



footure
academy

ESPECIALIZAÇÃO

CIÊNCIA NOVO
DE DADOS

NO FUTEBOL

Aula 6 - Expected Goals (xG)

Caio Batatinha

AGENDA

- Weighted Shots
- Definição
- Seleção de modelo e variáveis
- Diferentes modelos
- Aplicações

OBJETIVOS DE APRENDIZAGEM

- Aprender a definição de Expected Goals
- Como é calculado, a importância da seleção de variáveis e treinar modelos
- Entender as diferentes aplicações e boas práticas do Expected Goals

REVISÃO CONCEITOS AULA PASSADA

- Introdução às métricas utilizadas para realizar as análises no futebol.
- Identificação por tipo e aplicação de métricas para equipes e jogadores.

HISTÓRIA - WEIGHTED SHOTS

Em 1997, Richard Pollard e Charles Reep publicaram um artigo chamado *Measuring the effectiveness of playing strategies at soccer*. O trabalho visava examinar diferentes posses de um time e os eventos com a bola e medir a eficiência dessas posses em relação ao gol.

A ideia central do artigo gira em torno do *Weighted Shots*, um modelo de regressão logística utilizado para prever a probabilidade de gol. Para todas as posses analisadas que não terminaram em finalização, um valor 0 foi assinalado. Para todas as outras, um valor p foi assinalado e esse parâmetro p determinado via regressão logística.

Esse parâmetro p depende ainda do tipo de zona e tipo de posse (bola parada ou open play).

Mais de 489 posses terminaram em chute, sendo que 47 terminaram em gol.

EXPECTED GOALS

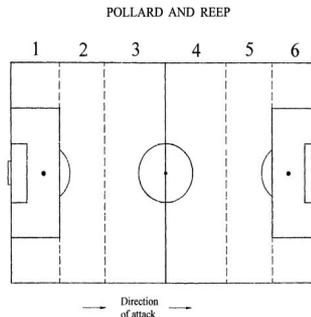


Fig. 1. Division of the field of play into six zones

$$p_{ij} = \sum_{k=1}^{n_{ij}} p_{ijk} / n_{ij}$$

TABLE 1
Comparison of values assigned to different outcome variables†

Outcome of team possession	Outcome variable				
	Goal	Shot	Weighted shot	Preliminary yield	Yield
Shot: goal	1	1	p	p	p
Shot: not a goal	0	1	p	p	p
Possession regained	0	0	0	p_y	y_y
Possession lost	0	0	0	$-p_y$	$-y_y$

† p is the estimated scoring probability of a shot, p_y is the estimated probability of scoring from possession originating as type j in zone i and y_y is the estimated yield from possession originating as type j in zone i .

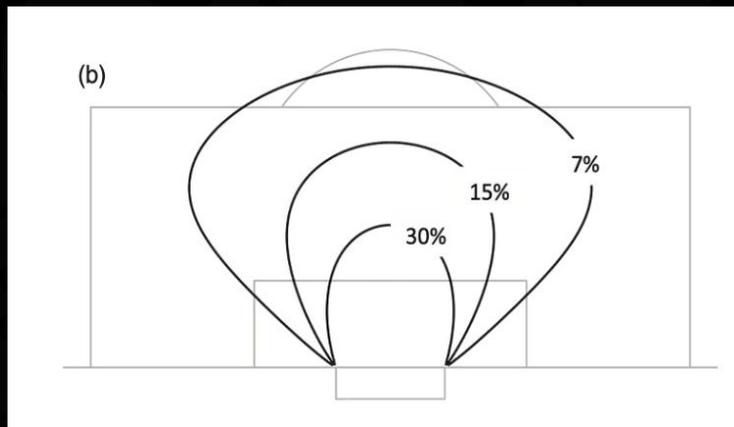
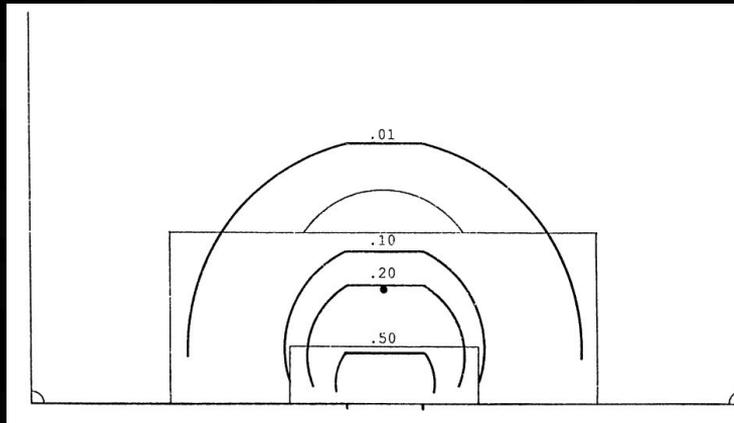
DEFINIÇÃO

Expected Goals, Gols Esperados ou popularmente abreviado para xG é um modelo matemático, cujo objetivo é medir a probabilidade de um chute se tornar um gol.

Exemplo: Da pequena área eu só preciso de 5 chutes para marcar 1 gol, ou seja, 0.2 chutes culminam em gol, essa é a probabilidade de marcar.

Tal métrica auxilia na tomada de decisões para avaliar se uma equipe está criando boas chances ou não, informações que apenas com o número de finalizações não é possível inferir.

EXPECTED GOALS



DEFINIÇÃO

EXPECTED GOALS

**SELEÇÃO DO
MODELO**



**SELEÇÃO DAS
VARIÁVEIS**

SELEÇÃO DE MODELOS

MODELOS

- Regressão Logística
- Árvore de decisão
- Random Forest
- Gradiente descendente (xG Boost)

REGRESSÃO LOGÍSTICA

Modelo de classificação binário (gol ou não gol)

Vantagens:

- Simples
- Menos processamento
- Bom para amostras pequenas

Desvantagens:

- Assume uma relação linear entre variáveis
- Limitação em relações mais complexas

RANDOM FOREST

Combinação de várias árvores de decisão

Vantagens:

- Bom para amostras grandes
- Reduz o overfitting de uma única árvore de decisão

Desvantagens:

- Mais difícil de interpretar
- Maior processamento

EXPECTED GOALS

ÁRVORE DE DECISÃO

Modelo de classificação não linear que divide os dados em galhos baseado nas variáveis.

Vantagens:

- Fácil de visualizar
- Captura relação não linear
- Adapta a dados contínuos e categóricos

Desvantagens:

- Problema de Overfitting
- Baixa estabilidade

GRADIENTE DESCENDENTE

Combinação de modelos simples sequenciais que minimizam o erro de previsão

Vantagens:

- Alta capacidade preditiva
- Excelente para dados não lineares

Desvantagens:

- Mais complexo
- Maior tempo de processamento
- Necessidade de ajustar os hiperparâmetros

SELEÇÃO DAS VARIÁVEIS

BÁSICAS

- Distância
- Ângulo
- Coordenadas do chute
- Parte do corpo
- Tipo de jogada
- Resultado do chute
- Assistência

AVANÇADAS

- Altura da bola
- Velocidade da bola
- Número de bloqueadores
- Posição do goleiro
- Pressão aplicada
- Tempo da partida
- Gamestate

MODELO STATSOMB

Um caso de modelo mais robusto de xG é o da empresa fornecedora de dados Statsbomb. Nos anos recentes eles adicionaram duas variáveis que fizeram diferença na forma como eles calculavam Expected Goals:

- Shot Freeze Frame: Uma fotografia do momento exato do chute envolvendo não só o chutador, mas variáveis relativas aos defensores e outros jogadores presentes no momento do chute.
- Shot Impact Height: O impacto da altura da bola no momento do chute.

O impacto da altura pode ser conferido no link a seguir:

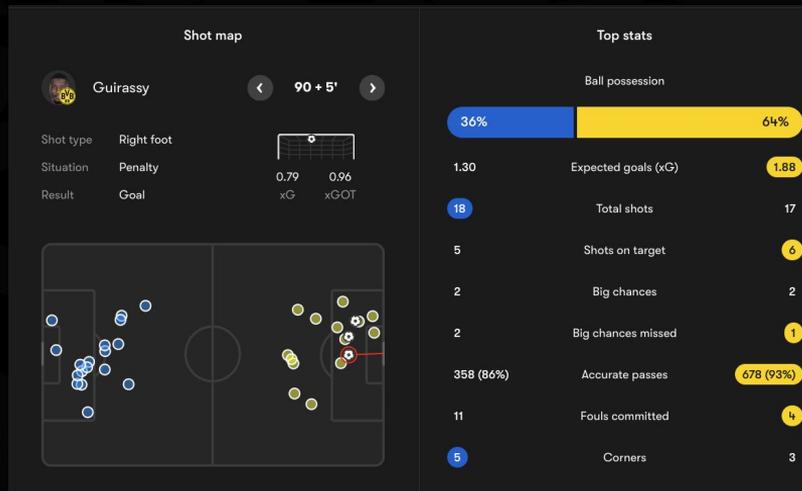
<https://statsbomb.com/news/statsbomb-release-expected-goals-with-shot-impact-height/>

EXPECTED GOALS



DIFERENÇA ENTRE MODELOS

Match Stats		StatsBomb
Club Brugge vs Borussia Dortmund 18/09/2024, UEFA Champions League		
Club Brugge	0-3	Borussia Dortmund
1.41	xG	2.15
18	Shots	17
5	Shots on Target	6
36%	Possession	64%
35	Passes Into Final Third	65



Match	Score	Time	Team	Formation	Expected Goals (xG)	Goals
Club Brugge 0:3 Borussia Dortmund	0-3	96'	Borussia Dortmund	4-2-3-1 91%	2.23	3
Club Brugge 0:3 Borussia Dortmund	0-3	96'	Club Brugge	4-1-4-1 100%	1.85	0

APLICAÇÕES

**ANÁLISE DE
EQUIPE**

**SCOUTING DE
ATLETAS**

**ANÁLISE DE
TREINADORES**

**PREVISÃO DE
TEMPORADA**

ANÁLISE EQUIPE

Resultados nem sempre refletem a performance de uma equipe. Fatores como sorte, desempenho do goleiro adversário tem impacto direto no resultado final. Expected Goals é uma forma de analisar a performance de uma equipe em uma perspectiva mais realista do que apenas gols.

Abaixo algumas boas práticas ao analisar xG de uma equipe em relação a amostragem:

1-2 Partidas: Análise individual das partidas. Quais táticas foram adotadas? Qual foi o desempenho dos jogadores?

3-6 Partidas: Aqui é uma ótima janela para analisar se um time está tendo mais sorte ou azar nas performances recentes. Quais pontos divergem para conquistar o resultado final?

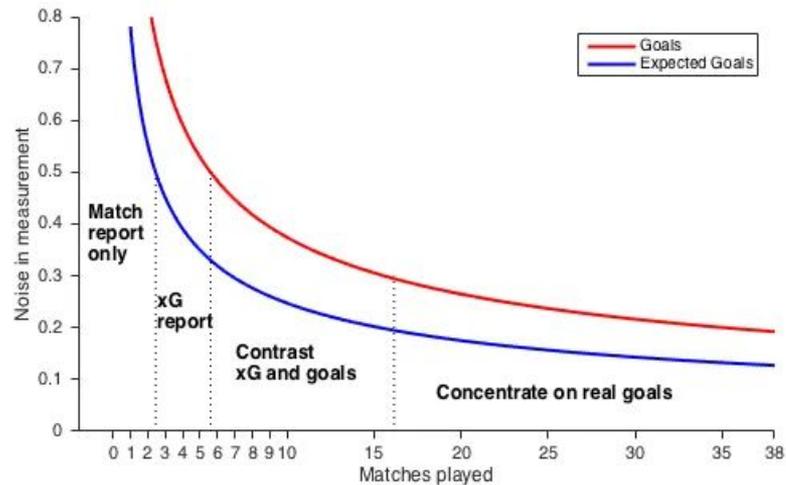
7-16 Partidas: Essa janela já é um ponto sensível na análise via xG. 10 partidas divergindo o número de gols x Chances criadas é preciso contrastar e verificar o modelo utilizado ou quais fatores que levam a tamanha diferença.

+16 Partidas : O ruído entre gols e xG já está mais diluído. Dificilmente um time vai estar performando muito acima do xG durante uma temporada completa.

xG AO LONGO DO TEMPO

A imagem ao lado contrasta o ruído nos dados entre gols e Expected Goals ao longo do tempo. Note-se que as duas com o aumento da amostra se aproximam.

EXPECTED GOALS



SCOUTING DE JOGADORES

Expected Goals e suas métricas derivadas auxiliam também no processo de scouting, principalmente de jogadores mais ofensivos e goleiros. Essa abordagem mais objetiva permite encontrar jogadores que se posiciona bem para finalizar, convertem e para goleiros avaliar quais conseguem prevenir chances de qualidade.

Métricas e estratégias para identificar jogadores:

Para Atacantes/Jogadores Ofensivos:

- xG por 90: Permite comparar com outros jogadores e avaliar a capacidade de se colocar em boas oportunidades de gol com frequência e estão envolvidos em jogadas perigosas.
- xG por chute: Separar jogadores que priorizam volume com baixa probabilidade de gol e pouco volume, porém com mais qualidade de chute.
- $np:Gols - xG$: Identificar jogadores que tenham performance abaixo por uma questão técnica ou acima sejam finalizadores acima da média.

Para Goleiros :

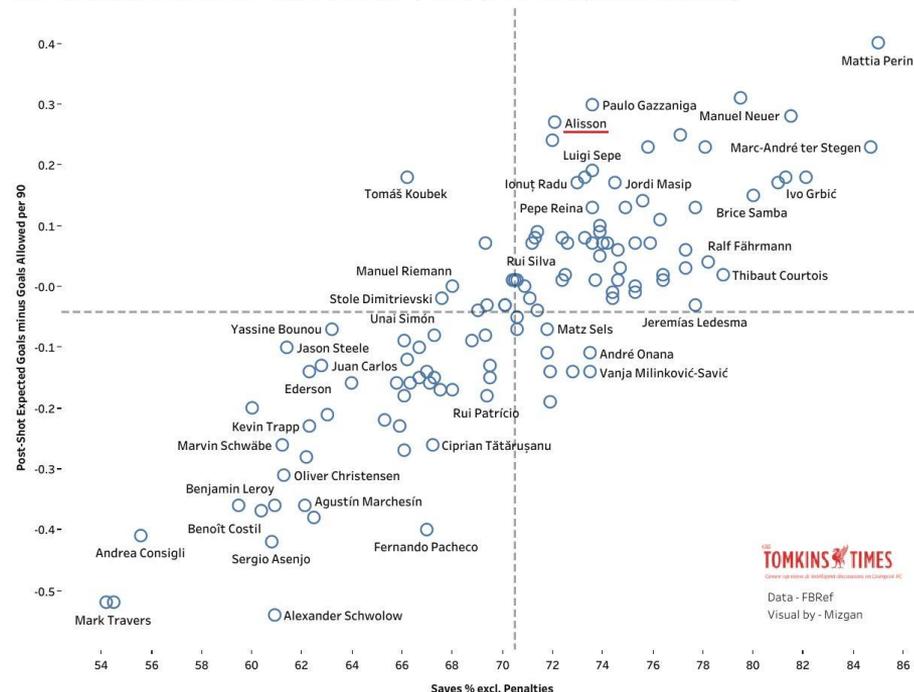
- PSxG: Medir a dificuldade das finalizações enfrentadas.
- PSxG/Finalizações no alvo: Identificar goleiros que destacam em defesa difíceis.
- PSxG - Gols : Identificar quais goleiros estão salvando mais finalizações de dificuldade elevada ou estão sofrendo mais gols do que esperado.

ALISSON REI DO PSxG

O gráfico ao lado mostra o cruzamento de dados entre PSxG - Gols Sofridos e % de defesas. O brasileiro Alisson mostra como teve uma excelente temporada em termos de prevenir gols de alta probabilidade, mesmo que sua % defesas não foi a melhor.

EXPECTED GOALS

Top-5 EU League Goalkeepers - Saves % & PSxG +/- : 2022/23 Season (min. 900 minutes)



ANÁLISE DE TREINADORES

De forma análoga a de equipes, reproduz a lógica para identificar treinadores que consigam implementar suas ideias de jogo e avaliar se sua performance condiz com os resultados ou aparentemente algum fator externo está influenciando positivamente ou negativamente.

Um bom exemplo é comparar a média de Expected Goals a favor e contra ao longo do tempo. Esse tipo de visualização permite ver se o desempenho é variável devido a condições de calendário, ou mudanças que ocorreram taticamente ou no próprio elenco.

Análise se a equipe tem um perfil mais ofensivo em criar várias chances ou defensivo de permitir poucas chances. xG/Chute é uma métrica derivada que revela também se a equipe prioriza mais volume e menos qualidade ou um volume menor de chutes, porém com probabilidade mais alta de ser gol.

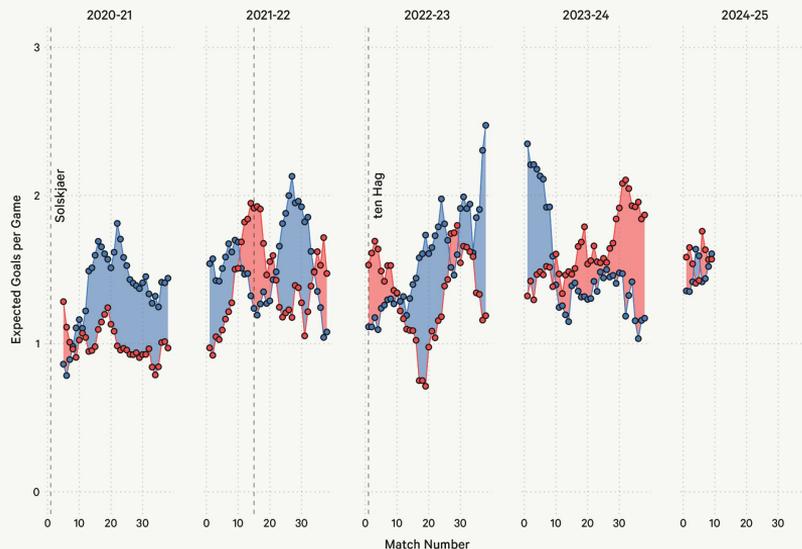
PANELA DE PRESSÃO

O gráfico ao lado descreve a tendência de Expected Goals a favor e contra do Manchester United ao longo das últimas temporadas. Percebe que o início de Ten Hag foi complicado, mas ao longo do tempo a equipe recuperou a performance e estabilizou mais a sua proposta de jogo. Na temporada 23-24 teve o efeito contrário, um bom início e depois uma queda sem voltar a forma desejada. Mesmo numa amostra pequena, a temporada 24-25 seguiu na mesma tendência e por isso que recentemente foi demitido do cargo.

EXPECTED GOALS

How Manchester United's non-penalty xG for and against trend over time

10-game rolling average for Manchester United in the Premier League since 2020-21



The Athletic

PREVISÃO DE TEMPORADA

Como demonstrado por vários pesquisadores, Expected Goals é um indicador melhor de performance do que o próprio gol. A partir do xG é possível analisar as finalizações totais de uma partida e calcular o Expected Points (xPts). A partir disso é possível simular cenários e incluir a própria variável Expected Goals em modelos preditivos para calcular resultados futuros e mais prováveis de ocorrer.

Metodologia: Todas as finalizações da partida tem um xG (probabilidade de gol) associada. Com isso é possível aplicar técnicas como Simulação de Monte Carlo para prever possíveis resultados decorrente dessa partida. A partir disso calcula-se a probabilidade de vitória, empate e derrota.

$$xPts = 3 * p(\text{Vitória}) + 1 * p(\text{Empate}) + 0 * p(\text{Derrota})$$

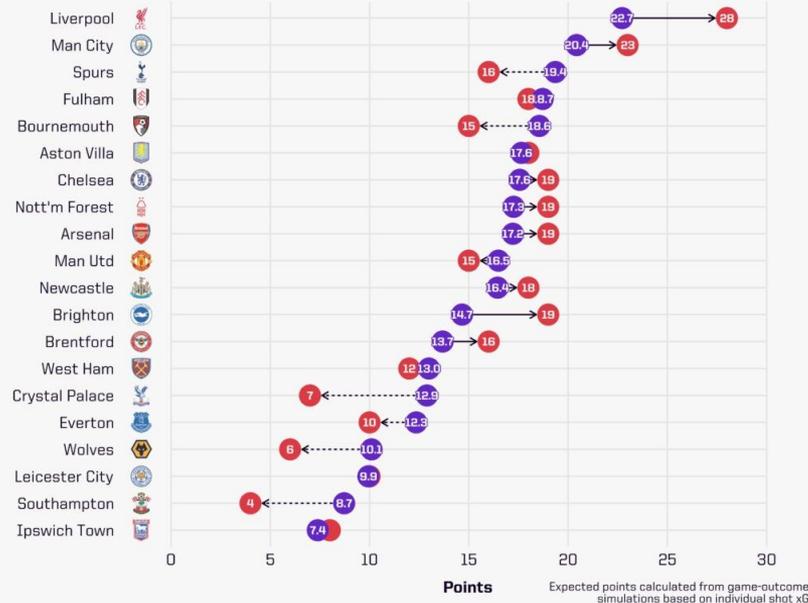
REGRESSÃO A MÉDIA?

A comparação da pontuação atual com xPts revela uma performance bem acima da média do Liverpool, com 2 vitórias de vantagem onde não se esperava. Já o Tottenham poderia estar ocupando a terceira posição da tabela, mas no entanto falhou em garantir pontos importantes.

EXPECTED GOALS

Which teams have over/under-performed?

Comparing **actual** and **expected** points | English Premier League 2024-25



MÉTRICAS DERIVADAS

MÉTRICA	DEFINIÇÃO	MELHOR USO
Expected Assists (xA)	Soma dos valores de xG das finalizações criadas diretamente por passes de um jogador	Medir criatividade e a capacidade de criação de jogadas de meio-campistas e jogadores de lado.
xG on Target (xGOT)	Gols esperados ajustados pela colocação do chute no alvo (ex.: canto superior vs. chute direto no goleiro)	Avaliar a qualidade do chute pós-contato, diferenciando finalizações "sortudas" das intencionais.
xG Chain	Valor total de xG de todas as chances em que um jogador participou durante uma sequência de posse.	Medir a contribuição geral de um jogador na criação de oportunidades em jogadas coletivas.
xG Buildup	xG Chain excluindo chutes e assistências.	Avaliar o papel de um jogador na progressão da bola ou na criação de jogadas iniciais.
Expected Point (xPts)	Pontos previstos com base no xG a favor e contra em um jogo ou temporada, usando modelagem probabilística.	Identificar equipes que estão performando acima ou abaixo da média; prever resultados futuros.

Dúvidas?

REFERÊNCIAS UTILIZADAS

Pollard, Richard, and Charles Reep. "Measuring the effectiveness of playing strategies at soccer." Journal of the Royal Statistical Society: Series D (The Statistician) 46, no. 4 (1997): 541-550.

<https://statsbomb.com/articles/soccer/upgrading-expected-goals/>

<https://statsbomb.com/news/statsbomb-release-expected-goals-with-shot-impact-height/>

<https://statsbomb.com/news/statsbomb-data-case-studies-freeze-frames-and-defender-locations/>

<https://statsbomb.com/pt-pt/metricas/o-que-e-xg/>

<https://theanalyst.com/eu/2024/11/premier-league-table-bunched-middle>

<https://statsbomb.com/articles/soccer/a-new-way-to-measure-keepers-shot-stopping-post-shot-expected-goals/>

<https://www.nytimes.com/athletic/5830756/2024/10/28/10-charts-that-explain-manchester-united-under-ten-hag/>

<https://soccermatics.medium.com/should-you-write-about-real-goals-or-expected-goals-a-guide-for-journalists-2cf0c7ec6bb6>

OBRIGADO!



@CcBatatinha



[linkedin.com/in/caio-batatinha/](https://www.linkedin.com/in/caio-batatinha/)



caio.batatinha@footure.com.br
caiocbatatinha@gmail.com