Tecnológico Nacional de México

INSTITUTO TECNOLÓGICO CAMPUS TIJUANA

ING. EN SISTEMAS COMPUTACIONALES

Subdirección Académica

Departamento de Sistemas y Computación

BDD-1704 TI9A - 6:00pm-7:00pm

ASIGNATURA:

Datos Masivos

SEMESTRE:

Septiembre- Enero 2020

Tarea:

Explicación de lo que yo entendí del código:

NaiveBayesExample.scala

MAESTRO:

JOSE CHRISTIAN ROMERO HERNANDEZ

09/12/2020

//Importaciones las librerías naivebayes y MulticlassClassificationEvaluator para hacer los calculos.

import org.apache.spark.ml.classification.NaiveBayes

import org.apache.spark.ml.evaluation.MulticlassClassificationEvaluator

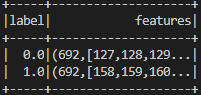
// creamos un datagrame llamado data y le cargamos los datos guardados en formato LIBSVM.

val data = spark.read.format("libsvm").load("C:/Users/DELL/Desktop/ejercicioequipo8NaiveBayes/sample\_libsvm\_data.txt")

//vemos dos elementos del dataframe

data.show(2)

//Resultado



//Separamos los datos en sets de entrenamiento y de prueba, 30% reservado para las pruebas y 70 al testeo

val Array(trainingData, testData) = data.randomSplit(Array(0.7, 0.3), seed = 1234L)

// Entrenamos al modelo Naive Bayes.

val model = new NaiveBayes().fit(trainingData)

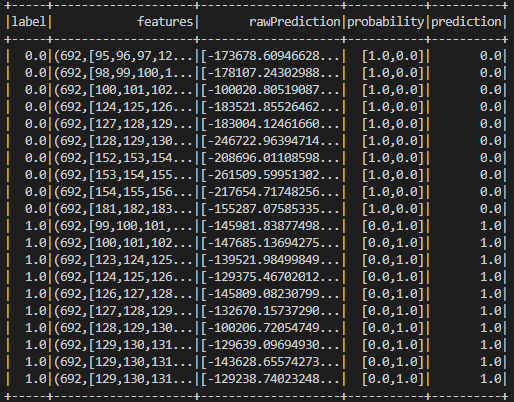
/\*creamos una variable llamada predictions para hacer una prediccion

utilizando al modelo y transformando las datos de testeo.\*/

val predictions = model.transform(testData)

predictions.show()

//Resultado



/\*creamos una variable evaluador y la igualamos a funcion MulticlassClassificationEvaluator

seleccionos prediccion, etiqueta de cierto y calculamos el error de prueba\*/

val evaluator = new MulticlassClassificationEvaluator().setLabelCol("label").setPredictionCol("prediction").setMetricName("accuracy")

//creamos la variable accuracy para evaluar las predicciones

val accuracy = evaluator.evaluate(predictions)

//imprimimos la prueba de accuracy

println(s"Test set accuracy = $accuracy")

//Resultado

