

SYLLABUS

Domaine :. Filière : Gestion et Analyse des Données Massives GADM
Spécialité : ...**Informatique**...Semestre : **M1** Année scolaire : 2021-2022

Identification de la matière d'enseignement

Intitulé :Fouille de données.....

Unité d'enseignement: ...UEF1.....

Nombre de Crédits: 06 Coefficient : 03

Volume horaire hebdomadaire total :

- Cours (nombre d'heures par semaine) :2h00.....
- Travaux dirigés (nombre d'heures par semaine) :
- Travaux pratiques (nombre d'heures par semaine) : ...1H00.....

Responsable de la matière d'enseignement

Nom, Prénom, Grade :Professeur AZIZI Nabih.....

Localisation du bureau (Bloc, Bureau) : Département d'informatique, bureau 12

Email :azizi@labged.net.....

Tel (Optionnel) :

Horaire du cours et lieu du cours :Dimanche 8h00 à 11h30

Description de la matière d'enseignement

Prérequis : Base de données, analyse de données, notions algorithmiques

Objectif général du la matière d'enseignement

et **Objectifs d'apprentissage :** (de 3 à 6 objectifs, n'inclure que les objectifs que vous pouvez évaluer)

- Maitriser les techniques statistiques pour explorer et valoriser les données, déduire des nouvelles connaissances à partir des donnée de départ
- Structuration et accès aux données, manipulations des logiciels
- Identifier la technique adéquate face à un problème à résoudre, de traiter les données à l'aide d'un classificateur, de produire des résultats commentés, et d'évaluer la portée réelle de ces résultats, notamment par la validation en analysant les mesures de performances.

Contenu de la matière d'enseignement

La fouille de données peut être définie comme l'ensemble d'approches statistiques permettant d'extraire de l'information de grands jeux de données dans une perspectives d'aide à la décision.

Ce concept a aujourd'hui une grande importance économique du fait qu'elle permet d'optimiser la gestion des ressources (humaines et matérielles).

L'aboutissement des avancées liées aux Bases de Données et à l'Intelligence Artificielle au sens de l'apprentissage. C'est une discipline émergente mais à fort potentiel en pratique.

Dans ce cours, nous nous concentrerons sur l'aspect algorithmique et logiciel de système de classification et de pré traitement de données , ainsi que leur utilisation. A la fin de ce cours, l'étudiant connaîtra les notions d'apprentissage non supervisé et supervisé, les algorithmes afférant, leur utilisation sur des cas pratiques ainsi que le mesures calculant les performances globales et locales. Un outil fédérateur sera particulièrement utilisé : le logiciel libre et évolutif Weka programmé en Java.

Le processus complet de fouille de données comprend plusieurs étapes :

1. nettoyage et pré traitement de la base de données : normalisation ;
2. extraction d'information pertinente eprésentant chaque individu de la base de dnnées
3. sélection des attributs utiles ;
5. Apprentissage et classification selon la technique adoptée; cette dernière dépend de la nature des données (étiquetée, non étiquetées ou semi étiquetées)
6. Evaluation des résultats Obtenus.

Un mini-projet représenté sous forme d'exposé permettront de concrétiser les aspects théorique étudiés et de rendre opérationnelles les connaissances enseignées

Modalités d'évaluation

Nature du contrôle	Pondération en %
Examen	60%
Micro – interrogation	
Travaux dirigés	
Travaux pratiques	
Projet personnel	40%
Travaux en groupe	
Sorties sur terrains	
Assiduité (Présence /Absence)	
Autres (à préciser)	
Total	100%

Références & Bibliographie

Textbook (Référence principale) :		
Titre de l'ouvrage	Auteur	Éditeur et année d'édition
Data Mining: Practical Machine Learning Tools and Techniques	Ian H. Witten	3 ème édition 20 décembre 2016
Les références de soutien si elles existent :		
Titre de l'ouvrage (1)	Auteur	Éditeur et année d'édition
DATAMINING Quel auteur doit-on éditer ?	Monsieur DIDAY	Johanna GARCIA 2005
Titre de l'ouvrage (2)	Auteur	Éditeur et année d'édition

Planning du déroulement du cours

Semaine	Titre du Cours	Date
01 semaine	Concepts introductifs :Des Bases de Données vers la Fouille de Données– Rappels et notion de connaissance:	
02 semaine	Processus général de fouille de données et type d'apprentissage : non supervisé , supervisé et semi supervisé	
03 semaine	- Etude de performance d'un modèle ou classifieur (1) - Approche classique base d'apprentissage et base de test- - La validation croisée	
04 semaine	Etude de performance d'un modèle ou classifieur (1) - Les mesures de performances globales - Les mesures de performances semi globales - Notion de rejet - Exercice de calcul des mesures locale dans le but d'analyser le comportement du modèle et proposer les amélioration nécessaires .	
05 semaine	Analyser du contenu des bases de données Pour la standardisation ou normalisation : - Notion de similarité entre données - Type de distances , - Application de l'approche z-score pour la normalisation	
06 semaine	Apprentissage Non Supervisé : Clustering 1- L'algorithme k-means: - Concept, algorithme, étude de cas uni, bi et multi dimensionnel de données - Avantage et limites Travaux dirigés	
07 ème semaine	Apprentissage Non Supervisé : Clustering (suite) 2- Classification ascendante hiérarchique Concept , notion de distance inter groupes et intr	

	groupe Calcul de distance entre groupe (distance de Ward) Construction de dendrogrammes Exercices pratiques	
8 ème semaine	Apprentissage Supervisé 1- Les réseaux de neurones Concepts généraux : notion de neurones , fonction de transfert et poids inter neuronal Le réseaux de neurones avec la fonction seuil	
9ème semaine	Les réseaux de neurones (suite) Le perceptron multi-couches (PMC) Apprentissage dans les PMC Les réseaux de neurone multi niveaux Exercices	
10ème semaine	Apprentissage Supervisé Les classifieurs à noyaux Les machine à vecteurs de support SVM	
11 et 12 semaines	soutenance orale du projet	
	Examen de fin de semestre	
	Examen de rattrapage	