Formation Text Mining

Objectifs : Découvrir comment décrire, comparer, classer, analyser des ensembles de textes. Il peut s'agir de textes littéraires, scientifiques (bibliométrie, recherche documentaire), économiques, sociologiques (réponses aux questions ouvertes dans des enquêtes socio-économiques, entretiens divers en marketing, psychologie appliquée, pédagogie, médecine), de textes historiques, politiques...

Compétences visées : - Se familiariser avec l'analyse textuelle
- Connaître les étapes de qualification et de formatage des données brutes en vue d'y appliquer les méthodes d'analyse statistique
- Mener une analyse descriptive du texte (fréquence des mots, bilan lexical, table de dissimilarité, co-occurences, nuage de mots,....)
- Approfondir l'analyse au moyen de techniques multidimensionnelles

Durée : 3 jour(s) (21 heures)

Public : Toute personne travaillant sur des données de type texte et désirant exploiter au mieux les méthodes statistiques exploratoires.

Pré-requis : Pour suivre ce stage dans de bonnes conditions, il est recommandé d'avoir suivi en amont la formation Statistique descriptive (exploratoire) : savoir décrire des observations

Méthode pédagogique : La présentation alterne exposés et illustrations par des applications en grandeur réelle, choisies dans des domaines divers, industriels et socio-économiques. On insistera plus sur le principe des méthodes et les règles d'interprétation des résultats qui en découlent, plutôt que sur les développements techniques ou mathématiques
Pédagogie active mêlant exposés, exercices et applications pratiques dans le logiciel 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 : 1950 € HT
- Distanciel : 1800 € HT
(-10% pour 2 inscrits, -20% dès 3 inscrits)

Option(s) :
- Forfait déjeuners : 60 € HT

Nos prochaines sessions

Distance
du 2 au 4 décembre 2024
du 2 au 4 juin 2025
du 1 au 3 décembre 2025

Lyon
du 2 au 4 avril 2025
du 29 sept. au 1 oct. 2025

Paris
du 23 au 25 avril 2025
du 3 au 5 novembre 2025

Toulouse
du 18 au 20 mars 2025
du 1 au 3 septembre 2025

Programme :

- Introduction

- Présentation de l'analyse statistique textuelle
- Domaines d'application
- Exemples d'utilisation
- Présentation du logiciel de traitement

- Importation des données

- Les diverses natures et sources de données textuelles utilisables
- Procédures d'importation selon la nature des données
- Exemples d'importation

- Codification : du texte brut au tableau statistique

- Données textuelles et données contextuelles
- Création du tableau lexical : la création des documents

- Prétraitement des données textuelles (mise en forme, lemmatisation)
- Dictionnaire des termes

-Analyse descriptive

- Calcul de la fréquence de mots : identifier les termes ou concepts les plus récurrents
- Bilan lexical par document, par variable de contexte
- Table de dissimilarité entre documents ou entre modalités de variable contexte
- Spécificités : termes sur- ou sous-représentés dans une modalité d'une variable de contexte
- Co-occurrences : termes spécifiques des documents qui contiennent un terme donné.
- Contexte dans lequel un mot est cité, permet d'éclairer le sens du texte
- Nuage de mots (« word_cloud »)

- Analyse multidimensionnelle

- Permet de révéler le sens profond des données textuelles et de synthétiser l'information contenue dans les données
- Analyse factorielle des correspondances (Examen multidimensionnel du lien des termes entre eux, avec les documents et avec les variables de contexte)
- Classification ascendante hiérarchique des documents en groupes homogènes au regard des termes et des variables de contexte

- Traitement complet d'un exemple réel

Date de dernière modification : 5 novembre 2024