

# Maestría en Exploración de Datos y Descubrimiento del Conocimiento

## Análisis Inteligente de Datos

### Trabajo Práctico N° 2

#### Análisis Multivariado de Datos

##### Descripción de la base de datos

1. Nombre de la base de datos: Conjunto de datos de calidad del vino
2. Área/Tema: Negocios
3. Link de la base: <https://archive.ics.uci.edu/ml/datasets/wine+quality>
4. Fecha de publicación: 2010-10-07
5. Fecha de relevamiento de los datos: mayo 2004 – febrero 2007
6. Fuente /Responsable/Propietarios de la base de datos: Paulo Cortez, Universidad de Minho, Guimarães, Portugal, <http://www3.dsi.uminho.pt/pcortez>. A. Cerdeira, F. Almeida, T. Matos y J. Reis, Comisión de Viticultura de la Región del Vino Verde (CVRVV), Oporto, Portugal
7. Base de datos abierta (sí/no): sí
8. Tipo de archivo (matriz/no matriz): matriz
9. Formato del archivo/s (csv/xlsx/txt/zip/otro (especificar)): xlsx
10. Tipo de datos (multivariado/univariado/serie de tiempo/texto/otro): multivariado
11. Características del atributo (categórico/numérico/mixto): numérico
12. Información de los atributos:

Variables de entrada (basadas en pruebas fisicoquímicas):
1 - acidez fija
2 - acidez volátil
3 - ácido cítrico
4 - azúcar residual
5 - cloruros
6 - anhídrido sulfuroso libre
7 - anhídrido sulfuroso total
8 - densidad
9 - pH

## Maestría en Exploración de Datos y Descubrimiento del Conocimiento

### Análisis Inteligente de Datos

10 - sulfatos
11 - alcohol
Variable de salida (basada en datos sensoriales):
12 - calidad (puntuación entre 0 - muy malo y 10 - excelente)
13 - variedad (blanco 1 / tinto 2)

#### Consignas:

A.- Para la base de datos seleccionada genere una muestra aleatoria estratificada y balanceada por variedad de vino de tamaño  $n = 2000$  utilizando como semilla los últimos tres dígitos del DNI/PASAPORTE.

B. De ahora en más trabaje con esta base de datos para el resto del trabajo práctico. Realice los procedimientos que se detallan a continuación acompañando los procedimientos de los gráficos que considere adecuados.

1.- Aplique el Análisis de Componentes Principales a la base de datos. Presente los resultados y gráficos que considere adecuados. Interprete los resultados.

2.- Realice el Análisis Discriminante para clasificar los vinos según la variable variedad de vino. Interprete los resultados.

3.- Aplique el algoritmo SVM al conjunto de datos. Interprete los resultados.

4.- Elija un método de Clasificación jerárquico y aplíquelo a los datos. Interprete los resultados.

5.- Aplique a los datos el método de clasificación no jerárquico K-means. Interprete los resultados.

C.- Presente un informe final de 2 carillas como máximo, no incluya gráficos, explicando las conclusiones del trabajo realizado, mencione si es necesario validar supuestos requeridos para aplicar el método. Compare los resultados de los métodos supervisados y establezca conclusiones. Por otro lado, compare los métodos no supervisados y presente sus conclusiones.

D.- Realice el trabajo en R y adjunte el código en R, el RMD y el html donde se puede observar el código y las salidas junto con la interpretación de cada método.

Fecha de entrega: 9 de julio antes de las 9 hs.

Formato de entrega: archivo.html en Entrega TP Aula Virtual.