Aplicación de técnicas de aprendizaje supervisado sobre datos biológicos

Objetivos

El objetivo de esta actividad es aplicar técnicas de aprendizaje supervisado para realizar la clasificación o predicción de datos biológicos, utilizando un conjunto de datos de ejemplo.

Pautas de elaboración

Para esta actividad necesitarás un ordenador con [R instalado](https://cran.r-project.org/bin/windows/base/) y los datos provistos por el profesor (data.csv y variables.csv), que consisten en un archivo llamado *data,* que contiene los datos que se van a utilizar, y un archivo llamado *variables,* que explica cada variable (los archivos están en formato csv).

Actividad

Con el conjunto de datos proporcionados, construye al menos dos modelos de aprendizaje automático utilizando métodos **supervisados**. Incluye los siguientes aspectos:

* Preparación del entorno de trabajo (instalación y carga de paquetes adecuados).
* División del conjunto de datos.
* Selección y entrenamiento del modelo.
* Cálculo de la precisión del modelo.
* Comparación entre modelos

Extensión y formato

Los resultados deberán ser entregados como un *script* de R.

Realiza los comentarios pertinentes para explicar el código.

No hay una extensión máxima del *script.*

Rúbrica

|  |  |  |  |
| --- | --- | --- | --- |
| Aplicación de técnicas de aprendizaje supervisado sobre datos biológicos | Descripción | Puntuación máxima  (puntos) | Peso  % |
| Criterio 1 | Adecuación del entorno de trabajo | 1 | 10 % |
| Criterio 2 | Preparación y manipulación de los datos | 2 | 20 % |
| Criterio 3 | Selección y entrenamiento del modelo | 4 | 40 % |
| Criterio 4 | Evaluación del modelo | 1,5 | 15 % |
| Criterio 5 | Orden y claridad del código | 1 | 10 % |
| Criterio 6 | Creatividad: se muestra un enfoque innovador o una solución creativa para resolver el problema planteado | 0,5 | 5 % |
|  |  | **10** | **100 %** |