

# Référentiel d'activités, de compétences et d'évaluation

Certification « Data Science : analyse et gestion de grandes masses de données »

Référentiel Compétences	Référentiel d'évaluation	
	Modalités d'évaluation	Critères d'évaluation
Proposer des axes de gestion et d'analyse de grandes masses de données	<ul style="list-style-type: none"> <li>- QCM</li> <li>- Projet sur données réelles</li> </ul>	<ul style="list-style-type: none"> <li>- Capacité à proposer une modélisation de gestion et d'analyse appropriée aux besoins et aux contraintes d'un système cible.</li> <li>- Capacité à combiner des outils et méthodes adaptées et tenant compte des évolutions récentes des techniques.</li> <li>- Capacité à synthétiser les résultats d'une analyse sous une ou les formes adaptées (rapport graphique, tableau de bord, indicateurs...) afin de permettre leur exploitation par les utilisateurs, clients, et/ou partenaires du système cible.</li> </ul>
Prendre en compte les contraintes techniques pour proposer une solution d'analyse de grandes masses de données de types pouvant être variés	<ul style="list-style-type: none"> <li>- Rapport sur une problématique donnée en phase avec le module</li> <li>- Projet sur données réelles</li> <li>- QCM</li> </ul>	<ul style="list-style-type: none"> <li>- Capacité à prendre en compte les contraintes techniques pour proposer une solution d'analyse de grandes masses de données.</li> <li>- Capacité à sourcer, rassembler et nettoyer l'ensemble des sources de données en utilisant les technologies adéquates afin de répondre aux besoins du système cible.</li> </ul>
Définir une architecture de traitement et une modélisation en déterminant les types de données, les outils statistiques et les méthodes de machine learning appropriées	<ul style="list-style-type: none"> <li>-QCM</li> <li>- Rapport sur une problématique donnée en phase avec le module</li> <li>- Projet sur données réelles.</li> </ul>	<ul style="list-style-type: none"> <li>-Capacité à utiliser les principales techniques de réduction de dimension</li> <li>-Capacité à traiter la problématique d'ingénierie de variables (feature engineering)</li> <li>-Capacité à proposer une modélisation de gestion et d'analyse appropriée aux besoins et aux contraintes d'un système cible</li> </ul>

		<ul style="list-style-type: none"> <li>-Capacité à combiner des outils et méthodes adaptées et tenant compte des évolutions récentes des techniques</li> <li>-Capacité à évaluer les différentes méthodes de passage à l'échelle</li> <li>-Capacité à synthétiser les résultats d'une analyse sous une ou les formes adaptées (rapport graphique, tableau de bord)</li> </ul>
Proposer une architecture de gestion et d'analyse appropriée aux besoins et aux contraintes d'un système cible.	<ul style="list-style-type: none"> <li>- QCM</li> <li>- Projet sur données réelles</li> </ul>	<ul style="list-style-type: none"> <li>- Capacité à combiner des outils et méthodes adaptées et tenant compte des évolutions récentes des techniques pour proposer une architecture de gestion et d'analyse appropriée.</li> </ul>
Combiner des outils et des méthodes adaptés et tenant compte des évolutions récentes des techniques	<ul style="list-style-type: none"> <li>- QCM</li> <li>- Projet sur données réelles</li> </ul>	<ul style="list-style-type: none"> <li>- Capacité à utiliser et adapter les algorithmes de Boosting.</li> <li>- Capacité à traiter des données avec des classes déséquilibrées</li> <li>- Capacité à évaluer les différents algorithmes et méthodes manipulés durant la formation</li> </ul>
Sourcer, rassembler et nettoyer l'ensemble des sources de données en utilisant les technologies adéquates afin de répondre aux besoins du système cible	<ul style="list-style-type: none"> <li>- QCM</li> <li>- Projet sur données réelles</li> </ul>	<ul style="list-style-type: none"> <li>- Capacité à sourcer, rassembler et nettoyer l'ensemble des sources de données en utilisant les technologies adéquates afin de répondre aux besoins du système cible.</li> </ul>
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	<ul style="list-style-type: none"> <li>- QCM</li> <li>- Projet sur données réelles</li> </ul>	<ul style="list-style-type: none"> <li>- Capacité à 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.</li> <li>- Capacité à identifier les obstacles lors de l'implémentation d'un modèle d'analyse et de gestion des masses de données afin</li> </ul>

		d'établir des préconisations d'amélioration.
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	<ul style="list-style-type: none"> <li>- QCM</li> <li>- Projet sur données réelles</li> </ul>	<ul style="list-style-type: none"> <li>- Capacité à 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.</li> </ul>
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	<ul style="list-style-type: none"> <li>- QCM</li> <li>- Projet sur données réelles</li> </ul>	<ul style="list-style-type: none"> <li>- Capacité à identifier les obstacles lors de l'implémentation d'un modèle d'analyse et de gestion des masses de données afin d'établir des préconisations d'amélioration.</li> </ul>
Synthétiser les résultats d'une analyse sous une ou des formes adaptées au besoin (rapport, graphique, tableau de bord, indicateur...) afin de permettre leur exploitation par les utilisateurs, clients et/ou partenaires du système cible	<ul style="list-style-type: none"> <li>- QCM</li> <li>- Projet sur données réelles</li> </ul>	<ul style="list-style-type: none"> <li>- Capacité à synthétiser les résultats d'une analyse sous une ou les formes adaptées (rapport graphique, tableau de bord, indicateurs...) afin de permettre leur exploitation par les utilisateurs, clients, et/ou partenaires du système cible.</li> </ul>
	<ul style="list-style-type: none"> <li>• <b>Projet de certification</b> Il est demandé au candidat dans le cadre d'un projet d'entreprise de réaliser un mémoire guidé par un tuteur. L'équipe pédagogique s'assure en amont que le contenu du mémoire intègre un maximum de compétences visées par la certification. Une soutenance orale devant un jury composé de formateurs et de professionnels d'entreprises a lieu. Les compétences sont évaluées selon un modèle de fiche de soutenance.</li> </ul>	<p>La certification finale vise à évaluer la capacité de chaque candidat à :</p> <ul style="list-style-type: none"> <li>-Gérer de grandes masses de données</li> <li>- Analyser de grandes masses de données</li> <li>- Analyser des données complexes</li> <li>- Développer un modèle d'analyse prédictive par l'apprentissage automatique ou machine learning</li> <li>- Utiliser les technologies de stockage distribué</li> <li>- Utiliser des technologies de calcul distribué</li> </ul>