

# Premier League via Dirichlet Regression

Nascimento, Igor Ferreira do

Departamento de Administração - UnB; Laboratório de Aprendizado de Máquina em Finanças

## Resumo

O artigo apresenta uma metodologia para estimar a probabilidades dos resultados dos jogos da Premier League utilizando covariáveis dos jogadores.

*Keywords:* Regularization, Dirichlet Regression, Premier League, Forecast

## Introdução

O fascínio por eventos incertos instiga o ser humano desde os remotos tempos dos antepassados. Os historiadores XXXX mostram que desde a época o ser humano demonstrava interesse por saber ao menos algo sobre o que é incerto. Ao longo dos anos, a teoria dos jogos proporcionou o desenvolvimento de modelos probabilísticos para determinar o nível de determinadas ocorrências.

O objetivo do trabalho é apresentar modelos probabilísticos para representar os resultados da Premier League por meio de covariáveis relacionado ao capital humano do time e os recursos financeiros relacionados ao time. Além disso, é apresentado uma estimativa para o resultado final da tabela de classificação.

## Sports Forecast

Sport is a human activity normally related to the playful environment but we can not dissemble the betting. The work like (Kain & Logan, 2014) analyse the dissonance between the outcome of sports and the sports betting market explained by fact the set betting lines is straight related with the profit goal instead the outcome probability. On the other hand, the paper (Štrumbelj & Šikonja, 2010) analysed some European soccer leagues and show that exist evidence that bookmakers learn over time about soccer prediction results. Anyway, this is interesting to either bettors or bookmakers researchers in this area.

Dito isso, pode-se destacar que os resultados esportivos possuem uma intenso interesse da sociedade como um toda, seja ela do ponto de vista financeiro ou a mera busca por entretenimento.

As redes de apostas como, XXX, YYY, GGG SÃO encontrados ao redor do mundo e fazem girar cifras ainda maiores de recursos no ato de acreditar em um determinado resultado. Esportes como o baisedball e o futebol americano possuem grande suporte dos cientistas de dados para apresentar pequenas melhoras nos resultados individuais dos atletas que aparentam ser inexpressíveis, ao menos aos olhos dos descuidados, porém representam um grande impacto no coletivo e conseqüentemente, em toda a cadeia produtiva associada ao esporte. De encontro com tais anseios está o vertiginoso desenvolvimento de tecnologias de captação, armazenamento, tratamento e análise de dados relacionados às atividades esportivas.

O futebol é, se não o mais, um dos esportes com as movimentações de destaque na atualidade. Atualmente, existem elenco de jogadores em times que passam a casa dos 10 dígitos,

sobre tudo na Europa Ocidental. Os times Espanhóis e Britânicos lideram os valores financeiros relacionados ao capital humano disponível. No entanto, essa aparente hegemonia é, cada vez mais, desafiada com o surgimento de novos mercados da bola. O mercado Norte Americano tem investido mais de XXXXX milhões de dólares em um projeto de incentivo à prática do futebol, desafiador em uma nação com outros esportes com importância sacramentada como é o caso do Futebol Americano e Basquete.

O evento mais recente e com maior destaque no mundo do esporte foi o valor total da transferência de Neymar do Barcelona para o Paris Saint German, um total de mais 400 milhões de euros. Seria ingenuidade não encarar tal transação como um investimento. É esperado que as externalidades dessa movimentação tenha reflexos comparadas com os maiores investimentos nos mercados de financeiros do mundo. Por exemplo, a venda de camisas aumentou 123123123321% no faturamento do clube apenas com vendas de camisas.

### Regressão Dirichlet

A distribuição Dirichlet permite modelar um vetor composicional de dimensão  $k \leq 2$   $Y = (y_1, y_2, \dots, y_K)$ , com parâmetros de escala  $\alpha_1, \alpha_2, \dots, \alpha_K \geq 0$ , sendo  $\sum_{j=1}^K y_j = 1$  um padrão  $K - 1$  simplex. Dessa forma, as probabilidades presentes na distribuição multinomial tem como priori conjugada tal distribuição, com densidade probabilidade:

The Dirichlet distribution is able to model a composite vector of dimension  $k \leq 2$   $Y = (y_1, y_2, \dots, y_K)$ , with scalings parameters  $\alpha_1, \alpha_2, \dots, \alpha_K \geq 0$ , being  $\sum_{j=1}^K y_j = 1$  a standard  $K - 1$  simplex. Thus the probabilities present in the multinomial distribution have as a priori conjugated such distribution, with probability density:

$$P(Y|\alpha) = \frac{\prod_{j=1}^K \Gamma(\alpha_j)}{\Gamma(\sum_{j=1}^K \alpha_j)} \prod_{j=1}^K y_j^{\alpha_j-1} \quad (1)$$

The work of (Hijazi & Jernigan, 2009) that observed a good ajust upon longitudial covariates changes using Dirichlet Regression. In field of sports the work of (Null, 2009) show a competitive accuracy modeling player abilities in Major League Baseball.

We object model the scale parameters by covariates  $X_r$ , and use this covariates to forecast the final classification result. Thus the model is:

$$g[\alpha_j(X)] = \beta_{j0} + \beta_{j1}X_1 + \beta_{j2}X_2 + \dots + \beta_{jR}X_R + \varepsilon_j \quad (2)$$

Being  $g(.) = \log(.)$  has a transformation possible. The sample  $Y$  with size  $N$  the likelihood is:

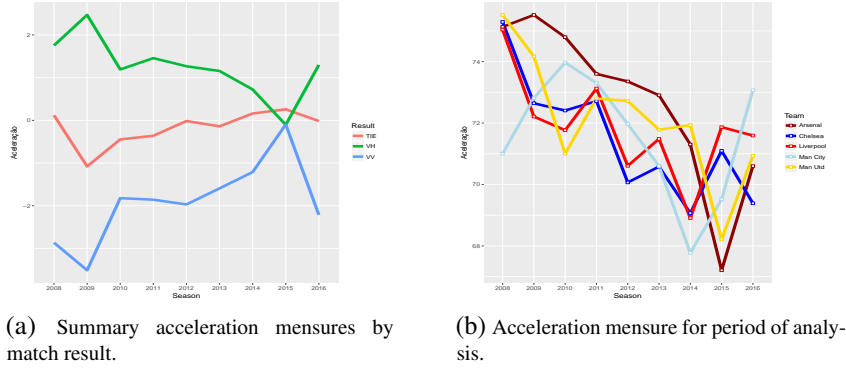
$$L(\beta|X, Y) = \prod_{i=1}^N \left[ \frac{\prod_{j=1}^K \Gamma(\alpha_j(X))}{\Gamma(\sum_{j=1}^K \alpha_c(X))} \prod_{j=1}^K y_{ij}^{\alpha_j(X)-1} \right]$$

The coefficients  $\beta_{jr}$  in each  $\alpha_j$  of Equation 2 could be mensured solvind the minimum problem as.

$$\underset{\beta}{\text{maximize}} \quad \log(L(\beta|X, Y)) - \lambda [\phi(||\beta||) + (1 - \phi)(||\beta^2||)] \quad (3)$$

Being

$$\log(L(\beta|X, Y)) = \sum_{i=1}^N \log \left[ \Gamma \left( \sum_{j=1}^K \alpha_j(X_{ri}) \right) \right] - \sum_{j=1}^K \log [\Gamma(\alpha_j(X_{ri}))] + \sum_{j=1}^K [1 - \alpha_j(X_{ri})] \log(Y_{ji})$$



**Figura 1.** (a) Summary acceleration measures by match result. Tie (TIE), Victory of Home team (VH) and Victory of visitor team (VV). (b) Skill acceleration measure for period of analysis.

Thus, we using a maximization procedure to find the  $\beta$  coefficients in equation 2.

Nosso modelo,  $K = 3$ , sendo  $\alpha_{Home}$  o parâmetro associado a probabilidade do time mandante vencer,  $\alpha_{Visitor}$  a probabilidade do time visitante vencer e  $\alpha_{Tie}$  a probabilidade para o empate dos times.

As variáveis regressoras são apresentadas a seguir.

### Base de dados

Foram consideradas como regressoras as habilidades encontradas no site do jogo virtual FIFA. São elas **Aceleração, Altura, Cabeceio, Carrinho, Ch., de, longe, Cobr., falta, Combativ., Contr., bola, Cruzamento, Div., em, pé, Dribles, Duração, Do, Contrato, Elast., GL, Finalização, Fôlego, Força, Força, chute, Idade, Lançamento, Manejo, Marcação, Passe, curto, Perna, boa, Perna, ruim, Peso, Pique, pos, Posicion., GL, Reação e Reflexos.**

Foram coletadas informações durante o período de 2008 e 2016.

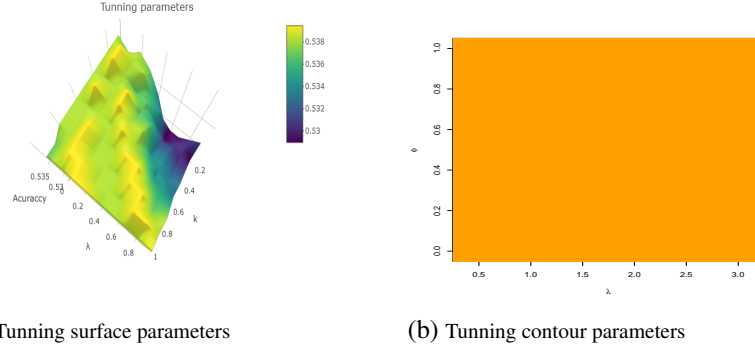
Similarmente ao *Premier League*, os dados dos disponibilizados no site da FIFA acompanham os times de cada temporada. Os dados são atualizados mais de uma vez durante semana de jogos e foram considerados os valores da versão mais atualizada de cada temporada.

Seja  $X_r^{home}$  o valor da variável regressora  $r$  para o time mandante do confronto e, analogamente,  $X_r^{visitor}$  o valor para a mesma variável do time visitante. A equação do modelo 2 é modelada pela diferença entre a habilidade na variável  $r$  entre o time mandante e o visitante.  $X_r = X_r^{home} - X_r^{visitor}$ . Dessa forma, valores muito positivos para a variável  $X_r$  significam superioridade do time mandante e os negativos significam que as habilidades do time visitante são superiores aos anfitriões.

O gráfico do painel (a) da imagem 1 mostra que, em média, os jogos em que a medida  $X_r$  é maior do que zero, houve vitória do time mandante. Do forma similar, nos casos em que  $X_r$  foi menor do que zero o time visitante venceu. Os casos tem empate, há equilíbrio entre os times e a medida  $X_r$  é próxima, em média, de zero. Em anexo estão as medidas para as demais variáveis.

O gráfico do painel (b) da imagem 1 mostra a evolução da variável aceleração ao longo do período de análise para os principais times da *Premier League*. É possível perceber que todos times tiveram entre 2008 e 2014 uma diminuição do valor médio da variável aceleração. O processo de renovação de elenco do Chelsea, Liverpool e Manchester City iniciou-se em 2015. Já para Arsenal e Manchester United a renovação ocorreu no ano seguinte em 2015.

O início de renovação do Chelse foi campeã pode ajudar a explicar o sucesso em 2015 e 2017. Uma queda dessas e outras informações das equipes inglesas podem ajudar a compreender a zebra de 2016, quando a equipe de pouca expressão Leicester foi campeã.



**Figura 2.** (a) Surface of parameters  $\lambda$  and  $k$  related with equation 3. (b) Contour of parameters  $\lambda$  and  $k$  related with equation 3.

Pretendemos explicar e prever o resultado dos jogos da Premier League, baseado nas habilidades auferidas pelo site FIFA dos jogadores de cada time.

A relação positiva é observada quando o aumentada variável regressora  $X_r$ , que representa o nível de superioridade do time mandante, contribui para a vitória do time mandante. Nesse raciocínio, é esperado nos jogos em que houve vitória do time mandante, a média dessa variável é maior do que os valores dos jogos empate, que por sua vez é maior do que os que ocorreram derrota.  $\bar{X}_r^{VH} \geq \bar{X}_r^{Tie} \geq \bar{X}_r^{VV}$

De maneira análoga, a relação negativa é observada quando o diminuição de  $X_r$  contribui para a derrota do time mandante. Nesse caso, é esperado que nos jogos em que houve vitória do time mandante, a média dessa variável é menor do que os valores dos jogos empate, que por sua vez é menor do que os que ocorreram derrota.  $\bar{X}_r^{VH} \leq \bar{X}_r^{Tie} \leq \bar{X}_r^{VV}$

Além desse resultado geral, é esperado que ao longo de **cada temporadas analisadas** esse comportamento se repita. O número de vezes em que esse resultado é observado pode sugerir o nível dessa relação.

A tabela 1 apresenta essas análises para cada variável e dá indícios para o processo de seleção das variáveis relevantes para o modelo. É possível identificar, que as variáveis “chute de longe” e “controle de bola” um comportamento adequado com a relação positiva. Também é possível notar que as variáveis “idade” e “classificação na tabela” têm relação negativa com a vitória do mandante.

## Método

A base de dados foi dividida em:

- **base de treino:** jogos das temporadas de 2008 e 2014. A base servirá para estimar os parâmetros do modelo da equação 2.
- **base de tuning:** jogos das temporadas de 2015. A base servirá para estimar os parâmetros  $\lambda$  e  $\phi$  do processo de **regularização** na equação 3.
- **base de teste:** jogos das temporadas de 2016. A base servirá para avaliar *out-of-sample* o modelo.

Os parâmetros de  $\phi$  e  $\lambda$  foram avaliados por meio do *grid-search*, sendo  $\phi = (0, 0.1, 0.2, 0.3, \dots, 0.9, 1)$  e  $\lambda = (0, 0.5, 1, 1.5, 2, 2.5, 3)$ . Os gráficos da figura 2 apresentam o desempenho do modelo para o *grid* de procura dos parâmetros.

### **Modelo 1**

O modelo 1 utiliza todas as variáveis disponíveis e utiliza medida de dissimilaridade para empate, vitória ou derrota do time visitante.

O resultado para o modelo proposto não consegue identificar os resultados empate.

A tabela 3 apresenta a simulação para o resultado da temporada de 2017/2018 baseada nos parâmetros estimados.

## Modelo 2

O modelo 2 adiciona medida de similaridade para modelar o empate.

Isso ocorre devido o valor de  $X_r$  representar uma medida de **dissimilaridade**, isto é, quanto maior o valor  $X_r$  mais diferente são as habilidades do time mandante e visitante. Caso as habilidades dos dois times sejam iguais o valor  $X_r$  zera, inviabilizando a regressão do parâmetro  $\alpha_{Tie}$ .

Dessa forma, propomos uma medida de **similaridade** para regressão do parâmetro  $\alpha_{Tie}$  relativo ao empate. Faremos:

$$X_r^{Tie} = \gamma \exp \left\{ -\sigma X_r^2 \right\} \quad (4)$$

Nessa estratégia, os parâmetros  $\lambda$ ,  $\phi$ ,  $\sigma$  e  $\gamma$  são determinados utilizando a procedimento *tuning*. O modelo para  $\alpha_{Tie}$  captará o nível de similaridade dos times.

Apesar de pequena, o tratamento para a regressão do coeficiente do empate contribui para melhorar o poder de previsibilidade do modelo.

A tabela 6 apresenta as medidas descritivas utilizando a transformação 4 com parâmetro  $\gamma = 1$  e  $\sigma = 1$ .

A tabela 6 dá indícios de que as variáveis “A”

### Modelo 3

O modelo 3 restringe as variáveis baseado nas análises preliminares do modelo 1 e 2.

O primeiro critério seleciona as variáveis para modelar a vitória e a derrota do time mandante. A partir da tabela 1 foram selecionados aqueles que possuem a variável NP e NN com valor superior ou igual a 80%, isso é, variáveis que possuem relação positiva ou negativa em mais de 80% das temporadas.

A figura 3 destaca as variáveis selecionadas.

Com esse critério são selecionadas as variáveis Aceleração, Ch. de longe, Cobr. falta, Contr. bola, Cruzamento, Div. em pé, Dribles, Duração Do Contrato, Força chute, Idade, Lançamento, Marcação, Passe curto, Pique, Posição e Reação.

O segundo critério seleciona as variáveis para modelar o empate do time mandante. A partir da tabela 6 foram selecionados aqueles que possuem a variável MNP com valor superior ou igual a 1, isso é, variáveis as quais a transformação 4 resulta em uma alta similaridade nas habilidades dos times que empataram e baixa similaridade dos resultados que não ocorreram empate.

A figura 3 destaca as variáveis selecionadas. Visualmente, esse filtro é realizado nos gráficos que possuem trajetórias em verde, empate, próxima de 1 (alta similaridade) e sempre superiores às trajetórias de vitória e derrota do mandante, estando esses próximos de 0 (baixa similaridade).

### Resultados

#### Comparação

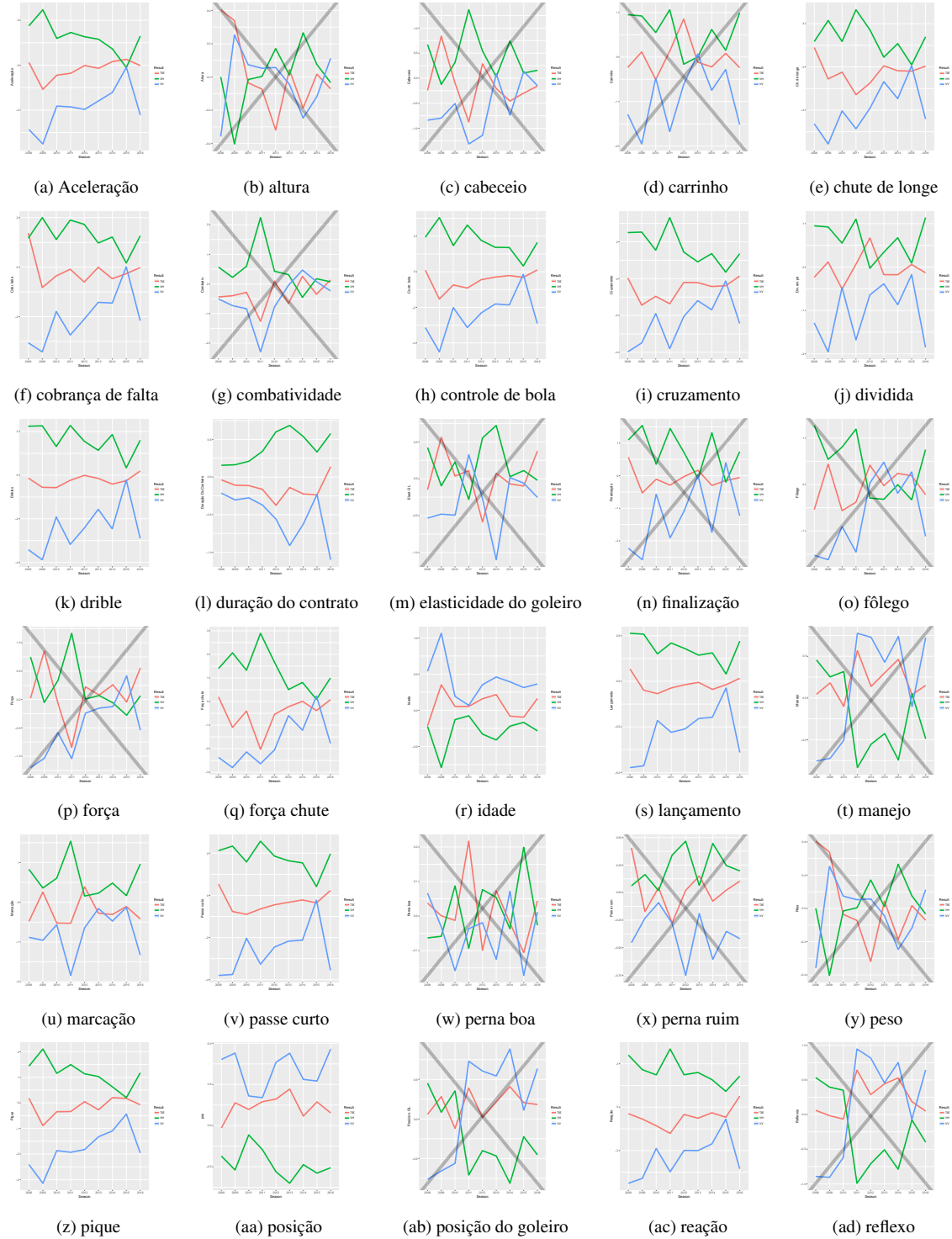
Com esse critério são selecionadas as variáveis Aceleração, Cobr. falta, Contr. bola, Cruzamento, Dribles, Lançamento, Passe curto, Pique e Reação.

Todas as variáveis selecionadas para o modelo do empate estão no modelo de vitória ou derrota do time mandante.

### Considerações finais

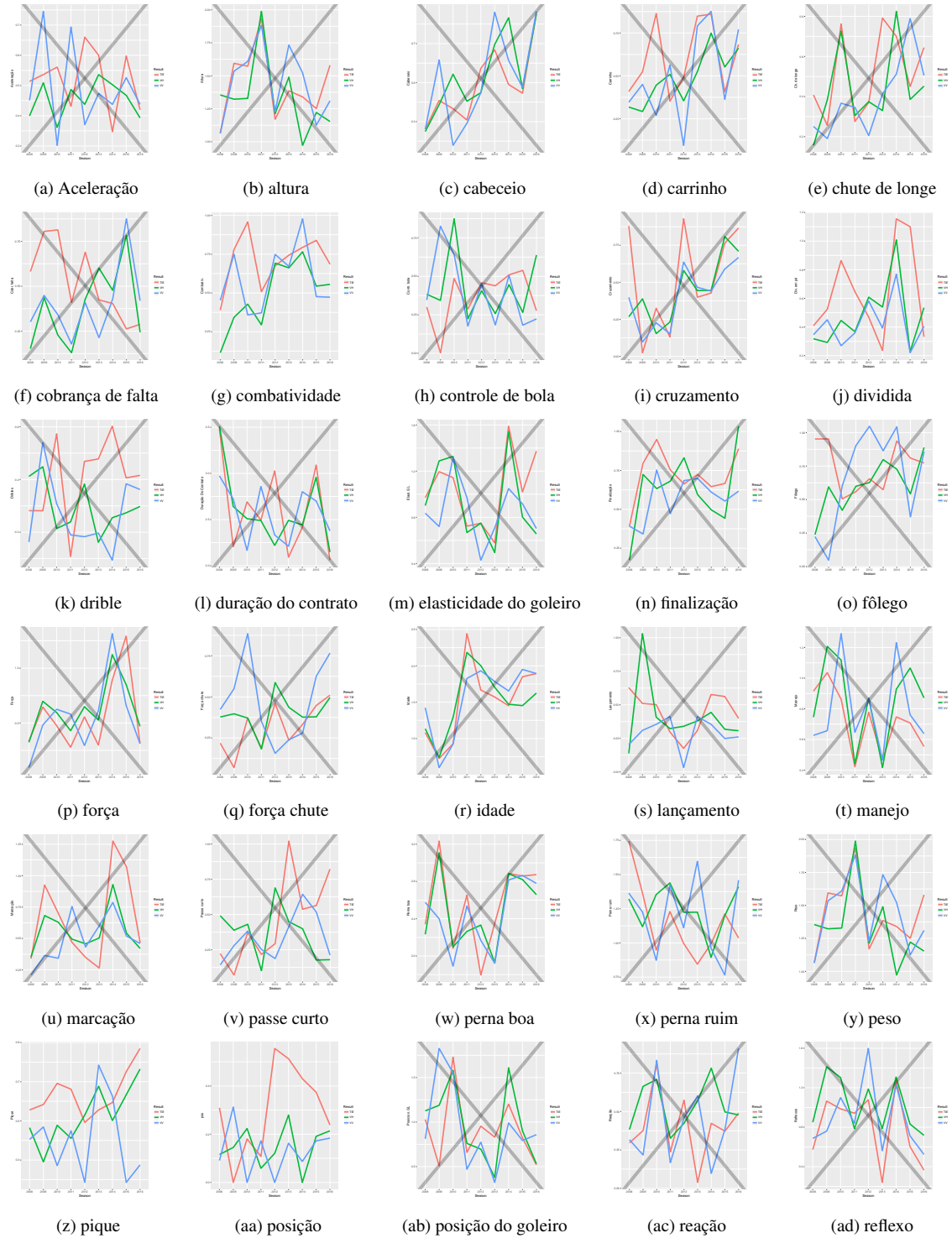
#### Referências

- Hijazi, R. H., & Jernigan, R. W. (2009). Modeling Compositional Data Using Dirichlet Regression Models. *Journal of Applied Probability and Statistics*.
- Kain, K. J., & Logan, T. D. (2014). Are Sports Betting Markets Prediction Markets? *Journal of Sports Economics*. doi: 10.1177/1527002512437744
- Null, B. (2009). Modeling Baseball Player Ability with a Nested Dirichlet Distribution. *Journal of Quantitative Analysis in Sports*. doi: 10.2202/1559-0410.1175
- Štrumbelj, E., & Šikonja, M. R. (2010). Online bookmakers' odds as forecasts: The case of European soccer leagues. *International Journal of Forecasting*. doi: 10.1016/j.ijforecast.2009.10.005

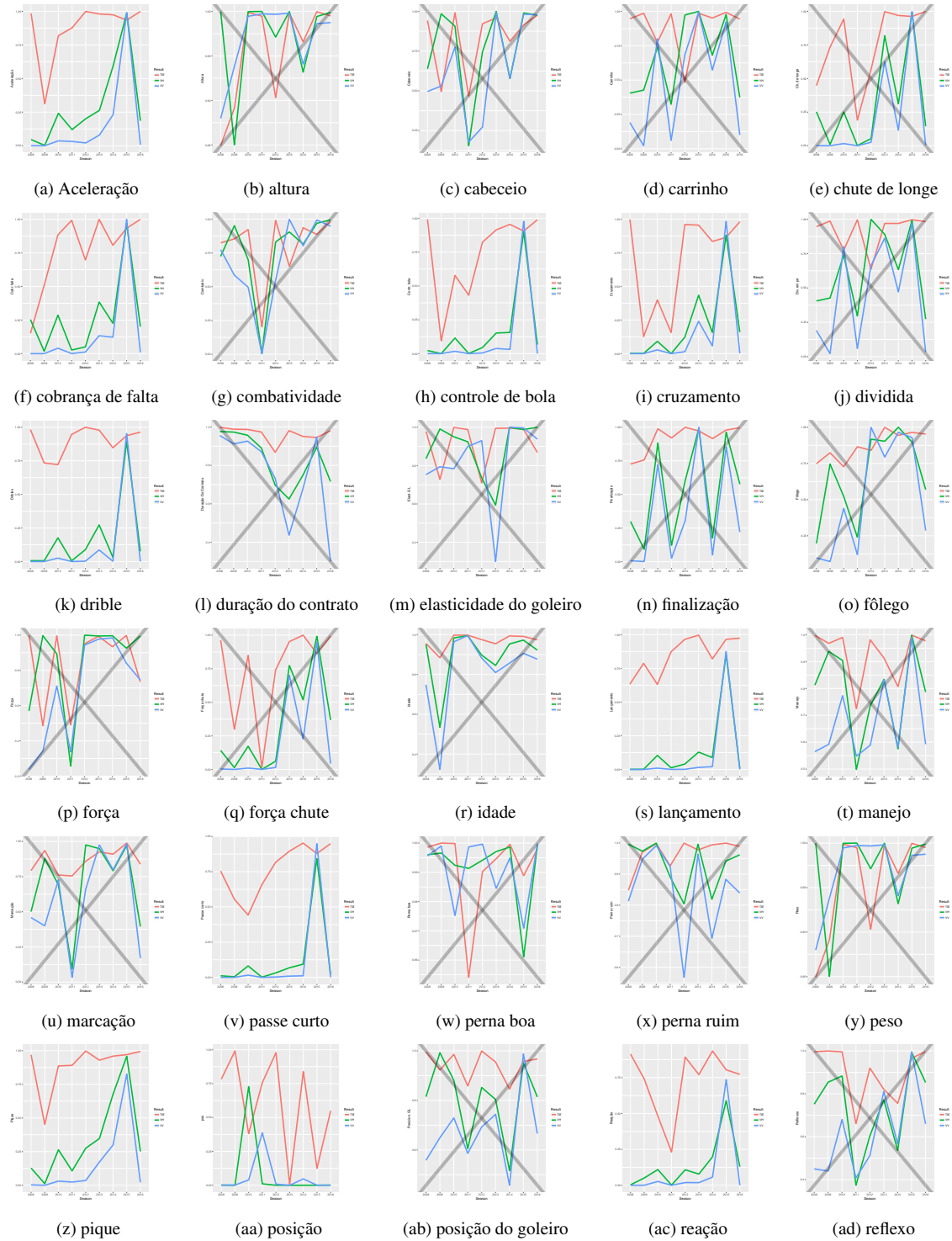


**Figura 3.** Medidas resumo ao longo das temporadas para cada variável disponível. O marcador *X* é inserido para os gráficos das variáveis não utilizadas para o modelo, por apresentar inconsistências nas relações positivas e negativas.

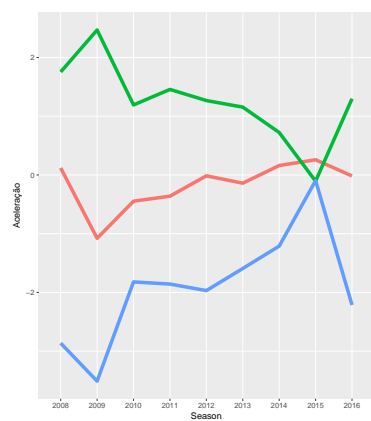




**Figura 4.** Medidas resumo transformada ao longo das temporadas para cada variável disponível. O marcador X é inserido para os gráficos das variáveis não utilizadas para o modelo, por apresentar inconsistências nas relações positivas e negativas.



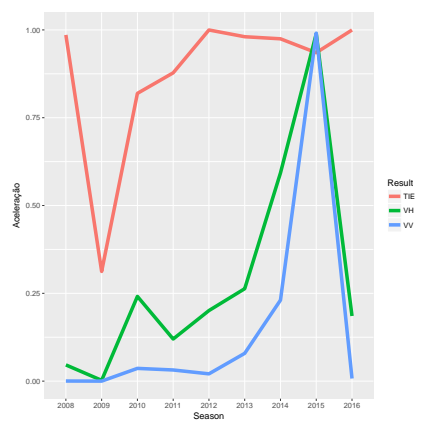
**Figura 5.** Medidas resumo média transformada ao longo das temporadas para cada variável disponível. O marcador  $X$  é inserido para os gráficos das variáveis não utilizadas para o modelo, por apresentar inconsistências nas relações positivas e negativas.



(a) Aceleração



(b) Aceleração transformado



(c) Aceleração transformado na média

**Figura 6.** Medidas resumo transformada ao longo das temporadas para cada variável disponível. O marcador *X* é inserido para os gráficos das variáveis não utilizadas para o modelo, por apresentar inconsistências nas relações positivas e negativas.

	Média			Indicador positivo				Indicador negativo			
Skill	$\bar{X}_r^{VH}$	$\bar{X}_r^{TIE}$	$\bar{X}_r^{VV}$	NVH	NVV	NTIE	NP	NVH	NVV	NTIE	NN
Aceleração	1.173	-0.095	-1.748	88.9	88.9	88.9	88.9	11.1	0.0	0.0	11.1
Altura	0.036	-0.017	-0.052	33.3	44.4	22.2	33.3	11.1	33.3	0.0	11.1
Cabeceio	0.458	-0.202	-0.557	66.7	66.7	55.6	66.7	0.0	22.2	0.0	0.0
Carrinho	0.549	-0.015	-0.852	77.8	77.8	66.7	77.8	0.0	11.1	0.0	0.0
Ch. de longe	1.266	-0.236	-1.781	100.0	88.9	88.9	100.0	0.0	0.0	0.0	0.0
Cobr. falta	1.240	-0.095	-1.878	88.9	88.9	77.8	88.9	0.0	0.0	0.0	0.0
Combativ.	0.479	-0.307	-0.513	77.8	66.7	55.6	77.8	11.1	11.1	11.1	11.1
Contr. bola	1.644	-0.402	-2.234	100.0	88.9	88.9	100.0	0.0	0.0	0.0	0.0
Cruzamento	1.651	-0.475	-2.181	100.0	88.9	88.9	100.0	0.0	0.0	0.0	0.0
Div. em pé	0.630	-0.040	-0.958	88.9	88.9	77.8	88.9	0.0	0.0	0.0	0.0
Dribles	1.587	-0.220	-2.306	100.0	88.9	88.9	100.0	0.0	0.0	0.0	0.0
Dur. contrato	0.419	-0.151	-0.535	100.0	88.9	88.9	100.0	0.0	0.0	0.0	0.0
Elast. GL	0.205	-0.032	-0.302	66.7	55.6	33.3	66.7	11.1	11.1	11.1	11.1
Finalização	0.720	-0.047	-1.091	77.8	77.8	77.8	77.8	22.2	11.1	11.1	22.2
Fôlego	0.405	-0.087	-0.567	55.6	66.7	55.6	55.6	33.3	22.2	22.2	33.3
Força	0.247	0.049	-0.445	33.3	88.9	33.3	33.3	11.1	11.1	11.1	11.1
Força chute	1.255	-0.447	-1.577	88.9	88.9	88.9	88.9	0.0	11.1	0.0	0.0
Idade	-0.258	0.039	0.364	0.0	0.0	0.0	0.0	100.0	100.0	100.0	100.0
Lançamento	1.761	-0.165	-2.630	100.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0
Manejo	-0.254	0.187	0.239	44.4	44.4	44.4	44.4	55.6	55.6	55.6	55.6
Marcação	0.618	-0.251	-0.757	88.9	88.9	77.8	88.9	0.0	0.0	0.0	0.0
Passe curto	1.756	-0.308	-2.494	100.0	88.9	88.9	100.0	0.0	0.0	0.0	0.0
Perna boa	0.019	0.010	-0.044	33.3	33.3	22.2	33.3	55.6	22.2	22.2	55.6
Perna ruim	0.223	0.039	-0.390	66.7	100.0	66.7	66.7	0.0	0.0	0.0	0.0
Peso	0.036	-0.017	-0.052	33.3	44.4	22.2	33.3	11.1	33.3	0.0	11.1
Pique	1.092	-0.082	-1.641	100.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0
Posição	-2.263	1.113	3.310	0.0	0.0	0.0	0.0	100.0	100.0	100.0	100.0
Posicion. GL	-0.264	0.165	0.267	22.2	33.3	22.2	22.2	66.7	55.6	55.6	66.7
Reação	1.647	-0.450	-2.218	100.0	100.0	100.0	100.0	0.0	0.0	0.0	0.0
Reflexos	-0.288	0.246	0.239	33.3	33.3	33.3	33.3	66.7	55.6	55.6	66.7

Tabela 1

Tabela indicativa do tipo de relação entre a habilidade dos times e o resultado do jogo. Na seção positiva, a coluna VH apresenta a proporção de temporadas em que o valor médio da habilidade para os casos em que o time mandante venceu foi superior às médias dos casos de empate e derrota, indicando uma relação **positiva** entre a variável  $X_r$  (superioridade do time mandante) e vitória do mandante. Na seção negativa, essa mesma coluna apresenta a proporção de temporadas em que o valor médio da habilidade foi **inferior** às médias de empate e derrota. De maneira análoga, nesse seção essa coluna indica uma relação **negativa** entre a variável  $X_r$ . Na seção positiva, a coluna Tie apresenta a proporção de temporadas em que o valor médio da habilidade foi superior às médias de empate e derrota, indicando uma relação **positiva** entre a variável  $X_r$ . A coluna TIE apresenta a proporção de temporadas em que o valor médio da habilidade foi superior às médias de empate e derrota.

Seasons	Global	Home	Tie	Visitor
2008-2014	47.11	79.65	0.00	36.52
2015	38.95	73.25	0.00	28.45
2016	56.32	85.56	0.00	49.54

Tabela 2

Medidas do modelo. A coluna Global é a acurácia geral para o modelo. A coluna Home, Tie e Visitor apresentam os valores da acurácia para os casos de vitória, empate e derrota do time mandante, respectivamente.

Team	PLC	CCL	CEL	RET
Man Utd	16.50	49.02	14.96	1.77
Man City	16.08	48.46	14.94	1.58
Tottenham	12.90	43.01	15.79	1.98
Chelsea	10.99	37.00	14.25	3.62
Arsenal	9.32	35.04	14.77	3.83
Southampton	8.70	36.16	14.13	3.08
Newcastle	4.51	22.01	12.36	7.78
Crystal Palace	3.85	19.93	11.86	9.30
Stoke City	3.31	17.91	11.30	9.36
Everton	2.79	16.15	10.55	11.90
Leicester	1.85	10.63	8.78	16.67
Liverpool	1.77	12.71	8.93	14.09
Watford	1.69	9.36	7.82	16.87
West Brom	1.52	10.38	8.18	17.85
Bournemouth	1.52	8.61	7.53	19.35
Swansea	1.14	8.78	7.97	19.50
Burnley	0.60	4.68	5.03	29.40
West Ham	0.46	4.25	4.81	31.67
Brighton	0.40	4.31	4.06	32.00
Huddersfield	0.10	1.62	1.98	48.40

Tabela 3

*Detailed table of competition where the values are probability of the simulation. Premier League Champion (PLC), Classify to UEFA Champions League (CCL), Classify to UEFA Europa League (CEL) Relegated team (RE).*

Seasons	Data	<b>Global</b>	Home	Tie	Visitor
2008-2014	Trainning	56.97	80.91	8.77	56.30
2015	Tuning	44.21	71.34	3.74	44.83
2016	Test	51.05	66.84	4.76	59.63

Tabela 4

*Medidas do modelo que inclui o ajuste para similaridade do empate. A coluna Global é a acurácia geral para o modelo. A coluna Home, Tie e Visitor apresentam os valores da acurácia para os casos de vitória, empate e derrota do time mandante, respectivamente.*

Team	PLC	CCL	CEL	RET
Liverpool	39.00	86.00	7.00	0.00
Man City	31.00	78.00	12.00	0.00
Chelsea	13.00	62.00	17.00	0.00
Leicester	6.00	34.00	24.00	0.00
Tottenham	5.00	32.00	24.00	2.00
Man Utd	4.00	26.00	19.00	0.00
Stoke City	1.00	26.00	26.00	0.00
Everton	1.00	14.00	8.00	3.00
West Ham	0.00	9.00	13.00	10.00
West Brom	0.00	11.00	17.00	4.00
Watford	0.00	0.00	3.00	21.00
Swansea	0.00	1.00	5.00	17.00
Southampton	0.00	6.00	6.00	16.00
Newcastle	0.00	3.00	3.00	9.00
Huddersfield	0.00	0.00	0.00	64.00
Crystal Palace	0.00	1.00	5.00	21.00
Burnley	0.00	0.00	0.00	84.00
Brighton	0.00	2.00	1.00	18.00
Bournemouth	0.00	0.00	1.00	27.00
Arsenal	0.00	9.00	9.00	4.00

**Tabela 5**

*Detailed table of competition where the values are probability of the simulation with similarity for tie. Premier League Champion (PLC), Classify to UEFA Champions League (CCL), Classify to UEFA Europa League (CEL) Relegated team (RE).*

	Média			Indicador positivo		Indicador negativo
Skill	$\bar{X}_r^{VH}$	$\bar{X}_r^{TIE}$	$\bar{X}_r^{VV}$	MNP	NP	NN
Aceleração	0.253	0.991	0.047	1.304	88.9	11.1
Altura	0.999	1.000	0.997	-0.030	22.2	77.8
Cabeceio	0.811	0.960	0.733	0.245	55.6	44.4
Carrinho	0.740	1.000	0.484	0.650	66.7	33.3
Ch. de longe	0.201	0.946	0.042	1.000	88.9	11.1
Cobr. falta	0.215	0.991	0.029	1.135	77.8	22.2
Combativ.	0.795	0.910	0.769	0.162	55.6	44.4
Contr. bola	0.067	0.851	0.007	1.204	88.9	11.1
Cruzamento	0.066	0.798	0.009	1.042	88.9	11.1
Div. em pé	0.673	0.998	0.399	0.755	77.8	22.2
Dribles	0.081	0.953	0.005	1.511	88.9	11.1
Duração Do Contrato	0.839	0.978	0.751	0.346	88.9	11.1
Elast. GL	0.959	0.999	0.913	0.121	44.4	55.6
Finalização	0.596	0.998	0.304	0.958	88.9	11.1
Fôlego	0.849	0.993	0.725	0.560	77.8	22.2
Força	0.941	0.998	0.820	0.157	44.4	55.6
Força chute	0.207	0.819	0.083	0.930	88.9	11.1
Idade	0.936	0.999	0.876	0.225	100.0	0.0
Lançamento	0.045	0.973	0.001	1.459	100.0	0.0
Manejo	0.937	0.966	0.944	0.397	100.0	0.0
Marcação	0.683	0.939	0.564	0.453	77.8	22.2
Passe curto	0.046	0.910	0.002	1.296	88.9	11.1
Perna boa	1.000	1.000	0.998	0.000	55.6	44.4
Perna ruim	0.951	0.998	0.859	0.155	55.6	44.4
Peso	0.999	1.000	0.997	-0.030	22.2	77.8
Pique	0.303	0.993	0.068	1.296	100.0	0.0
pos	0.006	0.289	0.000	0.580	88.9	11.1
Posicion. GL	0.933	0.973	0.931	0.408	77.8	22.2
Reação	0.066	0.817	0.007	1.218	100.0	0.0
Reflexos	0.920	0.941	0.945	0.453	88.9	11.1

Tabela 6

Tabela indicativa do tipo de relação entre a habilidade dos times e o resultado do jogo com ênfase no resultado de empate. A interpretação é similar àquela realizada na tabela 1. A transformação 4 permite que o efeito de similaridade das habilidades esperado para os resultados de empate seja captado. Essa transformação é exclusiva para a regressão do parâmetro  $\alpha_{Tie}$ .  $MNP = 2\bar{X}_r^{TIE} - \bar{X}_r^{VH} - \bar{X}_r^{VV}$  é o valor médio para a superioridade da similaridade do empate, variando entre  $-2$  e  $2$ . Valores próximos de  $2$  indicam que a variável possui nível de **similaridade máxima** das habilidades dos times nos empates e **similaridade mínima** para os casos de vitória ou derrota do mandante. A coluna NP indica a proporção de temporadas em que o valor do regressor transformado foi superior nos resultados de empate quando comparada com vitória e derrota do time mandante. O coluna NN representa indica a proporção de temporadas em que o valor do regressor transformado foi inferior nos resultados de empate quando comparada ou com vitória ou com derrota do time mandante. A primeira medida indica a consistência do método para modelar o empate e a segunda inconsistência.

Seasons	Data	Global	Home	Tie	Visitor
2008-2014	Training	53.11	83.05	1.80	50.43
2015	Tuning	47.11	83.44	3.74	37.93
2016	Test	58.16	84.49	3.57	55.05

time	campeao	champions	europa	rebaixa
Man City	18.00	55.00	10.00	6.00
Man Utd	17.00	51.00	15.00	0.00
Arsenal	15.00	35.00	15.00	3.00
Tottenham	14.00	43.00	18.00	0.00
Liverpool	7.00	31.00	9.00	3.00
Newcastle	6.00	21.00	19.00	13.00
Chelsea	6.00	32.00	12.00	6.00
Southampton	5.00	27.00	11.00	8.00
Everton	4.00	17.00	10.00	7.00
Crystal Palace	2.00	15.00	10.00	13.00
Bournemouth	2.00	10.00	5.00	24.00
West Brom	1.00	5.00	4.00	34.00
Swansea	1.00	8.00	10.00	18.00
Stoke City	1.00	12.00	12.00	12.00
Leicester	1.00	9.00	9.00	8.00
West Ham	0.00	14.00	14.00	18.00
Watford	0.00	4.00	6.00	27.00
Huddersfield	0.00	1.00	4.00	36.00
Burnley	0.00	5.00	3.00	33.00
Brighton	0.00	5.00	4.00	31.00