



Estudo sobre a Evolução da Eficiência das Odds no Mercado de Apostas de Futebol

Trabalho de Conclusão de Curso desenvolvido na Universidade Federal da Paraíba, Centro de Ciências Sociais Aplicadas, Coordenação de Ciência de Dados para Negócios.

Orientador: Jorge Henrique Norões Viana

Discente: Matheus Santos de Oliveira Flôr



Como a Casa Sempre Vence?

O "Juice" ou Margem

Taxa embutida nas odds que garante lucro independente do resultado. Em vez de odds 2.00, a casa oferece 1.90 para ambos os lados.

Volume Equilibrado

A casa ajusta as odds para receber apostas balanceadas em todos os resultados, garantindo lucro matemático. **Levitt (2004)**

Ineficiências de Mercado

Odds nem sempre coincidem com probabilidades reais dos eventos, criando oportunidades de análise. (Rodney D. Fort)



Regulamentação e Crescimento

Marco Legal

Lei 14.790/2023 formalizou o setor, trazendo segurança jurídica para operadores e confiança para usuários.

Regulamentação plena desde final de 2024 foi o principal catalisador do crescimento exponencial.

Impacto no Mercado

Todos os 20 clubes da Série A são patrocinados por casas de apostas em 2025.

Publicidade focada em futebol domina transmissões e mídias digitais.





O Boom das Apostas no Brasil

4B

Visitas no 1º Semestre

Tráfego para sites de apostas no Brasil
em 2025 (**Similarweb**)

R\$100B

Volume Movimentado

Apostas esportivas movimentaram entre
R\$ 60-100 bilhões em 2023 (**Strategy da
PWC**)

85%

Domínio do Futebol

Futebol representa 75-85% de todo
volume de apostas no país (**H2 Gambling
Capital e Kantar IBOPE Media**)

Hipótese Central e Relevância Teórica

1

Hipótese Central

A eficiência dos mercados de apostas evoluiu significativamente, tornando-se mais robusta devido a:

- Avanços tecnológicos
- Algoritmos de precificação sofisticados
- Explosão do Big Data

2

Relevância Teórica

- **Teoria Econômica:** Desafia a visão tradicional de que apostas são puramente jogos de azar, explorando como a eficiência do mercado se manifesta em contextos não financeiros.
- **Finanças Comportamentais:** Fornece um campo rico para analisar vieses cognitivos e irracionalidades de decisão em ambientes de alta incerteza e recompensa.

3

Demistificando as Apostas

Este estudo demistifica as apostas como um sistema baseado em informação e análise, contrastando-o com a percepção popular de que é meramente uma questão de sorte. A habilidade e a capacidade de processar informações desempenham um papel crucial no sucesso a longo prazo.



WEB SCRAPPING
EXTRAÇÃO DE DADOS DA WEB

Metodologia de Pesquisa

01

Coleta de Identificadores

Script Python com Selenium extraiu IDs únicos de todas as partidas do Flashscore (2018-2024).

02

Extração de Estatísticas

BeautifulSoup coletou dados detalhados: placares, posse de bola, xG, finalizações, escanteios.

03

Dados Climáticos

API Open-Meteo forneceu temperatura, umidade e velocidade do vento para cada partida.

Modelos de Machine Learning

1

Regressão Logística

Modelo baseline linear para transparência e interpretabilidade das previsões. (**Hosmer; Lemeshow; Sturdivant, 2013**)

2

Random Forest

Ensemble de árvores para capturar interações não lineares, resistente ao overfitting. (**Breiman, 2001**)

3

XGBoost

Algoritmo de gradient boosting, altamente eficaz para dados tabulares e competições. (**Chen; Guestrin, 2016**)



Validação Temporal dos Dados

1 Problema do Data Leakage

Informações futuras no treino geram avaliações irrealisticamente otimistas. (Kaufman et al., 2012)

2 Validação Walk-Forward

Modelo treinado em dados passados, validado no período subsequente. (Bergmeir; Hyndman, 2018)

3 Simulação Real

Processo iterativo simula fielmente operação em cenários reais de apostas.



**DATA LEAKAGE -
ERRO GRAVE**

Tabelas de algumas variáveis obtidas

campeonato_rodada		data_hora		mandante	visitante	finalizacoes_mandante	finalizacoes_visitante	finalizacoes_alvo_mandante
Brasileirão	Betano - Rodada 24	26.08.2024	21:00	Vasco	Athletico-PR			
Premier League	- Rodada 14	04.12.2024	17:15	Aston Villa	Brentford	6	8	4
Premier League	- Rodada 20	04.01.2025	12:00	Aston Villa	Leicester	20	9	10
Premier League	- Rodada 24	01.02.2025	12:00	Everton	Leicester	13	4	5
Premier League	- Rodada 24	01.02.2025	12:00	Everton	Leicester	13	9	7
Premier League	- Rodada 2	25.08.2024	12:30	Liverpool	Brentford	19	8	8
Premier League	- Rodada 2	25.08.2024	12:30	Liverpool	Brentford	11	15	4
Premier League	- Rodada 36	10.05.2025	11:00	Ipswich	Brentford	29	9	8
Brasileirão	Betano - Rodada 26	14.09.2024	18:30	Athletico-PR	Fortaleza	11	12	4
Premier League	- Rodada 10	03.11.2024	13:30	Manchester Utd	Chelsea	6	26	4
Premier League	- Rodada 14	04.12.2024	16:30	Southampton	Chelsea	25	4	6
Premier League	- Rodada 25	15.02.2025	12:00	Aston Villa	Ipswich			

gols_mandante	gols_visitante	posse_mandante	posse_visitante
2	1	48%	52%
3	1	49%	51%
2	1	61%	39%
4	0	47%	53%
2	0	62%	38%
0	1	50%	50%
1	1	51%	49%
1	1	46%	54%
1	5	45%	55%
1	1	75%	25%

Estádio		Cidade	Temperatura (°C)	Umidade (%)	Vento (km/h)
	Maracanã	Rio de Janeiro	22.3	93	1.8
	Neo Química Arena	Sao Paulo	18.4	93	2.9
	Morumbi	Sao Paulo	18.4	93	2.9
	Mineirão	Belo Horizonte	20.5	74	3.1
	Beira-Rio	Porto Alegre	21.1	87	15.1
	Arena Castelão	Fortaleza	26.6	89	7.2
	Pituaçu	Salvador	27.3	83	4.3
	Ligga Arena	Curitiba	16.1	97	7.9
	Arena Condá	Chapeco	16.3	85	5.2
	Arena Pantanal	Cuiaba	25.9	85	2.5

Objetivos da Pesquisa

Objetivo Principal

Analisar e comparar a eficiência informacional dos mercados de apostas entre Premier League e Brasileirão Série A ao longo do tempo.

Variáveis Externas

Identificar fatores climáticos e contextuais que influenciam a precificação das odds.

Retorno Esperado

Medir retornos de apostas ao longo de temporadas completas em diferentes tipos de mercado.

Impacto da Popularidade

Investigar se há sobrevalorização de times populares nas odds das casas de apostas.

Apostas vs. Loteria: Habilidade vs. Sorte

Loteria: Aleatoriedade Pura (Nassim Nicholas Taleb)

- Probabilidades fixas e independentes
- Histórico não oferece poder preditivo
- Habilidade é fator nulo
- Sistema sem memória

Apostas Esportivas: Mercado de Informação

- Probabilidades condicionais e dinâmicas
- Variáveis analisáveis influenciam resultados
- Desempenho histórico, lesões, táticas
- Fatores climáticos e psicológicos

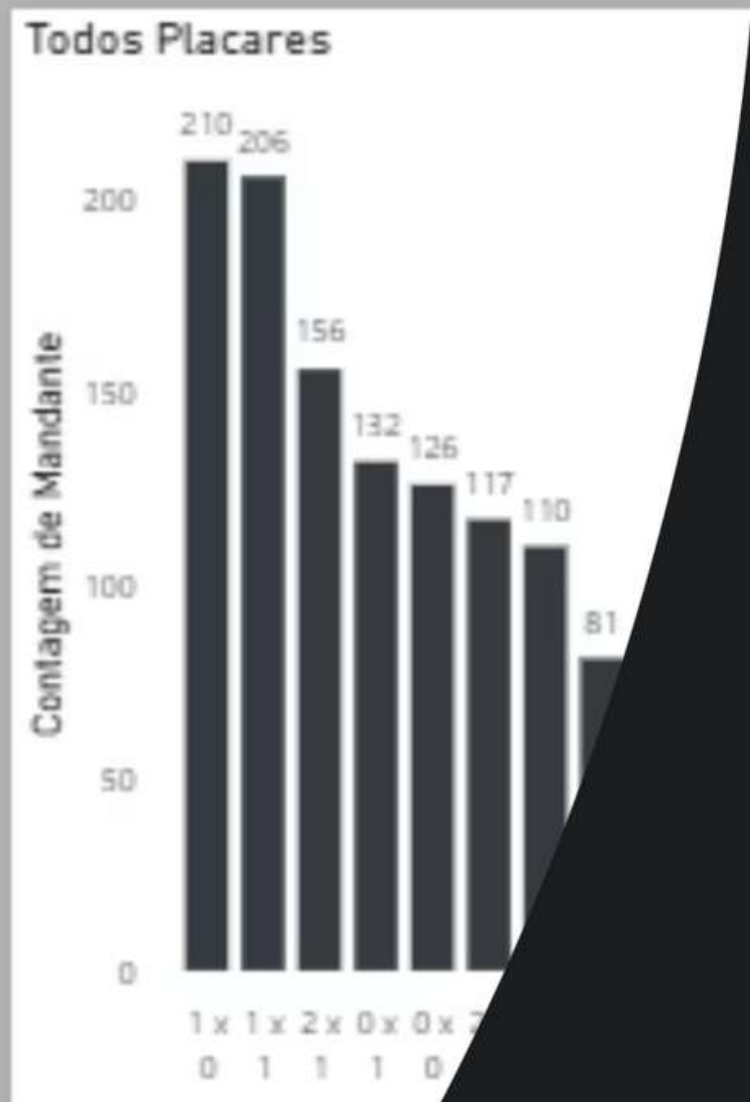
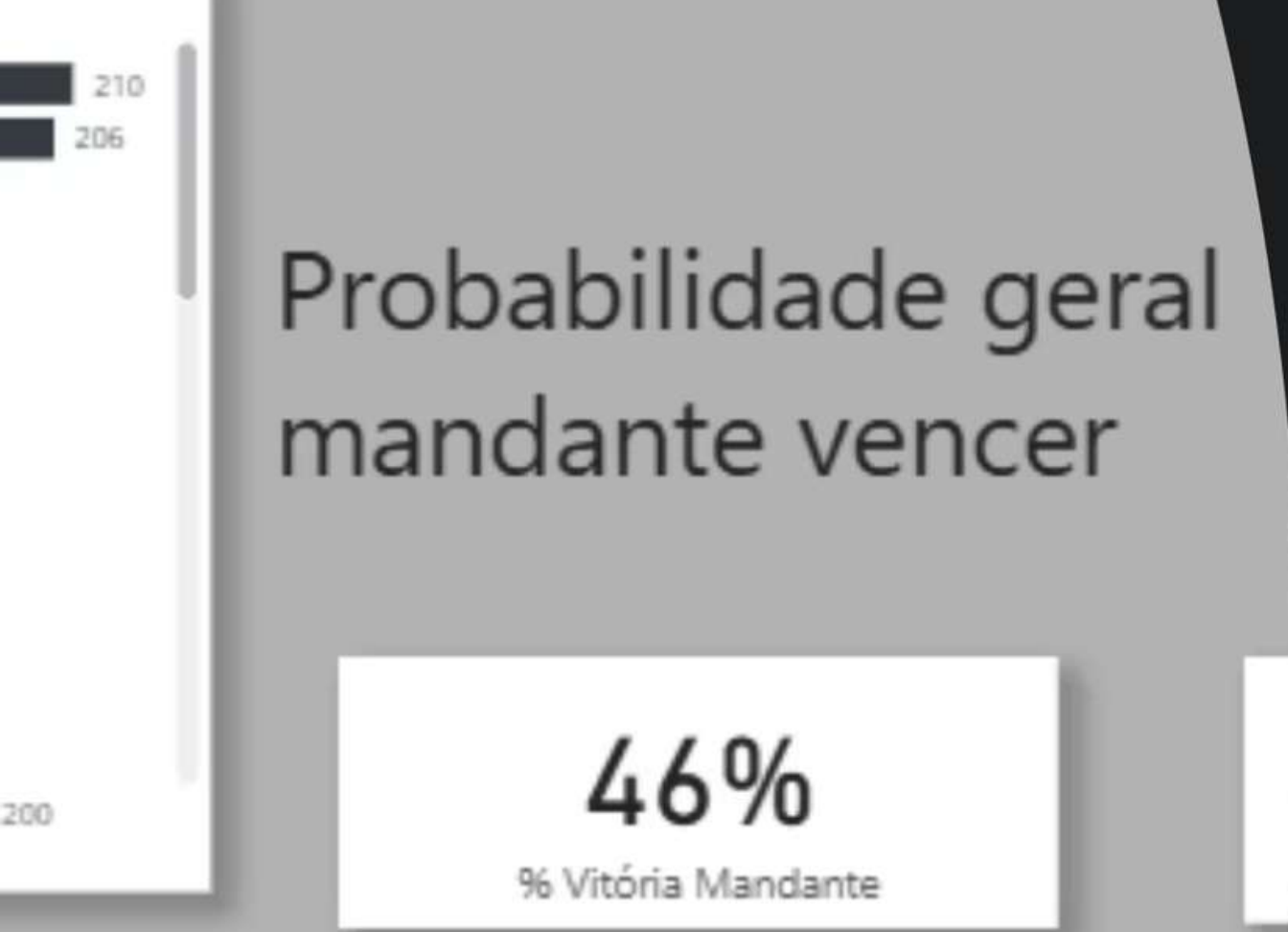
LOTERIA



APOSTA E



LOTERIA vs. APOSTA
SORTE vs. ESTRATÉGIA



Plano de Comunicação dos Resultados



Painéis Interativos

Dashboards temáticos com filtros dinâmicos por liga, temporada, equipe e rodada para análise personalizada.



Análise Comparativa

Visualizações lado a lado comparando probabilidades do modelo versus odds de mercado.



Métricas de Performance

Acurácia, matriz de confusão e importância das variáveis para cada modelo de ML.

Links

Países que mais acessaram

Countries That Most Access
Sports Betting Sites in the World

Vídeo mostra a curva de
crescimento dos países que mais
acessaram sites de apostas
esportivas

<https://youtu.be/JndihgoPmcs?si=AQqCdwO1kk1Q4L2Y>

[power bi](#)

**Power BI (visualização dos
dados)**

Power BI Report

