



Contact

Responsable de la Formation Professionnelle

pro.strasbourg-europe@icam.fr

03 90 40 09 63



Accessibilité universelle

Formation accessible aux personnes en situation de handicap

Pour tout renseignement, contacter referent-handicap.strasbourg-europe@icam.fr



Présentation générale

La formation 'Data Scientist : de l'intégration à l'industrialisation des données' prépare les stagiaires à la certification validant les cinq blocs de compétences nécessaires à la mise en œuvre d'un projet d'analyse de données de bout en bout :

- Analyse des besoins métiers de la structure en matière d'accompagnement décisionnel afin d'établir une stratégie globale de transformation de la donnée prenant en compte les contraintes sociales, économiques et environnementales.
- Pilotage de la remontée et de l'intégrité des données en collaboration avec les équipes techniques en vue de leur exploitation.
- Exploration et analyse des ensembles de données pour fournir des connaissances ("insights") exploitables et soutenir la prise de décision.
- Développement des modèles par l'utilisation des techniques avancées de statistiques et d'apprentissage automatique pour fournir des informations et accompagner le processus de décision.
- Pilotage des projets en communiquant avec les parties prenantes et en guidant les membres de l'équipe pour atteindre des objectifs spécifiques.



Suites possibles après cette formation

- **Métiers :** Reconversion dans le domaine du numérique, et plus spécifiquement dans les métiers de l'analyse de données. L'analyse de données a pour objectif de permettre aux entreprises et aux organisations de valoriser leur capital 'données', et de les administrer, les analyser et les visualiser. Les métiers : Data Scientist, Data Analyst, Data Miner, Data Manager, Architecte Big Data.
- **Formations :** Possibilités de suivre des formations spécialisées afin d'approfondir ces connaissances dans des domaines spécifiques tel que le Report BI, le Machine Learning, le Deep Learning, le Data Mining, le Data Engineering, Data Manager et l'Architecte Big Data.



« Opération financée par la Région Grand-Est »



Icam, site de Strasbourg-Europe, Association de Gestion

Association loi 1908 enregistrée au registre des Associations du Tribunal d'Instance de Schiltigheim sous vol. 40, folio 41, enregistrée en tant qu'organisme de formation sous le n° 42 67 04761 67 * auprès de la préfecture, dont le siège est sis 2 rue de Madrid CS 20013 - 67012 Strasbourg Cedex, identifiée sous les n° SIRET 51773337400022 APE 8542Z.

*ce numéro ne vaut pas agrément de l'état.

Page 1 sur 8 – Dernière mise à jour 15/02/2024



Blocs de compétences et capacités professionnelles

- **Bloc de compétence n°1 : Analyse des besoins métiers de la structure en matière d'accompagnement décisionnel afin d'établir une stratégie globale de transformation de la donnée prenant en compte les contraintes sociales, économiques et environnementales.**
 - Capacité C1 : Traduire de manière analytique le besoin métier en rédigeant un cahier des charges en collaboration avec les parties prenantes pour fixer les objectifs attendus.
 - Capacité C2 : Proposer une conception technique et technologique grâce à un état de l'art prenant en compte l'ensemble des contraintes (temporelles, budgétaires, techniques, managériales, juridiques...) afin de répondre au cahier des charges.
- **Bloc de compétence n°2 : Pilotage de la remontée et l'intégrité des données en collaboration avec les équipes techniques en vue de leur exploitation**
 - Capacité C3 : Cartographier les flux de données et leur structure par des modèles logiques afin de les préparer aux processus d'intégration.
 - Capacité C4 : Superviser le développement de systèmes d'acquisition, de transfert et de stockage des données en accord avec les besoins techniques et juridiques (RGPD) en collaborant avec des équipes techniques de façon transparente et inclusive pour pérenniser les flux de données.
 - Capacité C5 : Assurer l'intégrité des données tout au long de leur cycle de vie par le développement de pipelines ETL ("Extract, Transform, Load") afin de réduire les erreurs potentielles et d'assurer une gestion pérenne des flux de données.
- **Bloc de compétence n°3 : Exploration et analyse des ensembles de données pour fournir des connaissances ("insights") exploitables et soutenir la prise de décision**
 - Capacité C6 : Analyser des ensembles de données par l'application de techniques statistiques afin de mettre en valeur des tendances, des corrélations et des modèles significatifs.
 - Capacité C7 : Transmettre des informations issues des données transformées par le développement d'outils de visualisation tels que des tableaux de bords, des infographies et des KPI pour accompagner la prise de décision de l'entreprise.
- **Bloc de compétence n°4 : Développement des modèles par l'utilisation des techniques avancées de statistiques et d'apprentissage automatique pour fournir des informations et accompagner le processus de décision**
 - Capacité C8 : Déterminer une représentation des données à l'aide de techniques d'ingénierie de caractéristiques ("feature engineering") pour entraîner un modèle répondant à la problématique.
 - Capacité C9 : Entraîner un modèle d'apprentissage supervisé ou non-supervisé à l'aide d'un ensemble de données représentées et en ajustant ses hyper-paramètres pour améliorer ses performances et sa capacité à répondre de manière généralisée à une problématique.



« Opération financée par la Région Grand-Est »



Icam, site de Strasbourg-Europe, Association de Gestion

Association loi 1908 enregistrée au registre des Associations du Tribunal d'Instance de Schiltigheim sous vol. 40, folio 41, enregistrée en tant qu'organisme de formation sous le n° 42 67 04761 67 * auprès de la préfecture, dont le siège est sis 2 rue de Madrid CS 20013 - 67012 Strasbourg Cedex, identifiée sous les n° SIRET 51773337400022 APE 8542Z.

*ce numéro ne vaut pas agrément de l'état.

Page 2 sur 8 – Dernière mise à jour 15/02/2024

- Capacité C10 : Établir une stratégie pour la création d'un modèle d'apprentissage en concevant ou réutilisant des modèles pré-entraînés (utilisant le transfert d'apprentissage), et entraîner ces modèles dans le but de résoudre une problématique métier spécifique.
 - Capacité C11 : Évaluer les modèles d'apprentissage établis en utilisant des techniques de validations croisées et des métriques de performances afin de déterminer le modèle le plus adapté à la problématique métier.
 - Capacité C12 : Communiquer les résultats des modèles en détaillant les liens avec les objectifs et en faisant preuve d'éthique et de transparence sur leurs limites afin d'accompagner les processus décisionnels.
 - Capacité C13 : Assurer une veille technologique par des stratégies de surveillance des tendances et des innovations dans le domaine des sciences de la donnée afin de développer des idées novatrices et accompagner la résolution de problèmes.
- **Bloc de compétence n°5 : Pilotage des projets en communiquant avec les parties prenantes et en guidant les membres de l'équipe pour atteindre des objectifs spécifiques**
 - Capacité C14 : Collaborer avec des équipes pluridisciplinaires en appliquant les outils du management dans le respect des valeurs éthiques et inclusives afin d'améliorer la performance individuelle et collective.
 - Capacité C15 : Organiser les activités et les ressources par la mise en place d'une stratégie de gestion afin d'assurer la conduite du projet dans les critères coût-qualité-délais du cahier des charges.

Objectifs pédagogiques de la formation

- Permettre une reconversion vers les métiers du numérique.
- Permettre aux stagiaires de maîtriser la chaîne de valeur de la donnée depuis l'intégration jusqu'à sa visualisation, en passant par l'analyse.
- Donner aux stagiaires les compétences leur permettant d'être opérationnels dans la mise en œuvre de projets d'analyse de données : gestion de projet, technologies logicielles, méthodologie d'analyse des données.



« Opération financée par la Région Grand-Est »



Icam, site de Strasbourg-Europe, Association de Gestion

Association loi 1908 enregistrée au registre des Associations du Tribunal d'Instance de Schiltigheim sous vol. 40, folio 41, enregistrée en tant qu'organisme de formation sous le n° 42 67 04761 67 * auprès de la préfecture, dont le siège est sis 2 rue de Madrid CS 20013 - 67012 Strasbourg Cedex, identifiée sous les n° SIRET 51773337400022 APE 8542Z.

*ce numéro ne vaut pas agrément de l'état.

Page 3 sur 8 – Dernière mise à jour 15/02/2024



Modalités pratiques

- **Public** : Demandeurs d'emplois ayant une formation scientifique poussée (au minimum niveau Bac +3 avec de bonnes connaissances en mathématiques et statistiques, Bac+5 à Bac+8 en sciences, physique, biologie).
- **Prérequis** : Compétences en mathématiques, statistiques, logique, bases de données.
- **Conditions d'accès** : Recrutement en 3 étapes :
 - Participation à un réunion d'information
 - Test d'évaluation des compétences en mathématiques, statistiques, logique, bases de données,
 - Entretien de motivation
- **Conditions de démarrage** : Nombre de participants de 8 personnes (minimum) à 13 personnes (maximum) pour l'ensemble de la formation
- **Lieu** : Icam site de Strasbourg Europe, 2 Rue de Madrid, 67300 SCHILTIGHEIM
- **Nombre d'heures** : 1200 heures réparties de la manière suivante :
 - 600 heures en centre de formation
 - 600 heures en Entreprise
- **Horaires** : 9h-12h30 et 13h30-17h00 (accueil à partir de 8h30)
- **Type de formation** : Formation en présentiel
- **Date début de formation** : 01 octobre 2024
- **Examen de validation des compétences** : Chaque module de la formation ne peut être validé que si la présence du stagiaire est attestée sur l'ensemble des sessions de formation du module. La formation ne peut être validée que si l'ensemble des modules sont validés. L'atteinte des compétences de la formation est attestée par :
 - La présence à l'ensemble des sessions de formation.
 - La préparation d'un dossier professionnel sur la mission en entreprise et un support de présentation.
 - La présentation orale avec le support (diaporama) de la mission réalisée pendant la séquence en entreprise. La soutenance se déroule uniquement en présence du jury.
 - Suivie d'un entretien professionnel avec le jury, afin de s'assurer de la maîtrise des compétences.
 - Seul le jury, en raison de sa compétence spécifique, est habilité à délivrer le titre après une évaluation rigoureuse du candidat selon une grille de critères d'évaluation.
- **Validation** : Sous réserve de la réussite aux examens, une attestation de compétences délivrée par un jury d'au moins trois professionnels issus du domaine de l'informatique et/ou de l'industrie, dont au moins 50% de Data Scientists professionnels, et externes à la structure d'accueil et de la structure de certification et la Certification Professionnelle RNCP (en cours d'inscription).



« Opération financée par la Région Grand-Est »





Moyens pédagogiques

- **Alternance de théorie, travaux pratiques et projets d'études.**
- **Accompagnement :** Un module 'Techniques de Recherche d'Emploi' (TRE) vous permet de développer vos capacités professionnelles et de vous préparer à valoriser les compétences acquises durant la formation. Un accompagnement personnalisé vous est proposé, dont les modalités seront précisées par le formateur lors de la première séance du module.



Programme détaillé

- **Stratégie des besoins clients appliquées au data science** 42 h
 - **Accueil/Administratif ; prise en main des outils** 7 h
 - **Analyse Besoin** 21 h
 - Être capable de discuter des notions des spécificités du marché de l'analyse de données.
 - Savoir comment de manière analytique construire un cahier des charges.
 - Comprendre le besoin client.
 - Démarche de questionnement/écoute auprès des parties prenantes.
 - Déterminer des objectifs précis.
 - Capacité de synthèse.
 - Revue de projets d'entreprise/cas d'école pour acquérir des réflexes de conceptions avec les contraintes du cahier des charges.
 - **Architecture des systèmes de données** 14 h
 - Cours théorique garantir la cohérence des systèmes avec les besoins d'intégration des données, une compréhension approfondie de l'architecture des systèmes de données, y compris les bases de données distribuées et les entrepôts de données.
- **Intégration des données** 105 h
 - **Programmation SQL** 14h
 - Connaître les principes de fonctionnement des bases de données relationnelles.
 - Être capable de programmer des requêtes SQL complexes.
 - Être capable de concevoir une base de données.
 - Être capable de restructurer et d'optimiser une base de données.
 - Avoir des connaissances de bases dans la mise en œuvre d'une base de données
 - Conception des schémas relationnels : les dépendances entre données, la démarche de conception, les formes normales, le processus de normalisation.
 - **Outils pour le traitement des données** 21 h
 - Être capable de cartographier les flux de données et leur structure.
 - Être capable d'appréhender les modèles de données en vue de leur traitement.
 - Être capable de préparer des données pour le traitement.
 - Être capable d'extraire et de transformer des données à l'aide de l'outil de Talend (ETL).



« Opération financée par la Région Grand-Est »



- **Alteryx** 14h
 - Être capable de préparer des données pour le traitement.
 - Être capable d'extraire et transformer des données à l'aide de l'outil de Alteryx (ETL).
- **Projet 'Intégration des données'** 28 h
 - Être capable de chercher un jeu de données conséquent.
 - Être capable d'effectuer la conception d'un MCD.
 - Être capable de concevoir la base relationnelle correspondante.
 - Être capable d'implémenter le processus ETL à l'aide d'un outil adéquat pour répondre à la problématique.
- **Web + Webscrapping** 28 h
 - Introduction Web Html/CSS/XML pour objectif webscrapping.
 - Activités python pour le webscrapping.
 - BeautifulSoup.
- **Exploration et analyse de données décisionnelle** 196 h
 - **Introduction à la Data Science** 7h
 - Être capable d'explicité les spécificités des projets données en entreprise.
 - Connaître les enjeux de la maîtrise de la donnée pour les entreprises.
 - Avoir des notions des spécificités du marché de l'analyse de données.
 - Connaître les spécificités des projets données en entreprise.
 - **FOAD (Python, Base de données, ML)** 7h
 - **Programmation Objet Python** 14h
 - Être familier avec les concepts de la programmation objet
 - Être capable de programmer une application dans un langage orienté objet.
 - Être capable de programmer une application simple en python.
 - Être capable de mettre en œuvre de la programmation script, procédurale, objet, en python.
 - Être capable d'utiliser les structures de données spécifiques au langage.
 - **Python pour la science des données** 21h
 - Réussir de faire de python votre boîte à outil afin de : se connecter à une base de données, lire des données, effectuer des opérations d'analyse de base sur les données (tri, recherche, extraction, transformation simple...).
 - **Statistiques pour le big data** 35 h
 - Être capable d'analyser des données à l'aide de méthodes statistiques (PCA, Anova, Multivarié, etc....).
 - Comprendre comment cette analyse peut être exploitée dans la conception du modèle répondant à la problématique.
 - Être capable d'utiliser des approches statistiques adéquates pour effectuer des opérations de feature engineering (exemple : passage au log).
 - Être capable d'appliquer les modèles statistiques mis en œuvre dans le big data.
 - Être capable de manipuler les algorithmes associés à ces modèles.
 - Être capable de choisir un modèle statistique adapté aux besoins d'un projet big data.



« Opération financée par la Région Grand-Est »



Icam, site de Strasbourg-Europe, Association de Gestion

Association loi 1908 enregistrée au registre des Associations du Tribunal d'Instance de Schiltigheim sous vol. 40, folio 41, enregistrée en tant qu'organisme de formation sous le n° 42 67 04761 67 * auprès de la préfecture, dont le siège est sis 2 rue de Madrid CS 20013 - 67012 Strasbourg Cedex, identifiée sous les n° SIRET 51773337400022 APE 8542Z.

*ce numéro ne vaut pas agrément de l'état.

Page 6 sur 8 – Dernière mise à jour 15/02/2024

- **Base de l'Apprentissage Automatique** 35 h
 - Être capable d'effectuer du Feature Engineering.
 - Comprendre les différents types de modèles de Machine Learning.
 - Être capable d'entraîner un modèle.
 - Connaître et être capable d'expliquer les limites d'un modèle.
 - Être capable d'évaluer les performances d'un modèle, et déterminer des pistes d'optimisations.
 - Être capable de présenter à des néophytes le résultat d'un modèle et les informations qui en découlent.
 - Avoir une première connaissance sur le Deep Learning.
 - Sensibiliser sur l'utilisation de l'apprentissage automatique en terme éthique et écologique.
- **Visualisation des données** 21 h
 - Être capable d'avoir une démarche de story telling, en réalisant des supports de visualisation (Dashboard, Rapport) à l'aide de l'outil Tableau.
- **Visualisation des données avec PowerBI** 21 h
 - Être capable d'avoir une démarche de story telling, en réalisant des supports de visualisation (Dashboard, Rapport) à l'aide de l'outil PowerBi.
- **Projet 'Visualisation des données'** 35 h
 - Être capable d'appliquer les principes d'ETL et de visualisation pour répondre à une problématique à l'aide d'outils de visualisations.
- **Analyse approfondie avec le Machine Learning** 105 h
 - **Apprentissage automatique avancé** 35 h
 - Comprendre le fonctionnement du Deep Learning.
 - Avoir une première approche avec les fonctions d'activations.
 - Connaître les différents types de couches (FC, LSTM,...).
 - Connaître les différentes architectures principales dans le domaine.
 - Être capable de concevoir un modèle adapté à une problématique.
 - Être capable d'entraîner un modèle de DL.
 - Connaître et être capable d'expliquer les limites d'un modèle.
 - Être capable d'évaluer les performances d'un modèle et de déterminer les hyperparamètres permettant son optimisation (ex : GridSearch).
 - Être capable d'utiliser un modèle pré-existant.
 - Être capable de se servir de techniques de transfert learning pour profiter de connaissances préalables.
 - Comprendre et utiliser les IA génératives et les fines tuner.
 - Machine Learning Automatique (Pycaret).
 - Solution Cloud (Azure/Aws/Google).



« Opération financée par la Région Grand-Est »



Icam, site de Strasbourg-Europe, Association de Gestion

Association loi 1908 enregistrée au registre des Associations du Tribunal d'Instance de Schiltigheim sous vol. 40, folio 41, enregistrée en tant qu'organisme de formation sous le n° 42 67 04761 67 * auprès de la préfecture, dont le siège est sis 2 rue de Madrid CS 20013 - 67012 Strasbourg Cedex, identifiée sous les n° SIRET 51773337400022 APE 8542Z.

*ce numéro ne vaut pas agrément de l'état.

Page 7 sur 8 – Dernière mise à jour 15/02/2024

- **Projet 'Analyse de données'** 35 h
 - Être capable d'appliquer les outils d'analyses de données afin d'obtenir des informations permettant de répondre à une problématique donnée.
 - Être capable de présenter les données explorées à l'aide de différentes méthodes statistiques, la problématique et l'analyse dans un contexte entrepreneurial.
 - Être capable de présenter une analyse des résultats permettant de répondre à la problématique.
 - Être capable de choisir un algorithme de machine learning ou deep learning adapté à la problématique.
 - Être capable d'entraîner le modèle pour la problématique.
- **Solution Interactive** 28 h
 - Solution IBM pour le développement d'application IA..
- **Méthodologie de travail personnel** 7 h
 - Connaître les bonnes pratiques dans le domaine de la Data Science.
 - Connaître les techniques pour se maintenir à jour dans les technologies et tendances du domaine.
 - Être capable de parler de l'actualité dans le domaine.
- **Coordination et management d'un projet data science** 49 h
 - **Informatique et liberté** 14 h
 - Être capable d'expliquer et d'appliquer les principes de la protection des données privées.
 - Être capable de mettre en œuvre une politique Informatique et Liberté pour des applications de traitement de données.
 - **Travail en équipe** 14 h
 - Connaître les outils et méthodologie de la collaboration avec des équipes techniques et la gestion des conflits .
 - Développer son Leadership.
 - **Gestion de Projet Big Data/Informatique** 21 h
 - Organiser les activités et les ressources par la mise en place d'une stratégie de gestion afin d'assurer la conduite du projet dans les critères coût-qualité-délais du cahier des charges.
 - Connaître les outils de gestion de projet data.
 - Savoir choisir l'outil adapté selon le type de projet.
- **Module Projet Data Scientist : De l'intégration à l'industrialisation** 70 h
- **Module accompagnement vers l'emploi** 14 h
- **Stage en Entreprise** 600 h



« Opération financée par la Région Grand-Est »

