Project Proposal

Team Name: Sports Analytics Club - Gatech

Group Members:

Emerald Singh - esingh60@gatech.edu
Sagar Pun - spun7@gatech.edu

Data Set:

https://www.kaggle.com/datasets/orkunaktas/all-football-players-stats-in-top-5-leagues-2324
https://www.kaggle.com/datasets/orkunaktas/all-football-players-stats-in-top-5-leagues-2425

Project Title: Predicting a Player's Goal and Assist Tally for the 24/25 Season Using 23/24 Performance Data

Project Summary:

This project aims to solve the real-world challenge of forecasting a football player's future offensive performance. We will build a model that takes a player's detailed performance statistics from the 2023/24 season to predict their combined goals and assists for the 2024/25 season. This is an interesting and valuable problem because accurate performance prediction is crucial for clubs in areas like player recruitment, squad building, and salary negotiation. Our goal is to determine if deep learning techniques can capture the complex, non-linear relationships in player data to create more accurate forecasts than simple intuition or traditional statistical models.

Project Description and Approach:

Dataset Rationale: We chose the football dataset because our team's shared domain knowledge allows for better data interpretation, feature selection, and validation of the model's results.

High-Level Objective: To accurately predict a player's Goals + Assists (G+A) using a model that is transparent in its logic and audited for fairness.

Three-Stage Approach:

1. Modeling: Use a Multi-Layer Perceptron (MLP) to predict a player's Goal and Assist total and measure its average prediction error using Mean Absolute Error (MAE).
2. Interpretability (XAI): Employ the SHAP framework to explain the model's "thinking," showing exactly which stats influence any individual player's prediction.

3. Fairness Audit: Check the model for bias by comparing its prediction error across different player positions and age groups to ensure equitable performance.

Resources/Related work and papers:

Our project is grounded in the established field of sports analytics but extends it by focusing on model interpretability, a crucial topic in responsible AI. While many models predict performance, few are audited for fairness or made transparent. Our work is therefore inspired by recent advancements in Explainable AI, particularly the development of model-agnostic methods like SHAP by Lundberg and Lee. By applying these techniques, we aim to build a model that is not just a "black box," but a tool that can provide justifiable and fair insights, which is a significant step beyond standard predictive modeling.

Lundberg and Lee's paper on SHAP: https://arxiv.org/abs/1705.07874