

Fusionner les données récupérées pour les différents joueurs

20 novembre 2023

1 Basketball reference

Pour chaque joueur, on crée un vecteur d’une plus haute dimension en fusionnant les données récupérées de *basketball-reference*. J’utilise ici l’exemple de Jaren Jackson Jr. qui a gagné le DPOY en 2023 :

- 2023 :
 - Afin de pouvoir les données de certaines saison, on commence avec la saison : “2023”,
 - puis on ajoute le nom, son rang dans le vote pour DPOY : “1”,
 - puis ses statistiques “per Game” pour la même saison,
 - les statistiques “Per 100 Poss” en supprimant “Age”, “Tm”, “Lg”, “Pos”, “G”, “GS”, “MP” qui sont redondant avec l’information “per Game”,
 - les statistiques “Advanced” en supprimant “Age”, “Tm”, “Lg”, “Pos”, “G”, “GS”, “MP” qui sont redondant avec l’information “per Game”,
 - puis on identifie son équipe : “Mem” et on ajoute leurs statistiques “Team and Opponent Stats”, mais seulement les lignes “Team” et “Opponent”,
 - et enfin la ligne “Team” des statistiques “Team Misc”
 - 2022 : dans cette saison Jaren Jackson Jr avait le rang 5 dans le vote, on ajoute les autres statistiques comme vu avant
 - 2021 : dans cette saison Jaren Jackson Jr n’était pas dans la liste des joueurs ayant reçu des votes pour DPOYs, son rang sera donc -1 (comme pour tous les autres joueurs qui n’ont pas reçu des votes)
- A ces informations s’ajoutent les statistiques “narratives” :
- 2023 :
 - As-t-il gagné le DPOY pendant la saison précédente ? : “false”
 - Combien de fois as-t-il déjà gagné DPOY ? : “0”
 - Quel est le meilleur tour que son équipe a achevé pendant les playoffs d’une saison où il gagné DPOY (entre 0 et 5) ? : “1”
 - Combien d’autres joueurs dans son équipe ont reçu des votes pour DPOY pendant une saison précédente (ou la même saison) ? : “0”

2 Ajouter les informations des autres sources

Il faut utiliser les noms via une comparaison des chaînes de caractères afin de récupérer les bonnes lignes par saison et les aussi ajouter, en supprimant des informations comme “Age”, “GP”, “W”, “L” etc. qui sont redondantes avec celles déjà incluses.