

STA211 - Entreposage et fouille de données

Présentation

Prérequis

être admis en M2 du master STIC, mention statistique ou être agrégé.

La demande d'agrément est à faire uniquement pour les auditeurs qui souhaitent suivre STA211 en présentiel au semestre 1. Pour obtenir l'agrément, les auditeurs adresseront, à partir du 25 Août, par courrier électronique à l'enseignant responsable : ndeye.niang_keita@cnam.fr, un CV détaillé et une lettre de motivation indiquant les raisons de la demande et le projet pédagogique dans lequel elle s'inscrit. Une réponse sera donnée dans un délai d'une dizaine de jours.

Niveau requis : NFA008 (bases de données) et STA101 (analyse des données, méthodes descriptives). **Ces prérequis sont indispensables pour obtenir l'agrément. Vérifier avant d'envoyer une demande d'agrément.**

Programme

Contenu

Modèles prévisionnels et systèmes de gestion de l'entreprise

- structures spécifiques des bases de données de Data warehouse (star schema)
- OLAP

Méthodologies générales

- Méthodologies de Data Mining

Pré-traitement des données

- Analyses de la qualité des données,
- Techniques d'appréhension des valeurs manquantes ou aberrantes
- Techniques de construction de bases de travail (agrégations, etc. . .)

Données et techniques de fouille

Méthodes non supervisées :

- Cartes de Kohonen, Règles d'association

Méthodes supervisées :

- Rappels de théorie de l'apprentissage
- Arbres de décision, forêts aléatoires, Réseaux de neurones, deep learning
- Méta-algorithmes :
- boosting, bagging

Fouille dans de nouveaux types de données et méthodes associées :

- Données textuelles - Données multivues - Images et Multimedia

Outils :

- Environnements freeware : R, Python
- Outils spécifiques : SAS-EM, SPAD
- Data Mining et bases de données : OLAP Business Object

Modalités de validation

- Projet(s)
- Mémoire

Bibliographie

Titre	Auteur(s)
Analyse discriminante (Dunod, 2001)	M.BARDOS
Probabilités, analyse des données et statistique (Technip, 2006)	G.SAPORTA

Mis à jour le 09-04-2024



Code : STA211

Unité d'enseignement de type cours

9 crédits

Volume horaire de référence (+/- 10%) : **70 heures**

Responsabilité nationale :

EPN06 - Mathématiques et statistique / 1

Contact national :

EPN06 Mathématiques et statistiques

2 rue conté

Accès 35 3 ème étage porte 19
75003 Paris

Sabine Glodkowski

sabine.glodkowski@lecnam.net

Data mining et statistique décisionnelle (Technip, 2005)	S.TUFFERY
Etude de cas en statistique décisionnelle (Technip, 2009)	S.TUFFERY
Elements of Statistical Learning (Springer, 2009)	T.HASTIE, J.FRIEDMAN, F.TIBSHIRANI
Analyse des données (Hermes,2003)	G.GOVAERT (ed)
Statistique exploratoire multidimensionnelle (1995)	L. LEBART, A. MORINEAU, M. PIRON
Statistique explicative appliquée (Technip, 2003)	J.P.NAKACHE, J.CONFAIS
An Introduction to Statistical Learning (2013) Téléchargeable ici: http://web.stanford.edu/~hastie/local.ftp/Springer/ISLR_print1.pdf	James, Witten, Hastie, & Tibshirani