

Data Science : savoir collecter, décrypter, analyser et prédire à partir de mégadonnées

—

Référentiels de compétences et de certification

<i>Intitulé de la certification</i>
<i>Data Science : savoir collecter, décrypter, analyser et prédire à partir de mégadonnées</i>
<i>Description du métier, de l'activité ou de la situation professionnelle à partir duquel le dispositif de formation visant la certification est initié</i>
<p>Le Data Scientist convertit une masse de données, d'origine diverse et le plus souvent non structurées, en enseignements exploitables visant à améliorer le processus de décision des managers. Ses analyses offrent un levier et un relai de croissance à l'organisation qui l'emploie. Le certificat « Data Science : savoir collecter, décrypter, analyser et prédire à partir de mégadonnées » est ouvert à une population diplômée de niveau bac +5 (avec une majeure en mathématiques ou informatique) et au moins 3 années d'expérience professionnelle. Il met à disposition des professionnels (ingénieurs, data analysts, informaticiens, chef de projets ...) un dispositif attestant des compétences nécessaires pour exercer le métier de Data Scientist.</p>

<i>Référentiel de compétences</i>	<i>Référentiel de certification</i>	
	<i>Modalités</i>	<i>Critères</i>
C1 : Analyser une problématique data science et en extraire les enjeux métier, informatique et mathématique afin de formaliser une étude	Réalisation d'un projet pendant les 6 mois que dure le cursus puis soutenance de ce projet	Cr1 : Le candidat a explicité la problématique et mis en évidence les enjeux algorithmiques à relever
C2 : Extraire des données depuis des bases de type SQL et NoSQL en utilisant des API (web scraping) et des techniques distribuées (Hadoop) afin de pouvoir mener une analyse	Examen de 4h sous forme d'une étude de cas, basée sur des données réelles, à implémenter sur ordinateur + Réalisation d'un projet pendant les 6 mois que dure le cursus puis soutenance de ce projet	Cr2 : Le candidat a extrait correctement des données de bases SQL ou NoSQL via des API, et a vérifié la pertinence de ses extractions
C3 : Mener une analyse exploratoire d'un jeu de données afin de synthétiser l'information sous-jacente en choisissant la méthode adaptée au problème (analyse factorielle ou classification non supervisée)	Examen de 4h sous forme d'une étude de cas, basée sur des données réelles, à implémenter sur ordinateur + Réalisation d'un projet pendant les 6 mois que dure le cursus puis soutenance de ce projet	Cr3 : Le candidat a choisi une méthode d'analyse des données (analyse factorielle ou clustering) appropriée aux données et à l'objectif visé. Il a su justifier ses choix et, les résultats de l'analyse sont cohérents

<p>C4 : Mettre en œuvre un modèle de Machine Learning pour obtenir des prévisions d'une variable d'intérêt à partir de facteurs explicatifs connus en utilisant un algorithme approprié au contexte (random forest, gradient boosting, SVM, réseaux de neurones...)</p>	<p>Examen de 4h sous forme d'une étude de cas, basée sur des données réelles, à implémenter sur ordinateur + Réalisation d'un projet pendant les 6 mois que dure le cursus puis soutenance de ce projet</p>	<p>Cr4 : Le candidat a comparé des méthodes de prévision à l'aide d'un critère d'erreur via une « validation croisée ». Il a su restituer les bénéfices/inconvénients des méthodes testées</p>
<p>C5 : Explorer et analyser des données non structurées (textes ou graphes) pour mettre en lumière des arrangements (sacs de mots ou communautés) à l'aide d'algorithmes de text mining ou de graph mining</p>	<p>Réalisation d'un projet pendant les 6 mois que dure le cursus puis soutenance de ce projet</p>	<p>Cr5 : Le candidat a su tirer de l'information signifiante de données non structurées (textes ou graphes) et l'a restituée d'une manière intelligible</p>
<p>C6 : Produire des synthèses visuelles afin de rendre accessibles aux décideurs des informations complexes issues de jeux de données en utilisant des outils de visualisation (dataviz)</p>	<p>Réalisation d'un projet pendant les 6 mois que dure le cursus puis soutenance de ce projet</p>	<p>Cr6 : Le candidat a créé une interface pour lancer les importations de données et restituer les analyses sous forme visuelle</p>