

Universidad Nacional Autónoma de México
Facultad de Ciencias



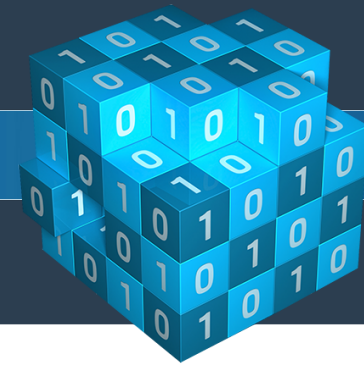
Profesores:

Dra. Amparo López Gaona
M. I. Gerardo Avilés Rosas

Laboratorista:

José Luis Vázquez Lázar





1. Objetivo

Al finalizar esta práctica el alumno será capaz de:

- Apreciar las distintas características de un conjunto de datos.
- Extraer un panorama general de un conjunto de datos.
- Apreciar las dificultades y bondades de conocer el conjunto de datos antes de cualquier transformación.

2. Marco teórico

La **Minería de Datos** (*Data Mining*) también conocida como **Exploración de Datos**, se define como el analista de (a menudo grandes) datos que se establece para encontrar relaciones insospechadas, intentando descubrir patrones para resumir los grandes volúmenes de éstos en formas novedosas que sean comprensibles y útiles para el titular de los datos. La entrada normalmente se administra en una tabla y para la salida puede haber reglas, clústers, estructuras de árbol, gráficos, ecuaciones, modelos, etc.

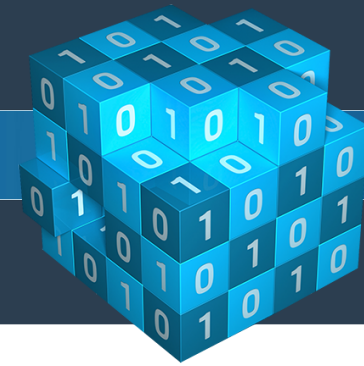
3. Instrucciones

- Descargar e instalar el software **RapidMiner Studio**
- Descargar el archivo **Adult.cvs** que se encuentra en la página del curso. La información de este dataset la puedes consultar en <https://archive.ics.uci.edu/ml/datasets/Adult>

4. Actividades

1. Da una descripción general del dataset. Para esto, deberás indicar el nombre del conjunto de datos, de dónde se obtuvo, quién lo donó, en qué fecha, con qué objetivo, número de registros y cualquier otra información que consideres relevante.





2. Carga el conjunto de datos usando el **Wizard** de la herramienta **RapidMiner Studio**. A continuación, clasifica cada uno de los atributos del conjunto de datos de acuerdo con las siguientes categorías: **nominal, numérico, binominal, polinomial, entero, real, texto, fecha, hora, fecha hora o ruta de archivo**. Posteriormente, deberás elaborar un diccionario de datos para el conjunto (con ayuda de **Rapidminer**), este deberá contener, para cada atributo: **nombre, definición, dominio de datos**, y además, identificar:
 - **Valores ausentes**. ¿Cómo identifica la herramienta a estos valores?, ¿coincide con el conocimiento que tiene sobre los atributos? (es decir, ¿es la representación que debería considerar la herramienta?).
 - **Media, mediana, desviación estándar y moda**, según corresponda.
3. Elabora **gráfica de barras, histograma, gráfica de burbujas, gráfica de desviación y gráfica de pastel** de 5 atributos adecuados que tu elijas (una gráfica por atributo). Deberás explicar qué indica cada gráfica y qué valores anómalos identifica.

5. Entregables

Deberás enviar un archivo .zip, con nombre <número de cuenta>_practica09, que contenga lo siguiente:

- Un archivo .pdf con las actividades solicitadas.
- Un archivo README.txt que contenga tu nombre completo, tu número de cuenta y tu correo.

a la dirección de correo luis_lazaro@ciencias.unam.mx con el asunto [A&MD2018-2] Practica09.

