



# PROMiDAT

IBEROAMERICANO

Programa Iberoamericano de  
Formación en Minería de Datos

**ML5000**

**Proyecto Final de Graduación**



(506) 4030.1205 - (506) 4030.1114



[info@promidat.com](mailto:info@promidat.com)



[facebook.com/promidat](https://facebook.com/promidat)



[Twitter.com/promidat](https://Twitter.com/promidat)



[www.promidat.com](http://www.promidat.com)

---

# Índice

Duración	2
Descripción	2
Objetivos	2
Metodología	2
Contenido	3
Evaluación	3
Bibliografía	3

---

## Duración

Cuatro semanas.

## Descripción



En este curso el estudiante desarrollará de principio a fin, siguiendo las buenas prácticas de gestión de proyectos en Machine Learning, su propio proyecto con datos de la empresa en donde trabaja o datos en los cuales está particularmente interesado(a).

## Objetivos

En este curso el estudiante será capaz de:

1. Utilizar adecuadamente la metodología CRISP-DM para el desarrollo de proyectos en Machine Learning.
2. Distinguir entre un proyecto de segmentación, clasificación, regresión y series de tiempo.
3. Elegir, para un problema concreto, qué técnicas de Machine Learning son las más apropiadas.
4. Determinar la herramienta de software más adecuada para el enfrentar el problema planteado.
5. Generar los modelos y patrones elegidos utilizando una herramienta o paquete de Machine Learning.
6. Evaluar la calidad de un modelo, utilizando técnicas sencillas de evaluación, validación cruzada.
7. Integrar sistemas de Bases de Datos desde Python.
8. Presentar resultados en aplicaciones Web.

## Metodología

Basado en la teoría y en la aplicación directa de los conceptos aprendidos. Para esto se dispondrán de las siguientes herramientas:

- 
1. Una vídeo conferencia semanal, las cuales quedarán grabadas en Zoom, para que los alumnos la puedan acceder en cualquier momento.
  2. Trabajos prácticos semanales.
  3. Foros para plantear dudas al tutor y compañeros.
  4. Aula virtual en Moodle.

Luego de este curso el estudiante será capaz de:

Desarrollar proyectos de Machine Learning de principio a fin, es decir, iniciando con la obtención y limpieza de datos hasta la generación de conocimiento derivados de los datos de la organización.

## Contenido

1. Fundamentos de la metodología para la implementación de proyectos de Machine Learning CRISP-DM.
2. Integración de sistemas de bases de datos en Python, para usarlos como fuente directa de datos.
3. Integración de los resultados de un proceso de Machine Learning en una aplicación web.
4. Estudio comparativo de diferentes tipos problemas en Machine Learning.

## Evaluación

El curso se evalúa con 4 tareas, una por semana, cada tarea tiene un valor de 25 puntos. La nota mínima de aprobación es de 70.

## Bibliografía

- [1] José Alberto Gallardo Arancibia. “Metodología para la definición de requisitos en proyectos de data mining (er-dm)”. <http://purl.org/dc/dcmitype/Text>. Universidad Politécnica de Madrid, 2009. URL: <https://dialnet.unirioja.es/servlet/tesis?codigo=20961> (visited on 10/19/2022).
- [2] Ana Azevedo and Manuel Filipe Santos. “KDD, SEMMA and CRISP-DM: a parallel overview”. In: *IADS - DM* (2008). Accepted: 2012-06-14T09:51:14Z. URL: <https://recipp.ipp.pt/handle/10400.22/136> (visited on 10/19/2022).
- [3] Davy Cielen and Arno Meysman. *Introducing Data Science: Big data, machine learning, and more, using Python tools*. Google-Books-ID: bTozEAAAQBAJ. Simon and Schuster, May 2, 2016. 475 pp. ISBN: 978-1-63835-249-5.
- [4] *CRISP-DM 1.0: Step-by-step Data Mining Guide*. Google-Books-ID: po7FtgAACAAJ. SPSS, 2000. book.

- 
- [5] Elias Dabbas. *Interactive Dashboards and Data Apps with Plotly and Dash: Harness the power of a fully fledged frontend web framework in Python – no JavaScript required*. Google-Books-ID: b\_gqEAAAQBAJ. Packt Publishing Ltd, May 21, 2021. 364 pp. ISBN: 978-1-80056-035-2.
- [6] Albert Lukaszewski and Antony Reynolds. *MySQL for Python*. Google-Books-ID: SWa1AAAAQBAJ. Packt Publishing Ltd, Sept. 21, 2010. 771 pp. ISBN: 978-1-84951-019-6.