



Predicción del Tipo de Agresión en Violencia Intrafamiliar

En esta tarea trabajé con la base de datos de violencia intrafamiliar del año 2024.

El objetivo fue **predecir el tipo de agresión sufrida por la víctima** a partir de variables como la edad, el sexo, la escolaridad y la situación laboral, tanto de la víctima como del agresor.

De esta forma, podemos identificar **patrones y perfiles de riesgo**, que pueden servir para orientar estrategias de prevención.

Preparación de los datos

Primero, cargué la base de datos y generé **nuevas variables derivadas**:

- universitaria: indica si la víctima tiene educación universitaria o superior.
- agresor_trabaja: indica si el agresor trabaja.

Esto permitió simplificar el análisis y obtener un modelo más interpretable.

Luego, agrupé la variable HEC_TIPAGRE (que tiene 15 códigos distintos) en tres categorías generales:

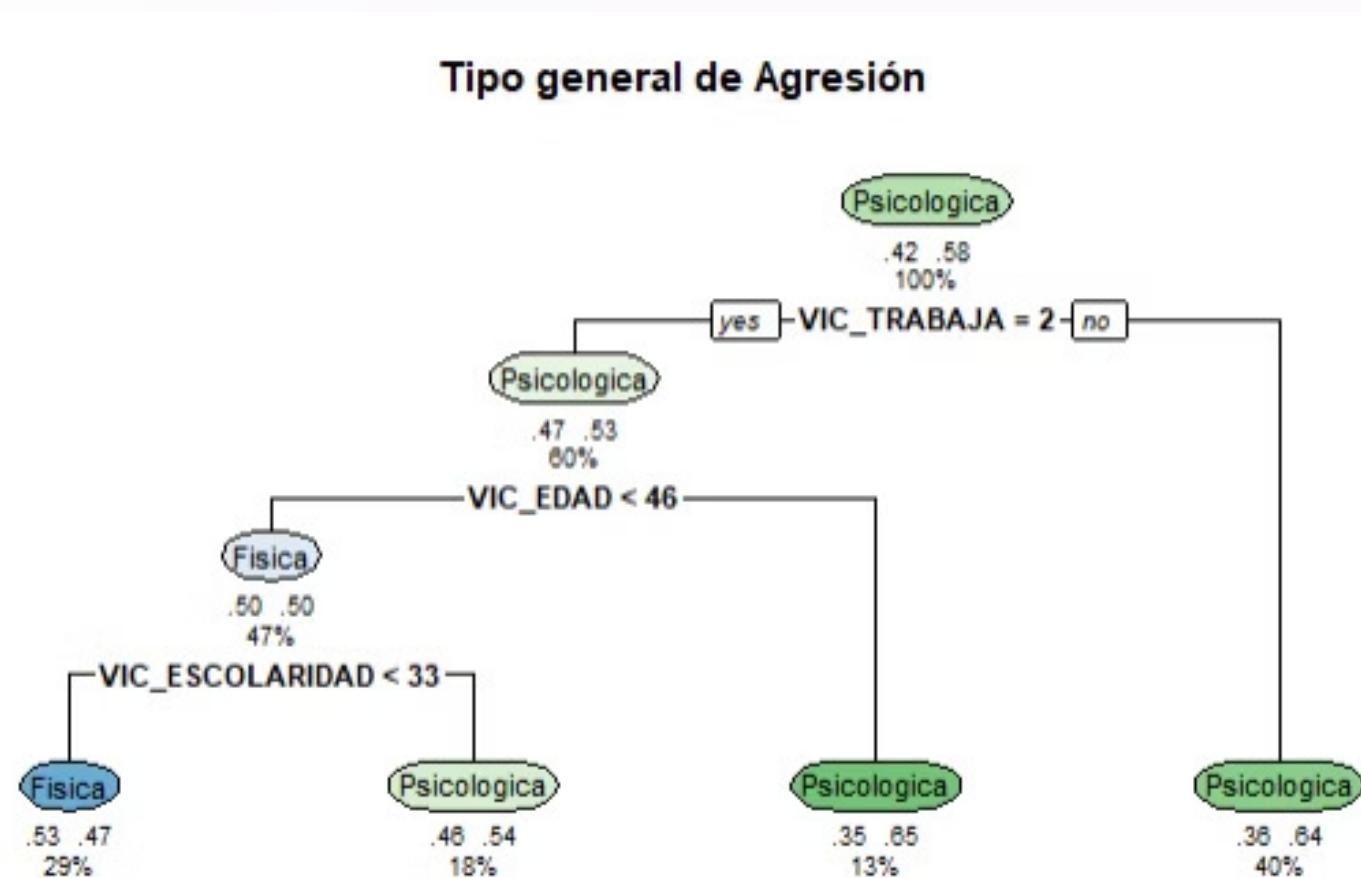
- **Física**
- **Psicológica**
- **Sexual**

Modelo del árbol de decisión

Entrené un árbol de decisión con la variable **tipo_general** como variable objetivo.

Las variables predictoras fueron: VIC_SEXO, VIC_EDAD, VIC_TRABAJA, VIC_ESCOLARIDAD, AGR_TRABAJA, AGR_ESCOLARIDAD y universitaria.

El árbol obtenido se muestra a continuación:



Interpretación

Víctima no trabaja

Si la víctima **no trabaja** (**VIC_TRABAJA = 2**), el modelo predice con alta probabilidad que el tipo de agresión es **psicológica**.

Víctima trabaja

Si la víctima **trabaja**, la predicción depende de la edad y la escolaridad:

- Menor de 46 años y con baja escolaridad → más probable **violencia física**.
 - Mayor de 46 años o con educación superior → más probable **violencia psicológica**.

El modelo indica que la **variable más importante** es la condición laboral de la víctima (VIC_TRABAJA), seguida por la edad y la escolaridad.

En general, la **violencia psicológica** es la más frecuente y aparece en la mayoría de ramas del árbol.

Predicción de un caso simulado

Para probar el modelo, se simuló una persona con las siguientes características:



Mujer



23 años



No trabaja

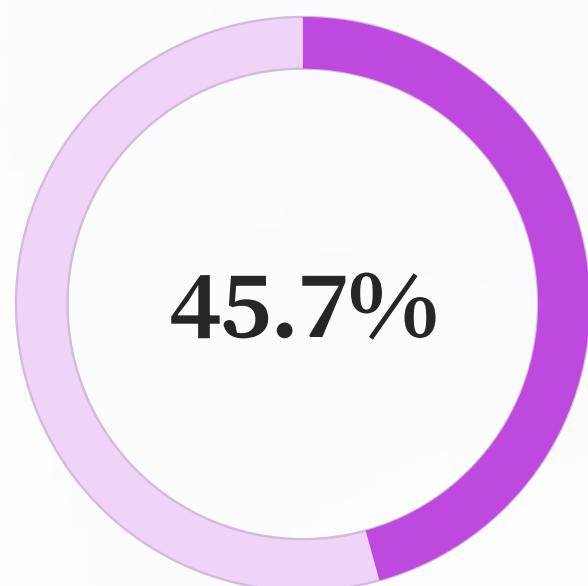


Educación universitaria

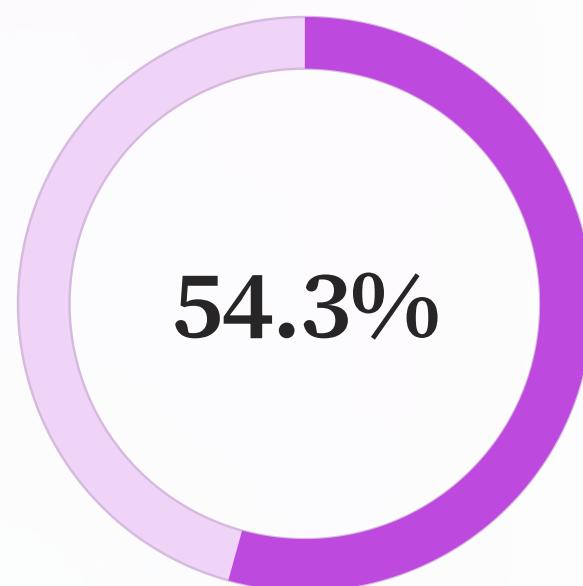


Agresor sí trabaja

El resultado de la predicción fue:



Física



Psicológica

Es decir, el modelo predice que **la probabilidad más alta (54.3%) corresponde a violencia psicológica**.

Conclusiones

La condición laboral de la víctima es la **variable más influyente** en el tipo de agresión.

Las mujeres jóvenes y sin empleo presentan una mayor probabilidad de sufrir **violencia psicológica**.

El modelo permite entender mejor cómo se relacionan las variables sociodemográficas con el tipo de violencia reportada.

En resumen, los árboles de decisión son una herramienta visual y práctica para **descubrir patrones de comportamiento en datos reales de carácter social**.



Reflexión final

Este análisis demuestra cómo la minería de datos puede aplicarse a temas sociales

ayudando a identificar factores de riesgo y posibles líneas de acción para políticas públicas y programas de prevención de la violencia intrafamiliar.