

Previsões de Resultados em Partidas do Campeonato Brasileiro de Futebol

João Marcos Amorim dos Santos
Moacyr Alvim Horta Barbosa da Silva
Rodrigo dos Santos Targino

FGV | EMap - Escola de Matemática Aplicada

27 de junho de 2019

Conteúdo

- 1 Introdução
- 2 Objetivos
- 3 Justificativa
- 4 Revisão de Literatura & Fundamentação Teórica
- 5 Metodologia
- 6 Resultados
- 7 Conclusão

- 1 Introdução
- 2 Objetivos
- 3 Justificativa
- 4 Revisão de Literatura & Fundamentação Teórica
- 5 Metodologia
- 6 Resultados
- 7 Conclusão

- O Futebol
 - É considerado o esporte mais popular do mundo;
- Campeonato Brasileiro de Futebol
 - Principal campeonato de futebol do Brasil;
- Cartola F.C.

Introdução - Cartola FC

O game

- O Cartola F.C. é um *fantasy game* onde a cada rodada os participantes escalam seus times com os jogadores reais do Brasileirão;
- Durante cada rodada do campeonato os jogadores pontuam de acordo com suas estatísticas na partida, cada tipo de estatística tem uma determinada pontuação.

Scouts	
Ataque	Defesa
Gol	Jogo sem sofrer gol
Assistência	Defesa de pênalti
Finalização na trave	Defesa difícil
Finalização defendida	Roubada de bola
Finalização para fora	Gol contra
Falta Sofrida	Cartão vermelho
Pênalti perdido	Cartão amarelo
Impedimento	Gol sofrido
Passé errado	Falta cometida

Figura: Scouts dos Cartola FC

Conteúdo

- 1 Introdução
- 2 **Objetivos**
- 3 Justificativa
- 4 Revisão de Literatura & Fundamentação Teórica
- 5 Metodologia
- 6 Resultados
- 7 Conclusão

Objetivo Geral

Gerar previsões de resultados e placares do campeonato brasileiro.

Objetivo Específico

Comparar a capacidade preditiva dos diferentes modelos de previsão de resultados do campeonato brasileiro.

Conteúdo

- 1 Introdução
- 2 Objetivos
- 3 Justificativa**
- 4 Revisão de Literatura & Fundamentação Teórica
- 5 Metodologia
- 6 Resultados
- 7 Conclusão

- Prever resultados do Campeonato Brasileiro
 - Mídia esportiva
 - Torcedores
 - Apostadores
 - Pesquisadores

- Perguntas que podem ser respondidas
 - Qual a probabilidade do Botafogo ir para Libertadores?
 - Qual a probabilidade do Fluminense ser rebaixado?
 - Qual a probabilidade do Palmeiras terminar na frente do Corinthians?
 - Qual a probabilidade do Flamengo não se classificar para competições internacionais?

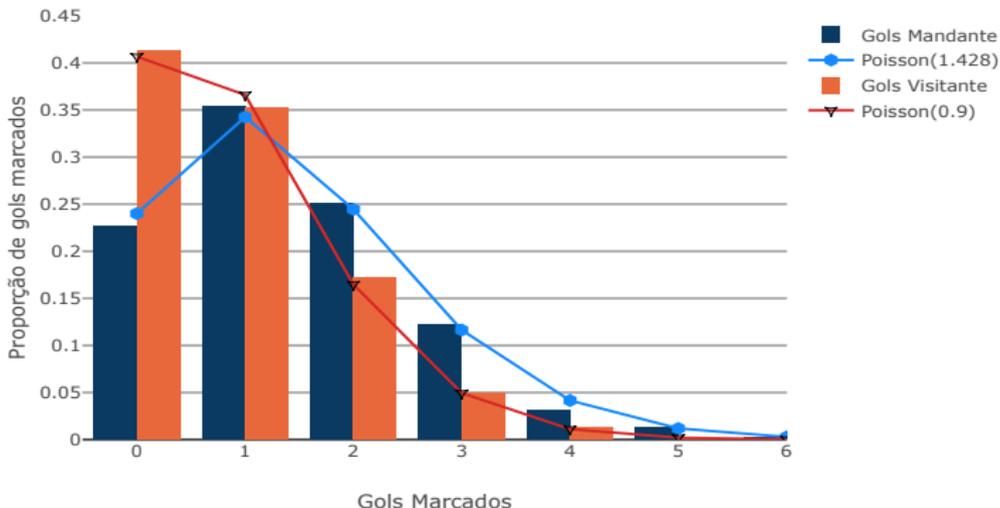
Conteúdo

- 1 Introdução
- 2 Objetivos
- 3 Justificativa
- 4 Revisão de Literatura & Fundamentação Teórica**
- 5 Metodologia
- 6 Resultados
- 7 Conclusão

- Modelagem por gols da partida
 - Poisson Independentes;
 - Poisson Modificada
 - Poisson Bivariada de Holgate;
 - Binomial - Poisson.
- Modelagem por resultado (vitória, empate, derrota)
 - Regressão logística multinomial;
 - Distribuição de Dirichlet.

Os gols serem modelados pela distribuição de Poisson ?

Distribuição de gols marcados no Brasileirão 2014 - 2018



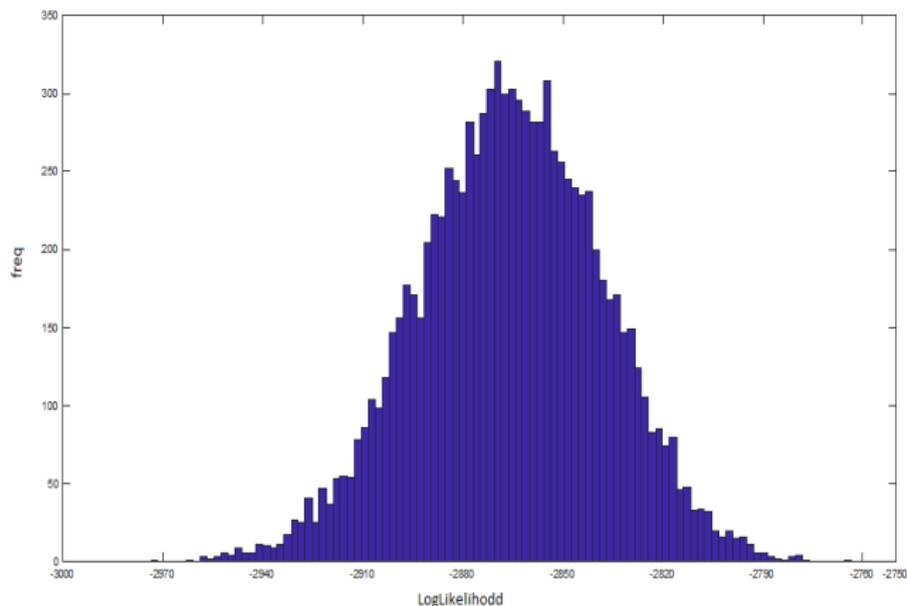
Os gols serem modelados pela distribuição de Poisson ?

Tabela: Distribuição dos gols e Poisson

Gols	0	1	2	3	4	5	6
Freq_obs	430	673	476	232	60	24	5
Freq_esp	456	651	464	221	79	22	5

Revisão de Literatura / Fundamentação Teórica

Poisson Independente



Estatística de teste = 7.71; P-valor = 0.17

Poisson Independente

X : Gols marcados pelo time mandante

Y : Gols marcados pelo time visitante

$X \perp Y$

Modelo

$$X_i \sim \text{Poi}(\lambda_{icv}); \quad E[X_i] = \lambda_{icv} = e^{\mathbf{w}_i^T \boldsymbol{\eta}}; \quad \ln(\lambda_{icv}) = \gamma_i + \alpha_{ic} + \beta_{iv}$$

$$Y_i \sim \text{Poi}(\lambda_{ivc}); \quad \ln(\lambda_{ivc}) = \alpha_{iv} + \beta_{ic}$$

- Trabalhos anteriores
 - Maher (1982) – Campeonato Inglês
 - Lee (1997) – Campeonato inglês
 - Dixon e Colles (1997) – Betting market
 - Rue e Salvesen (2000) - Bayesiana
 - Karllis e Noufras (2000) – Binomial x Poisson
 - Goddard (2005) – Modelar gols x Modelar placar
 - Farias (2008) - Brasileirão
 - Salazar (2017)

Dixon e Coles - Modelo

$$P(X = x, Y = y) = \tau_{\lambda, \mu}(x, y) \frac{\lambda^x \exp(-\lambda)}{x!} \frac{\mu^y \exp(-\mu)}{y!}$$

$$\tau_{\lambda, \mu}(x, y) = \begin{cases} 1 - \lambda\mu\rho & \text{se } x = y = 0 \\ 1 + \lambda\rho & \text{se } x = 0, y = 1 \\ 1 + \mu\rho & \text{se } x = 1, y = 0 \\ 1 - \rho & \text{se } x = y = 1 \\ 1 & \text{se c.c} \end{cases}$$

$$\lambda = \exp(\gamma\alpha_c\beta_v); \quad \mu = \exp(\alpha_v\beta_c)$$

$$\max(-1/\lambda, -1/\mu) \leq \rho \leq \min(1/\lambda\mu, 1)$$

Poisson Bivariada de Holgate

X : Gols marcados pelo time mandante

Y : Gols marcados pelo time visitante

$$(X, Y) \sim BP(\lambda_1, \lambda_2, \lambda_3)$$

Construção

$$Z_1 \perp\!\!\!\perp Z_2 \perp\!\!\!\perp Z_3; \quad \text{tal que } Z_i \sim Poi(\lambda_i); i = 1, 2 \text{ e } 3$$

$$X = Z_1 + Z_3; \quad Y = Z_2 + Z_3;$$

$$X \sim Poi(\lambda_1 + \lambda_3); Y \sim Poi(\lambda_2 + \lambda_3); Cov(X, Y) = \lambda_3$$

Poisson Bivariada - Modelo

$$P(X = x, Y = y) = e^{-(\lambda_1 + \lambda_2 + \lambda_3)} \sum_{i=0}^{\min(x,y)} \frac{\lambda_1^{x-i}}{(x-i)!} \frac{\lambda_2^{y-i}}{(y-i)!} \frac{\lambda_3^i}{(i)!}$$

$$\ln(\lambda_{1icv}) = \gamma_i + \alpha_{ic} + \beta_{iv}$$

$$\ln(\lambda_{2ivc}) = \alpha_{iv} + \beta_{ic}$$

$$\ln(\lambda_{3ivc}) = k_i$$

- Trabalhos anteriores
 - Arruda (2000)
 - Karlis e Ntzoufras (2003)
 - Suzuki (2007)
 - Silva(2014) - BFGS

Binomial - Poisson / Gols

X : Gols marcados pelo time mandante

Y : Gols marcados pelo time visitante

$X \perp Y$; $X_i \sim \text{Bin}(Z_i, p_i | Z_i)$; $Y_i \sim \text{Bin}(W_i, p_i | W_i)$

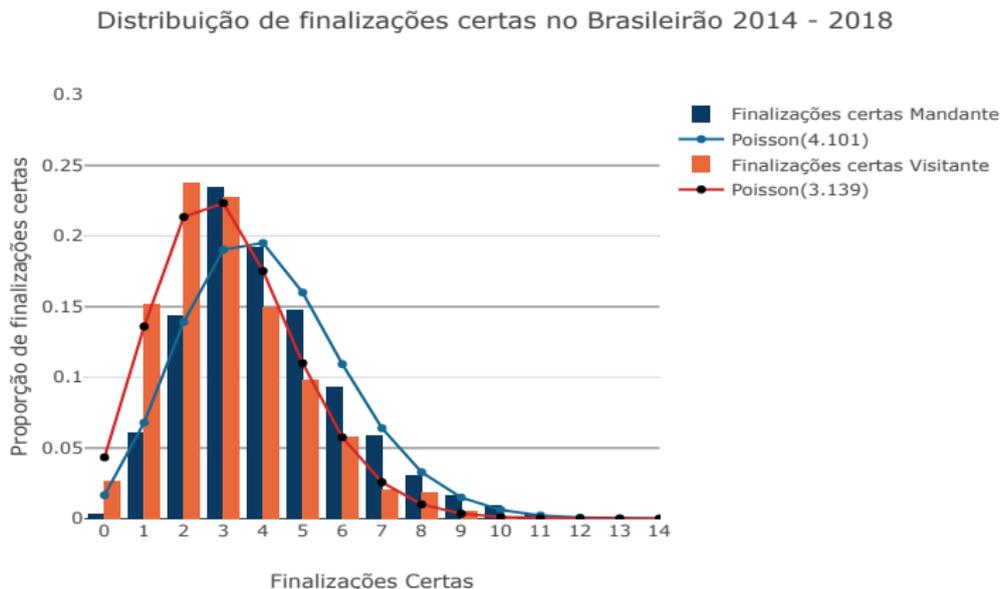
Finalizações certas

- Z : Número de finalizações certas do time mandante
- W : Número de finalizações certas do time visitante
- $Z_i \sim \text{Poisson}(\lambda_{icv})$; $\ln(\lambda_{icv}) = \gamma_i + \alpha_{ic} + \beta_{iv}$
- $W_i \sim \text{Poisson}(\lambda_{ivc})$; $\ln(\lambda_{ivc}) = \alpha_{iv} + \beta_{ic}$

Probabilidades de gols marcados - Lei da probabilidade total

$$P(X = x) = \sum_{Z=1}^{20} \text{Bin}(x; z, p) \times \text{Poi}(z; \lambda)$$

• Finalizações Certas



- Probabilidade de vitória = $P(X > Y)$
- Probabilidade de empate = $P(X = Y)$
- Probabilidade de derrota = $P(X < Y)$

Conteúdo

- 1 Introdução
- 2 Objetivos
- 3 Justificativa
- 4 Revisão de Literatura & Fundamentação Teórica
- 5 Metodologia**
- 6 Resultados
- 7 Conclusão

Metodologia

Modelos propostos

Modelos	Parâmetros						
	$\ln(E[X])$	Casa	Ataque	Defesa	Fina	FinC	RB
Lee	λ_i	γ_{1i}	α_{1ci}	β_{1vi}	0	0	0
Cartola Fin + RB	λ_i	γ_{2i}	α_{2ci}	β_{2vi}	$x_{4ci} * \delta_{2ci}$	0	$x_{6ci} * \theta_{2ci}$
Cartola FinC + RB	λ_i	γ_{3i}	α_{3ci}	β_{3vi}	0	$x_{5ci} * \eta_{3ci}$	$x_{6ci} * \theta_{3ci}$
Cartola Fin	λ_i	γ_{4i}	α_{4ci}	β_{4vi}	$x_{4ci} * \delta_{4ci}$	0	0
Cartola FinC	λ_i	γ_{5i}	α_{5ci}	β_{5vi}	0	$x_{5ci} * \eta_{5ci}$	0
Cartola Bin-Poisson	μ_i	γ_{6i}	α_{6ci}	β_{6vi}	0	0	0
Dixon e Coles	λ_i	γ_{7i}	α_{7ci}	β_{7vi}	0	0	0
Poisson Bivariada	λ_i	γ_{8i}	α_{8ci}	β_{8vi}	0	0	0

Métricas

- Medida de de Finetti
 - $\sum_{j=1}^3 (p_j - o_j)^2$
- Rank Probability Score (RPS)
 - $\frac{1}{r-1} \sum_{i=1}^r \left(\sum_{j=1}^i p_j - \sum_{j=1}^i o_j \right)^2$
- Proporção de acertos
 - $\sum_{i=1}^n \frac{W_i}{n}$
- Taxa de acerto de placar
 - $\sum_{j=1}^n (p_j \times o_j)$

Métricas - Exemplo

- Resultado Casa 3 x 2 Visitante

- Previsão Modelo 1 P (0.6, 0.2, 0.2)

- $Finetti = (0.6 - 1)^2 + (0.2 - 0)^2 + (0.2 - 0)^2 = 0.24$

- $RPS = (0.6 - 1)^2 + (0.8 - 1)^2 + (1 - 1)^2 = 0.10$

- Previsão Modelo 2 P (0.6, 0.3, 0.1)

- $Finetti = (0.6 - 1)^2 + (0.3 - 0)^2 + (0.1 - 0)^2 = 0.26$

- $RPS = (0.6 - 1)^2 + (0.9 - 1)^2 + (1 - 1)^2 = 0.09$

Palpites e valores referência

- Modelo ingênuo 1 $P = (1/3, 1/3, 1/3)$
 - Todos jogos são atribuídos $P = (1/3, 1/3, 1/3)$.
 - Finetti = 0.66
- Modelo ingênuo 2 - Palpite Bra $P = (0.51, 0.26, 0.23)$
 - Todos os jogos são atribuídos $P = (0.51, 0.26, 0.23)$.

Conteúdo

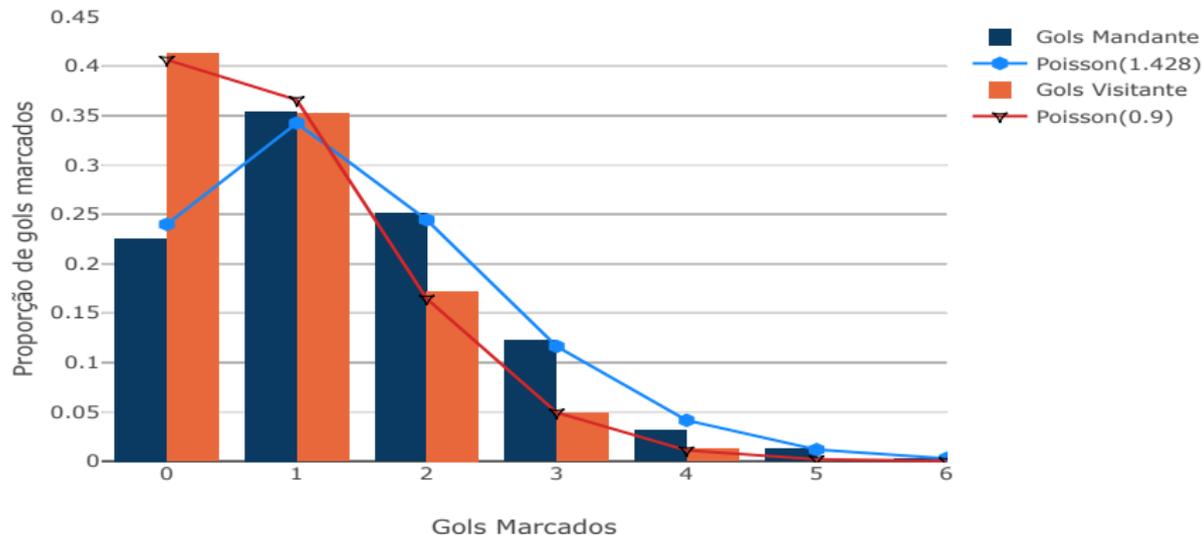
- 1 Introdução
- 2 Objetivos
- 3 Justificativa
- 4 Revisão de Literatura & Fundamentação Teórica
- 5 Metodologia
- 6 Resultados**
- 7 Conclusão

- Amostra da base de dados

Clube	Rodada	Ano	FS	FT	FD	FF	RB	FC	CA	FinC	Finalizacoes
Corinthians	18	2017	18.000	0.000	2.000	3.000	16.000	12.000	1.000	2.000	5.000
América-MG	28	2016	13.000	0.000	0.000	6.000	10.000	15.000	2.000	0.000	6.000
Coritiba	5	2014	12.000	0.000	2.000	1.000	13.000	12.000	1.000	2.000	3.000
Cruzeiro	23	2017	15.000	0.000	5.000	2.000	10.000	17.000	1.000	5.000	7.000
Chapecoense	7	2018	20.000	1.000	4.000	5.000	19.000	14.000	1.000	5.000	10.000
São Paulo	30	2018	13.000	2.000	5.000	5.000	20.000	20.000	3.000	7.000	12.000
Palmeiras	34	2016	18.000	0.000	2.000	3.000	10.000	9.000	0.000	2.000	5.000
Chapecoense	24	2017	6.000	0.000	4.000	2.000	20.000	18.000	5.000	4.000	6.000

- Vantagem de jogar em casa

Distribuição de gols marcados no Brasileirão 2014 - 2018



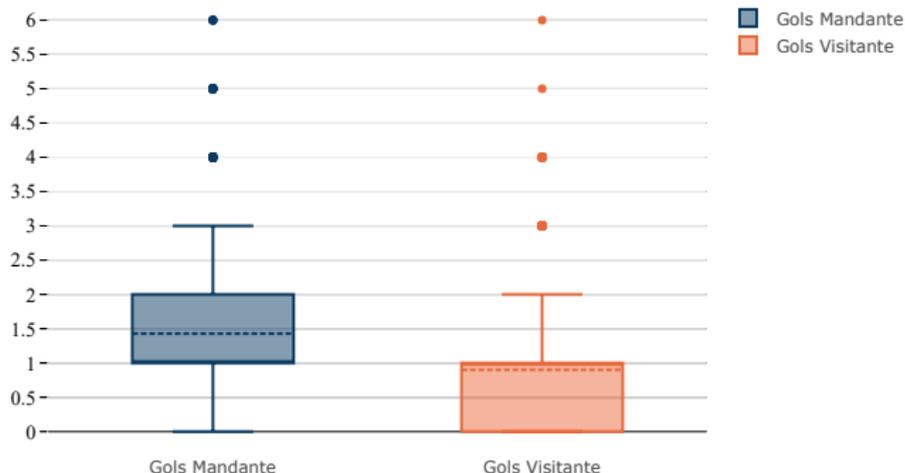
Estatística	Gols mandante	Gols visitante
Média	1,43	0,90
Variância	1,34	0,91
Min	0	0
Max	6	6
p25	1	0
p50	1	1
p75	2	1

Tabela: Estatísticas descritiva dos gols marcados

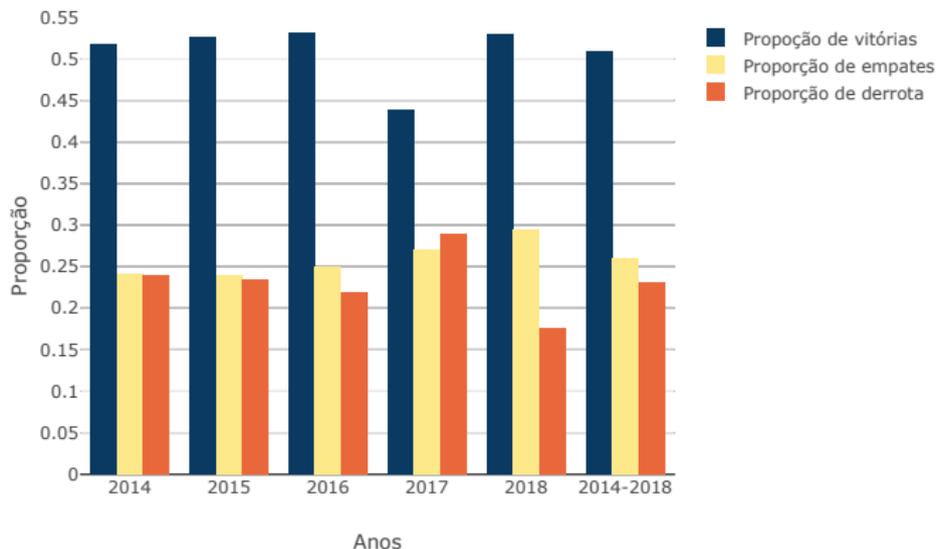
Resultados

Análise descritiva

Box Plot da distribuição de gols marcados no Campeonato Brasileiro 2014 - 2018



Proporção de vitórias, empates e derrotas no Brasileirão 2014 - 2018



Resultados

Análise dos modelos 1 - passo

Modelos	de Finetti				
	2014	2015	2016	2017	2018
Modelo 1	0.5873	0.6264	0.5842	0.6899	0.5595
Modelo 2	0.5854	0.6264	0.5813	0.6896	0.5575
Modelo 3	0.5889	0.6294	0.5861	0.6911	0.5625
Modelo 4	0.5856	0.6253	0.5811	0.6892	0.5560
Modelo 5	0.5893	0.6286	0.5853	0.6904	0.5611
Modelo 6	0.5916	0.6313	0.5949	0.6921	0.5700
Modelo 7	0.6101	0.6346	0.5924	0.6956	0.6065
Modelo 8	0.5988	0.6248	0.5850	0.6918	0.5576
Palpite Bra	0.5956	0.6181	0.6055	0.6554	0.6023

Figura: Tabela de comparação das medidas de de Finetti 1 passo

Resultados

Análise dos modelos 1 - passo

Modelos	RPS				
	2014	2015	2016	2017	2018
Modelo 1	0.2069	0.2246	0.1997	0.2421	0.1777
Modelo 2	0.2062	0.2245	0.1984	0.2415	0.1769
Modelo 3	0.2077	0.2257	0.2009	0.2428	0.1791
Modelo 4	0.2063	0.2241	0.1981	0.2413	0.1761
Modelo 5	0.2078	0.2254	0.2003	0.2425	0.1784
Modelo 6	0.2103	0.2233	0.2051	0.2421	0.1797
Modelo 7	0.2206	0.2292	0.2038	0.2445	0.1985
Modelo 8	0.2107	0.2237	0.1987	0.2435	0.1780
Palpite Bra	0.2135	0.2209	0.2109	0.2244	0.1953

Figura: Tabela de comparação das medidas RPS 1 passo

Resultados

Análise dos modelos 1 - passo

Modelos	Proporção de acertos				
	2014	2015	2016	2017	2018
Modelo 1	0.5211	0.4842	0.5368	0.3579	0.5526
Modelo 2	0.5211	0.4947	0.5263	0.3737	0.5632
Modelo 3	0.5211	0.4895	0.5421	0.3474	0.5526
Modelo 4	0.5158	0.5000	0.5368	0.3789	0.5632
Modelo 5	0.5211	0.4895	0.5316	0.3474	0.5526
Modelo 6	0.5368	0.5000	0.5474	0.4105	0.4737
Modelo 7	0.5316	0.4737	0.5211	0.3632	0.4895
Modelo 8	0.5000	0.4842	0.5421	0.3474	0.5842

Figura: Tabela de comparação das proporções de acertos 1 passo

Resultados

Análise dos modelos 1 - passo

Modelos	Taxa de acerto de placar				
	2014	2015	2016	2017	2018
Modelo 1	0.0894	0.0779	0.0874	0.0793	0.1006
Modelo 2	0.0894	0.0776	0.0881	0.0799	0.1011
Modelo 3	0.0893	0.0775	0.0875	0.0792	0.1005
Modelo 4	0.0895	0.0781	0.0879	0.0799	0.1011
Modelo 5	0.0894	0.0778	0.0874	0.0793	0.1005
Modelo 6	0.0847	0.0749	0.0869	0.0810	0.0944
Modelo 7	0.0796	0.0707	0.0829	0.0783	0.0853
Modelo 8	0.0887	0.0776	0.0884	0.0803	0.1043

Figura: Tabela de comparação das taxas de acerto de placar 1 passo

Resultados

Tabela brasileiro 2018 - 30°

Classificação	PG	J	V	E	D	GM	GS	SG
1° Palmeiras	62	30	18	8	4	47	19	28
2° Flamengo	58	30	17	7	6	48	22	26
3° Internacional	57	30	16	9	5	41	22	19
4° São Paulo	53	30	14	11	5	40	27	13
5° Grêmio	52	30	14	10	6	38	19	19
6° Atlético-MG	46	30	13	7	10	47	35	12
7° Santos	43	30	11	10	9	35	28	7
8° Atlético-PR	40	30	11	7	12	41	29	12
9° Fluminense	40	30	11	7	12	31	35	-4
10° Cruzeiro	40	29	10	10	9	25	25	0
11° Bahia	37	30	9	10	11	31	34	-3
12° Corinthians	36	30	9	9	12	30	29	1
13° Botafogo	35	30	8	11	11	29	40	-11
14° Vasco	34	30	8	10	12	35	42	-7
15° América-MG	34	30	8	10	12	26	35	-9
16° Vitória	33	30	9	6	15	30	51	-21
17° Ceará	31	29	7	10	12	23	31	-8
18° Chapecoense	31	30	7	10	13	29	45	-16
19° Sport	30	30	8	6	16	27	50	-23
20° Paraná	17	30	3	8	19	12	47	-35

Figura: Tabela Brasileiro 30° rodada

Resultados

Chances brasileiro 2018 - 30°

Clubes	Campeão	Libertadores	Rebaixamento
América - MG	0.00%	0.01%	19.07%
Atlético - MG	0.00%	67.57%	0.00%
Atlético - PR	0.00%	8.00%	0.02%
Bahia - BA	0.00%	0.72%	0.92%
Botafogo - RJ	0.00%	0.01%	19.85%
Ceará - CE	0.00%	0.00%	32.96%
Chapecoense - SC	0.00%	0.00%	59.77%
Corinthians - SP	0.00%	0.12%	2.51%
Cruzeiro - MG	0.00%	1.64%	0.06%
Flamengo - RJ	11.55%	100.00%	0.00%
Fluminense - RJ	0.00%	1.66%	0.08%
Grêmio - RS	0.12%	99.82%	0.00%
Internacional - RS	6.61%	100.00%	0.00%
Palmeiras - SP	81.61%	100.00%	0.00%
Paraná - PR	0.00%	0.00%	100.00%
Santos - SP	0.00%	20.97%	0.00%
Sport - PE	0.00%	0.00%	82.19%
São Paulo - SP	0.11%	99.49%	0.00%
Vasco da Gama - RJ	0.00%	0.00%	38.28%
Vitória - BA	0.00%	0.00%	44.29%

Figura: Chances Brasileiro 30° rodada

Conteúdo

- 1 Introdução
- 2 Objetivos
- 3 Justificativa
- 4 Revisão de Literatura & Fundamentação Teórica
- 5 Metodologia
- 6 Resultados
- 7 Conclusão

- Melhor modelo baseado na medida de de Finetti e RPS
 - O modelo 4 - Cartola Fin.
- Melhor modelo baseado na proporção de acertos
 - Modelo 6 - Bin-Poi
- Ganho no uso das variáveis do Cartola FC.

- Utilizar variáveis do Cartola FC no modelo Poisson Bivariado;
- Explorar o uso de técnicas de *machine learning* para prever resultados;
- Estimar o p do modelo Bin-Poisson através de algum tipo de modelo.

Obrigado!

 Michael J Maher.
Modelling association football scores.
Statistica Neerlandica, 36(3):109–118, 1982.

 Alan J Lee.
Modeling scores in the premier league: is manchester united really the best?
Chance, 10(1):15–19, 1997.

 Mark J Dixon and Stuart G Coles.
Modelling association football scores and inefficiencies in the football betting market.
Journal of the Royal Statistical Society: Series C (Applied Statistics), 46(2):265–280, 1997.

-  Wesley Bertoli da Silva et al.
Distribuição de poisson bivariada aplicada à previsão de resultados esportivos.
Master's thesis, 2014.
-  Snorre Gebhardt Stenerud.
A study on soccer prediction using goals and shots on target.
Master's thesis, NTNU, 2015.

Previsões de Resultados em Partidas do Campeonato Brasileiro de Futebol

João Marcos Amorim dos Santos
Moacyr Alvim Horta Barbosa da Silva
Rodrigo dos Santos Targino

FGV | EMap - Escola de Matemática Aplicada

27 de junho de 2019

