

## FORMATION ANALYSE TEXTUELLE AVEC PYTHON

REF. : PYTHONTEXT201

### DESCRIPTION

Pour les utilisateurs de Python en data science, le traitement du langage (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 de text mining.

### OBJECTIFS

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 DETAILLE

- 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
- Le NLP avec nltk et spaCy
  - Les packages nltk et spaCy : leurs principes et leurs différences
  - Applications avancées de nltk et spaCy pour du traitement du langage

### DUREE

2 jours

### PUBLIC ET PREREQUIS

Public connaissant le langage python ou ayant suivi la formation python pour la data science.

### MOYENS PEDAGOGIQUES

Alternance d'exposés et d'applications pratiques avec des exercices sur des données. Nous utiliserons Anaconda, Python 3 et les Jupyter Notebooks.

### PLUS D'INFORMATIONS

Plus d'infos :

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

Formation disponible en intra ou en inter-entreprises

- Vers le text mining : combiner nltk et des algorithmes de machine learning
  - Identification de thématiques à partir de textes (topic modeling)
  - Classifieurs Bayésien
  - L'utilisation de nltk et du machine learning avec scikit-learn pour l'analyse des sentiments
- L'utilisation du deep learning pour le NLP
  - Utilisation de keras pour construire un premier réseau
  - Les possibilités pour le traitement du langage avec le deep learning