Universidad Autónoma de Baja California 

Facultad de Contaduría y Administración

Maestría en Gestión de Tecnologías de la Información y la Comunicación



Materia: Big Data Clave: \_\_\_ \_\_\_\_ Grupo: \_\_\_ \_\_\_ Fecha: 22/11/2023

Docente: Dr. José Christian Romero Hernández Unidad(es) a evaluar: II Ciclo: \_\_\_\_\_ Alumno: Calificación:

Instrucciones

Desarrollar las siguientes instrucciones en Spark con el lenguaje de programación Scala, utilizando solo la documentación de la librería de Machine Learning Mllib de Spark y Google.

1. Cargar en un dataframe de la fuente de datos Iris.csv que se encuentra en https://github.com/jcromerohdz/iris, elaborar la limpieza de datos necesaria para ser procesado por el siguiente algoritmo (Importante, esta limpieza debe ser por medio de un script de Scala en Spark).

a. Utilice la librería Mllib de Spark el algoritmo de Machine Learning Multilayer Perceptron Classifier

2. ¿Cuáles son los nombres de las columnas?

3. ¿Cómo es el esquema?

4. Imprime las primeras 5 columnas.

5. Usa el método describe () para aprender más sobre los datos del DataFrame. 6. Haga la transformación pertinente para los datos categóricos los cuales serán nuestras etiquetas a clasificar.

7. Construya el modelo de clasificación y explique su arquitectura.

8. Imprima los resultados del modelo y de sus observaciones.