



Introducción a Data Mining

Dr. Juan M. Ale

Mg. Gaston Pezzuchi, MSc.

SAS SEMMA ©

- El *SAS Institute* © define a la **minería de datos** como el proceso de:
 - Sampling,
 - Exploring,
 - Modifying,
 - Modeling,
 - Assessing

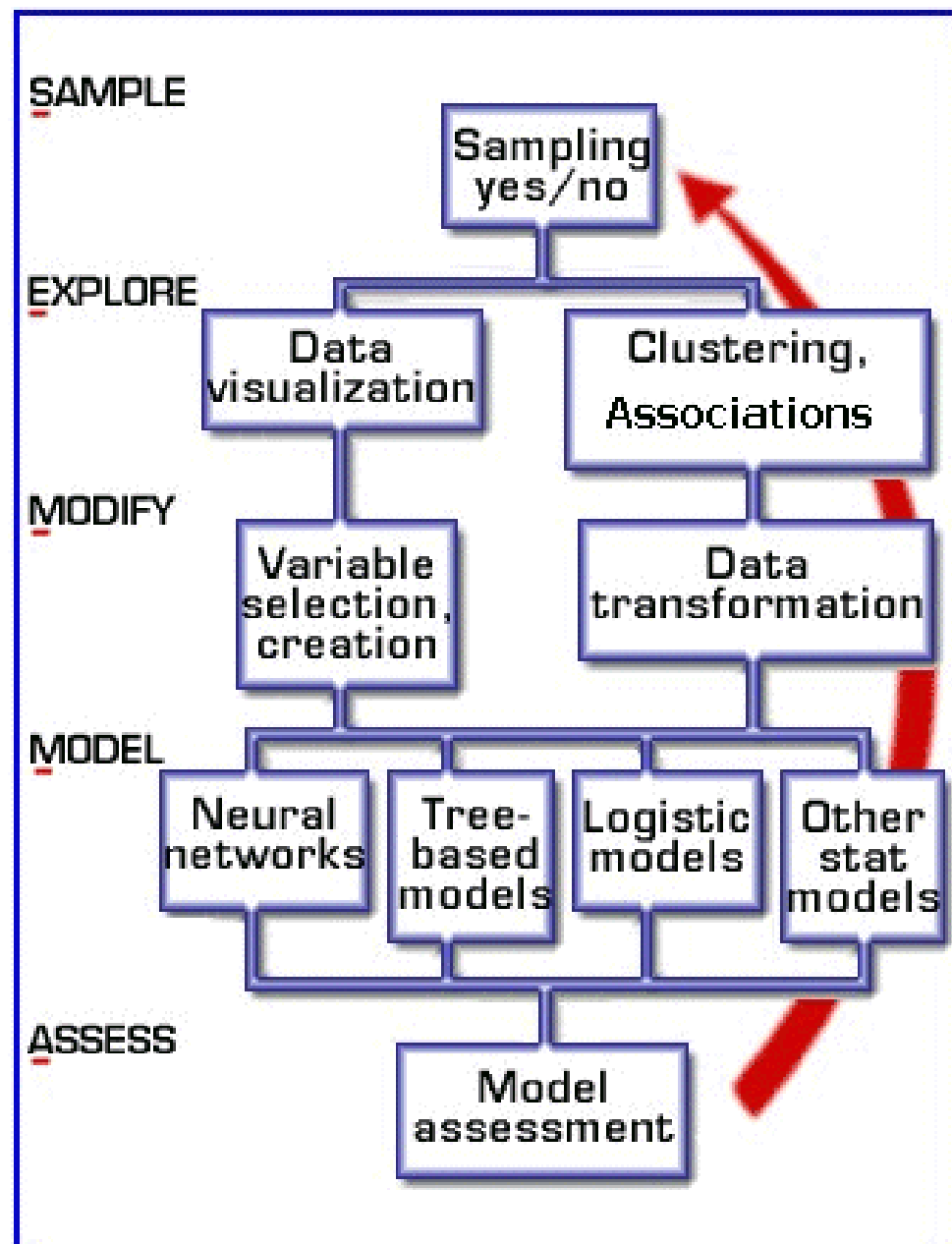
Aplicado sobre grandes cantidades de datos de manera de **descubrir** patrones previamente desconocidos que pueden ser empleados como ventaja de negocios.

Este proceso de minería de datos puede aplicarse a lo largo de una *gran variedad de industrias*, y provee metodologías para problemas de negocios tan diversos como la detección de fraudes, retención de clientes, marketing y segmentación de Mercado, análisis de riesgo, de afinidad, satisfacción del cliente, predicción de bancarrotas y análisis de portfolios.

SAS SEMMA ©

- **Sample** – Muestrear los datos de manera tal que el conjunto resultante contenga suficiente información significativa y a la vez sea suficientemente pequeño para poder ser procesado.
- **Explore** – Explorar los datos buscando relaciones anticipadas, tendencias no anticipadas y anomalías de manera de ganar un mayor entendimiento e ideas.
- **Modify** – Modificar los datos, creando, seleccionando y transformando las variables para focalizar el proceso de selección de modelos.
- **Model** - Modelar los datos empleando herramientas analíticas para buscar una combinación de los datos que predigan adecuadamente.
- **Assess** – Evaluar los datos, evaluando la utilidad y confiabilidad de los hallazgos resultantes del proceso de minería de datos.

SAS SEMMA ©



SAS SEMMA ©

- Se pueden incluir o no todos los pasos SEMMA en un análisis, y es probable repetir uno o varios de ellos antes de estar satisfecho con el resultado.