COMILLAS
UNIVERSIDAD PONTIFICIA
ICAI ICADE CIHS

**TEACHING GUIDE**
**2024 - 2025**

## SUBJECT INFO

| Datos de la asignatura | |
|---|---|
| **Nombre completo** | Machine Learning I |
| **Código** | DOI-MBD-514 |
| **Título** | Máster en Big Data. Tecnología y Analítica Avanzada/Master in Big Data Technologies and Advanced Analytics |
| **Impartido en** | Máster Universitario en Big Data [Primer Curso]<br>Máster en Big Data. Tec. y Analítica Avanzada/Master in Big Data Technologies and Advanced Analytics [Primer Curso] |
| **Nivel** | Master |
| **Cuatrimestre** | Semestral |
| **Créditos** | 6,0 ECTS |
| **Carácter** | Obligatoria |
| **Departamento / Área** | Departamento de Organización Industrial |
| **Responsable** | José Portela |
| **Horario de tutorías** | A consultar con el profesor |

| Datos del profesorado | |
|---|---|
| **Profesor** | |
| **Nombre** | Jaime Pizarroso Gonzalo |
| **Departamento / Área** | Departamento de Telemática y Computación |
| **Despacho** | Santa Cruz de Marcenado 26 |
| **Correo electrónico** | jpizarroso@comillas.edu |
| **Teléfono** | 2732 |
| **Profesor** | |
| **Nombre** | José Portela González |
| **Departamento / Área** | Departamento de Métodos Cuantitativos |
| **Despacho** | Santa Cruz de Marcenado 26 |
| **Correo electrónico** | Jose.Portela@iit.comillas.edu |
| **Teléfono** | 2741 |
| **Profesor** | |
| **Nombre** | Alejandro Polo Molina |
| **Departamento / Área** | Instituto de Investigación Tecnológica (IIT) |
| **Correo electrónico** | apolo@comillas.edu |

## SUBJECT DATA

# Context of the subject

## Contribution to the professional profile of the degree

The purpose of this course is to provide students with a fundamental understanding and an extensive practical experience of how to extract knowledge from an apparently unstructured set of data.

By the end of the course, students will:

- Understand the basic principles behind machine learning.
- Have practical experience with the most relevant machine learning algorithms.
- Have well-form criteria to choose the most appropriate techniques for a given application.

# Competences - Objectives

## Competences

**Results of the training and learning process**

**Knowledge or contents**

- CO1 Understand the fundamentals of data analytics and its application in various areas of artificial intelligence, highlighting the integration in complex and multidisciplinary solutions for advanced analysis of massive data addressing the diversity of specific problems in each area.
- CO2 Understand the most common and appropriate data processing techniques, architectures and tools for specific case conditions and requirements.
- CO3 Understand machine learning and artificial intelligence techniques, from the simplest to the most sophisticated, and be able to select the most appropriate and implement it in a manner appropriate to the conditions and requirements of the specific case.

**Competences**

- CP1 Integrate architectures, artificial intelligence techniques, advanced data analysis and visualization and legal compliance to offer the optimal global solution.
- CP3 Implement statistical analysis techniques and programming languages for Machine Learning in the context of Big Data, adjusting the methodology to the specificities of each dataset to optimize the results obtained.

- CP4 Implement data processing techniques and use the most common tools appropriate to the conditions and requirements of specific cases.
- CP5 Apply ethical principles related to the collection, storage, and analysis of data taking into account possible direct or indirect discriminations derived from decision making.
- CP7 Apply advanced knowledge in Big Data and data analytics to develop innovative solutions in projects and research, providing and evaluating optimal solutions for large-scale data processing and analysis.

**Abilities or skills**

- HA1 Communicate orally and in writing with technical rigor, expository clarity and argumentative coherence to all types of technical and non-technical interlocutors.
- HA2 Work in multidisciplinary and/or international teams and adequately organize and lead group dynamics.
- HA3 Develop the interpersonal skills required in today's professional environments (empathy, tolerance, respect, ability to bring together conflicting interests).
- HA4 Manage, organize and plan work and time adequately, meeting objectives and quality standards.
- HA5 Maintain continuous training and learning and adaptation to technological and scientific changes.

## THEMATIC BLOCKS AND CONTENTS

### Contents - Thematic Blocks

**CONTENTS**

1. 1. Types of machine learning and concepts common to the different Machine Learning techniques such as cross-validation, regularization or hyperparameter selection.

2. Advanced data preprocessing techniques for Machine Learning.

3. Classification techniques:

- Bayes classifier
- Decision trees
- K-neighbors (k-NN)
- Support Vector Machines (SVM)
- Neural networks with multilayer perceptrons (MLP)

3. Regression Techniques:

- Linear regression, regularization methods and variable selection.

- Non-Linear Models, k-NN,
- Regression Trees, GAM, SVM y MLP.

4. Time Series Models:

- ARIMA models for stochastic processes
- Dynamic regression models.
- Non-linear prediction models.

5. Dimensionality Reduction Techniques

- Principal Component Analysis (PCA)
- Independent Component Analysis (ICA)

6. Clustering techniques

- Hierarchical clustering
- k-means
- Gaussian Mixture (GMM) and DBSCAN
- Techniques for cluster validation.

## TEACHING METHODOLOGY

### General methodological aspects of the course

#### Face-to-Face Methodology: Activities

| Classroom activities |
| --- |
| • Theoretical classes: The professor will present the fundamental concepts of each unit, together with some practical recommendations, and will review solved examples to support the explanation. Active participation will be encouraged by posing open questions to stimulate discussion and proposing short application exercises to be solved in class, either on paper or using a software package.. |
| • Laboratory sessions: Under the supervision of the instructor, students, divided into small groups, will apply the concepts and techniques covered in the theoretical classes and become familiar with the practical application of the most relevant algorithms using software tools and libraries. |
| • Tutorials for groups or individual students will be arranged upon request. |

#### Non-attendance Methodology: Activities

| Out-of-class activities |
| --- |
| Personal study of the course material and resolution of the proposed exercises. |

| |
|---|
| Preparation of laboratory sessions, analysis of results and report writing. |

## SUMMARY OF STUDENT WORK HOURS

| STUDENT WORK-TIME SUMMARY | |
|---|---|
| **CLASS-HOURS** | |
| **Theory classes** | **Lab sessions** |
| 30 | 30 |
| **OUT OF CLASS** | |
| **Individual study** | **Practices and reports** |
| 60 | 60 |
| | |

## EVALUATION AND GRADING CRITERIA

| Activities | Weight |
|---|---|
| Midterm exam | 15% |
| Final Exam | 35% |
| Lab sessions | 50% |

## Grading

**Grading and rules**

| Grading |
|---|
| Ordinary exam |
| Theory accounts for 50% of grading:<br>• Midterm exam: 15% |

- Final Exam: 35%

Lab will be 50% of grading.

To pass the course, the weighted average grade must be greater than or equal to 5 out of 10 points, and the final exam grade must be greater than or equal to 4.5 out of 10 points. Otherwise, the final grade will be the lower of the two grades.

Retake exam

Laboratory grades will be retained. In addition, all students will be required to take a final exam. The resulting grade will be calculated as follows:

- Final exam: 50%
- Lab Practice: 50%

As in the ordinary evaluation period, in order to pass the course, the weighted average grade must be greater than or equal to 10 points, and the final exam grade must be greater than or equal to 4.5 out of 10 points. Otherwise, the final grade will be lower of the two marks.

Rules

Class attendance is mandatory in accordance with Article 93 of the General Regulations of Comillas Pontifical University and Article 6 of the Academic Regulations of the ICAI School of Engineering. Failure to comply with this requirement may have the following consequences:

- Students who do not attend more than 15% of the classes may lose the right to take the final exam during the ordinary evaluation period.
- As for laboratories, absence to more than 15% of the sessions may result in the loss of the right to take the final exam of the ordinary evaluation period and the make-up exam. Missed sessions must be made up in order to obtain the corresponding grade.
- Students who commit an irregularity in any evaluated activity will receive a grade of zero in that activity and a disciplinary procedure will be initiated (cf. Article 168 of the General Regulations of the Pontific Comillas University).

## SCHEDULE

| Activities | | | | | | Date | |
|---|---|---|---|---|---|---|---|
| | IN-CLASS ACTIVITIES | | | | | | |
| Session | h/s | LECTURE & PROBLEM SOLVING | Contents | LAB | ASSESMENT | | |
| 1 | 2 | Introduction (1h) | Introduction to Machine learning | Lab 1 - Intro python | | | |
| 2 | 2 | Classification I (1h) | The classification problem - EDA | Lab 2.1 - EDA | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 3 | 2 | Classification II (1h) | Logistic regression. Generalization | Lab 2.2 - LR | | | |
| 4 | 2 | Classification III (1h) | KNN. Validación | Lab 2.2 - KNN | | | |
| 5 | 2 | Classification IV (1h) | Decision trees. | Lab 2.3 | | | |
| 6 | 2 | Classification V (1h) | SVM | Lab 2.4 | | | |
| 7 | 2 | Classification VI (1h) | MLP | Lab 2.5 | | | |
| 8 | 2 | Classification VII (1h) | | Hackathon | | | |
| 9 | 2 | Regression I (1h) | The regression problem. Linear Regression. | Lab 3.1 - LR | Assignment 1 | | |
| 10 | 2 | Regression II (1h) | Model selection and Regularization | Lab 3.2 | | | |
| 11 | 2 | Regression III (1h) | Non-Linear regression (Polynomial Regression, Splines, GAMs) | Lab 3.3 | | | |
| 12 | 2 | Regression III (1h) | Non-Linear regression (SVM, Reg Tree, MLP) | Lab 3.4 | | | |
| 13 | 2 | Regression IV (1h) | Explainable Machine Learning | Lab 3.5 | | | |
| 14 | 2 | Regression V | | Hackathon | | | |
| 15 | 2 | Summary | | | | | |
| 16 | 2 | Mid-term exam I | | | Mid-term exam I | | |
| 17 | 2 | Forecasting I (1h) | Intro + Decomposition methods | Lab 4.1 | | | |
| 18 | 2 | Forecasting II (1h) | ARIMA + SARIMA | Lab 4.2 | | | |
| 19 | 2 | Forecasting III (1h) | | Hackaton streamlit - 15 series | | | |
| 20 | 2 | Forecasting III (1h) | SARIMAX - Dynamic Regression | Lab 4.4 | | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 21 | 2 | Forecasting V (1h) | Advanced Forecasting methods + Prophet | Lab 4.5 | | | |
| 22 | 2 | Forecasting VI (1h) | Recurrent Neural Networks | Lab 4.6 | | | |
| 23 | 2 | | | | Assignment 2 | | |
| 24 | 2 | Density estimation (1h) | Parametric & Non-parametric methods - Bayes Clasifiers | Lab 5.1 | | | |
| 25 | 2 | Dimensionality reduction (1h) | PCA + ICA + tSNE | Lab 5.2 | | | |
| 26 | 2 | Clustering I (1h) | Hierarquical & partitional clustering | Lab 5.3 | | | |
| 27 | | Clustering II | Vector Quantization. Neural Gas. Mixture Models. | Lab 5.4 | | | |
| 28 | 2 | Course summary | | | | | |

## BIBLIOGRAPHY

### Basic

#### Basic bibliography


- Slides prepared by the lecturer (available in Moodlerooms).
- James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). *An introduction to statistical learning: With applications python*. Springer Nature. https://hastie.su.domains/ISLP/ISLP_website.pdf.download.html
- Raschka, S., & Mirjalili, V. (2019). *Python machine learning: Machine learning and deep learning with Python, scikit-learn, a TensorFlow 2*. Packt publishing ltd.


### Complementary

#### Complementary bibliography


- James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). Statistical learning. In *An introduction to statistical learnin With applications in Python* (pp. 15-67). Cham: Springer International Publishing.
- M. Kuhn and K. Johnson, *Applied Predictive Modeling*, Springer, 2013.
- E. Alpaydin, *Introduction to Machine Learning*, 3rd Ed., MIT Press, 2014.

- S. Marsland, *Machine Learning: An Algorithmic Perspective*, 2nd Ed., Chapman & Hall/CRC Machine Learning & Patte Recognition, 2015.
- T. Mitchell, *Machine Learning*, McGraw-Hill, 1997.
- R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2nd Ed., Wiley-Interscience, 2000.
- C. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2007.
- S. Haykin, *Neural Networks. A Comprehensive Foundation*, 2nd Ed., Pearson, 1999.
- W. Wei, *Time Series Analysis. Univariate and Multivariate Methods*, 2nd Ed., Addison-Wesley, 2006.