023\)](#) foi realizada de maneira automatizada (com *crontrab* do Linux), envolvendo a execução de dois *scripts* distintos. O primeiro *script* operava diariamente às 21:05, correspondendo a 00:05 no horário Coordenado Universal (UTC). Nessa fase inicial, o *script* tinha a tarefa de identificar quais partidas de futebol estavam programadas para os próximos dias.

Para cada partida de futebol retornada como resultado da [API \(2023\)](#), o primeiro *script* agendava algumas tarefas para que fosse coletado dados específicos desse jogo, no primeiro momento era agendado duas tarefas, sendo 1h e 30 minutos antes do jogo acontecer. A execução dessas tarefas específicas era realizada por meio de um segundo *script*. O objetivo principal dessa operação era coletar dados detalhados relacionados a cada partida.

A estratégia de agendar tarefas em horários variados tinha como propósito capturar variações nas probabilidades das partidas, uma informação importante para identificar oportunidades de arbitragem.

Inicialmente, a coleta de dados estava limitada a cinco campeonatos específicos. No entanto, em decorrência da quantidade de *surebets* que estava sendo encontrado, decidiu-se por remover essa restrição e coletar dados das mais variadas ligas e torneios esportivos de futebol. Uma vez que o foco principal estava na capacidade de identificar oportunidades de arbitragem, independentemente da origem dos jogos.

4.2 Encontrando *surebets*

Após a coleta de dados usando a [API \(2023\)](#), o próximo passo consistiu em montar a base de dados que contivesse as possibilidades de arbitragens, o mercado escolhido foi o Mais/Menos gols. Devido à natureza da coleta de informações, era comum que a base de dados apresentasse registros de um mesmo jogo, isso porque a *surebet* se manteve aberta por uma janela de tempo, como exemplificado na [Tabela 7](#).

Tabela 7 – Exemplo de Registros na Base de Dados

Game ID	Bookmaker	Odds 1	Odds 2	Data de atualização
02b4ccd3b7848	MATCHBOOK	2.02	1.50	2023-09-01 18:44:00.000
02b4ccd3b7848	MATCHBOOK	2.12	1.45	2023-09-01 18:49:00.000
02b4ccd3b7848	MATCHBOOK	2.20	1.35	2023-09-01 18:54:00.000
02b4ccd3b7848	PINNACLE	1.25	1.99	2023-09-01 18:44:00.000
02b4ccd3b7848	PINNACLE	1.25	2.09	2023-09-01 18:49:00.000

Fonte: Próprio autor.

Nesse contexto, era necessário criar combinações válidas de registros antes de verificar a existência de *surebets*. A validação dessas combinações envolveu a verificação de duas condições:

Garantir que a casa de apostas não fosse a mesma, evitando a comparação de registros da mesma casa de apostas. Considerar a diferença de tempo na atualização dos registros, permitindo apenas combinações com uma diferença máxima de 5 minutos. Essa janela de tempo alinhava os testes com a realidade, uma vez que a [API \(2023\)](#) atualizava os dados a cada 5 minutos. Uma vez que combinações válidas foram geradas, o próximo passo foi aplicar um cálculo para determinar se havia oportunidades de arbitragem. Quando tais oportunidades eram identificadas, os dados correspondentes eram registrados em uma tabela específica de *surebets*.

Caso esse intervalo de 5 minutos fosse alterado para, por exemplo, 10 minutos, estaríamos modelando um cenário menos fiel à realidade. Isso porque poderíamos incluir na mesma análise uma *surebet* formada com odds da casa A coletadas às 10:01 e outra com odds da casa B coletadas às 10:11. Na prática, as odds da casa A podem ter sofrido alterações nesse período, inviabilizando a existência da *surebet*. Portanto, a escolha do intervalo de 5 minutos é importante para garantir uma representação mais precisa das condições que a [API \(2023\)](#) nos oferece.

4.3 Implementação do NSGA-II

4.3.1 Representação do indivíduo

A representação do cromossomo é um dos pontos mais importantes do algoritmo genético, é através dele que carregaremos as informações que traduzem o problema. Portanto, uma boa representação permitirá que o algoritmo explore o espaço de busca eficientemente sem perder informações necessárias, por isso ela deve estar ligada às funções objetivo, ser de fácil manipulação e interpretação.

O problema em questão requer uma solução que permita identificar, entre diversas opções de *surebets*, quais serão incluídas na solução, além de determinar a alocação de

dinheiro entre aquelas que forem selecionadas.

Optamos por uma representação híbrida do indivíduo, composta por duas partes distintas. A primeira parte é uma variável binária que sinaliza se uma *surebet* está incluída na solução (1) ou não (0). A segunda parte consiste em uma variável contínua, variando de 0 a 1, com duas casas decimais de precisão, indicando a porcentagem do montante a ser alocada para o investimento em uma determinada *surebet*. É importante ressaltar que, embora essa representação seja contínua em um modelo matemático, na realização computacional, podem ocorrer perdas de precisão. Essa precisão é mais do que suficiente, uma vez que uma das formas das casas de apostas identificarem o uso da estratégia de arbitragem é verificando investimentos com casas decimais. Isso acende um alerta, pois os centavos não trariam muito lucro, mas podem ser suficientes para manter a possibilidade de *surebet*.

Esta abordagem híbrida foi selecionada com base na metodologia apresentada por [Streichert, Ulmer e Zell \(2004\)](#), em seus estudos. Após uma análise, ficou evidente que essa abordagem se alinharia com as demandas do nosso problema, permitindo uma representação flexível e eficaz tanto para a seleção de *surebets* quanto para a alocação de recursos.

Inspirados pelo trabalho de [Streichert, Ulmer e Zell \(2004\)](#), notamos que, em determinadas situações, o montante total alocado para uma alocação inicial, operação de cruzamento ou mutação pode exceder o valor unitário (1). Para lidar com essa situação, utilizamos o processo de normalização também visto no trabalho, garantindo que a representação de cada valor esteja devidamente ajustada em uma escala que varia de 0 a 1.

Para exemplificar a representação do indivíduo, veja a tabela com 3 possibilidades de arbitragens.

Tabela 8 – Exemplo de três *surebets*

Surebet	Casa de Aposta	Odd	Lucro (%)
1	Betfair	2.30	5.2
	Bet365	1.95	
2	Betano	2.25	5.3
	Bet365	1.99	
3	Betfair	1.95	3.26
	Betsson	2.20	

Fonte: Próprio autor.

A [Figura 2](#), é a representação da codificação do cromossomo, e por consequência uma possível solução. Perceba que ela possui 3 posições (cada posição é um gene representativo de uma *surebet* da [Tabela 8](#)), cada posição é codificada de forma híbrida sendo representada por uma tupla que combina valores binários e reais. A parte binária indica se a *surebet* foi escolhida para fazer parte da solução (0 para não escolhida, 1 para escolhida), enquanto a

parte real representa a porcentagem alocada a essa *surebet*.

É importante observar que, mesmo que a *surebet* da primeira posição não tenha sido escolhida (valor binário 0), ainda assim ela possui um valor alocado na parte real. Essa informação é significativa para a etapa de mutação.

Figura 2 – Representação híbrida do cromossomo para 3 *surebets*.

Indivíduo		
(0, 0.72)	(1, 0.72)	(1, 0.28)

Fonte: Próprio autor.

4.3.2 Cruzamento

Para gerar novos indivíduos e promover a evolução da população, foram desenvolvidas três estratégias de cruzamento: *One Point Crossover*, *Two Points Crossover* e *Uniform Crossover*. Após avaliar o desempenho de cada uma delas, como descrito na 4.4 do projeto fatorial, optou-se por continuar com a estratégia de *Uniform Crossover*.

No *uniform Crossover*, cada gene é tratado individualmente, tendo igual probabilidade de contribuir para a construção do cromossomo filho. Após a seleção de dois pais, uma moeda é lançada para cada gene. Cada pai tem uma chance de 50% de que seu gene seja incluído no cromossomo filho. Neste trabalho, dois filhos são gerados em cada etapa de cruzamento, e o gene não selecionado para o primeiro filho é atribuído ao segundo cromossomo. Além disso, não foi o caso desse trabalho, mas é possível adicionar um viés, para um dos pais. Na Figura 3, podemos ver o resultado dessa operação.

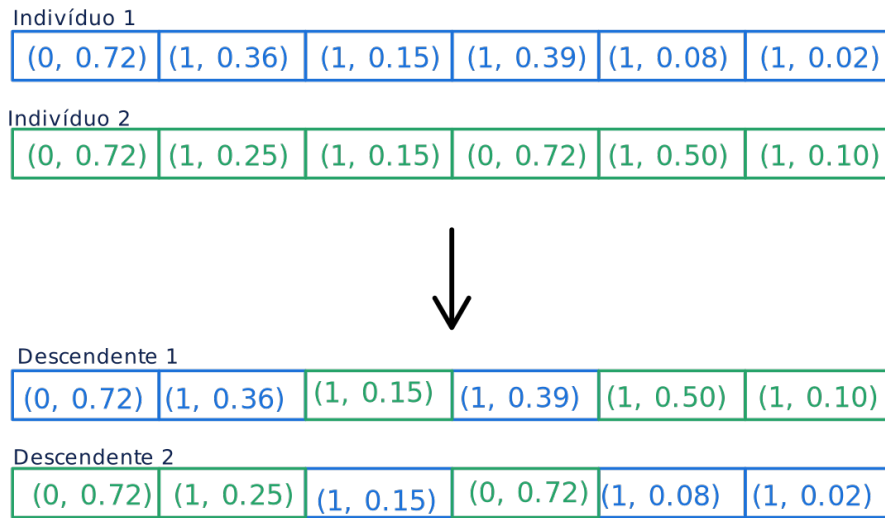
Durante o processo de cruzamento, surgiu a necessidade de realizar uma normalização. Isso ocorre devido à possibilidade de que a soma das alocações de recursos para todas as *surebets* escolhidas na solução exceda o valor de 1, o que representaria 100% da banca. Para corrigir essa situação, implementamos um processo de normalização, que será mostrado de forma mais detalhada na seção de mutação 4.3.3.

4.3.3 Mutação

Outro aspecto importante em algoritmos genéticos, é a mutação. Com essa estratégia, é possível introduzir uma diversidade no conjunto de soluções e como consequência, evitar que o algoritmo convirja prematuramente para uma solução sub-ótima. Além disso, ao inserir diversidades, é possível explorar regiões que ainda não foram investigadas.

Neste trabalho, a mutação foi aplicada na parte binária do gene. O processo de mutação envolve a seleção aleatória de um gene e a alteração de seu valor binário. Se o

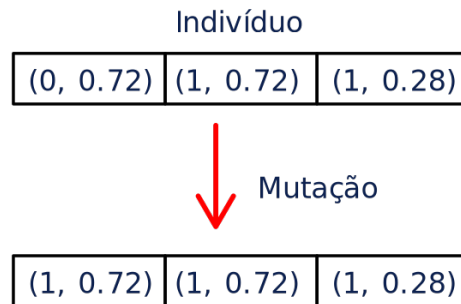
Figura 3 – Estratégia de *Uniform Crossover* para geração de novos indivíduos na população.



Fonte: Próprio autor.

valor binário anterior era 0, indicando que o gene não fazia parte da solução, ele é alterado para 1. Esse processo está representado na [Figura 4](#).

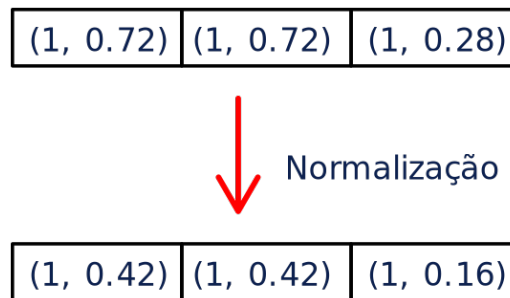
Figura 4 – Mutação alterando valor binário do primeiro gene.



Fonte: Próprio autor.

Uma consequência importante da alteração na parte binária é que a alocação total na parte real do gene também é modificada e se torna diferente de 1. Isso desencadeia um efeito cascata, resultando na necessidade de normalizar a alocação do investimento para garantir que ela seja adequada. Esse processo de normalização pode ser observado na [Figura 5](#). Vale ressaltar que, com a aplicação da normalização, o total da parte real retorna a 1. Essa mesma estratégia de normalização foi aplicada no trabalho [Streichert, Ulmer e Zell \(2004\)](#).

Figura 5 – Normalização sendo aplicada depois da mutação.



Fonte: Próprio autor.

4.3.4 Função objetivo: Maximização do lucro

Em qualquer análise de investimento, a consideração do potencial retorno, ou seja, o lucro, é de grande importância. Isso se aplica igualmente às apostas esportivas, onde a busca pela maximização do lucro desempenha um papel fundamental e reflete as necessidades do cenário real.

A função objetivo de maximização do lucro é uma componente fundamental dessa abordagem. Ela é construída diretamente: para cada *surebet* escolhida que faz parte da solução, multiplicamos o valor alocado de investimento pelo percentual de lucro associado a essa *surebet*, o lucro total se dá pela soma de todos os lucros. Essa abordagem nos permite avaliar e otimizar o lucro potencial em nosso processo de tomada de decisão.

$$\text{Lucro Total} = \sum_i x_i \cdot (\text{Lucro}_i/100) \cdot (y_i \cdot \text{Orçamento})$$

Nesta fórmula:

- \sum_i significa a soma sobre todas as *surebets* (representado por cada gene).
- x_i é uma variável binária que indica se a *surebet* i foi escolhida (1 para escolhida, 0 para não escolhida).
- Lucro_i é o lucro da *surebet* i em porcentagem (%).
- y_i representa a quantidade de dinheiro alocada na *surebet* i .
- Orçamento é o valor total do orçamento disponível para apostas.

Ainda que o desejo dos apostadores seja obter o maior lucro possível, apenas considerar esse objetivo pode ser arriscado, uma vez que as casas de aposta podem identificar o seu uso e aplicar penalizações. Conforme discutido na seção 2.1, a construção

de uma surebet envolve o uso de duas casas de apostas para cobrir todas as possibilidades e ter uma aposta segura. Ela é segura porque se perder em uma casa de aposta, ganharia na outra por se tratar de apostas complementares. No entanto, uma das ações, que acaba sendo uma penalização para o apostador, é o cancelamento da entrada por parte das casas de aposta. Então se uma das casas de aposta cancelar a entrada, e a outra permanecer aberta, o apostador corre o risco de perder o dinheiro caso a aposta vencedora não seja aquela que permaneceu aberta, tornando a surebet não totalmente segura.

Por esse motivo, não é seguro considerar apenas a maximização do lucro como função objetivo para modelar o problema. Para aumentar a segurança e evitar grandes prejuízos, incorporaremos outra função objetivo que visa maximizar a distribuição do investimento entre várias surebets, veja a seção 4.3.5.

4.3.5 Função objetivo: Maximização da quantidade de *surebets* escolhidas

Surebets são oportunidades de apostas esportivas que garantem lucro devido às discrepâncias nas odds oferecidas por diferentes casas de apostas. No entanto, essa estratégia não está isenta de riscos, como o cancelamento de apostas, limitação de valores investidos ou até mesmo o banimento de contas. As casas de apostas identificam a utilização dessa estratégia, que gera lucro para os apostadores, e impõem restrições, já que não querem perder dinheiro.

Visando evitar essas penalizações, desenvolvemos uma função objetivo específica. Sua finalidade é avaliar a distribuição do dinheiro entre as *surebets*, ou seja, além de quantas *surebets* fazem parte da solução, como é a divisão do recurso entre elas. Ao escolher várias oportunidades de arbitragem, evitamos concentrar o investimento em um pequeno número de *surebets*, reduzindo assim o risco de banimento e a possível perda¹ de todo o investimento e retorno potencial.

A base para a criação dessa função objetivo é o cálculo de entropia, uma abordagem utilizada em alguns estudos no campo da economia para avaliar a diversificação de uma carteira de investimento. Trabalhos anteriores também empregaram essa estratégia na montagem de carteiras de investimento, um desses é [Pola \(2016\)](#).

Veja a seguir a fórmula usada para calcular a entropia com base nas alocações normalizadas das *surebets* escolhidas. Quanto maior o valor absoluto de S , maior a diversificação das alocações de dinheiro.

$$S = - \sum_i y_i \log_2(y_i)$$

¹ A perda do investimento pode ocorrer no seguinte cenário: ao entrar em duas apostas em casas diferentes, se uma delas for cancelada, a *surebet* deixa de ser segura, e o apostador pode perder o valor investido.

Nesta fórmula:

- S representa a entropia.
- \sum_i denota a soma sobre todas as *surebets* escolhidas.
- y_i representa a proporção da alocação de dinheiro na *surebet* i .
- $\log_2(y_i)$ é o logaritmo na base 2 da proporção da alocação.

4.3.6 Escolha de uma solução dentre aquelas da fronteira de Pareto

Após a conclusão do processo de coleta de dados, a verificação da existência de *surebets* e a execução do algoritmo NSGA-II, obtemos um conjunto de soluções, cada uma com suas características distintas. Algumas dessas soluções priorizam o lucro máximo, mas podem ser arriscadas ao concentrar o investimento em um número limitado de *surebets*, reduzindo a segurança das apostas, caso tenha a entrada cancelada por uma das casas de aposta ou qualquer outro motivo como abordado na seção 4.3.5. Por outro lado, outras soluções optam por uma distribuição mais ampla do investimento, renunciando a lucros mais altos.

Na prática, a escolha da solução é geralmente feita por um especialista ou por alguém com interesses específicos que atendam às suas necessidades. Para efeitos de avaliação dos resultados em diferentes cenários, destacamos três soluções: uma focada na maximização do lucro, outra na maximização da quantidade de *surebets* selecionadas e uma terceira que busca um equilíbrio entre esses dois objetivos.

4.4 Calibração dos parâmetros do NSGA-II via projeto fatorial 2^k

O projeto fatorial empregado na calibração dos parâmetros incluiu quatro fatores:

1. Número de gerações
2. Tamanho da população
3. Taxa de mutação
4. Probabilidade de cruzamento

Cada um com dois níveis distintos. Isso resultou em um total de 16 configurações para conduzir os testes. Além disso, o experimento foi realizado em 10 iterações, totalizando 160 execuções para avaliar o desempenho do algoritmo.

Para obter uma calibração mais refinada dos parâmetros e determinar a estratégia de cruzamento mais eficaz dentre as desenvolvidas (*One Point Crossover*, *Two Points Crossover* e *Uniform Crossover*), as configurações e iterações foram aplicadas a cada uma das três estratégias de cruzamento durante três rodadas de testes.

A avaliação do desempenho dos testes foi conduzida com base em duas métricas principais: o *hypervolume* e o tempo de execução (milissegundos). Dado o número de iterações realizadas, foi necessária uma análise estatística para resumir o desempenho do algoritmo sob diversas configurações. Nesse contexto, optamos por calcular a mediana dos resultados em vez da média. A escolha da mediana se justifica pela sua robustez em relação a valores discrepantes, também conhecidos como *outliers* que a média pode sofrer. Assim, a mediana foi utilizada para obter um conjunto representativo das iterações realizadas, garantindo uma avaliação mais estável.

Na próxima subseção (4.4.1), detalharemos os parâmetros empregados na primeira rodada do projeto fatorial, destacando as configurações específicas para cada estratégia de cruzamento e trazendo a melhor.

4.4.1 Resultado comparativo primeiro projeto fatorial 2^k

Adotou-se na primeira rodada do projeto fatorial parâmetros uniformes para todas as estratégias de cruzamento, conforme detalhado na Tabela 9. Esses parâmetros foram definidos para estabelecer limites no qual, utilizando uma estratégia de busca binária, seriam ajustados à medida que fossem obtendo respostas do algoritmo para cada estratégia de cruzamento.

Tabela 9 – Configuração inicial do primeiro projeto fatorial para todas as estratégias.

k = 4	FATORES				FATORES				FIXO
	A		B		C		D		Iterações
	(nº de gerações)		(tam. população)		(taxa de mutação)		prob. de cruzamento		
NÍVEIS	1	50	1	10	1	0,05	1	0,70	10
	2	500	2	250	2	0,50	2	0,95	10

Fonte: Próprio autor.

4.4.1.1 *One Point Crossover*

Na Tabela 10, são destacados os dois melhores e dois piores resultados em termos de *hypervolume* e tempo de execução (em milissegundos) para cada configuração estabelecida no primeiro projeto fatorial para *One Point Crossover*, assim como o C.V (Coeficiente de Variação)² obtido por tal resultado. O algoritmo demonstrou melhores respostas de *hypervolume* quando o tamanho da população e a probabilidade de cruzamento foram

² O coeficiente de variação indica a variabilidade relativa, do *hypervolume* nesse caso, em relação à média. Ter um CV baixo, manifesta que os valores da distribuição estão próximos da média.

configurados em níveis mais elevados, enquanto a taxa de mutação foi mantida mais baixa. Quanto ao número de gerações, os resultados do primeiro e do segundo nível revelaram preferências distintas. Mesmo com essa discrepância, optou-se por reduzir o número total de gerações. Em relação ao tamanho da população, decidiu-se aumentar o valor para a configuração 1 (de 10 para 50) e diminuir para a configuração 2 (de 250 para 125) na próxima rodada do projeto fatorial. Dessa forma, a melhor configuração resultou em A (50 e 500), B (250), C (0,05) e D (0,95), representando os valores específicos A (50 e 500), B (250), C (0,05) e D (0,95).

Tabela 10 – Resultado do 1º projeto fatorial para o cruzamento *One Point Crossover*.

MELHORES E PIORES CONFIGURAÇÕES						
CONFIGURAÇÕES				MÉTRICAS		
A	B	C	D	Qualidade do Resultado (hypervolume)	C.V (%)	Tempo Gasto (ms)
1	2	1	2	31,0051	0,27	1982,15
2	2	1	2	30,9918	0,29	24447,01
2	1	2	1	26,2101	7,87	128
1	1	1	1	25,3019	9,09	14

Fonte: Próprio autor.

4.4.1.2 *Two Points Crossover*

De maneira análoga ao projeto fatorial do *One Point Crossover*, discutido na seção 4.4.1.1, os dois melhores e piores resultados do *Two Points Crossover* estão resumidos na Tabela 11. A estratégia *Two Points Crossover* também revelou um desempenho superior quando o tamanho da população e a probabilidade de cruzamento foram configurados nos níveis mais altos (250 e 0,95 respectivamente), enquanto a taxa de mutação permaneceu mais baixa (0,05). Mais uma vez, os dois níveis do fator número de gerações surgiram como os melhores resultados, e, assim como na seção anterior, houve uma redução no valor desse parâmetro. Em relação ao tamanho da população, o nível 1 foi aumentado de 10 para 50, enquanto o nível 2 foi reduzido de 250 para 125, buscando uma aproximação aos melhores parâmetros para esse fator. Portanto, os resultados mais relevantes desse projeto fatorial foram obtidos com os seguintes valores: fator A (500 e 50) para o número de gerações, B (250) para o tamanho da população, C (0,05) para a taxa de mutação e D (0,95) para a probabilidade de cruzamento.

4.4.1.3 *Uniform Crossover*

Ao analisarmos as configurações do primeiro projeto fatorial, conforme detalhado na Tabela 9, observamos que a estratégia *Uniform Crossover*, da mesma forma que *One Point Crossover* e *Two Points Crossover*, apresenta melhor desempenho com níveis mais elevados

Tabela 11 – Resultado do 1º projeto fatorial para o cruzamento *Two Points Crossover*.

MELHORES E PIORES CONFIGURAÇÕES							
CONFIGURAÇÕES				MÉTRICAS			
A	B	C	D	Qualidade do Resultado (hypervolume)	C.V (%)	Tempo Gasto (ms)	
2	2	1	2	30,9964	0,21	25788,01	
1	2	1	2	30,9590	0,25	2129,71	
1	1	2	1	25,8911	3,45	16	
1	1	1	1	25,0101	5,68	15	

Fonte: Próprio autor.

dos fatores B (tamanho da população) e D (probabilidade de cruzamento), além de um fator C (taxa de mutação) mais baixo, como as demais estratégias. Além disso, o fator A (número de gerações) também demonstrou os dois níveis (50 e 500) de configurações como sendo os melhores.

As configurações que apresentaram melhor resultado foram: A (50 e 500 para número de gerações), B (250 para o tamanho da população), C (0,05 para a taxa de mutação) e D (0,95 para a probabilidade de cruzamento). No próximo projeto fatorial, optou-se por reduzir o valor do fator A (número de gerações) para diminuir o tempo de execução. Além disso, para o fator B (tamanho da população), elevou-se o valor do nível 1 (de 10 para 50) e reduziu-se o do nível 2 (de 250 para 125).

Tabela 12 – Resultado do 1º projeto fatorial para o cruzamento *Uniform Crossover*.

MELHORES E PIORES CONFIGURAÇÕES							
CONFIGURAÇÕES				MÉTRICAS			
A	B	C	D	Qualidade do Resultado (hypervolume)	C.V (%)	Tempo Gasto (ms)	
2	2	1	2	31,0295	0,21	25797,40	
1	2	1	2	31,0146	0,33	2019,35	
2	1	2	1	26,1581	4,34	132	
1	1	1	1	26,0490	11,64	14	

Fonte: Próprio autor.

Uma análise inicial da primeira rodada do projeto fatorial revela que as três estratégias implementadas apresentaram resultados de *hypervolume* muito próximos, dificultando assim a escolha entre elas nesse estágio inicial. Todas as estratégias tiveram os dois níveis do fator A (número de gerações) entre as duas melhores respostas e observou-se um aumento no tempo de execução para o nível 2 (500 gerações). Por esse motivo, optou-se por reduzir esse valor, considerando que a diferença entre os dois melhores resultados não era significativa.

4.4.2 Resultado comparativo quarto (último) projeto fatorial 2^k

Embora tenham sido implementadas diferentes estratégias, a calibração dos parâmetros nos projetos fatoriais anteriores (2 e 3) revelou resultados bastante similares, como também abordado na seção 4.4.1. Essa tendência persistiu até a última rodada do projeto, na qual os mesmos parâmetros foram mantidos para avaliação.

4.4.2.1 One Point Crossover

A Tabela 13 apresenta as configurações utilizadas na quarta e última rodada do projeto fatorial para a estratégia de cruzamento *One Point Crossover*. As configurações marcadas em verde foram identificadas como as mais eficazes para essa rodada, considerando não só o *hypervolume*, mas também o tempo gasto para execução.

Tabela 13 – Configuração do 4º projeto fatorial para estratégia *One Point Crossover*.

k = 4	FATORES				FATORES				FIXO
	A (nº de gerações)		B (tam. população)		C (taxa de mutação)		D prob. de cruzamento		Iterações
NÍVEIS	1	150	1	40	1	0,01	1	0,96	10
	2	250	2	200	2	0,05	2	0,98	10

Fonte: Próprio autor.

De posse dos resultados (veja a Tabela 14), mesmo diminuindo o número de gerações, os valores de resposta do *hypervolume* não é reduzido. A mesma observação também vale para os fatores B (tamanho da população), C (taxa de mutação) e D (probabilidade de cruzamento, neste caso aumentando). Mas uma diferença notória, é o tempo de execução, já que houve uma redução do número de gerações e tamanho da população em comparação com o primeiro projeto fatorial (10).

Dessa forma, sem perder a qualidade da métrica *hypervolume*, os melhores parâmetros para *One Point Crossover* foram: A (2), B (2), C (1) e D (2) que se configuram em 250 gerações (A), população com 200 indivíduos (B), taxa de mutação com 1% (C) e probabilidade de cruzamento 98% (D).

Tabela 14 – Resultado do 4º projeto fatorial para estratégia *One Point Crossover*.

MELHORES E PIORES CONFIGURAÇÕES						
CONFIGURAÇÕES				MÉTRICAS		
A	B	C	D	Qualidade do Resultado (hypervolume)	C.V (%)	Tempo Gasto (ms)
2	2	1	2	31,2094	0,33	6702,70
2	2	2	2	31,2084	0,38	7141
1	1	1	2	30,2166	2,67	283
2	1	2	1	30,0218	1,83	447

Fonte: Próprio autor.

4.4.2.2 Two Points Crossover

Como ocorrido na seção 4.4.2.1, a estratégia *Two Points Crossover* também passou por uma mudança nos valores de alguns fatores, incluindo o número de gerações, o tamanho da população e a taxa de mutação, além do aumento na probabilidade de cruzamento. As configurações mais eficientes, considerando tanto o tempo quanto o *hypervolume*, estão destacadas em verde na Tabela 15.

Tabela 15 – Configuração do 4º projeto fatorial para estratégia *Two Points Crossover*.

k = 4	FATORES				FATORES				FIXO
	A		B		C		D		Iterações
	(nº de gerações)		(tam. população)		(taxa de mutação)		prob. de cruzamento		
NÍVEIS	1	150	1	40	1	0,01	1	0,96	10
	2	200	2	200	2	0,05	2	0,98	10

Fonte: Próprio autor.

Ao analisarmos a Tabela 16, observamos que a redução do tamanho da população, da taxa de mutação e do número de gerações não resultou em prejuízos para a métrica de *hypervolume*. Pelo contrário, a diminuição desses fatores contribuiu para a redução do tempo de execução do algoritmo.

Dessa forma, a configuração que apresentou melhores valores são: 150. Dessa forma, os parâmetros apresentaram melhores valores foram: 150 gerações, 200 indivíduos formando a população, 5% de mutação e 98% de probabilidade de cruzamento.

Tabela 16 – Resultado do 4º projeto fatorial para estratégia *Two Points Crossover*

MELHORES E PIORES CONFIGURAÇÕES						
CONFIGURAÇÕES				MÉTRICAS		
A	B	C	D	Qualidade do Resultado (hypervolume)	C.V (%)	Tempo Gasto (ms)
2	2	2	2	31,1997	0,35	7269
1	2	2	2	31,1575	0,63	4031
1	1	1	1	30,3307	1,98	277
1	1	1	2	30,3146	1,45	280

Fonte: Próprio autor.

4.4.2.3 Uniform Crossover

Na última estratégia de cruzamento analisada, mantivemos a mesma configuração de fatores das seções apresentadas anteriormente (4.4.2.1 e 4.4.2.2). Isso porque as três estratégias seguiram apresentando resultados semelhantes para fatores semelhantes. Os melhores parâmetros para essa estratégia estão destacados em verde na Tabela 17.

Na Tabela 18, são apresentados os dois melhores e piores resultados gerados do projeto fatorial para a estratégia *Uniform Crossover*. Em comparação com o primeiro

Tabela 17 – Configuração do 4º projeto fatorial para estratégia *Uniform Crossover*.

k = 4	FATORES				FATORES				FIXO
	A (nº de gerações)		B (tam. população)		C (taxa de mutação)		D prob. de cruzamento		Iterações
NÍVEIS	1	150	1	40	1	0,01	1	0,96	10
	2	250	2	200	2	0,05	2	0,98	10

Fonte: Próprio autor.

projeto fatorial 12, observamos que não houve perda no *hypervolume* ao reduzir o número de gerações, tamanho da população e taxa de mutação. Houve até uma leve melhoria na métrica do *hypervolume*, e o tempo de execução também foi reduzido. Os parâmetros que se destacaram, em termos de *hypervolume* e tempo de execução, foram 250 gerações (fator A), uma população composta por 200 indivíduos (fator B), 5% de taxa de mutação e 98% de probabilidade de cruzamento. É importante notar que, embora essa configuração tenha sido marcada como a melhor, o segundo melhor resultado não fica muito atrás quando comparado ao *hypervolume*, a diferença ocorre nas casas decimais.

Tabela 18 – Resultado do 4º projeto fatorial para estratégia *Uniform Crossover*.

MELHORES E PIORES CONFIGURAÇÕES							
CONFIGURAÇÕES				MÉTRICAS			
A	B	C	D	Qualidade do Resultado (hypervolume)	C.V (%)	Tempo Gasto (ms)	
2	2	2	2	31,2016	0,33	4021,29	
1	2	1	2	31,1776	0,40	7224,77	
1	1	1	1	29,7330	1,93	471	
2	1	2	1	29,6830	1,59	475	

Fonte: Próprio autor.

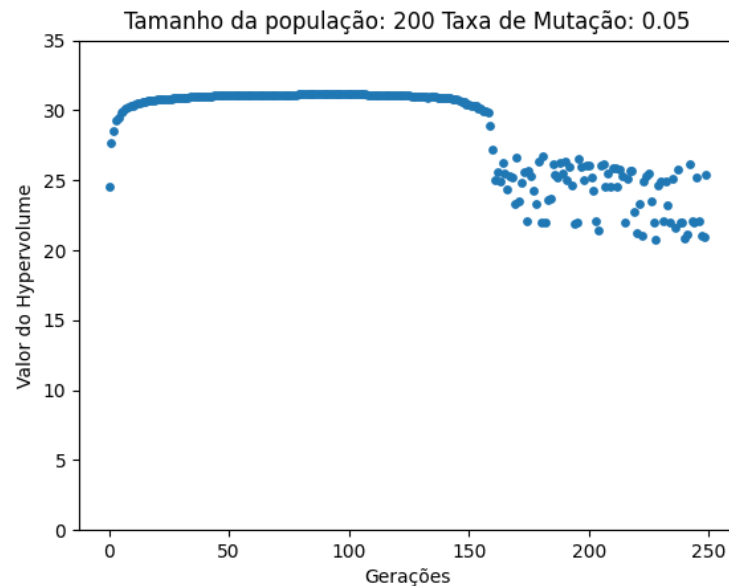
4.4.3 Definição dos parâmetros do NSGA-II

Embora tenham sido realizadas quatro rodadas do projeto fatorial para três estratégias de cruzamento implementadas (*One Point Crossover*, *Two Points Crossover* e *Uniform Crossover*), visando identificar qual delas proporcionaria a melhor resposta em termos de métricas como *hypervolume*, tempo e evolução, não foi possível determinar uma estratégia mais adequada com base nos resultados dos projetos fatoriais. Isso se deve ao fato de que todas as estratégias implementadas apresentaram valores muito próximos entre si.

A escolha da abordagem de cruzamento foi realizada por meio de uma análise do problema em questão. Nesse contexto, o método de recombinação escolhido foi o *Uniform Crossover*. Essa estratégia consegue promover uma maior variabilidade genética para os descendentes, uma vez que cada gene tem a oportunidade de contribuir para a formação de uma solução na próxima geração.

Na Figura 6, é apresentada a evolução do *hypervolume* ao longo das gerações, utilizando os parâmetros da quarta rodada do projeto fatorial com o *Uniform Crossover*, discutido na seção 4.4.2.3. Observa-se que não há mais evolução após 150 gerações; ao contrário, o algoritmo começa a piorar o *hypervolume* (involuir). Portanto, manter esse número de gerações seria não apenas um desperdício de recursos como um contrassenso.

Figura 6 – Estabilização do *hypervolume* com os parâmetros do projeto fatorial.

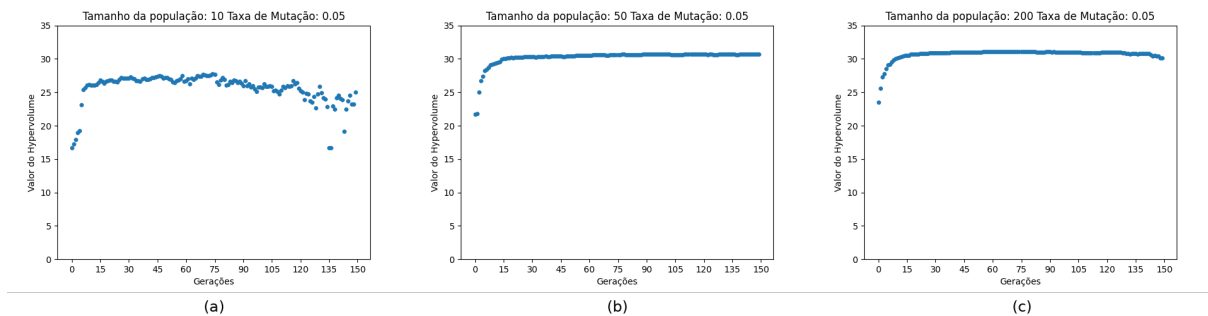


Fonte: Próprio autor.

O mesmo raciocínio sobre o desperdício de recursos se estende ao tamanho da população. Como mostrado na Figura 7, percebe-se que o tamanho da população não precisa necessariamente ser alto, como na Figura 7 (c) com 200 indivíduos, e um valor muito baixo, como na Figura 7 (a) com 10 indivíduos, também não é ideal. Os resultados obtidos para 50 e 200 indivíduos foram bastante próximos em relação ao melhor *hypervolume* alcançado, indicando que podemos reduzir o tamanho da população para economizar recursos de processamento. Isso é importante, uma vez que o processamento envolve não apenas o cálculo de avaliação para cada indivíduo, mas também a alocação na fronteira e outros passos que consomem recursos computacionais.

Portanto, com base nas várias rodadas do projeto fatorial e nos resultados obtidos em relação ao *hypervolume* e tempo de execução de outros testes, observamos que a combinação de parâmetros ideal para a estratégia *Uniform Crossover* é a seguinte:

- Número de gerações: 150
- Tamanho da população: 50
- Taxa de mutação: 5%

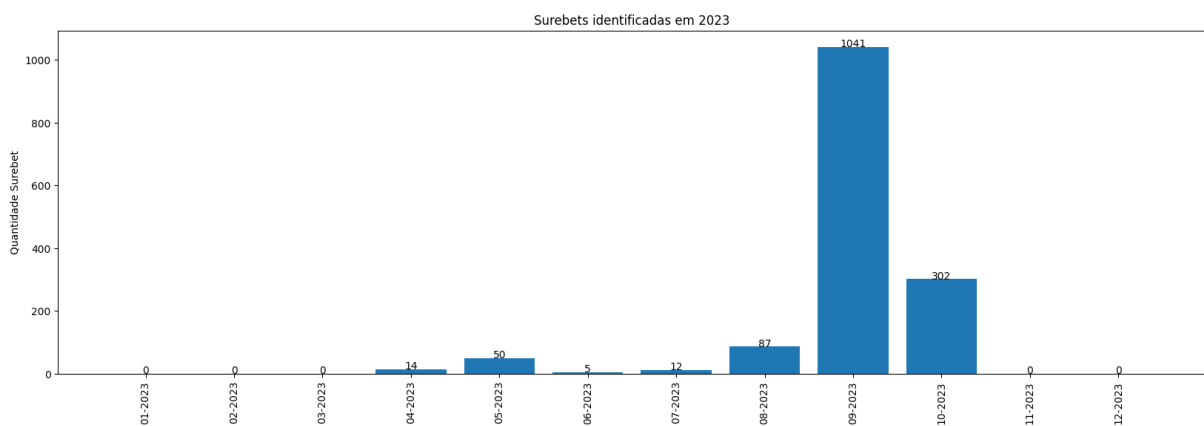
Figura 7 – Estabilização do *hypervolume* para diferentes tamanhos de população.

Fonte: Próprio autor.

- Probabilidade de cruzamento: 98%

4.5 Definição dos cenários para realização dos experimentos

Após calibrar os parâmetros do NSGA-II, para verificar o desempenho dessa heurística na otimização de soluções, definimos os cenários nos quais os experimentos seriam conduzidos. A Figura 8 ilustra a quantidade de *surebets* encontradas ao longo de cada mês em 2023, começando em janeiro e indo até outubro/2023, intervalo onde o *dataset* foi coletado.

Figura 8 – Gráfico com as *surebets* encontradas por mês.

Fonte: Próprio autor.

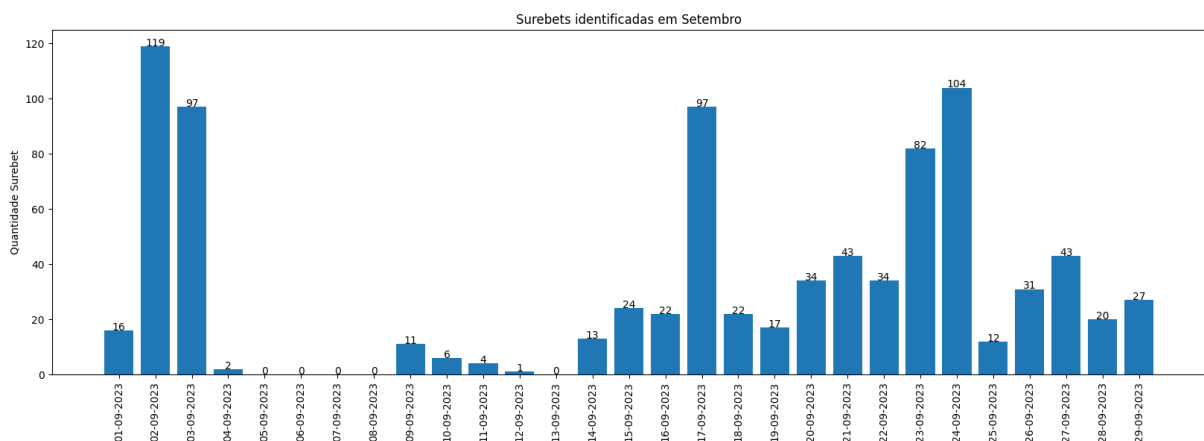
É evidente que vários meses tiveram um número reduzido de *surebets* identificadas. Isso ocorreu, em parte, devido à frequência inicial da coleta de dados, que se limitava a apenas dois horários para cada partida esportiva. Posteriormente, essa frequência foi aumentada. Para garantir que houvesse dados suficientes para a análise, optamos por focar em dois meses do *dataset*: setembro e outubro, por serem os meses que mais apresentaram oportunidades de arbitragem (mais *surebets*).

Um aspecto importante com relação à quantidade de possibilidade de arbitragens para cada mês é sobreposição desses eventos. Por exemplo, uma oportunidade de arbitragem poderia estar disponível às 14:00 e persistir até às 14:15, o que resultaria na mesma *surebet* para um jogo (mas com *odds* diferentes), sendo registrada em ambas as instâncias na base de dados. Para ilustrar melhor frequência de ocorrência das *surebets*, o gráfico exibido na [Figura 8](#) não exibe uma *surebet* que permaneceu aberta durante o tempo da próxima coleta, isto é, foi desconsiderado quantas janelas de tempo que ela ficou aberta. Uma análise que inclua as oportunidades mantidas em aberto será apresentada na seção [5.1](#), de caracterização do *dataset*.

4.5.1 Setembro

Para o mês de setembro, foram escolhidas duas datas específicas: 02/09/2023 e 24/09/2023. Essas datas foram selecionadas devido à quantidade de oportunidades de arbitragem identificadas. Na primeira data, registramos aproximadamente 119 *surebets*, enquanto na segunda, contabilizamos 104 oportunidades (conforme ilustrado na [Figura 9](#)).

Figura 9 – Gráfico com as *surebets* encontradas para o mês de setembro.



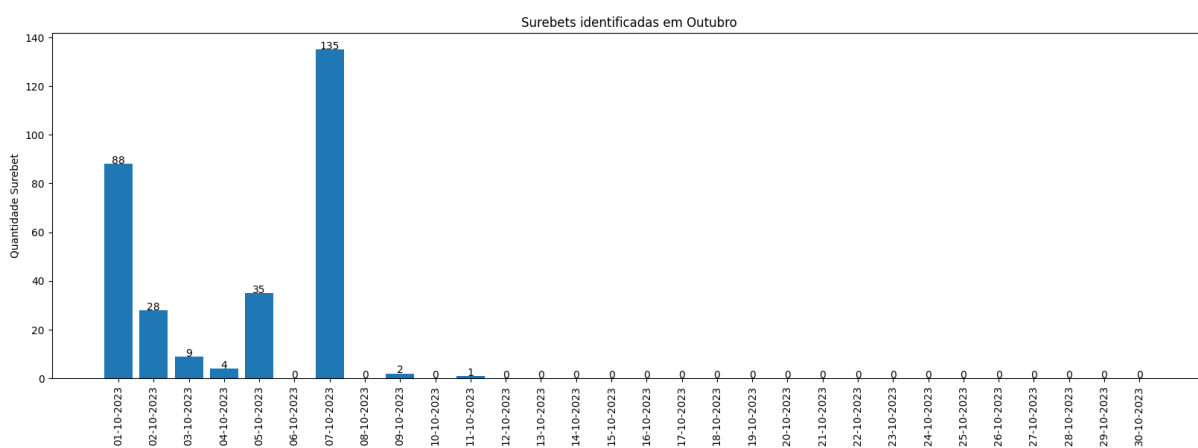
Fonte: Próprio autor.

4.5.2 Outubro

Da mesma forma, para o mês de outubro, optamos por realizar experimentos em duas datas propícias: 01/10/2023 e 07/10/2023. Essas datas foram escolhidas devido ao número de *surebets* identificadas. No dia 01/10/2023, registramos aproximadamente 80 oportunidades de arbitragem, enquanto no dia 07/10/2023, o número aumentou para cerca de 135 *surebets* (veja a [Figura 10](#)).

É relevante observar que, após o dia 07/10/2023, não houve registro de *surebets* devido a uma limitação de acesso à API decorrente do esgotamento dos créditos disponíveis.

Figura 10 – Gráfico com as *surebets* encontradas para o mês de outubro.



Fonte: Próprio autor.

5 Resultados e Análise

Nesta seção, apresentaremos a caracterização do conjunto de dados montado durante o período de coleta, logo depois serão apresentados os resultados do experimento conduzido na data 02/09/2023, de forma mais detalhada e por fim, será feita uma comparação dos resultados obtidos dentro de 4 datas, sendo duas em setembro (02/09/2023 e 24/09/2023) e duas em outubro (01/10/2023 e 07/10/2023) do ano de 2023.

Para o cenário de data 02/09/2023, descreveremos três tipos distintos de experimentos realizados. O primeiro experimento consiste na escolha de um intervalo curto de 5 minutos, numa janela de 1 hora (foi escolhido a faixa que contava com o maior número de *surebets*). Analisaremos a evolução da fronteira de Pareto e realizaremos a seleção de três indivíduos da fronteira de Pareto: um otimizado para maximizar o lucro, outro focado em maximizar a quantidade de *surebets* escolhidas com o intuito de diluir melhor o investimento, e o último que busca um equilíbrio entre essas duas otimizações.

O segundo experimento aborda a variação do lucro entre dois intervalos de tempo, enquanto o terceiro destaca o lucro acumulado ao longo de um período predefinido.

É importante relatar que todas as soluções apresentadas, foram retiradas da fronteira de Pareto, com o intuito de utilizar os melhores resultados identificados pelo algoritmo.

5.1 Caracterização do *dataset*

A elaboração da base de dados obtida ocorreu por meio da API, conforme detalhado na seção 3.2.1. Os dados adquiridos estão relacionados ao futebol, dada a maior familiaridade com esse esporte, o que facilitou a compreensão dos mercados oferecidos pelas casas de apostas.

A coleta ocorreu de janeiro até dia 11 de outubro sendo contabilizados 2446 jogos. Inicialmente estava dando uma preferência por alguns campeonatos (Ligue 1 (França), Bundesliga (Alemanha), serie A (Itália), La Liga (Espanha)), mas depois de um período essa restrição foi descartada, uma vez que ficou entendido que o foco era a identificação de *surebets* para servir de base para os experimentos.

Cada jogo representado pela API continha informações sobre a partida, casas de aposta e os mercados cobertos por cada casa de aposta. Cada mercado, por sua vez, apresentava valores a serem pagos (odds), também conhecidos como resultados (ou *outcomes*), e outras informações complementares. A estrutura desse processo pode ser visualizada na [Figura 11](#).

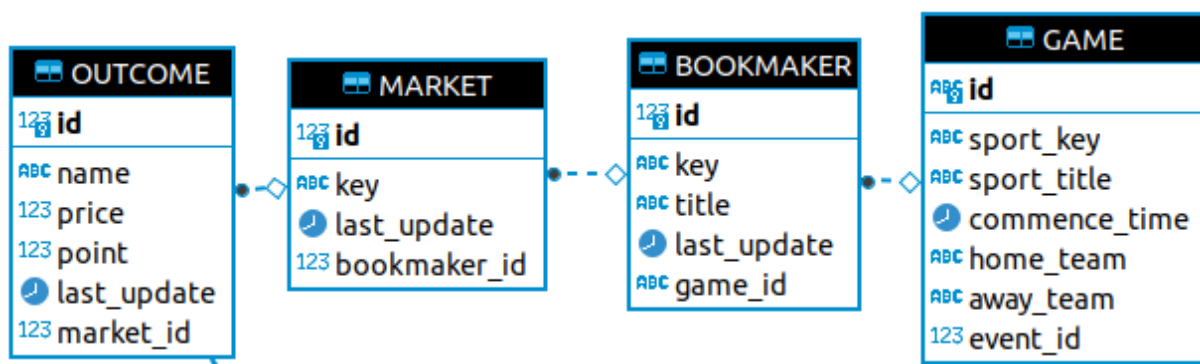


Figura 11 – Estrutura dos dados coletados da API.

Para otimizar o uso dos créditos fornecidos pela API, apenas dois mercados foram incluídos na base. O primeiro é o mercado h2h (head to head), que oferece três opções de aposta: no time da casa, no empate ou no time visitante. O segundo mercado coletado é o *totals*, que possui duas opções de aposta, como, por exemplo, apostar se haverá mais de 2,5 gols ou menos de 2,5 gols na partida.

Conforme explicado na seção 2, uma *surebet* deve envolver o mesmo evento/mercado e apresentar probabilidades que permitam cobrir todos os resultados possíveis. Nesse contexto, o mercado escolhido para avaliar as possibilidades de *surebet* foi o *totals*, que apresenta apenas duas opções de entrada.

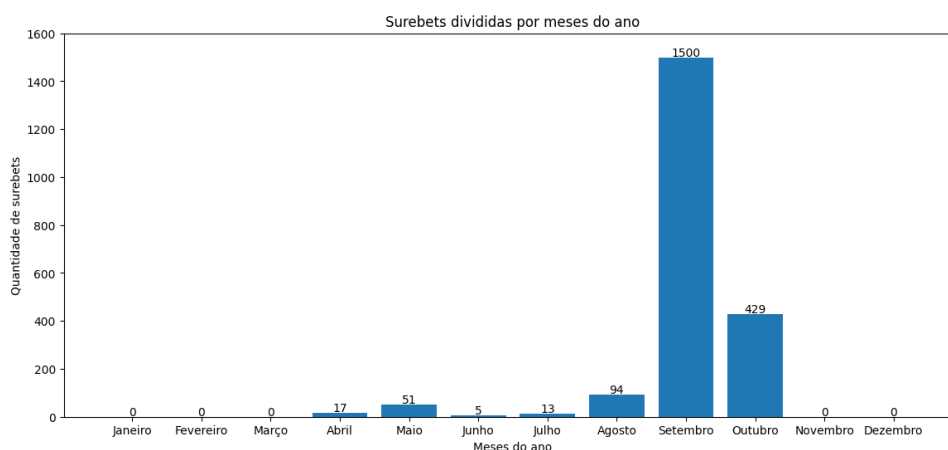


Figura 12 – Divisão das *surebets* encontradas em mês.

A Figura 12 ilustra a quantidade de *surebets* encontradas no mercado *totals*, divididas pelos meses do ano. Nessa análise, são consideradas também as arbitragens que permaneceram abertas ao realizar uma nova consulta na API. Por exemplo, se uma requisição foi feita às 14:00 e, ao realizar outra às 14:30, a *surebet* ainda existia, ela foi contabilizada no gráfico. Observa-se que os meses de setembro e outubro concentram a maior quantidade de *surebets*, sendo uma possível explicação a remoção da restrição de acompanhar campeonatos específicos.

Já a Figura 13 e 14 proporciona outra perspectiva da distribuição das arbitragens

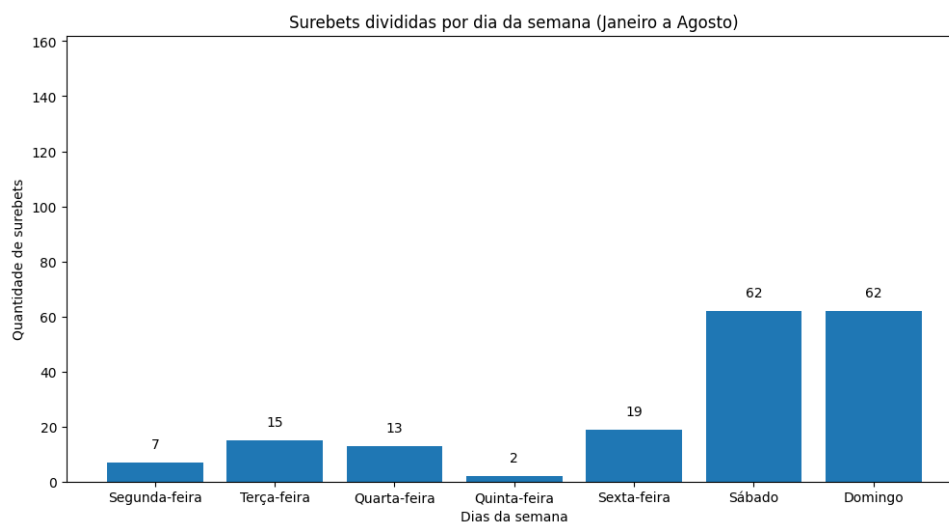


Figura 13 – *Surebets* encontradas por dia da semana, de janeiro até agosto/2023

encontradas. Nota-se um número moderado de *surebets* de segunda a sexta-feira, mas esse valor aumenta significativamente nos sábados e domingos. Uma análise rápida sugere que o final de semana pode ser mais propício para encontrar *surebets* e realizar apostas, devido a esse aumento significativo.

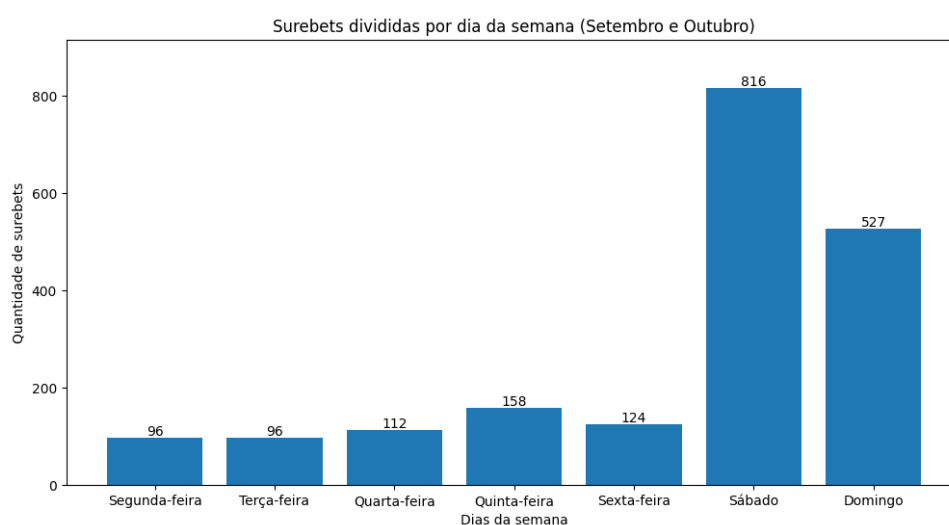


Figura 14 – *Surebets* encontradas por dia da semana, em setembro e outubro/2023

Devido ao aumento do número de coletas em setembro e outubro, que resultou em mais *surebets* nesse período, optou-se por apresentar a análise semanal separadamente. Isso foi feito para garantir que os meses com maior número de *surebets* não prejudicassem a visualização e análise dos meses com menor ocorrência desse tipo de oportunidade. E como podemos ver, o mesmo padrão se aplica nos dois, isto é, mais *surebets* aos sábados e domingos.

Na [Figura 15](#), é apresentada a distribuição de *surebets* encontradas por cada casa de apostas. Observa-se uma discrepância nos números entre algumas delas. O artigo de [Franck](#),

Verbeek e Nüesch (2013) sugere a existência de uma rivalidade competitiva entre as casas de apostas. Para atrair apostadores, essas empresas podem adotar diversas estratégias, como oferecer bônus e ajustar suas probabilidades para proporcionar ganhos mais atrativos. A desregulação das *odds* pode ser um dos fatores que propiciam o surgimento de *surebets*. Possivelmente, essa relação está refletida nas casas de apostas que apresentam mais *surebets* na figura analisada.

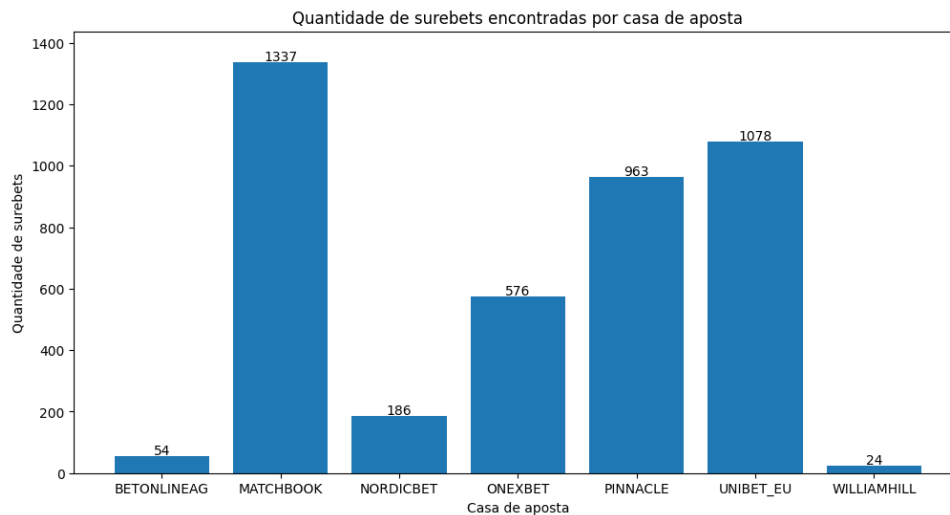


Figura 15 – Divisão de *surebets* encontradas por casa de aposta

5.2 Experimento para data 02/09/2023

A Figura 16 ilustra a distribuição das *surebets* ao longo de um intervalo de 1 hora para o dia 02/09. Optamos por realizar os experimentos no intervalo das 14h às 15h devido à maior concentração de *surebets* nesse período.

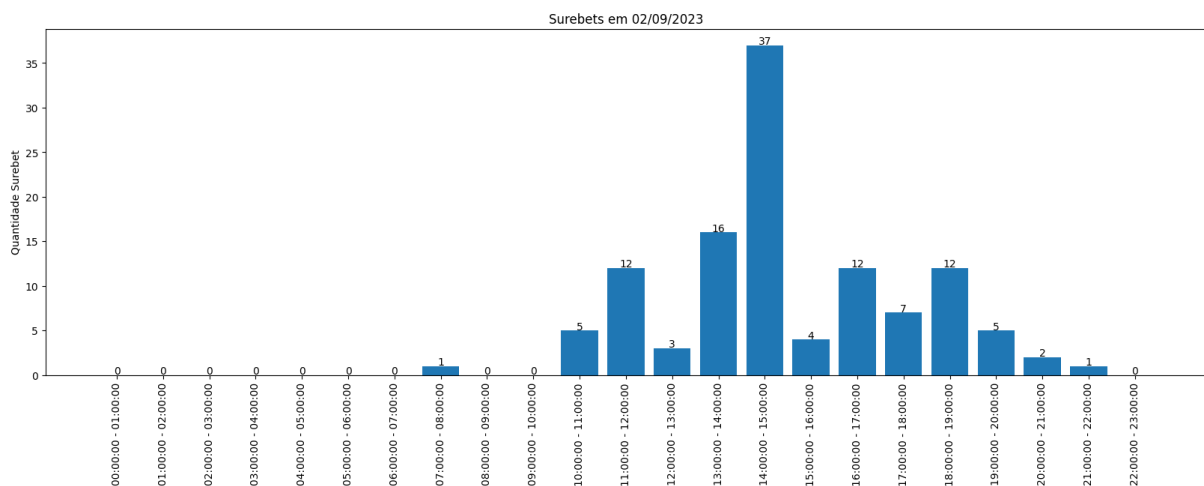


Figura 16 – Distribuição das *surebets* encontradas em intervalos de 1h

Explorando o intervalo específico de 14h às 15h, a Figura 17 representa a distribuição das *surebets* ao longo de um período de 5 minutos. Para o primeiro experimento, focaremos

nossa análise no intervalo de 14:20 às 14:25. É importante definir essa faixa para tornar os experimentos o mais próximo possível da realidade, considerando que os dados fornecidos pela API (2023) são atualizados a cada 5 minutos. Se a faixa escolhida fosse, por exemplo, de 14:00h às 15:00h, não seria possível garantir que uma arbitragem identificada às 14:01 permanecesse aberta até as 14:55h, o que comprometeria a fidelidade do experimento em relação à dinâmica real do mercado.

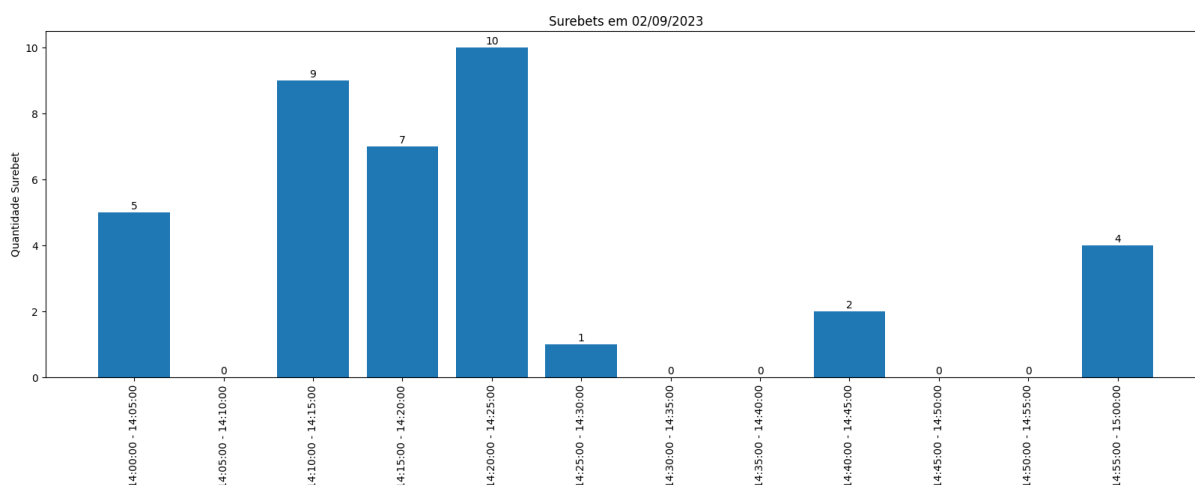


Figura 17 – Surebets encontradas a cada slot de 5 min., de 14h às 15h

A lista de *surebets*, consideradas genes do cromossomo, identificadas na faixa de tempo de 14:20 às 14:25, é apresentada na Tabela 19. Essas opções serão a base para condução do primeiro experimento. Ao todo foram identificadas 10 arbitragens.

Tabela 19 – Lista de *surebets* para o intervalo de 14:20 às 14:25

Índice	Time	Casa de aposta 1	Casa de aposta 2	odds acima x gols	odds abaixo x gols	Lucro
0	Ipswich Town x Cardiff City [Championship]	MATCHBOOK	PINNACLE	2.44	1.93	7.202922
1	Doncaster Rovers x Swindon Town [League 2]	MATCHBOOK	PINNACLE	2.36	1.91	5.271098
2	Ipswich Town x Cardiff City [Championship]	UNIBET_EU	PINNACLE	2.32	1.93	5.083080
3	Manchester City x Fulham [EPL]	UNIBET_EU	MATCHBOOK	1.85	2.18	0.074386
4	Bristol Rovers x Lincoln City [League 1]	MATCHBOOK	UNIBET_EU	1.66	2.55	0.543350
5	Mansfield Town x Bradford City [League 2]	MATCHBOOK	PINNACLE	2.34	1.90	4.633378
6	Doncaster Rovers x Swindon Town [League 2]	UNIBET_EU	PINNACLE	2.28	1.91	3.784330
7	Manchester City x Fulham [EPL]	MATCHBOOK	UNIBET_EU	1.94	2.14	1.724636
8	Mansfield Town x Bradford City [League 2]	UNIBET_EU	PINNACLE	2.23	1.90	2.525372
9	Chelsea x Nottingham Forest [EPL]	UNIBET_EU	MATCHBOOK	2.38	1.76	1.165011

5.2.1 Maximização do lucro

Essa solução tem o objetivo de maximizar o lucro. No Quadro 1, são apresentados os resultados para esse propósito.

O Quadro 1 mostra que o algoritmo fez a escolha de selecionar apenas 1 *surebet* em seu conjunto de soluções, investindo a quantia de R\$ 100,00, conforme previamente definido como a banca. A seleção feita se trata da arbitragem que prometia o maior retorno

financeiro, conforme detalhado na [Tabela 19](#), o que acabou gerando um lucro de R\$ 7,20. Essa otimização, portanto, representa o benefício financeiro máximo alcançável no intervalo de tempo estipulado e está conforme o propósito do objetivo que era de maximizar o lucro.

Quadro 1 – Solução para maximização do lucro

***** RELATÓRIO DO EXPERIMENTO *****

[Data]: 02/09/2023

[Intervalo das surebets]: 14:20 to 14:25

[Geração da Fronteira de Pareto]: 109^o geração

[Banca]: R\$ 100,00

Otimização do Lucro

Número de surebets escolhidas: 1

Surebets selecionadas:

('Partida: Ipswich Town x Cardiff City', 'Casa 1: MAT', 'Casa 2: PIN', 'Valor apostado R\$ 100.0', 'Lucro R\$ 7.20', '7.20%')

Lucro total: R\$ 7.20

5.2.2 Maximização de *surebets*

O propósito desta abordagem vai além de simplesmente escolher o maior número de *surebets* para compor o conjunto de soluções. A proposta é selecionar não só a maior quantidade de *surebets* possíveis, mas também alocar os recursos entre elas de maneira estratégica. Isso visa reduzir o risco¹ associado ao investir todo o montante em uma única *surebet*. Obviamente o lucro cairá utilizando essa estratégia, visto a seguir, mas o risco associado também diminui, conforme discutido na seção [4.3.5](#).

Para atingir esse objetivo, o algoritmo seguiu uma abordagem oposta à maximização do lucro. O resultado exibido no [Quadro 2](#) revela que o algoritmo selecionou o máximo de *surebets* disponíveis no intervalo (10 arbitragens), conforme apresentado na [Tabela 19](#). Além disso, a alocação de recursos, utilizando o cálculo da entropia (como visto em carteiras de investimento), foi estratégica para garantir uma distribuição eficiente e diminuir o risco associado à estratégia de *surebet*.

5.2.3 Equilíbrio entre os dois objetivos

Esta solução visa alcançar um equilíbrio entre as duas otimizações, traduzindo-se em uma solução que possui um lucro significativo, mas que também valoriza a maximização da quantidade de *surebets* e a distribuição eficiente de recursos.

Como demonstrado no [Quadro 3](#), foram selecionadas 7 arbitragens dentre as 10 apresentadas na [Tabela 19](#). Esse número, naturalmente, superou a quantidade selecionada

¹ O risco no contexto de *surebet* está relacionado ao fato de: ao dar entrada nas duas casas de aposta, uma poderia cancelar, deixando a outra em aberto, quando isso ocorre o resultado cuja operação foi finalizada poderia ser a vencedora, mas como foi cancelada, o apostador perderá o dinheiro investido no outro evento; outro risco seria de ser pego pelas casas de aposta e começar a ter a conta limitada ou bloqueada.

 Quadro 2 – Solução para maximização da quantidade de *surebets* e alocação do recurso.

***** RELATÓRIO DO EXPERIMENTO *****

[Data]: 02/09/2023

[Intervalo das surebets]: 14:20 to 14:25

[Geração da Fronteira de Pareto]: 109^o geração

[Banca]: R\$ 100,00

Otimização da quantidade de surebets

Número de surebets escolhidas: 10

Surebets selecionadas:

('Partida: Ipswich Town x Cardiff City', 'Casa 1: MAT', 'Casa 2: PIN', 'Valor apostado R\$ 9.0', 'Lucro R\$ 0.65', '7.20%')

('Partida: Manchester City x Fulham', 'Casa 1: UNI', 'Casa 2: MAT', 'Valor apostado R\$ 12.0', 'Lucro R\$ 0.01', '0.07%')

('Partida: Bristol Rovers x Lincoln City', 'Casa 1: MAT', 'Casa 2: UNI', 'Valor apostado R\$ 12.0', 'Lucro R\$ 0.07', '0.54%')

('Partida: Doncaster Rovers x Swindon Town', 'Casa 1: MAT', 'Casa 2: PIN', 'Valor apostado R\$ 10.0', 'Lucro R\$ 0.53', '5.27%')

('Partida: Mansfield Town x Bradford City', 'Casa 1: MAT', 'Casa 2: PIN', 'Valor apostado R\$ 12.0', 'Lucro R\$ 0.56', '4.63%')

('Partida: Manchester City x Fulham', 'Casa 1: MAT', 'Casa 2: UNI', 'Valor apostado R\$ 8.0', 'Lucro R\$ 0.14', '1.72%')

('Partida: Mansfield Town x Bradford City', 'Casa 1: UNI', 'Casa 2: PIN', 'Valor apostado R\$ 12.0', 'Lucro R\$ 0.30', '2.53%')

('Partida: Chelsea x Nottingham Forest', 'Casa 1: UNI', 'Casa 2: MAT', 'Valor apostado R\$ 7.0', 'Lucro R\$ 0.08', '1.17%')

('Partida: Doncaster Rovers x Swindon Town', 'Casa 1: UNI', 'Casa 2: PIN', 'Valor apostado R\$ 9.0', 'Lucro R\$ 0.34', '3.78%')

('Partida: Ipswich Town x Cardiff City', 'Casa 1: UNI', 'Casa 2: PIN', 'Valor apostado R\$ 13.0', 'Lucro R\$ 0.66', '5.08%')

Lucro total: R\$ 3.34

no objetivo de maximizar o lucro. Podemos estabelecer o mesmo paralelo frente ao outro objetivo que era de maximização de *surebets*, o lucro atingindo havia sido R\$ 3,34 contra R\$ 5,96 (ambos investimentos R\$ 100,00) garantido por essa solução que busca o equilíbrio entre os dois objetivos.

 Quadro 3 – Solução de equilíbrio entre os dois objetivos.

***** RELATÓRIO DO EXPERIMENTO *****

[Data]: 02/09/2023

[Intervalo das surebets]: 14:20 to 14:25

[Geração da Fronteira de Pareto]: 109^o geração

[Banca]: R\$ 100,00

Otimização da quantidade de surebets

Número de surebets escolhidas: 7

Surebets selecionadas:

('Partida: Ipswich Town x Cardiff City', 'Casa 1: MAT', 'Casa 2: PIN', 'Valor apostado R\$ 56.0', 'Lucro R\$ 4.03', '7.20%')

('Partida: Doncaster Rovers x Swindon Town', 'Casa 1: MAT', 'Casa 2: PIN', 'Valor apostado R\$ 10.0', 'Lucro R\$ 0.53', '5.27%')

('Partida: Mansfield Town x Bradford City', 'Casa 1: MAT', 'Casa 2: PIN', 'Valor apostado R\$ 10.0', 'Lucro R\$ 0.46', '4.63%')

('Partida: Manchester City x Fulham', 'Casa 1: MAT', 'Casa 2: UNI', 'Valor apostado R\$ 4.0', 'Lucro R\$ 0.07', '1.72%')

('Partida: Mansfield Town x Bradford City', 'Casa 1: UNI', 'Casa 2: PIN', 'Valor apostado R\$ 4.0', 'Lucro R\$ 0.10', '2.53%')

('Partida: Doncaster Rovers x Swindon Town', 'Casa 1: UNI', 'Casa 2: PIN', 'Valor apostado R\$ 7.0', 'Lucro R\$ 0.26', '3.78%')

('Partida: Ipswich Town x Cardiff City', 'Casa 1: UNI', 'Casa 2: PIN', 'Valor apostado R\$ 10.0', 'Lucro R\$ 0.51', '5.08%')

Lucro total: R\$ 5.96

5.2.4 Evolução da fronteira de Pareto

A [Figura 18](#) apresenta a evolução da fronteira de Pareto, com destaque para três fronteiras selecionadas, ao longo das gerações. O armazenamento das melhores fronteiras se deu guardando a próxima fronteira que apresentasse um valor de *hypervolume* (Hv) superior à anterior. No eixo das abscissas, observamos o lucro, enquanto no eixo das ordenadas, é representado o resultado obtido no cálculo de entropia, relacionado à distribuição de recursos entre as *surebets*. Cada ponto na fronteira representa uma solução com um

conjunto específico de *surebets* escolhidas e a alocação correspondente do capital disponível para investimento. Vale lembrar que os dois objetivos são de maximização, logo quanto maior o valor do *hypervolume* (indicado como Hv na legenda da Figura 18), melhor a fronteira.

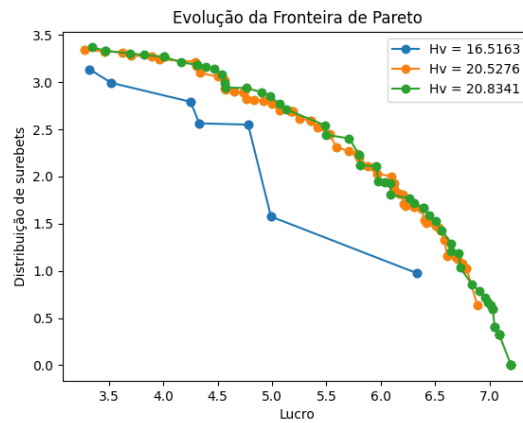


Figura 18 – Evolução da fronteira de Pareto para o intervalo das 14:20h às 14:25h.

5.2.5 Variação do lucro no intervalo das 14:00h às 15:00h

O objetivo principal deste experimento é evidenciar as variações no lucro ao longo da faixa de intervalo estabelecida. A alocação de R\$ 100,00 para investimento em cada faixa de 5 minutos foi considerada a banca inicial, e o valor no eixo y representa o lucro obtido com a banca definida.

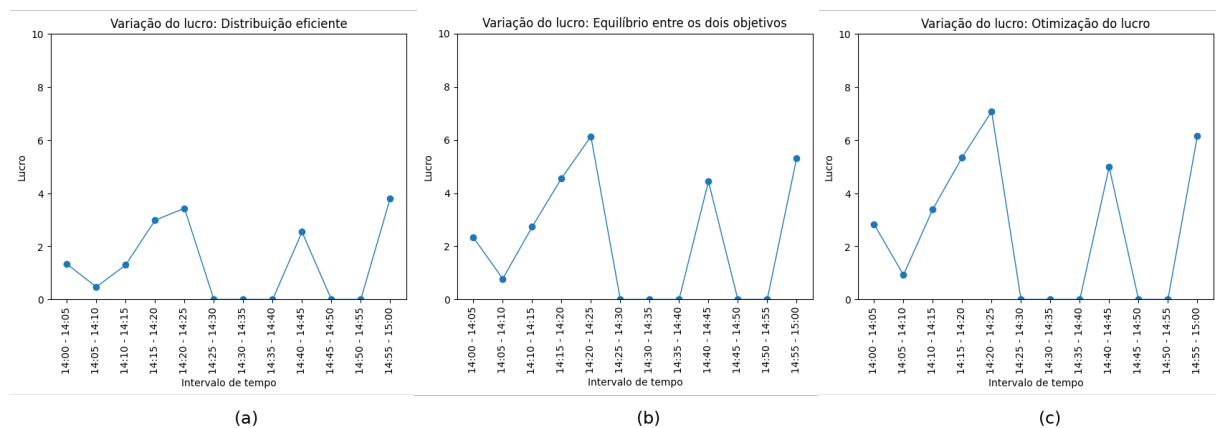


Figura 19 – Variação do lucro no intervalo das 14:00 às 15:00 para três finalidades (Maximização da quantidade de *surebets*, Equilíbrio entre os objetivos, Maximização do lucro).

As diferentes estratégias exploradas foram: Figura 19 (a) maximizará a quantidade de apostas, Figura 19 (b) procura um equilíbrio entre a maximização da quantidade de apostas e a maximização do lucro, enquanto Figura 19 (c) concentra-se na maximização do lucro propriamente dito. Observe que a Figura 19 exibe as oscilações no lucro obtido,

oferecendo uma representação visual da variação de desempenho ao longo do tempo. Para as faixas de horário que possuem lucro 0, é porque não houve oportunidade de *surebet* para avaliação. Ainda, sim, conforme esperado, devido aos objetivos específicos de cada abordagem, temos, em ordem decrescente, os maiores pico de lucro, (Figura 19c), (Figura 19b) e (Figura 19a).

Essa análise temporal poderia revelar não apenas padrões ou tendências, como os melhores horários para apostar com maior lucro, mas também como eventos específicos podem impactar diretamente o lucro obtido com *surebets*. Eventos esportivos de grande relevância ou até mesmo mudanças climáticas são exemplos que podem justificar alterações na quantidade de *surebet* encontradas ou mesmo no seu valor de retorno.

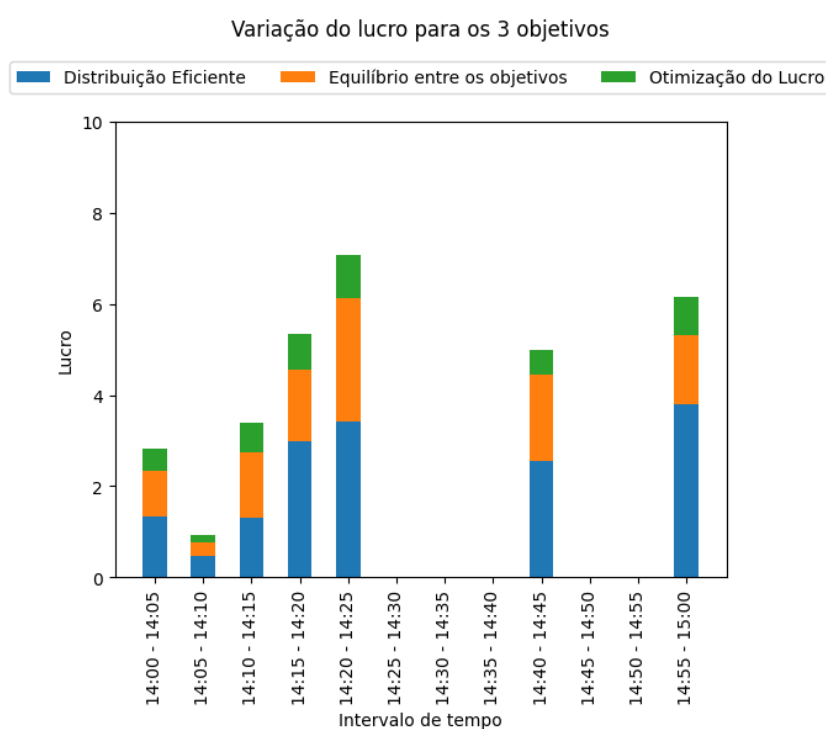


Figura 20 – Variação do lucro com os 3 objetivos sobrepostos.

Na Figura 20, apresentamos o lucro obtido por cada estratégia de maximização durante o intervalo das 14h às 15h. Cada segmento de 5 minutos representa o lucro alcançado, dentro dessa faixa definida. Quando o gráfico não possui barras, indica que não havia oportunidade de *surebet* disponível para investimento nesse período, resultando na ausência de lucro, como visto nas faixas 14:25 às 14:30, 14:30 às 14:35 e 14:35 às 14:40.

5.2.6 Lucro acumulado no intervalo das 14:00h às 15:00h

O propósito desta análise é compreender não apenas a rapidez com que cada objetivo de otimização aumenta o montante, mas também avaliar a expressividade do mercado de apostas esportivas *online*. Nesse experimento, foi considerado a existência de um valor fixo a ser investido em cada intervalo, com uma banca de R\$ 100,00. Dessa forma,

para cada faixa de 5 minutos, simulou-se o investimento total da banca e foi acumulando o lucro. É importante ressaltar que o lucro não foi alocado nas entradas posteriores, foi deixado fixo o valor da banca.

A Figura 21 mostra o acúmulo do lucro para as três estratégias. A estratégia representada por (a) utiliza a maximização da quantidade de *surebets* selecionadas; a segunda (b) busca o equilíbrio entre essa estratégia e a maximização do lucro, e a última (c) utiliza a estratégia de maximização do lucro. É possível observar que, para alguns horários, o lucro acumulado se estabiliza, indicando a ausência de *surebets* para avaliação.

Como era de se esperar, o lucro acumulado ao longo do tempo foi maior na última (Figura 21c), que tinha exatamente esse objetivo. Ainda assim, a Figura 21b), que busca o equilíbrio entre essa otimização e a distribuição do recurso, também apresentou um crescimento interessante. Por último, mesmo com o menor lucro acumulado, a Figura 21a) demonstrou ser possível obter lucro, embora com um crescimento mais lento.

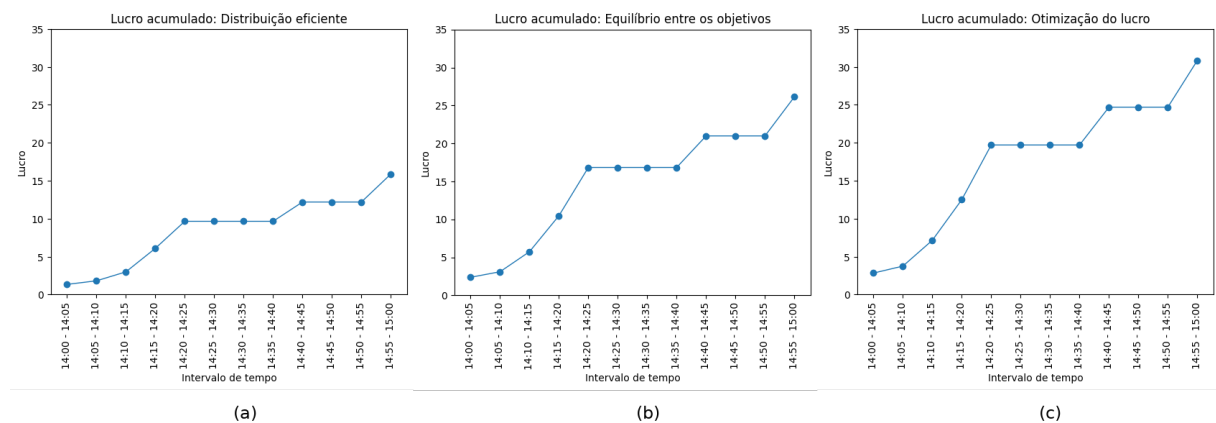


Figura 21 – Lucro acumulado no intervalo de 14h–15h conforme cada objetivo.

2

Para uma compreensão mais clara da magnitude do lucro acumulado entre os três objetivos, veja a Figura 22, onde os resultados dos três objetivos estão sobrepostos. À primeira vista, a estratégia que maximiza o lucro (curva representada pela cor azul) pode parecer a escolha mais vantajosa e óbvia, devido ao seu maior lucro. No entanto, considerando que as casas de apostas conseguem identificar o uso dessa estratégia e aplicar penalizações, como o cancelamento da aposta, a limitação diária do valor que o apostador pode investir ou até mesmo o bloqueio da conta, essa estratégia pode deixar de ser interessante.

A situação pode se agravar ainda mais quando apenas uma das casas de apostas cancela a entrada do apostador. Se isso ocorrer e a aposta restante permanecer aberta, o resultado vencedor pode ser aquele que teve a aposta cancelada, resultando na perda total do dinheiro investido. Por outro lado, se o apostador diversificar suas escolhas, optando por

² Maximização da quantidade de *surebets*, equilíbrio entre os objetivos e maximização do lucro

mais de uma *surebet*, o impacto do cancelamento de uma aposta seria menor, reduzindo assim o prejuízo potencial.

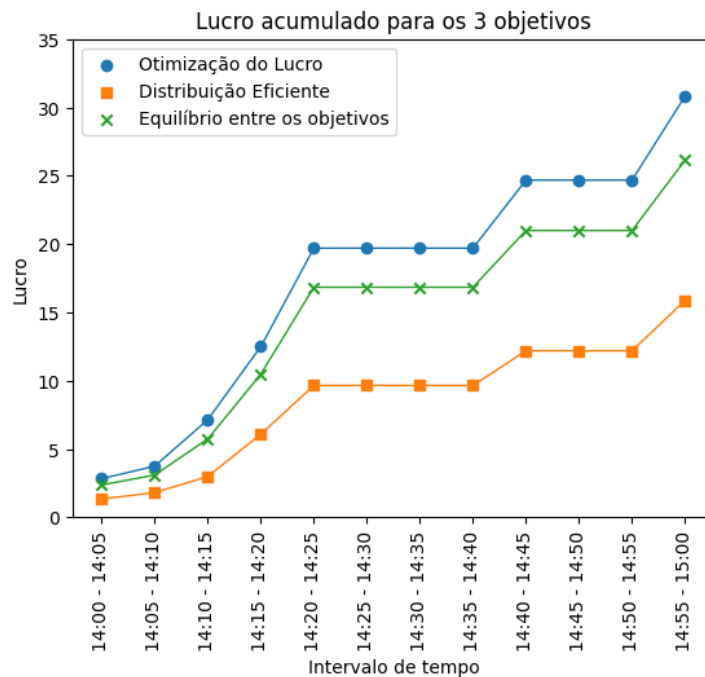


Figura 22 – Lucro acumulado com os 3 objetivos sobrepostos.

Esse experimento é relevante para termos uma ideia do ganho que seria obtido ao variar o objetivo do apostador. Se ele possui um perfil mais conservador, poderia optar por maximizar a distribuição do seu investimento. No entanto, se possuir um perfil arrojado e não se importar em arriscar (considerando limitações, cancelamento de apostas, bloqueio da conta, entre outros, conforme abordado na seção 4.3.5), poderia escolher otimizar o lucro e ter seu investimento acumulado, como representado na Figura 21c). Além disso, existe a opção de manter um equilíbrio entre os dois objetivos, como mostrado na Figura 21b).

5.3 Experimento comparativo

O propósito deste experimento é apresentar os resultados obtidos em diferentes datas, evidenciando o desempenho do algoritmo em diversos cenários. Para essa análise, foi elaborada a Tabela 20, composta por quatro colunas contendo as principais informações. Cada experimento (listado na primeira coluna) resultou em três soluções da fronteira de Pareto: uma visa maximizar o lucro, a segunda busca um equilíbrio entre os dois objetivos e a última distribuir os recursos de maneira eficiente. Cada estratégia possui a quantidade de *surebets* escolhidas e o lucro atingido.

De maneira semelhante à abordagem da seção anterior, o algoritmo mantém uma tendência para cada estratégia, mantendo o número de *surebets* escolhidas baixo para a maximização do lucro, um número mais elevado para o objetivo de distribuição

eficiente de recursos (com todas as *surebets* sendo selecionadas em todos os testes) e um equilíbrio intermediário ao buscar uma combinação dos dois objetivos. Observa-se que a proporção de ganhos se repete em outros cenários, dependendo do objetivo definido. Por exemplo, ao buscar a maximização do lucro, essa meta é alcançada, e não existe outra solução, encontrada pelo algoritmo, com um lucro superior. O mesmo princípio se aplica à distribuição eficiente de recursos.

A análise da tabela também revela que o lucro não está exclusivamente relacionado ao número de *surebets* disponíveis, mas sim ao percentual oferecido por elas. Isso é evidenciado no cenário do dia 01/10/2023, que possui o menor número de opções de *surebets*, mas resultou no maior lucro. Ainda no mesmo experimento, as estratégias de equilíbrio e distribuição eficiente consideraram o mesmo número de *surebets*, mas apresentaram lucros diferentes. Isso ocorreu porque a distribuição do dinheiro da banca foi mais eficaz na segunda estratégia, resultando em um lucro inferior.

Um aspecto importante a ser destacado, conforme estabelecido na seção 3.2.4, é que a banca utilizada para atingir esses valores foi de R\$ 100,00. Isso implica que os resultados podem ser interpretados em termos percentuais. Portanto, um lucro que inicialmente era de R\$ 15,09, poderia ser R\$ 150,90 se fossem investidos R\$ 1.000,00. Observe que, como o valor escolhido para a banca fictícia foi de R\$ 100,00, o valor do lucro auferido pode ser diretamente interpretado como representando também a porcentagem de lucro: 15,09%

Tabela 20 – Comparação entre as 4 datas definidas como cenários de experimento.

Quadro comparativo entre os cenários			
Cenário do experimento	Objetivo	Surebets selecionadas	Lucro (R\$)
02/09/2023 14:20 - 14:25 Surebets disponíveis: 10	Maximização do lucro	1	7,20
	Equilíbrio entre os dois objetivos	7	5,96
	Distribuição eficiente	10	3,34
24/09/2023 15:15 - 15:20 Surebets disponíveis: 11	Maximização do lucro	3	16,40
	Equilíbrio entre os dois objetivos	9	14,15
	Distribuição eficiente	11	11,21
01/10/2023 15:45 - 15:50 Surebets disponíveis: 4	Maximização do lucro	1	21,30
	Equilíbrio entre os dois objetivos	4	15,09
	Distribuição eficiente	4	6,01
07/10/2023 14:15 - 14:20 Surebets disponíveis: 10	Maximização do lucro	1	8,12
	Equilíbrio entre os dois objetivos	8	6,16
	Distribuição eficiente	10	3,38

6 Considerações Finais

Este trabalho teve como propósito o desenvolvimento de um protótipo de software capaz de identificar *surebets*, construir uma base de dados e utilizá-la em uma heurística. Além disso, visou apresentar um conjunto de soluções que contemplasse não apenas a otimização do lucro, mas também a realização de uma alocação de recursos eficiente.

O protótipo desenvolvido consegue realizar tarefas automatizadas, como (i) coletar dados de uma API de forma diária e automatizada, (ii) identificar oportunidades de arbitragem com base nesses dados, (iii) otimizar tanto o lucro quanto a quantidade de *surebets* escolhidas, assegurando uma distribuição eficiente do recurso entre o conjunto selecionado. Além disso, (iv) oferecer diferentes soluções da fronteira de Pareto, adaptadas ao objetivo e perfil do usuário. Isso significa que, seja o apostador mais arrojado, visando a maximização do lucro, ou o mais conservador, buscando uma distribuição mais equilibrada dos investimentos entre diversas *surebets*, o protótipo atende a ambos e outros, bastando escolher outra solução da fronteira de Pareto.

A realização do projeto fatorial 2^k permitiu não apenas a verificar a melhor estratégia de cruzamento, mas também calibrar os parâmetros do NSGA-II. Isso proporcionou a exploração das estratégias mais eficazes e a definição dos parâmetros ideais para os experimentos definidos.

Através dos experimentos executados e sua análise, confirmamos em ambiente simulado a hipótese de que seria possível explorar oportunidades de *surebets* no mercado brasileiro de apostas esportivas. Em todos os cenários, o algoritmo apresentou soluções capazes de maximizar o lucro, otimizar a quantidade de *surebets* e efetuar a distribuição eficiente do recurso, tudo isso a depender do objetivo preferido.

Caso fosse aplicado em ambiente real, as soluções geradas que alocam eficientemente as apostas (dispersando-as) provavelmente não infringiriam os termos de uso de casas de apostas *online*. Entretanto, registre-se que durante a realização deste trabalho não foram realizadas apostas no mundo real pelo autor, apenas a simulação de tal.

Ainda que não tenha sido firmadas apostas no ambiente real, o lucro obtido nas soluções reflete o que seria conquistado na realidade, se diferenciando na proporção que poderia ser investido e no potencial retorno, uma vez que ele é proporcional ao valor investido. Dessa forma, este trabalho contribui para a área, proporcionando uma forma de aproveitar as oportunidades de *surebets* de uma forma flexível, conforme o objetivo do apostador.

6.1 Principais contribuições

Destacam-se a seguir as principais contribuições alcançadas por meio deste trabalho:

- Formação e disponibilização de uma base de dados consistente e recente, resultante da coleta realizada durante a execução do projeto, que inclui pouco mais de 2.400 jogos de futebol catalogados e até 7 casas de apostas.
- Realização de experimentos utilizando o projeto fatorial, explorando três tipos de cruzamento (*One Point Crossover*, *Two Points Crossover* e *Uniform Crossover*) e apresentação da estratégia que obteve os melhores resultados para a modelagem proposta.
- Disponibilização do código-fonte, permitindo a realização de testes em cenários diversos ou sua utilização como ponto de partida para futuros trabalhos.
- Abordagem do problema visando maximizar dois objetivos e oferecer opções de aposta para diferentes perfis, o que até então não foi visto na literatura até onde pudemos verificar.

6.2 Trabalhos futuros

A aplicação da meta-heurística NSGA-II, como explorada neste trabalho, oferece a oportunidade de otimizar múltiplos objetivos. Durante o desenvolvimento, identificou-se a possibilidade de incorporar objetivos adicionais, como considerar taxas e valores mínimos exigidos pelas casas de apostas para saque dos ganhos. Quando uma aposta de *surebet* é finalizada, o dinheiro muitas vezes fica concentrado em uma das casas, sendo crucial e pode existir que seja atingido uma quantia mínima para o jogador conseguir fazer o saque, dessa forma, criar um objetivo que vise minimizar essa quantia, ou minimizar os dias que o dinheiro fica liberado, poderia ajudar a modelar mais o cenário real de apostas esportivas *online*.

Além desse ponto, abaixo são listados outros tópicos que poderiam enriquecer o trabalho:

- Explorar outros esportes e mercados para identificação de *surebets*. Este estudo se concentrou no futebol e em um mercado específico para identificar as oportunidades de *surebet*, mas as casas de apostas oferecem uma variedade de esportes e mercados. Ampliar essa análise para incluir mais opções pode proporcionar uma visão mais abrangente.

- Analisar padrões de surgimento de *surebets* em diferentes dias da semana, horários ou até mesmo em campeonatos específicos. Identificar se há alguma sazonalidade ou padrão específico pode melhorar a estratégia de seleção de *surebets*.
- Fazer uma análise do período que uma *surebet* normalmente permanece disponível, para investigar se há alguma correlação com a popularidade do evento. Isso permitirá entender se oportunidades de arbitragem em eventos relevantes são corrigidas mais rapidamente ou se há outros padrões identificáveis.
- Investigar a possibilidade de coleta de dados de forma mais independente, fazer a coleta de dados das casas de apostas que autorizem essa prática ou mesmo utilizar APIs fornecidas diretamente por essas casas. Isso pode reduzir a dependência de terceiros na coleta de dados.
- Desenvolver uma automação das entradas com base na variação do lucro, estabelecendo um valor mínimo aceitável. Considerar a espera para entrar no mercado até que o lucro atinja um ponto de inflexão (representando que atingiu um máximo e começou a cair) pode ser interessante.
- Desenvolver uma função objetivo que utilize a Teoria de Markowitz¹ ao mercado de apostas esportivas, onde o risco associado seria não só a variação das *odds*, como também as chances das casas de aposta cancelarem a entrada do apostador.
- Identificar o perfil do apostador (ex: lógica *fuzzy*) e retornar soluções que o atenda melhor.
- Ainda que não tenha sido pensado em um produto, desenvolver uma interface amigável poderia facilitar o seu uso pelos usuários.
- Utilizar ou desenvolver uma *API* para que, além de coletar os dados, também diretamente realize apostas em diferentes casas de apostas conforme as *surebets* identificadas e selecionadas.
- Os experimentos foram realizados em um ambiente simulado. Por isso, seria interessante investigar e analisar em quais condições as casas de apostas conseguiriam detectar o uso da estratégia de arbitragem e impor as suas penalizações. Além disso, seria relevante examinar como os *Termos de Uso* de cada casa de apostas se relaciona com a rapidez com que tal casa consegue identificar o uso dessa estratégia e aplicar as penalidades.

¹ A Teoria de Markowitz propõe a construção de portfólios de investimentos otimizados através da diversificação, equilibrando o retorno esperado com o risco, medido pelo desvio padrão dos retornos, para alcançar a máxima eficiência conforme as preferências de risco do investidor.

-
- Analisar as razões pelas quais as soluções retornadas pelo algoritmo começam a regredir o valor do *hypervolume* após de 150 gerações, aproximadamente. Isto poderia estar relacionado tanto ao *crowding distance* do NSGA-II (optar por indivíduos que mais se distanciem da população típica) quanto à taxa de mutação configurada e o operador de *crossover* adotado.

Referências

API, T. O. *The Odds API*. 2023. Disponível em: <<https://the-odds-api.com/>>. Citado 4 vezes nas páginas 31, 34, 35 e 56.

COELLO, C. A. C.; LAMONT, G. B.; VELDHUIZEN, D. A. V. *Evolutionary Algorithms for Solving Multi-Objective Problems*. 2. ed. [S.l.]: Springer, 2007. Citado na página 32.

DEB, K. et al. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE transactions on evolutionary computation*, IEEE, v. 6, n. 2, p. 182–197, 2002. Citado 2 vezes nas páginas 25 e 26.

EADINGTON, W. R. Roll the bones: The history of gambling. *Journal of Gambling Issues*, n. 21, p. 135–137, 2008. Citado na página 15.

FRANCK, E.; VERBEEK, E.; NÜESCH, S. Inter-market arbitrage in betting. *Economica*, v. 80, n. 318, p. 300–325, 2013. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/ecca.12009>>. Citado 2 vezes nas páginas 16 e 55.

GRANDVIEWRESEARCH. *Sports Betting Market Size Worth 182.12BillionBy2030.StateofCaliforniaatGrandViewResearch, Inc.201SpearStreet1100, San Francisco*. [//www.grandviewresearch.com/press-release/global-sports-betting-market](http://www.grandviewresearch.com/press-release/global-sports-betting-market). Citado na página 15.

JONES, D.; MIRRAZAVI, S.; TAMIZ, M. Multi-objective meta-heuristics: An overview of the current state-of-the-art. *European Journal of Operational Research*, v. 137, n. 1, p. 1–9, 2002. ISSN 0377-2217. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0377221701001230>>. Citado na página 25.

KAUNITZ, L.; ZHONG, S.; KREINER, J. Beating the bookies with their own numbers – and how the online sports betting market is rigged. *arXiv preprint arXiv:1710.02824*, 2017. Citado 3 vezes nas páginas 16, 24 e 29.

LEVITT, S. D. Why are gambling markets organised so differently from financial markets? *The Economic Journal*, Oxford University Press, v. 114, n. 474, p. 223–246, 2004. Citado na página 28.

POLA, G. On entropy and portfolio diversification. *Journal of Asset Management*, Macmillan Publishers Ltd., v. 17, n. 4, p. 218–228, 2016. Citado na página 40.

RESEARCH, G. V. *Sports Betting Market Size, Share Trends Analysis Report By Platform, By Betting Type (Fixed Odds Wagering, Exchange Betting, Live/In-Play Betting, eSports Betting), By Sports Type, By Region, And Segment Forecasts, 2023 - 2030*. State of California at Grand View Research, Inc. 201 Spear Street 1100, San Francisco, CA 94105, United States, 2023. <https://www.grandviewresearch.com/industry-analysis/sports-betting-market-report>. Citado na página 15.

STREICHERT, F.; ULMER, H.; ZELL, A. Evaluating a hybrid encoding and three crossover operators on the constrained portfolio selection problem. *IEEE Congress on Evolutionary Computation*, IEEE, v. 1, p. 932–939, 2004. Citado 2 vezes nas páginas 36 e 38.

TAX, C. for; ACCOUNTABILITY, B. *Gamblers' sports betting losses generate 122millionforIllinoisintwoyears.*[S.l.], 2022.*Https : //www.ctbaonline.org/press – room/gamblers – sports – betting – losses – generate – 122 – million – illinois – two – years.*Citado na página 16.

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. 1, p. 26–48, 2009. Citado na página 16.