

**Universidad Nacional de Rosario**

**Facultad de Ciencias Económicas y Estadística**

**Secretaría de Posgrado y Formación Continua**

**Escuela de Graduados y Extensión Universitaria**

###### CURSO-TALLER “MANEJO DE GRANDES BASES DE DATOS Y NOCIONES DE DATA MINING”

***CONTENIDOS:***

1. Introducción al Data Mining. Definiciones. Relación con conceptos adquiridos en la carrera. Técnicas utilizadas. Mención de softwares disponibles. Ejemplos de aplicaciones en diferentes áreas.
2. Manejo de base de datos: descripción de tablas de datos (diccionarios, estructuras, categorías de variables). Importación de datos. Unión de tablas. Identificación de la unidad de observación y sumarización según unidad de análisis.
3. Depuración de datos: chequeo y validación de la información. Evaluación de casos repetidos y duplicados. Análisis de outliers, valores nulos, valores perdidos. Labels de variables y categorías. Análisis por variable.
4. Creación de nuevas variables a partir de las existentes. Selección de muestras de validación y desarrollo. Mención de pasos a seguir dentro del proceso de Data Mining.

***REQUISITOS:***

Los asistentes deberán tener conocimientos previos en el manejo de la herramienta SAS.

***MODALIDAD DE EVALUACIÓN:***

Presentación de trabajo práctico en la última clase, en base a un caso similar al presentado durante el taller.

***BIBLIOGRAFÍA SUGERIDA:***

BERRY, M.J.A.; LINOFF, G. “Data Mining Techniques for Marketing Sales and Customer Relationship Management” Wiley Publishing, 2004.

DER, G.; EVERITT, B.S. “A handbook of Statistical using SAS” Chapman&Hall/CRC, 2002.

HAN, J.; KAMBER, M. “Data Mining. Concepts an techniques” Elservier Inc., 2006.

HERNÁNDEZ ORALLO, J; RAMÍREZ QUINTANA, M.J.; FERRI RAMÍREZ, C. “Introducción a la Minería de Datos” Pearson Educación, Madrid, 2004.