Traitement du langage : du text mining aux transformers

Extraire de l'information de textes (avis, courriels, comptes rendus, documents) : prétraitement, classification, analyse de sentiment, extraction d'entités, avec les méthodes classiques et les modèles transformers.

En pratique

DURÉE
14 h (2 jours)

MODALITÉS
Présentiel · Distanciel · Intra-entreprise · Inter-entreprises

LIEU
Orléans (45), dans vos locaux ou en classe virtuelle

EFFECTIF
3 à 8 participants

TARIF
Inter : 1 290 € HT par participant ; intra : sur devis

DÉLAI D'ACCÈS
30 jours ouvrés entre la demande et le début de la formation, sous réserve d'un nombre suffisant de participants

VALIDATION
Attestation de fin de formation

NIVEAU
Développeur

Objectifs pédagogiques

À l'issue de la formation, le stagiaire sera capable de :

- Prétraiter et représenter des données textuelles
- Classer des textes et analyser leur tonalité
- Extraire des thèmes et des entités nommées
- Utiliser des modèles transformers pré-entraînés avec Hugging Face
- Choisir entre méthodes classiques, transformers et grands modèles de langage

Public visé

- Data scientists, data analysts et développeurs
- Équipes qui exploitent de grands volumes de textes

Prérequis

- Pratique de Python et de pandas
- Notions de machine learning

Programme

Jour 1 — Le text mining

Préparer les textes

- Nettoyage, tokenisation, lemmatisation, mots vides (spaCy)
- Expressions régulières

Représenter les textes

- Sac de mots et TF-IDF
- Embeddings de mots et de phrases

Analyser

- Classification de textes, analyse de sentiment
- Modélisation de thèmes (topic modeling)

Jour 2 — Les transformers

Modèles pré-entraînés

- Architecture transformer et principe de l'attention
- Hugging Face : pipelines, modèles francophones (CamemBERT…)

Tâches avancées

- Extraction d'entités nommées, résumé, similarité sémantique
- Ajustement fin (fine-tuning) d'un classifieur

Et les grands modèles de langage ?

- Classification et extraction par requête (zero-shot, few-shot)
- Coûts, performances, confidentialité : bien choisir
- Projet final

Méthodes pédagogiques

- Possibilité de travailler sur les données des participants (en intra, après accord)
- Supports de cours, codes source et corrigés remis aux participants
- Formation dispensée par un consultant data et IA qui conçoit et met en production ce qu'il enseigne

Modalités d'évaluation

- Questionnaire de positionnement en début de formation
- Exercices pratiques corrigés tout au long de la formation
- Projet final : chaîne d'analyse de textes sur un corpus réel
- Questionnaire de satisfaction à chaud

Accessibilité

Formation accessible aux personnes en situation de handicap. Contactez-nous en amont (contact@dcid-consulting.fr) afin d'étudier ensemble les adaptations nécessaires.

Financement

DCID Consulting n'est pas certifié Qualiopi à ce jour : cette formation est financée directement par l'employeur et n'est éligible ni au CPF ni à une prise en charge par un OPCO.

Contact et inscription

contact@dcid-consulting.fr · 02 38 25 33 90 ·dcid-consulting.fr/formations/traitement-langage-nlp/