

DAS

*DSIC*

DEPARTAMENT DE SISTEMES  
INFORMÀTICS I COMPUTACIÓ



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA

# DATA SCIENCE

33965-MÁSTER UNIVERSITARIO EN INGENIERÍA Y  
TECNOLOGÍA DE SISTEMAS SOFTWARE

**M. José Ramírez-Quintana**

DSIC, UPV, [mramirez@dsic.upv.es](mailto:mramirez@dsic.upv.es)

# DESCRIPTION

- Credits: 3.0 (2.0: theory, 1.0: lab)
- Tuesdays (15:00 to 17:00)
- Thursdays(17:00 to 19:00)
- Theory and lab will be intertwined.
  - Everything in the same classroom.
- Lecturer
  - M.José Ramírez-Quintana ([mramirez@dsic.upv.es](mailto:mramirez@dsic.upv.es))
    - Office 2D39, 2nd floor DSIC (Bldg. 1F).
    - Attention/tutoring hours: on demand by email.

# LEARNING OUTCOMES

After completion of the course, the student will be able to understand the role of the data scientist in organisations, identify problems and opportunities and deploy solutions using off-the-shelf tools.

- Specific objectives:
  - ▣ Realise the value of data and data-driven products.
  - ▣ Know the process of converting data into knowledge.
  - ▣ Use tools to integrate, prepare and visualise data.
  - ▣ Use a data-analysis language or tool to obtain models.
  - ▣ Evaluate models.
  - ▣ Deploy and exploit knowledge.

# CONTENTS

- Unit 1: Introduction
  - Data science: the data scientist role.
  - The value of data: examples
  - The D2K process (Data to Knowledge).
- Unit 2: Data integration and manipulation
  - Types of data and repositories.
  - Data integration and cleansing.
  - Data ownership, privacy and security.
  - Introduction to R.
  - Visualisation and data understanding.
  - Data manipulation and cleaning using R. Data Visualisation in R.
- Unit 3: Data analysis
  - Types of predictive and descriptive tasks.
  - Supervised learning.
  - Solving classification and regression problems in R.
  - Unsupervised learning. An example in R.
  - Model evaluation.
  - Model evaluation in R. Visualising model performance.

# COURSE PLAN

Tuesday	Thursday
10Sep: U1	12Sep: U1
17Sep: Introduction R	19Sep: Project Guide R-Lab 1
24Sep: U2	26Sep: U2
1Oct: Graphs in R Lab 2	3Oct:U3
8Oct: Wednesday teaching	10Oct: Project Revision
15Oct: U3	17Oct: Lab 3
22Oct:U3	24Oct:U3
29Oct: Lab 4	
7 Nov: Exam	
8 Nov: Project Presentation	
28 Jan: Resit for the exam	

# EVALUATION

- Laboratory:
  - A1: practical exercises in R (20% in total)-Tests Lab 2, 3 & 4
- Written exam (November 7<sup>th</sup>, 15:00-17:00)
  - A2: Questionnaire (theory and R) (30% in total)
- Practical. (Friday November 8<sup>th</sup>)
  - A3: Project Assignment (40%)
    - Option 1: Guided assignment
      - Groups of 2-3 students.
      - There will be instructions to work with a dataset on Kaggle.
      - Oral presentation (questions-answers with the R code)
    - Option 2: Freelance data scientist project
      - Groups of 2-3 students.
      - Develop the **idea of a new product** from the use of data (open data, Internet, repositories, etc.) or that could improve an existing procedure.
      - Oral presentation.
- Co-evaluation. A4 (10% in total)

Final mark (0-10) = A1 + A2 + A3 + A4.

# ASSESSMENT OF LEARNING OUTCOME REQUIREMENTS FOR THE MASTER PROGRAM

- Innovation and creativity.
  - Project assignment

# RECOMMEND READINGS

- Theory:

- Lars Nielsen, Noreen Burlingame “A simple introduction to data science”, New Street Communications 2012 (ultra-short introduction)
- Rachel Schutt “Doing data science”, O’Reilly 2013
- Bing Liu Web Data Mining [Recurso electrónico-En línea] : Exploring Hyperlinks, Contents, and Usage Data
- Jiawei Han Data Mining [Recurso electrónico-En línea] : Concepts and Techniques.
- José Hernández-Orallo, M.José Ramírez-Quintana, Cèsar Ferri, “Introducción a la minería de datos”, Pearson 2004
- Peter Flach “Machine learning: the art and science of algorithms that make sense of data”, Cambridge University Press 2013.
- Foster Provost and Tom Fawcett Data Science for Business: Fundamental principles of data mining and data analytic thinking, O’Reilly Media, 2013
- Jeffrey Stanton “Introduction to Data Science”, 2012.  
[https://ischool.syr.edu/media/documents/2012/3/DataScienceBook1\\_1.pdf](https://ischool.syr.edu/media/documents/2012/3/DataScienceBook1_1.pdf)

- Lab (R):

- CRAN manuals: <http://cran.r-project.org/doc/manuals/R-intro.pdf>, <http://cran.r-project.org/doc/contrib/Torfs+Brauer-Short-R-Intro.pdf>
- Luis Torgo “Data Mining with R”, CRC Press 2010.
- Wikibooks: [http://en.wikibooks.org/wiki/Data\\_Mining\\_Algorithms\\_In\\_R](http://en.wikibooks.org/wiki/Data_Mining_Algorithms_In_R), [http://en.wikibooks.org/wiki/R\\_Programming](http://en.wikibooks.org/wiki/R_Programming).
- Graham Williams: Hands-On Data Science with R, <http://onepager.togaware.com/>
- Hadley Wickham & Garrett Grolemund: “R for Data Science”, 2017