Course guide
270539 - TMD - Data Mining Techniques

Last modified: 12/07/2022

Unit in charge:	Barcelona School of Informatics
Teaching unit:	715 - EIO - Department of Statistics and Operations Research.

Degree:	MASTER'S DEGREE IN INFORMATICS ENGINEERING (Syllabus 2012). (Optional subject).

Academic year: 2022   **ECTS Credits:** 3.0   **Languages:** Catalan

LECTURER

Coordinating lecturer:	CARINA GIBERT OLIVERAS
Others:	Primer quadrimestre: CARINA GIBERT OLIVERAS - 10

DEGREE COMPETENCES TO WHICH THE SUBJECT CONTRIBUTES

Specific:
CDG1. Capability to integrate technologies, applications, services and systems of Informatics Engineering, in general and in broader and multicisciplinary contexts.
CTE9. Capability to apply mathematical, statistical and artificial intelligence methods to model, design and develop applications, services, intelligent systems and knowledge-based systems.

Generical:
CG8. Capability to apply the acquired knowledge and to solve problems in new or unfamiliar environments inside broad and multidisciplinary contexts, being able to integrate this knowledge.

Transversal:
CTR2. SUSTAINABILITY AND SOCIAL COMMITMENT : Capability to know and understand the complexity of the typical economic and social phenomena of the welfare society. Capacity for being able to analyze and assess the social and environmental impact.
CTR3. TEAMWORK: Capacity of being able to work as a team member, either as a regular member or performing directive activities, in order to help the development of projects in a pragmatic manner and with sense of responsibility; capability to take into account the available resources.
CTR4. INFORMATION LITERACY: Capability to manage the acquisition, structuring, analysis and visualization of data and information in the area of informatics engineering, and critically assess the results of this effort.
CTR5. APPROPIATE ATTITUDE TOWARDS WORK: Capability to be motivated by professional achievement and to face new challenges, to have a broad vision of the possibilities of a career in the field of informatics engineering. Capability to be motivated by quality and continuous improvement, and to act strictly on professional development. Capability to adapt to technological or organizational changes. Capacity for working in absence of information and/or with time and/or resources constraints.
CTR6. REASONING: Capacity for critical, logical and mathematical reasoning. Capability to solve problems in their area of study. Capacity for abstraction: the capability to create and use models that reflect real situations. Capability to design and implement simple experiments, and analyze and interpret their results. Capacity for analysis, synthesis and evaluation.

Basic:
CB6. Ability to apply the acquired knowledge and capacity for solving problems in new or unknown environments within broader (or multidisciplinary) contexts related to their area of study.
CB7. Ability to integrate knowledges and handle the complexity of making judgments based on information which, being incomplete or limited, includes considerations on social and ethical responsibilities linked to the application of their knowledge and judgments.
CB8. Capability to communicate their conclusions, and the knowledge and rationale underpinning these, to both skilled and unskilled public in a clear and unambiguous way.
CB9. Possession of the learning skills that enable the students to continue studying in a way that will be mainly self-directed or autonomous.

TEACHING METHODOLOGY

The course uses a mixt methodology of case-based learning and project-based learning

The first week the fundamentals of the matter will be given and the activities to be developed by the student to guarantee the learning process will be assigned. Basically two activities: a paper reading activity regarding a Data Mining application and the developement of a Data Mining project in a working team.

In the following weeks, the structure will be as described bellow:
Every week two hours will be devoted to a case presentation, including the whole steps of development (preprocessing, analisys, postprocessing and validation). In part of the third hour the students will give synthetic presentations of complementary cases to be documented individually. The remaining part of third hour and forth hour, lab activities will be followed related with the project to be developed by every working team.

Together with the acquisition of tecnical skills directly related with Data Mining, an important goal of the course is to provide to the student transversal skills considered relevants for the professional developement, like team-working capacity, long-term planning skills, oral, visual and written communication skills, synthesis skills, justifying decisions made during the project, incidence management skills, knowledge integration for building solutions to high complex problems. The activities scheduled during the course have been especially designed to this purpose.

Last week of the course, every project will be presented and followed by a discussion, usefull as oral examination. The lecturer will use last hour of the course to highlight commonalities and particularities of the presented projects related with the basic schemes of a Data Mining project. Common discussionwill follow on what students understood about usefulness of Data Mining in Computer Engineering, this completing the general message of the course

LEARNING OBJECTIVES OF THE SUBJECT

STUDY LOAD

Type	Hours	Percentage
Hours large group	13,5	18.00
Hours small group	13,5	18.00
Self study	48,0	64.00

Total learning time: 75 h

CONTENTS

Introduction. Data Mining origins, steps, Statistics and Artificial Intelligence

Description:
Data Mining is placed in the historical context.
The overall process of Knowledge Discovery from Databases is presented, together with its steps and including Data Mining itself.
The disciplinary pillars of Data Mining are introduced: Statistics and Artificial Intelligence, Information Systems and Data Visualization

Scope and tools

Description:
Different natures of real problems and their different levels of complexity are discussed according to the classification proposed by Simpson. . Ill-structured domains are introduced, as well as a priori and implicit knowledge management, causes and consequences.
Some software tools for developing data mining tasks are introduced.

Method Selection. Typology of problems (DMMCM)

Description:
The course follows a problem-oriented KDD approach, where the nature of the problem mainly determines the analysis process. Factors determining a correct choice of data mining method in real cases are presented. The DMMCM typology of methods is presented as a conceptual basis for selection.

Data, Metadata

Description:
Main data structures analyzed by Data Mining techniques.
Importance of metadata, formats and contents

Preprocessing

Description:
Brief introduction of relevant aspects in data preparation step: Missing data, outliers detection and treatment, derived variables, transformed variables, filtering, sampling, feature weighting, dimensionality reduction. Good practice guidelines will be provided

Data Mining Descriptive methods

Description:
Statistical clustering: partitional methods, hierarchical methods, density-based, model-based, scalability; Conceptual Clustering (IA); Hybrid AI&Stats methods: clustering based on rules. Case OMS: mental health systems

Associative Data Mining methods

Description:
Association rules induction. Factorial methods. Bayesian Networks.

Predictive Data Mining methods

Description:
Regressión, statistical modelling in general. Temporal methods, Artificial Neural Networks, Swarm Intelligence.

Data Mining Discriminant methods

Description:
Decision trees, rule induction, support vector machines, Random Forest. discriminant analysis, hybrid methods. Case elderly people functioning and profiles assessment grid

Space-temporality

Description:
Introduction of some tools to manage data including simultanoeulsy spatial information changing over time. Case Quality of Life Guttmann

Post-processing and validation

Description:
Post-processing tools and validation tools for both models and results adapted to different Data Mining methods. Case wastewater treatment

Conclusion

Description:
All the elements seen during the course will be placed over the general scheme of the Knowledge Discovery process presented in section 1, as a global synthesis of the course

ACTIVITIES

Paper reading

Description:
A paper from an impact journal about a real data mining application will be selected. The paper can be proposed by both the student or the lecturer. The student must read and understand the process of Knowledge Discovery used in the applicationwith all its components. A form with this information must be filled-in.

Specific objectives:
6, 7

Related competencies :
CG8. Capability to apply the acquired knowledge and to solve problems in new or unfamiliar environments inside broad and multidisciplinary contexts, being able to integrate this knowledge.
CDG1. Capability to integrate technologies, applications, services and systems of Informatics Engineering, in general and in broader and multicisciplinary contexts.
CTR5. APPROPIATE ATTITUDE TOWARDS WORK: Capability to be motivated by professional achievement and to face new challenges, to have a broad vision of the possibilities of a career in the field of informatics engineering. Capability to be motivated by quality and continuous improvement, and to act strictly on professional development. Capability to adapt to technological or organizational changes. Capacity for working in absence of information and/or with time and/or resources constraints.
CTR6. REASONING: Capacity for critical, logical and mathematical reasoning. Capability to solve problems in their area of study. Capacity for abstraction: the capability to create and use models that reflect real situations. Capability to design and implement simple experiments, and analyze and interpret their results. Capacity for analysis, synthesis and evaluation.
CTR2. SUSTAINABILITY AND SOCIAL COMMITMENT : Capability to know and understand the complexity of the typical economic and social phenomena of the welfare society. Capacity for being able to analyze and assess the social and environmental impact.
CTR4. INFORMATION LITERACY: Capability to manage the acquisition, structuring, analysis and visualization of data and information in the area of informatics engineering, and critically assess the results of this effort.
CB8. Capability to communicate their conclusions, and the knowledge and rationale underpinning these, to both skilled and unskilled public in a clear and unambiguous way.
CB9. Possession of the learning skills that enable the students to continue studying in a way that will be mainly self-directed or autonomous.
CB7. Ability to integrate knowledges and handle the complexity of making judgments based on information which, being incomplete or limited, includes considerations on social and ethical responsibilities linked to the application of their knowledge and judgments.

Full-or-part-time: 8h
Laboratory classes: 4h
Self study: 4h

(ENG) Definició del projecte en equip

Description:
(ENG) Per grups, els estudiants triaran un tema i unes dades sobre les que resoldre un problema de Mineria de Dades

Specific objectives:
(ENG) 2, 8, 9

Material:
(ENG)

Delivery:
(ENG)

Related competencies :
CG8. Capability to apply the acquired knowledge and to solve problems in new or unfamiliar environments inside broad and multidisciplinary contexts, being able to integrate this knowledge.
CDG1. Capability to integrate technologies, applications, services and systems of Informatics Engineering, in general and in broader and multicisciplinary contexts.
CTR5. APPROPIATE ATTITUDE TOWARDS WORK: Capability to be motivated by professional achievement and to face new challenges, to have a broad vision of the possibilities of a career in the field of informatics engineering. Capability to be motivated by quality and continuous improvement, and to act strictly on professional development. Capability to adapt to technological or organizational changes. Capacity for working in absence of information and/or with time and/or resources constraints.
CTR3. TEAMWORK: Capacity of being able to work as a team member, either as a regular member or performing directive activities, in order to help the development of projects in a pragmatic manner and with sense of responsibility; capability to take into account the available resources.
CTR6. REASONING: Capacity for critical, logical and mathematical reasoning. Capability to solve problems in their area of study. Capacity for abstraction: the capability to create and use models that reflect real situations. Capability to design and implement simple experiments, and analyze and interpret their results. Capacity for analysis, synthesis and evaluation.
CTR2. SUSTAINABILITY AND SOCIAL COMMITMENT : Capability to know and understand the complexity of the typical economic and social phenomena of the welfare society. Capacity for being able to analyze and assess the social and environmental impact.
CTR4. INFORMATION LITERACY: Capability to manage the acquisition, structuring, analysis and visualization of data and information in the area of informatics engineering, and critically assess the results of this effort.
CB8. Capability to communicate their conclusions, and the knowledge and rationale underpinning these, to both skilled and unskilled public in a clear and unambiguous way.
CB6. Ability to apply the acquired knowledge and capacity for solving problems in new or unknown environments within broader (or multidisciplinary) contexts related to their area of study.

Full-or-part-time: 2h
Laboratory classes: 2h

(ENG) Presentació Control de projectes en equip

Description:
(ENG) Cada grup presentarà en públic el plantejament del seu projecte. Descripció del projecte, objectius, estructura, contingut i origen de les dades, disseny del procés de Data Mining a aplicar, pla de treball

Specific objectives:
(ENG) 1, 2, 3, 8, 9, 10

Material:
(ENG)

Delivery:
(ENG)

Related competencies :
CG8. Capability to apply the acquired knowledge and to solve problems in new or unfamiliar environments inside broad and multidisciplinary contexts, being able to integrate this knowledge.
CTE9. Capability to apply mathematical, statistical and artificial intelligence methods to model, design and develop applications, services, intelligent systems and knowledge-based systems.
CDG1. Capability to integrate technologies, applications, services and systems of Informatics Engineering, in general and in broader and multicisciplinary contexts.
CTR5. APPROPIATE ATTITUDE TOWARDS WORK: Capability to be motivated by professional achievement and to face new challenges, to have a broad vision of the possibilities of a career in the field of informatics engineering. Capability to be motivated by quality and continuous improvement, and to act strictly on professional development. Capability to adapt to technological or organizational changes. Capacity for working in absence of information and/or with time and/or resources constraints.
CTR3. TEAMWORK: Capacity of being able to work as a team member, either as a regular member or performing directive activities, in order to help the development of projects in a pragmatic manner and with sense of responsibility; capability to take into account the available resources.
CTR6. REASONING: Capacity for critical, logical and mathematical reasoning. Capability to solve problems in their area of study. Capacity for abstraction: the capability to create and use models that reflect real situations. Capability to design and implement simple experiments, and analyze and interpret their results. Capacity for analysis, synthesis and evaluation.
CTR2. SUSTAINABILITY AND SOCIAL COMMITMENT : Capability to know and understand the complexity of the typical economic and social phenomena of the welfare society. Capacity for being able to analyze and assess the social and environmental impact.
CTR4. INFORMATION LITERACY: Capability to manage the acquisition, structuring, analysis and visualization of data and information in the area of informatics engineering, and critically assess the results of this effort.
CB8. Capability to communicate their conclusions, and the knowledge and rationale underpinning these, to both skilled and unskilled public in a clear and unambiguous way.
CB6. Ability to apply the acquired knowledge and capacity for solving problems in new or unknown environments within broader (or multidisciplinary) contexts related to their area of study.

Full-or-part-time: 11h
Laboratory classes: 4h
Self study: 7h

(ENG) Presentació final del projecte en equip

Description:
(ENG) Cada grup entregarà l'informe de la pràctica i presentarà als seus companys els resultats de l'aplicació de mineria de dades desenvolupada. Hi haurà debat i discussió amb el professor sobre les decisions preses al llarg del projecte

Specific objectives:
(ENG) 1, 2, 3, 4, 5, 6, 8, 9

Material:
(ENG)

Delivery:
(ENG)

Related competencies :
CG8. Capability to apply the acquired knowledge and to solve problems in new or unfamiliar environments inside broad and multidisciplinary contexts, being able to integrate this knowledge.
CTE9. Capability to apply mathematical, statistical and artificial intelligence methods to model, design and develop applications, services, intelligent systems and knowledge-based systems.
CDG1. Capability to integrate technologies, applications, services and systems of Informatics Engineering, in general and in broader and multicisciplinary contexts.
CTR5. APPROPIATE ATTITUDE TOWARDS WORK: Capability to be motivated by professional achievement and to face new challenges, to have a broad vision of the possibilities of a career in the field of informatics engineering. Capability to be motivated by quality and continuous improvement, and to act strictly on professional development. Capability to adapt to technological or organizational changes. Capacity for working in absence of information and/or with time and/or resources constraints.
CTR3. TEAMWORK: Capacity of being able to work as a team member, either as a regular member or performing directive activities, in order to help the development of projects in a pragmatic manner and with sense of responsibility; capability to take into account the available resources.
CTR6. REASONING: Capacity for critical, logical and mathematical reasoning. Capability to solve problems in their area of study. Capacity for abstraction: the capability to create and use models that reflect real situations. Capability to design and implement simple experiments, and analyze and interpret their results. Capacity for analysis, synthesis and evaluation.
CTR2. SUSTAINABILITY AND SOCIAL COMMITMENT : Capability to know and understand the complexity of the typical economic and social phenomena of the welfare society. Capacity for being able to analyze and assess the social and environmental impact.
CTR4. INFORMATION LITERACY: Capability to manage the acquisition, structuring, analysis and visualization of data and information in the area of informatics engineering, and critically assess the results of this effort.
CB8. Capability to communicate their conclusions, and the knowledge and rationale underpinning these, to both skilled and unskilled public in a clear and unambiguous way.
CB6. Ability to apply the acquired knowledge and capacity for solving problems in new or unknown environments within broader (or multidisciplinary) contexts related to their area of study.
CB7. Ability to integrate knowledges and handle the complexity of making judgments based on information which, being incomplete or limited, includes considerations on social and ethical responsibilities linked to the application of their knowledge and judgments.

Full-or-part-time: 16h
Guided activities: 2h
Self study: 14h

(ENG) Conclusió Final del curs

Description:
(ENG) Integra tots els elements que s'han vist i treballat durant el curs, així com la posta en comú dels projectes desenvolupats per grups i articles llegits durant el curs

Specific objectives:
(ENG) 3, 6

Material:
(ENG)

Delivery:
(ENG)

Related competencies :
CG8. Capability to apply the acquired knowledge and to solve problems in new or unfamiliar environments inside broad and multidisciplinary contexts, being able to integrate this knowledge.
CTR6. REASONING: Capacity for critical, logical and mathematical reasoning. Capability to solve problems in their area of study. Capacity for abstraction: the capability to create and use models that reflect real situations. Capability to design and implement simple experiments, and analyze and interpret their results. Capacity for analysis, synthesis and evaluation.
CTR2. SUSTAINABILITY AND SOCIAL COMMITMENT : Capability to know and understand the complexity of the typical economic and social phenomena of the welfare society. Capacity for being able to analyze and assess the social and environmental impact.
CTR4. INFORMATION LITERACY: Capability to manage the acquisition, structuring, analysis and visualization of data and information in the area of informatics engineering, and critically assess the results of this effort.
CB8. Capability to communicate their conclusions, and the knowledge and rationale underpinning these, to both skilled and unskilled public in a clear and unambiguous way.
CB6. Ability to apply the acquired knowledge and capacity for solving problems in new or unknown environments within broader (or multidisciplinary) contexts related to their area of study.
CB7. Ability to integrate knowledges and handle the complexity of making judgments based on information which, being incomplete or limited, includes considerations on social and ethical responsibilities linked to the application of their knowledge and judgments.

Full-or-part-time: 2h
Theory classes: 2h

Introduction

Full-or-part-time: 2h
Theory classes: 2h

Scope, tools, Data, Metadata, Preprocessing

Full-or-part-time: 8h
Theory classes: 6h
Laboratory classes: 2h

DMMCM map, Data Mining methods

Full-or-part-time: 20h
Theory classes: 12h
Laboratory classes: 8h

Spatio-temporality

Full-or-part-time: 4h
Theory classes: 2h
Laboratory classes: 2h

Post-processing

Full-or-part-time: 4h
Theory classes: 2h
Laboratory classes: 2h

GRADING SYSTEM

Two scores corresponding to two activities developed during the course:
20% for Paper activity: It will evaluate the capacities of comprehension (0.5), synthesis (0.5), oral and visual communication (0.5), as well as argumentative capacity (0.5), which will be demonstrated through discussion

80% for a project developed by teams. There will be a single evaluation of the Data Mining project quality, considering the methodologic rigour (0.5), the correctness of the Knowledge Discovery process designed (0.5), the selected preprocessing methods (0,25), the selected data mining methods (0,25), the selected tools (0,5), correct application and results interpretation (1), the integration of several techniques in the project (0,5), the quality of the written report (1), and final public presentation (1). For the final scoring, it will be important the level of planning and coordination of the team, how the incidencies during the course have been solved (1). Additionally, invididual evaluation of the communication skills of every single student (0,5) will be taken into account, as well as its integration level to the working team (1).

RESOURCES

Hyperlink:
- https://www-eio.upc.edu/~karina/datamining/