

Data scientist

RNCP34545 · RNCP · Publiée

L'essentiel

Niveau de qualification	Niveau 7
Code(s) NSF	326
Formacode(s)	31026
Code(s) ROME	M1403
Type d'enregistrement	Enregistrement sur demande
Date d'enregistrement	30/03/2020
Date d'échéance	30/03/2023

Certificateur(s)

OPENCLASSROOMS · SIRET : 49386136300080

OPENCLASSROOMS · SIRET : 49386136300072

Partenaires habilités

9 organismes habilités, d'après la liste publiée par France compétences : 1 à former et organiser l'évaluation et 8 à former.

Organisme	SIRET	Habilitation
GRETA COTES NORMANDES	19501219000029	Former
GRETA DE LA REGION HAVRAISE	19760058800044	Former
GRETA DU CALVADOS	19140014200045	Former
GRETA ELBEUF VALLEE DE SEINE	19760030700049	Former
GRETA PORTES NORMANDES	19270016900027	Former
GRETA SUD-NORMANDIE	19610021800029	Former
LYCEE ENS GEN TECHNO GUSTAVE FLAUBERT	19760096800022	Former
LYCEE GENERAL ET TECHNOLOGIQUE CAMILLE JULLIAN	19330023300031	Former
OPENCLASSROOMS	49386136300072	Former et organiser l'évaluation

Résumé de la certification

OBJECTIFS ET CONTEXTE

L'accroissement des données disponibles et la multiplication des technologies permettant de les produire font de la valorisation des données un enjeu économique majeur et croissant.

La valorisation de ces données constitue, tant pour le secteur marchand que pour le secteur non marchand, un enjeu économique important, ouvrant pour les entreprises des perspectives de développement, de qualité et de compétitivité accrues. Les Big Data constituent un enjeu particulièrement stratégique dans les secteurs du numérique, de l'industrie ou de la santé.

L'essor des Big data entraîne des besoins en compétences expertes et fait émerger des nouveaux métiers.

Autour des data, et plus particulièrement des Big data, de nouveaux métiers émergent et se structurent.

A l'appui de nos partenaires, nous avons identifié les contours des emplois les plus recherchés autour des data et fait

le choix de les structurer autour de 4 certifications : data analyst, data architect, data scientist et ingénieur machine learning.

Cette certification vise à qualifier des Data scientists disposant des compétences expertes et transverses nécessaires à leur insertion immédiate sur le marché du travail.

ACTIVITÉS VISÉES

Le Data scientist est en mesure d'exercer les missions suivantes :

- agréger un ensemble de données pour ensuite les nettoyer et les structurer ;
- explorer ces données afin d'en déterminer la valeur et y repérer des motifs récurrents ou des irrégularités significatives ;
- produire de la valeur à partir des données par la création de modèles prédictifs ;
- modéliser le phénomène à l'origine des données pour répondre à un problème de l'entreprise ;
- évaluer et améliorer les performances d'un modèle d'apprentissage automatique (machine learning) ;
- rendre les données visuelles pour communiquer ses résultats ;
- établir des recommandations pour orienter les décisions stratégiques de l'entreprise.

COMPÉTENCES ATTESTÉES

La certification permet d'attester l'acquisition des compétences suivantes :

BLOC 1 – Prétraiter et analyser des données structurées pour répondre à un problème métier.

- Déterminer les objectifs d'une analyse de données à partir d'un problème métier.
- Effectuer des opérations de nettoyage sur des données structurées.
- Effectuer une analyse statistique univariée à partir de données structurées et nettoyées.
- Effectuer une analyse statistique multivariée à partir de données structurées et nettoyées.
- Représenter des données grâce à des graphiques clairs et pertinents.
- Formuler des préconisations pour un traitement automatisé des données.

BLOC 2 – Entraîner un modèle d'apprentissage automatique supervisé pour réaliser une analyse prédictive.

- Sélectionner et/ ou transformer les variables pertinentes pour la modélisation (feature engineering).
- Sélectionner et mettre en place un modèle d'apprentissage supervisé adapté à une problématique métier.
- Évaluer les performances d'un modèle d'apprentissage supervisé.
- Adapter les paramètres d'un modèle d'apprentissage supervisé afin de l'améliorer.

BLOC 3 – Entraîner un modèle d'apprentissage non supervisé adapté à une problématique de segmentation ou de réduction de données.

- Sélectionner, transformer et créer les variables pertinentes pour la modélisation (feature engineering).
- Sélectionner et mettre en place un modèle d'apprentissage non supervisé adapté à une problématique métier.
- Évaluer les performances d'un modèle d'apprentissage non supervisé.
- Adapter les paramètres d'un modèle d'apprentissage non supervisé afin de l'améliorer.

BLOC 4 – Prétraiter et analyser des données non structurées (texte, images) pour obtenir un jeu de données exploitable.

- Collecter des données répondant à des critères définis via une interface de programmation (API).
- Prétraiter des données textuelles non structurées pour obtenir un jeu de données exploitable.
- Prétraiter des données sous forme d'images non structurées pour obtenir un jeu de données exploitable.
- Réduire la dimension de données de grande dimension afin d'optimiser les temps de calcul.
- Représenter graphiquement des données à grandes dimensions afin d'en réaliser l'analyse exploratoire.

BLOC 5 – Présenter et déployer un modèle d'apprentissage automatique auprès de ses utilisateurs finaux.

- Déployer un modèle via une interface de programmation (API) dans le web.
- Réaliser un tableau de bord (dashboard) pour présenter son travail de modélisation.
- Réaliser la présentation orale d'une démarche de modélisation à un client interne/ externe.
- Rédiger une note méthodologique afin de communiquer sa démarche de modélisation.
- Assurer l'intégration du modèle auprès de collaborateurs en utilisant un logiciel de version de code.

BLOC 6 – Déployer un modèle d'apprentissage automatique à l'échelle en utilisant les technologies du Big data.

- Sélectionner les outils du Cloud permettant de disposer d'un environnement Big Data.
- Prétraiter, analyser et modéliser des données dans un environnement Big data en utilisant les outils du Cloud.
- Réaliser des calculs distribués sur des données massives en utilisant les outils adaptés.

Blocs de compétences

RNCP34545BC01 — Prétraiter et analyser des données structurées pour répondre à un problème métier.

LISTE DE COMPÉTENCES

- Déterminer les objectifs d'une analyse de données à partir d'un problème métier.
- Effectuer des opérations de nettoyage sur des données structurées.
- Effectuer une analyse statistique univariée à partir de données structurées et nettoyées.
- Effectuer une analyse statistique multivariée à partir de données structurées et nettoyées.
- Représenter des données grâce à des graphiques clairs et pertinents.
- Formuler des préconisations pour un traitement automatisé des données.

MODALITÉS D'ÉVALUATION

Chaque bloc de compétences est évalué sous la forme de projet, accompagné et validé par un mentor, puis présenté au jury du titre.

RNCP34545BC02 — Entraîner un modèle d'apprentissage automatique supervisé pour réaliser une analyse prédictive.

LISTE DE COMPÉTENCES

- Sélectionner et mettre en place un modèle d'apprentissage supervisé adapté à une problématique métier.
- Évaluer les performances d'un modèle d'apprentissage supervisé.
- Adapter les paramètres d'un modèle d'apprentissage supervisé afin de l'améliorer.

MODALITÉS D'ÉVALUATION

Chaque bloc de compétences est évalué sous la forme de projet, accompagné et validé par un mentor, puis présenté au jury du titre.

RNCP34545BC03 — Entraîner un modèle d'apprentissage non supervisé adapté à une problématique de segmentation ou de réduction de données.

LISTE DE COMPÉTENCES

- Sélectionner, transformer et créer les variables pertinentes pour la modélisation (feature engineering).
- Sélectionner et mettre en place un modèle d'apprentissage non supervisé adapté une problématique métier.
- Évaluer les performances d'un modèle d'apprentissage non supervisé.
- Adapter les paramètres d'un modèle d'apprentissage non supervisé afin de l'améliorer.

MODALITÉS D'ÉVALUATION

Chaque bloc de compétences est évalué sous la forme de projet, accompagné et validé par un mentor, puis présenté au jury du titre.

RNCP34545BC04 — Prétraiter et analyser des données non structurées (texte, images) pour obtenir un jeu de données exploitable.

LISTE DE COMPÉTENCES

- Collecter des données répondant à des critères définis via une interface de programmation (API).
- Prétraiter des données textuelles non structurées pour obtenir un jeu de données exploitable.
- Prétraiter des données sous forme d'images non structurées pour obtenir un jeu de données exploitable.
- Réduire la dimension de données de grande dimension afin d'optimiser les temps de calcul.
- Représenter graphiquement des données à grandes dimensions afin d'en réaliser l'analyse exploratoire.

MODALITÉS D'ÉVALUATION

Chaque bloc de compétences est évalué sous la forme de projet, accompagné et validé par un mentor, puis présenté au jury du titre.

RNCP34545BC05 — Présenter et déployer un modèle d'apprentissage automatique auprès de ses utilisateurs.

LISTE DE COMPÉTENCES

- Déployer un modèle via une interface de programmation (API) dans le web.
- Réaliser un tableau de bord (dashboard) pour présenter son travail de modélisation.
- Réaliser la présentation orale d'une démarche de modélisation à un client interne/ externe.
- Rédiger une note méthodologique afin de communiquer sa démarche de modélisation.
- Assurer l'intégration du modèle auprès de collaborateurs en utilisant un logiciel de version de code.

MODALITÉS D'ÉVALUATION

Chaque bloc de compétences est évalué sous la forme de projet, accompagné et validé par un mentor, puis présenté au jury du titre.

RNCP34545BC06 — Déployer un modèle d'apprentissage automatique à l'échelle en utilisant les technologies du Big data.

LISTE DE COMPÉTENCES

- Sélectionner les outils du Cloud permettant de disposer d'un environnement Big Data.
- Prétraiter, analyser et modéliser des données dans un environnement Big data en utilisant les outils du Cloud.
- Réaliser des calculs distribués sur des données massives en utilisant les outils adaptés.

MODALITÉS D'ÉVALUATION

Chaque bloc de compétences est évalué sous la forme de projet, accompagné et validé par un mentor, puis présenté au jury du titre.

Secteur d'activité et type d'emploi

SECTEURS D'ACTIVITÉS

Tous secteurs d'activité

TYPE D'EMPLOIS ACCESSIBLES

Data scientist ; Chef de projet Data

; Consultant Data ; Chief Data scientist ; Ingénieur data ; Ingénieur Big data

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)	Modérée (3/5) Moyenne pondérée par l'emploi ; les familles ci-dessous s'échelonnent de 3 à 4 sur 5.
Facteurs principaux	Manque de main-d'œuvre disponible
Salaire médian (mensuel)	2 718 €
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)
Projets de recrutement (BMO 2026)	2 271
Jugés difficiles à recruter	36 %
Emplois saisonniers	2 %
Familles professionnelles (FAP)	L4X82 — Techniciens et agents de maîtrise chargés d'études socio-économiques L5X93 — Chargés d'études socio-économiques

Voies d'accès

MODALITÉS D'ACCÈS

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

PRÉREQUIS À L'ENTRÉE EN FORMATION

classe préparatoire scientifique ou Bac + 3 en mathématiques ou master/ diplôme d'ingénieur en informatique.

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

3 personnes :

- Le/la responsable pédagogique
- 2 professionnels

COMPOSITION DU JURY — FORMATION CONTINUE

3 personnes :

- Le/la responsable pédagogique
- 2 professionnels

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	30/03/2020
Durée de l'enregistrement	3 ans
Date d'échéance de l'enregistrement	30/03/2023

Statistiques

Année	Certifiés	Dont VAE	Insertion 6 mois	Métier 2 ans
2019	29	0	100 %	—
2018	19	0	93 %	—

Liens avec d'autres certifications

NOUVELLE(S) CERTIFICATION(S)
RNCP37431

Pour plus d'informations

LIEN VERS LE DESCRIPTIF
<https://openclassrooms.com/fr>

Fiche officielle (France Compétences)

<https://www.francecompetences.fr/recherche/rncp/34545/>

Entrées et sorties de formation

Entrées en formation	75
Parcours menés à terme	86
Taux d'achèvement	115 %
Période	02/2022 – 06/2024

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

Demande CPF — formations engagées

Dossiers CPF engagés	246
Montant total engagé	817 329 €
Coût moyen par dossier	3 322 €
Durée moyenne	415 h
Période	04/2020 – 03/2023
France compétences	55 %
Reste à charge titulaire	18 %
France Travail	15 %
Entreprise	9 %
Régions	1 %
État	1 %
Compte d'engagement citoyen	0 %

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