



Boulogne-Billancourt, 24 avril 2019

TEST TECHNIQUE DATA SCIENTIST JUNIOR/CONFIRME

Auteur	MP DATA Faïlor ELFASSI Directeur Technique Faïlor.elfassi@mpdata.fr
---------------	--

MP DATA - 39 rue de la Saussière - 92100 Billancourt

SAS at capital of 25 000 euros - SIRET 812 051 704 00011 RCS NANTERRE

1 – Contexte et données

Vous trouverez ci-joint un dataset *nba_logreg.csv* qui contient des statistiques sportives sur les joueurs débutants de la NBA. L'objectif est de fournir un classifieur permettant de prédire qu'un joueur vaut le coup d'investir sur lui car il va durer plus de 5 ans en NBA en s'appuyant sur ses statistiques sportives. Ce modèle vise à conseiller des investisseurs cherchant à capitaliser sur de futurs talents de la NBA.

Le descriptif des paramètres du jeu de données est le suivant :

	Description
Name	Name
GP	Games Played
MIN	Minutes Played
PTS	Points Per Game
FGM	Field Goals Made
FGA	Field Goal Attempts
FG%	Field Goal Percent
3P Made	3 Point Made
3PA	3 Point Attempts
3P%	3 Point Attempts
FTM	Free Throw Made
FTA	Free Throw Attempts
FT%	Free Throw Percent
OREB	Offensive Rebounds
DREB	Defensive Rebounds
REB	Rebounds
AST	Assists
STL	Steals
BLK	Blocks
TOV	Turnovers
TARGET_5Yrs	Outcome: 1 if career length >= 5 yrs, 0 if < 5...

Table 1 – description des features

2 – Questions

Question 1 : Training

Vous trouverez ci-joint un fichier template (*test.py*) à compléter qui lit et décode les données, et propose une fonction de scoring.

Votre but sera de proposer, d'entraîner et de valider un classifieur répondant le mieux possible à l'objectif des investisseurs. Le fichier fournit également une fonction de scoring en recall, que vous êtes libre de modifier tant que vous justifiez pourquoi. Plus que le résultat, c'est la démarche analytique qui nous intéresse. Il n'y a pas de restrictions sur le format de remise, si un jupyter notebook vous semble plus pertinent.

Question 2 : Intégration

Une fois votre classifieur entraîné, vous devrez l'intégrer sous forme de requête unitaire dans un webservice. Vous êtes libre de choisir la librairie qui vous convient (flask, django ou autre). Ce web service au format d'API REST devra prendre en entrée tous les paramètres pertinents que vous aurez identifiés comme s'il était mis à disposition d'un utilisateur voulant faire une requête sur un seul joueur au modèle que vous aurez entraîné.