

DATA SCIENCE AVEC « R » « POUR LES NULS »



1 jour

Bulletin de participation
sur www.caritat.fr

Thèmes transversaux



À qui s'adresse cette formation ?

À toute personne susceptible d'intervenir dans l'élaboration et la mise en œuvre d'une stratégie Big Data, et ce dans les domaines : produits, marketing, distribution et tarification. Ainsi qu'à toute personne désireuse de se familiariser avec ce sujet en devenant.

Pour obtenir quoi ?

Une compréhension globale de la problématique Big Data et des outils disponibles à ce jour et nécessaires à son traitement.

Comment ?

Des exemples pratiques animeront chaque étape du processus de traitement du Big Data, et compléteront la définition et les perspectives de ce nouveau métier.

Quels sont les prérequis ?

Aucun en particulier. L'intérêt et l'utilisation des outils numériques sont cependant fortement recommandés.

Chaque participant se munira d'un ordinateur portable pour les travaux pratiques.

Qui anime cette formation ?

Kezhan SHI est membre certifié de l'IA. Il est titulaire d'un master en actuariat de l'Université Paris Dauphine. Il a également fait l'École Centrale Paris. Il a travaillé chez Prim'Act, Direct Assurance et Aviva, avant de rejoindre Allianz en 2017, au titre de Data Scientist.



La formation en pratique...

Quand et où ?

4 octobre 2018

9 h 00 - 12 h 30 et 14 h 00 - 17 h 30
Caritat, Paris 8°

Combien ça coûte ?

1 220 € HT + TVA 20 %, soit 1 464 € TTC.
Les frais de participation couvrent la journée de formation, la documentation complète, le déjeuner et les pauses café.

Où vous loger ?

Si vous venez de province ou de l'étranger, pour toute réservation d'hôtel, contactez notre partenaire :

Elysées West Hôtel - 01 85 34 72 00
Précisez que vous venez de la part de Caritat.
<http://elysees.hotusa.com/caritat/>

Qu'allez-vous apprendre ?

Introduction

- Processus général du traitement de données
- Analyse des données textes : principe et intérêts
- Présentation des outils pour la collecte et traitements des données
- Prise en main de « R »
- Packages à connaître pour la data science
- Exemples de text mining
- Lien avec speech analytics et intérêt pratique

Collecte des données web avec « R »

- Bases de la structure des pages web
- Scraping des données html avec « R »
- Cas d'application : analyse des avis et notations des produits
- Visualisation de l'évolution des notations
- Import et analyse des données de twitter
- Construction d'un outil de veille pour les produits ou les services
- AB testing

Nettoyage de données html et textes

- Traitement automatique du langage naturel (Natural language processing)
- Manipulation des corpus avec « R »
- Exploitation des documents term matrix (DTM)
- Association des mots
- Visualisation des nuages de mots
- Extraction d'informations et analyse des sentiments
- Clustering de textes

Détection de spams et machine learning avec « R »

- Analyse prédictive des mails indésirables
- Pratique de différents algorithmes avec « R »
- Application d'un arbre de classification
- Application de SVM (Support Vector Machine)
- Application de la régression logistique
- Agrégation de modèle : bagging et forêt aléatoire
- Comparaison des différents algorithmes de classification

Performance des algorithmes

- Matrice de confusion et taux d'erreur
- Courbe ROC et AUC
- Coefficient de Gini

Conclusions

- Intérêt de l'analyse des données textes
- Intérêt du processus d'analyse prédictive

Qu'en disent les stagiaires ?

« Le programme est très intéressant et le formateur est passionné. »

GL, Responsable de l'actuariat – ADDING

« Le programme est en cohérence avec mes attentes. L'exposé a été très clair. »

MK, Actuaire – AXA

« Formateur à l'écoute et pédagogie. »

NP, Chargé d'études statistiques – MACIF

