

## À qui s'adresse cette formation ?

À toute personne souhaitant analyser les textes et tirer des conclusions intéressantes.

## Pour obtenir quoi ?

Des compétences de text mining et des cas d'applications.

## Comment ?

Les différentes techniques de collecte, de nettoyage et d'analyse de données seront illustrées avec des exemples pratiques, codés sous « R ».

## 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 ?

**11 et 12 juin 2018**

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

### Combien ça coûte ?

2 000 € HT + TVA 20%, soit 2 400 € TTC.  
Les frais de participation couvrent les deux journées de formation, la documentation complète, les deux déjeuners 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 ?

*Apprendre à récupérer, nettoyer et intégrer les données textes afin de mieux connaître les besoins clients.*

### Source de données

- Données internes
- Reconnaissance optique de caractères
- Données des sites d'avis
- Manipulation des documents html
- Technique de webscraping

### Transformation des données textes

- Manipulation des chaînes de caractères
- Manipulation des expressions régulières
- Construction de corpus
- Matrice Documents-termes
- TD IDF

### Nettoyage des données textes

- Distance entre les chaînes de caractères et correction de fautes d'orthographe
- Principe de géocodage
- Détection d'anomalies

### Machine learning

- Clusterint des chaînes de caractères
- Classification de spams
- Caractérisation des avis par une note et sa prédiction
- Segmentation des avis

### Visualisation

- Tableau de bord
- Nuage de mots
- Fréquence des mots ou bigrammes

### Cas d'application

- Webscraping
- Analyse des avis clients
- Analyse des CV et des annonces de postes
- Correction des fautes d'orthographe
- Géocodage

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

## Qu'en disent les stagiaires ?

Cette formation est une nouveauté du catalogue Caritat.