

Expert en data science

RNCP37431 · RNCP · Publiée

L'essentiel

Niveau de qualification	Niveau 7
Code(s) NSF	326
Formacode(s)	31026
Code(s) ROME	M1806, M1805, M1802, M1403
Type d'enregistrement	Enregistrement sur demande
Date d'enregistrement	27/03/2023
Date d'échéance	27/03/2025

Certificateur(s)

OPENCLASSROOMS · SIRET : 49386136300080

OPENCLASSROOMS · SIRET : 49386136300072

Partenaires habilités

1 organisme habilité, d'après la liste publiée par France compétences : 1 à former.

Organisme	SIRET	Habilitation
LYCEE TECHNIQUE REGIONAL LES EUCALYPTUS	19060075900020	Former

Résumé de la certification

OBJECTIFS ET CONTEXTE

La certification professionnelle d'Expert en data science, a initialement été enregistrée au Répertoire national des certifications professionnelles le 30 mars 2020 au niveau 7 sous l'intitulé de Data scientist.

Le périmètre de la certification professionnelle couvre les quatre disciplines - pour partie imbriquées - que constituent la data science (traitement et analyse de données), l'intelligence artificielle (reproduction par des machines de processus cognitifs), l'apprentissage automatique et l'apprentissage automatique profond - machine learning et deep learning -, visant à donner aux machines les capacités d'apprendre à partir de données. Elle s'inscrit dans la continuité du périmètre du Data scientist, tout en intégrant des évolutions en phase avec les besoins du marché.

Elle vise à certifier des experts en statistique, en intelligence artificielle et en apprentissage automatique ; des professionnels capables de conduire des projets complexes d'analyse et de modélisation des données, pour répondre à des problématiques métier (visualisation, appui à la décision, prédiction, classification, traitement de données non structurées tel que du texte ou des images, déploiement de modèles d'apprentissage automatique dans un environnement Cloud, etc.).

La hausse exponentielle des données produites, concomitamment à la disponibilité des technologies permettant de les traiter et à l'intégration croissante de l'intelligence artificielle par les entreprises créent en effet des besoins en recrutement et en compétences pour les experts en data science - besoins que les entreprises peinent actuellement à pourvoir. Cette certification professionnelle s'inscrit dans ce contexte et permet de certifier des professionnels de haut niveau, en réponse à ces besoins.

ACTIVITÉS VISÉES

Collecter, pré-traiter et analyser des données et créer des variables pertinentes (feature engineering) pour l'entraînement d'un modèle

Définition de la stratégie de collecte de données et réalisation d'une collecte via une API

Définition et gestion du nettoyage de données structurées

Réalisation d'analyses statistiques et de représentations graphiques de données structurées

Conception et adaptation de variables pertinentes pour les modèles d'apprentissage (feature engineering)

Concevoir des modèles d'apprentissage à partir de données structurées et de données non structurées (texte et images)

Définition et mise en oeuvre de la stratégie d'élaboration d'un modèle d'apprentissage supervisé avec des données structurées, pour réaliser une analyse prédictive

Définition et mise en oeuvre de la stratégie d'élaboration d'un modèle d'apprentissage non supervisé avec des données structurées, pour segmenter ou réduire des données

Définition et mise en oeuvre de la stratégie d'élaboration d'un modèle d'apprentissage profond (deep learning) pour réaliser une analyse prédictive

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

Gestion du cycle de vie et déploiement continu en production d'un modèle dans le cadre d'une démarche orientée MLOps

Conception, préparation, et déploiement d'une application ou d'un modèle d'apprentissage en utilisant les technologies du big data

Manager des projets IA et de data science complexes en intégrant les contraintes légales et les valeurs d'éthique

Réalisation d'une veille sur les outils et tendances en data science et IA

Organisation de l'expression du besoin des métiers et formalisation de user stories

Cadrage et gestion d'un projet IA ou de data science, selon une approche agile de type SCRUM

Présentation des projets de data science et restitution des résultats aux métiers

COMPÉTENCES ATTESTÉES

Définir la stratégie de collecte de données en recensant les API disponibles, et réaliser la collecte des données répondant à des critères définis via une API (interface de programmation) en prenant en compte les normes RGPD, afin de les exploiter pour l'élaboration d'un modèle.

Déterminer les objectifs du nettoyage des données et de la démarche de mise en oeuvre, afin de construire un jeu de données adapté à une problématique métier.

Effectuer des opérations de nettoyage sur des données structurées, notamment l'identification d'outliers et le remplissage de données manquantes, dans le respect des normes en vigueur (RGPD) afin de proposer des données exploitables pour une problématique métier.

Effectuer des analyses statistiques univariées et multivariées, notamment des tests statistiques, à partir de données structurées et nettoyées afin de présenter leur distribution, valider leur cohérence et détecter des corrélations.

Représenter des données grâce à des graphiques afin d'explicitier et justifier les analyses réalisées.

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.

Sélectionner, transformer et créer les variables pertinentes d'un modèle d'apprentissage supervisé ou non supervisé (feature engineering) afin de réaliser un apprentissage optimal.

Pré-traiter des données non structurées de type texte en prenant en compte les normes liées à la propriété intellectuelle, et réaliser un feature engineering adapté aux modèles d'apprentissage afin d'obtenir un jeu de données exploitables.

Pré-traiter des données non structurées de type image en veillant au respect du droit à l'image et réaliser un feature engineering adapté aux modèles d'apprentissage afin d'obtenir un jeu de données exploitables.

Réduire la dimension de données de grande dimension, afin d'optimiser les temps d'entraînement des modèles.

Définir la stratégie d'élaboration d'un modèle d'apprentissage supervisé, sélectionner et entraîner des modèles adaptés à une problématique métier, afin de réaliser une analyse prédictive.

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 performant pour la problématique métier.

Définir la stratégie d'élaboration d'un modèle d'apprentissage non supervisé, sélectionner et entraîner des modèles adaptés à une problématique métier afin de réaliser une segmentation ou une réduction de données.

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.

Définir la stratégie d'élaboration d'un modèle d'apprentissage profond, concevoir ou ré-utiliser des modèles pré-entraînés (transfer learning) et entraîner des modèles afin de réaliser une analyse prédictive.

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.

Utiliser des techniques d'augmentation des données afin d'améliorer la performance des modèles.

Définir et mettre en œuvre un pipeline 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.

Mettre en œuvre un logiciel de version de code afin d'assurer en continu l'intégration et la diffusion du modèle auprès de collaborateurs.

Concevoir et assurer un déploiement continu d'un moteur d'inférence (modèle de prédiction encapsulé dans une API) sur une plateforme Cloud afin de permettre à des applications de réaliser des prédictions via une requête à l'API.

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.

Sélectionner les outils du Cloud permettant de traiter et stocker les données d'un projet Big Data conforme aux normes RGPD en vigueur, afin de concevoir une application de qualité supportant le traitement de données massives. Pré-traiter, analyser et modéliser des données (en veillant à leur conformité RGPD) dans un environnement Big data en utilisant les outils du Cloud, afin de concevoir une application sécurisée de qualité supportant le traitement de données massives.

Réaliser des calculs distribués sur des données massives en utilisant les outils adaptés et en prenant en compte le RGPD, afin de permettre la mise en œuvre d'applications à l'échelle.

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.

Organiser l'identification des besoins métiers d'un projet de Data Science / IA et formaliser ces besoins sous forme de user stories afin de réaliser des solutions adaptées aux besoins des clients.

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.

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.

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.

Piloter ou participer au pilotage du projet en planifiant les différents sprints afin de permettre la mobilisation des parties prenantes et le bon déroulement du projet.

Réaliser un tableau de bord afin de présenter son travail de modélisation à un public non technique, et adapté à des personnes en situation de handicap en appliquant certains critères d'accessibilité du WCAG.

Rédiger une note méthodologique, contenant notamment le choix des algorithmes testés, les métriques utilisées et l'interprétabilité du modèle proposé, afin de communiquer sa démarche de modélisation.

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.

Blocs de compétences

RNCP37431BC01 — Collecter, pré-traiter, analyser des données et créer des variables pertinentes (feature engineering) pour l'entraînement d'un modèle

LISTE DE COMPÉTENCES

Définir la stratégie de collecte de données en recensant les API disponibles, et réaliser la collecte des données répondant à des critères définis via une API (interface de programmation) en prenant en compte les normes RGPD, afin de les exploiter pour l'élaboration d'un modèle.

Déterminer les objectifs du nettoyage des données et de la démarche de mise en œuvre, afin de construire un jeu de données adapté à une problématique métier.

Effectuer des opérations de nettoyage sur des données structurées, notamment l'identification d'outliers et le remplissage de données manquantes, dans le respect des normes en vigueur (RGPD) afin de proposer des données exploitables pour une problématique métier.

Effectuer des analyses statistiques univariées et multivariées, notamment des tests statistiques, à partir de données structurées et nettoyées afin de présenter leur distribution, valider leur cohérence et détecter des corrélations.
Représenter des données grâce à des graphiques afin d'explicitier et justifier les analyses réalisées.
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.
Sélectionner, transformer et créer les variables pertinentes d'un modèle d'apprentissage supervisé ou non supervisé (feature engineering) afin de réaliser un apprentissage optimal.
Pré-traiter des données non structurées de type texte en prenant en compte les normes liées à la propriété intellectuelle, et réaliser un feature engineering adapté aux modèles d'apprentissage afin d'obtenir un jeu de données exploitables.
Pré-traiter des données non structurées de type image en veillant au respect du droit à l'image et réaliser un feature engineering adapté aux modèles d'apprentissage afin d'obtenir un jeu de données exploitables.
Réduire la dimension de données de grande dimension, afin d'optimiser les temps d'entraînement des modèles.

MODALITÉS D'ÉVALUATION

Projets (mises en situation professionnelle) et soutenances.

RNCP37431BC02 — Concevoir des modèles d'apprentissage automatique à partir de données structurées et de données non structurées (texte et images)

LISTE DE COMPÉTENCES

Définir la stratégie d'élaboration d'un modèle d'apprentissage supervisé, sélectionner et entraîner des modèles adaptés à une problématique métier, afin de réaliser une analyse prédictive.
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 performant pour la problématique métier.
Définir la stratégie d'élaboration d'un modèle d'apprentissage non supervisé, sélectionner et entraîner des modèles adaptés à une problématique métier afin de réaliser une segmentation ou une réduction de données.
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.
Définir la stratégie d'élaboration d'un modèle d'apprentissage profond, concevoir ou ré-utiliser des modèles pré-entraînés (transfer learning) et entraîner des modèles afin de réaliser une analyse prédictive.
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.
Utiliser des techniques d'augmentation des données afin d'améliorer la performance des modèles.

MODALITÉS D'ÉVALUATION

Projets (mises en situation professionnelle) et soutenances.

RNCP37431BC03 — 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

LISTE DE COMPÉTENCES

Définir et mettre en œuvre un pipeline 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.
Mettre en œuvre un logiciel de version de code afin d'assurer en continu l'intégration et la diffusion du modèle auprès de collaborateurs.
Concevoir et assurer un déploiement continu d'un moteur d'inférence (modèle de prédiction encapsulé dans une API) sur une plateforme Cloud afin de permettre à des applications de réaliser des prédictions via une requête à l'API.
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.
Sélectionner les outils du Cloud permettant de traiter et stocker les données d'un projet Big Data conforme aux normes RGPD en vigueur, afin de concevoir une application de qualité supportant le traitement de données massives.
Prétraiter, analyser et modéliser des données (en veillant à leur conformité RGPD) dans un environnement Big data en utilisant les outils du Cloud, afin de concevoir une application sécurisée de qualité supportant le traitement de données massives.
Réaliser des calculs distribués sur des données massives en utilisant les outils adaptés et en prenant en compte le

RGPD, afin de permettre la mise en œuvre d'applications à l'échelle.

MODALITÉS D'ÉVALUATION

Projets (mises en situation professionnelle) et soutenances.

RNCP37431BC04 — Manager des projets d'intelligence artificielle et de data science complexes en intégrant les contraintes légales et les valeurs d'éthique

LISTE DE COMPÉTENCES

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.

Organiser l'identification des besoins métiers d'un projet de Data Science / IA et formaliser ces besoins sous forme de user stories afin de réaliser des solutions adaptées aux besoins des clients.

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.

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.

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.

Piloter ou participer au pilotage du projet en planifiant les différents sprints afin de permettre la mobilisation des parties prenantes et le bon déroulement du projet.

Réaliser un tableau de bord afin de présenter son travail de modélisation à un public non technique, et adapté à des personnes en situation de handicap en appliquant certains critères d'accessibilité du WCAG.

Rédiger une note méthodologique, contenant notamment le choix des algorithmes testés, les métriques utilisées et l'interprétabilité du modèle proposé, afin de communiquer sa démarche de modélisation.

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.

MODALITÉS D'ÉVALUATION

Projets (mises en situation professionnelle) et soutenances.

Secteur d'activité et type d'emploi

SECTEURS D'ACTIVITÉS

L'expert en data science peut exercer en tant qu'indépendant, dans un cabinet de conseil, en entreprise ou encore dans le secteur public.

Les employeurs sont de toutes tailles et de tous secteurs, mais le besoin est plus important pour les grandes entreprises, exception faite des petites et moyennes entreprises (TPE- PME) technologiques, qui offrent un produit technique, et des cabinets de conseil, qui offrent des prestations d'expertise.

Les secteurs de prédilection pour le recrutement d'un expert en data science sont la banque/ finance, l'assurance, les entreprises et services du numérique (ESN), les entreprises de télécommunication, la logistique et la distribution et, enfin, les grands groupes de tous secteurs, qui recrutent des experts en data science pour leurs besoins d'optimisation en finance, marketing, vente ou produit.

Les principaux secteurs adressés par les entreprises et services du numérique sont la distribution, l'industrie et les services financiers.

TYPE D'EMPLOIS ACCESSIBLES

L'expert en data science

peut exercer plusieurs fonctions, à titre non exhaustif :

- Data Scientist
- Responsable data/ lead data
- Chef de projet data
- Data analyst senior/ Senior Data analyst
- Business analyst senior/ Head of business analysis
- Ingénieur machine learning/ Machine learning engineer
- Ingénieur deep learning/ Deep learning engineer
- Ingénieur vision assistée par ordinateur/ Computer vision scientist
- Ingénieur traitement automatique du langage naturel/ Natural Language processing scientist

- Compliance officer data scientist
- Analyste quantitatif
- Ingénieur Big data
- Data manager

RÉGLEMENTATIONS D'ACTIVITÉ

L'exercice professionnel de l'Expert en data science n'est pas un métier réglementé en termes d'accès ; néanmoins, les activités digitales doivent se conformer au droit applicable au numérique (traitement de données, respect de la vie privée, RGPD, déclarations CNIL...), et plus généralement au droit de la communication et de la propriété intellectuelle.

Marché du travail

Indicateurs au niveau des familles professionnelles (FAP) des métiers visés, agrégés au niveau national — France Travail / DARES (open data).

Tension de recrutement (DARES 2024)	Très élevée (5/5) Moyenne pondérée par l'emploi ; les familles ci-dessous s'échelonnent de 3 à 5 sur 5.
Facteurs principaux	Manque de main-d'œuvre disponible, Lien formation-emploi
Salaire médian (mensuel)	2 912 €
Diplôme le plus fréquent	Diplôme de niveau bac+5 ou plus (doctorat, master ou équivalent) en communication et information (41 % des actifs en emploi)
Familles professionnelles (FAP)	L4X82 — Techniciens et agents de maîtrise chargés d'études socio-économiques · tension 4/5 (Élevée) L5X93 — Chargés d'études socio-économiques · tension 3/5 (Modérée) M1X80 — Techniciens d'étude et de développement en informatique · tension 4/5 (Élevée) M2X90 — Ingénieurs et cadres d'étude, recherche et développement en informatique et télécom · tension 5/5 (Très élevée) M2X93 — Experts et consultants en systèmes d'information · tension 5/5 (Très élevée)
Projets de recrutement (BMO 2026)	34 487
Jugés difficiles à recruter	50 %
Emplois saisonniers	1 %
Familles professionnelles (FAP)	L4X82 — Techniciens et agents de maîtrise chargés d'études socio-économiques L5X93 — Chargés d'études socio-économiques M1X80 — Techniciens d'étude et de développement en informatique M2X90 — Ingénieurs et cadres d'étude, recherche et développement en informatique et télécom M2X93 — Experts et consultants en systèmes d'information

Voies d'accès

MODALITÉS D'ACCÈS

Formation initiale, Formation continue, Apprentissage, Contrat de professionnalisation, VAE

PRÉREQUIS À L'ENTRÉE EN FORMATION

Pour accéder à la certification Expert en data science, les candidats doivent obligatoirement justifier d'un des prérequis suivants :

- être titulaire d'une certification professionnelle de niveau 6 en mathématiques, informatique, technologie, économie, finance ou analyse de données ;
- ou, être titulaire d'une certification professionnelle de niveau 5 avec une expérience professionnelle d'au moins 2

ans en mathématiques, informatique, technologie, économie, finance ou analyse de données.

Tout autre profil relève d'une procédure dérogatoire.

Pour tous les candidats, la procédure de sélection s'effectue au travers d'un formulaire de candidature validant les prérequis académiques ou dérogatoires, ainsi que l'adéquation du projet professionnel avec les objectifs de la formation. Sont collectées les pièces justificatives constituant le dossier d'admission (CV, éléments de motivation et le cas échéant, copie du diplôme, attestations employeur, et/ou test de positionnement). Si nécessaire, un entretien individuel de motivation est organisé.

Pour les candidats dont le français n'est pas la langue maternelle, un justificatif d'un niveau de français B2 minimum (niveau courant) sur l'échelle européenne du CECRL sera également demandé.

PRÉREQUIS À LA VALIDATION

Pour bénéficier de la certification professionnelle complète, le candidat doit valider les 4 blocs de compétences.

Le candidat souhaitant obtenir la certification par la VAE devra faire la preuve qu'il a acquis les compétences de chacun des blocs.

COMPOSITION DU JURY — FORMATION INITIALE, FORMATION CONTINUE, APPRENTISSAGE, CONTRAT DE PROFESSIONNALISATION, VAE

Le jury de délivrance de la certification est composé du responsable pédagogique de la filière Data et deux représentants du monde professionnel (data scientists /ingénieurs IA) extérieurs à l'organisme.

Financement en alternance (NPEC)

Niveau de prise en charge d'un contrat d'apprentissage par les OPCO, fixé par France compétences par branche professionnelle (CPNE).

Prise en charge usuelle	7 085 €
Selon la branche	6 700 € – 7 188 €
Prise en charge	Branches concernées
7 085 €	Valeur par défaut · 206 branches
7 188 €	CPNEF de la Publicité, CPNEFP Mutualité, CPNEFP de la branche des industries de produits alimentaires élaborés, CPNEFP de la branche professionnelle des Entreprises du courtage d'assurances et/ou réassurances, CPNIEFP dans la coopération agricole
7 162 €	CPNE chimie
6 700 €	CPNEFP du négoce des matériaux de construction

Base légale

Date de décision	27/03/2023
Durée de l'enregistrement	2 ans
Date d'échéance de l'enregistrement	27/03/2025
Date de dernière délivrance possible	27/03/2029

Statistiques

Année	Certifiés	Dont VAE	Insertion 6 mois	Métier 2 ans
2021	165	0	84 %	75 %
2020	72	0	91 %	70 %
2019	38	0	100 %	87 %

Liens avec d'autres certifications

ANCIENNES VERSIONS

RNCP34545

NOUVELLE(S) CERTIFICATION(S)

RNCP39775

Pour plus d'informations

LIEN VERS LE DESCRIPTIF

<https://openclassrooms.com/fr/paths/164-data-scientist>

Fiche officielle (France Compétences)

<https://www.francecompetences.fr/recherche/rncp/37431/>

Entrées et sorties de formation

Entrées en formation	193
Parcours menés à terme	151
Taux d'achèvement	78 %
Période	05/2023 – 03/2026

Entrées et sorties — Caisse des Dépôts (open data), tous organismes confondus.

Demande CPF — formations engagées

Dossiers CPF engagés	196
Montant total engagé	1 186 385 €
Coût moyen par dossier	6 053 €
Durée moyenne	610 h
Période	04/2023 – 03/2025
France compétences	42 %
France Travail	28 %
Reste à charge titulaire	19 %
Entreprise	4 %
OPCO	4 %
Régions	2 %
Compte d'engagement citoyen	0 %

Formations engagées via le CPF — Caisse des Dépôts (open data), tous organismes confondus.