

ANNEXE – 1

Nom du cours : Certification en science des données avec Python

Objectif du cours : L'objectif de ce cours est de créer une expertise en programmation Python pour développer les applications de science des données.

Prérequis : Le candidat doit maîtriser les bases de l'informatique, les bases de données de base et les concepts de programmation avec approche logique.

Module 1: Introduction au langage de programmation Python

1. Installation et configuration,
2. Création et sauvegarde d'un fichier script,
3. Bases de python,
4. Types de données,
5. Variables.
6. Syntaxe et commentaires en Python : comprendre la syntaxe et commenter le code.
7. Manipulation de chaînes de caractères Python : type de données de chaîne, indexation de chaîne, découpage, concaténation et formatage.
8. Introduction aux opérateurs python :
 - Opérateurs arithmétiques
 - Opérateurs de comparaison
 - Opérateurs logiques
 - Opérateurs d'affectation

Module 2: comprendre le concept de structure de données en Python et leur implémentation

1. Listes Python,
2. Tuples, lignes
- 3 ensembles,
4. Dictionnaires
5. Instructions conditionnelles et boucle en python
 - Sinon
 - Boucle While
 - Pour la boucle
 - Boucles imbriquées,
6. Introduction aux fonctions python
 - Définir les fonctions
 - Arguments
 - Déclaration de retour
 - Portée des variables

7. Introduction à la manipulation des données

Introduction à NumPy, package NumPy en Python, importation de NumPy, création de différents tableaux à l'aide de NumPy, fonctions et méthodes de tableau, différentes fonctions mathématiques, différentes opérations matricielles, nombres aléatoires, génération de nombres entre des plages.

Module 3: Notions de base de Pandas pour la manipulation de données en Python

1. Comprendre les opérations de la bibliothèque Pandas, des séries et des trames de données

- Concepts de séries dans Pandas, création à l'aide de Pandas, différents attributs de séries, séries vs listes, opérations sur les séries, séries à partir de fichiers CSV, différentes fonctions dans les séries, différents algorithmes de tri dans les séries, extraction de valeurs à partir de séries, .value counts (), . appliquer () les méthodes,

2. Analyse exploratoire des données

- Lecture de fichiers, Nettoyage des données en python : Gestion des valeurs manquantes et remplissage NA
- Préparation et prétraitement des données
- Ingénierie des caractéristiques des données : gestion des données catégorielles
- Techniques de validation de données en python
- Ingénierie des caractéristiques des données : suppression des colonnes et des lignes des données brutes

Module 4 : Visualisation des données sur l'utilisation des bibliothèques matplotlib et seaborn

- Nuage de points
- Graphique linéaire
- Diagramme à barres
- Histogramme
- Boîte à moustaches
- Parcelle de paires

Module 5 : Introduction à l'apprentissage automatique en Python

- Moyenne, médiane et mode
- Techniques d'apprentissage automatique pour l'apprentissage supervisé et non supervisé
- Analyse de régression dans l'apprentissage automatique
- Régression linéaire simple en Python
- Régression linéaire multiple en Python
- Erreur quadratique moyenne, score R2, corrélation
- Matrice de confusion, score de precision

Projet final : Présentez toutes vos compétences dans un projet d'analyse de données de bout en bout. Vous choisirez l'ensemble de données, ferez le tri des données, poserez les

questions de recherche, visualiserez les données, tirerez des conclusions et présenterez vos résultats