

Advanced Metrics : Développement de métriques innovantes pour l'analyse de performances (Football)

Description du projet et objectifs :

Le projet « **Advanced Metrics** » est conçu pour explorer et développer des indicateurs innovants qui vont au-delà des statistiques traditionnelles.

L'objectif est de créer des indicateurs de performance plus sophistiqués à partir de nos données brutes (tracking et Events), permettant une analyse approfondie des aspects individuels et collectifs du jeu.

Données et Ressources Disponibles :

Vous trouverez dans le repo ci-dessous les event data brutes de nombreuses compétitions : <https://github.com/statsbomb/open-data/tree/master>

Pour rappel, **les données d'évènements ou "Event Data"** : Ces données représentent tous les évènements survenant pendant un match, tels que les passes, les tacles et les tirs. Ces données permettent de mesurer les performances techniques des joueurs et fournissent également des informations tactiques précieuses.

Conseil : Garder les données d'une grande compétition pour faire les tests et valider vos métriques (Par exemple : "competition_id" : 43 qui correspond à la dernière coupe du monde)

Description des métriques à créer :

1) Expected Threat (xT)

Description : <https://soccerment.com/expected-threat/>

Le terrain est découpé en un cadrillage de zone de danger. Plus on est proche du gardien, plus on considère que la zone est dangereuse. Pour chaque event (passes, dribbles, centres...), on prend les coordonnées de départ de l'action et on les compare avec les coordonnées d'arrivée. Si le ballon est amené dans une zone plus dangereuse, le joueur sera accrédité pour cette action d'un xT positif (avec une valeur dépendant de à quel point la zone d'arrivée est dangereuse).

Bonus :

Créer une métrique similaire qui évalue la qualité d'un défenseur à annihiler une action de but en suivant une approche similaire

2) VAEP (Valuing Actions by Estimating Probabilities)

Description : <https://dtai.cs.kuleuven.be/sports/vaep>

Approche similaire à l'xT, mais en enrichissant chaque action du contexte des actions passés. Approche très intéressante pour extraire un delta offensif (proba que l'action augmente les chances de marquer) et d'un delta défensif (proba que l'action diminue les chances de marquer) pour chaque event.

Deux modèles de Machine Learning (xgboost ou catboost) sont construits indépendamment pour les delta offensif et défensif.

L'objectif de créer 3 métriques :

- VAEP (générale) des actions d'un joueur
- VAEP (Dribbles) qui mesure l'impact des dribbles d'un joueur
- VAEP (Carries) qui mesure l'impact des percées ballon au pied d'un joueur

Bonus :

Améliorer la métrique en prenant en compte le fait que l'action soit réussie ou non (Par exemple : Ne pas attribuer un VAEP important à un dégagement s'il conduit à une perte de balle)

3) Rating

Description :

L'objectif de cette métrique est de vous laisser le champ libre afin de construire une note de la performance globale d'un joueur durant un match.

Piste à explorer :

Une option possible serait de créer une note à partir du VAEP (normalisation pour se ramener à une note entre 0 et 10).

Or, une limitation est que le VAEP met trop en avant les buts. Pour un event classique, la note de VAEP est environ de 0.1/0.2, alors que pour un but on est proche de 1. Le classement des joueurs ordonné par ce rating est donc fortement corrélé au classement des buteurs.

Autre limitation du VAEP : les notes des défenseurs risquent d'être lissées et donc toutes similaires (vu que moins d'impacts directs et prononcés sur les chances de buts et d'en concéder).

Il est donc important de :

- Réduire l'impact des buts sur l'xG
- Réduire l'impact des dégagements et jeu long qui rend la balle à l'adversaire.

4) Passes Allowed Per Defensive Action (PPDA)

Description: <https://statsbomb.com/articles/soccer/defensive-metrics-measuring-the-intensity-of-a-high-press/>

L'objectif de cette métrique est de mesurer l'intensité du pressing d'une équipe : plus la note est basse, plus le pressing de l'équipe est intense.

En effet, le PPDA (pass per defensive actions) peut être considéré comme une notion de pressing (au moins pressing réussi). Plus la valeur est faible mieux c'est. Mais cette notion est fortement corrélée à la possession (si une équipe a une dominance marquée sur le jeu, les adversaires vont faire moins de passes).

D'autres métriques peuvent être ajoutés au PPDA tel que le % de passes réussies par l'adversaire.

Missions et livrables attendus :

- **Rapport de description des métriques :** Pour chaque nouvel indicateur créé, décrire l'approche, les limites identifiées et les pistes d'améliorations envisagées.
- **Analyse de Pertinence des Métriques :** Un rapport d'analyse comparative des nouvelles métriques, incluant des études de cas avec des joueurs connus pour valider leur efficacité et leur pertinence dans des contextes réels.

Possibilité de s'inspirer des modèles suivants :



