

Réf : ET_DEV_PYTIADAS

Python pour l'Intelligence Artificielle et la Data Science

Objectif

Cette e-formation sur le langage Python permet d'expérimenter les techniques d'Intelligence Artificielle et de Data Science mises en œuvre dans la réalisation de modèles de type NLP, l'analyse de gros volumes de données ou encore le développement de modèles neuronaux avec TensorFlow.

Public

Data Scientist

Prérequis

Une maîtrise du langage Python et un bon niveau en mathématiques est recommandé.

Durée estimée

pour le suivi des modules indispensables

13 à 15 heures

Contenu pédagogique

► Les modules indispensables

Le Machine Learning avec Python - De la théorie à la pratique



Livre

Ce module vous propose la consultation d'un support de cours numérique.

- La Data Science
- Le langage Python
- La bibliothèque NumPy
- La bibliothèque Pandas
- Travailler avec Jupyter
- Statistiques
- La régression linéaire et polynomiale
- La régression logistique
- Arbres de décision et Random Forest
- L'algorithme k-means
- Support Vector Machine
- Analyse en composantes principales
- Les réseaux de neurones
- Le Deep Learning avec TensorFlow
- Le traitement automatique du langage
- La programmation orientée objet avec Python

Python pour la Data Science



Ce module vous propose la consultation d'une vidéo d'une durée de 2h35.

- **Python et la Data Science**
 - Choisir Python pour la Data Science
 - La bibliothèque NumPy
 - Le type et la taille des vecteurs NumPy
 - Initialisation et tableau NumPy
 - Accéder aux données d'un tableau NumPy à une dimension
 - Accéder aux données d'un tableau NumPy à deux dimensions
 - Algèbre linéaire avec NumPy
 - Tableau NumPy versus liste Python
 - Statistique descriptive avec NumPy
- **Visualisation de données**
 - Installer Anaconda et Jupyter
 - Travailler avec Jupyter
 - La bibliothèque Pandas
 - Accéder aux données d'un Data Frame
 - Filtrer les données d'un Data Frame
 - Trier les données d'un Data Frame
 - Les statistiques de base avec un Data Frame
 - Lecture des fichiers de grandes tailles avec Pandas
- **Les statistiques inférentielles avec Python**
 - Utilisation des méthodes melt et apply
 - Extraire des informations à partir des données existantes
 - Créer de nouvelles variables à partir des données existantes
 - Visualiser les données avec Matplotlib
 - La loi Normale
 - Introduction aux tests d'hypothèses
 - Test statistique de comparaison de deux moyennes
- **Modélisation multivariée avec Python**
 - Introduction à la régression linéaire
 - Exemple de modélisation avec la régression linéaire
 - Introduction à l'algorithme Support Vector Machine
 - Exemple de modélisation avec un Support Vector Machine
 - Introduction à l'algorithme de K-Means
 - Exemple de l'algorithme K-Means
 - Conclusion

Natural Language Processing (NLP) avec Python



Vidéo

Ce module vous propose la consultation d'une vidéo d'une durée de 2h18.

- **Introduction au Natural Language Processing (NLP)**
 - Introduction et définition
 - Quelques cas d'application
 - Les étapes de modélisation en NLP
 - Evaluer un modèle
 - La validation croisée d'un modèle
 - La matrice de confusion d'un modèle NLP
 - Repérer l'overfitting
 - Installer la distribution Anaconda
 - Découvrir l'environnement Jupyter
- **Manipulation et préparation des données textuelles avec Python**
 - Chargement des données textuelles
 - Les expressions régulières
 - Travailler avec le module Re
 - Fonctionnalités avancées avec le module Re
 - Nettoyer un texte
 - Supprimer les Stopwords
 - Le stemming
 - La lemmatisation
 - Le stemming versus la lemmatisation
- **La 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
 - Implémenter TF-IDF
 - Le Feature Engineering en NLP
 - Ajout de nouvelles Features
 - Analyse et comparaison des Features
- **Modélisation NLP avec Random Forest**
 - Introduction à Random Forest
 - Développer un modèle NLP avec Random Forest
 - Mesurer les performances d'un modèle
 - 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
 - Conclusion générale

Les réseaux de neurones avec TensorFlow



Vidéo

Ce module vous propose la consultation d'une vidéo d'une durée de 2h39.

- **Les réseaux de neurones**
 - Introduction aux réseaux de neurones
 - Quelques cas d'application
 - Les étapes de modélisation en Data Science
 - Du neurone biologique au neurone artificiel
 - Les fonctions d'assemblage
 - Les fonctions d'activation
 - Introduction au perceptron
 - Du perceptron au réseau de neurones
 - Entraîner un réseau de neurones
- **L'algorithme de rétropropagation**
 - Présentation de l'algorithme
 - La propagation avant
 - Le calcul des erreurs à la sortie d'un réseau de neurones
 - La propagation arrière
 - La mise à jour d'un réseau de neurones
 - Apprentissage par itérations
 - Critères d'arrêt de la phase d'apprentissage
- **Introduction à TensorFlow 2**
 - Installer Anaconda
 - Créer un environnement virtuel et installer TensorFlow
 - Architecture de TensorFlow
 - Architecture de tf.keras
 - Introduction aux Tensors
 - Manipuler les variables avec les Tensors
 - Initialisation des Tensors
 - Les opérations algébriques avec les Tensors
- **Les réseaux de neurones avec TensorFlow 2**
 - Charger les données pour entraîner un réseau de neurones
 - Créer un MLP avec TensorFlow
 - Accès aux informations et initialisation des paramètres d'un réseau de neurones
 - Exécuter un réseau de neurones sur des données
 - Évaluer un réseau de neurones
 - Prédire avec un réseau de neurones
 - Utiliser le Callback
 - Configurer le critère d'arrêt
 - Sauvegarder un réseau de neurones dans un fichier
 - Gérer les hyperparamètres d'un réseau de neurones
 - Conclusion