**Escuela de Estudios de Postgrado**

**Facultad de Ingeniería**

**Universidad San Carlos de Guatemala**

**Maestría en Ingeniería para la Industria con**

**Especialización en Ciencias de la Computación.**

**Minería de Datos**

**Ing. Ms. Kevin Lajpop**

**Tarea 5**

**ARBOL DE PREDICCIÓN**

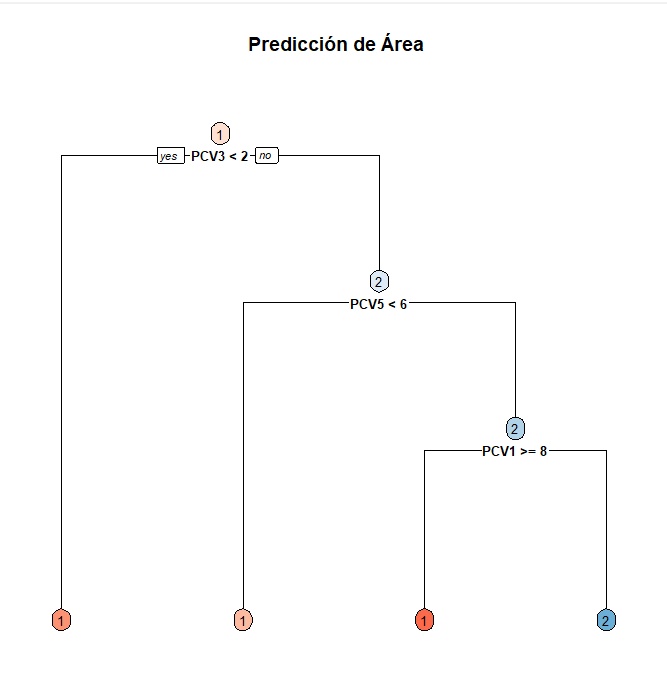
[**https://github.com/obduliogonzalez/tarea\_5**](https://github.com/obduliogonzalez/tarea_5)

**Néstor Obdulio González López**

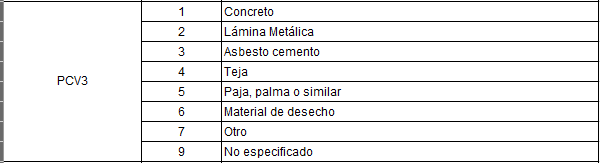
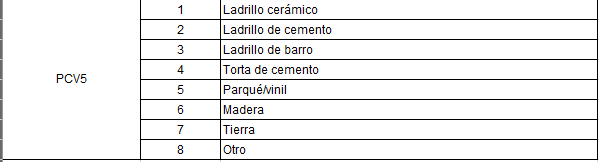
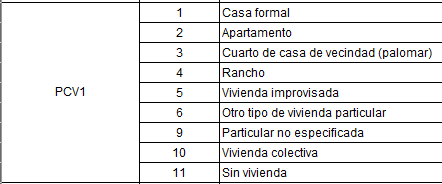
**Carné: 999013150**

Tarea 5

Árbol de Predicción de Área, archivo VIVIENDA\_BDP.csv



Simbología



Sabiendo la simbología anterior

La primera decisión en el árbol representa que, si el material de la construcción es menor a dos, ya sea ladrillo cerámico o ladrillo de cemento el área al que pertenece la vivienda tiene a ser Urbana

La segunda decisión no indica que en el caso de que en la primera decisión PCV3 fue mayor a dos hay otro factor de PCI5 a tomar en cuenta, si PCV5 es menor a 6 y PVC3 es mayor, entonces el área al que pertenece la vivienda es urbana si no pues sigue a la siguiente decisión

La última decisión nos indica que, si PCV1 es mayor o igual a 8 pues que todavía es área urbana, pero si no definitivamente es un área rural

Con el árbol podemos ir identificando que la tendencia que van teniendo las viviendas a tender hacia un área rural, según el material con lo que se trabaja en ellas.  
  
código

*# Configura el espejo de CRAN por defecto*

options(repos = c(CRAN = "https://cran.rstudio.com/"))

*# Instala los paquetes necesarios si no están instalados*

if (!require("rpart")) install.packages("rpart")

if (!require("rpart.plot")) install.packages("rpart.plot")

*# Carga las librerías*

library(rpart)

library(rpart.plot)

*# Asegúrate de que la ruta al archivo CSV es correcta*

vivienda <- read.csv("c:\\Users\\diego\\Desktop\\II\\Tarea5\\VIVIENDA\_BDP.csv", sep=',')

*#Construye el modelo de árbol de decisión*

arbol <- rpart(AREA ~ PCV1+

                      PCV2+

                      PCV3+

                      PCV4+

                      PCV5,

                      data=vivienda, method="class")

rpart.plot(arbol, type=2, extra=0, under=TRUE, fallen.leaves=TRUE, box.palette="RdBu", main=" Predicción de Área", cex=0.80)