

# RESUMEN EJECUTIVO

AUTOR: Diego Hernández Jiménez

Fecha: marzo 2020

## PROBLEMA ABORDADO

---

Los accidentes cerebrovasculares constituyen un grave problema de salud pública, siendo la tercera causa de muerte en nuestro país, siendo además la primera causa de muerte en mujeres. Ante esta situación se hace necesario hacer uso de los avances en estadística para conseguir pronosticar con precisión qué personas tienen más riesgo de sufrir ictus, y qué factores están implicados, para poder prevenir más eficazmente sus efectos.

## ESTUDIO REALIZADO

---

De las diversas técnicas de clasificación disponibles, se propone comparar la capacidad predictiva de tres, un modelo de **análisis cuadrático discriminante**, un **árbol de decisión** y una **máquina de vector soporte**.

- La **muestra** empleada está compuesta por 43000 personas y sus puntuaciones en diversos atributos como edad, nivel de glucosa en sangre, si se padece hipertensión...
- El **objetivo** es utilizar algunos de estos atributos para clasificar con precisión a los sujetos en la categoría de "ha sufrido ictus" y "no ha sufrido ictus" minimizando además los errores consistentes en pronosticar que no se ha sufrido ictus cuando realmente sí se ha sufrido (falsos negativos). Para lograrlo se establece un coste que penaliza los falsos negativos producidos por el modelo. Se emplea además la validación cruzada, una técnica que permite estimar la capacidad de generalización de las predicciones de los modelos a nuevas muestras.
- Para la realización de todos los análisis se utiliza el lenguaje de programación Matlab.

# SOLUCIÓN PROPUESTA

De los modelos comparados, se elige como más adecuado el **árbol de decisión**.

- Consigue clasificar correctamente un **75%** de los casos.
- De manera más concreta, pronostica adecuadamente los casos de ictus en un **82%** de las ocasiones y en un **75%** los casos de no ictus.
- Además, obtiene un valor de AUC, una medida global del rendimiento de **0,81**, siendo el máximo posible de 1.
- Los otros modelos consiguen puntuaciones no muy alejadas, pero el árbol de decisión posee la ventaja de ser fácilmente interpretable, pues el modelo consiste básicamente en reglas de decisión.

