Big Data Using SparkR for predictions (DePaul University Assignment #4 subset)

Nora Vega: Used Microsoft Azure to run these commands

# Create a SQL context using SparkR context.

sc <- sparkR.init()

sqlContext <- sparkRSQL.init(sc)

# Create a Data Frame called "cars" from the mtcars dataset

cars <- createDataFrame(sqlContext, mtcars)

# Create a GLM of the Gaussian family of models, using the formula that

# has "mpg" as the response variable and "hp" and "cyl" as the predictors.

model <- SparkR::glm(mpg ~ hp + cyl, data = cars, family = "gaussian")

# check the data for this model in an easy-to-read manner using the **summary** function.  
 SparkR::summary(model)

# Create predictions based on the model created.   
predictions <- SparkR::predict(model, newData = cars)

SparkR::head(select(predictions, "mpg", "prediction"))

