



Le Campus Numérique in the Alps

Formation Data Scientist

Objectifs de la formation

La formation Data Scientist est un parcours de 770 heures formant aux métiers de la Data en permettant à des Diplômés en reconversion de développer des compétences complémentaires à leur spécialité en analyse de données et en développement de l'intelligence artificielle.

La formation vise à offrir à des diplômés en recherche d'emploi une formation adaptée aux compétences recherchées par les entreprises en analyse de données, en vue de favoriser leur insertion professionnelle sur des métiers en tension. Nous formons des experts métiers de haut niveau avec des compétences complémentaires à leur cœur de métier en récolte, analyse et exploitation des données.

Le Data Scientist est responsable de l'extraction et l'interprétation de données mises à disposition par les services web et les autres canaux digitaux (big data), dans l'objectif d'aider l'entreprise à prendre des décisions stratégiques.

Les métiers de Data Scientist s'articule autour de 5 activités principales :

- Récupérer et analyser les données pertinentes liées au processus de décision de l'entreprise, à la vente ou encore aux données client ;
- Construire les algorithmes permettant d'améliorer les résultats de recherches et de ciblage ;
- Elaborer les modèles prédictifs permettant d'anticiper l'évolution des données et les tendances relatives à l'activité de l'entreprise ;
- Modéliser les résultats d'analyse des données pour les rendre lisibles et exploitables aux décideurs ;
- Fournir des outils d'aide à la décision et des recommandations stratégiques aux équipes métiers.

Compétences développées

Le programme aborde les technologies et compétences indispensables pour l'analyse de données : algorithmie, Python, Machine Learning, bases de données, outils de visualisation des données, R, outils statistiques... et les soft skills nécessaires en entreprise. Les méthodes d'apprentissage « apprendre à apprendre » permettent aux apprenants de se former en autonomie tout au long de leur carrière pour suivre les évolutions technologiques.

Compétences développées à l'issue de la formation :

- Maîtrise des outils statistiques pour l'analyse de données
- Maîtrise des outils analytiques pour l'analyse et l'exploration des données (R)
- Maîtrise des langages de programmation utilisés en analyse de données (Python)
- Maîtrise des méthodes de Machine Learning
- Gestion de données non structurées
- Récolte, analyse et exploitation de grands jeux de données
- Interprétation des données et formulation de recommandations stratégiques

Méthodes pédagogiques

Le Campus adopte une approche par compétences, les modules de formation étant construits sur la base de compétences à acquérir : savoirs, savoir-faire, savoir-être, outils, méthodologies... Cette approche permet d'assurer une formation professionnalisante avec le développement de compétences opérationnelles correspondant aux exigences des métiers visés.

Le modèle de pédagogie par projet mis en place dans les modules de formation permet de rendre les stagiaires acteurs de la formation et de l'apprentissage. L'autonomie développée dans les modules (e-learning et classes inversées), la transversalité marquée des enseignements et une pratique pédagogique hautement professionnalisante (développement d'applications dès les premières heures passées au Campus et apprentissage en alternance) sont au cœur du projet pédagogique.

Modalités d'évaluation et diplôme visé

La formation donne lieu à une Attestation de compétences délivrée par Le Campus Numérique in the Alps.

La formation Data Scientist est une formation labellisante par l'Institut MIAI de l'UGA, sous le label « [Applications de l'IA](#) ».

Modalités d'évaluation :

- **Validation des compétences** (auto-validation et validation par un formateur professionnel de l'ensemble des compétences associées à chacun des modules de formation)
- **Réalisation de 4 rapports d'activité en entreprise**
- **Réalisation d'un rapport de mission professionnelle**
- **Soutenance devant un jury mixte** (professionnels du secteur numérique et de représentant académiques).

Durée de la formation

La formation se décompose en deux phases :

- **3 mois de formation intensive au Campus (399 heures) permettant d'acquérir le socle de compétences techniques et de travail en équipe ;**
- **7 mois en entreprise comprenant 371 heures de formation afin d'élargir et d'approfondir les compétences techniques et les soft skills développées.**

Modalités d'accès et prérequis

Tous les candidats titulaires d'un diplôme de niveau Master ou Doctorat (mathématiques, informatique, ingénierie, physique, chimie, statistiques, matériaux, sciences de la terre, économie, génie civil, génie industriel, mécanique, nanotechnologies, traitement du signal, astrophysique, biomédical, sciences humaines et sociales...), possédant des notions d'au moins un langage de programmation et étant demandeurs d'emploi inscrits à Pôle Emploi peuvent s'inscrire à la formation. Les critères de sélection sont basés sur :

- Motivation, cohérence et clarté du projet professionnel : vouloir développer des compétences complémentaires au cœur de métier en analyse de données
- Aptitudes scientifiques (mathématiques, probabilités, apprentissage de la programmation)
- Aptitudes métier, qualités professionnelles et aptitudes au travail en équipe.
- Niveau d'anglais (B1-B2)

Processus de sélection :

- Envoi d'un CV, d'une lettre de motivation et d'une vidéo pitch
- Immersion dans un travail collaboratif non numérique (piscine de 3 à 4 jours)

Prix de la formation

Gratuit grâce aux partenaires financeurs de la formation et les fonds de financement de la formation.

Lieu de formation

La formation Data Scientist est réalisée :

Campus de Grenoble

Le Totem

16 boulevard Maréchal Lyautey

38000 Grenoble

07 50 66 51 51 / audrey.graffagnino@le-campus-numerique.fr

Accessibilité aux personnes en situation de handicap

La formation est accessible aux personnes en situation de handicap, avec une adaptation possible du rythme pédagogique et des modalités d'évaluation.

Programme Pédagogique

Socle		
Positionnement	4 heures	Bilan, diagnostic et définition du parcours individualisé de formation
Outils numériques pour formation à distance	3 heures	Mise en place de l'environnement de formation à distance, formation sur les outils numériques utilisés à distance (installation de logiciels, vérification de connexion internet, tests techniques)
TRE	21 heures	<p>Objectifs pédagogiques :</p> <ul style="list-style-type: none"> - Maîtriser les techniques de recherche d'emploi, de rédaction du CV et de la lettre de motivation - Connaître les techniques de création d'un profil professionnel numérique (LinkedIn notamment) - Connaître les techniques et stratégies nécessaires pour s'engager dans un processus actif de recherche d'emploi - Maîtriser les techniques de valorisation des profils et de présence numérique.
Développement Python & Algorithmie	91 heures	<p>Objectifs pédagogiques :</p> <ul style="list-style-type: none"> • Maîtriser les concepts algorithmiques de la programmation (variables, opérateurs, conditions, boucles et tableaux) • Connaître la méthodologie de conception d'une solution informatique • Analyser la complexité d'un algorithme. • Maîtriser les bases de la programmation par la pratique à l'aide du langage Python (variables, instructions, conditions, boucles, fonctions, structures de données, programmation orientée objet). • Maîtrise de Jupyter Notebooks et de la bibliothèque Python Panda. <p>Compétences à acquérir :</p> <ul style="list-style-type: none"> - Programmer avec des éléments algorithmiques de base (boucle, condition, variable et leur scop) - Utiliser les outils de visualisation de base de Python - Implémenter des traitements en base de données
Machine Learning	119 heures	<p>Objectifs pédagogiques :</p> <p>Implémentation et apprentissage sur des jeux de données réels pour appréhender les principaux outils de Machine Learning (classification,</p>

		<p>décomposition en composantes principales, K-Mean, K-Folds, Descentes de gradients, réseaux de neurones).</p> <p>Compétences à acquérir :</p> <ul style="list-style-type: none"> - Implémenter des algorithmes d'apprentissage statistique après analyse de la problématique du cahier des charges, en différenciant les types d'apprentissages automatiques supervisés et non-supervisés, les notions de régression et de classification - Utiliser les algorithmes d'apprentissage statistique avec la bibliothèque de visualisation des données pertinente. - Mettre en place une procédure de sélection de modèles par grid-search et cross-validation.
Outils statistiques pour la donnée	77 heures	<p>Objectifs pédagogiques : Maîtrise des outils mathématiques pour l'analyse de données (statistiques descriptives, tests d'hypothèses, nettoyage de données, régressions).</p> <p>Compétences à acquérir :</p> <ul style="list-style-type: none"> - Etablir un plan d'analyse statistique avec différentes hypothèses - Définir un modèle statistique sans apprentissage pour l'analyse de données - Définir un ensemble d'algorithmes transposés dans un langage spécifique à l'analyse de données pour construire un programme de traitement statistique sans apprentissage.
Bonnes pratiques du développement logiciel en équipe	21 heures	<p>Objectifs pédagogiques : Utilisation des méthodes de gestion de projet Agile et des outils de travail collaboratifs dans une équipe de développement informatique.</p> <p>Compétences à acquérir :</p> <ul style="list-style-type: none"> - Découper un projet en tâches et réaliser une estimation temporelle des tâches - Piloter un projet avec la méthode Agile avec des revues de projets et une rétrospective de fin de projet - Coordonner un projet entre différents acteurs en fédérant les membres de l'équipe de développement informatique et en utilisant la méthodologie Agile, des outils de

		rétrospectives et des systèmes de gestion des flux de travaux (workflows).
Système d'exploitation et bases de données	42 heures	<p>Objectifs pédagogiques :</p> <ul style="list-style-type: none"> • Maîtrise du système d'exploitation open source Linux et construction et manipulation de bases de données. • Maîtrise des concepts et l'utilisation d'un système de gestion de base de données (SGBD). <p>Compétences à acquérir :</p> <ul style="list-style-type: none"> - Maîtrise du système d'exploitation Linux (installation de paquet, utilisation de la ligne de commande, gestion automatisée de fichiers, redirection d'entrées / sorties) - Exploiter un SGBD à l'aide du langage de programmation SQL - Se connecter à un SGBD à l'aide d'un client GUI et du langage de programmation Python - Concevoir une base de données relationnelle.
Outils de visualisation	14 heures	<p>Objectifs pédagogiques :</p> <p>Cartographier et adapter les outils de visualisation aux jeux de données pour créer une représentation simplifiée.</p> <p>Compétences à acquérir :</p> <ul style="list-style-type: none"> - Utilisation de bibliothèques de données avancées - Réalisation d'un tableau de bord (<i>dashboard</i>) pour présenter les données sélectionnées et le travail de modélisation réalisé - Restituer des données claires et synthétiques au commanditaire et à l'équipe en adaptant la présentation et en sélectionnant la figure adaptée pour restituer des résultats compréhensibles par un public non-expert.
Compétences douces	7 heures	<p>Objectifs pédagogiques :</p> <ul style="list-style-type: none"> - Maîtrise des outils et techniques de communication - Vulgarisation et présentation de résultats scientifiques <p>Compétences à acquérir :</p> <ul style="list-style-type: none"> - Utiliser des bibliothèques et des méthodes de visualisation des données avancées

		<ul style="list-style-type: none"> - Restituer des données claires et synthétiques - Maîtriser les techniques et outils de communication
--	--	--

Contenu pédagogique Phase alternée		
Machine Learning avancé	21 heures	Objectifs pédagogiques : Savoir reconnaître et résoudre des problèmes d'optimisation
Gestion de projet Data	21 heures	Objectifs pédagogiques : Maîtrise des méthodologies et des outils de gestion de projet appliquées à l'analyse de données (méthodologies agiles, projets cycle en V)
Optimisation	14 heures	Objectifs pédagogiques : <ul style="list-style-type: none"> • Modélisation des modèles d'optimisation par contrainte • Résolution des contraintes • Optimisation
Mathématiques pour l'informatique	21 heures	Objectifs pédagogiques : <ul style="list-style-type: none"> • Outils permettant d'implémenter les articles de la littérature • Intégration & dérivation numérique multi-dimensionnelle • Calcul matriciel
Visualisation et UI	14 heures	Objectifs pédagogiques : <ul style="list-style-type: none"> • Outils et méthodes de visualisation de grands jeux de données • Vulgarisation et représentation simplifiée • Création d'un Dashboard interactif de visualisation de données
Advanced computer vision (ACV)	42 heures	Objectifs pédagogiques : <ul style="list-style-type: none"> • Traitement de l'image • Deep learning (augmentation de données, GAN) • Deep learning vidéo
Apprentissage par renforcement	21 heures	Objectifs pédagogiques : <ul style="list-style-type: none"> • Création d'algorithmes de prise de décision sans jeux de données à priori • Prise de décision itérative • Optimisation sous contrainte
Natural Language Processing	21 heures	Objectifs pédagogiques : Traitement naturel du langage : compréhension du langage humain grâce à l'intelligence artificielle
Ethique de la donnée	17.5 heures	Objectifs pédagogiques : Compréhension des problématiques de données personnelles et services numériques, d'éthique et de gouvernance des données et

		d'anonymisation des données personnelles.
Open Source	7 heures	Objectifs pédagogiques : Maîtrise des problématiques liées aux aspects communautaire du libre, aux différents types de licences et aux aspects légaux des projets d'analyse de données.
Big Data	35 heures	Objectifs pédagogiques : Application de gestion de projet / Machine Learning & Deep Learning sur un grand jeu de données
Dev Ops	14 heures	Objectifs pédagogiques : Technologies et outils du DevOps : <ul style="list-style-type: none"> - GIT et GIT flow - Tests et déploiement - Conteneurisation - Linux avancé - Docker & containers
Conception Objet	14 heures	Objectifs pédagogiques : Réalisation d'un projet Orienté Objet avancé : Conception UML.
Support Projet	28 heures	Objectifs pédagogiques : Accompagnement sur les problématiques entreprises par un expert en Machine Learning
Préparation Soutenance	7 heures	Objectifs pédagogiques : Préparation des rapports professionnels Préparation soutenances et présentations professionnelles.
Green IT	14 heures	Objectifs pédagogiques : Sensibilisation à l'éco-conception Elaboration d'un guide des bonnes pratiques environnementale Elaboration de la fresque du numérique
Culture mathématiques et informatique	21 heures	Objectifs pédagogiques : <ul style="list-style-type: none"> • Choix des structures de données • Fonctionnement des machines • Algorithmique avancée • Langage C++ • Bases de données NoSQL • Statistiques avancées • Epistémologie • Scrapping/ API
Laïcité, citoyenneté et valeurs de la République	3.5 heures	Objectifs pédagogiques : Maîtrise des principes et valeurs de laïcité et des notions liées aux valeurs de la République Française (devise, laïcité, citoyenneté).
Anglais	7 heures	Objectifs pédagogiques :

		<p>Utilisation de l'anglais technique pour évoluer dans un contexte professionnel anglophone et utiliser la documentation spécifique en anglais.</p> <p>Compétences à acquérir :</p> <ul style="list-style-type: none"> - Maîtrise du vocabulaire technique et scientifique lié à l'analyse de données et au développement de l'intelligence artificielle - Gérer les supports techniques en anglais - Maîtrise des méthodologies de rédaction de documentations techniques en anglais - Maîtrise du vocabulaire technique permettant de réaliser des présentations professionnelles. <p>Modalités d'apprentissage :</p> <p>Réalisation de projets d'analyse de données et présentation de productions en anglais (synthèse d'analyses de données, présentations de résultats).</p> <p>Travail avec un formateur sur la réalisation du rapport de mission professionnelle en anglais présenté devant le jury du label IA & Application du Multidisciplinary Institute in Artificial Intelligence (MIAI) de l'Université Grenoble Alpes.</p>
RGPD	7 heures	<p>Objectifs pédagogiques :</p> <p>Compréhension du RGPD, les différents principes à respecter (minimisation de la collecte, finalité, consentement, conservation limitée des données, sécurité et confidentialité des données), les enjeux du RGPD pour les développeurs web et l'analyse de données.</p>
Cloud	21 heures	<p>Objectifs pédagogiques :</p> <p>Maîtrise des technologies Cloud (Azure, GCP, AWS) pour développer, mettre en place et maintenir les outils et infrastructures des applications Data. Maîtrise des concepts généraux du cloud computing (différents providers, avantages...)</p>

En partenariat avec

