

RÉFÉRENTIEL

[EXPERT EN DATA SCIENCE]

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. »

Évaluation des candidats en situation de handicap

Avec l'accord du candidat, l'identification d'une situation de handicap peut être déclarée dans le dossier académique destiné au jury final qui le prendra en compte dans son évaluation. Tout candidat en situation de handicap peut également bénéficier d'un aménagement de ses modalités d'évaluation au cours de sa formation (notamment un tiers-temps pour les soutenances).

Conditions d'octroi de la certification

L'octroi de la certification professionnelle de **Expert en data science** d'OpenClassrooms est conditionnée à la validation des 4 blocs de compétences. Les blocs de compétences sont les suivants :

- **Bloc 1** – Collecter, pré-traiter, analyser des données et créer des variables pertinentes (*feature engineering*) pour l'entraînement d'un modèle
- **Bloc 2** – Concevoir des modèles d'apprentissage automatique à partir de données structurées et de données non structurées (texte et images)
- **Bloc 3** – Gérer le cycle de vie et le déploiement en production d'un modèle (MLOps) ou d'une application dans des environnements classiques et Big Data
- **Bloc 4** – Manager des projets IA et de data science complexes en intégrant les contraintes légales et les valeurs d'éthique

RÉFÉRENTIEL D'ACTIVITÉS <i>décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	RÉFÉRENTIEL DE COMPÉTENCES <i>identifie les compétences et les connaissances, y compris transversales, qui découlent du référentiel d'activités</i>	RÉFÉRENTIEL D'ÉVALUATION <i>définit les critères et les modalités d'évaluation des acquis</i>	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION
Bloc de compétences 1 – Collecter, pré-traiter et analyser des données et créer des variables pertinentes (<i>feature engineering</i>) pour l'entraînement d'un modèle			
A1.1 – Définition de la stratégie de collecte de données et réalisation d'une collecte <i>via</i> une API	C1.1 – Définir la stratégie de collecte de données en recensant les API disponibles et réaliser la collecte des données répondant à des critères définis <i>via</i> une API (interface de programmation), en prenant en compte les normes RGPD, afin de les exploiter pour l'élaboration d'un modèle.	<p>L'évaluation de la compétence est réalisée au moyen d'une mise en situation professionnelle reconstituée.</p> <p>Le candidat doit définir sa stratégie de collecte de données, réaliser le travail de collecte ciblée de données en exécutant une requête à une API et stocker les informations pour réutilisation ultérieure.</p> <p>Le travail est présenté lors d'une soutenance à un évaluateur (expert technique). L'évaluateur posera des questions au candidat pour vérifier l'acquisition des compétences sur sa stratégie de collecte des données et les opérations de collecte de données <i>via</i> une API.</p>	<p>CE1 Le candidat a défini sa stratégie de collecte de données et recensé les API disponibles pour le besoin de son projet</p> <p>CE2 Le candidat a écrit et testé une requête pour obtenir les données <i>via</i> l'API.</p> <p>CE3 Le candidat a récupéré les seuls champs nécessaires.</p> <p>CE4 Le candidat applique au moins un filtre sur un des champs nécessaires pour ne collecter que les données ayant les valeurs correspondantes sur ce ou ces champ(s).</p> <p>CE5 Le candidat a stocké les données collectées <i>via</i> l'API dans un fichier utilisable (ex. : fichier CSV ou pickle).</p> <p>CE6 Le candidat veille au respect des normes RGPD dans toutes phases de la collecte et du stockage des données.</p> <p>CE7 La présentation orale est claire et synthétique.</p>
A1.2 - Définition et gestion du nettoyage de données structurées	C1.2.1 Déterminer les objectifs du nettoyage des données et de la démarche de mise en œuvre afin de construire un jeu de	<p>L'évaluation de la compétence est réalisée au moyen d'une mise en situation professionnelle reconstituée.</p> <p>Le candidat doit répondre à un appel à projet qui inclut un jeu de données, en proposant une idée d'application et en détaillant les</p>	<p>CE1 Le candidat a défini les objectifs de son nettoyage de données en fonction de la problématique métier</p> <p>CE2 Le candidat a défini sa démarche de préparation et nettoyage des données</p> <p>CE3 La présentation orale est claire et synthétique.</p>

	données adapté à une problématique métier.	objectifs et les moyens pour nettoyer le jeu de données adapté au contexte de l'application. Le candidat doit simplement présenter et défendre son idée d'application, qui serait réalisée ultérieurement. Le travail est présenté lors d'une soutenance à un évaluateur (expert technique). L'évaluateur posera des questions au candidat pour vérifier l'acquisition des compétences sur la détermination des objectifs d'un nettoyage de données.	
	C1.2.2 Effectuer des opérations de nettoyage sur des données structurées, notamment l'identification d' <i>outliers</i> et le remplissage de données manquantes, dans le respect des normes en vigueur (RGPD) afin de proposer des données exploitables pour une problématique métier.	L'évaluation de la compétence est réalisée au moyen d'une mise en situation professionnelle reconstituée. Le candidat doit réaliser un travail de filtrage et de nettoyage complet des données afin qu'elles puissent être exploitées dans le contexte d'une future application : suppression de variables non utiles ou non exploitables, traitement des valeurs aberrantes, des valeurs manquantes, des doublons. Le travail est présenté lors d'une soutenance à un évaluateur (expert technique). L'évaluateur posera des questions au candidat pour vérifier l'acquisition des compétences sur les opérations de nettoyage.	CE1 Le candidat a éliminé les variables non pertinentes pour la problématique de l'application. CE2 Le candidat a proposé et justifié au minimum trois méthodes de traitement des valeurs manquantes, adaptées aux variables concernées (médiane, mise à 0, IterativeImputer, KNN, suppression...). CE3 Le candidat a identifié, quantifié et traité les valeurs aberrantes de chaque variable, en tenant compte du contexte métier. CE4 Le candidat a traité les duplicats de variables et d'enregistrements. CE5 Le candidat a mis en œuvre une automatisation de certains traitements, à l'aide de fonctions et méthodes adaptées. CE6 Le candidat veille au respect des normes RGPD pendant les opérations de nettoyage CE7 La présentation orale est claire et synthétique.
A1.3 - Réalisation d'analyses statistiques et de représentations graphiques de données	C1.3.1 Effectuer des analyses statistiques univariées et multivariées, notamment des tests statistiques, à partir de données structurées et nettoyées afin de présenter leur	L'évaluation de la compétence est réalisée au moyen d'une mise en situation professionnelle reconstituée. Le candidat doit réaliser un travail d'analyse statistique sur les principales variables, numériques et catégorielles, afin de détecter	<i>Pertinence de l'analyse statistique uni-variée :</i> CE1 Le candidat a mis en évidence et analysé les éventuels <i>outliers</i> d'un point de vue statistique, afin de s'assurer s'il s'agit de valeurs aberrantes ou atypiques CE2 Le candidat a caractérisé correctement les distributions observées (uni, bi, multi-modale)

structurées	distribution, valider leur cohérence et détecter des corrélations.	<p>des valeurs atypiques et présenter leurs densités et leurs distributions statistiques.</p> <p>Il doit réaliser également un travail d'analyse statistique bi-variée et multivariée entre les principales variables qui seront utilisées par l'application. Il doit réaliser notamment des tests statistiques pour vérifier la significativité des résultats d'analyse et identifier des arguments justifiant la faisabilité (ou non) de l'application à partir des données.</p> <p>Le travail est présenté lors d'une soutenance à un évaluateur (expert technique). L'évaluateur posera des questions au candidat pour vérifier l'acquisition des compétences sur les opérations d'analyses statistiques univariées et multivariées.</p>	<p>CE3 Le candidat a utilisé des métriques adaptées (moyenne ou médiane selon la dispersion)</p> <p>CE4 Le candidat a pu correctement définir le terme "quantiles".</p> <p><i>Pertinence de l'analyse statistique bi-variées et multivariées :</i></p> <p>CE5 Le candidat a présenté et analysé au minimum trois graphiques d'analyse bi-variée (pairplot ou graphique variables numérique / numérique, graphique variables numérique / catégorielle, heatmap de corrélation).</p> <p>CE6 Le candidat a expliqué, justifié et appliqué au minimum une méthode d'analyse multivariée descriptive sur le jeu de données (par exemple une ACP).</p> <p>CE7 Le candidat a expliqué, justifié et appliqué au minimum une méthode d'analyse multivariée explicative sur le jeu de données (par exemple une ANOVA, un test statistique).</p> <p>CE8 La présentation orale est claire et synthétique.</p>
	C1.3.2 Représenter des données grâce à des graphiques afin d'expliquer et justifier les analyses réalisées.	<p>L'évaluation de la compétence est réalisée au moyen d'une mise en situation professionnelle reconstituée.</p> <p>Le candidat doit réaliser des graphiques dès que nécessaire, afin de représenter les données et justifier les analyses réalisées : il doit démontrer sa capacité à mettre en œuvre de manière appropriée les principaux types de graphiques à sa disposition.</p> <p>Le travail est présenté lors d'une soutenance à un évaluateur (expert technique). L'évaluateur posera des questions au candidat pour vérifier l'acquisition des compétences sur les opérations de représentation des données grâce à des graphiques.</p>	<p>CE1 Le candidat a identifié les cas où il était nécessaire de créer un graphique (en associant par exemple un graphe à la description de la distribution d'une variable).</p> <p>CE2 Le candidat a réalisé des graphiques lisibles (taille de texte suffisante, définition lisible).</p> <p>CE3 Le candidat a mis en œuvre au moins un de chacun des graphiques suivants : boxplot, barplot, pie chart, histogramme, scatter plot.</p> <p>CE4 La présentation orale est claire et synthétique.</p>

	<p>C1.3.3 Utiliser des techniques appropriées de réduction en deux dimensions de données de grande dimension et les représenter graphiquement afin d'en réaliser l'analyse exploratoire.</p>	<p>L'évaluation de la compétence est réalisée au moyen d'une mise en situation professionnelle reconstituée.</p> <p>Le candidat doit, pour des données de grande dimension, réaliser un travail de réduction en deux dimensions, les représenter sur un graphique en 2D, et réaliser une analyse. L'analyse consiste à vérifier graphiquement la bonne séparation des données selon leur catégorie d'origine pour ainsi statuer sur la faisabilité d'une classification automatique des données selon la catégorie.</p> <p>Le travail est présenté lors d'une soutenance à un évaluateur (expert technique). L'évaluateur posera des questions au candidat pour vérifier l'acquisition des compétences sur les opérations de réduction de la dimension en 2D de son jeu de données, de sa représentation sous forme de graphique et de son analyse.</p>	<p>CE1 Le candidat a mis en œuvre au moins une technique de réduction de dimension en 2D (via T-SNE ou autre technique).</p> <p>CE2 Le candidat a réalisé au moins un graphique représentant les données réduites en 2D .</p> <p>CE3 Le candidat a réalisé et formalisé une analyse du graphique en 2D</p> <p>CE4 La présentation orale est claire et synthétique.</p>
<p>A1.4 - Conception et adaptation de variables pertinentes pour les modèles d'apprentissage (<i>feature engineering</i>)</p>	<p>C1.4.1 Sélectionner, transformer et créer les variables pertinentes d'un modèle d'apprentissage supervisé ou non supervisé (<i>feature engineering</i>) afin de réaliser un apprentissage optimal.</p>	<p>L'évaluation de la compétence est réalisée au moyen d'une mise en situation professionnelle reconstituée.</p> <p>Le candidat doit réaliser un travail de préparation des variables qui seront utilisées pour l'entraînement du modèle, variables catégorielles et/ou numériques, à partir d'un jeu de données brut.</p> <p>Le travail est présenté lors d'une soutenance à un évaluateur (expert technique). L'évaluateur posera des questions au candidat pour vérifier l'acquisition des compétences sur les opérations de <i>feature engineering</i>.</p>	<p>CE1 Le candidat a identifié les variables catégorielles (qualitatives).</p> <p>CE2 Les variables catégorielles identifiées ont été transformées en fonction du besoin (par exemple <i>via</i> OneHotEncoder ou TargetEncoder).</p> <p>CE3 Le candidat a créé de nouvelles variables à partir de variables existantes.</p> <p>CE4 Le candidat a réalisé des transformations mathématiques lorsque cela est requis pour transformer les distributions de variables.</p> <p>CE5 Le candidat a normalisé les variables lorsque c'est requis.</p> <p>CE6 La présentation orale est claire et synthétique.</p>
	<p>C1.4.2 Prétraiter des données non structurées de type texte en prenant en compte les normes liées à la propriété intellectuelle</p>	<p>L'évaluation de la compétence est réalisée au moyen d'une mise en situation professionnelle reconstituée.</p>	<p>CE1 Le candidat a nettoyé les champs de texte (suppression de la ponctuation et des mots de liaison, mise en minuscules).</p>

	<p>et réaliser un <i>feature engineering</i> adapté aux modèles d'apprentissage afin d'obtenir un jeu de données exploitables.</p>	<p>Le candidat doit réaliser un travail de pré-traitement de données non structurées de type texte (suppression ponctuation et des mots de liaison, suppression de types de mots non pertinents dans le contexte du projet, racinisation ou lemmatisation - forme canonique -, etc) et créer des variables adaptées aux modèles d'apprentissage (<i>feature engineering</i>), de type bag-of-word, word embedding ou sentence embedding.</p> <p>Le travail est présenté lors d'une soutenance à un évaluateur (expert technique). L'évaluateur posera des questions au candidat pour vérifier l'acquisition des compétences sur les opérations de pré-traitement de données non structurées de type texte et de <i>feature engineering</i>.</p>	<p>CE2 Le candidat a écrit une fonction permettant de "tokeniser" une phrase.</p> <p>CE3 Le candidat a écrit une fonction permettant de "stemmer" une phrase.</p> <p>CE4 Le candidat a écrit une fonction permettant de "lemmatiser" une phrase.</p> <p>CE5 Le candidat a construit des features ("feature engineering") de type bag-of-words (bag-of-words standard : comptage de mots, et Tf-idf), avec des étapes de nettoyage supplémentaires : seuil de fréquence des mots, normalisation des mots.</p> <p>CE6 Le candidat a testé une phrase ou un court texte d'exemple, pour illustrer la bonne réalisation des 5 étapes précédentes.</p> <p>CE7 Le candidat, en complément de la démarche de type "bag-of-words", a mis en oeuvre 3 démarches de word/sentence embedding : Word2Vec (ou Doc2Vec ou Glove ou FastText), BERT, et USE (Universal Sentence Encoder).</p> <p>CE8 Le candidat s'assure que le texte traité ne relève pas d'une propriété intellectuelle dont l'utilisation ou la modification est interdite.</p> <p>CE9 La présentation orale est claire et synthétique.</p>
	<p>C1.4.3 Prétraiter des données non structurées de type image en veillant au respect du droit à l'image et réaliser un <i>feature engineering</i> adapté aux modèles d'apprentissage afin d'obtenir un jeu de données exploitables.</p>	<p>L'évaluation de la compétence est réalisée au moyen d'une mise en situation professionnelle reconstituée.</p> <p>Le candidat doit réaliser un travail de pré-traitement de données non structurées de type image (passage en gris, filtrage du bruit, égalisation, floutage), et créer des variables adaptées aux modèles d'apprentissage (<i>feature engineering</i>), par exemple <i>via</i> une approche de type descripteur comme SIFT, et <i>via</i> une approche de type CNN Transfer Learning.</p>	<p>CE1 Le candidat a utilisé des bibliothèques spécialisées pour un premier traitement du contraste (ex. : openCV).</p> <p>CE2 Le candidat a présenté des opérations de retraitement d'images (par exemple passage en gris, filtrage du bruit, égalisation, floutage) sur un ou plusieurs exemples.</p> <p>CE3 Le candidat a élaboré une fonction permettant d'extraire des features ("feature engineering") de type</p>

		<p>Le travail est présenté lors d'une soutenance à un évaluateur (expert technique). L'évaluateur posera des questions au candidat pour vérifier l'acquisition des compétences sur les opérations de pré-traitement de données non structurées de type image et de feature engineering.</p>	<p>"bag-of-images" via la génération de descripteurs (algorithmes ORB, ou SIFT, ou SURF). CE4 Le candidat a élaboré et exécuté une fonction permettant d'extraire des features ("feature engineering") via un algorithme de Transfer Learning basé sur des réseaux de neurones, comme par exemple CNN). CE5 Le candidat s'assure que les images utilisées ne relèvent pas d'une propriété intellectuelle. CE6 La présentation orale est claire et synthétique.</p>
	<p>C1.4.4 Réduire la dimension de données de grande dimension, afin d'optimiser les temps d'entraînement des modèles.</p>	<p>L'évaluation de la compétence est réalisée au moyen d'une mise en situation professionnelle reconstituée.</p> <p>Le candidat doit réaliser un travail adapté de réduction de la dimension d'un jeu de données, qui conserve l'intégrité des données et sa variance expliquée, permettant ainsi, tout en maintenant la pertinence du jeu de données, de diminuer les temps d'entraînement des modèles.</p> <p>Le travail est présenté lors d'une soutenance à un évaluateur (expert technique). L'évaluateur posera des questions au candidat pour vérifier l'acquisition des compétences sur les opérations de réduction de la dimension de son jeu de données.</p>	<p>CE1 Le candidat a justifié la nécessité de la réduction de dimension. CE2 Le candidat a appliqué une méthode de réduction de dimension adaptée à la problématique (ex. : ACP). CE3 Le candidat a justifié le choix des valeurs des paramètres dans la méthode de réduction de dimension retenue (ex. : le nombre de dimensions conservées pour l'ACP). CE4 La présentation orale est claire et synthétique.</p>

RÉFÉRENTIEL D'ACTIVITÉS décrit les situations de travail et les activités exercées, les métiers ou emplois visés	RÉFÉRENTIEL DE COMPÉTENCES identifie les compétences et les connaissances, y compris transversales, qui découlent du référentiel d'activités	RÉFÉRENTIEL D'ÉVALUATION définit les critères et les modalités d'évaluation des acquis	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION
Bloc de compétences 2 – Concevoir des modèles d'apprentissage automatique à partir de données structurées et de données non structurées (texte et images)			
A2.1 – Définition et mise en œuvre de la stratégie d'élaboration d'un modèle d'apprentissage supervisé avec des données structurées, pour réaliser une analyse prédictive	C2.1.1 Définir la stratégie d'élaboration d'un modèle d'apprentissage supervisé et sélectionner et entraîner des modèles adaptés à une problématique métier afin de réaliser une analyse prédictive.	<p>L'évaluation de la compétence est réalisée au moyen d'une mise en situation professionnelle reconstituée.</p> <p>Le candidat doit réaliser un travail de définition de sa stratégie d'élaboration d'un modèle d'apprentissage supervisé en choisissant sa ou ses cibles et différents algorithmes adaptés, et en mettant en œuvre l'entraînement de ces algorithmes, pour répondre à la problématique métier.</p> <p>Le travail est présenté lors d'une soutenance à un évaluateur (expert technique). L'évaluateur posera des questions au candidat pour vérifier l'acquisition des compétences sur la définition et la mise en œuvre de l'élaboration d'un modèle d'apprentissage supervisé.</p>	<p>CE1 Le candidat a défini sa stratégie d'élaboration d'un modèle pour répondre à un besoin métier</p> <p>CE2 Le candidat a choisi la ou les variables cibles pertinentes.</p> <p>CE3 Le candidat a vérifié qu'il n'y a pas de problème de data leakage (c'est-à-dire, des variables trop corrélées à la variable cible et inconnues <i>a priori</i> dans les données en entrée du modèle).</p> <p>CE4 Le candidat a testé plusieurs algorithmes de façon cohérente, en partant des plus simples vers les plus complexes (au minimum un linéaire et un non linéaire).</p> <p>CE5 La présentation orale est claire et synthétique.</p>
	C2.1.2 Évaluer les performances des modèles d'apprentissage supervisé selon différents critères (scores, temps d'entraînement, etc.) en adaptant les paramètres afin de choisir le modèle le plus	<p>L'évaluation de la compétence est réalisée au moyen d'une mise en situation professionnelle reconstituée.</p> <p>Le candidat doit réaliser un travail d'évaluation des performances des différents algorithmes d'apprentissage supervisé, testés, à l'aide d'une ou plusieurs mesures adaptées à la problématique métier.</p>	<p>CE1 Le candidat a choisi une métrique adaptée pour évaluer la performance d'un algorithme (par exemple : R2 ou RMSE en régression, accuracy ou AUC en classification, etc.)</p> <p>CE2 Le candidat a exploré d'autres indicateurs de performance que le score pour comprendre les résultats (coefficients des variables en fonction de la</p>

	<p>performant pour la problématique métier.</p>	<p>Le candidat doit également réaliser un travail d'optimisation des hyper-paramètres des algorithmes d'apprentissage supervisé sélectionnés, en particulier de l'algorithme qui est choisi pour élaborer le modèle final, afin d'améliorer sa performance.</p> <p>Le travail est présenté lors d'une soutenance à un évaluateur (expert technique). L'évaluateur posera des questions au candidat pour vérifier l'acquisition des compétences sur les opérations d'évaluation des performances et sur les opérations d'optimisation des hyperparamètres des algorithmes d'apprentissage supervisé.</p>	<p>pénalisation, visualisation des erreurs en fonction des variables du modèle, temps de calcul...)</p> <p>CE3 Le candidat a séparé les données en train/test pour les évaluer de façon pertinente et détecter l'overfitting</p> <p>CE4 Le candidat a mis en place un modèle simple de référence pour évaluer le pouvoir prédictif du modèle choisi (dummyRegressor ou dummyClassifier)</p> <p>CE5 Le candidat a pris en compte dans sa démarche de modélisation l'éventuel déséquilibre des classes (dans le cas d'une classification).</p> <p>CE6 Le candidat a optimisé les hyper-paramètres pertinents dans les différents algorithmes.</p> <p>CE7 Le candidat a mis en place une validation croisée (via GridsearchCV, RandomizedSearchCV ou équivalent) afin d'optimiser les hyperparamètres et comparer les modèles.</p> <p>CE8 Le candidat a présenté l'ensemble des résultats en allant des modèles les plus simples aux plus complexes. Il a justifié le choix final de l'algorithme et des hyperparamètres.</p> <p>CE9 Le candidat a réalisé l'analyse de l'importance des variables (feature importance) globale sur l'ensemble du jeu de données et locale sur chaque individu du jeu de données.</p> <p>CE10 La présentation orale est claire et synthétique.</p>
<p>A2.2 – Définition et mise en oeuvre de la stratégie d'élaboration d'un modèle d'apprentissage non supervisé avec des</p>	<p>C2.2.1 Définir la stratégie d'élaboration d'un modèle d'apprentissage non supervisé, sélectionner et entraîner des modèles adaptés à une problématique métier afin de</p>	<p>L'évaluation de la compétence est réalisée au moyen d'une mise en situation professionnelle reconstituée.</p> <p>Le candidat doit réaliser un travail de définition de sa stratégie d'élaboration d'un modèle d'apprentissage non supervisé, et de mise en œuvre de l'entraînement d'algorithmes appropriés, pour répondre à la problématique métier.</p>	<p>CE1 Le candidat a défini sa stratégie d'élaboration d'un modèle pour répondre à un besoin métier</p> <p>CE2 Le candidat a proposé un nombre de segments et une répartition adaptés à la problématique métier.</p> <p>CE3 Le candidat a explicité la stratégie d'ajout de nouveaux clients.</p>

données structurées, pour segmenter ou réduire des données	réaliser une segmentation ou une réduction de données.	Le travail est présenté lors d'une soutenance à un évaluateur (expert technique). L'évaluateur posera des questions au candidat pour vérifier l'acquisition des compétences sur la définition et la mise en œuvre de l'élaboration d'un modèle d'apprentissage non supervisé.	<p>CE4 Le candidat a pris en compte la nature des variables d'entrée dans le choix des algorithmes et de la distance.</p> <p>CE5 Le candidat a testé et comparé plusieurs algorithmes.</p> <p>CE6 La présentation orale est claire et synthétique.</p>
	C2.2.2 Évaluer les performances des modèles d'apprentissage non supervisé selon différents critères (scores, temps d'entraînement, etc.) en adaptant les paramètres afin de choisir le modèle le plus performant pour la problématique métier.	<p>L'évaluation de la compétence est réalisée au moyen d'une mise en situation professionnelle reconstituée.</p> <p>Le candidat doit réaliser un travail d'évaluation des performances des différents algorithmes d'apprentissage non supervisé, testés, à l'aide d'une ou plusieurs mesures adaptées à la problématique métier.</p> <p>Le candidat doit également réaliser un travail d'optimisation des hyper-paramètres des algorithmes d'apprentissage non supervisé sélectionnés, en particulier de l'algorithme qui est choisi pour élaborer le modèle final, afin d'améliorer sa performance.</p> <p>Le travail est présenté lors d'une soutenance à un évaluateur (expert technique). L'évaluateur posera des questions au candidat pour vérifier l'acquisition des compétences sur les opérations d'évaluation des performances et sur les opérations d'optimisation des hyper-paramètres des algorithmes d'apprentissage non supervisé.</p>	<p>CE1 Le candidat a choisi des métriques adaptées pour évaluer les valeurs candidates pour le nombre de segments (selon algorithme, par exemple : coefficient silhouette / elbow pour le KMeans, eps/min samples pour le DBScan).</p> <p>CE2 Le candidat a évalué la forme des clusters.</p> <p>CE3 Le candidat a évalué la stabilité des clusters.</p> <p>CE4 Le candidat a optimisé les hyper-paramètres pertinents dans les différents algorithmes.</p> <p>CE5 Le candidat a justifié le choix final d'algorithme et d'hyper-paramètres.</p> <p>CE6 La présentation orale est claire et synthétique.</p>
A2.3 – Définition et mise en oeuvre de la stratégie d'élaboration d'un modèle d'apprentissage profond (deep	C2.3.1 Définir la stratégie d'élaboration d'un modèle d'apprentissage profond, concevoir ou ré-utiliser des modèles pré-entraînés (<i>transfer learning</i>) et entraîner des	<p>L'évaluation de la compétence est réalisée au moyen d'une mise en situation professionnelle reconstituée.</p> <p>Le candidat doit réaliser un travail de définition de sa stratégie d'élaboration d'un modèle d'apprentissage profond, et de mise en œuvre de l'entraînement d'algorithmes appropriés, pour répondre à la problématique métier.</p>	<p>CE1 Le candidat a défini sa stratégie d'élaboration d'un modèle pour répondre à un besoin métier (par exemple : choix de conception d'un modèle ou ré-utilisation de modèles pré-entraînés).</p> <p>CE2 Le candidat a identifié la ou les cibles.</p>

<p>learning) pour réaliser une analyse prédictive</p>	<p>modèles afin de réaliser une analyse prédictive.</p>	<p>Le travail est présenté lors d'une soutenance à un évaluateur (expert technique). L'évaluateur posera des questions au candidat pour vérifier l'acquisition des compétences sur la définition et la mise en œuvre de l'élaboration d'un modèle d'apprentissage profond.</p>	<p>CE3 Le candidat a réalisé la séparation du jeu de données en jeu d'entraînement, jeu de validation et jeu de test. CE4 Le candidat s'est assuré qu'il n'y a pas de fuite d'information entre les deux jeux de données (entraînement, validation et test). CE5 Le candidat a testé plusieurs modèles d'apprentissage profond (par exemple à l'aide de la librairie Tensorflow / Keras) en partant du plus simple vers les plus complexes. CE6 Le candidat a mis en oeuvre des modèles à partir de modèles pré-entraînés (technique de Transfer Learning) CE7 La présentation orale est claire et synthétique.</p>
	<p>C2.3.2 Évaluer la performance des modèles d'apprentissage profond selon différents critères (scores, temps d'entraînement, etc.) afin de choisir le modèle le plus performant pour la problématique métier.</p>	<p>L'évaluation de la compétence est réalisée au moyen d'une mise en situation professionnelle reconstituée.</p> <p>Le candidat doit réaliser un travail d'évaluation des performances des différents algorithmes d'apprentissage profond testés, à l'aide d'une ou plusieurs mesures adaptées à la problématique métier.</p> <p>Le travail est présenté lors d'une soutenance à un évaluateur (expert technique). L'évaluateur posera des questions au candidat pour vérifier l'acquisition des compétences sur les opérations d'évaluation des performances des algorithmes d'apprentissage profond.</p>	<p>CE1 Le candidat a choisi une métrique adaptée à la problématique métier, et sert à évaluer la performance des modèles CE2 Le candidat a explicité le choix de la métrique d'évaluation CE3 Le candidat a évalué la performance d'un modèle de référence et sert de comparaison pour évaluer la performance des modèles plus complexes CE4 Le candidat a calculé, hormis la métrique choisie, au moins un autre indicateur pour comparer les modèles (par exemple : le temps nécessaire pour l'entraînement du modèle) CE5 Le candidat a optimisé au moins un des hyperparamètres du modèle choisi (par exemple : le choix de la fonction Loss, le Batch Size, le nombre d'Epochs)</p>

			<p>CE6 Le candidat a présenté une synthèse comparative des différents modèles, par exemple sous forme de tableau.</p> <p>CE7 La présentation orale est claire et synthétique.</p>
	<p>C2.3.3 Utiliser des techniques d'augmentation des données afin d'améliorer la performance des modèles.</p>	<p>L'évaluation de la compétence est réalisée au moyen d'une mise en situation professionnelle reconstituée.</p> <p>Le candidat doit réaliser un travail de mise en œuvre d'une fonction d'augmentation des données, permettant de générer un jeu de données plus important pour l'entraînement du modèle et ainsi en améliorer sa performance.</p> <p>Le travail est présenté lors d'une soutenance à un évaluateur (expert technique). L'évaluateur posera des questions au candidat pour vérifier l'acquisition des compétences sur les opérations d'augmentation des données.</p>	<p>CE1 Le candidat a utilisé plusieurs techniques d'augmentation des données (ex. pour des images : rotation, changement d'échelle, ajout de bruit...).</p> <p>CE2 Le candidat a présenté une synthèse comparative des améliorations de performance grâce aux différentes techniques d'augmentation de données utilisées (maîtrise de l'overfitting, meilleure performance).</p> <p>CE3 La présentation orale est claire et synthétique.</p>

RÉFÉRENTIEL D'ACTIVITÉS <i>décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	RÉFÉRENTIEL DE COMPÉTENCES <i>identifie les compétences et les connaissances, y compris transversales, qui découlent du référentiel d'activités</i>	RÉFÉRENTIEL D'ÉVALUATION <i>définit les critères et les modalités d'évaluation des acquis</i>	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION

Bloc de compétences 3 – Gérer le cycle de vie et le déploiement en production d'un modèle (MLOps) ou d'une application dans des environnements classiques et Big Data

A3.1 - Gestion du cycle de vie et déploiement continu en production d'un modèle dans le cadre d'une démarche orientée MLOps	C3.1.1 Définir et mettre en œuvre un <i>pipeline</i> d'entraînement des modèles, avec centralisation du stockage des modèles et formalisation des résultats et mesures des différentes expérimentations réalisées, afin d'industrialiser le projet de Machine Learning.	<p>L'évaluation de la compétence est réalisée au moyen d'une mise en situation professionnelle reconstituée.</p> <p>Le candidat doit définir et mettre en œuvre le pipeline d'entraînement des modèles afin de suivre l'exécution de l'entraînement de ses modèles et de stocker ses modèles.</p> <p>Le travail est présenté lors d'une soutenance à un évaluateur (expert technique). L'évaluateur posera des questions au candidat pour vérifier l'acquisition des compétences sur la définition et la mise en œuvre du pipeline d'entraînement des modèles.</p>	<p>CE1 Le candidat a mis en oeuvre un pipeline d'entraînement des modèles reproductible</p> <p>CE2 Le candidat a sérialisé et stocké les modèles créés dans un registre centralisé afin de pouvoir facilement les réutiliser.</p> <p>CE3 Le candidat a formalisé des mesures et résultats de chaque expérimentation, afin de les analyser et de les comparer</p> <p>CE4 La présentation orale est claire et synthétique.</p>
	C3.1.2 Mettre en œuvre un logiciel de version de code afin d'assurer en continu l'intégration et la diffusion du modèle auprès de collaborateurs.	<p>L'évaluation de la compétence est réalisée au moyen d'une mise en situation professionnelle reconstituée.</p> <p>Le candidat doit gérer son code en version via un logiciel de gestion de version, et pouvoir le partager aux autres collaborateurs.</p> <p>Le travail est présenté lors d'une soutenance à un évaluateur (expert technique). L'évaluateur posera des questions au candidat pour vérifier l'acquisition des compétences sur les opérations de gestion en version et de partage du code réalisé.</p>	<p>CE1 Le candidat a créé un dossier contenant tous les scripts du projet dans un logiciel de version de code (ex : Git) et l'a partagé (ex : Github).</p> <p>CE2 Le candidat a présenté un historique des modifications du projet qui affiche au moins trois versions distinctes, auxquelles il est possible d'accéder.</p> <p>CE3 Le candidat a tenu à jour et mis à disposition la liste des packages utilisés ainsi que leur numéro de version .</p> <p>CE4 Le candidat a rédigé un fichier introductif permettant de comprendre l'objectif du projet et le découpage des dossiers.</p>

			<p>CE5 Le candidat a commenté les scripts et les fonctions facilitant une réutilisation du travail par d'autres personnes et la collaboration.</p> <p>CE6 La présentation orale est claire et synthétique.</p>
	<p>C3.1.3 Concevoir et assurer un déploiement continu d'un moteur d'inférence (modèle de prédiction encapsulé dans une API) sur une plateforme <i>Cloud</i> afin de permettre à des applications de réaliser des prédictions <i>via</i> une requête à l'API.</p>	<p>L'évaluation de la compétence est réalisée au moyen d'une mise en situation professionnelle reconstituée.</p> <p>Le candidat doit concevoir et assurer un déploiement continu d'un moteur d'inférence sur une plateforme Cloud via une API qui pourra être appelée via une requête par une ou des applications qui exploiteront les résultats de prédiction transmis par l'API. Le pipeline de déploiement mis en œuvre doit permettre le déploiement continu en production de l'API, dans le cadre d'une démarche orientée MLOps.</p> <p>Le travail est présenté lors d'une soutenance à un évaluateur (expert technique). L'évaluateur posera des questions au candidat pour vérifier l'acquisition des compétences sur les opérations de conception, réalisation et déploiement sur le Cloud d'une API.</p>	<p>CE1 Le candidat a défini et préparé un pipeline de déploiement continu.</p> <p>CE2 Le candidat a déployé le modèle de machine learning sous forme d'API (via Flask par exemple) et cette API renvoie bien une prédiction correspondant à une demande.</p> <p>CE3 Le candidat a mis en œuvre un pipeline de déploiement continu, afin de déployer l'API sur un serveur d'une plateforme Cloud.</p> <p>CE4 Le candidat a mis en oeuvre des tests unitaires automatisés (par exemple avec pyTest)</p> <p>CE5 Le candidat a réalisé l'API indépendamment de l'application qui utilise le résultat de la prédiction.</p> <p>CE6 La présentation orale est claire et synthétique.</p>
	<p>C3.1.4 Définir et mettre en œuvre une stratégie de suivi de la performance d'un modèle en production et en assurer la maintenance afin de garantir dans le temps la production de prédictions performantes.</p>	<p>L'évaluation de la compétence est réalisée au moyen d'une mise en situation professionnelle reconstituée.</p> <p>Le candidat doit définir et mettre en oeuvre une stratégie de suivi de la performance d'un modèle en production, notamment en stockant les événements de production et données associées, en formalisant les indicateurs de suivis, les alertes, les analyses de performances et leur évolution dans le temps, et proposant des actions d'amélioration du modèle.</p> <p>Le travail est présenté lors d'une soutenance à un évaluateur (expert technique). L'évaluateur posera des questions au candidat</p>	<p>CE1 Le candidat a défini une stratégie de suivi de la performance du modèle.</p> <p>CE2 Le candidat a réalisé un système de stockage d'événements relatifs aux prédictions réalisées par l'API et une gestion d'alerte en cas de dégradation significative de la performance.</p> <p>CE3 Le candidat a analysé la stabilité du modèle dans le temps et défini des actions d'amélioration de sa performance.</p> <p>CE4 La présentation orale est claire et synthétique.</p>

		pour vérifier l'acquisition des compétences sur la définition et la mise en œuvre de la stratégie de suivi de la performance d'un modèle en production.	
A3.2 - Conception, préparation, et déploiement d'une application ou d'un modèle d'apprentissage en utilisant les technologies du big data	C3.2.1 Sélectionner les outils du <i>Cloud</i> permettant de traiter et stocker les données d'un projet <i>Big Data</i> conforme aux normes RGPD en vigueur afin de concevoir une application de qualité supportant le traitement de données massives.	<p>L'évaluation de la compétence est réalisée au moyen d'une mise en situation professionnelle reconstituée.</p> <p>Le candidat doit définir l'environnement Big Data en sélectionnant les outils du Cloud nécessaires à un traitement de données massives.</p> <p>Le travail est présenté lors d'une soutenance à un évaluateur (expert technique). L'évaluateur posera des questions au candidat pour vérifier l'acquisition des compétences sur les opérations de sélection des outils du Cloud.</p>	<p>CE1 Le candidat a identifié les différentes briques d'architecture nécessaires pour la mise en place d'un environnement Big Data.</p> <p>CE2 Le candidat a identifié les outils du cloud permettant de mettre en place l'environnement Big Data conforme aux normes RGPD en vigueur.</p> <p>CE3 La présentation orale est claire et synthétique.</p>
	C3.2.2 Prétraiter, analyser et modéliser des données (en veillant à leur conformité RGPD) dans un environnement <i>Big Data</i> et en utilisant les outils du <i>Cloud</i> afin de concevoir une application sécurisée de qualité supportant le traitement de données massives.	<p>L'évaluation de la compétence est réalisée au moyen d'une mise en situation professionnelle reconstituée.</p> <p>Le candidat doit réaliser une gestion des données dans un environnement Big Data du début à la fin du traitement.</p> <p>Le travail est présenté lors d'une soutenance à un évaluateur (expert technique). L'évaluateur posera des questions au candidat pour vérifier l'acquisition des compétences sur les opérations de gestion des données dans un environnement Big Data.</p>	<p>CE1 Le candidat a chargé les fichiers de départ et ceux après transformation dans un espace de stockage cloud conforme à la réglementation RGPD.</p> <p>CE2 Le candidat a exécuté les scripts en utilisant des machines dans le cloud.</p> <p>CE3 Le candidat a réalisé un script qui permet d'écrire les sorties du programme directement dans l'espace de stockage cloud.</p> <p>CE4 La présentation orale est claire et synthétique.</p>
	C3.2.3 Réaliser des calculs distribués sur des données massives en utilisant les outils adaptés et en prenant en compte le RGPD afin de permettre la	<p>L'évaluation de la compétence est réalisée au moyen d'une mise en situation professionnelle reconstituée.</p> <p>Le candidat doit réaliser une parallélisation du traitement via des calculs distribués afin de traiter des données massives.</p>	<p>CE1 Le candidat a identifié les traitements critiques lors d'un passage à l'échelle en termes de volume de données.</p> <p>CE2 Le candidat veille à ce que l'exploitation des données soit conforme au RGPD</p>

	<p>mise en œuvre d'applications à l'échelle.</p>	<p>Le travail est présenté lors d'une soutenance à un évaluateur (expert technique). L'évaluateur posera des questions au candidat pour vérifier l'acquisition des compétences sur les opérations de calcul distribué.</p>	<p>CE3 Le candidat a développé les scripts s'appuyant sur Spark. CE4 Le candidat s'est assuré que toute la chaîne de traitement est exécutée dans le cloud. CE5 La présentation orale est claire et synthétique.</p>
--	--	--	---

RÉFÉRENTIEL D'ACTIVITÉS <i>décrit les situations de travail et les activités exercées, les métiers ou emplois visés</i>	RÉFÉRENTIEL DE COMPÉTENCES <i>identifie les compétences et les connaissances, y compris transversales, qui découlent du référentiel d'activités</i>	RÉFÉRENTIEL D'ÉVALUATION <i>définit les critères et les modalités d'évaluation des acquis</i>	
		MODALITÉS D'ÉVALUATION	CRITÈRES D'ÉVALUATION
Bloc de compétences 4 – Manager des projets IA et de data science complexes en intégrant les contraintes légales et les valeurs d'éthique			
A4.1 Réalisation d'une veille sur les outils et tendances en data science et IA	C4.1 Réaliser une veille sur les outils et tendances en data science et IA afin de mettre à jour son expertise et de s'assurer que les méthodes utilisées mobilisent bien les techniques en vigueur.	<p>L'évaluation de la compétence est réalisée au moyen d'une mise en situation professionnelle reconstituée.</p> <p>Le candidat doit réaliser une veille sur l'état de l'art et ses évolutions récentes : les nouveaux outils, les nouveaux concepts, les nouveaux modèles, les nouvelles démarches et tendances.</p> <p>Le travail est présenté lors d'une soutenance à un évaluateur (expert technique). L'évaluateur posera des questions au candidat pour vérifier l'acquisition des compétences sur la réalisation d'une veille technique en data science.</p>	<p>CE1 Le candidat a consulté des sources reconnues d'informations, produites récemment, (blogs reconnus, articles de recherche de journaux et conférences reconnues dans le domaine).</p> <p>CE2 Le candidat a présenté les points clés de chacune des sources bibliographiques, y compris des détails mathématiques.</p> <p>CE3 Le candidat a bien articulé les différentes sources entre elles.</p> <p>CE4 Le candidat a mis en place une preuve de concept pour tester le nouvel outil / le nouveau modèle, la nouvelle démarche, comparée à une approche classique.</p> <p>CE6 La présentation orale est claire et synthétique.</p>
A4.2 - Organisation de l'expression du besoin des métiers et formalisation de user stories	C4.2 Organiser l'identification des besoins métiers d'un projet de data science / IA et formaliser ces besoins sous forme de <i>user stories</i> afin de réaliser des solutions adaptées aux besoins des clients.	<p>L'évaluation de la compétence est réalisée au moyen d'une mise en situation professionnelle reconstituée.</p> <p>Le candidat doit définir et mettre en œuvre une démarche de recueil du besoin auprès des métiers. Il doit ensuite, à partir de ces besoins recensés, formaliser des user stories (approche agile) décrivant les macro-fonctionnalités à développer, les prioriser et définir leur complexité afin d'en estimer la charge de réalisation.</p>	<p>CE1 Le candidat a défini la démarche de recueil des besoins auprès des métiers (ateliers de sensibilisation ou brainstorming, interviews, documents d'expression de besoin des métiers...)</p> <p>CE2 Le candidat a formalisé dans le backlog les user stories décrivant les macro-fonctionnalités de l'application</p> <p>CE3 Le candidat a estimé la complexité des user-stories</p>

		<p>Le travail est présenté lors d'une soutenance à un évaluateur (expert technique). L'évaluateur posera des questions au candidat pour vérifier l'acquisition des compétences sur l'organisation de l'identification des besoins métiers et la formalisation de user stories.</p>	<p>CE4 La présentation orale est claire et synthétique.</p>
<p>A4.3 Cadrage et gestion d'un projet IA ou de data science, selon une approche agile de type SCRUM</p>	<p>C4.3.1 Identifier les ressources humaines, techniques et financières d'un projet de data science afin de mettre en œuvre la conception et la réalisation de modèles ou d'applications mettant en œuvre des modèles.</p>	<p>L'évaluation de la compétence est réalisée au moyen d'une mise en situation professionnelle reconstituée.</p> <p>Le candidat doit réaliser une identification et une estimation des ressources humaines, techniques et financières d'un projet data science.</p> <p>Le travail est présenté lors d'une soutenance à un évaluateur (expert technique). L'évaluateur posera des questions au candidat pour vérifier l'acquisition des compétences sur les opérations d'identification et d'estimation des ressources.</p>	<p>CE1 Le candidat a qualifié qualitativement et décrit les ressources humaines, techniques et financières requises pour l'ensemble du projet (profils et rôles, type d'environnement technique).</p> <p>CE2 Le candidat a qualifié quantitativement et décrit les ressources humaines, techniques et financières requises pour l'ensemble du projet (nombre de personnes des différents profils, estimation du budget d'infrastructure, séparation des coûts de mise en place et des coûts récurrents).</p> <p>CE3 La présentation orale est claire et synthétique.</p>
	<p>C4.3.2 Intégrer le respect des contraintes légales et éthiques relatives à la collecte de données personnelles afin de respecter les règles du RGPD et maîtriser les problématiques d'éthique des données personnelles traitées.</p>	<p>L'évaluation de la compétence est réalisée au moyen d'une mise en situation professionnelle reconstituée.</p> <p>Le candidat doit analyser et présenter les contraintes légales sur la gestion des données personnelles d'un projet de data science, ainsi que les enjeux éthiques associés. Il doit notamment formaliser un registre des traitements pour au moins un traitement orienté IA, à partir d'un modèle officiel de registre fourni par la CNIL, et en déduire les problématiques de biais dans l'utilisation des données personnelles dans un contexte IA.</p> <p>Le travail est présenté lors d'une soutenance à un évaluateur (expert technique). L'évaluateur posera des questions au candidat</p>	<p>CE1 Le candidat a dressé une liste des sources de données personnelles impliquées dans le projet. Il a notamment formalisé au moins un traitement de type IA dans un registre des traitements RGPD.</p> <p>CE2 Le candidat a listé et brièvement décrit les obligations réglementaires, à destination d'un public non expert.</p> <p>CE3 Le candidat a listé et brièvement décrit les enjeux éthiques, à destination d'un public non expert.</p> <p>CE4 La présentation orale est claire et synthétique.</p>

		pour vérifier l'acquisition des compétences sur l'intégration du respect des contraintes légales et éthiques.	
	C4.3.3 Analyser les risques inhérents au projet et à la gestion des données personnelles et planifier des réponses adéquates aux risques significatifs afin de garantir la maîtrise et la réussite du projet.	<p>L'évaluation de la compétence est réalisée au moyen d'une mise en situation professionnelle reconstituée.</p> <p>Le candidat doit identifier les risques liés à la réalisation du projet, ainsi que les risques induits par la gestion de données personnelles. Il doit également analyser ces risques, leur impact et probabilité, et définir des plans d'action afin de les maîtriser.</p> <p>Le travail est présenté lors d'une soutenance à un évaluateur (expert technique). L'évaluateur posera des questions au candidat pour vérifier l'acquisition des compétences sur la gestion des risques du projet et des risques liés à la gestion de données personnelles.</p>	<p>CE1 Le candidat a recensé les risques liés à la réalisation du projet IA et les risques liés à la gestion des données personnelles</p> <p>CE2 Le candidat a analysé les impacts des risques recensés, leur probabilité et leur criticité</p> <p>CE3 Le candidat a défini un plan de mitigation des risques afin de maîtriser leur impact</p> <p>CE4 Le candidat a réalisé une synthèse des principaux risques</p> <p>CE5 Le candidat prend en compte les impacts de son activité sur l'environnement.</p> <p>CE6 La présentation orale est claire et synthétique.</p>
	C4.3.4 Piloter ou participer au pilotage du projet en planifiant les différents <i>sprints</i> afin de permettre la mobilisation des parties prenantes et le bon déroulement du projet.	<p>L'évaluation de la compétence est réalisée au moyen d'une mise en situation professionnelle reconstituée.</p> <p>Le candidat doit définir la planification et l'organisation du suivi du projet selon une démarche agile. Il doit notamment planifier les sprints de réalisation des fonctionnalités du projet IA, et organiser la gestion des sprints (daily scrum, sprint review, sprint rétrospective, tableaux d'avancement des réalisations) en lien avec les parties prenantes.</p> <p>Le travail est présenté lors d'une soutenance à un évaluateur (expert technique). L'évaluateur posera des questions au candidat pour vérifier l'acquisition des compétences sur le pilotage du projet IA.</p>	<p>CE1 Le candidat a décrit le fonctionnement général du projet selon la démarche agile SCRUM</p> <p>CE2 Le candidat a formalisé la priorisation des besoins (backlog par sprints) et leur planification dans les différents sprints du projet</p> <p>CE3 Le candidat a formalisé le planning des sprints</p> <p>CE4 Le candidat a formalisé l'organisation du suivi et pilotage du projet SCRUM : daily scrum, sprint review, sprint retrospective, tableaux d'avancement des réalisations (burndown chart, ...)</p> <p>CE5 La présentation orale est claire et synthétique.</p>

<p>A4.4 Présentation des projets de data science et restitution des résultats aux métiers</p>	<p>C4.4.1 Réaliser un tableau de bord afin de présenter son travail de modélisation à un public non technique et de manière adaptée à des personnes en situation de handicap en appliquant certains critères d'accessibilité du WCAG.</p>	<p>L'évaluation de la compétence est réalisée au moyen d'une mise en situation professionnelle reconstituée.</p> <p>Le candidat doit réaliser un tableau de bord présentant une analyse des données, une modélisation et des moyens d'interpréter les résultats par un utilisateur non expert. Ce tableau de bord doit être adapté à des personnes en situation de handicap en couvrant certains critères d'accessibilité définis dans le WCAG par le W3C.</p> <p>Le travail est présenté lors d'une soutenance à un évaluateur (expert technique). L'évaluateur posera des questions au candidat pour vérifier l'acquisition des compétences sur la réalisation du tableau de bord.</p>	<p>CE1 Le candidat a décrit et conçu un parcours utilisateur simple permettant de répondre aux besoins des utilisateurs (les différentes actions et clics sur les différents graphiques permettant de répondre à une question que se pose l'utilisateur).</p> <p>CE2 Le candidat a développé au moins deux graphiques interactifs permettant aux utilisateurs d'explorer les données.</p> <p>CE3 Le candidat a réalisé des graphiques lisibles (taille de texte suffisante, définition lisible).</p> <p>CE4 Le candidat a réalisé des graphiques qui permettent de répondre à la problématique métier.</p> <p>CE5 Le candidat a pris en compte le besoin des personnes en situation de handicap dans la réalisation des graphiques : le candidat a pris en compte au minimum les critères d'accessibilité du WCAG suivants :</p> <ul style="list-style-type: none"> ● Critère de succès 1.1.1 Contenu non textuel ● Critère de succès 1.4.1 Utilisation de la couleur ● Critère de succès 1.4.3 Contraste (minimum) ● Critère de succès 1.4.4 Redimensionnement du texte ● Critère de succès 2.4.2 Titre de page <p>CE6 Le candidat a déployé le dashboard sur le web afin qu'il soit accessible pour d'autres utilisateurs sur leurs postes de travail.</p> <p>CE7 La présentation orale est claire et synthétique.</p>
	<p>C4.4.2 Rédiger une note méthodologique contenant notamment le choix des algorithmes testés, les métriques utilisées et l'interprétabilité du modèle</p>	<p>L'évaluation de la compétence est réalisée au moyen d'une mise en situation professionnelle reconstituée.</p> <p>Le candidat doit rédiger une note technique qui décrit l'ensemble de sa démarche et de ses résultats de modélisation : sa stratégie d'élaboration du modèle, les modèles testés, les mesures de</p>	<p>CE1 Le candidat a présenté la démarche de modélisation de manière synthétique dans une note.</p> <p>CE2 Le candidat a explicité la métrique d'évaluation retenue et sa démarche d'optimisation.</p> <p>CE3 Le candidat a explicité l'interprétabilité globale et locale du modèle.</p>

	<p>proposé, afin de communiquer sa démarche de modélisation.</p>	<p>performance du modèle, l'interprétabilité du modèle afin de la partager à un public non expert.</p> <p>Le travail est présenté lors d'une soutenance à un évaluateur (expert technique). L'évaluateur posera des questions au candidat pour vérifier l'acquisition des compétences sur la rédaction de la note technique.</p>	<p>CE4 Le candidat a décrit les limites et les améliorations envisageables pour gagner en performance et en interprétabilité de l'approche de modélisation. CE5 La présentation orale est claire et synthétique.</p>
	<p>C4.4.3 Réaliser la présentation orale d'une démarche de modélisation à un client interne/externe afin de partager les résultats et faciliter la prise de décision de l'interlocuteur.</p>	<p>L'évaluation de la compétence est réalisée au moyen d'une mise en situation professionnelle reconstituée.</p> <p>Le candidat doit réaliser une présentation orale complète de la modélisation mise en œuvre dans son projet de data science. Il doit notamment décrire l'ensemble de sa démarche et de ses résultats de modélisation : sa stratégie d'élaboration du modèle, les modèles testés, les mesures de performance du modèle, l'interprétation de l'importance des variables du modèle.</p> <p>Le travail est présenté lors d'une soutenance à un évaluateur (expert technique). L'évaluateur posera des questions au candidat pour vérifier l'acquisition des compétences sur sa présentation orale.</p>	<p>CE1 Le candidat a expliqué de manière compréhensible par un public non technique la méthode d'évaluation de la performance du modèle de machine learning, la façon d'interpréter les résultats du modèle, et la façon d'interpréter l'importance des variables du modèle. CE2 Le candidat a su répondre de manière simple (compréhensible par un public non technique) à au moins une question portant sur sa démarche de modélisation. CE3 Le candidat a présenté une démarche de modélisation et une évaluation complète des modèles, en particulier la comparaison de plusieurs modèles CE4 La présentation orale est claire et synthétique.</p>