INSTITUT NATIONAL
DES SCIENCES
APPLIQUÉES
LYON

DPT INFORMATIQUE
INFORMATIQUE APPRENTISSAGE
INSA Campus LyonTech - 7 Avenue Jean Capelle
Bâtiment Blaise Pascal - 69100 VILLEURBANNE

Knowledge Engineerin

Data mining

IDENTIFICATION

CODE :	IFA-4-S2-EC-FD
ECTS :	2.0

HOURS

Lectures :	10.0 h
Seminars :	4.0 h
Laboratory :	12.0 h
Project :	0.0 h
Teacher-student contact :	26.0 h
Personal work :	25.0 h
Total :	51.0 h

ASSESSMENT METHOD

Final exam (1h30) where every document on paper can be used.

Programming project where a real dataset has to be mined. A report is expected. Implemented KNIME workflows can be asked as well.

TEACHING AIDS

Slides of lessons are disseminated.

Two seminars are dedicated to KNIME practice on toy data sets.

Two weeks are targeted to a programming project on a real data set. Tutoring and a forum can help on Moodle.

TEACHING LANGUAGE

French

CONTACT

M. BENTO Alexandre
alexandre.bento@insa-lyon.fr
MME NURBAKOVA Diana
diana.nurbakova@insa-lyon.fr

AIMS

Data mining was identified as one of the top ten emerging technologies for the 21st century (MIT Technology Review, 2001). The goal of this discipline is to support the discovery of knowledge from a large volume of data, typically data warehouses. Its development was built at the intersection of several existing disciplines in data processing, for example, machine learning, database management, visual display and statistics. The main data mining techniques are introduced (statistical techniques like PCA, supervised classification or unsupervised classification, pattern discovery methods).

We expect that after this module, students are able to explore real data sets, perform cleaning tasks, looking for patterns with an emphasis on cluster discovery within real data. We expect that they understand how to choose a given algorithm and how to determine relevant parameters for them. This involves also the practice of discovery processes by means of the open source platform KNIME. Students are expected to understand, use and adapt typical data analysis workflows prepared for KNIME.

As such, you will acquire the following skills:
- Learn the basic of Knowledge discovery in all its aspects, from data cleaning to model interpretation
- Learn several techniques for supervised classification, clustering and pattern discovery
- Be able to discuss the choice of a data analysis algorithm and its parameters
- Be able to use a data analysis platform (KNIME) for a real-world knowledge discovery problem

CONTENT

The main data mining tasks are introduced. The concepts are illustrated during two exercise sessions (1 on data exploration and aand 1 on data mining, both based on the use of the open source platform KNIME) and a 2-weeks project.

Class 1. Motivations and terminology
Class 2. Data exploration
Class 3. Clustering
Class 4. Prediction and supervised classification
Class 5. Computing pattern and descriptive rules
Class 6. Knowledge Discovery Processes

Some popular data mining algorithms are detailed like K-Means, DBSCAN, C4.5, NB, APRIORI (non exhaustive list). Important issues related to predictive tasks and machine learning are just sketched with decision trees (advanced concepts and methods for Big Data Analytics are studied during the 5IF first semester).

The project concerns geo-localized data analysis where localized objects are photos associated to tags. By computing clusters of photos, we expect to be able to discover automatically points of interest within a city. Pattern mining techniques will be used to help interpreting the found clusters.

BIBLIOGRAPHY

INSA LYON
Campus LyonTech La Doua
20, avenue Albert Einstein - 69621 Villeurbanne cedex - France
Phone +33 (0)4 72 43 83 83 - Fax +33 (0)4 72 43 85 00
www.insa-lyon.fr

- U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy. Advances in Knowledge Discovery and Data Mining. AAAI/MIT Press, 1996.
- J. Han and M. Kamber. Data Mining: Concepts and Techniques. Morgan Kaufmann, 2000.
- D. Hand, H. Mannila, P. Smyth. Principles of Data Mining. MIT Press, 2001.
- P. N. Tan, M. Steinbach, and V. Kumar. Introduction to Data Mining. Addison-Wesley, 2006.
- M. R. Berthold, C. Borgelt, F. Hoppner, F. Klawonn. Guide to Intelligent Data Analysis, Springer, 2010.
- M. J. Zaki, W. Meira Jr. Fundamentals of Data Mining Algorithms. Cambridge Univeristy Press, 2013.
- A. Cornuéjols et L. Miclet. Apprentissage Artificiel. Concepts et Algorithmes. Seconde version, Eyrolles, 2010.

PRE-REQUISITE

Basic statistics and mathematics, relational databases, SQL, programming