

INTITULE DE LA MATIERE : **ANALYSE DE DONNEES**

CODE : UE-64

SEMESTRE : 6

NOM DE L'ENSEIGNANT RESPONSABLE : P. BELLOT

NOMBRE D'HEURES ENCADREES CM : 16 TD : 8 TP : 16

NOMBRE D'HEURES DE TRAVAIL PERSONNEL DE L'ELEVE – estimation : 20

OBJECTIF GENERAL (pas plus de 5 lignes) :

Ce cours introduit l'analyse de données et le domaine du Big Data aussi bien du point de vue algorithmique que pratique par l'intermédiaire d'approches statistiques descriptives (ACP, partitionnement...) ou prédictives (arbres de décision, classification bayésienne) et des environnements logiciels R, Weka et Python. Un objectif est de sensibiliser aux besoins de l'industrie et de la société tout en réfléchissant aux limites des approches en terme de généralité, d'interprétabilité et de respect de la vie privée.

CONTENU – PLAN DU COURS (pas plus de 15 lignes) :

- 1- Présentation de l'analyse de données au travers d'applications concrètes exploitant des jeux de données réels (données du Web, graphes, données tabulées issues d'entrepôts ouverts). Illustrations via des interfaces humain-machine destinées à la visualisation des données et des résultats des analyses.
- 2- Approches descriptives (description uni-dimensionnelle, représentations vectorielles, corrélation, régression linéaire et logistique, méthodes de partitionnement, classification hiérarchique par similarité deux à deux).
- 3- Présentation des environnements R (TP sur les méthodes de classification)
- 4- Analyse en composantes principales (TP avec R)
- 5- Méthode de classification supervisée (arbres de décision): TP avec Weka
- 6- Initiation au langage Python et à la bibliothèque SciKit
- 7- Réalisation d'un mini-projet comparant les approches vues.

FORME DE L'EVALUATION :

Contrôle continu Contrôle terminal Mémoire/rapport Soutenance
préciser si nécessaire, nombre d'épreuves : et type oral / écrit

ACQUIS DE LA FORMATION ATTENDUS, ET QUI SONT EVALUES (5 à 10 items)

A l'issue de cet enseignement, l'élève est capable de :

Utilisez les termes comprendre, analyser, exploiter, interpréter, calculer, utiliser, mettre en œuvre, choisir, communiquer, rédiger,...

- Analyser et modéliser un problème réel sous la forme d'une analyse de données descriptive ou prédictive (choix des données, représentation machine à partir de données brutes, définition des sorties attendues)
- Choisir une approche en fonction des objectifs et déterminer les métriques pour l'évaluer
- Représenter et interpréter des résultats obtenus (des données à la connaissance)
- Utiliser et exploiter des environnements logiciels d'analyse de données

PREREQUIS (pas plus de 5 lignes): Bases de statistiques et de calcul matriciel

REFERENCES, BIBLIOGRAPHIE (pas plus de 5 lignes) :

- Data Mining et Statistique décisionnelle : La science des données. Editions Technip, 2017
- Data Science : fondamentaux et études de cas: Machine Learning avec Python et R. Eyrolles, 2015
- Data mining et statistique décisionnelle, S. Tuffery, TECHNIP, 2012.

- Le langage R au quotidien : Traitement et analyse de données volumineuses. Mise en pratique avec exemples en Open Data (Applications et métiers), O. Decourt, Dunod, 2018
- Analyse de données textuelles sous R, N. Turenne, iSTE Editions, 2015