

Documentation du projet Player Performance en Français

Introduction au projet

Nous analyserons les performances d'un joueur au cours de la saison 24/25 en fonction de différents facteurs (buts, passes décisives, tacles) et prismes (âge, prix, nationalité, etc.).

Pour réaliser ce sujet, nous utiliserons deux sources : Fbref pour les statistiques et TransferMarkt pour la valeur marchande de chaque joueur.

Extraction des données / Pré-traitement des données

Après avoir importé les librairies importantes pour cette partie, on va dans un 1^{er} temps récupérer les données provenant d'utilisateurs de Kaggle ayant réalisé du web scraping sur les sites Fbref : <https://www.kaggle.com/datasets/hubertsidorowicz/football-players-stats-2024-2025> et TransferMarkt : <https://www.kaggle.com/datasets/davidcariboo/player-scores> .

Concernant le fichier de TransfertMarkt, on filtrera les données afin de ne conserver uniquement les joueurs ayant participé à la saison 2024/2025, et ceux dans l'un des 5 grands championnats européens (Ligue 1, Premier League, Bundesliga, Serie A et Liga). Pour le fichier de Fbref, on effectuera notamment une transformation des données nécessitant d'être calculer par 90 minutes. De plus, nous prendrons en compte uniquement les joueurs ayant joué plus de 600 minutes. À noter que les joueurs ayant joué dans 2 clubs au cours de la saison auront leurs statistiques cumulées ou agrégées selon le type de variable.

Une étape majeure du pré-traitement des données a été d'associer les deux fournisseurs de données pour chaque joueur. Pour cela, nous avons utilisé plusieurs stratagèmes pour englober le maximum de joueur sans se tromper. La 1^{ère} a été d'associer les joueurs ayant le même nom + même championnat + même année de naissance + poste si c'était un gardien (seule poste où aucune différence ne subsiste entre les fournisseurs). Cette phase a englobé la majorité des joueurs mais étant donné que nous avons remarqué certaines variations dans le nom entre les deux fournisseurs, il a fallu aller plus loin que cette 1^{ère} étape. Ainsi, nous utiliserons une fonction de ressemblance de caractère (fuzzy) en suivant cette logique : si le % de ressemblance est satisfaisant + même championnat + même année de naissance + poste si c'était un gardien, alors on est considéré que c'est bien le même joueur. Nous avons choisi de séparer les phases en plusieurs parties afin de limiter le nombre d'erreur, en voici le détail avec le nombre de joueurs récupérer dans chaque phase :

- Noms identique (même année, ligue, poste gardien) : 1645
- Matches à 90% (même année, ligue, poste gardien) : 39
- Matches à 75% (même année, ligue, poste gardien) : 42
- Matches à 90% (même ligue, poste gardien) : 4
- Matches à 65% (même année, ligue, poste gardien) : 13
- Matches à 80% (même ligue, poste gardien) : 4
- Matches à 60% (même année, ligue, poste gardien) : 8
- Total appariés : 1751.

À noter que logiquement, à mesure que l'on baisse le % de matches sur le nom, moins il y a de chances que l'on procède à la bonne association entre les deux fournisseurs. Nous avons ainsi pris la décision de ne pas descendre en dessous des 65 % de matches, conservant une association sans erreur.

Cette liste de joueurs, accompagné des colonnes d'intérêt pour la suite sera sauvegardé dans un fichier csv nommé `database_player.csv`.

Liste des pages

L'application sera composée de 4 pages :

- L'accueil : Présentation du projet et de ses ressources
- Analyse d'un Joueur : Analyse du joueur de votre choix à travers plusieurs statistiques
- Comparaison entre Joueurs : Analyse comparative entre deux joueurs du même poste
- Classement des joueurs : Classement des joueurs par performance selon une statistique choisie

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

```
Application
├── image
├── CV
├── data
├── documentation
├── Player_performance.py: Main
├── README.md
└── requirements.txt
```

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, CV)

Section analyse

En tête du projet

Plusieurs fonctions seront créées comme l'adaptation du format de la valeur des joueurs sur le marché, la traduction des pays ou des postes, la catégorisation des postes en sous-catégorie (avec la liste des statistiques qu'on utilisera), ou encore la création d'une fonction calculant la

ressemblance entre chaque joueur ayant le même poste à partir de leurs statistiques avancées sur toute la saison

Affichage de l'application

Concernant la page d'accueil, on affichera le logo du projet, suivi de l'explication du projet, et de ressources associées. Plusieurs fichiers seront téléchargeables comme la documentation et le CV.

Pour l'analyse du joueur, premièrement, on demandera à l'utilisateur de choisir le joueur de son choix. Cette page contiendra le profil du joueur avec ses informations de base (Nom, Photo, Poste, Club ect...), son radar statistique, et une table des 5 joueurs le ressemblant le plus statistiquement. À noter que le radar peut être adaptable selon le pays, championnat, tranche d'âge. Les statistiques affichées sur le radar sont celles semblant d'intérêt par rapport au poste du joueur (en bleu), et sera comparé avec la médiane du groupe choisie (en rouge). Par ailleurs, un glossaire des statistiques sera disponible, et dépliable si besoin.

La page de comparaison de joueur suivra la même logique avec la possibilité de choisir deux joueurs (du même poste) que l'utilisateur souhaite analyser. Leur profil sera affiché côte à côte, suivi de leur radar respectif (en bleu et rouge).

À propos de la page de classement, il sera demandé à l'utilisateur la statistique qu'il souhaite analyser. À partir de ce choix, un classement des meilleurs joueurs selon cette métrique sera affiché avec un podium pour les 3 premiers, suivie par la suite du classement sous forme de tableau avec les informations de base sur ces derniers. La sidebar contiendra notamment des filtres facultatifs (Poste, Club, Championnat, Tranche d'Age, Valeur sur le Marché), ainsi qu'un glossaire des statistiques. Par ailleurs, une image sera affichée avant le choix de l'utilisateur, afin de ne pas laisser vide la page (comme cela est le cas pour chaque page).