

## Traitement des données

CATEGORIE : C

### Vue d'ensemble

Domaine(s) d'activité professionnel dans lequel(s) est utilisé la certification :

Transverse : ■ **Tous domaines concernés**

Code(s) NAF : **58.29A**, **49.10Z**, **64.11Z**, **65.11Z**,  
**61.10Z**

Code(s) NSF : **320**, **313**

Code(s) ROME : **M1804**, **M1701**, **C1102**, **C1101**, **C1104**

Formacode : **31054**

Date de création de la certification : **01/09/2015**

Mots clés : **Evaluation**, **qualité**, **Modelisation**,  
**Traitement de données**

### Identification

Identifiant : **2965**

Version du : **11/07/2017**

### Références

Consensus, reconnaissance ou recommandation :

Formalisé :

- [Projet de loi : République numérique](#)

Non formalisé :

- [traitement des données : enjeu crucial des entreprises](#)

### Descriptif

#### Objectifs de l'habilitation/certification

Les procédés utilisés dans le traitement des données sont de nature statistiques, probabilistes et doivent servir aux objectifs métiers pour l'entreprise. Les moyens et techniques utilisés sont divers. L'approche retenue est une approche itérative d'amélioration continue. Elle vise à réaliser plusieurs itérations, afin de s'assurer que l'objectif métier a été bien atteint. Cette approche est aussi collaborative puisqu'elle fait appel autant à des connaissances informatiques

qu'à des connaissances métiers. La dimension relative à la communication et la présentation des résultats sont aussi importantes. L'objectif visé est de faire monter en compétences les collaborateurs sur l'analyse critique des résultats et sur la capacité de communiquer sur ces derniers. Il est attendu une communication pertinente autant au niveau d'une vision analytique que d'une vision synthétique.

#### Lien avec les certifications professionnelles ou les CQP enregistrés au RNCP

- Néant

#### Descriptif général des compétences constituant la certification

##### Analyse qualitative et quantitative des données.

##### T1 : Identification du périmètre des données.

Classer les données, préalablement listées, en distinguant les sources internes et externes pour qualifier leurs fiabilités en fonction de leurs typologies

Appréhender les enjeux de chaque métier et des cas d'usage associés pour intégrer les objectifs à viser et les besoins à satisfaire

Créer les liens entre les données mises à disposition et les différents métiers concernés pour définir le périmètre de l'analyse

Exploiter la connaissance sur les données détenue par leurs

### Public visé par la certification

Tous publics

propriétaires et gestionnaires pour les enrichir et mieux les interpréter

## **T2 : Qualification des données.**

Appliquer les procédés statistiques pour dénombrer le contenu de l'échantillon des données.

Appliquer les techniques d'évaluation de la qualité des données pour apprécier celle de l'échantillon et y accorder un niveau de confiance.

Analyser les écarts constatés entre les résultats intermédiaires obtenus et les objectifs associés pour évaluer la nécessité de compléter ou non l'analyse, en y attribuant des moyens supplémentaires.

Formaliser un rapport de description des données, incluant des représentations graphiques afin de restituer un état des lieux sur la qualité et la quantité des données.

Présenter le rapport d'analyse de données en s'adaptant à l'auditoire.

Déterminer les corrélations entre les données pour les aligner aux enjeux des métiers existants. (Gestion des risques - Réduction des coûts - Génération de revenus additionnels)

## **Préparation des données.**

### **T1 : Sélection des données**

Sélectionner les enregistrements et les attributs de données en lien avec les objectifs métiers pour permettre l'analyse de l'exploitation des données.

Arbitrer suivant une logique d'ensemble, entre la décision d'inclure ou non les données, en fonction des objectifs métiers, pour clarifier le traitement qui sera mis en place.

### **T2 - Nettoyage des données.**

Exploiter le rapport de description des données mettant en exergue l'état des lieux sur leur qualité pour choisir les techniques de nettoyage les plus appropriées.

Appliquer les procédés de nettoyage de la donnée pour fiabiliser les résultats.

Elaborer un rapport sur le nettoyage des données pour présenter le niveau d'utilité des données dans leur futur traitement.

### **T3 - Construction de nouvelles données.**

Appliquer des techniques de calcul de nouveaux attributs et de normalisation des données (création de champ calculé à partir de champ existant) pour permettre l'exécution des algorithmes de modélisation retenus.

### **T4 - Agrégation de la donnée.**

Appliquer les deux méthodes d'agrégation des données (fusion des données - ajout des données) pour créer le référentiel des données à modéliser excluant les doublons.

Évaluer le niveau de confiance relatif au référentiel de données préalablement élaboré pour confirmer l'intérêt d'engager ou non les traitements futurs.

### **T5 - Formatage de la donnée**

Mettre en adéquation des formats de données au regard des algorithmes de modélisation qui seront appliqués pour optimiser le traitement. (chemin d'utilisation - délais de traitement- durée de modélisation).

## **Traitement des données.**

Appliquer les critères de modélisation (types de données disponibles pour l'exploration – objectifs d'analyse de données – exigences de modélisation particulières) pour sélectionner les techniques de modélisation adaptées.

Transformer le problème Métier initial en un problème mathématique pour le traduire en algorithme

Concevoir le modèle mathématique et estimer les paramètres afin de choisir les techniques de modélisation les plus appropriées.

Appliquer les techniques de modélisation des données à travers la création de modèles qui répondent aux problématiques posées tout en s'assurant de la traçabilité des actions menées.

## **T2 - Evaluation des résultats du traitement des données**

Evaluer la pertinence des résultats au niveau de leurs utilités et de leurs exploitabilités, à partir de critères de réussite préalablement prédéfinis pour valider l'atteinte des objectifs.

## **T3- Evaluation des modèles et techniques de modélisation**

Evaluer la pertinence des modèles et techniques en fonction des objectifs d'analyse préalablement définis pour une réexécution possible des modèles et/ou un ajustement des valeurs de paramètres.

Mettre en œuvre les procédés d'amélioration et d'adaptation des modèles et techniques de modélisation pour correspondre aux évolutions futures des données.

## **Restitution des résultats.**

## **T1 - Interprétation des résultats du traitement de la donnée et communication**

Interpréter les résultats en lien avec les objectifs métiers visés et les marges de progression identifiées pour aboutir à une analyse finale circonstanciée.

Formaliser un rapport des résultats d'exploitation des données, incluant des représentations graphiques afin de restituer un état des lieux sur les résultats

Communiquer sur le rapport d'exploitation des données, en élaborant une "histoire à raconter" mettant en lumière les principaux points pertinents aux yeux des Métiers.

## **T2 - Déploiement des résultats du projet.**

Accompagner les équipes-métiers dans l'utilisation des résultats du traitement de la donnée, en faisant preuve de pédagogie, pour mettre en place les améliorations possibles.

Relever les retours d'expériences obtenus auprès des équipes-métiers corrélés à sa propre expérience pour réaliser une révision finale du projet.

## *Modalités générales*

*La formation se dispense en continue ou en alternance. En intra ou en inter.*

*4 modules de formation : Durée de 70 heures à 84 heures (de 10 à 12 jours)*

*Module 1 : Analyse qualitative et quantitative des données.*

*Module 2 : Préparation de données*

*Module 3 : Traitement des données*

## Module 4 : Restitution des résultats

### Liens avec le développement durable

niveau 1 : Certifications et métiers qui internalisent le développement durable. Les activités et compétences mobilisées mettent en oeuvre des matériaux et produits moins polluants

## Valeur ajoutée pour la mobilité professionnelle et l'emploi

### Pour l'individu

L'individu bénéficie d'un apport en compétences axées sur l'identification du périmètre des données en fonction des objectifs visés par le projet. Il devra être capable de créer un modèle mathématique débouchant sur une représentation graphique qui amènera une analyse compréhensive des résultats avec les équipes métiers. La communication devra se faire de manière fluide afin de créer un lien permettant une mise à jour des données en partenariat avec les équipes métiers. Analyse, qualification, sélection, nettoyage des données constituent les étapes essentielles pour aboutir à une modélisation et une interprétation des données. Toutes ces compétences participent à l'employabilité des personnes ainsi que la sécurisation de leur parcours professionnels.

### Pour l'entité utilisatrice

Pour l'entreprise, la certification « Traitement des données » est nécessaire car elle s'agit là, d'une activité participant à la stratégie de l'entreprise. Savoir détecter les données fiables en créant une modélisation mathématique de celles-ci apporte à l'entreprise, un appui stratégique qui va lui permettre de rebondir en fonction des résultats de l'analyse. La coopération entre l'analyste et les équipes métiers est un réel plus pour tirer les conclusions des actions menées ainsi que pour la stratégie future à bâtir. Plus cette activité s'appuiera sur le professionnalisme de collaborateurs immergés dans le traitement des données et plus le développement pourra être envisagé par les entreprises.

## Evaluation / certification

### Pré-requis

Aucun prérequis particulier mais une connaissance en informatique est favorable

### Compétences évaluées

#### **Analyse qualitative et quantitative des données.**

##### **T1 : Identification du périmètre des données.**

Classer les données, préalablement listées, en distinguant les sources internes et externes pour qualifier leurs fiabilités en fonction de leurs typologies

Décrire les enjeux de chaque métier et des cas d'usage associés pour intégrer les objectifs à viser et les besoins à satisfaire

Décrire les liens entre les données mises à disposition et les différents métiers concernés pour définir le périmètre de l'analyse.

##### **T2 : Qualification des données.**

Appliquer les procédés statistiques pour dénombrer le contenu de l'échantillon des données.

Appliquer les techniques d'évaluation de la qualité des données pour apprécier celle de l'échantillon et y accorder un niveau de confiance.

Analyser les écarts constatés entre les résultats intermédiaires obtenus et les objectifs associés pour évaluer la nécessité de

### Centre(s) de passage/certification

- Ecole Supérieure d'Assurances 13 rue Fernand Léger - 75020 Paris

compléter ou non l'analyse, en y attribuant des moyens supplémentaires.

Formaliser un rapport de description des données, incluant des représentations graphiques afin de restituer un état des lieux sur la qualité et la quantité des données.

Présenter le rapport d'analyse de données en s'adaptant à l'auditoire.

## **Préparation des données.**

### **T1 : Sélection des données.**

Sélectionner les enregistrements et les attributs de données en lien avec les objectifs métiers pour permettre l'analyse de l'exploitation des données.

Justifier de la décision d'inclure ou non des données pour clarifier le traitement qui sera mis en place.

### **T2 - Nettoyage des données.**

Exploiter le rapport de description des données mettant en exergue l'état des lieux sur leur qualité pour choisir les techniques de nettoyage les plus appropriées.

Appliquer les procédés de nettoyage de la donnée pour fiabiliser les résultats.

Elaborer un rapport sur le nettoyage des données pour présenter le niveau d'utilité des données dans leur futur traitement.

### **T3 - Construction de nouvelles données.**

Appliquer des techniques de calcul de nouveaux attributs et de normalisation des données (création de champ calculé à partir de champ existant) pour permettre l'exécution des algorithmes de modélisation retenus.

### **T4 - Agrégation de la donnée**

Choisir et justifier les méthodes d'agrégation des données pour créer le référentiel des données à modéliser excluant les doublons.

### **T5 - Formatage de la donnée**

Créer un nouveau jeu de données en adéquation avec les nouveaux formats de données pour optimiser le traitement.

## **Traitement des données.**

### **T1 - Modélisation des données.**

Appliquer les critères de modélisation (types de données disponibles pour l'exploration (objectifs d'analyse de données) exigences de modélisation particulières) pour sélectionner les techniques de modélisation adaptées.

Transformer le problème Métier initial en un problème mathématique pour le traduire en algorithme.

Concevoir le modèle mathématique et estimer les paramètres afin de choisir les techniques de modélisation les plus appropriées.

Appliquer les techniques de modélisation des données à travers la création de modèles qui répondent aux problématiques posées tout en s'assurant de la traçabilité des actions menées.

### **T2 - Evaluation des résultats du traitement des données**

Evaluer la pertinence des résultats au niveau de leurs utilités et de leurs exploitabilités, à partir de critères de réussite préalablement prédéfinis pour valider l'atteinte des objectifs.

## **T3- Evaluation des modèles et techniques de modélisation**

Evaluer la pertinence des modèles et techniques en fonction des objectifs d'analyse préalablement définis pour une réexécution possible des modèles et/ou un ajustement des valeurs de paramètres. Mettre en œuvre les procédés d'amélioration et d'adaptation des modèles et techniques de modélisation pour correspondre aux évolutions futures des données.

### **Restitution des résultats.**

## **T1 - Interprétation des résultats du traitement de la donnée et communication.**

Interpréter les résultats en lien avec les objectifs métiers visés et les marges de progression identifiées pour aboutir à une analyse finale circonstanciée.

Formaliser un rapport des résultats d'exploitation des données, incluant des représentations graphiques afin de restituer un état des lieux sur les résultats

Communiquer sur le rapport d'exploitation des données, en élaborant une "histoire à raconter" mettant en lumière les principaux points pertinents aux yeux des Métiers.

## **T2 - Déploiement des résultats du projet.**

*Niveaux délivrés le cas échéant (hors nomenclature des niveaux de formation de 1969)*

Aucun

La validité est Permanente

**Possibilité de certification partielle :** non

Matérialisation officielle de la certification :

Certificat de compétences

## Plus d'informations

### *Statistiques*

120 stagiaires depuis 2015

### *Autres sources d'information*

[www.esassurances.com](http://www.esassurances.com)