

Nombre candidato:

Fecha:

Cargo: Científico de Datos

Prueba técnica

Objetivo

Cree un regresor que recomiende el tamaño de la "tripulación" para los posibles compradores de barcos.

Instrucciones

Esta prueba desee ser desarrollada por con el lenguaje de programación python. El resultado debe ser cargado en un Jupyter notebook y debe ser cargado en un repositorio de github para su entrega.

Datos

Los datos para la prueba deben ser descargados de [aquí](#).

Pasos a seguir para el desarrollo de la prueba

1. Leer el archivo
2. Mostrar las columnas
3. Calcular estadísticas básicas de los datos, interpretarlos y enunciar sus observaciones
4. Seleccionar las columnas que se consideran importantes para predecir el tamaño de la tripulación (crew)
5. Si eliminas columnas explica la razón de la decisión tomada
6. Utiliza "one-hot encoding" para las características categóricas
7. Crea conjuntos de entrenamiento y prueba (utilice el 60 % de los datos para el entrenamiento y el restante para prueba).
8. Cree un modelo de aprendizaje automático para predecir el tamaño de la tripulación.
9. Calcula el coeficiente de correlación de Pearson para el conjunto de entrenamiento y los conjuntos de datos de prueba.

10. Describa los hiperparámetros en su modelo y cómo los cambiaría para mejorar el rendimiento del modelo.
11. ¿Qué es la regularización? ¿Cuál es el parámetro de regularización en su modelo?
12. Trazar el valor del parámetro de regularización frente a la correlación de Pearson para los conjuntos de prueba y entrenamiento, y ver si su modelo tiene un problema de sesgo o de varianza

Tiempo

Esta prueba tiene un tiempo estimado de 4 horas. Luego de haber recibido la prueba tienes 3 días para hacer la entrega de la solución.