



# Et si la performance d'un gardien ne se résumait pas à son taux d'arrêt ?

Analyse et développement de nouveaux indicateurs de performance pour les gardiens de but au handball



## Introduction

Le handball, en plein essor depuis son entrée aux JO en 1972, met souvent les joueurs de champ en avant, tandis que le rôle du gardien reste sous-évalué. Pilier défensif dont les interventions peuvent changer l'issue d'un match, sa performance est mesurée par des statistiques simples, qui ne prennent pas en compte la difficulté des tirs ou leur contexte.

Cette étude propose de nouveaux indicateurs pour mieux évaluer l'impact réel des gardiens sur le jeu en combinant analyse descriptive et apprentissage supervisé. L'objectif est d'améliorer l'analyse de la performance et le recrutement de talents.

## Les données

Les données de ce projet, collectées manuellement par la start-up Data7&Match, couvrent tous les matchs de la saison 2020-2021 de la plus haute ligue française de handball.



239 rencontres



16 équipes



21 894 tirs tentés  
91 tirs par match en moyenne



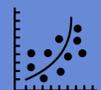
13 202 buts  
52 buts par match en moyenne



24 % d'arrêt en moyenne sur les tirs cadrés

→ 60,3 % de buts

## Méthodes d'apprentissage supervisée pour la création de plusieurs indices de performance des gardiens



### Régression logistique bayésienne

Cette méthode permet d'estimer la probabilité qu'un tir aboutisse à un but en prenant en compte l'incertitude liée aux joueurs et aux autres paramètres d'un match. Cette approche permet d'obtenir des prédictions plus nuancées et robustes, particulièrement utiles dans un contexte de faible nombre d'observations (les matchs sur une seule saison).



### Arbres de décision

Les arbres de décision (basé sur l'algorithme CART) permettent de séparer les actions qui mène à un tir en groupes selon des paramètres comme la zone de tir ou la posture du tireur, pour prédire la probabilité qu'une action se termine par un but. La méthode est utile pour construire un modèle d'expected goals.

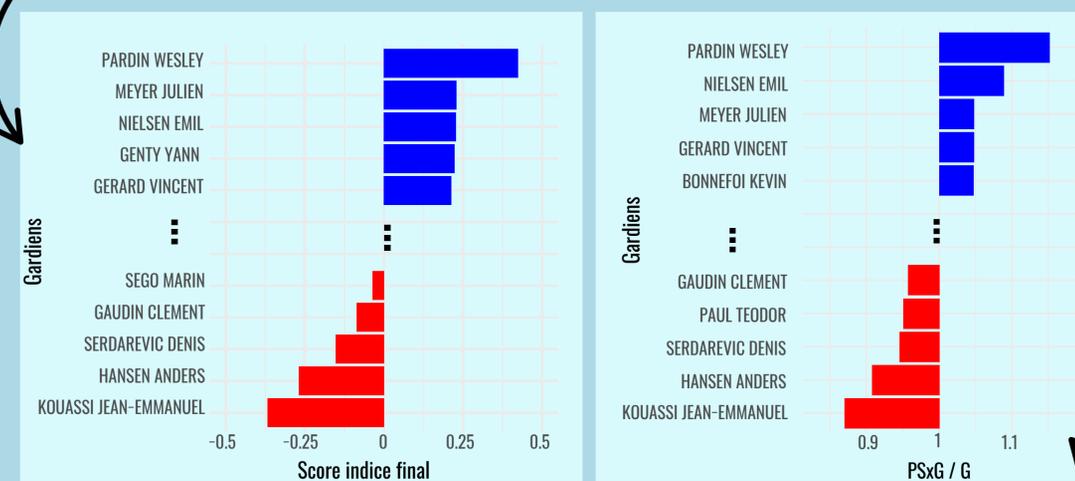
## Des indicateurs de performance ayant des résultats similaires

### Un premier indicateur : l'indice final

- Deux méthodes utilisées : la régression logistique bayésienne et les arbres CART, ce qui crée :
  - L'indice bayésien, modélisant la probabilité d'arrêt en fonction des caractéristiques du tir, avec effets aléatoires pour capturer la variabilité individuelle des gardiens.
  - L'indice CART, prédisant le pourcentage d'arrêts attendu selon les caractéristiques des tirs subis, et le comparant au pourcentage réel, ce qui donne un score reflétant la sur- ou sous-performance du gardien.
- Optimisation des poids  $\lambda_1$  et  $\lambda_2$  en minimisant la RMSE\* pour procéder à une combinaison linéaire des deux indices et obtenir un indice final :  $\text{Indice Final} = \lambda_1 \times \text{Indice bayésien} - \lambda_2 \times \text{Indice CART}$ .
- Les **5 meilleurs** et les **5 moins bons** gardiens selon cet indice final sont représentés dans le graphique de gauche ci-dessous.



\*RMSE = Root Mean Square Error

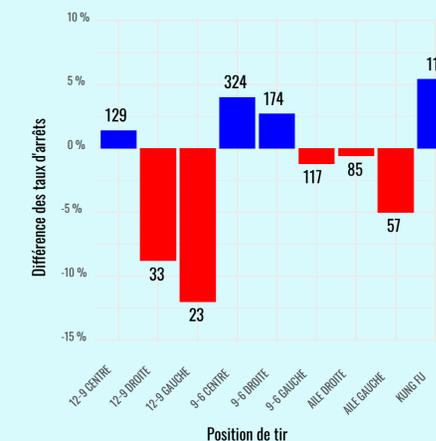


### Un second indicateur: le PSxG/G

- La méthode utilisée : les modèles **Post-Shot expected Goals (PSxG)**, inspirés des modèles d'expected Goals (xG) qui calculent la probabilité qu'un tir se transforme en but. Ses caractéristiques :
  - Spécifiquement conçu pour évaluer les gardiens.
  - Prise en compte de la difficulté d'un tir, notamment le type du tir (direct, avec rebond, lob) et sa localisation dans le but.
- Addition, pour un gardien donné, des probabilités de but associées à chaque tir cadré qu'il encaisse au cours de la saison, pour estimer le nombre de buts qu'il était censé encaisser.
- Calcul du rapport **PSxG / Buts encaissés (G)** :
  - Si  $> 1$  : le gardien a arrêté plus de tirs que prévu compte tenu de leur difficulté : il a donc **surperformé**.
  - Si  $< 1$  : le gardien a arrêté moins de tirs que prévu compte tenu de leur difficulté : il a donc **sous-performé**.
- Les **5 meilleurs** et les **5 moins bons** gardiens selon cet indice final sont représentés dans le graphique de droite ci-dessus.

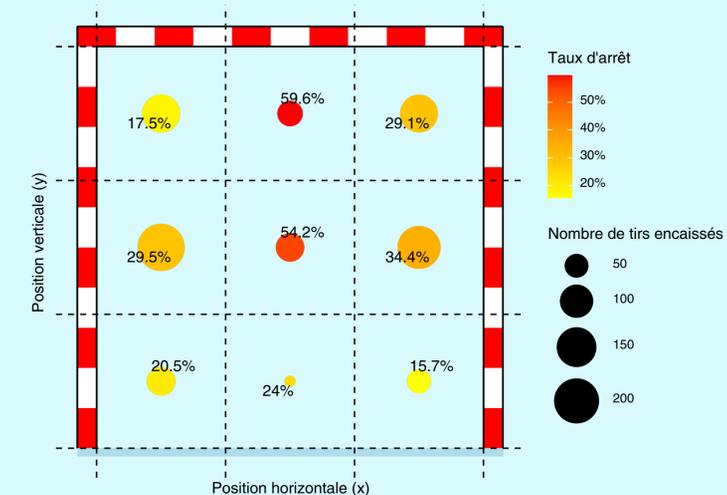
## L'évaluation de la performance sur une saison

Prenons l'exemple de Mate Šunjić, gardien de l'US Ivry Handball lors de la saison 2020-2021. Ses performances sont comparées aux moyennes des autres gardiens par position de tir pour identifier ses points forts et faibles.



Note de lecture: Dans le secteur 12-9 Gauche (entre 9 et 12 mètres à gauche de la cage), la différence entre le taux d'arrêt de Mate Šunjić et le taux d'arrêt moyen de tous les gardiens est de -12 % sur les 23 tirs reçus dans cette zone.

Le graphique suivant présente les taux d'arrêt de Mate Šunjić, répartis sur les 9 compartiments de la cage, permettant de visualiser sa performance dans différents secteurs.



Note de lecture : Pour la case de la position verticale du milieu et de la position horizontale à droite, Mate Šunjić a réalisé un taux d'arrêt de 34.4 % pour un nombre de tirs reçu d'environ 150. (attention à l'interprétation : tirs encaissés = tirs reçus et non tirs se concrétisant par un but)