

Plan Formativo: Ciencia de Datos	Nivel de Dificultad
Módulo: Fundamentos del Big Data	Bajo / Medio
Tema: Machine learning	
Intención del aprendizaje o aprendizaje esperado:	
<ul style="list-style-type: none"> • Elabora un modelo de aprendizaje de máquina utilizando MLlib para resolver un problema de grandes volúmenes de datos 	
Ejercicios planteados	
<p>PARTE II</p> <p>Continuando con los datos del Corazón.</p> <ol style="list-style-type: none"> 1. Use Standard Scaler para poner todos los números en la misma escala. Esto toma la observación y resta la media, y luego la divide por la desviación estándar. 2. Imprima el estado actual de sus datos 3. Divida los datos en conjuntos de datos de entrenamiento y prueba. Use 50/50. 4. Cree un modelo de regresión logística y entrénalo. 	

5. Finalmente Use la función F de Spark SQL para crear una nueva columna correcta cuando ENFERMO() es igual a la predicción, lo que significa que el resultado predicho es igual a los resultados reales.

Caso

Preguntas guía

- Machine Learning
- MLlib

Recursos Bibliográficos:

Referencias

[1] Desde cero: Machine learning

<https://medium.com/datos-y-ciencia/introduccion-al-machine-learning-una-gu%C3%ADa-desde-cero-b696a2ead359>

[2] MLlib

<https://www.datahack.es/mlib-machine-learning-spark/#:~:text=MLlib%20o%20Spark%20MLlib%20es,de%20algoritmos%20de%20Machine%20Learning.>

[3] Características MLlib

<https://www.codetd.com/es/article/10369560>

[4] Aprendizaje automático:



AWAKELAB

<https://docs.microsoft.com/es-es/azure/hdinsight/spark/apache-spark-machine-learning-mllib-ipython>

[5] Ejemplo – Algoritmo

<https://www.baeldung.com/spark-mllib-machine-learning>