Data Analysis

Academic Year: (2020 / 2021) **Review date: 10-07-2020**

Department assigned to the subject: Computer Science and Engineering Department

Coordinating teacher: GARCIA HERRERO, JESUS

Type: Electives ECTS Credits : 6.0

Year : 4 Semester :

REQUIREMENTS (SUBJECTS THAT ARE ASSUMED TO BE KNOWN)

Programming
Artificial Intelligence

OBJECTIVES

General competences:
- Analysis (PO a)
- Abstraction (PO a)
- Problem solving (PO c)
- Capacity to apply theoretical concepts (PO c)

Specific competences
- Cognitive

1. Evaluation based on multiple Theoretical machine learning tasks (PO a)
2. Knowledge about several model building techniques working on data (PO a)
3. Knowledge about practical techniques to deal with uncertainty and errors in data to take advantage of them (PO c)

- Procedimental/Instrumental

2. Students should use different Data Mining techniques, compare them through experiments, and analyze the results (PO b)
3. Students should apply the right and appropriate Data Mining technique and parameters to solve a task (objective) (PO c)

- Attitudinal

4. Students should work on the homeworks in teams (PO d)
5. Students are required to use Data Mining tools and provide solutions to real-world problems through computer engineering (PO e)
6. Students must present a written summary for each homework, the final homework should be orally presented, and the final exam is written (PO g)
7. Students should be able to use state of the art Data Mining tools to solve homework tasks (PO k)

DESCRIPTION OF CONTENTS: PROGRAMME

1. Introduction to Data Analysis and Data Mining
2. Machine learning with numeric techniques
2.1. Statistical analysis and causal relations
2.2. Bayesian classifiers. Numeric and symbolic attributes
3. Numerical learning
3.1. Regression
3.2. Clustering with numeric techniques: K-means, Expectation Maximization
4. Evaluation of Machine Learning Models
4.1. Confusion matrices
4.2. Comparison of alternatives, significance contrasts
5. Attribute analysis
5.1. Non-supervised selection
5.2. Attribute transformation
5.3. Supervised selection
6. Methodology of data mining projects
6. Introduction to other advanced techniques (combination, SVM , Fuzzy systems, GAs)

LEARNING ACTIVITIES AND METHODOLOGY

Theoretical lectures: 2 ECTS. To achieve the specific cognitive competences of the course (PO a).
Practical lectures: 2,5 ECTS. To develop the specific instrumental competences and most of the general competences, such as analysis, abstraction, problem solving and capacity to apply theoretical concepts. Besides, to develop the specific attitudinal competences. (PO a, c, d, f, g).
Guided academic activities (present teacher): 1,5 ECTS. The student proposes a project according to the teachers guidance to go deeply into some aspect of the course, followed by public presentation (PO a, c, d, g, k).

ASSESSMENT SYSTEM

Exercises and examinations are both learning and evaluation activities. The evaluation system includes the assessment of guided academic activities and practical cases, with the following weights:
Examination of concepts taught in course: 30% (PO a)
Practical exercises and presentation to proposed problems: 20% (PO b, c, d, e)
Practical case: 50% (PO a, c, d, g). The evaluation of this subjet will follow a continuous process, covered by a practical project developed along the course, where theoretical concepts will be applied to a practical situation. The work teams will be formed by groups of two people.

% end-of-term-examination:	30
% of continuous assessment (assigments, laboratory, practicals…):	70

BASIC BIBLIOGRAPHY

- I. Witten y E. Frank Data Mining: Practical Machine Learning Tools and Techniques (Third Edition) , Morgan Kaufmann, 2011
- Jesús García, Antonio Berlanga, José M. Molina, Miguel A. Patricio Ciencia de datos: Técnicas analíticas y aprendizaje estadístico en un enfoque práctico, Altaria, 2018

ADDITIONAL BIBLIOGRAPHY

- David Hand, Heikki Mannila Principles of data mining, MIT Press, 2002
- Jesus Garcia, José M. Molina Apuntes de la asignatura, http://www.giaa.inf.uc3m.es/docencia/II/ADatos/apuntesAD.pdf, 2004
- Pérez López, César Estadística aplicada a través de Excel, Prentice Hall, 2002