

Caso práctico | Aprendizaje automático

Parte 1

Introducción

Los desfibriladores externos automáticos o DEA son dispositivos electrónicos que se usan de manera habitual para tratar ciertos tipos de arritmias. Son capaces de analizar el ritmo cardíaco y detectar si es susceptible de recibir un tratamiento cardíaco o no mediante una descarga eléctrica.

En RiseData interesa mucho este tipo de dispositivos salvavidas, que se suelen encontrar en lugares públicos de forma estratégica, como estadios deportivos. Su ubicación y las instrucciones que proporciona permiten que sean utilizados por cualquier persona que presencie una parada cardíaca, de forma que puedan iniciar de manera inmediata la reanimación de la persona afectada.

Planteamiento del problema

Para que estos dispositivos sean realmente efectivos, deben usarse lo antes posible. Por ello, es de suma importancia que cualquier persona sea capaz de usarlo sin necesidad de conocimientos previos. Para que esto pueda suceder, el dispositivo cuenta con un algoritmo predictivo de clasificación que detecta si existe o no alguna anomalía cardíaca.

Desarrollo

Un DEA ha llegado a las instalaciones de RiseData junto a algunos ficheros, para que se compruebe que estos dispositivos funcionan correctamente y están bien configurados antes de colocarlos en las instalaciones de los clientes.

Para ello, recibes un fichero con datos de pacientes (**datos_reto.csv**) con variables asociadas a los cambios en el patrón de ritmo en ventanas de 4 segundos etiquetados con 0 en caso de ritmo normal y 1 en caso de ritmo afectado. Debes encontrar un modelo supervisado capaz de clasificar correctamente el conjunto de datos **datos_onu.csv**.

El dataset consiste en un conjunto de 30 parámetros calculados a partir del Electrocardiograma (ECG). Estos parámetros recogen distintas características del ECG, desde su comportamiento temporal como la amplitud, pendiente, etc. Hasta características un poco más complejas como variables espectrales. A continuación, puedes ver la lista completa de parámetros:

Parámetros temporales	TCI = threshold crossing interval TCSC = threshold crossing sample count Exp = exponencial estandard Expmad = exponencial modificada MAV = valor medio absoluto count1 count2 count3 x1 x2 bCP
Parámetros espectrales	vFleak = VF filter M A1 A2 A3 x3 x4 x5 bWT
Parámetros de tipo tiempo-frecuencia	Li = basada en el análisis de ondas del ECG. Parámetros de complejidad del ECG CM = medida de la complejidad CVbin = covarianza abin = area Frqbin = frecuencia Kurt = Curtosis de una señal binaria extraída del ECG PSR = La reconstrucción del espacio de fase HILB = La transformación de Hilbert SamEn = Muestra de entropía

Ejercicio 1

Carga los datos y realiza una primera exploración de la información. Remata con una visualización de la misma.

Ejercicio 2

Realiza un análisis exploratorio de datos (EDA).

Ejercicio 3

Encuentra un modelo que haga una predicción por encima del 85% en accuracy.

Requisitos de la entrega

Debes entregar un notebook ejecutado que incluya:

- Las decisiones que vas tomando a lo largo del desarrollo del ejercicio.
- Las conclusiones a las que has llegado tras la realización del ejercicio.

Parte 2

Introducción

La segmentación de clientes es una herramienta muy empleada en marketing para agrupar clientes similares para poder dirigir políticas comerciales diferenciadas por grupo. Llevar a cabo este proceso permite fidelizar a clientes e intentar que se vayan moviendo a grupos más rentables para la compañía.

Planteamiento del problema

En RiseData necesitáis identificar el perfil de los deportistas individuales que no pertenecen a grandes equipos, para así poder desarrollar políticas de marketing para cada uno de ellos.

Desarrollo

Dispones de un fichero que contiene datos de clientes (**Clientes.csv**) basado en su histórico de ventas y con las siguientes variables:

- *ID*: ID que RiseData le ha otorgado al cliente.
- *Sexo*: sexo del cliente
- *Edad*: edad del cliente
- *Ingresos*: ingresos declarados por el cliente
- *Score*: puntuación que le ha otorgado RiseData al cliente en base a su comportamiento de compra.

Ejercicio 1

Carga los datos y realiza una primera exploración de la información. Remata con una visualización de la misma.

Ejercicio 2

Realiza un análisis exploratorio de los datos (EDA).

Ejercicio 3

Identifica el número de clusters más adecuado.

Ejercicio 4

Perfila y califica los clusters identificados.

Requisitos de la entrega

Debes entregar un notebook ejecutado que incluya:

- Las decisiones que vas tomando a lo largo del desarrollo del ejercicio.
- Las conclusiones a las que has llegado tras la realización del ejercicio.