TALLER DE CIENCIA DE DATOS

**Parte I 50 pts.**

Para este taller tenga en cuenta las siguientes variables que se tomaron de una base de datos, sobre el salario de varios individuos (hoja Taller5). Para la imputación de los datos utilice IterativeImputer.

Y: (Education) Nivel de educación del individuo.

X1: (Age) Edad del individuo.

X2: (Gender) Sexo del individuo.

X3: (YearofExperience) Años de experiencia laboral del individuo.

X4: (Job) Tipo de trabajo del individuo.

X5: (Salary) Salario en unidades monetarias (UM) del individuo.

Los Árboles de decisión permiten hacer clasificación de los individuos

Etapa 1

Estime los siguientes modelos:

|  |  |  |
| --- | --- | --- |
| # | Modelo | Métricas |
| (1) |  |  |
| (2) |  |  |
| (3) |  |  |
| (4) |  |  |
| (5) |  |  |
| (6) |  |  |

Que modelo seleccionó e interprételo y realice un pronóstico con datos inventados

**Parte II 50 pts.**

Desarrolla una actividad de análisis predictivo aplicando la metodología CRISP-DM al dataset del Titanic, con el objetivo de construir y evaluar un modelo de clasificación que permita predecir la supervivencia de los pasajeros. Para ello, deberás comprender el contexto del problema, explorar y preparar adecuadamente los datos mediante la imputación de valores faltantes y la transformación de variables, iterando sobre diferentes estrategias de imputación multivariable para comparar su impacto en el rendimiento del modelo, aplicar y evaluar diversos algoritmos de clasificación supervisada vistos en clases, utilizando métricas correspondientes, identificando las variables más influyentes en la predicción, y justificar la elección del modelo final con base en su rendimiento y valor explicativo.