

# Documentation du projet axé sur les modèles d'évaluation de qualité des actions en Français

## Introduction au projet

Dans le cadre du challenge PySport X SkillCorner Analytics Cup, nous disposons des données de tracking et de Game Intelligence Dynamic Events pour 10 matchs d'Australian A-League sur la saison 24/25.

L'objectif de ce projet sera de visualiser les performances des joueurs à travers l'utilisation des données de 2 modèles liés à l'évaluation des actions de jeu :

- xThreat : Probabilité qu'un but soit marqué dans les 10 secondes si un joueur donné était la cible d'une passe à un moment donné.
- Expected Possession Value : Détermine les chances qu'une équipe marque dans les 90 secondes suivantes ou avant que le ballon ne sorte du terrain.

## Extraction des données

À partir de la liste des identifiants de match, nous allons dans un premier temps récupérer l'ensemble des événements de jeu de ces matchs, en filtrant selon une liste définie de variables d'intérêt concernant les informations de base (joueur impliqué, type d'événement, réussite de l'action...) ainsi que les variables liées aux modèles xThreat (xthreat, player\_targeted\_xthreat) ou Expected Possession Value (possession\_danger, beaten\_by\_possession, beaten\_by\_movement, stop\_possession\_danger, reduce\_possession\_danger).

En parallèle, via les fichiers JSON des matchs associés, nous récupérerons les informations personnelles de chaque joueur ayant participé à ces matchs (nom, poste, équipe), ainsi que leur temps de jeu cumulé sur l'ensemble de la base de données.

Nous extrairons également les informations de base de chaque match présentes dans la base de données (identifiant, noms des équipes, date), qui nous seront utiles plus tard pour lister les matchs disponibles. Par ailleurs, ces données figurent dans le dossier info\_matches.

## Pré-traitement des données

Maintenant que nous disposons de ces informations, nous allons agréger les données par match. Ce travail repose sur la construction de quatre indicateurs statistiques issus de ces deux modèles :

- **Top Movement** : indicateur cumulant le danger généré (via xThreat) par un joueur grâce à son positionnement comme option de passe pour ses coéquipiers.
- **Top Off Ball Runs** : indicateur cumulant le danger généré (xThreat) par un joueur via ses courses sans ballon.
- **Top Choice** : représente le pourcentage de bons choix de passe effectués par le porteur du ballon au regard des options disponibles, mesurées via le xThreat associé

aux coéquipiers. À noter qu'un joueur doit disposer d'au moins deux options de passe pour être comptabilisé.

- **Top Press** : indicateur évaluant la qualité de la pression exercée par un joueur, en mesurant son influence sur le danger de but adverse. Une pression annulant une action dangereuse est valorisée, tandis qu'une pression manquée est pénalisée.

En plus du calcul de ces différentes statistiques pour l'ensemble des joueurs et pour chaque match, nous les déduirons également selon plusieurs situations de jeu :

- **Type de phase de jeu** : "build\_up", "create", "finish", "direct", "quick\_break", "transition", "set\_play", "chaotic"
- **Type de course sans ballon** (uniquement pour *Top Off Ball Runs*) : "behind", "coming\_short", "cross\_receiver", "dropping\_off", "overlap", "pulling\_half\_space", "pulling\_wide", "run\_ahead\_of\_the\_ball", "support", "underlap"
- **État de jeu** : "winning", "drawing", "losing"
- **Type de pression** (uniquement pour *Top Press*) : "pressing", "pressure", "counter\_press", "recovery\_press", "other"
- **Vitesse du joueur au moment de l'action** : "jogging", "running", "hsr", "sprinting"
- **Zone du terrain** : "left", "center", "right", "middle\_third", "attacking\_third", "penalty"

L'ensemble de ces statistiques sera disponible dans le dossier *matches*, pour chaque rencontre selon son identifiant.

À partir de ces données, nous déduirons également les mêmes variables agrégées par équipe (dossier *teams*) et par joueur (dossier *players*), en tenant compte du temps de jeu de chaque joueur dans l'agrégation de ces statistiques, ainsi que du nombre de choix effectués par match pour l'indicateur *Top Choice*.

## Liste des pages

Ainsi, nous disposerons de quatre pages : une page d'accueil récapitulant brièvement l'objectif du projet, ainsi que trois pages dédiées aux différents types d'analyses possibles : par match, par équipe et par joueur.

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

### Application

- data\_processing.ipynb : Fichier utilisé pour le traitement des données
- application.py: Main : Fichier utilisé pour la mise en page de l'application
- src : image, data, documentation
- README.md
- requirements.txt : Fichier de dépendances de librairies

## Mise en place de l'application

Il est important de préciser que l'application contiendra 2 versions : Français et Anglais.

## Accueil

Comme expliqué précédemment, la partie de l'accueil présentera brièvement les composantes du projet, et donne accès à diverses ressources (Documentation, Code source du projet, Lien pour accéder aux données de SkillCorner).

## Section analyse

### En tête du projet

Nous créerons au préalable plusieurs fonctions (construction des tableaux d'analyse, glossaire, etc.) ainsi que diverses variables (glossaire, listes de données par catégorie) afin d'éviter de surcharger la longueur du projet, compte tenu des deux langues disponibles et des trois types d'analyses présentant de nombreuses similarités.

### Affichage de l'application

Selon le type d'analyse, la barre latérale demandera à l'utilisateur de sélectionner le match, l'équipe ou le joueur de son choix. Cette sélection déterminera les tableaux affichés, répartis en cinq sections :

- **Vue d'ensemble** : présentant les quatre indicateurs globaux.
- **Top Movement / Top Off-Ball Runs / Top Choice / Top Press** : chaque section contient la statistique globale correspondante, ainsi qu'un filtre au choix (phase de jeu, état du match, course sans ballon, type de pression ou vitesse du joueur).

Pour l'analyse des joueurs et des matchs, nous afficherons les informations des cinq meilleurs joueurs : leur poste, leur équipe et leur valeur pour la statistique demandée. En ce qui concerne l'analyse des équipes, si le joueur sélectionné ne fait pas partie du Top 5, nous afficherons le Top 4 ainsi que le joueur choisi, accompagné de sa position dans le classement pour cette statistique. Pour l'analyse des équipes, nous afficherons l'ensemble des équipes disponibles dans la base de données, soit douze, classées selon la statistique sélectionnée. Par ailleurs, l'équipe ou le joueur choisi sera mis en couleur afin de mettre en valeur la sélection de l'utilisateur. Enfin, un glossaire adapté à la section affichée sera disponible dans la barre latérale.