

Problématique 1

Description : Dans une équipe de football, les joueurs sont répartis en 4 catégories selon le poste occupé: gardien de but, défenseur, milieu de terrain et attaquant.

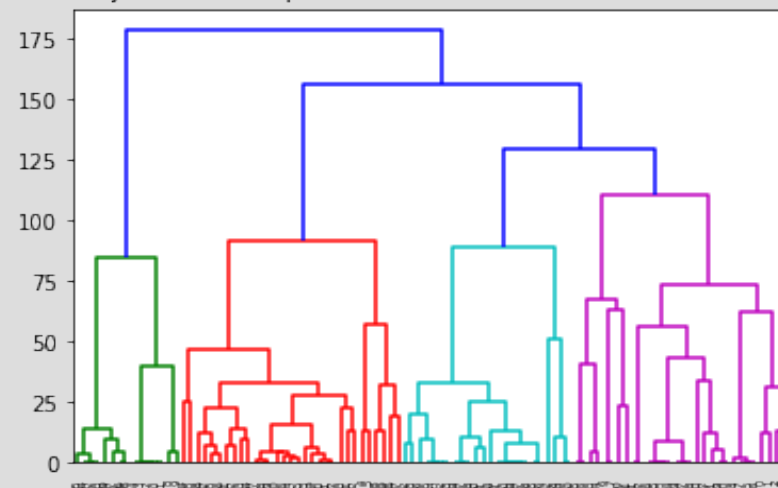
L'objectif final est de déterminer pour un joueur donné (même généré aléatoirement) sa catégorie.

Méthodes : Clustering hiérarchique et algorithme des k-moyennes puis prédiction à partir des clusters obtenus.

Résultats :

- Clustering hiérarchique sur les joueurs.

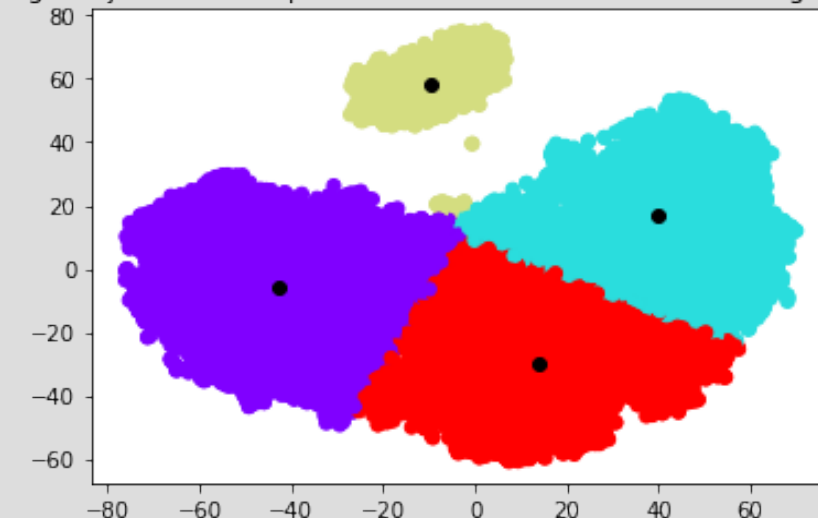
Dendrogramme des joueurs: attaquants, défenseurs, milieu de terrain et gardiens de buts



Ce clustering fait apparaître 4 groupes : gardien, défenseur, milieu de terrain et attaquant (de gauche à droite).

- Résultat de l'algorithme des k-moyennes (k=4).

Clustering des joueurs: attaquants, défenseurs, milieu de terrain, gardiens de buts



Formation de 4 clusters correspondant aux 4 catégories précédemment définies.

Problématique 2

Description : Utilisation des caractéristiques des joueurs pour la prédiction du résultat d'un match (gagnant, perdant, match nul).

Méthodes :

Un classifieur (PlayerPrediction) entraîné sur un ensemble de matchs et de joueurs (Caractéristiques) pour calculer un score pour chaque équipe qui est la moyenne des scores des joueurs qui la composent.

$$1. Score_{joueur} = \frac{1}{N} \sum_{i=1}^N Potentiel_{joueur}[i]$$

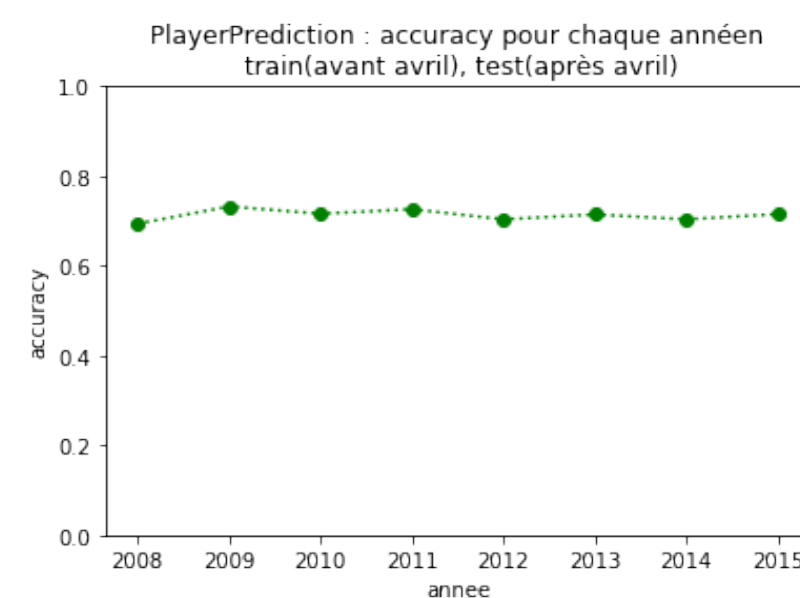
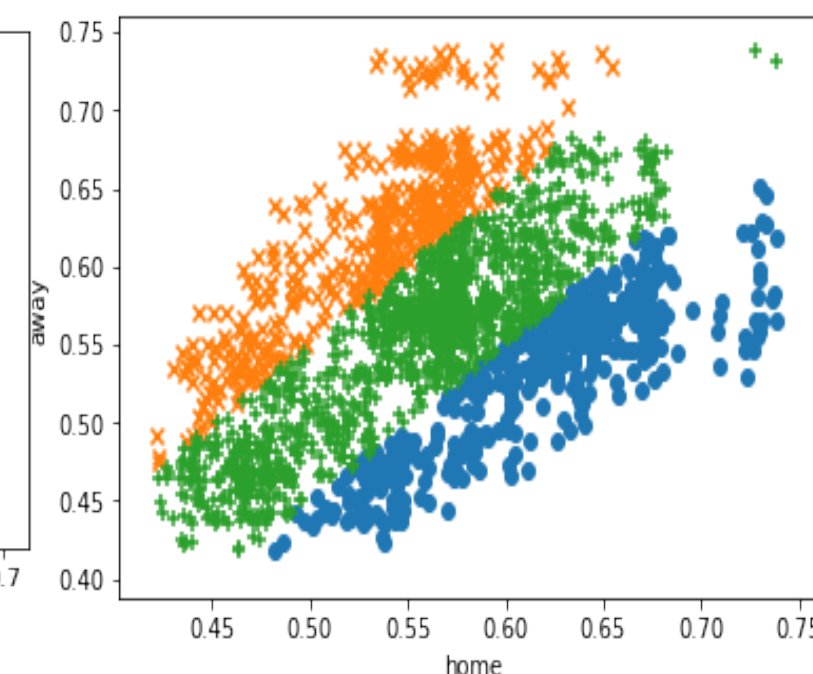
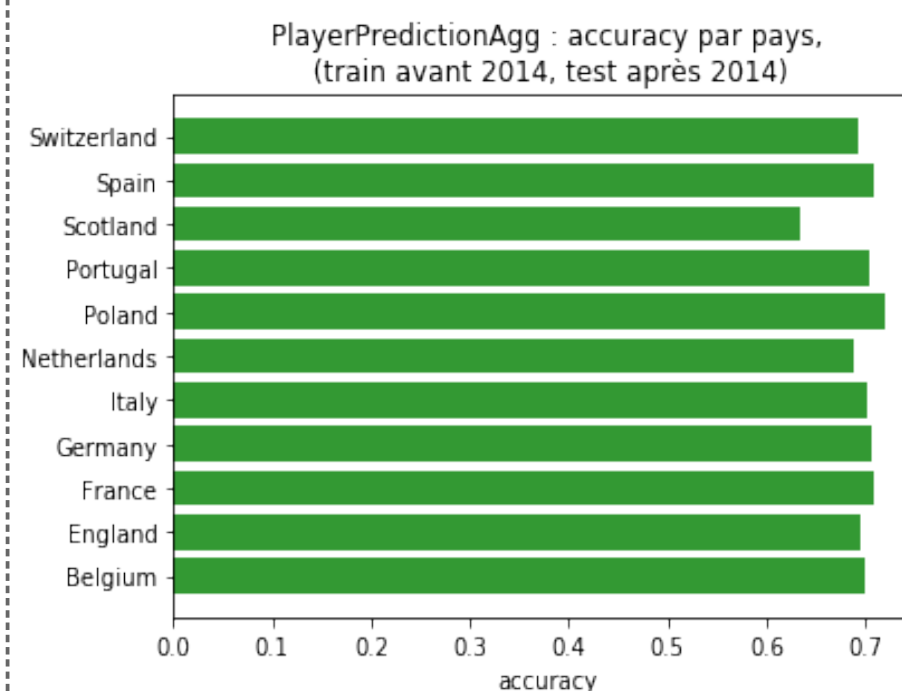
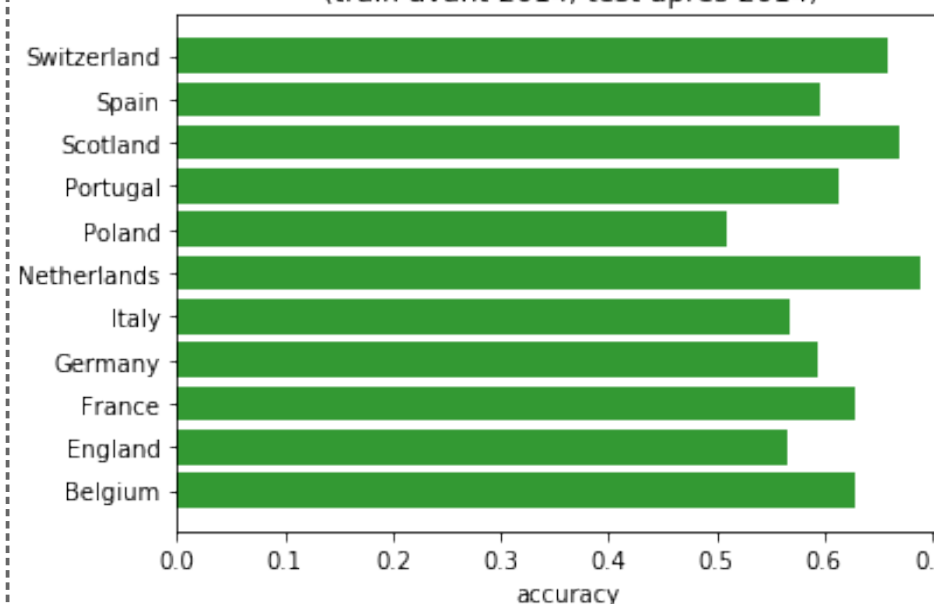
$$2. Score_{equipe} = \frac{1}{N} \sum_{i=1}^N Score_{joueur}[i]$$

Pour la prédiction du résultat d'un match, on calcule un score pour chaque équipe (home, away) qui est :

$$Score_{match}^{equipe} = Score_{equipe} \times \frac{1}{11} \sum_{i=1}^{11} Potentiel_{joueur_sur_le_terrain}[i]$$

On compare les scores des deux équipes pour avoir le résultat prédit (1 : match nul, 2 : victoire home, -2 : victoire away).

Résultats :



Problématique 3

Description : Étude de la fiabilité des côtes avant match des sites de paris sportifs.

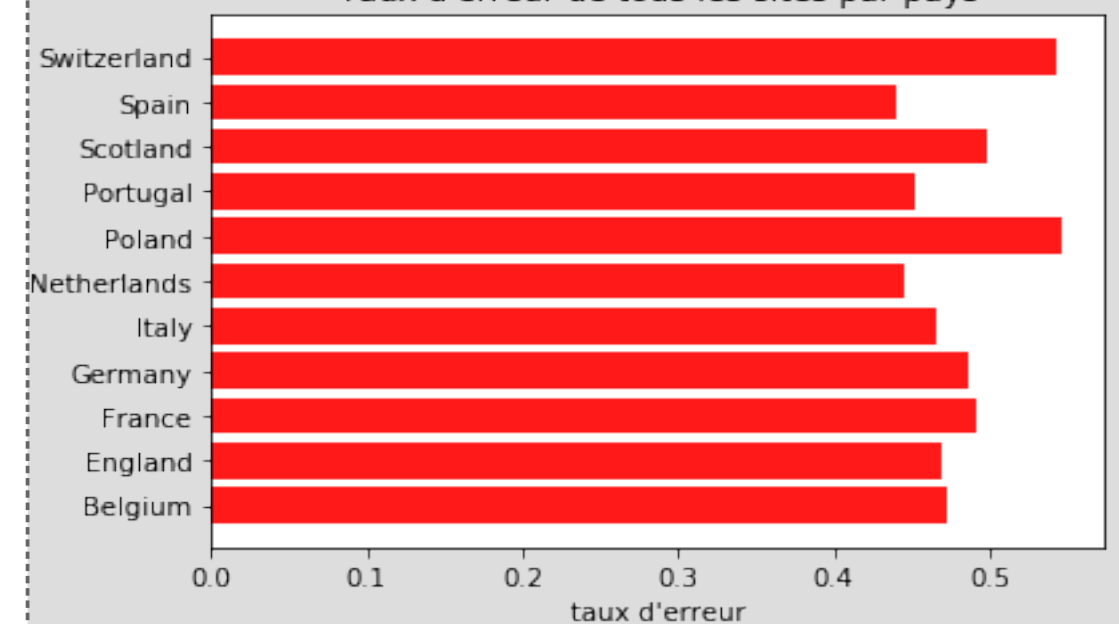
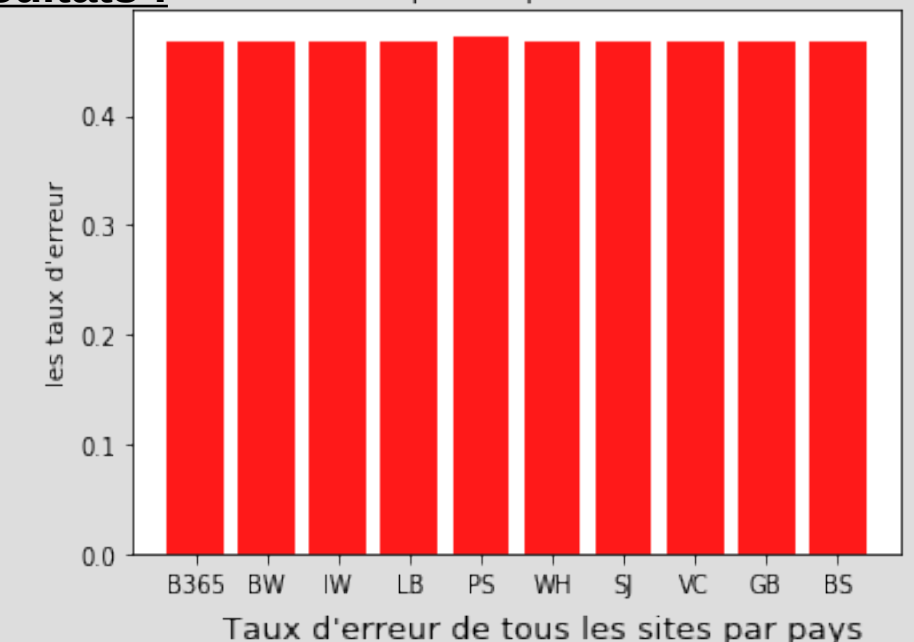
Méthodes :

- Pour **chaque site** et pour **chaque** match, on récupère les côtes d'avant match. Selon la côte minimale on déduit le résultat probable et on le compare avec le résultat de fin du match.
- Pour **tous les sites** et pour **chaque** match, on prend les côtes d'avant match. Pour chaque site on récupère la côte minimale et on déduit son résultat probable, puis on prend le résultat probable majoritaire parmi ceux de tous les sites et on le compare avec le résultat de fin de match.

Dans les 2 cas, pour tous les matchs, on compte le nombre d'erreurs pour avoir à la fin un taux d'erreur.

$$taux_{erreur} = \frac{nombre_{erreurs}}{nombre_{total_matches}}$$

Résultats :



On conclut que les côtes d'avant match ne sont pas fiables. Donc on ne peut pas construire un modèle prédictif basé dessus.