

20/02/2017



Emetteur :
P.PARREND



Formation Métier Spécialisée

« Data scientist »

L'objectif de ce document est de présenter la proposition de formation 'Analyse de données' proposée par l'ECAM Strasbourg-Europe en partenariat avec les entreprises leaders dans le domaine du traitement de données sur le territoire Alsace et la commission Adaptation des Compétences de Numéric Emploi Grand-Est.

OBJECTIF DE LA FORMATION

La formation métier proposée a pour objectif de préparer au métier de Data Scientist, c'est-à-dire aux personnes **qui donnent du sens à la valeur pour l'organisation**.

L'objectif est de former des personnes ayant une compétence opérationnelle de bout en bout sur la chaîne de traitement de la donnée : intégration, exploitation et analyse, visualisation, déploiement et industrialisation. Les personnes formées seront capables de porter des missions de valorisation des données de leur entreprise, mais également de s'insérer dans des projets numériques mettant en œuvre des données.

Il ne s'agit pas de former des spécialistes des théories de l'analyse de données, ni des spécialistes des infrastructures informatiques souvent associées à l'analyse de données.

La formation cible en priorité les publics suivants :

- Candidats de Niveau Bac+4, Bac+5 ou plus ayant une expérience sérieuse et des connaissances en informatique souhaitant se spécialiser en analyse de données (notamment les spécialistes de Business Intelligence voulant acquérir les compétences Analytics)
- Candidats de Niveau Bac+4, Bac+5 ou plus ayant une expérience d'analyse de données dans un autre domaine métier que l'informatique
- Candidats de Niveau Bac+4, Bac+5 ou plus en reconversion professionnelle ayant l'aptitude pour acquérir les connaissances d'analyse de données et ayant un minimum de connaissances et de pratique de base en informatique

1/ PROGRAMME DETAILLE EN PRECISANT POUR CHAQUE SUJET

DUREE PREVUE PAR MODALITES PEDAGOGIQUES (COURS MAGISTRAL, TRAVAUX PRATIQUES / DIRIGES)

Dispositif POEI

Le volume total de la formation est :

- 266h de cours (Cours magistraux, travaux dirigés, projets), soit 38j (7,6 semaines)
- 133h de stage en entreprise, soit 19j (3,8 semaines)

La formation comprend 4 domaines de compétences :

- Intégrer les données
- Exploiter et analyser les données
- Visualiser les données
- Déployer et industrialiser l'analyse de données

Dont les volumes horaires respectifs de cours (hC), travaux dirigés (hTD), travaux pratiques (hTP), projets (hP), ainsi que le total des heures encadrées sont récapitulés ci-dessous.

UE (Unité d'enseignements)	ECTS	Coef.	Ref module	EC (Elément Constitutif)	hC	hTD	hTP	hP	Heures encadrées	
Intégrer les données		2	PO	Programmation Objet	7		14		21	
		2	PSQL	Programmation SQL	3,5	3,5	21		28	
		1	OTD	Outils pour le traitement des données	3,5		3,5		7	
		2	IntProj	Projet 'Intégrer les données'				21	21	
					7	14	3,5	38,5	21	77

UE (Unité d'enseignements)	ECTS	Coef.	Ref module	EC (Elément Constitutif)	hC	hTD	hTP	hP	Heures encadrées	
Exploiter et analyser les données		1	SBD	Statistiques pour le Big Data	7		7		14	
		1	IA	Apprentissage automatique	7		7		14	
		1	OBD	Les outils du Big Data	7		7		14	
		1	AN	Analytics	3,5		3,5		7	
		1	AnaProj	Projet 'analyser les données'				28	28	
					5	24,5	0	24,5	28	77

UE (Unité d'enseignements)	ECTS	Coef.	Ref module	EC (Elément Constitutif)	hC	hTD	hTP	hP	Heures encadrées	
Visualiser les données		1	TW	Technologies Web	3,5		3,5		7	
		2	VD	Visualisation des données	3,5		3,5	14	21	
		1	VisProj	Projet 'visualiser les données'				28	28	
					4	7	0	7	42	56

UE (Unité d'enseignements)	ECTS	Coef.	Ref module	EC (Elément Constitutif)	hC	hTD	hTP	hP	Heures encadrées
Déployer et industrialiser l'analyse de données		1	CA	Culture analytique	3,5		3,5		7
		1	IL	Protection des données	7	7			14
		1	PGA	Analyse fonctionnelle et métier d'un projet Big Data	3,5	7	3,5		14
		3			14	14	7	0	35

UE (Unité d'enseignements)	ECTS	Coef.	Ref module	EC (Elément Constitutif)	hC	hTD	hTP	hP	Heures encadrées
Ressources Humaines		2	CV	CV et valorisation des compétences	7	14			21
					7	14	0	0	21

COMPETENCES CIBLES DE LA FORMATION

Intégrer les données

Module

Compétences

Programmation Objet

- Etre familier avec les concepts de la programmation objet
- Savoir programmer une application dans un langage orienté objet
- Etre capable de mener un projet de développement de bout en bout
- Connaître les facteurs impactant la performance des applications

Programmation SQL

- Connaître les principes de fonctionnement des bases de données relationnelles
- Etre capable de programmer des requêtes SQL complexes
- Connaître les domaines d'application des bases de données SQL et NoSQL

Outils pour le traitement des données

- Connaître les modèles de données en vue de leur traitement
- Savoir préparer des données pour le traitement
- Savoir extraire et transformer des données

Exploiter et analyser les données

<i>Module</i>	<i>Compétences</i>
<i>Statistiques pour le Big data</i>	<p>Connaitre les modèles statistiques mis en œuvre dans le big data</p> <p>Connaitre et savoir manipuler les algorithmes associés à ces modèles</p> <p>Savoir choisir un modèle statistique adapté aux besoins d'un projet big data</p> <p>Etre capable de mener un projet de mise en œuvre des statistiques pour le big data</p>
<i>Apprentissage automatique</i>	<p>Connaitre les principes théoriques du machine learning, de l'optimisation, de la détection d'anomalie</p> <p>Savoir choisir un algorithme en fonction du besoin de traitement des données</p> <p>Savoir développer et paramétrer ses propres algorithmes d'analyse de données</p> <p>Apprentissage supervisé : Arbre de décision, Régression linéaire, Régression logistique, SVM et méthodes à Noyau, Ensemble de modèles, Réseau de Neurones (& deep learning)</p> <p>Apprentissage non-supervisé : Clustering, Reduction de dimension,</p> <p>Ensembles de modeles ; Ré-enforcement ; Evaluation des modèles ; Application sur des données réelles ; Modèles prédictifs</p> <p>Les domaines traités sont: Statistical Learning Theory, Instance-based Learning ou Memory based Reasoning, Ensemble Methods, Deep Machine Learning</p>
<i>Les outils du Big Data</i>	<p>Connaitre les différents outils utilisé pour les traitements big data et leurs domaines d'application</p> <p>Connaitre les composants constitutifs d'une base de données</p> <p>Connaitre les différentes couches d'une infrastructure big data: systèmes distribués, plates-formes d'analyse</p>
<i>Analytics</i>	<p>Connaitre les outils de l'analytics</p> <p>Savoir choisir un outil adapté à ses besoins</p> <p>Savoir mettre en œuvre un outil spécifique</p>

Savoir croiser des données d'analytics avec des données tiers

Visualiser les données

Module

Compétences

Technologies Web

Connaitre des langages permettant de réaliser une application web

Etre capable de mener un projet web de bout en bout

Visualisation des données

Connaitre les particularités des outils de visualisation pour le big data

Savoir manipuler un outil représentatif de la visualisation pour le big data

Savoir mettre en œuvre une démarche de data story telling

Etre capable de mener un projet de mise en œuvre de la visualisation

Etre capable de convaincre avec les données et la visualisation

Déployer et industrialiser l'analyse de données

Module

Compétences

Culture analytique

Connaitre les mécanismes permettant de générer de l'engagement auprès des utilisateurs

Connaitre les bonnes pratiques de l'animation de communautés

Connaitre les bonnes pratiques de la relation entre IT et utilisateurs

Protection des données

Connaitre les principes de la protection des données privées

Savoir mettre en œuvre une politique Informatique et Liberté pour des applications de traitement de données

Savoir définir des solutions techniques de protection des données

Savoir réagir en cas de divulgation des données

**Analyse fonctionnelle et métier d'un projet
Big Data**

Connaitre les outils de gestion de projet data

Savoir choisir l'outil adapté selon le type de projet

LES SUPPORTS DE FORMATION FOURNIS

- Les supports de formation seront mis à disposition sous format numérique aux participants

LES MOYENS TECHNIQUES UTILISES (PLATEFORME, LOGICIELS,)

Une salle de formation de l'ECAM Strasbourg-Europe sera mise à disposition du groupe de stagiaires. Elle sera équipée d'ordinateurs ainsi que des logiciels suivants, utilisés dans les modules de formation :

- Domaine 'Intégrer les données'
 - Java v7 + eclipse
 - MS SQL, MySQL
- Domaine 'Exploiter et analyser les données'
 - Python 2,3 + R
- Domaine 'visualiser les données'
 - WAMP
 - Tableau Software

Cette liste est non exhaustive.

LES CONTENUS MIS A DISPOSITION EN AUTOFORMATION A DISTANCE

Les ressources suivantes pourront être utilisées par les stagiaires pour de l'auto-formation à distance. Elles couvrent des thématiques proches du sujet des modules, mais restent incomplètes dans la couverture des compétences identifiées par les professionnels ayant participé à la définition du syllabus.

Les cours sélectionnés sont accessibles et gratuits au moment où ce dossier est réalisé, et semblent le rester au-delà de la période d'interaction avec les enseignants (pour les MOOCs). Nous précisons quand l'accès nécessite la création d'un compte.

Les cours sélectionnés sont en français quand des ressources ont été identifiées, en anglais sinon.

Intégrer les données

Module**Ressource pour formation à distance****Programmation Objet**

Les cours suivants sont des introductions de référence :

<https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-092-introduction-to-programming-in-java-january-iap-2010/> (nécessite la création d'un compte)

<https://fr.coursera.org/learn/object-oriented-java> (accès en auditeur libre possible ; attention, cours synchrone donc la disponibilité permanente des ressources n'est pas assurée)

Programmation SQL

Administrez vos bases de données avec MySQL :

<https://openclassrooms.com/courses/administrez-vos-bases-de-donnees-avec-mysql>

Plus de ressources sont disponibles ici (en anglais ; création de compte nécessaire):

<https://lagunita.stanford.edu/courses/DB/2014/SelfPaced/about>

Outils pour le traitement des données**Exploiter et analyser les données**

Module**Ressource pour formation à distance****Statistiques pour le big data**

Effectuez vos études de statistiques avec R :

<https://openclassrooms.com/courses/effectuez-vos-etudes-statistiques-avec-r>

Plus de ressources sont disponibles ici ; création de compte nécessaire

<https://lagunita.stanford.edu/courses/HumanitiesSciences/StatLearning/Winter2016/info>

Apprentissage automatique

Programmez en Python :

<https://openclassrooms.com/courses/apprenez-a-programmer-en-python>

<https://openclassrooms.com/courses/demarrez-votre-projet-avec-python>

Plus de ressources sont disponibles ici (en anglais):

<http://openclassroom.stanford.edu/MainFolder/CoursePage.php?course=MachineLearning>

et ici (création de compte nécessaire) :

<https://lagunita.stanford.edu/courses/course-v1:ComputerScience+MMDS+Fall2016/info>

**Les outils du Big
Data**

Analytics

En cours de préparation : Optimisez votre site avec le web analytics.

<https://openclassrooms.com/courses/optimizez-votre-site-avec-le-web-analytics>

Visualiser les données

Module

Ressource pour formation à distance

Technologies Web

<https://openclassrooms.com/courses/creez-des-pages-web-interactives-avec-javascript>

Visualisation des données

Déployer et industrialiser l'analyse de données

Module

Ressource pour formation à distance

Culture analytique

Le big data transforme ma vie et celle des entreprises :

<https://openclassrooms.com/courses/le-big-data-transforme-ma-vie-et-celles-des-entreprises>

Le machine learning est-il l'avenir de l'homme :

<https://openclassrooms.com/courses/le-machine-learning-est-il-l-avenir-de-l-homme>

Protection des données

Le site de la CNIL contient de nombreux documents de référence sur le sujet : <https://www.cnil.fr>

**Analyse fonctionnelle et métier d'un projet
Big Data**

<https://openclassrooms.com/courses/decouvrez-les-bases-de-la-gestion-de-projet>

<https://openclassrooms.com/courses/apprenez-a-bien-cadrer-un-projet-multimedia>

LE RYTHME D'ALTERNANCE ENTRE SEQUENCES PEDAGOGIQUES (COURS, TP, ...) AU COURS D'UNE JOURNEE ET D'UNE SEMAINE TYPE

La formation est planifiée selon les principes suivants :

- Journées de 7h de formation
 - 9h-12h30
 - 14h-17h30
- Chaque cours est composé d'une séance de cours et applications (1 demi-journée à une journée) et de séances de travaux pratiques (1/2 journée à 3 jours)
- Chaque semaine commence par un cours lié à la culture informatique (domaine 'Déployer et Industrialiser l'analyse de données'), très importante pour des personnes en reconversion, puis se poursuit par des cours plus techniques
- Pour chacun des domaines (hors industrialisation), un projet d'intégration clos les modules de formation technique, au bout de 1 à 3 semaines selon le volume horaire du domaine et l'avancement de la formation

2/ MODALITES DE DEROULEMENT DU STAGE EN ENTREPRISE

Le suivi du stage en entreprise est organisé comme suit :

- La montée en compétence en vue du stage est soutenue dans le cadre de la formation par la réalisation de trois projets sur les domaines de formation 'Intégrer les données', 'Exploiter et analyser les données', 'Visualiser les données'
- Selon les possibilités administratives liées au mode de financement de la formation, une convention de stage ou un contrat équivalent sera établie entre l'ECAM Strasbourg-Europe et l'entreprise
- Durant le stage, chaque stagiaire sera encadré par un maître de stage en entreprise
- Chaque stagiaire sera accompagné par un tuteur qui l'accompagnera dans la préparation de son stage (entretien d'une heure) et sera disponible pour répondre aux questions de l'entreprise et du stagiaire durant le stage.
- En cas de difficulté particulière, l'équipe de l'ECAM Strasbourg-Europe se tiendra naturellement à disposition pour intervenir auprès du stagiaire ou de l'entreprise.

3/ ACCOMPAGNEMENT VERS L'EMPLOI

Un module d'Accompagnement vers l'Emploi est intégré à la formation. Il a pour objectif de préparer les stagiaires à leur intégration dans les entreprises de services du numérique.

4/ LES PREREQUIS DE CONNAISSANCES POUR INTEGRER LA FORMATION ET LE DISPOSITIF PROPOSE POUR VERIFIER QUE LE CANDIDAT LES POSSEDE

Les prérequis pour intégrer la formation sont des bases solides dans au moins 2 des domaines suivants :

- Culture scientifique et numérique
 - Statistiques
 - Logique
 - Culture informatique
- Outils informatiques
 - Algorithmique et programmation
 - Données

Les candidats seront évalués lors de leur sélection par le biais d'un test écrit par QCM sur chacun de ces domaines.

5/ LE(S) LIEU(X) DE LA FORMATION

Le lieu de formation est :

- ECAM Strasbourg-Europe, 2, Rue de Madrid, 67300 Schiltigheim