Introduction to Data Mining for Business Intelligence

Academic Year: (2021 / 2022) **Review date: 24-06-2021**

Department assigned to the subject: Statistics Department

Coordinating teacher: MUÑOZ GARCIA, ALBERTO

Type: Electives ECTS Credits : 6.0

Year : Semester :

REQUIREMENTS (SUBJECTS THAT ARE ASSUMED TO BE KNOWN)

This course assumes that the student knows the contents of

a) Statistics I (http://www3.uc3m.es/reina/Fichas/Idioma_2/204.13154.html),
b) Statistics II (http://www3.uc3m.es/reina/Fichas/Idioma_2/204.13160.html),

and the lesson of Properties of Matrices in
c) Mathematics for Economics II (http://www3.uc3m.es/reina/Fichas/Idioma_2/204.13156.html)

in the Business Administration degree.

Some notions in Multivariate Statistics are welcome.

OBJECTIVES

SPECIFIC COMPETENCES:
1. To know and use advanced statistical techniques, with last generation software support.
2. To extract and analyze information from large data sets.
3. Learning the basic Statistical skills for the analysis of multivariate socio-economical data such as those coming from a market research.
4. Being able to describe and analyze real data sets using the techniques mentioned above.
5. Being able to elaborate reports with the results of the analysis of real case studies.

CROSS COMPETENCES
1. Information analysis and synthesis capacity on data mining problems.
2. Solving real problems.
3. Learning and training in the use of Statistical software to solve real case studies.
4. Critical and selective reasoning to solve

DESCRIPTION OF CONTENTS: PROGRAMME

1. Learning the R Statistical Language.
   - 1.1 Basic commands.
   - 1.2 Graphics in R.
   - 1.3 Statistical functions in R and basic programming.
2. Visualization Techniques for complex business data.
   - 2.1 Principal component analysis theory.
   - 2.2 Basic examples with R code.
   - 2.3 Case studies.
3. Multidimensional Scaling.
   - 3.1 Metric scaling theory.
   - 3.2 Examples with R code.
   - 3.3 Perceptual mappings in R.
4. Cluster Analysis.
   - 4.1 Hierarchical methods.
   - 4.2 Centroid methods: k-means.
   - 4.3 Case studies.
5. Classification Trees.
   - 5.1 Information theory.
   - 5.2 Classification trees algorithms.
   - 5.3 Real case: credit scoring.

6. Real Case Studies.
   6.1 Comprehensive real cases involving all the studied techniques.

LEARNING ACTIVITIES AND METHODOLOGY

1. Theoretical lectures (4 ECTS)
2. Computer labs (2 ECTS)
3. Final project.

ASSESSMENT SYSTEM

50%: Final exam oriented to practice.
10%: Continuous evaluation (*).
40%: Handing a final project.

(*) Continuous evaluation consists of handing several case studies (homework) along the course.

% end-of-term-examination:	60
% of continuous assessment (assigments, laboratory, practicals…):	40

BASIC BIBLIOGRAPHY

- Avril Coghlan A little book of R for multivariate analysis, Internet, 2014
- Johannes Ledolter Data Mining and Business Analytics with R, Wiley, 2013
- Paul L. Hewson Multivariate Statistics with R, Internet, 2009

ADDITIONAL BIBLIOGRAPHY

- Y. Zhao R and Data Mining. Examples and Case Studies, Elsevier, 2012