CURSUS - ACTUAIRES : DEVENEZ DATA SCIENTIST

Se former aux langages R & Python

8 JOURS, 56 HEURES

INTELLIGENCE ARTIFICIELLE (IA) ET STRATÉGIE DIGITALE

CODE : ACU03

Objectifs de la formation

Définir l'environnement de développement en R

Coder des opérations sur R

Distinguer les méthodes d'apprentissage

Appréhender l'environnement python

Coder sur Pyhton

Parmi nos formateurs

- **SHI Kezhan**
  Data scientist et actuaire certifié,
  DIOT-SIACI

Public concernés

- Toute personne ayant une première expérience confirmée dans le milieu de l'actuariat et souhaitant avoir des missions de data scientist

Critères d'admission

- Pour participer à chacun des modules, les stagiaires doivent être munis d'un ordinateur portable et avoir téléchargé les packages R et Pyhton (logiciels Open Source et gratuits)

Prérequis

- Aucun prérequis n'est nécessaire

MODULE 1 - PROGRAMMER EN PYTHON : NIVEAU 1

Appréhender les principes de bases du développement

- Présentation du langage Python
- Installer le logiciel Python
- Savoir utiliser Python : Jupyter Notebook
- Maîtriser les différentes fonctionnalités : les imports, les variables, les fichiers Script, les boucles, les fonctions

Connaître les formats de données et réaliser des opérations basiques

- Qu'est-ce qu'un Iterables

- Mener des opérations sur les itérables
- Comprendre l'utilité des tableaux de données (dataframes) et savoir les créer
- Apprendre à créer et utiliser des classes

Manipulation avancée des données / étude de cas

- Description du dataset utilisé
- Lire des fichiers .csv
- Manipulation des tableaux de données : utiliser le module package Panda
- Manipuler les graphiques

Le module inclut de nombreuses mises en pratique sur le logiciel Python

MODULE 2 - ACQUERIR LES BASES DU MACHINE LEARNING AVEC PYTHON : NIVEAU 2

Cerner les outils et techniques nécessaires aux méthodes de l'analyse de données et au machine learning

- Assimiler les techniques de data cleaning
- Comprendre le fonctionnement du formatage de données

Savoir gérer l'acquisition et la préparation des données : open data (INSEE, météo… API, scrapping web…)

Maîtriser les algorithmes supervisés

- Comment les algorithmes s'entraînent-ils ? Illustration d'exemples de régressions et de l'arbre de décision
- Comment optimiser les paramètres des algorithmes
- Tester des algorithmes supervisés : méthodologie
- Comprendre les hyperparamètres et leurs fonctionnements
- Gérer le surapprentissage
- Connaître les principaux algorithmes : SVM, bayésien naïf, réseaux baysésiens, réseaux de neurones

Savoir utiliser les algorithmes non supervisés

- Comprendre le clustering

Comprendre la datavisualisation : Matplotlib

Assimiler les méthodes de text mining

- Procéder à une extraction de texte dans un document
- Construction de variables explicatives à partir du texte
- Application de modèles supervisés sur des données textuelles

Le module inclut de nombreuses mises en pratique sur le logiciel Python

MODULE 3 - PROGRAMMER EN R : NIVEAU 1

Appréhender les principes de bases du développement en R

- Présentation du langage R
- Installer le logiciel R
- Savoir utiliser R : premiers pas avec R
- Maîtriser les différentes fonctionnalités : les variables, les fichiers Script, les boucles, les fonctions

Connaitre les formats de données et réaliser des opérations basiques

- Maîtriser les structures de données : Vecteurs, Listes, Matrices, Dataframes
- Exercices pratiques sur les opérations de structures de données *Facteurs

Engager des manipulations avancées des données

- Savoir décrire le dataset utilisé
- Lire des fichiers .csv
- Package dplyr, manipulation des tableaux de données
- Créer des graphiques
- Étude de cas : entraînement de premiers modèles et interprétations

Le module inclut de nombreuses mises en pratique sur le logiciel R

Module 4 - Acquérir les bases du Machine Learning avec R : Niveau 2

Cerner les outils et techniques nécessaires aux méthodes de l'analyse de données et au machine learning

- Assimiler les techniques de data cleaning
- Comprendre le fonctionnement du formatage de données

Savoir gérer l'acquisition et la préparation des données : open data (INSEE, météo… API, scrapping web…)

Maîtriser les algorithmes supervisés

- Gérer l'implémentation concrète
- Comment les algorithmes s'entraînent-ils ? Illustration d'exemples de régressions et de l'arbre de décision
- Comment optimiser les paramètres des algorithmes
- Tester des algorithmes supervisés : méthodologie
- Comprendre les hyperparamètres et leurs fonctionnements
- Gérer le surapprentissage
- Connaître les principaux algorithmes : SVM, bayésien naïf, réseaux baysésiens, réseaux de neurones

Savoir utiliser les algorithmes non supervisés

- Comprendre le clustering

Comprendre la datavisualisation : Application RShiny

Assimiler les méthodes de text mining

- Procéder à une extraction de texte dans un document

- Construction de variables explicatives à partir du texte
- Application de modèles supervisés sur des données textuelles

Le module inclut de nombreuses mises en pratique sur le logiciel R

Dates

Modalités pédagogiques, d'évaluation et techniques

- **Modalités pédagogiques:**

Pour les formations synchrones-présentiel ou classes virtuelles (formations à distance, en direct), les stages sont limités, dans la mesure du possible, à une douzaine de participants, et cherchent à respecter un équilibre entre théorie et pratique. Chaque fois que cela est possible et pertinent, des études de cas, des mises en pratique ou en situation, des exercices sont proposées aux stagiaires, permettant ainsi de valider les acquis au cours de la formation. Les stagiaires peuvent interagir avec le formateur ou les autres participants tout au long de la formation, y compris sur les classes virtuelles durant lesquelles le formateur, comme en présentiel peut distribuer des documents tout au long de la formation via la plateforme. Un questionnaire préalable dit 'questionnaire pédagogique' est envoyé aux participants pour recueillir leurs besoins et attentes spécifiques. Il est transmis aux intervenant(e)s avant la formation, leur permettant de s'adapter aux publics. Pour les formations en E-learning (formations à distance, asynchrones), le stagiaire peut suivre la formation à son rythme, quand il le souhaite. L'expérience alterne des vidéos de contenu et des activités pédagogiques de type quizz permettant de tester et de valider ses acquis tout au long du parcours. Des fiches mémos reprenant l'essentiel de la formation sont téléchargeables. La présence d'un forum de discussion permet un accompagnement pédagogique personnalisé. Un quizz de validation des acquis clôture chaque parcours. Enfin, le blended-learning est un parcours alternant présentiel, classes virtuelles et/ou e-learning.

- **Modalités d'évaluation:**

Toute formation se clôture par une évaluation à chaud de la satisfaction du stagiaire sur le déroulement, l'organisation et les activités pédagogiques de la formation. Les intervenant(e)s évaluent également la session. La validation des acquis se fait en contrôle continu tout au long des parcours, via les exercices proposés. Sur certaines formations, une validation formelle des acquis peut se faire via un examen ou un QCM en fin de parcours. Une auto-évaluation des acquis pré et post formation est effectuée en ligne afin de permettre à chaque participant de mesurer sa progression à l'issue de la formation. Une évaluation à froid systématique sera effectuée à 6 mois et 12 mois pour s'assurer de l'ancrage des acquis et du transfert de compétences en situation professionnelle, soit par téléphone soit par questionnaire en ligne.

- **Modalités techniques FOAD:**

Les parcours sont accessibles depuis un simple lien web, envoyé par Email aux stagiaires. L'accès au module de E-learning se fait via la plateforme 360Learning. La durée d'accès au module se déclenche à partir de la réception de l'invitation de connexion. L'accès aux classes virtuelles se fait via la plateforme Teams. Le(a) stagiaire reçoit une invitation en amont de la session lui permettant de se connecter via un lien. Pour une bonne utilisation des fonctionnalités multimédia, vous devez disposer d'un poste informatique équipé d'une carte son et d'un dispositif vous permettant d'écouter du son (enceintes ou casque). En ce qui concerne la classe virtuelle, d'un microphone (éventuellement intégré au casque audio ou à la webcam), et éventuellement d'une webcam qui permettra aux autres participants et au formateur de vous voir. En cas de difficulté technique, le(a) stagiaire pourra contacter la hotline au 01 70 72 25 81, entre 9h et 17h ou par mail au logistiqueformations@infopro-digital.com et la prise en compte de la demande se fera dans les 48h.