

Conception des modèles analytiques d'IA et de machine-learning

RS5335 · RS · Publiée

L'essentiel

Code(s) NSF	326t, 326
Formacode(s)	31028, 31026, 31052, 31025, 11014
Type d'enregistrement	Enregistrement sur demande
Date d'enregistrement	20/01/2021
Date d'échéance	20/01/2026

Certificateur(s)

DIGITAL SERVICES SCHOOL · SIRET : 38183776400095<https://ecoleladataetia.com>**ORANGE · SIRET : 38012986648625****BUSINESS ET DECISION UNIVERSITY · SIRET : 38183776400087****ORANGE · SIRET : 38012986646850**

Résumé de la certification

OBJECTIFS ET CONTEXTE

Depuis les cinq dernières années, les projets Data Science et d'intelligence artificielle connaissent une forte croissance dans des domaines très variés, tels que l'industrie pétrolière, la téléphonie, la distribution, la banque, la finance et les services. Les récentes évolutions technologiques permettent de faire évoluer ces projets vers la mise en place de véritables solutions d'intelligence artificielle. La réalisation de ces projets nécessite de former des équipes de « Data-Engineers » et de « Data Scientists » pour lesquelles il est nécessaire de conjuguer l'acquisition des compétences techniques et méthodologiques, l'expérience des cas concrets et surtout, un parcours d'apprentissage jalonné.

Les principales compétences visées par cette certification sont celles nécessaires à la mise en œuvre des activités suivantes :

- Réaliser une Analyse exploratoire et multivariée complète des données émettre et tester des hypothèses, faire des inférences sur les données. Réaliser des jeux de données de modélisation et de test
- Identifier et sélectionner des algorithmes de base à utiliser pour la modélisation en adéquation avec les jeux de données et l'objectif souhaité. Adapter la complexité des algorithmes à la complexité des données
- Prototyper les modèles de DataScience en utilisant les algorithmes et les jeux de données sélectionnés
- Restituer aux interlocuteurs métiers les perspectives et les interprétations basées sur les premiers résultats des modèles

COMPÉTENCES ATTESTÉES

- Maîtriser les principales méthodes de réduction de dimension non supervisées, et en particulier l'Analyse en composantes principales et l'analyse en correspondances multiples
- Maîtriser les principales méthodes de clustering et de typologies non supervisées, en particulier les méthodes hiérarchiques et les méthodes de type K-Means
- Maîtriser l'analyse statistique des corrélations, de leur validité et des corrélations partielles
- Savoir créer un jeu de données pertinent dans une stratégie de cross validation, savoir utiliser les différentes méthodes d'échantillonnage associées, et savoir éviter un biais dans les données d'apprentissage
- Maîtriser les tests de Student, les différentes méthodes de test d'hypothèses et la notion de p-value associée à

différents tests statistiques

- Connaître les différentes formes d'apprentissage machine, les grandes familles d'algorithmes qui en dépendent, et leurs cas d'application métier
- Savoir utiliser les lois d'apprentissage des algorithmes et les règles de Vapnik associées, savoir éviter l'under fitting et l'over fitting
- Connaître les principaux algorithmes de DataScience (du domaine public) et leurs applications dans chaque grande famille d'algorithmes
- Savoir utiliser les principales bibliothèques de DataScience en Python avec les environnements de développement intégrés et les workflows pour prototyper les modèles
- Être capable d'assembler la chaîne de traitement des différents modèles en utilisant les principaux outils de DataScience que ce soit en mode workflow, en mode notebook ou en mode programmation
- Savoir optimiser le compromis explicabilité et traçabilité des algorithmes versus sa précision
- Savoir documenter les modèles avec Jupyter Notebook
- Être capable d'illustrer les résultats des modèles grâce à des graphiques spécialisées compréhensibles par les métiers
- Savoir sélectionner les observations qui illustrent le comportement typique des modèles
- Savoir interpréter les résultats des différents modèles et en expliquer les résultats
- Savoir utiliser des modèles explicatifs pour pouvoir rendre traçables et explicables des modèles complexes

Voies d'accès

MODALITÉS D'ACCÈS

Formation continue

COMPOSITION DU JURY — FORMATION CONTINUE

Le jury est présidé par un data-scientist confirmé. Il établit la composition du jury qui sera composée d'au minimum 4 consultants-experts des domaines couverts par le certificat et du directeur de Business & Decision University. Les membres du jury peuvent être des formateurs du certificat, mais dans tous les cas plus de 50% des membres du jury seront des personnes extérieures à l'organisation et si des formateurs participent au jury, leur voix sera strictement consultative.

Base légale

Date de décision	20/01/2021
Durée de l'enregistrement	5 ans
Date d'échéance de l'enregistrement	20/01/2026

Pour plus d'informations

Fiche officielle (France Compétences)

<https://www.francecompetences.fr/recherche/rs/5335/>

Demande CPF — formations engagées

Dossiers CPF engagés	9
Montant total engagé	12 200 €
Coût moyen par dossier	1 356 €
Durée moyenne	78 h
Période	05/2021 – 12/2021
France compétences	100 %

