

Master Mathématiques appliquées, statistique

Statistique textuelle

Informations

Composante : Faculté des Sciences

Langue(s) d'enseignement

Français

Contenu

- Prétraitement de données textuelles :
 - Nettoyage de données, tokenisation, lemmatisation et racinisation
 - Extraction de caractéristiques et vectorisation des textes (sacs de mots, TF-IDF)
- Analyse exploratoire de données textuelles :
 - Visualisation de données textuelles (nuages de mots, distribution de fréquences)
 - Mesures de similarité et de distance entre textes
- Modélisation statistique de données textuelles
 - Modèles de régression appliqués aux textes, Analyse de la variance
 - Modèles de comptage et modèles probabilistes
- Analyse de sentiments
 - Techniques d'extraction des sentiments
 - Modèles de classification des sentiments
- Thèmes spéciaux en statistique textuelle
 - Analyse de réseaux textuels
 - Topic modeling (LDA, NMF)
 - Modèles de langues et N-grammes



Dernière modification le 07/10/2024

Compétences à acquérir

- Traiter et analyser de grands volumes de données textuelles
- Maîtriser un langage de programmation (R ou Python ou...) pour l'analyse de données textuelles
- Extraire des informations pertinentes à partir de textes non structurés
- Créer des visualisations et des résumés de données efficaces pour représenter des informations textuelles

Bibliographie, lectures recommandées

Ignatow, G., & Mihalcea, R. (2016). Text mining: A guidebook for the social sciences. Sage Publications.

Guérin-Pace, F. (1997). La statistique textuelle. Un outil exploratoire en sciences sociales. Population (french edition), 865-887.

Lebart, L., Pincemin, B., & Poudat, C. (2019). Analyse des données textuelles. PUQ.

Silge, J., & Robinson, D. (2017). Text mining with R: A tidy approach. O'Reilly Media

VOLUME HORAIRE

- Volume total: 24 heures
- Cours magistraux: 12 heures
- Travaux dirigés: 12 heures

Codes Apogée

- SMSBU34C [ELP]

Pour plus d'informations

[Aller sur le site de l'offre de formation...](#)