



Actividad 1

Utilizar una base de datos de un problema de clasificación para aplicar las funciones vistas en clase y analizar los resultados.

Se recomienda que la base de datos solo contenga datos numéricos (no que sean todos binarios), y también que contenga más de 10 atributos(columnas).

Utilizar las siguientes funciones para entender el significado de información de las bases de datos.

- `head()`.
 - `shape`.
 - `dtypes`.
 - `describe()`.
 - `groupby()`.
- Funciones para la visualización de los datos.
 - `hist()`.
 - Función de densidad.
 - Boxplot.
 - Matriz de correlación.
 - Matriz de dispersión por clase.
- Transformación de datos.
 - Escalamiento.
 - Estandarización
 - Normalización.
 - Box-cox, Yeo-johnson (si se requiere).
- Métodos de Validación (Técnicas de Remuestreo) aplicando Regresión logística.
 - Validación cruzada.
 - División por porcentaje.
 - Validación cruzada con repeticiones.
 - División por porcentaje repetidos aleatoriamente.
- Métricas para evaluación de Algoritmos.
 - Matriz de confusión.
 - Kappa de Cohen.
 - Curva ROC-AUC (para clasificación binaria)



Puntos para Evaluación de la Actividad.

- Entregar en archivo PDF (con portada).
- No hay un formato en especial, por lo que el alumno diseñara su propio formato.
- Las bases de datos deben de contener una explicación con respecto a las características de los atributos y si es utilizada para algún problema en específico.
- Descripción teórica de cada una de las funciones utilizadas.
- **Análisis de los resultados Obtenidos.**
- Las imágenes y las tablas deben de tener descripción (pie de figura), en donde se explique lo que se está mostrando.
- Al final del documento, poner las referencias de las funciones utilizadas y la base de datos.

Se revisará el contenido y organización del documento.