



ITESO, Universidad  
Jesuita de Guadalajara

# Ciencia de Datos e Inteligencia de Negocios

.....

Dr. Gaddiel Desirena López

Primavera 2021

# Descripción de la asignatura

La ciencia de datos es un campo interdisciplinario que involucra a los procesos y sistemas para extraer conocimiento o un mejor entendimiento de grandes volúmenes de datos en sus diferentes formas y formatos.

Por otra parte, la inteligencia empresarial, inteligencia de negocios o BI (del inglés business intelligence), es el conjunto de estrategias, aplicaciones, datos, productos, tecnologías y arquitectura técnica, los cuales están enfocados a la administración y creación de conocimiento a través del análisis de los datos existentes en una organización o empresa.

# Propósito general



Analizar bases de datos mediante el uso de algoritmos de aprendizaje máquina para la toma de decisiones en fenómenos financieros.

1. Minería de Datos: Datos y su clasificación
  - 1.1 Datos y su Clasificación
  - 1.2 Exploratory Data Analysis (EDA)
  - 1.3 Data Quality Report (DQR)
  - 1.4 Data Cleaning
  - 1.5 Medidas de Similitud
2. Preprocesamiento de Datos y Aprendizaje no Supervizado
  - 2.1 Integración de datos
  - 2.2 Reducción de datos PCA
  - 2.3 Búsqueda de Patrones (Clustering)
3. Aprendizaje Supervizado
  - 3.1 Análisis de discriminantes lineales
  - 3.2 Regresión logística
  - 3.3 Máquina de vector soporte (SVM)
  - 3.4 Introducción a Redes Neuronales

Actividad	Ponderación
Exámenes	30 %
Tareas y Actividades	30 %
Proyecto	40 %

- ▶ Los seres humanos estamos creando y almacenando información constantemente en cantidades astronómicas. Si todos los bits y bytes de datos del último año fueran guardados en CD's, se generaría una torre de la tierra a la luna y de regreso.
- ▶ La tecnología ha abierto las puertas hacia un nuevo enfoque de entendimiento y toma de decisiones, la cual es utilizada para describir enormes cantidades de datos. El concepto de Big Data (Ciencia de datos para negocios) aplica para toda aquella información que no puede ser procesada o analizada utilizando métodos o herramientas tradicionales.

- ▶ Toda esta información recabada, permitiría a las empresas mejorar sus servicios, logística, líneas de producción, abrir oportunidades de mercado, etcétera.
- ▶ El objetivo de la materia de Ciencia de Datos e Inteligencia de Negocios es que cuenten con conocimientos y habilidades en la modelación, interpretación e incorporación de datos en la toma de decisiones, evaluando y tomando en cuenta grandes volúmenes de información en un contexto de negocios, esto mediante métodos cuantitativos que ayuden al análisis de información.



**Datos:** son observaciones de fenómenos del mundo real.

**Aprendizaje automático:** Es un subconjunto de la Inteligencia Artificial y se refiere al estudio de algoritmos que mejoran a través de la experiencia.

**Característica (features):** Es una propiedad individual medible de un fenómeno que se observa. Éstas pueden ser numéricas o estructurales.

**Conocimiento del dominio:** Consiste en saber con precisión cómo se hace algo, tener una intuición de los principios subyacentes e integrarlos en la red de conocimiento existente.

**Minería de datos:** Es un conjunto de técnicas que permiten explorar grandes bases de datos, de manera automática o semiautomática, con el objetivo de encontrar patrones que expliquen el comportamiento de estos datos.

**Sobreajuste:** Es la situación en la que un modelo se ajusta muy bien a los datos actuales pero falla al predecir nuevas muestras. Suele ocurrir cuando el modelo se ha basado demasiado en patrones y tendencias en el conjunto de datos actual que no ocurrirían de otra manera.



<https://www.python.org/>



**ANACONDA®**

<https://www.anaconda.com/products/individual>



<https://jupyter.org/>