

Statistique

TEXT MINING - CREATION D'UN NUAGE DE MOTS (WORDCLOUD)

Référence : 2020-STAT-N2-TextMiningWordCloud

1 jours (7h) – Présentiel ou À distance

Population visée : Technicien – Employé – Cadre – Chercheur – Etudiant

Public concerné : Toute personne souhaitant mettre en œuvre un wordcloud et avoir un premier contact avec les analyses de type Text Mining.

La formation en quelques mots

Cette formation s'adresse à des personnes souhaitant comprendre et réaliser un nuage de mots (word cloud) analyse de classique de type Text Mining.

Au-delà de la création proprement dite du nuage, une partie sera consacrée aux étapes en amont concernant la préparation des données (homogénéisation, tokenisation, stopwords,...).

Cette formation peut être vue comme une finalité en tant que telle ainsi qu'un premier contact avec les analyses liées à la fouille de données textuelles que l'on appelle le Text Mining.

A la fin de la formation, une ouverture sur des analyses statistiques Text Mining plus avancées peut-être envisagée sous forme d'échanges et de démonstrations.

Outil logiciel

Cette formation n'est pas dédiée à la pratique d'un logiciel particulier mais nous proposons de nous appuyer sur le logiciel R pour les exercices et les illustrations.

L'usage de Jmp peut également être réalisé.

Prérequis

Il n'y a pas de prérequis particulier pour le suivi de cette formation. La connaissance de base de R est nécessaire si les participants souhaitent insister sur le volet logiciel.

Objectifs pédagogiques

À l'issue de cette formation, l'apprenant sera capable de :

-  Comprendre le cheminement de la réalisation d'un nuage de mots.
-  Comprendre les étapes de préparation du dictionnaire.
-  Préparer les données en effectuant les principaux pré-traitements spécifiques aux données textuelles
-  Mettre en œuvre les procédures de tokenisation
-  Traduire l'information des mots par le nuage de mots (word cloud)

Programme

Introduction au Text Mining

- Contexte
- Exemple de données
- Application du Text Mining
- Finalité du word cloud

Particularité des données textuelles

- Multiplicité des représentations
- Fléau de la dimension
- Subjectivité des données

Représentations des données textuelles

- Sac de mots
- Vectorisation

Pré-traitements

- Tokenization
- Nettoyage (majuscule, nombre, ponctuations...)
- Stopwords
- Stemming

Visualisation

- Histogramme
- Nuage de mots

Ouverture sur les analyses avancées liées au Text Mining

- Recherche d'information
- Analyse de sentiments
- Classification de documents
- ...

Méthodes d'évaluation des acquis

L'acquisition des compétences de la formation se fait à travers le suivi du formateur tout au long de la formation (séquences synchrones et asynchrones). Elle s'appuie également sur la réalisation d'exercices et de TP. Enfin, des quiz s'ajoutent aux différents outils de validation de l'acquisition des compétences visées. Une évaluation est systématiquement réalisée par chaque stagiaire, à l'issue de la formation.

Profil formateur

Nos formateurs sont certifiés à l'issue d'un parcours organisé par nos soins. Ils bénéficient d'un suivi de maintien et d'évolution de leurs compétences aussi bien au niveau technique que pédagogique. Chacun de nos formateurs a bénéficié d'une formation spécifique à l'animation de classe virtuelle et à l'utilisation des solutions de formation à distance du Groupe ARKESYS.

Support stagiaire

À l'issue de la formation, les exercices et travaux pratiques réalisés, leurs corrigés ainsi qu'un support de cours dématérialisé sera fourni à chaque stagiaire par e-mail ou via la plate-forme FOAD.

Accessible à tous

Cette formation est accessible à toute personne en situation de handicap. Notre référent handicap prendra contact avec les stagiaires concernés pour adapter l'animation à leurs besoins et rendre l'apprentissage accessible à tous. Enfin, nos centres de formation sont accessibles aux personnes à mobilité réduite.