EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA

TECNOLÓGICO NACIONAL DE MÉXICO

TECNOLÓGICO NACIONAL DE MÉXICO

INSTITUTO TECNOLÓGICO DE TIJUANA
SUBDIRECCIÓN ACADÉMICA
Departamento de Sistemas y Computación
EXAMEN

Carrera: Ingeniería En Sistemas Computacionales/ Tecnologías de la información/ Informatíca
Materia: Datos Masivos
Unidad (es) a evaluar: Unidad 3
Catedrático: Jose Christian Romero Hernandez
Grupo:
Tipo de examen:Práctico
Firma del maestro:
Período: **Enero-Diciembre 2021**
Salón:
Fecha:
Calificación:

Alumno: _____
No. Control: _____

Instrucciones
Desarrolle las siguientes instrucciones en Spark con el lenguaje de programación Scala.

Objetivo:
El objetivo de este examen practico es tratar de agrupar los clientes de regiones específicas de un distribuidor al mayoreo. Esto en base a las ventas de algunas categorías de productos.

Las fuente de datos se encuentra en el repositorio:
https://github.com/jcromerohdz/BigData/blob/master/Spark_clustering/Wholesale%20customers%20data.csv

1. Importar una simple sesión Spark.
2. Utilice las lineas de código para minimizar errores
3. Cree una instancia de la sesión Spark
4. Importar la librería de Kmeans para el algoritmo de agrupamiento.
5. Carga el dataset de Wholesale Customers Data
6. Seleccione las siguientes columnas: Fresh, Milk, Grocery, Frozen, Detergents_Paper, Delicassen y llamar a este conjunto feature_data
7. Importar Vector Assembler y Vector
8. Crea un nuevo objeto Vector Assembler para las columnas de caracteristicas como un conjunto de entrada, recordando que no hay etiquetas
9. Utilice el objeto assembler para transformar feature_data
10. Crear un modelo Kmeans con K=3
11. Evalúe los grupos utilizando Within Set Sum of Squared Errors WSSSE e imprima los centroides.

Instrucciones de evaluación

- Tiempo de entrega 3 dias
- Al terminar poner el codigo y la explicación en la rama (branch) correspondiete de su github asi mismo realizar su explicación de la solución en su google drive.
- Finalmente defender su desarrollo en un video de 8-10 min el cual servira para dar su calificación, este video debe subirse a youtube para ser compartido por un link púbico .