

UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"

FACULDADE DE CIÊNCIAS - CAMPUS BAURU

DEPARTAMENTO DE COMPUTAÇÃO

BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

GABRIEL JOOJI YAMASHIRO

**APLICAÇÃO DE APRENDIZADO DE MÁQUINA E
EQUILÍBRIO DE NASH NO JOGO DE PÔQUER**

BAURU

Novembro/2025

GABRIEL JOOJI YAMASHIRO

APLICAÇÃO DE APRENDIZADO DE MÁQUINA E EQUILÍBRIO DE NASH NO JOGO DE PÔQUER

Trabalho de Conclusão de Curso do Curso
de Ciência da Computação da Universidade
Estadual Paulista “Júlio de Mesquita Filho”,
Faculdade de Ciências, Campus Bauru.

Orientador: Prof. Dr. Leandro Aparecido Passos
Junior

BAURU

Novembro/2025

Gabriel Jooji Yamashiro

APLICAÇÃO DE APRENDIZADO DE MÁQUINA E EQUILÍBRIO DE NASH NO JOGO DE PÔQUER

Trabalho de Conclusão de Curso do Curso de Ciência da Computação da Universidade Estadual Paulista "Júlio de Mesquita Filho", Faculdade de Ciências, Campus Bauru.

Banca Examinadora

**Prof. Dr. Leandro Aparecido Passos
Junior**

Orientador

Universidade Estadual Paulista "Júlio de
Mesquita Filho"

Faculdade de Ciências

Departamento de Ciência da Computação

**Profa. Dra. Simone das Graças
Domingues Prado**

Universidade Estadual Paulista "Júlio de
Mesquita Filho"

Faculdade de Ciências

Departamento de Ciência da Computação

Prof. Dr. Douglas Rodrigues

Universidade Estadual Paulista "Júlio de
Mesquita Filho"

Faculdade de Ciências

Departamento de Ciência da Computação

Bauru, 11 de novembro de 2025.

Y19a

Yamashiro, Gabriel Jooji

Aplicação de aprendizado de máquina e equilíbrio de Nash no jogo de pôquer / Gabriel Jooji Yamashiro. -- Bauru, 2025

44 f. : il.

Trabalho de conclusão de curso (Bacharelado - Ciência da Computação) - Universidade Estadual Paulista (UNESP), Faculdade de Ciências, Bauru

Orientador: Leandro Aparecido Passos Junior

1. Pôquer. 2. Inteligência artificial. 3. Equilíbrio Nash. 4. CFR. 5. Agrupamento de dados. I. Título.

Agradecimentos

Gostaria de expressar minha profunda gratidão a todas as pessoas que fizeram parte dessa jornada. Primeiramente, à minha família — aos meus pais, Gilberto e Gina, pelo amor, incentivo e suporte incondicional em todas as etapas da minha vida — e ao meu irmão, Fernando, pelos conselhos e ensinamentos que sempre levo comigo.

Aos meus amigos da República Sem Parar, pelas conversas e por todas as experiências compartilhadas que tornaram essa caminhada mais leve e marcante.

Aos meus sócios de pool (Swapool), cuja parceria foi essencial e sem os quais eu talvez não reconhecesse minha verdadeira paixão pelo pôquer.

Ao meu orientador, pela paciência, dedicação e pelas orientações que foram fundamentais para a realização deste trabalho.

“A mente tranquila vem quando você deixa de tentar controlar o que está fora do seu alcance.”

— *Marco Aurélio*

Resumo

O pôquer *Heads-Up No-Limit Texas Hold'em*, por ser um jogo de informação imperfeita, apresenta um desafio complexo para a inteligência artificial devido à sua vasta combinatória, além de elementos como blefe e adaptação ao oponente. Este trabalho propõe um agente de Inteligência Artificial (IA) que integra o algoritmo *Counterfactual Regret Minimization* (CFR) para aproximar o Equilíbrio de Nash aliado a técnicas de aprendizado de máquina para abstrações de cartas e ações. As abstrações foram realizadas por agrupamento de dados usando o K-Means, baseado em distribuições de equidade, gerando 169 grupos no *pré-flop*, 50 no *flop*, 100 no *turn* e 200 no *river*. Um dataset do tamanho de 1 milhão de mãos foi treinado de forma *offline*, com treinamentos separados para *pré-flop* e *pós-flop*. O agente foi avaliado contra o *Slumbot* em cerca de 10.000 mãos, registrando perdas de -41.47 *Big Blind* a cada 100 mãos (BB/100) (Dataset de 250.000 mãos), -32.125 BB/100 (Dataset de 500.000 mãos), -47.025 BB/100 (Dataset de 750.000 mãos) e -47.19 BB/100 (Dataset de 1.000.000 mãos). Apesar da não linearidade dos resultados, possivelmente devido à alta variância no pôquer, a estratégia CFR superou consistentemente heurísticas predefinidas, exibindo perdas menores em todos os testes. A implementação demonstrou viabilidade em *hardware* modesto, contribuindo para avanços em jogos de informação imperfeita e teoria dos jogos.

Palavras-chave: Pôquer, Inteligência Artificial, CFR, Equilíbrio de Nash, Abstração, Agrupamento de dados.

Abstract

The Heads-Up No-Limit Texas Hold'em poker, as an imperfect information game, poses a significant challenge for artificial intelligence due to its vast combinatorial complexity, coupled with elements such as bluffing and opponent adaptation. This study proposes an AI agent that integrates the Counterfactual Regret Minimization (CFR) algorithm to approximate the Nash Equilibrium, combined with machine learning techniques for card and action abstractions. Abstractions were implemented using K-Means clustering based on equity distributions. A dataset of 1 million hands was generated offline, with separate training for pre-flop and post-flop phases. Despite the non-linear results, likely due to poker's inherent high variance, the CFR-based strategy consistently outperformed predefined heuristics, showing lower losses across all scenarios. The implementation demonstrated feasibility on modest hardware, contributing to advancements in imperfect information games and game theory. This work was conducted as part of the undergraduate thesis in Computer Science at Universidade Estadual Paulista "Júlio de Mesquita Filho", Faculty of Sciences, Bauru, under the supervision of Prof. Dr. Leandro Aparecido Passos Junior, and completed on October 30, 2025.

Keywords: Poker, Artificial Intelligence, CFR, Nash Equilibrium, Abstraction, Clustering.

Lista de figuras

Figura 1 – Hierarquia das mãos	19
Figura 2 – Árvore de Ações	24
Figura 3 – Estratégia 0: Aposto Tudo em BB/100 em 10.000 mãos.	34
Figura 4 – Estratégia 1: Mesa ou paga em BB/100 em 10.000 mãos.	35
Figura 5 – Estratégia 2: Equidade Simples em BB/100 em 10.000 mãos.	36
Figura 6 – Estratégia 3: Equidade mais avançada em BB/100 em de 10.000 mãos. . .	37
Figura 7 – Desempenho da estratégia CFR com abstração em termos de BB/100 ao longo de 10.000 mãos simuladas.Dataset = 250.000 mãos	38
Figura 8 – Desempenho da estratégia CFR com abstração em termos de BB/100 ao longo de 10.000 mãos simuladas.Dataset =500.000 mãos	39
Figura 9 – Desempenho da estratégia CFR com abstração em termos de BB/100 ao longo de 10.000 mãos simuladas.Dataset = 750.000 mãos	40
Figura 10 – Desempenho da estratégia CFR com abstração em termos de BB/100 ao longo de 10.000 mãos simuladas.Dataset = 1.000.000 mãos	41

Lista de abreviaturas e siglas

API	Application Programming Interface
BB	Big Blind
BB/100	Big Blinds a cada 100 mãos
CFR	Counterfactual Regret Minimization
CMU	Carnegie Mellon University
EV	Expected Value
IA	Inteligência Artificial
ML	Machine Learning
SB	Small Blind
WSOP	World Series of Poker

Sumário

1	INTRODUÇÃO	12
1.1	Problemática e Justificativa	12
1.2	Objetivos	13
1.2.1	Objetivo Geral	13
1.2.2	Objetivos Específicos	13
2	TRABALHOS CORRELATOS	15
2.1	Libratus (2017)	15
2.2	Pluribus (2019)	15
2.3	Modicum (2017–2019)	15
2.4	Slumbot (2020–presente)	16
2.5	Relação com o presente trabalho	16
3	REFERENCIAL TEÓRICO	17
3.1	Elementos Históricos do Pôquer e da Modalidade <i>Texas Hold'em</i>	17
3.2	O Jogo de Pôquer <i>Heads-Up</i> e a Dimensão Psicológica	17
3.2.1	Como Jogar <i>Heads-Up Texas Hold'em</i>	17
3.2.2	Hierarquia de mãos no pôquer	19
3.2.3	Jogo Psicológico no Pôquer	20
3.3	Equilíbrio de Nash	20
3.4	<i>Counterfactual Regret Minimization</i>	21
3.4.1	<i>Regret Matching</i>	21
3.4.1.1	Exemplo ilustrativo	21
3.4.2	Introdução ao CFR	22
3.4.3	Como o CFR funciona	22
3.4.4	Exemplo: Kuhn Poker com CFR	23
3.4.5	Resumo do CFR	27
3.4.6	Minimização de Arrependimento Contrafactual	27
4	METODOLOGIA	30
4.1	Ambiente de Desenvolvimento	30
4.2	Ferramentas Utilizadas	30
4.3	Geração do Dataset	31
4.4	Geração do Conjunto de Dados	31
4.5	Processo de Treinamento	31
4.6	Resumo do Fluxo Experimental	31

5	RESULTADOS EXPERIMENTAIS	33
5.1	Estratégia 0: Aposto Tudo (Bruto -227.12 BB/100)	33
5.2	Estratégia 1: Mesa ou paga (Bruto -123.335 BB/100)	34
5.3	Estratégia 2: Equidade Simples (Bruto -112.045 BB/100)	35
5.4	Estratégia 3: Equidade mais avançada (Bruto -204,292 BB/100) . .	36
5.5	Estratégia 4: CFR com abstração Dataset 250.000 (Bruto -41.47 BB/100)	37
5.6	Estratégia 4: CFR com abstração Dataset 500.000 (Bruto -32.125 BB/100)	38
5.7	Estratégia 4: CFR com abstração Dataset 750.000 (Bruto -47.025 BB/100)	39
5.8	Estratégia 4: CFR com abstração Dataset 1.000.000 (Bruto -47.19 BB/100)	40
5.9	Análise Geral	41
6	CONCLUSÃO	42
	REFERÊNCIAS	43

1 Introdução

A Teoria dos Jogos é um ramo da matemática aplicada e economia que estuda a tomada de decisão estratégica em situações de interação entre agentes racionais. Um dos seus conceitos centrais é o Equilíbrio de Nash, introduzido por John Forbes Nash Jr. em 1950. Um perfil de estratégias está em equilíbrio de Nash quando nenhum jogador pode aumentar sua recompensa esperada alterando unilateralmente sua estratégia, assumindo que as estratégias dos demais jogadores permanecem fixas (NASH, 1950). Em outras palavras, no equilíbrio de Nash todos os jogadores estão fazendo o “melhor possível” dadas as escolhas dos adversários, tornando-o um ponto de estabilidade preditiva em jogos competitivos.

Esse conceito possui aplicações que vão muito além da economia clássica, sendo hoje fundamental em Inteligência Artificial (IA) (RUSSELL; NORVIG, 2013), ciência política e, especialmente, no projeto de agentes autônomos capazes de jogar jogos complexos.

Diferentemente de jogos de informação perfeita como xadrez e Go — em que ambos os jogadores conhecem o estado completo do tabuleiro a todo momento —, o pôquer (em particular o *No-Limit Texas Hold'em*) é um jogo de informação imperfeita: cartas privadas, blefes, ordens de apostas e incerteza sobre as intenções do oponente geram uma explosão combinatória no espaço de estados e ações (CARMICHAEL, 2018). Essa característica torna o desenvolvimento de agentes de IA um desafio significativo.

A variante *Heads-Up* (um contra um) reduz significativamente a complexidade em relação a mesas com mais jogadores, mas ainda preserva todas as dificuldades essenciais de informação imperfeita, tornando-a o cenário ideal para o desenvolvimento e teste de algoritmos que buscam aproximar o equilíbrio de Nash em jogos de grande escala.

Este trabalho propõe a construção de um agente computacional para *No-Limit Texas Hold'em Heads-Up* baseado em algoritmos modernos como o *Counterfactual Regret Minimization* (CFR) e suas variantes. O objetivo é obter uma estratégia que converja para um equilíbrio de Nash (ou seja, uma estratégia que garanta um desempenho arbitrariamente próximo ao ótimo contra qualquer oponente), permitindo que o agente explore padrões comportamentais, adapte-se a diferentes estilos de jogo e sirva como base para avanços futuros em IA aplicada a jogos e a problemas reais de decisão estratégica sob incerteza.

1.1 Problemática e Justificativa

A natureza de informação imperfeita do pôquer, aliada à sua elevada complexidade combinatória, torna desafiador o desenvolvimento de agentes computacionais capazes de tomar decisões estratégicas consistentes por meio de abordagens tradicionais de IA e aprendizado

de máquina (em inglês, *Machine Learning*, ML). Cada rodada do jogo envolve múltiplos estados possíveis, ações estocásticas e interações estratégicas entre jogadores, o que dificulta a aproximação de soluções racionais, especialmente em cenários com múltiplos jogadores (CARMICHAEL, 2018).

Diante desse contexto, formula-se a seguinte problemática central desta pesquisa: **É possível desenvolver um agente de pôquer, baseado em técnicas de ML, capaz de tomar decisões alinhadas ao conceito de Equilíbrio de Nash no formato *Heads-Up*?**

A relevância da questão decorre de três fatores principais. Em primeiro lugar, o pôquer permanece um problema aberto na área de IA: embora avanços como o CFR e sistemas como *DeepStack* tenham demonstrado progresso, a tomada de decisão ótima em ambientes de informação imperfeita ainda não está plenamente resolvida, sobretudo no que diz respeito à convergência de estratégias robustas.

Em segundo lugar, a restrição ao formato *Heads-Up* reduz a complexidade computacional do problema sem comprometer o rigor científico. Esse cenário de dois jogadores permite controlar a explosão combinatória e possibilita análises mais precisas sobre convergência, equilíbrio e comportamento estratégico, elementos essenciais para avaliar a aderência das ações do agente ao Equilíbrio de Nash.

Por fim, a aplicação prática deste estudo transcende o domínio do pôquer. Técnicas utilizadas para lidar com incerteza, blefes, probabilidades e adaptação estratégica são diretamente aplicáveis a áreas como negociação automatizada, análise de riscos, mercados financeiros e sistemas que operam sob adversarialidade.

Dessa forma, este esforço se justifica por investigar uma abordagem que integra modelos de aprendizado de máquina com princípios da teoria dos jogos, especialmente por meio do algoritmo CFR. Ao explorar essa interseção entre IA e estratégia, o trabalho busca contribuir tanto para o avanço científico quanto para aplicações práticas em ambientes de tomada de decisão sob incerteza.

1.2 Objetivos

1.2.1 Objetivo Geral

Desenvolver um agente de IA baseado em ML que utilize o Equilíbrio de Nash para otimizar decisões no pôquer *Texas Hold'em Heads-Up*.

1.2.2 Objetivos Específicos

- Implementar um modelo de ML para identificar padrões de jogo e ajustar estratégias automaticamente;

- Aplicar algoritmos de Teoria dos Jogos para encontrar estratégias otimizadas baseadas no Equilíbrio de Nash;
- Avaliar a performance do agente em simulações contra adversários computacionais.

2 Trabalhos Correlatos

O pôquer, especialmente a modalidade *No-Limit Texas Hold'em*, consolidou-se como um dos principais *benchmarks* para avaliar algoritmos de decisão em jogos de informação imperfeita. A seguir, apresentam-se os principais agentes que influenciaram diretamente este trabalho.

2.1 Libratus (2017)

Desenvolvido pela Carnegie Mellon University, o Libratus foi o primeiro sistema a derrotar consistentemente profissionais de elite em *Heads-Up No-Limit Texas Hold'em* (BROWN; SANDHOLM, 2018). Baseado em CFR com abstrações sofisticadas e refinamento em tempo real, o sistema consumiu cerca de 15–19 milhões de *core-hours* no supercomputador Bridges, evidenciando que o estado da arte ainda exigia recursos computacionais extremos.

2.2 Pluribus (2019)

O trabalho mais avançado até o momento é o Pluribus, também da CMU/Facebook AI Research, que estendeu o sucesso do CFR para mesas de seis jogadores (*6-max*), algo considerado praticamente impossível poucos anos antes (BROWN; SANDHOLM, 2019b). Diferentemente do Libratus, o Pluribus introduziu técnicas inovadoras de busca em tempo real com profundidade limitada (*depth-limited search*) e estratégias baseadas em *blueprints* pré-computadas, alcançando desempenho super-humano contra cinco profissionais simultaneamente usando apenas cerca de 12.500 *core-hours* — uma fração do custo do Libratus. O Pluribus reforça que variantes modernas do CFR como a abordagem dominante para jogos de grande escala.

2.3 Modicum (2017–2019)

Publicado como contraponto ao Libratus, o Modicum demonstrou que resultados competitivos em *Heads-Up* podem ser obtidos com hardware comum (uma única máquina com GPU) utilizando abstrações agressivas de ações e baralhos, além de CFR com amostragem (BROWN; SANDHOLM, 2019a). Seu objetivo explícito foi democratizar a pesquisa em IA para pôquer.

2.4 Slumbot (2020–presente)

Desenvolvido pela Universidade de Alberta, o Slumbot é o agente de *Heads-Up No-Limit* mais conhecido disponível publicamente. Vencedor de diversas edições da Annual Computer Poker Competition, ele utiliza CFR com abstrações de cartas e histórico, e pode ser enfrentado online por qualquer pessoa (TEAM, 2020). Por sua acessibilidade e estabilidade ao longo dos anos, constitui um *benchmark* aberto amplamente aceito pela comunidade.

2.5 Relação com o presente trabalho

Este trabalho posiciona-se na linha do Modicum e do espírito do Pluribus: explorar até onde é possível avançar em *Heads-Up No-Limit Texas Hold'em* utilizando apenas recursos computacionais modestos (máquina pessoal ou pequena nuvem) e implementações relativamente simples de CFR vanilla, CFR+, Discounted CFR e Monte Carlo CFR. Assim como o Modicum, busca viabilidade prática; assim como o Pluribus, valoriza a pureza do autjogo sem conhecimento heurístico humano. Por fim, o agente desenvolvido será avaliado empiricamente contra o Slumbot, permitindo comparação direta com um adversário público consolidado e posicionando o desempenho alcançado em relação aos três eixos principais da literatura atual:

Esse conjunto de comparações permite contextualizar o trabalho dentro de três eixos:

1. **Estado da arte** (Libratus/Pluribus – escala massiva).
2. **Viabilidade em Hardware** (Modicum – abstrações + CFR).
3. **Benchmark aberto e reproduzível** (Slumbot – referência prática).

Dessa forma, o presente estudo contribui tanto para o entendimento prático da convergência do CFR em recursos limitados quanto para a disponibilização de um novo agente comparável a baselines públicas.

3 Referencial Teórico

3.1 Elementos Históricos do Pôquer e da Modalidade *Texas Hold'em*

O pôquer, um jogo de cartas com origens no início do século XIX, combina sorte, habilidade e estratégia, possivelmente derivado de jogos europeus como o “Poque” francês e o “Pochen” alemão, influenciados pelo persa “As Nas” (DALLA; ALVAREZ, 2009). Popularizou-se nos barcos fluviais do Rio Mississippi por volta de 1830, como entretenimento e aposta entre comerciantes e viajantes (MCMANUS, 2009), expandindo-se pelos Estados Unidos e gerando variações como *Five-Card Draw*, *Seven-Card Stud* e *Texas Hold'em*, esta última tornando-se a forma mais competitiva. O *Texas Hold'em* surgiu no início do século XX em Robstown, Texas, destacando-se pelo uso de cartas comunitárias, que aumentaram a profundidade estratégica e reduziram a dependência da sorte (DALLA; ALVAREZ, 2009). Introduzido em Las Vegas nas décadas de 1960 e 1970 por profissionais como Doyle Brunson, Amarillo Slim e Crandell Addington nos cassinos *Golden Nugget* e *Binion's Horseshoe* (BRUNSON, 1978), tornou-se a modalidade principal do *World Series of Poker* (WSOP) em 1970. Addington observou: “O *Texas Hold'em* é o jogo que leva um minuto para aprender e uma vida inteira para dominar” (DALLA; ALVAREZ, 2009). Historicamente, o pôquer evoluiu para uma disciplina estratégica, vinculada à Teoria dos Jogos por Von Neumann e Morgenstern (NEUMANN; MORGENSTERN, 1944), que o classificam como um jogo de soma não nula, ideal para estudos em matemática aplicada, psicologia e inteligência artificial. O formato *Heads-Up*, disputado entre dois jogadores, ganhou destaque em torneios e pesquisas, enfatizando leitura psicológica, blefe e adaptação probabilística (CARO, 2003). Com o avanço tecnológico na década de 2000, projetos como Libratus e Pluribus, da *Carnegie Mellon University*, superaram humanos usando CFR para aproximar o Equilíbrio de Nash (BROWN; SANDHOLM, 2018), consolidando o pôquer como plataforma para pesquisa em decisão racional e aprendizado de máquina.

3.2 O Jogo de Pôquer *Heads-Up* e a Dimensão Psicológica

3.2.1 Como Jogar *Heads-Up Texas Hold'em*

O *Heads-Up Texas Hold'em*, jogado entre dois jogadores, exige estratégia matemática, leitura comportamental e adaptação dinâmica. Cada jogador recebe duas cartas privadas (*hole cards*) e utiliza, em conjunto, cinco cartas comunitárias dispostas na mesa para formar a melhor combinação possível de cinco cartas. As cartas comunitárias são reveladas em três etapas:

- **Flop:** três cartas abertas simultaneamente;

- **Turn:** uma quarta carta comunitária;
- **River:** a quinta e última carta comunitária.

O jogo é dividido em quatro rodadas de apostas, nas quais os jogadores podem escolher entre as ações **fold** (desistir ou fugir), **check** (mesa ou passar a vez), **call** (pagar a aposta ou igualar a aposta), **bet/raise** (apostar ou aumentar a aposta) ou **all-in** (apostar todas as fichas):

1. **Pré-flop:** apostas realizadas após o recebimento das cartas privadas e antes da revelação das comunitárias.
2. **Flop:** apostas após a revelação das três primeiras cartas comunitárias.
3. **Turn:** apostas após a quarta carta.
4. **River:** apostas após a quinta carta.

No poker Texas Hold'em, as blinds (ou "cegas") constituem apostas obrigatórias postas por dois jogadores antes da distribuição das cartas, com o objetivo de iniciar o pote e estimular a ação no jogo. No formato heads-up (confronto direto entre dois jogadores), essa estrutura é adaptada para otimizar a dinâmica do jogo, eliminando posições intermediárias e garantindo alternância equitativa das responsabilidades. Nessa configuração, identificam-se apenas duas posições fixas, que rotacionam a cada mão no sentido horário:

- **Small Blind (SB):** O jogador no botão do dealer (*button*) posta a *small blind*, equivalente a metade do valor da *big blind* (por exemplo, em uma estrutura 1/2, posta 1 unidade). Essa posição confere vantagem posicional, atuando primeira no pré-flop (pode completar a aposta para igualar a *big blind*, aumentar ou desistir) e por último nas rodadas pós-flop (*flop*, *turn* e *river*).
- **Big Blind (BB):** O jogador imediatamente no sentido horário do *button* aposta a *big blind*, no valor integral da aposta mínima (exemplo: 2 unidades). Atua por último no pré-flop (após a ação do SB) e atua primeiro nas rodadas pós-flop

Essa configuração assegura que o pote inicie com o equivalente a 1,5 *big blinds* (SB+BB), fomentando agressividade desde o início e evitando inatividade. A rotação do *button* a cada mão equilibra as posições, de modo que cada jogador ocupe alternadamente o SB e o BB.

Fase final, também chamada de *Showdown*, ocorre quando ambos os jogadores permanecem na mão até o final da rodada de apostas, e as cartas são reveladas e vence quem tiver a melhor combinação de cinco cartas segundo a hierarquia oficial do pôquer.

3.2.2 Hierarquia de mãos no pôquer

A hierarquia das mãos, da mais fraca para a mais forte é apresentada na figura 1.

Figura 1 – Hierarquia das mãos



Fonte: (GIPSYTEAM, 2022)

1. **Carta Alta (High Card):** Quando o jogador não forma nenhuma combinação, a força é determinada pela carta de maior valor. Exemplo: Ás é mais forte que Rei.
2. **Par (Pair):** Duas cartas do mesmo valor. Exemplo: dois Damas.
3. **Dois Pares (Two Pair):** Duas combinações de pares distintos. Exemplo: dois Reis e dois Setes.
4. **Trinca (Three of a Kind):** Três cartas do mesmo valor. Exemplo: três Três.
5. **Sequência (Straight):** Cinco cartas em ordem numérica, independentemente do naipe. Exemplo: A-2-3-4-5.
6. **Flush:** Cinco cartas do mesmo naipe, não necessariamente em sequência.

7. **Full House**: Combinação de uma trinca e um par. Exemplo: três Três e dois Dois.
8. **Quadra(Four of a Kind)**: Quatro cartas do mesmo valor. Exemplo: quatro Quatros.
9. **Straight Flush**: Cinco cartas em sequência, todas do mesmo naipe. Exemplo: 5-6-7-8-9 de paus.
10. **Royal Flush**: A sequência mais alta possível: 10-J-Q-K-A do mesmo naipe. É a mão mais rara e imbatível no Texas Hold'em.

Em situações em que dois jogadores possuem mãos do mesmo tipo, o valor das cartas é utilizado como critério de desempate ("kicker"). Por exemplo, um par de Reis vence um par de Damas; se ambos têm um par de Reis, vence quem possuir a carta extra de maior valor. Caso mesmo assim continue empatado, então o pote é dividido igualmente entre os jogadores restantes.

3.2.3 Jogo Psicológico no Pôquer

O componente psicológico, crucial no *Heads-Up*, envolve interpretar padrões de apostas e manipular percepções via blefe, semi-blefe e *trap plays*, explorando vieses cognitivos (CARO, 2003). Na Teoria dos Jogos, isso reflete a informação imperfeita, modelada como incerteza sobre o oponente (NEUMANN; MORGENSTERN, 1944). Computacionalmente, inspira modelos adaptativos, como o CFR, que ajusta estratégias via aprendizado iterativo baseado em arrependimento, aproximando intuitivamente o Equilíbrio de Nash (NASH, 1951).

3.3 Equilíbrio de Nash

O equilíbrio de Nash é um conceito fundamental na teoria dos jogos, nomeado em homenagem ao matemático John Nash (NASH, 1951). Ele representa uma situação em um jogo onde nenhum jogador pode melhorar seu resultado alterando unilateralmente sua estratégia, assumindo que as estratégias dos outros jogadores permaneçam inalteradas.

Em um jogo com múltiplos jogadores, um perfil de estratégias está em equilíbrio de Nash se, para cada jogador, a estratégia escolhida for a melhor resposta às estratégias dos demais jogadores. Isso significa que, dada a escolha dos outros, nenhum jogador tem incentivo para desviar de sua estratégia atual.

Por exemplo, no clássico jogo do Dilema do Prisioneiro, o equilíbrio de Nash ocorre quando ambos os prisioneiros confessam, mesmo que a cooperação mútua (não confessar) leve a um resultado melhor para ambos coletivamente. Aqui, confessar é a estratégia dominante, levando ao equilíbrio.

Em jogos de soma zero, como o xadrez ou o poker, o equilíbrio de Nash garante que cada jogador maximize seu payoff mínimo esperado, assumindo que o oponente joga de forma ótima.

Este conceito é amplamente aplicado em economia, biologia evolutiva e ciências sociais para modelar interações estratégicas.

Para ilustrar, considere o jogo Pedra-Papel-Tesoura. A estratégia de equilíbrio de Nash é cada jogador escolher aleatoriamente cada opção com probabilidade $1/3$, pois qualquer desvio pode ser explorado pelo oponente.

Em contextos computacionais, como na tese discutida, algoritmos como o CFR aproximam equilíbrios de Nash em jogos extensos de informação imperfeita.

Para uma definição formal, em um jogo em forma normal com jogadores $i = 1, \dots, n$, conjuntos de estratégias S_i e funções de utilidade u_i , um perfil $\sigma^* = (\sigma_1^*, \dots, \sigma_n^*)$ é um equilíbrio de Nash se:

$$u_i(\sigma_i^*, \sigma_{-i}^*) \geq u_i(\sigma_i, \sigma_{-i}^*) \quad \forall \sigma_i \in S_i, \forall i$$

onde σ_{-i}^* denota as estratégias dos outros jogadores.

3.4 *Counterfactual Regret Minimization*

3.4.1 *Regret Matching*

O *Regret Matching* é um dos conceitos fundamentais que compõem o algoritmo CFR, amplamente utilizado para resolver jogos de decisão sequencial imperfeita, como o *poker heads-up*. O termo *regret* pode ser traduzido como “arrependimento” e representa o quanto um jogador teria obtido um resultado melhor caso tivesse escolhido uma ação diferente em situações passadas. Assim, o *Regret Matching* consiste em ajustar as probabilidades das ações futuras de acordo com os arrependimentos acumulados ao longo do tempo (ZINKEVICH et al., 2007).

De forma intuitiva, quanto maior o arrependimento associado a não ter escolhido uma ação no passado, maior será a probabilidade de o jogador optar por essa ação no futuro. Esse processo permite que a estratégia do jogador evolua de forma adaptativa, com base em experiências anteriores, aproximando-se gradualmente de um equilíbrio ótimo.

3.4.1.1 Exemplo ilustrativo

Considere o jogo Pedra - Papel - Tesoura. Suponha que um jogador dispute diversas rodadas contra o mesmo oponente. Após várias partidas, ele observa que:

- quando joga *pedra*, costuma perder;
- quando joga *papel*, costuma ganhar;
- quando joga *tesoura*, o resultado é equilibrado.

Nesse cenário, o jogador percebe um “arrependimento” por ter escolhido com frequência a ação *pedra*, pois ela gerou resultados negativos. O *Regret Matching* utiliza esse histórico de arrependimentos para aumentar a probabilidade de escolha da ação *papel*, que teria produzido melhores resultados nas jogadas anteriores. Assim, à medida que o processo se repete, o jogador passa a tomar decisões baseadas em seus arrependimentos acumulados, desenvolvendo uma estratégia mais equilibrada e menos explorável.

3.4.2 Introdução ao CFR

Enquanto o *regret matching* ajusta estratégias com base em arrependimentos de jogadas passadas em jogos simples, o CFR foi desenvolvido para lidar com jogos mais complexos, como pôquer, onde há informações ocultas e múltiplas decisões. Ele permite que um agente aprenda uma estratégia ótima jogando contra si mesmo, considerando não só o que aconteceu, mas também o que poderia ter acontecido — daí o termo “contrafactual”.

Em jogos de informação imperfeita, como o pôquer, as decisões ocorrem em pontos denominados *information sets* (nós). O CFR calcula o arrependimento contrafactual para cada ação possível em cada nó, levando em conta a probabilidade de chegar até aquele ponto específico do jogo. Com isso, o algoritmo converge para o Equilíbrio de Nash, em que nenhum jogador pode melhorar seu desempenho unilateralmente, dado o comportamento do oponente (ZINKEVICH et al., 2007).

3.4.3 Como o CFR funciona

1. **Estrutura da Árvore do Jogo:** O jogo é representado como uma árvore, em que cada nó indica um ponto de decisão (por exemplo: “apostar” ou “passar” com uma determinada carta).
2. **Cálculo do Arrependimento Contrafactual:** Para cada ação s em um nó i , o arrependimento contrafactual é calculado conforme:

$$R_i(s) = [u(s', I) - u(s, I)] \cdot P(I) \quad (3.1)$$

onde:

- $R_i(s)$: arrependimento contrafactual da ação s no nó i ;
- $u(s', I)$: utilidade se a melhor ação alternativa s' fosse escolhida;

- $u(s, I)$: utilidade da ação realmente escolhida s ;
- $P(I)$: probabilidade de alcançar o nó I , que depende das ações do oponente.

3. **Arrependimentos Cumulativos:** Os arrependimentos são somados ao longo das iterações:

$$R_i^{\text{cum}}(s) = R_i^{\text{cum}}(s) + R_i(s) \quad (3.2)$$

4. **Escolha da Próxima Ação (Regret Matching):** A próxima ação é escolhida proporcionalmente aos arrependimentos positivos acumulados:

$$P(s) = \frac{R_i^{\text{cum}}(s)}{\sum R_i^{\text{cum}}(s^+)} \quad (3.3)$$

onde $R_i^{\text{cum}}(s^+)$ são apenas os arrependimentos positivos. Caso não existam arrependimentos positivos, adota-se uma estratégia uniforme.

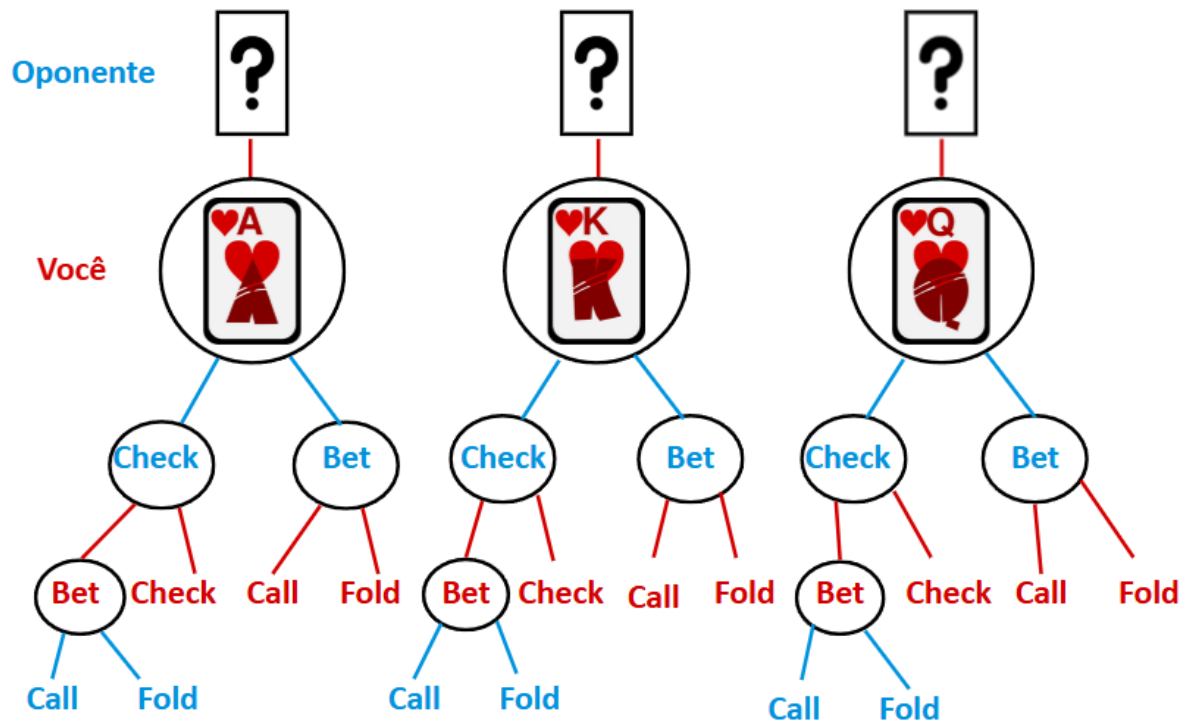
5. **Estratégia Média:** Após muitas iterações, a estratégia final é obtida pela média das escolhas realizadas, aproximando-se do *Equilíbrio de Nash*.

Essa abordagem é especialmente eficiente em jogos como o pôquer, pois considera todas as possibilidades, mesmo quando as informações são parciais. Na próxima seção, esse funcionamento será ilustrado na prática por meio do *Kuhn Poker*.

3.4.4 Exemplo: Kuhn Poker com CFR

Para ilustrar o funcionamento do algoritmo *Counterfactual Regret Minimization* (CFR), apresenta-se a seguir um exemplo completo e coerente baseado no jogo Kuhn Poker. Nesta versão, utiliza-se um baralho reduzido contendo três cartas: **Q**, **K** e **A**, onde $\mathbf{Q} < \mathbf{K} < \mathbf{A}$. O jogo segue as seguintes regras: como visto na figura 2.

Figura 2 – Árvore de Ações



Fonte: o autor

- Cada jogador recebe uma carta privada.
- Ambos iniciam colocando uma unidade no pote.
- O Jogador 1 pode *passar* (*check*) ou *apostar* (*bet*).
- Se o Jogador 1 passar, o Jogador 2 pode passar ou apostar.
- Caso o Jogador 2 aposte, o Jogador 1 pode *pagar* (*call*) ou *desistir* (*fold*).
- Se a rodada termina em *check-check* ou *bet-call*, ocorre o *showdown* e vence a carta mais alta.

O objetivo do CFR é ajustar iterativamente a estratégia dos jogadores por meio do cálculo dos arrependimentos contrafactuais. A seguir, apresentam-se três iterações exemplificativas totalmente consistentes com as regras do jogo.

Iteração 1

Situação: Jogador 1 recebe **Q** e Jogador 2 recebe **A**. A sequência de ações é:

J1: passar → J2: apostar → J1: desistir.

O Jogador 1 perde sua aposta inicial, resultando em utilidade real:

$$u = -1.$$

Arrependimento contrafactual do Jogador 1 ao escolher *não apostar* no início: se o Jogador 1 tivesse apostado e o Jogador 2 desistisse (situação possível em equilíbrios com blefe), sua utilidade seria:

$$u' = +2.$$

Assim, o arrependimento contrafactual é dado por:

$$R_1(\text{apostar}) = (u' - u) \cdot P(I) = (2 - (-1)) \cdot 0,5 = 1,5. \quad (3.4)$$

Arrependimentos acumulados após a iteração:

$$R^{\text{cum}}(\text{apostar}) = 1,5, \quad R^{\text{cum}}(\text{passar}) = 0.$$

Iteração 2

Situação: Jogador 1 recebe **K** e Jogador 2 recebe **Q**. A sequência de ações é:

$$J1: \text{apostar} \rightarrow J2: \text{pagar}.$$

Como **K** > **Q**, o Jogador 1 vence o pote final de duas unidades, resultando em:

$$u = +2.$$

Se o Jogador 1 tivesse passado e o Jogador 2 também passasse, o valor de utilidade seria:

$$u' = +1.$$

Assim, o arrependimento contrafactual para a ação alternativa *passar* é:

$$R_2(\text{passar}) = (u' - u) \cdot 0,5 = (1 - 2) \cdot 0,5 = -0,5. \quad (3.5)$$

Como arrependimentos negativos não contribuem no *regret matching*, temos:

$$R^{\text{cum}}(\text{passar}) = 0.$$

Os valores acumulados permanecem:

$$R^{\text{cum}}(\text{apostar}) = 1,5, \quad R^{\text{cum}}(\text{passar}) = 0.$$

Iteração 3

Situação: Jogador 1 recebe **A** e Jogador 2 recebe **K**. A sequência é:

J1: passar → J2: passar.

Como **A** > **K**, o Jogador 1 vence o pote inicial (2 unidades), resultando em:

$$u = +1.$$

Se o Jogador 1 tivesse apostado e o Jogador 2 desistido, sua utilidade seria:

$$u' = +2.$$

O arrependimento contrafactual é:

$$R_3(\text{apostar}) = (2 - 1) \cdot 0,5 = 0,5. \quad (3.6)$$

Arrependimento acumulado:

$$R^{\text{cum}}(\text{apostar}) = 1,5 + 0,5 = 2,0.$$

Atualização de Estratégia

Pelo método de *regret matching*, a probabilidade de escolher cada ação é proporcional apenas aos arrependimentos positivos:

$$P(\text{apostar}) = \frac{R^{\text{cum}}(\text{apostar})}{R^{\text{cum}}(\text{apostar}) + R^{\text{cum}}(\text{passar})} = \frac{2,0}{2,0 + 0} = 1.$$

$$P(\text{passar}) = 0.$$

Estratégia Média

Ao longo de diversas iterações (por exemplo, 1000), calcula-se a estratégia média. Se a ação *apostar* foi escolhida 750 vezes, então:

$$\sigma(\text{apostar}) = \frac{750}{1000} = 0,75, \quad \sigma(\text{passar}) = 0,25. \quad (3.7)$$

Com o acúmulo das estratégias médias em toda a árvore do jogo, o CFR converge para o equilíbrio de Nash do Kuhn Poker.

3.4.5 Resumo do CFR

O CFR é um algoritmo iterativo da Teoria dos Jogos que minimiza o arrependimento contrafactual — o ganho perdido por não escolher uma ação alternativa em pontos de decisão — para aproximar o Equilíbrio de Nash em jogos de informação imperfeita como o pôquer (ZINKEVICH et al., 2007; SUTTON; BARTO, 2018). Em cada iteração de auto-jogo, o CFR calcula o arrependimento de ações, atualizando estratégias via *regret matching*, priorizando ações com arrependimento positivo (HART; MAS-COLELL, 2000). Com o tempo, estabiliza-se em uma estratégia onde mudanças unilaterais não trazem benefícios (NASH, 1951). O valor contrafactual, que avalia sub-jogos como se o jogador os alcançasse com certeza, permite eficiência computacional, viabilizando sua aplicação em jogos complexos como o *Texas Hold'em* (BROWN; SANDHOLM, 2018; MORAVČÍK et al., 2017). Assim, o CFR conecta Teoria dos Jogos e aprendizado de máquina, oferecendo estratégias robustas em incerteza, sendo central neste trabalho (RUSSELL; NORVIG, 2013).

3.4.6 Minimização de Arrependimento Contrafactual

O arrependimento contrafactual de não jogar a ação $a \in A(I)$ em I é a diferença na utilidade esperada de jogar a em I versus jogar $\sigma(I)$ em I , ponderada pela probabilidade dos oponentes de alcançar I (ZINKEVICH et al., 2007). O CFR minimiza o arrependimento contrafactual máximo (sobre todas as ações) em todo conjunto de informações.

Seja Z_I o subconjunto de histórias terminais onde um prefixo dessa história está em I ; para $z \in Z_I$, denote $z[I]$ como esse prefixo. Defina o valor contrafactual como:

$$v_i(\sigma, I) = \sum_{z \in Z_I} \pi_{-i}^\sigma(z[I]) \pi^\sigma(z[I], z) u_i(z) \quad (3.8)$$

que é a soma do pagamento esperado por ter alcançado I ponderado pela probabilidade do oponente de alcançar I .

O arrependimento contrafactual imediato para um conjunto de informações específico é definido como:

$$R_{i,imm}^T(I) = \frac{1}{T} \max_{a \in A(I)} \sum_{t=1}^T \left(v_i(\sigma_{I \rightarrow a}^t, I) - v_i(\sigma^t, I) \right) \quad (3.9)$$

onde $\sigma_{I \rightarrow a}^t$ é idêntico a σ^t exceto que a é tomado em I . O algoritmo produz novas estratégias em cada iteração empregando um minimizador de arrependimento em cada conjunto de informações sobre os valores contrafactuais.

O algoritmo CFR, na etapa de tempo t , computa um valor de arrependimento:

$$r^t(I, a) = v_i(\sigma_{I \rightarrow a}^t, I) - v_i(\sigma^t, I) \quad (3.10)$$

para cada conjunto de informações I e cada ação $a \in A(I)$. Esse valor de arrependimento representa o quanto um jogador i se arrepende de jogar σ_i em vez de escolher jogar $\sigma_{I \rightarrow a}$.

Na minimização de arrependimento contrafactual, o *regret-matching* é aplicado em cada conjunto de informações para atualizar as estratégias. Define o arrependimento contrafactual cumulativo no conjunto de informações I para a ação a até o tempo T como

$$R_i^T(I, a) = \sum_{t=1}^T r^t(I, a). \quad (3.11)$$

Para produzir novas estratégias, o *regret-matching* normaliza as porções positivas do arrependimento acumulado para cada ação (HART; MAS-COLELL, 2000). Mais precisamente, a estratégia que o *regret-matching* produz em I para $a \in A(I)$ é(NELLER; LANCTOT, 2013):

$$\sigma^{T+1}(I, a) = \begin{cases} \frac{R_i^{T,+}(I, a)}{\sum_{a' \in A(I)} R_i^{T,+}(I, a')} & \text{se o denominador for positivo;} \\ \frac{1}{|A(I)|} & \text{caso contrário} \end{cases} \quad (3.12)$$

onde $R_i^{T,+}(I, a) = \max\{R_i^T(I, a), 0\}$. O *regret-matching* produz arrependimento sublinear como resultado de satisfazer as condições do teorema de acessibilidade de Blackwell (BLACKWELL, 1956). É tentador perguntar por que misturar sobre essas ações em vez de escolher de forma gulosa a ação com o maior arrependimento? (ZINKEVICH et al., 2007). também mostraram que minimizar o arrependimento contrafactual imediato também minimiza o arrependimento geral médio $\bar{R}^T$. Portanto, o perfil médio se aproxima de um equilíbrio de Nash ao longo do tempo no caso de jogos de soma zero.

O algoritmo é resumido no Algoritmo 1. Nos referimos a esse algoritmo como Vanilla CFR quando não há amostragem aplicada, nem mesmo nos nós de chance. Como visto em um exemplo do algoritmo 1.

Algoritmo 1 Minimização de Arrependimento Contrafactual (Vanilla CFR)

1. *Inicialize tabelas de arrependimento:* $\forall I, \forall a \in A(I), r_I[a] \leftarrow 0$.
2. *Inicialize tabelas de estratégia cumulativa:* $\forall I, \forall a \in A(I), s_I[a] \leftarrow 0$.
3. *Inicialize perfil inicial:* $\sigma_1(I, a) \leftarrow 1/|A(I)|$
- 4.
5. **Função** $CFR(h, i, t, \pi_1, \pi_2)$:
6. **Se** h **é final** **Então**
7. **Retorne** $u_i(h)$
8. **Caso Contrário** h **for um nó de chance** **Então**
9. $\pi'_1 \leftarrow \pi_1$ **se** $i = 1$ **Senão** $\sigma_c(h, a) \cdot \pi_1$

10. $\pi'_2 \leftarrow \pi_2$ **se** $i = 2$ **Senão** $\sigma_c(h, a) \cdot \pi_2$
11. $\text{Retorne } \sum_{a \in A(h)} \sigma_c(h, a) \text{ CFR}(ha, i, t, \pi'_1, \pi'_2)$
12. **Finaliza Se**
13. *Seja I o conjunto de informações contendo h .*

4 Metodologia

4.1 Ambiente de Desenvolvimento

Os experimentos foram conduzidos em um ambiente controlado, utilizando um computador com as seguintes especificações técnicas:

- **Processador:** 13th Gen Intel® Core™ i5-13400 (2,50 GHz);
- **Memória RAM:** 32,0 GB;
- **Placa de vídeo:** 9 GB VRAM;
- **Armazenamento:** 1,82 TB SSD.

*Para garantir a reprodutibilidade e o isolamento das dependências, foi configurado um ambiente virtual **Conda**, permitindo o controle preciso das versões das bibliotecas e a execução dos experimentos sem interferência de pacotes externos.*

4.2 Ferramentas Utilizadas

*O projeto foi desenvolvido integralmente em **Python 3.9**, devido à sua versatilidade e ao amplo ecossistema de bibliotecas voltadas à inteligência artificial e análise numérica.*

As principais bibliotecas empregadas foram:

- **NumPy:** utilizada para operações matemáticas vetorizadas e manipulação de matrizes de arrendimento e estratégias;
- **Scikit-learn (K-means):** aplicada para a realização da abstração de estados, agrupando mãos semelhantes e reduzindo a dimensionalidade do espaço de decisão;
- **Joblib:** responsável pelo armazenamento e recuperação eficiente de dados intermediários, como batches de treinamento e estratégias parciais.

*Além disso, foi implementada uma Application Programming Interface (API) de comunicação com o agente **Slumbot**, um dos bots de pôquer de código aberto mais reconhecidos, possibilitando que o agente treinado interagisse em tempo real contra uma estratégia externa de referência.*

4.3 Geração do Dataset

4.4 Geração do Conjunto de Dados

O conjunto de dados utilizado neste trabalho foi produzido por meio de simulações offline de mãos de pôquer, seguindo a mesma abordagem empregada em repositórios de referência na área, como apresentado por Gongsta (GONGSTA, 2024). Cada mão é amostrada de forma independente, incluindo a distribuição das cartas privadas, das cartas comunitárias e da sequência completa de ações observadas nas rodadas de apostas. Para cada estado visitado durante a simulação, são registradas informações essenciais, tais como a Expected Hand Strength (EHS) e sua respectiva abstração por meio de algoritmos de agrupamento, especificamente o k-means, que agrupa estados estrategicamente semelhantes.

O processo de geração é computacionalmente intensivo devido ao cálculo da equidade das mãos por meio de simulações e à aplicação da abstração em cada estado avaliado. Como resultado, cada lote de aproximadamente 50.000 mãos consumiu entre 4,5 e 6 horas de processamento na máquina utilizada neste estudo, valor compatível com o reportado em implementações acadêmicas recentes. Ao final, obtém-se um conjunto de dados estruturado, anonimizado e adequado para o treinamento do algoritmo CFR.

O total do treinamento foi de 1000000 (um milhão de mãos) e um total de 20 batches

4.5 Processo de Treinamento

O agente CFR foi treinado utilizando a versão vanilla (versão simples do algoritmo), com atualização iterativa dos arrependimentos e das estratégias médias acumuladas. Durante o treinamento, o algoritmo alternou entre as seguintes etapas:

- 1. Cálculo do arrependimento contrafactual de cada ação;*
- 2. Atualização das estratégias com base nos arrependimentos positivos acumulados;*
- 3. Normalização e armazenamento das estratégias médias para posterior avaliação.*

A convergência foi monitorada por meio da variação do valor esperado(Expected Value – EV) entre iterações consecutivas e pela estabilização das probabilidades de escolha das ações.

4.6 Resumo do Fluxo Experimental

- 1. Configuração do ambiente Conda e inicialização do agente;*
- 2. Geração do dataset com 1.000.000 de mãos;*

3. *Treinamento iterativo com batches de 50.000 (pré-flop);*
4. *Treinamento iterativo com batches de 50.000 (pós-flop);*
5. *Armazenamento intermediário via Joblib;*
6. *Avaliação contra heurísticas e Slumbot;*
7. *Coleta dos resultados e geração dos gráficos de desempenho.*

5 Resultados Experimentais

Após o treinamento, o agente CFR foi comparado com um conjunto de estratégias heurísticas pré-definidas, representando diferentes perfis de jogadores humanos.

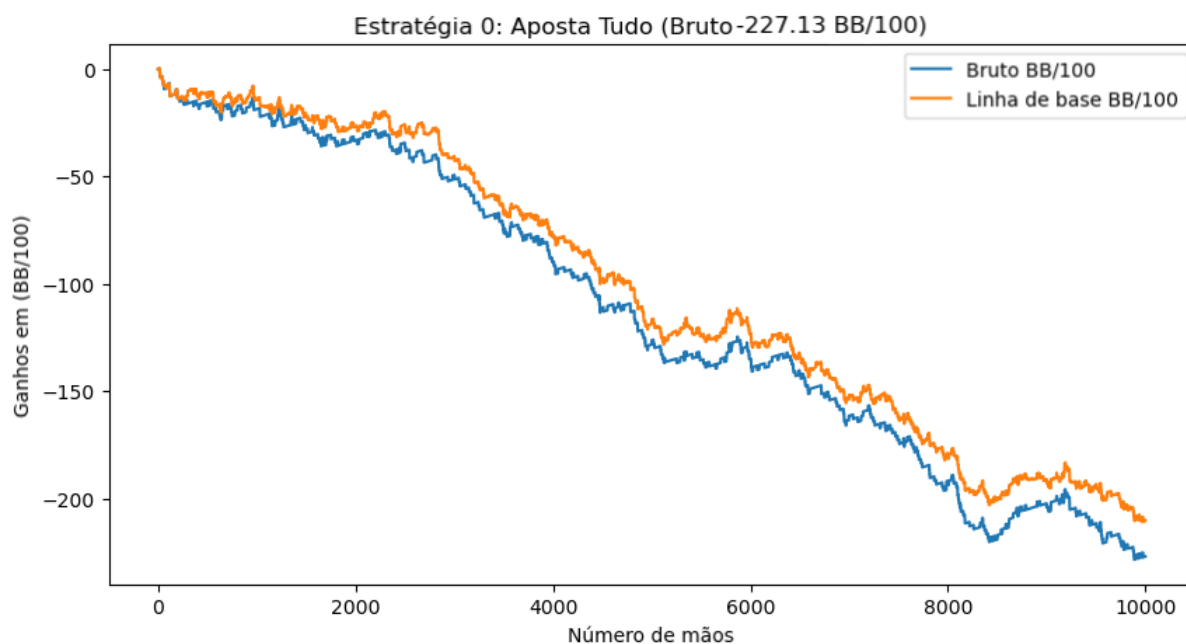
1. **Estratégia 0: Aposta Tudo** - Estratégia onde o jogador joga sempre de all-in.
2. **Estratégia 1: Mesa ou paga** - Também conhecido como Calling Station é o perfil de jogador que não aumenta a aposta e também não foge, apenas joga check/call
3. **Estratégia 2: Equidade Simples** - Ele aposta se possui equidade maior que 0.5 e check/fold caso contrário
4. **Estratégia 3: Equidade mais avançada** - Quanto maior a equidade, maior a probabilidade de apostar ou pagar.
5. **Estratégia 4: CFR com abstração**: Estratégia após o treinamento CFR, com diferentes tamanhos de Dataset.

As figuras 3,4,5..., ilustram simulações de jogos contra o Slumbot, e foi analisado uma amostragem de 10.000 mãos para cada estratégia. No eixo x tem-se o número de mãos e no eixo y tem-se os ganhos em BB/100 (quantos Big Blinds a cada 100 mãos o jogador ganha ou perde). O ideal é a estratégia chegar o mais próximo do 0, já que o Slumbot joga no equilíbrio de Nash. A linha de base é uma referência de comparação, usada para avaliar o desempenho do agente, nesse caso ele representa um modelo anterior no caso da estratégia após treinamento CFR e no caso das estratégias fixas (heurísticas) e a linha azul é o resultado sem tratamento.

5.1 Estratégia 0: Aposta Tudo (Bruto -227.12 BB/100)

Esta estratégia heurística apresenta o pior resultado, perdendo cerca de 227 BB/100. Isso sugere que jogar de all-in todas as mãos contra o Slumbot é altamente ineficaz. O desvio-padrão da linha Bruto foi equivalente a 98.89 BB/100 e o desvio-padrão da linha base foi equivalente a 100.93 BB/100. É possível notar que a linha azul (Bruto) não superou a linha base nenhuma vez, o que pode indicar que a estratégia Bruta não está performando melhor que o comportamento padrão. Também é possível notar que com o desvio-padrão desse tamanho, a variação do resultado BB/100 pode variar muito. Como visto na figura 3.

Figura 3 – Estratégia 0: Aposta Tudo em BB/100 em 10.000 mãos.

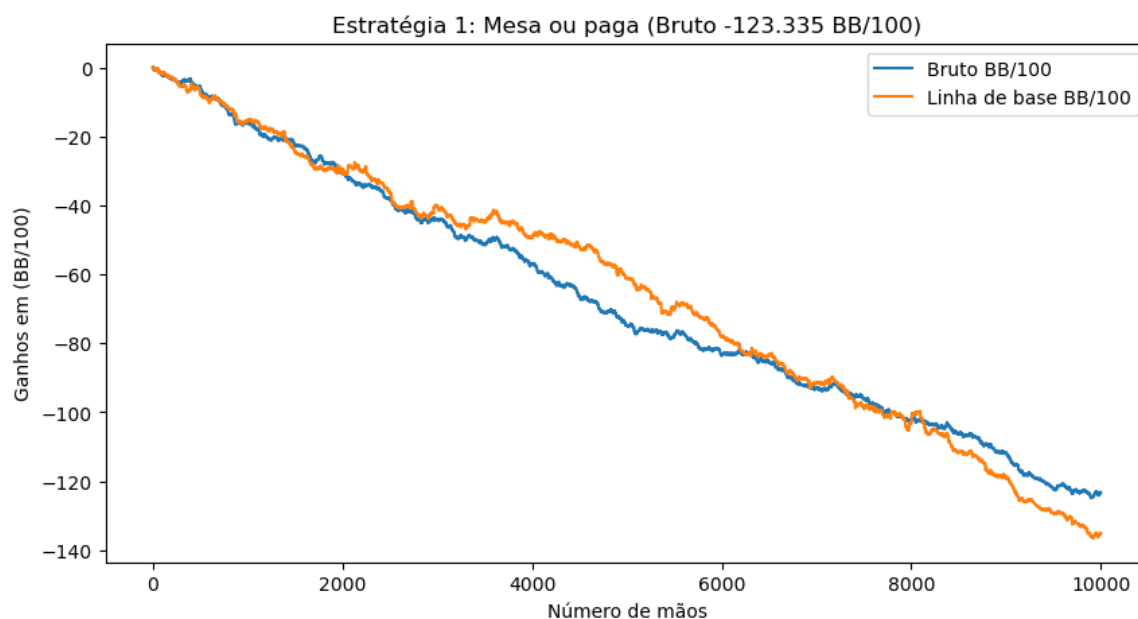


Fonte: o autor

5.2 Estratégia 1: Mesa ou paga (Bruto -123.335 BB/100)

Outra estratégia heurística, com desempenho melhor que a estratégia 0, mas ainda com perdas consideráveis (123 BB/100). A abordagem de check ou então call mostra uma melhoria mas continua sendo pouco competitiva. O desvio-padrão da linha Bruto foi equivalente a 35.35 BB/100 e o desvio-padrão da linha base foi equivalente a 37,69 BB/100 . É possível notar que a linha azul superou praticamente apenas uma vez a linha laranja (linha de base) o que pode indicar que a estratégia superou a linha base ou então isso faz parte da variância natural do jogo. Como visto na figura 4.

Figura 4 – Estratégia 1: Mesa ou paga em BB/100 em 10.000 mãos.

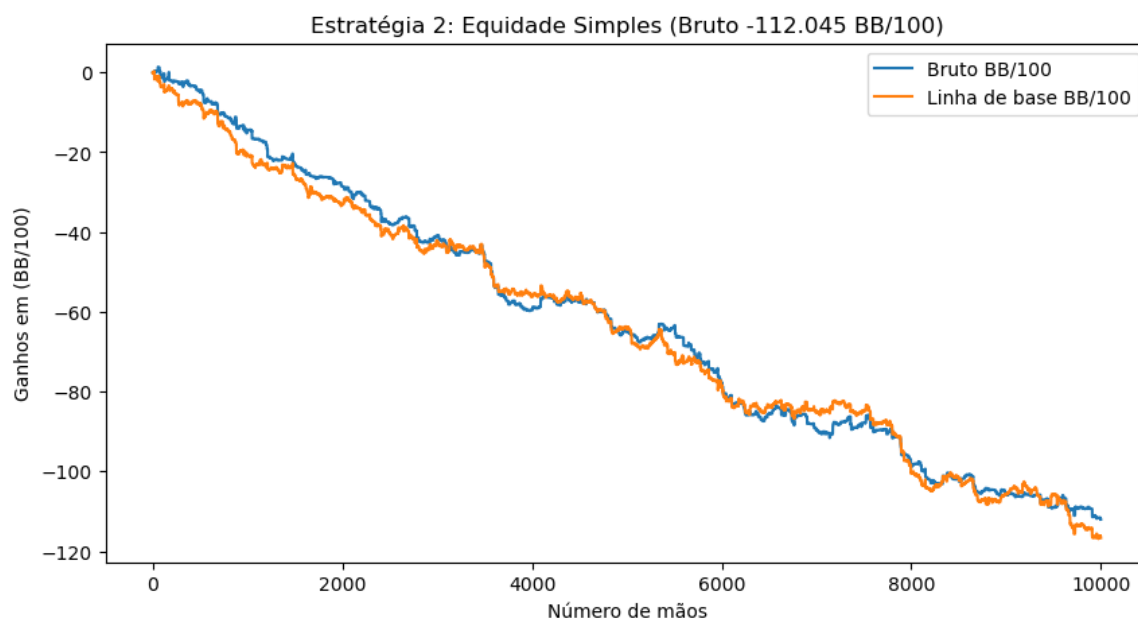


Fonte: o autor

5.3 Estratégia 2: Equidade Simples (Bruto -112.045 BB/100)

Esta heurística resulta em uma perda média de 112 BB/100. Apesar de ser a melhor entre as heurísticas, o desempenho permanece ruim, indicando que uma análise simples de equidade não é suficiente para superar o Slumbot. O desvio-padrão da linha Bruto foi equivalente a 32.72 BB/100 e o desvio-padrão da linha base foi equivalente a 31.54 BB/100. É possível notar que a linha bruta oscilou praticamente o tempo todo a linha de base, o que pode indicar que a estratégia praticamente se igualou a linha base. Como visto na figura 5.

Figura 5 – Estratégia 2: Equidade Simples em BB/100 em 10.000 mãos.

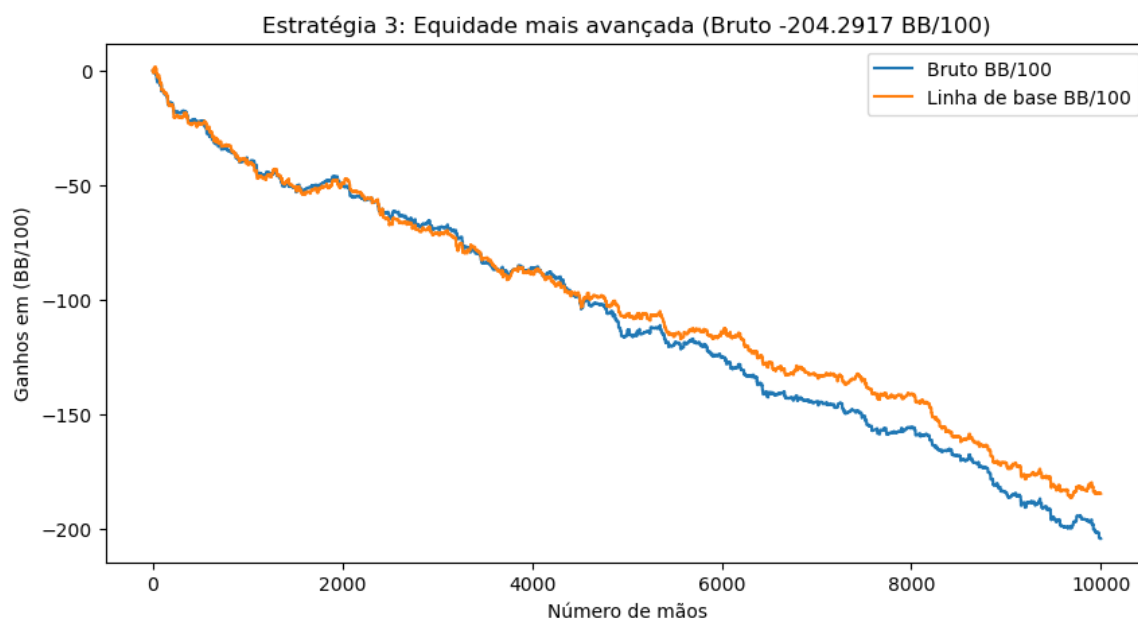


Fonte: o autor

5.4 Estratégia 3: Equidade mais avançada (Bruto -204,292 BB/100)

Apesar de ser uma heurística mais sofisticada, o resultado é uma perda de aproximadamente 204 BB/100. Isso sugere que a complexidade adicional não melhorou o desempenho e, na verdade, pode ter levado a decisões menos eficazes. O desvio-padrão da linha Bruto foi equivalente a 53.00 BB/100 e o desvio-padrão da linha base foi equivalente a 47.44 BB/100. É possível notar que a linha bruta novamente ficou praticamente abaixo da linha de base o que pode indicar que a estratégia é pior do que o esperado. Como visto na figura 6.

Figura 6 – Estratégia 3: Equidade mais avançada em BB/100 em de 10.000 mãos.



Fonte: o autor

5.5 Estratégia 4: CFR com abstração Dataset 250.000 (Bruto - 41.47 BB/100)

A Estratégia 4 CFR com abstração demonstrou uma boa diferença com relação às estratégias anteriores, com uma perda de 41.5 BB/100. É possível notar que no início até por volta da mão 2.300 o agente estava com ganhos positivos, algo até agora que não havia sido demonstrado nas outras estratégias. Como visto na figura 7.

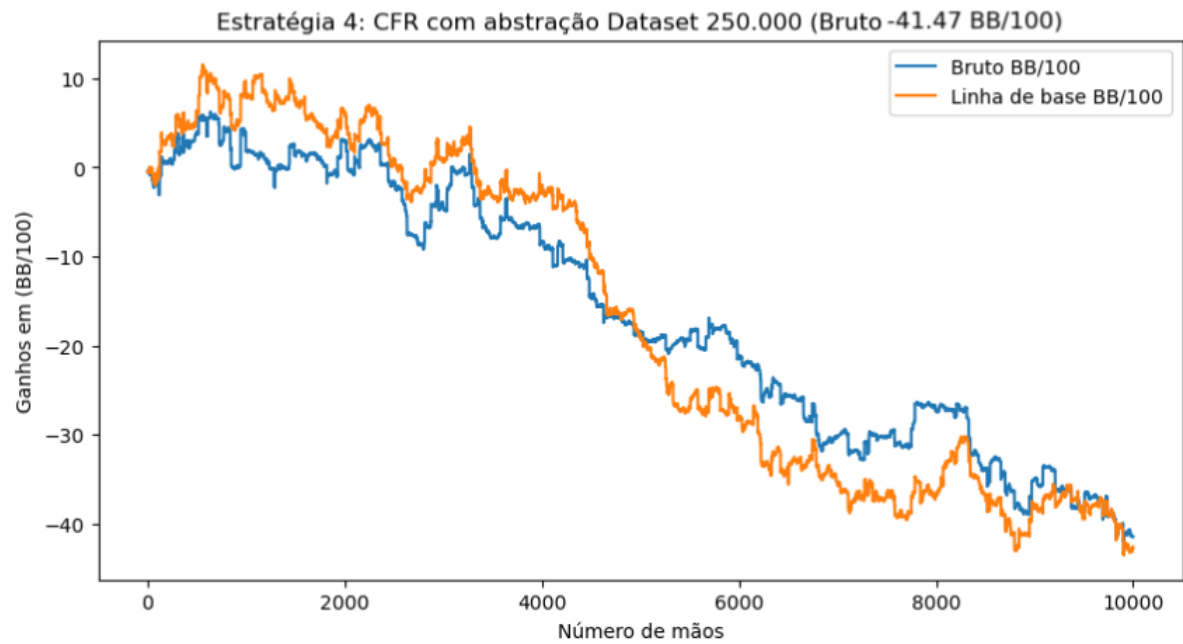


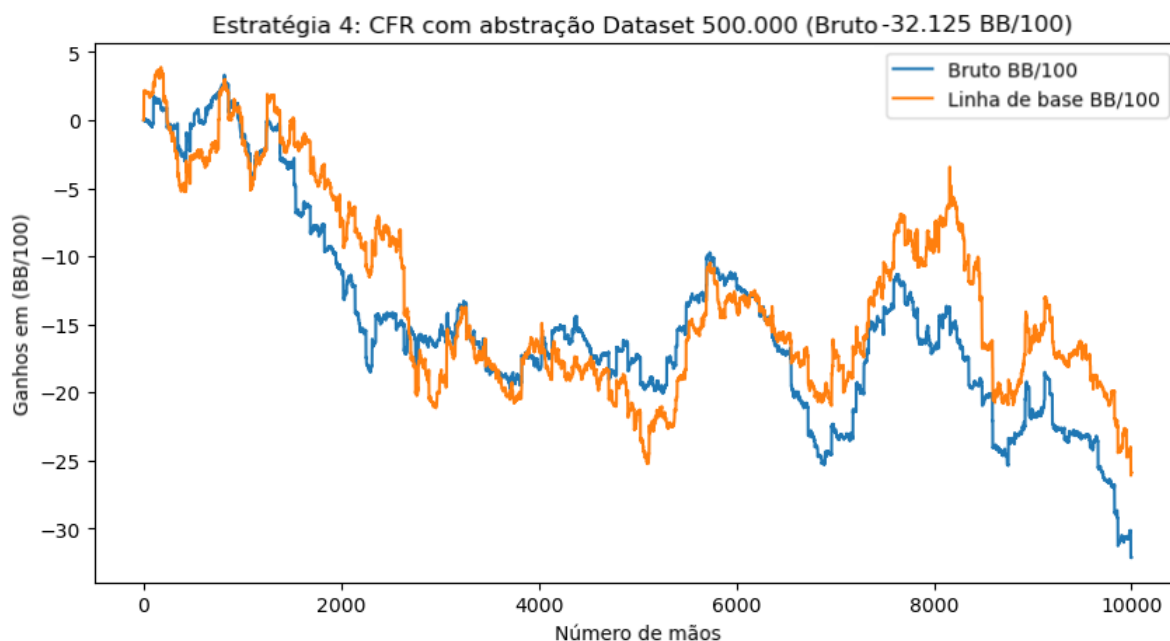
Figura 7 – Desempenho da estratégia CFR com abstração em termos de BB/100 ao longo de 10.000 mãos simuladas. Dataset = 250.000 mãos

Fonte: o autor

5.6 Estratégia 4: CFR com abstração Dataset 500.000 (Bruto - 32.125 BB/100)

Perda diminuiu para 32.125 BB/100, mostrando uma melhoria com um dataset maior, sugerindo que o modelo está convergindo para o equilíbrio. Como visto na figura 8.

Figura 8 – Desempenho da estratégia CFR com abstração em termos de BB/100 ao longo de 10.000 mãos simuladas. Dataset = 500.000 mãos

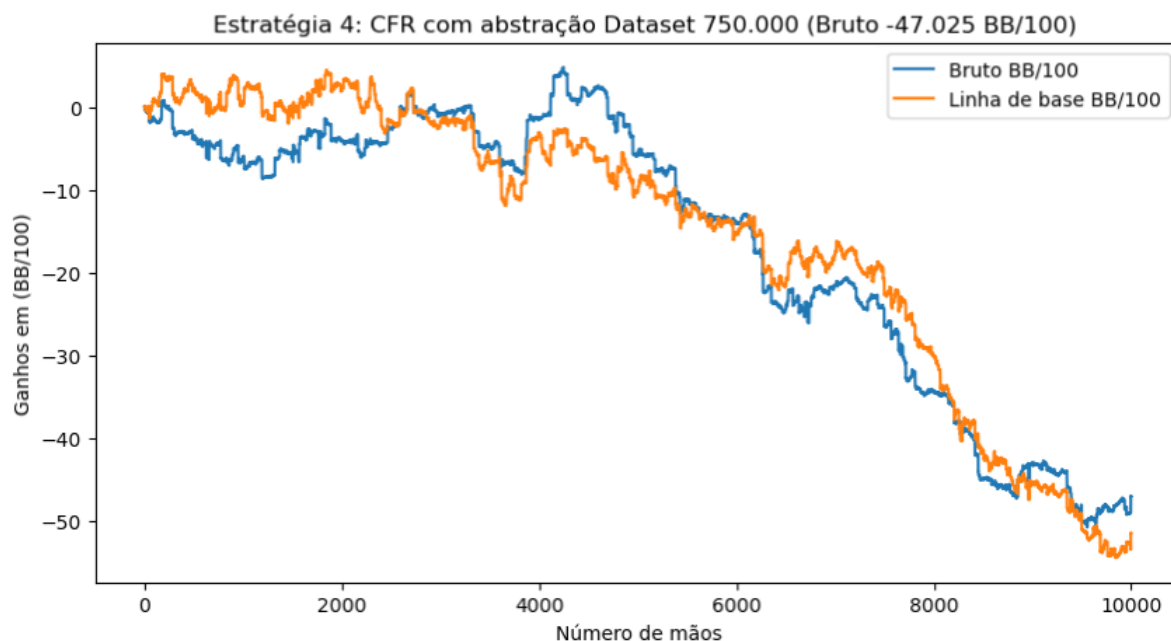


Fonte: o autor

5.7 Estratégia 4: CFR com abstração Dataset 750.000 (Bruto - 47.025 BB/100)

Perda aumentou para 47.025 BB/100, um leve retrocesso em relação ao dataset de 500.000, possivelmente devido à variância natural do pôquer. Podemos observar no gráfico que até por volta da mão 4400 ele oscilou o suficiente ficar com ganhos positivos por um breve momento contra o Slumbot. Como visto na figura 9.

Figura 9 – Desempenho da estratégia CFR com abstração em termos de BB/100 ao longo de 10.000 mãos simuladas. Dataset = 750.000 mãos

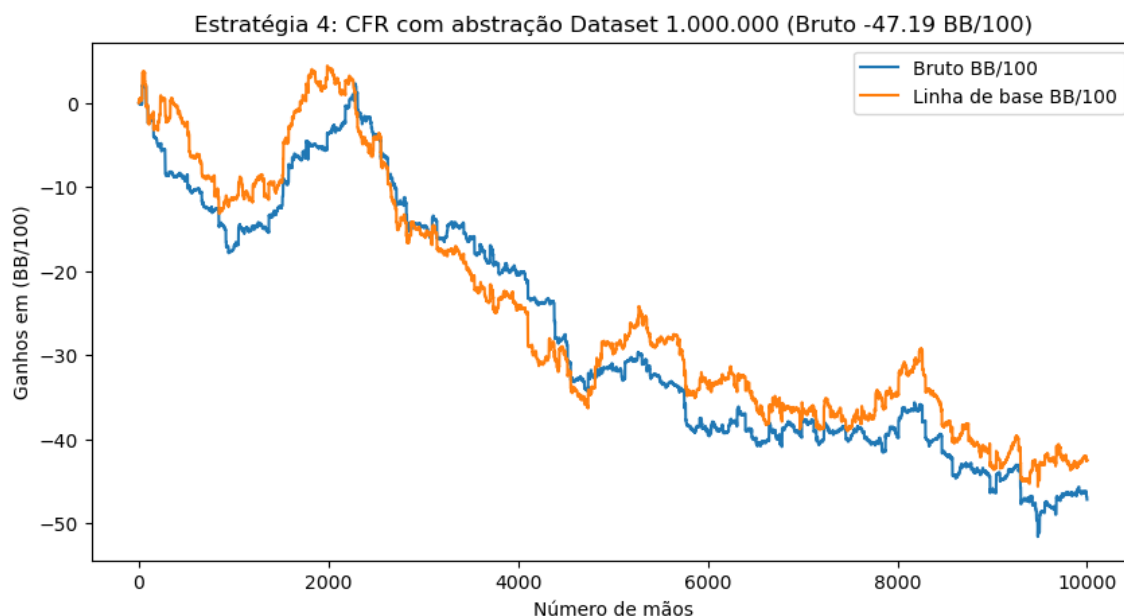


Fonte: o autor

5.8 Estratégia 4: CFR com abstração Dataset 1.000.000 (Bruto -47.19 BB/100)

Perda semelhante ao dataset de 750.000, indicando que o aumento para 1 milhão de mãos não trouxe uma melhoria significativa sobre o dataset 500.000. Como visto na figura 10.

Figura 10 – Desempenho da estratégia CFR com abstração em termos de BB/100 ao longo de 10.000 mãos simuladas. Dataset = 1.000.000 mãos



5.9 Análise Geral

As estratégias heurísticas (0 a 3) apresentaram perdas médias muito altas, variando de -112 a -227 BB/100, confirmando que abordagens baseadas em regras simples são ineficazes contra o Slumbot. O desvio-padrão de todas as estratégias heurísticas foram superiores ao desvio-padrão da estratégia 4, o que indica que a Estratégia 4 perde menor BB/100 para o Slumbot do que as estratégias (0 a 3) e isso contando com a variância do pôquer. A Estratégia 4, baseada em CFR, mostra uma redução significativa nas perdas médias de (-32,125 a - 47,19 BB/100) com desvio-padrão Bruto de 14,69 BB/100 enquanto o desvio-padrão da linha de base foi equivalente a 14,34, indicando que o treinamento com abstração é mais eficaz. O melhor resultado foi com o dataset de 500.000 (-32,125 BB/100), sugerindo um ponto ótimo no tamanho do dataset para este contexto. Contudo, como o resultado do dataset de 500.000 está dentro do desvio-padrão, não é possível concluir que esse é o tamanho de dataset ótimo ou se apenas fez parte da variância natural do pôquer. Assim seria necessário outro estudo ou outras técnicas mais avançadas de convergência para essa estratégia.

6 Conclusão

Este trabalho desenvolveu um agente de IA para o pôquer Heads-Up No-Limit Texas Hold'em, combinando o algoritmo CFR com técnicas de aprendizado de máquina para abstrações de cartas e ações. Os resultados mostram que a estratégia CFR superou abordagens heurísticas, apesar das flutuações causadas pela variância do jogo, reforçando a hipótese de que o Equilíbrio de Nash pode ser aproximado em cenários de informação imperfeita. A criação de um dataset de 1 milhão de mãos e o uso de abstração reduziram a complexidade computacional, viabilizando a execução em hardware modesto. O treinamento CFR destacou-se com o dataset de 500k, alcançando o melhor desempenho (-32.125 BB/100), porém esse resultado está dentro do desvio-padrão calculado (14,69 BB/100), o que pode indicar apenas a variância natural do jogo, embora as perdas persistam, indicando que o agente ainda não rivaliza com o Slumbot. As variações entre datasets (-32.125 a -47.19 BB/100) e a não linearidade dos resultados apontam para a necessidade de ajustes no treinamento, ou outra abordagem na implementação e otimização das variáveis. Contudo, é um resultado que abre portas para novas pessoas interessadas em desenvolver ou estudar jogos imperfeitos. Para o futuro, sugere-se explorar algoritmos como CFR+ ou Monte- Carlo CounterFactual Regret Minimization (MCCFR), além de aprimorar as abstrações, visando maior robustez e aplicabilidade em formatos multi-jogador, ou então outros jogos de informação imperfeita, como o boardgame Hanabi.

Referências

BLACKWELL, D. *An analog of the minimax theorem for vector payoffs*. Pacific Journal of Mathematics, v. 6, p. 1–8, 1956.

BROWN, N.; SANDHOLM, T. *Superhuman AI for heads-up no-limit poker: Libratus beats top professionals*. Science, v. 359, n. 6374, p. 418–424, 2018.

BROWN, N.; SANDHOLM, T. *Solving imperfect-information games via discounted regret minimization*. Proceedings of the AAAI Conference on Artificial Intelligence, v. 33, n. 01, p. 1829–1836, 2019.

BROWN, N.; SANDHOLM, T. *Superhuman AI for multiplayer poker*. Science, v. 365, n. 6456, p. 885–890, 2019.

BRUNSON, D. *Super System: A Course in Power Poker*. Las Vegas: Cardoza Publishing, 1978.

CARMICHAEL, T. *Artificial Intelligence in Poker: Theory and Applications*. Cambridge: MIT Press, 2018.

CARO, M. *Caro's Book of Poker Tells*. Las Vegas: Cardoza Publishing, 2003.

DALLA, N.; ALVAREZ, A. *One of a Kind: The Rise and Fall of Stuey 'The Kid' Ungar, the World's Greatest Poker Player*. New York: Atria Books, 2009.

GIPSYTEAM. Sequências de poker, venha conhecer de A a K! 2022. <<https://www.gipsyteam.com.br/poker/sequencias-poker>>. Acessado em 29 de outubro de 2025.

GONGSTA. *Poker-AI GitHub Repository*. 2024. Acesso em 27 de Outubro de 2025. Dispon

HART, S.; MAS-COLELL, A. *A simple adaptive procedure leading to correlated equilibrium*. Econometrica, v. 68, n. 5, p. 1127–1150, 2000.

MCMANUS, J. *Cowboys Full: The Story of Poker*. New York: Farrar, Straus and Giroux, 2009.

MORAVČÍK, M.; SCHMID, M.; BURCH, N.; LISÝ, V.; MORRILL, D.; BARD, N.; DAVIS, T.; WAUGH, K.; JOHANSON, M.; BOWLING, M. *Deepstack: Expert-level artificial intelligence in heads-up no-limit poker*. Science, v. 356, n. 6337, p. 508–513, 2017. Acesso em 27 de Outubro de 2025. Dispon

NASH, J. *Equilibrium points in n-person games*. Proceedings of the National Academy of Sciences, v. 36, p. 48–49, 1950. Acesso em 27 de Outubro de 2025. Disponível em: <<https://www.pnas.org/doi/abs/10.1073/pnas.36.1.48>>.

NASH, J. *Non-cooperative games*. Annals of Mathematics, v. 54, n. 2, p. 286–295, 1951.

NELLER, T. W.; LANCTOT, M. *An introduction to counterfactual regret minimization*. In: 2013 AAAI Spring Symposium on Computer Poker and Imperfect Information Games. Stanford University, CA: [s.n.], 2013. Acesso em 27 de Outubro de 2025. Disponível em: <<https://www.aaai.org/ocs/index.php/SSS/SSS13/paper/view/5790>>.

NEUMANN, J. von; MORGENTERN, O. *Theory of Games and Economic Behavior*. Princeton: Princeton University Press, 1944.

RUSSELL, S.; NORVIG, P. *Inteligência Artificial: Uma Abordagem Moderna*. 3. ed. São Paulo: Pearson, 2013.

SUTTON, R. S.; BARTO, A. G. *Reinforcement Learning: An Introduction*. 2. ed. Cambridge: MIT Press, 2018.

TEAM, S. *Slumbot AI Poker Agent*. 2020. Acesso em 27 de Outubro de 2025. Disponível em: <<https://www.slumbot.com>>.

ZINKEVICH, M.; JOHANSON, M.; BOWLING, M.; PICCIONE, C. Regret minimization in games with incomplete information. In: *Advances in Neural Information Processing Systems*. [s.n.], 2007. v. 20, p. 1729–1736. Acesso em 27 de Outubro de 2025. Disponível em: <<https://proceedings.neurips.cc/paper/2007/file/0afd7f3741c8f7b3b04ed3c3f2db729c-Paper.pdf>>.