

Nombre participante:

Base de datos (en excel)

link =

https://docs.google.com/spreadsheets/d/e/2PACX-1vTrtwaovqjbmsalswS453k4jzR1FzKvLVjkFijoxeQthQ5XMx0OnAt6_Xq0CB-7Cg/pub?output=xlsx

código Python

```
df = pd.read_excel(link)
```

1 Realizar un modelo de clasificación con árboles (max_depth = 4)

X: utilizar las 2 variables que consideramos más importante (importance_feature: Ranking)

Plotear la frontera de clasificación y determinar el score y sensibilidad (trp)

y: diagnosis

Determina cuál es la clase que tiene más datos (0,1)

2 Balancear la data utilizando el balanceo ROS (oversampling)

Utilizar las 2 variables que consideras más importantes en la pregunta 1

Plotear la frontera de clasificación y determinar el score y sensibilidad

3 Balancear la data utilizando el balanceo SMOTE

Utilizar las 2 variables que consideras más importantes en la pregunta 1

Plotear la frontera de clasificación y determinar el score y sensibilidad


4 Comparamos los 3 modelos previos y recomendar el mejor y por que?


Los entregables 1 colab por cada pregunta colgarlo en paideia


En su defecto poner títulos para separar sus respuestas

Hora máxima de entrega 12:00

Sesión 8 - 30/10 EV final

 Por favor llenar la ENCUESTA FINAL, la cual es obligatoria para asegurar la calidad de la enseñanza.

 sesion 8

 Evaluación fin

Disponibles pero no visibles en la página del curso
