

PROGRAMMING R

AREA: HUMAN SCIENCES

MASTER IN BUSINESS
ANALYTICS AND BIG DATA

Nº OF SESSIONS: 20

Professor: **JESUS PRADA ALONSO**

E-mail: jprada@faculty.ie.edu

MBD Professor Bio

Summary

Professor Prada possesses a double-degree program in informatics and mathematics, a master's double-degree of applied mathematics + master of investigation and innovation in information and communications technology, machine learning specialty and is currently finalizing a PhD. in machine learning. He has 6+ years of experience as a researcher in machine learning applied to real-world problems, including renewable energies and healthcare. He has published five papers on this topic and plan to publish another one in the coming months. He also has knowledge of the corporate world, having 4+ years of experience as a Data Scientist working for different companies ranging from healthcare startups to airline companies. In his freetime, his main hobbies are sports and travelling.

Experience

Teacher, DevAcademy

Freelance Data Scientist, Hospital Clínico San Carlos

Freelance Data Scientist, Iberia Express

Machine Learning Engineer, Madrid-MIT M+Vision Consortium

Data Scientist, Kernel Analytics

Research intern, Machine Learning Group (GAA) at Universidad Autónoma de Madrid (UAM)

Research intern, Instituto de Ingeniería del Conocimiento (IIC)

Education

PhD in Informatics and Telecommunications, Universidad Autónoma de Madrid (UAM) – In progress.

Master's double-degree of investigation and innovation in information and communications technology, ICT (computational intelligence specialty) + Master of Applied Mathematics, Universidad Autónoma de Madrid (UAM).

Double-degree program in informatics and mathematics, Universidad Autónoma de Madrid (UAM).

LEARNING OBJECTIVES

R is one of the most popular programming languages for data science and statistics.

In this course you will learn how to program in R and how to use it to perform proper data analysis.

The goals of this course include learning about:

- How to install R and RStudio.
- Variables and data types in R.
- Control structures.
- Functions.
- Plotting.
- Performing analysis over real datasets.
- Basic Machine Learning models using R.

PROGRAM

SESSION 1 (FACE TO FACE)

Introduction to R:

- Why R?
- Installing R and RStudio.
- Introduction to RStudio: Running your first commands.
- R scripts.
- Variable assignments.

SESSIONS 2 - 3 (FACE TO FACE)

Variable classes and basic data types:

- Variable classes.
- Dates and datetimes.
- Vectors.
- Matrices.

SESSIONS 4 - 5 (FACE TO FACE)

Advanced data types:

- Data frames.
- Data tables.
- Working with data tables: filter, merge, group...
- Lists.

SESSION 6 (LABORATORY)

Quiz 1.

SESSION 7 (FACE TO FACE)

Functions in R:

- Load libraries.
- Create custom functions.
- Source functions.
- Function scope.

SESSIONS 8 - 9 (FACE TO FACE)

Control structures and vectorization:

- If - else
- For loop.
- While loop.
- Repeat, next, break.
- Vectorization: apply, sapply, lapply.

SESSION 10 (FACE TO FACE)

R markdown:

- What is markdown?
- Creating an R markdown.
- Text blocks.
- Code blocks.

SESSIONS 11 - 12 (FACE TO FACE)

Plots in R:

- Line and scatter plots.
- Barplots.
- Histograms.
- Boxplots.
- Advanced plots: ggplot, shiny...

SESSION 13 (LABORATORY)

Quiz 2.

SESSIONS 14 - 15 (LABORATORY)

Working with real-world datasets:

- Reading tabular data.
- Querying SQL.
- Data cleaning.
- Sampling.
- EDA: Exploratory Data Analysis.
- Debugging.

SESSIONS 16 - 17 (FACE TO FACE)

Supervised Learning:

- Classification vs Regression.
- Bias-variance tradeoff. Underfitting and overfitting.
- Accuracy-'explainability' tradeoff.
- Your first model in R.

SESSIONS 18 - 19 (FACE TO FACE)

Unsupervised Learning.

- Unsupervised problems: Recommender systems, dimensionality reduction, and clustering.
- K-means using R.

SESSION 20 (LABORATORY)

Final Exam.

EVALUATION CRITERIA

Criteria	Percentage	Comments
Final Exam	50 %	
Intermediate Tests	30 %	
Workgroups	20 %	