Universidad de San Carlos de Guatemala

Escuela de estudios de postgrado

Maestría en ingeniería para la industria con especialización

en ciencias de la computación

Introducción al análisis de datos.

No carne: 999014035

**Documentación**

**Resultados**

**Predicción por medio de árboles de decisión**

**Árbol 1**

**Escala de tiempo

Descripción generada automáticamente**

**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 ≥ 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.

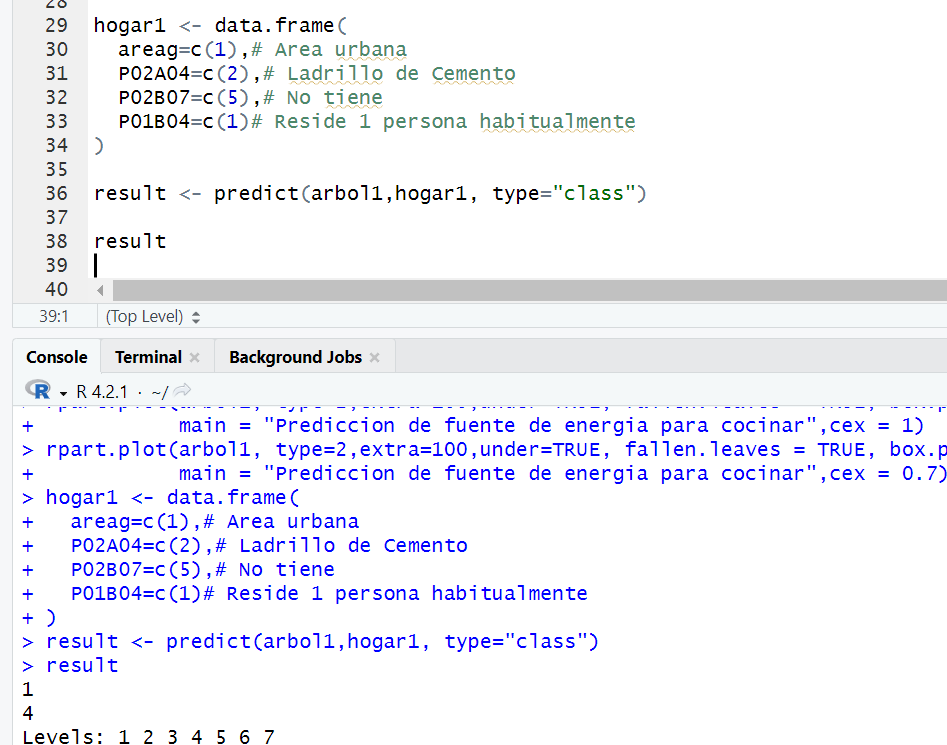
9% de los hogares usan propano si el material del piso es Cerámico, Ladrillo de Cemento, Ladrillo de Barro, Torta cemento o Granito en áreas rurales.

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

Nodo terminal 4 (Leña):

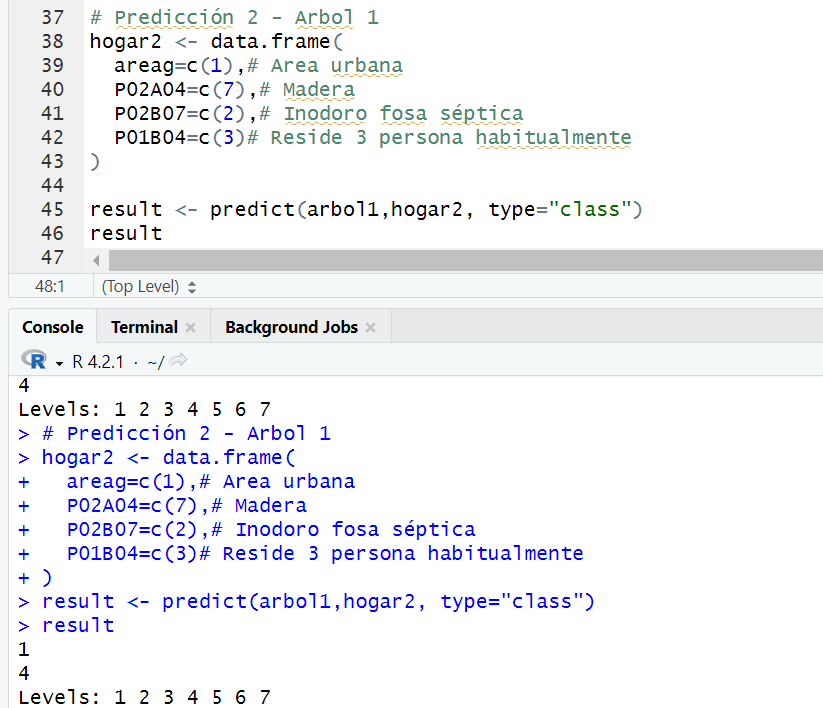
40% de los hogares usan Leña si el tipo de sanitario es Inodoro fosa séptica, Excusado, Letrina o No tiene y el área es urbana.

**Primera predicción**



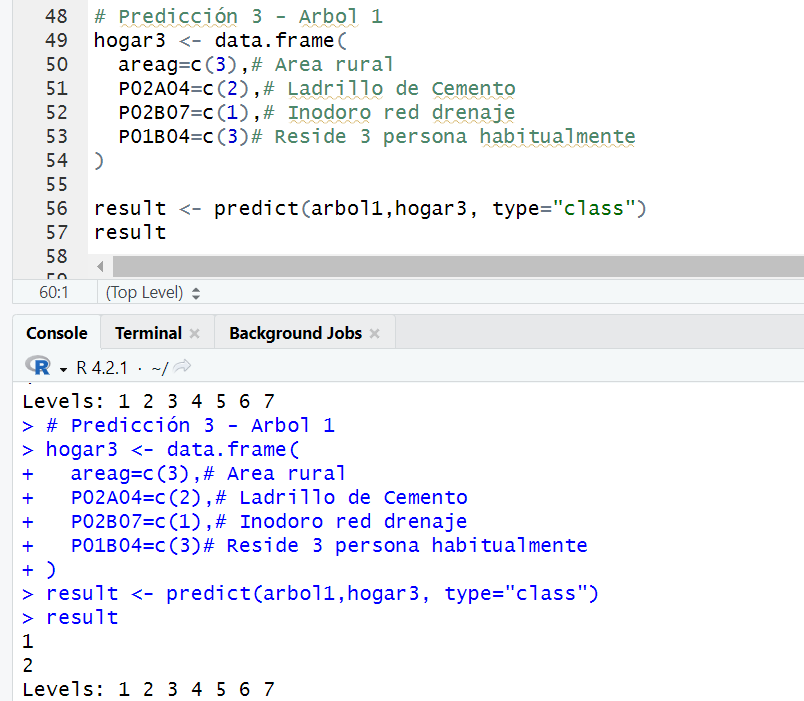
Para un hogar que pertenece al área urbana, que el material predominante del piso es ladrillo de cemento, que habitualmente reside 1 persona y que no tiene ningún tipo de servicio sanitario, la predicción muestra que la fuente de energía para cocinar es leña.

**Segunda predicción**



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**



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**

**Diagrama

Descripción generada automáticamente**

**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 ≥ 4 (Leña, Carbón, No cocina, Otro):**

areag (Área geográfica):

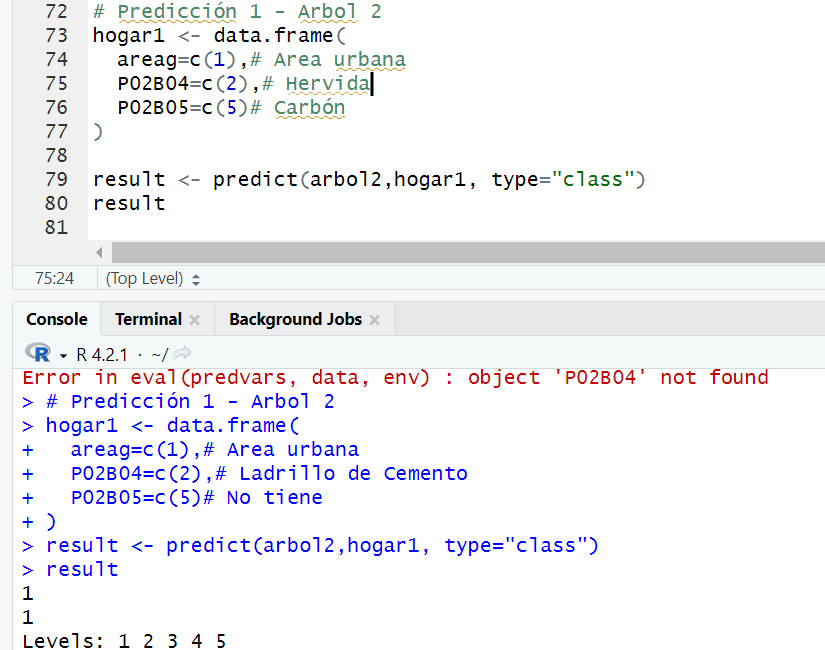
Si areag es menor que 2 (es decir, rural):

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**



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.

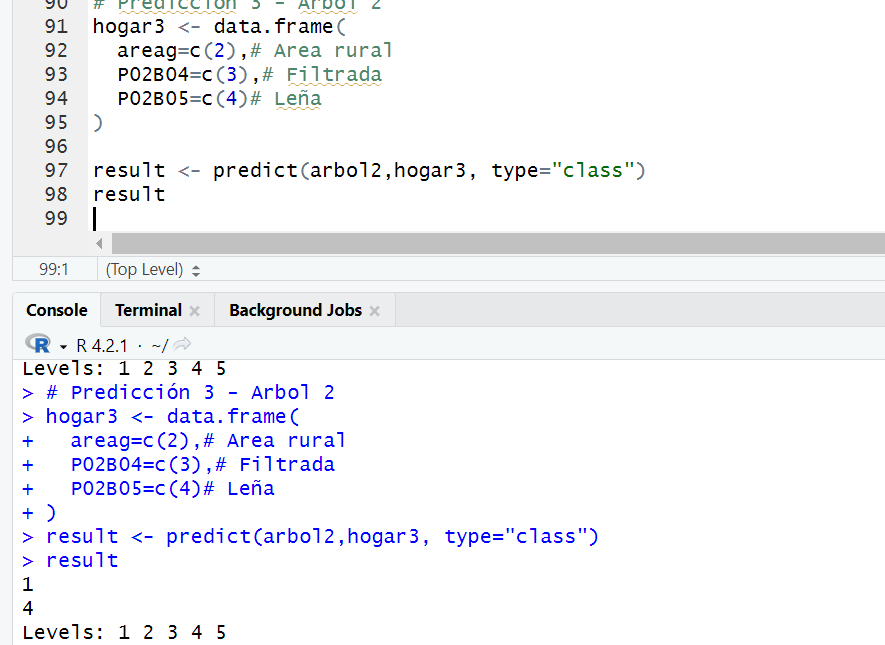
**Segunda predicción**

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

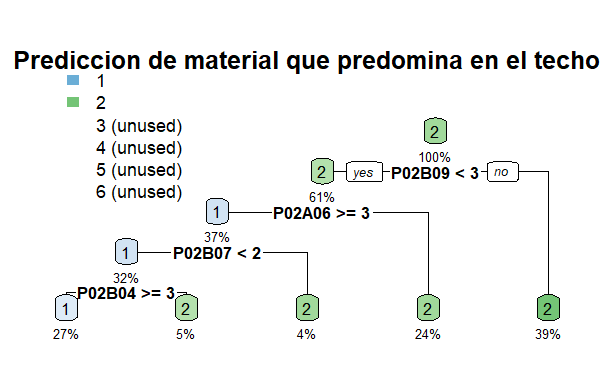
Para un hogar que pertenece al área rural, que el tratamiento al agua para beber es clorado y que tiene una fuente de energía para cocinar el kerocene, la predicción muestra que el tipo de servicio de sanitario es inodoro de red de drenaje.

**Tercera predicción**



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 el leña, la predicción muestra que el tipo de servicio de sanitario es letrina.

**Árbol 3**

****

**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**

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

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**

Texto

Descripción generada automáticamente

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.

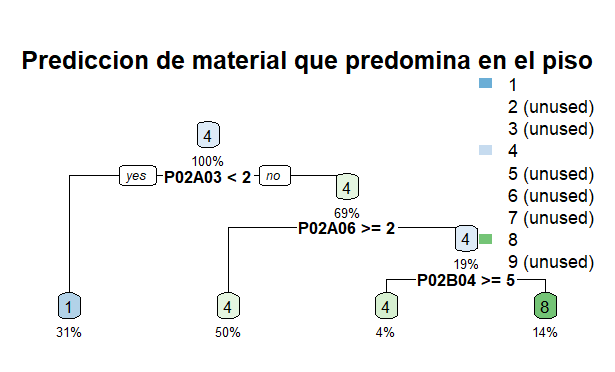
**Tercera predicción**

**Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente**

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**

****

**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).

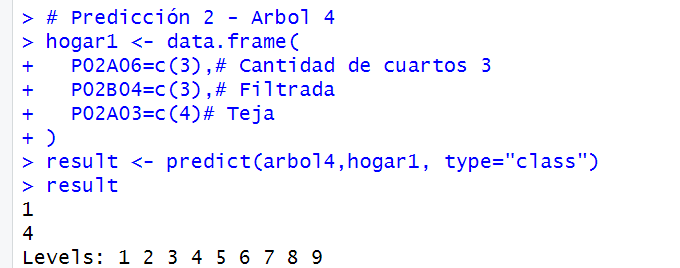
**Primera predicción**

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

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**



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**

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

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**

Gráfico

Descripción generada automáticamente

**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**

Texto, Carta

Descripción generada automáticamente

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**

**Texto

Descripción generada automáticamente**

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

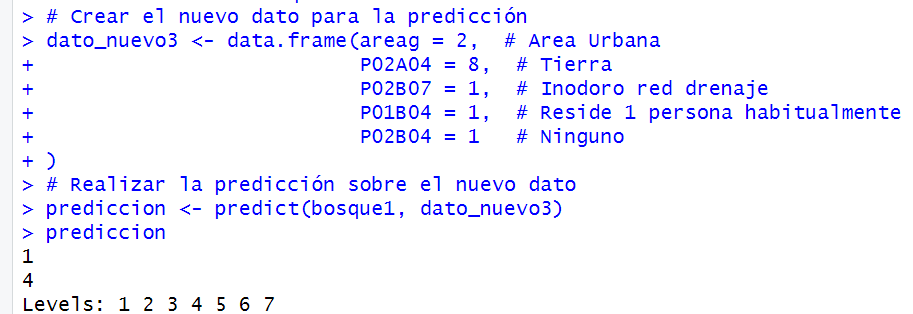
El hogar tiene una población pequeña, lo que podría influir en las necesidades y hábitos de consumo.

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**



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**

Gráfico

Descripción generada automáticamente

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**

Texto

Descripción generada automáticamente

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)

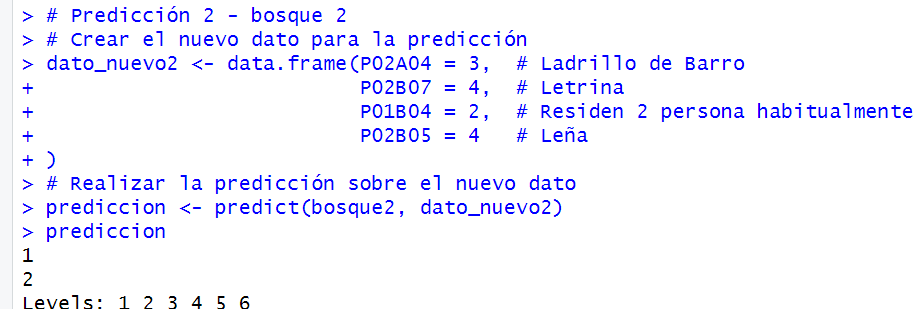
Este hogar tiene una familia de tamaño mediano, lo que podría influir en las necesidades de agua potable y tratamiento.

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**



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.

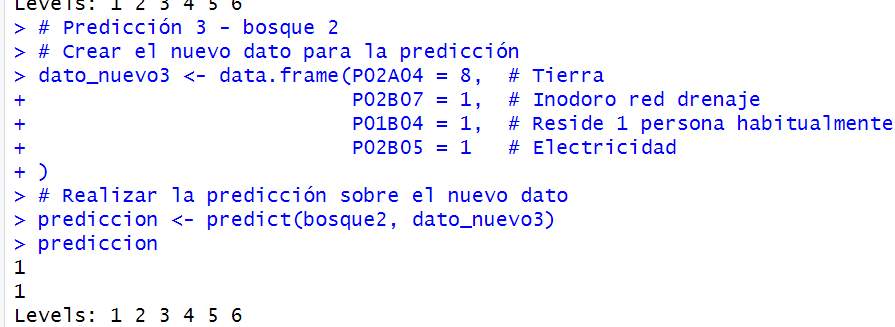
Utiliza leña como fuente de energía para cocinar (P02B05 = 4)

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**



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**

**Reglas de asociación Apriori**

**1er patrón**









Los hogares con piso de cerámica y servicio sanitario conectado a una red de drenaje utilizan propano o kerosene como fuente de energía para cocinar en un 86.97% de los casos.

**2do patrón**



****

****

En los hogares que eliminan la basura mediante métodos como quemarla, enterrarla, tirarla, o utilizando aboneras u otros medios, sin contar con servicio municipal ni privado, el 76.1% emplea leña, carbón o no cocina como fuente de energía. Estos hogares no cuentan con electricidad, propano ni kerosene para cocinar.

**3er patrón**





En los hogares donde la fuente de energía para cocinar es propano o kerosene, sin incluir aquellos que utilizan electricidad, leña, carbón o que no cocinan, el 59% compra agua purificada para beber, sin recurrir a tratamientos como hervir, filtrar, clorar el agua, o no aplicar ningún tratamiento.

**4to patrón**







En los hogares donde el servicio sanitario está conectado a una red de drenaje, el 79.76% utiliza propano o kerosene como fuente de energía para cocinar.

**Kmeans**

Gráfico, Gráfico de dispersión

Descripción generada automáticamente

Clúster 1 (Verde)

Este grupo parece estar compuesto por hogares con materiales de piso predominantemente básicos o tradicionales, como tierra o concreto, y que poseen una mayor cantidad de cuartos en la vivienda. Las fuentes de energía utilizadas para cocinar son variadas, lo cual indica que estos hogares podrían tener acceso a diferentes tipos de combustible. Este clúster podría representar hogares con recursos moderados en áreas semiurbanas.

Clúster 2 (Azul)

Los hogares en este clúster muestran un uso de materiales más diversos para el piso, pero tienden a tener un número de cuartos moderado. Las fuentes de energía para cocinar también son variadas, pero el patrón indica una posible dependencia de combustibles como leña o carbón. Este grupo podría corresponder a hogares rurales o semiurbanos que tienen acceso limitado a electricidad o gas.

Clúster 3 (Naranja)

Este grupo agrupa hogares con materiales de piso de mayor calidad, probablemente como cerámica, y tienden a tener menos cuartos en comparación con los otros clústeres. La fuente de energía para cocinar en estos hogares es más avanzada, sugiriendo el uso de propano o electricidad. Esto podría representar hogares de recursos económicos ligeramente superiores, con mejor acceso a infraestructura y servicios.

Clúster 4 (Rosa)

Este clúster está formado por hogares con pocos cuartos y materiales básicos en el piso. Estos hogares tienden a usar fuentes de energía para cocinar que podrían ser menos eficientes o más contaminantes, como leña o carbón. Es probable que estos hogares representen áreas rurales o de bajos recursos, donde el acceso a servicios básicos y energías limpias es limitado.

**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**

Para mejorar el acceso a fuentes de energía seguras y reducir el uso de métodos contaminantes como la leña, se propone la implementación de un programa de acceso a energía renovable y educación en el uso sostenible de recursos domésticos, en particular en comunidades rurales.

La propuesta incluye:

**Instalación de estufas mejoradas**

Estas estufas están diseñadas para ser más eficientes en el consumo de leña y producir menos humo, mejorando la salud de las familias y reduciendo la deforestación. Estudios han demostrado que el uso de estufas mejoradas en Guatemala reduce el consumo de leña y mejora la salud respiratoria (Torres et al., 2019).

**Iniciativas de acceso a energía solar**

Implementar paneles solares para iluminación y pequeñas estufas eléctricas en áreas rurales. Esto puede reducir la dependencia de fuentes de energía no renovables y mejorar la calidad de vida de las familias (Rodríguez & Pérez, 2020).

**Educación en el manejo de residuos y agua**

Programas de educación para enseñar a las comunidades cómo tratar el agua y gestionar los residuos de manera segura, con el fin de reducir el riesgo de enfermedades. Además, promover el uso de técnicas de compostaje y otras prácticas de eliminación de residuos que disminuyan la necesidad de quemar basura (López et al., 2021).

**Subsidios para combustibles menos contaminantes**

Facilitar el acceso a combustibles como el gas propano mediante subsidios o créditos, de modo que sea más asequible para las familias rurales, especialmente aquellas con menos recursos (García & Martínez, 2018).

**Repositorio**

https://github.com/Mruiz-99/Proyecto1

**Bibliografía**

García, J., & Martínez, M. (2018). Eficacia de subsidios en energías alternativas para zonas rurales. Revista de Economía Ambiental, 25(3), 45-67.

López, F., Gómez, H., & Pérez, C. (2021). Manejo de residuos y educación ambiental en comunidades rurales. Estudios de Medio Ambiente, 19(1), 31-47.

Rodríguez, P., & Pérez, R. (2020). Energía solar como alternativa en zonas de difícil acceso. Tecnologías Renovables, 10(4), 113-126.

Torres, L., Méndez, D., & Cruz, A. (2019). Impacto de estufas mejoradas en comunidades rurales. Salud y Ambiente, 14(2), 55-69.