

Analyse textuelle avec Python

Description

Pour les utilisateurs de Python en data science, le traitement automatique du langage (TAL / NLP) est un domaine de premier plan.

Cette formation analyse textuelle avec python se décompose en deux parties : une présentation des principes du NLP et des outils de traitement des chaînes de caractères avec python et une étude plus approfondie des packages spécialisés avec notamment nltk et les packages tels que Spacy.

Formation disponible en intra ou en inter-entreprises

Inscription : <https://www.stat4decision.com/fr/formations/formation-analyse-textuelle-python/>

Durée

2.00 jours (14.00 heures)

Profils des stagiaires et prérequis

Profils :

- Data scientist désirant s'initier au traitement des données textuelles

Prérequis :

- Connaissance de Python
- Ou ayant suivi la formation Python pour la Data science

Objectifs pédagogiques

- Maîtrise des outils pour traiter des données textuelles avec python
- Identifier la problématique et les outils à utiliser pour traiter des problèmes de NLP
- Mener à bien un projet de traitement NLP avec Python

Programme détaillé

- Les principes de l'analyse textuelle
 - Qu'est-ce que le traitement du langage naturel (NLP) ?
 - A quelles questions répond le NLP ?
- Les outils de python pour le NLP
 - Le travail sur les chaînes de caractères avec python (string, expressions régulières, transformation vectorisées dans des DataFrame)
 - Les API pour récupérer des données textuelles (Réseaux sociaux, Web scrapping, Open Data)
 - La préparation des données textuelles en fonction de la problématique (Données d'enquêtes, Automatisation de processus)
 - Les descripteurs en NLP : BOW, TF-IDF, Word Embedding
- Le NLP avec nltk et spaCy
 - Les packages nltk et spaCy : leurs principes et leurs différences

Stat4decision

37-39 avenue Ledru Rollin Paris 75012

Tel. 01.72.25.40.82 | E-mail : info@stat4decision.com | www.stat4decision.com

Numéro SIRET: 81048985600015 | Numéro de déclaration d'activité: 11755352275 (auprès du préfet de région de: 75)

Organisme de formation certifié Qualiopi pour ses actions de formation

- Combiner analyse textuelle et algorithmes de machine learning
 - L'utilisation de nltk et du machine learning avec scikit-learn pour l'analyse des sentiments
 - Identification de thématiques à partir de textes (topic modeling) en utilisant le package gensim
- L'utilisation du deep learning pour le NLP
 - Les possibilités pour le traitement du langage avec le deep learning
 - Les réseaux de neurones récurrents pour le NLP

Organisation de la formation

Moyens pédagogiques et techniques

- Documents supports de formation projetés.
- Exposés théoriques
- Etude de cas concrets
- Mise à disposition en ligne de documents supports à la suite de la formation
- Outils utilisés : Anaconda, python 3, SpaCy et nltk

Dispositif de suivi de l'exécution et d'évaluation des objectifs de la formation

- Mises en situation.
- Cas pratiques validés par le formateur
- Évaluation des connaissances (quizz / tests)

Stat4decision

37-39 avenue Ledru Rollin Paris 75012

Tel. 01.72.25.40.82 | E-mail : info@stat4decision.com | www.stat4decision.com

Numéro SIRET: 81048985600015 | Numéro de déclaration d'activité: 11755352275 (auprès du préfet de région de: 75)

Organisme de formation certifié Qualiopi pour ses actions de formation