

Module : Apprentissage et Fouille de données

Code

ING-5-GLSI-S9-P3

Période	S9	Volume horaire	21h	ECTS	2
---------	-----------	----------------	------------	------	----------

Responsable	Sonia Gharsalli	email	Sonia.gharsalli@tek-up.tn
-------------	------------------------	-------	---------------------------

Equipe pédagogique	Sonia Gharsalli
--------------------	-----------------

1. Objectifs de Module (Savoirs, aptitudes et compétences)

L'objectif de ce cours est d'introduire la fouille des données et de présenter les techniques de base et les techniques avancées pour l'analyse des ensembles fréquents et des règles d'association. Ainsi, que la reconnaissance des motifs séquentiels fréquents.

Acquis d'apprentissage :

A la fin de cet enseignement, l'élève sera capable de :

- Maîtriser les notions de bases de fouille de données et savoir les différentes techniques sur lesquels se basent le processus de fouille de données (**C1.2**)
- Concevoir des systèmes de recommandation de type (item based filtering) (**C1.1**)
- Concevoir des solutions pour l'extraction de connaissance (**C1.2**)
- Développer un sens de critique pour le choix des algorithmes utilisés (**C1.3**)
- Appliquer les connaissances acquises pour le développement des systèmes de recommandation spécialisés (**C1.4**)

2. Pré-requis(autres UE et compétences indispensables pour suivre l'UE concernée)

- programmation

3. Répartition d'Horaire de Module

Intitulé de l'élément d'enseignement	Total	Cours	TD	Atelier	PR
Module : Data Mining	21	12	9	0	0

4. Méthodes pédagogiques et moyens spécifiques au Module

(pédagogie d'enseignement, ouvrages de références, outils matériels et logiciels)

- Supports de Cours
- Projecteur et Tableau

- Travaux dirigés

Bibliographie

Titre	Auteur(s)	Edition
Introduction to Data Mining	P. Tan, M. Steinbach, V. Kumar	Pearson, 2005
Data Mining and Data Science	R. Rakotomalala	springer
PrefixSpan: Mining Sequential Patterns Efficiently by Prefix-Projected Pattern Growth	J. Jian, P. Jiawei Han, B. Mortazavi-Asl, H. Pinto, C. Qiming, U. Dayal, H. Mei-Chun	Proceedings 17th International Conference on Data Engineering IEEE, 2002

5. Contenu (Descriptifs et plans des cours / Déroulement / Détail de l'évaluation de l'activité pratique)

Durée allouée

Séance 1
Objectif 1 : Introduction à la fouille des données

Cours 3H

Eléments de contenu :

- Introduction à la fouille des données par des applications introductives
- Spécificité de la fouille des données
- Typologies des données

Séance 2

TD 3H

- Applications des notions de bases

Séance 3 :

Cours 3H

Objectif 2 : Maîtriser les techniques de détection des itemsets fréquents et de la construction des règles d'association

Eléments de contenu :

- Introduction à la notion de règles d'association et aux critères d'évaluation
- Présentation des algorithmes à priori
 - Présentation de l'algorithme FP-Growth

Séance 4

TD 3H

- Exercice Sur les deux algorithmes (A priori et FP-Growth)

Séance 5

Cours 3H

Objectif 3 : Maitriser les techniques de détection des motifs séquentiels fréquents

Eléments de contenu :

- Introduction à la notion de motifs séquentiels fréquents
- Algorithme Apriori : GSP algorithm
- Algorithme Pattern-Growth :
 - Free-Span
 - PrefixSpan

Séance 6

TD 3H

- Exercices sur la recherche de motif fréquents par les deux algorithmes (GSP, PrefixSpan)

Séance 7

Application des règles d'association dans un système de recommandation.

Cours

3H

6. Mode d'évaluation de Module(nombre, types et pondération des contrôles)

Eléments d'enseignement	Coeff	DS	EX	TP	PR
Apprentissage et Fouille de données	1	40%	60%	0%	0%

Pour valider le module, les étudiants passeront un examen dont le coefficient est de 60%, un DS dont le coefficient est de 40%.

La durée de tous les examens (Examen, DS...) est de 1h30.

Le DS est planifié 4 semaines après le début du module et portera sur les thématiques suivantes :

- la fouille des données par des applications introducives
- Spécificité de la fouille des données
- Typologies des données
- les techniques de détection

Quant à l'examen, il est planifié après l'écoulement des 7 semaines et portera sur toutes les thématiques enseignées tout au long des 21 heures.

Le module est validé si l'étudiant obtient une moyenne supérieure ou égal à 10 sur 20.