

VWPI (Valorant Win Probability Index): um contraponto analítico às odds das casas de apostas



VWPI (Valorant Win Probability Index):
an analytical counterpoint to betting odds.

Renan Martins 

Fatec Santos
qs.zero@gmail.com

Andrey Watanuki 

Fatec Santos
andreysw123@gmail.com

Vitor Soares 

Fatec Santos
vitor33c50@gmail.com

Vagner dos Santos Macedo 

Fatec Santos
josepazetti@gmail.com

Guilherme Onorio 

Fatec Santos
guiopsilva@gmail.com

Revista Processando o Saber

eISSN 2179-5150 · Vol 18, n. 01, 2026
Multidisciplinar · DOI · Revisão por pares

Faculdade de Tecnologia Praia Grande – FATEC

Períodicidade: Anual
revista@fatecpg.edu.br

Recebido: Jan 2026

Aceito: Mar 2026

Publicado: Jun 2026

URL: <https://www.fatecpg.edu.br/revista/index.php/ps/article/view/462>

DOI: <https://doi.org/10.5281/zenodo.20076441>



RESUMO

Este estudo propõe o desenvolvimento de modelo híbrido de machine learning, utilizando os algoritmos XGBoost e Skellam, para a previsão de vitórias no jogo eletrônico Valorant. Diante da ascensão dos eSports e da viabilidade de abordagens data-driven, a pesquisa introduz o modelo VWPI (Valorant Win Probability Index). O objetivo central é estabelecer um indicador para mensurar a contribuição individual de cada jogador, conforme sua função específica, no cálculo da probabilidade de vitória da equipe. A metodologia baseou-se na análise de dados de 1.360 partidas, permitindo que o classificador estime o valor posicional e o impacto de cada integrante no resultado final. Os resultados demonstram que o modelo é capaz de superar as probabilidades estipuladas por bookmakers ao detectar variáveis de desempenho. Conclui-se que o modelo VWPI apresenta-se como uma ferramenta viável e robusta para a modelagem estatística e computacional de placares em competições.

PALAVRAS-CHAVE: Apostas; Ciência de Dados; Valorant; VWPI; XGBoost.

ABSTRACT

This study proposes the development of a hybrid machine learning model, based on the XGBoost and Skellam algorithms, for predicting victories in the electronic game Valorant. Given the rise of eSports and the feasibility of data-driven approaches, the research introduces the VWPI (Valorant Win Probability Index) model. The central objective is to establish an indicator to measure the individual contribution of each player, according to their specific role, in calculating the team's win probability. The methodology was based on the analysis of data from 1,360 matches, allowing the classifier to estimate positional value and the impact of each member on the final result. The results demonstrate that the model is capable of outperforming the odds set by bookmakers by detecting key performance variables. It is concluded that the VWPI model presents itself as a viable and robust tool for the statistical and computational modeling of scores in competitive gaming.

KEY-WORDS: Betting; Data Science; Valorant; VWPI, XGBoost.

INTRODUÇÃO

As apostas esportivas configuram-se como prática amplamente difundida na atualidade, marcada por significativo impacto social e econômico (Williams et al., 2017). Dada a contemporaneidade, o crescimento desse setor decorre da flexibilização regulatória, da expansão das plataformas digitais e do uso intensivo de estratégias de *marketing*, transformando um palpite casual em um ecossistema bruto entre o entretenimento e o ajuste financeiro (Russel et al., 2018).

Levando em consideração o crescimento no mercado global, o cenário de *betting* alcançou valor estimado em 243 bilhões de dólares em 2023, tendo o futebol como principal segmento (Ibisworld apud Torrance et al., 2023; Feng et al., 2017).

Contudo, trata-se de um modelo de negócio fundamentado em princípios matemáticos e estatísticos para as casas de apostas (*bookmakers*), no qual atuam como intermediárias, calculando probabilidades por meio de análises quantitativas e avaliações especializadas para reduzir seus riscos financeiros (Dmochowski, 2023), assumindo o papel de contraparte das apostas, baseando-se tanto em métodos estatísticos quanto em avaliações subjetivas de especialistas para minimizar erros de avaliações (Mandadapu, 2024; Terawong et al., 2024).

Nesse contexto, uma aposta é considerada justa quando o retorno líquido esperado é nulo, o que ocorreria se a *odd* correspondesse exatamente ao inverso da probabilidade real do evento. Contudo, são ajustadas sistematicamente essas probabilidades, de modo que as *odds* oferecidas incorporam margem de lucro, assegurando vantagem à instituição como se fosse a precificação de ativos do mercado financeiro (Jiménez, 2023; Walsh; Joshi, 2024).

Em um ambiente lúdico caracterizado pela presença de apostas, a busca por abordagens mais realistas e aderentes às dinâmicas dos jogos e aos modelos probabilísticos evidencia a necessidade de novas ferramentas, tanto para aprimorar a inteligência do jogo quanto para estimar os retornos das apostas com base nas *odds* definidas pelos *bookmakers*, vendo a ascensão dos esportes eletrônicos (*eSports*), representa uma transformação no cenário competitivo global (Mangat et al., 2024), consolidando-se como setor econômico relevante (Galekwa et al., 2024; Hing et al., 2023).

A compreensão sobre essa lógica para os *eSports* é essencial pelos desafios matemáticos, probabilísticos e computacionais que se apresentam. Sob uma abordagem data-driven, o novo indicador deve estimar a contribuição posicional e o valor efetivo do jogador para a vitória, utilizando dados granulares (eventos *in-game*) que capturem o impacto de ações específicas além de métricas estatísticas convencionais (Pelechrinis; Winston, 2018). A seleção

do modelo de *machine learning* deve priorizar a calibração (proximidade da probabilidade real), o que resulta em maior retorno sobre o investimento (ROI) do apostador (Walsh; Joshi, 2024).

Portanto, este estudo busca identificar ineficiências no mercado de apostas em *eSports* por meio de modelos de *machine learning*, visando detectar “apostas de valor” e superar as probabilidades das casas a partir de calibração de probabilidades e avaliação das métricas de desempenho propostas, como o classificador de jogadores capaz de estimar valor posicional e integração em equipes, apoiando previsões de desempenho com base em análises quantitativas e preditivas.

1. FUNDAMENTAÇÃO TEÓRICA

A expansão da tecnologia e da ciência de dados transformou a forma como o esporte e as apostas são compreendidos, resultando em uma integração crescente entre estatística, comportamento humano e economia digital. O uso de modelos probabilísticos para prever eventos esportivos surgiu da aplicação de princípios financeiros de precificação e gestão de risco, ganhando visibilidade no início dos anos 2000 e redefiniu a maneira de interpretar o desempenho em competições (Galekwa *et al.*, 2024).

Nos *eSports*, especialmente em títulos como *Valorant*, esse processo ainda é embrionário. Pesquisas sobre comportamento de apostadores mostram que o mercado digital tende a ser dominado por decisões influenciadas pela emoção e pela narrativa, e não por análises baseadas em dados concretos (Greer *et al.*, 2021).

Essa relação emocional com o risco é interpretada pela literatura como uma manifestação da chamada gambificação, fenômeno em que o entretenimento competitivo é entrelaçado com atividades de risco financeiro (Hing *et al.*, 2023). Dessa forma, o cenário evidencia como a aposta em *eSports*, apesar de integrar tecnologia, carece de métodos quantitativos e rigor analítico, favorecendo ineficiências cognitivas e decisões enviesadas.

Também é notável o crescimento do mercado global de apostas digitais, que movimenta bilhões de dólares e estimula inovação em algoritmos de precificação e plataformas interativas (Kim *et al.*, 2023).

A expansão do uso de dados nas apostas esportivas introduz riscos metodológicos, uma vez que maior volume informacional não garante aplicação estruturada, dificultando a padronização analítica. Nesse cenário, o emprego de *machine learning* em sistemas preditivos surge como alternativa técnica para elevar a acurácia e mitigar vieses humanos. Estudos

recentes indicam que a integração de variáveis históricas, contextuais e comportamentais amplia substancialmente a precisão em relação a abordagens tradicionais (Galekwa *et al.*, 2024).

O *Valorant Win Probability Index* (VWPI) propõe-se, portanto, a representar a materialização desse avanço, utilizando fundamentos estatísticos e computacionais para estruturar previsões objetivas e transparentes, alinhadas à evolução dos estudos científicos sobre apostas digitais.

1.1 A REVOLUÇÃO ANALÍTICA DE MATTHEW BENHAM E O PARADIGMA MONEYBALL

Matthew Benham, físico e analista financeiro, introduziu no futebol técnicas estatísticas e de gestão de risco inspiradas no mercado financeiro. Sua abordagem propôs tratar partidas e desempenhos como eventos probabilísticos, substituindo a intuição por modelagem matemática. Essa visão foi um marco na evolução do esporte, permitindo o uso de dados quantitativos para mensurar desempenho e prever resultados (Manoel, 2025).

O método de Benham se alinhou à crescente profissionalização da análise esportiva, que visava eliminar vieses cognitivos e subjetividades nas decisões (Rao; Hastie, 2022), compondo técnicas semelhantes às usadas em finanças e consolidando o futebol como campo no uso probabilístico de dados, bem como o cenário de apostas digitais e dos *eSports* (Greer *et al.*, 2021).

1.2 O MERCADO DE APOSTAS ESPORTIVAS COMO SISTEMA PROBABILÍSTICO E O CONCEITO DE COTAÇÕES JUSTAS

O mercado de apostas esportivas, conforme Galekwa *et al.* (2024), funciona como um sistema econômico baseado em probabilidades objetivas e nas leis da estatística aplicada à precificação e gestão de risco. Cada cotação (*odd*) representa matematicamente o preço de uma probabilidade implícita, definida pela seguinte relação:

$$P = \frac{1}{O}$$

onde:

- P é a probabilidade implícita;
- O é a cotação decimal (*odd*).

Por exemplo, uma *odd* de 2,50 reflete a probabilidade de 0,40, ou seja, 40% de chances de ocorrência do evento (Galekwa *et al.*, 2024). A partir dessa lógica, define-se a probabilidade verdadeira (*true odd*), que representa o equilíbrio entre risco e retorno em uma aposta neutra, calculada por:

$$P_r = \frac{1}{P(e)}$$

A cotação justa (*fair odd*) é, portanto, o valor matematicamente neutro de uma aposta, em que o valor esperado (*expected value*) é nulo. Esse valor pode ser expresso pela equação:

$$EV = (P_r \times G) - ((1 - P_r) \times C)$$

em que:

- EV é o valor esperado;
- P_r é a probabilidade real;
- G é o ganho potencial;
- C é o custo da aposta.

Quando $EV > 0$, identifica-se uma aposta de valor positivo (*value bet*), ou seja, uma oportunidade de vantagem estatística no longo prazo (Rao; Hastie, 2022).

Contudo, as casas de apostas adicionam uma margem operacional ao sistema, denominada *overround*, calculada pela soma das probabilidades implícitas de todos os resultados possíveis, que ultrapassa 100%:

$$O_v = \sum_{i=1}^n \frac{1}{O_i}$$

em que:

- O_i : odd (cotação) oferecida para cada resultado i ;
- n : número total de resultados possíveis (por exemplo, vitória, empate e derrota);
- O_v : resultado da soma das probabilidades implícitas ($\frac{1}{O_i}$) de todos os resultados.

Pode-se exemplificar:

$$\text{Vitória da Equipe A: } O_{t1} = \frac{1}{2,50} = 0,40 \text{ (40\%)}$$

$$\text{Empate: } O_{t2} = \frac{1}{3,20} = 0,3125 \text{ (31,25\%)}$$

$$\text{Vitória da Equipe B: } O_{t3} = \frac{1}{2,80} = 0,3571 \text{ (35,71\%)},$$

A soma dessas probabilidades é $0,40 + 0,3125 + 0,3571 = 1,0696$ (ou seja, 106,96%). O excesso de 6,96% representa o *overround* (Ov), a margem de lucro que garante retorno à casa, assegurando lucro esperado sobre qualquer resultado (Alves de Freitas, 2025).

Esse princípio é essencial para o cálculo de apostas de valor positivo, que ocorrem quando a probabilidade estimada pelo modelo analítico do apostador supera a probabilidade implícita na cotação ofertada:

$$P(e) \text{ modelo} > \frac{1}{O_m}$$

Dessa forma, compreender e estimar corretamente as *true odds* permite construir modelos preditivos de alta precisão e detectar ineficiências nos mercados de apostas.

Esse sistema aproxima as apostas esportivas da lógica dos mercados financeiros, em que se busca eficiência informacional - todas as informações disponíveis refletem-se no preço das *odds*. Contudo, estudos apontam que o comportamento humano viola essa eficiência através de vieses cognitivos, como a falácia do apostador, a tendência de superestimar sequências de sucesso e o excesso de confiança (Greer *et al.*, 2021).

Tais distorções criam ineficiências exploráveis por modelos preditivos e técnicas de *machine learning*, capazes de estimar probabilidades reais de forma mais precisa que a intuição

humana. Assim, o mercado de apostas pode ser interpretado como um sistema probabilístico dinâmico, em que o conhecimento estatístico e a modelagem de dados se tornam instrumentos centrais de vantagem competitiva e precisão preditiva (Kim *et al.*, 2023).

1.3 A GAMBLIFICAÇÃO DOS ESPORTS E SUAS IMPLICAÇÕES

A gamblificação dos *eSports* pode ser compreendida como o processo pelo qual elementos típicos do jogo de azar são incorporados ao ecossistema dos jogos competitivos digitais. Nesse contexto, práticas originalmente voltadas ao entretenimento passam a incluir mecanismos de aposta, risco e recompensa financeira, aproximando-se de dinâmicas especulativas. Essa transformação ocorre por meio de diferentes modalidades, como as apostas em partidas de *eSports*, o *skin betting* — no qual itens virtuais dos jogos são utilizados como moeda de aposta — e o *token wagering*, que envolve o uso de criptomoedas ou ativos digitais (Kim *et al.*, 2023). Assim, a experiência do jogador deixa de ser exclusivamente recreativa e passa a integrar lógicas econômicas baseadas em incerteza, expectativa de ganho e exposição ao risco (Mattinen; Macey; Hamari, 2023).

Pesquisas empíricas identificam que apostadores em *eSports* apresentam maior vulnerabilidade a comportamentos problemáticos em comparação aos apostadores esportivos tradicionais, especialmente entre jovens adultos com alta imersão em ambientes digitais (Greer *et al.*, 2021).

O mercado global de apostas em *eSports* consolidou-se como um dos segmentos mais dinâmicos da indústria do entretenimento digital, movimentando, em 2025, aproximadamente US\$ 12,66 bilhões e com projeções de alcançar US\$ 20,39 bilhões até 2029, impulsionado por inovações tecnológicas, integração transnacional e expansão das plataformas de *betting* digitais (Research and Markets, 2025).

A estrutura desse mercado abrange apostas pré-jogo e em tempo real, baseadas em fluxos contínuos de dados e na interação com o público, integrando algoritmos probabilísticos e previsões fundamentadas em dados históricos e contextuais. Essa abordagem apoia-se em modelos preditivos análogos à gestão de portfólios financeiros, empregando regressão logística e *random forest* para estimar valores esperados e otimizar retornos (Lessmann *et al.*, 2010).

No caso específico de *Valorant*, a consolidação do título como disciplina de *eSports* competitivos desde 2020 catalisou a criação de mercados paralelos de apostas, especialmente pela presença de torneios licenciados e forte audiência digital. As plataformas oferecem linhas

de apostas para eventos, utilizando métricas de desempenho individual e de equipes para precificação dinâmica das *odds* a partir de aprendizado supervisionado e redes neurais para detectar padrões táticos e probabilísticos com elevada acurácia (Terawong, Cliff, 2024).

1.4 CIÊNCIA DE DADOS E MODELAGEM ESTATÍSTICA EM APOSTAS ESPORTIVAS

A ciência de dados aplicada às apostas esportivas fundamenta-se na combinação entre mineração de dados (*data mining*), estatística inferencial e técnicas de aprendizado supervisionado, visando à predição probabilística de resultados e à identificação de ineficiências de mercado. Modelos como regressão logística, *random forest* e *gradient boosting* são empregados para estimar probabilidades condicionais de vitória, derrota ou empate, analisando atributos de equipes, jogadores, contexto de jogo e histórico de desempenho (Galekwa *et al.*, 2024).

A síntese teórica dessas abordagens revela forte analogia com modelos econométricos de precificação de ativos, nos quais o princípio de maximização de retorno esperado é ajustado pelo risco (Matej *et al.*, 2021).

O *data mining* permite a extração de padrões ocultos em grandes volumes de dados esportivos, integrando fontes heterogêneas como estatísticas de desempenho, variáveis contextuais (clima, local e arbitragem) e até sentimento público medido por redes sociais. Técnicas como *Support Vector Machines* (SVM), *random forests* e Redes Neurais Artificiais oferecem robustez na modelagem de sistemas não lineares e altamente correlacionados, favorecendo análises dinâmicas de precificação de *odds* e gerenciamento de risco. Contudo, essas ferramentas enfrentam restrições associadas à qualidade dos dados, presença de ruído, vieses amostrais e volatilidade contextual, que limitam a generalização das previsões (Galekwa *et al.*, 2024).

Os modelos supervisionados são otimizados por métricas como *Area Under the Curve* (AUC), *Brier Score* e *Rank Probability Score* (RPS), mas a acurácia também depende de ajustes de calibração e da capacidade de atualização contínua a partir de dados ao vivo. O aprendizado adaptativo, por meio de algoritmos como *XGBoost* e *LSTM* (*Long Short-Term Memory*), tem sido testado na predição esportiva em tempo real, com destaque para o uso de arquiteturas híbridas que combinam aprendizado supervisionado e técnicas bayesianas de atualização. Ainda que esses modelos apresentem desempenho superior aos métodos estatísticos

tradicionais, persistem desafios éticos e metodológicos, sobretudo na transparência algorítmica e na gestão de predição (Thabtah *et al.*, 2019).

Portanto, a interseção entre estatística inferencial e inteligência artificial redefine a lógica das previsões em apostas esportivas. Ao transformar o ruído estocástico em informação estruturada, a modelagem baseada em *machine learning* não apenas amplia a eficiência dos mercados de apostas, mas também impõe a necessidade de regulação técnica e epistemológica sobre o uso de dados massivos em contextos de risco e lucro.

1.5 VALORANT

Valorant é um jogo digital do gênero *First-Person Shooter* (FPS), desenvolvido pela *Riot Games* em 2020, inserido no cenário de *eSports*. O título combina mecânicas clássicas de tiro tático em equipe com habilidades exclusivas atribuídas a personagens jogáveis denominados agentes, integrando desempenho individual e estratégia coletiva. Sua estrutura competitiva e a mensuração de métricas de performance têm fundamentado pesquisas em ciência de dados, inteligência artificial e engenharia de software, ampliando investigações em estatística aplicada, psicométrica e análise de partidas (Riot Games, 2026).

No modo competitivo, duas equipes compostas por cinco jogadores alternam-se entre ataque e defesa ao longo de até 25 rodadas no formato “*plant/defuse*”, vencendo quem alcançar 13 rodadas. A equipe ofensiva deve instalar o artefato denominado “*spike*” em áreas específicas do mapa, enquanto a defensora busca impedir a ação ou desativá-lo. Embora inspirado em modelos tradicionais do gênero, o jogo distingue-se pela exigência de uso estratégico das habilidades singulares de cada agente, além da proficiência no manuseio de armamentos (Riot Games, 2026).

Os agentes distribuem-se em quatro funções: Duelistas, focados em confrontos diretos; controladores, responsáveis por modular áreas do mapa; Sentinelas, voltados à contenção e defesa; e Iniciadores, especializados em geração de informação e criação de oportunidades táticas. O balanceamento dessas funções é investigado por meio de métricas como taxa de vitória e frequência de escolha, bem como por algoritmos de agrupamento aplicados à análise de interações e dinâmicas coletivas (Riot Games, 2026).

1.5.1 Estatística e Métricas de Desempenho

A mensuração do desempenho em *Valorant* permeia diversas métricas, tanto individuais quanto coletivas. A *Kills/Deaths/Assists (KDA)* representa a relação entre abates, mortes e assistências, fundamental para caracterizar o impacto do jogador; o *Average Combat Score (ACS)* funciona como indicador composto que agrega contribuições por dano, eliminações, assistências e impacto tático; e o percentual de *headshot* mede a precisão e eficiência das ações ofensivas.

Além dessas métricas fundamentais, outras variáveis com elevada relevância para predição de resultados incluem:

- a) ***kd_diff_total (Diferença de Kills-Deaths)***: Captura o desempenho relativo do jogador em relação às quedas, representando o saldo líquido de eliminações. Essa métrica isolada remove a interferência das assistências, focando especificamente na capacidade de derrotar adversários sem ser eliminado;
- b) ***rating_total (Rating Total)***: Sintetiza a contribuição geral do jogador através de uma métrica composta que pondera múltiplos fatores de desempenho. O *rating* total normalmente incorpora eliminações, dano infligido, objetivos completados e sobrevivência, oferecendo uma visão holística do impacto do jogador na partida;
- c) ***KAST (Kill, Assist, Survival, Trade)***: Indicador que avalia a participação ativa em rodadas, medindo a proporção de rodadas nas quais o jogador obteve pelo menos uma das quatro ações: eliminou um inimigo, assistiu uma eliminação, sobreviveu à rodada ou trocou sua morte por um ou mais inimigos. Essa métrica é particularmente relevante para avaliar consistência e engajamento tático;
- d) ***econ_rating (Economic Rating)***: Reflete o impacto das decisões econômicas do jogador, considerando o desempenho em relação aos recursos investidos em armas e utilidades. Valor elevado indica eficiência ao converter gastos econômicos em resultados ofensivos ou defensivos;
- e) ***first_kills_total (Eliminações Iniciais)***: Determinam o curso tático da rodada, pois uma eliminação inicial desequilibra numericamente o confronto e oferece vantagem estratégica significativa. Essa métrica é crítica para entender a capacidade do jogador em criar oportunidades no início dos engajamentos;
- f) ***first_deaths_total (Mortes Iniciais)***: Igualmente determinantes para o curso tático da rodada, uma morte inicial coloca o time em desvantagem numérica imediata. O registro permite avaliar vulnerabilidades posicionais e decisões táticas arriscadas;
- g) ***multikill_2k, multikill_3k, multikill_4k, multikill_5k (Multikills)***: Indicadores de potência ofensiva que medem a capacidade do jogador de eliminar múltiplos adversários em sequência rápida, representando momentos de dominância individual e que resultam em viradas de rodada ou consolidação de vantagens;
- h) ***ADR (Average Damage Rating)***: Quantifica o dano médio infligido por rodada, independentemente de resultar em eliminações. Essa métrica é especialmente relevante para avaliar a eficiência do posicionamento, escolha de armas e precisão, oferecendo uma medida contínua de contribuição ofensiva mesmo em rodadas sem eliminações confirmadas;

- i) **plants (Plantios de Spike)**: Ação objetiva realizada pelo time atacante. Registra quantas vezes o jogador instalou o spike, refletindo participação em objetivos primários da equipe e exposição a risco tático ao realizar ação crítica;
- j) **defuses (Desativações de Spike)**: Ação objetiva realizada pelo time defensor. Registra quantas vezes o jogador realizou a desativação do *spike*, indicando participação crucial em defesa de objetivo e frequentemente exigindo tomada de decisão estratégica sob pressão.

Tais indicadores são utilizados em análises comparativas e ajustes de modelos preditivos em *machine learning* e inteligência artificial. Estudos recentes têm recorrido a conjuntos de dados extraídos de ligas profissionais, como a *Valorant Champions Tour* (VCT), para estudar o impacto das decisões táticas na dinâmica da partida. Análises por *cluster*, métricas de desempenho coletivo e modelos de predição de resultados que são capazes de capturar nuances que métodos agregados simples não identificam.

Adicionalmente, o gerenciamento econômico, a seleção de agentes e as composições de equipe configuram variáveis centrais em modelos matemáticos de predição e balanceamento competitivo. Embora situações de *clutch* (inferioridade numérica) possuam relevância estratégica, sua baixa variância discriminativa em grandes amostras indica que a performance individual tende a ser melhor representada por métricas compostas, como KAST e *rating* total.

2. PROCEDIMENTOS METODOLÓGICOS

Trata-se de um estudo descritivo, exploratório e quantitativo-transversal, baseado em técnicas computacionais de extração, processamento e análise de dados aplicadas à previsão de resultados em partidas de *eSports*, mais especificamente, partidas de *Valorant*. O objetivo da pesquisa foi estruturar uma arquitetura de coleta, atualização e modelagem de dados que possibilitasse identificar padrões de desempenho e estimar a probabilidade de vitória ou derrota das equipes, utilizando informações estatísticas das partidas registradas na base *vlr.gg*.

O ambiente de execução foi configurado em uma infraestrutura virtualizada via Docker, integrando a plataforma *Coolify* e uma REST API responsável pela obtenção sistemática dos dados. Essa API foi composta por endpoints específicos: `/match`, que retorna todas as partidas disponíveis, e `/match/:id`, que fornece os detalhes estatísticos, incluindo indicadores de desempenho dos jogadores, equipes e seus respectivos `vlr_id`.

A automação da coleta foi conduzida pelo Apache Airflow, que operou como orquestrador de fluxos de atualização em intervalos de 24 horas, garantindo a integridade dos

registros temporais. Os dados foram armazenados em um banco PostgreSQL, estruturado em formato relacional, com dados consolidados de jogadores, equipes e resultados de partidas.

Na etapa analítica, utilizou-se *Jupyter Notebook* para a condução da análise exploratória de dados (EDA) e subsequente aplicação de modelos preditivos. Inicialmente, os dados foram tratados para cálculo de métricas agregadas com base nas dez últimas partidas dos jogadores de cada equipe, representando o desempenho recente e a forma atual.

Para a modelagem preditiva, aplicou-se o *Extreme Gradient Boosting (XGBoost)* a fim de estimar o peso estatístico de cada posição de jogador no desempenho coletivo. Em seguida, utilizou-se o algoritmo *Skellam*, adaptado para prever resultados, modelando a diferença de pontuações médias entre as equipes. A combinação entre os modelos permitiu gerar probabilidades de vitória com maior precisão e consistência, considerando os aspectos individuais e coletivos das equipes.

Essa metodologia possibilitou a construção de um sistema capaz de prever resultados de forma objetiva e probabilística, propondo práticas contemporâneas de inteligência artificial aplicada à análise de desempenho esportivo, com potencial de contribuição em previsão estatística e gestão de risco em contextos digitais.

Abaixo, serão descritas as principais ferramentas e tecnologias empregadas, bem como suas respectivas funções dentro da arquitetura computacional utilizada na pesquisa.

a) Coleta e Estruturação de Dados

Fundamenta-se nos princípios da engenharia de dados e no modelo de integração Extração, Transformação e Carga (*ETL*), conforme delineado por Kimball (2008) e Inmon (2005), que ressaltam a importância da padronização e automação dos fluxos de dados para análises sistemáticas e reproduzíveis.

A coleta foi realizada por meio de *endpoints* REST, */match* e */match/:id*, responsáveis, respectivamente, por retornar as partidas disponíveis e detalhar estatísticas individuais de cada jogador. O sistema foi orquestrado pelo *Apache Airflow*, que coordenou o fluxo ETL em intervalos de 24 horas e armazenou os dados processados em um banco *PostgreSQL*. A virtualização foi conduzida em *Docker*, empregando o *Coolify* para gerenciar os serviços, garantindo reprodutibilidade e escalabilidade ao processo.

b) Análise Exploratória de Dados (EDA)

A análise exploratória de dados constitui etapa fundamental para a extração sistemática de informações, possibilitando a identificação de padrões e modelos plausíveis a serem posteriormente submetidos à análise confirmatória ou à inferência estatística. Esse processo envolve abordagens sistemáticas tanto na coleta de observações empíricas quanto na análise dos dados, contemplando descrições quantitativas e qualitativas e a delimitação das inter-relações entre as propriedades do fenômeno investigado (Medri, 2011; Marconi; Lakatos, 2023).

A análise exploratória foi aplicada com o uso das bibliotecas *Pandas*, *NumPy*, *Matplotlib* e *Seaborn* no *Python*, permitindo a identificação de padrões, distribuição de variáveis e tratamento de dados ausentes (*missing values*). Esses processos subsidiam a preparação dos dados para a etapa preditiva, garantindo a consistência e qualidade das observações estatísticas extraídas.

c) Análise Preditiva e Modelagem Estatística

A aplicação da Análise Preditiva possibilita a identificação de padrões nos dados, permitindo prever desfechos e direcionamentos futuros. Essa abordagem capacita as empresas e instituições a agirem de forma proativa diante das circunstâncias e a assumirem uma posição estratégica em relação ao seu crescimento e evolução a longo prazo (Turban, 2017 apud Soares, 2018).

Neste trabalho, a análise preditiva foi orientada por modelos de aprendizado supervisionado, utilizando o algoritmo *XGBoost* como técnica principal de previsão. Com base nas dez últimas partidas de cada equipe, o modelo calculou pesos estatísticos para as variáveis de desempenho, gerando uma previsão binária de vitória ou derrota.

d) Otimização de Hiperparâmetros com *Optuna*

O processo de otimização de hiperparâmetros consiste na busca sistemática pelos melhores valores de parâmetros que maximizem o desempenho de um modelo de aprendizado de máquina (Hutter, Kotthoff & Vanschoren, 2019). Um modelo otimizado tende a apresentar melhor capacidade de generalização, reduzindo o erro de validação e evitando tanto o *underfitting* quanto o *overfitting*.

Neste estudo, empregou-se a ferramenta *Optuna*, uma biblioteca de código aberto em *Python* desenvolvida por Akiba *et al.* (2019), baseada no paradigma de otimização bayesiana sequencial (*Sequential Model-Based Optimization – SMBO*). O *Optuna* conduz a otimização por meio de busca adaptativa, ajustando os hiperparâmetros conforme o histórico dos resultados anteriores — diferentemente de métodos tradicionais, como *grid search* e *random search*, que realizam explorações cegas do espaço de busca.

No contexto da pesquisa, o *Optuna* foi aplicado para otimizar os parâmetros do modelo *XGBoost*, especificamente o número de estimadores, a profundidade máxima das árvores, a taxa de aprendizado e o parâmetro de regularização.

A interação entre o *Optuna* e o *XGBoost* permitiu obter um modelo mais eficiente, ajustado aos padrões estatísticos e computacionais do conjunto de dados.

e) Aprendizado de Máquina com *XGBoost*

O *XGBoost* é uma técnica de aprendizado supervisionado baseada no método de construção aditiva de árvores de decisão, em que cada nova árvore é treinada para corrigir os erros do modelo anterior (Chen & Guestrin, 2016). Essa abordagem de *boosting* permite combinar múltiplos modelos fracos para formar um modelo robusto, com maior capacidade preditiva e menor viés estatístico.

Do ponto de vista teórico, o *XGBoost* se destaca pelo uso de funções de custo regularizadas, o que controla a complexidade do modelo e reduz o risco de sobreajuste. Além disso, a técnica implementa otimizações no uso da memória e do paralelismo, favorecendo o alto desempenho em grandes volumes de dados (Bishop, 2006). A seguir, a fórmula do modelo do *XGBoost*:

O valor predito pelo *XGBoost* (\hat{y}_i) é a soma das saídas de K árvores de decisão treinadas sequencialmente:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i)$$

Onde:

- \hat{y}_i : Valor predito para o dado i . Representa a previsão ou saída final do modelo para a amostra i .
- K : Número total de árvores (na iteração final do modelo). Em algoritmos como *XGBoost*, são as árvores construídas durante o processo de *boosting*.
- $f_k(x_i)$: Saída da k -ésima árvore para o dado i . Cada f_k é uma árvore de decisão que faz uma previsão parcial para x_i ; a soma dessas previsões gera a predição final.

A função objetivo do *XGBoost* é composta por dois termos: função de perda e regularização:

$$obj(\theta) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k)$$

onde:

- $obj(\theta)$: Função objetivo a ser minimizada pelo modelo durante o treinamento.
- $\sum_{i=1}^n l(y_i, \hat{y}_i)$: Soma da função de perda para todas as amostras do conjunto de dados.
- $l(y_i, \hat{y}_i)$: Função de perda que mede o erro entre o valor verdadeiro (y_i) e o valor predito \hat{y}_i , podendo ser o erro quadrático (regressão) ou *log-loss* (classificação).
- $\sum_{k=1}^K \Omega(f_k)$: Termo de regularização que penaliza a complexidade das árvores construídas.
- $\Omega(f_k)$: Termo de regularização para a k -ésima árvore, projetado para evitar o ajuste excessivo (*overfitting*) ao penalizar árvores muito complexas.

Na prática deste estudo, o *XGBoost* foi utilizado para prever o resultado das partidas com base nas dez últimas estatísticas dos jogadores de cada equipe. As variáveis de entrada englobam métricas de dano médio, eficiência de abate, impacto tático e consistência estatística.

O modelo foi implementado em *Python* por meio da biblioteca *xgboost*, integrando-se ao *Scikit-learn* para rotinas de avaliação cruzada, cálculo de métricas de desempenho e validação do modelo otimizado via *Optuna*.

f) Modelo Estatístico de *Skellam*

A distribuição *Skellam* é um modelo estatístico que descreve a diferença entre duas variáveis aleatórias independentes de distribuição *Poisson* (Skellam, 1946). Em contextos esportivos, essa distribuição é frequentemente aplicada para calcular probabilidades de vitória ou derrota, considerando o número esperado de pontos, rounds de cada equipe (Karlis & Ntzoufras, 2003).

Teoricamente, esse modelo permite representar o resultado líquido de um confronto, ajustando a probabilidade de vitória com base na diferença esperada de desempenho das equipes.

A fórmula da função de massa de Probabilidade (PMF) é descrita abaixo:

$$P(K = k) = e^{-(\mu_1 + \mu_2)} \left(\frac{\mu_1}{\mu_2} \right)^{k/2} I_{|k|} (2\sqrt{\mu_1 \mu_2})$$

onde:

- $P(K=k)$: Probabilidade de a diferença observada entre dois eventos Poisson ser igual a k .
- k : Diferença observada ($K=N_1-N_2$), ou seja, o resultado dos eventos (como coeficiente ponderado dos dois times).
- μ_1, μ_2 : Médias das distribuições Poisson (por exemplo, médias dos coeficientes do *XGBoost* ponderados dos times).
- $I_{|k|}$: Função de Bessel modificada de primeira espécie e ordem k , uma função especial usada na estatística de distribuições discretas.
- $\left(\frac{\mu_1}{\mu_2} \right)^{k/2}$: Fator de normalização exponencial.
- $I_{|k|} (2\sqrt{\mu_1 \mu_2})$: Avaliação da função de Bessel no ponto $2\sqrt{\mu_1 \mu_2}$.

Portanto, para prever o resultado final (vitória, empate ou derrota), somam-se as probabilidades para os valores positivos de k (indicando vitória), valores negativos de k (indicando derrota) e para $k=0$ (indicando empate). Isso é feito porque a distribuição *Skellam* estima diretamente as probabilidades de cada diferença de gols ou pontos entre as equipes, permitindo calcular as chances de cada resultado por meio da soma desses eventos (Alves de Freitas, 2025).

Neste estudo, o algoritmo de *Skellam* foi implementado na biblioteca *SciPy*, sendo parametrizado com as médias de desempenho geradas pelas equipes nas suas últimas dez partidas. O modelo retornou à probabilidade de vitória e derrota considerando o equilíbrio entre os adversários, funcionando como complemento ao *XGBoost* para transformar previsões estatísticas em probabilidades diretas aplicáveis às partidas.

A metodologia descrita integra, de forma sistemática, processos de engenharia de dados, análise exploratória, modelagem estatística e aprendizado de máquina. Essa combinação permitiu a construção de um modelo analítico completo, capaz de transformar dados brutos de desempenho esportivo em previsões estruturadas de resultados, com base em séries temporais curtas e variáveis individuais de performance.

Os métodos de coleta, tratamento e otimização, apoiados em ferramentas como *Python*, *Airflow*, *Optuna*, *XGBoost* e *Skellam*, proporcionam um ambiente analítico reproduzível e alinhado às práticas contemporâneas de ciência de dados aplicada ao esporte eletrônico.

As atualizações constantes aplicadas no *Valorant* não afetam o algoritmo construído por conta da utilização de métricas baseadas em posicionamento médio da equipe a partir das funções dos jogadores e estatísticas graduais para a elaboração das probabilidades. Dessa maneira, a adaptação para novas ferramentas e funções *in-game*, bem como a adoção de novos cenários, não afeta diretamente o VWPI.

Dessa forma, os resultados obtidos a partir dessa arquitetura metodológica são apresentados e discutidos na próxima seção, no qual, são examinadas as evidências empíricas e o desempenho estatístico do modelo proposto.

3. ANÁLISE E RESULTADO

A presente análise teve como objetivo principal aplicar técnicas de aprendizado de máquina, especificamente o algoritmo *XGBoost*, para a modelagem preditiva de resultados em partidas de *eSports*, utilizando dados detalhados de desempenho individual dos jogadores. Os

dados compreendem aproximadamente 1.360 partidas e possuem um conjunto de 57 variáveis relacionadas à performance dos atletas em partidas. Entre as variáveis analisadas destacam-se estatísticas de combate, como número de *kills*, mortes, assistências, bem como métricas avançadas, incluindo diferencial de *kills* e mortes (*K/D diff*), participação em *rounds* decisivos (*KAST*), dano médio por *round* (*ADR*), percentual de tiros na cabeça (*HS%*), além de indicadores de *multikills*, *clutch plays* e métricas econômicas (*econ_rating*, números de *plants* e *defuses*).

Para assegurar que as análises capturassem a dinâmica recente de desempenho, foram calculadas as médias móveis das dez partidas anteriores para cada jogador, criando um panorama temporal que destaca tendências atuais ao invés de dados acumulados históricos, segundo cada função na equipe, para que a modelagem segmentada reflita as particularidades de cada ação.

Modelos preditivos de classificação binária foram treinados separadamente para cada papel, com a finalidade de prever o resultado no próximo jogo. Os dados foram divididos em conjunto de treinamento (70%) e de teste (30%), mantendo a proporcionalidade das classes para que o modelo aprendesse adequadamente a distinguir ambas as situações. A validação dos modelos ocorreu por meio da análise das métricas de precisão, revocação, *F1-score* e acurácia geral, oferecendo uma visão equilibrada da capacidade preditiva para vitórias e derrotas.

A performance obtida foi consistente e robusta com acurácia geral variando entre 80% para Duelistas e 83% para Sentinelas. Nos demais papéis, Controladores e Iniciadores apresentaram acurácia de 82%, indicando eficácia adequada nas previsões para todas as funções analisadas. O equilíbrio entre as métricas de precisão e revocação corrobora a eficácia dos modelos em capturar as duas classes de resultados, minimizando viés para uma simples predição.

A análise das importâncias relativas das variáveis para cada papel revelou padrões coerentes com o contexto tático dos jogadores. Para Duelistas, métricas diretamente relacionadas ao impacto ofensivo e sobrevivência, como o diferencial de *K/D* e taxas de *deaths*, foram as mais relevantes, destacando a importância do desempenho individual na capacidade ofensiva e resistência. Para as funções de suporte (Controlador, Iniciador e Sentinela), as métricas ligadas à sobrevivência e impacto no conjunto de rounds (*deaths_total* e *KAST*) assumiram maior peso, refletindo a importância estratégica para a estabilidade coletiva. Outros atributos como assistências, *kills* e métricas econômicas também contribuíram significativamente para a predição, porém em menor escala.

Além da avaliação preditiva individual, os coeficientes gerados para cada jogador foram agregados para compor um índice preditivo médio para as equipes. A distribuição de *Skellam* foi utilizada como modelo probabilístico para estimar as chances de vitória em partidas futuras, considerando os coeficientes agregados dos times. As probabilidades estimadas mostraram-se condizentes com o cenário competitivo atual, como no exemplo dos confrontos de FNATIC *versus* NRG e DRX contra Paper Rex, onde os valores previstos indicaram probabilidades realistas de vitória em torno de 55% para FNATIC e 62% para Paper Rex, respectivamente.

O coeficiente preditivo individual também permitiu a identificação do desempenho previsto de atletas-chave em cada função, possibilitando uma avaliação objetiva e quantitativa sobre a influência potencial dos jogadores no resultado coletivo. Essa abordagem pode fornecer ferramentas úteis para equipes técnicas na avaliação de jogadores e para o desenvolvimento de estratégias, além de contribuir para análises de mercado relacionadas à apostas e *scouting*.

Em síntese, a aplicação do *XGBoost* demonstrou-se eficiente para modelagem preditiva em esportes eletrônicos utilizando dados complexos de desempenho individual, alcançando altas taxas de acurácia e produzindo *insights* precisos sobre a importância relativa dos fatores de performance para cada função tática. A combinação de estatísticas descritivas detalhadas com modelagem supervisionada e técnicas probabilísticas para previsão de resultados constitui uma metodologia viável e replicável para análises futuras, desde que atualizada continuamente para acompanhar alterações no *metagaming* competitivo.

Limitações importantes foram observadas, como a dependência da correta classificação das funções dos jogadores e da representatividade das amostras utilizadas, além da necessidade de constante atualização do modelo para manter a aderência ao ambiente dinâmico dos *eSports*. Ressalta-se que a metodologia proposta favorece análises e aplicações em contexto competitivo, constituindo uma contribuição relevante para o avanço do uso de ciência de dados e *machine learning* na esfera dos esportes eletrônicos.

4. CONSIDERAÇÕES FINAIS

A mensuração das diversas métricas de desempenho em *Valorant* revela a complexidade e a riqueza da análise quantitativa aplicada a jogos eletrônicos competitivos.

A multiplicidade de indicadores, que vão desde medidas tradicionais como KDA e ACS até métricas situacionais como confrontos iniciais, *multikills*, e ações objetivas como plantios e desativações de *spike*, mostra que o desempenho do jogador e o sucesso coletivo dependem

tanto de habilidades individuais quanto de estratégias integradas e decisões táticas ao longo da partida.

Dito isso, a aplicação dessas métricas em modelos preditivos de *machine learning* e inteligência artificial contribui para uma melhor compreensão do impacto de variáveis técnicas e estratégicas, permitindo o avanço na análise e no desenvolvimento de treinamentos e estratégias em *eSports*.

Assim, o estudo das métricas de desempenho no *Valorant* não apenas aprimora a competitividade do jogo, mas também serve como base para pesquisas interdisciplinares em ciência de dados, psicologia do esporte e tecnologia digital, promovendo inovação e eficácia no campo emergente dos esportes eletrônicos.

Essa convergência entre análise estatística rigorosa e dinâmica de jogos destaca a importância de indicadores precisos e contextualizados para capturar nuances do desempenho em ambientes em tempo real, revelando que o sucesso no *Valorant* é uma síntese complexa entre habilidade, estratégia e gerenciamento de recursos.

Além disso, a incorporação de métricas como KAST e *rating* total evidencia que processos compostos e interdependentes são mais indicativos da eficácia do jogador do que métricas isoladas, como *clutches* individuais, consolidando assim um entendimento mais profundo de desempenho em *eSports* competitivos.

Os próximos passos consistem em utilizar dados provenientes diretamente da desenvolvedora do jogo, a *Riot Games*, para garantir maior precisão e abrangência das informações analisadas.

Ademais, será realizada uma análise mais específica e detalhada focada na montagem ideal de equipes, identificando os melhores jogadores para cada função dentro do time, bem como os encaixes mais adequados para preencher vagas de jogadores faltantes.

Esse estudo busca aprofundar a compreensão das sinergias entre agentes, habilidades individuais e estratégias coletivas, permitindo comparar a eficácia de diferentes combinações de jogadores e identificar perfis ideais para compor equipes competitivas.

Até o momento, houve satisfação com os resultados obtidos pelo *Valorant Win Probability Index* (VWPI) pela sua eficiência em captar nuances de desempenho em partidas. Essa continuidade permitirá aprimorar os modelos preditivos e expandir as práticas em treinamentos, composição de times e estratégias no âmbito dos *eSports*, potencializando o impacto e a aplicabilidade do estudo na área.

REFERÊNCIAS

- AKIBA, T. et al. **Optuna**: A Next-generation Hyperparameter Optimization Framework. [s. l.], 2019. Disponível em: <https://arxiv.org/pdf/1907.10902>. Acesso em: 14 dez. 2025.
- ALVES DE FREITAS, A. A.; DOS SANTOS, A. **Probabilistic model based on the Skellam distribution**: application in sports betting. *Brazilian Journal of Biometrics*, [s. l.], v. 43, n. 3, e-43773, 2025. Disponível em: <https://doi.org/10.28951/bjb.v43i3.773>. Acesso em: 14 dez. 2025.
- BISHOP, C. **Pattern Recognition and machine learning**. *Journal of Electronic Imaging*, [s. l.], v. 16, n. 4, p. 049901, 2007. Disponível em: <https://www.researchgate.net/publication/221995544>. Acesso em: 14 dez. 2025.
- CHEN, T.; GUESTRIN, C. **XGBoost**: A Scalable Tree Boosting System. In: *ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING*, 22., 2016, [s. l.]. Proceedings [...]. [s. l.]: ACM, 2016. Disponível em: <https://arxiv.org/pdf/1603.02754>. Acesso em: 14 nov. 2025.
- FENG, Guanhao; POLSON, Nicholas; XU, Jianeng. **The Market for English Premier League (EPL) Odds**. *Booth School of Business, University of Chicago*, [s. l.], p. 1-25, jan. 2017. Disponível em: <https://arxiv.org/pdf/1604.03614>. Acesso em: 15 nov. 2025.
- GALEKWA, René Manassé et al. **A systematic review of machine learning in sports betting**: techniques, challenges and future directions. *arXiv preprint*, [s. l.], 28 out. 2024. Disponível em: <https://arxiv.org/abs/2410.21484>. Acesso em: 13 dez. 2025.
- GREER, N. et al. **Are eSports bettors a new generation of harmed gamblers?** A comparison with sports bettors on gambling involvement, problems, and harm. *Journal of Behavioral Addictions*, [s. l.], v. 10, n. 3, p. 435-446, 2021. Disponível em: <https://akjournals.com/view/journals/2006/10/3/article-p435.xml>. Acesso em: 22 out. 2025.
- HING, Nerilee. **A bad bet for sports fans**: the case for ending the “gambification” of sport audiences. *Sport Management Review*, [s. l.], v. 26, n. 5, p. 788-812, 2023. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/14413523.2023.2260079>. Acesso em: 22 out. 2025.
- HUTTER, Frank; KOTTHOFF, Lars; VANSCHOREN, Joaquin (ed.). **Automated Machine Learning**: Methods, Systems, Challenges. Cham: Springer International Publishing, 2019. Disponível em: <https://link.springer.com/book/10.1007/978-3-030-05318-5>. Acesso em: 14 dez. 2025.
- INMON, W. H. **Building the data warehouse**. Indianapolis: Wiley Pub, 2005.
- JIMÉNEZ, V. et al. **Sports Betting**: an application of neural networks and modern portfolio theory to the English Premier League. [s. l.], 2023. Disponível em: <https://arxiv.org/abs/2307.13807>. Acesso em: 14 dez. 2025.
- KARLIS, D.; NTZOUFRAS, I. **Analysis of sports data by using bivariate Poisson models**. *Journal of the Royal Statistical Society: Series D (The Statistician)*, [s. l.], v. 52, n. 3, p. 381–393, 2003. Disponível em: http://www2.stat-athens.aueb.gr/~jbn/papers2/08_Karlis_Ntzoufras_2003_RSSD.pdf. Acesso em: 11 dez. 2025.

KIM, Hyoun S. et al. **A scoping review of the association between loot boxes, eSports, skin betting, and token wagering with gambling and video gaming behaviours.** *Journal of Behavioral Addictions*, [s. l.], v. 12, n. 2, p. 309-351, 2023. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/34419935/>. Acesso em: 22 out. 2025.

KIMBALL, Ralph et al. **The Data Warehouse Lifecycle Toolkit.** 2. ed. [s. l.]: Wiley, 2008. Disponível em: <https://www.kimballgroup.com/data-warehouse-business-intelligence-resources/books/data-warehouse-dw-lifecycle-toolkit/>. Acesso em: 15 nov. 2025.

LESSMANN, Stefan; SUNG, Ming-Chien; JOHNSON, Johnnie E.V. **Alternative methods of predicting competitive events: An application in horserace betting markets.** *International Journal of Forecasting*, [s. l.], v. 26, n. 3, p. 518-536, 2010. Acesso em: 13 out. 2025.

MANDADAPU, Purnachandra. **The Evolution of Football Betting: A Machine Learning Approach to Match Outcome Forecasting and Bookmaker Odds Estimation.** arXiv preprint, [s. l.], 2024. Disponível em: <https://arxiv.org/abs/2403.16282>. Acesso em: 13 out. 2025.

MANGAT, Harshdeep S. et al. **Understanding Esports-related Betting and Gambling: A Systematic Review of the Literature.** *Journal of Gambling Studies*, [s. l.], v. 40, n. 2, p. 893-914, 2024. Disponível em: <https://doi.org/10.1007/s10899-023-10256-5>. Acesso em: 10 nov. 2025.

MANOEL, H. **Matthew Benham: Dono do Brentford fez história no mundo das apostas.** *MKT Esportivo*, maio 2025. Disponível em: <https://www.mktesportivo.com/2025/05/matthew-benham-dono-do-brentford-fez-historia-no-mundo-das-apostas>. Acesso em: 16 set. 2025.

MARCONI, Marina de Andrade; LAKATOS, Eva Maria. **Técnicas de pesquisa: planejamento e execução de pesquisa, amostragens e técnicas de pesquisa, elaboração, análise e interpretação de dados.** 9. ed. Rio de Janeiro: Atlas, 2021.

MATEJ, U. et al. **Optimal sports betting strategies in practice: an experimental review.** *IMA Journal of Management Mathematics*, [s. l.], v. 32, n. 4, p. 465-489, 2021. Disponível em: <https://arxiv.org/abs/2107.08827>. Acesso em: 11 dez. 2025.

MATHESON, Victor. **An Overview of the Economics of Sports Gambling and an Introduction to the Symposium.** *Eastern Economic Journal*, [s. l.], v. 47, n. 1, p. 1-8, 2021. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC7780080/>. Acesso em: 15 nov. 2025.

MATTINEN, Topias; MACEY, Joseph; HAMARI, Juho. **Gambling in digital games and esports: a scoping review.** Tampere: Tampere University, 2022. Disponível em: <https://trepo.tuni.fi/bitstream/handle/10024/208524/0383.pdf>. Acesso em: 22 out. 2025.

MEDRI, Waldir. **Análise Exploratória de Dados.** Londrina: UEL, 2011. 82 f. Apostila (Especialização em Estatística). Disponível em: <https://docs.ufpr.br/~benitoag/apostilamedri.pdf>. Acesso em: 12 dez. 2025.

PELECHRINIS, Konstantinos; WINSTON, Wayne. **A Skellam Regression Model for Quantifying Positional Value in Soccer.** arXiv preprint, [s. l.], jul. 2018. Disponível em: <https://arxiv.org/pdf/1807.07536>. Acesso em: 21 out. 2025.

RAO, K.; HASTIE, R. **Predicting Outcomes in a Sequence of Binary Events: Belief Updating and Gambler's Fallacy Reasoning.** *Cognitive Science*, [s. l.], v. 47, n. 1, e13211, jan. 2023. DOI: 10.1111/cogs.13211. Acesso em: 10 dez. 2025.

RESEARCH AND MARKETS. **Sports Betting Market Report 2025.** [s. l.]: Research and Markets, mar. 2025. Disponível em: <https://www.researchandmarkets.com/reports/5939406/sports-betting-market-report>. Acesso em: 15 nov. 2025.

RIOT GAMES. **Agentes do Valorant.** Valorant, 2026. Disponível em: <https://playvalorant.com/pt-br/agents/>. Acesso em: 20 mar. 2026.

RIOT GAMES. **Guia do Premier.** Valorant Support, 2026. Disponível em: <https://support-valorant.riotgames.com/hc/pt-br/articles/42089963186451-Guia-do-Premier>. Acesso em: 20 mar. 2026.

SKELLAM, J. G. **The Frequency Distribution of the Difference Between Two Poisson Variates Belonging to Different Populations.** *Journal of the Royal Statistical Society*, [s. l.], v. 109, n. 3, p. 296, 1946. Disponível em: [suspicious link removed]. Acesso em: 14 dez. 2025.

SOARES, Josiel Ferreira. **Análise Preditiva: Importância e Vantagens para os Negócios.** [s. l.], 2018. Disponível em: <https://repositorio-api.animaeducacao.com.br/server/api/core/bitstreams/b78507ee-3666-4c90-ad73-c9b693d96cf7/content>. Acesso em: 15 dez. 2025.

TERAWONG, Chawin; CLIFF, Dave. **XGBoost Learning of Dynamic Wager Placement for In-Play Betting on an Agent-Based Model of a Sports Betting Exchange.** In: INTERNATIONAL CONFERENCE ON AGENTS AND ARTIFICIAL INTELLIGENCE (ICAART), 16., 2024, Rome. Proceedings [...]. Setúbal: SCITEPRESS, 2024. p. 1-12. Acesso em: 18 nov. 2025.

THABTAH, Fadi; ZHANG, Li; ABDELHAMID, Neda. **NBA Game Result Prediction Using Feature Analysis and Machine Learning.** *Annals of Data Science*, [s. l.], v. 6, n. 1, p. 103–116, 2019. Acesso em: 13 nov. 2025.

TORRANCE, J. et al. **The structural characteristics of online sports betting: a scoping review of current product features and utility patents as indicators of potential future developments.** *Addiction Research & Theory*, [s. l.], v. 32, n. 3, 2023. Acesso em: 13 out. 2025.

WALSH, Conor; JOSHI, Alok. **Machine learning for sports betting: should model selection be based on accuracy or calibration?** arXiv preprint, [s. l.], 10 mar. 2023. Disponível em: <https://arxiv.org/pdf/2303.06021>. Acesso em: 21 out. 2025.

WILLIAMS, R. et al. **The Definition, Dimensionalization, and Assessment of Gambling Participation.** [s. l.], 2017. Disponível em: <https://opus.uleth.ca/server/api/core/bitstreams/ebb3250e-0d19-426c-b649-70e99941450b/content>. Acesso em: 16 nov. 2025.