



**Syllabus**  
**Introducción al Data Mining y**  
**Algoritmos de Machine Learning**

PROGRAMA EJECUTIVO EN  
INTELIGENCIA ARTIFICIAL &  
DEEP LEARNING  
REPÚBLICA DOMINICANA

David Kremer  
dkremerg@gmail.com

Curso académico 2018 - 2019

## **1. Presentación de la materia**

Aunque el Data Mining y Machine Learning no son disciplinas nuevas, en los últimos años están adquiriendo especial relevancia y se están aplicando a una gran variedad de problemas de negocio.

En este módulo se introduce el concepto de Data Mining y el desarrollo de modelos con algoritmos de Machine Learning. Se explorará el concepto de modelo y sus distintos tipos, así como las estrategias para su evaluación y validación. Finalmente, se describirán algunos de los algoritmos de Machine Learning más comunes.

Durante la semana, el alumno trabajará en la aplicación de estas técnicas a un problema de negocio.

## **2. Objetivos de aprendizaje**

1. Comprender el proceso general de Data Mining.
2. Entender qué es un modelo, como se entrena, evalúa y valida.
3. Conocer algunos de los algoritmos de Machine Learning más utilizados y algunos ejemplos de su aplicación.
4. Aplicar estos conceptos a un caso de uso.

## **3. Programa de la materia: estructura y contenido**

1. Introducción: Data Mining, Machine Learning, modelos y sus clases.
2. Proceso de Data Mining, metodología CRISP-DM: entendimiento de negocio, entendimiento de los datos, limpieza y preparación de datos, modelización, validación, despliegue y mantenimiento.
3. Modelado con algoritmos de Machine Learning: algunos de los algoritmos más comunes y ejemplos de su aplicación.

## 4. Metodología y Actividades

El programa se desarrollará en formato on-line, bajo la plataforma Blackboard proporcionada por la EOI, de acuerdo con el siguiente detalle de actividades.

### Descripción de las actividades:

Actividades	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo	Lunes
Videoconferencia Inicial (hora: 20:00 duración: 1h)	<ul style="list-style-type: none"> <li>• Presentación del profesor</li> <li>• Clase: introducción al Data Mining</li> <li>• Descripción y guía de los ejercicios a realizar</li> </ul>						
Tablero de Discusión: Foro de Dudas	Espacio de dudas transversal a la semana lectiva						
Lecturas y estudio de documentación	X	X	X	X	X	X	Día de descanso
Tablero de Discusión: Foro de Debate		Presentación de planteamientos de debate	Debate y discusión	Debate y discusión	Cierre del foro de debate		
Ejercicio individual		Presentación de ejercicios	Resolución de ejercicios	Resolución de ejercicios	Resolución de ejercicios	Entrega de ejercicios para evaluación	
Videoconferencia de cierre (hora: 20:00 duración: 1h)						<ul style="list-style-type: none"> <li>• Clase: algoritmos de Machine Learning</li> <li>• Cierre del tema de la semana</li> </ul>	

## 5. Evaluación

- Participación en el foro 20 %
- Ejercicio individual 80 %

## 6. Bibliografía y materiales de consulta

- “La metodología CRISP-DM”, disponible en: [https://eva.fing.edu.uy/pluginfile.php/202013/mod\\_resource/content/1/Metodologia%20CRISP-DM%20para%20mineria%20de%20datos.pdf](https://eva.fing.edu.uy/pluginfile.php/202013/mod_resource/content/1/Metodologia%20CRISP-DM%20para%20mineria%20de%20datos.pdf)
- “Manual CRISP-DM de IBM SPSS Modeler”, disponible en: <ftp://ftp.software.ibm.com/software/analytics/spss/documentation/modeler/15.0/es/CRISP-DM.pdf>
- “Sobre el uso responsable de datos y algoritmos”, disponible en: <https://www.bbva.com/es/uso-responsable-datos-algoritmos/>
- Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani (2014). An Introduction to Statistical Learning: With Applications in R. Springer Publishing Company, Incorporated. Disponible en: <https://www-bcf.usc.edu/~gareth/ISL/ISLR%20First%20Printing.pdf>

- Jiawei Han, Micheline Kamber, Jian Pei (2012). Data Mining Concepts and Technics (Third Edition). Elsevier. Disponible en: <http://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf>

## 7. CV del Profesor

### David Kremer

David es físico y matemático de formación. Trabajó dos años con IBM Research en proyectos de investigación en machine learning y sistemas cognitivos. Actualmente es científico de datos en IBM trabajando como Data Lead en diversos proyectos de analítica. Trabaja en diversas áreas como el mantenimiento predictivo, la optimización, el marketing digital o el análisis de datos de sensores, y en diversas industrias como la industria eléctrica, la petrolífera, el retail o la industria sanitaria.