

**Référentiel d'activité, de compétences et d'évaluation**  
**Diplôme d'ingénieur diplômé par le Conservatoire national des arts et métiers,**  
**Spécialité Sciences de la donnée et IA**

Référentiel d'activités	Référentiel de compétences	Modalités d'évaluation
<ul style="list-style-type: none"> <li>• Identification et analyse des besoins d'un client en matière de collecte et exploitation des données et d'intelligence artificielle d'une organisation, d'une entreprise industrielle et de service.</li> <li>• Élaboration d'un plan d'action concerté en équipe intégrant une planification de conseil et traitement des besoins visant l'optimisation de la qualité des données et de l'usage de l'intelligence artificielle.</li> <li>• Présentation et négociation avec le client des solutions les plus adaptées via l'utilisation des sciences de la donnée et l'intelligence artificielle au traitement de ses données.</li> <li>• Conceptualisation de la réponse aux besoins d'un client en matière d'outils informatiques pour la gestion de ses données massives.</li> <li>• Modélisation de l'architecture informatique et logicielle pour le traitement des données massives.</li> <li>• Élaboration de l'architecture informatique et logicielle du traitement des données massives et leur intégration au système informatique du client.</li> <li>• Nettoyage et pré-traitement des données (qualité, complétude, etc.) en vue de les rendre exploitables dans le cadre d'un projet de stockage et d'analyse des bases de données et d'intelligence artificielle.</li> <li>• Conception et validation des algorithmes d'apprentissage statistique et automatique.</li> <li>• Analyse des données à l'aide des algorithmes d'apprentissage statistique et automatique conçu.</li> </ul>	<ul style="list-style-type: none"> <li>• Identifier la demande du client en termes de budget, temps, contraintes logicielles et humaines, utilisateurs finaux, technologies à utiliser et la modéliser sous forme d'un cahier des charges précis.</li> <li>• Valider le cahier des charges produit avec le client et corriger et mettre en adéquation avec des nouvelles demandes du client.</li> <li>• Identifier les risques internes ou externes du projet pouvant impacter de façon significative et les classer à l'aide d'une matrice des risques en prenant en compte la gravité, l'occurrence et la détectabilité.</li> <li>• Définir un plan d'action pour chaque risque afin d'en réduire la criticité.</li> <li>• Développer la relation client en instaurant une relation de confiance via des contacts réguliers (réunion, tests, retour d'expérience, validation) entre le client et le chargé du projet et en intégrant le client dans l'évolution du projet permettant de valider ses demandes et leur évolution.</li> <li>• Prévoir avec le client une possible extension du projet en identifier des perspectives futures.</li> <li>• Parfaire la relation client par les contacts réguliers entre le client et le chargé et les responsables des corps de métiers aux différents moments du planning d'intervention.</li> <li>• Choisir la méthode de gestion de projet (méthode en V, méthode itérative, méthode agile, combinaison de méthodes) adaptée au projet et à l'équipe en place.</li> <li>• Déployer les outils de gestion de projet (planning, outils de communication, gestion des tâches, évaluation) en lien avec la méthode choisie.</li> <li>• Adapter et appliquer sa gestion de projet (SCRUM, agile, etc.) au sein d'une équipe pluridisciplinaire de développeurs, architectes de bases de données, de statisticiens, gestionnaires de données, etc., en prenant en compte l'hétérogénéité des niveaux d'expertise.</li> </ul>	<p><b>Évaluation des séquences académiques</b></p> <ul style="list-style-type: none"> <li>• Mises en situations, analyse d'articles, études de cas, examens sur table et questionnaires de vérification du savoir.</li> <li>• Travaux et projets individuels ou en groupe.</li> </ul> <p><b>Évaluation des séquences entreprises</b></p> <ul style="list-style-type: none"> <li>• Évaluation des rapports et travaux conduits en entreprise : présentation des projets conduits en entreprise</li> <li>• Évaluation par le maître d'apprentissage des savoir-faire et savoir être tout au long de la formation.</li> </ul> <p><b>Critères d'évaluation</b></p> <ul style="list-style-type: none"> <li>• Rigueur des analyses techniques proposées.</li> <li>• Analyse critique de solutions techniques existantes.</li> <li>• Qualité de la formalisation écrite et orale en français ou en anglais (fond scientifique et technique, forme des documents).</li> </ul>

<ul style="list-style-type: none"> <li>• Interprétation des résultats de l'analyse et tirage des conclusions utiles pour aider à la décision.</li> <li>• Composition de l'équipe projet en fonction des compétences internes disponibles dans l'entreprise industrielle et de service.</li> <li>• Organisation et planification des activités des membres de l'équipe projet en fonction de leurs compétences et dans le respect de la réglementation HSE et RSE.</li> <li>• Inscription dans une démarche d'amélioration continue l'équipe projet et ses performances en confrontant les travaux à une veille prospective métier (évolution technologique, méthodologique, sectorielle...) et en élaborant le plan de formation de ses membres.</li> </ul>	<ul style="list-style-type: none"> <li>• Adapter le projet en fonction du contexte international du client et de sa demande et se renseigner sur les différences de normes, législations, utilisations par les utilisateurs finaux, types de données, etc. pouvant apparaître dans des pays étrangers.</li> <li>• Comprendre la problématique du client en analysant l'environnement dans lequel doit être développé le projet de données massives : type, taille, fréquence d'acquisition des données.</li> <li>• Retranscrire en termes techniques adaptés les besoins exprimés en interagissant avec les experts métier : analyse des outils existants, architecture de communication des données, architecture d'acquisition des données.</li> <li>• Proposer au client une description schématisée et modélisée précise de sa problématique de traitement des données massives. Intégrer à celle-ci et schématiser dans l'environnement requis les points faisant intervenir les contraintes techniques, de qualité, de sécurité, environnementales, sociétales et réglementaires.</li> <li>• Faire intervenir les responsables de chaque spécificité du réseau informatique, des systèmes informatiques et de développeurs logiciels pour l'élaboration du projet.</li> <li>• Construire le cahier des charges en respectant une forme d'optimum de Pareto pour les spécifications et contraintes du S.</li> <li>• Élaborer le schéma directeur du SI en évaluant les risques / opportunités en utilisant une matrice SWOT et en intégrant un cycle de formation continue.</li> <li>• Proposer au client une description schématisée et modélisée précise et aboutie de sa problématique de traitement des données massives. Intégrer à celle-ci et schématiser dans l'environnement requis les points faisant intervenir les contraintes techniques, de qualité, de (cyber)sécurité, environnementales, sociétales, réglementaires et d'optimisation des données.</li> <li>• Faire l'inventaire des différentes sources de données potentielles, qu'elles soient internes ou externes (sites web, réseaux sociaux, capteurs IoT, etc.) et identifier la nature des données (tableurs, fichiers de texte, images, etc.).</li> <li>• Mettre en place une infrastructure adaptée pour la collecte des données ; cela peut impliquer l'utilisation de technologies telles que des API, des connecteurs de base de données, des scripts d'extraction, etc.</li> <li>• Utiliser des systèmes de gestion des données pour stocker et organiser les données collectées</li> <li>• Choisir le système le plus adapté (Bases de données relationnelles, Bases de données NoSQL, Entrepôts de données, Cloud, systèmes de stockage</li> </ul>	<ul style="list-style-type: none"> <li>• Pertinence des outils et des méthodologies utilisés.</li> <li>• Adéquation entre données collectées et projet.</li> <li>• Qualité de synthèse et de communication des principaux résultats.</li> </ul>
---	---	---

	<p>distribué) selon le volume et la nature des données (structurées ou non structurées).</p> <ul style="list-style-type: none"><li>• Identifier et traiter les valeurs manquantes dans les données et les valeurs aberrantes, détecter et gérer les doublons et les erreurs de saisie dans les données.</li><li>• Traiter les variables catégorielles en les encodant de manière appropriée, supprimez les variables ou les colonnes qui ne sont pas pertinentes pour l'analyse ou qui contiennent des informations redondantes.</li><li>• Décrire les données par des statistiques descriptives univariées, bivariées et multivariées.</li><li>• Visualiser les données par des graphiques appropriés.</li><li>• Utiliser de techniques de fouille de données pour réduire la dimensionalité de données (réduction des variables et segmentation des observations) afin d'identifier des relations entre les variables et les similarités parmi les observations</li><li>• Utiliser des techniques de modélisation et apprentissage statistique pour expliquer et prédire les valeurs d'une variable ou bien les classes d'appartenance des observations à partir d'un ensemble de prédicteurs.</li><li>• Analyser les résultats obtenus à partir des modèles pour tirer des conclusions sur les relations entre les variables et la capacité prédictive du modèle.</li><li>• Sélectionner le modèle prédictif approprié en fonction de la nature des données et des objectifs de l'analyse. Entraîner le modèle sur l'échantillon d'apprentissage. Optimiser le modèle en ajustant les hyperparamètres et/ou en essayant différentes architectures.</li><li>• Déployer le modèle dans le système numérique de destination.</li><li>• Choisir des mesures appropriées pour évaluer les performances des algorithmes d'IA, en fonction de la nature de la tâche (classification, régression, clustering, etc.).</li><li>• Évaluer la compatibilité du temps d'exécution et de la consommation de ressources avec les besoins spécifiques de l'utilisateur.</li><li>• Exploiter les retours des utilisateurs finaux pour comprendre comment les modèles fonctionnent dans des conditions réelles et pour identifier les domaines d'amélioration potentiels. Surveiller les évolutions des données au fil du temps.</li><li>• Planifier des périodes de réentraînement régulières des modèles en utilisant des données récentes.</li></ul>	
--	--	--

	<ul style="list-style-type: none"><li>• Expérimenter de nouvelles architectures de modèles et techniques d'entraînement issues des dernières avancées en matière de modélisation et d'apprentissage automatique.</li><li>• Identifier et mobiliser les ressources humaines pluridisciplinaires internes et externes, en vue de constituer l'équipe projet.</li><li>• Manager l'équipe projet en tenant compte des enjeux de la relation au travail, en tenant compte des responsabilités éthiques et professionnelles.</li><li>• Inscrire dans une démarche d'amélioration continue les équipes métier en informatique et mathématiques, en confrontant les travaux à une veille prospective métier (évolution technologique, méthodologique, sectorielle).</li><li>• Piloter les activités de recherche développement, mettre en place et évaluer des dispositifs expérimentaux dans un cadre collaboratif et pluridisciplinaire.</li><li>• Organiser et planifier les activités des membres de l'équipe projet en fonction de leurs compétences</li><li>• Définir des indicateurs de suivi des actions et d'atteintes des résultats à chaque étape du projet et si besoin réajuster les moyens mis en œuvre.</li><li>• Proposer des méthodes appropriées pour une analyse collaborative de ces indicateurs.</li><li>• Prendre en compte les enjeux des relations au travail, de sécurité et de santé au travail et de la diversité.</li><li>• Déployer des outils métiers de mesure de performances et d'évaluation des équipes.</li><li>• Participer à la définition du plan de formation de l'entreprise afin d'ajuster les compétences avec les activités et tâches.</li></ul>	
--	---	--