

TALLER ANALISIS EXPLORATORIO Y PROCESAMIENTO

Primer Parcial DATA SCIENCE

Fecha de entrega: 12 de Octubre 2020

Formato: Utilizar Markdown para escribir informes reproducibles y dinámicos con R.

Grupos de máximo 2 Estudiantes

1. Obtenga el conjunto **German Credit Data** del repositorio UCI Machine Learning Repository y cárguelo en Rstudio o Phyton
[https://archive.ics.uci.edu/ml/datasets/statlog+\(german+credit+data\)](https://archive.ics.uci.edu/ml/datasets/statlog+(german+credit+data))

Cargar Dataset en Phyton o Rstudio

2. Preprocesamiento de Datos .

- 2.1 Número de instancias
- 2.2 Número de atributos
- 2.3 ¿El conjunto de datos está etiquetado? ¿Cuántas clases tiene el conjunto de datos?
- 2.4 ¿Cuántos atributos son numéricos y cuántos categóricos?
- 2.5 Reporte la moda para cada atributo categórico
- 2.6 Reporte la media, rango y desviación estándar para cada atributo numérico
- 2.7 Determine la distribución de las clases (Diagrama de Densidad)
- 2.8 Escoja una técnica para la detección de datos atípicos y aplíquela sobre el conjunto de datos
- 2.9 Aplique al menos dos estrategias diferentes para manejar los datos faltantes
- 2.10 Convierta todas los atributos numéricos a categóricos utilizando dos estrategias diferentes.
- 2.11 Transforme el conjunto de datos de manera que todos los atributos sean numéricos
- 2.12 Discretice los atributos numéricos en máximo 5 bins de igual tamaño.

3 Responda las siguientes preguntas:

- 3.1 ¿Cuál es lo propósito predominante de los préstamos?
- 3.2 ¿Qué tipo de estatus tienen las personas que más hacen préstamos? ¿Y el perfil de la de menos préstamos? ¿Cuál es el perfil de las personas que hacen los préstamos más costoso? ¿Y el de los menos costosos?
- 3.3 ¿Puede establecer alguna relación entre edad, estatus personal y la clase?
- 3.4 ¿Puede establecer alguna relación entre clase de trabajo, el número de créditos, estatus personal y la clase?
- 3.5 ¿Existe alguna relación entre la cantidad solicitada y el número de meses del préstamo?
- 3.6 ¿Existe alguna relación entre la edad, el estatus, la clase y la cantidad del préstamo?
- 3.7 Pruebe diferentes combinaciones entre los atributos y establezca las relaciones entre ellos, reporte la herramienta de visualización que utilizó para tal fin.

4 Realice los siguientes procedimientos sobre alguno de los atributos del conjunto de datos, analice los resultados y extraiga resultados

- 4.1 Análisis de rangos intercuartiles
- 4.2 Histogramas
- 4.3 Gráfica de dispersión
- 4.4 Gráficas de coordenadas paralelas

5 Proponga tres preguntas y resuélvalas a partir de técnicas de visualización