Web Mining

Course title – Intitulé du cours	Web Mining
Level / Semester – Niveau /semestre	M2/S2
School – Composante	Ecole d'Economie de Toulouse
Teacher – Enseignant responsable	TAMINE-LECHANI Lynda
Other teacher(s) – Autre(s) enseignant(s)	
Other teacher(s) – Autre(s) enseignant(s)	
Other teacher(s) – Autre(s) enseignant(s)	
Other teacher(s) – Autre(s) enseignant(s)	
Other teacher(s) – Autre(s) enseignant(s)	
Lecture Hours – Volume Horaire CM	18
TA Hours – Volume horaire TD	
TP Hours – Volume horaire TP	
Course Language – Langue du cours	English
TA and/or TP Language – Langue des TD et/ou TP	English

Teaching staff contacts – Coordonnées de l'équipe pédagogique :

Lynda Tamine – Lechani, lynda.lechani@irit.fr

Office number : IRIT- 410

Course Objectives – Objectifs du cours :

Web Mining refers to data mining techniques specifically applied on web data. Web mining techniques are commonly classified into 3 subsets based on the facets that characterize web data, namely content, structure and usage. This course will focus on the social web (web 3.0), viewed as a combination of information space and social space. It will provide an overview of the following techniques illustrated through the concepts and algorithms underlying usual and well-known problems and/or applications.

1. ***Web Content Mining***: is concerned with the analysis of the web content. In this course, we will focus on textual content and will address applications such as information retrieval, opinion mining and sentiment analysis.
2. ***Web Structure Mining***: refers to the computational models of web social networks structure in terms of both the entities (eg., users, documents, items, etc.) that populate those networks and the links that exist between them (eg., friendship, citation, likes, etc.). We will address problems such as the measurement of entity influence in the network, the extraction of user communities and the prediction of information diffusion over the network.
3. ***Web Usage Mining***: consists in the analysis of traces left by users on the Web (eg., clicks, likes, comments, etc.) to support personalized services. We will particularly address the problem of user profiling based on navigational data including clicks and likes.

Prerequisites – Pré requis :

- Machine Learning (basics)
- Graph theory
- Python for Data Science

Practical information about the sessions – Modalités pratiques de gestion du cours:

Most sessions will begin with a presentation of the session topic and end with a practical application of the concepts introduced.

It is assumed that students have the prerequisites in Python for data science.

Students are allowed to make use of laptops and tablets.

Grading system – Modalités d'évaluation :

50% of the grade is a final exam on the theoretical part with exercises. The remaining 50% of the grade is a project carried out per group of students.

Bibliography/references – Bibliographie/références :

- Web Data Mining - Exploring Hyperlinks, Contents, and Usage Data by Bing Liu
- Mining the Social Web by Matthew A Russell and Mikhail Klassen
- Python Data Science Handbook by Jake VanderPlas

Session planning – Planification des séances

1. Introduction (2 sessions X 2H)
   Web content mining 1
   - Definitions, applications
   - Text processing
   - Basics of information retrieval
2. Web content mining 2 (1 session X 2H)
   - Opinion mining and sentiment analysis
3. Web structure mining 1 (2 sessions X 2H)
   - Basic concepts
   - Social influence
4. Web structure mining 2 (3 sessions X 2H)
   - Community detection
   - Information diffusion
5. Web usage mining (1 session X 2H)
   - User profiling

Distance learning – *Enseignement à distance :*

Distance learning can be provided when necessary by implementing:

- *Interactive virtual classrooms*

- *Remote (online) tutorials (classes)*
- *Chatrooms*
- *Forum...*