

Ingenieur en science des données spécialisé en infrastructure data ou en apprentissage automatique

RNCP39586 · RNCP · Publiée

L'essentiel

Niveau de qualification	Niveau 7
Code(s) NSF	326
Formacode(s)	31054
Code(s) ROME	M1801, M1802, M1805, M1806
Type d'enregistrement	Enregistrement sur demande
Date d'enregistrement	01/10/2024
Date d'échéance	01/10/2029

Certificateur(s)

YNOV · SIRET : 53056211500101

Partenaires habilités

17 organismes habilités, d'après la liste publiée par France compétences : 17 à former et organiser l'évaluation.

Organisme	SIRET	Habilitation
AIROBJECT	53970902200039	Former et organiser l'évaluation
AIX YNOV CAMPUS	52831294500033	Former et organiser l'évaluation
BORDEAUX YNOV CAMPUS	53275739000036	Former et organiser l'évaluation
CONNECT YNOV CAMPUS	92042238300017	Former et organiser l'évaluation
ECOLE EUROPEENNE DES METIERS DE L'INTERNET	52966176100021	Former et organiser l'évaluation
ISTEC	90023572200023	Former et organiser l'évaluation
LILLE YNOV CAMPUS	90310559100023	Former et organiser l'évaluation
LYON YNOV CAMPUS	79906530500036	Former et organiser l'évaluation
MONTPELLIER YNOV CAMPUS	90311060900034	Former et organiser l'évaluation
NANTES YNOV CAMPUS	80442673200033	Former et organiser l'évaluation

Liste complète des 17 partenaires habilités : <https://data.aktaconseil.fr/certifications/rncp39586-ingenieur-en-science-des-donnees-specialise-en-infrastructure-data-ou-en-apprentissage-automatique/partenaires>

Résumé de la certification

OBJECTIFS ET CONTEXTE

Le secteur stratégique de la data et de l'intelligence artificielle continue sa croissance. Les quantités de données numériques produites et traitées dans le monde sont exponentielles. L'enjeu est crucial pour les entreprises. Il s'agit de les valoriser et les exploiter pour produire des connaissances, faire des prédictions, analyser des comportements etc.

Cette certification professionnelle s'inscrit dans ce contexte et permet d'acquérir un socle de compétences

communes nécessaires pour collecter, stocker, traiter et valoriser des données. Elle vise aussi à fournir les compétences nécessaires pour piloter des projets complexes, intégrer les problématiques des entreprises et élaborer des solutions adaptées qui permettront d'atteindre les objectifs fixés. Ces compétences sont complétées de compétences liées à l'apprentissage automatique ou à l'infrastructure data :

- Véritable expert en apprentissage automatique, il conçoit des modèles de Machine Learning et des indicateurs pour aider la prise de décision dans la stratégie d'entreprise. Il utilise ces modèles prédictifs pour faire des recommandations et aider les dirigeants dans leur pilotage d'entreprise.
- ou expert en architecture des données, il conçoit et administre les systèmes de collecte de données. Il assure la récupération, la mise à disposition et la sécurisation des données afin de permettre leur exploitation.

ACTIVITÉS VISÉES

Collecte de données structurées et non structurées - Stockage des données structurées et non structurées - Structuration, transformation et enrichissement des données - Sécurisation des données - Analyse des données - Visualisation des données, interprétation et communication des résultats - Support et accompagnement des utilisateurs - Élaboration et cadrage du projet DATA - Pilotage du projet DATA - Management d'équipe - Veille, éthique et gouvernance des données

Pour la spécialisation en infrastructure data s'ajoutent : Analyse des besoins et définition de l'architecture - Conception et déploiement de l'infrastructure DATA - Supervision et exploitation de l'infrastructure DATA - Maintenance et sécurisation de l'infrastructure DATA

Pour la spécialisation en apprentissage automatique s'ajoutent : Analyse du besoin et résolution de problèmes complexes - Développement de modèles d'apprentissage automatique - Déploiement et automatisation des modèles d'apprentissage automatique

COMPÉTENCES ATTESTÉES

Élaborer une stratégie de collecte de données en définissant les données utiles et nécessaires pour répondre à une problématique, et en identifiant les sources des données afin de cadrer le travail à réaliser pour collecter les données ciblées.

Mettre en œuvre des techniques de collecte de données en exploitant les API externes et les bases de données disponibles, des techniques de web crawling et de web scraping afin de recueillir les données ciblées.

Automatiser la collecte de données en mettant en place des tâches planifiées et/ou des flux temps réel, en utilisant des logiciels d'automatisation afin de garantir l'actualisation des données.

Élaborer la stratégie de stockage des données avec un modèle de données adéquat en intégrant les différents types de données, l'utilisation envisagée (analyse, stockage, disponibilité, accessibilité) et le volume à stocker afin d'organiser le stockage des données.

Construire une base de données en sélectionnant la technologie (SQL, NoSQL), un système de gestion de base de données (SGBD) et/ou une solution de stockage BIG DATA, en assurant le paramétrage et l'implémentation afin de mettre en œuvre le modèle de données qui garantit la disponibilité et l'intégrité des données.

Sélectionner les technologies et les outils de traitement de données en identifiant les solutions existantes et en comparant leurs avantages et leurs inconvénients afin de traiter efficacement les données collectées.

Transformer les données à l'aide de langage de programmation ou en utilisant des outils dédiés (Talend, Spark...) afin d'obtenir des données nettoyées exploitables.

Développer un processus ETL en identifiant les bénéfices des technologies ETL (ex : facilité de développement), en exploitant la technologie ETL préalablement sélectionnée afin d'automatiser l'extraction, la transformation et le chargement de données.

Définir la politique de sécurisation des données en évaluant les risques, en qualifiant leur niveau de sensibilité, en identifiant les droits d'accès selon les rôles des différentes parties prenantes et en respectant les exigences légales (ex : RGPD) afin de garantir la bonne utilisation et l'intégrité des données.

Concevoir une architecture sécurisée et robuste, en intégrant des mesures de sécurité multicouches et des contrôles d'accès stricts, en mettant en place des solutions de protection des données (ex : chiffrement) pour sécuriser les données en transit et au repos, en veillant à l'anonymisation des données personnelles afin de répondre aux normes de protection de la vie privée et de sécurité des données.

Analyser les besoins métier et les enjeux exprimés par un commanditaire en réalisant des entretiens exploratoires et en récupérant les informations stratégiques nécessaires afin de cadrer le travail d'analyse des données à produire.

Définir les axes d'analyse et les métriques en identifiant les données à exploiter, celles disponibles et pertinentes pour traduire la problématique d'entreprise énoncée en problème numérique.

Réaliser des requêtes et des calculs en utilisant des outils de dashboarding, des tableurs, des requêtes SQL ou

scripts Python afin de produire une analyse des données préalablement collectées.

Élaborer des modèles statistiques et des tests d'hypothèses en modélisant des relations entre les variables, en évaluant la pertinence des résultats des simulations afin de valider ou réfuter des hypothèses.

Représenter les données en choisissant les modèles de représentation les plus adaptés (ex : histogramme, Heat map, nuage de points) et en utilisant des outils de représentation adaptés (ex : Office, power BI) afin de permettre la compréhension et l'exploitation des données par le public visé.

Présenter des recommandations, en préparant son discours et des arguments, en structurant son analyse sur les données représentées afin d'aider les décideurs à établir leurs stratégies.

Former les utilisateurs à l'utilisation des données et des outils de visualisation en analysant le besoin de montée en compétences et en élaborant des supports de formation et de sensibilisation adaptés afin de permettre aux utilisateur de maîtriser l'exploitation des données.

Rédiger la documentation technique d'utilisation du système d'analyse de données en identifiant le public concerné, en détaillant le fonctionnement du système d'analyse de données afin d'assurer la traçabilité et la transmission aux utilisateurs.

Définir les objectifs à atteindre et le périmètre du projet, en analysant les contraintes techniques et réglementaires, en étudiant le contexte et les enjeux afin de dimensionner le projet en termes de délai et budget.

Dimensionner le projet en évaluant la charge de travail et les ressources nécessaires (humaines, matérielles) au regard des exigences attendues et des contraintes préalablement définies afin d'estimer le temps et le budget nécessaires à la faisabilité du projet.

Rédiger la documentation projet, en identifiant les parties prenantes concernées, en prenant en compte l'ensemble des caractéristiques du projet, afin de clarifier et formaliser les attendus.

Planifier l'exécution du projet en organisant la répartition et l'ordonnancement des activités, le planning prévisionnel de réalisation et les ressources nécessaires à son exécution, en prenant en considération les personnes en situation de handicap afin de suivre les différentes phases du projet.

Suivre l'avancement du projet en mettant en place un outil de suivi (logiciel de suivi, tableau de bord), en définissant les indicateurs (qualitatifs et/ou quantitatifs) pour chaque jalon défini dans le planning, en réalisant des reportings et des comptes rendus de réunion afin d'anticiper les aléas éventuels.

Évaluer les besoins en compétences de l'équipe projet, en collaborant avec le service Ressources Humaines, en établissant un plan de développement des compétences et en orientant les membres de l'équipe vers des formations adaptées, afin de renforcer l'équipe responsable de mener à bien le projet DATA.

Piloter l'équipe projet en affectant les missions à réaliser, en prenant en compte les spécificités des membres de l'équipe, en intégrant les spécificités d'un contexte multiculturel, international, en utilisant les différentes techniques de communication et d'animation managériale pour favoriser le bon fonctionnement de l'équipe.

Procéder aux arbitrages et aux réajustements nécessaires à partir de l'analyse des écarts entre le prévisionnel et l'état du projet à date, en utilisant des outils d'aide à la décision (ex : logigramme) afin de garantir la performance du projet dans le respect des objectifs de qualité, coûts et délai.

Mettre en place un système de veille technologique et réglementaire en matière de science des données et d'Intelligence Artificielle à l'aide de recherches documentaires, de plateformes de partage, de webinars afin d'être alerté des évolutions qui impacteraient les pratiques métier.

Intégrer dans ses pratiques métier les enjeux en termes de données responsables, de responsabilité sociétale et environnementale (RSE), de sécurité, d'éthique et de confidentialité des données en se tenant informé des évolutions du cadre juridique, à travers une recherche documentaire ou en étant accompagné par des juristes afin d'agir dans le respect de la législation.

Pour la spécialisation en infrastructure data :

Analyser l'environnement du projet en recueillant les besoins métiers, les volumes de données à traiter, en réalisant un état des lieux des composants existants afin d'orienter le choix de conception de l'architecture DATA à mettre en œuvre.

Sélectionner l'ensemble des composants et technologies de l'infrastructure en étudiant les solutions existantes, en vérifiant leur compatibilité et les normes en vigueur afin de concevoir une architecture DATA correctement dimensionnée pour le projet.

Concevoir une architecture d'entrepôt de données en s'appuyant sur le cahier des charges, en sélectionnant les composants appropriés afin d'optimiser le stockage des données en termes de rapidité, de sécurité et d'accessibilité.

Mettre en place des pipelines DATA temps réel ou asynchrones à l'aide d'outils BIG DATA (ex : DATA brick, plateforme DBT Snowflake), Cloud et/ou On premise afin d'automatiser la transformation et la transmission des données.

Automatiser l'intégration et le déploiement des composants en utilisant des outils d'intégration et de développement

continu, afin d'industrialiser la mise en production de l'architecture DATA.

Mettre en place un système de supervision et d'alertes en déployant des outils de supervision, et en déterminant les indicateurs de suivi pertinents afin de s'assurer du bon fonctionnement des composants de l'architecture DATA.

Exploiter les systèmes et les équipements de l'infrastructure DATA en respectant les procédures d'administration et de maintien en condition opérationnelle afin de garantir l'intégrité et la disponibilité des données de l'organisation.

Rédiger la documentation fonctionnelle et technique en identifiant le public concerné et les objectifs attendus, en prenant en compte l'ensemble des caractéristiques de l'infrastructure DATA afin de transférer les procédures et les modalités de fonctionnement aux différents utilisateurs.

Élaborer le cahier de recette en rédigeant les scénarios de tests et les résultats attendus afin de détecter les anomalies de fonctionnement et les régressions éventuelles.

Résoudre un incident technique en investiguant la source du problème et en déployant une méthode de résolution afin de rétablir la disponibilité du service.

Pour la spécialisation en apprentissage automatique :

Analyser la problématique et le contexte d'un commanditaire en réalisant des entretiens exploratoires, des questionnaires et une analyse de l'existant afin de lui apporter une réponse appropriée.

Cadrer la stratégie de résolution du problème, en utilisant des algorithmes, en traduisant le problème en un problème d'optimisation afin de le résoudre avec les outils des modèles d'apprentissage automatique.

Sélectionner les technologies, les outils et les algorithmes en identifiant les différentes solutions disponibles et en comparant leurs avantages et leurs inconvénients afin de répondre à la problématique du commanditaire au regard des contraintes du projet.

Construire des variables en utilisant des langages de programmation (ex : Python, Scala, R, Julia...) en exploitant des bibliothèques d'analyse de données afin de fournir les meilleures variables au modèle d'apprentissage automatique.

Sélectionner les variables en identifiant les différentes méthodes de sélection de variables possibles, en utilisant des méthodologies d'apprentissage automatique afin d'optimiser la performance du modèle.

Entraîner un modèle d'apprentissage automatique à l'aide de bibliothèques (ex : Scikit-learn, XGBoost, TensorFlow, PyTorch) afin d'obtenir des modèles capables de prédictions sur de nouvelles données inconnues.

Optimiser la performance des modèles d'apprentissage automatique en modifiant les hyperparamètres et en analysant les prédictions afin de répondre au mieux à la problématique du commanditaire.

Sauvegarder le modèle d'apprentissage automatique entraîné à l'aide d'outils de sérialisation, virtualisation, conteneurisation, versioning afin de pouvoir le déployer dans des environnements de production.

Déployer des modèles d'apprentissage automatique en utilisant des API et des outils CI/CD afin de le mettre en production.

Superviser le système Machine Learning en sélectionnant des outils de monitoring et en les exploitant afin de détecter les dérives et les bugs du modèle d'apprentissage automatique.

Automatiser les tâches inhérentes au cycle de vie d'un système d'apprentissage automatique à l'aide de pipelines et des outils adaptés afin de maintenir la performance du modèle d'apprentissage automatique.

Blocs de compétences

RNCP39586BC01 — Collecter, transformer et sécuriser des données

LISTE DE COMPÉTENCES

Élaborer une stratégie de collecte de données en définissant les données utiles et nécessaires pour répondre à une problématique, et en identifiant les sources des données afin de cadrer le travail à réaliser pour collecter les données ciblées.

Mettre en œuvre des techniques de collecte de données en exploitant les API externes et les bases de données disponibles, des techniques de web crawling et de web scraping afin de recueillir les données ciblées.

Automatiser la collecte de données en mettant en place des tâches planifiées et/ou des flux temps réel, en utilisant des logiciels d'automatisation afin de garantir l'actualisation des données.

Élaborer la stratégie de stockage des données avec un modèle de données adéquat en intégrant les différents types de données, l'utilisation envisagée (analyse, stockage, disponibilité, accessibilité) et le volume à stocker afin d'organiser le stockage des données.

Construire une base de données en sélectionnant la technologie (SQL, NoSQL), un système de gestion de base de données (SGBD) et/ou une solution de stockage BIG DATA, en assurant le paramétrage et l'implémentation afin de mettre en œuvre le modèle de données qui garantit la disponibilité et l'intégrité des données.

Sélectionner les technologies et les outils de traitement de données en identifiant les solutions existantes et en comparant leurs avantages et leurs inconvénients afin de traiter efficacement les données collectées.

Transformer les données à l'aide de langage de programmation ou en utilisant des outils dédiés (Talend, Spark...) afin d'obtenir des données nettoyées exploitables.

Développer un processus ETL en identifiant les bénéfices des technologies ETL (ex : facilité de développement), en exploitant la technologie ETL préalablement sélectionnée afin d'automatiser l'extraction, la transformation et le chargement de données.

Définir la politique de sécurisation des données en évaluant les risques, en qualifiant leur niveau de sensibilité, en identifiant les droits d'accès selon les rôles des différentes parties prenantes et en respectant les exigences légales (ex : RGPD) afin de garantir la bonne utilisation et l'intégrité des données.

Concevoir une architecture sécurisée et robuste, en intégrant des mesures de sécurité multicouches et des contrôles d'accès stricts, en mettant en place des solutions de protection des données (ex : chiffrement) pour sécuriser les données en transit et au repos, en veillant à l'anonymisation des données personnelles afin de répondre aux normes de protection de la vie privée et de sécurité des données.

MODALITÉS D'ÉVALUATION

Mise en situation professionnelle réelle ou fictive.

: À partir de l'analyse d'une organisation réelle ou fictive de son choix, le candidat propose une stratégie de collecte et de transformation des données. Il remet au jury un dossier écrit comprenant : Une stratégie de collecte de données - Un exemple de collecte de données - Une méthode d'automatisation de collecte - Une stratégie de stockage des données - Un modèle de données - Une base de données - Une solution de stockage Big Data - Une présentation des outils et des technologies de traitement des données - Une présentation des données transformées, des méthodes et outils utilisés - Un processus ETL - Les solutions utilisés pour l'automatisation et l'orchestration du traitement des données

- Une politique de sécurité des données - Un schéma d'architecture de sécurité

RNCP39586BC02 — Analyser, organiser et valoriser des données

LISTE DE COMPÉTENCES

Analyser les besoins métier et les enjeux exprimés par un commanditaire en réalisant des entretiens exploratoires et en récupérant les informations stratégiques nécessaires afin de cadrer le travail d'analyse des données à produire.

Définir les axes d'analyse et les métriques en identifiant les données à exploiter, celles disponibles et pertinentes pour traduire la problématique d'entreprise énoncée en problème numérique.

Réaliser des requêtes et des calculs en utilisant des outils de dashboarding, des tableurs, des requêtes SQL ou scripts Python afin de produire une analyse des données préalablement collectées.

Élaborer des modèles statistiques et des tests d'hypothèses en modélisant des relations entre les variables, en évaluant la pertinence des résultats des simulations afin de valider ou réfuter des hypothèses.

Représenter les données en choisissant les modèles de représentation les plus adaptés (ex : histogramme, Heat map, nuage de points) et en utilisant des outils de représentation adaptés (ex : Office, power BI) afin de permettre la compréhension et l'exploitation des données par le public visé.

Présenter des recommandations, en préparant son discours et des arguments, en structurant son analyse sur les données représentées afin d'aider les décideurs à établir leurs stratégies.

Former les utilisateurs à l'utilisation des données et des outils de visualisation en analysant le besoin de montée en compétences et en élaborant des supports de formation et de sensibilisation adaptés afin de permettre aux utilisateurs de maîtriser l'exploitation des données.

Rédiger la documentation technique d'utilisation du système d'analyse de données en identifiant le public concerné, en détaillant le fonctionnement du système d'analyse de données afin d'assurer la traçabilité et la transmission aux utilisateurs.

MODALITÉS D'ÉVALUATION

Mise en situation professionnelle réelle ou fictive. À partir de l'analyse d'une organisation réelle ou fictive de son choix, le candidat propose une stratégie d'analyse, de mise en forme et de valorisation des données. Il remet au jury un dossier écrit comprenant : Une analyse du besoin - Une présentation d'un plan d'analyse - Une présentation des requêtes et des résultats - Une méthodologie de tests statistiques - La visualisation des résultats de l'analyse - Une présentation de recommandations - Un support de formation - Une documentation technique

RNCP39586BC03 — Elaborer et piloter un projet data

LISTE DE COMPÉTENCES

Définir les objectifs à atteindre et le périmètre du projet, en analysant les contraintes techniques et réglementaires, en étudiant le contexte et les enjeux afin de dimensionner le projet en termes de délai et budget.

Dimensionner le projet en évaluant la charge de travail et les ressources nécessaires (humaines, matérielles) au regard des exigences attendues et des contraintes préalablement définies afin d'estimer le temps et le budget nécessaires à la faisabilité du projet.

Rédiger la documentation projet, en identifiant les parties prenantes concernées, en prenant en compte l'ensemble des caractéristiques du projet, afin de clarifier et formaliser les attendus.

Planifier l'exécution du projet en organisant la répartition et l'ordonnancement des activités, le planning prévisionnel de réalisation et les ressources nécessaires à son exécution, en prenant en considération les personnes en situation de handicap afin de suivre les différentes phases du projet.

Suivre l'avancement du projet en mettant en place un outil de suivi (logiciel de suivi, tableau de bord), en définissant les indicateurs (qualitatifs et/ou quantitatifs) pour chaque jalon défini dans le planning, en réalisant des reportings et des comptes rendus de réunion afin d'anticiper les aléas éventuels.

Évaluer les besoins en compétences de l'équipe projet, en collaborant avec le service Ressources Humaines, en établissant un plan de développement des compétences et en orientant les membres de l'équipe vers des formations adaptées, afin de renforcer l'équipe responsable de mener à bien le projet DATA.

Piloter l'équipe projet en affectant les missions à réaliser, en prenant en compte les spécificités des membres de l'équipe, en intégrant les spécificités d'un contexte multiculturel, international, en utilisant les différentes techniques de communication et d'animation managériale pour favoriser le bon fonctionnement de l'équipe.

Procéder aux arbitrages et aux réajustements nécessaires à partir de l'analyse des écarts entre le prévisionnel et l'état du projet à date, en utilisant des outils d'aide à la décision (ex : logigramme) afin de garantir la performance du projet dans le respect des objectifs de qualité, coûts et délai.

Mettre en place un système de veille technologique et réglementaire en matière de science des données et d'Intelligence Artificielle à l'aide de recherches documentaires, de plateformes de partage, de webinars afin d'être alerté des évolutions qui impacteraient les pratiques métier.

Intégrer dans ses pratiques métier les enjeux en termes de données responsables, de responsabilité sociétale et environnementale (RSE), de sécurité, d'éthique et de confidentialité des données en se tenant informé des évolutions du cadre juridique, à travers une recherche documentaire ou en étant accompagné par des juristes afin d'agir dans le respect de la législation.

MODALITÉS D'ÉVALUATION

Mise en situation professionnelle réelle ou fictive. À partir de l'analyse d'une organisation réelle ou fictive de son choix, le candidat propose la gestion d'un projet Data. Il remet au jury un dossier écrit comprenant : Le cadrage du projet - Le dimensionnement du projet - La documentation projet - Le planning projet - Un outil de suivi de projet - Un tableau de bord - Un plan de développement des compétences - Les outils de communication et managériaux utilisés - La présentation d'un cas d'arbitrage rencontré au cours du projet - Une méthodologie de veille - Un plan d'actions relatif aux enjeux RSE, de sécurité, d'éthique et de confidentialité

RNCP39586BC04 — Spécialisation infrastructure data - Concevoir et opérer une infrastructure data

LISTE DE COMPÉTENCES

Analyser l'environnement du projet en recueillant les besoins métiers, les volumes de données à traiter, en réalisant un état des lieux des composants existants afin d'orienter le choix de conception de l'architecture DATA à mettre en œuvre.

Sélectionner l'ensemble des composants et technologies de l'infrastructure en étudiant les solutions existantes, en vérifiant leur compatibilité et les normes en vigueur afin de concevoir une architecture DATA correctement dimensionnée pour le projet.

Concevoir une architecture d'entrepôt de données en s'appuyant sur le cahier des charges, en sélectionnant les composants appropriés afin d'optimiser le stockage des données en termes de rapidité, de sécurité et d'accessibilité.

Mettre en place des pipelines DATA temps réel ou asynchrones à l'aide d'outils BIG DATA (ex : DATA brick, plateforme DBT Snowflake), Cloud et/ou On premise afin d'automatiser la transformation et la transmission des données.

Automatiser l'intégration et le déploiement des composants en utilisant des outils d'intégration et de développement continu, afin d'industrialiser la mise en production de l'architecture DATA.

Mettre en place un système de supervision et d'alertes en déployant des outils de supervision, et en déterminant les

indicateurs de suivi pertinents afin de s'assurer du bon fonctionnement des composants de l'architecture DATA.

Exploiter les systèmes et les équipements de l'infrastructure DATA en respectant les procédures d'administration et de maintien en condition opérationnelle afin de garantir l'intégrité et la disponibilité des données de l'organisation.

Rédiger la documentation fonctionnelle et technique en identifiant le public concerné et les objectifs attendus, en prenant en compte l'ensemble des caractéristiques de l'infrastructure DATA afin de transférer les procédures et les modalités de fonctionnement aux différents utilisateurs.

Élaborer le cahier de recette en rédigeant les scénarios de tests et les résultats attendus afin de détecter les anomalies de fonctionnement et les régressions éventuelles.

Résoudre un incident technique en investiguant la source du problème et en déployant une méthode de résolution afin de rétablir la disponibilité du service.

MODALITÉS D'ÉVALUATION

Mise en situation professionnelle réelle ou fictive. À partir de l'analyse d'un projet réel ou fictif, le candidat présente la stratégie de conception et de maintien d'une architecture DATA. À l'aide d'un support de présentation de son choix, il présente lors d'une soutenance orale : Un rapport d'analyse - Une présentation des composants de l'architecture DATA - Un schéma de données - Des pipelines de traitement de la donnée

- Un pipeline CI/CD - Une présentation d'un système de supervision - Une feuille de route d'exploitation - Une documentation technique - Un cahier de recettes et de tests - Une méthodologie d'investigation et de traitement d'un incident.

RNCP39586BC05 — Spécialisation apprentissage automatique - Concevoir et déployer des modèles d'apprentissage automatique

LISTE DE COMPÉTENCES

Analyser la problématique et le contexte d'un commanditaire en réalisant des entretiens exploratoires, des questionnaires et une analyse de l'existant afin de lui apporter une réponse appropriée.

Cadrer la stratégie de résolution du problème, en utilisant des algorithmes, en traduisant le problème en un problème d'optimisation afin de le résoudre avec les outils des modèles d'apprentissage automatique.

Sélectionner les technologies, les outils et les algorithmes en identifiant les différentes solutions disponibles et en comparant leurs avantages et leurs inconvénients afin de répondre à la problématique du commanditaire au regard des contraintes du projet.

Construire des variables en utilisant des langages de programmation (ex : Python, Scala, R, Julia...) en exploitant des bibliothèques d'analyse de données afin de fournir les meilleures variables au modèle d'apprentissage automatique.

Sélectionner les variables en identifiant les différentes méthodes de sélection de variables possibles, en utilisant des méthodologies d'apprentissage automatique afin d'optimiser la performance du modèle.

Entraîner un modèle d'apprentissage automatique à l'aide de bibliothèques (ex : Scikit-learn, XGBoost, TensorFlow, PyTorch) afin d'obtenir des modèles capables de prédictions sur de nouvelles données inconnues.

Optimiser la performance des modèles d'apprentissage automatique en modifiant les hyperparamètres et en analysant les prédictions afin de répondre au mieux à la problématique du commanditaire.

Sauvegarder le modèle d'apprentissage automatique entraîné à l'aide d'outils de sérialisation, virtualisation, containerisation, versioning afin de pouvoir le déployer dans des environnements de production.

Déployer des modèles d'apprentissage automatique en utilisant des API et des outils CI/CD afin de le mettre en production.

Superviser le système Machine Learning en sélectionnant des outils de monitoring et en les exploitant afin de détecter les dérives et les bugs du modèle d'apprentissage automatique.

Automatiser les tâches inhérentes au cycle de vie d'un système d'apprentissage automatique à l'aide de pipelines et des outils adaptés afin de maintenir la performance du modèle d'apprentissage automatique.

MODALITÉS D'ÉVALUATION

Mise en situation professionnelle réelle ou fictive. À partir de l'analyse d'un projet réel ou fictif, le candidat présente la stratégie de conception et de déploiement de modèles de Machine Learning. À l'aide d'un support de présentation de son choix, il présente lors d'une soutenance orale : Une analyse du besoin - Une présentation de la stratégie de résolution du problème - Une présentation des technologies et des outils sélectionnés - Un jeu de données exploitables pour le Machine Learning - Des méthodes de sélection de variables - Un entraînement d'un modèle d'apprentissage automatique - Une méthode d'optimisation des modèles d'apprentissage automatique - Une méthode de sauvegarde - Un processus CI/CD - Un système de monitoring de la performance - Un système de collecte de données

Secteur d'activité et type d'emploi

SECTEURS D'ACTIVITÉS

L'ingénieur en science des données, qu'il soit spécialisé en infrastructure data ou en apprentissage automatique, peut intégrer tous types de structures (entreprise, organisation, administration) de tous secteurs confondus (industrie, énergie, télécommunications, acteurs du e-commerce, constructeurs informatiques, administration, banques, services, transport, santé, éducation, ...).

TYPE D'EMPLOIS ACCESSIBLES

Data Scientist - Data Engineer - Machine Learning Engineer - Consultant en Data - Consultant informatique

RÉGLEMENTATIONS D'ACTIVITÉ

Il n'y a pas de réglementation particulière pour exercer ce métier qu'il concerne la spécialisation infrastructure data ou la spécialisation apprentissage automatique. En revanche, ces experts en science des données sont des acteurs clés du respect par l'entreprise de la législation française et européenne sur la protection des données (RGPD, Patriot Act, CNIL).

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)	2 781 €
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)	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)	41 765
Jugés difficiles à recruter	50 %
Emplois saisonniers	2 %
Familles professionnelles (FAP)	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, VAE

PRÉREQUIS À L'ENTRÉE EN FORMATION

Le dispositif de certification est accessible aux titulaires d'une certification de niveau 4 pour un parcours en 5 ans.

Pour une admission en 3ème année : être titulaire d'une certification de niveau 5 dans un domaine en lien avec la certification.

Pour une admission en 4ème année : être titulaire d'une certification de niveau 6 dans un domaine en lien avec la certification.

Toute demande ne répondant pas aux prérequis de la formation est étudiée par une commission placée sous l'autorité du certificateur.

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

Le jury de certification est composé de 5 membres : 2 représentants du groupe YNOV, 3 professionnels externes exerçant dans le métier visé ou l'un des métiers visés par la certification ou supervisant des personnes qui l'exercent.

COMPOSITION DU JURY — VAE

Le jury VAE est composé de 5 membres : 2 représentants du groupe YNOV, 3 professionnels externes exerçant dans le métier visé ou l'un des métiers visés par la certification ou supervisant des personnes qui l'exercent.

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	9 350 €
Selon la branche	8 195 € – 9 996 €
Prise en charge	Branches concernées
9 350 €	Valeur par défaut · 190 branches
9 996 €	CPNE Casinos, CPNE Travail Temporaire, CPNE de l'édition de livres, de l'édition phonographique et de l'édition de musique, CPNE des Télécoms, CPNEF de l'audiovisuel
8 500 €	CPNE des MFR, CPNE des Métiers de la Transformation des Grains, CPNE des Vins et Spiritueux, CPNE des centres de gestion agréés et habilités agricoles, CPNEFP Transformation Laitière
9 992 €	CPNEF branche retraite complémentaire et prévoyance, CPNEFP Mutualité
8 837 €	CPNE du Commerce de détail et de gros à prédominance alimentaire
8 775 €	CPNEFP de la branche des entreprises du bureau et du numérique (commerces et services)
8 750 €	CPNEFP Commerce à distance
8 195 €	CPNE FP Branche des magasins du Bricolage

Base légale

Date de décision	01/10/2024
Durée de l'enregistrement	5 ans
Date d'échéance de l'enregistrement	01/10/2029
Date de dernière délivrance possible	01/10/2033

Statistiques

Année	Certifiés	Dont VAE	Insertion 6 mois	Métier 2 ans
--------------	------------------	-----------------	-------------------------	---------------------

2022	52	0	92 %	—
2021	61	0	91 %	79 %
2020	59	0	85 %	69 %

Liens avec d'autres certifications

ANCIENNES VERSIONS

RNCP35078

CORRESPONDANCES (PAR BLOCS)

RNCP34408, RNCP35078, RNCP36398, RNCP37137, RNCP37172, RNCP37431, RNCP37624, RNCP37843

Pour plus d'informations

LIEN VERS LE DESCRIPTIF

École de formation Informatique - débutant à technicien | Ynov

Fiche officielle (France Compétences)

<https://www.francecompetences.fr/recherche/rncp/39586/>

Se former en apprentissage

Formations en apprentissage	57
Organismes gestionnaires (CFA)	9
Organismes formateurs	14
Régions couvertes	10
Départements couverts	13
Organismes certifiés Qualiopi	14
Formations à distance	1
Sessions	de 09/2023 à 09/2029

Catalogue des formations en apprentissage — réseau des Carif-Oref. Sans frais pour l'alternant : formation financée par l'OPCO de la branche (NPEC).