

# Programas Asíncronos

## SÍLABO

### I. DATOS GENERALES

<b>Curso</b>	:	Técnicas de imputación de datos en Machine Learning
<b>Área</b>	:	B.I.
<b>Docente</b>	:	John Caballero
<b>Duración</b>	:	3 horas académicas

### II. SUMILLA

El curso se desarrollará mediante la metodología asíncrona, que incluye la revisión de los videos grabados, las lecturas y la resolución de la evaluación planteada. Le permitirá al participante comprender la imputación de datos tanto desde el punto de vista conceptual y su finalidad en un proyecto de machine learning, como desde las diferentes técnicas aplicables a un dataset. El curso comprende los siguientes temas generales a) Introducción a la imputación de datos, y b) Principales técnicas de imputación de datos con R y Python.

### III. REQUISITOS

- Conocimiento de programación básica en Lenguaje R y/o Python.

### IV. DIRIGIDO A

- Profesionales y estudiantes que trabajen con datasets y modelos predictivos que deseen iniciarse en las técnicas de imputación de datos.

### V. OBJETIVOS – LOGROS

El curso le permitirá al participante:

- Comprender los fundamentos de la imputación de datos como parte del proceso general de un proyecto de machine learning.
- Comprender cómo aplicar la imputación de datos empleando los lenguajes R y Python de acuerdo a diversas casuísticas típicas.

## VI. METODOLOGÍA

Los contenidos se desarrollarán mediante sesiones grabadas, lecturas como recursos didácticos y una evaluación final.

## VII. CONTENIDOS

Total horas	3 horas
<b>Sesión 1: Introducción a la imputación de datos</b>	
1.1. El proceso de Machine Learning: CRISP-DM	15 min.
1.2. La etapa de preparación de datos. Finalidad, técnicas generales.	15 min.
1.3. Imputación de datos: Definición, justificación en un proyecto de machine learning.	20 min.
1.4. Imputación de datos: Casos de uso vs. Técnicas.	20 min.
1.5. Análisis de un dataset: Identificación de variables candidatas a imputación.	10 min.
<b>Sesión 2: Principales técnicas de imputación de datos con R y Python</b>	
2.1. Imputación de variables cualitativas: Uso de moda. Demostración en R y Python.	15 min.
2.2. Imputación de variables cuantitativas: Uso de media, mediana. Demostración en R y Python.	15 min.
2.3. Imputación multivariada: Uso de modelos lineales, no-lineales, combinaciones lineales. Demostración en R y Python.	30 min.
2.4. Imputación de datos basado en Machine Learning. Demostración en R y Python.	20 min.
2.5. Demostración de la aplicación de una técnica imputación de datos para mejorar el performance de un modelo predictivo.	20 min.

## VIII. SISTEMA DE EVALUACIÓN

La nota final del curso se obtendrá a partir de los siguientes rubros de evaluación y pesos asociados:

Rubro	Peso
Evaluación	100%

## IX. LECTURAS COMPLEMENTARIAS

**X. | DOCENTE****John Caballero**

## Resumen:

- Gerente adj Data Scientist y consultor, con 8 años de experiencia en la aplicación de metodologías de analytics y machine learning. Estudios de pregrado en Ingeniería Estadística en la Universidad Nacional de Ingeniería y cursando un Master en Business Analytics, Big Data and Artificial Intelligence Management en EADA Business School.