



INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO PIAUÍ
CAMPUS TERESINA CENTRAL
TECNOLOGIA EM ANÁLISE E DESENVOLVIMENTO DE SISTEMAS

JOÃO CARLOS PEREIRA SILVA

MONEYBALL: plataforma analítica com inteligência artificial para predição de probabilidades e avaliação de valor em apostas de futebol

TERESINA, PIAUÍ
2025

JOÃO CARLOS PEREIRA SILVA

MONEYBALL: plataforma analítica com inteligência artificial para predição de probabilidades e avaliação de valor em apostas de futebol

Trabalho de Conclusão de Curso (monografia) apresentado como exigência parcial para obtenção do diploma do Curso de Análise e Desenvolvimento de Sistemas do Instituto Federal de Educação, Ciência e Tecnologia do Piauí, Campus Teresina Central.

Orientador: Prof. M^e. Fernando Castelo Branco Goncalves Santana

TERESINA, PIAUÍ
2025

Dados Internacionais de Catalogação na Publicação (CIP) de acordo com ISBD

Silva, João Carlos Pereira

S586m Moneyball : plataforma analítica com inteligência artificial para predição de probabilidades e avaliação de valor em apostas de futebol / João Carlos Pereira Silva. - 2025.
69 f.: il. color.

Trabalho de Conclusão de Curso (Tecnologia em Análise e Desenvolvimento de Sistemas) -
Instituto Federal de Educação, Ciência e Tecnologia do Piauí, Campus Teresina Central, 2025.
Orientador : Prof Me. Fernando Castelo Branco Goncalves Santana.

1. Futebol . 2. Apostas esportivas . 3. Inteligência artificial . 4. Ciência de dados. I. Título.

CDD - 004.21

JOÃO CARLOS PEREIRA SILVA

MONEYBALL: plataforma analítica com inteligência artificial para predição de probabilidades e avaliação de valor em apostas de futebol

Trabalho de Conclusão de Curso (monografia) apresentado como exigência parcial para obtenção do diploma do Curso de Análise e Desenvolvimento de Sistemas do Instituto Federal de Educação, Ciência e Tecnologia do Piauí, Campus Teresina Central.

Aprovada em 08 / 12 / 2025 .

BANCA EXAMINADORA:

Prof. M^e. Fernando Castelo Branco Goncalves
Santana(Orientador)
Instituto Federal do Piauí (IFPI)

Prof. Dr. Ricardo Martins Ramos
Instituto Federal do Piauí (IFPI)

Prof. Dra^e. Nádia Mendes dos Santos
Instituto Federal do Piauí (IFPI)

*Dedico este trabalho a meus pais e, principalmente,
minha filha, que sempre foi e será a minha
principal fonte de força e energia.*

AGRADECIMENTOS

Agradeço, em primeiro lugar, a Deus, que nunca me abandona nos momentos mais difíceis.

Agradeço a minha mãe, Maria do Carmo, meu pai, José de Arimatéia, aos ensinamentos, apoio, conselhos e por doarem parte de sua vida para proporcionarem uma vida melhor aos seus filhos, e também a minha irmã Vanessa Baptista por ser meu maior exemplo na infância e por moldar boa parte da minha personalidade.

Agradeço ao meu professor e orientador Me. Fernando Castelo Branco Gonçalves Santana pela paciência e acolhimento durante todo o período acadêmico.

Agradeço também aos professores Ricardo Martins e Nádia Mendes por integram a banca avaliadora deste trabalho e pelas contribuições para o seu aprimoramento.

Agradeço a todos os professores e servidores do IFPI Campus Central, cuja dedicação, paciência e profissionalismo foram essenciais para o meu crescimento acadêmico e pessoal.

Agradeço a todos os meus amigos que estiveram ao meu lado de alguma forma durante essa trajetória. A amizade de vocês foi uma fonte constante de apoio e motivação.

*Não nos tornamos sábios pela lembrança do nosso
passado, mas pela responsabilidade pelo nosso futuro”.*

Bertrand Russell

RESUMO

O futebol é, sem dúvidas, o esporte mais popular e influente do mundo, movendo paixões e gerando impacto direto na economia, mídia e cultura. Nos últimos anos, esse cenário abriu caminho para a explosão do mercado de apostas esportivas, especialmente com o avanço da tecnologia e da presença massiva das chamadas "Bets". Em 2025, por exemplo, mais de 90% dos clubes da Série A do Campeonato Brasileiro possuem casas de apostas como patrocinadores master. Essas empresas utilizam modelos estatísticos sofisticados e algoritmos para definir as odds de cada evento esportivo, ajustando-as conforme o comportamento do mercado e sempre incorporando sua margem de lucro. Assim, mesmo quando um resultado parece "justo", a casa de apostas tende a pagar menos do que a probabilidade real do evento. Esse modelo garante que, no longo prazo, a casa tenha vantagem matemática. Nesse ambiente, muitos apostadores entram motivados pela paixão pelo esporte ou por achar que possui entendimento elevado no esporte, mas na verdade o conhecimento matemático e estatístico se sobressai ao conhecimento esportivo para ganhar dessas Bets. Estima-se que apenas cerca de 3% dos apostadores em todo o mundo conseguem obter lucro constantemente nesse mercado. Com isso, o objetivo deste trabalho é oferecer uma plataforma inteligente para análise de competições, times e resultados voltados para apostas esportivas em futebol utilizando da tecnologia de análise de dados e inteligência artificial para encontrar potencial valor em apostas, apostando com razão, cálculos e probabilidades reais antes da emoção, visando de fato lucrar antes que simplesmente acertar apostas.

Palavras-chaves: futebol; apostas esportivas; inteligência artificial; ciência de dados.

ABSTRACT

Football is undoubtedly the most popular and influential sport in the world, stirring passions and generating a direct impact on the economy, media and culture. In recent years, this scenario has paved the way for the explosion of the sports betting market, especially with the advancement of technology and the massive presence of so-called "Bets". In 2025, for example, more than 90% of the clubs in the Brazilian Championship Series A will have bookmakers as master sponsors. These companies use sophisticated statistical models and algorithms to define the odds for each sporting event, adjusting them according to market behavior and always incorporating their profit margin. Thus, even when a result seems "fair", the bookmaker tends to pay less than the real probability of the event. This model ensures that, in the long run, the bookmaker has a mathematical advantage. In this environment, many bettors enter motivated by their passion for the sport or because they believe they have a high understanding of the sport, but in fact mathematical and statistical knowledge outweighs sports knowledge to win these Bets. It is estimated that only about 3% of bettors worldwide are able to consistently make a profit in this market. Therefore, the objective of this work is to offer an intelligent platform for analyzing competitions, teams and results aimed at sports betting on football using data analysis technology and artificial intelligence to find potential value in bets, betting with reason, calculations and real probabilities before the emotion, aiming to actually make a profit rather than simply getting bets right.

Keywords: football; sports betting; artificial intelligence; data science.

LISTA DE ILUSTRAÇÕES

Figura 1 – Cotações em tempo real em uma casa de apostas(Em azul temos as odds disponíveis para BACK e em rosa temos as odds disponíveis para LAY. Os valores em dinheiro abaixo corresponde a liquidez de mercado	23
Figura 2 – Simulação da Lei dos Grandes Números	25
Figura 3 – Arquitetura de pilha de dados moderna, onde o Streamlit é utilizado para fornecer aplicativos móveis e da web para o usuário final . . .	33
Figura 4 – Lógica que carrega os dados diretamente do site e filtra por temporada e liga	34
Figura 5 – Processamento de dados para o processo de transformação em features	37
Figura 6 – Exemplo de processo de filtragem de Features	41
Figura 7 – Representação da curva de calibração do modelo – probabilidade prevista vs. frequência observada de vitória do visitante – idealmente próxima da diagonal que representa perfeita calibração	42
Figura 8 – Processo de votação da floresta aleatória, com exemplo de 3 árvores.	43
Figura 9 – Linha do tempo de aprendizado com Walk-Foward	45
Figura 10 – Histórico da simulação da temporada 24/25 (Premier League) com o método back visitante	47
Figura 11 – Resultado da simulação da temporada 24/25 (Premier League) após a rodada 7 com o método back visitante	48
Figura 12 – Recomendações de apostas para jogos futuros no método back visitante.	51
Figura 13 – Tela de seleção de quais jogos o usuário irá realizar a predição back visitante.	52
Figura 14 – Tela de classificação do MoneyBall	53
Figura 15 – Gráfico de posição do Man City ao decorrer da temporada	54
Figura 16 – Rodadas e jogos disponíveis para detalhar	54
Figura 17 – Estatísticas da partida entre Tottenham e Brighton	55
Figura 18 – Estatísticas do Arsenal	57
Figura 19 – Comparativo entre Arsenal e Man City na temporada	58
Figura 20 – Jogos por acontecer na data	59
Figura 21 – Evolução temporal da taxa de acerto e ROI entre ligas.	60
Figura 22 – Evolução temporal do lucro entre ligas.	61
Figura 23 – Ranking geral entre ligas	61
Figura 24 – Principais métricas por temporada.	62

Figura 25 – Resumo por temporadas.	62
Figura 26 – Visão geral de resultados.	62
Figura 27 – Evolução temporal da taxa de acerto e ROI entre ligas.	63
Figura 28 – Evolução temporal do lucro entre ligas.	64
Figura 29 – Ranking geral entre ligas	64
Figura 30 – Principais métricas por temporada.	65
Figura 31 – Resumo por temporadas.	65
Figura 32 – Visão geral de resultados.	65

LISTA DE TABELAS

Tabela 1 – Comparação entre os trabalhos relacionados	29
---	----

LISTA DE ABREVIATURAS E SIGLAS

API	Application Programming Interface
CSV	Comma-Separated Values
CRUD	Create, Read, Update, Delete
EV	Expected Value
RESTful	Representational State Transfer (com boas práticas REST)
ROI	Return on Investment

SUMÁRIO

1	INTRODUÇÃO	16
1.1	O TERMO MONEYBALL	16
1.2	JUSTIFICATIVA	16
1.3	OBJETIVOS	17
1.3.1	Objetivo Geral	17
1.3.2	Objetivos Específicos	17
2	FUNDAMENTAÇÃO TEÓRICA	18
2.1	A ASCENSÃO DAS APOSTAS ESPORTIVAS NO BRASIL	18
2.2	A REALIDADE DAS APOSTAS	19
2.3	APOSTAS ESPORTIVAS NA PRÁTICA	20
2.3.1	Apostas BACK	20
2.3.2	Apostas LAY	21
2.4	A CASA SEMPRE VENCE	23
2.5	EXPLORANDO INEFICIÊNCIAS NO MERCADO DAS ODDS	26
3	TRABALHOS RELACIONADOS	28
3.1	PREDIÇÃO DE RESULTADOS DE UMA PARTIDA DE FUTEBOL UTILIZANDO APRENDIZADO DE MÁQUINA - FERNANDES	28
3.2	MODELO DE PREDIÇÃO DE RESULTADOS DE PARTIDAS DE FUTEBOL - TAKAYANAGI ET AL.	28
3.3	PREDIÇÃO DE RESULTADOS DE PARTIDAS DO CAMPEONATO BRASILEIRO DE FUTEBOL - MACIEL	29
3.4	COMPARATIVO	29
4	MONEYBALL	31
4.1	VISÃO GERAL E FERRAMENTAS UTILIZADAS	31
4.2	VISUALIZAÇÃO DA INTERFACE	32
4.3	COLETA DE DADOS	33
4.4	DADOS COLETADOS	34
4.4.1	Informações gerais da partida	35
4.4.2	Estatísticas de desempenho	35
4.4.3	Odds da Bet365	35
4.4.4	Odds agregadas de mercado	35
4.5	PROCESSAMENTO ESTATÍSTICO	36
4.6	ENGENHARIA DE ATRIBUTOS (FEATURES)	37

4.6.1	Indicadores estatísticos de desempenho	37
4.6.2	Indicadores de mercado (Odds)	38
4.7	NORMALIZAÇÃO DOS DADOS	39
4.7.1	Normalização de proporções e frequências	39
4.7.2	Conversão de odds em probabilidades implícitas	39
4.7.3	Padronização de métricas contínuas	39
4.7.4	Tratamento de dados ausentes e times recém-promovidos	40
4.7.5	Justificativa técnica	40
4.8	MODELO PREDITIVO MONEYBOT	40
4.8.1	Engenharia de Atributos (Feature Engineering)	40
4.8.2	Biblioteca <i>Scikit-Learn</i>, Pipeline e Calibração	41
4.8.3	Classificador Random Forest	42
4.8.4	Validação Walk-Forward com Refit Periódico	44
4.8.5	Simulação de Apostas, Avaliação de Resultados e Critérios de Valor Esperado	45
4.8.5.1	Simulação de Temporadas	46
4.8.5.2	Simulando Temporadas “Lay Mandante”	48
4.8.5.3	Simulando Temporadas “Back Visitante”	49
4.8.5.4	Simulação de Jogos Futuros	49
4.8.5.5	Simulando Jogos “Lay Mandante”	51
4.8.5.6	Simulando Jogos “Back Visitante”	52
4.9	FERRAMENTAS ANALÍTICAS	52
4.9.1	Classificação	53
4.9.2	Jogos por rodada	54
4.9.3	Estatísticas avançadas por time	55
4.9.4	Comparativo entre times	57
4.9.5	Jogos do dia	58
4.10	RESULTADOS ALCANÇADOS	59
4.10.1	Resultados Lay Mandante	60
4.10.1.1	Resultados por Ligas	60
4.10.1.2	Resultados gerais	61
4.10.1.3	Conclusões sobre o modelo Lay Mandante	62
4.10.2	Resultados Back Visitante	62
4.10.2.1	Resultados por Ligas	63
4.10.2.2	Resultados gerais	64
4.10.2.3	Conclusões sobre o modelo Back Visitante	65
5	CONCLUSÃO	66

REFERÊNCIAS 67

1 INTRODUÇÃO

Este capítulo irá fazer uma breve explicação sobre o que é MoneyBall e o porque este trabalho ter adotado esse nome. Em seguida fornecer dados e informações sobre esta metodologia e o que conseguimos absorver com tudo isso para criar um conceito de MoneyBall voltado para análise preditiva e apostas esportivas com futebol.

1.1 O TERMO MONEYBALL

O termo Moneyball ganhou notoriedade após a publicação do livro homônimo de Michael Lewis (LEWIS, 2004) e sua posterior adaptação para o cinema (MILLER, 2011), estrelada por Brad Pitt. A obra relata a revolução silenciosa protagonizada por Billy Beane, gerente do time de beisebol Oakland Athletics, ao desafiar o modelo tradicional de recrutamento esportivo. Em vez de seguir a intuição de olheiros experientes, Beane adotou uma abordagem baseada em estatísticas avançadas e evidências objetivas, buscando identificar jogadores subvalorizados pelo mercado. O sucesso dessa estratégia provou que, mesmo com orçamentos limitados, era possível competir em alto nível se as decisões fossem orientadas por dados.

Esse conceito rompeu as fronteiras do beisebol e passou a influenciar diversos esportes ao redor do mundo, especialmente o futebol. Um exemplo notável ocorreu em 2016, quando o Leicester City surpreendeu o mundo ao conquistar a Premier League (HARINGEY, 2016). O clube inglês aplicou os princípios do Moneyball ao utilizar análises de dados para identificar e contratar jogadores como Riyad Mahrez, N'Golo Kanté e Jamie Vardy, todos subvalorizados à época. Além da montagem do elenco, o Leicester aplicou estratégias táticas baseadas em estatísticas para maximizar o desempenho coletivo, desafiando os grandes clubes com uma abordagem eficiente, racional e orientada por evidências.

Outros casos notórios de relativo sucesso com a abordagem do MoneyBall são o Brighton, clube médio britânico que conseguiu se classificar pela primeira vez em sua história para uma competição europeia (BISPO, 2024). Também temos o caso do Brentford, time britânico que ficou 74 anos sem conseguir o acesso a Premier League, e após a saída do manager da equipe Mark Warburton em 2015 e adoção ao MoneyBall, já na temporada 2021-22 a equipe consegue o acesso a elite do futebol inglês, onde está até hoje (FLANAGAN, 2021).

1.2 JUSTIFICATIVA

Inspirado no princípio em que de que tomar decisões esportivas com base em dados objetivos, e não apenas em opiniões ou achismo é o caminho mais con-

sistente, este trabalho apresenta o MoneyBall, uma plataforma analítica desenvolvida para análise estatística e predição de resultados no futebol, com foco em avaliar potenciais valores para apostas esportivas. A proposta se alinha à filosofia do livro ao aplicar inteligência artificial, estatísticas históricas e probabilidades de mercado (odds) para encontrar oportunidades de apostas com valor esperado positivo, priorizando a racionalidade matemática em detrimento da intuição e emoção.

Assim como Billy Beane desafiou os dogmas do beisebol, e como o Leicester City rompeu a probabilidade de 5000-1 (MACHADO, 2024) com o uso de dados no futebol, este projeto busca mostrar que também é possível desafiar o mercado das apostas com um olhar estatístico e técnico, substituindo o palpite pelo cálculo, e o achismo pela evidência.

1.3 OBJETIVOS

1.3.1 Objetivo Geral

O objetivo geral deste trabalho é desenvolver uma ferramenta de apoio para apostadores esportivos na área de futebol, com foco em ferramentas técnicas e estatísticas de apoio para análise de apostas e um bot de inteligência artificial com análise preditiva de apostas com potencial lucrativo.

1.3.2 Objetivos Específicos

- Fornecer tabelas das principais ligas atualizadas, oferecendo filtros personalizados para apoio de análises;
- Carregar as estatísticas de uma determinada partida para análise;
- Oferecer estatísticas avançadas de uma determinada equipe para que o usuário possa analisar os pontos fracos e fortes desta equipe, além dos valores médios de suas Odds;
- Oferecer uma lista de jogos do dia das principais ligas, as Odds das equipes para o jogo, sem a necessidade de procurar em uma casa de apostas;
- Fazer uma comparação direta frente a frente de dois times de uma determinada liga, oferecendo assim uma análise completa para auxiliar na decisão de uma aposta;
- Oferecer um bot que faz cálculos e recomenda apostas em partidas futuras com Odds lucrativas, utilizando os métodos Lay mandante(apostar contra o mandante) e Back visitante(apostar a favor do visitante), podendo também simular resultados de temporadas utilizando os métodos de apostas em que o modelo considera com potencial lucrativo e informando seus resultados ao fim.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, serão abordados os principais conceitos e estudos que fundamentam a proposta deste trabalho. Inicialmente, será apresentada uma visão geral sobre a ascensão das apostas esportivas no Brasil e os fatores que contribuíram para sua popularização. Em seguida, serão discutidos os desafios enfrentados por apostadores, a matemática por trás das odds, como funcionam as apostas esportivas e a vantagem estrutural das casas de apostas. Além disso, será explorado o conceito de valor esperado (Expected Value) como pilar para decisões racionais em apostas, bem como estratégias que buscam superar a margem da casa.

Por fim, será discutido como a inteligência artificial e a análise preditiva podem ser aplicadas nesse contexto, estabelecendo a base teórica para o desenvolvimento do sistema preditivo MoneyBot, proposto neste trabalho.

2.1 A ASCENSÃO DAS APOSTAS ESPORTIVAS NO BRASIL

Segundo informações publicadas pela pesquisa feita sobre a indústria de apostas esportivas no Brasil (ESPORTE, 2024), desde que as apostas esportivas foram autorizadas no Brasil em 2018, o setor passou por um crescimento notável. Estima-se que movimentou R\$120bi em apostas online em 2023 e até R\$150bi em 2024, o que representa 12% do PIB, impulsionado por regulamentação, smartphone, Pix e apostas ao vivo (BRASIL, 2025). A regulamentação oficial só veio em dezembro de 2023, por meio da Lei nº 14.790/23, que estabeleceu diretrizes como o modelo de cobrança de impostos, as regras para concessão de licenças e as limitações quanto ao público-alvo.

O perfil predominante entre os apostadores online é de homens jovens, de classe média baixa, concentrados principalmente na região Sudeste. De acordo com uma pesquisa do Instituto Locomotiva (STRATEGYAND, 2024) realizada em setembro de 2023, cerca de 33 milhões de pessoas de baixa renda já realizaram apostas esportivas. Dentro desse grupo, aproximadamente 22 milhões (ou seja, 20% da população de baixa renda) afirmam apostar pelo menos uma vez por mês.

Em pesquisa sobre o impacto das apostas esportivas (STRATEGYAND, 2024), a principal razão citada para apostar é o desejo de obter lucro. No entanto, muitos dos que ganham acabam reinvestindo o valor conquistado em novas apostas. A percepção mais comum entre os apostadores é a de que, no longo prazo, perderam mais dinheiro do que ganharam.

A forte cultura esportiva do Brasil, especialmente a paixão pelo futebol, foi um fator decisivo para o avanço das apostas esportivas. O entusiasmo nacional pelo esporte, aliado ao aumento no uso da internet e de smartphones, facilitou o acesso às

plataformas de apostas (BRASIL, 2025). Além disso, problemas econômicos como a desigualdade de renda e a busca por fontes extras de renda impulsionaram ainda mais o interesse da população. Campanhas publicitárias intensas feitas pelas empresas do setor também contribuíram para tornar as apostas uma prática cada vez mais comum.

2.2 A REALIDADE DAS APOSTAS

Enquanto uma parcela da população aceita que é impossível ganhar dinheiro com apostas esportivas, outra parcela afirma que sim é possível, porém na prática nunca teve ganhos relevantes de fato que fizessem valer a pena o seu investimento de tempo e dinheiro. De acordo com uma pesquisa realizada pela BBC News Brasil (GALLAS, 2024) após consultar diversos especialistas que já dedicaram mais de 20 anos de pesquisa sobre o assunto, a conclusão deles é que: sim, teoricamente é possível ganhar dinheiro com apostas esportivas no longo prazo. No entanto, a tarefa é tão difícil que só estaria ao alcance de pouquíssimas pessoas.

A grande maioria inevitavelmente vai perder dinheiro se continuar apostando. É só uma questão de tempo.

"Se você não tem fortes habilidades matemáticas e nunca parou para construir modelos matemáticos para vencer as casas de apostas, você não vai conseguir. Você não vai ganhar dinheiro a longo prazo a menos que trabalhe duro. E trabalhar não significa assistir a muitas partidas de futebol ou ter conhecimento do jogo. Trabalhar duro significa realmente entender as probabilidades", disse o matemático David Sumpter à BBC News Brasil (GALLAS, 2024).

Sumpter é professor de matemática na universidade de Uppsala, na Suécia, fundador de uma empresa que trabalha com matemática e futebol e autor do livro (SUMPTER, 2016) *Soccermatics: Mathematical Adventures in the Beautiful Game* ("*Soccermatics: Aventuras matemáticas no jogo bonito*", em tradução livre).

Ele diz que a habilidade matemática necessária para derrotar as casas de apostas é tamanha, que um apostador que eventualmente consiga isso lucraria mais vendendo seu conhecimento matemático no mercado, inclusive para as próprias casas de apostas ou fundando sua própria empresa de estatísticas.

Outro especialista, o autor Joseph Buchdahl, diz que a maioria dos apostadores são como "macacos jogando dardos", ou seja, contam apenas com pura sorte em suas apostas (GALLAS, 2024).

Recentemente ele publicou uma análise em que concluiu que há apostadores com técnicas tão avançadas que são capazes de vencer o mercado no longo prazo. No entanto, eles ainda são casos raros.

2.3 APOSTAS ESPORTIVAS NA PRÁTICA

2.3.1 Apostas BACK

A modalidade *Back* é o tipo de aposta mais tradicional e intuitiva das apostas esportivas (ASHBOURNE, 2016). Apostar *Back* significa apostar *a favor* de um determinado resultado acontecer, por exemplo, apostar a favor da vitória do mandante ou do visitante.

Em uma aposta *Back*, o apostador lucra caso o resultado escolhido ocorra, e perde sua *stake* (valor apostado) caso o resultado não aconteça. Assim, o apostador atua da mesma forma que um cliente de uma casa de apostas convencional: ele investe uma quantia em um evento e recebe um retorno proporcional à odd (cotação) caso acerte o resultado.

Estrutura da aposta Back

Sejam:

- S : o valor apostado (*stake*);
- O : a odd decimal do evento;
- G : o lucro líquido obtido caso o resultado apostado ocorra.

O ganho líquido é calculado por:

$$G = (O - 1) \times S$$

Assim:

- Se o evento **ocorre**, o apostador *Back* ganha o lucro G e recebe de volta sua *stake*, totalizando $O \times S$;
- Se o evento **não ocorre**, o apostador perde a *stake* S .

Portanto, o risco máximo de uma aposta *Back* é o valor da *stake* apostada, enquanto o potencial de lucro cresce linearmente conforme a odd aumenta (THOMAS, 2025).

Exemplo 1 — Aposta Back em odd 2,50

Suponha que o apostador aposte R\$100,00 a favor da vitória do visitante (*Back Visitante*) em odd $O = 2,50$.

$$G = (2,50 - 1) \times 100 = 1,5 \times 100 = \text{R\$ } 150,00$$

Logo:

- Se o visitante **vencer**: o apostador recebe R\$250,00 (R\$150,00 de lucro + R\$100,00 da stake);
- Se o visitante **não vencer**: o apostador perde sua stake de R\$100,00.

Exemplo 2 — Aposta Back em odd 1,80

Suponha agora uma aposta *Back Mandante* de R\$100,00 em odd $O = 1,80$.

$$G = (1,80 - 1) \times 100 = 0,8 \times 100 = \text{R\$ } 80,00$$

Neste caso:

- Se o mandante **vencer**: o apostador ganha R\$80,00 de lucro e recebe R\$180,00 no total;
- Se o mandante **não vencer**: o apostador perde R\$100,00.

Note que, quanto menor a odd, menor o lucro potencial, mas maior a probabilidade implícita de vitória. Já odds mais altas representam eventos menos prováveis, mas com ganhos potenciais maiores.

2.3.2 Apostas LAY

Já a modalidade *Lay* representa o ato de apostar *contra* um determinado resultado, por exemplo, apostar contra a vitória do mandante. Diferente da aposta *Back*, em que o apostador lucra caso o evento aconteça, na aposta *Lay* o apostador lucra caso o evento *não* ocorra.

Em bolsas de apostas como a Betfair Exchange, o apostador *Lay* atua como se fosse a própria casa de apostas, aceitando a aposta de quem está em *Back* (SMARKETS, 2024). Assim, ele oferece uma odd ao mercado e assume um risco (*liability*) proporcional à cotação oferecida. Esse risco é o valor que ele perde caso o resultado em que apostou contra aconteça.

Estrutura da aposta Lay

Sejam:

- S : o valor apostado (*stake*);
- O : a odd decimal;
- L : a responsabilidade (*liability*) do apostador.

A relação entre *stake* e *liability* é dada por:

$$L = (O - 1) \times S$$

Assim:

- Se o evento *não* ocorre, o apostador *Lay* **ganha** o valor da *stake* S ;
- Se o evento *ocorre*, o apostador *Lay* **perde** o valor da responsabilidade L .

O ganho no sucesso é fixo e igual à *stake* apostada, enquanto a perda é variável e cresce proporcionalmente à odd oferecida.

Exemplo 1 — Aposta Lay em odd 3,00

Suponha que o apostador decida apostar *contra* a vitória do mandante em uma odd $O = 3,00$, com uma *stake* de R\$100,00.

A responsabilidade será:

$$L = (3,00 - 1) \times 100 = 2 \times 100 = \text{R\$ } 200,00$$

Portanto:

- Se o mandante **não vencer** (empate ou derrota): o apostador *Lay* ganha R\$100,00;
- Se o mandante **vencer**: o apostador *Lay* perde R\$200,00.

Esse exemplo evidencia que, em odds mais altas, o risco financeiro é maior do que o possível ganho. O apostador precisa estar ciente de que está “bancando” uma aposta com *payout* elevado.

Exemplo 2 — Aposta Lay em odd 1,50

Agora, suponha que o apostador *Lay* aposte contra o mandante em odd $O = 1,50$, novamente com *stake* de R\$100,00.

A responsabilidade será:

$$L = (1,50 - 1) \times 100 = 0,5 \times 100 = \text{R\$ } 50,00$$

Neste caso:

- Se o mandante **não vencer**: o apostador *Lay* ganha R\$100,00;
- Se o mandante **vencer**: o apostador *Lay* perde apenas R\$50,00.

Quanto menor a odd, menor o risco assumido na aposta *Lay*. Contudo, uma odd mais baixa indica uma alta probabilidade implícita de vitória, o que reduz a chance de sucesso.

Figura 1 – Cotações em tempo real em uma casa de apostas (Em azul temos as odds disponíveis para BACK e em rosa temos as odds disponíveis para LAY. Os valores em dinheiro abaixo corresponde a liquidez de mercado)

Futebol		1		X		2		
Hoje às 16:00	Espanha França	2.86 R\$76552	2.88 R\$204449	2.92 R\$95911	2.94 R\$142821	3.2 R\$156333	3.25 R\$223873	i
qua 16:00	Holanda Inglaterra	3.3 R\$82631	3.4 R\$115533	2.92 R\$64326	2.94 R\$39191	2.76 R\$47783	2.8 R\$66101	i
Hoje às 21:00	Argentina Canadá	1.36 R\$106313	1.37 R\$4510	5.2 R\$17610	5.3 R\$22056	12.5 R\$1194	13 R\$3833	i
jul 11 19:30	Palmeiras Atlético-GO	1.46 R\$554	1.47 R\$248	4.5 R\$135	4.7 R\$2184	10 R\$1735	11 R\$1454	i

Fonte: <https://www.netflu.com.br/apostas/betfair-exchange/>

2.4 A CASA SEMPRE VENCE

As cotações (odds) oferecidas nas apostas são definidas pelas próprias casas de apostas. E é justamente aí que surge uma das principais dificuldades para quem aposta: essas casas sempre incluem uma margem de lucro embutida nas odds.

Uma maneira prática de identificar essa margem é somar as probabilidades implícitas de cada possível resultado de um evento esportivo (overround). O overround é o excedente acima de 100% na soma das probabilidades implícitas, representando a margem de lucro da casa de apostas (SOHAIL, 2024).

Se estivermos lidando apenas com três desfechos possíveis (vitória, empate ou derrota), a soma dessas probabilidades deveria ser exatamente 100%, já que um dos três eventos obrigatoriamente ocorrerá. Mas nas casas de apostas essa soma é sempre maior que 100%, e o valor que ultrapassa essa marca representa a vantagem da casa, ou seja, o quanto ela espera lucrar a longo prazo.

Vamos ver isso na prática com um exemplo de jogo entre Vasco x Flamengo. Suponha que uma casa de apostas esteja oferecendo as seguintes odds:

Vitória do Vasco: 4,410

Empate: 3,740

Vitória do Flamengo: 1,787

Ao calcular as probabilidades implícitas:

$$\text{Vasco: } \frac{1}{4,410} = 0,2268 = 22,68\%$$

$$\text{Empate: } \frac{1}{3,740} = 0,2674 = 26,74\%$$

$$\text{Flamengo: } \frac{1}{1,787} = 0,5596 = 55,96\%$$

Somando essas porcentagens, temos:

$$22,68\% + 26,74\% + 55,96\% = 105,38\%$$

Ou seja, há uma margem de 5,38% embutida nas odds. Esse percentual extra é o que garante lucro para a casa, independentemente do resultado.

Agora imagine que um apostador resolva dividir seu dinheiro proporcionalmente entre os três resultados:

R\$22,68 na vitória do Vasco

R\$26,74 no empate

R\$55,96 na vitória do Flamengo

Total apostado: R\$105,38

Como ele apostou em todas as possibilidades, o esperado seria que recuperasse exatamente o valor investido, qualquer que fosse o resultado. Mas o que aconteceria de fato é o seguinte:

$$\text{Se o Vasco vencer: } 22,68 \times 4,410 = 100,00$$

$$\text{Se der empate: } 26,74 \times 3,740 = 100,00$$

$$\text{Se o Flamengo vencer: } 55,96 \times 1,787 = 100,00$$

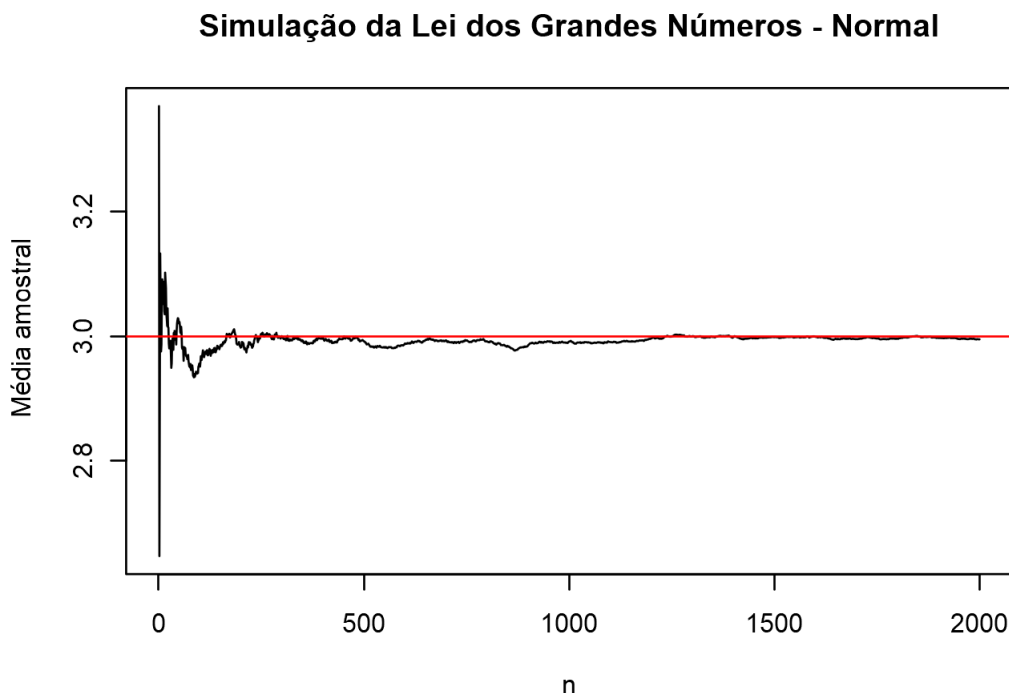
Ou seja, em todos os cenários o retorno seria menor do que o valor apostado, o que significa prejuízo garantido de cerca de R\$5,37 a R\$5,39.

Isso mostra que as casas de apostas não pagam exatamente o valor correspondente às verdadeiras chances de um evento ocorrer (KNIGHT, 2024), elas sempre pagam um pouco menos para garantir seu lucro. Esse detalhe, muitas vezes ignorado pelos iniciantes, é um dos grandes desafios para quem deseja lucrar com apostas esportivas.

Esse “pequeno desconto” presente nas odds das casas de apostas representa, no longo prazo, um grande prejuízo para os apostadores e, por outro lado, um lucro garantido para as casas de apostas.

Essa lógica é sustentada por um princípio matemático conhecido como Lei dos Grandes Números, que afirma que quanto mais vezes um experimento aleatório é repetido, mais os resultados se aproximam da probabilidade teórica esperada.

Figura 2 – Simulação da Lei dos Grandes Números



Fonte: <https://www.inf.ufsc.br/~andre.zibetti/probabilidade/teoremas-limite.html>

Imagine uma situação simples: um jogo de cara ou coroa, onde a chance de sair qualquer um dos lados é de 50%. Em um cenário justo, o pagamento ideal seria 2.00 para cada resultado, ou seja, se você apostar R\$100 em "cara" e acertar, receberia R\$200 (R\$100 de lucro + R\$100 apostado).

$$100 \times 2,00 = 200$$

No entanto, as casas de apostas geralmente oferecem odds como 1.95 para "cara" e 1.95 para "coroa". Isso significa que, se você apostar R\$100 e acertar, receberá apenas R\$195 no total — R\$95 de lucro, e não R\$100.

$$100 \times 1,95 = 195$$

Agora, considere que você faz 1.000 apostas de R\$100, sempre em cara. Suponha que os resultados estejam bem equilibrados: você acerta 500 vezes e erra outras 500.

Total apostado:

$$100 \times 1000 = 100.000$$

Acertos (500 apostas com odds 1.95):

$$500 \times 195 = 97.500$$

Erros (500 apostas perdidas):

$$500 \times 0 = 0$$

Retorno total:

$$97.500 + 0 = 97.500$$

Prejuízo acumulado:

$$100.000 - 97.500 = 2.500$$

Mesmo acertando exatamente 50% das apostas, ou seja, jogando bem dentro da expectativa estatística, o apostador termina com prejuízo. Isso acontece porque o valor pago pelas odds não reflete a real probabilidade de acerto, sendo sempre um pouco inferior, o que garante margem de lucro para a casa (WIKIPEDIA, 2024).

Essa diferença, muitas vezes imperceptível em poucas apostas se torna evidente e inevitável conforme o número de apostas aumenta. Por isso, mesmo que alguém tenha lucros pontuais em apostas isoladas, no longo prazo a tendência é que perca dinheiro.

É exatamente esse padrão que sustenta o modelo de negócios das casas de apostas: milhares de apostas sendo feitas continuamente, onde a margem embutida nas odds garante lucro constante para o operador, graças à Lei dos Grandes Números.

2.5 EXPLORANDO INEFICIÊNCIAS NO MERCADO DAS ODDS

Ganhar dinheiro com apostas esportivas não é sobre sorte, é sobre ter uma vantagem real e constante sobre as casas de apostas.

Para vencer no longo prazo, o apostador precisa encontrar situações em que as odds oferecidas pelas casas não refletem corretamente as chances reais de um resultado acontecer (BETBETTER, 2024). Ou seja: apostar quando a casa está pagando mais do que deveria.

Vamos a um exemplo prático. Imagine um clássico entre Internacional e Grêmio, com as seguintes odds oferecidas por uma casa de apostas:

2,50 para vitória do Internacional

3,20 para empate

2,80 para vitória do Grêmio

Convertendo essas odds em probabilidades implícitas (basta fazer 1 dividido pela odd), temos:

$$\text{Internacional: } \frac{1}{2,50} = 0,40 = 40\%$$

$$\text{Empate: } \frac{1}{3,20} \approx 0,3125 = 31,25\%$$

$$\text{Grêmio: } \frac{1}{2,80} \approx 0,3571 = 35,71\%$$

A soma dessas probabilidades dá:

$$40\% + 31,25\% + 35,71\% = 106,96\%$$

o que, como vimos antes, representa a margem de lucro da casa de apostas.

Agora imagine que, por meio de um bom modelo estatístico, o apostador avalia que o Internacional tem 50% de chances reais de vencer o jogo, e não os 40% sugeridos pela odd.

Se ele está certo, a odd de 2,50 oferecida pela casa está acima do que deveria ser, o que cria uma aposta de valor, ou seja, uma oportunidade matemática de lucro.

No entanto, identificar essas oportunidades de forma pontual não é suficiente. Para superar a margem da casa, o apostador precisa ser capaz de fazer previsões mais precisas do que as casas de forma consistente ao longo de muitas apostas.

O grande desafio, é que as casas de apostas utilizam bancos de dados gigantescos, algoritmos avançados e equipes de analistas. Competir contra isso exige preparo, disciplina, análise estatística profunda e até o controle emocional para suportar as inevitáveis sequências de perdas.

E diante disso, o algoritmo de inteligência artificial do MoneyBot foi treinado justamente para encontrar apostas lucrativas que desafiam as odds das casas de apostas.

3 TRABALHOS RELACIONADOS

3.1 PREDIÇÃO DE RESULTADOS DE UMA PARTIDA DE FUTEBOL UTILIZANDO APRENDIZADO DE MÁQUINA - FERNANDES

O trabalho desenvolvido por Luiz Fernando Rodrigues Fernandes, apresentado na Universidade Federal de Ouro Preto, tem como objetivo aplicar algoritmos de aprendizado de máquina à predição de resultados da La Liga, o campeonato espanhol. O autor adotou uma abordagem técnica e exploratória, utilizando como algoritmos principais o Random Forest, o Naive Bayes e o XGBoost. A base de dados foi construída a partir de temporadas recentes com estatísticas obtidas via web scraping, e o pré-processamento incluiu a criação de variáveis como médias móveis e atributos derivados de desempenho anterior.

O foco do estudo foi a análise da capacidade preditiva dos modelos com base em métricas tradicionais como acurácia, precisão e recall. O modelo de melhor desempenho foi o Naive Bayes, que alcançou acurácia próxima a 53%. Além disso, o trabalho aplicou técnicas de interpretabilidade como LIME, com o objetivo de entender a influência das variáveis nas decisões dos modelos. Não houve aplicação prática voltada a apostas esportivas, nem simulações de ROI, valor esperado ou odds. O estudo destaca-se pela profundidade no tratamento dos dados e avaliação técnica rigorosa, mas permanece no campo acadêmico, sem direcionamento operacional para o mercado.

3.2 MODELO DE PREDIÇÃO DE RESULTADOS DE PARTIDAS DE FUTEBOL - TAKAYANAGI ET AL.

Gabriel Takayanagi, Guilherme Bastos e Tomas Caldeira, alunos da Escola Politécnica da Universidade de São Paulo, desenvolveram um projeto voltado à construção de uma plataforma de predição de resultados de futebol com interface mobile. O backend foi desenvolvido em Supabase e o frontend em Flutter, demonstrando proficiência tanto em ciência de dados quanto em engenharia de software. O sistema prevê resultados de partidas com base em modelos de aprendizado de máquina, priorizando a integração e entrega das previsões ao usuário.

No aspecto técnico, o trabalho implementou modelos como Random Forest, AdaBoost, árvores de decisão e XGBoost, com foco na escolha do melhor algoritmo preditivo baseado em métricas estatísticas de desempenho. A base de dados envolvia estatísticas gerais dos confrontos, sendo utilizadas práticas de validação cruzada e engenharia de atributos. O modelo escolhido final foi o XGBoost. Apesar da menção ao potencial uso do sistema para apostas esportivas, o trabalho não incorporou qualquer

estratégia de apostas, nem cálculo de ROI, EV (valor esperado) ou odds ajustadas. O escopo permanece centrado na previsão estatística e experiência do usuário via aplicativo.

3.3 PREDIÇÃO DE RESULTADOS DE PARTIDAS DO CAMPEONATO BRASILEIRO DE FUTEBOL - MACIEL

Vítor Maciel, em sua monografia apresentada na Universidade Federal do Rio Grande do Sul, propôs a construção de um modelo de previsão voltado ao mercado de apostas esportivas, com foco no Campeonato Brasileiro. Diferentemente dos trabalhos anteriores, Maciel deu ênfase à simulação prática de apostas com base nos resultados gerados pelos modelos. A base de dados foi obtida via Sofascore, abrangendo variáveis estatísticas detalhadas como desempenho recente, estatísticas acumuladas, rating Elo e histórico de confrontos.

O trabalho incorporou etapas robustas de pré-processamento, como normalização, transformação de variáveis e aplicação do algoritmo de clusterização K-means para agrupar partidas em subconjuntos de similaridade. Na modelagem preditiva, foram testados sete algoritmos: Regressão Logística, Random Forest, KNN, Naive Bayes, Redes Neurais, Árvore de Decisão e XGBoost. Os modelos foram comparados com base em métricas como acurácia, F1-score e lucro simulado em apostas.

Os dois melhores modelos de classificação e o modelo de regressão foram então avaliados em simulações de apostas com odds reais. O autor utilizou estratégias baseadas em previsões binárias (vitória, empate ou derrota) sem um critério fixo de aposta. Os resultados obtidos indicaram ROI positivo em alguns cenários, com destaque para o XGBoost, que chegou a atingir ROI superior a 20% em determinadas simulações, com acurácia variando entre 45% e 59%. Ainda assim, o trabalho não estabeleceu um método unificado ou replicável de apostas, mantendo foco na comparação técnica entre algoritmos em diferentes clusters de dados.

3.4 COMPARATIVO

Tabela 1 – Comparação entre os trabalhos relacionados

Trabalho	Método de Aposta	Acurácia (%)	ROI (%)	Funcionalidades da Plataforma
Modelo de Fernandes	Não aplica	~53%	Não avaliado	Apenas modelo preditivo
Modelo de Takayanagi et al	Não aplica	Não informado	Não avaliado	App móvel com previsões
Modelo de Maciel	Genérico	~45–59%	Variável	Scripts e relatórios offline
MoneyBall	Back Visitante e Lay Mandante	~40%	>20%	Ferramentas analíticas e modelo preditivo de apostas

Fonte: Elaborado pelo autor (2025)

A Tabela resume as principais diferenças entre o MoneyBall e os demais trabalhos. O MoneyBall se diferencia ao oferecer uma plataforma interativa voltada ao apostador, com funcionalidades como estatísticas filtráveis, comparação entre equipes, odds médias, lista de jogos do dia e o bot preditivo MoneyBot. O foco é o lucro, não a acurácia. O MoneyBot adota estratégias específicas Back Visitante e Lay Mandante (apostar na vitória do visitante e apostar contra o mandante quando há valor esperado positivo), resultando em acurácia média de 40% e ROI médio superior a 20%.

4 MONEYBALL

Neste capítulo são descritas as etapas e técnicas utilizadas no desenvolvimento da plataforma analítica MoneyBall, incluindo a coleta e preparação de dados, a construção do modelo preditivo (MoneyBot) e a implementação das funcionalidades propostas. O foco da metodologia é garantir que as decisões e análises sejam baseadas em dados estatísticos confiáveis e modelos de inteligência artificial, fornecendo suporte objetivo aos apostadores. A seguir detalhamos a arquitetura do sistema, as fontes de dados empregadas, os procedimentos de processamento e modelagem e, por fim, como cada funcionalidade da plataforma foi implementada para atender aos objetivos propostos.

4.1 VISÃO GERAL E FERRAMENTAS UTILIZADAS

A plataforma MoneyBall foi desenvolvida com base em uma arquitetura modular, composta por três núcleos fundamentais: (I) Módulo de Dados, responsável pela coleta, tratamento e atualização dos dados históricos e em tempo real das partidas de futebol; (II) Módulo de Inteligência Artificial (MoneyBot), que executa os cálculos estatísticos e aplica o modelo preditivo treinado para estimar probabilidades e sugerir apostas com valor esperado positivo; e (III) Interface do Usuário, ambiente visual acessível onde o usuário interage com o sistema, explora estatísticas e consulta recomendações de apostas geradas pelo modelo.

A escolha da linguagem Python como base do desenvolvimento se deve à sua robustez, simplicidade sintática e vasta gama de bibliotecas voltadas à ciência de dados e aprendizado de máquina. Entre as bibliotecas utilizadas, destacam-se:

- **Pandas(v2.2.3)** e **NumPy(v2.2.2)** para leitura, manipulação e análise de estruturas tabulares e vetores numéricos;
- **Matplotlib(v3.10.0)** e **Seaborn(v0.13.2)** para construção de gráficos explicativos e visualizações estatísticas;
- **Plotly(v6.0.1)** para criação de gráficos interativos e visualizações dinâmicas, permitindo maior exploração dos dados diretamente na interface da aplicação;
- **Scikit-learn(v1.6.1)** como núcleo do pipeline de machine learning, incluindo validação walk forward, calibração, métricas de performance e o algoritmo de classificação **Random Forest**;
- **Requests(v2.32.3)** para automação da coleta de dados via leitura remota de arquivos CSV;

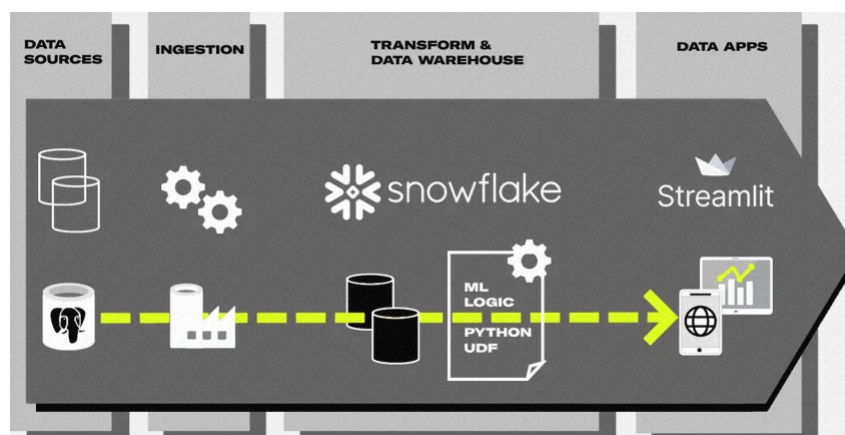
- **Streamlit(v1.44.1)** como framework de interface para construção da aplicação analítica acessível ao usuário final.

4.2 VISUALIZAÇÃO DA INTERFACE

A adoção do Streamlit para a camada de apresentação da plataforma baseia-se em uma decisão técnica estratégica fundamentada em três pilares principais:

1. **Compatibilidade com a Estrutura de Dados do Projeto:** A base do projeto utiliza arquivos CSV obtidos de forma pública e remota, o que dispensou a necessidade de integrações complexas com bancos de dados relacionais ou APIs com autenticação. Como o Streamlit opera de maneira reativa sobre *DataFrames* carregados local ou remotamente, ele se mostrou ideal para prototipação rápida e visualização dinâmica dos dados.
2. **Facilidade na Criação de Interfaces Interativas para Simulação e Consulta:** O Streamlit permite a inserção direta de formulários, seletores, filtros, tabelas e gráficos com poucos comandos, integrando-se de forma nativa aos fluxos de *machine learning* escritos em Python. Isso possibilitou, por exemplo, que o usuário simulasse uma partida específica, selecionando data, times e parâmetros, e recebesse em tempo real as probabilidades estimadas, odds ajustadas e valor esperado (EV) de uma aposta.
3. **Evita a Complexidade de um CRUD Tradicional e Frameworks Web de Propósito Geral:** Como o foco do projeto está na análise de dados e predição de resultados com base em *datasets* atualizados frequentemente, não houve necessidade de um sistema persistente com cadastro de usuários (PARIMI, 2024), armazenamento em banco de dados ou rotas *RESTful* típicas de aplicações web tradicionais (como as desenvolvidas em Django ou Flask). O Streamlit, por sua vez, permite desenvolver *dashboards* e aplicações de dados com apenas um script Python, tornando a manutenção e implantação mais simples.

Figura 3 – Arquitetura de pilha de dados moderna, onde o Streamlit é utilizado para fornecer aplicativos móveis e da web para o usuário final



Fonte: <https://www.recordlydata.com/blog/introduction-to-streamlit-and-data-applications-on-snowflake-with-winter-sports-examples>

Além disso, por ser uma ferramenta de código aberto voltada à ciência de dados, o Streamlit também contribui para a transparência e interpretabilidade dos resultados. Assim, a escolha pelas ferramentas utilizadas, especialmente pelo Streamlit na camada de visualização, se mostrou não apenas adequada ao escopo do projeto, mas também eficiente do ponto de vista técnico, garantindo uma solução funcional, escalável e simples (ASHISH, 2024) alinhada aos objetivos da plataforma MoneyBall.

4.3 COLETA DE DADOS

A primeira etapa do projeto consistiu na coleta de dados históricos de partidas de futebol, abrangendo resultados dos jogos, estatísticas da partida e as cotações (odds) oferecidas pelas casas de apostas para cada evento. O objetivo dessa etapa foi reunir uma base estatística confiável e diversificada, essencial para o treinamento e a validação do modelo preditivo da plataforma MoneyBall.

A principal fonte utilizada para os dados históricos foi o repositório público <https://www.football-data.co.uk/>, que disponibiliza arquivos CSV organizados por temporada e por liga. A coleta dos dados foi realizada de forma automatizada via código Python, utilizando a biblioteca requests para acessar diretamente os arquivos CSV hospedados no site (BETFAIR, 2022).

Essa abordagem permitiu carregar automaticamente os dados das temporadas desejadas por meio de um mapeamento dinâmico de códigos, sem a necessidade de downloads manuais, facilitando a manutenção e a escalabilidade do projeto, já que o repositório é atualizado diariamente sempre que ocorre uma partida de uma determinada liga.

Para a funcionalidade de “Jogos do Dia”, o projeto adotou uma estratégia complementar de coleta: os jogos agendados para o dia corrente foram obtidos a

partir de um repositório público no GitHub que integra dados da API do FootyStats. A estrutura utilizada acessa diretamente os arquivos CSV hospedados no endereço.

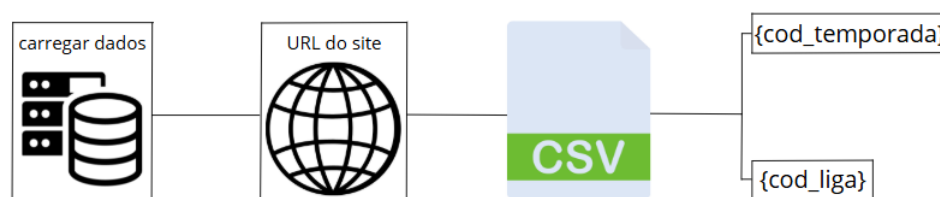
Os dados são carregados dinamicamente na interface da aplicação, permitindo ao usuário consultar os jogos do dia das ligas presentes na plataforma, com cotações e estatísticas pré jogo atualizadas.

A junção dessas duas fontes públicas, o histórico robusto do football-data.co.uk e os dados em tempo real do FootyStats via GitHub oferece uma base atualizada para alimentar o modelo preditivo. O uso de arquivos CSV também simplifica o tratamento dos dados via bibliotecas como pandas, eliminando a necessidade de bases relacionais complexas como SQLite. A estratégia adotada garantiu uma coleta de dados ética, gratuita e de programação mais simples

4.4 DADOS COLETADOS

Os arquivos são fornecidos no formato CSV que foram carregados diretamente da fonte de dados e contêm uma série de colunas com estatísticas técnicas das partidas, resultados, bem como cotações das principais casas de apostas.

Figura 4 – Lógica que carrega os dados diretamente do site e filtra por temporada e liga



Fonte: Elaborado pelo autor (2025)

Para este projeto, foram consideradas exclusivamente as odds da casa de apostas Bet365, além dos mercados Over 2.5 e Under 2.5 gols, por representarem os tipos de aposta mais populares e com maior liquidez no mercado. Os dados extraídos podem ser agrupados nas seguintes categorias:

4.4.1 Informações gerais da partida

- **Div:** Código da divisão (liga) a que pertence o jogo;
- **Date:** Data da partida;
- **Time:** Horário da partida;
- **HomeTeam / AwayTeam:** Nome do time mandante e visitante, respectivamente;
- **HTHG / HTAG:** Gols no primeiro tempo para mandante e visitante;
- **FTHG / FTAG:** Quantidade de gols marcados pelo mandante e pelo visitante (*Full Time Home/Away Goals*);
- **HTR:** Resultado parcial da partida ao término do primeiro tempo;
- **FTR:** Resultado final da partida (H = vitória do mandante, D = empate, A = vitória do visitante).

4.4.2 Estatísticas de desempenho

- **HS / AS:** Total de chutes efetuados pelo mandante e pelo visitante;
- **HST / AST:** Chutes no alvo por mandante e visitante;
- **HC / AC:** Número de escanteios cobrados por cada equipe;
- **HF / AF:** Faltas cometidas;
- **HY / AY:** Cartões amarelos;
- **HR / AR:** Cartões vermelhos.

4.4.3 Odds da Bet365

- **B365H / B365D / B365A:** Cotações (odds decimais) da Bet365 para os desfechos vitória do mandante, empate e vitória do visitante;
- **B365>2.5 / B365<2.5:** Odds da Bet365 para os mercados de gols: acima (*over*) e abaixo (*under*) de 2.5 gols.

4.4.4 Odds agregadas de mercado

- **P>2.5 / P<2.5:** Probabilidades médias do mercado para *Over* e *Under* 2.5 gols;
- **Max>2.5 / Max<2.5:** Maior odd disponível entre diversas casas para os mercados *Over/Under*;

- **Avg>2.5 / Avg<2.5:** Média aritmética das odds disponíveis no mercado para os respectivos desfechos.

Esses dados são carregados diretamente nos scripts Python do projeto por meio da biblioteca pandas. O sistema realiza a conversão adequada de tipos de dados (por exemplo, transformação de datas no formato datetime e conversão de odds para float) e procede à limpeza inicial dos registros com campos ausentes ou inválidos. Os dados passam, então, a compor o conjunto base para as etapas de processamento estatístico, análise exploratória e posterior aplicação dos modelos de predição.

4.5 PROCESSAMENTO ESTATÍSTICO

Após a coleta dos dados brutos, o sistema realiza um processo de agregação e síntese de informações com o objetivo de extrair estatísticas descritivas relevantes para cada equipe. Essas estatísticas são fundamentais tanto para alimentar o modelo preditivo quanto para possibilitar visualizações informativas na interface da plataforma.

Para cada time, o sistema irá considerar tanto os jogos anteriores à data de referência, quanto o desempenho geral e calcula indicadores separados conforme a equipe tenha atuado como mandante ou visitante. Essa distinção é necessária tendo em vista que o desempenho de muitas equipes apresenta diferenças significativas conforme o mando de campo (LAGO-PEÑAS, 2011). A metodologia aplicada contempla, portanto, uma visão segmentada e temporalmente coerente com o histórico disponível até o momento da predição. Os principais indicadores estatísticos extraídos por equipe incluem:

- **Média de gols marcados e sofridos:** representa a eficiência ofensiva e a solidez defensiva da equipe, servindo como base para avaliar tendências de placar e impacto potencial nos mercados de gols e resultado final;
- **Número de partidas disputadas como mandante e visitante:** garante o controle sobre o tamanho da amostra utilizada no cálculo das métricas agregadas;
- **Média de chutes totais e chutes no alvo (a favor e contra):** oferece uma perspectiva sobre a produtividade ofensiva e o volume de oportunidades criadas e sofridas;
- **Média de escanteios (pró e contra):** pode indicar o padrão de pressão ofensiva ou estilo de jogo adotado pela equipe;
- **Faltas cometidas e sofridas:** variáveis que, além de refletirem a intensidade defensiva, ajudam a compreender a propensão ao jogo físico;

- **Número de jogos com ao menos um gol marcado:** indica a regularidade ofensiva da equipe;
- **Número de jogos sem sofrer gols (clean sheets):** representa a consistência defensiva;
- **Conversão de finalizações em gols:** percentual calculado pela razão entre gols marcados e chutes totais, refletindo a eficiência ofensiva da equipe;
- **Cartões amarelos:** indicam propensão a sanções disciplinares, o que pode afetar o desempenho da equipe em partidas futuras.

Figura 5 – Processamento de dados para o processo de transformação em features



Fonte: <https://medium.com/@repositoriothiago/engenharia-de-dados-transformando-informa%C3%A7%C3%B5es-em-insights-valiosos-35a321b2af66>

4.6 ENGENHARIA DE ATRIBUTOS (FEATURES)

A engenharia de atributos é uma etapa fundamental na construção de modelos preditivos eficazes. Consiste na criação, seleção e transformação de variáveis que melhor representam os padrões históricos de desempenho das equipes. No contexto deste projeto, foram desenvolvidos atributos que capturam o comportamento estatístico recente e histórico dos times, a partir dos dados coletados.

Os atributos foram estruturados com base em dois grupos principais: desempenho estatístico agregado e dados de mercado (odds). A seguir, detalham-se os principais atributos utilizados no modelo e os critérios para sua inclusão.

4.6.1 Indicadores estatísticos de desempenho

Os indicadores foram calculados separando-se os jogos em que o time atuou como mandante ou visitante. As variáveis resultantes são:

- **Média de Gols Pró e Contra:** reflete a eficiência ofensiva e solidez defensiva do time.

- **Chutes Totais e no Alvo (Pró e Contra):** indicadores de pressão ofensiva e vulnerabilidade defensiva.
- **Escanteios a favor e contra:** sinalizam domínio territorial e presença ofensiva na área adversária.
- **Cartões Amarelos:** indicam agressividade e perfil disciplinar da equipe.
- **Porcentagem de Jogos com Gol Marcado:** avalia a consistência ofensiva.
- **Porcentagem de Jogos sem Sofrer Gols:** indica solidez defensiva.
- **Conversão de Chutes em Gols:** mede a eficiência na finalização.
- **Total de Jogos Disputados:** garante a validade estatística das médias.
- **Forma Recente:** representa o desempenho da equipe nos últimos 5 jogos (vitórias, empates e derrotas).

4.6.2 Indicadores de mercado (Odds)

Foram utilizadas exclusivamente as odds oferecidas pela casa de apostas Bet365, devido à sua ampla cobertura, confiabilidade e relevância no mercado. As cotações foram utilizadas tanto na forma atual (do jogo a ser previsto) quanto como médias históricas, conforme abaixo:

- **Odd Atual de Vitória do Visitante (B365A);**
- **Odds Históricas Médias do Time Visitante** nas últimas partidas como visitante (média das odds Bet365 para vitória, empate e derrota);
- **Odds Históricas Médias do Time Mandante** nas últimas partidas como mandante (idem);
- **Odds para Over e Under 2.5 gols (Bet365):** representam a expectativa de gols do mercado, que pode estar correlacionada com a tendência ofensiva de ambas as equipes.

Essas variáveis de odds são posteriormente transformadas em probabilidades implícitas, utilizando a fórmula $P=1$, permitindo que o modelo utilize estimativas de mercado como uma feature contínua. A presença de odds no vetor de entrada também permite o cálculo posterior do valor esperado (Expected Value) de cada aposta, combinando a probabilidade estimada pelo modelo com a odd real oferecida pela casa.

4.7 NORMALIZAÇÃO DOS DADOS

A normalização dos dados é uma etapa essencial no desenvolvimento de modelos de aprendizado de máquina, especialmente quando diferentes atributos possuem escalas e unidades variadas. Quando não tratadas adequadamente, essas discrepâncias podem influenciar de maneira desproporcional o processo de aprendizado, prejudicando a generalização do modelo (CAPITAL, 2022). No projeto MoneyBall, adotou-se uma estratégia de normalização leve e pontual, ajustada à natureza dos dados disponíveis, sem recorrer a transformações paramétricas agressivas. As transformações aplicadas seguiram critérios específicos conforme o tipo e comportamento das variáveis.

4.7.1 Normalização de proporções e frequências

Indicadores como "% de jogos com gol" e "% de jogos sem sofrer gol" foram calculados pela razão entre ocorrências e total de jogos válidos, garantindo valores no intervalo [0,1]. Essa abordagem permite comparabilidade entre equipes que disputaram número diferente de partidas, para que o modelo não leve em consideração o fato de time x ter mais jogos disputados que time y seja necessariamente uma vantagem.

4.7.2 Conversão de odds em probabilidades implícitas

As odds fornecidas pela casa de apostas Bet365 foram convertidas em probabilidades implícitas, aplicando-se a fórmula:

$$\text{Probabilidade Implícita} = \frac{1}{\text{Odd Decimal}}$$

Esse procedimento tornou possível comparar as probabilidades previstas pelo modelo com as probabilidades oferecidas pelo mercado, etapa indispensável para o cálculo do valor esperado (Expected Value) de uma aposta. A transformação também traz os valores para a escala probabilística [0,1], homogênea com as demais variáveis estatísticas.

4.7.3 Padronização de métricas contínuas

Variáveis contínuas como gols marcados, escanteios, chutes e cartões foram padronizadas com base em médias por jogo, em vez de totais acumulados. Isso permite ao modelo captar o desempenho relativo da equipe, independentemente do número absoluto de partidas, o que é especialmente importante em janelas curtas.

4.7.4 Tratamento de dados ausentes e times recém-promovidos

Para times com histórico recente insuficiente (geralmente recém-promovidos), foi adotada a regra de exclusão de variáveis estatísticas caso o número de partidas anteriores à data da previsão fosse inferior a um mínimo (tipicamente 6 partidas). Nessas situações, os jogos foram ignorados na geração das features, e, quando aplicável, os campos resultantes foram deixados como nulos e descartados do vetor de entrada. Essa abordagem visou preservar a integridade do modelo, evitando introdução de ruído ou inferências artificiais baseadas em amostras estatisticamente irrelevantes.

4.7.5 Justificativa técnica

Optou-se por evitar transformações paramétricas (como `StandardScaler` ou `MinMaxScaler`) da biblioteca `scikit-learn`, dado que os modelos baseados em árvores de decisão (como o `Random Forest` utilizado) não são sensíveis à escala das variáveis (CHAWLA, 2023). Entretanto, a normalização aplicada foi suficiente para garantir coerência estatística, consistência semântica e evitar distorções por escala, mantendo a interpretabilidade do modelo e sua capacidade de generalização.

4.8 MODELO PREDITIVO MONEYBOT

Para desenvolver o Modelo Preditivo MoneyBot foi implementado um sistema de aprendizado de máquina que opera em dois mercados complementares de apostas no futebol: *Back Visitante* (apostar a favor da vitória do visitante) e *Lay Mandante* (apostar contra a vitória do mandante). Ambos os modelos compartilham a mesma base metodológica: (I) engenharia de atributos cronologicamente consistente; (II) pipeline com *Random Forest* e filtragem por variância; (III) calibração probabilística; (IV) validação temporal do tipo *walk-forward* com refit periódico; e (V) decisão de entrada pautada por valor esperado (EV). Adicionalmente, o sistema suporta tanto *backtesting* por temporada quanto simulações prospectivas de jogos futuros mediante o preenchimento manual de odds, data e confronto. Os modelos são capazes de simular e predir os jogos das 5 principais ligas europeias (**Alemanha, Espanha, França, Itália e Inglaterra**). A escolha dos métodos de apostas (apostar a favor da vitória do visitante e apostar contra a vitória do mandante) se dá pelo fato do mercado geralmente superestimar o mandante, possibilitando assim apostas com EV (valor esperado) realmente positivo.

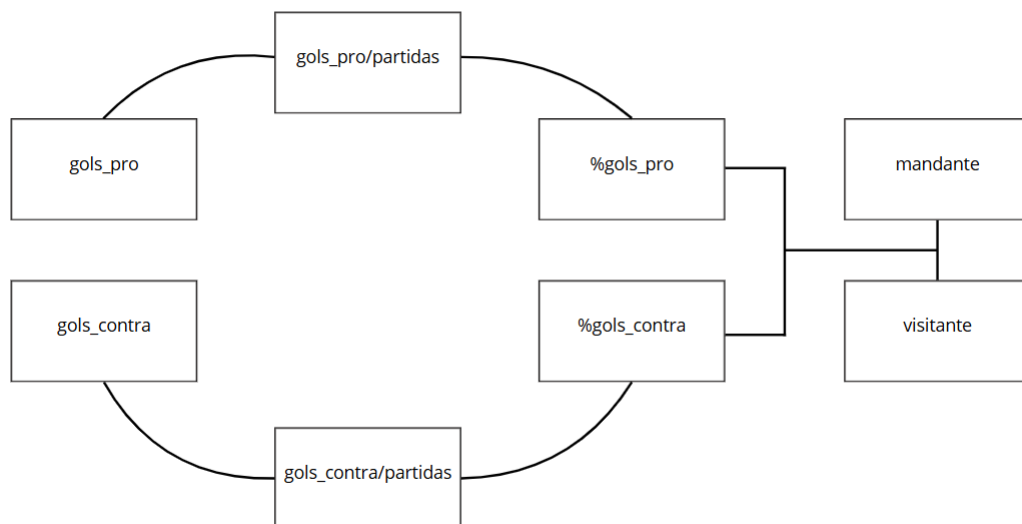
4.8.1 Engenharia de Atributos (Feature Engineering)

A construção de atributos segue o princípio de causalidade temporal: cada vetor de *features* de uma partida é calculado apenas com dados anteriores à data do

jogo. São contempladas estatísticas históricas agregadas (gols pró/contra, finalizações, escanteios, pontos, aproveitamento em casa/fora), forma recente (últimos 5 jogos), e *features* de expectativa de mercado via *odds* históricas por mando, segmentadas por mandante/visitante. O conjunto é prefixado com **Mandante_** e **Visitante_** para distinguir os lados do confronto.

Para preservar a qualidade dos atributos, partidas sem histórico suficiente são ignoradas no treino: exigem-se mínimo de 6 jogos prévios e ausência de times recém-promovidos no instante da previsão. Tais regras são aplicadas com a mesma lógica nas rotinas de previsão/simulação de ambos os modelos (Back e Lay), garantindo consistência e evitando vazamento de informação.

Figura 6 – Exemplo de processo de filtragem de Features



Fonte: Elaborado pelo autor (2025)

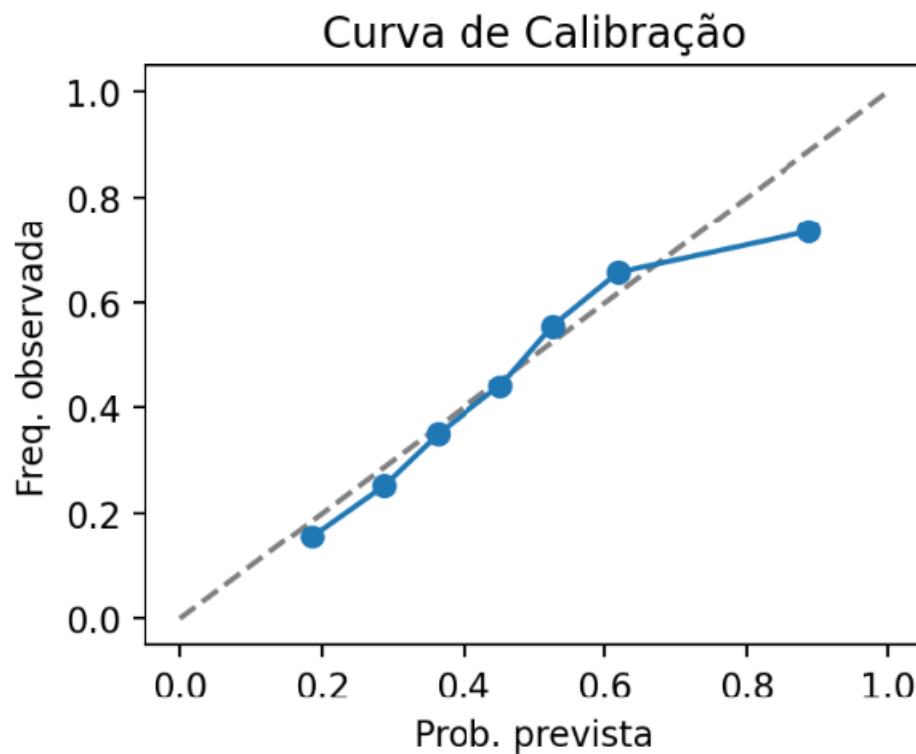
4.8.2 Biblioteca *Scikit-Learn*, Pipeline e Calibração

O núcleo de aprendizado do projeto MoneyBot foi estruturado sobre a biblioteca *Scikit-Learn*, uma das ferramentas mais consolidadas e amplamente utilizadas em projetos de aprendizado de máquina com Python. A escolha dessa biblioteca foi estratégica por três motivos principais: (I) sua modularidade e padronização, que permitem compor fluxos complexos de modelagem em etapas reproduzíveis e transparentes; (II) sua integração nativa com estruturas de dados (*DataFrames*) do *pandas*, viabilizando a manipulação direta dos atributos estatísticos de cada partida; e (III) a disponibilidade de classes robustas para pré-processamento, ajuste de hiperparâmetros, calibração de probabilidades e avaliação temporal, elementos essenciais para a confiabilidade preditiva exigida em apostas esportivas.

O fluxo de modelagem foi encapsulado em um *pipeline* com **VarianceThreshold**(remoção de *features* quase constantes) seguido de um **RandomForestClas-**

sifier configurado para robustez e generalização. Para lidar com desbalanceamento (especialmente no Back Visitante), utiliza-se **class_weight='balanced'**. A calibração é feita com **CalibratedClassifierCV**, adotando *sigmoid* ou *isotonic* conforme o caso e a disponibilidade de *holdout/folds*. No Back Visitante, a calibração é treinada por validação cruzada estratificada dinâmica (2–3 *folds*, conforme suporte de classes no instante); no Lay Mandante, há uma rotina de calibração *profit* em *holdout* temporal (priorizando *isotonic* e caindo para *sigmoid* se necessário).

Figura 7 – Representação da curva de calibração do modelo – probabilidade prevista vs. frequência observada de vitória do visitante – idealmente próxima da diagonal que representa perfeita calibração



Fonte: Elaborado pelo autor (2025)

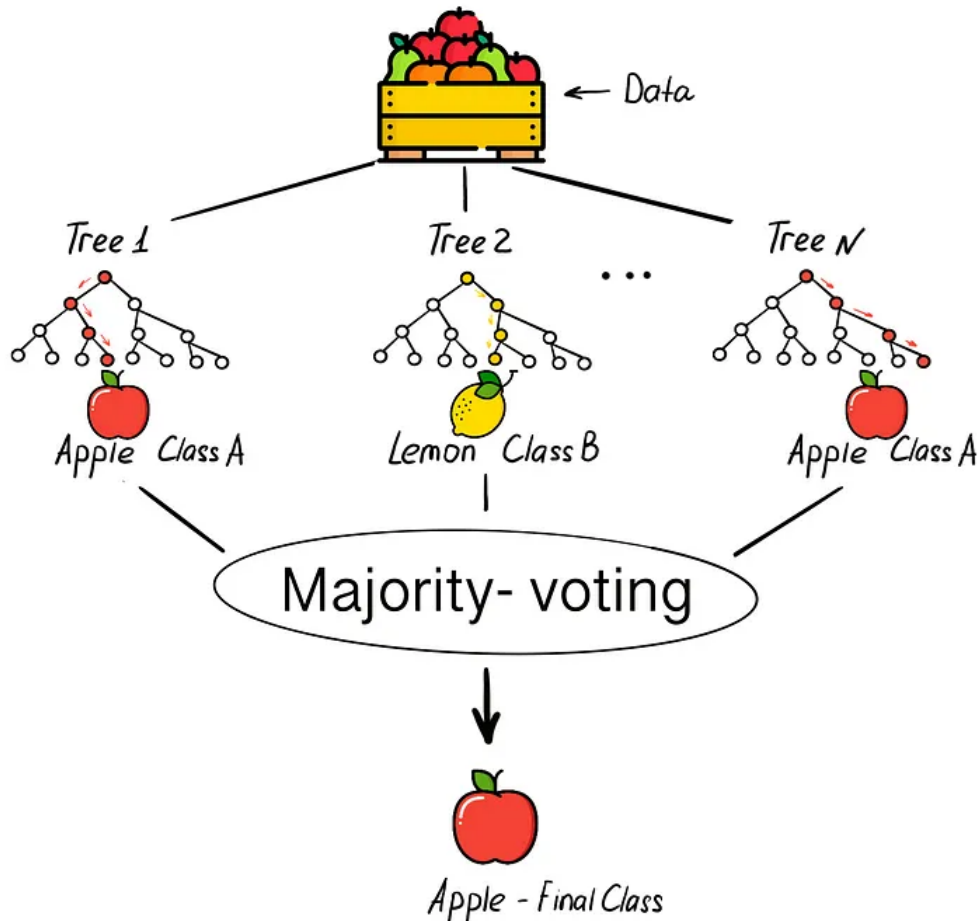
4.8.3 Classificador Random Forest

Para realizar a predição de resultados, optou-se por um algoritmo de classificação do tipo Random Forest (Floresta Aleatória). O RandomForestClassifier foi escolhido por diversas razões técnicas e de desempenho:

- É um método de ensemble de árvores de decisão, combinando previsões de múltiplas árvores para melhorar a generalização. Cada árvore é treinada em amostras aleatórias do conjunto de dados (com reposição, no processo de bagging (GUSAROVA, 2022)) e seleciona aleatoriamente um subconjunto de features em cada divisão. Esse mecanismo tende a reduzir overfitting (ineficácia para

prever novos resultados) de árvores individuais e a capturar relações não lineares entre as variáveis.

Figura 8 – Processo de votação da floresta aleatória, com exemplo de 3 árvores.



Fonte: <https://medium.com/%40data.science.enthusiast/understanding-random-forest-hyperparameters-tunning-python-code-136c3b1ce9e>

- Apresenta boa robustez a outliers (valor atípico) e escalas distintas das features, já que as árvores de decisão internamente dividem os dados por critérios hierárquicos e não são sensíveis a distribuições não normalizadas. Isso dispensa a necessidade de normalização estrita de todas as variáveis, simplificando o pré-processamento.
- Suporta naturalmente features categóricas ou numéricas e lida bem com conjuntos de dados com muitas variáveis explicativas, realizando seleção implícita de atributos importantes durante a construção das árvores. No caso do MoneyBot, há dezenas de atributos (estatísticas de desempenho, forma, odds etc.), e o Random Forest consegue identificar automaticamente quais são mais relevantes para prever a vitória do visitante.

- O hiperparâmetro `class_weight='balanced'` foi empregado no modelo Back Visitante para lidar com o possível desbalanceamento de classes. No problema em questão, a proporção de vitórias do time visitante geralmente é menor que a de vitórias do mandante ou empates (ou seja, partidas tendem a ter o mandante como vencedor com mais frequência no futebol, tornando “vitória do visitante” uma classe minoritária). Ao usar ponderação balanceada, o algoritmo dá peso maior às instâncias da classe menos frequente durante o treino, equilibrando a contribuição de acertos em ambas as classes (visitante vence vs. não vence). Isso ajuda a evitar que o modelo tenha viés excessivo em prever sempre a classe majoritária.

Os hiperparâmetros do Random Forest foram ajustados de forma a controlar a complexidade do modelo e prevenir sobreajuste. Utilizou-se, por exemplo, 100 árvores na floresta (`n_estimators=100`), o que geralmente é suficiente para estabilizar as estimativas sem incorrer em custo computacional desnecessário (SAXENA, 2024). Limitou-se a profundidade máxima de cada árvore (`max_depth=10`) e o mínimo de amostras em cada folha (`min_samples_leaf=3`), forçando as árvores a não se especializarem demais nos dados de treino (evitando divisões muito profundas ou ramos terminal com pouquíssimas instâncias). Além disso, um `random_state=42` foi fixado para garantir reprodutibilidade dos resultados, já que o treinamento do Random Forest envolve aleatoriedade.

Por fim, o classificador foi integrado em um pipeline de processamento junto com um filtro de variância: antes de treinar o Random Forest, aplica-se a transformação `VarianceThreshold(threshold=0.01)`. Este passo remove quaisquer features quase constantes (variância menor que 0,01), que pouco acrescentam em termos de informação. Por exemplo, se alguma variável tiver praticamente o mesmo valor para todas as partidas, ela é eliminada para não atrapalhar o treino (poupando recursos computacionais e evitando ruído). Na prática, esse filtro assegura que apenas atributos com variação suficiente (e potencial informativo) sejam considerados pelo modelo.

4.8.4 Validação Walk-Forward com Refit Periódico

Para simular condições reais, as previsões seguem um esquema walk-forward: o histórico cresce jogo a jogo e o classificador é periodicamente re-treinado (refit) com todos os dados disponíveis até cada ponto no tempo.

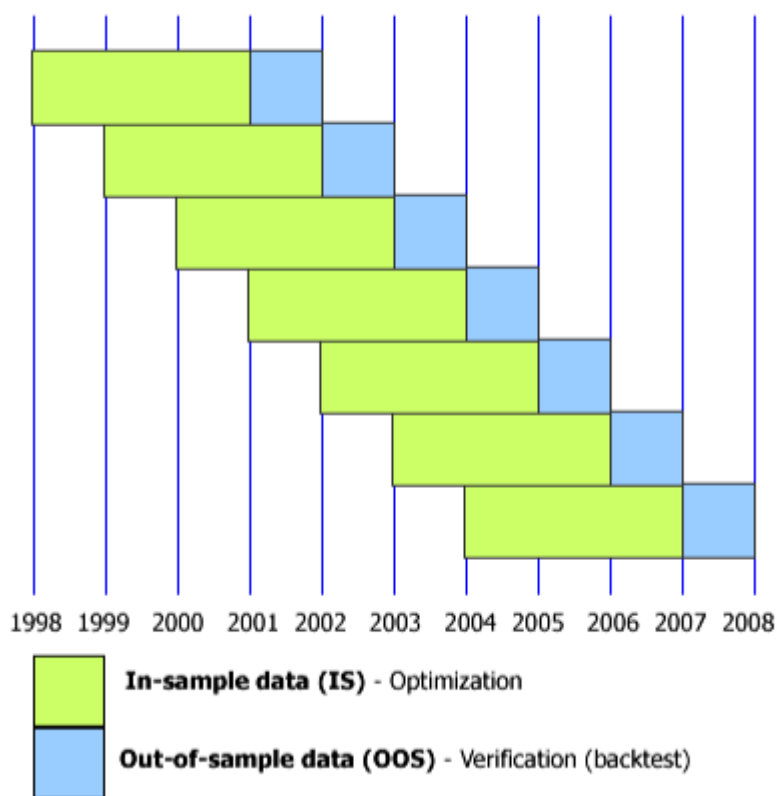
Back Visitante (temporada): **simular_periodo** executa walk-forward com parâmetro **refit_every = 7**, recalibrando o modelo a cada 7 jogos(dados) e removendo do treino quaisquer colunas ligadas a odds reais de mercado (**Odd_Real/Odd Atual**) para evitar vazamento.

Lay Mandante (temporada): **simular_periodo_lay** aplica um walk-forward

rápido com `refit_every= 25`. O treinamento+calibração é feito em uma etapa do tipo *prefit+holdout temporal* (com `holdout_frac` e `min_cal`) para estabilidade e velocidade.

Coerência com simulação de jogos futuros: as rotinas **simular_jogos** Back e Lay replicam exatamente o estado que o walk-forward de temporada teria até a data do jogo futuro, respeitando o último ponto de *refit* e o *seed* temporal (início da temporada). Isso garante que a decisão prospectiva use o mesmo conhecimento que estaria disponível no fluxo sequencial real (FAHAD, 2024).

Figura 9 – Linha do tempo de aprendizado com Walk-Foward
Walk-Forward Test procedure



Fonte: https://www.amibroker.com/guide/h_walkforward.html

4.8.5 Simulação de Apostas, Avaliação de Resultados e Critérios de Valor Esperado

A fase de simulação do *MoneyBot* representa a etapa em que o modelo preditivo é efetivamente colocado à prova, aplicando as previsões geradas para tomar decisões de aposta e mensurar os resultados financeiros. Essa etapa foi dividida em duas modalidades principais de execução: a simulação de **temporadas passadas completas** e a simulação de **jogos futuros**, sendo ambas compatíveis com as estratégias de *Back Visitante* e *Lay Mandante*. O objetivo central dessa arquitetura é reproduzir o comportamento de um apostador automatizado, garantindo que as previsões sejam

feitas apenas com base em informações anteriores à data da partida, por meio de um ciclo de validação walk-forward com refit sequencial do modelo.

O processo é inteiramente reproduzido tanto para o modelo *Back Visitante* quanto para o modelo *Lay Mandante*, com pequenas diferenças de lógica e cálculo entre eles.

4.8.5.1 Simulação de Temporadas

A simulação de temporadas é conduzida pelas funções principais implementadas **simular_jogos_back** e **simular_jogos_lay**. Ambas executam a simulação completa de uma liga (por exemplo, a Premier League), partida a partida, respeitando a ordem cronológica e aplicando o método **walk-forward com refit(7 para back e 25 para lay)**. Isso significa que o modelo é reentrenado a cada jogo utilizando todo o histórico disponível até aquele momento, incorporando inclusive os jogos anteriores da própria temporada, o que impede o vazamento de dados e simula o aprendizado contínuo (BROWNLEE, 2019).

A função percorre o conjunto de partidas em sequência temporal. Para cada uma:

1. Gera o vetor de atributos da partida utilizando a rotina de engenharia de *features*;
2. Reajusta o modelo de classificação (Random Forest calibrado) com todos os jogos anteriores à data atual;
3. Calcula a probabilidade predita do evento (vitória do visitante ou não vitória do mandante, conforme o método);
4. Obtém a *odd* de mercado correspondente e calcula o **Valor Esperado (EV)**;
5. Avalia se os critérios mínimos de aposta são atendidos e, em caso positivo, executa a aposta simulada.

Esse processo é registrado em um *DataFrame*, contendo colunas com data, times, odds, probabilidade prevista, valor esperado, resultado da aposta, lucro/prejuízo e saldo acumulado. O fluxo é projetado para ser iterativo e transparente, permitindo visualizar a evolução da banca ao longo da temporada.

Figura 10 – Histórico da simulação da temporada 24/25 (Premier League) com o método back visitante

MoneyBot - Selecione a Liga Liga: Premier League	Brentford 2x3 Fulham 2025-05-18 Odd Casa: 1.91 Odd Empate: 3.9 Odd Fora: 3.6 Odd Fora Ajustada: 2.94 Prob (modelo): 34.02% Implícita: 27.78% EV: 0.2247 Resultado final: A → ACERTO Banca: R\$ 2522.50
	Crystal Palace 4x2 Wolves 2025-05-20 Odd Casa: 2.25 Odd Empate: 3.3 Odd Fora: 3.2 Odd Fora Ajustada: 2.62 Prob (modelo): 38.16% Implícita: 31.25% EV: 0.2211 Resultado final: H → ERRO Banca: R\$ 2472.50
	Ipswich 1x3 West Ham 2025-05-25 Odd Casa: 3.0 Odd Empate: 3.8 Odd Fora: 2.2 Odd Fora Ajustada: 1.48 Prob (modelo): 67.34% Implícita: 45.45% EV: 0.4815 Resultado final: A → ACERTO Banca: R\$ 2532.50
	Newcastle 0x1 Everton 2025-05-25 Odd Casa: 1.33 Odd Empate: 5.5 Odd Fora: 8.5 Odd Fora Ajustada: 3.48 Prob (modelo): 28.74% Implícita: 11.76% EV: 1.4431 Resultado final: A → ACERTO Banca: R\$ 2907.50
	Wolves 1x1 Brentford 2025-05-25 Odd Casa: 2.7 Odd Empate: 3.7 Odd Fora: 2.45 Odd Fora Ajustada: 2.06 Prob (modelo): 48.61% Implícita: 40.82% EV: 0.1909 Resultado final: D → ERRO Banca: R\$ 2857.50

Fonte: Elaborado pelo autor (2025)

Durante o *loop* principal, o código também realiza o controle de consistência de dados. Caso algum jogo não possua informações completas (por exemplo, ausência de odds ou histórico insuficiente para gerar *features*), a função ignora aquele confronto e segue para o próximo. Essa decisão garante robustez e evita distorções estatísticas, já que previsões em condições de incerteza extrema tenderiam a gerar ruído.

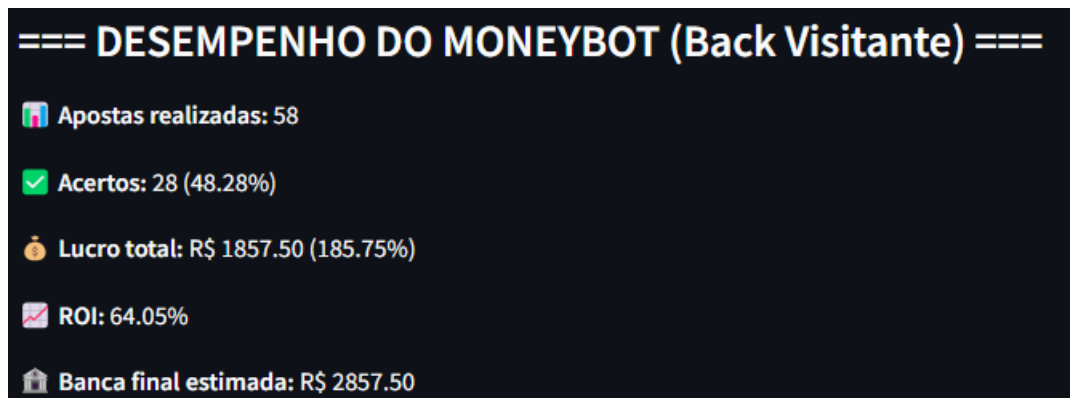
Ao término da simulação, são calculadas métricas agregadas como:

- Apostas realizadas
- Acertos e porcentagem de taxa de acertos
- Lucro líquido total;
- ROI (Retorno sobre o investimento);
- Banca final;

O resultado da figura 11 foi obtido através de uma simulação da temporada 2024/2025 da Premier League após a rodada 7 utilizando o método Back Visitante. Foram feitos testes no começo da temporada e após a rodada 7 e, a diferença entre o ROI foi de quase 25%, e a diferença no total do lucro foi maior que 50%, obtendo

um lucro médio de 130% treinando a partir da primeira rodada, contra um lucro médio de 185% após a rodada 7. A decisão de recomendar a predição do MoneyBot após 7 rodadas vem do trecho do livro do renomado treinador Pep Guardiola (PERARNAU, 2015), que afirma que a partir de 8 rodadas, podemos tentar cravar quais os times que conseguem lutar pelo título até o fim da temporada. Ele afirma que até a rodada 8, os times que tiverem mais que 4 pontos atrás do líder do campeonato possuem pouquíssimas chances de lutar pelo título e de fato, se formos analisar o campeonato brasileiro desde 2007, em 63% dos casos isso realmente ocorreu, e se formos levar essa pesquisa para campeonatos europeus o número será ainda maior. Baseado nesses dados, concluímos que será bem mais eficaz o uso do MoneyBot após a rodada 7.

Figura 11 – Resultado da simulação da temporada 24/25 (Premier League) após a rodada 7 com o método back visitante



Fonte: Elaborado pelo autor (2025)

4.8.5.2 Simulando Temporadas “Lay Mandante”

A função responsável pela simulação de temporadas na estratégia *Lay Mandante* implementa uma lógica de cálculo e decisão adaptada ao perfil dessa modalidade. Nesse caso, o modelo prevê a probabilidade de **não vitória do mandante** ($P(\text{empate ou vitória do visitante})$), utilizando a odd do mandante como referência.

Após a previsão da probabilidade calibrada q , o sistema calcula o valor esperado adaptado para apostas *lay*:

$$EV_{\text{lay}} = q \times O_H - (O_H - 1)$$

onde O_H representa a odd do mandante. O valor esperado é positivo quando a probabilidade estimada de não vitória supera a expectativa implícita da odd de mercado (SMITH, 2025).

Os critérios de decisão seguem três condições principais:

1. $EV_{\text{lay}} > 0,05$;

2. Odd do mandante mínima de 1,20;
3. Odd do mandante máxima de 3,00.

Quando essas condições são atendidas, a função registra uma aposta lay e calcula o lucro ou prejuízo conforme o resultado da partida:

$$\text{Lucro}_{lay} = \begin{cases} +1 & \text{se o mandante não vence,} \\ -(O_H - 1) & \text{se o mandante vence.} \end{cases}$$

A simulação completa das temporadas *Lay Mandante* permite observar padrões complementares aos do modelo *Back Visitante*, reduzindo a volatilidade da banca simulada.

4.8.5.3 Simulando Temporadas “Back Visitante”

A função de simulação *Back Visitante* segue a mesma arquitetura de iteração temporal, porém voltada à previsão da probabilidade de vitória do visitante $p = P(\text{visitante vence})$. O valor esperado é calculado pela equação clássica:

$$EV_{back} = (p \times O_A) - 1$$

onde O_A é a odd do visitante. A função realiza apostas apenas quando:

1. $EV_{back} > 0,10$;
2. $p \geq 0,25$.

Para cada aposta realizada, o resultado é calculado como:

$$\text{Lucro}_{back} = \begin{cases} (O_A - 1) & \text{se o visitante vence,} \\ -1 & \text{caso contrário.} \end{cases}$$

A simulação de temporadas nesse modo é essencial para avaliar a robustez do modelo sob diferentes janelas de treino e identificar as condições de maior eficiência estatística.

4.8.5.4 Simulação de Jogos Futuros

O módulo de simulação de jogos futuros foi criado para permitir a aplicação prospectiva do *MoneyBot*, simulando apostas em confrontos ainda não realizados.

O funcionamento segue a mesma lógica do ciclo *walk-forward*, mas restrita a um conjunto de até 10 jogos que ainda não estão salvos no banco de dados. A função recebe como entrada:

- Data e times do confronto;

- Odds disponíveis (mandante, empate e visitante);
- Temporada e base de dados histórica já consolidada.

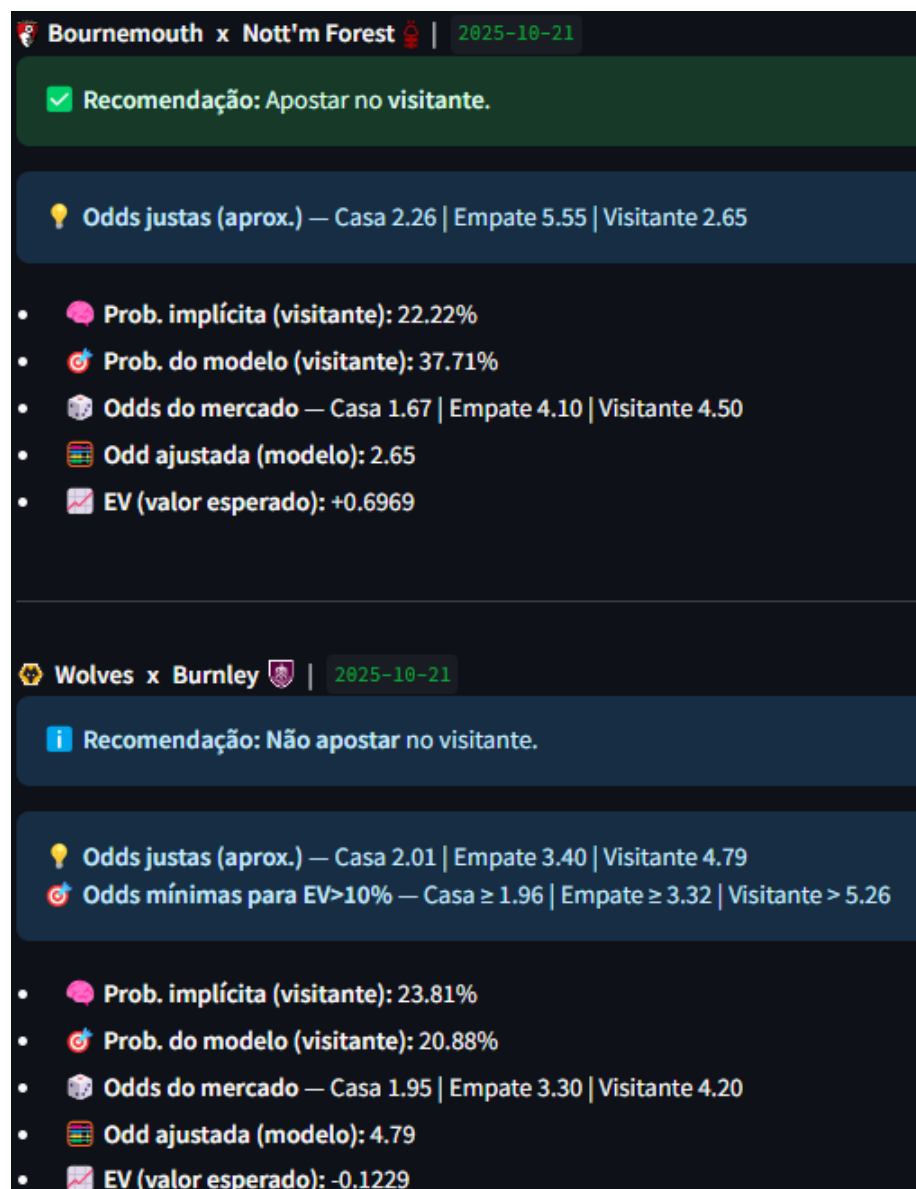
A rotina identifica automaticamente todas as partidas anteriores à data informada e treina o modelo calibrado com aquele histórico. Em seguida, gera o vetor de *features* atualizadas para o jogo em questão e produz a probabilidade calibrada de ocorrência do evento-alvo (*vitória do visitante* ou *não vitória do mandante*). O valor esperado é calculado com as mesmas fórmulas e critérios definidos anteriormente no módulo de simular temporadas.

O resultado da simulação é exibido em formato tabular, contendo:

- Probabilidade estimada pelo modelo;
- Odd de mercado e odd ajustada (aproximada);
- Odds mínimas para a aposta ser matematicamente lucrativa ($EV > 10\%$)
- Valor esperado (EV);
- Decisão de aposta (sim/não).

Esse mecanismo permite que o usuário simule jogos futuros manualmente, preenchendo as odds e o confronto desejado, obtendo em tempo real a análise probabilística e a recomendação de aposta, conforme ilustrado na Figura 14.

Figura 12 – Recomendações de apostas para jogos futuros no método back visitante.



Fonte: Elaborado pelo autor (2025)

4.8.5.5 Simulando Jogos “Lay Mandante”

A função de simulação de jogos *Lay Mandante* realiza o mesmo fluxo do modo *temporadas*, mas aplicado a uma única partida. Após o treinamento incremental do modelo, calcula-se $q = P(\text{mandante não vence})$ e o valor esperado EV_{lay} com base na odd do mandante. O retorno esperado, responsabilidade e decisão final de aposta são apresentados ao usuário com base nas mesmas restrições de simular temporadas.

Essa função foi projetada para uso interativo, permitindo testar hipóteses em tempo real, por exemplo, avaliar se determinado jogo futuro apresenta valor estatístico para uma aposta contra o mandante.

4.8.5.6 Simulando Jogos “Back Visitante”

De modo análogo, a função de simulação de jogos *Back Visitante* calcula $p = P(\text{visitante vence})$ e o valor esperado EV_{back} , fornecendo as probabilidades ajustadas, o valor esperado e a decisão final de aposta. A lógica reflete fielmente o comportamento de uma aposta real de mercado, em que o modelo avalia se as condições estatísticas e de valor estão alinhadas com os critérios de entrada.

Esse módulo é amplamente utilizado para análises prospectivas e monitoramento das previsões do modelo, fornecendo subsídios para comparação entre as probabilidades calculadas e o resultado real após o jogo. O uso combinado das duas modalidades (*Back* e *Lay*) proporciona ao *MoneyBot* flexibilidade analítica e capacidade de decisão em múltiplos contextos de mercado. Para realizar a predição de jogos futuros, o usuário precisará selecionar a quantidade de jogos que o modelo irá predir, escolher os times (mandante e visitante), selecionar até qual data os dados devem ser carregados e imputar as odds de mercados para o determinado confronto que irá acontecer.

Figura 13 – Tela de seleção de quais jogos o usuário irá realizar a predição back visitante.

A interface 'Simular jogo (Back Visitante)' apresenta os seguintes elementos:

- Quantidade de jogos (1 a 10):** Um menu suspenso com o valor '10' selecionado.
- Seleção até que data os jogos devem ser carregados:** Um campo de data com o valor '2025/10/21'.
- Partida 1:**

Mandante	Visitante	Odd Casa	Odd Empate	Odd Visitante
Leeds	West Ham	1,83 - +	3,40 - +	4,33 - +
- Partida 2:**

Mandante	Visitante	Odd Casa	Odd Empate	Odd Visitante
Chelsea	Sunderland	1,42 - +	4,75 - +	7,00 - +

Fonte: Elaborado pelo autor (2025)

4.9 FERRAMENTAS ANALÍTICAS

Além do modelo preditivo MoneyBot, a plataforma MoneyBall oferece diversos painéis analíticos complementares. Essas funcionalidades visam prover ao usuário análises estatísticas detalhadas e contextuais, fortalecendo a tomada de decisão racional em apostas esportivas. Nesta seção iremos descrever as funcionalidades implementadas, detalhando seu propósito e abordagem técnica.

4.9.1 Classificação

A funcionalidade de Classificação fornece ao usuário uma tabela classificatória dinâmica de cada liga selecionada, permitindo visualizar a posição dos times e seus pontos em determinado recorte de campeonato. O propósito principal é contextualizar o desempenho relativo das equipes ao longo da temporada, sem a necessidade de recorrer a outros sites e aplicativos. Na plataforma, o usuário pode filtrar a classificação por liga, temporada e por intervalo de datas específicas. Também é possível filtrar a tabela por mando de campo exibindo o desempenho geral, somente em casa ou apenas fora, bem como por turno (primeira ou segunda metade da temporada). Esses filtros interativos fornecem uma visão refinada da performance, por exemplo destacando se uma equipe acumula mais pontos como mandante ou como visitante.

Figura 14 – Tela de classificação do MoneyBall

Filtros

Liga

Inglterra

Temporada

2025/2026

Mando de Campo

Mando

Geral

Mandante

Visitante

Período

Data inicial

2025/08/15

Data final

2025/10/20

Tabela Classificatória

Classificação Inglaterra - 2025/2026

Posição	Time	Pts	PJ	V	E	D	Aprv%	GP	GC	SG
1	Arsenal	19	8	6	1	1	79.2	15	3	12
2	Man City	16	8	5	1	2	66.7	17	6	11
3	Liverpool	15	8	5	0	3	62.5	14	11	3
4	Bournemouth	15	8	4	3	1	62.5	14	11	3
5	Tottenham	14	8	4	2	2	58.3	14	7	7
6	Chelsea	14	8	4	2	2	58.3	16	9	7
7	Sunderland	14	8	4	2	2	58.3	9	6	3
8	Crystal Palace	13	8	3	4	1	54.2	12	8	4
9	Man United	13	8	4	1	3	54.2	11	12	-1
10	Brighton	12	8	3	3	2	50.0	12	11	1
11	Aston Villa	12	8	3	3	2	50.0	8	8	0
12	Everton	11	8	3	2	3	45.8	9	9	0

Fonte: Elaborado pelo autor (2025)

Do ponto de vista técnico, a plataforma recalcula a classificação em tempo real a partir dos dados históricos armazenados. Os resultados das partidas são agregados para computar vitórias, empates, derrotas, gols pró e contra, e pontos de cada time, replicando a lógica de pontuação tradicional (3 pontos por vitória, 1 por empate) dentro do período filtrado. Uma característica adicional é a visualização gráfica da evolução da posição de um time ao longo das rodadas: ao selecionar uma equipe, a aplicação plota um gráfico de linha mostrando em que colocação aquele time esteve após cada rodada da temporada. Essa visualização temporal ajuda a identificar tendências de performance como asensão e queda na tabela.

5 Evolução de um Time na Tabela

Selecione um Time

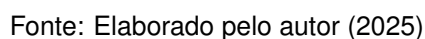
Man City

Evolução do Man City na Tabela

Rodada	Posição
1	4
2	1
3	1
4	1
5	1
6	2
7	3
8	1
9	2
10	2
11	2
12	5
13	4
14	4
15	6
16	6
17	7
18	6
19	6
20	7
21	6
22	4
23	6
24	4
25	4
26	4
27	5
28	5
29	5
30	6
31	5
32	5
33	4
34	3
35	4
36	2
37	2
38	2

4.9.2 Jogos por rodada

Figura 16 – Rodadas e jogos disponíveis para detalhar



No aspecto técnico, ao selecionar uma liga, temporada e um número de rodada, a aplicação lista todas as partidas daquela rodada com seus placares finais. O usuário então pode selecionar uma partida específica para ver suas estatísticas. Os dados exibidos incluem o placar final e o placar ao intervalo, indicando a evolução do resultado, bem como uma comparação de estatísticas de jogo entre mandante e visitante: número

de escanteios de cada lado, total de finalizações e finalizações no alvo, faltas cometidas por cada equipe, cartões e etc.

Esses dados são extraídos diretamente do dataset histórico de partidas, onde cada jogo possui essas estatísticas coletadas. A interface apresenta esses valores em colunas lado a lado, similar ao comparativo, facilitando a leitura do domínio de cada indicador pelo mandante ou visitante. Um complemento importante é que, junto a essas estatísticas, o sistema exibe a posição de cada time na tabela antes daquela partida, isto é, consultando a classificação acumulada até a rodada anterior. Essa informação contextual (por exemplo, o mandante entrou na rodada como 3º colocado vs. o visitante como 15º) ajuda a julgar o resultado à luz da expectativa prévia. Internamente, isso é implementado recalculando a classificação parcial até a rodada anterior e extraíndo a colocação de cada um dos dois clubes naquela partida. Em resumo, a arquitetura combina dados de desempenho do jogo específico com dados históricos acumulados até então.

Figura 17 – Estatísticas da partida entre Tottenham e Brighton



Fonte: Elaborado pelo autor (2025)

4.9.3 Estatísticas avançadas por time

A ferramenta de Estatísticas Avançadas por Time aprofunda a análise ao fornecer um panorama estatístico completo de um único time em relação a diversos indicadores de performance. O propósito é permitir que o usuário estude detalhadamente uma equipe específica, entendendo seus pontos fortes e fracos através de métricas quantitativas, ao invés de se basear apenas em resultados ou impressões gerais. Essa funcionalidade está alinhada com a ideia de análise exploratória e fundamentada em dados: o apostador pode, por exemplo, selecionar um time que considere promissor e verificar se os números sustentam essa percepção.

Em termos de metodologia, a plataforma compila uma série de indicadores de desempenho agregados daquele time na temporada (ou período filtrado) e ranqueia o time nesses indicadores em comparação aos demais da liga. Dentre as estatísticas

calculadas estão: média de gols marcados e gols sofridos por jogo, que refletem sua eficiência ofensiva e solidez defensiva; média de chutes totais e chutes no alvo por jogo (a favor e contra), indicando volume de oportunidades criadas e concedidas; média de escanteios a favor e contra, que pode sinalizar domínio territorial ou estilo de jogo; média de faltas cometidas e recebidas, relacionada à intensidade de marcação e jogo físico; e percentual de conversão de chutes em gol, que mede a eficácia ofensiva. Ademais, são apresentados indicadores de regularidade, como o número de jogos em que marcou gol e jogos sem sofrer gol, evidenciando consistência defensiva.

A ferramenta também computa as “séries” ou sequências da equipe: maiores sequências invictas, de vitórias consecutivas ou de derrotas ao longo da temporada, bem como as séries ativas atuais (ex: está há 5 jogos sem perder). Essas métricas de forma sequencial dão ideia do momento vivido pela equipe. Por fim, incorporam-se as odds médias históricas envolvendo o time, similares às do comparativo: qual odd média o mercado ofereceu para vitória, empate e derrota desse time, tanto no geral quanto separando quando atuou em casa ou fora. Todos esses cálculos são derivados diretamente do conjunto de dados histórico filtrado para aquela equipe, fazendo uso intenso de agregações e estatísticas descritivas via Pandas. Para contextualizar cada número, o sistema ranqueia o time na liga para aquela métrica – por exemplo, ao exibir “média de gols: 1,8 (pos 5º)”, indica que o time tem o 5º melhor ataque do campeonato em média de gols. Essa comparação relativa é crucial para interpretar a significância do número bruto.

Figura 18 – Estatísticas do Arsenal



Fonte: Elaborado pelo autor (2025)

4.9.4 Comparativo entre times

A ferramenta de Comparativo entre Times permite analisar lado a lado duas equipes, evidenciando diferenças estatísticas importantes entre elas. O propósito é fornecer uma ferramenta de análise pré-jogo: ao selecionar, por exemplo, os dois times que irão se enfrentar em uma rodada, o usuário obtém uma comparação direta de seus desempenhos médios, potenciais vantagens e desvantagens de cada lado. Isso apoia decisões de aposta mais embasadas, pois possibilita identificar desequilíbrios ou convergências entre as equipes que podem não estar evidentes apenas pelas cotações de apostas.

Figura 19 – Comparativo entre Arsenal e Man City na temporada

Comparativo: Arsenal x Man City		
Classificação Geral		
Arsenal	Posição na Tabela	Man City
2º		3º
Arsenal	Pontos por Jogo	Man City
1.95		1.87
Arsenal	Aproveitamento Casa	Man City
68.4%		73.7%
Arsenal	Aproveitamento Fora	Man City
61.4%		50.9%
Gols e Finalizações		
Odds Médias		
Séries Atuais		
Outros Indicadores		
Arsenal	Escanteios a favor	Man City
6.61		6.66
Arsenal	Escanteios sofridos	Man City
2 11		2 76

Fonte: Elaborado pelo autor (2025)

Tecnicamente, o comparativo extrai diversas métricas de desempenho de cada time (usando a base de dados histórica filtrada para a liga e temporada escolhidas) e as apresenta de forma comparativa. São computadas estatísticas ofensivas e defensivas, como média de gols marcados e sofridos por jogo, finalizações realizadas e cedidas, aproveitamento (%) de pontos em casa e fora e outros indicadores chaves. Além disso, são exibidas métricas de desempenho recente, como a sequência de resultados (por exemplo, últimos 5 jogos) e séries em andamento – número de vitórias consecutivas, jogos sem perder etc.

Importante notar que o comparativo também incorpora dados de odds médias históricas: para cada time, calcula-se a odd média que o mercado ofereceu para suas vitórias, empates e derrotas, tanto em geral quanto separadas por mandante/visitante. Dessa forma, além dos dados de desempenho esportivo, o usuário vê como o mercado de apostas tende a precificar cada equipe. A implementação utiliza funções de agregação em Pandas para computar as estatísticas por time de forma eficiente, e a interface organiza os resultados em colunas paralelas, facilitando a comparação visual número a número.

4.9.5 Jogos do dia

A funcionalidade Jogos do Dia foca nos eventos agendados para o dia corrente (ou para uma data escolhida), oferecendo uma visão das partidas futuras iminentes junto com informações relevantes para apostas. Diferentemente das demais funcionali-

dades que trabalham sobre dados históricos finalizados, esta opera em tempo quase real, preparando o usuário para os jogos prestes a acontecerem. O propósito é central para uma plataforma de suporte a apostas: informar o apostador sobre os jogos do dia, com estatísticas pré-jogo e odds atualizadas, de forma integrada e filtrada para as principais ligas, sem que o usuário precise recorrer a sites ou outros aplicativos.

Para obter os jogos do dia, o projeto adotou uma estratégia de coleta complementar via API, utilizando um repositório público no GitHub que integra dados da API FootyStats. A aplicação, ao ser acessada, faz a leitura direta de um arquivo CSV atualizado diariamente nesse repositório, contendo a lista de partidas do dia com diversos atributos pré-partida. Esse processo dinâmico permite que, a cada data selecionada, sejam carregadas informações atualizadas sem necessidade de manutenção manual de dados. Após carregar os dados do dia, a plataforma filtra apenas as ligas suportadas.

Para cada jogo listado, são exibidas colunas como: data e horário da partida, times mandante e visitante, odds de aposta pré-jogo (tanto para resultado do primeiro tempo quanto resultado final, incluindo odds de vitória mandante, empate e visitante) e estatísticas preditivas como indicadores de desempenho recente dos times. Por exemplo, são fornecidos PPG (pontos por jogo) médio do mandante e do visitante, e xG (gols esperados) pré-jogo tanto de cada equipe quanto total estimado para a partida. Essas métricas servem como um instantâneo estatístico do confronto antes de sua realização. Os nomes das colunas são traduzidos para português na interface para melhor compreensão do usuário. Os dados carregados desta forma são apresentados em formato tabular interativo, permitindo ordenar e filtrar se necessário no dashboard.

Figura 20 – Jogos por acontecer na data

	Liga	Data	Hora	Rodada	Mandante	Visitante	Odd Mandante (1º Tempo)	Odd Empate (1º Tempo)	Odd Visitante (1º Tempo)	Odd Mandante (Final)	Odd Empate (Final)	Odd Visitante (Final)
12	SPAIN 1	2025-10-25	09:00:00	10	Girona FC	Real Oviedo	2.3	2.2	5.45	1.82	3.3	
30	ITALY 1	2025-10-25	10:00:00	8	Parma	Como	4.2	2.15	2.45	4	3.3	
31	ITALY 1	2025-10-25	10:00:00	8	Udinese	Lecce	2.73	1.92	5.3	1.85	3.2	
32	GERMANY 1	2025-10-25	10:30:00	8	Eintracht Frankfurt	St. Pauli	2.2	2.62	4	1.73	3.9	
33	GERMANY 1	2025-10-25	10:30:00	8	Augsburg	RB Leipzig	3.25	2.66	2.57	3.1	3.7	
34	GERMANY 1	2025-10-25	10:30:00	8	Hamburger SV	Wolfsburg	2.75	2.25	3.2	2.25	3.65	
35	GERMANY 1	2025-10-25	10:30:00	8	Hoffenheim	Heidenheim	2.08	2.52	4.8	1.56	4.33	
36	GERMANY 1	2025-10-25	10:30:00	8	Borussia M'gladbach	Bayern München	8.76	3.51	1.53	11	7	
63	ENGLAND 1	2025-10-25	11:00:00	9	Chelsea	Sunderland	1.91	2.4	6.5	1.45	4.75	
71	ENGLAND 1	2025-10-25	11:00:00	9	Newcastle United	Fulham	2.15	2.2	5	1.57	4	

Fonte: Elaborado pelo autor (2025)

4.10 RESULTADOS ALCANÇADOS

Durante o período de testes da plataforma MoneyBall, foram realizadas diversas simulações com os modelos preditivos desenvolvidos, visando validar sua consistência

e desempenho estatístico. Os resultados obtidos mostraram-se satisfatórios, especialmente considerando que se trata de um sistema baseado em machine learning, cujo conjunto de dados passou a ser estruturado apenas a partir do segundo semestre de 2021.

Esse recorte temporal implica que o volume de dados disponíveis ainda não é suficientemente robusto para consolidar totalmente os modelos. Contudo, mesmo diante dessa limitação, os resultados apresentaram uma tendência de desempenho progressivo ao longo dos anos analisados, evidenciando que o modelo tende a aprimorar sua precisão conforme acumula mais dados históricos e realiza novos ciclos de treinamento.

4.10.1 Resultados Lay Mandante

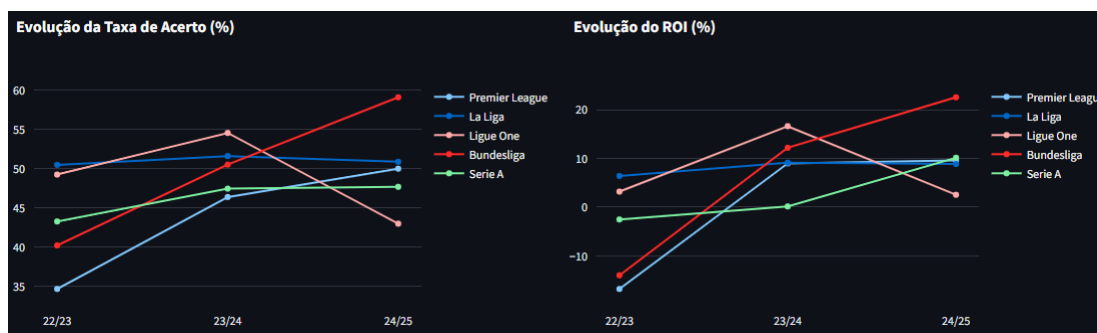
Neste tópico será apresentado os resultados do modelo Lay Mandante. Toda a simulação foi realizada após a rodada 7 de cada temporada de liga e com uma banca de R\$ 1.000,00 utilizando uma stake fixa de R\$ 50,00.

4.10.1.1 Resultados por Ligas

De forma geral, os resultados do modelo Lay Mandante demonstram um desempenho estável e em crescimento ao longo das temporadas analisadas. Observou-se um aumento progressivo tanto na taxa de acerto quanto no ROI, indicando que o modelo se tornou mais assertivo e eficiente à medida que acumulou mais dados históricos e passou por novos ciclos de calibração.

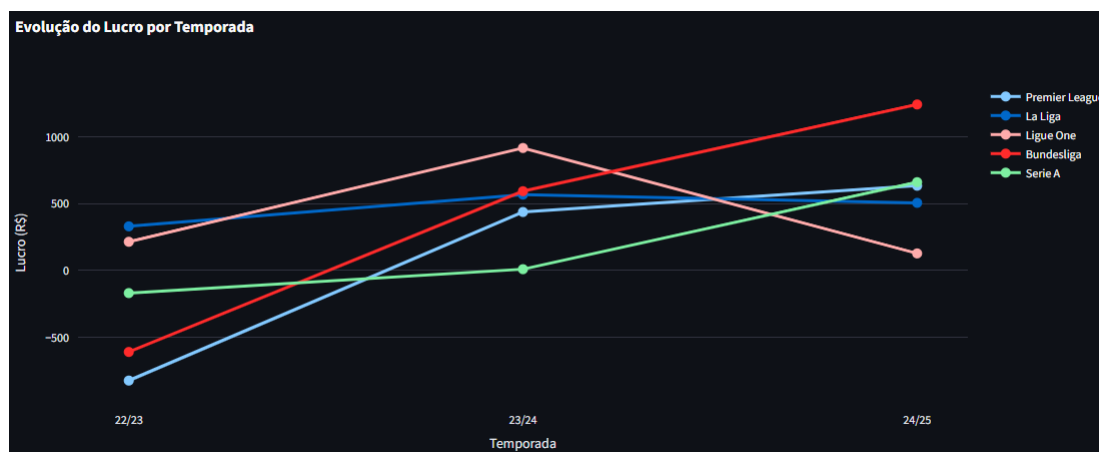
Entre as ligas avaliadas, La Liga, Ligue One e Bundesliga apresentaram os melhores retornos financeiros, com lucros e ROIs positivos e consistentes. Já a Serie A e a Premier League tiveram resultados mais moderados, porém ainda dentro de uma faixa considerada satisfatória. Essa variação de desempenho entre ligas é natural, refletindo diferenças estruturais como equilíbrio competitivo, comportamento de odds e frequência de vitórias dos mandantes.

Figura 21 – Evolução temporal da taxa de acerto e ROI entre ligas.



Fonte: Elaborado pelo autor (2025)

Figura 22 – Evolução temporal do lucro entre ligas.



Fonte: Elaborado pelo autor (2025)

Figura 23 – Ranking geral entre ligas

	Ranking	Liga	Apostas	Taxa de Acerto	Lucro	ROI
1	1	La Liga	341	51.03%	R\$ 1401.00	8.13%
2	2	Ligue One	344	49.13%	R\$ 1256.00	7.46%
0	3	Bundesliga	294	50.68%	R\$ 1223.00	6.93%
4	4	Serie A	382	46.07%	R\$ 499.00	2.59%
3	5	Premier League	327	44.34%	R\$ 246.00	0.59%

Fonte: Elaborado pelo autor (2025)

4.10.1.2 Resultados gerais

De forma agregada, o modelo realizou 1 688 apostas, alcançando uma taxa média de acerto de 48,16 %, um lucro total de R\$ 4 625,00 e um ROI médio de 5,14 %. Esses indicadores demonstram que a estratégia se mostrou lucrativa mesmo após um período inicial de adaptação, na temporada 2022/2023 o sistema apresentou resultado negativo (−R\$ 1 059,00 e ROI de −4,75 %), mas já na temporada seguinte o modelo reagiu, obtendo R\$ 2 519,00 de lucro (ROI de 9,43 %) e mantendo o crescimento na temporada 2024/2025, quando o lucro chegou a R\$ 3 165,00 (ROI de 10,74 %).

Esse progresso evidencia o processo de aprendizado acumulativo do modelo, típico de sistemas de machine learning, que se tornam mais precisos à medida que incorporam novos dados e passam por ciclos sucessivos de calibração e validação temporal (walk-forward). A taxa de acerto, inicialmente em 44 %, estabilizou-se em torno de 50 %, indicando uma melhora na identificação de jogos em que o mandante apresentava odds superavaliadas.

Figura 24 – Principais métricas por temporada.

Evolução das Métricas Principais		
Total de Apostas por Ano	Taxa de Acerto Geral (%)	Lucro Total (R\$)
22/23	22/23	22/23
556	44.06%	R\$ -1059.00
23/24	23/24	23/24
546	50.18%	R\$ 2519.00
↓ -10	↑ 6.12%	↑ R\$ 3578.00
24/25	24/25	24/25
586	50.17%	R\$ 3165.00
↑ 40	↓ -0.01%	↑ R\$ 646.00

Fonte: Elaborado pelo autor (2025)

Figura 25 – Resumo por temporadas.

Tabela Resumo por Temporada										
	Temporada	Total Apostas	Total Acertos	Taxa (%)	Lucro (R\$)	ROI (%)	Banca Final (R\$)	Odd 0-45% (%)	Odd 45-60% (%)	Odd 60-100% (%)
0	22/23	556	245	44.06	-1059.00	-4.75	788.20	24.04	42.11	48.97
1	23/24	546	274	50.18	2519.00	9.43	1503.80	29.23	39.88	58.85
2	24/25	586	294	50.17	3165.00	10.74	1633.00	38.68	42.08	60.06

Fonte: Elaborado pelo autor (2025)

Figura 26 – Visão geral de resultados.

Visão Geral dos Resultados - Lay Mandante			
Total de Apostas	Taxa de Acerto	Lucro Total	ROI Médio
1688	48.16%	R\$ 4625.00	5.14%
		↑ 4625.00	

Fonte: Elaborado pelo autor (2025)

4.10.1.3 Conclusões sobre o modelo Lay Mandante

Em síntese, os resultados gerais confirmam que a estratégia Lay Mandante evolui de forma sustentável, apresentando retorno financeiro positivo, robustez estatística crescente e tendência de consolidação à medida que a base histórica é ampliada, reforçando seu potencial como componente estável dentro do portfólio preditivo do projeto MoneyBall.

4.10.2 Resultados Back Visitante

Neste tópico será apresentado os resultados do modelo Back Visitante. Toda a simulação foi realizada após a rodada 7 de cada temporada de liga e com uma banca de R\$ 1.000,00 utilizando uma stake fixa de R\$ 50,00.

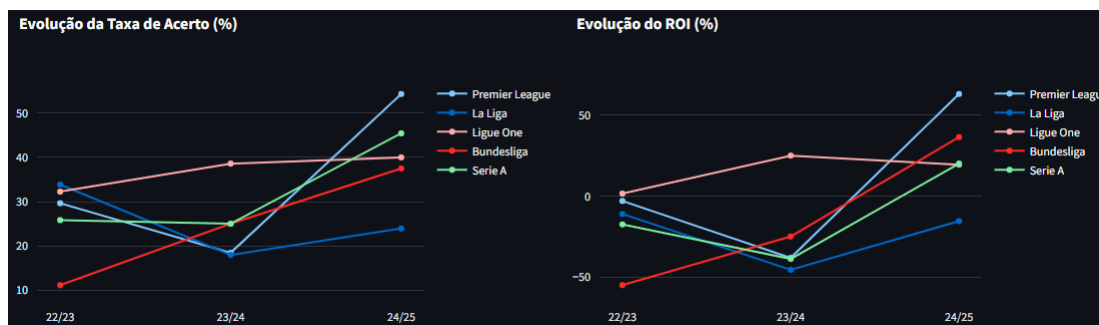
4.10.2.1 Resultados por Ligas

De modo geral, o desempenho variou significativamente entre as ligas, refletindo as diferenças estruturais no equilíbrio competitivo e nas probabilidades de vitória dos visitantes. A Ligue One destacou-se como a liga mais lucrativa, alcançando um ROI de 15,24 % e um lucro acumulado de R\$ 1.094,00, demonstrando boa sensibilidade do modelo para identificar jogos com valor esperado positivo nesse mercado. Em seguida, a Premier League também apresentou desempenho consistente, com ROI de 7,24 % e lucro de R\$ 640,00, indicando boa estabilidade preditiva em um campeonato de alta competitividade.

Por outro lado, as ligas Bundesliga, Serie A e La Liga registraram resultados negativos, com ROIs de -14,56 %, -12,11 % e -24,07 %, respectivamente. Esses resultados refletem uma maior dificuldade do modelo em ajustar as probabilidades em contextos onde o visitante tende a ter desempenho historicamente inferior ou onde as odds de vitória do visitante são superavaliadas pelo mercado.

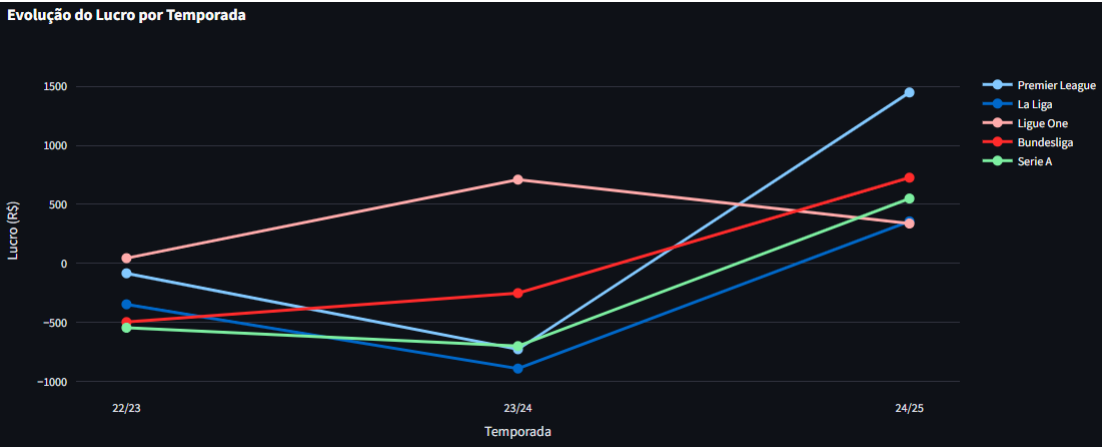
A análise temporal mostra que, embora o início da simulação (temporada 2022/2023) tenha sido marcado por resultados abaixo do esperado, houve uma recuperação substancial nas temporadas seguintes. O aumento nas taxas de acerto e no ROI em 2024/2025 confirma que o modelo está se tornando mais eficiente com o avanço do aprendizado e a incorporação de novos dados.

Figura 27 – Evolução temporal da taxa de acerto e ROI entre ligas.



Fonte: Elaborado pelo autor (2025)

Figura 28 – Evolução temporal do lucro entre ligas.



Fonte: Elaborado pelo autor (2025)

Figura 29 – Ranking geral entre ligas

Ranking das Ligas							
	Ranking	Liga	Apostas	Taxa de Acerto	Lucro	ROI	
2	1	Ligue One	154	36.36%	R\$ 1094.00	15.24%	
3	2	Premier League	138	34.78%	R\$ 640.00	7.24%	
0	3	Bundesliga	78	28.21%	R\$ -18.50	-14.56%	
4	4	Serie A	153	32.68%	R\$ -691.50	-12.11%	
1	5	La Liga	147	26.53%	R\$ -876.50	-24.07%	

Fonte: Elaborado pelo autor (2025)

4.10.2.2 Resultados gerais

Com base nos resultados gerais do modelo Back Visitante, é possível observar que, embora o desempenho inicial tenha sido negativo, o modelo apresentou evolução significativa ao longo das temporadas, indicando um processo de aprendizado e amadurecimento do sistema preditivo.

No total, foram 670 apostas simuladas, com uma taxa média de acerto de 32,09 %, lucro total de R\$ 147,50 e um ROI médio de -5,65 %. Apesar do ROI ainda levemente negativo, o modelo mostrou uma clara curva de recuperação. Nas duas primeiras temporadas (2022/2023 e 2023/2024), os resultados foram deficitários, com prejuízos de R\$ 1.420,00 e R\$ 1.852,00, respectivamente — reflexo de um período de calibração inicial, em que o volume de dados e a estabilidade probabilística ainda eram limitados.

A partir da temporada 2024/2025, o modelo apresentou um salto expressivo de performance, atingindo 40,54 % de acerto, R\$ 3.419,50 de lucro e ROI de 24,62 %, consolidando um comportamento positivo e validando a hipótese de que o acúmulo de dados históricos e o aprimoramento contínuo do processo de calibração aumentam a eficiência do sistema.

Figura 30 – Principais métricas por temporada.

Evolução das Métricas Principais		
Total de Apostas por Ano	Taxa de Acerto Geral (%)	Lucro Total (R\$)
22/23	22/23	22/23
258	29.07%	R\$ -1420.00
23/24	23/24	23/24
190	26.32%	R\$ -1852.00
↓ -68	↓ -2.75%	↑ R\$ -432.00
24/25	24/25	24/25
222	40.54%	R\$ 3419.50
↑ 32	↑ 14.22%	↑ R\$ 5271.50

Fonte: Elaborado pelo autor (2025)

Figura 31 – Resumo por temporadas.

Tabela Resumo por Temporada										
	Temporada	Total Apostas	Total Acertos	Taxa (%)	Lucro (R\$)	ROI (%)	Banca Final (R\$)	Odd 0-45% (%)	Odd 45-60% (%)	Odd 60-100% (%)
0	22/23	258	75	29.07	-1420.00	-17.05	716.00	19.85	33.83	27.78
1	23/24	190	50	26.32	-1852.00	-24.52	629.60	18.83	36.28	50.67
2	24/25	222	90	40.54	3419.50	24.62	1541.10	33.79	48.54	42.17

Fonte: Elaborado pelo autor (2025)

Figura 32 – Visão geral de resultados.

Visão Geral dos Resultados - Back Visitante			
Total de Apostas	Taxa de Acerto	Lucro Total	ROI Médio
670	32.09%	R\$ 147.50	-5.65%
		↑ 147.50	

Fonte: Elaborado pelo autor (2025)

4.10.2.3 Conclusões sobre o modelo Back Visitante

De forma geral, os resultados demonstram que a estratégia Back Visitante, apesar de mais volátil e sensível a variações nas odds, possui potencial de lucratividade no longo prazo. A trajetória de melhora entre as temporadas reforça a importância do aprendizado incremental e do refinamento contínuo do modelo, elementos fundamentais para a consolidação do MoneyBall como ferramenta de apoio em decisões probabilísticas e estratégias de apostas esportivas baseadas em valor esperado.

5 CONCLUSÃO

O desenvolvimento da plataforma MoneyBall demonstrou a viabilidade técnica e analítica de aplicar métodos de machine learning e estatística preditiva ao mercado de apostas esportivas de futebol. A integração entre análise de dados históricos, critérios matemáticos de valor esperado e validação temporal realista permitiu construir um sistema coerente, interpretável e operacionalmente aplicável.

Os resultados obtidos evidenciam que o modelo é capaz de reconhecer padrões de ineficiência nas odds de mercado e gerar decisões de aposta fundamentadas em vantagem matemática. O método Lay Mandante apresentou estabilidade e lucratividade sustentada, enquanto o Back Visitante revelou maior volatilidade, mas com tendência de crescimento consistente após os ciclos de calibração e expansão de dados.

A evolução do ROI e do lucro ao longo das temporadas confirma o aprendizado progressivo do sistema e reforça o potencial da plataforma como ferramenta de apoio a decisões probabilísticas. Ainda que o conjunto de dados utilizado não seja suficientemente extenso para conclusões definitivas, os resultados alcançados indicam uma direção promissora para pesquisas futuras, especialmente no aprimoramento de calibração, diversificação de ligas e integração com modelos probabilísticos mais avançados.

Assim, o MoneyBall cumpre sua proposta de unir ciência de dados e racionalidade estatística ao universo das apostas esportivas, demonstrando que a aplicação de inteligência artificial pode transformar o ato de apostar em uma decisão técnica, mensurável e orientada por evidências.

REFERÊNCIAS

ASHBOURNE, N. **What Does Bet Back Mean: A Beginner's Guide to Back and Lay Betting**. [S.l.], 2016. Disponível em: <<https://rg.org/guides/sportsbetting-guides/back-betting>>. Acesso em: 29 ago. 2025.

ASHISH. **Streamlit • A faster way to build and share data apps**. [S.l.], 2024. Disponível em: <<https://www.analyticsvidhya.com/blog/2021/07/streamlit-quickly-turn-your-ml-models-into-web-apps/>>. Acesso em: 13 jun. 2025.

BETBETTER. Value betting: Identifying bets with the best risk-reward ratios. **BetBetter**, 2024. Disponível em: <<https://www.betbetter.world/BettingAcademy/value-betting.aspx>>. Acesso em: 18 set. 2025.

BETFAIR. **EPL Machine Learning Walkthrough**. [S.l.], 2022. Disponível em: <<https://betfair-datascientists.github.io/modelling/EPLmlPython/>>. Acesso em: 01 mar. 2025.

BISPO Ícaro. **Moneyball no futebol O Brighton como case de sucesso**. [S.l.], 2024. Disponível em: <<https://www.meer.com/pt/77507-moneyball-no-futebol>>. Acesso em: 11 dez. 2025.

BRASIL, R. V. **Mudanças no mercado de apostas esportivas no Brasil: regulamentação, impactos e crescimento**. [S.l.], 2025. Disponível em: <<https://www.vavel.com/br/futebol/2025/03/14/1216914-mudancas-no-mercado-de-apostas-esportivas-no-brasil-regulamentacao-impactos-e-cre>>. Acesso em: 04 out. 2025.

BROWNLEE, J. **How To Backtest Machine Learning Models for Time Series Forecasting**. [S.l.], 2019. Disponível em: <<https://machinelearningmastery.com/backtest-machine-learning-models-time-series-forecasting/>>. Acesso em: 01 maio 2025.

CAPITAL, F. **Feature Scaling is NOT Always Necessary**. [S.l.], 2022. Disponível em: <<https://fastercapital.com/pt/tema/introdu%C3%A7%C3%A3o-%C3%A0-normaliza%C3%A7%C3%A3o-dos-dados.html>>. Acesso em: 08 ago. 2025.

CHAWLA, A. **Feature Scaling is NOT Always Necessary**. [S.l.], 2023. Disponível em: <<https://blog.dailydoseofds.com/p/feature-scaling-is-not-always-necessary>>. Acesso em: 24 mar. 2025.

ESPORTE, R. máquina do. **Por Dentro da Indústria de Apostas Esportivas do Brasil: O Terceiro Maior Mercado do Mundo**. [S.l.], 2024. Disponível em: <<https://maquinadoesporte.com.br/apostas/por-dentro-da-industria-de-apostas-esportivas-do-brasil-o-terceiro-maior-mercado-do-mundo/>>. Acesso em: 29 maio 2025.

FAHAD, I. A. **Understanding Walk Forward Validation in Time Series Analysis: A Practical Guide**. [S.l.], 2024. Disponível em: <<https://medium.com/%40ahmedfahad04/>>

understanding-walk-forward-validation-in-time-series-analysis-a-practical-guide-ea3814015abf>. Acesso em: 23 jun. 2025.

FLANAGAN, C. **Brentford back in the Premier League: how the Bees' Moneyball philosophy overcame play-off heartbreak to make it back to the promised land.** [S.l.], 2021. Disponível em: <<https://www.fourfourtwo.com/features/brentford-back-in-the-premier-league-how-the-bees-moneyball-philosophy-overcame-play-off-hea>>. Acesso em: 26 set. 2025.

GALLAS, D. Por que você quase sempre vai perder dinheiro com bets, segundo a matemática. **BBC NEWS Brasil**, 2024. Disponível em: <<https://www.bbc.com/portuguese/articles/c981g2n1dm9o>>. Acesso em: 07 set. 2025.

GUSAROVA, M. **Understanding Random Forest.** [S.l.], 2022. Disponível em: <<https://medium.com/%40data.science.enthusiast/understanding-random-forest-hyperparameters-tunning-pathon-code-136c3b1ce9e>>. Acesso em: 31 ago. 2025.

HARINGEY, R. **Leicester City are using "Moneyball" to win the Premier League.** [S.l.], 2016. Disponível em: <<https://playingfor90.com/2016/02/06/leicester-city-is-using-moneyball-to-win-the-premier-league/>>. Acesso em: 23 out. 2025.

KNIGHT, J. **What Is Overround in Betting & How Does it Work Online?** [S.l.], 2024. Disponível em: <<https://www.oddsportal.com/reviews/betting/guide/overround/>>. Acesso em: 09 fev. 2025.

LAGO-PEÑAS, C. **Game Location and Team Quality Effects on Performance Profiles in Professional Soccer.** [S.l.], 2011. Disponível em: <<https://pmc.ncbi.nlm.nih.gov/articles/PMC3737821/>>. Acesso em: 29 jul. 2025.

LEWIS, M. **Moneyball: The Art of Winning an Unfair Game.** [S.l.]: W. W. Norton & Company, 2004.

MACHADO, R. **Moneyball- uma nova abordagem no desporto.** [S.l.], 2024. Disponível em: <<https://www.fink.pt/moneyball-uma-nova-abordagem-no-desporto/>>. Acesso em: 23 jun. 2025.

MILLER, B. **Moneyball: Playing the Game.** [S.l.], 2011. Disponível em: <<https://www.imdb.com/pt/title/tt1210166/>>. Acesso em: 14 set. 2025.

PARIMI, S. **Ultimate Guide: Streamlit vs. Flask vs. Dash vs. Voila — A Comprehensive Comparison.** [S.l.], 2024. Disponível em: <<https://sureshparimi.medium.com/ultimate-guide-streamlit-vs-flask-vs-dash-vs-voila-a-comprehensive-comparison-5f6150ca3714>>. Acesso em: 23 jan. 2025.

PERARNAU, M. **Guardiola Confidencial: Um Ano Dentro Do Bayern De Munique Acompanhando De Perto O Técnico Que Mudou O Futebol Para Sempre.** [S.l.]: Editora Grande Área, 2015.

SAXENA, S. **A Beginner's Guide to Random Forest Hyperparameter Tuning.** [S.l.], 2024. Disponível em: <<https://www.analyticsvidhya.com/blog/2020/03/beginners-guide-random-forest-hyperparameter-tuning/>>. Acesso em: 17 jan. 2025.

SMARKETS. **What is lay betting on an exchange?** [S.l.], 2024. Disponível em: <<https://help.smarkets.com/hc/en-gb/articles/115003374849-What-is-lay-betting-on-an-exchange>>. Acesso em: 23 set. 2025.

SMITH, S. **Calculating Expected Value in Sports Betting: Expected Return Explained.** [S.l.], 2025. Disponível em: <<https://www.oddsshopper.com/articles/betting-101/calculating-expected-value-in-sports-betting-explained-y10>>. Acesso em: 26 abr. 2025.

SOHAIL, S. **The Math Behind Betting Odds and Gambling.** [S.l.], 2024. Disponível em: <<https://www.investopedia.com/articles/dictionary/042215/understand-math-behind-betting-odds-gambling.asp>>. Acesso em: 13 maio 2025.

STRATEGYAND. **O impacto das apostas esportivas no consumo.** [S.l.], 2024. Disponível em: <<https://www.strategyand.pwc.com/br/pt/relatorios/o-impacto-das-apostas-esportivas-no-consumo.html>>. Acesso em: 21 ago. 2025.

SUMPTER, D. **Soccermatics: Mathematical Adventures in the Beautiful Game.** [S.l.]: Bloomsbury Sigma, 2016.

THOMAS, H. What is a back bet? (+ how it works & benefits). **Pinnacle Odds Dropper**, 2025. Disponível em: <<https://www.pinnacleoddsdropper.com/blog/back-bet>>. Acesso em: 25 nov. 2025.

WIKIPEDIA. Gambling mathematics. **Wikipedia**, 2024. Disponível em: <https://en.wikipedia.org/wiki/Gambling_mathematics>. Acesso em: 27 jul. 2025.