

Aplicação de Métodos Baseados em Concept Drift para Previsão de Gols no Futebol Profissional

Ana G. Araújo¹, Matheus Melo¹, Lucas G. Tavares¹, Jorge Soares¹

¹Cefet/RJ – Rio de Janeiro/RJ – Brasil

{ana.araujo.4, matheus.melo, lucas.tavares}@aluno.cefet-rj.br

jorge.soares@cefet-rj.br

Abstract. *Predicting goals in real time is a difficult and relevant task in sports analytics. In this work, we investigate whether unsupervised drift detectors can anticipate football goals. To this end, we model pass-event time series as a virtual drift problem and evaluate Page-Hinkley, KSWIN, and ADWIN on StatsBomb data (La Liga 2015/2016) using an asymmetric SoftED variant that rewards early alarms. Page-Hinkley achieved the highest MCC, suggesting that drift detection is a viable and interpretable alternative for real-time tactical alert systems in football.*

Resumo. *Prever gols em tempo real é uma tarefa difícil e relevante no esporte. Neste trabalho, investigamos se detectores de drift não supervisionados conseguem antecipar gols. Para isso, modelamos séries de passes como um problema de drift virtual e avaliamos Page-Hinkley, KSWIN e ADWIN em dados StatsBomb (La Liga 2015/2016) com uma variante assimétrica do SoftED que valoriza alarmes antecipados. O Page-Hinkley obteve o maior MCC, sugerindo que a detecção de drift é uma alternativa viável e interpretável para alertas táticos em tempo real no futebol.*

1. Introdução

A análise de desempenho no futebol tem evoluído rapidamente com a disponibilização de dados de rastreamento e eventos em alta resolução temporal. Identificar com antecedência situações que culminam no principal objetivo do jogo (o gol) é relevante tanto para sistemas de apoio a treinadores em tempo real, para modelos de predição tática e para o mercado de apostas esportivas [Lang et al., 2025]. Como gols são eventos raros [Sarmiento et al., 2018] e as partidas são compostas por séries de eventos não estacionários, propõe-se a hipótese de que a ocorrência de um gol é precedida por uma mudança no padrão de eventos, configurando um problema de detecção de *drift* virtual [Gama et al., 2014]. Diante disso, a pergunta de pesquisa que orienta este trabalho é: “em que medida detectores de concept drift não supervisionados, aplicados a séries temporais de passes, são capazes de antecipar a ocorrência de gols no futebol profissional?”.

Assim, o objetivo deste trabalho é entender se mudanças na distribuição do comportamento de jogo podem funcionar como alarme preditor para a ocorrência de gols. Avaliamos se padrões temporais e oscilações em métricas de passes, capturadas em janelas de tempo intra-partida e processadas sequencialmente, são capazes de sinalizar antecipadamente um aumento na probabilidade de marcar. Se confirmada, detectores de *drift*

não supervisionados poderiam ser empregados como alertas táticos sem necessidade de dados rotulados históricos.

Nesse contexto, o presente trabalho traz como contribuições: i) formulação do problema de detecção antecipada de gols como detecção de *drift* virtual em séries de passes; ii) adaptação da métrica SoftED para avaliação assimétrica de alarmes antecipados; iii) avaliação empírica de detectores não supervisionados em dados reais de futebol. O restante do artigo discute referencial teórico, trabalhos relacionados, metodologia, resultados e conclusão.

2. Referencial Teórico

Em aprendizado de máquina em *streams*, ocorre *drift* quando a distribuição dos dados muda ao longo do tempo [Gama et al., 2014]. Há dois tipos principais: o *drift* real, quando $P(Y|X)$ se altera (o conceito-alvo muda), e o *drift* virtual (tendo *data drift* como um possível sinônimo), em que apenas $P(X)$ muda sem alterar $P(Y|X)$. A detecção de *concept drift* também pode ser utilizada para detectar eventos em séries temporais, pois mudanças persistentes indicam novos regimes do processo, enquanto anomalias representam desvios transitórios seguidos de retorno ao estado anterior [Hinder et al., 2024; Tavares et al., 2025; Kore et al., 2024].

No contexto deste trabalho, a série de passes por minuto é a representação de X ; o gol é um evento discreto, não um rótulo associado a cada amostra. A hipótese investigada é, portanto, de *drift* virtual: mudanças na distribuição de passes precedem o evento-gol, sem que haja um modelo explícito de $P(\text{gol}|\text{passes})$.

Entre os detectores de *drift*, destacam-se: (1) o *Page-Hinkley (PH)* [Page, 1954], teste sequencial paramétrico que monitora desvios acumulados em relação a uma média de referência e dispara alarmes quando excede um limiar λ , operando em modo *up*, *down* ou *both*; (2) o *KSWIN* [Raab et al., 2020], detector não paramétrico baseado no teste de Kolmogorov-Smirnov entre sub-janelas deslizantes; e (3) o *ADWIN* [Bifet and Gavaldà, 2007], que utiliza janelas adaptativas e detecta mudanças por diferenças estatísticas entre médias. Os três estão disponíveis na biblioteca `River`¹.

Para avaliação de detectores de eventos em séries temporais, Salles et al. [2024] propõem o *SoftED* (Soft Error Detection), que atribui pontuações parciais a alarmes próximos ao evento-alvo em contraposição a coincidências exatas. Os alarmes são classificados em TP, FP, FN e TN com valores fracionários, tornando a avaliação mais adequada a cenários desbalanceados. Neste trabalho, utilizam-se as métricas *recall*, precisão, F1 e MCC (Matthew's Correlation Coefficient).

3. Trabalhos Relacionados

Para localizar a literatura relevante ao tema do trabalho, realizou-se uma revisão *ad hoc* na base Scopus utilizando a string de pesquisa: TITLE-ABS-KEY ((*“sport”* OR *“football”* OR *“soccer”*) AND (*“event prediction”* OR *“goal prediction”* OR *“goal forecast*”* OR *“event forecast*”* OR *“concept drift”* OR *“drift detection”*)). A consulta retornou um total de 73 artigos, que foram refinados seguindo o modelo PRISMA [Page et al., 2021]. Destes, apenas 16 relacionam-se de forma mais estreita ao escopo do presente trabalho.

¹Disponível em: <https://riverml.xyz/dev/api/drift/ADWIN/>

E, considerando-se esse subconjunto de artigos com alta aderência ao objetivo desta pesquisa, nenhum utilizou a abordagem de previsão de eventos com concept drift.

Os principais temas encontrados na literatura são predição em tempo real, previsão do próximo evento e mudança de desempenho. Entre os trabalhos de predição em tempo real, destacam-se abordagens baseadas em análise de sobrevivência [Dutta et al., 2024] e modelos estatísticos [Yao et al., 2022; Capobianco et al., 2019]. Modelos sequenciais de próximo evento ajudam a quantificar potencial ofensivo, mas não necessariamente a prever gols [Simpson et al., 2022; Mendes-Neves et al., 2024].

Por fim, Lang et al. [2025] avaliam como 28 indicadores de desempenho antecipam eventos relacionados ao gol (como chutes, escanteios e entradas na área adversária) na Bundesliga, utilizando cinco modelos supervisionados. O melhor resultado para previsão de gols atingiu $MCC_{\text{médio}} = 0,061$, evidenciando a dificuldade intrínseca do problema.

4. Metodologia

Utilizamos os dados abertos da StatsBomb² correspondentes à temporada 2015/2016 da La Liga espanhola, totalizando 380 partidas. Os dados incluem eventos classificados por tipo (passe, chute, duelo etc.), time, período e minuto exato. A variável-alvo corresponde à ocorrência de gols oriundos de jogadas abertas (*open play*), excluindo bolas paradas (escanteios, faltas diretas e pênaltis) por não refletirem a dinâmica de passes que o trabalho busca avaliar. Ao todo, o conjunto contém 916 gols *open play*. A Figura 1 resume a abordagem metodológica proposta.

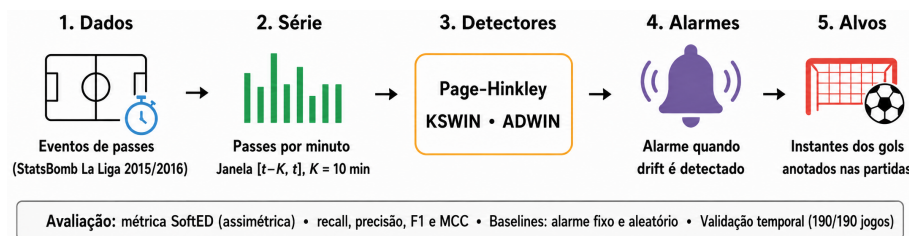


Figura 1. Metodologia de detecção de gols em função do volume médio de passes distoantes (imagem gerada por ferramenta de Inteligência Artificial).

Os eventos são agregados por minuto em formato tabular (*wide*), com colunas separadas para mandante e visitante. Os atributos correspondem a métricas de passes (totais, certos e errados). Cada tempo da partida é tratado como uma série temporal independente, iniciada em zero, e processada sequencialmente sem acesso a informações futuras. Avalia-se apenas a tarefa de ataque: a série de passes do time e os gols por ele marcados.

A aplicação de detectores de *drift* a séries esportivas impõe um desafio de avaliação: métricas pontuais (precisão, *recall*) são inadequadas para alarmes antecipados. Alarmes próximos ao gol são mais valiosos que tardios, e há falsos negativos (FN) estruturais quando o gol ocorre nos primeiros K minutos (sem janela prévia de alerta). Para lidar com isso, adaptamos o SoftED [Salles et al., 2024] para uma janela de tolerância assimétrica $[t-K, t]$, com pontuação máxima em $t-K$ e decaimento linear até zero em t .

²Disponível em: <https://github.com/statsbomb/statsbombpy>

O *pipeline* (Figura 2) opera em quatro etapas para cada série temporal de uma partida: (i) suavização (*warmup*) por média móvel de janela W ; (ii) alimentação do detector de *drift* com a série suavizada; (iii) aplicação de *cooldown* (arrefecimento) de C minutos após cada alarme; (iv) avaliação com *SoftED* assimétrico ($K = 10$).

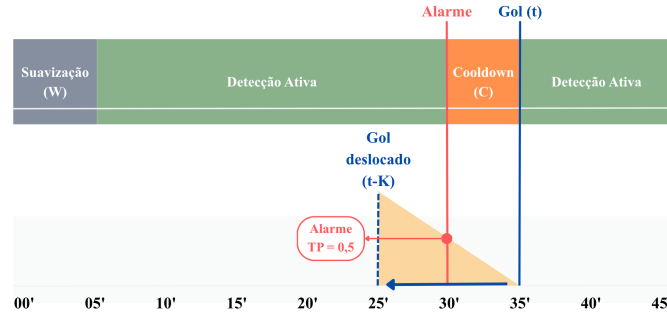


Figura 2. Pipeline de detecção de *drift* em três fases: suavização, fase ativa e *cooldown* (imagem elaborada pelos autores).

Adota-se $K = 10$ minutos como janela de tolerância. Tal escolha se deveu por representar um ponto intermediário entre o maior valor de janela (K), descontando-se o valor da suavização da média móvel (W) possíveis no contexto deste trabalho. Os parâmetros são avaliados em *grid search* com $W \in \{3, 5\}$ e $C \in \{3, 5, 10\}$, respeitando duas restrições: $C \leq K$, para que o silenciamento pós-alarme não exceda a janela de tolerância; e $W + K \leq \bar{t}_{\text{gol}}$, onde $\bar{t}_{\text{gol}} = 18$ min é o tempo médio até o primeiro gol de cada período, garantindo que o detector já tenha concluído a suavização e a janela $[t - K, t]$ esteja integralmente disponível para um gol ocorrendo no instante médio.

Os três detectores de *drift* também são avaliados em *grid search*:

- **Page-Hinkley:** $\lambda \in \{3, 5, 7, 10\}$, $\delta \in \{0,001; 0,005; 0,01\}$, $\alpha \in \{0,90; 0,95; 0,99\}$, $n_{\min} \in \{1, 3\}$, modo = both;
- **KSWIN:** $\alpha \in \{0,001; 0,005; 0,01; 0,05\}$, w e s definidos em função de K : para $K=10$, $s \in \{5, 10, 20\}$ e $w \in \{10, 20, 30, 40, 60\}$ com a restrição $w \geq 2s$;
- **ADWIN:** $\delta \in \{0,0001; 0,0005; 0,001; 0,002; 0,005; 0,01; 0,05\}$.

Para todos, $W \in \{3, 5\}$ e $C \in \{3, 5, 10\}$. A combinação de hiperparâmetros com maior F_1 agregado (soma acumulada de TP, FP e FN sobre todas as partidas do time) é selecionada independentemente por time e tarefa. Nos experimentos principais, essa seleção é realizada sobre todos os 380 jogos disponíveis (*in-sample*). Para verificar a estabilidade dos hiperparâmetros selecionados, adota-se também uma divisão temporal: os 190 jogos mais antigos são usados para seleção e os 190 jogos mais recentes para avaliação *out-of-sample*.

Dois baselines são comparados: (i) o *alarme fixo*, que dispara alarmes a cada 22 minutos de cada período, por representar o instante médio aproximado de um tempo de jogo padrão (45 minutos), desconsiderando acréscimos concedidos pela arbitragem (elemento aleatório), independentemente do sinal observado; e (ii) o *alarme aleatório*, baseado em um processo de Bernoulli com $p = 0,5$ por minuto, com média sobre $N = 10$ execuções independentes.

Duas características da estrutura temporal dos gols influenciam o *recall* independentemente da qualidade do detector: 196 gols (21,4%) ocorrem nos primeiros $K = 10$ minutos de algum período, intervalo em que, por definição da janela de avaliação, não há tempo hábil para um alarme antecipado; e 170 gols (18,6%) ocorrem a menos de K minutos após outro gol no mesmo período, com janelas de avaliação sobrepostas. Com sobreposição de 23 gols entre os dois casos, ao todo 343 gols (37,4%) apresentam ao menos uma dessas condições.

5. Resultados

A Tabela 1 avalia a robustez da seleção de hiperparâmetros: uma única configuração global é escolhida no conjunto de treino (190 partidas) e aplicada diretamente no teste (190 partidas), sem nova otimização³. A coluna *in-sample* serve como referência: é a melhor configuração global encontrada sobre as 380 partidas completas, podendo diferir da selecionada no treino.

Tabela 1. MCC por período de avaliação — robustez da seleção de hiperparâmetros (divisão temporal 190/190, tarefa de ataque, $K=10$).

Detector	Treino (190)	Teste (190)	In-sample (380)
Page-Hinkley	0,133	0,132	0,133
KSWIN	0,092	0,088	0,090
ADWIN	0,010	0,030	0,022

Para referência, o *baseline* aleatório ($p=0,5$, média de 10 execuções) obteve $MCC = 0,065$ e *baseline* fixo (alarme a cada 22 min), $MCC = 0,037$; como não possuem hiperparâmetros ajustados aos dados, seus valores permanecem estáveis entre períodos de avaliação.

A Tabela 2 apresenta os melhores resultados agregados por time para MCC, F_1 , precisão, *recall* e alarmes por partida. Diferentemente da tabela anterior, que usa uma configuração global única, cada time adota a configuração que maximiza seu próprio F_1 , produzindo valores agregados de MCC mais elevados.

Tabela 2. Desempenho agregado — tarefa de ataque, $K=10$. Métricas calculadas in-sample (380 partidas); alarmes/partida com melhor configuração/modelo.

Detector	MCC	F1	Recall	Precisão	Alarmes/partida
Page-Hinkley	0,1517	0,1242	0,4256	0,0727	6,6
KSWIN	0,1196	0,1266	0,1983	0,0930	2,4
ADWIN	0,0344	0,0155	0,0081	0,1683	0,1
Baseline Fixo	0,0373	0,0500	0,1080	0,0325	4,0
Aleatório	0,0691	0,0397	0,8043	0,0203	47,7

Em termos absolutos, F1 e MCC ainda são baixos, refletindo a dificuldade da tarefa. PH e KSWIN superam ambos os baselines em MCC (Tabela 2). O PH obteve o maior MCC (0,1517) com $W=3$ e $C=3$, indicando que suavização mínima e *cooldown*

³Disponível em: <https://github.com/anavibiga/sbbd-drift-detection>

curto favorecem reatividade, ao custo de mais alarmes (35,2/partida). O KSWIN alcançou F1 semelhante ao PH (0,1266 vs. 0,1242) com menos alarmes (2,4/partida) e maior precisão (0,0930 vs. 0,0727), indicando que seleciona momentos mais informativos, embora com menor *recall*.

O ADWIN foi excessivamente conservador: apenas 0,1 alarmes por partida, com $F1 = 0,016$ e $MCC = 0,035$. A alta variabilidade das séries de passes possivelmente impede que diferenças entre sub-janelas excedam o limiar estatístico do detector. O baseline aleatório ($p=0,5$) gera 47,7 alarmes por partida, produzindo *recall* artificialmente alto (0,80), mas precisão muito baixa (0,02) e MCC inferior aos detectores (0,069). O baseline fixo (22 min) apresenta o pior MCC (0,037).

Lang et al. [2025] reportam $MCC_{\text{médio}} = 0,06$ para o melhor modelo supervisionado na Bundesliga. O PH alcançou $MCC = 0,13$, aproximadamente o dobro desse valor. No entanto, a comparação não é direta, devido a diferenças nos dados (La Liga vs. Bundesliga), na definição da tarefa (alarme pontual vs. predição por janela) e no protocolo experimental (divisão temporal vs. validação cruzada). Ainda assim, os resultados sugerem que detectores de *drift* não supervisionados podem alcançar desempenho competitivo.

6. Considerações finais

Este trabalho investigou se detectores de *drift* não supervisionados antecipam gols no futebol, formulando o problema como *drift* virtual em séries de passes e adotando a variante assimétrica *SoftED* para privilegiar alarmes antecipados. Nos experimentos, Page-Hinkley e KSWIN superaram ambos os baselines em MCC, com PH alcançando $MCC = 0,1517$ e $F1 = 0,1242$. ADWIN apresentou desempenho limitado dada a alta variabilidade das séries de passes.

Delimitações metodológicas residem na restrição dos experimentos a: (i) um único campeonato (La Liga), (ii) uma única *feature* (contagem de passes) e (iii) um único valor de janela de antecipação ($K=10$ min). Generalizações para outros contextos requerem validação adicional e serão futuramente explorados.

Como trabalhos futuros, destacam-se: validação temporal entre temporadas; exploração de atributos além de passes; combinações de detectores (*ensemble*) e controle do orçamento de alarmes; análise separada dos tempos da partida; investigação de causalidade reversa entre passes e gol; avaliação do impacto do *cooldown*, da exclusão de gols de bola parada na calibração de W e C e de um terceiro *baseline* baseado em limiar absoluto de passes; além da análise de outros valores de K e campeonatos. Os resultados sugerem que detectores de *drift* podem atuar como mecanismos simples e interpretáveis de alerta antecipado em sistemas táticos em tempo real sem demandar treinamento.

Uso de Inteligência Artificial

Ferramentas de IA generativa foram utilizadas para revisão linguística e para gerar a Figura 1. Os autores permanecem responsáveis pelo conteúdo do artigo.

Referências

Bifet, A. and Gavaldà, R. (2007). *Learning from Time-Changing Data with Adaptive Windowing*, pages 443–448.

- Capobianco, G., Giacomo, U. D., Mercaldo, F., Nardone, V., and Santone, A. (2019). Can machine learning predict soccer match results? In *ICAART 2019 - Proceedings of the 11th International Conference on Agents and Artificial Intelligence*, volume 2, pages 458–465. SciTePress.
- Dutta, A., Saikia, H., Gogoi, J., and Bhattacharjee, D. (2024). Forecasting the opening goal in second-half of a football match: Bayesian and frequentist perspectives. *Computational Statistics*.
- Gama, J., Zliobaite, I., Bifet, A., Pechenizkiy, M., and Bouchachia, A. (2014). A survey on concept drift adaptation.
- Hinder, F., Vaquet, V., and Hammer, B. (2024). One or two things we know about concept drift—a survey on monitoring in evolving environments. Part A: detecting concept drift. *Frontiers in Artificial Intelligence*, 7:1330257.
- Kore, A., Abbasi Babil, E., Subasri, V., Abdalla, M., Fine, B., Dolatabadi, E., and Abdalla, M. (2024). Empirical data drift detection experiments on real-world medical imaging data. *Nature Communications*, 15:1887.
- Lang, S., Wimmer, T., Erben, A., and Link, D. (2025). Which indicators matter? using performance indicators to predict in-game success-related events in association football. *International Journal of Computer Science in Sport*, 24:16–44.
- Mendes-Neves, T., Meireles, L., and Mendes-Moreira, J. (2024). Towards a foundation large events model for soccer. *Machine Learning*, 113:8687–8709.
- Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, 41(1/2):100–115.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., and Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372:n71.
- Raab, C., Heusinger, M., and Schleif, F.-M. (2020). Reactive soft prototype computing for concept drift streams. *Neurocomputing*, 416.
- Salles, R., Lima, J., Reis, M., Coutinho, R., Pacitti, E., Masegla, F., Akbarinia, R., Chen, C., Garibaldi, J., Porto, F., and Ogasawara, E. (2024). SoftED: Metrics for soft evaluation of time series event detection. *Computers and Industrial Engineering*, 198.
- Sarmiento, H., Figueiredo, A., Lago-Peñas, C., Milanovic, Z., Barbosa, A., Tadeu, P., and Bradley, P. S. (2018). Influence of tactical and situational variables on offensive sequences during elite football matches. Technical report.
- Simpson, I., Beal, R. J., Locke, D., and Norman, T. J. (2022). Seq2event: Learning the language of soccer using transformer-based match event prediction. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 3898–3908. Association for Computing Machinery.
- Tavares, L. G., Lima, J., Melo, M., Chen, C., Garibaldi, J., Scatena, G. d. S., Costa, A. H. R., Gomi, E. S., Salles, R., Pacitti, E., Santos, I., Siqueira, I. G. a., Carvalho, D., Coutinho, R., Porto, F., and Ogasawara, E. (2025). Fuzzy-based ensemble method for robust concept drift detection in multivariate time series. In *2025 International Joint Conference on Neural Networks (IJCNN)*.
- Yao, W., Wang, Y., Zhu, M., Cao, Y., and Zeng, D. (2022). Goal or miss? a bernoulli distribution for in-game outcome prediction in soccer. *Entropy*, 24.