

Formation NLP (Natural Language Processing)

Améliorez les capacités de votre équipe avec notre formation Natural Language Processing ! Ce cours permet aux professionnels d'acquérir des compétences essentielles en matière de techniques de prétraitement NLP, de représentations vectorielles de texte et d'architectures de réseaux neuronaux avancés, leur permettant ainsi d'intégrer de manière transparente l'apprentissage automatique et l'IA dans leurs flux de travail. Les participants acquerront une expérience pratique avec des bibliothèques populaires telles que spaCy et NLTK, ce qui leur permettra de relever des défis concrets tels que l'analyse des sentiments et le résumé de texte. Investissez dans l'avenir de votre équipe - inscrivez-vous dès aujourd'hui et transformez leur approche de la data !

Durée

3 jours

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

Data scientists, data analysts,
ingénieurs Big Data...

Prérequis

Connaissance préalable de la
programmation Python, des bases en
machine learning et en statistiques
générales.

Programme de formation

Phase d'inclusion

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

Exemples de travaux 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

Exemples de travaux 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

Exemples de travaux 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

Exemples de travaux 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

Exemples de travaux 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

Exemples de travaux 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.

Moyens et méthodes pédagogiques

- La formation alterne entre présentations des concepts théoriques et mises en application à travers d'ateliers et exercices pratiques (hors formation de type séminaire).
- Les participants bénéficient des retours d'expérience terrains du formateur ou de la formatrice
- Un support de cours numérique est fourni aux stagiaires

Modalités d'évaluation

- **En amont de la session de formation**, un questionnaire d'auto-positionnement est remis aux participants, afin qu'ils situent leurs connaissances et compétences déjà acquises par rapport au thème de la formation.
- **En cours de formation**, l'évaluation se fait sous forme d'ateliers, exercices et travaux pratiques de validation, de retour d'observation et/ou de partage d'expérience, en cohérence avec les objectifs pédagogiques visés.
- **En fin de session**, le formateur évalue les compétences et connaissances acquises par les apprenants grâce à un questionnaire reprenant les mêmes éléments que l'auto-positionnement, permettant ainsi une analyse détaillée de leur progression.