

Universidad de San Carlos de Guatemala

Facultad de Ingeniería

Escuela de Estudios de Postgrado

Maestría en ingeniería para la industria con especialización
en ciencia de la computación

Proyecto fase 2

Mynor René Ruiz Guerra

Guatemala, diciembre 2024

ÍNDICE GENERAL

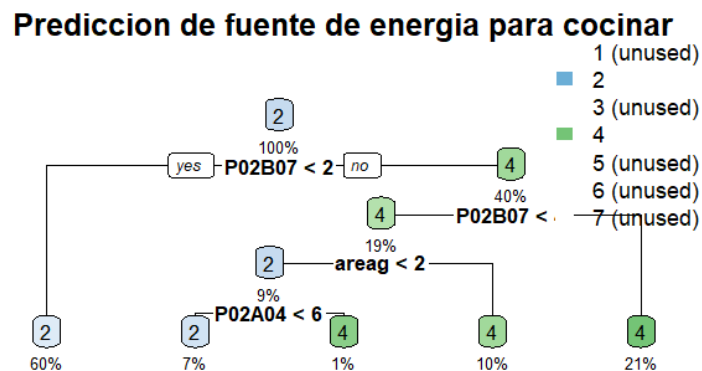
Contenido

Resultados	3
Predicción por medio de árboles de decisión	3
Árbol 1	3
Árbol 2	6
Árbol 3	8
Árbol 4	11
Bosques aleatorios	13
Bosque 1	13
Bosque 2	16
Redes Neuronales	19
Modelo 1	19
Modelo 2	21
Problema identificado	24
Propuestas	25
Promoción de cocinas eficientes y seguras	25
Fortalecimiento de infraestructura sanitaria	25
Educación sobre tratamiento de agua	25
Reducción de costos de agua purificada	26
Repositorio	27
Bibliografía	27

Resultados

Predicción por medio de árboles de decisión

Árbol 1



Interpretación del árbol de decisión

Nodo Raíz

P02B07 (Tipo de servicio sanitario):

Si P02B07 es menor a 2 (es decir, Inodoro de red de drenaje), el árbol predice que el 60% de los hogares utilizan propano (P02B05 = 2).

Si P02B07 es mayor o igual a 2 (es decir, Inodoro fosa séptica, Excusado, Letrina o No tiene), se realiza una división adicional.

Subdivisión cuando $P02B07 \geq 2$:

Si P02B07 es mayor o igual a 2:

P02B07 = Inodoro fosa séptica, Excusado, Letrina, o No tiene:

Se evalúa areag (Área geográfica).

Subdivisión cuando $areag < 2$ (Área rural):

P02A04 (Material predominante del piso):

Si P02A04 es menor a 6 (es decir, el material predominante es Cerámico, Ladrillo de Cemento, Ladrillo de Barro, Torta cemento o Granito):

Predicción: 9% de los hogares usan propano (P02B05 = 2).

Si P02A04 es mayor o igual a 6 (es decir, el material predominante es Parqué, Madera o Tierra):

Predicción: 1% de los hogares usan propano (P02B05 = 2).

Subdivisión cuando areag ≥ 2 (Área urbana):

Si areag es urbano (valor 1):

Predicción: 40% de los hogares usan Leña (P02B05 = 4).

Nodos Terminales:

En los nodos terminales, los números indican la predicción de la fuente de energía para cocinar:

Nodo terminal 2 (Propano):

60% de los hogares usan propano si P02B07 es Inodoro de red de drenaje.

9% de los hogares usan propano si el material del piso es Cerámico, Ladrillo de Cemento, Ladrillo de Barro, Torta cemento o Granito en áreas rurales.

1% de los hogares usan propano si el material del piso es Parqué, Madera o Tierra en áreas rurales.

Nodo terminal 4 (Leña):

40% de los hogares usan Leña si el tipo de sanitario es Inodoro fosa séptica, Excusado, Letrina o No tiene y el área es urbana.

Primera predicción

```
29 hogar1 <- data.frame(  
30   areag=c(1),# Area urbana  
31   P02A04=c(2),# Ladrillo de Cemento  
32   P02B07=c(5),# No tiene  
33   P01B04=c(1)# Reside 1 persona habitualmente  
34 )  
35  
36 result <- predict(arbol1,hogar1, type="class")  
37  
38 result  
39  
40  
39:1 (Top Level) ↕  
  
Console Terminal Background Jobs  
R 4.2.1 - /  
+ main = "Prediccion de fuente de energia para cocinar",cex = 1)  
> rpart.plot(arbol1, type=2,extra=100,under=TRUE, fallen.leaves = TRUE, box.p  
+ main = "Prediccion de fuente de energia para cocinar",cex = 0.7)  
> hogar1 <- data.frame(  
+   areag=c(1),# Area urbana  
+   P02A04=c(2),# Ladrillo de Cemento  
+   P02B07=c(5),# No tiene  
+   P01B04=c(1)# Reside 1 persona habitualmente  
+ )  
> result <- predict(arbol1,hogar1, type="class")  
> result  
1  
4  
Levels: 1 2 3 4 5 6 7
```

Para un hogar que pertenece al área urbana, que el material predominante del piso es ladrillo de cemento, que habitualmente reside 1 persona y que no tiene ningún tipo de servicio sanitario, la predicción muestra que la fuente de energía para cocinar es leña.

Segunda predicción

```
37 # Predicción 2 - Arbol 1
38 hogar2 <- data.frame(
39   areag=c(1),# Area urbana
40   P02A04=c(7),# Madera
41   P02B07=c(2),# Inodoro fosa séptica
42   P01B04=c(3)# Reside 3 persona habitualmente
43 )
44
45 result <- predict(arbol1,hogar2, type="class")
46 result
47
```

48:1 (Top Level) ▾

Console Terminal x Background Jobs x

R 4.2.1 · ~/

```
4
Levels: 1 2 3 4 5 6 7
> # Predicción 2 - Arbol 1
> hogar2 <- data.frame(
+   areag=c(1),# Area urbana
+   P02A04=c(7),# Madera
+   P02B07=c(2),# Inodoro fosa séptica
+   P01B04=c(3)# Reside 3 persona habitualmente
+ )
> result <- predict(arbol1,hogar2, type="class")
> result
1
4
Levels: 1 2 3 4 5 6 7
```

Para un hogar que pertenece al área urbana, que el material predominante del piso es madera, que habitualmente reside 3 persona y que tipo de servicio sanitario es inodoro de fosa séptica, la predicción muestra que la fuente de energía para cocinar es leña.

Tercera predicción

```
48 # Predicción 3 - Arbol 1
49 hogar3 <- data.frame(
50   areag=c(3),# Area rural
51   P02A04=c(2),# Ladrillo de Cemento
52   P02B07=c(1),# Inodoro red drenaje
53   P01B04=c(3)# Reside 3 persona habitualmente
54 )
55
56 result <- predict(arbol1,hogar3, type="class")
57 result
58
59
```

60:1 (Top Level) ▾

Console Terminal x Background Jobs x

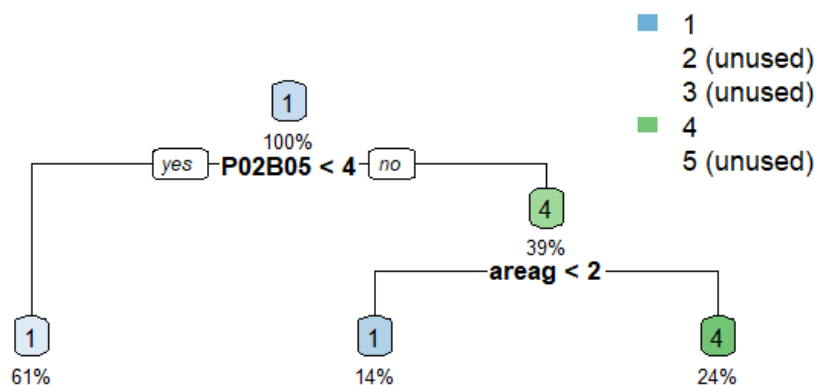
R 4.2.1 · ~/

```
Levels: 1 2 3 4 5 6 7
> # Predicción 3 - Arbol 1
> hogar3 <- data.frame(
+   areag=c(3),# Area rural
+   P02A04=c(2),# Ladrillo de Cemento
+   P02B07=c(1),# Inodoro red drenaje
+   P01B04=c(3)# Reside 3 persona habitualmente
+ )
> result <- predict(arbol1,hogar3, type="class")
> result
1
2
Levels: 1 2 3 4 5 6 7
```

Para un hogar que pertenece al área rural, que el material predominante del piso es ladrillo de cemento, que habitualmente reside 3 persona y que tipo de servicio sanitario es inodoro de red de drenaje, la predicción muestra que la fuente de energía para cocinar es Propano.

Árbol 2

Predicción de tipo de servicio sanitario



Interpretación del árbol de decisión

Nodo Raíz:

P02B05 (Fuente de energía para cocinar):

Si P02B05 es menor que 4 (es decir, la fuente de energía es Electricidad, Propano o Kerosene):

Predicción: El 61% de los hogares tienen Inodoro de red de drenaje (P02B07 = 1).

Si P02B05 es mayor o igual a 4 (es decir, la fuente de energía es Leña, Carbón, No cocina o Otra), el árbol realiza una división adicional.

Subdivisión cuando $P02B05 \geq 4$ (Leña, Carbón, No cocina, Otro):

areag (Área geográfica):

Si areag es menor que 2 (es decir, rural):

Predicción: 39% de los hogares tienen Letrina (P02B07 = 4).

Si areag es mayor o igual a 2 (es decir, urbano):

Predicción: 24% de los hogares tienen Letrina (P02B07 = 4).

Primera predicción

```
72 # Predicción 1 - Arbol 2
73 hogar1 <- data.frame(
74   areag=c(1),# Area urbana
75   P02B04=c(2),# Hervida
76   P02B05=c(5)# Carbón
77 )
78
79 result <- predict(arbol2,hogar1, type="class")
80 result
81
```

75:24 (Top Level) ↕

Console	Terminal x	Background Jobs x
R 4.2.1 · ~/		
Error in eval(predvars, data, env) : object 'P02B04' not found		
> # Predicción 1 - Arbol 2		
> hogar1 <- data.frame(
+ areag=c(1),# Area urbana		
+ P02B04=c(2),# Ladrillo de Cemento		
+ P02B05=c(5)# No tiene		
+)		
> result <- predict(arbol2,hogar1, type="class")		
> result		
1		
1		
Levels: 1 2 3 4 5		

Para un hogar que pertenece al área urbana, que el tratamiento al agua para beber es hervido y que tiene una fuente de energía para cocinar el carbón, la predicción muestra que el tipo de servicio de sanitario es inodoro de red de drenaje.

Segunda predicción

```
81 # Predicción 2 - Arbol 2
82 hogar2 <- data.frame(
83   areag=c(2),# Area rural
84   P02B04=c(4),# Clorada
85   P02B05=c(3)# Kerocene
86 )
87
88 result <- predict(arbol2,hogar2, type="class")
89 result
90
```

90:1 (Top Level) ↕

Console	Terminal x	Background Jobs x
R 4.2.1 · ~/		
Levels: 1 2 3 4 5		
> # Predicción 2 - Arbol 2		
> hogar2 <- data.frame(
+ areag=c(2),# Area rural		
+ P02B04=c(4),# Clorada		
+ P02B05=c(3)# Kerocene		
+)		
> result <- predict(arbol2,hogar2, type="class")		
> result		
1		
1		
Levels: 1 2 3 4 5		

Para un hogar que pertenece al área rural, que el tratamiento al agua para beber es clorado y que tiene una fuente de energía para cocinar el kerocene, la predicción muestra que el tipo de servicio de sanitario es inodoro de red de drenaje.

Tercera predicción

```

90 # Predicción 3 - Arbol 2
91 hogar3 <- data.frame(
92   areag=c(2),# Area rural
93   P02B04=c(3),# Filtrada
94   P02B05=c(4)# Leña
95 )
96
97 result <- predict(arbol2,hogar3, type="class")
98 result
99 |

```

99:1 (Top Level) ↕

Console Terminal Background Jobs

R 4.2.1 · ~/

Levels: 1 2 3 4 5

```

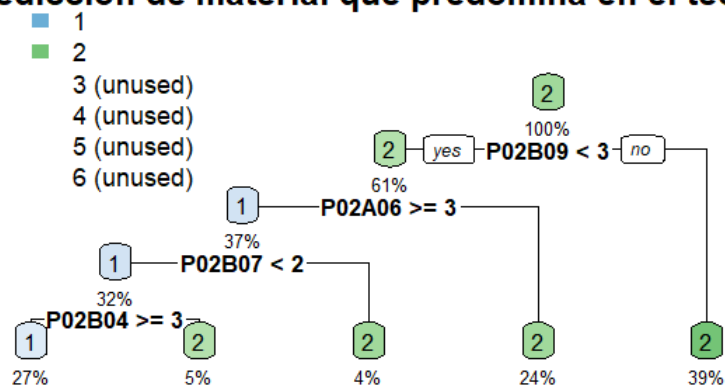
> # Predicción 3 - Arbol 2
> hogar3 <- data.frame(
+   areag=c(2),# Area rural
+   P02B04=c(3),# Filtrada
+   P02B05=c(4)# Leña
+ )
> result <- predict(arbol2,hogar3, type="class")
> result
1
4
Levels: 1 2 3 4 5

```

Para un hogar que pertenece al área rural, que el tratamiento al agua para beber es filtrado y que tiene una fuente de energía para cocinar la leña, la predicción muestra que el tipo de servicio de sanitario es letrina.

Árbol 3

Predicción de material que predomina en el techo



Interpretación del árbol de decisión

Nodo Raíz:

P02B09 (Forma para eliminar la basura):

Si P02B09 es menor a 3 (es decir, Servicio municipal o Servicio privado):

Se evalúa P02A06 (Total de cuartos en la vivienda).

Si P02B09 es mayor o igual a 3 (es decir, Se quema, Se entierra, Se tira, Aboneras, o Otro), el árbol predice P02A03 = 2 (material predominante del techo = Lámina).

Subdivisión cuando P02B09 < 3 (Servicio municipal o Servicio privado):

P02A06 (Total de cuartos en la vivienda):

Si P02A06 es mayor o igual a 3 (es decir, el hogar tiene 3 o más cuartos):

Predicción: 61% de los hogares tienen material predominante del techo = Concreto (P02A03 = 1).

Si P02A06 es menor que 3 (es decir, el hogar tiene menos de 3 cuartos):

Se evalúa P02B07 (Tipo de servicio sanitario).

Subdivisión cuando P02A06 < 3:

P02B07 (Tipo de servicio sanitario):

Si P02B07 es menor que 2 (es decir, Inodoro red de drenaje):

Predicción: 32% de los hogares tienen material predominante del techo = Concreto (P02A03 = 1).

Si P02B07 es mayor o igual a 2 (es decir, Inodoro fosa séptica, Excusado, Letrina, o No tiene):

Predicción: 37% de los hogares tienen material predominante del techo = Duralita (P02A03 = 3).

Primera predicción

```
> hogar1 <- data.frame(  
+   P02B07=c(3),# Excusado  
+   P02A06=c(2),# Cantidad de cuartos 2  
+   P02B04=c(2),# Hervida  
+   P02B09=c(1)# Servicio municipal  
+ )  
> result <- predict(arbol3,hogar1, type="class")  
> result  
1  
2  
Levels: 1 2 3 4 5 6
```

El árbol de decisión predice que el material predominante en el techo es Lámina (P02A03 = 2) para el hogar1, que tiene 2 cuartos (P02A06 = 2), utiliza agua hervida para beber (P02B04 = 2) y tiene un servicio sanitario tipo Excusado (P02B07 = 3) y Servicio municipal para eliminar la basura (P02B09 = 1). El árbol sigue una ruta que, dado el número de cuartos y el tipo de servicio sanitario, lleva a la predicción de Lámina como material del techo. Esta es una opción más económica comparada con otros materiales como Concreto.

Segunda predicción

```
> hogar1 <- data.frame(  
+   P02B07=c(1),# Inodoro red drenaje  
+   P02A06=c(1),# Cantidad de cuartos 1  
+   P02B04=c(5),# Compran purificada  
+   P02B09=c(4)# Se entierra  
+ )  
> result <- predict(arbol3,hogar1, type="class")  
> result  
1  
2  
Levels: 1 2 3 4 5 6
```

El árbol evalúa que el hogar tiene 1 cuarto y utiliza agua purificada (P02B04 = 5), y el tipo de servicio sanitario es Inodoro de red de drenaje (P02B07 = 1). Además, el hogar utiliza una forma de eliminar la basura que es Se entierra (P02B09 = 4), lo cual lleva a la predicción de que el material predominante en el techo es Lámina (P02A03 = 2). Esto indica que el hogar probablemente tiene un techo de material económico o menos duradero comparado con materiales como Concreto.

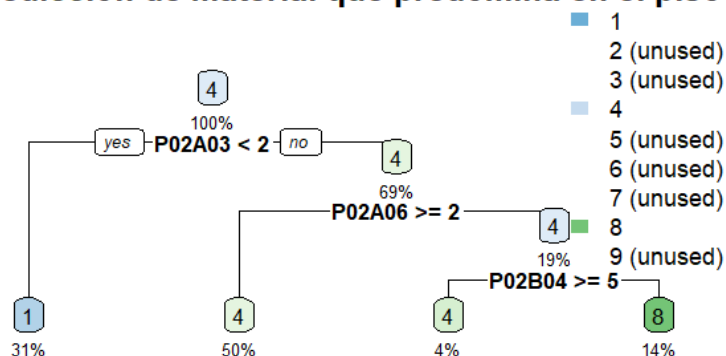
Tercera predicción

```
> hogar1 <- data.frame(  
+   P02B07=c(4),# Letrina  
+   P02A06=c(3),# Cantidad de cuartos 3  
+   P02B04=c(4),# Clorada  
+   P02B09=c(3)# Se quema  
+ )  
> result <- predict(arbol3,hogar1, type="class")  
> result  
1  
2  
Levels: 1 2 3 4 5 6
```

El árbol de decisión evalúa que el hogar tiene 3 cuartos (P02A06 = 3), usa agua clorada para beber (P02B04 = 4), y utiliza la forma de eliminar la basura "Se quema" (P02B09 = 3). Además, el tipo de servicio sanitario es Letrina (P02B07 = 4). A partir de estas características, el árbol predice que el material predominante en el techo es Lámina (P02A03 = 2), lo que indica que el hogar probablemente tiene un techo de material económico o menos duradero en comparación con otros materiales como Concreto.

Árbol 4

Predicción de material que predomina en el piso



Interpretación del árbol de decisión

Nodo Raíz:

P02A03 (Material predominante del techo):

Si el material del techo es menor a 2 (es decir, el techo es Lámina o Duralita):

Predicción: El 50% de los hogares tienen material predominante en el piso = Teja (P02A03 = 4).

Si el material del techo es mayor o igual a 2 (es decir, el techo es Concreto, Teja, Paja o Palma o Otro), el árbol realiza una subdivisión adicional.

Subdivisión cuando P02A03 >= 2 (Material del techo diferente a Lámina o Duralita):

P02A06 (Total de cuartos en la vivienda):

Si el hogar tiene 2 o más cuartos (P02A06 >= 2):

Predicción: El 69% de los hogares tienen material predominante en el piso = Teja (P02A03 = 4).

Si el hogar tiene menos de 2 cuartos (P02A06 < 2), se evalúa el tratamiento al agua.

Subdivisión cuando P02A06 < 2 (Hogar con menos de 2 cuartos):

P02B04 (Tratamiento al agua para beber):

Si el hogar compra agua purificada (P02B04 = 5):

Predicción: El 14% de los hogares tienen material predominante en el piso = Paja o palma (P02A03 = 5).

Primera predicción

```
> # Predicción 1 - Arbol 4
> hogar1 <- data.frame(
+   P02A06=c(4),# Cantidad de cuartos 4
+   P02B04=c(1),# Ninguno
+   P02A03=c(2)# Lámina
+ )
> result <- predict(arbol4,hogar1, type="class")
> result
1
4
Levels: 1 2 3 4 5 6 7 8 9
```

El árbol evalúa que el hogar tiene 4 o más cuartos ($P02A06 = 4$), utiliza agua sin tratar ($P02B04 = 1$), y el material predominante en el techo es Lámina ($P02A03 = 2$). Según el árbol, con estas características, la predicción es que el material predominante en el piso es Teja ($P02A03 = 4$).

Segunda predicción

```
> # Predicción 2 - Arbol 4
> hogar1 <- data.frame(
+   P02A06=c(3),# Cantidad de cuartos 3
+   P02B04=c(3),# Filtrada
+   P02A03=c(4)# Teja
+ )
> result <- predict(arbol4,hogar1, type="class")
> result
1
4
Levels: 1 2 3 4 5 6 7 8 9
```

El árbol evalúa que el hogar tiene 3 o más cuartos ($P02A06 = 3$), utiliza agua filtrada ($P02B04 = 3$) para beber, y el material predominante en el techo es Teja ($P02A03 = 4$). Según el árbol, con estas características, la predicción es que el material predominante en el piso es también Teja ($P02A03 = 4$).

Tercera predicción

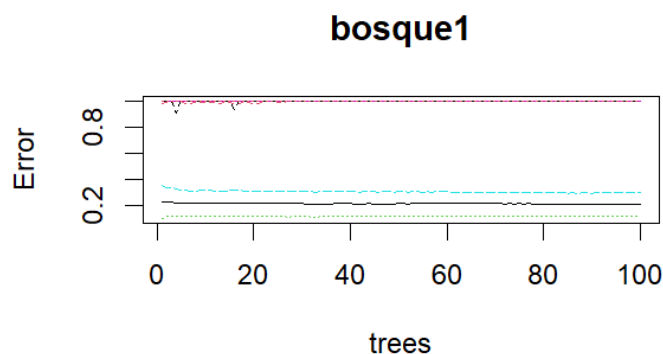
```
> # Predicción 3 - Arbol 4
> hogar1 <- data.frame(
+   P02A06=c(1),# Cantidad de cuartos 1
+   P02B04=c(5),# Compran purificada
+   P02A03=c(1)# Concreto
+ )
> result <- predict(arbol4,hogar1, type="class")
> result
1
1
Levels: 1 2 3 4 5 6 7 8 9
```

El árbol evalúa que el hogar tiene 1 cuarto ($P02A06 = 1$), utiliza agua purificada para beber ($P02B04 = 5$), y el material predominante en el techo es Concreto ($P02A03 = 1$). Según el árbol, con estas características, la predicción es que el material predominante en el piso es Concreto ($P02A03 = 1$).

Bosques aleatorios

Bosque 1

Interpretación del bosque aleatorio



Convergencia

El modelo alcanza un nivel estable de error con aproximadamente 30 a 40 árboles, lo que significa que agregar más árboles después de este punto no mejora significativamente el desempeño del modelo.

Desempeño global

La línea negra baja y estable sugiere que el modelo tiene un buen desempeño global para predecir la fuente de energía para cocinar (P02B05).

Primera predicción

```
> # Crear el nuevo dato para la predicción
> dato_nuevo <- data.frame(areag = 1, # Area Urbana
+                           P02A04 = 2, # Ladrillo de Cemento
+                           P02B07 = 5, # No tiene
+                           P01B04 = 4, # Reside 4 persona habitualmente
+                           P02B04 = 2 # Hervida
+ )
> # Realizar la predicción sobre el nuevo dato
> prediccion <- predict(bosque, dato_nuevo)
> # Mostrar la predicción
> prediccion
1
2
Levels: 1 2 3 4 5 6 7
```

Para un hogar que:

Está ubicado en un área urbana, Tiene un piso de ladrillo de cemento, No cuenta con servicio sanitario, Aloja habitualmente a 4 personas, y Hierve el agua para beber.

El modelo considera que la fuente de energía predominante para cocinar es Gas. Esto podría deberse a que las áreas urbanas tienden a tener mejor acceso a gas licuado (GLP) o gas natural, y el perfil socioeconómico del hogar con estas características probablemente coincide con aquellos que usan gas como fuente de energía principal.

Segunda predicción

```
> # Predicción 2 - bosque 1
> # Crear el nuevo dato para la predicción
> dato_nuevo2 <- data.frame(areag = 2, # Area rural
+                           P02A04 = 3, # Ladrillo de Barro
+                           P02B07 = 4, # Letrina
+                           P01B04 = 2, # Residen 2 persona habitualmente
+                           P02B04 = 4 # Clorada
+ )
> # Realizar la predicción sobre el nuevo dato
> prediccion <- predict(bosque1, dato_nuevo2)
> prediccion
1
4
Levels: 1 2 3 4 5 6 7
```

Para un hogar que:

Está ubicado en un área rural

El hogar se encuentra en una zona rural, donde los servicios y acceso a recursos pueden ser más limitados que en áreas urbanas

Tiene un piso de ladrillo de barro

Este tipo de material suele ser utilizado en construcciones más tradicionales o de menor costo en áreas rurales.

Cuenta con letrina como servicio sanitario

Este es un tipo básico de servicio sanitario, común en zonas rurales.

Aloja habitualmente a 2 personas

El hogar tiene una población pequeña, lo que podría influir en las necesidades y hábitos de consumo.

Trata el agua para beber con cloración

El uso de cloro para el agua sugiere un esfuerzo por mantener estándares básicos de salubridad en el hogar.

El modelo predice que la fuente de energía predominante para cocinar en este hogar es Leña. Esto podría deberse a que, en áreas rurales, los hogares con pisos de ladrillo de barro y servicios básicos como letrinas tienden a depender más de recursos naturales disponibles localmente, como la leña, para cocinar.

Tercera predicción

```
> # Crear el nuevo dato para la predicción
> dato_nuevo3 <- data.frame(areag = 2, # Area Urbana
+                             P02A04 = 8, # Tierra
+                             P02B07 = 1, # Inodoro red drenaje
+                             P01B04 = 1, # Reside 1 persona habitualmente
+                             P02B04 = 1 # Ninguno
+ )
> # Realizar la predicción sobre el nuevo dato
> prediccion <- predict(bosque1, dato_nuevo3)
> prediccion
1
4
Levels: 1 2 3 4 5 6 7
```

Para un hogar que:

Está ubicado en un área urbana

Aunque se encuentra en una zona urbana, el hogar podría estar en un sector marginado o de bajos recursos.

Tiene un piso de tierra

Este tipo de piso sugiere condiciones de vivienda precarias, lo que puede ser inusual para áreas urbanas.

Cuenta con inodoro conectado a red de drenaje como servicio sanitario

A pesar de tener un piso de tierra, el hogar dispone de un sistema sanitario moderno, lo que indica cierta infraestructura disponible.

Aloja habitualmente a 1 persona

El hogar tiene un solo residente, lo que podría influir en sus hábitos de consumo y necesidades energéticas.

No trata el agua para beber

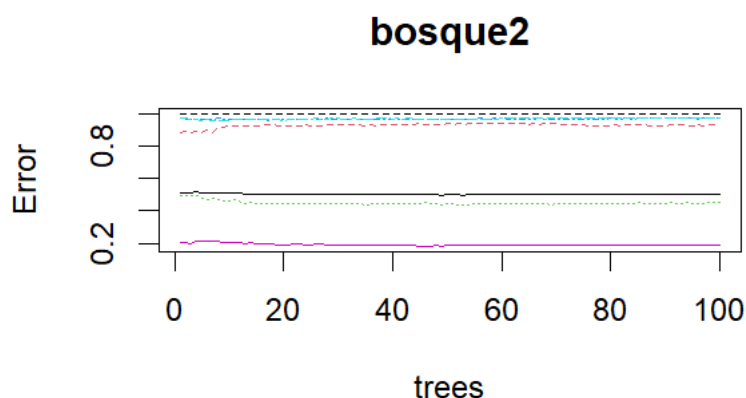
La falta de tratamiento de agua podría estar relacionada con la falta de acceso o conciencia sobre la calidad del agua potable.

El modelo predice que la fuente de energía predominante para cocinar en este hogar es Leña.

A pesar de estar en un área urbana y contar con un inodoro conectado a drenaje, la presencia de un piso de tierra y la falta de tratamiento de agua para beber sugieren que este hogar tiene características asociadas a niveles socioeconómicos bajos. En estos casos, el uso de leña para cocinar podría ser una opción económica y accesible, incluso en áreas urbanas.

Bosque 2

Interpretación del bosque aleatorio



La gráfica evidencia que el modelo de bosque aleatorio es efectivo para predecir el tratamiento del agua para beber (P02B04), logrando un error global bajo y estable con un número moderado de árboles.

El error global de 0.2 es bajo y demuestra que el modelo tiene un buen desempeño en términos generales.

Si una clase tiene un error de 0.8, significa que el modelo solo clasifica correctamente el 20% de las observaciones para esa clase específica.

Primera predicción

```
> # Predicción 1 - bosque 2
> # Crear el nuevo dato para la predicción
> dato_nuevo1 <- data.frame(P02A04 = 2, # Ladrillo de Cemento
+                           P02B07 = 5, # No tiene
+                           P01B04 = 4, # Reside 4 persona habitualmente
+                           P02B05 = 2  # Propano
+ )
> # Realizar la predicción sobre el nuevo dato
> prediccion <- predict(bosque2, dato_nuevo1)
> prediccion
1
5
Levels: 1 2 3 4 5 6
```

Para un hogar que:

Tiene un piso de ladrillo de cemento (P02A04 = 2)

Esto indica que el hogar tiene un nivel de infraestructura intermedio, dado que el ladrillo de cemento es común en construcciones formales.

No cuenta con servicio sanitario (P02B07 = 5)

La ausencia de servicio sanitario sugiere condiciones de vivienda limitadas en cuanto a infraestructura básica.

Aloja habitualmente a 4 personas (P01B04 = 4)

Este hogar tiene una familia de tamaño mediano, lo que podría influir en las necesidades de agua potable y tratamiento.

Utiliza gas propano como fuente de energía para cocinar (P02B05 = 2)

El uso de gas propano como fuente de energía indica acceso a combustibles más limpios y modernos, aunque podría ser limitado dependiendo de la región.

El modelo predice que el tratamiento del agua para beber para este hogar es Filtración.

Segunda predicción

```
> # Predicción 2 - bosque 2
> # Crear el nuevo dato para la predicción
> dato_nuevo2 <- data.frame(P02A04 = 3, # Ladrillo de Barro
+                           P02B07 = 4, # Letrina
+                           P01B04 = 2, # Residen 2 persona habitualmente
+                           P02B05 = 4  # Leña
+ )
> # Realizar la predicción sobre el nuevo dato
> prediccion <- predict(bosque2, dato_nuevo2)
> prediccion
1
2
Levels: 1 2 3 4 5 6
```

Para un hogar que:

Tiene un piso de ladrillo de barro (P02A04 = 3)

Este tipo de piso refleja una vivienda con infraestructura básica, común en zonas rurales o de menor nivel socioeconómico.

Cuenta con letrina como servicio sanitario (P02B07 = 4)

La letrina es una solución sanitaria básica, especialmente en áreas con infraestructura limitada.

Aloja habitualmente a 2 personas (P01B04 = 2)

Este hogar es pequeño, probablemente ocupado por una pareja o una familia pequeña.

Utiliza leña como fuente de energía para cocinar (P02B05 = 4)

El uso de leña sugiere acceso limitado a fuentes de energía modernas y es común en áreas rurales donde los recursos naturales están disponibles.

El modelo predice que el tratamiento del agua para beber para este hogar es Hervir el agua. Hogares con infraestructura básica, como pisos de ladrillo de barro y letrinas, tienden a optar por métodos de tratamiento accesibles y económicos como hervir el agua.

El uso de leña como fuente de energía facilita el proceso de hervir el agua, ya que el combustible está fácilmente disponible.

Tercera predicción

```
Levels: 1 2 3 4 5 6
> # Predicción 3 - bosque 2
> # Crear el nuevo dato para la predicción
> dato_nuevo3 <- data.frame(P02A04 = 8, # Tierra
+                           P02B07 = 1, # Inodoro red drenaje
+                           P01B04 = 1, # Reside 1 persona habitualmente
+                           P02B05 = 1  # Electricidad
+ )
> # Realizar la predicción sobre el nuevo dato
> prediccion <- predict(bosque2, dato_nuevo3)
> prediccion
1
1
Levels: 1 2 3 4 5 6
```

Para un hogar que:

Tiene un piso de tierra (P02A04 = 8)

Esto indica condiciones de vivienda precarias, ya que el piso de tierra es característico de hogares con infraestructura muy básica.

Cuenta con inodoro conectado a red de drenaje (P02B07 = 1)

La disponibilidad de un servicio sanitario moderno contrasta con la precariedad del piso de tierra, lo que sugiere que el hogar podría estar en un sector marginado dentro de una zona urbana.

Aloja habitualmente a 1 persona (P01B04 = 1)

Este hogar tiene un solo ocupante, probablemente una persona con recursos económicos limitados que vive sola.

Utiliza electricidad como fuente de energía para cocinar (P02B05 = 1)

El uso de electricidad para cocinar sugiere acceso a energía moderna, lo cual es un indicador positivo de infraestructura.

El modelo predice que el tratamiento del agua para beber para este hogar es Ninguno.

Con un solo residente, podría no haber suficiente motivación o recursos para implementar métodos de tratamiento de agua. La presencia de un inodoro con red de drenaje sugiere que el hogar podría estar conectado a un sistema de agua municipal, lo que disminuye la percepción de necesidad de tratar el agua.

Redes Neuronales

Modelo 1

```

v Definición de variables dependientes e independientes

# Material predominante del piso
# Personas que reside habitualmente
# Tratamiento al agua para beber
x = data[['P02A04', 'P01B04', 'P02B04']]
y = data['P02A07']
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=42)
y_train = y_train - 1
y_test = y_test - 1

v Diseño de la red neuronal

model = Sequential()
model.add(Dense(15, input_dim=3, activation='relu'))
model.add(Dense(8, activation='relu'))
model.add(Dense(5, activation='softmax'))

```

El modelo predice la tenencia de la propiedad. Recibe 3 características como entrada: P02A04, P01B04 y P02B04. Utiliza una red neuronal con 1 capa oculta y una capa de salida. La capa de salida tiene 5 neuronas, una para cada tipo de fuente de energía.

Primera predicción

```

# 1ra predicción
# Material predominante del piso -> ladrillo de Cemento
# Reside 4 persona habitualmente
# Tratamiento al agua para beber -> Hervida
hogar = np.array([[2,4,2]])
p = model.predict(hogar)
print(p)

1/1 ————— 0s 25ms/step
[[0.717912  0.03786336 0.18548952 0.05316513 0.00557  ]]
```

Para un hogar que:

Tiene un piso de ladrillo de cemento (P02A04 = 2)

Esto indica que el hogar tiene un nivel de infraestructura intermedio, dado que el ladrillo de cemento es común en construcciones formales.

Aloja habitualmente a 4 personas (P01B04 = 4)

Este hogar tiene una familia de tamaño mediano, lo que podría influir en las necesidades de agua potable y tratamiento.

Utiliza hervir el agua como tratamiento del agua para beber (P02B04 = 2)

El modelo predice que la casa es propia con una probabilidad del 71.79%.

Segunda predicción

```
# 2da prediccion
# Material predominante del piso -> Ladrillo de Barro
# Reside 2 persona habitualmente
# Tratamiento al agua para beber -> Clorada
hogar = np.array([[3,2,4]])
p = model.predict(hogar)
print(p)
```

```
1/1 ————— 0s 35ms/step
[[0.65257245 0.03324347 0.24934883 0.06080176 0.00403348]]
```

Para un hogar que:

Tiene un piso de ladrillo de barro (P02A04 = 3)

Aloja habitualmente a 2 personas (P01B04 = 2)

Utiliza cloro como tratamiento del agua para beber (P02B04 = 4)

El modelo predice que la casa es propia con una probabilidad del 71.79%.

Tercera predicción

```
# 3ra prediccion
# Material predominante del piso -> Tierra
# Reside 1 persona habitualmente
# Tratamiento al agua para beber -> Ninguno
hogar = np.array([[8,1,1]])
p = model.predict(hogar)
print(p)
```

```
1/1 ————— 0s 23ms/step
[[0.7773569 0.00537622 0.09851986 0.11527006 0.00347689]]
```

Para un hogar que:

Tiene un piso de tierra (P02A04 = 8)

Aloja habitualmente a 1 personas (P01B04 = 1)

No utiliza ningún tratamiento del agua para beber (P02B04 = 1)

El modelo predice que la casa es propia con una probabilidad del 71.79%.

Modelo 2

```
[17] # Material predominante del piso
      # Tipo de servicio sanitario
      # Fuente de energia para cocinar
      # area geografico
x = data[['P02A04', 'P02B07', 'P02B05', 'areag']]
y = data['P02B04']
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=42)
y_train = y_train - 1
y_test = y_test - 1
```

✓ Diseño de la red neuronal

```
model = Sequential()
model.add(Dense(15, input_dim=4, activation='relu'))
model.add(Dense(8, activation='relu'))
model.add(Dense(6, activation='softmax'))
```

El modelo predice la tenencia de la propiedad. Recibe 4 características como entrada: Material predominante del piso, tipo de servicio sanitario, fuente de energia para cocinar y área geográfica. Utiliza una red neuronal con 1 capa oculta y una capa de salida. La capa de salida tiene 6 neuronas, una para cada tipo de tratamiento de agua.

```

# 1ra prediccion
# Material predominante del piso -> ladrillo de Cemento
# Tipo de servicio sanitario ->Inodoro red drenaje
# Fuente de energia para cocinar ->Propano
# Area Urbano
hogar = np.array([[2,1,2,1]])
p = model.predict(hogar)
print(p)

# 2da prediccion
# Material predominante del piso -> Ladrillo de Barro
# Tipo de servicio sanitario ->Letrina
# Fuente de energia para cocinar ->Leña
# Area rural
hogar = np.array([[3,4,4,2]])
p = model.predict(hogar)
print(p)

# 3ra prediccion
# Material predominante del piso -> Tierra
# Tipo de servicio sanitario -> No tiene
# Fuente de energia para cocinar ->Electricidad
# Area rural
hogar = np.array([[8,5,1,2]])
p = model.predict(hogar)
print(p)

1/1 _____ 0s 57ms/step
[[0.06040111 0.11220074 0.14243816 0.04568375 0.6356979 0.00357832]]
1/1 _____ 0s 20ms/step
[[0.18203261 0.4343425 0.06440654 0.16165341 0.14942673 0.00813822]]
1/1 _____ 0s 20ms/step
[[0.11705028 0.21641406 0.12030015 0.09162835 0.45026374 0.00434341]]

```

Primera predicción

Material predominante del piso: Ladrillo de Cemento (P02A04 = 2)

Tipo de servicio sanitario: Inodoro conectado a red de drenaje (P02B07 = 1)

Fuente de energía para cocinar: Propano (P02B05 = 2)

Área geográfica: Urbano (areag = 1)

El modelo predice que la categoría más probable con 63.57% un hogar con el tratamiento de agua es comprar agua purificada.

Segunda predicción

Material predominante del piso: Ladrillo de Barro (P02A04 = 3)

Tipo de servicio sanitario: Letrina (P02B07 = 4)

Fuente de energía para cocinar: Leña (P02B05 = 4)

Área geográfica: Rural (areag = 2)

El modelo predice que la categoría más probable es con 43.43% un hogar con el tratamiento del agua es hervir el agua.

Tercera predicción

Material predominante del piso: Tierra (P02A04 = 8)

Tipo de servicio sanitario: No tiene (P02B07 = 5)

Fuente de energía para cocinar: Electricidad (P02B05 = 1)

Área geográfica: Rural (areag = 2)

El modelo predice que la categoría más probable es con 45.02% un hogar con el tratamiento de agua es comprar agua purificada.

Problema identificado

Falta de acceso a fuentes de energía seguras y eficientes para cocinar en hogares rurales y de bajos recursos en Guatemala. Los patrones muestran una dependencia de fuentes como leña, carbón, y propano, especialmente en hogares sin servicios municipales de eliminación de residuos ni acceso a electricidad. Esto expone a las familias a riesgos de salud respiratoria y contribuye a la deforestación. Además, muchos hogares compran agua purificada en lugar de utilizar métodos de tratamiento, reflejando preocupaciones sobre la calidad del agua disponible. La combinación de estos factores revela una situación que afecta tanto la salud pública como el medio ambiente.

Propuestas

Promoción de cocinas eficientes y seguras

Implementar programas de subsidios o créditos para facilitar el acceso a estufas mejoradas que reduzcan el consumo de leña y minimicen los riesgos de enfermedades respiratorias.

La dependencia de la leña como fuente de energía en áreas rurales indica la necesidad de soluciones más sostenibles y seguras.

Estudios en Guatemala han demostrado que el uso de estufas eficientes puede reducir hasta en un 50% la cantidad de leña utilizada, mejorando la calidad del aire en interiores.

Fortalecimiento de infraestructura sanitaria

Invertir en sistemas de alcantarillado y tratamiento de aguas residuales en comunidades rurales, priorizando hogares que utilizan letrinas o no tienen acceso a servicios sanitarios.

Los modelos identificaron que el tipo de servicio sanitario está estrechamente relacionado con la calidad de vida y las prácticas de tratamiento del agua.

Mejores servicios sanitarios están vinculados a una reducción en enfermedades hídricas.

Aprovechar programas como los Fondos de Desarrollo Social (FODES) para priorizar estos proyectos.

Educación sobre tratamiento de agua

Realizar campañas de concientización en comunidades rurales sobre métodos económicos de tratamiento de agua, como hervirla, y sus beneficios.

El modelo muestra que una proporción significativa de hogares no trata el agua debido a la falta de conocimiento o recursos.

Campañas previas en Guatemala han demostrado ser efectivas en cambiar comportamientos relacionados con la higiene y el consumo de agua.

Las municipalidades pueden trabajar con el Ministerio de Salud Pública y Asistencia Social (MSPAS) para distribuir materiales educativos y equipos básicos.

Reducción de costos de agua purificada

Incentivar la producción local de filtros de agua económicos y eficientes para áreas rurales.

El modelo evidencia una alta dependencia de la compra de agua purificada en áreas con infraestructura básica.

Los filtros cerámicos han sido exitosos en comunidades rurales de países vecinos como Honduras.

Instituciones como el Instituto de Ciencia y Tecnología Agrícola (ICTA) pueden liderar investigaciones para crear filtros adaptados al contexto guatemalteco.

Repositorio

https://github.com/Mruiz-99/Proyecto_parte2

Bibliografía

HELPS International. (2020). Improving health and the environment with clean cookstoves in Guatemala. Recuperado de <https://www.helpsintl.org>

Instituto Nacional de Estadística. (2022). Encuesta Nacional de Condiciones de Vida 2022. Guatemala: INE.

Ministerio de Salud Pública y Asistencia Social. (2020). Impacto de la calidad del agua en la salud de los guatemaltecos. Recuperado de <https://www.mspas.gob.gt>

World Health Organization. (2021). Household air pollution and health. Recuperado de <https://www.who.int>

Rojas, L., & Martínez, P. (2020). Efficient use of resources in rural Guatemala: A case study. *Journal of Environmental Studies*, 18(3), 45-59.