

Sessions garanties disponibles

Formation NLP (Natural Language Processing)

Réf. SF33903

3 jours de formation pour apprendre les techniques de NLP !

3 j Durée	2 400 € HT Inter / participant	3 950 € HT Intra (≤12 pers.)	1 Sessions à venir
--------------	-----------------------------------	---------------------------------	-----------------------

Objectifs pédagogiques

- ✓ Résumer les principales techniques de prétraitement NLP telles que la tokenisation, le stemming, la lemmatisation et la suppression des mots vides, et les appliquer à l'aide de bibliothèques telles que spaCy et NLTK.
- ✓ Construire des représentations vectorielles de texte (par exemple, sac de mots, tf-idf, word embeddings) et comparer leurs performances à l'aide de modèles simples d'apprentissage automatique.
- ✓ Mettre en œuvre des architectures de réseaux neuronaux, y compris des RNN et des modèles basés sur des transformateurs.
- ✓ Évaluer les performances des modèles de TAL à l'aide de métriques et de techniques de rééchantillonnage sur des ensembles de données réels.

Public visé & prérequis

- Public visé** Data analysts, Data scientists, Ingénieurs Big Data
- Prérequis** Connaissance préalable de la programmation Python, des bases en machine learning et en statistiques générales.

Programme détaillé

Introduction à la formation Natural Language Processing et aux techniques de prétraitement

- Aperçu des principes fondamentaux du NLP
- Encodage des données textuelles
- Expressions régulières
- Techniques de nettoyage de texte :
 - Suppression des mots vides
 - Troncature
 - Lemmatisation
 - Suppression de la ponctuation
- Outils : spaCy, NLTK

Activités pratiques :

- Nettoyage de texte sur un échantillon de données à l'aide de NLTK et spaCy
- Comparaison des résultats du stemming et de la lemmatisation

Représentation du texte et ingénierie des caractéristiques

- Sac de mots
- TF-IDF
- N-grammes
- Enchâssement de mots (Word2Vec, GloVe, FastText)
- Créer des représentations vectorielles
- Évaluer les performances du modèle

Activités pratiques :

- Appliquer des techniques de vectorisation sur un corpus de textes
- Comparer les résultats de différentes méthodes de vectorisation

- Implémenter des modèles de classification de texte de base
- Évaluer les performances du modèle avec différentes représentations

Applications de l'apprentissage automatique dans le domaine du NLP

- Construction de modèles prédictifs
- Techniques classiques d'apprentissage automatique
- Méthodes de classification
- Méthodes de clustering
- Analyse du sentiment
- Modélisation des sujets
- Outils : scikit-learn

Activités pratiques :

- Analyse de sentiments sur des données de médias sociaux
- Modélisation thématique d'articles de presse à partir de données textuelles vectorisées

Réseaux neuronaux et apprentissage profond pour le NLP

- Aperçu des réseaux neuronaux
- Réseaux neuronaux récurrents (RNN)
- Mécanisme d'attention
- Transformateurs
- Grands modèles de langage (LLM) : GPT, BERT
- Applications :
- Analyse de sentiments
- Résumés de textes
- Génération de contenu
- Outils : TensorFlow, PyTorch

Activités pratiques :

- Entraînement d'un RNN pour le résumé de texte
- Entraînement d'un modèle basé sur un transformateur pour l'analyse des sentiments en utilisant des ensembles de données ouvertes.

Évaluation et déploiement de modèles

- Mesures d'évaluation : précision, rappel, score F1
- Matrices de confusion
- Stratégies de validation
- Méthodes de validation croisée
- Techniques de déploiement de modèles
- Applications réelles des modèles NLP

Activités pratiques :

- Évaluer un modèle NLP à l'aide de mesures spécifiques
- Effectuer une validation croisée
- Déployer le modèle via l'API
- Créer une application légère pour le déploiement

Considérations éthiques et pratiques dans le domaine du NLP

- Questions éthiques dans le NLP
- Biais dans les ensembles de données
- Transparence dans les grands modèles de langage (LLM)
- Implications du contenu généré
- Considérations sur l'équité
- Préoccupations relatives à la confidentialité des données

Activités pratiques :

- Analyse d'une étude de cas sur les défis éthiques dans le domaine de la PNL
- Réfléchir à des solutions pour réduire les risques de modèles biaisés ou non transparents.

Prochaines sessions

DATE	LIEU	STATUT
30/11/2026	à distance	Sur confirmation

Pourquoi choisir Sparks Formation ?

20 ans d'expertise IT, des formateurs experts et des sessions garanties contre l'annulation.

20 ans
d'expertise

1 068+
formations

11 178
apprenants formés/an

+67
Net Promoter Score

Avis des participants

Note moyenne des participants de la filière Data Science & Machine Learning : 4,9/5 — 21 avis

Anthony G. — Formation très bien organisée, très claire Alison explique très clairement et est très à l'écoute, avec toujours une approche pratique essentielle. je n'avais jamais fait de Python et...

François H. — François-Marie est un excellent formateur ! Un grand merci à lui d'avoir été aussi pédagogue et clair dans ses explications du début à la fin de la formation,...

Votre formateur

Semeh B. — Formateur Machine Learning & Data Science

Semeh intervient sur des formations portant sur la Data Science et le Machine Learning. Parmi les formations qui lui sont confiées figurent notamment « Computer Vision » et « Machine...

Ils nous font confiance

Cette année notre filière Data Science & Machine Learning a accompagné plus de 70 entreprises dont :

SOPRA STERIA RTE AMADEUS SAS CEGEDIM LBC FRANCE SNCF - E.SNCF SOLUTIONS AMADEUS DATA PROCESSING GMBH CSTB GERS SA
ADENTIS +60 autres

Financement & certifications

Cette formation est éligible aux financements OPCO et FNE-Formation. Nos équipes vous accompagnent dans le montage de votre dossier.



Réservez votre place dès maintenant

20 ans d'expertise · 1 000+ formations · NPS +67 · réponse sous 24h ouvrées

0 805 950 800

demande@sparks-formation.com