



---

## DIPLOMATURA UNIVERSITARIA EN CIENCIA DE DATOS

### Modalidad a Distancia

### Plan de Estudios

#### Contenidos Mínimos

#### **MODULO 1: Introducción a la Ciencia de Datos**

Ciencia de Datos: definición, origen y evolución. Para qué se utiliza la ciencia de datos. Diferencia entre Ciencia de Datos y otras disciplinas. La Ciencia de Datos y su importancia en las empresas y en nuestra vida. Cuáles son las diferentes tecnologías de la Ciencia de Datos. Aplicación de la Ciencia de Datos: presentación de casos reales. Etapas en la formulación de un proyecto para la resolución de un problema realista de ciencia de datos. Definición de .datos. Recolección y fuentes de datos. Formato de almacenamiento de datos (CSV, TSV, ASCII). Formato de los datos (numérico, string, fechas y otros). Formateo y limpieza de datos. Selección, descubrimiento y diseño de features relevantes para la resolución de un problema dado.

Principales lenguajes de programación para Ciencia de Datos: Python, R, SQL, MySQL, entre otros. Breve descripción de cada uno. Ventajas y desventajas.

#### **MÓDULO 2 - Lenguaje de Programación Python**

Presentación del lenguaje de programación Python. Diferentes entornos de programación (consola, IDE, notebooks). Sintaxis del lenguaje. Tipos de datos básicos. Funciones y documentación. Estructuras de datos: diccionarios, listas, tuplas, vectores, matrices y árboles. Módulos y paquetes. Introducción a la programación orientada a objetos. Estructuras de lectura y escritura de datos.

#### **MODULO 3: Bases de Datos: (MySQL)**

Funciones de los sistemas de Base de Datos. Modelos de datos. Lenguajes de consulta. Diseño de Base de Datos. Estructuras físicas y lógicas de datos. Optimización de consultas. Transacciones. Concurrencia y recuperación. Implementadores.

## **MODULO 4 Análisis de Datos**

Introducción práctica al análisis de datos. Técnicas descriptivas de los datos. Estimadores. Concepto de correlación y covarianza. Coeficiente de correlación. Coeficiente de Pearson. Coeficiente de Spearman. Outliers Modelado de los datos. Modelos estadísticos básicos. Parámetros de los Modelos. Regresión lineal simple y múltiple. Interpretación. Evaluación del modelo. Introducción básica a test estadísticos. Extensión de los Modelos. Modelos aditivos. El análisis diagnóstico, el análisis descriptivo, el análisis predictivo y el análisis prescriptivo.

## **MODULO 5: Visualización de Datos**

Visualización. Conceptos. Definiciones. Proceso de exploración de datos. Operaciones en una exploración. Tipos de visualizaciones. Sistemas de Visualización de Información (SVI): requerimientos, componentes, categorías. Objetivos y requerimientos. Enfoques. Múltiples visiones. Dimensionalidad: técnicas 2D y 3D. Ambientes de propósito específico. Vinculación con los datos fuente. Implementación de técnicas de visualización especializadas.

## **MODULO 6 Aprendizaje Automático**

Funcionamiento de los algoritmos de aprendizaje automático. Validez, aplicabilidad, y limitaciones. Conjunto de datos de entrenamiento, validación y testeo. Funciones de error. Sobreajuste. Regularización. Regresión ridge y lasso. Aplicación de algoritmos de aprendizaje automático: regresión y clasificación lineal; discriminante lineal de Fischer; perceptrón; support vector machine; árboles de decisión; naive Bayes.

## **MODULO 7 Introducción a la Minería de Datos**

Introducción a la minería de datos. Origen y motivación. El proceso de descubrimiento de conocimiento en bases de datos. Tipos de datos. Relación con otras áreas. Utilidad de la minería de datos. Aplicaciones. Repercusión, desafíos y tendencias en minería de datos. Técnicas - Árboles de decisión. Reglas de clasificación y asociación. Técnicas de Agrupamiento. Redes Neuronales Evaluación de Modelos}



## **MODULO 8 Trabajo Final Integrador**

Durante este módulo los estudiantes planificarán y ejecutarán su trabajo final integrador. El objetivo es contribuir a la formación del diplomado, donde deba aplicar los conocimientos básicos de la ciencia de datos y de las técnicas métodos lógicas estudiadas.

El trabajo final consistirá en el desarrollo de un proyecto individual creativo e innovador en el que pone en práctica de forma integral los conocimientos y experiencias adquiridos en las diferentes asignaturas, pro poniendo soluciones a problemáticas relacionadas a la sistematización y visualización de grandes volúmenes de datos relacionados con un área de aplicación (Economía, Informática, Geografía, Biología, etc.).