

Test Technique Quantmetry

Cet exercice vise à évaluer vos compétences en data science et se compose de 3 étapes : Statistiques Descriptives, Machine Learning, Restitution des résultats. Vous serez évalués sur ces trois axes ainsi que sur un quatrième axe correspondant à la qualité et la propreté de votre code. Ne négligez pas les parties de restitution des résultats et de qualité du code ☺

Votre tâche consiste à nous fournir 3 fichiers :

- Les réponses aux questions ci-dessous dans un fichier texte (Word, LibreOffice ou PDF) nommé *NOM_Prénom_Réponses*
Le fichier ne devra pas faire plus de 6 pages.
- Le code produit durant l'exercice dans un fichier nommé *NOM_Prénom_Code.R* ou *NOM_Prénom_Code.py*
La qualité et la propreté de votre code seront prises en compte
- Les slides demandées en partie III dans un fichier pdf ou Powerpoint, nommé *NOM_Prénom_Présentation.ppt*
Le présentation ne doit pas comporter plus de 4 slides.

Le jeu de données fourni est un ensemble de relevés concernant un système de vélo partagé (type Vélib). On cherche à prédire le nombre de vélos loués par heure dans la ville (variable **count**).

Voici une brève description des colonnes du jeu de données :

datetime - date et heure du relevé

season - 1 = printemps , 2 = été, 3 = automne, 4 = hiver

holiday – indique si le jour est un jour de vacances scolaires

workingday - indique si le jour est travaillé (ni week-end ni vacances)

weather - 1: Dégagé à nuageux, 2 : Brouillard, 3 : Légère pluie ou neige, 4 : Fortes averses ou neiges

temp – température en degrés Celsius

atemp – température ressentie en degrés Celsius

humidity – taux d'humidité

windspeed – vitesse du vent

casual - nombre de locations d'utilisateurs non abonnés

registered – nombre de locations d'utilisateurs abonnés

count – nombre total de locations de vélos

Le code fourni doit être écrit de préférence en R ou en Python. Si vous souhaitez utiliser un autre langage, merci de bien vouloir nous faire confirmer votre choix au préalable. Vous êtes libres d'utiliser les bibliothèques de Machine Learning que vous estimerez nécessaires.

Votre script devra pouvoir être rejoué facilement. N'hésitez pas à fournir un fichier README ou à spécifier dans le corps du script les actions nécessaires à la relance de votre code.

Partie I – Statistiques descriptives

1. Sans commencer la partie modélisation, quels sont les facteurs qui semblent influencer la demande en vélos ? Justifiez vos choix. Présentez quelques graphiques (6 maximums) pertinents pour illustrer votre réponse et interprétez-les.

Partie II – Machine Learning

1. Concevez un modèle permettant de prédire la variable **count** et expliquez votre choix d'algorithme. Si votre modèle comporte des spécificités de paramétrage, justifiez également vos choix de paramètres.
2. Décrivez le critère de performance utilisé lors de la conception de votre algorithme, et justifiez-en le choix.
3. Proposez deux à trois pistes d'amélioration de votre modèle.

Partie III - Restitution

Préparez une présentation décrivant votre démarche générale de prise en main des données et votre modèle, ses performances et ses pistes d'amélioration.

La présentation sera livrée sous forme d'un fichier PowerPoint, LibreOffice ou PDF de 2 à 4 slides :

- 1 à 2 slides expliquant votre démarche générale de prise en main des données.
- 1 à 2 slides de restitution du modèle et de ses performances et ses pistes d'amélioration. Rappelez vos choix et justifiez les.