

Maestría en Exploración de Datos y Descubrimiento del Conocimiento

Análisis Inteligente de Datos

Trabajo Práctico N° 2 Obligatorio

Descripción de la base de datos

1. Nombre de la base de datos: Sleep Health and Lifestyle Dataset
2. Año de publicación: 2023
3. Fuente /Responsable/Propietarios de la base de datos: Laksika Tharmalingam (modificación de un dataset de Google)
4. Base de datos abierta (sí/no): sí
5. Tipo de archivo (matriz/no matriz): matriz
6. Formato del archivo/s: csv
7. Tipo de datos (multivariado/univariado/serie de tiempo/texto/otro): multivariado
8. Información de los atributos y base de datos: [[Data card](#)]

Consignas:

1. Para la base de datos seleccionada genere una muestra aleatoria estratificada por "Sleep Disorder" de tamaño $n = 300$ utilizando como semilla los últimos tres dígitos del DNI/PASAPORTE. Para que la muestra sea proporcional, se sugiere seleccionar la muestra con los siguientes tamaños: "None"= 180; "Sleep Apnea"= 60; "Insomnia"= 60. Guarde los datos en un archivo y realice todo el trabajo práctico con la muestra generada.

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:

2. Considere la variable "Blood Pressure" y separe los datos dos variables: "cystolic pressure" y "diastolic pressure" .
3. Aplique el Análisis de Componentes Principales a la base de datos. Presente los resultados y gráficos que considere adecuados. Interprete los resultados.
4. Elija dos variables y aplique Análisis de Correspondencias Simple. Interprete los resultados.

5. Elija tres variables y aplique Análisis de Correspondencias Múltiples. Interprete los resultados.
6. Realice el Análisis Discriminante para clasificar los pacientes según la variable trastorno de sueño. Interprete los resultados.
7. Aplique el algoritmo SVM al conjunto de datos. Interprete los resultados.
8. Aplique a los datos un método de clasificación no jerárquico. Interprete los resultados.
9. Aplique a los datos el método de clasificación jerárquico KNN. Interprete los resultados.
10. 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.

Fecha de entrega: 10 de Julio antes de las 23.55 hs.

Formato de entrega:

- A. **Código en R** (.Rmd y .html) donde se puede observar el código y las salidas junto con la interpretación de cada método.
- B. **Informe** (.doc/.pdf, siguiendo las indicaciones del punto 10). La entrega se realiza con el nombre "**Apellido_nombre_TP2_AID.**" en "Entrega Trabajo Práctico 2" en Aula Virtual.