

Utiliser Natural Language Processing (NLP) avec Python

Tutorat en option

Formation en ligne - 02h15

Réf : 4EU - Prix 2024 : 95CHF HT

Ce cours en ligne a pour objectif de vous apprendre à réaliser des modèles de type NLP en utilisant le langage Python. La pédagogie s'appuie sur un auto-apprentissage séquentiel par actions de l'utilisateur sur l'environnement à maîtriser.

OBJECTIFS PÉDAGOGIQUES

À l'issue de la formation l'apprenant sera en mesure de :

Construire un modèle basé sur NLP

Gérer la vectorisation des données et le réglage des hyperparamètres

Utiliser l'application de Random Forest pour la construction de modèles de classification basés sur NLP

PÉDAGOGIE ET PRATIQUES

Une évaluation tout au long de la formation grâce à une pédagogie active mixant théorie, exercice, partage de pratique et gamification. Un service technique est dédié au support de l'apprenant. La formation est diffusée au format SCORM (1.2) et accessible en illimité pendant 1 an.

ACTIVITÉS DIGITALES

Démonstrations, cours enregistrés, partages de bonnes pratiques, fiches de synthèse.

LE PROGRAMME

dernière mise à jour : 06/2023

1) Introduction au Natural Language Processing (NLP)

- Définition du Natural Language Processing.
- Les étapes de modélisation en NLP.
- L'évaluation d'un modèle.
- La validation croisée d'un modèle.
- La matrice de confusion d'un modèle NLP.
- L'overfitting.
- La distribution Anaconda.
- L'environnement Jupyter.

2) Manipulation et préparation des données textuelles avec Python

- Charger des données textuelles.
- Connaître les expressions régulières.
- Travailler avec le module Re.
- Utiliser les fonctionnalités avancées avec le module Re.
- Nettoyer un texte.
- Supprimer les stopwords.
- Utiliser le stemming.
- Utiliser la lemmatisation.
- Choisir entre le stemming et la lemmatisation.

PARTICIPANTS

Développeurs Python souhaitant réaliser des modèles de type NLP.

PRÉREQUIS

Bonnes connaissances du langage Python.

COMPÉTENCES DU FORMATEUR

Les experts qui ont conçu la formation et qui accompagnent les apprenants dans le cadre d'un tutorat sont des spécialistes des sujets traités. Ils ont été validés par nos équipes pédagogiques tant sur le plan des connaissances métiers que sur celui de la pédagogie, et ce pour chaque cours. Ils ont au minimum cinq à dix années d'expérience dans leur domaine et occupent ou ont occupé des postes à responsabilité en entreprise.

MODALITÉS D'ÉVALUATION

La progression de l'apprenant est évaluée tout au long de sa formation au moyen de QCM, d'exercices pratiques, de tests ou d'échanges pédagogiques. Sa satisfaction est aussi évaluée à l'issue de sa formation grâce à un questionnaire.

MOYENS PÉDAGOGIQUES ET TECHNIQUES

Les moyens pédagogiques et les méthodes d'enseignement utilisés sont principalement : documentation et support de cours, exercices pratiques d'application et corrigés des exercices, études de cas ou présentation de cas réels. ORSYS fournit aux participants un questionnaire d'évaluation du cours qui est ensuite analysé par nos équipes pédagogiques. Une attestation de fin de formation est fournie si l'apprenant a bien suivi la totalité de la formation.

MODALITÉS ET DÉLAIS D'ACCÈS

L'inscription doit être finalisée 24 heures avant le début de la formation.

ACCESSIBILITÉ AUX PERSONNES HANDICAPÉES

Vous avez un besoin spécifique d'accessibilité ? Contactez Mme FOSSE, référente handicap, à l'adresse suivante psh-accueil@orsys.fr pour étudier au mieux votre demande et sa faisabilité.

3) Vectorisation de données

- Introduction aux techniques de vectorisation.
- La vectorisation d'un texte avec CountVectorizer.
- La vectorisation contextuelle avec N-grams.
- La technique TF-IDF.
- L'implémentation TF-IDF.
- Le Feature Engineering en NLP.
- L'ajout de nouvelles Features.
- L'analyse et la comparaison des Features.

4) Modélisation NLP avec Random Forest

- Introduction à Random Forest.
- Développer un modèle NLP avec Random Forest.
- Mesurer les performances d'un modèle.
- Gérer la validation k-fold avec Random Forest.
- Tester plusieurs paramètres du Random Forest.
- Appliquer le GridSearchCV avec Random Forest.
- Homogénéiser les données avec TF-IDF.
- Calculer l'impact des Features.
- Sélectionner un modèle.