

IA - Traitement du langage naturel NLP

Référence : PYDS021

Durée : 2 jours

Certification : **Aucune**

CONNAISSANCES PREALABLES

- Connaissances de base de Python et du deep learning.

PROFIL DES STAGIAIRES

- Toutes personnes intéressées par le NLP: Ingénieurs, Analystes, Data Scientists, Data Analysts, Data Scientists, Data Steward, Développeurs....

OBJECTIFS

- Comprendre les principes du NLP et savoir les mettre en œuvre avec Python.

CERTIFICATION PREPAREE

Aucune

METHODES PEDAGOGIQUES

- Mise à disposition d'un poste de travail par stagiaire
- Remise d'une documentation pédagogique papier ou numérique pendant le stage
- La formation est constituée d'apports théoriques, d'exercices pratiques, de réflexions et de retours d'expérience
- Le suivi de cette formation donne lieu à la signature d'une feuille d'émargement

FORMATEUR

Consultant-Formateur expert Bigdata

METHODE D'EVALUATION DES ACQUIS

- Auto-évaluation des acquis par le stagiaire via un questionnaire
- Attestation des compétences acquises envoyée au stagiaire
- Attestation de fin de stage adressée avec la facture

CONTENU DU COURS

NLP Introduction

- Comprendre le langage humain, et savoir générer des réponses
- Différentes étapes : reconnaissance de caractères, ou de la voix, conversion des données en texte, décomposition en éléments de phrase, nettoyage des données, traitement de l'ambiguïté d'un mot, reconnaissance d'une entité nommée (NEM), traitement des multiples références pour une entité, extraction des informations subjectives, etc.
- Les outils de NLP et historique : outils statistiques, de machine learning, de deep learning, Watson NLU, Python et le NLTK
- Applications actuelles : solutions de détection de spam, traduction automatique, assistants virtuels, chatbots, analyses d'opinions, de sentiments, etc.

Python et le NLTK

- Plate-formes supportées, versions de Python

- Présentation des textes et modèles fournis avec le NLTK
- **Exemple d'Atelier : installation du package NLTK et des datasets**

Traitements de textes

- Etude des différentes fonctions fournies par le NLTK
- Découpage d'un texte en mots ou en phrases, avec `nltk.tokenize()`, nettoyage de textes avec le filtrage de mots, stemming avec `nltk.stem`, alertes sur les risques d'un mauvais usage, étiquetage des différentes parties d'un texte avec `nltk.pos-tag()`, lemmatisation, pour identifier les formes canoniques des mots, identification de phrases avec le chunking
- **Exemple d'Atelier : réalisation d'exemples sur des datasets simples**

Analyses de textes

- Description de `nltk.ne_chunk()` pour la reconnaissance d'identités nommées

- Présentation des fonctions `concordance()`, `dispersion_plot()`, `FreqDist`
- **Exemple d'Atelier : Import de corpus de textes, analyse, mise en évidence de l'utilisation de termes caractéristiques**

Etude de cas

- Analyse de sentiments avec `nlk.sentiment`
- Présentation des fonctions disponibles
- **Exemple d'Atelier : mise en oeuvre sur un corpus. Utilisation de `polarity.scores()`**

Intégration de scikit-learn

- Import des algorithmes de classification de scikit-learn
- **Exemple d'Atelier : exemple d'utilisation des algorithmes de scikit-learn depuis nltk**

Notre **référent handicap** se tient à votre disposition au 01.71.19.70.30 ou par mail à referent.handicap@edugroupe.com pour recueillir vos éventuels besoins d'aménagements, afin de vous offrir la meilleure expérience possible