

Data Science : analyse et gestion de grandes masses de données

CATEGORIE : C

Vue d'ensemble

Domaine(s) d'activité professionnel dans lequel(s) est utilisé la certification :

- Spécifique : ■ **Support à l'entreprise - Systèmes d'information et de télécommunication**

Les compétences en Data Science peuvent être mises en œuvre dans de nombreux domaines d'application spécifiques et transverses, qui revêtent des enjeux économiques de taille. Leur valeur ajoutée n'est pas circonscrite à l'industrie informatique mais imprègne tous les secteurs économiques et l'ensemble des fonctions de l'entreprise, en particulier, les technologies et usages liés au Big data : (industrie, santé, énergie, transport, assurances, marketing, prédiction de modèles économiques, avantages concurrentiels liés à l'amélioration des systèmes d'information etc.)

(source : rapport «Industrial Internet Insights Report for 2015 (https://www.accenture.com/tw-en/_acnmedia/Accenture/next-gen/reassembling-industry/pdf/Accenture-Industrial-Internet-Changing-Competitive-Landscape-Industries.pdf) », General Electric / Accenture)

(source : Etude Deloitte Tendances Data et Analytics (http://www2.deloitte.com/fr/fr/pages/technology/articles/tendances-data-et-analytics-2016.html))

Code(s) NAF : —

Code(s) NSF : **326t**, **326n**, **326m**

Code(s) ROME : **M1803**, **M1805**, **M1806**, **M1810**

Formacode : **31054**

Date de création de la certification : **05/05/2015**

Mots clés : **Analyse de données**, **Big data**, **Data scientist**, **traitement de données massives**

Identification

Identifiant : **1918**

Version du : **23/05/2016**

Références

Consensus, reconnaissance ou recommandation :

Formalisé :

- [L'enjeu du plan Big Data est de créer 137 000 emplois](#)
- [Priorité gouvernementale, le Big Data est l'objet d'un des 34 plans de la Nouvelle France Industrielle](#)

Non formalisé :

- [Les meilleurs programmes pour se spécialiser en Big Data](#)
- [Devenir Data-Scientist : les formations et les MOOC Big Data](#)

Descriptif

Objectifs de l'habilitation/certification

La certification en data science vise à vérifier l'acquisition des compétences nécessaires à l'exploitation de grandes masses de données et la réalisation d'analyses prédictives de systèmes

Lien avec les certifications professionnelles ou les CQP enregistrés au RNCP

- aucun

Descriptif général des compétences constituant la certification

C.1 Proposer des axes de gestion et d'analyse de grandes masses de données

C2. Définir une architecture de traitement et une modélisation en déterminant les types de données, les outils statistiques et les méthodes appropriées

C.3 Sourcer, rassembler et nettoyer l'ensemble des sources de données

C.4 Utiliser des outils d'analyse et de gestion de bases de données de types variés dans de grands volumes en faisant preuve de réactivité et d'adaptation afin de surmonter les obstacles rencontrés

C.5 Synthétiser les résultats d'une analyse sous une ou des formes adaptées au besoin

Public visé par la certification

Tous publics

Modalités générales

Méthode Pédagogique et Organisation

La formation alterne cours théoriques et travaux pratiques. Elle a lieu en présentiel et à distance (FOAD).

Elle est d'une durée totale de 25 jours (175 heures en présentiel), dont 1 journée d'évaluation finale. Les jours de formation sont répartis par sessions de 2 jours toutes les deux ou trois semaines.

Entre chacune des 12 sessions de formation en présentiel, les participants doivent réaliser des travaux en ligne (nécessitant environ 3h de travail individuel) permettant de mettre en application les compétences acquises.

La formation est d'une durée totale de 12 mois (dont 3 mois dédiés à la réalisation d'un projet personnel tutoré présenté lors d'une soutenance finale). A titre exceptionnel, 2 mois supplémentaires peuvent être octroyés pour améliorer le projet personnel.

Actions pré-formatives

La formation est précédée d'une action pré-formative constituée :

d'une session de formation / évaluation type MOOC

d'un entretien de motivation et de positionnement

Ces actions pré-formatives peuvent donner lieu à des ajustements du contenu de la formation, notamment dans le cadre d'un dispositif de formation intra-entreprise.

Liens avec le développement durable

niveau 1 : Certifications et métiers qui internalisent le développement durable. Les activités et compétences mobilisées mettent en oeuvre des matériaux et produits moins polluants

Valeur ajoutée pour la mobilité professionnelle et l'emploi

Pour l'individu

Construction de la certification autour d'une double dimension : **informatique distribuée** d'une part, et **machine learning/apprentissage statistique** d'autre part

Volet projets et travaux pratiques garantissant le caractère opérationnel de la certification

Caractère scientifique pointu permettant aux certifiés de développer des capacités de maîtrise des outils nécessaires pour s'adapter à leurs

Pour l'entité utilisatrice

Emploi de professionnels opérationnels et capables de s'adapter aux évolutions d'un secteur hyperconcurrentiel

Légitimité, professionnalisme et partenariat d'une école d'ingénieurs de renom développant depuis plus d'un siècle la recherche et l'enseignement dans les domaines des mathématiques appliquées, informatique et réseaux. Ces travaux ont convergé

évolutions dans le temps

naturellement vers le big data et ont permis de proposer la première formation professionnalisante dans ce domaine en France il y a 3 ans.

Evaluation / certification

Pré-requis

Diplôme de niveau bac + 5 ou expérience professionnelle équivalente
Bonnes connaissances en mathématiques (optimisation, probabilités/statistique, algèbre linéaire)
Bonne expérience de la programmation
Validation de l'action pré-formative (MOOC)

Centre(s) de passage/certification

- Locaux de Telecom Paristech

Compétences évaluées

CE1. Prendre en compte les contraintes techniques pour proposer une solution d'analyse de grandes masses de données de types pouvant être variés

CE2a. Proposer une architecture de gestion et d'analyse appropriée aux besoins et aux contraintes d'un système cible.

CE2b. Combiner des outils et des méthodes adaptés et tenant compte des évolutions récentes des techniques.

CE3. Sourcer, rassembler et nettoyer l'ensemble des sources de données en utilisant les technologies adéquates afin de répondre aux besoins du système cible

CE4a. Utiliser et adapter des outils statistiques, de calcul distribué ainsi que des techniques de passage à l'échelle afin de procéder à l'analyse ou à la gestion des grandes masses de données

CE4b. Identifier les obstacles lors de l'implémentation d'un modèle d'analyse et de gestion des

grandes masses de données afin d'établir des préconisations d'amélioration

CE5. Synthétiser les résultats d'une l'analyse sous un ou des formes adaptées (rapport,

graphique, tableau de bord, indicateur...) afin de permettre leur exploitation par les utilisateurs, clients et/ou partenaires du système cible

Niveaux délivrés le cas échéant (hors nomenclature des niveaux de formation de 1969)

pas de niveau

La validité est Permanente

Possibilité de certification partielle : non

Matérialisation officielle de la certification :

Certificat d'Etudes Spécialisées de Telecom ParisTech

Plus d'informations

Statistiques

8 certifiés et 43 en cours de certification

Autres sources d'information

https://www.telecom-paristech.fr/fileadmin/documents/pdf/formation_continue/CES/Data-Scientist.pdf

<http://www.telecom-evolution.fr/fr/formations-certifiantes/ces-data-scientist>