



DataScientest • com

Paris sportifs

Cursus concerné : Data Scientist

Niveau de difficulté : 05/10

Description du projet :

L'objectif de ce projet est d'essayer de battre les algorithmes des bookmakers sur l'estimation de la probabilité d'une équipe gagnant un match.

C'est un problème parfait pour traiter l'ensemble des étapes d'un projet de Data Science. Une première étape est d'étudier et appliquer des méthodes pour nettoyer le jeu de données. Une fois le jeu de données propre, une deuxième étape est d'extraire à partir de l'historique des matchs les caractéristiques qui semblent être pertinentes pour estimer la performance d'un(e) joueur/équipe (classement, winning streak ...). Ensuite, à partir de ces features, le but est d'estimer la probabilité qu'une équipe/joueur à battre une équipe/joueur B.

Une fois l'entraînement du modèle terminé, vous allez comparer votre modèle avec le modèle des bookmakers.

Ressources à consulter :

- **Données :**
 - Pour le tennis, match ATP:
<https://www.kaggle.com/edouardthomas/atp-matches-dataset>
- **Bibliographie :**
 - <https://towardsdatascience.com/making-big-bucks-with-a-data-driven-sports-betting-strategy-6c21a6869171>
 - https://www.researchgate.net/publication/331218530_Exploiting_sports-betting_market_using_machine_learning

Conditions de validation du projet :

- un **rapport** d'exploration, de data visualisation et de pre-processing des données ;
- un **rapport** de modélisation ;
- un **rapport** final et le **code** associé.

[DataScientest.com](https://www.data-scientest.com)

Agrément organisme de formation 11755665975

09 80 80 79 49

2 place de Barcelone, 75016 Paris