

# Métiers de la donnée big data et de l'intelligence artificielle

## PROGRAMME DE FORMATION

### ACCUEIL, DECOUVERTE METIER

### DEVELOPPEMENT DURABLE

- Connaissance de base du développement durable appliqué dans l'entreprise

### TECHNIQUES DE RECHERCHE D'EMPLOI

### CITOYENNETE

- Développer ses connaissances en base en ce qui concerne le fonctionnement de la société et la santé en société, en entreprise et en formation
- Favoriser la prise de conscience citoyenne des stagiaires notamment à l'occasion des séquences collectives

### BILAN

### NUMÉRIQUE INITIATION INFORMATIQUE

- Remise à niveau des savoirs fondamentaux en informatique en fonction des pré-acquis de chacun des apprenants, acquisition d'une méthode d'une recherche d'emploi sur internet, sensibilisation à la e-réputation.

### ANGLAIS – VOCABULAIRE TECHNIQUE

- Comprendre et se faire comprendre en anglais technique
- Lire et comprendre une documentation technique savoir écrire un compte rendu / un e-mail professionnel en anglais
- Développer ses aptitudes à communiquer en anglais dans un contexte professionnel, à l'oral et à l'écrit.
- Bilan linguistique
- Acquisition et révision du vocabulaire professionnel
- Révision de la syntaxe et grammaire
- Travail sur les différentes sections du test

### ACCOMPAGNEMENT A LA FAD

- Apprendre à apprendre
- les stratégies d'apprentissage à distance
- Se créer un cadre propice à l'apprentissage à distance
- Utiliser moodle et webex
- Aller chercher l'information sur internet
- Gérer son temps et mettre en place un planning

### GESTION DE PROJET

- Organisation et gestion de projet  
Comprendre les fondamentaux d'ITIL et se certifier
- Adopter les meilleurs pratiques d'organisation d'un service informatique - Passer la certification ITIL
- Gestion de projet informatique  
Comprendre les points essentiels des techniques de conduite de projets informatiques
- Gérer un projet avec MS Project
- Utiliser l'outil Project pour planifier, suivre et contrôler vos projets.

### PROJET PERSONNEL

- Maîtriser l'ensemble des étapes de la création d'un projet

### FONDAMENTAUX DE LA STATISTIQUE

- L'intégration des données
- Importer des données
- Créer des connexions avec des sources externes
- Contrôler la saisie avec la validation des données d'Excel
- Savoir réaliser un questionnaire de satisfaction
- Données à caractère discret ou continu
- L'échantillonnage avec l'utilitaire d'analyse
- RGRRRRRRRRRR
- Les tableaux statistiques
- Les tableaux de contingence avec les tableaux croisés dynamiques
- Utiliser les fonctions NB, NBVAL, NB.ENS.SI, NB.VIDE (caractères quantitatifs, qualitatifs, discrets) pour compter les valeurs
- Caractères continus :
- Classes

- Bornes
- Centres de classe
- Déterminer empiriquement le nombre de classes avec RACINE(NB(()))
- Regrouper les valeurs continues dans les TCD
- Les tableaux de fréquence ("fréquence")
- Choisir le bon graphique pour présenter les données
- Graphiques à échelle arithmétique
- Diagrammes de hauteur :
  - Tuyaux d'orgue (barres, colonnes)
  - Radar
  - Diagrammes angulaires : secteurs
  - Un graphique semi-circulaire : l'hémicycle
- Diagrammes de surface : les histogrammes avec l'utilitaire d'analyse
- Séries chronologiques : lignes et courbes
- Les graphiques avec échelle logarithmique (outils graphiques, axes)
- Quand les échelles arithmétiques déforment les évolutions relatives
- Représenter des variations relatives entre plusieurs séries en paramétrant les axes
- Représenter des séries de grandeurs différentes sur un même graphique
- Savoir gérer des axes saturés
- Les indicateurs de position, de valeur centrale et les pièges de la moyenne
- Le piège des moyennes
- Moyenne, médiane, mode, quantiles, quartiles
- La méthode "Shift and share" pour isoler l'effet de structure
- Analyse de position (RANG) avec l'utilitaire d'analyse
- Autres indicateurs de position :
  - Grande valeur
  - Petite valeur
  - MINA
  - MAXA
  - COEFFICIENT.ASYMETRIE
  - MOYENNE.RANG
- Autres indicateurs de valeurs centrales :
  - MOYENNE.REDUITE
  - CENTREE.REDUITE
  - AVERAGEA
  - MOYENNE.GEOMETRIQUE
- La mise en forme conditionnelle pour les indicateurs de position
- Règles des valeurs plus ou moins élevées
- Barres de données
- Nuances de couleurs
- Jeux d'icônes
- Les indicateurs de dispersion autour de la valeur centrale
  - MIN et MAX
  - Intervalles
  - Interquartiles
  - Construire une boîte à moustaches
  - Variance, écart type et coefficient de variation
  - Les variances expliquées
  - Autres indicateurs de dispersion
    - COVARIANCE.PEARSON
    - KURTOSIS
  - Richesses réparties ?
- Les indicateurs de concentration avec la médiale
- Calculer la médiale
- Construire le carré de Gini
- Statistiques économiques : les indices
  - Indice simple de prix et volume
  - Indice synthétique prix et volume
  - Laspeyres et Paasche avec SOMMEPROD
  - Indice synthétique et calcul matriciel
  - L'indice IPC de l'INSEE
- Analyser les séries chronologiques pour les prévisions
- Les tableaux à double décomposition
- Composante structurelle : la tendance
- Composante conjoncturelle : saisonnalité
- Lissage par moyennes mobiles :
  - Utilitaire d'analyse
  - Moyennes mobiles
  - Les séries corrigées des variations saisonnières
  - CVS
  - Le modèle additif

- Le modèle multiplicatif
- Série ajustée prévisionnelle
- La corrélation, la liaison entre les événements
- Visualisation : nuage de point et tableaux
- Calcul du centre de gravité du nuage de point
- Liaisons nulles, totales, relatives
- La droite de régression (DROITEREG, PENTE, ORDONNEE.ORIGINE)
- Coefficient de détermination (COEFFICIENT.DETERMINATION)
- Le piège des liaisons paraboliques
- Le piège du facteur caché : les liaisons indirectes

#### **GÉRER UN PROJET BIG DATA**

- Bien comprendre les challenges dans la mise en place d'un projet Big Data
- Mesurer l'impact politique et stratégique sur l'entreprise et la pertinence d'un reporting régulier
- Bien comprendre les spécificités de ce genre de projets

#### **MAITRISER HADOOP ET SPARK**

- Savoir comment Hadoop Distributed File System (HDFS) et YARN / MapReduce fonctionnent
- Explorer HDFS
- Suivre l'exécution d'une application YARN
- Utiliser les différents outils de manipulation de la donnée tels que : Hue, Hive, Pig, Tez, Sqoop et Oozie.

#### **LE NOSQL AVEC MONGODB ET NEO4J**

- Comprendre et raisonner avec des modèles de bases de données graphes
- Maîtriser les API unitaires et d'indexation
- Exprimer et diagnostiquer des requêtes, même complexes
- Interagir avec Neo4j à distance
- Etendre les fonctionnalités de Neo4j
- Importer des données existantes
- Partager des Proofs Of Concept (POC) sous forme de Graph Gist.

#### **LA LEGISLATION AUTOUR DE LA DONNÉE : DEVENEZ DPO DATA PROTECTION OFFICIER)**

- Comprendre les enjeux, gérer les risques et obligations légales du RGPD
- Développer une méthodologie pour l'implémentation d'une RGPD

#### **SÉCURISER LA DONNÉE**

- Apprendre les techniques de sécurité pour mieux protéger les SI dans le monde du Big Data

#### **DATA VISUALISATION**

- Développer des applications avec Spark Streaming
- Utiliser la programmation parallèle sur un cluster
- Exploiter des données avec Spark SQL
- Comprendre et exploiter les algorithmes standard de Machine Learning et permettre une visualisation avec Graphx.