Course guide
240310 - 240IIT31 - Data Science for Industry

Last modified: 12/06/2023

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

Degree: MASTER'S DEGREE IN INDUSTRIAL ENGINEERING (Syllabus 2014). (Optional subject).

Academic year: 2023 **ECTS Credits:** 4.5 **Languages:** Catalan

LECTURER

Coordinating lecturer: Puig De Dou, Ignacio

Others: Corral Lopez, Jesus

REQUIREMENTS

Linear Algebra and Calculus at the level of GETI courses.
Probability and Inference knowledge at the GETI Statistics level.
Programming fundamentals.

DEGREE COMPETENCES TO WHICH THE SUBJECT CONTRIBUTES

Specific:
CEEORG2. Design, develop and apply analytical methods (quantitative methods, statistics, models and decision tools) in order to make strategic, tactical and operational decisions for organizations.
CEETI3. (ENG) Analitzar dades de gran volumn mitjançant models estadístics i algoritmes d'intel.ligència artificial. (Competència específica associada a l'especialitat en Tecnologies de la Informació per a la Industria).
CEETI4. (ENG) Determinar la fiabilitat i seguretat en sistemes complexos mitjançant mètodes de manteniment predictiu i seguretat ciberfísica. (Competència específica associada a l'especialitat en Tecnologies de la Informació per a la Industria).

Transversal:
CT4. EFFECTIVE USE OF INFORMATION RESOURCES: Managing the acquisition, structuring, analysis and display of data and information in the chosen area of specialisation and critically assessing the results obtained.

Basic:
CB 6. (ENG) Tenir i comprendre coneixements que aportin una base o oportunitat de ser originals en el desenvolupament i/o aplicació d'idees, sovint en un context d'investigació
CB 9. (ENG) Que els estudiants sàpiguen comunicar les seves conclusions i coneixements (i darrers raonaments que els sostentin), a públics especialitzats i no especialitzats de manera clara i sense ambigüitats.
CB10. (ENG) Que els estudiants poseeixin les habilitats d'aprenentatge qur els permetin continuar estudiant d'una manera d'una forma que haurà de ser en gran mesura autodirigit o autonònom

TEACHING METHODOLOGY

Theoretical concepts and methods reviewed in formal hour and a half blackboard classes.
Pratical deployment of methods and algorithms using R in hour and a half computer room classes.
Development of a real case to be delivered at the end of the course.

LEARNING OBJECTIVES OF THE SUBJECT

The student will learn the basic characteristics of statistical and machine learning models and will know which model to use depending on the problem at hand.
He will understand the training and validation methodologies of machine learning models and will be able to plan the training process of a model.
He will know the different machine learning models available and will be able to implement them in real cases.
He will know and be able to use R software to implement and validate machine learning models.

STUDY LOAD

Type	Hours	Percentage
Hours large group	40,5	36.00
Self study	72,0	64.00

Total learning time: 112.5 h

CONTENTS

Intro to Machine Learning and R

Description:
Data Science in Industry 4.0.
Basic concepts of Machine Learning
Introduction to R:
- exploratory analysis
- commands and programming
- data wrangling

Full-or-part-time: 18h
Practical classes: 9h
Self study : 9h

Supervised Learning: Regression and Classification problems

Description:
Linear and non-linear Multivariate Regression
Logistics Regression
Goodness-of-fit for Classification Models
Model Fitting and Selection. Resampling.

Full-or-part-time: 27h
Practical classes: 13h 30m
Self study : 13h 30m

Non-Supervised Machine Learning Methods

Description:
Dimensionality Reduction Techniques
Distance based cluster methods: hierarchical and agglomerative

Full-or-part-time: 18h
Practical classes: 9h
Self study : 9h

Advanced Machine Learning Methods: Tree-Based and SVM

Description:
Tree-based methods: bagging, random forest and boosting.
Support Vector Machines

Full-or-part-time: 18h
Practical classes: 9h
Self study : 9h

GRADING SYSTEM

There will be three regular class sessions devoted to do an exam on the material covered so far. The final exam will be comprehensive of the whole course with special focus on the last classes.
0.30 In-class Exams + 0.25 Term Project + 0.45 Final Exam

BIBLIOGRAPHY

Basic:
- James, Gareth. An Introduction to statistical learning : with applications in R [on line]. 2nd ed. New York: Springer, 2021 [Consultation: 11/01/2022]. Available on: https://ebookcentral.proquest.com/lib/upcatalunya-ebooks/detail.action?pq-origsite=primo&docID=6686746. ISBN 9781071614174.
- Bertsekas, Dimitri P; Tsitsiklis, John N. Introduction to probability : Dimitri P. Bertsekas and John N. Tsitsikis. 2nd ed. Belmont, Massachussets: Athena Scientific, cop. 2008. ISBN 9781886529236.
- Manly, Bryan F. J; Navarro Alberto, Jorge A. Multivariate statistical methods : a primer. Fourth edition. Boca Raton: CRC Press, Taylor & Francis Group, [2017]. ISBN 9781498728966.

Complementary:
- Hastie, T.; Tibshirani, R.; Friedman, J. The elements of statistical learning : data mining, inference and prediction [on line]. 2nd ed. New York [etc.]: Springer, cop. 2009 [Consultation: 25/08/2022]. Available on: https://link-springer-com.recursos.biblioteca.upc.edu/book/10.1007/978-0-387-84858-7. ISBN 0387952845.
- Peña, Daniel. Análisis de datos multivariantes [on line]. 1a ed. Madrid: McGraw-Hill/Interamericana de España, S.L, 2013 [Consultation: 05/04/2023]. Available on: http://www.ingebook.com/ib/NPcd/IB_BooksVis?cod_primaria=1000187&codigo_libro=4203. ISBN 9788448191849.