

Ven y sana mi dolor Tienes la cura de este amor

José Vidal Cardona Rosas
Tecnologías para la Información en Ciencias
ENES, UNAM Morelia
vrosas832@gmail.com

Brian Kalid García Olivo
Tecnologías para la Información en Ciencias
ENES, UNAM Morelia
briankalid2000@gmail.com



Figure 1: Heart failure.

ABSTRACT

En el presente documento se realizará un análisis de grupos haciendo uso del algoritmo de aprendizaje no supervisado, k-medios (k-means en inglés) para determinar las características de las personas que son más propensas a sufrir insuficiencia cardíaca.

CCS CONCEPTS

• **Data Mining, Clustering** → **K-means**.

KEYWORDS

Data Mining, Clustering, K-means

1 DESCRIPCIÓN DE LOS DATOS

Los datos a utilizar cuentan con 13 columnas (descritas en la tabla) y con un total de 299 registros.

1.1 Fuente

La versión original de los datos fue recopilada por:

- Tanvir Ahmad
- Assia Munir
- Sajjad Haider Bhatti
- Muhammad Aftab

- Muhammad Ali Raza

(Government College University, Faisalabad, Pakistán) y fueron puestos a disposición por las mismas personas en FigShare bajo los derechos de autor *Attribution 4.0 International (CC BY 4.0: libertad para compartir y adaptar el material)* en julio de 2017.

La versión actual de los datos fue elaborada por:

- Davide Chicco (Instituto de Investigación Krembil, Toronto, Canadá)

Donada al Repositorio de Aprendizaje Automático de Irvine de la Universidad de California bajo los mismos derechos de autor *Attribution 4.0 International (CC BY 4.0)* en enero de 2020.

1.2 Información de atributos

- **age**: edad del paciente (años)
- **anaemia**: disminución glóbulos rojos (booleana)
0=No tiene|1=Si tiene
- **high blood pressure**: paciente con hipertensión (booleano)
0=No tiene|1=Si tiene
- **creatinine phosphokinase (CPK)**: nivel de la enzima CPK en sangre (mcg/L)

- **diabetes:** paciente con diabetes (booleano)
0=No tiene|1=Sí tiene
- **ejection fraction:** porcentaje de sangre que sale del corazón en cada contracción (porcentaje)
- **platelets:** plaquetas en la sangre (kiloplaquetas/ml)
- **sex:** Mujer u Hombre 0=Mujer|1=Hombre
- **serum creatinine:** nivel de creatinina sérica en sangre (mg/dl)
- **serum sodium:** nivel de sodio sérico en sangre (mEq/L)
- **smoking:** si el paciente fuma o no (booleano)
0=No fuma|1=Sí fuma
- **time:** período de seguimiento (días)
- **death event (target):** si el paciente falleció durante el período de seguimiento (booleano) 0=Sobrevivió|1=Murió

Una vez que los datos fueron analizados mediante un mapa de calor, se corroboró la inexistencia de valores nulos.

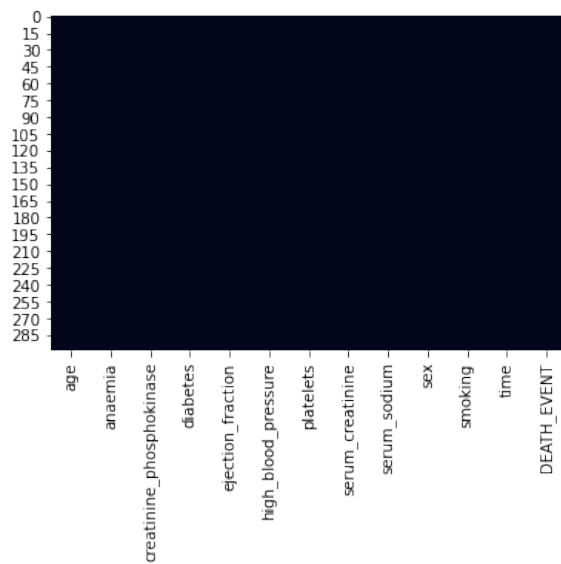


Figure 2: Mapa de calor para la detección de valores nulos.

2 DESCRIPCIÓN DE LA TAREA DE APRENDIZAJE NO SUPERVISADO

Contamos con un conjunto de datos de múltiples personas que tuvieron insuficiencia cardíaca, dicho conjunto de datos describe a cada persona con múltiples estadísticas, por lo tanto se procedió a usar un algoritmo de clustering, específicamente el algoritmo **k-means**, para agrupar a estas personas por grupos, de tal manera que podamos buscar similitudes en sus estadísticas de un grupo frente a otro y determinar cuáles son más proclives a morir de insuficiencia cardíaca.

3 DESCRIPCIÓN DE EXPERIMENTOS

4 CONCLUSIÓN

ACKNOWLEDGMENTS

The authors would like to thank Dr. Yuhua Li for providing the MATLAB code of the *BEPS* method. The authors would also like

to thank the anonymous referees for their valuable comments and helpful suggestions.

REFERENCES