

Expert en sciences des données

RNCP34262 · RNCP · Publiée

L'essentiel

Niveau de qualification	Niveau 7
Code(s) NSF	114g
Formacode(s)	31026
Code(s) ROME	M1403, M1802, M1805, M1806
Type d'enregistrement	Enregistrement sur demande
Date d'enregistrement	10/10/2019
Date d'échéance	10/10/2024

Certificateur(s)

DATA SCIENCETECH INSTITUTE · SIRET : 81023913700039

Partenaires habilités

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

Organisme	SIRET	Habilitation
AMTALENTS	9002108400001 4	Former
ECOLE NATIONALE SUPERIEURE D'ARTS ET METIERS (ENSAM)	1975347200001 0	Former et organiser l'évaluation

Résumé de la certification

OBJECTIFS ET CONTEXTE

Le métier visé est celui d'expert (« Data Scientist » en Anglais) dans le domaine des Sciences des Données. Ces dernières peuvent être de toute taille, jusqu'à la qualification de « massives » (« Big Data » en Anglais). Ces données partagent une complexité nécessitant l'utilisation de techniques mathématiques avancées avec leurs implémentations informatiques correspondantes. Ces implémentations sont localisées sur des infrastructures informatiques locales et/ou dans les nuages (« cloud computing » en Anglais).

Le Data Scientist conçoit des indicateurs pour l'aide à la décision à partir de sources de données multiples et dispersées, mises en place par des experts en ingénierie des données (« Data Engineers » en Anglais). Ces indicateurs sont construits à l'aide d'analyses mathématiques et de prototypes d'implémentations informatiques. Ces prototypes sont ensuite industrialisés à l'échelle de l'organisation par les Data Engineers.

Le métier de Data Scientist est donc spécialisé en modélisation mathématique -qu'elle soit de nature probabiliste/statistique, déterministe ou stochastique-, en informatique scientifique et en architecture des systèmes d'information. Ce spectre large de compétences débouche, en partie, sur ce que l'on appelle aujourd'hui et communément « l'intelligence artificielle – IA »

Référentiel Utilisé : Data Scientist – Référentiel Métiers OPIIEC - La fiche métier « Data Scientist » du référentiel métier OPIIEC a été choisie délibérément, car elle est la seule en France sur ce métier précis. Elle est par ailleurs attachée à la branche SYNTEC, qui est la principale source d'emploi de nos certifiés.

ACTIVITÉS VISÉES

Identifier les problématiques « décisionnelles » des directions métiers.

Définir une modélisation mathématique qui permette de répondre à la problématique.

Construire des outils d'analyse pour collecter les données de l'entreprise.
Rassembler l'ensemble des sources de données pertinentes (structurées ou non structurées) liées aux processus de production, de vente ou de la connaissance du client/usager.
Organiser, étudier et synthétiser ces sources de données sous forme de résultats exploitables.
Modéliser les comportements et en extraire de nouveaux usages utilisateurs.

COMPÉTENCES ATTESTÉES

Intitulé : Analyser, concevoir et développer des modélisations mathématiques pour enrichir les systèmes d'aide à la décision

Descriptif :

Explorer, décrire et interpréter des données dans leur aspect multidimensionnel.

Produire des traitements de données simples en utilisant les concepts de la statistique descriptive, en vue de présenter les résultats obtenus à l'aide de tableaux, de graphiques et d'indicateurs numériques.

Mesurer la liaison entre deux variables.

Choisir les outils et déterminer leurs limites afin de communiquer les conclusions et leur interprétation.

Produire des études statistiques en mettant en œuvre les techniques d'apprentissage automatique par la machine (machine learning) pour des données avec ou sans caractères temporels.

Intitulé : Concevoir et déployer des infrastructures informatiques distribuées pour stocker et traiter des données massives.

Descriptif :

Participer, avec des Data Engineers, à la rédaction d'un cahier des charges technique pour la conception et la mise en place d'une solution d'analyse des données volumineuses.

Exploiter un système distribué d'entrepôt de données structurées et non structurées, notamment via Apache Hadoop & Apache Spark.

Contribuer à la définition de l'architecture de stockage de données, en guidant les Data Engineers à prendre en compte les besoins de traitements algorithmiques de l'intelligence artificielle.

Concevoir et implémenter des algorithmes pour le traitement des grands volumes de données dans le respect des bonnes pratiques de l'ingénierie logicielle (incluant les approches SQL et NoSQL).

Exploiter des infrastructures informatiques distribuées en mode cloud sur la plateforme Amazon AWS pour y déporter des traitements algorithmiques de l'intelligence artificielle (préparation à la certification Amazon AWS Solution Architect – Associate)

Intitulé : Détecter et évaluer l'état de l'art de l'intelligence artificielle pour développer des projets innovants.

Descriptif :

Cibler et faire ressortir des modèles linéaires, non-linéaires, généralisés ou mixtes et construire des modèles complexes ou novateurs de prévisions en utilisant les techniques de modélisation, estimations, tests et diagnostics.

Identifier les variables ou composantes importantes d'un modèle afin de tester des hypothèses en utilisant les techniques de modélisation, estimations, tests et diagnostics.

Modéliser des données observées, les interpréter pour en faire de la prédiction en utilisant des outils appropriés et en tenant compte de leur nature aléatoire.

Déterminer et analyser des données complexes en utilisant les méthodes de fouille et de visualisation de séries temporelles et fonctionnelles, de textes (clustering, classification), de réseaux (statique ou dynamiques) ou d'images.

Intégrer les technologies complémentaires (web sémantique, modélisation multi-agents) pour l'hybridation des approches intelligence artificielle connectivistes (réseaux de neurones) et symboliques (moteurs de raisonnement).

Opérationnaliser une publication scientifique en implémentation informatique pour réaliser un prototype innovant.

B3C6 - Utiliser l'écosystème des solutions SAS pour la modélisation prédictive (préparation à la certification SAS® Certified Predictive Modeler Using SAS® Enterprise Miner™)

Intitulé : Conduire des projets complexes en intégrant les contraintes légales et les valeurs d'éthique pour diffuser les bonnes pratiques dans les organisations.

Descriptif

:

Intégrer les philosophies, problématiques et contraintes des réglementations françaises et européennes (CNIL, RGPD) sur la protection des données dans un projet Data Science.

Prendre la mesure de l'environnement international sur la réglementation de la protection des données, notamment aux Etats-Unis.

Intégrer les deux grandes familles de pratiques de gestion de projets informatiques complexes (en cascade et en

agilité) ainsi que les approches du « juste-à-temps ».

Blocs de compétences

RNCP34262BC01 — Analyser, concevoir et développer des modélisations mathématiques pour enrichir les systèmes d'aide à la décision

LISTE DE COMPÉTENCES

Explorer, décrire et interpréter des données dans leur aspect multidimensionnel.

Produire des traitements de données simples en utilisant les concepts de la statistique descriptive, en vue de présenter les résultats obtenus à l'aide de tableaux, de graphiques et d'indicateurs numériques.

Mesurer la liaison entre deux variables.

Choisir les outils et déterminer leurs limites afin de communiquer les conclusions et leur interprétation.

Produire des études statistiques en mettant en œuvre les techniques d'apprentissage automatique par la machine (machine learning) pour des données avec ou sans caractères temporels.

MODALITÉS D'ÉVALUATION

Examens écrits.

Projets d'application.

Mise en situation professionnelle réelle au travers du stage obligatoire d'application.

RNCP34262BC02 — Concevoir et déployer des infrastructures informatiques distribuées pour stocker et traiter des données massives.

LISTE DE COMPÉTENCES

Participer, avec des Data Engineers, à la rédaction d'un cahier des charges technique pour la conception et la mise en place d'une solution d'analyse des données volumineuses.

Exploiter un système distribué d'entrepôt de données structurées et non structurées, notamment via Apache Hadoop & Apache Spark.

Contribuer à la définition de l'architecture de stockage de données, en guidant les Data Engineers à prendre en compte les besoins de traitements algorithmiques de l'intelligence artificielle.

Concevoir et implémenter des algorithmes pour le traitement des grands volumes de données dans le respect des bonnes pratiques de l'ingénierie logicielle (incluant les approches SQL et NoSQL).

Exploiter des infrastructures informatiques distribuées en mode cloud sur la plateforme Amazon AWS pour y déporter des traitements algorithmiques de l'intelligence artificielle (préparation à la certification Amazon AWS Solution Architect – Associate)

MODALITÉS D'ÉVALUATION

Questionnaires à choix multiples

Projets d'applications

Examen externe de certification industrielle « Amazon AWS Solutions Architect – Associate »

Mise en situation professionnelle réelle au travers du stage obligatoire d'application.

RNCP34262BC03 — Détecter et évaluer l'état de l'art de l'intelligence artificielle pour développer des projets innovants

LISTE DE COMPÉTENCES

Cibler et faire ressortir des modèles linéaires, non-linéaires, généralisés ou mixtes et construire des modèles complexes ou novateurs de prévisions en utilisant les techniques de modélisation, estimations, tests et diagnostics.

Identifier les variables ou composantes importantes d'un modèle afin de tester des hypothèses en utilisant les techniques de modélisation, estimations, tests et diagnostics.

Modéliser des données observées, les interpréter pour en faire de la prédiction en utilisant des outils appropriés et en tenant compte de leur nature aléatoire.

Déterminer et analyser des données complexes en utilisant les méthodes de fouille et de visualisation de séries temporelles et fonctionnelles, de textes (clustering, classification), de réseaux (statique ou dynamiques) ou d'images.

Intégrer les technologies complémentaires (web sémantique, modélisation multi-agents) pour l'hybridation des approches intelligence artificielle connectivistes (réseaux de neurones) et symboliques (moteurs de raisonnement).

Opérationnaliser une publication scientifique en implémentation informatique pour réaliser un prototype innovant.

MODALITÉS D'ÉVALUATION

Questionnaires à choix multiples.

Projets d'applications.

Examen externe de certification industrielle « SAS® Certified Predictive Modeler Using SAS® Enterprise Miner™ ».

Mise en situation professionnelle réelle au travers du stage obligatoire d'application.

RNCP34262BC04 — Conduire des projets complexes en intégrant les contraintes légales et les valeurs d'éthique pour diffuser les bonnes pratiques dans les organisations

LISTE DE COMPÉTENCES

Intégrer les philosophies, problématiques et contraintes des réglementations françaises et européennes (CNIL, RGPD) sur la protection des données dans un projet Data Science.

Prendre la mesure de l'environnement international sur la réglementation de la protection des données, notamment aux États-Unis.

Intégrer les deux grandes familles de pratiques de gestion de projets informatiques complexes (en cascade et en agilité) ainsi que les approches du « juste-à-temps ».

MODALITÉS D'ÉVALUATION

Mise en situation en travaux dirigés et pratiques.

Questionnaires à choix multiples.

Mise en situation professionnelle réelle au travers du stage obligatoire d'application.

Secteur d'activité et type d'emploi

SECTEURS D'ACTIVITÉS

Toutes entreprises industrielles et de services, organismes publics, associations, sociétés de services et de conseil en ingénierie (y compris informatique).

Toutes ces structures ont comme point commun de générer et/ou d'avoir accès à des volumes importants de données numériques.

TYPE D'EMPLOIS ACCESSIBLES

Data Scientist

Machine Learning Engineer

Data Analyst

M1403 : Études et perspectives socio-économiques

M1802 : Expertise et support en systèmes d'information

M1805 : Études et développement informatique

M1806 : Conseil et maîtrise d'ouvrage en systèmes d'information

RÉGLEMENTATIONS D'ACTIVITÉ

Non Applicable

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, Candidature individuelle, VAE

PRÉREQUIS À L'ENTRÉE EN FORMATION

Formation initiale et continue à ce jour et VAE dans le futur

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

Cette information a été supprimée en application de l'article 17.1 du Règlement général sur la protection des données (RGPD)

COMPOSITION DU JURY — VAE

La composition du futur jury VAE répond aux critères spécifiés ci-après :

3 représentants de DSTI « Enseignement » :

- le Responsable VAE
- deux enseignants DSTI ayant des activités d'applications industrielles.

1 représentant DSTI « Conseil Scientifique et Pédagogique ».

3 représentants du monde industriel.

Ce jury est nommé par le Président de l'École, sur propositions respectives des Directeurs Généraux à l'Enseignement et aux Relation Extérieures.

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	8 741 €
Selon la branche	7 000 € – 8 950 €
Prise en charge	Branches concernées
8 741 €	Valeur par défaut · 203 branches

8 950 €	CPNE de l'ingénierie, des services informatiques et du conseil, CPNE des Télécoms, CPNEFP ICGV entreprises de la viande
8 500 €	CPNE Branche Caoutchouc, CPNE des industries de santé (CPNEIS)
8 800 €	CPNEF de l'audiovisuel
8 792 €	CPNE de l'édition de livres, de l'édition phonographique et de l'édition de musique
8 690 €	CPNEFP dans les établissements d'enseignement privés sous contrat
8 100 €	CPNE Banque
7 000 €	CPNEFP de la métallurgie

Base légale

Date de décision	10/10/2019
Durée de l'enregistrement	5 ans
Date d'échéance de l'enregistrement	10/10/2024

Statistiques

Année	Certifiés	Dont VAE	Insertion 6 mois	Métier 2 ans
2017	16	0	88 %	100 %
2016	22	0	86 %	100 %
2015	7	0	86 %	86 %

Pour plus d'informations

Fiche officielle (France Compétences)

<https://www.francecompetences.fr/recherche/rncp/34262/>

Entrées et sorties de formation

Entrées en formation	92
Parcours menés à terme	116
Taux d'achèvement	126 %
Période	02/2022 – 01/2026

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

Demande CPF — formations engagées

Dossiers CPF engagés	160
Montant total engagé	428 746 €
Coût moyen par dossier	2 680 €
Durée moyenne	635 h
Période	02/2020 – 09/2024
France compétences	78 %

France Travail	11 %
Entreprise	6 %
Reste à charge titulaire	2 %
État	1 %
Régions	1 %
Compte d'engagement citoyen	0 %

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