

Creado por:

Isabel Maniega

Iris Dataset con MLlib

Realizar el siguiente ejercicio basandonos en la explicación de Titanic dataset usando PySpark:

Usaremos el archivo iris.csv adjunto a esta actividad.

- Transformar en vector los datos numéricos usando VectorAssembler
- Transformar los datos correspondientes a la columna especies usando StringIndexer
- Calcular la predicción y la precisión del modelo empleando Decision Tree Classifier
- Calcular la predicción y la precisión del modelo empleando Gradient-boosted tree classifier
- Calcular la predicción y la precisión del modelo empleando Random Forest Classifier

En el caso de poder usar estos algoritmos razona el porqué y usa otra alternativas.

Creado por:

Isabel Maniega