

UNIVERSIDAD DEL VALLE DE GUATEMALA



Informe proyecto 2

Bryan Carlos Roberto España Machorro - 21550

Javier Alejandro Prado Ramirez - 21486

Diego Valdez - 21328

GUATEMALA, 23 de mayo de 2024

Introducción

Guatemala es un país donde muchas personas sufren de violencia intrafamiliar de distintos tipos, como grupo se interesó en esta problemática y deseamos investigar acerca de esta misma. A lo largo de la investigación se llegó a observar que hay varias variables y patrones a lo largo de los casos que son repetitivos entre ellos por lo cual mediante el método de regresión logística deseamos hacer un modelo predictivo en cada tipo de violencia para poder predecir si va a ocurrir un caso de violencia intrafamiliar o no.

1. Configurar y cargar datos

Primero, incorporamos las herramientas para el análisis de datos. Usamos numpy y pandas para trabajar con datos, y matplotlib y seaborn para mostrar los resultados. Cargamos los datos y miramos las primeras filas para comprender su estructura y contenido.

2. Comprender los datos

Recopilamos detalles sobre el conjunto de datos y descubrimos que tiene 1000 filas y 36 columnas. La función info() nos proporciona los tipos de datos y la cantidad de valores nulos en cada columna.

3. Encontrar variables categóricas

Encontramos 31 variables categóricas. Estos son importantes para el análisis, ya que pueden necesitar un manejo especial para la visualización y el modelado.

4. Manejo de los valores nulos

Para manejar los valores nulos, completamos los espacios con interpolación lineal para columnas numéricas y para columnas específicas para mantener la coherencia de los datos.

5. Limpieza de datos

Intercambiamos valores incorrectos o no aptos en varias columnas categóricas para estandarizar las respuestas.

6. Mapeo de variables

Cambiamos variables categóricas a números para facilitar el análisis y la visualización.

7. Visualización de resultados

Por último, creamos varias imágenes para comprender mejor la distribución de los tipos de violencia y su conexión con el género de la víctima. Estas imágenes ayudan a analizar cómo se difunden y vinculan los diferentes tipos de violencia con el género de la víctima, proporcionando información clave sobre la dinámica de la violencia doméstica.

Resultados

Violencia Física: De las 1000 muestras, 700 (70%) reportan violencia física, mientras que 300 (30%) no la reportan.

Violencia Verbal: Predomina con un 96.7% de los casos, lo que indica que casi todas las víctimas experimentan algún tipo de abuso verbal.

Violencia Económica: Es menos común, con solo el 6.3% de las víctimas reportándose.

Violencia Psicológica: Es bastante prevalente, con un 83.3% de las víctimas que la sufren.

Violencia Sexual: Es la menos reportada, con solo un 0.9% de los casos.

Las víctimas son predominantemente mujeres (84.4%), mientras que los hombres representan el 15.5% de los casos. Solo un caso fue categorizado como otro género. La escolaridad de las

víctimas varía desde primaria hasta niveles de postgrado, con la mayoría teniendo educación secundaria o técnica. Se observa una mayor prevalencia de violencia en mujeres en comparación con hombres en todos los tipos de violencia. Esto destaca la necesidad de enfoques específicos de género en la intervención y prevención de la violencia intrafamiliar. Para evaluar la eficacia de los modelos de clasificación en la predicción de los diferentes tipos de violencia, se utilizaron matrices de confusión. Estas matrices permiten comprender mejor el rendimiento de los modelos en términos de verdaderos positivos, falsos positivos, verdaderos negativos y falsos negativos.

Violencia Física:

Verdaderos Positivos : Casos correctamente identificados como violencia física.

Falsos Positivos : Casos incorrectamente identificados como violencia física.

Verdaderos Negativos : Casos correctamente identificados como no violencia física.

Falsos Negativos: Casos incorrectamente identificados como no violencia física.

Violencia Verbal:

Similar a la física, pero con una mayor tasa de verdaderos positivos debido a la alta prevalencia.

Violencia Económica:

Menos casos en comparación con otras categorías, lo que puede resultar en una mayor proporción de falsos negativos y falsos positivos.

Violencia Psicológica:

Alta tasa de verdaderos positivos, pero también una notable cantidad de falsos negativos debido a la dificultad de identificación de este tipo de violencia.

Violencia Sexual:

Baja prevalencia y, por tanto, puede tener una alta tasa de falsos negativos.

Matriz de confusión



Gráfica 1: Matriz de confusión Económica

Violencia Económica:

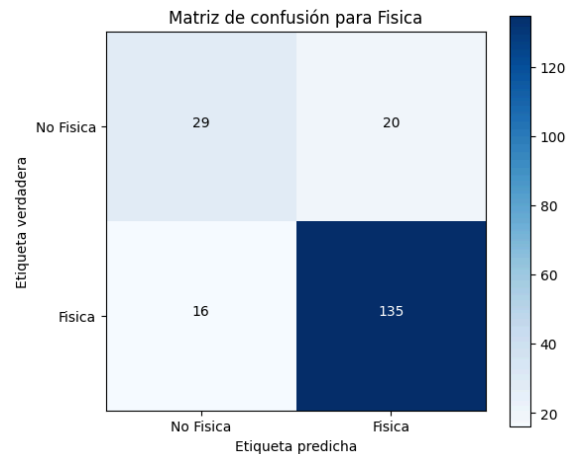
Verdaderos Positivos (VP): 186

Falsos Negativos (FN): 2

Falsos Positivos (FP): 4

Verdaderos Negativos (VN): 8

El modelo identificó correctamente 186 casos de violencia económica y clasificó incorrectamente 4 casos que no son violencia económica como positivos. Sin embargo, también dejó de identificar 2 casos de violencia económica y clasificó correctamente 8 casos como negativos.



Gráfica 2: Matriz de confusión Física

Violencia Física:

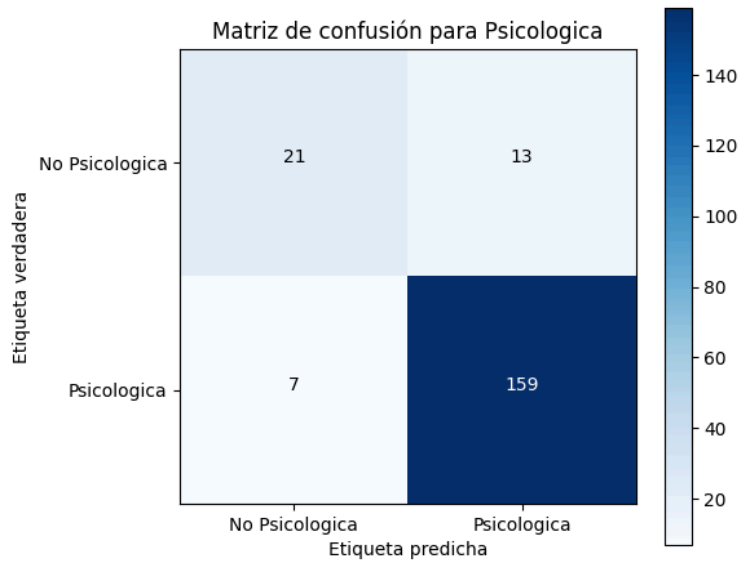
VP: 29

FN: 20

FP: 16

VN: 135

Se identificaron correctamente 29 casos de violencia física, pero 20 casos fueron clasificados incorrectamente como negativos. Además, 16 casos que no son violencia física se clasificaron erróneamente como positivos. Sin embargo, 135 casos fueron correctamente clasificados como negativos.



Gráfica 3: Matriz de confusión Psicologica

Violencia Psicológica:

VP: 21

FN: 13

FP: 7

VN: 159

El modelo identificó correctamente 21 casos de violencia psicológica, pero dejó de identificar 13 casos que eran violencia psicológica. Clasificó erróneamente 7 casos como positivos cuando no lo eran, pero 159 casos se clasificaron correctamente como negativos.



Gráfica 4: Matriz de confusión Sexual

Violencia Sexual:

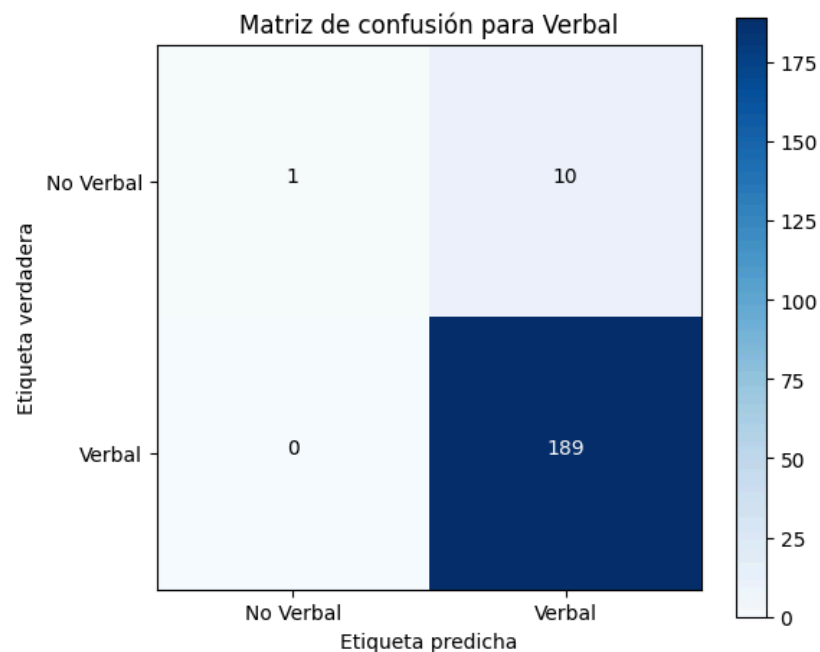
VP: 198

FN: 0

FP: 1

VN: 1

El modelo identificó correctamente 198 casos de violencia sexual y no dejó de identificar ningún caso. Sin embargo, clasificó erróneamente 1 caso que no era violencia sexual como positivo. Además, un caso se clasificó correctamente como negativo.



Gráfica 5: Matriz de confusión Verbal

Violencia Verbal:

VP: 1

FN: 10

FP: 0

VN: 189

Solo se identificó correctamente 1 caso de violencia verbal, mientras que 10 casos de violencia verbal no fueron identificados. No hubo casos que se clasificaron erróneamente como positivos, pero 189 casos se clasificaron correctamente como negativos.

Conclusiones

El análisis de datos sobre violencia doméstica muestra patrones claros en los tipos de violencia y su conexión con las características de la víctima y el perpetrador. Las decisiones tomadas durante el análisis, como la limpieza de datos, el manejo de valores faltantes y la transformación de variables, garantizan resultados precisos y útiles para futuras investigaciones y políticas.

Link a github: <https://github.com/DiegoValdez10/Proyecto2mineria>