

Universidad Católica "Nuestra señora de la Asunción"
Facultad de ciencias y tecnología
Complementos de Informática



Análisis exploratorio y selección de los datos a utilizar para los modelos.

Entrega 2 - Data mining

Alumno: Alain Vega

Profesor: Wilfrido Felix Inchausti Martínez

16 de Octubre del 2023

Índice

1. Datos a utilizar	2
2. Analisis exploratorio	3
2.1. Missing values	3
2.2. Ruidos y anomalias	3
2.3. Redundancia	3
2.4. Correlacion	3
3. Conclusiones	3

1. Datos a utilizar

Los datos a utilizar en el proyecto sera un archivo .csv (*Comma Separated Values*) llamado *heart.csv* que se obtuvo de la plataforma web *Kaggle* [1] el cual de inicio contiene 12 columnas, las cuales:

1. ***Age***. edad del paciente
2. ***Sex***. sexo del paciente, donde:
 - **M**: Masculino
 - **F**: Femenino
3. ***ChestPain***. tipo de dolor en el pecho el cual puede ser:
 - **TA**: *Typical Angina*
 - **ATA**: *Atypical Angina*
 - **NAP**: *Non-Anginal Pain*
 - **ASY**: *Asymptomatic*
4. ***RestingBP***. presión arterial en reposo medido en mililitros de mercurio [mmHg]
5. ***Cholesterol***. Colesterol serico medido en miligramos por decilitro de sangre [mm/dl]
6. ***FastingBS***. azúcar en sangre en ayunas medido en miligramos por decilitro [mg/dl]
7. ***RestingECG***. resultados del electrocardiograma en reposo, puede ser:
 - **Normal**: normal
 - **ST**: tener anomalía de la onda ST-T (inversiones de la onda T y/o elevación o depresión del ST $> 0,05$ mV)
 - **LVH**: muestra probable o definitiva hipertrofia ventricular izquierda según los criterios de Estes
8. ***MaxHR***. frecuencia cardíaca máxima alcanzada medido en pulsaciones por minuto [ppm]

9. ***ExerciseAngina***. angina inducida por el ejercicio, puede ser:
- **Y**: Si
 - **N**: No
10. ***OldPeak***. pico antiguo = ST
11. ***ST_Slope***. pendiente del segmento ST durante un ejercicio físico máximo en una prueba de esfuerzo cardíaco. Puede ser:
- **Up**: ascendente
 - **Flat**: plano
 - **Down**: descendente
12. ***HeartDisease***. sufrio de insuficiencia cardiaca
- **1**: insuficiencia cardiaca
 - **0**: normal

2. Analisis exploratorio

2.1. Missing values

2.2. Ruidos y anomalias

2.3. Redundancia

2.4. Correlacion

3. Conclusiones

Referencias

- [1] fedesoriano. *Heart Failure Prediction Dataset*. <https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction>. 2021.