



### Durée de la formation ?

1 jour.

### 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.

### 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 ?

**13 avril 2019**

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

#### Combien ça coûte ?

1 250 € HT + TVA 20 %, soit 1 500 € 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 à la programmation avec « R »

- Installation de R et de RStudio
- Présentation de RStudio

#### Collecte des données web avec « R »

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

#### Nettoyage de données html et textes

- Text mining
- Manipulation des corpus avec R
- Exploitation des document term matrix (DTM)
- Association des mots
- Visualisation des nuages de mots

#### 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 modèles
- Comparaison des différents algorithmes

#### 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

01 44 51 04 00  
[info@caritat.fr](mailto:info@caritat.fr)

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

### 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 pédagogue. »

*NP, Chargé d'études statistiques – MACIF*