

# Minería de Datos

## Lab. #6

Ivan Saavedra, Ph.D.

[saavedrai@uninorte.edu.co](mailto:saavedrai@uninorte.edu.co)

Universidad del Norte  
División de Ingenierías  
Dpto. Ingeniería de Sistemas



202130

# 1. Clasificación de reorden de compra

Aplique el clasificador de Bayes para clasificar si una línea de producto será ordenada nuevamente basados en la sucursal, ciudad, tipo y sexo del cliente. Considere el conjunto de datos adjunto para esta implementación.

Data Source: [https://github.com/sushantag9/Supermarket-Sales-Data-Analysis/blob/master/supermarket\\_sales%20-%20Sheet1.csv](https://github.com/sushantag9/Supermarket-Sales-Data-Analysis/blob/master/supermarket_sales%20-%20Sheet1.csv).

## Metodología

- Realizar el análisis exploratorio que responda las siguientes preguntas:
  - Cual es la línea de producto que mas se compra?, Cual es la frecuencia diaria de ese producto? En que cantidad se compra ese producto?
- Preparar y adecuar el conjunto de datos de entrada para poder alimentar los modelos
- Entrenar los modelos
- Encontrar predicciones
- Evaluar precisión del modelo
- Comparar precisiones de los modelos. Cual es el modelo que presenta mejor precisión ?



[https://www.vhv.rs/viewpic/hxhibmR\\_free-download-of-shopping-cart-icon-png-grocery/](https://www.vhv.rs/viewpic/hxhibmR_free-download-of-shopping-cart-icon-png-grocery/)

## 2. Clasificador de pronóstico de cáncer

Aplique los diferentes clasificadores vistos en clase para clasificar si una persona puede tener cáncer o no dependiendo del tipo de la masa encontrada afectada (esta puede ser benigna o maligna). Esta decisión se encuentra basada en atributos como Uniformidad del tamaño y forma de las células, la cromatina, los núcleos, entre otros.

Considere el conjunto de datos suministrado para entrenar al modelo.

<https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Original%29>

### Metodología

- Realizar un análisis descriptivo exploratorio
  - Analizar frecuencias, diagrama de cajas para analizar atributos, histogramas, y matriz de correlación
- Entrenar los modelos
  - Visualizar el árbol de decisión encontrado para el modelo de Árbol de Decisiones
- Encontrar predicciones
- Evaluar precisión de los modelos
- Comparar precisiones de los modelos. Cual es el modelo que presenta mejor precisión ?

# 3. Clasificación de Dígitos

Aplice los diferentes clasificadores vistos en clase para clasificar dígitos que han sido escritos a mano. Esta decisión se encuentra basada en la codificación de las imágenes de los dígitos en una matriz de pixeles de 28x28, representando dígitos de 0 a 9.

Considere los conjuntos de datos (MNIST) suministrados para entrenar y testear el modelo.

## Metodología

- Realizar un análisis descriptivo exploratorio
  - Visualice un ejemplo de imagen de cada dígito
  - Analizar por medio de un diagrama de cajas los valores de los atributos en general
  - Analizar matriz de correlación de los atributos con la variable de respuesta
- Entrenar los modelos
  - Visualizar el árbol de decisión encontrado para el modelo de Árbol de Decisiones
- Encontrar predicciones
- Evaluar precisión de los modelos
- Comparar precisiones de los modelos. Cual es el modelo que presenta mejor precisión ?

# Modelos de Clasificación

## **Entregable:**

- Un archivo de Jupyter Notebook con el desarrollo de los análisis
- Las conclusiones y respuestas al objetivo del análisis deben ser contestadas en el mismo notebook.
- Se sugiere que comente las secciones de manera adecuada para una mejor interpretación de su análisis.
- La fecha de entrega es el Sábado 23 de Octubre del 2021 vía catalogo web enlace de laboratorios.