

Project Dynamic Player-Based Match Outcome Model

1. Introduction & Motivation

Traditional football prediction models focus on team-level statistics such as win rates, goals scored, goals conceded, league position. While these approaches capture historical performance, they ignore *who* will actually step on the pitch and *how each player is performing recently*.

But match outcomes depend heavily on player availability, form, interactions. For example:

- Removing one playmaker changes expected goals (xG) value drastically.
- A striker's recent finishing form affects expected conversion.
- Changing center-back pairings influences expected goals against (xGA).

This project aims to build a **player-centric prediction system** that updates match expectations based on *the specific lineup, player form, and how players combine together*.

2. Goal of the Project

1. **Predict match outcomes and team strength based on selected players**, not just historical team statistics.
2. Use **dynamic player performance metrics**, blending:
 - Career/season averages
 - Recent form / rolling averages
 - Player supporter score with sentiment analysis on social media
 - Chemistry indicators (e.g., number of matches played together)
3. Provide an **interactive UI** where users can:
 - Select 11 players for each team
 - Simulate match strength / predicted xG / win probability

- Compare different lineups

4. Data Features (Dynamic Performance Metrics)

Player Statistics

Specific data per position , for example for attackers expected Goals (xG) , shots ,...
Rolling average or average + form

Chemistry Metrics

Number of matches played together

5. Modeling Approach

Possible Model type:

Tabular machine learning model (XGBoost / LightGBM)

Training Input:

Each match → transformed into aggregated player-level features for both teams, e.g.:

- Specific metrics for each player according to his position :
- Or average per position
- Only consider starting player
- Chemistry scores
- Form indicators

Further Improvement if possible : **player embedding model**, where each player is represented as a learned vector encoding style, strengths, and role. Combined with an **attention mechanism**, the model could better capture non-linear interactions between players in the same team, improving prediction accuracy beyond traditional tabular features