

Ejercicios Prácticos y Aplicados



Ejercicio 1

En este ejercicio vamos a importar y preprocesar los datos de 'data/heart.csv' que utilizaremos para entrenar un modelo de clasificación binaria con PySpark. Para ello, tendrás que inicializar una sesión de Spark, cargar los datos con el esquema correcto y analizar su distribución.

Es decir, debes completar la parte de importación y análisis exploratorio de los datos.

Solución: *Ejercicio_Solucion_Machine Learning con Spark.ipynb*



Ejercicio 2

En este ejercicio vas a aplicar un pre-procesamiento mínimo en el conjunto de datos del ejercicio anterior, de manera que puedas utilizarlos para entrenar un modelo.

Solución: *Ejercicio_Solucion_Machine Learning con Spark.ipynb*

Ejercicio 3

En este ejercicio vas a entrenar un modelo de clasificación binaria con la librería de machine learning de PySpark, con el conjunto de datos pre-procesados del ejercicio anterior.

Una vez entrenado el modelo, vas a tener que realizar una predicción con el conjunto de datos de test y comparar los resultados. A continuación, deberás obtener diferentes métricas de evaluación para determinar si el modelo es adecuado o no.

Solución: *Ejercicio_Solucion_Machine Learning con Spark.ipynb*