

# MODELO PREDICTIVO DE DESERCIÓN ESTUDIANTIL

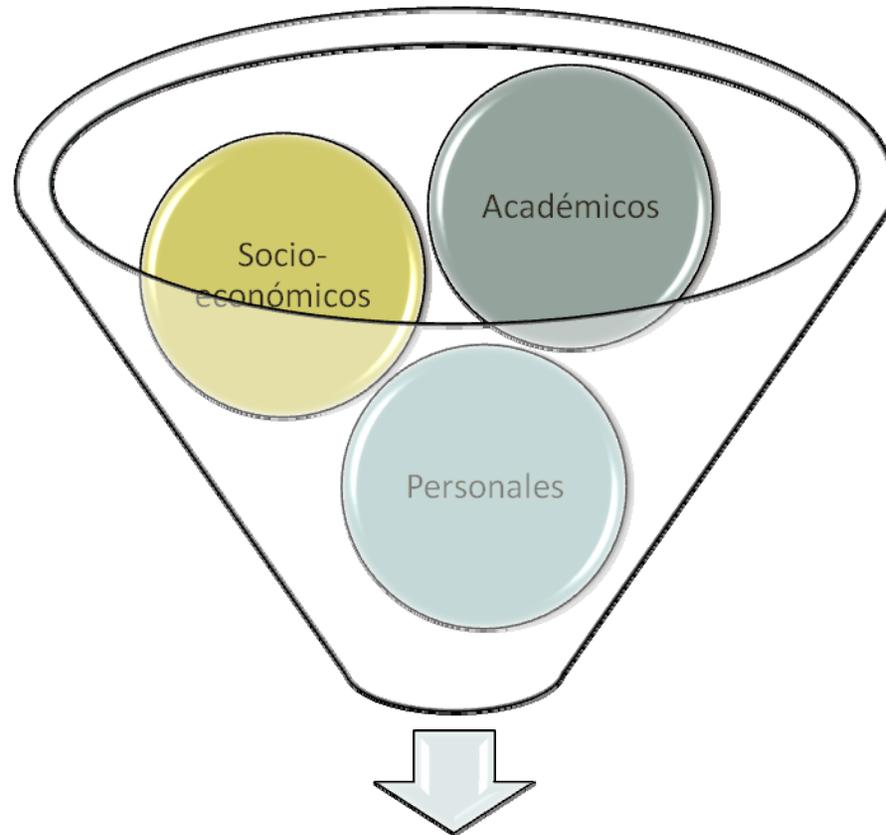
## APLICACIÓN DE TÉCNICAS DE MINERÍA DE DATOS

Yegny Amaya, Edwin Barrientos, Universidad Francisco de  
Paula Santander, Colombia

Diana Heredia Vizcaíno, Universidad Simón Bolívar,  
Colombia

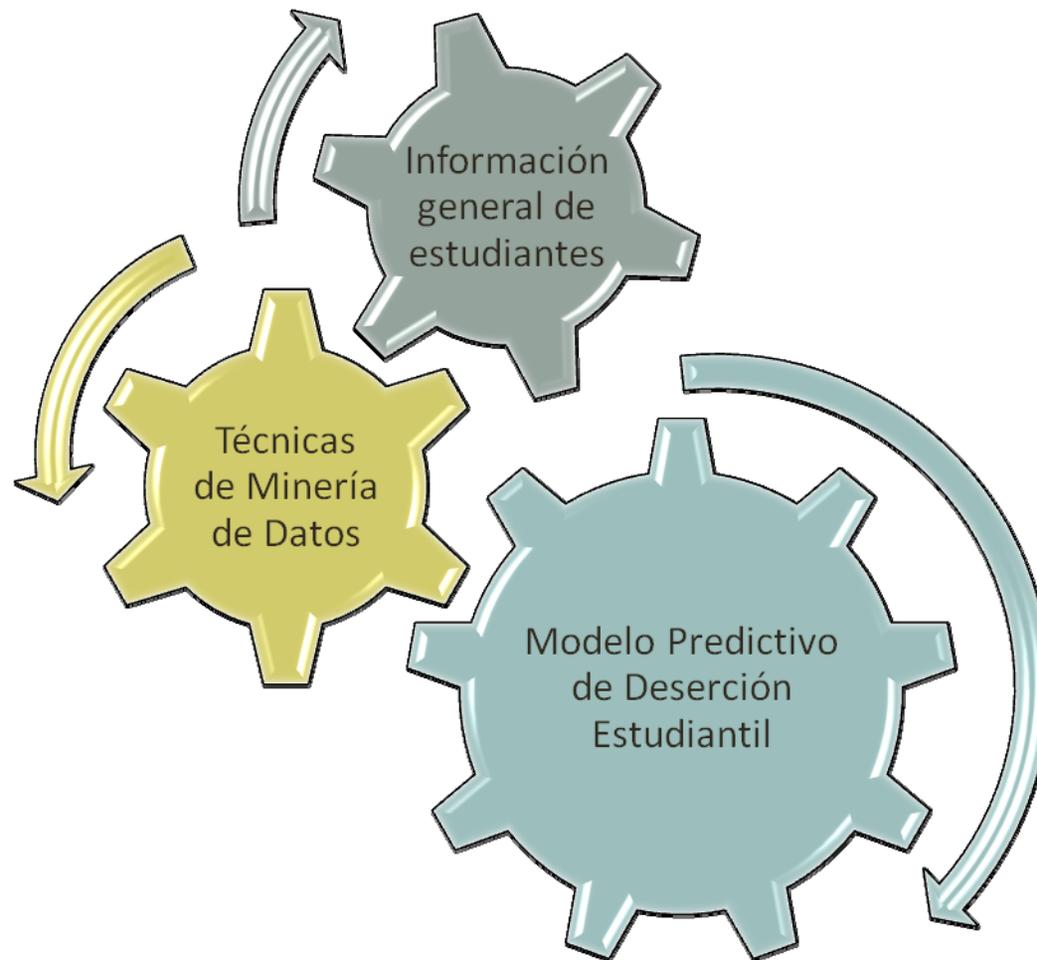
Mayo de 2014

# SISTEMAS DE INFORMACIÓN EN INSTITUCIONES DE EDUCACIÓN SUPERIOR



Informes /Reportes generales  
de los estudiantes

# APLICACIÓN DE MINERÍA DE DATOS



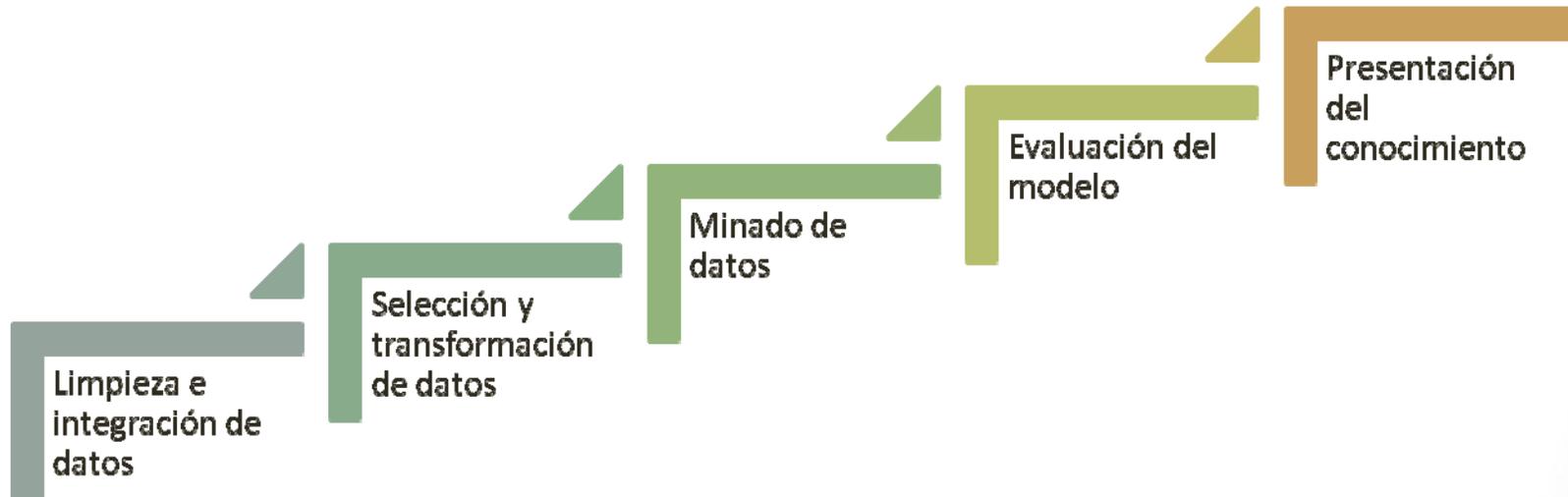
# DESERCIÓN EN LAS IES EN COLOMBIA

- Tasa de cobertura de la educación superior en Colombia (2009): 35.5%
- Tasa de deserción: En 2004 fue de 48.4% y en 2008, 44.9%. Se desea disminuirla a 25% en 2019 (Esta tasa se estima como promedio de abandono de estudios por cohorte de ingreso).
- Causas de deserción observadas:
  - Académicas
  - Socio económicas
  - Orientación vocacional y profesional

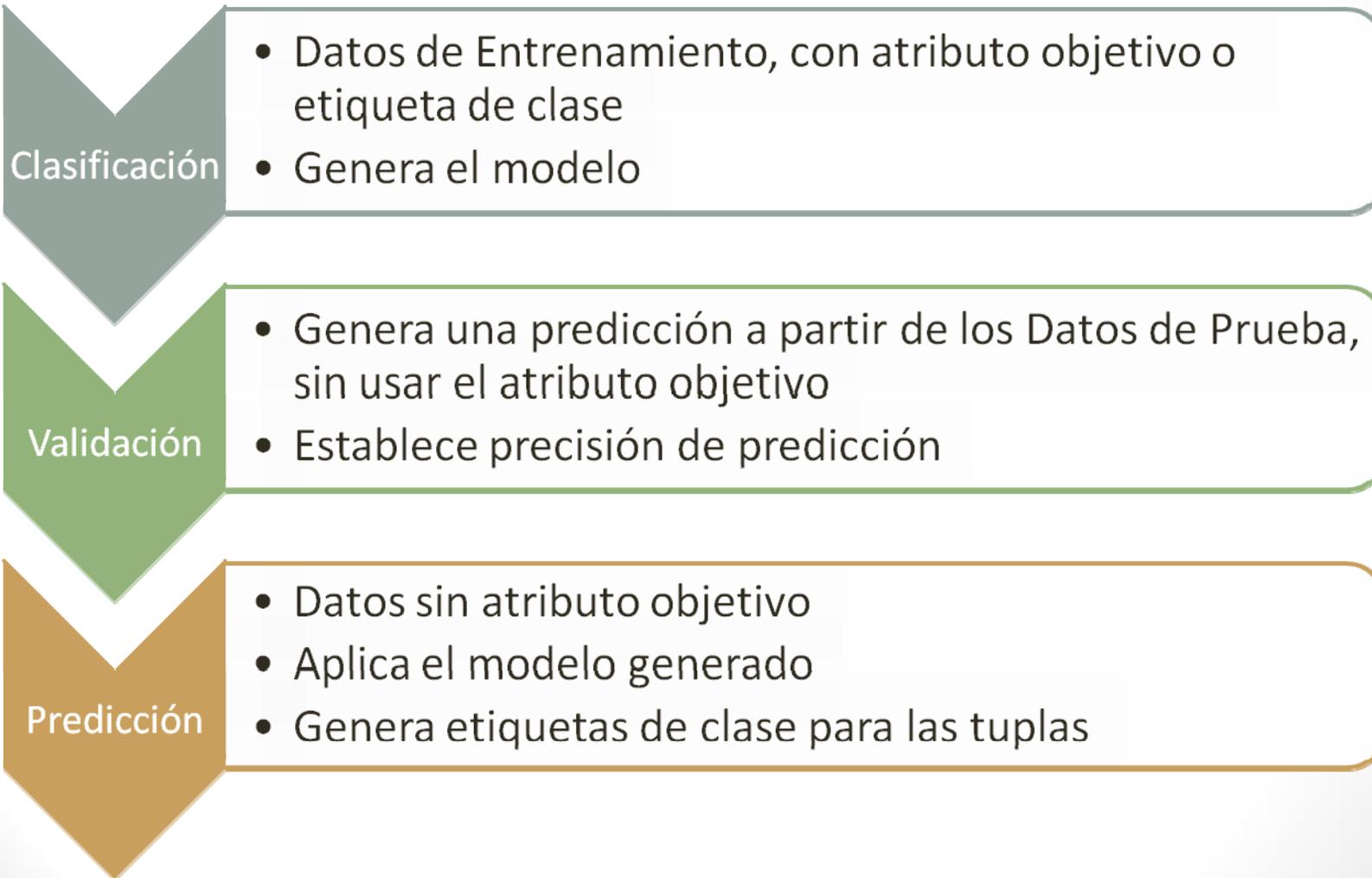
Fuente: Deserción estudiantil en la educación superior colombiana. Ministerio de Educación, 2009.



# PROCESO DE DESCUBRIMIENTO DEL CONOCIMIENTO



# MODELADO PREDICTIVO POR CLASIFICACIÓN



# ALGUNOS ANTECEDENTES

PAIS	TÍTULO DEL ESTUDIO	TÉCNICA DE MINERÍA DE DATOS UTILIZADA
Colombia	Una lectura sobre deserción universitaria en estudiantes de pregrado desde la perspectiva de la minería de datos.	árboles de decisión C4.5
México	Minería de datos: predicción de la deserción escolar mediante el algoritmo de árboles de decisión y el algoritmo de los k vecinos más cercanos.	Árboles de decisión C4.5 Técnica de los k vecinos más cercanos
Estados Unidos	A Model to Predict Ohio University Student Retention From Admissions and Involvement Data.	Regresión lineal, regresión logística, árboles de decisión
Estados Unidos	New Directions in Education Research: Using Data Mining Techniques to Explore Predictors of Grade Retention.	árboles de clasificación y regresión logística
Croacia	Student dropout analysis with application of data mining methods.	regresión logística, árboles de decisión y redes neuronales
India	Mining Education Data to Predict Student's Retention: A comparative Study.	Árbol de decisión ID3 y C4.5

# OBJETIVO DEL ESTUDIO

- Este estudio fue realizado por los estudiantes de la Maestría en Ingeniería de Sistemas y Computación de la Universidad Simón Bolívar, Yegny Amaya y Edwin Barrientos, dirigidos por la docente Diana Heredia Vizcaíno.
- Su objetivo inicial fue aplicar técnicas de minería de datos para generar un modelo predictivo y observar su comportamiento bajo ciertas circunstancias.
- El modelo busca predecir la probabilidad de que un estudiante abandone sus estudios, dadas sus características socio-económicas y académicas.

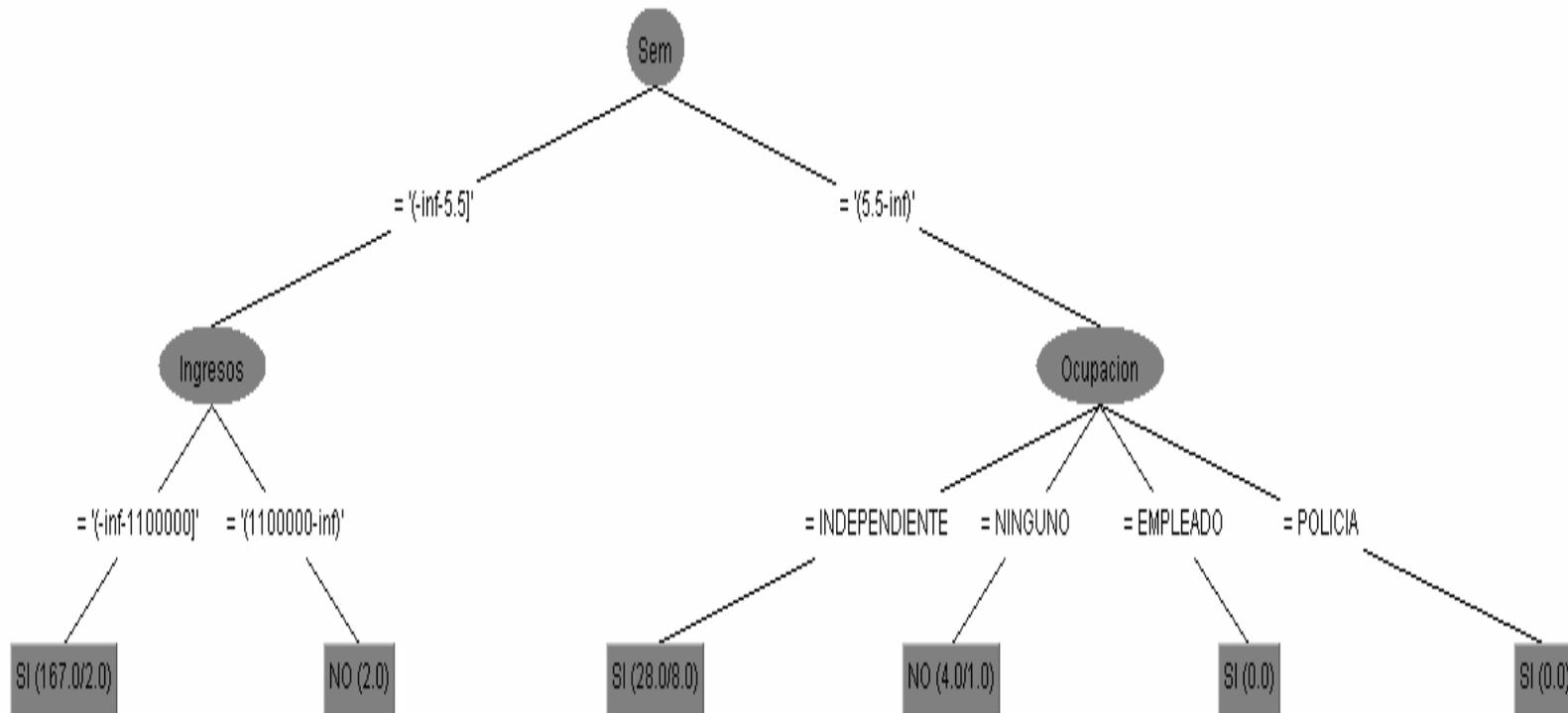
# MODELO PREDICTIVO DE DESERCIÓN

- Técnicas de minería utilizadas: Árboles de decisión ID3 y C4.5
- Fuente de datos: Sistema de información académico – administrativo de la Universidad Simón Bolívar (Barranquilla – Colombia)
- Muestra de datos: Registros correspondientes a estudiantes de Ingeniería de Sistemas, en el periodo comprendido entre 2007 y 2012.
- La muestra inicial consta de 40 atributos (Características), que incluyen datos personales y familiares, socio-económicos y académicos.
- El atributo para realizar clasificación es Desertado, según registro en el sistema de información.

# MODELO PREDICTIVO DE DESERCIÓN

- Atributos finales:
  - Semestre cursado actualmente
  - Edad actual
  - Ciudad de procedencia
  - Estrato
  - Jornada
  - Sexo
  - Valor de la matrícula
- Ocupación
- Materias cursadas
- Materias perdidas
- Promedio
- Estado Civil
- Nivel de estudios del Padre
- Nivel de estudios de la Madre
- Ingresos
- Desertado

# MODELO PREDICTIVO DE DESERCIÓN



# CONCLUSIONES

- El modelo obtenido como resultado de este estudio es solamente una muestra del poder de los procesos de descubrimiento de conocimiento basado en herramientas de aprendizaje automático y minería de datos.
- El modelo predice con aceptable grado de precisión la probabilidad de que una persona abandone sus estudios, dados los valores de sus atributos en un momento dado. Con base en ello, las instituciones pueden identificarlos e implementar planes para evitarlo.
- Cada institución puede presentar características muy particulares, las cuales generarían modelos diferentes; sin embargo, el proceso si es replicable.

GRACIAS....