

Anais do



Football
Analytics
Modelling &
Experience

2024

5 de setembro de 2024
Belo Horizonte | MG | Brasil

Adriano C. M. Pereira (UFMG), Gabriel Reis G. Barbosa (UFMG), Gabriel Valadão Meira (UFMG), Hugo Rios Neto (UFMG), Leo M. Sá Freire (UFMG), Lucca Carvalho Augusto (UFMG), Ricardo Furbino (UFMG), Thiago C. Porto (UFMG), Wagner Meira Jr. (UFMG)

Patrocinador:



Organização:



NYU

SCHOOL OF
PROFESSIONAL STUDIES



Apresentação

O *Football Analytics and Modelling Experience* (FAME), em sua terceira edição, integrou em 2024 uma programação ainda mais ampla, como parte do **Future of Football Conference**, iniciativa conjunta da Universidade Federal de Minas Gerais (UFMG) e da New York University (NYU). Esta expansão resultou em dois dias de atividades complementares: no primeiro dia, o **Business of Global Football**, com foco nos aspectos estratégicos, comerciais e regulatórios do futebol internacional; e, no segundo dia, o **FAME'24**, dedicado à ciência de dados, modelagem e aplicações analíticas no futebol.

Este compêndio reúne os artigos apresentados no **FAME'24**, que permaneceu como uma atividade acadêmico-científica independente e essencial dentro do Future of Football. A diversidade dos temas, abordagens e autores reflete o crescente interesse pela interseção entre tecnologia, dados e o esporte mais popular do mundo.

Agradecemos especialmente à **GEMINI Sports Analytics**, patrocinadora do FAME'24, cujo apoio foi decisivo para viabilizar esta edição. Da mesma forma, expressamos nossa gratidão à **ONEFAN**, patrocinadora do Business of Global Football, que tornou possível o fortalecimento do evento como um todo.

Reconhecemos também o importante suporte institucional do Departamento de Ciência da Computação da UFMG (DCC/UFMG) e da equipe do Centro de Atividades Didáticas 3 (CAD 3), onde os encontros presenciais ocorreram.

Por fim, estendemos nosso agradecimento a todas as pessoas que participaram, divulgaram e colaboraram na realização deste evento — em especial, ao time de voluntários da UFMG, que compôs a comissão organizadora do FAME'24 e do Business of Global Football, conforme listado a seguir.

Comissão Organizadora

FAME'24 – Football Analytics and Modelling Experience

Time de voluntários da UFMG – Universidade Federal de Minas Gerais (em ordem alfabética):

- Adriano C. Machado Pereira
- Ana Clara Vasconcelos
- Bernardo Ribeiro
- Bruno Martins
- Gabriel Reis
- Gabriel Valadão Meira
- Hugo Rios-Neto
- João Lucas Lage Gonçalves
- Larissa Dolabella Gomide

- Leo M. Sá Freire
- Lucca Carvalho Augusto
- Luiza Campos
- Ricardo Furbino
- Thiago Costa Porto
- Wagner Meira Jr.

Business of Global Football

Time de voluntários da NYU – New York University (em ordem alfabética):

- Brian Hall
- Marcella Medaglia
- Rafael Gleicher
- Mr. Dean Vince Gennaro

Comitê de Programa

A segunda edição do FAME - Football Analytics and Modelling Experience – teve uma chamada de trabalhos técnicos, que foram apresentados no evento e, posteriormente, convidados para publicação dos trabalhos neste compêndio (Anais / Proceedings) de trabalhos do evento.

Este comitê de programa foi composto pelos seguintes colaboradores:

- Adriano C. Machado Pereira
- Gabriel Reis
- Gabriel Valadão Meira
- Hugo Rios-Neto
- Leo M. Sá Freire
- Lucca Carvalho Augusto
- Ricardo Furbino
- Thiago C. Porto
- Wagner Meira Jr.

Os trabalhos submetidos constituíram de textos ou artigos acadêmicos, aplicações industriais e produtos tecnológicos de *startups* que utilizassem métodos quantitativos para solucionar problemas no esporte, preferencialmente no futebol. Sendo que foram sugeridos os seguintes temas, com possibilidade de outros temas relacionados:

- Análise tática e estratégica
- Análise de desempenho de jogadores e equipes
- Recrutamento e *valuation* de atletas
- Previsão e análise de resultados de jogos e competições

- Prevenção de lesões
- Otimização de treinamento e *performance* física

Cada trabalho enviou um resumo de até 1000 palavras, descrevendo o problema tratado e o objetivo geral, a metodologia, os resultados e as conclusões. E posteriormente convidados os(as) autores(as) a gerar uma versão final, com possibilidade de ser um resumo ou um artigo com mais detalhes do trabalho, a ser publicado em um compêndio do evento, que é este documento aqui apresentado. Esperamos que possam desfrutar da leitura dos trabalhos e aprender mais sobre essa temática do evento.

Cordialmente,

O Comitê de Programa – FAME'24.

Sumário

Apresentamos, a seguir, o conjunto de trabalhos que compõem este compêndio, fruto das contribuições submetidas e selecionadas no âmbito do FAME'24 – *Football Analytics and Modelling Experience*. Os trabalhos aqui publicados passaram por um processo de revisão final após o evento, e incluem artigos completos e resumos expandidos que refletem a diversidade de temas e abordagens presentes na interseção entre ciência de dados e futebol.

Nesta seção, listamos os trabalhos que aceitaram o convite para publicação neste compêndio.

	Título do Trabalho	Pág.
Resumo		
1	Utilização de técnicas de machine learning para identificação e predição das posições táticas dos jogadores Paulo Caixeta de Oliveira	6
2	Como técnicas de Visualização de Dados e Medidas de Centralidade em Grafos podem ajudar a entender o jogo: Copa do Mundo Feminina de 2023 Daniel Barros, Balthazar Paixão, Pedro Henrique González, Glauco Fiorott Amorim	10
3	Análise do Efeito da Idade Relativa sobre o Sucesso de Atletas Jovens de Futebol Lucas Carneiro Calmon, Rodrigo Marques Pau Ferro, Lucas Giusti Tavares, Pedro Henrique González, Glauco Fiorott Amorim	19
4	Comparação entre o equilíbrio competitivo dos jogos em dois formatos de competição Guilherme Machado, Renato Almeida	26

	Título do Trabalho	Pág.
Resumo		
5	Transferência de Jogadores - Estatísticas Agregadas como Ferramenta de Scouting Eduardo Fiuza Costa e Cruz, Gabriel Lima Barros, Gabriel Teixeira Carvalho, Isabella Vignoli Gonçalves	30
6	Escanteios: uma análise dos fatores que influenciam a sua efetividade Bernardo A. Ribeiro, Bruno L. Oliveira, Guilherme D. Rosa, Otavio M. M. Zucheratto	34
7	Análise de Contra-Ataques na Premier League 22/23 Artur F. Souza, Daniel B. Torres, Henrique C. A. Souza, Lucas A. M. Silva	40
8	Simplified Pass Expected Possession Value in PFF Data Implementation and Analysis Diogo Tuler Chaves, Rafael Martins Gomes, Vitor Mendes, João Marcos T. S. Campos	47
9	Análise de Momentos Críticos em Partidas de Futebol a partir de Séries Temporais Gabriel Vitor Lopes Gonçalves, Júlio Cesar de Paula Ferreira, Pablo Correa Costa, Pedro Dias Pires	55
10	Quais os indicadores-chaves de desempenho técnico-táticos discriminam a vitória e a derrota na 1a divisão das cinco principais ligas da CONMEBOL? André L. B. Jacinto, Guilherme P. Sousa, Filipe S. Guaceroni, Varley T. Costa	61
11	Quais Indicadores-Chave de desempenho discriminam equipes que vencem e perdem na Bundesliga (2014 a 2022)? André L. B. Jacinto, Matheus O. Silva, Guilherme P. Sousa, Filipe S. Guaceroni, Varley T. Costa	65
12	Quais indicadores-chaves de desempenho técnico-táticos discriminam a vitória e a derrota na primeira divisão no Big Five Europeu? André L. B. Jacinto, Guilherme P. Sousa, Filipe S. Guaceroni, Varley T. Costa	69
Sessão de Best Papers		74
13	Análise Espacial de Interpretações no Futebol: A Abordagem Voronoi Antônio Caetano Neves Neto, Bernardo Dutra Lemos, João Lucas Simões Moreira, Raphael Aroldo Carreiro Mendes	75
14	Seleção Ótima de Equipes: Uma Abordagem Integrada de IA e Programação Inteira Letícia Maia Távora, Glauco Fiorott Amorim, Pedro Henrique González	80
15	Ferramenta VAR e seu impacto no desempenho da Arbitragem Paulista Bernardo Martins	86

Utilização de técnicas de *machine learning* para identificação e predição das posições táticas dos jogadores

Paulo Caixeta de Oliveira

pcoliveira00@gmail.com

1. Introdução

O futebol é um esporte baseado bastante em tática [Rico-González et al 2022], onde diferentes estratégias são treinadas à exaustão, e os melhores técnicos criam legados e são estudados. Ao mesmo tempo, o futebol é um esporte bastante dinâmico, com uma grande sorte de elementos que influenciam o jogo, tanto dentro de campo (como erros individuais, dribles inesperados, lesões etc) quanto fora de campo (clima, torcida, invasões, problemas técnicos etc), forçando uma mudança e adaptação nestas táticas.

Enquanto isto, temos que uma das melhores aplicação de machine learning é identificação e reconhecimento de padrões [Singh 2023]. Com isto em mente, este artigo se propõe a utilizar técnicas de machine learning supervisionadas e não-supervisionadas para aprender estes padrões e idealmente replicá-los para sugerir uma adaptação aos técnicos nestes momentos de não-linearidade do futebol (uma substituição forçada, uma mudança nas condições de jogo), ou até mesmo para contratação de jogadores com características desejadas similares.

2. Metodologia

Para isto, vamos utilizar uma abordagem mais agnóstica, baseando apenas em dados de eventos (o que realmente foi executado em campo), sem utilizar dados pré estabelecidos ou pessoais dos jogadores, juntamente com dados posicionais. Foram utilizados dados de evento da *Statsbomb* de todos os jogos da *Premiere League* da temporada 15/16 [Statsbomb 2024], das colunas “*type*” e “*position*”.

Na coluna “*type*”, foram removidos os eventos que não são relacionados à toques na bola (como início do jogo, por exemplo). A coluna “*position*”, originalmente em coordenadas (x,y) foi transformada em 96 áreas quadradas de 10m de lado cada (12 horizontais e 8 verticais). Além disto, as posições originais dos jogadores foram remapeadas para as posições clássicas do futebol, conforme tabela abaixo (Tabela 1). Finalmente, as variáveis foram categorizadas antes de serem colocadas para treinar o modelo.

Para a modelagem, os dados iniciais (aproximadamente 1.2M de linhas) foram divididos na proporção 80% para treino e 20% para teste.

Table 1. Variables to be considered on the evaluation of interaction techniques

'Center Attacking Midfield'	'meia-atacante',
'Right Midfield'	'ala-direita',
'Left Back'	'lateral-esquerdo',
'Right Back'	'lateral-direito',
'Left Defensive Midfield'	'volante',
'Left Center Back'	'zagueiro',
'Right Defensive Midfield'	'volante',
'Right Center Back'	'zagueiro',
'Left Center Midfield'	'meia-armador',
'Left Wing'	'ala-esquerda',
'Center Forward'	'atacante',
'Right Center Midfield'	'meia-armador',
'Right Wing'	'ala-direita',
'Left Center Forward'	'atacante',
'Left Midfield'	'ala-esquerda',
'Right Center Forward'	'atacante',
'Center Defensive Midfield'	'volante',
'Right Wing Back'	'lateral-direito',
'Right Attacking Midfield'	'meia-atacante',
'Left Wing Back'	'lateral-esquerdo',
'Center Back'	'zagueiro',
'Left Attacking Midfield'	'meia-atacante',
'Center Midfield'	'meia-armador'

3. Resultados

3.1. Modelo não supervisionado

Para o modelo não supervisionado, foi escolhido o método K-Modes, por ser adequado para clusterização de dados categorizados [Huang ,2009]. A representação das posições dos *clusters* no campo ficaram conforme a Figura 1, e a concentração de cada posição em cada *cluster* ficou conforme a Tabela 2.

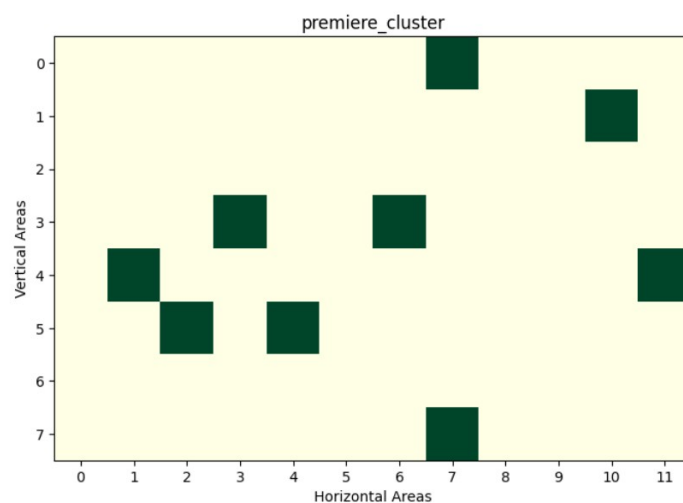


Figura 1. Disposição dos clusters no campo de jogo, sendo o lado esquerdo a defesa, e o direito, o ataque do time

Tabela 2. Concentração das posições nos clusters

		CLUSTER									
		0	1	2	3	4	5	6	7	8	
POSIÇÃO	ala-direita	10%	13%	7%	10%	11%	9%	17%	8%	2%	
	ala-esquerda	10%	21%	4%	11%	10%	8%	17%	8%	3%	
	atacante	11%	26%	2%	15%	9%	7%	14%	44%	2%	
	lateral-direito	9%	3%	14%	8%	10%	13%	8%	5%	2%	
	lateral-esquerdo	9%	12%	6%	10%	9%	11%	6%	1%	3%	
	meia-armador	12%	7%	10%	11%	12%	11%	9%	4%	13%	
	meia-atacante	7%	11%	3%	8%	8%	7%	10%	9%	4%	
	volante	15%	6%	15%	13%	15%	15%	12%	5%	21%	
	zagueiro	16%	2%	40%	13%	16%	18%	7%	16%	49%	

3.2. Modelo supervisionado

Para o modelo supervisionado, foram testados 15 técnicas diferentes, sendo a Light-GBM a que melhor performou. Os resultados completos estão na Tabela 3.

Tabela 3. Resultados do treinamento das diversas técnicas supervisionadas

Modelo	Acurácia	AUC	Recall	Prec.	F1
Light Gradient Boosting Machine	0.36	0.73	0.36	0.29	0.29
Extreme Gradient Boosting	0.36	0.73	0.36	0.29	0.29
Decision Tree Classifier	0.35	0.73	0.35	0.30	0.29
Random Forest Classifier	0.35	0.73	0.35	0.30	0.29
Extra Trees Classifier	0.35	0.73	0.35	0.30	0.29
Gradient Boosting Classifier	0.35	0.00	0.35	0.28	0.28
Ada Boost Classifier	0.29	0.00	0.29	0.22	0.20
Ridge Classifier	0.28	0.00	0.28	0.17	0.16
Logistic Regression	0.28	0.00	0.28	0.18	0.16
Linear Discriminant Analysis	0.28	0.00	0.28	0.18	0.16
K Neighbors Classifier	0.27	0.63	0.27	0.26	0.25
Dummy Classifier	0.22	0.50	0.22	0.05	0.08
SVM - Linear Kernel	0.21	0.00	0.21	0.14	0.12
Quadratic Discriminant Analysis	0.09	0.00	0.09	0.01	0.01
Naive Bayes	0.04	0.60	0.04	0.20	0.03

4. Conclusão

Os resultados das posições dos *clusters* nos mostram que de forma geral o modelo aprendeu uma formação tática parecida com a real, enquanto os *clusters* mostraram um nível de separação. Existem posições melhor separadas, e isto se dá pelas características fundamentais delas (ex: zagueiro e atacante), que demandam mais obediência tática e área de atuação mais bem definida. Outras (como lateral) são bem mais dinâmicas em campo, sendo mais difícil contê-las em *clusters*. Portanto, é possível se basear em modelos de *machine learning* não supervisionados para agrupar jogadores em *clusters* por similaridade de ações, sem colocar nenhuma informação prévia sobre eles.

Para o modelo supervisionado, os resultados (com apenas 36% de acurácia e 29% de precisão) mostram que talvez esta abordagem ainda não seja a melhor para encontrar as posições dos jogadores em campo, pelo menos não com os dados desta forma.

Como trabalhos futuros, podemos refinar os dados de entrada, e isolar a

modelagem por time, por país ou por competição. Podemos também introduzir características físicas ou históricos de jogadores para verificar se melhora a separação ou classificação.

5. References

Rico-González M, Pino-Ortega J, Castellano J, Oliva-Lozano JM, Los Arcos A. Reference values for collective tactical behaviours based on positional data in professional football matches: a systematic review. *Biol Sport*. 2022 Jan;39(1):110-114. doi: 10.5114/biolSport.2021.102921. Epub 2021 Mar 6. PMID: 35173369; PMCID: PMC8805357.

Singh, C. 2023. Machine Learning in Pattern Recognition. *European Journal of Engineering and Technology Research*. 8, 2 (Apr. 2023), 63–68. DOI:<https://doi.org/10.24018/ejeng.2023.8.2.3025>.

<https://statsbomb.com/what-we-do/hub/free-data/>, acessado em 02/08/2024

Huang, Joshua Zhexue. "Clustering categorical data with k-Modes." *Encyclopedia of Data Warehousing and Mining*, Second Edition. IGI Global, 2009. 246-250.

Como técnicas de Visualização de Dados e Medidas de Centralidade em Grafos podem ajudar a entender o jogo: Copa do Mundo Feminina de 2023

Daniel Barros^{b,*}, Balthazar Paixão^a, Pedro Henrique González^b, Glauco Fiorott Amorim^a

^aPPCIC - Centro Federal de Educação Tecnológica Celso Suckow da Fonseca,
CEFET/RJ, , Rio de Janeiro, , RJ, Brazil

^bPESC - Universidade Federal do Rio de Janeiro, UFRJ/RJ, , Rio de
Janeiro, , RJ, Brazil

Abstract

Este artigo explora o uso de técnicas de visualização de dados e medidas de centralidade em grafos para a análise do jogo final da Copa do Mundo Feminina de 2023. Através dessas ferramentas, é possível identificar padrões e dinâmicas de jogo que contribuem para o entendimento tático e estratégico das equipes e obter *insights* detalhados sobre o desempenho das equipes, estratégias táticas e padrões de jogo. Este estudo visa demonstrar como tais ferramentas podem ser usadas para aprofundar a compreensão do jogo, contribuir para o desenvolvimento do esporte e fornecer informações valiosas para treinadores, analistas e fãs.

Keywords: Ferramentas Computacionais; Visualização de Dados; Medidas de Centralidade; Futebol Feminino;

1. Introdução

A análise de desempenho esportivo tem evoluído significativamente com o advento de novas ferramentas computacionais, particularmente na aplicação

*Corresponding author.

Email addresses: danielreboucas.20231@poli.ufrj.br (Daniel Barros),
balthazar.paixao@eic.cefet-rj.br (Balthazar Paixão), pegonzalez@cos.ufrj.br
(Pedro Henrique González), glauco.amorim@cefet-rj.br (Glauco Fiorott Amorim)

de técnicas de visualização de dados e análise de redes complexas [5]. No contexto do futebol, essas técnicas permitem uma percepção mais clara dos padrões de jogo, ajudando a identificar não apenas as jogadoras-chave, mas também as interações e dependências entre diferentes jogadoras e setores do campo [1].

Este artigo investiga como a aplicação dessas técnicas pode oferecer uma visão mais detalhada de jogos da Copa do Mundo Feminina de 2023. Focaremos em como métodos de visualização de dados e medidas de centralidade em grafos podem ser usados para destacar os fatores críticos que influenciaram os resultados das partidas.

O objetivo é demonstrar que o emprego de ferramentas computacionais nesse contexto pode fornecer informações úteis para tomadas de decisão, informações essas que não seriam facilmente obtidas com a análise feita de forma manual. Essas ferramentas podem ajudar na análise, tanto da equipe quanto de uma jogadora individual.

2. Metodologia

Os dados utilizados para analisar a Final da Copa do Mundo de 2023 foram acessados da base da StatsBomb¹. Para a análise foram utilizadas três principais ferramentas: filtros, visualizações e estatísticas.

Inicialmente, o tratamento dos dados é feito através de filtros, desenvolvidos pela manipulação de *dataframes* com a biblioteca *pandas*, em Python, a partir do entendimento de quais colunas são relevantes e agregam informação para o tipo de evento a ser estudado. Foram criados filtros específicos para conduções de bola, chutes, passes, ações defensivas e ações progressivas. Com essas ferramentas, é possível separar os eventos relevantes e realizar uma análise mais direcionada.

Utilizando outras bibliotecas como *matplotlib* e *mplsoccer*, em conjunto com os *dataframes* filtrados por tipo de evento, é possível criar visualizações para analisá-los com mais clareza. Por exemplo, com um *dataframe* filtrado por passes, foram desenvolvidos gráficos como o Mapa de Sonar de Passes, que facilita o entendimento das frequências dos passes de um jogador de acordo com a direção do passe, o *Heatmap* de Passes, que permite assimilar rapidamente as áreas do campo onde mais ocorrem passes por um time, e o

¹<https://statsbomb.com>

Mapa de Ações Progressivas, que também leva em conta conduções de bola que tenham essa característica progressiva.

3. Estudo de Caso: Copa do Mundo Feminina 2023

Para o trabalho proposto, técnicas de visualização de dados e medidas de centralidade em grafos foram utilizadas para apresentar informações mais significativas ao observador. Por isso, esta seção será dividida em duas sub-seções: uma para apresentar o exemplo da aplicação das técnicas de visualização de dados e outra para apresentar o exemplo da aplicação das centralidades em uma rede de passes.

3.1. Visualização de Dados

Como as visualizações podem ser mais sugestivas do que a análise estatística tradicional, o número de visualizações de dados esportivos cresceu rapidamente nas últimas décadas [4]. Neste contexto, novas formas de apresentar dados como diagramas vértice-enlace, linhas de série temporais e *heatmaps* são ferramentas muito úteis para ajudar os profissionais nas tomadas de decisão [3]. Nesta seção, foram utilizadas algumas dessas ferramentas para ajudar a entender como foi a campanha de Espanha e Inglaterra na Copa do Mundo Feminina de 2023.

Na Figura 1, fica claro que a seleção inglesa costuma trocar uma quantidade significativa de seus passes atrás da linha do meio-campo, com as zagueiras sendo bastante participativas em uma posição bem recuada. Por outro lado, a seleção da Espanha concentra a troca de passes mais no meio de campo, o que sugere um estilo de jogo mais ofensivo com relação às suas adversárias.

A Figura 2 apresenta os mapas de chutes que servem para analisar o desempenho ofensivo das seleções finalistas durante a competição. No mapa da Espanha (Figura 2a), é possível comprovar o seu estilo agressivo, sendo a equipe com mais gols e mais chutes por jogo (22.1) entre todas na competição. Por outro lado, na equipe da Inglaterra (Figura 2b), é importante destacar a eficiência nos chutes, já que teve um saldo positivo de 3.18 na diferença entre gols e gols esperados. Além disso, vale comentar que as inglesas finalizam bem menos que suas adversárias na final, tendo uma média de 13.2 chutes por jogo.

A Figura 3 apresenta um Radar Comparativo (levando em consideração o desempenho das atletas durante toda a competição) entre duas das jogadoras

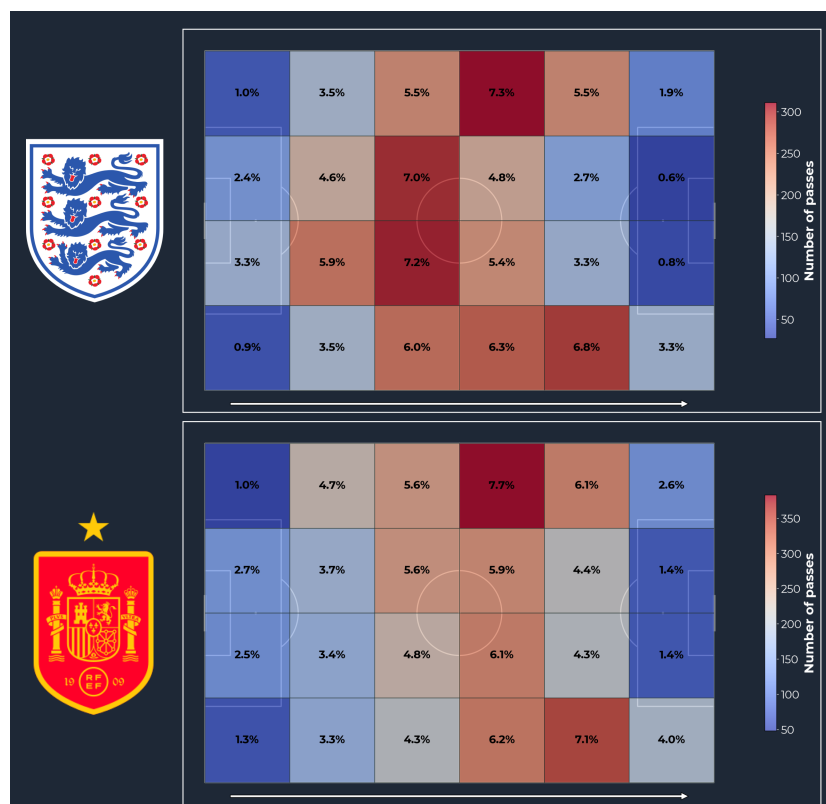
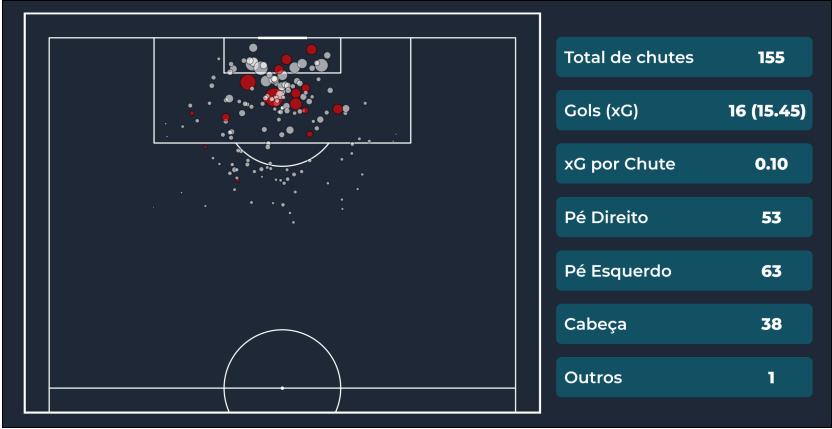


Figura 1: Comparação entre os *Heatmaps* de Passes das Seleções da Espanha e Inglaterra durante toda a competição

mais importantes para as duas finalistas: Aitana Bonmati e Alex Greenwood. O motivo da escolha dessas duas jogadoras vem do resultado das medidas das centralidades que serão destacadas mais adiante. Essa visualização permite uma comparação quase instantânea entre os atributos das jogadoras. Nesse exemplo, ficam bem claras as similaridades e as diferenças nas funções das duas jogadoras. A jogadora espanhola participa bastante da finalização das jogadas, tendo uma quantidade alta de gols, chutes e xG. A inglesa Alex Greenwood, apesar de ser uma defensora, se destaca também em métricas ofensivas, como passes para chute/gol e ações progressivas.

As métricas do Radar são normalizadas de acordo com a quantidade de minutos jogados pela atleta. Além disso, os limites inferiores e superiores são calculados a partir de estatísticas de outras jogadoras. Assim, valores próximos ao limite superior, por exemplo, representam um destaque em uma



(a) Mapa de chutes da Seleção da Espanha



(b) Mapa de chutes da Seleção da Inglaterra

Figura 2: Mapas de Chutes da Seleção da Espanha e da Inglaterra durante toda a competição

métrica.

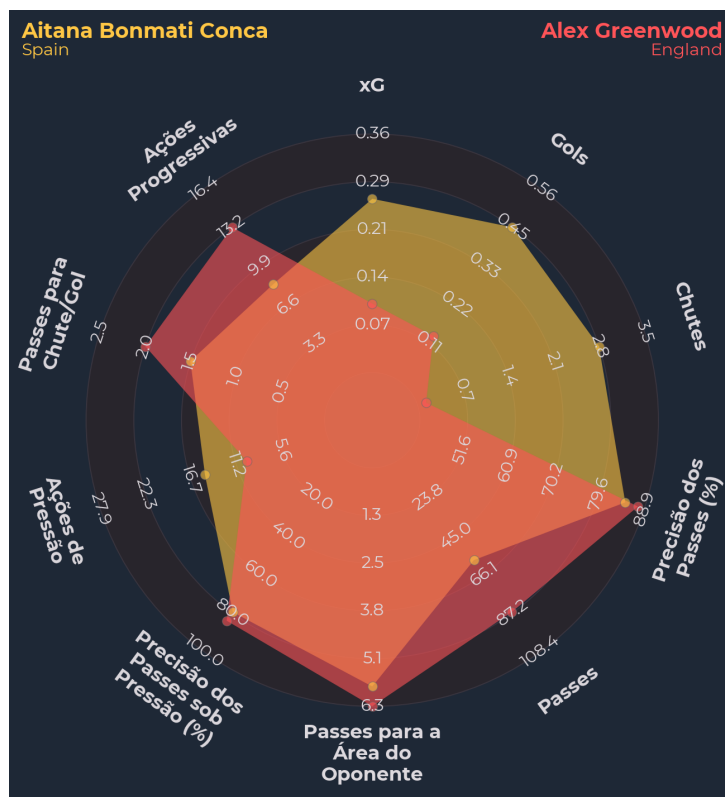


Figura 3: Radar Comparativo entre as jogadoras Aitana Bonmati e Alex Greenwood

3.2. Métricas de Centralidade

Com os dados obtidos sobre a interação entre cada jogador em uma partida é possível criar um grafo denominado de rede de passes. A partir desta rede é possível obter vários indicadores que são muito importantes para a análise da equipe em um jogo. A Figura 4 mostra a rede de passes das duas seleções na final do campeonato.

Um dos indicadores mais importantes em uma rede de passes é o valor da centralidade de um vértice. A medida de centralidade pode ser entendida como a importância daquele vértice em função de uma característica selecionada para aquela rede. Existem várias medidas de centralidade, como: grau de um vértice, *closeness*, *betweenness*, *coeficiente de clustering* e *Page Rank* [2].

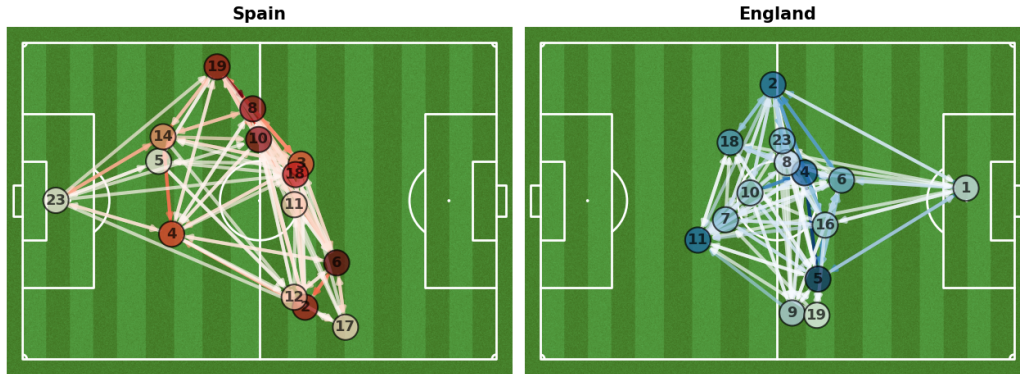


Figura 4: Redes de Passes das Seleções da Espanha e da Inglaterra durante a final. A variação das cores dos nós e das arestas indicam, respectivamente, os valores da centralidade do Pagerank e a quantidade de passes entre os nós. Quanto mais escuro maior o valor.

Neste trabalho, realizou-se um estudo para calcular as medidas de centralidade de cada jogadora que atuou na final do campeonato. Um resumo desses valores encontram-se nas Tabelas 1 e 2. Na seleção da Inglaterra, é possível perceber que as três melhores jogadoras ranqueadas, usando a centralidade de *Betweenness*, foram jogadoras que fazem parte do sistema defensivo. Por isso, é possível dizer que a seleção tem um jogo cadenciado com uma construção que começa com suas zagueiras, que participam efetivamente do jogo. Vale destacar que a jogadora que mais influenciou na partida (maior pagerank) para a seleção da Inglaterra foi Alex Greenwood, que também é a melhor ranqueada na centralidade de *Betweenness*.

Tabela 1: Tabela com os valores de Betweenness e PageRank das três jogadoras mais bem ranqueadas da Seleção da Inglaterra

Nome	Posição	betweenness
Alex Greenwood	Zagueira esquerda	0.168
Lucy Bronze	Ala direita	0.067
Keira Walsh	Meio campo defensiva	0.060
Nome	Posição	pagerank
Alex Greenwood	Zagueira esquerda	0.120
Keira Walsh	Meio campo defensiva	0.105
Lauren Hemp	Atacante	0.101

Já na seleção espanhola, as jogadoras mais influentes (maior *PageRank*) estão no ataque. Entretanto, as laterais também participam do jogo sendo distribuidoras de passes (maior *Betweenness*). É possível verificar que o jogo da Espanha se mostrou mais incisivo e isso foi refletido nos valores de gols esperados, como descrito anteriormente.

Tabela 2: Tabela com os valores de Betweenness e PageRank das três jogadoras mais bem ranqueadas da Seleção da Espanha

Nome	Posição	betweenness
Olga Carmona Garcia	Lateral esquerda	0.092
Aitana Bonmati Conca	Meia	0.049
Ona Batlle Pascual	Lateral direita	0.038
Nome	Posição	pagerank
Aitana Bonmati Conca	Meia	0.108
Jennifer Hermoso Fuentes	Atacante	0.103
María Francesca Caldentey	Ponta esquerda	0.101

As métricas de centralidade em redes de passes fornecem importantes informações para o entendimento do comportamento da equipe no momento em que estava de posse da bola e pode ser usada pelo comissão técnica para várias tomadas de decisão.

4. Conclusão

Este estudo demonstrou a eficácia das ferramentas computacionais na análise de jogos de futebol, especificamente na final da Copa do Mundo Feminino de 2023. Através da aplicação de técnicas avançadas de análise de dados, visualização de dados e técnicas de centralidade em grafos, foi possível obter uma compreensão mais profunda do jogo e identificar fatores-chave que contribuíram para o resultado final. Esses *insights* podem ser utilizados para melhorar o treinamento das equipes, desenvolver novas estratégias e tornar o futebol feminino ainda mais competitivo.

Referências

- [1] Arriaza-Ardiles, E., Martín-González, J., Zuniga, M., Sánchez-Flores, J., de Saa, Y., García-Manso, J., 2018. Applying graphs and complex

- networks to football metric interpretation. *Human Movement Science* 57, 236 – 243. doi:<https://doi.org/10.1016/j.humov.2017.08.022>.
- [2] Barabási, A.L., Pósfai, M., 2016. *Network science*. Cambridge University Press. URL: <http://barabasi.com/networksciencebook/>.
 - [3] Du, M., Yuan, X., 2021. A survey of competitive sports data visualization and visual analysis. *Journal of Visualization* 24, 47–67.
 - [4] Perin, C., Vuillemot, R., Stolper, C.D., Stasko, J.T., Wood, J., Carpendale, S., 2018. State of the art of sports data visualization, in: *Computer Graphics Forum*, Wiley Online Library. pp. 663–686.
 - [5] Uchoa Maia Rodrigues, D.C., Moura, F.A., Cunha, S.A., da S. Torres, R., 2019. Graph visual rhythms in temporal network analyses. *Graphical Models* 103, 101021. doi:<https://doi.org/10.1016/j.gmod.2019.101021>.

Análise do Efeito da Idade Relativa sobre o Sucesso de Atletas Jovens de Futebol

Lucas Calmon^a, Rodrigo Ferro^a, Lucas Giusti Tavares^a, Pedro Henrique González^b, Glauco Fiorott Amorim^a

^aPPCIC - Centro Federal de Educação Tecnológica Celso Suckow da Fonseca,
CEFET/RJ, , Rio de Janeiro, 20271-110, RJ, Brazil

^bPESC - Universidade Federal do Rio de Janeiro, UFRJ/RJ, , Rio de Janeiro, , RJ, Brazil

Keywords: Idade Relativa; Futebol; Atletas da base

1. Resumo

O Efeito da Idade Relativa (EIR) nos esportes é tema de diversos estudos, que conflitam em alguns resultados. Esse artigo investiga a influência do EIR sobre o sucesso de atletas das categorias de base dos clubes de futebol. Foram realizados testes de qui-quadrado para verificar se os quartis de nascimento são distribuídos igualmente entre todos os níveis de Valor de Mercado de jogadores aos 21 anos. Os dados de valor de mercado foram retirados do Transfermarkt¹. Os resultados mostraram que o EIR não é presente na categoria de jogadores mais caros da nossa base de dados e os jogadores jovens nascidos no último quartil estão concentrados na maior categoria de valor de mercado, o que corrobora com a teoria do azarão.

2. Introdução

Nas academias de futebol, é prática comum que as categorias competitivas sejam organizadas com base no ano de nascimento dos atletas. Dessa forma,

*Corresponding author.

Email addresses: lucas.calmon@aluno.cefet-rj.br (Lucas Calmon), (Rodrigo Ferro), lucas.giusti@eic.cefet-rj.br (Lucas Giusti Tavares), pegonzalez@cos.ufrj.br (Pedro Henrique González), glauco.amorim@eic.cefet-rj.br (Glauco Fiorott Amorim)

¹<https://transfermarkt.com/>

cada categoria é composta por jogadores nascidos em um período de um ano, que pode variar conforme o calendário da temporada da liga de futebol do país. Nesse contexto, os atletas nascidos nos primeiros meses desse período anual possuem uma idade cronológica superior em comparação aos outros jogadores da mesma categoria [7, 6]. Esse fenômeno é conhecido como Efeito da Idade Relativa (EIR).

Atletas de futebol com maior idade relativa demonstram níveis maturacionais mais elevados e melhores resultados em testes antropométricos, motores e táticos, o que pode justificar o EIR [3, 8, 4]. Estudos mostram que jogadores nascidos nos primeiros meses do ano têm vantagens em potência aeróbica, agilidade e velocidade [10]. Atletas mais velhos são mais frequentemente selecionados para clubes de elite, possuindo melhores níveis de maturação e habilidades físicas [1]. No entanto, alguns resultados são conflitantes, como o estudo de Skorski [9], que encontrou diferenças significativas apenas para a velocidade na categoria Sub-20 e para a estatura no Sub-16, destacando que o impacto do EIR ainda é tema de debate.

Nesse contexto, algumas questões são levantadas. Um bom exemplo é a teoria do azarão [5], que afirma que atletas nascidos nos últimos semestres precisam manter ou desenvolver habilidades superiores para se manterem no processo de formação.

Sendo assim, apesar da grande atenção da comunidade científica às diferenças antropométricas, motoras e táticas associadas a EIR, as evidências ainda são controversas. Portanto, o objetivo desta pesquisa é analisar a relação do EIR com o valor de mercado de atletas aos 21 anos. Para isso, propomos um método que integra diversas fontes de dados e utiliza uma separação dos jogadores por Categorias de valor de mercado e Quartil de Nascimento para identificar a relação de EIR no valor de mercado em jogadores contratados profissionalmente.

Além desta introdução, o artigo está estruturado em quatro seções. A Seção 3 detalha os métodos utilizados no estudo. A Seção 4 descreve os testes estatísticos utilizados. A Seção 5 descreve os resultados obtidos. Finalmente, a Seção 6 discute as principais conclusões e sugere direções para pesquisas futuras.

3. Metodologia

Esta seção descreve a abordagem metodológica utilizada no estudo. É apresentado o processo de coleta de dados, detalhando as fontes e técnicas

empregadas para colher as informações necessárias. Os dados deste artigo foram originalmente coletados como parte de um projeto que visa estudar a Idade Relativa e as Funções Executivas (FE) de atletas da base de três clubes de elite do Rio de Janeiro, contemplando jogadores entre 13 a 21 anos de idade[11]. O projeto original conta com a aprovação do comitê de ética da Universidade do Estado do Rio de Janeiro com o número de registro CRD42017077640. A base de dados utilizada consiste na junção de: (i) Quartis de Nascimento: Extraídos do *Dataset* com os valores de IR e FE dos atletas acompanhados pelo projeto; (ii) Alvo: *Dataset* com categorias para cada atleta de acordo com seu valor de mercado aos 21 anos de idade. A metodologia para definição das categorias será explicada a seguir.

O valor de mercado dos jogadores foi discretizado e categorizado da seguinte forma para melhor análise: Categoria 0 (C0) corresponde aos jogadores que não têm um perfil cadastrado no Transfermarkt e que não se tornaram profissionais; Categoria 1 (C1) foi discretizada em um limite fixo definido como 0, o que corresponde a todos os jogadores que têm um perfil cadastrado no Transfermarkt, mas que não têm um valor de mercado atribuído, refletindo sua menor visibilidade ou relevância no mercado, especialmente em ligas menores, onde as transferências gratuitas são comuns e o foco está nos salários. Categoria 2 (C2) foi discretizada no 50º percentil dos valores de mercado dos jogadores que são maiores que 0 até €250.000; Categoria 3 (C3) foi discretizada no 100º percentil dos valores de mercado dos jogadores que são maiores que €250.000 até €32.000.000.

A base de dados contém 232 jogadores, todos com 21 anos ou mais no ano de 2024, cujos dados foram coletados do Transfermarkt, um site de estatísticas e transações de futebol amplamente utilizado como fonte de dados para estudos de futebol profissional [2]. Cada jogador teve seu perfil verificado e atualizado perante o Transfermarkt. Aqueles que não possuem perfil no site aos 21 anos foram classificados como categoria zero. A etapa de junção das bases envolve a inclusão dos atributos alvos descrito acima na base de dados de Quartis de Nascimento.

4. Análise Estatística

Testes de qui-quadrado foram realizados para verificar a relação da distribuição dos quartis de nascimento dos jogadores com o valor de mercado dos jogadores aos 21 anos de idade. Os quartis de nascimento dos jogadores foram categorizados em quatro trimestres de idade relativa (sendo: Q1 =

janeiro a março; Q2 = abril a junho; Q3 = julho a setembro; e Q4 = outubro a dezembro). O nível de significância de 0.05 foi adotado neste estudo em consonância com práticas amplamente estabelecidas na literatura científica, que reconhecem esse valor como um equilíbrio adequado entre a sensibilidade do teste estatístico e o controle do risco de cometer um erro Tipo I.

5. Testes e Resultados

A Tabela 1 mostra a distribuição proporcional do valor de mercado dos jogadores em cada quartil de EIR. Na C0, que inclui os jogadores que não se tornaram profissionais, há uma alta proporção no Q1, mas ela diminui significativamente nos quartis seguintes, atingindo 0,033 no Q4. Os Q3 e Q4 aumentam gradualmente da C0 até a C3, onde as categorias de valor de mercado representam os valores mais altos, indicando uma possível tendência de aumento do valor de mercado em jogadores nascidos nos dois últimos trimestres do ano.

Tabela 1: Tabela de Contingência Proporcional por Categoria de Valor de Mercado

Quartil	C0	C1	C2	C3
Q1	0.623	0.633	0.488	0.375
Q2	0.197	0.178	0.244	0.200
Q3	0.148	0.144	0.171	0.225
Q4	0.033	0.044	0.098	0.200

A Tabela 2 mostra os resultados dos testes de qui-quadrado realizados para avaliar a associação entre os quartis da IR e as categorias de valor de mercado. Nos quartis Q1, Q2 e Q3, os valores de p são superiores ao nível de significância de 0,05, indicando que não há evidências de independência entre o EIR e as categorias de valor de mercado nesses quartis. No Q4, o teste de qui-quadrado resultou em um valor de p de 0,008, sugerindo uma associação significativa entre o EIR dos jogadores nascidos no último trimestre do ano e as categorias de valor de mercado.

Tabela 2: Resultados dos Testes Qui-Quadrado por Quartil

Quartil	Qui-Quadrado	p -valor
Q1	3.86	0.277
Q2	0.72	0.868
Q3	0.84	0.840
Q4	11.75	0.008

Os resultados apresentados na Tabela3 mostram a aplicação do teste de qui-quadrado para diferentes categorias de valor de mercado dos jogadores. A análise revela que as categorias C0, C1 e C2 possuem p -valores abaixo de 0,05, indicando uma diferença estatisticamente significativa entre as frequências observadas e esperadas para essas categorias. A categoria C3 apresentou um p -valor de 0,334, revelando que apenas a categoria C3 não sofre com o EIR.

Tabela 3: Resultados dos Testes Qui-Quadrado por Categoria de valor de mercado

Categoria	Qui-Quadrado	p -valor
C0	48.70	1.51×10^{-10}
C1	74.00	5.93×10^{-16}
C2	14.12	0.0027
C3	3.40	0.334

6. Conclusão

Os resultados deste trabalho mostram que o EIR existe no futebol, até certo ponto. Os resultados dos testes estatísticos corroboram com a teoria do azarão: na 2 é observado que a maior quantia de jogadores nascidos no último quartil do ano estão na categoria mais alta de valor de mercado. Além disso, a 3 mostra que a melhor categoria não é influenciada pelo EIR. Esse resultado mostra que jogadores nascidos no último trimestre do ano podem ser favorecidos para chegar no nível mais alto do futebol profissional, porque precisam desenvolver habilidades superiores para continuar no processo de formação competindo com jogadores mais velhos.

Referências

- [1] António J. Figueiredo, Carlos E. Gonçalves, M.J.C.e.S., Malina, R.M., 2009. Characteristics of youth soccer players who drop out, persist or move up. *Journal of Sports Sciences* 27, 883–891. URL: <https://doi.org/10.1080/02640410902946469>, doi:10.1080/02640410902946469, arXiv:<https://doi.org/10.1080/02640410902946469>.
- [2] Bezuglov, E., Morgans, R., Butovskiy, M., Emanov, A., Shagiakhmetova, L., Pirmakhanov, B., Waśkiewicz, Z., Lazarev, A., 2023. The relative age effect is widespread among European adult professional soccer players but does not affect their market value. *PLoS ONE* 18. doi:10.1371/journal.pone.0283390.
- [3] Bozděch, M., Agricola, A., Zháněl, J., 2023. The relative age effect at different age periods in soccer: A meta-analysis. *Perceptual and Motor Skills* 130, 2632–2662. URL: <https://doi.org/10.1177/00315125231210585>, doi:10.1177/00315125231210585, arXiv:<https://doi.org/10.1177/00315125231210585>.
- [4] Gabriele Morganti, Adam L. Kelly, G.A.L.P.M.E.A.G., Ruscello, B., 2023. Relative age effects and the youth-to-senior transition in italian soccer: the underdog hypothesis versus knock-on effects of relative age. *Science and Medicine in Football* 7, 406–412. URL: <https://doi.org/10.1080/24733938.2022.2125170>, doi:10.1080/24733938.2022.2125170, arXiv:<https://doi.org/10.1080/24733938.2022.2125170>.
- [5] Gibbs, B.G., Jarvis, J.A., Dufur, M.J., 2012. The rise of the underdog? the relative age effect reversal among canadian-born nhl hockey players: A reply to nolan and howell. *International Review for the Sociology of Sport* 47, 644–649. URL: <https://doi.org/10.1177/1012690211414343>, doi:10.1177/1012690211414343, arXiv:<https://doi.org/10.1177/1012690211414343>.
- [6] Malina, R.M., Rogol, A.D., Cumming, S.P., Coelho e Silva, M.J., Figueiredo, A.J., 2015. Biological maturation of youth athletes: assessment and implications. *British Journal of Sports Medicine* 49, 852–859. URL: <https://bjsm.bmj>.

com/content/49/13/852, doi:10.1136/bjsports-2015-094623, arXiv:<https://bjsm.bmj.com/content/49/13/852.full.pdf>.

- [7] Musch, J., Grondin, S., 2001. Unequal competition as an impediment to personal development: A review of the relative age effect in sport. *Developmental Review* 21, 147–167. URL: <https://www.sciencedirect.com/science/article/pii/S0273229700905161>, doi:<https://doi.org/10.1006/drev.2000.0516>.
- [8] Romann, M., Javet, M., Cobley, S., Born, D.P., 2021. How relative age effects associate with football players’ market values: Indicators of losing talent and wasting money. *Sports* 9. URL: <https://www.mdpi.com/2075-4663/9/7/99>, doi:10.3390/sports9070099.
- [9] Skorski, S., Skorski, S., Faude, O., Hammes, D., Meyer, T., 2016. The relative age effect in elite german youth soccer: Implications for a successful career. *International Journal of Sports Physiology and Performance* 11, 370 – 376. URL: <https://journals.humankinetics.com/view/journals/ijsp/11/3/article-p370.xml>, doi:10.1123/ijsp.2015-0071.
- [10] Susana Maria Gil, Aduna Badiola, I.B.L.J.Z.L.L.G.J.S.C.J.A.L., Granados, C., 2014. Relationship between the relative age effect and anthropometry, maturity and performance in young soccer players. *Journal of Sports Sciences* 32, 479–486. URL: <https://doi.org/10.1080/02640414.2013.832355>, doi:10.1080/02640414.2013.832355, arXiv:<https://doi.org/10.1080/02640414.2013.832355>.
- [11] Tavares, L.G., 2019. Funções executivas e idade relativa em atletas de futebol de elite de 13 a 17 anos. Master’s thesis. Programa de Pós-Graduação em Ciências do Exercício e do Esporte. URL: <http://www.btd.uerj.br/handle/1/8266>. centro de Educação e Humanidades::Instituto de Educação Física e Desporto.

Comparação entre o equilíbrio competitivo dos jogos em dois formatos de competição

Guilherme Machado^{1,2}, Renato Almeida^{1,2}

¹ Departamento Científico e Departamento de Integração e Desenvolvimento de Atletas, Federação Paulista de Futebol (FPF), São Paulo, Brasil.

² Núcleo de Pesquisa e Estudos em Futebol (NUPEF), Universidade Federal de Viçosa (UFV), Brasil.

1. Introdução

O equilíbrio competitivo é um conceito central no estudo das competições esportivas, referindo-se ao grau em que as equipes participantes têm chances similares de vencer uma partida ou campeonato. Em um ambiente onde o equilíbrio competitivo é alto, as partidas tendem a ser mais imprevisíveis e disputadas, o que pode aumentar o engajamento dos torcedores e o desenvolvimento dos jogadores (Basini et al., 2023). No contexto do futebol de base, particularmente na categoria sub-20, o equilíbrio competitivo adquire uma importância adicional. Esta fase de transição para o futebol profissional exige que os jogadores sejam expostos a desafios que simulem a intensidade e a competitividade que enfrentarão em níveis mais altos. Estudos sugerem que quanto mais competitivos forem os jogos na base, melhor será a preparação dos atletas para o ambiente profissional (Plumley et al., 2023; Serrano & Espitia-Escuer, 2024; Machado et al., 2020).

Pesquisas anteriores têm investigado o impacto dos diferentes formatos de competição sobre o equilíbrio competitivo. Por exemplo, a divisão de equipes em diferentes níveis de habilidade, como ocorre em competições com múltiplas divisões, tem sido associada a um maior equilíbrio competitivo, pois equipes de habilidades semelhantes são agrupadas, resultando em partidas mais equilibradas (Peeters & Szymanski, 2023). Por outro lado, competições organizadas em formato de divisão única, onde equipes de diferentes níveis de habilidade competem entre si, podem apresentar maior disparidade nos resultados, o que reduz o equilíbrio competitivo.

Compreender como o formato da competição afeta o equilíbrio competitivo é crucial para as federações e confederações que organizam torneios de futebol. Ajustes nesses formatos podem criar um ambiente mais propício ao desenvolvimento dos atletas, garantindo que todos os participantes tenham a oportunidade de competir em condições mais equitativas e formativas (Teoldo, Machado & Cardoso, 2023; Machado et al., 2024). Esse entendimento pode orientar políticas esportivas que visem não apenas o sucesso das competições, mas também a formação de atletas capazes de se destacarem no cenário profissional.

2. Objetivo

O objetivo deste trabalho foi comparar o equilíbrio competitivo dos jogos em dois formatos de competição sub-20.

3. Métodos

Este estudo analisou a competitividade dos jogos em dois formatos distintos de competição de um campeonato estadual Sub-20, utilizando dados de cinco temporadas: 2016, 2017, 2019, 2022 e 2023. Foram analisadas duas condições de competição: Divisão Única e Duas Divisões.

Na condição de Divisão Única, foram analisados os anos de 2022 e 2023. Em 2022, a competição contou com 518 jogos e 72 equipes participantes, enquanto em 2023, foram 548 jogos com 78 equipes. O formato dessa competição foi dividido em várias fases: na primeira fase, as equipes foram divididas em grupos de 6, com cada grupo disputando 10 jogos em sistema de ida e volta. Na segunda fase, as equipes classificadas foram redistribuídas em grupos de 4, disputando 6 jogos em sistema de ida e volta. Na terceira fase, novamente, as equipes foram reorganizadas em grupos de 4, realizando mais 6 jogos em sistema de ida e volta. As melhores equipes avançaram para as quartas de final, seguidas pelas semifinais e a final, todas disputadas em jogos de ida e volta.

Na condição de Duas Divisões, foram analisados os anos de 2016, 2017 e 2019, tanto para a Primeira Divisão quanto para a Segunda Divisão. Na Primeira Divisão, em 2016 e 2017, a competição contou com 500 e 511 jogos, respectivamente, com 46 equipes participantes em ambos os anos. Em 2019, houve uma redução para 413 jogos e 40 equipes participantes. O formato da competição foi estruturado da seguinte maneira: na primeira fase, as equipes foram divididas em grupos de 11, com cada grupo disputando 20 jogos em sistema de ida e volta. As equipes classificadas avançaram diretamente para as oitavas de final, seguidas pelas quartas de final, semifinais e a final, todas disputadas em jogos de ida e volta.

Na Segunda Divisão, em 2016, a competição contou com 125 jogos e 25 equipes participantes. Em 2017, foram 149 jogos com 27 equipes, e em 2019, 149 jogos com 30 equipes. O formato dessa divisão foi semelhante à Primeira Divisão, mas com um número menor de jogos. Na primeira fase, as equipes foram divididas em grupos de 5, com cada grupo disputando 8 jogos em sistema de ida e volta. As equipes classificadas avançaram diretamente para as oitavas de final, seguidas pelas quartas de final, semifinais e a final, todas disputadas em jogos de ida e volta.

A competitividade dos jogos foi avaliada com base na diferença de gols no placar final dos jogos. Eles foram classificadas em três categorias: i) competitivo (com diferença de até 2 gols); ii) pouco competitivo (com diferença de 3 a 4 gols); e iii) muito pouco competitivo (com diferença de 5 ou mais gols). Os dados foram coletados de maneira online por meio das súmulas eletrônicas de acesso público disponibilizadas no site oficial da Federação Paulista de Futebol (www.futebolpaulista.com.br). Essas súmulas contêm informações detalhadas sobre os resultados das partidas, incluindo o placar final, que foi utilizado para a categorização da competitividade dos jogos. A análise estatística foi conduzida utilizando o teste do qui-quadrado para comparar a distribuição das categorias de competitividade entre os dois formatos de competição (Divisão Única e Duas Divisões).

4. Resultados e Discussão

A análise dos resultados apresentou diferenças significativas na distribuição da competitividade dos jogos entre os formatos de competição avaliados ($p < 0,001$) para todos os formatos (ver Tabela 1). Os resultados mostraram que o formato de Divisão Única apresentou o menor percentual de jogos competitivos, com 73,1% das partidas tendo uma diferença de até 2 gols, e o maior percentual de jogos muito pouco competitivos, com 9,6% das partidas tendo uma diferença de 5 ou mais gols. Em contraste, o formato de Duas Divisões, especialmente na

Segunda Divisão, apresentou o maior equilíbrio competitivo, com 79,7% das partidas sendo competitivas e apenas 3,0% classificadas como muito pouco competitivas.

Tabela 1. Frequência do placar final para os diferentes formatos competitivos.

Formato Competitivo	Placar final			p
	<i>até 2 gols (Competitivo)</i>	<i>3 ou 4 gols (Pouco Competitivo)</i>	<i>5 ou mais gols (Muito Pouco competitivo)</i>	
	% (valor absoluto)	% (valor absoluto)	% (valor absoluto)	
Divisão Única	73,1% (779)	17,4% (185)	9,6% (102)	< 0,001
Duas Divisões - Primeira Divisão	75,8% (1080)	18,3% (261)	5,83% (83)	< 0,001
Duas Divisões - Segunda Divisão	79,7% (337)	17,3% (73)	3,0% (13)	< 0,001

Os resultados sugerem que o formato de Duas Divisões, tanto na Primeira quanto na Segunda Divisão, proporciona um equilíbrio competitivo maior em comparação ao formato de Divisão Única. A maior proporção de jogos competitivos na Segunda Divisão do formato de Duas Divisões (79,7%) indica que, ao agrupar equipes de níveis semelhantes, o formato de Duas Divisões reduz a disparidade entre as equipes, resultando em partidas mais equilibradas. Esses achados corroboram estudos anteriores que apontam para a eficácia de formatos de competição que segmentam as equipes com base em seu desempenho ou nível de habilidade, proporcionando um ambiente mais competitivo e, portanto, mais propício para o desenvolvimento dos jogadores (Peeters & Szymanski, 2023; Serrano & Espitia-Escuer, 2024; Teoldo et al., 2023). Além disso, a menor proporção de jogos muito pouco competitivos nas Duas Divisões, especialmente na Primeira e Segunda Divisão, sugere que esse formato pode minimizar a ocorrência de partidas com grande disparidade de resultados, o que é benéfico para manter o engajamento tanto dos jogadores (Machado et al., 2024).

A maior frequência de jogos muito pouco competitivos no formato de Divisão Única (9,6%) pode ser atribuída à maior heterogeneidade entre as equipes participantes, uma vez que este formato não segmenta as equipes por nível de habilidade ou desempenho anterior. Isso pode resultar em uma frequência maior de encontros desiguais, onde equipes mais fortes enfrentam adversários significativamente mais fracos, levando a resultados desequilibrados (Basini et al., 2023). Diante desses resultados, sugere-se a estratégia de considerar a implementação de formatos de competição que segmentem as equipes em diferentes divisões ou categorias, de modo a promover um maior equilíbrio competitivo. Isso não só pode melhorar a qualidade das partidas, mas também oferecer um ambiente mais adequado para o desenvolvimento dos atletas (Machado et al., 2024).

5. Conclusão

Os resultados deste estudo indicam que o formato de Duas Divisões, tanto na Primeira quanto na Segunda Divisão, promove um maior equilíbrio competitivo em comparação ao formato de Divisão Única na categoria sub-20. A segmentação das equipes com base em níveis de habilidade ou desempenho parece reduzir a disparidade entre os jogos, resultando em partidas mais equilibradas e competitivas. Esses achados sugerem que federações e organizadores de competições devem considerar a implementação de formatos que segmentem as equipes, a fim de proporcionar um ambiente mais propício ao desenvolvimento dos atletas.

6. Referências

- Basini, F., Tsouli, V., Ntzoufras, I., & Friel, N. (2023). Assessing competitive balance in the English Premier League for over forty seasons using a stochastic block model. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 186(3), 555-575.
- Machado, G. F., Teoldo, I., Cardoso, F. S. L., & Serra-Olivares, J. (2020). Development of Tactical Decision-making Skills in Youth Soccer Players: Macro- and Microstructure of Soccer Developmental Activities as a Discriminant of Different Skill Levels. *Journal of Sports Sciences*, 38(5), 515-523.
- Machado, G. F., Teoldo, I., & Costa, I. T. (2024). Mapping talent pathways: A comparative study of developmental activities and practice structure in Brazilian and Spanish U-18 elite youth male soccer players. *Journal of Sports Sciences*, 42(2), 233-245.
- Peeters, T., & Szymanski, S. (2023). Competitive balance and demand for European men's football. *Journal of Global Sport Management*.
- Plumley, D., Ramchandani, G., & Wilson, R. (2023). Assessing competitive balance in the English Premier League using network modelling. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 186(3), 530-553.
- Serrano, R., & Espitia-Escuer, M. (2024). The competitive balance of UEFA and CONMEBOL football leagues: A comparative analysis. *Managerial and Decision Economics*.
- Teoldo, I., Machado, G. F., & Cardoso, F. S. L. (2023). *Caderno Metodológico para a Formação de Jogadores Mais Inteligentes e Criativos para o Jogo: Case UFV Soccer Academy*. Appris.

Transferência de Jogadores - Estatísticas Agregadas como Ferramenta de Scouting

Eduardo Fiuza Costa e Cruz¹, Gabriel Lima Barros², Gabriel Teixeira Carvalho³,
Isabella Vignoli Gonçalves⁴

¹Departamento de Ciência da Computação - Universidade Federal de Minas Gerais
Belo Horizonte – MG – Brasil

¹eduardofiuza6@ufmg.br, ²gabrielimab@ufmg.br,
³gabrieltc@ufmg.br, ⁴isabellavignoli@ufmg.br

Resumo. Este artigo descreve o processo realizado para a construção de um modelo de avaliação de contratação de jogadores dentro de um contexto de estatísticas agregadas. O projeto aprofunda nas etapas de Data Understanding, Data Preparation, Modelo e nos Resultados Obtidos. A ideia principal é desenvolver uma ferramenta de Scouting para que os times de futebol consigam buscar por jogadores na janela de transferência que se assemelham com a demanda do clube. Para isso, são empregadas técnicas de aprendizado de máquina e mineração de dados, como Matriz de Correlação, Variance Inflation Factor, Random Forest, entre outros.

1. Introdução

O impacto mais significativo no desempenho de um clube geralmente ocorre após uma janela de transferências, entretanto, é comum que os jogadores contratados não alcancem as expectativas estabelecidas. Refinar o processo de contratação durante essas janelas, não apenas melhoraria o desempenho dos clubes, mas também fortaleceria sua saúde financeira

Assim, a proposta do projeto foi desenvolver um modelo de avaliação de contratações de jogadores, juntamente com a seleção de atributos e métricas relevantes que influenciam na agregação ou desagregação de valor de um jogador dentro do contexto de um novo clube. Esta iniciativa tem como propósito auxiliar o departamento de scouting dos clubes a direcionar seus esforços para jogadores que melhor se encaixam dentro do perfil desejado.

2. Data Understanding

Os conjuntos de dados utilizados foram obtidos do site FBref [FBref 2024]. Estas bases contam com estatísticas agregadas das últimas 3 temporadas da Premier League. Optamos por utilizar estatísticas agregadas para estudar as limitações e vantagens desse tipo de dado.

3. Data Preparation

Realizamos uma seleção inicial das features mais adequadas para a construção modelo de avaliação. Assim, escolhemos 4 métricas de cada tipo dos dados da Premier League: estatísticas de passe, estatísticas de finalização, estatísticas de posse de bola e estatísticas de defesa. As features selecionadas foram: Cmp, xA, xAG e KP; npxG, Sh/90, SoT/90 e G/SoT; Touches, Succ%, Carries e Rec; e Tkl%, TklW, Clr, Int, respectivamente.

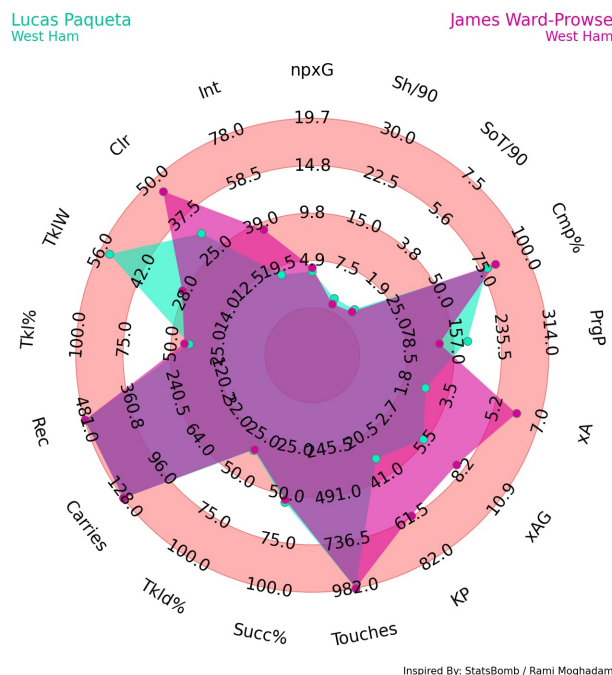


Figure 1. Comparação entre Lucas Paquetá e James Ward-Prowse

4. Modelo

Para tornar o modelo menos instável e mais interpretável, permitindo isolar melhor o efeito individual de cada coluna, filtramos os atributos mais importantes do conjunto de dados utilizando diferentes estratégias.

Calculamos uma matriz de correlação entre os atributos considerados, permitindo identificar aqueles que eram mais correlacionados. Em seguida, calculamos o Fator de Inflação de Variância (VIF) para cada coluna do dataframe. O VIF é uma métrica utilizada para identificar a presença de multicolinearidade entre as colunas independentes em um modelo de regressão, o que permite identificar colunas não relevantes para o modelo. Por fim, calculamos a importância de cada feature simulando um problema de classificação binária, através da geração de um vetor aleatório com valores 0 e 1 para cada linha do conjunto de dados e então classificando-o com um modelo Random Forest. Ao final obtemos uma medida de quão útil cada característica foi para o nosso modelo.

De posse das colunas altamente correlacionadas, com alto VIF ou pouco importantes, pudemos remover com assertividade as colunas menos relevantes para o modelo, restando apenas as que seriam utilizadas, as quais foram: Cmp%, xA, npxG, Sh/90, G/SoT, Tkl%, Succ%, Tkld%, como pode ser visto na figura 2.

A partir desses atributos, construímos um modelo que calcula a similaridade entre um dado jogador e todos os outros jogadores do conjunto de dados, combinando as métricas de distância cosseno e distância euclidiana. Ao combinar essas duas métricas, o modelo busca capturar tanto o estilo quanto o impacto dos jogadores, através da direção e da magnitude de suas características, visando obter uma medida de similaridade mais robusta e representativa. Por fim, a similaridade é filtrada por um limiar mínimo e os jogadores mais similares são ordenados e retornados.

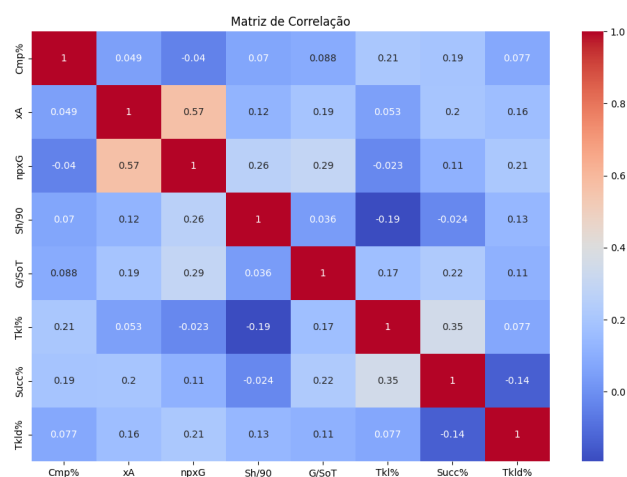


Figure 2. Matriz de Correlação das colunas utilizadas no modelo

5. Resultados Obtidos

Para iniciar a apresentação dos resultados obtidos com o projeto, faremos um estudo de caso de um jogador conhecido que atua na Premier League. O jogador em questão é o meio-campista Bruno Fernandes.

5.1. Estudo de Caso: Bruno Fernandes

As principais características do Bruno Fernandes, como jogador, que devem ser consideradas com o intuito de encontrar jogadores parecidos com ele são: criatividade, deve refletir em estatísticas como assistências, passes e carregadas, chute de fora da área, está representada pelos gols esperados e chutes ao gol, importância defensiva, deve haver uma valorização das estatísticas de cunho defensivo e é uma figura central do time, espera-se que tenha um número alto de toques na bola por partida.

Quando usamos o modelo para retornar os dez jogadores mais parecidos com o Bruno Fernandes obtivemos o seguinte resultado:

1. Bruno Fernandes - MC - Manchester United
2. Martin Odegaard - MC - Arsenal
3. Dwight McNeil - PD - Everton
4. Dejan Kulusevski - PD - Tottenham
5. Morgan Gibbs-White - MC - Nottingham Forest
6. Pascal Gross - MC - Brighton
7. Bukayo Saka - PD - Arsenal
8. Andreas Pereira - MC - Fulham
9. Pablo Sarabia - PD - Wolves
10. Dominik Szoboszlai - MC - Liverpool

5.2. Validação do Modelo

Com o objetivo de validar a aplicabilidade do modelo foram desenvolvidas duas estratégias, verificar a simetria entre os jogadores gerados pelo modelo como semelhantes e testar o modelo para as três últimas temporadas da Premier League, usando apenas jogadores que estiveram no mesmo time em todas as temporadas e contando quantas vezes ele aparece como semelhante a si mesmo.

A estratégia de simetria foi basicamente gerar, para cada jogador, a sua lista de dez jogadores mais semelhantes e contar o número de jogadores dessa lista que contam com o jogador original em suas listas de semelhantes.

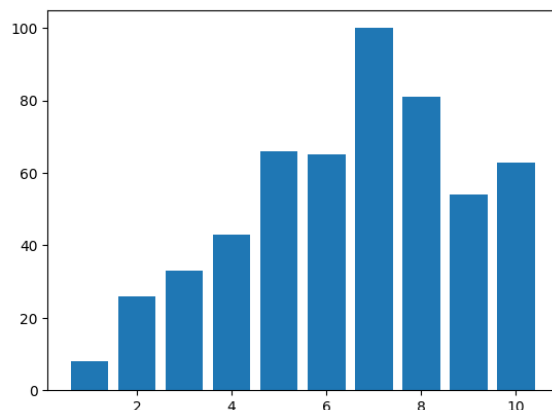


Figure 3. Número de jogadores simétricos

Os resultados obtidos pela primeira estratégia foram satisfatórios, pois para a maioria do jogador era possível observar simetria entre os seus semelhantes (figura 1).

A segunda estratégia requer a integração dos dados das temporadas 2021-2022 e 2022-2023 ao dataframe utilizado pelo modelo. Para essa etapa foram usados os 153 jogadores que jogaram todas as três temporadas no mesmo time, eliminando uma possível inconsistência relacionada à troca de times. Essa validação teve como objetivo definir os pesos para as distâncias cosseno e euclidiana e também avaliar a eficácia do modelo.

A partir do estudo de caso e dos resultados obtidos nas duas estratégias de validação foi possível concluir que a simplicidade do formato de dados escolhido e a complexidade da tarefa de reconhecer jogadores semelhantes tornaram o modelo desenvolvido claramente limitado.

References

FBref (2024). 2024-2025 premier league stats. Accessed: 2024-08-06.

Escanteios: uma análise dos fatores que influenciam a sua efetividade

Bernardo A. Ribeiro¹, Bruno L. Oliveira², Guilherme D. Rosa³, Otávio M. M. Zucheratto⁴

¹Departamento de Ciência da Computação – Universidade Federal de Minas Gerais (UFMG)
Caixa Postal 702 – 31.270-901 – Belo Horizonte – MG – Brazil

{bernardoaribeiro, brunolima, gdrumondrosa, ommz21}@ufmg.br

Abstract. *This study investigates the factors influencing the effectiveness of corner kicks in football, focusing on the 22/23 Premier League season. By analyzing tracking data and event data from approximately 1,800 corner kicks, the research identifies key elements contributing to successful outcomes. Only 73 corner kicks resulted in goals, leading to the development of an Expected Keep Possession (xKP) metric, which predicts the likelihood of maintaining possession after a corner. Significant factors include player positioning, ball speed, and subsequent event types. The study concludes that strategic adjustments can enhance a team's ability to retain possession and create scoring opportunities.*

Resumo. *Este estudo investiga os fatores que influenciam a efetividade dos escanteios no futebol, com foco na temporada 22/23 da Premier League. Analisando os dados de tracking e de eventos resultando em aproximadamente 1.800 escanteios, a pesquisa identifica os principais elementos que contribuem para resultados bem-sucedidos. Apenas 73 escanteios resultaram em gols, levando ao desenvolvimento da métrica de Manutenção Esperada de Posse (xKP), que prevê a probabilidade de manter a posse após o escanteio. Fatores significativos incluem o posicionamento dos jogadores, a velocidade da bola e os tipos de eventos subsequentes. Conclui-se que ajustes estratégicos podem melhorar a capacidade de uma equipe de manter a posse e criar oportunidades de gol.*

1. Introdução

Os escanteios são jogadas que ocorrem em quantidades consideráveis na maioria das partidas de futebol. Times que utilizam bem as oportunidades que as jogadas de escanteios proporcionam, podem fazer dessas jogadas um fator determinante para suas vitórias. Dessa maneira, o escanteio se torna um evento muito interessante de se analisar e destrinchar os fatores que compõem um escanteio bem executado.

2. Data Understanding e Data Preparation

Foi iniciado o trabalho explorando os dados de tracking da temporada 22/23 da Premier League, recebidos como amostra inicial pela PFF. Para esse entendimento realizou-se uma manipulação dos dados, com o intuito de compreender melhor os mesmos.

Então, foram criados filtros para capturar apenas os lances de escanteios das partidas, resultando em cerca de 1.817 escanteios válidos que geraram 73 gols após 5 eventos depois da batida de escanteio.

3. Modelo

Inicialmente o grupo tinha optado por trabalhar com a efetividade dos escanteios na geração de gol, ideia que foi descartada à medida que o trabalho avançou, devido à escassez de gols e a outras variáveis que não foram possíveis de estimar. Dessa forma, as features foram pensadas com o intuito de analisar a manutenção de posse da equipe após a cobrança do escanteio, ou seja, deseja-se estimar quais configurações/estilos de cobrança são os ideais para que a posse seja mantida após o evento em questão.

3.1. Features Seleccionadas

Diferença entre o número de jogadores de cada equipe na área

Diferença entre a média de altura dos jogadores de cada equipe na área

Tipo da batida (aberta ou fechada)

Porção da área em que a bola foi cruzada

Velocidade da bola cruzada

Altura em que a bola chega no evento subsequente ao cruzamento

Distância entre o jogador do evento subsequente ao cruzamento e o marcador mais próximo

Tipo do evento subsequente ao cruzamento (chute, passe...)

3.2. Divisão de Zonas

Analisando o heat-map foram criadas as seguintes zonas:

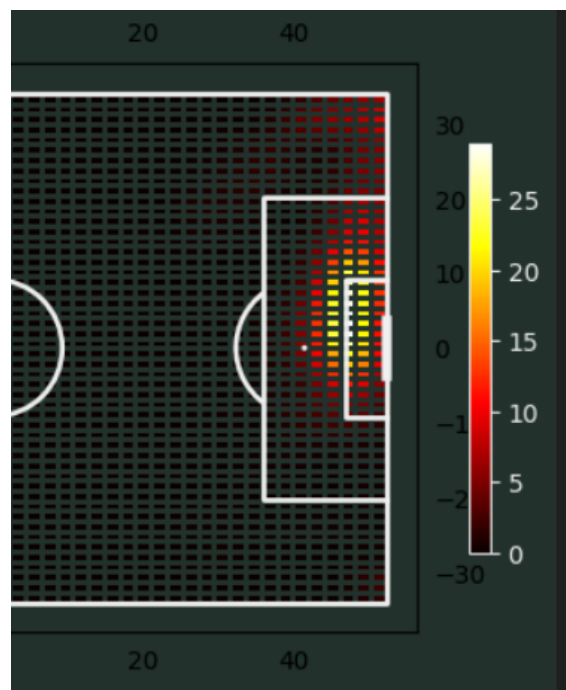


Figura 1. HeatMap de escanteios.



Figura 2. Divisão das zonas.

4. Resultados

4.1. Boxplot das alturas

O boxplot das alturas mostra que a média da altura relativa dos jogadores em lances que resultam em gol é de 0,6, enquanto que a média para os lances que não resultam em gol é de -0,32. Esses resultados estão de acordo com o esperado, sugerindo que, quanto maior a altura dos jogadores do time atacante em comparação aos defensores, maior é a probabilidade de marcar um gol.

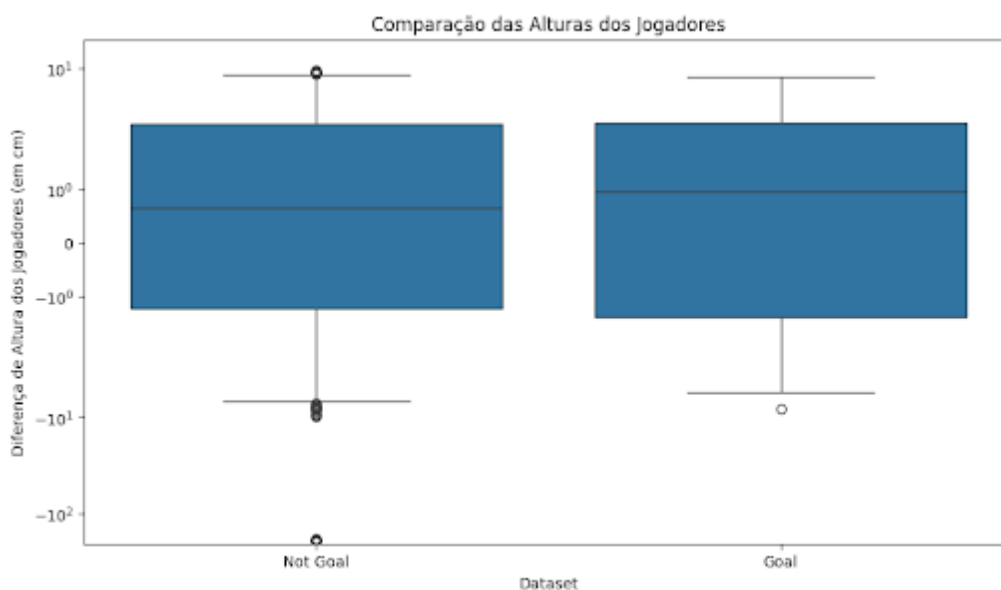


Figura 3. Boxplot das alturas.

4.2. Distribuição dos eventos após cobrança

O gráfico ilustra a frequência dos eventos subsequentes a uma cobrança de escanteio, considerando apenas aqueles que mantêm a posse de bola para o time atacante. O evento mais frequente é o passe (PA), com mais de 250 ocorrências, seguido por chutes (SH). Eventos como rebotes (RE) e condução de bola (BC) ocorrem com pouca frequência, sendo os menos observados. Isso indica que, após um escanteio, o time atacante tende a realizar passes e finalizar mais frequentemente antes de tentar um cruzamento ou conduzir a bola.

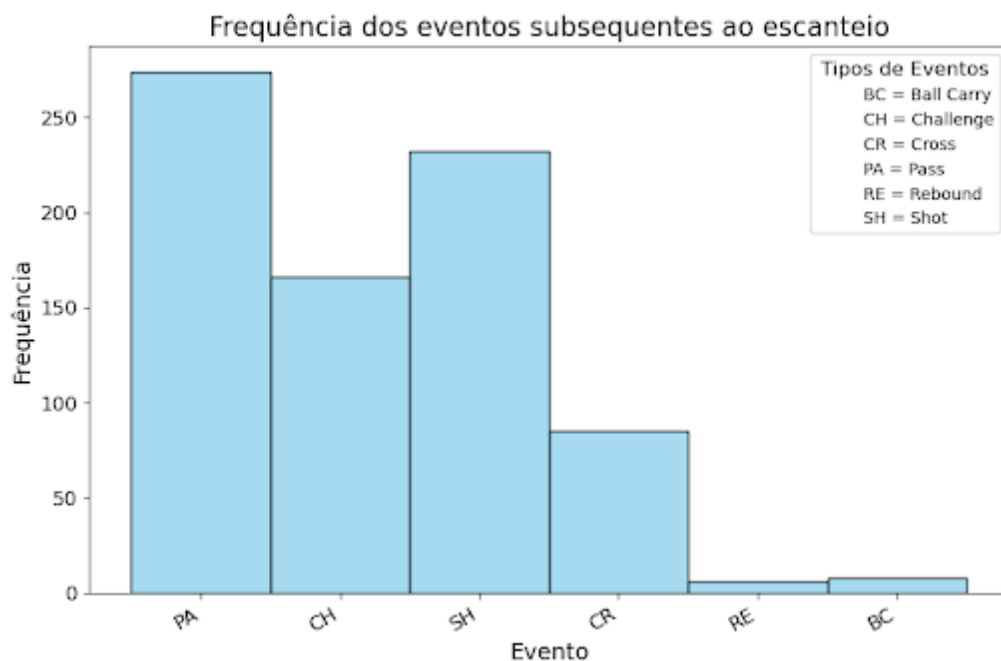


Figura 4. Distribuição dos eventos subsequentes ao escanteio.

4.3. Distribuição das zonas da cobrança

O gráfico mostra a distribuição das cobranças de escanteio por zona, destacando as áreas mais visadas. A Grande Área é a zona mais frequente, com cerca de 450 ocorrências, seguida pela Entrada da Pequena Área. Esses dados sugerem que quando a bola é lançada na área, as zonas mais procuradas incluem o centro da área (zona de maior congestionamento, porém com maior chances de finalização) e a porção mais distante do gol dentro da área (zona mais vazia, com maior chance de manter a posse). Escanteios curtos também são comuns, como evidenciado pela frequência na zona Fora da Área.

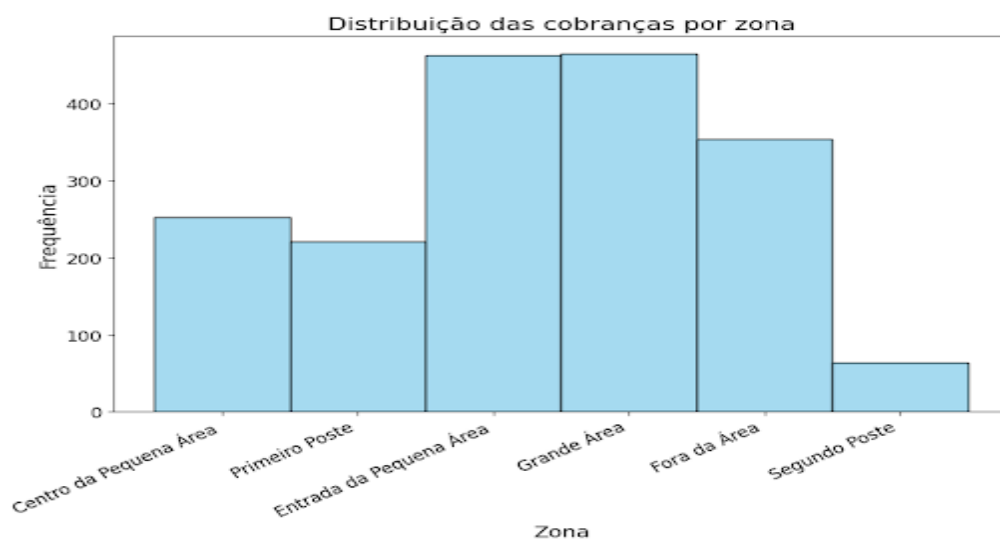


Figura 5. Distribuição das cobranças por zona.

4.4. xKP

Foi desenvolvido o Expected Keep Possession (xKP), uma métrica para avaliar a probabilidade de uma equipe manter a posse de bola após a cobrança de um escanteio. Utilizando várias features relacionadas ao contexto da cobrança, o xKP estima as chances de o time que está atacando continuar com a posse de bola no próximo evento. Para implementar essa métrica, foi criado um classificador que prevê se a posse de bola será mantida ou não pelo time atacante. O desempenho desse classificador é visualizado por meio de uma matriz de confusão, que permite avaliar sua precisão na tarefa de distinguir entre cenários de posse.

A matriz de confusão revela um desempenho sólido do classificador. Entre os 1.046 casos em que a posse não foi mantida (classe 0), o modelo previu todos corretamente, resultando em 0 falsos positivos. Dos 771 casos de posse mantida (classe 1), 752 foram identificados corretamente, com 19 erros, gerando falsos negativos. O classificador mostra alta precisão em prever a perda de posse, e uma boa, mas não perfeita, capacidade de prever sua manutenção. A ausência de falsos positivos destaca a confiabilidade do modelo em não prever a posse quando ela não ocorre.

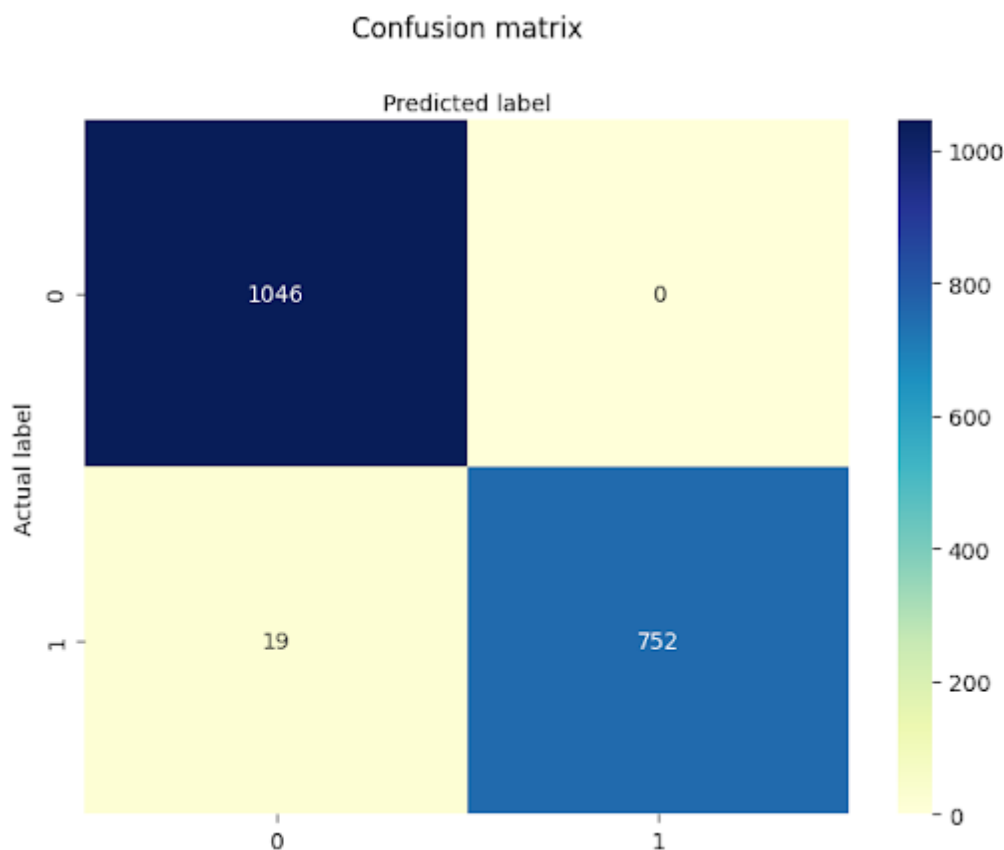


Figura 6. Matriz de Confusão.

4.5. Correlação

Com base na matriz de correlação, observa-se a relação entre diversas variáveis e a métrica xKP. Nota-se que a métrica xKP é positivamente influenciada por características como a

maior distância do marcador mais próximo e cobranças realizadas fora da área, indicando que essas situações favorecem a manutenção da posse de bola. Em contrapartida, variáveis como altura e velocidade da bola têm uma correlação negativa com a xKP, sugerindo que cobranças mais rasteiras e lentas tendem a manter a posse com o time atacante.

Além disso, outras zonas do campo influenciam a xKP de formas variadas, assim como a diferença no número de jogadores na área após a cobrança. Apesar de a métrica xKP estar associada à manutenção da posse, sua influência sobre as chances de gol, ainda que sutil, não deve ser subestimada. Comparada individualmente a outras features isoladas, a xKP oferece uma visão mais abrangente do potencial ofensivo, capturando aspectos que isoladamente poderiam passar despercebidos. Portanto, mesmo que a correlação direta entre xKP e gols não seja tão evidente, a métrica desenvolvida sugere uma influência significativa nas oportunidades de gol.

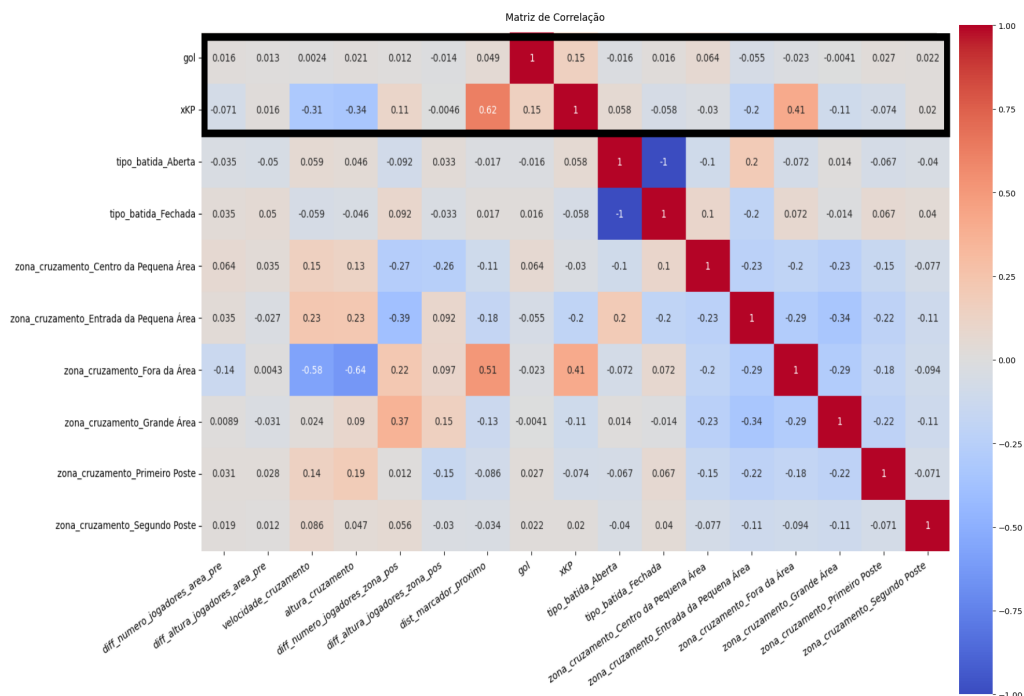


Figura 7. Matriz de Correlação.

5. Conclusão

Dado que o objetivo ao longo do projeto foi alterado, conclui-se que uma análise mais robustas dos escanteios necessita de dados mais confiáveis. Apesar de numerosos as features não foram suficientemente expressivas para capturar os gols.

Porém, foi possível criar uma métrica que captura um aspecto relevante dos escanteios e que tem o maior grau de influência sobre o gol entre as features.

Análise de Contra-Ataques na Premier League 22/23

Artur F. Souza¹, Daniel B. Torres¹, Henrique C. A. Souza¹, Lucas A. M. Silva¹

¹ Departamento de Ciência da Computação
Universidade Federal de Minas Gerais (UFMG) – Horizonte, MG – Brazil

arturfsouza@ufmg.br, daniboybt@ufmg.br, henriquecas@ufmg.br, lucasaraujoms3@ufmg.br

Abstract. *This article presents an analysis of counterattacks during the 2022/23 Premier League season. The focus of our work was on identifying key patterns in ball possession sequences, filtering the plays that meet our definition of "counterattack", aiming to find patterns that lead to successful counterattacks. We examine several metrics such as ball recovery locations, transition times to the offensive field and final results, considering the opponent's penalty area as the ultimate objective.*

Resumo. *Este artigo apresenta uma análise de contra-ataques durante a temporada 2022/23 da Premier League. O foco de nosso trabalho se deu na identificação de padrões chave em sequências de posse de bola, filtrando as jogadas que se encontram em nossa definição de "contra-ataque", visando encontrar padrões que levam a contra-ataques bem-sucedidos. Examinamos diversas métricas como localizações de recuperação de bola, tempos de transição para o campo ofensivo e resultados finais, considerando a grande área adversária como o objetivo final.*

1. Introdução

A análise de contra-ataques no futebol é crucial para entender a dinâmica do jogo e a eficácia das equipes em momentos de transição. O presente estudo foca na temporada 22/23 da Premier League, utilizando dados de tracking fornecidos pela PFF, com ênfase nas características que distinguem contra-ataques bem-sucedidos, como médias de ações (passes, dribles, entre outros), porcentagem de sucesso e análise de ações decisivas, por exemplo. A análise inclui a identificação de eventos de posse de bola que se enquadram em nossa definição de contra-ataque.

2. Metodologia

2.1. Definição de Contra-Ataque

Para definir o que caracteriza um contra-ataque no contexto do nosso trabalho, foram estabelecidos os seguintes critérios, com base em [Sahasrabudhe and Bekkers 2023]:

- **Recuperação de Bola no Campo Defensivo:** A sequência de posse deve iniciar-se com a recuperação da bola no campo defensivo.
- **Transição Rápida para o Campo Ofensivo:** A bola deve alcançar a metade ofensiva em menos de 10 segundos.
- **Movimento Avançado Significativo:** A bola deve avançar pelo menos 10 metros durante a transição da metade defensiva para a ofensiva.
- **Sucesso do Contra-Ataque:** O contra-ataque é considerado bem-sucedido se a bola alcançar a grande área adversária.

2.2. Estratégias de Processamento de Dados

Para analisar os dados da temporada 2022/23 da Premier League, fornecidos pela PFF, várias estratégias de processamento de dados foram empregadas em Python, garantindo a qualidade e a consistência dos dados.

1. Carregamento e Limpeza de Dados:

Os dados foram carregados e organizados em DataFrames, contendo informações detalhadas sobre jogadores, quadros de jogo (frames), eventos de jogo e eventos de posse. Durante o processo de carregamento, foram removidas duplicatas e padronizadas colunas, assegurando a consistência dos dados ao longo das partidas analisadas.

2. Estruturação dos Dados de Jogadores:

Os dados dos jogadores foram extraídos e categorizados por time (casa ou visitante), permitindo o cálculo de métricas adicionais, como velocidade e aceleração, ao longo do tempo. Essa estruturação possibilitou uma análise dinâmica do desempenho dos jogadores.

3. Organização dos Eventos do Jogo:

Eventos do jogo, como posse de bola e eventos significativos, foram associados aos respectivos frames, vinculando as ações específicas do jogo aos momentos em que ocorreram. Essa associação foi essencial para uma análise detalhada dos eventos em relação ao comportamento dos jogadores.

4. Integração dos Dados:

Após a extração e organização, os dados de frames, jogadores e eventos foram integrados em conjuntos de dados consolidados. Essa etapa garantiu que todas as informações estivessem sincronizadas e prontas para análises posteriores, como a identificação de sequências de posse de bola e eventos críticos.

5. Cálculo de métricas avançadas:

Com os dados organizados, foram calculadas métricas avançadas, como velocidades e acelerações dos jogadores, com base nas coordenadas espaciais em cada frame. Essas métricas forneceram insights adicionais sobre o comportamento dos jogadores durante diferentes momentos do jogo.

2.3. Processamento de Sequências de Posse

Nesta etapa, foi descrito o procedimento para identificar e organizar as sequências de posse de bola em um DataFrame denominado "eventos posse com id". O objetivo era identificar as diferentes sequências de posse ao longo do jogo e associá-las ao respectivo ID de sequência.

1. Inicialização do DataFrame e Variáveis:

Um DataFrame vazio chamado "eventos posse com id" foi criado com as colunas necessárias, e a variável "sequence id" foi inicializada para rastrear os IDs de sequência.

2. Identificação de Mudanças na Posse de Bola:

A função "possession changed" foi implementada para detectar mudanças na posse de bola entre eventos consecutivos, comparando o time atual com o novo time.

3. Iteração sobre os Eventos de Posse:

O processo iterou sobre todos os eventos de posse contidos em "possession events", buscando informações correspondentes no DataFrame de eventos do jogo e filtrando por frame num e game id.

4. **Deteção de Novas Sequências:** Novas sequências foram detectadas quando o evento correspondia a uma saída de bola (kick-off) ou quando havia uma mudança na posse de bola. Nesses casos, o DataFrame foi atualizado com o ID da sequência atual, e uma nova sequência foi iniciada.
5. **Construção do DataFrame de Sequências:** À medida que novos eventos eram processados, eles eram adicionados a uma lista temporária, que, ao detectar uma nova sequência, era transformada em um DataFrame e concatenada ao DataFrame principal "eventos posse com id".
6. **Adição de Colunas e Integração com Dados de Posição:** Após a construção do DataFrame principal, foi realizada uma fusão (merge) com o DataFrame "frames", que continha as posições da bola em cada frame. Isso permitiu associar a cada evento de posse as coordenadas espaciais e o período do jogo.

Esse processo resultou em um DataFrame final que facilitou a análise das sequências de posse de bola, permitindo identificar a dinâmica do jogo e avaliar aspectos táticos, como contra-ataques e transições rápidas.

3. Análise

3.1. Análise Descritiva

A partir dos dados obtidos no dataframe de contra ataques, conseguimos obter estatísticas descritivas, como:

- Média de contra-ataques por jogo: 8
- Média de eventos por contra-ataque: 6
- Média de passes por contra-ataque: 4
- Porcentagem de contra-ataques que tiveram finalizações: 13%
- Média de passes por contra-ataques de sucesso (que alcançaram a grande área): 6
- A porcentagem de contra-ataques com sucesso foi de: 13%

3.2. Plots de um Contra Ataque

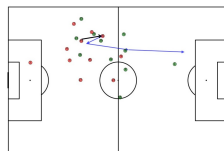


Figure 1. Sequência de Eventos em um Contra Ataque (1)

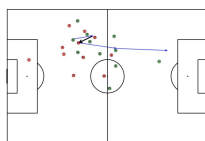


Figure 2. Sequência de Eventos em um Contra Ataque (2)

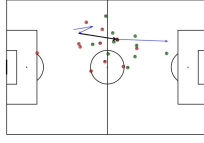


Figure 3. Sequência de Eventos em um Contra Ataque (3)

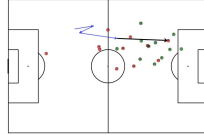


Figure 4. Sequência de Eventos em um Contra Ataque (4)

4. Modelos

Utilizamos dois principais modelos para classificação e análise de contra-ataques: K-means para clusterização e Multilayer Perceptron (MLP) para aprendizado supervisionado.

Como *features*, utilizamos:

- Tempo total do contra-ataque (em segundos).
- Número de atacantes: *Calculado pelo número de jogadores do time atacante que, no começo da jogada, estão a uma distância de até 5 metros da bola ou à frente da linha da bola.*
- Número de defensores: *Calculado pelo número de jogadores do time defensor que, no começo da jogada, estão a uma distância de até 5 metros da bola ou atrás da linha da bola.*
- Velocidade média da bola.
- Pressão da defesa:

$$\frac{1}{\sum_{i=1}^n d_i} \quad \text{onde } d_i$$

representa a distância do i-ésimo defensor até a bola.

- Pressão do ataque:

$$\frac{1}{\sum_{i=1}^n d_i} \quad \text{onde } d_i$$

representa a distância do i-ésimo atacante até a bola.

- Distância percorrida nos eixos x e y: *Distância total percorrida da bola.*
- Lado do campo onde houve a recuperação da bola.
- Verificação binária que verifica se a jogada chegou ou não na grande área: *Será utilizada como nossa identificação de sucesso do contra ataque.*

A extração de features resultou em um conjunto de 58 amostras válidas, das quais 12 alcançaram a grande área. Uma rede neural MLP com arquitetura 16-8-1 foi treinada, empregando funções de ativação ReLU nas camadas intermediárias e sigmoide na camada de saída. A fim de mitigar o overfitting, foi utilizada a técnica de regularização EarlyStopping.

5. Resultados

5.1. Clusters

Utilizando o método do cotovelo podemos ver que bons números de número de centróides giram entre 2 à 4 centróides.

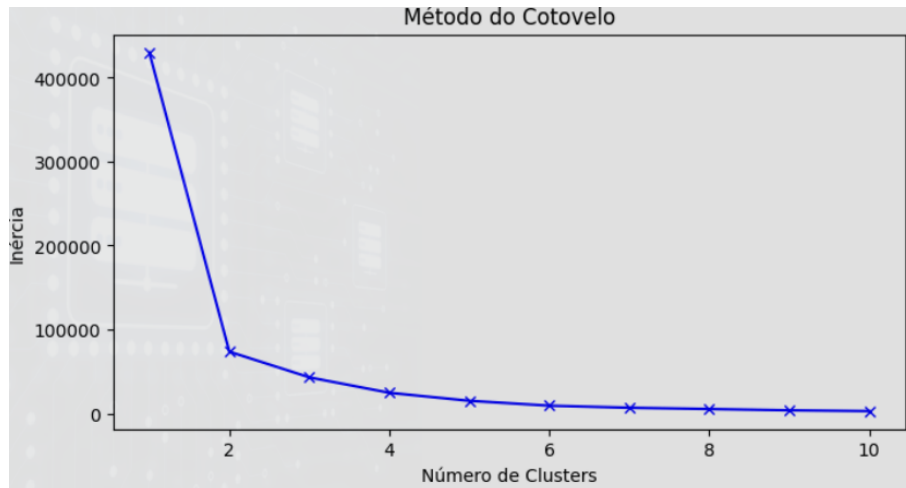


Figure 5. Método do Cotovelo

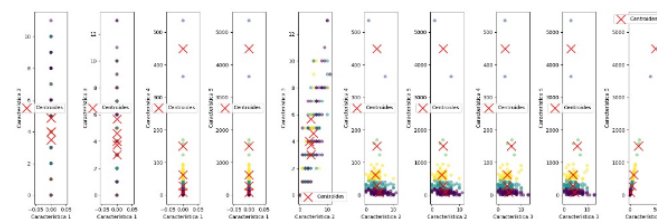


Figure 6. Centróides

Com base na análise dos clusters identificados pelo algoritmo *KMeans*, podemos extrair algumas conclusões sobre os padrões de contra-ataques:

- **Identificação de Padrões de Contra-Ataques:** O método de *clustering* foi utilizado para agrupar sequências de posse de bola que compartilham características semelhantes, como tempo total da sequência, distância percorrida, velocidade média da bola, e presença na grande área. Clusters bem definidos sugerem que existem diferentes tipos de contra-ataques que ocorrem em situações semelhantes no jogo.
- **Número de Clusters Otimizado:** O método do cotovelo sugere que 3 clusters são adequados para segmentar as sequências de posse analisadas. Isso pode indicar que os contra-ataques podem ser classificados em três categorias principais, cada uma representando um padrão distinto de jogo.
- **Aplicação Prática:** A categorização dos contra-ataques em clusters pode ser utilizada para criar estratégias de jogo mais eficazes, ajudando a equipe a identificar e replicar os padrões mais eficientes, ou a preparar-se melhor contra tipos específicos de contra-ataques adversários.

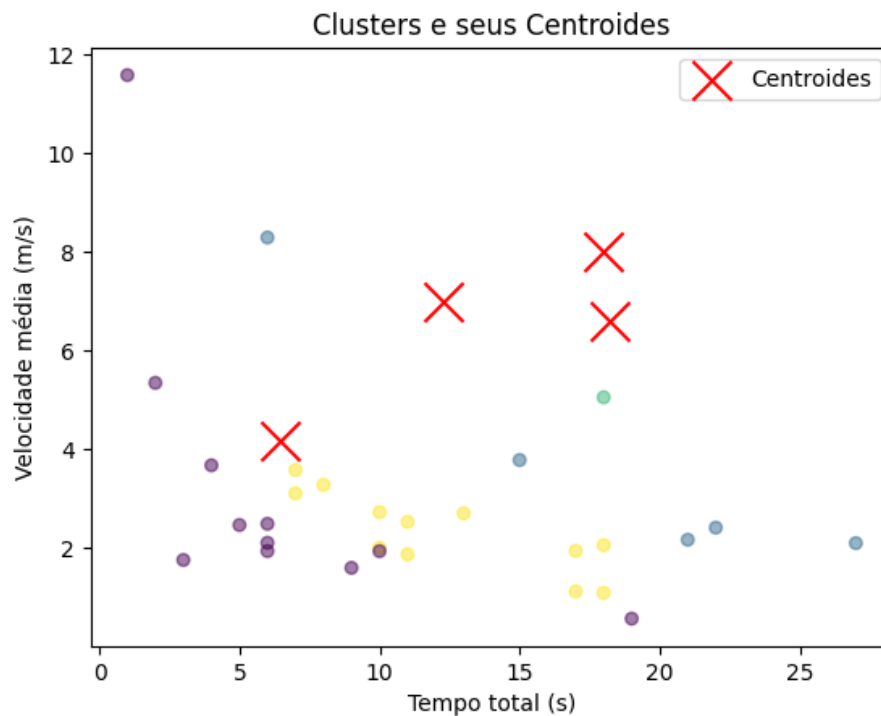


Figure 7. Velocidade x Tempo

5.2. Multilayer Perceptron - MLP

O treinamento do modelo para 50 épocas gerou os seguintes resultados:

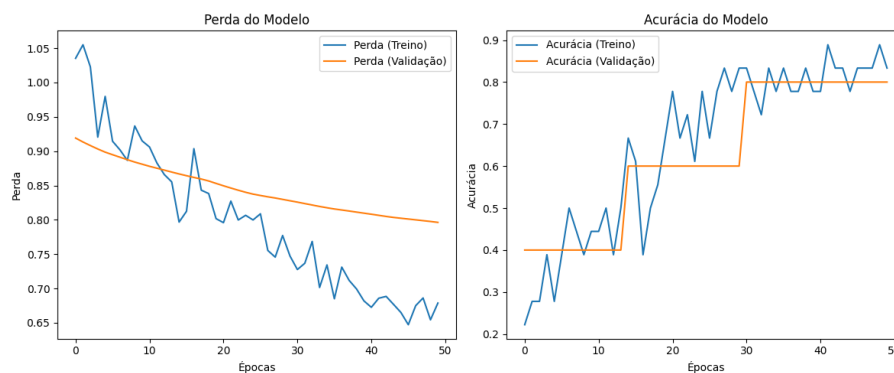


Figure 8. Perda e acurácia do modelo

A curva de erro durante o treinamento e a validação apresentou proximidade, o que indica um desempenho satisfatório do modelo. No entanto, ao analisar a matriz de confusão, observamos que o modelo tende a produzir respostas negativas, independentemente das características das amostras. É importante destacar que a escassez de amostras torna inviável a construção de um modelo robusto e eficaz.

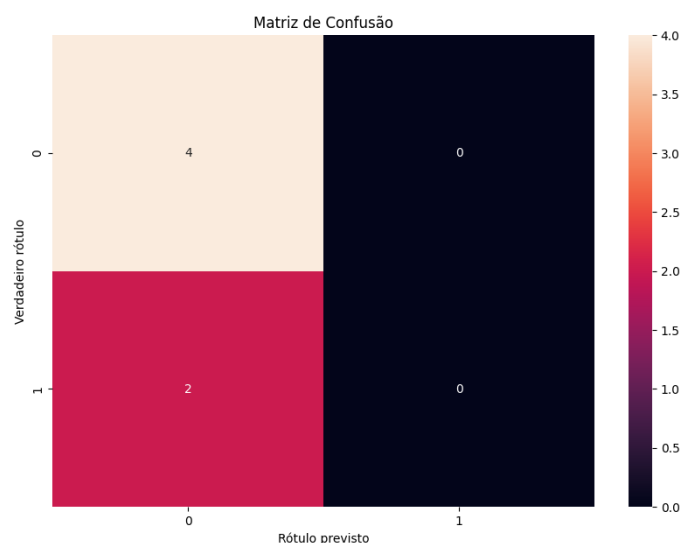


Figure 9. Matriz de confusão

6. Conclusão

Este estudo sobre contra-ataques na temporada 2022/23 da Premier League revela insights valiosos sobre a dinâmica dessas jogadas e os fatores que contribuem para seu sucesso. Identificamos que as sequências de contra-ataque tendem a ser rápidas e envolvem menos eventos, com uma média de passes mais alta em contra-ataques bem-sucedidos. Isso sugere que a qualidade da troca de passes é um componente crucial para a eficácia dessas jogadas, possibilitando manter a posse de bola e criar oportunidades de gol.

A taxa de sucesso de 13% dos contra-ataques destaca a importância do treinamento e da preparação tática focada em transições rápidas e eficientes. No entanto, a análise também indica que há espaço para melhorias. A compreensão detalhada dos fatores que levam ao sucesso ou fracasso de um contra-ataque pode orientar estratégias mais eficazes, como a adaptação das táticas de pressão e a otimização das ações de recuperação de bola.

Para trabalhos futuros, recomenda-se a correção de problemas de processamento em análises com grandes volumes de dados, o que permitirá uma exploração mais robusta e detalhada das sequências de contra-ataque. Outra linha de investigação importante seria a definição mais precisa do sucesso de um contra-ataque, levando em conta a zona final do campo onde a jogada se encerra. Adicionalmente, uma avaliação detalhada de cada evento dentro de um contra-ataque pode oferecer novas perspectivas sobre como melhorar a eficácia dessas jogadas.

Uma possível extensão deste estudo seria a análise comparativa de contra-ataques entre diferentes equipes, visando identificar padrões e nuances associados a estilos de jogo, treinadores e características específicas de cada time.

References

Sahasrabudhe, A. and Bekkers, J. (2023). *A Graph Neural Network deep-dive into successful counterattacks*. 42 Analytics.

Simplified Pass Expected Possession Value in PFF Data Implementation and Analysis

Diogo T. Chaves¹, João Marcos T. S. Campos¹, Rafael M. Gomes¹, Vitor Mendes²

¹Departamento de Ciência da Computação – Universidade Federal de Minas Gerais (UFMG)
Belo Horizonte – MG – Brazil

²Departamento de Computação – Centro Federal de Educação Tecnológica de Minas Gerais
(CEFET - MG)
Belo Horizonte – MG – Brazil

diogochaves@dcc.ufmg.br, joaomarcoscampos@dcc.ufmg.br,
rafamg@ufmg.br, vitormabreu23@gmail.com

Abstract. *Passing is one of the most frequent and impactful actions in a football match, essential for game progression and team success. This study aims to implement and apply the Expected Possession Value (EPV) model, as proposed by Javier Fernández, to evaluate plays and player performance. In addition to the standard EPV model, two complementary models are introduced: one for estimating the success probability of passes and another for evaluating the probability of pass selection. The combination of these models allows for detailed and precise analysis, revealing special conditions that influence pass success and, consequently, on-field performance.*

Resumo. *O passe é uma das ações mais frequentes e impactantes em uma partida de futebol, sendo essencial para a progressão do jogo e o sucesso de um time. Este trabalho tem como objetivo implementar e aplicar o modelo de Valor Esperado da Posse de Bola (EPV), proposto por Javier Fernández, para avaliar jogadas e o desempenho dos jogadores. Além do modelo padrão de EPV, são introduzidos dois modelos complementares: um para estimar a probabilidade de sucesso dos passes e outro para avaliar a probabilidade de escolha dos mesmos. A combinação desses modelos permite uma análise detalhada e precisa, revelando condições especiais que influenciam o sucesso dos passes e, consequentemente, o desempenho em campo.*

1. Introdução

Dentro de uma partida de futebol, o passe é o evento mais comum e, quando bem executado, impacta significativamente a progressão do time. Times e jogadores que dominam essa ação se destacam, tornando a análise detalhada dos passes essencial para otimizar o desempenho coletivo e individual.

A motivação deste trabalho é entender o que faz uma ação individual agregar valor a uma partida e medir esse valor. O objetivo final é implementar o modelo de Valor Esperado da Posse de Bola (EPV), conforme proposto por Javier Fernández [Fernández et al. 2021], para avaliar jogadas e jogadores, considerando passes, chutes e conduções de bola.

Também apresentamos dois modelos complementares: um que estima a probabilidade de sucesso de um passe, considerando posição, velocidade dos jogadores e a dinâmica do jogo; e outro que avalia a probabilidade de escolha de um passe, considerando alternativas e o contexto. A combinação desses modelos permite análises mais precisas sobre a importância dos passes e a performance em campo.

O foco do trabalho é analisar detalhadamente as jogadas, buscando padrões e insights que contribuam para uma avaliação mais profunda do desempenho dos jogadores. Com o apoio dos monitores Gabriel Valadão e Bruno Martins, do professor Leo Freire e do mentor Hugo Rios Neto, foi possível desenvolver uma pesquisa satisfatória.

2. Trabalhos Relacionados

Diversos estudos têm explorado a análise quantitativa e qualitativa de eventos em jogos de futebol, com foco especial na avaliação de passes e sua contribuição para o desempenho de uma equipe. Dentre os trabalhos mais relevantes, destaca-se o modelo de *Expected Possession Value* (EPV) desenvolvido por Fernández et al. [Fernández et al. 2021], que oferece uma metodologia abrangente para estimar o valor esperado de uma posse de bola em diferentes contextos utilizando o soccermap descrito em [Fernández and Bornn 2020]. Este modelo se tornou uma referência na análise de futebol por sua capacidade de integrar diferentes tipos de eventos, como passes, finalizações e conduções de bola, oferecendo uma métrica quantitativa para avaliar a eficiência das ações de jogo. Para avaliar esses modelos pegamos como base as análises feitas em [Fernández 2021].

3. Metodologia

3.1. Compreensão e preparação dos dados

O pré-processamento dos dados envolveu a filtragem e organização dos eventos e tracking para preparar as entradas dos modelos. Para cada jogo, os dados foram divididos em dataframes específicos: passes, conduções de bola, finalizações e informações de tracking, como posições, velocidades e acelerações dos jogadores no momento dos eventos.

Após padronizar os IDs dos jogadores para garantir a consistência entre jogos, geramos dataframes onde cada linha representa um passe, incluindo atributos relevantes de todos os jogadores em campo. O resultado foram arquivos CSV contendo dados de passes, conduções e finalizações para cada jogo, usados como entrada para os modelos.

Na fase de engenharia de features, criamos 17 atributos essenciais para capturar a dinâmica de uma partida de futebol. Essas features incluem posições dos jogadores, distâncias até a bola e o gol, ângulos entre a bola e o gol, além de velocidades nos eixos X e Y. Também consideramos o número de jogadores entre a bola e cada ponto do campo, e a distância até o início do evento. Cada feature é representada por uma matriz que fornece ao modelo uma visão detalhada e contextualizada, capaz de capturar nuances táticas e estratégicas essenciais para a análise do jogo.

3.2. Modelos

Em relação à modelos nosso objetivo foi implementar os modelos de passe do descrito no artigo [Fernández et al. 2021] de forma simplificada nos dados da PFF. Cada

instância dos modelos era um evento de algum dos jogos, tendo sido usado conjunto de 52 jogos, onde 45 foram reservados para treino (80% dos eventos) e validação (20% dos eventos) e os 7 restantes foram destinados ao teste. Todas as instancias foram convertidas nas 17 features, resultando em uma entrada com 17 canais.

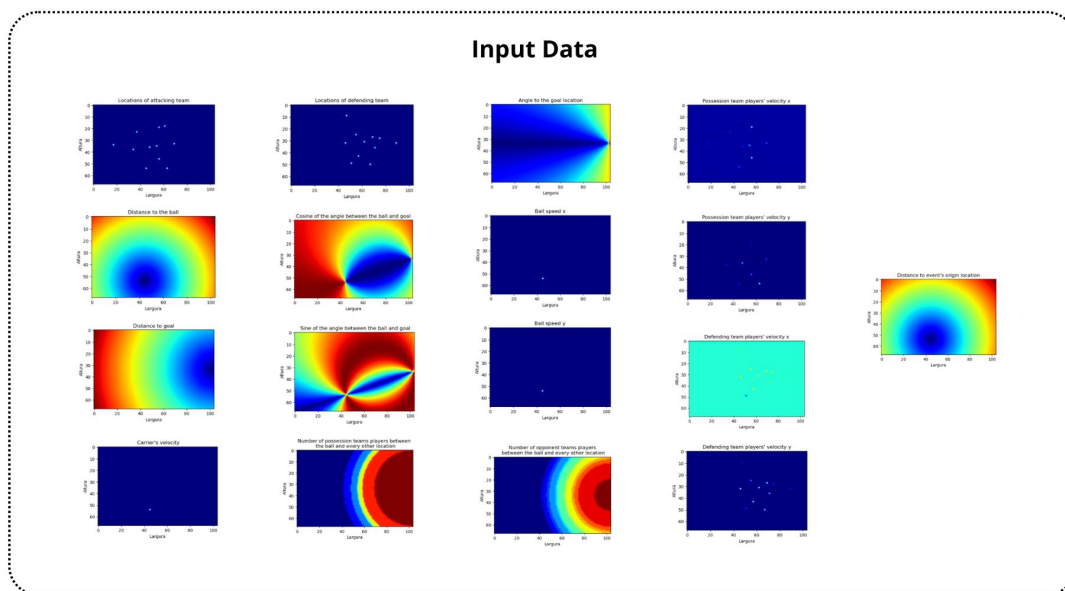


Figure 1. Exemplo de entrada do modelo

3.2.1. Probabilidade de Sucesso de Passes

O "modelo de probabilidade de sucesso de passes" utiliza a rede "Soccermap", descrita no artigo [Fernández and Bornn 2020], em junção com uma função sigmoide na saída para converter essas matrizes de entrada em um tensor onde cada elemento representa uma posição do campo e seu valor corresponde à probabilidade de um passe ser bem-sucedido nesse local. Vale ressaltar que essa rede possui uma arquitetura totalmente convolucional e utiliza a convergência de uma função de perda através do método de gradiente descendente para conseguir aprender sobre os dados.

Para realizar o treinamento é aplicado a função de perda de entropia cruzada binária na posição onde foi feita a tentativa de passe. Essa função de perda recebe como entrada o valor gerado pelo modelo com o valor "0", caso o passe tenha sido um sucesso, ou "1", caso por algum motivo o passe tenha falhado.

3.2.2. Probabilidade de Escolha de Passes

O modelo de probabilidade de escolha estima a probabilidade de um jogador optar por um determinado passe em determinada localidade do campo, baseando-se nas matrizes geradas.

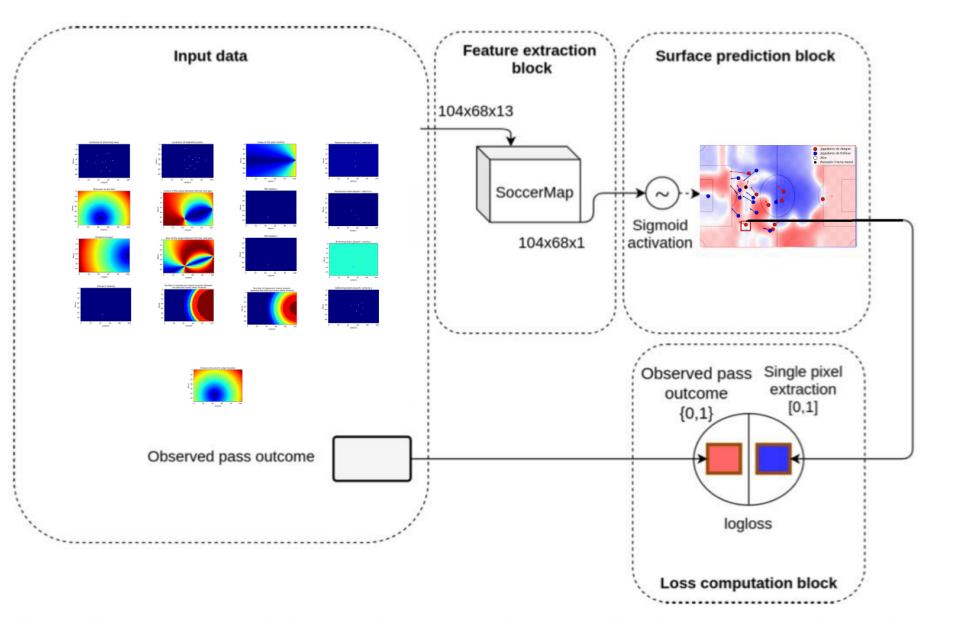


Figure 2. Imagem do artigo [Fernández et al. 2021] com as adaptações do nosso modelo mostrando o treinamento funciona

Este modelo é bem parecido com o modelo de probabilidade de passes, com apenas algumas diferenças. Diferentemente do modelo de probabilidade de passes, que utiliza uma função sigmoide para suas estimativas, o modelo de probabilidade de escolha adota uma abordagem distinta. Primeiramente, ele transforma a matriz de saída em um vetor, facilitando a aplicação da função de entropia cruzada. Essa transformação é essencial para garantir que a distribuição dos resultados seja adequada, de modo que a soma de todos os elementos do vetor seja igual a "1", refletindo a probabilidade total de todas as opções de passe possíveis.

A função de perda do modelo também é entropia cruzada binária aplicada na posição onde a tentativa de passe foi realizada entre o valor predito pelo modelo e 1, pois aquela posição foi a selecionada para a realização do passe.

3.2.3. Valor de posse esperado

O modelo de valor esperado da posse de bola quantifica o valor esperado de uma posse de bola, proporcionando uma métrica para avaliar a eficácia das ações em campo. Para realizar seu treinamento este modelo adiciona mais um canal na matriz de entrada de cada uma instancia, esse canal é o resultado do modelo de probabilidade de sucesso de passes desse evento.

A função de perda do modelo é erro quadrático médio (MSE) na posição onde a tentativa de passe foi realizada entre o valor predito pelo modelo e a recompensa de longo prazo associada ao resultado observado da posse de bola após o passe. Esse modelo de recompensa também é descrito no artigo [Fernández et al. 2021] e se baseia na ideia de analisar e recompensar, positivamente e negativamente, os eventos entre o passe e o próximo gol. Conseguimos implementar o modelo; no entanto, a extração dessas rec-

ompensas encontrou um contratempo e ainda não foi concluída, dado que, em alguns dos dados de evento, os gols não estão devidamente sinalizados.

4. Resultados

Para avaliar o modelo, primeiramente os valores probabilísticos foram convertidos em 0 ou 1 com base em um threshold, permitindo o cálculo das métricas apresentadas na Figura 3. Escolhemos um threshold de 0,75, já que os passes são comuns durante o jogo e geralmente têm alta taxa de sucesso. Com um threshold elevado, apenas os passes com alta probabilidade são classificados como bem-sucedidos. No entanto, as métricas obtidas indicaram que essa abordagem não era ideal.

Accuracy	Precision	Recall	F1 Score
0.8384781048097631	0.4192390524048816	0.5	0.4560718469347911

Figure 3. Métricas utilizadas – probabilidade de sucesso de passes

Reconhecendo as limitações dessa metodologia, optamos por utilizar plots para uma análise mais detalhada dos resultados do modelo. Após testes exaustivos com diferentes hiperparâmetros, obtivemos os melhores resultados.

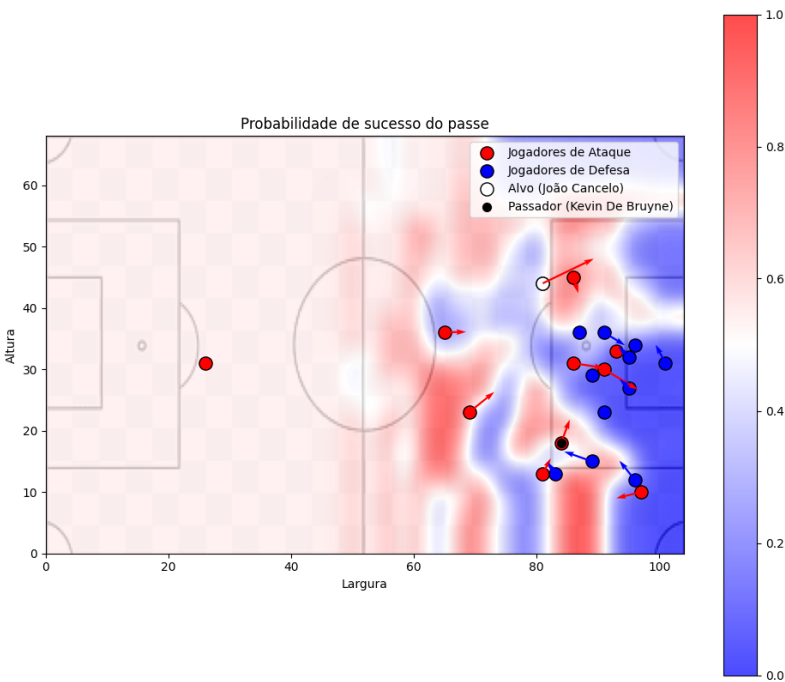


Figure 4. Probabilidade de sucesso de um passe

A Figura 4 mostra um heatmap gerado pelo modelo de probabilidade de sucesso de passes. O modelo produziu um mapeamento coerente, destacando zonas com altas e baixas probabilidades de sucesso. As áreas de maior e menor chance de sucesso coincidem com a ocupação dos espaços pelos jogadores do mesmo time e do time adversário. Vale destacar que fatores como a velocidade do passador e dos demais jogadores em campo foram levados em conta para determinar as probabilidades em cada local do campo.

Análises sobre o posicionamento dos jogadores em campo, como as sugeridas em [Fernández 2021], indicam que otimizações na disposição dos jogadores (se deslocando para zonas de menor pressão adversária e/ou para mais perto do jogador com a bola) podem aumentar as chances de sucesso do passe, conforme ilustrado na Figura 5.

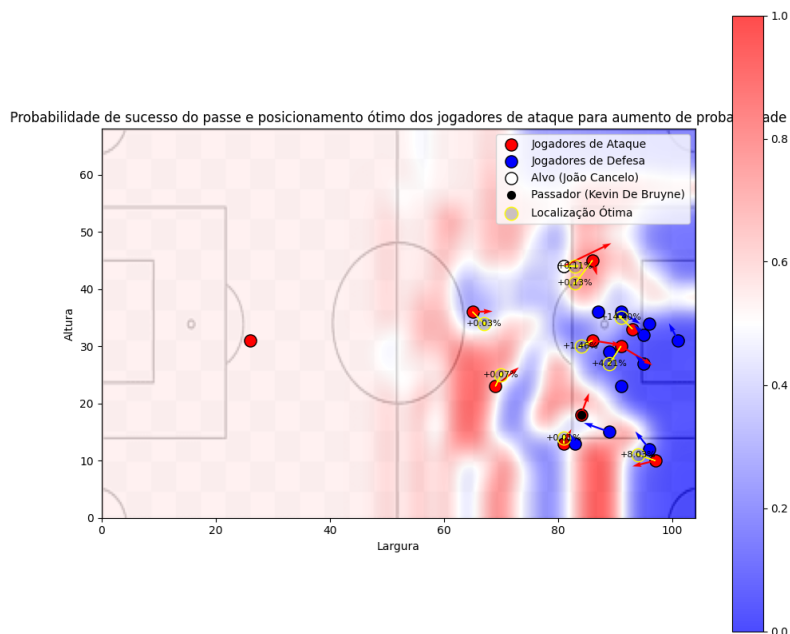


Figure 5. Probabilidades de sucesso de um passe dadas otimizações no posicionamento dos jogadores

Por outro lado, o modelo de probabilidade de escolha de passes apresentou resultados insatisfatórios, mesmo com thresholds baixos. A análise visual, como mostrada na Figura 6, se mostra mais apropriada e sugere que, em situações de ataque, onde o passador está dentro de uma área adversária populada, ele tende a escolher colocar a bola no miolo da grande área.

Estudos adicionais, como o proposto em [Fernández 2021], exploram como otimizar o posicionamento dos jogadores para aumentar a probabilidade de que eles recebam um passe, conforme ilustrado na Figura 7.

Essas visualizações são apenas pequenas amostras do potencial dos modelos apresentados. Metrificar a qualidade do posicionamento dos jogadores e de suas escolhas quando detém a bola, por exemplo, pode ser muito útil para treinadores que desejam mitigar a perda da posse de bola.

5. Conclusão

A análise dos modelos de probabilidade de sucesso e de escolha de passes forneceu insights valiosos sobre as dinâmicas de jogo, destacando a importância do posicionamento dos jogadores para a efetividade dos passes. O modelo de sucesso se mostrou eficaz ao mapear áreas com maior e menor probabilidade de um passe bem-sucedido, refletindo a realidade do jogo. Embora o modelo de escolha de passes tenha apresentado resultados menos satisfatórios, ele ainda sugeriu direções para otimizações táticas.

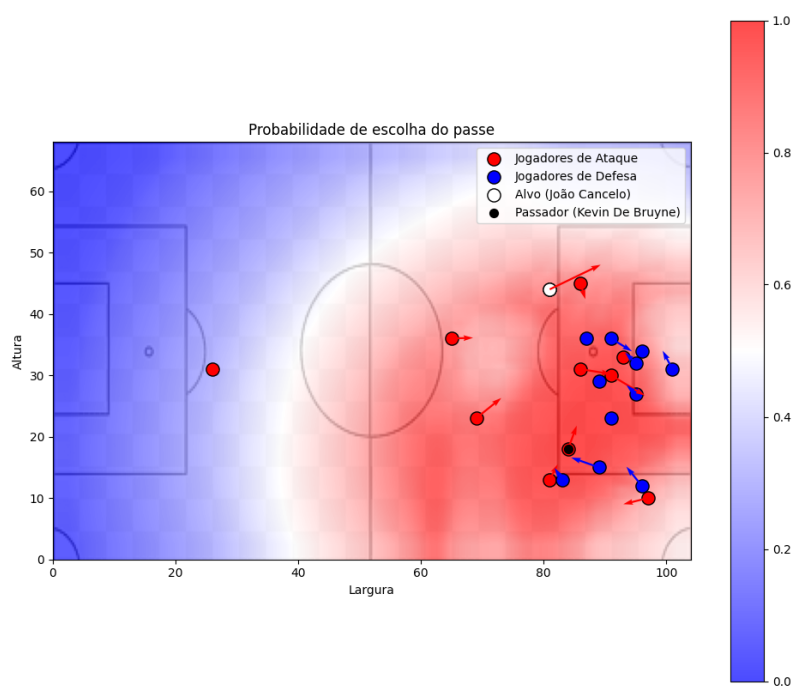


Figure 6. Probabilidade de escolha de um passe

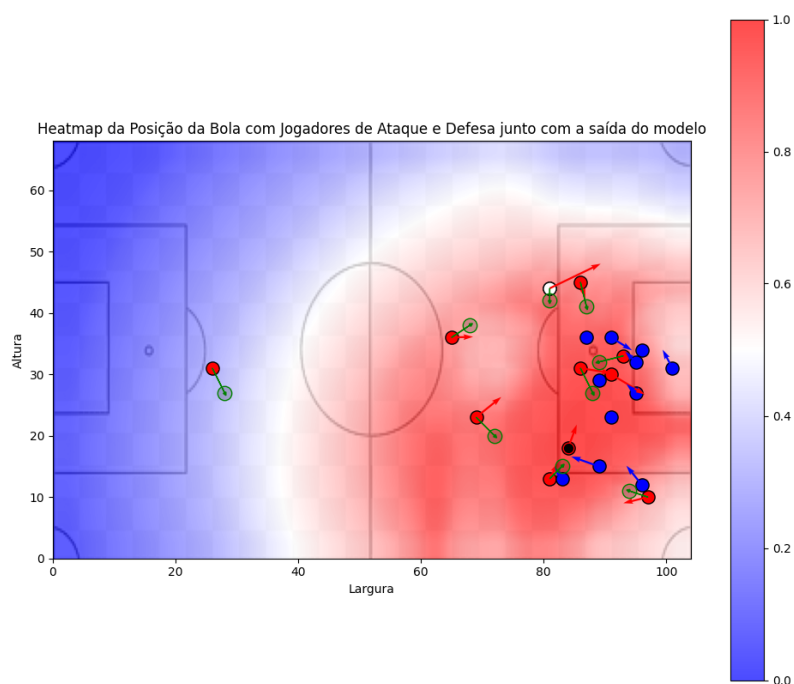


Figure 7. Probabilidades de escolha de um passe, dadas otimizações no posicionamento dos jogadores

Esses resultados indicam que, apesar das melhorias necessárias, a implementação simplificada dos modelos de Fernández contribuiu para uma melhor compreensão das decisões tomadas em campo. Ao integrar e expandir seu uso para situações reais, treinadores podem tomar decisões mais informadas, ajustando suas táticas para aumentar a eficácia

dos passes e, conseqüentemente, a performance geral do time.

Com isso, o intuito do trabalho foi cumprido, pois os resultados obtidos reforçam a relevância de métricas e visualizações detalhadas para aprimorar o desenvolvimento de estratégias no futebol.

References

Fernández, J. (2021). A framework for the analytical and visual interpretation of complex spatiotemporal dynamics in soccer. *Machine Learning*.

Fernández, J. and Bornn, L. (2020). Soccermap: A deep learning architecture for visually-interpretable analysis in soccer. *Machine Learning*.

Fernández, J., Bornn, L., and Cervone, D. (2021). A framework for the fine-grained evaluation of the instantaneous expected value of soccer possessions. *Machine Learning*.

Análise de Momentos Críticos em Partidas de Futebol a partir de Séries Temporais

Gabriel Vitor Lopes Gonçalves¹, Júlio César¹, Pablo Correa¹, Pedro Dias Pires¹

¹Departamento de Ciência da Computação
Universidade Federal de Minas Gerais (UFMG)
Belo Horizonte – MG – Brazil

{gabrielvlg, juliocesarp20, pablocorreacosta, pdp2019}@ufmg.br

Abstract. *This study focuses on the analysis of critical moments in football matches by leveraging time series data. Despite technological advancements and the increasing availability of sports data, the precise identification of these moments remains underexplored in the literature. The complexity lies in detecting these subtle yet decisive events, influenced by various factors such as player emotions, game pressure, and tactical strategies. By applying anomaly detection algorithms to specific game metrics, this research aims to identify and understand these critical moments, providing deeper insights into match dynamics. Keywords: critical moments, football, time series, anomaly detection, game metrics*

Resumo. *Este estudo concentra-se na análise de momentos críticos em partidas de futebol a partir do uso de séries temporais. Apesar dos avanços tecnológicos e da crescente disponibilidade de dados esportivos, a identificação precisa desses momentos ainda é pouco explorada na literatura. A complexidade está em detectar esses eventos sutis, porém decisivos, que são influenciados por fatores como o estado emocional dos jogadores, a pressão do jogo e as estratégias táticas. Aplicando algoritmos de detecção de anomalias em métricas específicas do jogo, esta pesquisa busca identificar e compreender esses momentos críticos, oferecendo uma compreensão mais profunda das dinâmicas das partidas. Palavras-chave: momentos críticos, futebol, séries temporais, detecção de anomalias, métricas de jogo*

1. Introdução

Com o aumento contínuo da disponibilidade de dados esportivos, a análise quantitativa no futebol tem ganhado crescente relevância, proporcionando uma nova dimensão ao entendimento do jogo. A identificação e compreensão dos momentos críticos durante uma partida — aqueles instantes que têm o potencial de alterar o resultado ou o desenrolar do jogo — tornam-se cada vez mais viáveis por meio dessa riqueza de dados.

Com isso em mente, o objetivo principal desta pesquisa se concentra em identificar e analisar esses momentos críticos, explorando a detecção de anomalias em diversas séries temporais das métricas de jogo. Essa abordagem permite uma compreensão mais aprofundada dos fatores que influenciam o resultado e a dinâmica das partidas, extraindo informações sobre o comportamento das equipes e dos jogadores em diferentes situações.

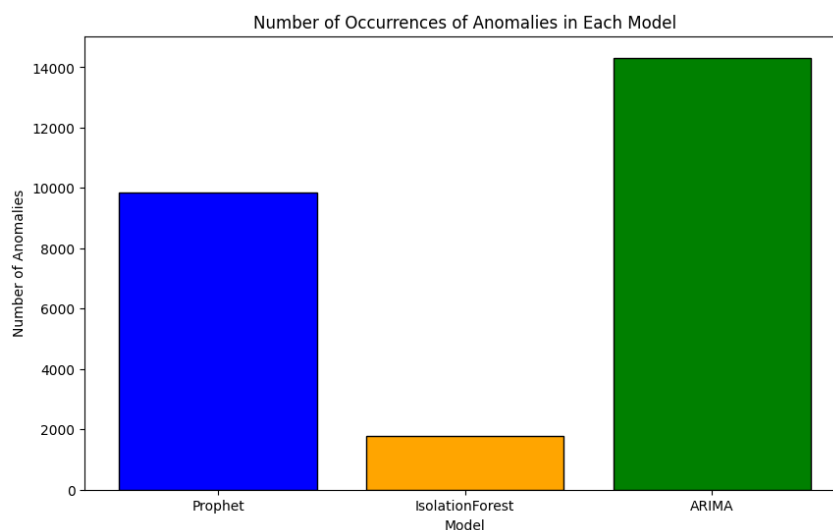


Figura 1. Comparação da quantidade de anomalias detectadas por algoritmo

2. Metodologia

A metodologia adotada no estudo envolve a modelagem das partidas de futebol como um conjunto de séries temporais, representando diferentes métricas do jogo ao longo do tempo. As métricas selecionadas foram: velocidade média dos jogadores, diferença no número de finalizações, diferença na posse de bola e diferença nas disputas de bola (desarmes). Os dados utilizados foram dados de *tracking* de partidas do Campeonato Brasileiro de 2023 [LLC 2023], fornecidos pela PFF, com cálculos realizados para cada *frame* dos dados de *tracking*. A detecção de momentos críticos ocorre também em nível de *frame*, com a aplicação de algoritmos de detecção de anomalias.

Para suavizar a volatilidade na série de velocidade média dos jogadores, foi utilizada uma média móvel com uma janela de um minuto (1800 *frames*), dado que os dados são capturados a 30 *frames* por segundo (fps). As métricas que envolvem diferenças, como posse de bola e finalizações, foram calculadas considerando a diferença entre os valores para o time da casa e o visitante. Além disso, foi gerada uma série temporal dos eventos ocorridos em cada *frame* do jogo, o que facilita a correlação entre as anomalias detectadas e os eventos críticos reais, como gols ou faltas importantes.

O algoritmo principal utilizado foi o *Isolation Forest*, escolhido por sua eficiência na detecção de anomalias e consistência entre diferentes partidas. Em comparação com outros métodos, como *ARIMA* e *Prophet*, o *Isolation Forest* destacou-se ao identificar um menor número de anomalias (Figura 1), sugerindo menor interferência de ruídos e maior precisão na identificação de momentos críticos. Este comportamento homogêneo é essencial para garantir que as anomalias detectadas sejam relevantes e não meramente o resultado de flutuações aleatórias nos dados.

3. Resultados

A aplicação do *Isolation Forest* permitiu identificar frames anômalos que indicaram a presença de eventos críticos (Figura 2). Esses *frames* foram então correlacionados com eventos reais nas partidas, facilitando uma análise mais profunda dos momentos que influenciam a dinâmica do jogo. Por exemplo, a diferença na posse de bola revelou padrões

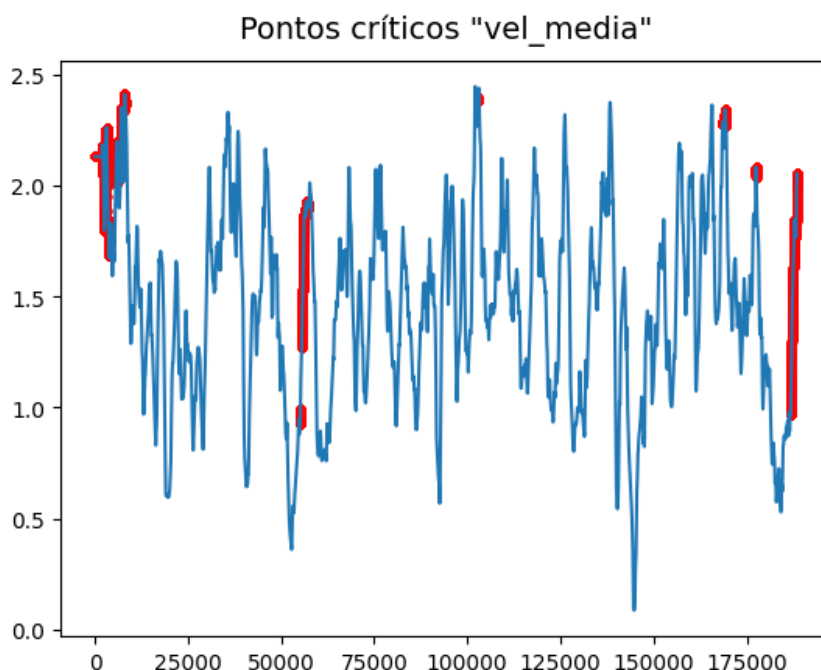


Figura 2. Série Temporal da Diferença da média móvel

anômalos que precederam gols decisivos, indicando mudanças estratégicas das equipes (Figura 3).

Uma análise detalhada foi realizada na partida 12828 do Campeonato Brasileiro de 2023, onde uma anomalia na posse de bola foi detectada minutos antes de um gol crucial (Figura 4), sugerindo uma mudança tática que levou ao evento crítico. Da mesma forma, a diferença no número de finalizações também apresentou anomalias que antecederam ataques perigosos e decisões táticas importantes (Figura 5).

Outro exemplo significativo foi a partida 12500 entre Fluminense e América-MG, onde o estudo identificou um momento crítico associado a um pênalti perdido pelo Fluminense. Apesar de ser um evento negativo, ele aumentou a intensidade do jogo para o Fluminense, que, em resposta, aumentou o número de finalizações e marcou três gols após o incidente. Um dos gols foi marcado pelo próprio jogador que perdeu o pênalti, Germán Cano, destacando a importância da resiliência e dos fatores emocionais no desempenho dos jogadores (Figura 6).

A análise global das anomalias em todas as partidas estudadas indicou que o número de *flags* por partida se estabilizou em torno de 14 mil. A visualização das séries temporais e a correlação com eventos reais permitiram uma compreensão mais profunda dos momentos que impactaram as partidas.

4. Conclusões

Analisando os resultados ao decorrer do estudo, no contexto de jogos de futebol, podemos definir eventos anômalos como aqueles que ocorrem de maneira rara ou inesperada, diferindo significativamente do padrão estabelecido de eventos “normais”. Esses eventos foram identificados como estando fora do *cluster* de eventos típicos, de acordo com

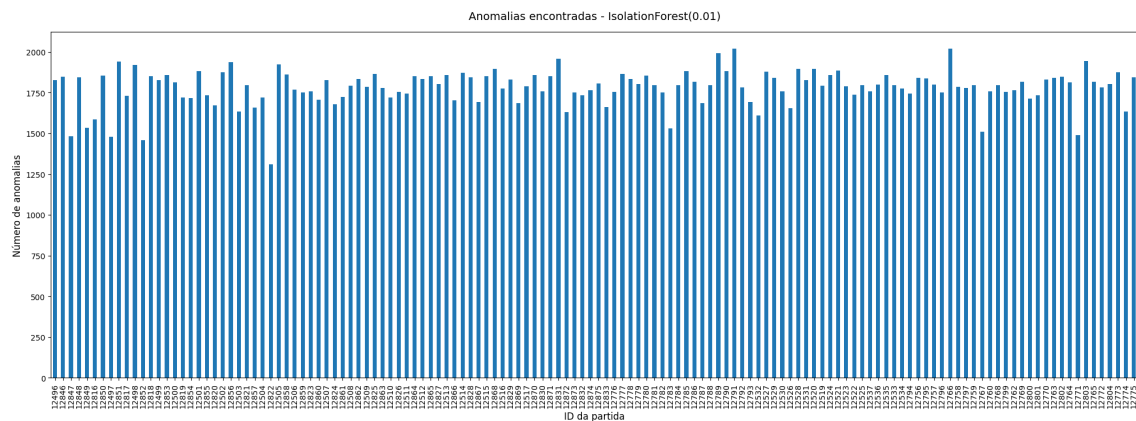


Figura 3. Quantidade de anomalias encontradas por partida

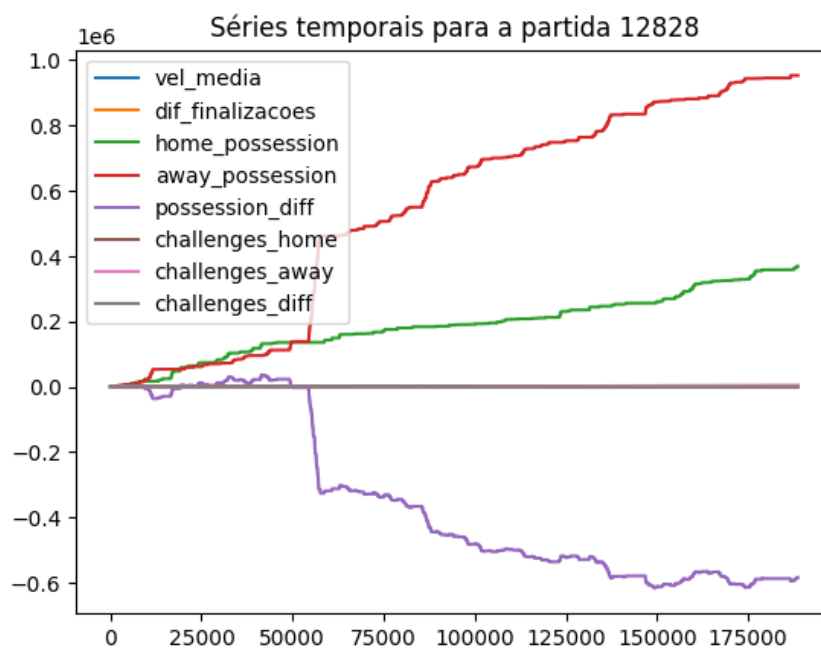


Figura 4. Séries temporais para a partida 12828

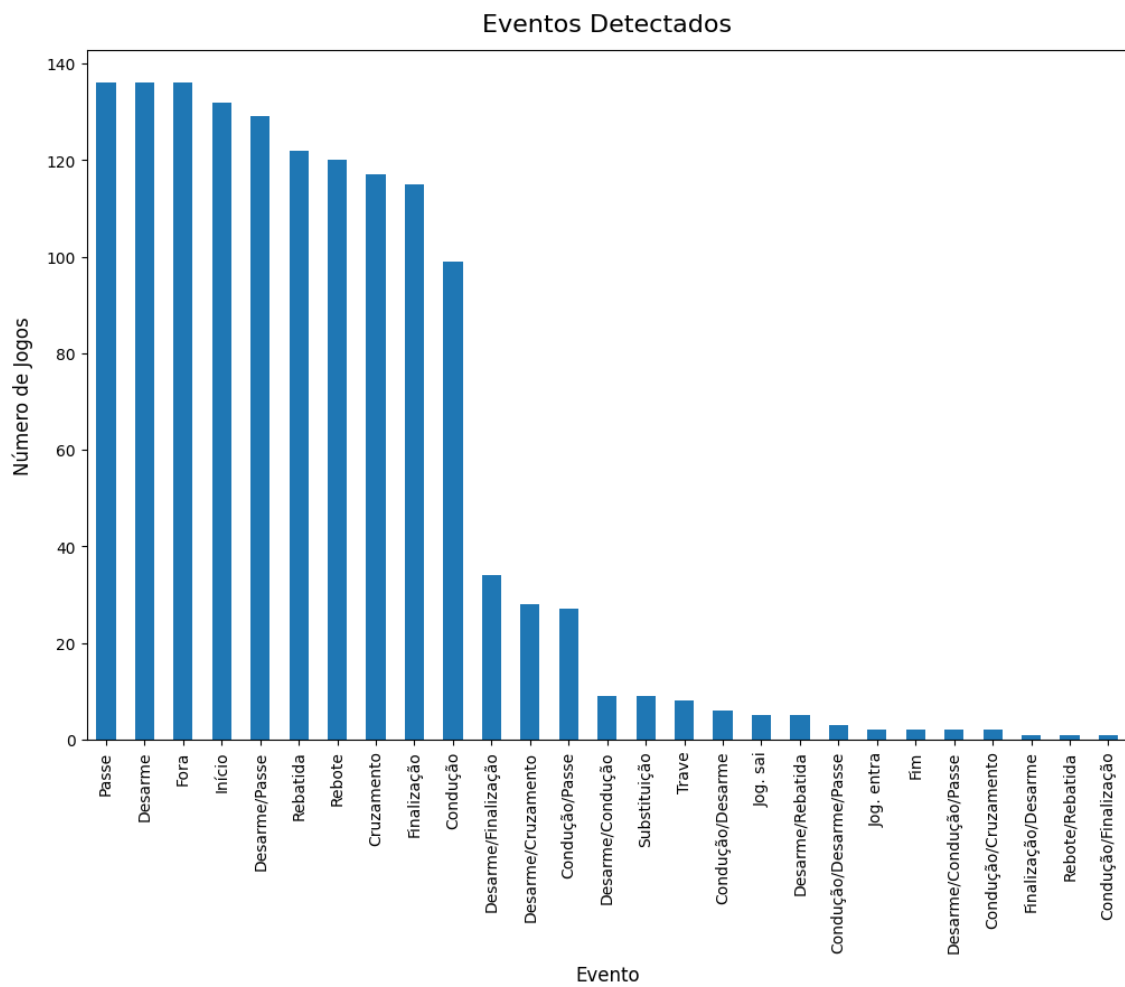


Figura 5. Eventos Críticos Detectados Globalmente

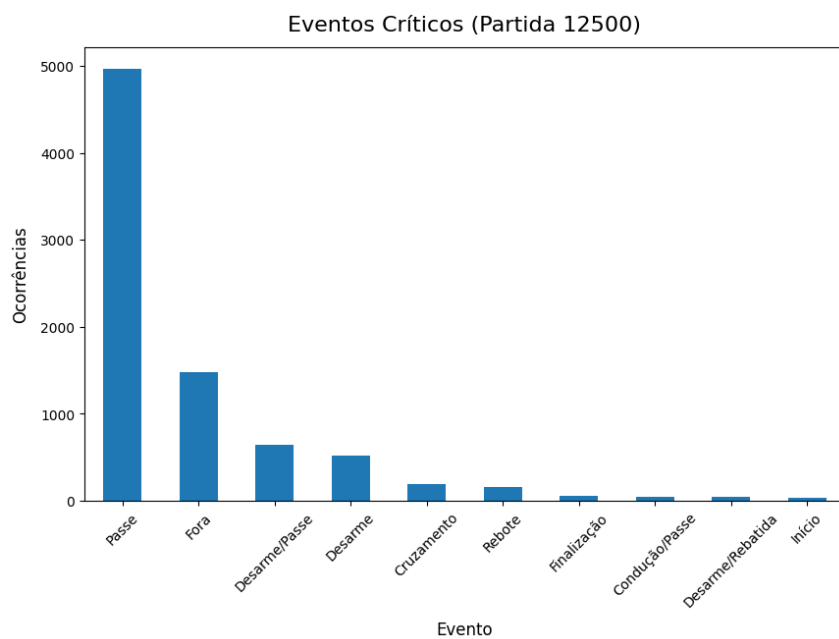


Figura 6. Eventos Críticos Detectados da partida 12500

o algoritmo do *Isolation Forest*, sugerindo uma discrepância significativa em relação ao comportamento esperado. A criticidade de um evento, por sua vez, está diretamente associada ao impacto psicológico que ele pode ter sobre os jogadores, influenciando não apenas o andamento da partida, mas também o desempenho subsequente da equipe.

Referências

LLC, P. A. (2023). PFF Tracking Data. <https://fc.pff.com/>.

Quais os indicadores-chaves de desempenho técnico-táticos discriminam a vitória e a derrota na 1ª divisão das cinco principais ligas da CONMEBOL?

André L. B. Jacinto¹, Guilherme P. Sousa², Filipe S. Guaceroni¹, Varley T. Costa¹

¹ UFMG SOCCER SCIENCE CENTER – Escola de Educação Física, Fisioterapia e Terapia Ocupacional (EEFFTO/UFMG). Av. Pres. Antônio Carlos, 6627 – Pampulha - Belo Horizonte (MG) - CEP 31270-901.

² Chair of Performance Analysis and Sports Informatics, Technical University of Munich, Munich, Germany.

`andreluiz.braga@hotmail.com, guilherme.gsp11@hotmail.com,
filipe.guaceroni@gmail.com, varleycosta@gmail.com.`

Abstract

The study aimed to identify which technical-tactical Key Performance Indicators (KPIs) discriminate between victory and defeat in the five main CONMEBOL leagues. A total of 5,536 matches from the top divisions of the five major CONMEBOL leagues were analysed during the 2017 to 2022 seasons. Fifty-five technical-tactical variables were examined, collected using InStat Scout®. Data analysis involved the Kolmogorov-Smirnov test, Levene's test, Box's test, MANOVA, and partial eta squared. The study identified seven offensive KPIs and one defensive KPI that differentiate between victory and defeat. It was concluded that offensive actions such as passing, finishing, efficiency in positional attacks, and counter-attacks discriminate victory in the top five CONMEBOL leagues.

Resumo

O estudo objetivou identificar quais Indicadores-Chave de Desempenho (ICD) técnico-tático discriminam a vitória e a derrota nas cinco principais ligas da CONMEBOL. Foram analisadas 5.536 partidas da 1ª divisão das cinco principais ligas da CONMEBOL, durante as temporadas de 2017 a 2022. Analisou-se 55 variáveis técnico-táticas, coletados no InStat Scout®. Para a análise dos dados, utilizou-se Kolmogorov-Smirnov, Teste de Levene, Box Teste, MANOVA e *eta* parcial ao quadrado. Identificou-se 7 ICD ofensivos e 1 ICD defensivo discriminante entre vitória e derrota. Conclui-se que, as ações ofensivas a passe, finalizações, eficiência no ataque posicional e contra-ataques, discriminam a vitória nas 5 principais ligas da CONMEBOL.

Introdução

Pouco se sabe sobre os indicadores-chave de desempenho (ICD) no futebol sul-americano. As principais ligas nacionais da 1ª divisão da CONMEBOL são as ligas do Brasil, Argentina, Chile, Paraguai e Uruguai [IFFHS, 2021]. Porém, a literatura científica a respeito das ligas da CONMEBOL é incipiente. Os ICD são métricas quantificáveis que permitem compreender o desempenho de uma equipe e analisar os aspectos contribuintes para a vitória da equipe na partida [HEROLD *et al.*, 2021].

Objetivo

Identificar quais ICD técnico-tático discriminam a vitória e a derrota na 1ª divisão das cinco principais ligas da CONMEBOL.

Materiais e Métodos

Trata-se de um estudo de caso longitudinal. Foram analisadas o total de 5.536 partidas da 1ª divisão das cinco principais ligas da CONMEBOL, durante as temporadas de 2017 a 2022. Foram excluídos 2.276 partidas com resultado final de empate. Analisou-se 55 variáveis de desempenho técnico-táticos ofensivos ($n=40$) e defensivos ($n=15$). A coleta dos dados foi realizada através do InStat Scout®. Os procedimentos estatísticos utilizados foram: Kolmogorov-Smirnov, Teste de Levene, Box Teste e MANOVA ($p < 0,05$). O tamanho de efeito das diferenças foi calculado por meio do *eta* parcial ao quadrado: efeito pequeno: $0.02 < h^2p < 0.13$; efeito médio $0.13 < h^2p < 0.26$; efeito grande $h^2p > 0.26$ [Pierce, Block e Aguinis, 2004].

Resultados

Tabela 1. ICD ofensivos e defensivo para as 5 principais ligas da CONMEBOL

	VITÓRIA		DERROTA		<i>P</i> valor	Tamanho de efeito
	Média	Desvio Padrão	Média	Desvio Padrão		
OFENSIVOS						
Chances	5.9	2.7	3.8	2.3	<0.001*	0.144
Passes Chave	6.9	3.8	5.0	3.2	<0.001*	0.065
Chutes	12.5	4.5	11.0	4.5	<0.001*	0.026
Chutes no Alvo	43.7	14.9	32.0	15.7	<0.001*	0.117
Passes Chave com sucesso	3.4	2.2	2.0	1.7	<0.001*	0.099
Eficiência em Ataque posicional	9.9	4.7	8.6	4.2	<0.001*	0.023
Eficiência em Contra-ataques	17.3	10.6	12.4	9.8	<0.001*	0.053
DEFENSIVO						
Bolas interceptadas	48.3	12.6	44.6	11.4	<0.001*	0.023

Discussão

Os resultados encontrados sobre ICD estão de acordo com estudos prévios [Jacinto *et al.*, 2024; Lago-Peñas *et al.*, 2011]. Entretanto, destaca-se que, as diferenças encontradas com tamanho de efeito pequeno podem não representar implicações no contexto prático.

Apresentar maior volume e eficácia das finalizações aumentam as chances de gols [Liu *et al.*, 2016] e o número de chutes e chutes no gol discriminam equipes vencedoras [Lago-Peñas *et al.*, 2011]. Ataques posicionais são ações ofensivas mais lentas, utilizando mais passes curtos [Lago-Peñas *et al.*, 2017]. Contra-ataques são mais eficazes e aumentam as chances de vitória na partida [Sarmiento *et al.*, 2018]. Passes chaves estão relacionados com a vitória na CONMEBOL [Jacinto *et al.*, 2024]. Bolas interceptadas é importante para retomada da posse e contribuí para a vitória [Almeida *et al.*, 2014].

Conclusão

Neste estudo da 1ª divisão das cinco principais ligas da CONMEBOL, as ações ofensivas relacionadas a passe, finalizações, eficiência no ataque posicional e contra-ataques, apresentaram diferenças com tamanho de efeito pequeno e discriminam a vitória nas 5 principais ligas da CONMEBOL.

Referências

- Almeida, C. H., et al. (2014). Effects of match location, match status and quality of opposition on regaining possession in UEFA champions league. *Journal of Human Kinetics*, v. 41, n. 1, p. 203–214.
- Herold, M. et al. (2021). Attacking key performance indicators in soccer: Current practice and perceptions from the elite to youth academy level. *Journal of Sports Science and Medicine*, v. 20, n. 1, p. 158–169, 1 mar.
- Jacinto, A. L. B., et al., (2024). Identification of indicators that predict victory in the five main CONMEBOL and UEFA leagues. *International Journal of Sports Science & Coaching*, p. 1–12.
- Lago-Peñas, et al., (2011). Differences in performance indicators between winning and losing teams in the UEFA Champions League. *Journal of Human Kinetics*, v. 27, n. 1, p. 135–146, 1 mar.
- Lago-Peñas, C. et al., (2017). Styles of play in professional soccer: an approach of the Chinese Soccer Super League. *International Journal of Performance Analysis in Sport*, v. 17, n. 6, p. 1073–1084, 2 nov.
- Liu, H., et al. (2016). A. Modelling relationships between match events and match outcome in elite football. *European Journal of Sport Science*, v. 16, n. 5, p. 516–525, 3 Jul.
- Sarmento, H., et al. (2017). Influence of tactical and situational variables on offensive sequences during elite football matches. *Journal of Strength and Conditioning Research*, 32(8), 2331-2339.

Quais Indicadores-Chave de desempenho discriminam equipes que vencem e perdem na Bundesliga (2014 a 2022)?

**André L. B. Jacinto¹, Matheus O. Silva¹, Guilherme P. Sousa², Filipe S. Guaceroni¹,
Varley T. Costa¹**

¹ UFMG SOCCER SCIENCE CENTER – Escola de Educação Física, Fisioterapia e Terapia Ocupacional (EEFFTO/UFMG). Av. Pres. Antônio Carlos, 6627 – Pampulha - Belo Horizonte (MG) - CEP 31270-901.

² Chair of Performance Analysis and Sports Informatics, Technical University of Munich, Munich, Germany.

**matolsilva@gmail.com, andreluiz.braga@hotmail.com,
guilherme.gsp11@hotmail.com, filipe.guaceroni@gmail.com,
varleycosta@gmail.com.**

Abstract

The study investigated which performance indicators discriminate between winning and losing teams in the Bundesliga from 2014 to 2022. The sample consisted of 1,843 matches. Fifty-six technical-tactical performance variables were analysed, collected using InStat Scout®. Statistical procedures included the Kolmogorov-Smirnov test, One-way ANOVA, and Cohen's d. The study identified nine offensive technical-tactical KPIs that discriminate between winners and losers in Bundesliga matches. No defensive KPIs were found to discriminate between winners and losers in the Bundesliga. It was concluded that winning teams in Bundesliga matches demonstrated moderate to high effect size differences for nine offensive technical-tactical variables.

Resumo

O estudo investigou quais indicadores de desempenho discriminam equipes que vencem e equipes que perdem na Bundesliga (2014 a 2022). A amostra foi composta por 1.843 jogos. Foram analisadas 56 variáveis de desempenho técnico-táticas, coletados do InStat Scout®. Procedimentos estatísticos: Kolmogorov-Smirnov, Anova One-way e D de Cohen. Identificou-se 9 ICD técnico-táticos ofensivos discriminante entre vencedores e perdedores na partida na Bundesliga. Não foram encontrados ICD defensivos discriminantes entre vencedores e perdedores na Bundesliga. Conclui-se que as equipes vencedoras nos jogos da Bundesliga apresentaram diferenças com tamanho de efeito moderado/alto para 9 variáveis ofensivas técnico táticas.

Introdução

Os indicadores chave de desempenho (ICD) auxiliam na compreensão do sucesso do jogo e têm sido utilizados com frequência com objetivo de estudar próprios atletas e adversários [HEROLD *et al.*, 2021]. A Bundesliga é a única liga europeia disputada por dezoito equipes e com interrupção no calendário esportivo em virtude das baixas temperaturas. Estas particularidades contribuem para uma maior importância de cada confronto na Bundesliga [Praça *et al.*, 2023]. Compreender quais são os parâmetros técnico-táticos que discriminam as vitórias no futebol alemão é o foco deste estudo longitudinal.

Objetivo

Investigar quais ICD técnico-táticos discriminam equipes vencedoras e perdedoras na Bundesliga entre as temporadas de 2014 a 2022.

Materiais e Métodos

Trata-se de um estudo de caso longitudinal. Foram analisadas o total de 1.843 partidas da 1ª divisão da Bundesliga (2014 a 2022). Foram excluídos 605 partidas com resultados de empate. Analisou-se 56 variáveis de desempenho técnico-táticos ofensivos ($n=41$) e defensivos ($n=15$). A coleta dos dados foi realizada através do InStat Scout®. Os procedimentos estatísticos utilizados foram: Kolmogorov-Smirnov, Anova One-Way e D de Cohen ($p < 0.05$). As comparações pareadas foram testadas com o D de Cohen: efeito mínimo ($0,2 < d < 0,5$), efeito moderado ($0,5 < d < 0,8$) e efeito forte ($d > 0,8$) [Cohen, 1998].

Resultados

Tabela 1. ICD ofensivos e defensivo para as 5 principais ligas da UEFA

	VITÓRIA		DERROTA		<i>P</i> valor	D de Cohen
	Média	Desvio Padrão	Média	Desvio Padrão		
OFENSIVOS						
Gols	2.6	±1.2	0.6	±0.7	<0.001*	1.963
Chances	6.4	±3.6	3.7	±2.6	<0.001*	0.849
Eficiência nas Chances	37.9	±21.9	15.5	±21.9	<0.001*	1.022
Finalizações no gol	6.1	±2.6	3.6	±2.2	<0.001*	1.029
Passes Chave com sucesso	4.3	±2.6	2.4	±1.8	<0.001*	0.832
Finalizações	13.4	±4.8	10.5	±4.5	<0.001*	0.614
Eficiência dos chutes no gol	48.5	±16.0	35.1	±17.6	<0.001*	0.795
Passes chave	9.0	±4.6	6.1	±3.6	<0.001*	0.692
Eficiência dos Contra-ataques	17.5	±10.8	12.2	±9.6	<0.001*	0.513

Discussão

Os resultados encontrados sobre ICD estão de acordo com estudos prévios na Bundesliga [Praça et al., 2023; Baskaya et al., 2023]. Destaca-se que, as diferenças encontradas apresentaram tamanho de efeito forte a moderado e representam importantes implicações no contexto prático.

Tendo em vista que o gol é o principal objetivo do jogo, nossos dados contribuem que a capacidade de marcar gols aumenta a probabilidade de vitória [Casal *et al.*, 2021]. Apresentar maior número de finalizações e chances de gols está fortemente ligado ao aumento das chances de vitória [Collet, 2013]. Melhor desempenho em passes certos discriminaram equipes bem-sucedidas na Bundesliga [Praça et al., 2023]. Também foi demonstrado que contra ataques geram número de gols importantes para times que vencem [Li & Zhao, 2021] e são mais eficientes para marcar gols [Tenga *et al.*, 2010].

Conclusão

Conclui-se que na 1ª divisão das Bundesliga (2014 a 2022), que as equipes vencedoras, finalizavam mais a gols. Em função de criarem mais chances de gol e serem eficientes marcavam mais gols. Destaca-se também a eficiência nos passes chaves (assistências) e nos contra-ataques como ações importantes para a vitória.

Referências

- BAŞKAYA, G (2023). What Are The Key Performance Indicators for A Top Position in the German Bundesliga? 2022-2023 Bundesliga Season Example. *Journal of ROL Sports Sciences Cilt*, n. 1, p. 184–199.
- Casal, C. A, et al. (2021). Multivariate Exploratory Comparative Analysis of LaLiga Teams: Principal Component Analysis. *Int J Environ Res Public Health*, v.18, n.6, p.3176.
- Cohen, J. *Statistical power analysis for the behavioural sciences* (2nd ed.), 1988. Hillsdale, NJ: Lawrence Earlbaum Associates.
- Collet, C. (2013). The possession game? A comparative analysis of ball retention and team success in European and international football, 2007–2010. 2013. *Journal of Sports Sciences*, v.31, n.2, p.123–136.
- Herold, M. et al. (2021). Attacking key performance indicators in soccer: Current practice and perceptions from the elite to youth academy level. *Journal of Sports Science and Medicine*, v. 20, n. 1, p. 158–169, 1 mar.
- Li, C., Zhao, Y. (2021). Comparison of Goal Scoring Patterns in “The Big Five” European Football Leagues. *Frontiers in Psychology*.
- Praça, G. M. (2023). What are the key performance indicators related to winning matches in the German Bundesliga? *International Journal of Performance Analysis in Sport*, v. 23, n. 4, p. 284–295.
- Tenga, A. et al., (2010). Effect of playing tactics on achieving score-box possessions in a random series of team possessions from Norwegian professional soccer matches. *Journal of Sports Sciences*, v. 28, n. 3, p. 245–255

Quais indicadores-chaves de desempenho técnico-táticos discriminam a vitória e a derrota na primeira divisão no Big Five Europeu?

André L. B. Jacinto¹, Guilherme P. Sousa², Filipe S. Guaceroni¹, Varley T. Costa¹

¹ UFMG SOCCER SCIENCE CENTER – Escola de Educação Física, Fisioterapia e Terapia Ocupacional (EEFFTO/UFMG). Av. Pres. Antônio Carlos, 6627 – Pampulha - Belo Horizonte (MG) - CEP 31270-901.

² Chair of Performance Analysis and Sports Informatics, Technical University of Munich, Munich, Germany.

`andreluiz.braga@hotmail.com, guilherme.gsp11@hotmail.com,
filipe.guaceroni@gmail.com, varleycosta@gmail.com.`

Abstract

The study aimed to identify which technical-tactical Key Performance Indicators (KPIs) discriminate between victory and defeat in the top division of the five major UEFA leagues. A total of 6,753 matches from the top divisions of the five major UEFA leagues were analysed during the 2017 to 2022 seasons. Fifty-five technical-tactical performance variables were investigated, collected using InStat Scout[®]. Statistical procedures included the Kolmogorov-Smirnov test, Levene's test, Box's test, MANOVA, and partial eta squared. The study revealed nine offensive technical-tactical KPIs that discriminate between victory and defeat. It was concluded that nine offensive technical-tactical variables discriminate victory within the Big Five European leagues.

Resumo

O estudo objetivou identificar quais indicadores-chave de desempenho (ICD) técnico-tático discriminam a vitória e a derrota na 1ª divisão das cinco principais ligas da UEFA. Foram analisadas o total de 6.753 partidas da 1ª divisão das cinco principais ligas da UEFA, durante as temporadas de 2017 a 2022. Investigou-se 55 variáveis de desempenho técnico-táticos, coletados no InStat Scout[®]. Procedimentos estatísticos: Kolmogorov-Smirnov, Teste de Levene, Box Teste, MANOVA e *eta* parcial ao quadrado. Revelou-se 9 ICD técnico-tático ofensivos discriminantes entre vitória e derrota. Conclui-se que, 9 variáveis técnico-táticos ofensivas discriminam a vitória dentro das 5 ligas no Big Five europeu.

Introdução

O “*Big Five*” é formado pelas 5 principais ligas nacionais europeia, sendo elas: a Bundesliga (Alemanha), La Liga (Espanha), Ligue 1 (França), Premier League (Inglaterra) e Serie A (Itália). Devido ao *Big Five*, o futebol europeu é uma das áreas mais lucrativas para o esporte [Li, C e Zhao, Y, 2021]. Para a investigação de indicadores-chave de desempenho (ICD) no futebol, é necessário uma quantidade substancial de dados [Carling, 2013]. Entretanto, estudos longitudinais investigando o Big Five são escassos.

Objetivo

Identificar quais ICD técnico-tático discriminam a vitória e a derrota na 1ª divisão das cinco principais ligas da UEFA.

Materiais e Métodos

Trata-se de um estudo de caso longitudinal. Foram analisadas o total de 6.753 partidas da 1ª divisão das cinco principais ligas da UEFA, durante as temporadas de 2017 a 2022. Foram excluídos 2.205 partidas com resultados de empate. Analisou-se 55 variáveis de desempenho técnico-táticos ofensivos ($n=40$) e defensivos ($n=15$). A coleta dos dados foi realizada através do InStat Scout®. Os procedimentos estatísticos utilizados foram: Kolmogorov-Smirnov, Teste de Levene, Box Teste e MANOVA ($p<0,05$). O tamanho de efeito das diferenças foi calculado por meio do *eta* parcial ao quadrado: efeito pequeno: $0.02 < h^2p < 0.13$; efeito médio $0.13 < h^2p < 0.26$; efeito grande $h^2p > 0.26$ [Pierce, Block e Aguinis, 2004].

Resultados

Tabela 1. ICD ofensivos e defensivo para as 5 principais ligas da UEFA

	VITÓRIA		DERROTA		<i>P</i> valor	Tamanho de efeito
	Média	Desvio Padrão	Média	Desvio Padrão		
OFENSIVOS						
Chances	6.8	3.0	4.1	2.4	<0.001*	0.202
Passes Chave	8.4	4.5	5.6	3.4	<0.001*	0.109
Passes	508.2	138.8	470.1	110.1	<0.001*	0.023
Posse de bola	51.5	10.0	48.4	10.0	<0.001*	0.022
Chutes	13.1	4.8	10.3	4.6	<0.001*	0.080
Chutes no Alvo	45.7	14.8	33.8	17.0	<0.001*	0.123
Passes Chave com sucesso	4.0	2.5	2.3	1.8	<0.001*	0.129
Eficiência em Ataque posicional	10.7	4.9	8.5	4.4	<0.001*	0.051
Eficiência em Contra-ataques	19.1	12.0	13.0	10.4	<0.001*	0.070

Discussão

Os resultados do presente estudo demonstraram quais os ICD técnico-tático ofensivo discriminam as equipes que vencem no Big Five europeu, apesar de apresentarem tamanho de efeito pequeno nas diferenças.

Melhor desempenho em ações relacionadas aos passes e posse de bola discriminaram equipes bem-sucedidas na Bundesliga [Praça et al., 2023], na La Liga [Sousa et al., 2019] e na Premier League [Oberstone, 2009]. Realizar maior número de chutes e chutes no gol também discriminaram equipes vencedoras na La liga [Lago-Peñas *et al.*, 2011]. As ações de contra-ataque criam mais oportunidades e aumentam as chances de gols [Tenga *et al.*, 2010]. Ataques posicionais são ações ofensivas e apresentar maior efetividade ofensivas é determinante para o sucesso [Kite e Nevill, 2017].

Conclusão

Neste estudo de caso da 1ª divisão das cinco principais ligas da UEFA, foram encontradas 9 variáveis técnico-táticas ofensivas que discriminam a vitória nas ligas do Big Five europeu.

Referências

- Carling, C (2013). Interpreting physical performance in professional soccer match-play: Should we be more pragmatic in our approach? *Sports Medicine*, v. 43, n. 8, p. 655–663.
- Kite, C. S.; Nevill, A. (2017). The Predictors and Determinants of Inter-Seasonal Success in a Professional Soccer Team. *Journal of Human Kinetics*, v. 58, n. 1, p. 157–167.
- Lago-Peñas, et al., (2011). Differences in performance indicators between winning and losing teams in the UEFA Champions League. *Journal of Human Kinetics*, v. 27, n. 1, p. 135–146, 1 mar.
- LI, C.; ZHAO, Y. Comparison of Goal Scoring Patterns in “The Big Five” European Football Leagues. *Frontiers in Psychology*, v. 11, 13 jan. 2021.
- Oberstone, J. (2009). Differentiating the Top English Premier League Football Clubs from the Rest of the Pack: Identifying the Keys to Success. *Journal of Quantitative Analysis in Sports*, v. 5, n. 3, 20 Jul.
- Praça, G. M. (2023). What are the key performance indicators related to winning matches in the German Bundesliga? *International Journal of Performance Analysis in Sport*, v. 23, n. 4, p. 284–295.
- Souza, D. B. (2019). A new paradigm to understand success in professional football: analysis of match statistics in LaLiga for 8 complete seasons. *International Journal of Performance Analysis in Sport*, v. 19, n. 4, p. 543–555, 4 Jul.
- Tenga, A. et al., (2010). Effect of playing tactics on achieving score-box possessions in a random series of team possessions from Norwegian professional soccer matches. *Journal of Sports Sciences*, v. 28, n. 3, p. 245–255

Sessão de Best Papers

Análise Espacial de Interceptações no Futebol: A Abordagem Voronoi

Antônio C. N. Neto¹, Bernardo D. Lemos¹, João Lucas S. Moreira¹,
Raphael Aroldo C. Mendes¹

¹Departamento de Ciência da Computação
Universidade Federal de Minas Gerais (UFMG)
Belo Horizonte – MG – Brazil

{antonioneto, bernardolemos, joaomoreira, raphaelmendes}@dcc.ufmg.br

1. Introdução

O objetivo deste estudo é explorar uma abordagem não convencional para análise espacial, utilizando uma interpretação geométrica do campo de futebol e dos jogadores. A ideia é extrair features relevantes para a predição de interceptação de passes por meio da análise do espaço.

Nessa via, os Diagramas de Voronoi são amplamente usados na literatura para controle de campo e interceptações, como mostrado por Spearman (2016). Nesse sentido, com o crescimento das informações de *tracking*, esses diagramas se tornam mais relevantes para análises geométricas precisas durante o jogo. Além disso, o avanço da computação permite a criação de modelos interpretáveis, como no trabalho dos autores Wang e Veličković (2024).

Portanto, o objetivo formal é construir uma métrica de xInt (Expected Interception), baseada exclusivamente de uma engenharia de features do Diagrama de Voronoi, através de modelos eficientes e interpretáveis.

2. Metodologia

Para o trabalho, foram utilizados dados de *tracking* e *eventos* da PFF (Pro Football Focus), extraídos de jogos da Premier League 2022/2023. Apenas instâncias de passe foram selecionadas, com informações do estado do campo no momento pré-passe. A definição de interceptação adotada foi a da PFF, que também fornece o receptor provável do passe.

3. Features Voronoi

Baseando em uma execução do Diagrama de Voronoi em uma instância do campo pré-passe, visualizado na Figura 1, a seguinte engenharia de features foi realizada, visando adicionar mais contexto espacial para o modelo, totalizando 10 features geradas.

- Área da região do receptor e do passador;
- Ângulo de oportunidade, formado pelas retas tangentes em relação ao passador ao transmissor. A Figura 2 é um exemplo dessa feature;
- Ângulo de oportunidade multiplicado por um fator k ;
- Proporção da área adversário sobre a área total do triângulo formado pelas retas tangentes do ângulo de oportunidade;

- Número de áreas adversárias atravessadas pela bola;
- Distância mínima e média entre os adversários em relação ao passador e receptor.

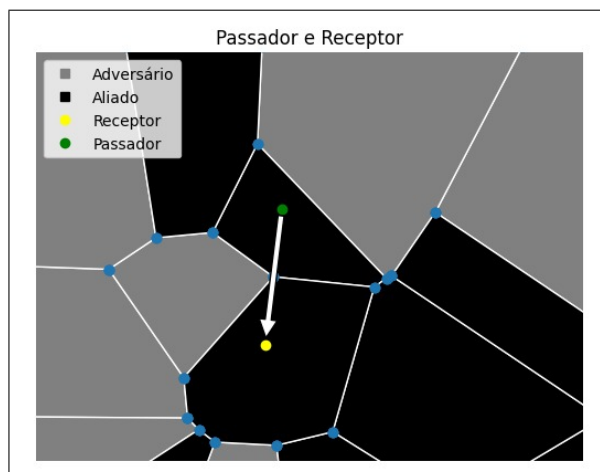


Figura 1. Imagem que mostra o passador, o receptor e suas áreas.

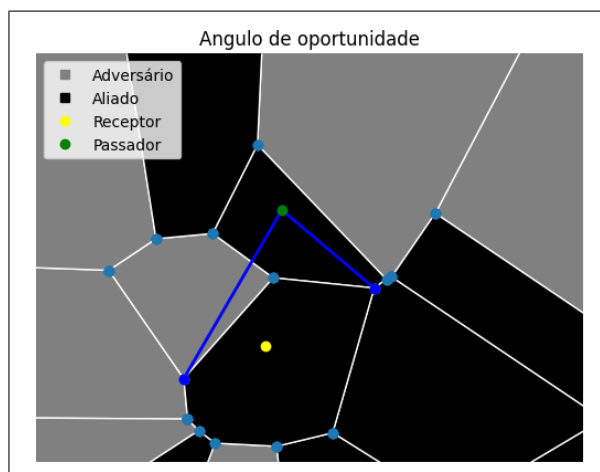


Figura 2. Retas tangentes que formam o ângulo em questão.

4. Resultados

O grupo extraiu as features propostas utilizando dados de *tracking* e eventos de 24 jogos. Foi considerado bloqueios como intercepções, totalizando 18.663 passes, dos quais 2.963 são intercepções, resultando em um notável desbalanceio de classes.

4.1. Modelos

O baseline da análise considera features simples, sendo a posição do receptor e do passador, o deslocamento provável da bola em x e y e a distância entre eles. O modelo utilizado para o baseline foi o RandomForest.

Para os testes comparativos, foram utilizados os modelos XGBoost e Random Forest, que apresentaram melhor performance geral em relação aos modelos MLP testados.

Para resolver o desbalanceamento, utilizou-se *class weights* no XGBoost e, no Random Forest, foram empregadas 100 árvores com a quantidade máxima de features por árvore de $\lfloor \sqrt{F} \rfloor$, sendo F o número de features.

4.2. Curvas de calibração

Para visualizar o âmbito probabilístico do modelo, foram utilizadas curvas de calibração, permitindo comparar o desempenho dos modelos com a reta ideal. Na Figura 3, é possível observar o comportamento do modelo Random Forest e das features simples.

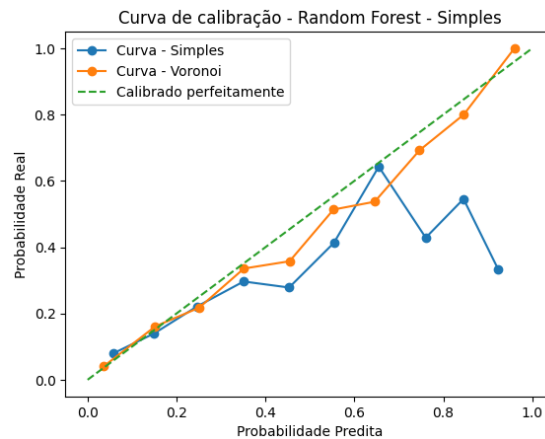


Figura 3. Curva de calibração dos modelos.

Nota-se que o comportamento do modelo Voronoi tende a ser próximo a reta ideal, indicando um bom aprendizado do modelo no viés probabilístico.

4.3. Métricas e Testes Práticos

Abraçando a natureza probabilística do problema, as métricas utilizadas foram o *Brier Score* e o *AUC*. A tabela 1 contém o desempenho para o baseline, XGBoost e RandomForest. De tal forma, o RandomForest é significativamente o melhor modelo, a partir do intervalo de confiança.

Modelo	Brier Score	AUC
Baseline	0.126 +- 0.002	0.673 +- 0.014
XGBoost	0.120 +- 0.005	0.771 +- 0.022
RandomForest	0.113 +- 0.004	0.790 +- 0.020

Tabela 1. Métricas por *cross-validation* de 5 folds com I.C. de 95%

Para visualizar a eficiência do modelo RandomForest, as figuras 4 e 5 mostram o campo pré-passe, com o cálculo do xInt.

4.4. Metrificando importância das features

Para metrificar a importância de cada feature, foi determinado o quanto cada uma reduz a impureza nas árvores do RandomForest, onde quanto maior a redução, mais significativa a feature é. Na Figura 6 é visualizada a estratégia no modelo.

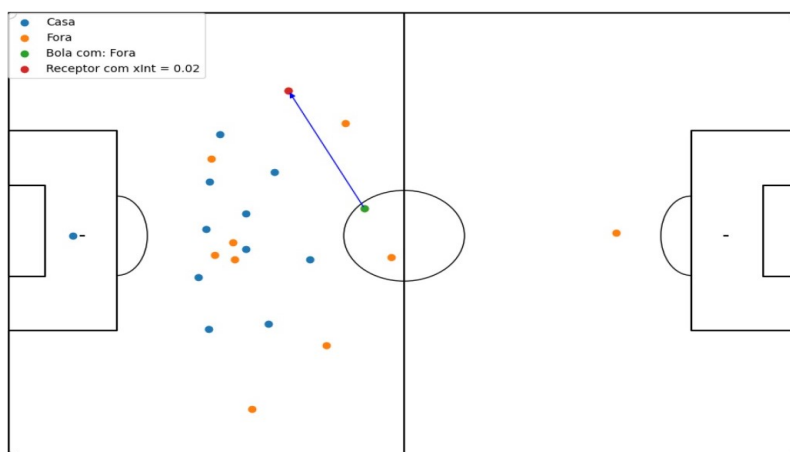


Figura 4. Passe com probabilidade 0.02 de interceptação.

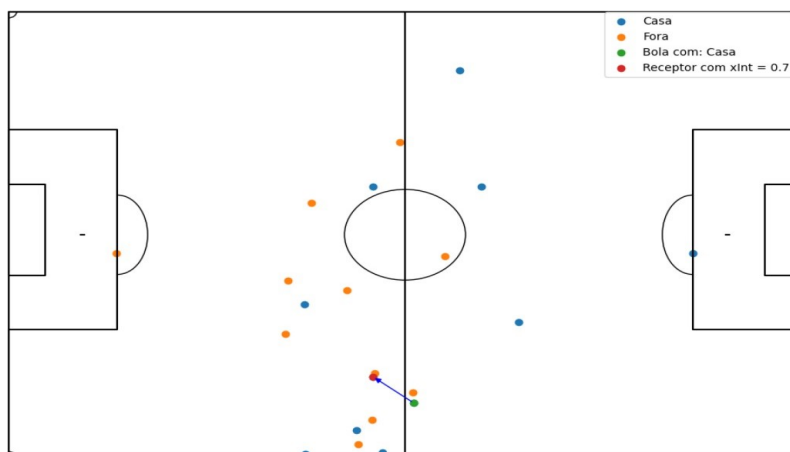


Figura 5. Passe com probabilidade 0.71 de interceptação.

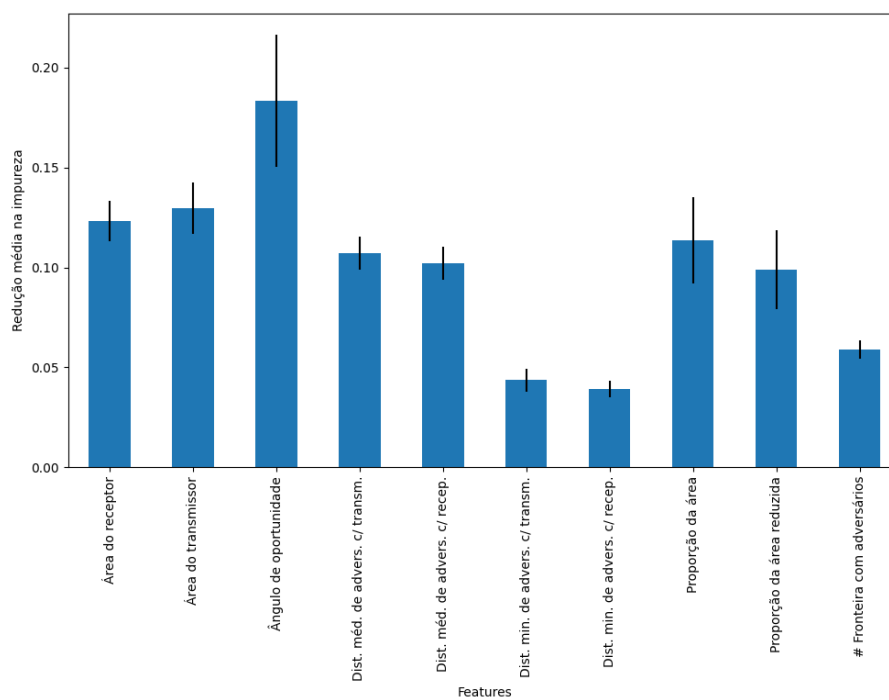


Figura 6. MDI do modelo Random Forest.

Observa-se que o ângulo de oportunidade é a feature significativamente mais importante do modelo, em sequência as features de área das entidades.

5. Conclusão

O estudo, focado em análises espaciais, demonstrou que, apesar das limitações intrínsecas ao passe, a extração de features usando o algoritmo aplicado ao campo de futebol trouxe bons resultados em comparação a um baseline acurado, especialmente nas métricas Brier Score e AUC. Isso sugere que o Voronoi enriquece o contexto espacial.

Além disso, incorporar features como velocidade e aceleração, bem como informações adicionais sobre o passador e características do passe, pode melhorar ainda mais a precisão do modelo.

Referências

[Spearman 2016] Spearman, William. (2016). Quantifying Pitch Control. 10.13140/RG.2.2.22551.93603.

[Wang and Veličković and Hennes 2024] Wang, Z., Veličković, P., Hennes, D. et al. TacticAI: an AI assistant for football tactics. Nat Commun 15, 1906 (2024).

[Ravid Shwartz-Ziv and Amitai Armon 2021] IT AI Group - Tabular Data: Deep Learning Is Not All You Need. November 24, 2021.

Seleção Ótima de Equipes: Uma Abordagem Integrada de IA e Programação Inteira

Letícia Maia Távora^{a,*}, Glauco Fiorott Amorim^a, Pedro Henrique González^b

^aPPCIC - Centro Federal de Educação Tecnológica Celso Suckow da Fonseca,
CEFET/RJ, Rio de Janeiro, RJ, Brazil

^bPESC - Universidade Federal do Rio de Janeiro, UFRJ, Rio de Janeiro, RJ, Brazil

Abstract

A composição estratégica de equipes esportivas é um desafio crucial para otimizar o desempenho em competições. Neste estudo, propomos um modelo de formação de equipes que combina métodos de previsão e otimização, focado em antecipar o desempenho futuro dos jogadores e selecionar as melhores escalas possíveis. Utilizando dados históricos do Campeonato Brasileiro, no contexto do Cartola FC de 2020, como base de validação, foram aplicados métodos de previsão como Médias Móveis, ARMA e Redes Neurais, em combinação com uma técnica de Programação Inteira. Este modelo leva em consideração restrições de custo e posição, replicando condições que poderiam ser aplicadas ao mundo real. A eficácia da abordagem foi validada por meio de simulações, indicando que a metodologia pode ser eficaz na otimização de formações reais, oferecendo um processo sistemático para decisões mais estratégicas na composição de equipes.

Keywords: Composição de Equipes, Inteligência Artificial, Otimização Combinatória, Séries Temporais.

1. Introdução

A definição de quais jogadores farão parte da equipe em um determinado jogo é uma das tarefas mais desafiadoras para os técnicos nos dias atuais.

*Corresponding author.

Email addresses: leticia.tavora@cefet-rj.br (Letícia Maia Távora),
glauco.amorim@cefet-rj.br (Glauco Fiorott Amorim), pegonzalez@cos.ufrj.br
(Pedro Henrique González)

Várias são as informações que podem auxiliar na escolha de um determinado jogador, como: dados fisiológicos, desempenho nos treinos e dados coletados em partidas anteriores. Essas informações podem ser agrupadas e, a partir desse agrupamento, servirem como referencial para estabelecer um histórico de desempenho de cada jogador, o que influencia diretamente no desempenho da equipe.

Portanto, o objetivo deste trabalho é usar Redes Neurais para tentar prever a melhor escalação possível de uma equipe esportiva baseado no histórico de desempenho individual dos atletas. Como esse histórico é uma informação sigilosa das equipes, utilizou-se um *Fantasy Game* como prova de conceito. Este *Fantasy Game* é baseado no futebol brasileiro e estabelece o desempenho do atleta utilizando informações de ações que ocorreram em partidas anteriores.

Espera-se que esta análise possa auxiliar o corpo técnico de equipes esportivas facilitando a tomada de decisão no momento da escolha dos jogadores para as partidas.

2. Metodologia

Neste estudo, propomos o uso do Cartola FC, um *Fantasy Game* baseado no futebol brasileiro, como uma ferramenta de simulação para composição real de equipes utilizando uma abordagem de previsão e otimização para obter a melhor escalação possível. O problema é formulado como uma otimização mono-objetivo, buscando a melhor combinação de jogadores dentro das restrições de orçamento e regras do jogo.

O processo inicia com a preparação dos dados das rodadas anteriores, que inclui desempenho individual dos jogadores, eventos em campo e estatísticas. Esses dados são analisados usando três métodos de previsão: Médias Móveis Simples (MMS), ARMA (*AutoRegressive Moving Average*) e Redes Neurais(LSTM). A MMS suaviza flutuações de curto prazo, o ARMA modela dependências temporais complexas e as Redes Neurais, como a LSTM (Long Short-Term Memory), capturam padrões não lineares e dependências de longo prazo nos dados. As previsões obtidas são usadas para selecionar a melhor combinação de jogadores, respeitando as restrições de orçamento e posições. [3]

O modelo de otimização é formulado como um problema de programação matemática, onde o objetivo é maximizar a pontuação total da equipe formada no Cartola FC, após a previsão do desempenho. As variáveis de decisão

incluem a escolha dos jogadores e o esquema tático, com restrições que garantem que o custo total não exceda o orçamento disponível (Cartoletas), que a quantidade de jogadores por posição respeite o esquema tático, e que cada jogador seja selecionado para no máximo uma posição. Cada restrição pode ser adaptada a situações reais.

O *Framework* proposto é cíclico: após cada rodada, os dados são atualizados e o modelo é reavaliado, ajustando a seleção de jogadores conforme o valor de Cartoletas disponível e o desempenho recente. A integração contínua de métodos de previsão, otimização e atualização dinâmica permite uma abordagem adaptativa, maximizando a adaptabilidade e sucesso nas escalações.

3. Fundamentação Teórica

As médias móveis (MA - *Moving Averages*) são amplamente utilizadas em séries temporais para suavizar flutuações e identificar tendências subjacentes. A principal vantagem das médias móveis é sua simplicidade, proporcionando uma visão clara das mudanças de desempenho ao eliminar ruídos temporais. Ao suavizar as pontuações dos jogadores, as médias móveis auxiliam na identificação de tendências consistentes ou declínios de forma, facilitando a escolha de jogadores para futuras rodadas com base em seu desempenho recente.[4]

Os modelos ARMA (*AutoRegressive Moving Average*) combinam dois componentes: autoregressão (AR) e médias móveis (MA). A parte AR captura a dependência linear entre uma variável e suas versões defasadas, enquanto a parte MA modela o erro da previsão como uma combinação linear de erros passados. Ao aplicar modelos ARMA ao Cartola FC, busca-se prever o desempenho futuro com base em padrões históricos, levando em consideração a autocorrelação inerente às séries temporais de pontuação dos jogadores. [1]

As LSTMs (*Long Short-Term Memory*), por sua vez, são eficazes na previsão de séries temporais, pois conseguem capturar padrões complexos e dependências de longo prazo. Elas utilizam mecanismos de portas que controlam o que deve ser lembrado ou esquecido ao longo do tempo, o que é crucial em esportes como o futebol, onde o desempenho de um jogador pode ser influenciado por diversos fatores acumulados.[5]

4. Manipulação dos Dados

A análise do desempenho dos jogadores no Campeonato Brasileiro começa com a obtenção e pré-processamento dos dados do repositório GitHub 'caR-

tola' [2]. Este repositório já coletou e pré-processou os dados. Antes da análise, os dados passam por um pré-processamento para garantir consistência e qualidade, incluindo a verificação de valores ausentes, correção de erros e padronização de formatos.

5. Testes e Resultados

Projetando no Cartola FC um possível cenário real do futebol resultante do nosso *Framework*, utilizando Médias Móveis, ARMA e LSTM para prever o desempenho das equipes, observou-se que os modelos conseguem aproximar as previsões dos desempenhos reais, demonstrando como cada método se comporta ao longo das rodadas. Para testar a eficácia desses métodos, comparamos os resultados com um cenário ideal, onde as previsões correspondem exatamente aos desempenhos reais dos jogadores. Este cenário ideal serve como um *Benchmark* para medir a precisão dos modelos.

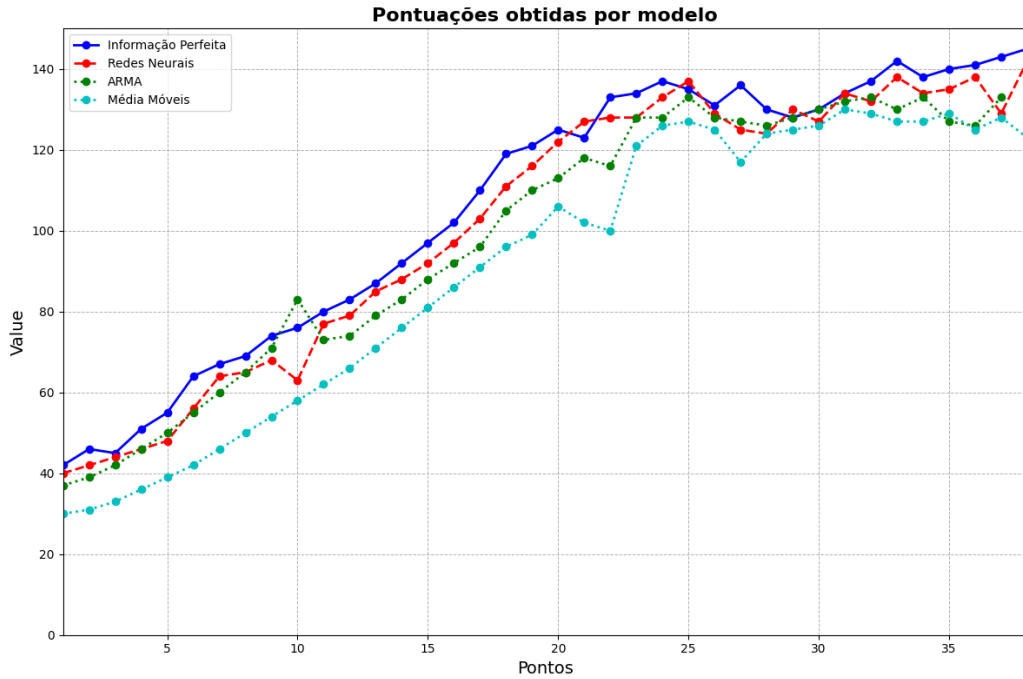


Figura 1: Pontuações das equipes

A análise revelou que a maior diferença entre as atuações obtidas pelos métodos foi de 25 pontos, destacando-se na rodada 21, quando o método uti-

lizando Redes Neurais alcançou 127 pontos, enquanto o método de Médias Móveis registrou apenas 102 pontos. Esse foi o maior desvio entre os métodos avaliados. Em contraste, na rodada 23, as pontuações dos métodos foram mais próximas, com o ARMA igualando ao valor resultando do método LSTM, evidenciando a similaridade dos métodos.

Além disso, durante todo o estudo, o método ARMA demonstrou uma capacidade superior de adaptação às mudanças bruscas no desempenho dos jogadores entre rodadas. Por outro lado, no mesmo período, o método LSTM apresentou resultados mais próximos do real valor da *Performance* dos jogadores, como pode ser analisado na figura 1.

6. Conclusão

Os resultados obtidos neste estudo demonstram claramente a superioridade do modelo LSTM em comparação com os outros métodos avaliados. Ao analisar o desempenho dos métodos ao longo das rodadas do Cartola FC, observamos que este foi capaz de capturar com maior precisão os padrões temporais complexos presentes nos dados históricos, resultando em previsões mais assertivas. Por outro lado, o método de Médias Móveis teve o pior desempenho, com uma precisão de cerca de 60%. Isso evidencia a limitação desse método em capturar as flutuações dinâmicas do desempenho dos jogadores ao longo do tempo, o que resultou em previsões menos precisas e, consequentemente, em pontuações totais mais baixas.

Referências

- [1] Brockwell, P.J., Davis, R.A., 2016. Introduction to Time Series and Forecasting. Springer.
- [2] Gomide, H., Gualberto, A., 2022. CaRtola: Extração de dados da API do CartolaFC, análise exploratória dos dados e modelos preditivos em R e Python. URL: <https://github.com/henriquepgomide/caRtola>.
- [3] Hansun, S., 2013. A new approach of moving average method in time series analysis, in: 2013 Conference on New Media Studies (CoNMedia), pp. 1–4. doi:10.1109/CoNMedia.2013.6708545.
- [4] Simão, T.D., Laroche, R., des Combes, R.T., 2020. Safe policy improvement with an estimated baseline policy. *arXiv:1909.05236*.

- [5] Wan, A., Chang, Q., AL-Bukhaiti, K., He, J., 2023. Short-term power load forecasting for combined heat and power using cnn-lstm enhanced by attention mechanism. *Energy* 282, 128274. URL: <https://www.sciencedirect.com/science/article/pii/S0360544223016687>, doi:<https://doi.org/10.1016/j.energy.2023.128274>.

Ferramenta VAR e seu impacto no desempenho da Arbitragem Paulista

Bernardo Martins¹

¹Departamento científico e Desenvolvimento de Dados da Arbitragem -
Federação Paulista de Futebol (FPF), São Paulo, Brasil

{bernardo.martins@fpf.org.br}

Abstract. *The study aims to understand the impact of VAR in the 104 games of the 2024 PaulistãoA1, highlighting the correction of clear and obvious errors, based on transparency and efficiency. The initial (without VAR) and final (with VAR) decisions of the referee were statistically compared, via the McNemar test. The results indicate that the use of VAR significantly increases the assertiveness of protocol decisions, with an odds ratio of 15.5 and a prevalence ratio of 4.97. In addition, there is a VAR impact in 58 seconds/match, contributing to sports justice without significant delays. It is concluded that VAR assisted in decisions, increasing accuracy.*

Resumo. *O estudo visa compreender o impacto do VAR nos 104 jogos do PaulistãoA1 2024, destacando correção de erros claros e óbvios, alicerçado em transparência e eficiência. Comparou-se estatisticamente as decisões iniciais(semVAR) e finais(comVAR) do árbitro, via teste de McNemar. Os resultados indicam que a utilização do VAR aumenta significativamente a assertividade das decisões protocolares, com razão de chance 15,5 e razão de prevalência de 4,97. Além disso, há impacto VAR em 58segundos/partida, contribuindo à justiça esportiva sem atrasos significativos. Conclui-se que o VAR auxiliou nas decisões, elevando acertos.*

1. Introdução

A ferramenta Árbitro Assistente de Vídeo(VAR) é discutida entre especialistas e amantes do futebol([Oliveira et al. 2020]).Este artigo avalia a efetividade do VAR, mensurando impacto das ocorrências protocolares nos 104 jogos do PaulistãoA1 2024, via protocolo institucional interno.

Segundo [FPF 2021], os árbitros tomam inúmeras decisões na partida, com ou sem infração. Para maior justiça pela filosofia de *mínima_interferência-máximo_benefício*, a ferramenta visa auxiliar *erros_claros_e_óbvios* ou *incidentes_graves_perdidos*, em quatro situações:

- gols ;
- pênaltis ;
- cartões vermelhos ;
- confusão de identidade.

O VAR, que verifica todos os incidentes protocolares, é supervisionado pelo Observador, que traz informações e dados sobre as ocorrências referentes ao VAR e seu desempenho, que servem de análise dos acertos da arbitragem.

2. Metodologia

Abordagem quantitativa e exploratória, mediado pela análise documental com estatística descritiva sobre quantidade, tempo, impacto e acerto da arbitragem no PaulistãoA1-2024, referente ao VAR.

2.1. Teste de McNemar

O *Teste McNemar* é utilizado para comparar proporções em dados pareados, ou seja, em situações onde amostras são avaliadas em condições distintas, como as decisões com/sem VAR. Sendo adequado para dados binários, que classificam os resultados em *Acerto* e *Erro*.

Neste estudo, os dados pareados foram organizados(Tabela 1):

Table 1. Contingência mostrando as contagens n_{ij} para acertos/erros com/sem VAR

	Sem VAR: Acerto	Sem VAR: Erro	Total
Com VAR: Acerto	n_{11}	n_{12}	$n_{1.}$
Com VAR: Erro	n_{21}	n_{22}	$n_{2.}$
Total	$n_{.1}$	$n_{.2}$	N

Sendo, p_1 e p_2 probabilidades de *Acerto* com e sem a ferramenta, respectivamente,o teste é a Equação 1

$$H_0 : p_1 = p_2 \quad \text{versus} \quad H_1 : p_1 \neq p_2 \tag{1}$$

Como apenas os pares discordantes (n_{12} e n_{21}) ofertam informações divergentes entre as condições, o teste foca nesses casos. Sob hipótese nula(H_0),espera-se que as

discordâncias observadas entre os grupos sejam aleatórias e que a quantidade de discordâncias seja igualmente distribuída entre ambos.

A estatística de teste com correção de continuidade para dados discretos (Equação 2):

$$X_{McN}^2 = \frac{(|n_{12} - n_{21}| - 1)^2}{n_{12} + n_{21}} \quad (2)$$

O X_{McN}^2 segue distribuição qui-quadrado com 1 grau de liberdade(gl). Assim, rejeita-se H_0 , se $X_{McN}^2 > \chi_{1,1-\alpha}^2$. Onde, $\chi_{1,1-\alpha}^2$ representa o percentil de ordem $1 - \alpha$ da qui-quadrado, com 1 gl.

2.2. Odds Ratio

Após teste, é interessante quantificar a magnitude da diferença entre condições pareadas. A razão de chances(*odds ratio*) compara as probabilidades relativas de sucesso entre os dois grupos, pares discordantes. OR é calculado como a razão entre as discordâncias(Equação 3).

$$OR = \frac{n_{12}}{n_{21}} \quad (3)$$

Sendo, n_{12} número de pares em que a decisão comVAR obteve sucesso e semVAR falhou, e n_{21} número de pares em que a condição comVAR falhou e a condição semVAR obteve sucesso. Ofertando interpretação quantitativa para entender a magnitude da divergência comparativa([Agresti 2007]).

2.3. Razão de prevalência

Razão de prevalência([Thompson et al. 1998]) também é de interesse pois auxilia na comparação entre as proporções de acerto com/semVAR, para entender o quão impactante é (Equação 4).

$$RP = \frac{\text{Prevalência comVAR}}{\text{Prevalência semVAR}}. \quad (4)$$

3. Resultados

Para os cálculos, 9 observadores foram os responsáveis por analisar os 104 jogos, informando 621 ocorrências(580 checagens e 41 revisões), aproximadamente 6 por partida, em média. Sendo 282 gols, 196 pênaltis, 142 cartões vermelhos e 1 por erro de identificação.

3.1. Ocorrências com Impacto

Impactaram em tempo 87 ocorrências, menos de 1 por jogo. Com 58segundos de média, dos jogos com impacto, vide Tabela 2. 42 segundos em checagens e 86 em revisão.

Table 2. Ocorrências com impacto

Ocorrências c/ impacto	Quantidade	Tempo médio	DP
Gol	36	64s	± 34
Pênalti	37	55s	± 29
Cartão Vermelho	13	45s	± 13
Erro de identidade	1	95s	

Observação importante são as 53 linhas de impedimento construídas. Levando em média 50 ± 13 segundos.

3.2. Revisão

Das 41 revisões, 7 *factualis* e 34 *On-Field Review (OFR)*, quando o árbitro rever as imagens. Em 80% houve mudança na decisão inicial do árbitro (Tabela 3).

Table 3. Revisões

Revisão	OFR	Factual
Gol	4	6
Penalti	22	1
Cartão Vermelho	8	-
Erro de identidade	-	-
Total	34	7

3.3. Percentual de acerto

Além do controle de qualidade e tempo, algo que auxilia são os acertos. Com viés comparativo obtém-se informe estratificado de acertos (Tabela 4), comparado com o desempenho em cada situação protocolar com/sem interferência VAR.

Table 4. Acerto final por incidente(%)

Gol	Penalti	Cartão Vermelho	Erro de Identificação
99,6%	98%	97,2%	100%

De forma geral, das 621 ocorrências protocolares, 98,6% foram acertos. Em comparativo, considerando a decisão inicial desses lances, seria 93,9% correto.

3.4. Comparação via Teste de McNemar

A Tabela 5 representa a relação de acerto/erro das ocorrências analisadas com/sem VAR, para analisarmos possíveis diferenças estatísticas entre acertos com/sem a ferramenta, via *teste de McNemar*.

Table 5. Contingência dos Resultados com e sem VAR

Tabela de Contingência	Sem VAR: Acerto	Sem VAR: Erro
Com VAR: Acerto	581	31
Com VAR: Erro	2	7

O p-valor 1.309×10^{-7} do teste exato de McNemar indica que há diferenças nas proporções da acertos protocolares da arbitragem com/sem a utilização da ferramenta VAR.

3.5. Resultado da razão de chances

Através da Tabela 5 é possível mensurar a magnitude da diferença via *odds ratio*, entre pares discordantes $n_{12}=31$ e $n_{21}=2$. $OR= 15.5$ com $IC95\%[3,9440;133,6649]$, que sugere forte associação entre o uso do VAR e o aumento na precisão das decisões protocolares, com intervalo de confiança confirmando significância estatística.

3.6. Resultado Razão de prevalência

Table 6. Acerto e erro das decisões com e sem VAR

Tabela de Acerto/Erro	Acerto	Erro
ComVAR	612	9
SemVAR	583	38

O resultado entre prevalência de acerto comVAR(0,9855) e acerto semVAR(0,9388), 1.0497, indica que ferramenta VAR aumenta prevalência de acertos protocolares em mais de quatro vezes nas decisões tomadas no cenário de estudo.

4. Conclusão

Os resultados reforçam importância do VAR na justiça do Paulistão. A alta taxa de acerto, com e sem VAR, destaca capacidade da arbitragem paulista, desconsiderando magnitude de cada situação. Embora mesmo com altos custos financeiros e condições estruturais insuficientes([Esporte 2018]), a introdução do VAR contribuiu positivamente, aumentando a prevalência de acertos em mais de quatro vezes.

A ferramenta tem um impacto médio menor que uma ocorrência por partida, com média de 58 segundos cada. Esse tempo fortalece a contribuição, também no tempo. Lembrando que qualquer impacto do VAR é acrescido ao jogo.

Através do estudo, o VAR não só aprimorou a justiça esportiva, como elevou nível de eficiência e precisão das decisões arbitrais no PaulistãoA1 2024.

References

Agresti, A. (2007). *An Introduction to Categorical Data Analysis*. Wiley, Hoboken, NJ, 2nd edition.

- Esporte, G. (2018). Quanto custa o VAR? tecnologia pode ser mais barata do que se pensa. Acesso em: 14 de agosto de 2024.
- FPF (2021). *Manual para Árbitros Assistentes de Vídeo (VAR)*. São Paulo, Brasil. Versão 8.0 - Resumo das principais diretrizes, ajustado pela FPF.
- Oliveira, A. C., Schmitz Filho, A. G., dos Santos, B. C., da Silva Machado, B., da Silva, D. D., and Cairrão, M. R. (2020). A nova tecnologia no futebol: diálogos sobre a influência do var. *RBFF-Revista Brasileira de Futsal e Futebol*, 12(47):94–102.
- Thompson, M. L., Myers, J. E., and Kriebel, D. (1998). Prevalence odds ratio or prevalence ratio in the analysis of cross sectional data: what is to be done? *Occupational and Environmental Medicine*, 55(4):272–277.



III Football Analytics Modelling & Experience

5 de setembro de 2024

Belo Horizonte, Brasil

Compêndio de Artigos

Realização

SALab - Sports Analytics Lab
NYU - New York University

Organização

Departamento de Ciência da Computação - DCC - UFMG