Project Documentation Focused on Action Quality Evaluation Models

Project Introduction

As part of the PySport X SkillCorner Analytics Cup challenge, we have access to tracking data and Game Intelligence Dynamic Events for 10 Australian A-League matches from the 24/25 season.

The objective of this project is to visualize player performance using data from two models related to the evaluation of on-ball and off-ball actions:

- **xThreat**: The probability that a goal will be scored within 10 seconds if a given player were the target of a pass at a specific moment.
- **Expected Possession Value**: The probability that a team will score within the next 90 seconds or before the ball goes out of play.

Data Extraction

Using the list of match identifiers, we first retrieve all game events from these matches, filtered through a predefined list of variables of interest covering basic information (involved player, event type, action success, etc.), as well as variables linked to the xThreat model (xthreat, player_targeted_xthreat) and the Expected Possession Value model (possession_danger, beaten_by_possession, beaten_by_movement, stop_possession_danger, reduce_possession_danger).

In parallel, using the associated match JSON files, we extract personal information for each player who participated in these matches (name, position, team), along with their total accumulated playing time across the database.

We also extract basic match information available in the database (identifier, team names, date), which will later be useful for listing available matches. These data are stored in the info_matches directory.

Data Preprocessing

With these elements collected, we aggregate all data at the match level. This work relies on the computation of four statistical indicators derived from the two evaluation models:

- **Top Movement**: Measures the total danger generated (via xThreat) by a player through their positioning as a passing option for teammates.
- **Top Off Ball Runs**: Measures the danger generated (xThreat) by a player through their off-ball runs.
- **Top Good Decisions**: Represents the percentage of optimal passing decisions made by the ball carrier relative to available passing options, evaluated using xThreat values. A player must have at least two pass options to be included.
- **Top Press**: Evaluates the quality of a player's pressing by measuring its influence on reducing the opponent's scoring threat. Effective pressure that cancels dangerous actions is rewarded, while unsuccessful pressure is penalized.

In addition to computing these indicators for each player and match, we break them down by different contextual situations:

- **Phase of play type**: "build_up", "create", "finish", "direct", "quick_break", "transition", "set_play", "chaotic"
- **Off-ball run type** (Top Off Ball Runs only): "behind", "coming_short", "cross_receiver", "dropping_off", "overlap", "pulling_half_space", "pulling_wide", "run_ahead_of_the_ball", "support", "underlap"
- **Match state**: "winning", "drawing", "losing"
- **Pressing type** (Top Press only): "pressing", "pressure", "counter_press", "recovery_press", "other"
- **Player speed at action time**: "jogging", "running", "hsr", "sprinting"
- **Pitch zone**: "left", "center", "right", "middle_third", "attacking_third", "penalty"

All statistics are stored in the matches directory, one file per match identifier.

From this data, we also compute the same variables aggregated by team (directory *teams*) and by player (directory *players*), weighting statistics by each player's playing time and by the number of decisions made in the case of the *Top Choice* indicator.

Page Structure

The application contains four pages: a home page briefly summarizing the project's objective, and three pages dedicated to the available types of analysis: match analysis, team analysis, and player analysis.

Arborescence du projet pour la mise en place de l'application

Application
```
├── data_processing.ipynb : File used for data processing
├── application.py: Main : Application layout file
├── src : image, data, documentation
├── README.md
└── requirements.txt : List of library dependencies
```

Application Setup

The application is available in two languages: French and English.

Home page

As mentioned earlier, the home page provides a brief overview of the project components and includes access to several resources (documentation, project source code, link to SkillCorner datasets).

Analysis section

Project Header

We first create several functions (analysis table builders, glossary construction, etc.) as well as various variables (glossary text, categorized data lists) to avoid cluttering the main application code, especially considering the availability of two languages and three analysis modes with many structural similarities.

Application display

Depending on the selected analysis type, the sidebar prompts the user to choose a match, team, or player. This selection determines the tables displayed across five sections:

- **Overview**: displays the four global indicators.
- **Top Movement / Top Off-Ball Runs / Top Good Decisions / Top Press**: each section displays the global statistic, along with a filter (phase of play, match state, off-ball run type, pressing type, or player speed).

For player and match analysis, the application displays the top five players, including their position, team, and value for the selected metric.

For team analysis, if the selected player does not appear in the Top 5, the application displays the Top 4 plus the selected player, showing their ranking for the statistic.

For team-based analysis, all 12 teams in the database are displayed, ranked according to the chosen metric.

Additionally, the selected team or player is highlighted to emphasize the user's selection. A glossary specific to the active section is also available in the sidebar.