

# CURSOS DE VERANO 2014

**TÍTULO DEL CURSO: APROXIMACIÓN PRÁCTICA A LA CIENCIA DE DATOS Y BIG DATA: HERRAMIENTAS KNIME, R, HADOOP Y MAHOUT**

**TÍTULO PONENCIA: KNIME. RESOLUCIÓN DE CASOS PRÁCTICOS**

**NOMBRE PROFESOR: CRISTÓBAL J. CARMONA**

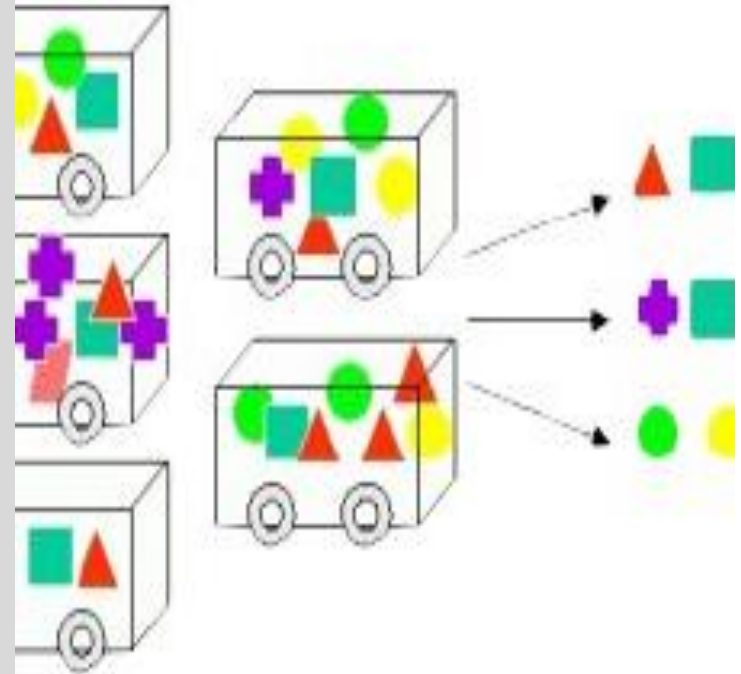
## MATERIAL ADICIONAL

- <http://tech.knime.org/getting-started>
- <https://www.youtube.com/user/KNIMETV>
- knime\_tutorial.pdf
- <http://informationandvisualization.de/blog/knime-interactive-views>
- <http://www.dataminingreporting.com/blog/category/knime>
- <http://tech.knime.org/forum/knime-general>

# Caso 3 – Segmentación y Asociación

## Objetivo:

- Utilización de técnicas de aprendizaje no supervisado.
- Análisis mediante clustering, asociación y subgrupos.
- Justificar estrategias de actuación.



## CASO 3: Segmentación y Asociación

### DESCRIPCIÓN

- Utilizarán técnicas de aprendizaje no supervisado para el análisis descriptivo de los datos.
- Vamos a trabajar con un conjunto de datos de captación de clientes de activo en banca.
- Banco tiene clientes:
  - Pasivo (ahorro y/o inversión)
  - Activo (financiación) – CRECER AQUÍ.

## CASO 3: Segmentación y Asociación

### DESCRIPCIÓN

- El principal objetivo que pretende el banco es la manera de convertir sus clientes de pasivo hacia activo.
- En el año anterior el banco realizó una campaña de marketing con conversión del 9% y este año persigue:
  - Grupos con características semejantes a las conversiones del año anterior.
  - Relación entre determinados productos y la conversión
  - Modelar el comportamiento de las conversiones del año anterior para conseguir que un cliente acepte un préstamo.

## CASO 3: Segmentación y Asociación

### DESCRIPCIÓN

- Contamos con un conjunto de datos:
  - 5000 instancias
  - 13 variables (categóricas y numéricas)
  
- Es necesaria la preparación de los datos para esta práctica con los conocimientos adquiridos previamente:

## CASO 3: Segmentación y Asociación

### DESCRIPCIÓN

- Utilizar aquellas variables que caracterizan al cliente.
- Ignorar aquellas variables que no se consideren necesarias para el estudio descriptivo del problema.
- Comprobar que todas las variables sean numéricas.
- Aplicar normalización para que las métricas de distancia y visualización funcionen correctamente.

## CASO 3: Segmentación y Asociación

### TAREAS A REALIZAR

#### CLUSTERING

- Analizar el uso de k-means con  $k=\{2,\dots,5\}$  y visualizar mediante nubes de puntos, gráficos de burbujas, etc.



## CASO 3: Segmentación y Asociación

### TAREAS A REALIZAR

#### CLUSTERING

- Analizar el uso del algoritmo jerárquico aglomerativo con distintos valores de corte.
- Analizar la utilización de un muestreo previo del conjunto de datos. Es importante mantener el 9% de clientes que aceptaron un préstamo en la muestra.

## CASO 3: Segmentación y Asociación

### TAREAS A REALIZAR

#### CLUSTERING

- Trabajar la visualización de este modelo mediante:
  - Dendograma.
  - Mapas de temperatura.

## CASO 3: Segmentación y Asociación

### TAREAS A REALIZAR

#### REGLAS DE ASOCIACIÓN

- Analizar los datos mediante el algoritmo Apriori:
  - Utilizar el *Association Rule Learner (Borgelt)*
  - Es necesario como entrada una lista de transacciones mediante el *Create Collection Column*.

## CASO 3: Segmentación y Asociación

### TAREAS A REALIZAR

#### REGLAS DE ASOCIACIÓN

- No se utilizará la variable de la aceptación del préstamo.
- Necesaria una discretización (basada en cuartiles o mediante visualización) con frecuencias relativamente equilibradas.

## CASO 3: Segmentación y Asociación

### TAREAS A REALIZAR

#### REGLAS DE ASOCIACIÓN

- Analizar los resultados mediante gráficas:
  - Diagrama de líneas con el número de reglas.
  - Nubes de puntos con soporte y confianza.
- Analizar las posibles ventas cruzadas.

## CASO 3: Segmentación y Asociación

### TAREAS A REALIZAR

#### DESCUBRIMIENTO DE SUBGRUPOS

- Es una técnica descriptiva pero mediante aprendizaje supervisado.
- Necesitamos establecer una variable objetivo.
- Utilizar el algoritmo Apriori con el objetivo de la concesión de préstamo.

## CASO 3: Segmentación y Asociación

### TAREAS A REALIZAR

#### DESCUBRIMIENTO DE SUBGRUPOS

- Es necesaria preparar los datos:
  - Discretización
  - Balancear los valores del consecuente
  
- Analizar las reglas obtenidas.

# CURSOS DE VERANO 2014

**TÍTULO DEL CURSO: APROXIMACIÓN PRÁCTICA A LA CIENCIA DE DATOS Y BIG DATA: HERRAMIENTAS KNIME, R, HADOOP Y MAHOUT**

**TÍTULO PONENCIA: KNIME. RESOLUCIÓN DE CASOS PRÁCTICOS**

**NOMBRE PROFESOR: CRISTÓBAL J. CARMONA**