



COMILLAS

UNIVERSIDAD PONTIFICIA

ICAI

ICADE

CIHS

GUÍA DOCENTE

2024 - 2025

FICHA TÉCNICA DE LA ASIGNATURA

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] 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

DATOS ESPECÍFICOS DE LA ASIGNATURA



Contextualización de la asignatura

Aportación al perfil profesional de la titulación

El propósito de este curso es proporcionar a los estudiantes una comprensión fundamental y una amplia experiencia práctica sobre cómo extraer conocimiento de un conjunto de datos aparentemente no estructurado.

Al final del curso, los estudiantes:

- Comprenderán los principios básicos detrás del aprendizaje automático.
- Tendrán experiencia práctica con los algoritmos de aprendizaje automático más relevantes.
- Tendrán criterios bien formados para elegir las técnicas más apropiadas para una aplicación determinada.

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.

Competencias - Objetivos

Competencias

Resultados del proceso de formación y aprendizaje

Conocimientos o contenidos

- CO1 Entender los fundamentos de la analítica de datos y su aplicación en diversas áreas de la inteligencia artificial, destacando la integración en soluciones complejas y multidisciplinares para el análisis avanzado de datos masivos atendiendo a la diversidad de problemas específicos de cada área.
- CO2 Comprender las técnicas de procesamiento de datos, las arquitecturas y herramientas más habituales y ser capaz de implementarlas de una manera adecuada a las condiciones y requisitos del caso específico. Comprender las técnicas de procesamiento de datos, las arquitecturas y herramientas más habituales y apropiadas para condiciones y requisitos de casos específicos
- CO3 Comprender las técnicas de machine Learning e Inteligencia artificial, desde las más sencillas a las más sofisticadas y ser capaz de seleccionar la más adecuada e implementarla de una manera adecuada a las condiciones y requisitos del caso específico.

Competencias

- CP1 Integrar las arquitecturas, técnicas de inteligencia artificial, análisis avanzado de datos y de visualización y de cumplimiento legal para ofrecer la solución global óptima.



- CP3 Implementar técnicas de análisis estadístico y lenguajes de programación para Machine Learning en el contexto de Big Data, ajustando la metodología a las especificidades de cada conjunto de datos para optimizar los resultados obtenidos
- CP4 Implementar las técnicas y herramientas más adecuadas para la visualización de datos complejos. Implementar las técnicas de procesado de datos y usar las herramientas más habituales y apropiadas a las condiciones y requisitos de casos específicos
- CP5 Aplicar los principios éticos relativos a la recogida, almacenamiento, y análisis de datos teniendo en cuenta las posibles discriminaciones directas o indirectas derivadas de la toma de decisiones.
- CP7 Aplicar conocimientos avanzados en Big Data y analítica de datos para desarrollar soluciones innovadoras en proyectos y en investigación, aportando y evaluando soluciones óptimas para el procesamiento y análisis de datos a gran escala.

Habilidades o destrezas

- HA1 Comunicar de manera oral y escrita con rigor técnico, claridad expositiva y coherencia argumentativa a todo tipo de interlocutores, técnicos y no técnicos.
- HA2 Trabajar en equipos de carácter pluridisciplinar y/o internacional y organizar y liderar adecuadamente las dinámicas de grupo.
- HA3 Desarrollar las habilidades interpersonales que requieren los entornos profesionales actuales (empatía, tolerancia, respeto, capacidad para aunar intereses contrapuestos).
- HA4 Gestionar, organizar y planificar adecuadamente el trabajo y el tiempo, cumpliendo objetivos y estándares de calidad.
- HA5 Mantener una formación y aprendizaje continuo y adaptación a los cambios tecnológicos y científicos.

BLOQUES TEMÁTICOS Y CONTENIDOS

Contenidos – Bloques Temáticos

CONTENIDOS

1. Tipos de aprendizaje automático y conceptos comunes a las distintas técnicas de Machine Learning como validation cruzada, regularización o selección de hiperparámetros.
2. Técnicas de preprocesado avanzado de datos para Machine Learning.
3. Técnicas de clasificación:
 - Clasificador de Bayes
 - Árboles de decisión
 - K-vecinos (k-NN)
 - Máquinas de soporte vectorial (SVM)
 - Redes neuronales con perceptrones multicapa (MLP)



3. Técnicas de regresión:

- Regresión Lineal, métodos de regularización y selección de variables
- Modelos no lineales como k-NN,
- Árboles de regresión, GAM, SVM y MLP.

4. Técnicas de análisis predicción de series temporales:

- Modelos ARIMA para procesos estocásticos
- Modelos de regresión dinámica
- Modelos de predicción no lineales.

5. Técnicas de reducción de la dimensión:

- Análisis de Componentes Principales (PCA)
- Análisis de Componentes Independientes (ICA)

6. Técnicas de clustering incluye:

- Clustering jerárquico
- k-medias
- Mixtura de Gaussianas (GMM) y DBSCAN
- Técnicas para la validación de los clusters.

METODOLOGÍA DOCENTE

Aspectos metodológicos generales de la asignatura

Metodología Presencial: Actividades

Actividades en clase

- **Clases teóricas:** El profesor presentará los conceptos fundamentales de cada unidad, junto con algunas recomendaciones prácticas, y revisará ejemplos resueltos para apoyar la explicación. Se fomentará la participación activa planteando preguntas abiertas para estimular la discusión y proponiendo ejercicios cortos de aplicación que se resolverán en clase, ya sea en papel o utilizando un paquete de software.
- **Sesiones de laboratorio:** Bajo la supervisión del instructor, los estudiantes, divididos en pequeños grupos, aplicarán los conceptos y técnicas abordados en las clases teóricas y se familiarizarán con la aplicación práctica de los algoritmos más relevantes utilizando herramientas y bibliotecas de software.
- Se organizarán tutorías para grupos o estudiantes individuales bajo solicitud.

Metodología No presencial: Actividades

Actividades fuera de clase



Estudio personal del material del curso y resolución de los ejercicios propuestos.

Preparación de las sesiones de laboratorio, análisis de los resultados y redacción de informes.

RESUMEN HORAS DE TRABAJO DEL ALUMNO

STUDENT WORK-TIME SUMMARY	
HORAS DE CLASE	
Clases magistrales expositivas y participativas	Sesiones Laboratorio
30	30
FUERA DE CLASE	
Estudio individual	Preparación de prácticas y reports
60	60

EVALUACIÓN Y CRITERIOS DE CALIFICACIÓN

Actividades de evaluación	Peso en la nota
Examen intermedio	15%
Examen final	35%
Sesiones de laboratorio y entregas	50%

Calificaciones

Calificaciones y Reglas del Curso

Calificaciones

Convocatoria ordinaria



La teoría representará el 50% de la calificación total, distribuido de la siguiente manera:

- Examen parcial: 15%
- Examen final: 35%

El laboratorio representará el 50% restante de la calificación.

Para aprobar el curso, la calificación media ponderada debe ser mayor o igual a 5 sobre 10 puntos, y la nota del examen final debe ser mayor o igual a 4.5 sobre 10 puntos. De lo contrario, la calificación final será la menor de las dos notas.

Convocatoria extraordinaria

Las notas de laboratorio se conservarán. Además, todos los estudiantes deberán realizar un examen final. La calificación resultante se calculará de la siguiente manera:

- Examen final: 50%
- Prácticas de laboratorio: 50%

Al igual que en el periodo de evaluación ordinaria, para aprobar el curso, la calificación media ponderada debe ser mayor o igual a 5 sobre 10 puntos, y la nota del examen final debe ser mayor o igual a 4.5 sobre 10 puntos. De lo contrario, la calificación final será la menor de las dos notas.

Reglas del curso

La asistencia a clases es obligatoria de acuerdo con el Artículo 93 del Reglamento General de la Universidad Pontificia Comillas y Artículo 6 de las Normas Académicas de la Escuela Técnica Superior de Ingeniería ICAI. No cumplir con este requisito puede tener las siguientes consecuencias:

- Los estudiantes que no asistan a más del 15% de las clases pueden perder el derecho a presentar el examen final durante el periodo de evaluación ordinaria.
- En cuanto a los laboratorios, la ausencia a más del 15% de las sesiones puede resultar en la pérdida del derecho a presentar el examen final del periodo de evaluación ordinaria y el examen de recuperación. Las sesiones perdidas deben recuperarse para obtener la calificación correspondiente.
- Los estudiantes que cometan una irregularidad en cualquier actividad evaluada recibirán una calificación de cero en dicha actividad y se iniciará un procedimiento disciplinario (cf. Artículo 168 del Reglamento General de la Universidad Pontificia Comillas).

PLAN DE TRABAJO Y CRONOGRAMA

Actividades						Fecha de realización	Fecha de entrega
IN-CLASS ACTIVITIES							
Session	h/s	LECTURE & PROBLEM SOLVING	Contents	LAB	ASSESSMENT		
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 Advanced	Lab 4.4	
21	2	Forecasting V (1h)	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)	Hierarchical & partitional clustering	Lab 5.3	
27	2	Clustering II	Vector Quantization. Neural Gas. Mixture Models.	Lab 5.4	
28	2	Course summary			

BIBLIOGRAFÍA Y RECURSOS

Bibliografía Básica

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, and TensorFlow 2*. Packt publishing ltd.

Bibliografía Complementaria

Complementary bibliography



COMILLAS

UNIVERSIDAD PONTIFICIA

ICAI

ICADE

CIHS

GUÍA DOCENTE

2024 - 2025

- James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). Statistical learning. In *An introduction to statistical learning 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 & Pattern 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.

En cumplimiento de la normativa vigente en materia de **protección de datos de carácter personal**, le informamos y recordamos que puede consultar los aspectos relativos a privacidad y protección de datos [que ha aceptado en su matrícula](#) entrando en esta web y pulsando "descargar"

[https://servicios.upcomillas.es/sedelectronica/inicio.aspx?csv=02E4557CAA66F4A81663AD10CED66792](https://servicios.upcomillas.es/sedeelectronica/inicio.aspx?csv=02E4557CAA66F4A81663AD10CED66792)