



# An AI-driven approach to adapting the Expected Goals (xG) model to women's football

Tomasz Lipowski, Tomasz Piłka

Adam Mickiewicz University, Faculty of Mathematics and Computer Science, 61-614 Poznań, Uniwersytetu Poznańskiego 4 Street, Poland

Project implementation related to research project IDUB No. 134.

7-9 April 2025

## Introduction

Football analysis helps teams improve performance and maximize profits. Analysts study details of their own and rival teams, using tracking data to draw insights. Their work shapes trends that evolve the game. Expected Goals (xG) is a metric based on historical data, estimating the probability (0-1) of a shot resulting in a goal. It outperforms raw goal counts in predicting scoring potential and serves as a foundation for new metrics. Built on multiple factors, xG is key to assessing a team's true performance. The problem is that xG models are not very gender-specific, making them less accurate in reality. Men's and women's football differ significantly, so the model should account for this in its calculations. This project aims to analyze these differences.

## AI-Based Model

This model is called My xG

- Logistic regression as the baseline model
- Multilayer Perceptron with 3 hidden layers 64, 64, 32
- Output layer with 1 neuron using a sigmoid activation function
- Optimizer - Adam
- Loss function - binary cross-entropy
- Evaluation is performed using log loss

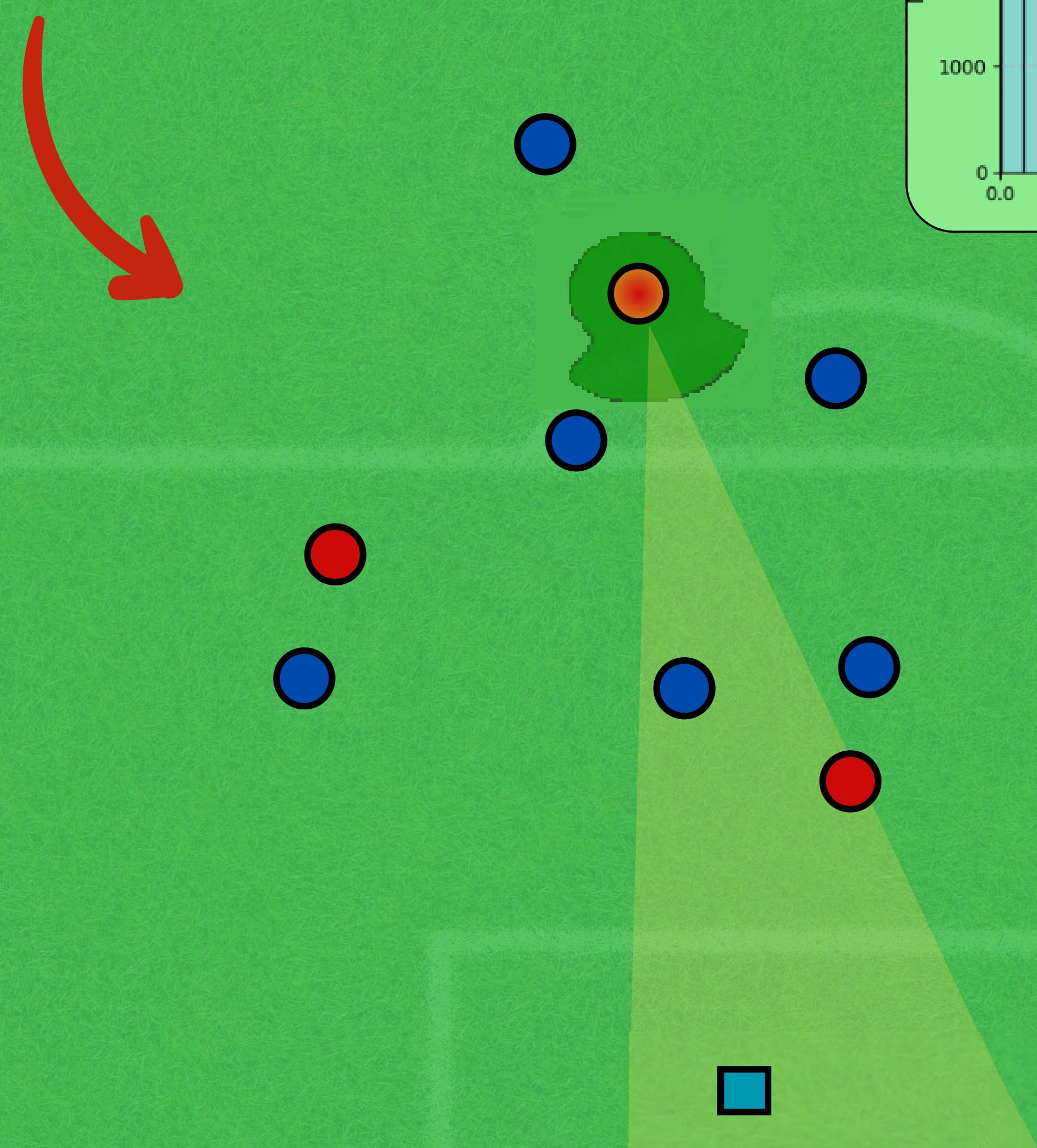
## Hudl StatsBomb Open Data

Competition Name	Year/Season	Matches	Shots
La liga	2015/2016	380	9071
Fa. Women's Super League	2018-2021	326	8239
FIFA World Cup	2022	128	3068
Women's World Cup	2019, 2023	116	2891
National Women's Soccer League	2018/2019	36	1034
UEFA Women's Euro	2023	31	871

## Shot parameters

The model uses both direct StatsBomb data, such as shot type, shot body part, shot technique, shot aerial, under pressure, distance, and angle, as well as newly developed attributes like goalkeeper in shot keeper cone, players in shot keeper cone, enemy players in shot keeper cone, players in the box, and pressing. These additional parameters improve the model's accuracy and are illustrated below.

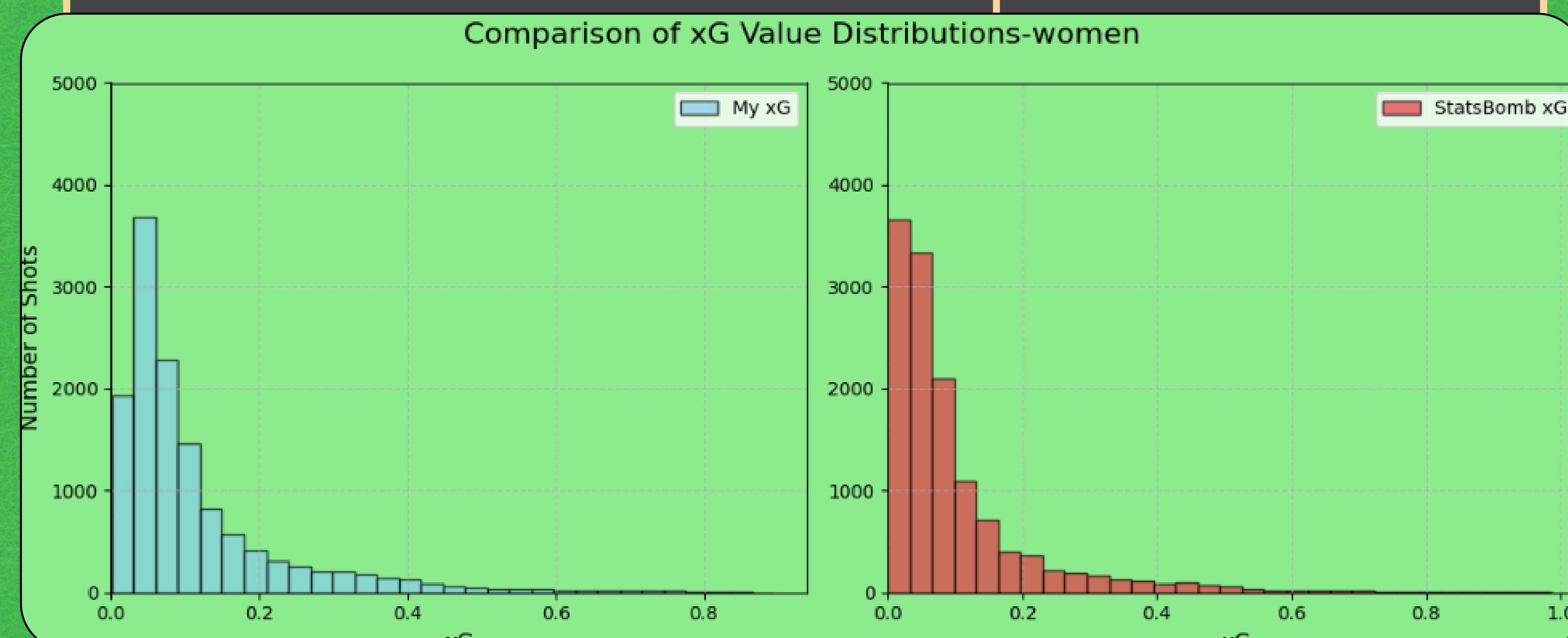
- Shooter
- Pressing Area
- Keeper Cone
- Defender
- Goalkeeper
- Attacker



## Results

My xG, after adding additional parameters, reduced the log loss from 0.273 to 0.260, while StatsBomb's is 0.265. Four out of the five additional parameters turned out to be among the most significant features for the model.

Features	Importance (%)
angle degrees	33.47
distance	28.30
pressing	2.94
players in the box	2.72
players in shot keeper cone	2.66
opponent players in shot keeper cone	1.45



Statistics	Female	Male
Long-range shots	3507	3554
Long-range goals	129	87
Long-range accuracy	3.68 %	2.45 %
Long-range frequency	26.90 %	29.28 %
Long-range goals/all	0.99 %	0.72 %

## Bibliography

- Bransen, L. and Davis, J. Women's football analyzed: interpretable expected goals models for women, 2021.  
 Statsbomb. What are expected goals (xG)? <https://statsbomb.com/soccer-metrics/expected-goals-xg-explained>, 2024. Accessed: 15-February-2025.  
 Anzer, G. and Bauer, P. A goal scoring probability model for shots based on synchronized positional and event data in football (soccer). *Frontiers in Sports and Active Living*, 3, 2021. doi:10.3389/fspor.2021.624475.