

Expert en sciences des données (MS)

RNCP35197 · RNCP · Publiée

L'essentiel

Niveau de qualification	Niveau 7
Code(s) NSF	326m
Formacode(s)	31011, 31052
Code(s) ROME	M1802, M1403, M1801, M1805
Type d'enregistrement	Enregistrement sur demande
Date d'enregistrement	20/01/2021
Date d'échéance	20/01/2026

Certificateur(s)

INST NAT SCIENCES APPLIQUEES ROUEN · SIRET : 19760165100023

Résumé de la certification

OBJECTIFS ET CONTEXTE

Cette certification, accréditée par la CGE depuis sa mise en œuvre en 2015 (et réaccréditée cette année), a été créée afin de former des professionnels capables de répondre à l'enjeu actuel de l'exploitation des données massives. Ces données ouvrent de nouvelles perspectives, d'analyse, de recherche, d'étude... qui réclament des compétences en ressources humaines en mesure de les exploiter. Le défi n'est plus d'accéder à l'information mais de traiter la masse d'informations hétérogènes. Pour y répondre, le métier d'expert en Sciences des données (ou Data Scientist) est devenu une nécessité. Cette nécessité a été mise en exergue dans le rapport Villani sur la stratégie nationale de l'intelligence artificielle. Cette certification permet de répondre aux besoins des grands groupes industriels, des sociétés de conseils en systèmes d'information confrontées à l'épineuse question du traitement de l'information, des start-up, des institutions et collectivités publiques. Ces entreprises et institutions peuvent ainsi former leurs collaborateurs (actuels ou futurs) aux différentes techniques de la Science des données, leur permettant de créer de nouvelles opportunités sur le marché national et international.

La certification s'adosse aux savoir-faire des laboratoires LITIS et LMI de l'INSA Rouen Normandie et permet ainsi de répondre à de nombreux enjeux économiques et sociétaux, dans tous les secteurs, au niveau national comme au niveau régional, pour les entreprises, cabinets de conseils et services, start-up et PME, services publics et collectivités : exploiter les données massives pour en tirer un avantage compétitif à travers une meilleure connaissance des usagers et de leurs tendances, et ainsi mieux orienter les décisions et la stratégie de l'entreprise ou des politiques publiques. La certification permet également d'accompagner la transformation numérique et la convergence entre big data et intelligence artificielle.

ACTIVITÉS VISÉES

L'expert en sciences des données réalise la gestion et l'analyse pointue des données massives pour la stratégie et l'opérationnel de l'entreprise.

Dans ce cadre, il

réalise plusieurs activités :

- dans ce premier grand champ d'activité, il va analyser les données, en procédant à leur analyse dans le cadre de la modélisation des données massives hétérogènes ou non d'une entreprise puis il va élaborer un modèle prédictif de traitement de données afin de modéliser ces données.
- dans un second grand champ d'activités, il va s'atteler à extraire et collecter des données via une technique d'analyse et de stockage pour chaque donnée ; il va réaliser le traitement et l'exploitation des données collectées et veiller à l'amélioration continue du traitement des données structurées ou non structurées.

- enfin, il va avoir pour mission d'élaborer une stratégie d'exploitation et de valorisation de la donnée lui permettant d'exploiter les données hétérogènes de la structure, puis d'organiser et planifier des projets de la structure, de façon collaborative et avec les différents services de la structure concernés par la donnée.

COMPÉTENCES ATTESTÉES

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

C1.1. Maîtriser les outils d'analyse pour décrire les données en tenant compte du contexte.

C1.2. Identifier l'outil de modélisation statistique en fonction du type de donnée, structurée ou non structurée, afin de la traiter.

C1.3. Effectuer l'analyse exploratoire et la description des masses de données au travers d'une méthode statistique pour déterminer les problématiques de description, de projection, de visualisation et de regroupement des données en catégories homogènes.

C2.1. Sélectionner une méthodologie de traitement statistique au regard du problème posé, du type de donnée pour la modéliser.

C2.2. Concevoir et réaliser un prototype du modèle à partir de logiciels de programmation haut niveau pour montrer la faisabilité et la pertinence du modèle.

C2.3. Evaluer la performance du modèle au regard des contraintes de mise en production pour le mettre en œuvre.

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

C1.1. Maîtriser et utiliser les outils d'interrogation des bases de données pour extraire et collecter les données.

C1.2. Concevoir un système d'ingénierie de données en définissant une approche méthodologique de collecte, de stockage et d'interrogation pour l'analyse de masse de données réelles.

C2.1. Elaborer et mettre en place une stratégie de traitement distribué en ligne ou hors ligne de la donnée structurée et non structurée par un langage de programmation fonctionnel afin de les exploiter.

C2.2. Développer des implémentations au regard des données à traiter pour cadrer les algorithmes et performances prédictives.

C3.1. Analyser les retours pour mettre à jour l'exploitation et le stockage des données.

C3.2. Repérer le type de problème rencontré, d'apprentissage ou de fouille de données, en utilisant des algorithmes d'optimisation pour le résoudre.

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

C1.1. Maîtriser les fonctionnalités d'une architecture informatique robuste pour le stockage et le traitement distribué des données.

C1.2. Traduire un problème business en problème mathématiques/statistiques afin de définir une stratégie de valorisation des données pour exploiter les informations décisionnelles de la structure.

C1.3. Concevoir une stratégie de valorisation des données en proposant des « entrepôts de données » tenant compte de l'écosystème économique, réglementaire et des contraintes de sécurité liées au traitement des données.

C2.1 Formuler et présenter des analyses statistiques décisionnels à viser des différents services de la structure pour les accompagner dans leur prise de décision.

C2.2 Communiquer sur le potentiel des outils innovants de représentation, de modélisation et de visualisation pour contribuer à la plus-value et la prise de décisions utiles à la structure.

C2.3 Dimensionner les actions du projet en définissant les choix techniques et méthodologiques des interventions pour évaluer et planifier les ressources techniques, budgétaires et humaines.

C2.4 Communiquer sur la stratégie de valorisation des données de la structure pour sensibiliser les acteurs, accompagner au changement et améliorer le fonctionnement interne.

Blocs de compétences

RNCP35197BC01 — Modéliser les données massives d'une structure

LISTE DE COMPÉTENCES

C1.1. Maîtriser les outils

d'analyse pour décrire les données en tenant compte du contexte.

C1.2. Identifier l'outil de modélisation statistique en fonction du type de donnée, structurée ou non structurée, afin de la traiter.

C1.3. Effectuer l'analyse exploratoire et la description des masses de données au travers d'une méthode statistique pour déterminer les problématiques de description, de projection, de visualisation et de regroupement des données en catégories homogènes.

C2.1. Sélectionner une méthodologie de traitement statistique au regard du problème posé, du type de donnée pour la modéliser.

C2.2. Concevoir et réaliser un prototype du modèle à partir de logiciels de programmation haut niveau pour montrer la faisabilité et la pertinence du modèle.

C2.3. Evaluer la performance du modèle au regard des contraintes de mise en production pour le mettre en œuvre.

MODALITÉS D'ÉVALUATION

E1 - Mise en situation professionnelle reconstituée d'analyse de données publiquement disponibles : chaque candidat réalise un support (code, rapport) de son modèle d'analyse exploratoire et de représentation des données

E2 - Thèse professionnelle

E3 - Projet en groupe

E4 : QCM de 10 questions sur les concepts théoriques et pratiques.Exemple de thématiques:

- modèle SVM et Bayes
- modèle d'apprentissage supervisé vs non supervisé
- modèle convolutif et récurrent (IA)
- optimisation

RNCP35197BC02 — Collecter, structurer et traiter les données d'une structure

LISTE DE COMPÉTENCES

C1.1. Maîtriser et utiliser les outils d'interrogation des bases de données pour extraire et collecter les données.

C1.2. Concevoir un système d'ingénierie de données en définissant une approche méthodologique de collecte, de stockage et d'interrogation pour l'analyse de masse de données réelles.

C2.1. Elaborer et mettre en place une stratégie de traitement distribué en ligne ou hors ligne de la donnée structurée et non structurée par un langage de programmation fonctionnel afin de les exploiter.

C2.2. Développer des implémentations au regard des données à traiter pour cadrer les algorithmes et performances prédictives.

C3.1. Analyser les retours pour mettre à jour l'exploitation et le stockage des données.

C3.2. Repérer le type de problème rencontré, d'apprentissage ou de fouille de données, en utilisant des algorithmes d'optimisation pour le résoudre.

MODALITÉS D'ÉVALUATION

E2 - Thèse professionnelle

E3 - Projet en groupe

E5 : Mise en situation professionnelle reconstituée sur le stockage de la donnée:Chaque candidat réalise la conception, la configuration et l'implémentation de systèmes de stockage de la donnée via des outils d'interrogation des bases de données et à partir de données publiquement disponibles.

E6 : Mise en situation professionnelle reconstituée portant sur l'architecture de traitement distribuée de données:Chaque candidat conçoit et réalise une architecture de traitement distribuée des données à partir de données publiquement disponibles.

RNCP35197BC03 — Concevoir une stratégie de valorisation de la donnée

LISTE DE COMPÉTENCES

C1.1. Maîtriser les fonctionnalités d'une architecture informatique robuste pour le stockage et le traitement distribué des données.

C1.2. Traduire un problème business en problème mathématiques/statistiques afin de définir une stratégie de valorisation des données pour exploiter les informations décisionnelles de la structure.

C1.3. Concevoir une stratégie de valorisation des données en proposant des « entrepôts de données » tenant compte de l'écosystème économique, réglementaire et des contraintes de sécurité liées au traitement des données.

C2.1 Formuler et présenter des analyses statistiques décisionnels à viser des différents services de la structure pour les accompagner dans leur prise de décision.

C2.2 Communiquer sur le potentiel des outils innovants de représentation, de modélisation et de visualisation pour contribuer à la plus-value et la prise de décisions utiles à la structure.

C2.3 Dimensionner les actions du projet en définissant les choix techniques et méthodologiques des interventions pour évaluer et planifier les ressources techniques, budgétaires et humaines.

C2.4 Communiquer sur la stratégie de valorisation des données de la structure pour sensibiliser les acteurs,

accompagner au changement et améliorer le fonctionnement interne.

MODALITÉS D'ÉVALUATION

E2 - Thèse professionnelle

E3 - Projet en groupe

E7 - Mise en situation professionnelle réelle sur la réglementation, la sécurisation des données: Chaque apprenant(e) traite l'aspect réglementaire sur une problématique de l'entreprise dans laquelle il (elle) réalise son alternance

Secteur d'activité et type d'emploi

SECTEURS D'ACTIVITÉS

L'expert en sciences des données travaille sur des technologies numériques à haute valeur ajoutée intellectuelle et technique dans des entreprises de toutes tailles : grands groupes, sociétés de prestations (consultants), collectivités territoriales, PME, startups.

Il peut exercer son activité dans les secteurs suivants : les banques, les compagnies d'assurances, les sociétés de transport, les cabinets de conseil, les entreprises spécialisées dans l'énergie, le web marketing, les plateformes d'e-commerce, l'industrie, la R&D.

TYPE D'EMPLOIS ACCESSIBLES

L'expert en sciences des données (data scientist) détermine et met en œuvre une stratégie de collecte des données massives, complexes et hétérogènes (structurées ou non structurées), afin d'en retirer l'information pour une prise de décision optimale. Il

maîtrise les méthodes efficaces de collecte, stockage, mise en forme, visualisation, modélisation et traitement de ces données. Il

explore les données de multiples sources dispersées.

Suivant les secteurs et les structures, il

assure les postes et fonctions suivantes :

- Data analyst et data scientist : procède à des analyses sur les approches statistiques possibles en relation avec des enjeux métiers. Conçoit et déploie des modèles prédictifs. Maintient, améliore et conçoit des algorithmes d'extraction d'information. Traduit les données en leviers opérationnels, accompagne les développeurs jusqu'au déploiement en production des solutions et assure une veille technologique.

- Consultant valorisation des données : analyse les besoins clients big data, définit les architectures applicatives et techniques nécessaires au traitement des données, conçoit des modèles de données, participe au développement et à l'implémentation des algorithmes de traitement, à la mise en production, rédige la documentation associée et effectue la veille technologique sur les nouvelles tendances autour de l'exploitation des données.

- Business analyst/business developer : interface indispensable entre les départements opérationnels de l'entreprise et le service informatique, il

analyse le processus d'information et les stratégies au service de la prise de décision quotidienne de sa société afin d'en évaluer l'efficacité ou d'apporter les ajustements nécessaires.

Autres appellations de fonctions possibles :

- Ingénieur d'étude et développement de données massives

- Strategy manager/project manager

L'expert en sciences des données est un cadre détenant des compétences techniques, scientifiques et transversales faisant de lui

un vecteur facilitant la prise de décision en entreprise. Il est à même d'intervenir au plus haut niveau de la stratégie de l'entreprise.

Il est l'interlocuteur privilégié du Directeur des Systèmes Informatiques, comme du Directeur Marketing, du Directeur R&D ou Bureaux d'études, comme encore du Président de start up. Son niveau de responsabilité varie en fonction du positionnement de l'entreprise et de sa stratégie de valorisation des données.

Il

est amené à travailler dans un environnement multi-échelle, à l'interface entre différents services : R&D, informatique, marketing, ressources humaines, direction financière, direction générale, etc.

RÉGLEMENTATIONS D'ACTIVITÉ

La certification tient compte du règlement n° 2016/679, dit règlement général sur la protection des données (RGPD).

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)	Élevée (4/5) 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)	2 775 €
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) M1X81 — Techniciens de production, d'exploitation, d'installation, et de maintenance, support et services aux utilisateurs en informatique · tension 4/5 (Élevée) M2X90 — Ingénieurs et cadres d'étude, recherche et développement en informatique et télécom · tension 5/5 (Très élevée) M2X92 — Responsables et cadres de la production, de l'exploitation et de la maintenance informatique et télécom · tension 3/5 (Modérée) M2X93 — Experts et consultants en systèmes d'information · tension 5/5 (Très élevée)
Projets de recrutement (BMO 2026)	44 036
Jugés difficiles à recruter	49 %
Emplois saisonniers	2 %
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 M1X81 — Techniciens de production, d'exploitation, d'installation, et de maintenance, support et services aux utilisateurs en informatique M2X90 — Ingénieurs et cadres d'étude, recherche et développement en informatique et télécom M2X92 — Responsables et cadres de la production, de l'exploitation et de la maintenance 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, Candidature individuelle, VAE

PRÉREQUIS À L'ENTRÉE EN FORMATION

Les conditions d'admission sont fixées par la CGE dans le cadre de Mastère Spécialisé labélisé.

Sont donc recevables les candidatures de titulaires d'un des diplômes suivants :

- ? Diplôme d'ingénieur habilité par la Commission des Titres d'Ingénieur (liste CTI),
- ? Diplôme d'une école de management habilitée à délivrer le grade national de Master (liste CEFDG),
- ? Diplôme de 3ème cycle habilité par les autorités universitaires (DEA, DESS, Master...) ou diplôme professionnel de niveau BAC + 5,
- ? Diplôme de M1 ou équivalent, pour des auditeurs justifiant d'aux moins trois années d'expérience professionnelle,
- ? Titre inscrit au RNCP niveau 7,

? Diplôme étranger équivalent aux diplômes français exigés ci-dessus dans les domaines des mathématiques, de l'informatique et du traitement de l'information.

Par dérogation, pour 30 % maximum du nombre d'étudiants suivant la formation MASTERE SPECIALISE® concernée :

? Niveau M1 validé ou équivalent sans expérience professionnelle,

? Diplôme de L3 justifiant d'une expérience adaptée de 3 ans minimum,

? VAP, Valorisation des Acquis Professionnels

(10 % maximum de l'effectif de la formation).

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

2 Responsables de la formation et 4 professionnels extérieurs.

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	9 157 €
Selon la branche	7 900 € – 12 000 €
Prise en charge	Branches concernées
9 157 €	Valeur par défaut · 175 branches
12 000 €	CPNE Entreprises des activités du déchet et de la propreté urbaine, CPNE autoroutes, CPNE de l'ingénierie, des services informatiques et du conseil, CPNE de l'édition de livres, de l'édition phonographique et de l'édition de musique, CPNEF de la Branche des organismes de formation
8 696 €	CPNEFP des commerces et services de l'audiovisuel, de l'électronique et de l'équipement ménager, CPNEFP des jardineries et graineteries, CPNEFP du Négocio de l'ameublement
10 683 €	CPNEFP de l'enseignement privé indépendant, CPNEFP de la Restauration rapide
10 164 €	CPNE Casinos, CPNE de la branche de la Distribution Directe
9 992 €	CPNEF branche retraite complémentaire et prévoyance, CPNEFP Mutualité
9 500 €	CPNE Branche Caisse d'Epargne, CPNE de la branche Banque Populaire
9 000 €	CPNE Banque, CPNEFP des industries des tuiles et briques
8 500 €	CPNE de la Plasturgie et des Composites, CPNEFP de la métallurgie
11 447 €	CPNE des Télécoms
11 446 €	CPNE Travail Temporaire
11 363 €	CPNEFP de la Mutualité Sociale Agricole
10 163 €	CPNEF de l'Exploitation cinématographique et de la Distribution de films
10 000 €	CPNE des industries de santé (CPNEIS)
9 842 €	CPNEFP Grands Magasins/ Magasins Populaires
9 800 €	CPNEF de l'audiovisuel
8 775 €	CPNEFP de la branche des entreprises du bureau et du numérique (commerces et services)
8 701 €	CPNE FP Branche des magasins du Bricolage
8 613 €	CPNEFP de la Branche des Caisses régionales de Crédit Agricole et autres organismes
8 000 €	CPNE chimie

Base légale

Date de décision	20/01/2021
Durée de l'enregistrement	5 ans
Date d'échéance de l'enregistrement	20/01/2026

Statistiques

Année	Certifiés	Dont VAE	Insertion 6 mois	Métier 2 ans
2018	16	0	100 %	—
2017	12	0	83 %	64 %

Pour plus d'informations

Fiche officielle (France Compétences)

<https://www.francecompetences.fr/recherche/rncp/35197/>