

Projet 31026

Introctution

Avec la progression technologique et la révolution numérique le foot a évolué. De nos jours les équipes s'intéressent de plus en plus à la science des données, donnent de l'importance à la statistique. Dans ce projet J'ai essayé d'analyser les chiffres des équipes des joueurs afin de chercher des liaisons entre les caractéristiques des joueurs et de produire modèles statistiques qui peuvent prédire les résultats des matchs.

Objectifs

Découvrir les caractéristiques qui définissent l'équipe/joueur idéal.

Prédire les résultats des matchs entre les équipes.

Data fournis

Les données fournis pour ce projet ont été regroupés par league . Pour chaque league, il existe un ensemble de match avec leur résultat et les équipes opposés. Pour chaque match, on peut récupérer les valeurs des attributs pour les équipes et leurs joueurs correspondants à partir des tables "player", "playerAttributes", "team" et "teamAttributes". On a remarqué que les données ne sont pas complètes donc on a pensé à nettoyer les données : On traite les informations manquantes avec 3 méthodes différentes selon le type de l'information et la logique : la valeur moyenne , substitution et lorsque cela peut affecter les autres données on efface la ligne

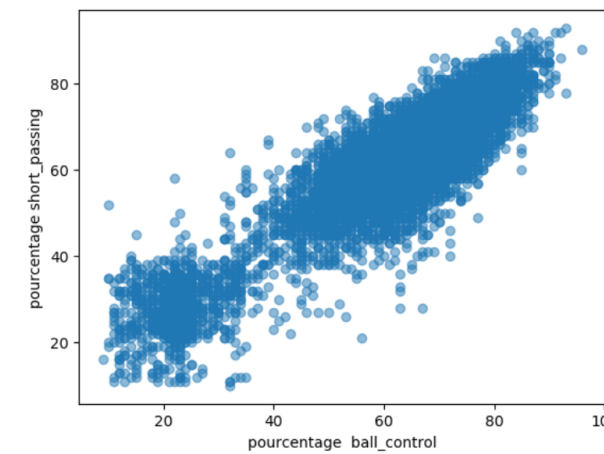
Conclusion et Amélioration

On a essayé de faire des hypothèses selon les clichés de foot et la logique après on a essayé d'analyser ces hypothèses selon les données qu'on a . Donc au début on a commencé a regardé les données qu'on a , à décider comment les traiter et après extraire ce que peut être utiles selon les résultats de l'analyse pour la prédiction mais on a utilisé aussi des attributs de "teamAttributes" qu'on comprend pas trop mais on a pensé que cela pour aider pour l'apprentissage. Cela n'empêche que les informations qu'on ne mènent pas nécessairement à une bonne production car les résultats peuvent être affecté par des autres facteurs dont on n'a pas des informations sur.

Essayer toutes combinaisons possibles pour les vecteurs et les tester et les comparer puisque le facteur de corrélation ne reflète pas essentiellement que cet attribut affecte la résultat et peut-être contrairement. On peut aussi exploiter les données des paris et extraire des informations utiles pour l'apprentissage. On peut aussi extraire selon les dates des matchs des informations sur les météo le jour du match et regardez si ça affecte les résultats .

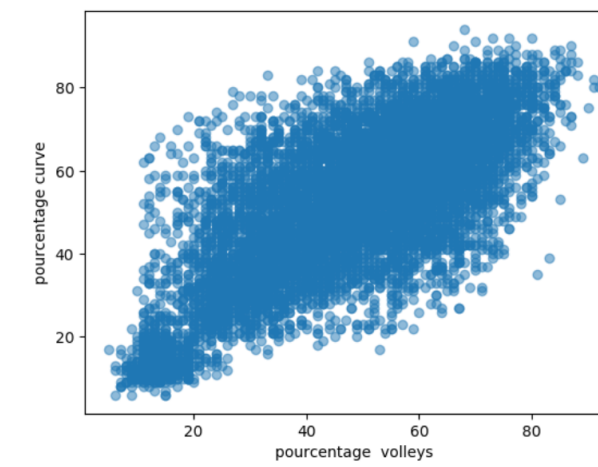
Partie 1: les clichés du foot face aux stats

Les passes courtes influencent-elles le contrôle sur le ballon ?

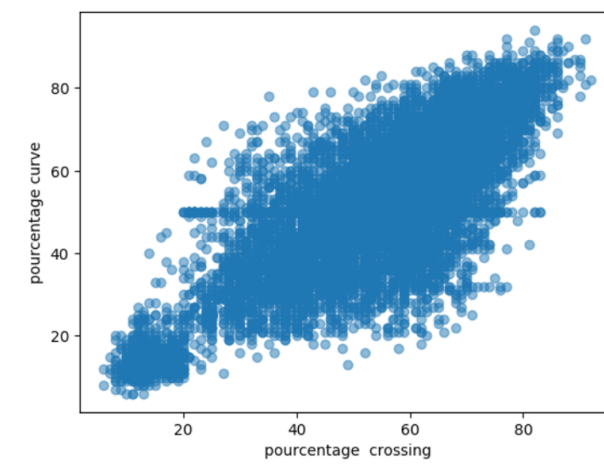


le coefficient de correlation est = 0.9

Plus un joueur sait tirer en courbe, plus il réussit sont crossing et volleys

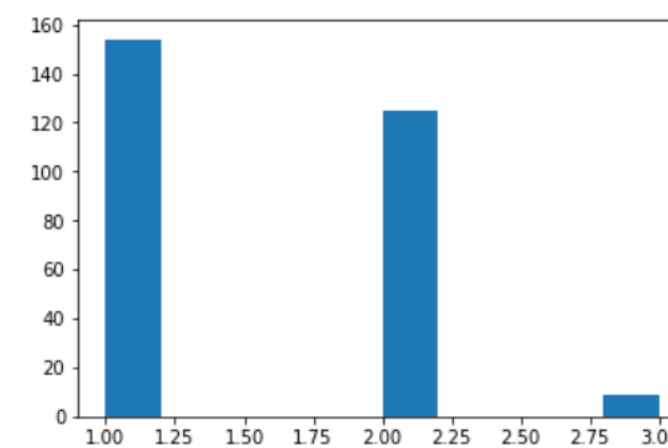


le coefficient de correlation est = 0.75

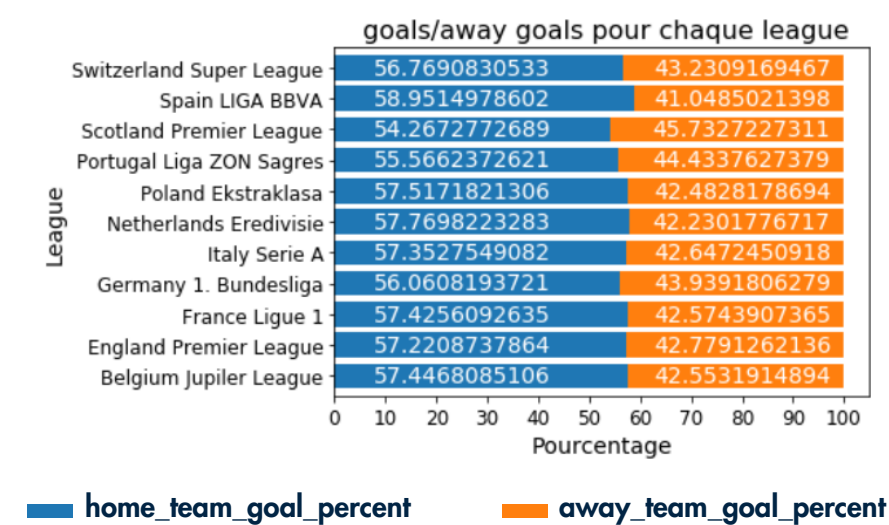


le coefficient de correlation est = 0.8

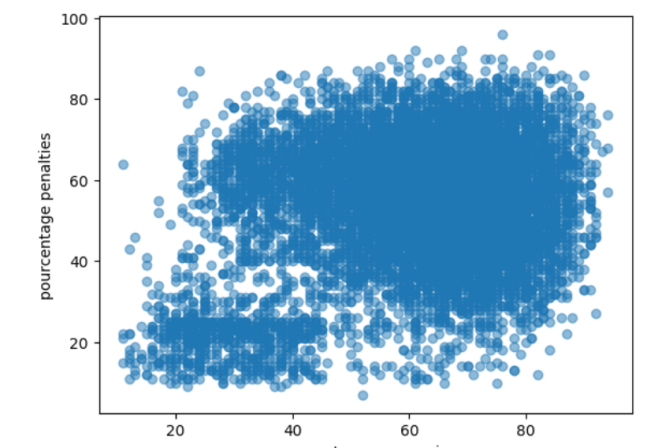
En équipe, les joueurs suivent-ils toujours la même stratégie ?



Est-ce qu'on gagne plus quand on joue chez nous ?



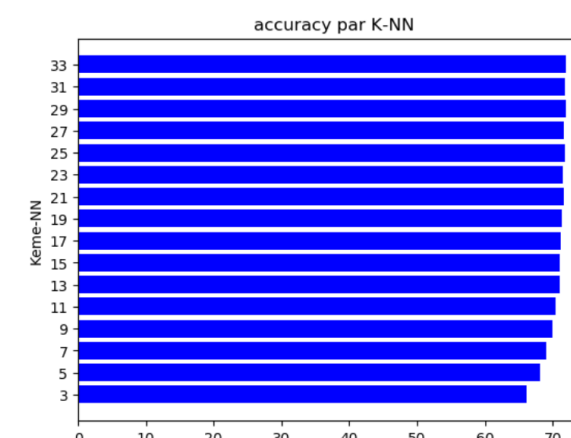
Pas de panique! Etre agressif ne vous donne pas forcément plus de penalties:



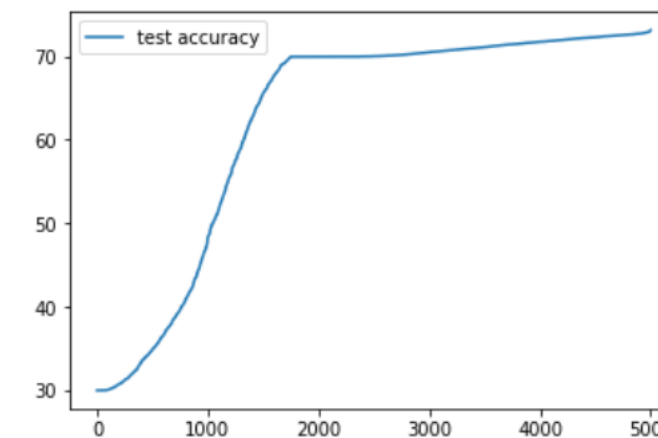
le coefficient de correlation est = 0.2

Partie 2: Prédiction des résultats des match

Learning et Résultats KNN



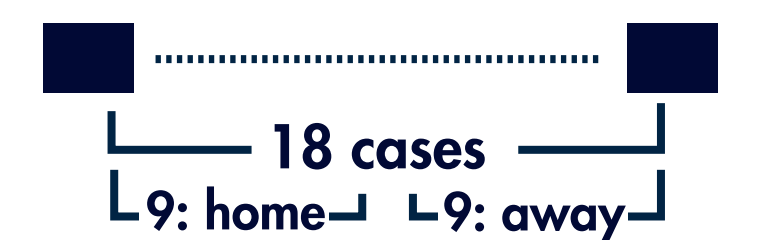
Perceptron



Arbre de Décision

accuracy atteinte = %62

Choix des attributs
Représentation d'un match



Equipe

["agility", "ball_control", "stamina", "potential", "overall_rating"]
["buildUpPlayPassing", "chanceCreationPassing", "chanceCreationShooting", "defencePressure"]