# Prédiction d'un match de football de Premier League 2023 dans le cadre d'une régression multinomiale

## Table des matières

\* [Description](#Description)

\* [Utilisation](#Utilisation)

\* [Auteurs](#Auteurs)

## Description

Le projet a pour objectif de prédire le résultat d'un match de football ( en Premier League) en utilisant une régression multinomiale. Concrètement, il s'agit de prédire au vu des informations les plus récentes en 2023, la probabilité de gagner, de perdre ou de faire un match nul d'une équipe lorsqu'elle rencontre un autre adversaire en Premier League. Une telle modélisation nécessite des informations sur les résultats des matchs des équipes et celles sur les variables explicatives à inclure dans le modèle. De plus, dans la mesure où nous voulons prédire les résultats des matchs du championnat en cours (2023), il est nécessaire d'entrainer notre modèle à partir des résultats des championnats de Premier League des années précédentes. Pour ce faire, notre approche méthodologique est la suivante:

- En premier, à partir du web scrapping, nous avons utilisé le site officiel de la Premier League et celui de BeSoccer pour obtenir en codes html bruts, les informations sur les résultats et les différentes statistiques des équipes depuis le championnat 2020 au championnat 2023.

- Ensuite, dans un premier temps, nous avons travaillé sur les statistiques des équipes (les variables explicatives de notre modèle). Nous avons créé des fonctions pour obtenir ces statistiques et les mettre dans un DataFrame. Ceci a été fait pour le championnat 2020, 2021, 2022 et 2023. Dans un second temps, nous avons crée des fonctions pour obtenir directement sous format DataFrame, les informations concernant les résultats des rencontres entre les équipes pour le championnat 2020, 2021 et 2022.

- En troisième lieu, étant donné que nous avons obtenu les informations sur les équipes à travers des sites différents, nous avons eu à harmoniser les noms des équipes. Ensuite, nous avons procéder à des merges pour avoir sur une même ligne au sein d'un DataFrame, les informations de l'équipe dont le match est prédit et celles de l'ensemble des adversaires possibles. Toutes les informations de 2020 à 2022 sont alors regroupées dans un seul DataFrame. Cela nous a permis d'avoir sur chaque ligne, les statistiques de l'équipe et d'un adversaire ainsi que l'issue du match entre ces deux équipes.

- En quatrième lieu, nous définissons et estimons un modèle dans lequel, la variable dépendante est représentée par les issues de chaque match depuis le championnat 2020 au championnat 2022. Les variables indépendantes sont alors les statistiques sur chaque équipe à savoir le nombre de match joué, la position au sein du championnat, le nombre de point à l'issue du championnat, le nombre de victoire, le nombre de défaite, le nombre de match nul, le nombre de but marqué, le nombre de but encaissé et une variable indicatrice indiquant si l'équipe dont le match est prédit avait joué à domicile ou à l'extérieur.

- Enfin, pour l'estimation du modèle définit, nous avons spécifié 11% de notre échantillon pour les tests et 89% pour les entrainements. De plus, nous avons spécifié une graine de 100 pour la séparation aléatoire de l'échantillon afin d'assurer la reproductibilité de nos résultats. Ces derniers indiquent un pouvoir prédictif de notre modèle tournant autour de 75%. Par la suite, nous régressons les valeurs prédites du modèle sur les statistiques de 2023 de l'équipe et de l'adversaire afin de prédire la probabilité des différentes issues possibles du match. Cette dernière est ensuite exposée à travers une figure en secteur indiquant la probabilité de l'équipe de gagner, de faire un math nul, ou de perdre le match.

## Utilisation

L'utilisateur doit juste télécharger le fichier 'Project\_Match\_Prediction.ipynb' et l'ouvrir soit à partir de Jupyter notebook soit directement à partir de Python.

Ensuite, il exécute l'ensemble du code via l'option run. Le programme lui demandera de renseigner le nom de l'équipe dont le résultat de match doit être prédit, le nom de l'adversaire, et de mettre un code 0 si l'équipe jouera le match à l'extérieur ou un code 1 si l'équipe jouera le match à domicile. L'utilisateur doit appuyer la touche "Entrée" à chaque fois qu'il renseigne ces trois informations. Ensuite, le programme se déroule automatiquement et fournit en fin de code, une figure en secteur indiquant la probabilité pour les différentes issues du match.

## Auteurs

\* DJAFON Kokouvi Joseph

\* DOBOROSSY Samuel

\* RAZAFIMANANTENA Iriantsoa