Soccer Expected Possession Value

A model that converts player actions and game state into expected goals using computer vision and machine learning

Mario Suazo, Fernando Arroyo, Melissa Martínez

Sports Analytics Research Group under Dr. Anton Dahbura and Tad Berkery

JOHNS HOPKINS
WHITING SCHOOL of ENGINEERING

Introduction and Motivation

Soccer decision-making is complex and dynamic. This EPV tool aims to quantify the best move in soccer at any moment, making decisions more intuitive. Our main motivation was how exactly we wanted to quantify this best move and how to relay that information to the user.

Project Overview

We developed a computer vision pipeline that detects players, estimates their location, and integrates this into an Expected Possession Value (EPV) model to recommend optimal actions for players in a given game state.

Input Photo and Player Roster » Player & field identification and evaluation » Input to model and predict

Data and Pipeline

AWS Database

SkillCorner MLS 2023 data

SoccerNet Soccer videos

STEP ONE: PROCESS AND ALIGN TRACKING AND EVENT DATA

STEP TWO: STORE AND QUERY DATA THROUGH AWS

STEP THREE: TRAIN MODELS USING PROBABILITY-BASED TARGETS

STEP FOUR: INTEGRATE MODEL OUTPUTS INTO DASHBOARD AND TOOLS

Results and Demo

1 Player Identification

Andi Zeqiri	17	Sion	#fafdfc
Bradley	38	Sion	#fafdfc
Pepe Reina	25	Luzern	#30338c
Mauro	22	Luzern	#30338c

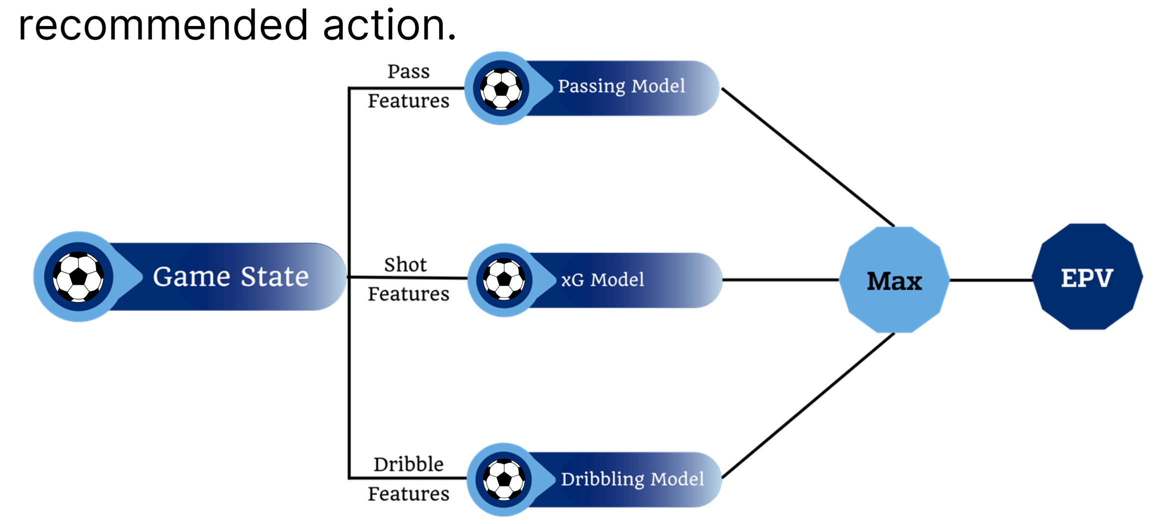
Team Roster and Jersey Color

Game Image

Player Identification
Player Coordinate CSV

The user first feeds in a game image along with a CSV of the team roster and their jersey hex code. The user also has the option to include player statistics, if available, to make the model more accurate. Then, YOLO is used to identify the players in the image, which the user will identify along with one feature of the field. Once complete, a CSV can be generated with the players' coordinates on the field.

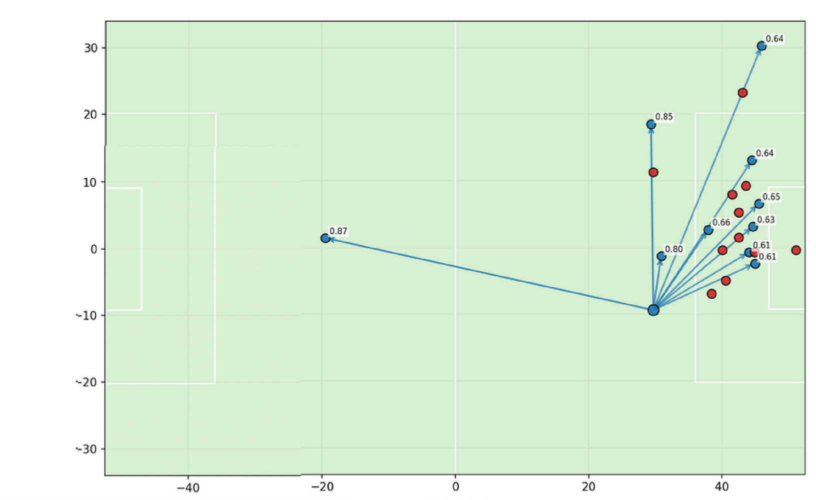
The EPV model uses a reinforcement learning approach by implementing a Markov Decision Process with states that are dependent on all players' locations and movement. For each state, there are three actions a player can do: shoot, pass, and dribble. The reward was defined as xG. The policy is therefore the action out of the three that maximizes xG. All three actions have their own trained models. The current model used is Random Forest; other models tested were: Kernelized SVMs, regularized logistic regression, Adaboost, and neural networks. The max EPV out of the three decisions is the recommended action.

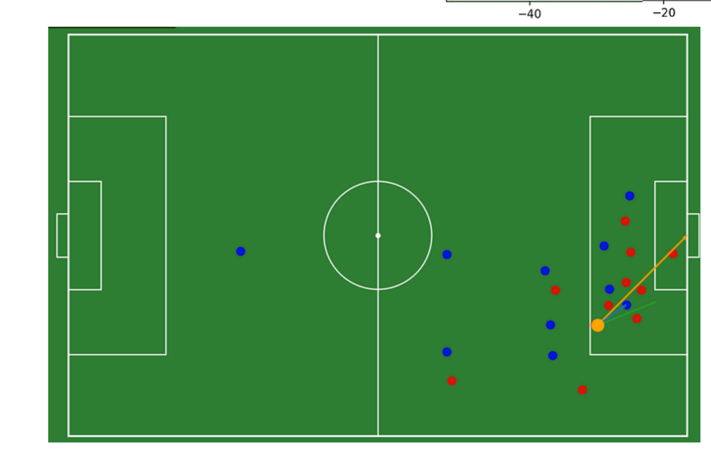


2 The Model

3 The Output

Pass Probabilities





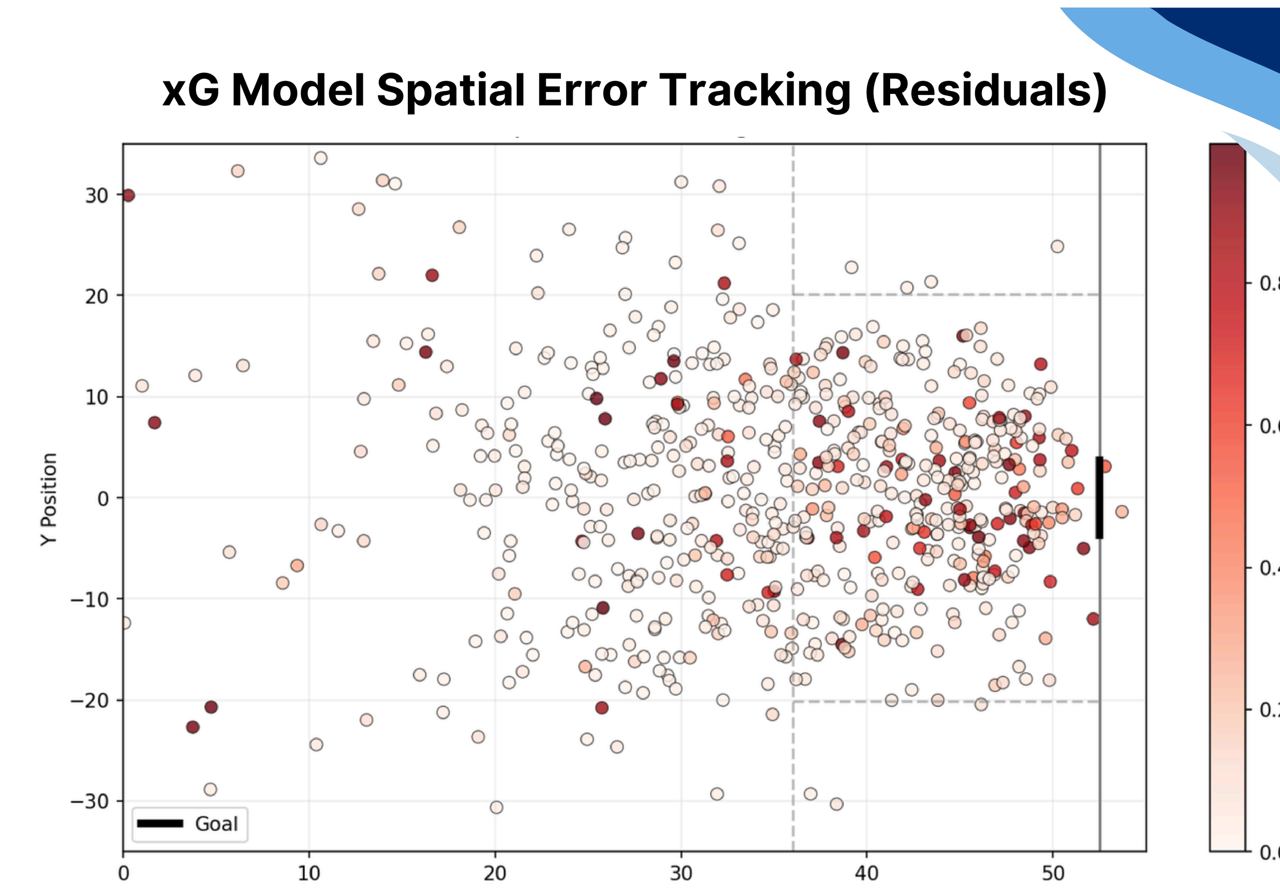
Shot Simulation with Tracking

The dashboard offers a number of features for teams to use, such as a replay tool that pinpoints pivotal game moments and a heat map of top kicking locations. Users also have access to a tactical board where players can be 'dragged and dropped' onto the field to test possible outcomes and formations. Two possible recommendations are listed above, with the top one showing the probabilities of a successful pass.

4 Performance Diagnostics

Diagnostic	Score	Target
Brier Score	0.0931	< 0.10
Log Loss	0.3261	0
ROC AUC	0.7640	> 0.75
F2 Score	0.3175	< 0.15
ECE	0.0305	< 0.02

The Brier score measures the mean squared error of the probabilistic predictions, scaling from 0.0 (perfect) to 1.0 (completely wrong). The model has a Brier Score of around 0.09. Spatial distribution of absolute error reveals the model struggles most with high-variance events inside the crowded 18-yard box.



Impact and Future Work

We expect this dashboard to help coaches analyze their players' decisions. The accessibility of our dashboard is ideal for smaller teams, such as university or high school teams, who may not have a group of analysts at their disposal. We hope to improve the performance of this model by using more advanced modeling and automating player identification. Our project concluded that decision-making in soccer is not about making the most aggressive play but about choosing the action that best fits the situation and the player's strengths to maximize long term scoring potential in a game.