

## ELEMENTS COMPLEMENTAIRES RELATIFS A LA DEMANDE

### 5 - REFERENTIELS

#### Article L6113-1 En savoir plus sur cet article... Créé par LOI n°2018-771 du 5 septembre 2018 - art. 31 (V)

« Les certifications professionnelles enregistrées au répertoire national des certifications professionnelles permettent une validation des compétences et des connaissances acquises nécessaires à l'exercice d'activités professionnelles. Elles sont définies notamment par un **référentiel d'activités** qui décrit les situations de travail et les activités exercées, les métiers ou emplois visés, un **référentiel de compétences** qui identifie les compétences et les connaissances, y compris transversales, qui en découlent et un **référentiel d'évaluation** qui définit les critères et les modalités d'évaluation des acquis. »

#### Bloc de compétences 1 : Modéliser les données massives d'une structure

REFERENTIEL D'ACTIVITES <i>décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	REFERENTIEL DE COMPETENCES <i>identifie les compétences et les connaissances, y compris transversales, qui découlent du référentiel d'activités</i>	REFERENTIEL D'ÉVALUATION <i>définit les critères et les modalités d'évaluation des acquis</i>	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION
<p>Activité 1.</p> <p><b>Dans le cadre de la modélisation des données massives hétérogènes ou non d'une entreprise, l'expert en sciences des données procède à l'analyse par :</b></p> <ul style="list-style-type: none"> <li>-L'extraction des données disponibles.</li> <li>-L'identification du type de données : structurées ou non structurées.</li> <li>-La sélection d'une technique d'approche et d'analyse adaptée.</li> </ul>	<p>C1.1. Maîtriser les outils d'analyse pour décrire les données en tenant compte du contexte.</p> <p>C1.2. Identifier l'outil de modélisation statistique en fonction du type de donnée, structurée ou non structurée, afin de la traiter.</p> <p>C1.3. Effectuer l'analyse exploratoire et la description des masses de données au travers d'une méthode statistique pour déterminer les problématiques de description, de projection, de visualisation et de regroupement des données en catégories homogènes.</p>	<p><b>E1 - Mise en situation professionnelle reconstituée d'analyse de données publiquement disponibles :</b></p> <ul style="list-style-type: none"> <li>-Chaque candidat réalise un support (code, rapport) de son modèle d'analyse exploratoire et de représentation des données.</li> </ul> <p><b>E2 - Thèse professionnelle</b></p> <p><b>E3 - Projet en groupe</b></p>	<p>Dans le cadre de l'analyse, de la thèse et du projet (<b>E1, E2, E3</b>), le (la) candidat(e) doit satisfaire les critères suivants :</p> <ul style="list-style-type: none"> <li>-La méthodologie d'analyse et de traitement employée est décrite et cohérente eu égard du modèle sélectionné. (<b>C1.1, C1.2, C1.3, C2.1</b>)</li> <li>-Les résultats du modèle sont codés et interprétés. (<b>C2.2, C2.3</b>)</li> <li>-Les performances du prototype sont cohérentes avec la réalité observée et l'état de l'art. (<b>C1.3, C2.3</b>)</li> </ul>

## ELEMENTS COMPLEMENTAIRES RELATIFS A LA DEMANDE

<p><i>Activité 2.</i>  <b>Afin de modéliser les données massives hétérogènes de l'entreprise, l'expert en sciences des données élabore un modèle prédictif de traitement de données :</b></p> <ul style="list-style-type: none"> <li>-Identification et formulation du problème technique.</li> <li>-Résolution, mise en œuvre du modèle sur les données.</li> <li>-Identification des problèmes d'optimisation des données massives.</li> <li>-Analyse des performances des résultats obtenus.</li> <li>-Exploitation des données non structurées venant des flux vidéo ou audio et des réseaux sociaux.</li> <li>-Exploitation des données structurées.</li> </ul>	<p>C2.1. Sélectionner une méthodologie de traitement statistique au regard du problème posé, du type de donnée pour la modéliser.</p> <p>C2.2. Concevoir et réaliser un prototype du modèle à partir de logiciels de programmation haut niveau pour montrer la faisabilité et la pertinence du modèle.</p> <p>C2.3. Evaluer la performance du modèle au regard des contraintes de mise en production pour le mettre en œuvre.</p>	<p><b>E4 : QCM de 10 questions sur les concepts théoriques et pratiques.</b>          Exemple de thématiques :          -Modèle SVM et Bayes          -Modèle d'apprentissage supervisé vs non supervisé          -Modèle convolutif et récurrent (IA)          -Optimisation</p>	<p>-Le comportement du modèle est caractérisé en termes de performances de prédiction. <b>(C2,3)</b></p> <p>Pour le QCM <b>(E4)</b>, le (la) candidat(e) doit satisfaire les critères suivants :          -Les concepts théoriques et pratiques sont maîtrisés (la moyenne au QCM est atteinte).  <b>(C1.1, C1.2, C2.1, C2.2)</b></p>
--	---	---	---

## ELEMENTS COMPLEMENTAIRES RELATIFS A LA DEMANDE

### Bloc de compétences 2 : Collecter, structurer et traiter les données d'une structure

REFERENTIEL D'ACTIVITES <i>décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	REFERENTIEL DE COMPETENCES <i>identifie les compétences et les connaissances, y compris transversales, qui découlent du référentiel d'activités</i>	REFERENTIEL D'ÉVALUATION <i>définit les critères et les modalités d'évaluation des acquis</i>	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION
<p><b>Activité 1.</b> <b>Afin de traiter les données, l'expert (e) en sciences des données réalise l'extraction et la collecte des données via une technique d'analyse et de stockage pour chaque donnée :</b></p> <ul style="list-style-type: none"> <li>-Extraction et analyse en ligne et hors ligne des données.</li> <li>-Identification du type de données hétérogènes : structurées ou non structurées.</li> <li>-Sélection d'une technique d'analyse et de stockage pour chaque donnée.</li> </ul>	<p>C1.1. Maîtriser et utiliser les outils d'interrogation des bases de données pour extraire et collecter les données.</p> <p>C1.2. Concevoir un système d'ingénierie de données en définissant une approche méthodologique de collecte, de stockage et d'interrogation pour l'analyse de masse de données réelles.</p>	<p><b>E5 : Mise en situation professionnelle reconstituée sur le stockage de la donnée :</b></p> <ul style="list-style-type: none"> <li>-Chaque candidat réalise la conception, la configuration et l'implémentation de systèmes de stockage de la donnée via des outils d'interrogation des bases de données et à partir de données publiquement disponibles.</li> </ul>	<p><u>Le (la) candidat(e) doit satisfaire les critères suivants :</u></p> <ul style="list-style-type: none"> <li>-Les différents types de données sont distingués.</li> <li>-La technique d'analyse et de stockage des données sélectionnée est pertinente au regard du type de donnée.</li> </ul>
<p><b>Activité 2.</b> <b>L'expert (e) en sciences des données réalise le traitement et l'exploitation des données collectées par :</b></p> <ul style="list-style-type: none"> <li>-L'analyse des données massives collectées.</li> <li>-La sélection des variables explicatives les plus pertinentes en vue de la modélisation.</li> <li>-La modélisation de la donnée, programmation.</li> </ul>	<p>C2.1. Elaborer et mettre en place une stratégie de traitement distribué en ligne ou hors ligne de la donnée structurée et non structurée par un langage de programmation fonctionnel afin de les exploiter.</p> <p>C2.2. Développer des implémentations au regard des données à traiter pour cadrer les algorithmes et performances prédictives.</p>	<p><b>E6 : Mise en situation professionnelle reconstituée portant sur l'architecture de traitement distribuée de données :</b></p> <p>Chaque candidat conçoit et réalise une architecture de traitement distribuée des données à partir de données publiquement disponibles.</p>	<p><u>Le (la) candidat(e) doit satisfaire les critères suivants :</u></p> <ul style="list-style-type: none"> <li>-La donnée est interrogée et agrégée dans le modèle via des langages haut niveau.</li> <li>-Les variables explicatives les plus pertinentes sont sélectionnées.</li> <li>-L'implémentation de modèle de traitement statistique distribué</li> </ul>

## ELEMENTS COMPLEMENTAIRES RELATIFS A LA DEMANDE

<p>-La formulation et l'exécution des requêtes.</p> <p>-L'exploitation des données structurées ou non structurées venant des flux.</p> <p>-La visualisation de la donnée.</p>			<p>est réalisé sur les données traitées.</p> <p>-Les requêtes sont formulées et exécutées.</p>
<p><i>Activité 3.</i></p> <p><b>L'expert (e) en science des données veille à l'amélioration continue du traitement des données structurées ou non structurées via :</b></p> <ul style="list-style-type: none"> <li>-L'analyse des retours, des bases de données de la structure.</li> <li>-L'identification des problèmes d'optimisation des données massives.</li> <li>-Le traitement des données comportant des imperfections.</li> <li>-L'analyse des différents types de problème liés aux retours : apprentissage supervisé ou non supervisé, batch ou online</li> </ul>	<p>C3.1. Analyser les retours pour mettre à jour l'exploitation et le stockage des données.</p> <p>C3.2. Repérer le type de problème rencontré, d'apprentissage ou de fouille de données, en utilisant des algorithmes d'optimisation pour le résoudre.</p>	<p><b>E2 - Thèse professionnelle</b></p> <p><b>E3 - Projet en groupe</b></p>	<p>Les critères d'évaluation du rapport écrit et de la soutenance pour la thèse professionnelle et le projet de groupe sont présentés dans le tableau de synthèse des modalités d'évaluation.</p>

## ELEMENTS COMPLEMENTAIRES RELATIFS A LA DEMANDE

### Bloc de compétences 3 : Concevoir une stratégie de valorisation de la donnée

REFERENTIEL D'ACTIVITES <i>décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	REFERENTIEL DE COMPETENCES <i>identifie les compétences et les connaissances, y compris transversales, qui découlent du référentiel d'activités</i>	REFERENTIEL D'ÉVALUATION <i>définit les critères et les modalités d'évaluation des acquis</i>	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION
<p><b>Activité 1.</b> Afin d'exploiter les données hétérogènes de la structure, l'expert en sciences des données élabore une stratégie d'exploitation et de valorisation de la donnée :</p> <ul style="list-style-type: none"> <li>-Identification des sources de données pertinentes.</li> <li>-Evaluation des données à traiter.</li> <li>-Prise en compte de l'écosystème économique et réglementaire lié au traitement des données massives.</li> <li>-Réalisation d'analyses prédictives.</li> </ul> <p><b>Activité 2.</b> L'expert en sciences des données organise et planifie des projets de la structure, de manière collaborative avec les différents services de la structure concernés par la donnée. Cette situation de travail comprend :</p>	<p>C1.1. Maîtriser les fonctionnalités d'une architecture informatique robuste pour le stockage et le traitement distribué des données.</p> <p>C1.2. Traduire un problème business en problème mathématiques/statistiques afin de définir une stratégie de valorisation des données pour exploiter les informations décisionnelles de la structure.</p> <p>C1.3. Concevoir une stratégie de valorisation des données en proposant des « entrepôts de données » tenant compte de l'écosystème économique, réglementaire et des contraintes de sécurité liées au traitement des données.</p> <p>C.2.1 Formuler et présenter des analyses statistiques décisionnelles à viser des différents services de la structure pour les accompagner dans leur prise de décision.</p> <p>C.2.2 Communiquer sur le potentiel des outils innovants de représentation, de modélisation et de visualisation</p>	<p><b>E7 - Mise en situation professionnelle réelle sur la réglementation, la sécurisation des données :</b> Chaque candidat(e) traite l'aspect réglementaire sur une problématique de l'entreprise dans laquelle il (elle) réalise son alternance.</p> <p><b>E2 - Thèse professionnelle</b></p> <p><b>E3 - Projet en groupe</b></p>	<p>Lors de la mise en situation (E7), le (la) candidat(e) est évalué(e) sur :</p> <ul style="list-style-type: none"> <li>-La problématique de sécurisation des données est formulée eu égard du contexte de l'entreprise (C1.1).</li> <li>-Les dispositions réglementaires (RGPD, innovation, PI) sont identifiées et prises en compte eu égard de l'environnement, de l'écosystème de l'entreprise (C1.3).</li> <li>-Les vulnérabilités, menaces, attaques potentielles sur les données de l'entreprise sont repérées (C1.1, C1.3, C2.1, C2.4).</li> <li>-Des outils de protection des données sont proposés au regard des vulnérabilités repérées et du contexte réglementaire de l'entreprise (C1.1).</li> </ul> <p>Les critères d'évaluation du rapport écrit et de la soutenance</p>

## ELEMENTS COMPLEMENTAIRES RELATIFS A LA DEMANDE

<ul style="list-style-type: none"> <li>-L'identification des solutions et technologies existantes.</li> <li>-L'analyse de la performance et des retours des outils existants.</li> <li>-La création d'outils d'aide à la décision ou pour un besoin exprimé par un service.</li> <li>-La participation à l'innovation et l'évolution de la structure.</li> <li>-La sensibilisation des utilisateurs aux outils.</li> <li>-L'élaboration de contenus de présentation du projet.</li> <li>-L'animation de réunion d'équipe.</li> <li>-La planification et la budgétisation du projet.</li> <li>-L'évaluation des ressources nécessaires au projet.</li> </ul>	<p>pour contribuer à la plus-value et la prise de décisions utiles à la structure.</p> <p>C.2.3 Dimensionner les actions du projet en définissant les choix techniques et méthodologiques des interventions pour évaluer et planifier les ressources techniques, budgétaires et humaines.</p> <p>C.2.4 Communiquer sur la stratégie de valorisation des données de la structure pour sensibiliser les acteurs, accompagner au changement et améliorer le fonctionnement interne.</p>	<p>pour la thèse professionnelle (E2) et le projet de groupe (E3) sont présentés dans le tableau de synthèse des modalités d'évaluation. Ici on s'attardera sur :</p> <ul style="list-style-type: none"> <li>- Le recueil des besoins de l'entreprise (C1.2)</li> <li>- Les contraintes techniques et économiques (C1.1, C.2.3)</li> <li>- Les contraintes réglementaires (C1.3.)</li> <li>- La conception d'une solution (C1.2, C.2.1, C.2.4)</li> <li>- Sa réalisation et sa valorisation (C1.2, C1.3., C.2.1, C.2.2, C.2.4)</li> </ul>
---	--	---

## ELEMENTS COMPLEMENTAIRES RELATIFS A LA DEMANDE

Tableau de synthèse des modalités d'évaluation :

<b>Modalités d'évaluation</b>	<b>Description de la modalité</b>
E1 - Mise en situation professionnelle reconstituée d'analyse de données	Chaque candidat réalise un support (code, rapport) d'un modèle d'analyse exploratoire et de représentation des données. L'analyse de données porte sur des données publiquement disponibles.
E2 - Thèse professionnelle	<p>La thèse professionnelle est à réaliser par chaque candidat, de manière individuelle. Elle porte sur une problématique liée à leur entreprise d'alternance. La problématique est validée en amont par le tuteur entreprise et les responsables du programme.</p> <p>Chaque candidat(e) doit rendre un rapport écrit, mémoire et effectuer une soutenance orale devant le jury de certification.</p> <p>Le mémoire doit comprendre a minima 30 pages sur le sujet traité en précisant le contexte, les problématiques du sujet, les enjeux et intérêts de ces problématiques et de relier si possible ces problèmes à des thématiques du programme. La démarche utilisée pour traiter le problème est présentée (recherches bibliographiques, développement des éléments théoriques, conception de solutions, étapes de la procédure, démarche expérimentale menée, résultats obtenus, analyse critique, etc.).</p> <p>La soutenance a pour but de vérifier les capacités de synthèse, de prise de parole en public, d'argumentation, de recul critique. Elle comprend deux parties : un exposé oral, d'une durée de 30 minutes, suivi de questions et discussions, d'une durée de 15 à 20 minutes. L'exposé</p>

## ELEMENTS COMPLEMENTAIRES RELATIFS A LA DEMANDE

	doit permettre de présenter de manière synthétique les points clés de la thèse : la problématique, les enjeux, les approches de solution, les résultats et leur analyse.
E3 - Projet de conception	<p>Les candidats réalisent, par groupe de 2-3 candidats, un rapport écrit sur un projet de conception au choix.</p> <p>Chaque groupe présente son projet au jury de certification durant une soutenance orale. Chaque groupe présente les données traitées, le prototype fonctionnel de l'architecture de traitement des données fonctionnelles, la méthodologie d'évaluation employée, les résultats obtenus, la répartition des tâches entre les membres de l'équipe projet, la gestion du projet (planification, versionning du code) ...</p>
E4 : QCM	<p>Le QCM comprend 10 questions portant sur les concepts théoriques et pratiques.</p> <p>Exemples de thématiques :</p> <ul style="list-style-type: none"> <li>-Modèle SVM et Bayes</li> <li>-Modèle d'apprentissage supervisé vs non supervisé</li> <li>-Modèle convolutif et récurrent (IA)</li> <li>-Optimisation</li> </ul>
E5 : Mise en situation professionnelle reconstituée sur un système de stockage de la donnée.	<p>Chaque candidat réalise la conception, la configuration et l'implémentation de systèmes de stockage de la donnée via des outils d'interrogation des bases de données.</p> <p>L'évaluation porte sur des données publiquement disponibles.</p>



## ELEMENTS COMPLEMENTAIRES RELATIFS A LA DEMANDE

E6 : Mise en situation professionnelle reconstituée portant sur l'architecture de traitement distribuée de donnée.	Chaque candidat conçoit et réalise une architecture de traitement distribuée des données. Cette évaluation porte sur des données publiquement disponibles.
E7 : Mise en situation professionnelle réelle sur la réglementation, la sécurisation des données.	Pour évaluer l'aspect réglementaire, chaque candidat(e) traite une problématique réglementaire, de sécurisation des données de l'entreprise dans laquelle il (elle) réalise son alternance.

Le cas échéant, description de tout autre document constitutif de la certification professionnelle

La certification est délivrée à la suite de la validation de l'ensemble des blocs de compétences. Un certificat de compétences sera remis à l'issu de l'obtention du bloc de compétences.