Caso Grupal: Detección de anomalías y técnicas de agrupamiento

**Objetivos**

Mediante este trabajo se pretende que ponga en práctica la aplicación de los algoritmos de detección de anomalías u *outliers* y las técnicas de agrupamiento. El objetivo es detectar anomalías, generar y analizar agrupaciones aplicando algoritmos de aprendizaje máquina, con el fin de elegir el mejor método para el conjunto de datos dado. Debe detallar los pasos que hay que realizar para detección automática de valores inusuales y, por otro lado, analizar los *clúster* o grupos resultado de aplicar un algoritmo de agrupamiento.

**Descripción de la actividad**

El conjunto de datos con el cual vamos a trabajar se encuentra en el siguiente link y se puede descargar del aula virtual.

<https://www.kaggle.com/akshat0007/cardiotocology/data?select=CTG.csv>

<<CTG.csv>>

**Tareas a realizar**

* Análisis descriptivo de los datos:
  + De las variables numéricas, hallar datos estadísticos.
  + De las variables categóricas, listar las diferentes categorías y hallar la frecuencia de cada una de ellas.
  + Crear matriz de correlaciones existentes entre las variables numéricas del conjunto de datos, analizar los resultados.
* Tratamiento de missing. Si existen valores faltantes, decidir si eliminar los registros, llenarlos con valores como la media, la mediana o la moda y justifique su respuesta.
* Aplicar una técnica de detección de anomalías y explicar los resultados obtenidos.
* Aplicar una técnica de agrupamiento, explicar la técnica y los resultados obtenidos.
* Comentar las ventajas y desventajas de cada modelo.

**Rúbrica**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Descripción | Puntuación máxima  (puntos) | Peso  % |
| Criterio 1 | Realiza el análisis descriptivo de los datos de manera adecuada | 2 | 20% |
| Criterio 2 | Aplica adecuadamente el algoritmo de clúster | 3 | 30% |
| Criterio 3 | Maneja correctamente la detección de anomalías | 3 | 30% |
| Criterio 4 | Los comentarios sobre los resultados son pertinentes y adecuados | 2 | 20% |
|  |  | **10** | **100 %** |

**Extensión** **máxima:** 20 páginas.