CERTIFICAT DE DATA SCIENTIST DU GENES

CERTIFICAT

Les technologies de traitement des données massives (Big Data) sont maintenant disponibles pour toutes les entreprises, que ce soit sous forme de services Cloud, de plateformes open source ou encore de solutions proposées par les éditeurs de logiciels propriétaires. **Mais le manque de compétences nécessaires à la valorisation de cette manne de données est palpable.** Tandis que la demande de data scientists s'accroît, l'offre de talents est insuffisante.

Le data scientist est un statisticien expert de la gestion et de l'analyse pointue de données massives ; il lui est demandé de concevoir de nouveaux usages des données, par le croisement de sources multiples, par leur accumulation et par la mise en œuvre de nouvelles techniques. Il manipule des quantités massives de données souvent hétérogènes, non structurées et non nettoyées. À la croisée de plusieurs métiers, il est à la fois un spécialiste en statistique, en informatique et connaît la fonction d'application des données analysées.

À la connaissance des méthodes traditionnelles d'analyse statistique, s'ajoute celle des techniques de machine learning et de deep learning. De plus, les volumes de données en jeu rendent essentielle une forte compétence technique des outils informatiques permettant de les traiter efficacement.

C'est pourquoi **l'Ensae-Ensai Formation Continue** a créé une formation certifiante, dispensée par les meilleurs spécialistes du domaine, s'adressant spécifiquement aux professionnels en activité.

LA FORMATION

La formation s'étend sur 18 jours – 126 heures (hors examen et soutenance) d'enseignements théoriques et pratiques.

Le programme est construit autour :

- Des bases de données SQL, NoSQL, et des logiciels de requêtage, traitement et manipulation.
- Des algorithmes de **machine learning** et de **deep learning** : régression, régression sous contrainte, clustering, forêts aléatoires, boosting, réseaux de neurones...
  Algorithmes présentés d'un point de vue théorique mais mis en pratique immédiatement sur des données réelles avec R ou Python.
- Des outils de visualisation, de **cloud computing**, de calcul parallèle (**Spark**)…

La formation est complétée par une sensibilisation aux **questions éthiques** et d'environnement juridique.

Cette formation permet d'appréhender une multitude de méthodes et de pratiques. Elle s'adresse à toute personne souhaitant acquérir une culture solide dans ce domaine, en interaction avec des professionnels éclairés, tous praticiens expérimentés ou enseignants-

ENSAE-ENSAI Formation continue (Cepe)

GENES

chercheurs de renom. Les participants au « certificat de Data Scientist » développent des compétences pour commanditer et réaliser des analyses en Big Data en utilisant une méthodologie pertinente et en apprécier la qualité et la portée.

La motivation et l'engagement des candidats à consacrer un temps régulier de travail personnel entre chaque session est un facteur clé de réussite.

La maîtrise des langages R ou Python (suivant le langage utilisé dans la promotion auquel le candidat postule) constitue un prérequis.

LA CERTIFICATION

À la fin du parcours de formation, les candidats sont soumis à deux épreuves.

L'une consiste en un **examen** qui requiert l'implémentation sur ordinateur d'études de cas sur des données réelles. Le candidat dispose de tous les documents qu'il juge nécessaires afin de se rapprocher au maximum des conditions de travail réelles. L'autre repose sur **la soutenance, devant un jury**, d'un projet implémenté tout au long du cursus de formation. Cette mise en situation professionnelle vise à mettre en pratique les fondamentaux de la data science, sur un cas réel.

La validation du certificat est soumise à la réussite à l'examen, à la soutenance ainsi qu'à la présence obligatoire aux cours. En cas de réussite, le participant se voit décerner le **« certificat de data scientist du Genes ».**

L'ENSAE-ENSAI FORMATION CONTINUE

L'Ensae-Ensai Formation Continue (Cepe) est le centre de formation continue du **Genes, Groupe des écoles nationales d'économie et statistique.**

Fort de plus de 60 ans d'expérience, son objectif principal est de délivrer des formations exigeantes et de qualité via des **contenus scientifiques innovants et une centaine de formateurs de premier plan, toujours experts dans leur domaine d'intervention.**

La formation aura lieu dans les locaux de l'Ensae-Ensai Formation Continue, situés au 21 rue Camille Desmoulins à Issy-les-Moulineaux : **informations d'accès**.

Pour en savoir plus, consultez www.lecepe.fr

RENSEIGNEMENTS : Envoyer un courriel à conseil@ensae.fr

CANDIDATURES : Déposer votre dossier

Le dossier de candidature est constitué d'une lettre de motivation et d'un Curriculum Vitae. Il permettra d'examiner les prérequis, les objectifs professionnels et l'engagement du candidat.

ENSAE-ENSAI Formation continue (Cepe)

GENES

Planning du CERTIFICAT DE DATA SCIENTIST du GENES – BD34

MODULES	PERIODE	DURÉE
Module 1	**Du 1 au 3 octobre 2025**	3 jours
1. Python et SQL 3. Python et introduction au NoSQL 4. Webscraping et pipeline de traitements	1 octobre 2025 2 octobre 2025 3 octobre 2025	
Module 2	**Du 5 au 7 novembre 2025**	3 jours
1. Machine learning : MCO & régressions pénalisées 2. Machine learning : Régression logistique 3. Machine learning : Arbres et random forest	5 novembre 2025 6 novembre 2025 7 novembre 2025	
Module 3	**Du 1 au 3 décembre 2025**	3 jours
1. Machine learning : Boosting & SVM 1. Réduction de dimension : Clustering & Analyse factorielle 3. Cloud Computing (matin) 4. Data camp (après-midi)	1 décembre 2025 2 décembre 2025 3 décembre 2025	
Module 4	**Du 5 au 7 janvier 2026**	3 jours
1. Visualisation 2. Deep Learning 3. Deep Learning (matin) 4. Projet tutoré #1 (après-midi)	5 janvier 2026 6 janvier 2026 7 janvier 2026	
Module 5	**Du 2 au 4 février 2026**	3 jours
1. Deep Learning : Application aux textes ou images 2. Deep Learning : Application aux textes ou images (matin) 3. Projet tutoré #2 (après-midi) 4. Pipeline de traitements Python	2 février 2026 3 février 2026 4 février 2026	
Module 6	**Du 2 au 4 mars 2026**	3 jours
1. Big Data Processing avec Spark (1/2) 2. Big Data Processing avec Spark (2/2) 3. Projet tutoré #3 (matin) 4. Droit et éthique de la donnée (après-midi)	2 mars 2026 3 mars 2026 4 mars 2026	

L'examen aura lieu le 31 mars 2026 en présentiel à l'Ensae-Ensai Formation Continue (Cepe).
L'ordre et le détail des modules peuvent être amenés à évoluer légèrement

ENSAE-ENSAI Formation continue (Cepe)

GENES