

Machine Learning avec Python du POC à la production

Python au service de la Data Science

Cours Pratique de 3 jours - 21h
Réf : PYD - Prix 2024 : 1 870€ HT

Le cours vous enseigne l'utilisation de Python pour la Data Science : la préparation des données, l'entraînement et la mise à disposition du modèle et des résultats. Les participants apprennent à utiliser divers outils et bibliothèques Python pour effectuer des tâches courantes de data science et de machine learning.

OBJECTIFS PÉDAGOGIQUES

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

Mettre en place les différentes étapes de préprocessing avec Python

Savoir choisir le modèle approprié pour une problématique donnée

Appliquer et évaluer des modèles sur des données réelles

Mettre à disposition un modèle dans le cloud et permettre son interrogation par le biais du API

MÉTHODES PÉDAGOGIQUES

Conçue par des experts, suite à leur retours d'expérience, cette formation passe en revue, les différentes étapes d'un projet de Machine Learning de sa conception à sa mise en production.

TRAVAUX PRATIQUES

La formation met l'accent sur la mise en pratique, afin d'assurer l'autonomie des participants.

LE PROGRAMME

dernière mise à jour : 05/2023

1) Import des données et preprocessing

- L'environnement de développement Python / Anaconda / Jupyter Notebook.
- Pandas : l'analyse de données tabulaires (CSV, Excel...), statistiques, pivots, jointures, filtres.
- Traitement des valeurs manquantes : imputation par la moyenne, médiane, interpolation, knn...
- Traitement des Outliers : analyse graphique, méthode de l'IQR, Z-score.
- Standardisation.
- Normalisation : Skewness et Kurtosis.
- Données non balancées : Undersampling, Oversampling, SMOTE.

Travaux pratiques : Manipulation de Python dans un notebook Jupyter. Exercice de mise en pratique avec pandas.

Mise en place de l'ensemble des pré-traitements à l'aide des bibliothèques python spécifiques.

2) Entraînement de modèles et évaluation

- Modèles d'apprentissage supervisés et non-supervisés les plus courants.
- Entraînement de modèles avec Scikit-learn.
- Méthodes d'évaluations : savoir choisir les bonnes métriques pour chaque problématique.

Travaux pratiques : Entraînement de plusieurs modèles supervisés et non supervisés, comparaison des performances et choix du meilleur modèle.

PARTICIPANTS

Toutes personnes intéressées par l'apprentissage de Python et son application à la data science et au machine learning.

PRÉREQUIS

Connaissance du langage Python.
Connaissances théoriques sur le machine learning.

COMPÉTENCES DU FORMATEUR

Les experts qui animent la formation sont des spécialistes des matières abordées. 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 qu'ils enseignent. 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

Le formateur évalue la progression pédagogique du participant tout au long de la formation au moyen de QCM, mises en situation, travaux pratiques...

Le participant complète également un test de positionnement en amont et en aval pour valider les compétences acquises.

MOYENS PÉDAGOGIQUES ET TECHNIQUES

- Les moyens pédagogiques et les méthodes d'enseignement utilisés sont principalement : aides audiovisuelles, documentation et support de cours, exercices pratiques d'application et corrigés des exercices pour les stages pratiques, études de cas ou présentation de cas réels pour les séminaires de formation.
- À l'issue de chaque stage ou séminaire, ORSYS fournit aux participants un questionnaire d'évaluation du cours qui est ensuite analysé par nos équipes pédagogiques.
- Une feuille d'émargement par demi-journée de présence est fournie en fin de formation ainsi qu'une attestation de fin de formation si le stagiaire a bien assisté à la totalité de la session.

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) Optimisation des modèles et log des performances

- Présentation des bibliothèques Optuna, Hyperopt.
- Présentation de l'approche Grid Search pour identifier les meilleurs hyper paramètres d'un modèle.
- Log des hyper paramètres et des performances dans Mlflow.

Travaux pratiques : Optimisation des modèles élaborés dans la partie précédente et log des métriques / hyperparamètres dans Mlflow.

4) Modèle et Data Drift

- Intérêt de vérifier le modèle Drift et le Data Drift.
- Présentation des bibliothèques Evidently et Streamlit.

Travaux pratiques : Mise en place d'un Dashboard Evidently pour monitorer le Drift des données.

5) Industrialisation : déploiement dans le cloud

- Présentation du service AWS EC2.
- Présentation de Flask pour la mise à disposition d'un modèle de machine learning par le biais d'une API.
- Présentation de divers outils de connexion à l'environnement virtuel tels que Putty, Visual Studio Code...
- Déploiement du code par le biais de GitHub.

Travaux pratiques : Déploiement d'un modèle sur un environnement cloud avec la bibliothèque Flask.

LES DATES

CLASSE À DISTANCE
2024 : 26 juin, 07 oct., 02 déc.

PARIS
2024 : 19 juin, 30 sept., 25 nov.