

Formation Détection d'anomalies - Outlier Detection

Objectifs : Approfondir la connaissance de la détection d'anomalies dans un contexte de données numériques et/ou fonctionnelles à l'aide de méthodes principalement non-supervisées

Compétences visées : - Différencier une anomalie des valeurs influentes et extrêmes

- Comprendre l'intérêt de détecter les anomalies
- Mettre en œuvre les différentes méthodes statistiques pour la détection selon le contexte des données : méthodes univariées, méthodes multivariées basées sur un modèle probabiliste, sur la détermination d'un sous-espace ou sur la notion de proximité, méthodes pour des données fonctionnelles, méthodes pour des données en HDLSS

Durée : 2 jour(s) (14 heures)

Public : Data Scientists ou statisticiens, chercheurs, ingénieurs, ...

Pré-requis : Pour suivre ce stage dans de bonnes conditions, il est recommandé d'avoir suivi en amont les formations [Statistique descriptive \(exploratoire\) : savoir décrire des observations](#) et [Statistique décisionnelle \(inférentielle\) : savoir décider au vu des observations](#)

Méthode pédagogique : Pédagogie active mêlant exposés, exercices et applications pratiques.

Chaque participant pourra mettre en œuvre les applications dans le logiciel de son choix parmi JMP ou R.

Modalités d'évaluation : Un formulaire d'auto-évaluation proposé en amont de la formation nous permettra d'évaluer votre niveau et de recueillir vos attentes. Ce même formulaire soumis en aval de la formation fournira une appréciation de votre progression.

Des exercices pratiques seront proposés à la fin de chaque séquence pédagogique pour l'évaluation des acquis.

En fin de formation, vous serez amené(e) à renseigner un questionnaire d'évaluation à chaud.

Une attestation de formation vous sera adressée à l'issue de la session.

Trois mois après votre formation, vous recevrez par email un formulaire d'évaluation à froid sur l'utilisation des acquis de la formation.

Accessibilité : Vous souhaitez suivre notre formation Formation par ville et êtes en situation de handicap ? Merci de nous contacter afin que nous puissions envisager les adaptations nécessaires et vous garantir de bonnes conditions d'apprentissage

Tarifs :

- Présentiel : 1300 € HT

- Distanciel : 1200 € HT

(-10% pour 2 inscrits, -20% dès 3 inscrits)

Option(s) :

- Forfait déjeuners : 40 € HT

Nos prochaines sessions

Distance

du 28 au 29 novembre 2024

du 3 au 4 avril 2025

du 27 au 28 novembre 2025

Lyon

du 17 au 18 avril 2025

du 29 au 30 septembre 2025

Paris

du 2 au 3 décembre 2024

du 23 au 24 juin 2025

du 1 au 2 décembre 2025

Toulouse

du 20 au 21 mai 2025

du 6 au 7 novembre 2025

Programme :

- Introduction

- Qu'est-ce qu'une anomalie ? une valeur influente ? une valeur extrême ?
- Les différentes motivations à la détection d'anomalies
- La notion de robustesse

- Les méthodes univariées

- La règle du k-sigma et test de Grubbs
- Règles du boxplot
- Tests en fonction de la distribution

- Les méthodes multivariées - généralités

- Les grandes approches dans la détection d'anomalies
- Évaluation des méthodes
 - Caractéristiques souhaitées
 - Métriques
 - Contributeurs ou signature des défauts

- Les méthodes multivariées basées sur un modèle probabiliste

- T2 de Hotelling, la distance de Mahalanobis et sa version robuste
- Notion de profondeur et notion d'angles

- Les méthodes multivariées basées sur la détermination d'un sous-espace

- L'ACP et sa version robuste
- Les réseaux de neurones

- Les méthodes multivariées basées sur la notion de proximité

- La classification non supervisée
- Le LOF basé sur la densité

- Les méthodes pour des données fonctionnelles

- Introduction au contexte de données fonctionnelles
- L'analyse de données fonctionnelles
 - Méthode de lissage
 - Spline cubique
 - Polynômes locaux
 - Réduction de dimension
 - Grandeurs statistiques (moyenne...)
 - Décomposition en coefficients d'ondelettes
- Méthodes de détection d'anomalies dans un contexte univarié
- Méthodes de détection d'anomalies dans un contexte multivarié

- Les méthodes pour des données en HDLSS (grande dimension, faible taille d'échantillonnage)

- Présentation des challenges induits par ce contexte
- Les principales méthodes

Date de dernière modification : 5 novembre 2024