Revista Codex Sapientia
ISSN-e 0054-9800
Vol. I N° 6 Agosto (2026) 45-70

MODELO DE APRENDIZAJE AUTOMÁTICO EXPLICABLE PARA EL ANÁLISIS DE LA VIOLENCIA CONTRA LAS MUJERES: DIÁLOGO ENTRE DATOS Y TEORÍA FEMINISTA

EXPLAINABLE MACHINE LEARNING MODEL FOR THE ANALYSIS OF VIOLENCE AGAINST WOMEN: A DIALOGUE BETWEEN DATA AND FEMINIST THEORY

Valeria Maeda Gutiérrez[1] , Norma Angélica Becerra Cuevas[2]

Resumen

La violencia durante la infancia constituye un determinante de múltiples consecuencias adversas en la vida adulta de las mujeres y se asocia con experiencias posteriores de violencia en la pareja. Este estudio identifica los factores predictores de la exposición a la violencia en la infancia en mujeres mexicanas mediante un enfoque de aprendizaje automático explicable, articulado con el modelo ecológico y la teoría feminista de la violencia por razones de género. Se utilizó una base de datos integrada a partir de la Encuesta Nacional sobre la Dinámica de las Relaciones en los Hogares (ENDIREH) 2021 y nueve fuentes complementarias, con 61,205 observaciones y 67 predictores

Abstract

Childhood violence is a well-established determinant of multiple adverse outcomes in women's adult lives and is consistently associated with subsequent experiences of intimate partner violence. This study identifies the predictive factors of exposure to childhood violence among Mexican women using an explainable machine learning approach articulated with the ecological model and feminist theory of gender-based violence. The analysis is based on an integrated dataset derived from the National Survey on the Dynamics of Household Relationships (ENDIREH) 2021 survey and nine complementary sources, comprising 61,205 observations and 67

[1] Valeria Maeda-Gutiérrez
Académica de Ingeniería Eléctrica/Universidad Autónoma de Zacatecas
valeria.maeda@uaz.edu.mx 0000-0002-0852-8315
[2] Norma Angélica Becerra Cuevas
Unidad Académica de Ciencia Política, Universidad Autónoma de Zacatecas
28902580@uaz.edu.mx 0009-0008-7947-1396

organizados en cuatro niveles. Se entrenó un modelo *CatBoost* con partición estratificada, optimización mediante validación cruzada, calibración del umbral e interpretación con valores SHAP. En el conjunto de prueba, el modelo alcanzó un AUC de 0.771 (IC 95%: 0.760-0.781), sensibilidad de 0.606 y especificidad de 0.792. El bloque historia de violencia concentró la mayor contribución predictiva, con la violencia presenciada, sexual y de la pareja en la infancia como predictores dominantes. Los hallazgos respaldan la hipótesis de transmisión intergeneracional de la violencia y el potencial de los modelos explicables para el análisis de fenómenos sociales complejos.

Palabras Clave: violencia en la infancia; violencia de género; aprendizaje automático; SHAP.

predictors organized across four ecological levels. A CatBoost model was trained using stratified partitioning, cross-validation based hyperparameter optimization, threshold calibration, and SHAP-based interpretation. On the test set, the model achieved an AUC of 0.771 (95% CI: 0.760-0.781), with a sensitivity of 0.606 and a specificity of 0.792. The history of violence block accounted for the largest predictive contribution, with witnessed, sexual, and partner-related childhood violence emerging as dominant predictors. These findings support the hypothesis of intergenerational transmission of violence and highlight the potential of explainable models for analyzing complex social phenomena

Keywords: childhood violence; gender-based violence; machine learning; SHAP.

Recibido: 18 de marzo 2026 **Aceptado:** 14 de agosto de 2026 **Publicado:** 28 de agosto 2026

Introducción

La violencia contra las mujeres por razones de género (VMG) es un fenómeno que persiste en las sociedades modernas como signo de estructuras sociales, políticas y económicas de signo patriarcal.

Entendemos por estructuras patriarcales el conjunto de relaciones de poder, prácticas, normas, instituciones y representaciones que producen y reproducen una organización jerárquica de las relaciones de género. En ello se otorga a los hombres y a lo masculino posiciones diferenciadas de autoridad, poder y privilegio, y sitúa a las mujeres y a lo femenino en posiciones de subordinación, limitada autonomía y desigual acceso a recursos, ejercicio y espacios de poder.

Dichas estructuras se reproducen mediante mecanismos sociales, culturales, económicos, políticos y jurídicos que naturalizan y legitiman la desigualdad. Además, en determinadas circunstancias, las estructuras patriarcales utilizan la violencia contra las mujeres como un mecanismo de expresión, disciplinamiento y reproducción del orden de género.

La VMG, malentendida durante larga data como un problema de la esfera privada, ha sido invisibilizada y normalizada, es decir, el ejercicio de las violencias suele concebirse como parte intrínseca (que no es necesaria modificar) de las dinámicas de la sociedad, producto de mandatos de poder en torno a roles socio-culturales.

Desde hace décadas, estudios formulados a través de las epistemologías y teorías feministas, han sabido explicar su génesis y consecuencias. En tal sentido, Lagarde (2008) sostiene que la violencia contra las mujeres no se puede entender como hechos aislados, resultado de dinámicas de los ámbitos privados, sino que se fraguan en la desigualdad estructural establecida entre mujeres y hombres.

Ello responde a los mecanismos de reproducción de la opresión hacia las mujeres, que se basan en relaciones de poder desiguales que sitúan lo masculino sobre lo femenino. Todo lo anterior crea sistemas sociales en los que las mujeres son vistas como seres “usables, prescindibles, maltratables y desechables” (Lagarde, 2008, p.16).

A la par, Segato (2003) afirma que la violencia contra las mujeres es un problema que constantemente aumenta, tanto en número de casos como en formas de ejercerla (cada vez más crueles).

Por otro lado, si bien todas las mujeres, en tanto sujetas constituidas en un orden de género desigual, se encuentran atravesadas por condiciones de vulnerabilidad, estas no se manifiestan de manera uniforme en las experiencias individuales.

La literatura muestra que la exposición a la violencia se configura a partir de la intersección de múltiples ejes de desigualdad (como la edad, la posición socioeconómica, el origen étnico y el territorio que se habita), produciendo formas diferenciadas y acumulativas de vulnerabilidad, mismas que intensifican el riesgo de sufrir violencia, e incluso, condicionan o limitan las posibilidades de acción frente a los actos de la misma (Granados-Martínez, 2022).

En consonancia con lo anterior, se entiende que la violencia contra las mujeres se expresa de formas diversas y en todos los ámbitos de sus vidas. Actualmente existen clasificaciones surgidas de posturas teóricas feministas y adoptadas por los marcos jurídicos, que permiten comprender la distinción entre ellas.

En México, la Ley General de Acceso de las Mujeres a una Vida Libre de Violencia (LGAMVLV, 2007) establece que los tipos de violencia son: psicológica, física, patrimonial, económica, sexual, vicaria y cualquier otra forma análoga que dañe su dignidad, integridad o libertad. Los instrumentos estadísticos que genera el Instituto Nacional de Estadística y Geografía (INEGI) sobre la violencia contra las mujeres toman en cuenta tal clasificación.

Dichos instrumentos jurídicos y estadísticos son a la par el resultado de contextos de suma violencia y una respuesta hasta ahora limitada por parte del Estado. En ese sentido, esta investigación surge con el objetivo de profundizar en las explicaciones en torno a un fenómeno que priva a las mujeres de una vida libre de violencias.

Lo anterior, considerando que toda acción que tenga por objetivo responder a la violencia contra las mujeres debe tener como punto de partida dos aspectos fundamentales.

El primero de ellos refiere la necesidad de comprender la violencia como consecuencia de sistemas de dominación patriarcal (tal como la hemos enmarcado al inicio). Ello permitirá que las políticas de atención, prevención, sanción o erradicación creadas por las instituciones estatales no recurran a explicaciones o intervenciones reduccionistas y/o revictimizantes.

Se busca evitar la justificación de las violencias como el resultado de dinámicas puramente privadas e interpersonales, dejando de lado los factores estructurales que las sostienen.

No obstante, el segundo aspecto revela la urgencia de identificar los elementos que en contextos específicos aumentan las situaciones de vulnerabilidad de las mujeres. Es decir, es necesario construir herramientas que, a partir de una comprensión profunda de la violencia, sepan actuar en torno a su prevención directa en entornos particulares.

En este marco, una de las dimensiones de la violencia contra las mujeres que requiere atención sostenida es la violencia experimentada durante la infancia. La literatura ha documentado de manera consistente que la exposición temprana a violencia en el hogar de origen no constituye un evento aislado en la trayectoria vital, sino que se asocia con un mayor riesgo de experimentar violencia en la vida adulta, particularmente en el ámbito de pareja (Guedes et al., 2016).

Esta continuidad ha sido conceptualizada bajo la noción de transmisión intergeneracional de la violencia, la cual sostiene que las experiencias tempranas configuran disposiciones, vínculos y entornos que tienden a reproducir, en distintas formas, las dinámicas de dominación patriarcal. En México, encuestas nacionales, como la Encuesta Nacional sobre la Dinámica de las Relaciones en los Hogares (ENDIREH), permiten analizar este fenómeno con rigor estadístico a escala poblacional.

Las aproximaciones empíricas al estudio de la violencia de pareja en México han recurrido principalmente a modelos econométricos clásicos como regresiones logísticas o modelos lineales generalizados (Casique & Castro, 2019; Torres & Martínez-Zargoso, 2022). Si bien estos enfoques han sido fundamentales, presentan limitaciones relevantes.

Por un lado, suelen asumir relaciones funcionales lineales que pueden subestimar interacciones complejas entre factores de distintos niveles. Por otro, privilegian la parsimonia interpretativa sobre la capacidad predictiva.

Frente a esto, los métodos de aprendizaje automático han adquirido relevancia en el análisis de fenómenos sociales complejos, debido a su capacidad para modelar relaciones no lineales, capturar interacciones y operar en espacios predictivos de alta dimensionalidad.

Hasta ahora, su aplicación en estudios sobre violencia contra las mujeres ha sido limitada, en parte por la percepción de que estos modelos funcionan como cajas negras de difícil interpretación, sin embargo, el desarrollo reciente de técnicas de aprendizaje automático explicable, en particular los valores *SHapley Additive exPlanations* (SHAP) (Lundberg & Lee, 2017), permite identificar de manera transparente la contribución de cada variable a las predicciones del modelo, abriendo la posibilidad de articular métodos cuantitativos avanzados con marcos interpretativos propios de las ciencias sociales.

Frente a estos antecedentes, la contribución de este estudio se plantea en dos planos complementarios:

- En el primero, un plano metodológico que integra un modelo de aprendizaje automático con técnicas de interpretabilidad para identificar no sólo la capacidad predictiva de un conjunto amplio de variables, sino también su contribución relativa en distintos niveles del modelo ecológico.
- Un segundo plano teórico, en el que los patrones identificados mediante el modelo se interpretan a la luz del enfoque ecológico y de la teoría feminista de la violencia por razones de género, de modo que los resultados no se consideran de manera aislada, sino como un diálogo con explicaciones sobre las relaciones de poder, desigualdades estructurales y la transmisión intergeneracional de la violencia.

Esta articulación busca aprovechar la capacidad analítica del aprendizaje automático sin desvincular los resultados del contexto social en el que se produce la violencia contra las mujeres.

Es por ello que este estudio tiene como objetivo identificar y caracterizar los factores asociados a la exposición a la violencia durante la infancia en mujeres mexicanas, mediante un modelo de aprendizaje automático explicable articulado con el marco ecológico y la perspectiva feminista de la violencia por razones de género.

Se plantea como hipótesis que las variables relacionadas con antecedentes de violencia en el entorno familiar presentarán una mayor contribución predictiva en la identificación de la exposición a violencia durante la infancia, en concordancia con el planteamiento de la transmisión intergeneracional de la violencia.

Metodología

Descripción del conjunto de datos

Los datos utilizados provienen del conjunto integrado por Strasser y Haensch (2025), construido a partir de 10 fuentes oficiales mexicanas.

La fuente principal fue la ENDIREH 2021, levantada por el INEGI (2021a), complementada con registros municipales y estatales sobre homicidios, migración, características censales, desigualdad, desarrollo humano, gobiernos municipales, marginación, victimización y calidad gubernamental (INEGI, 2019, 2020a, 2020b, 2021b, 2021c, 2021d; Consejo Nacional de Evaluación de la Política de Desarrollo Social [CONEVAL], 2020; Consejo Nacional de Población [CONAPO], 2020; Programa de las Naciones Unidas para el Desarrollo [PNUD], 2023).

El conjunto original fue filtrado, vinculado con fuentes complementarias, imputado, sometido a verificaciones de plausibilidad y depurado de observaciones atípicas por las autoras originales, resultando en una base de datos de 61,205 observaciones y 68 variables.

A diferencia de Strasser y Haensch (2025), quienes modelaron la violencia psicológica de pareja, este estudio tomó como variable objetivo la violencia experimentada en la infancia. Esta variable dicotómica toma el valor de “1” si la mujer reportó haber sido golpeada o insultada por alguna persona del hogar antes de los 15 años, y “0” en caso contrario.

En la muestra, 24,467 mujeres reportaron exposición (39.98%) y 36,738 no la reportaron (60.02%). Los 67 predictores se organizaron de acuerdo con el E*cological Model for Understanding Violence* (Modelo Ecológico para Comprender la Violencia) (Krug et al., 2002), estructurado en cuatro niveles: individual, relacional, comunitario y societal.

El nivel individual incluyó variables sociodemográficas, antecedentes de violencia y edades relevantes de la trayectoria vital; el relacional, características de la pareja y dinámicas del hogar; el comunitario, indicadores municipales como homicidios, migración, Gini, Índice de Desarrollo Humano (IDH), marginación y tipo de comunidad; y el societal, variables estatales sobre delitos, no denuncia, percepción de corrupción y satisfacción con servicios públicos.

Particionamiento de los datos

El conjunto se dividió en tres particiones mediante muestreo estratificado por la variable objetivo, preservando la proporción de casos positivos. El conjunto de entrenamiento (70%, n = 42,843) se utilizó para el ajuste del modelo y la optimización de hiperparámetros mediante validación cruzada.

El conjunto de validación (15%, n = 9,181) se reservó para la calibración del umbral de clasificación, mientras que el conjunto de prueba (15%, n = 9,181) se mantuvo completamente aislado y se utilizó exclusivamente para la evaluación final del desempeño.

Algoritmo de clasificación y ajuste de hiperparámetros

Categorical Boosting

Categorical Boosting (CatBoost) es un algoritmo de aprendizaje supervisado basado en árboles de decisión dentro del enfoque *gradient boosting*, que construye modelos mediante la combinación secuencial de árboles poco profundos para corregir errores previos (Prokhorenkova et al., 2018).

En clasificación binaria, estima probabilidades optimizando una función de pérdida logística, permitiendo capturar relaciones no lineales e interacciones entre variables. El algoritmo presenta tres características clave:

1. Manejo nativo de variables categóricas mediante *ordered target statistics*,

evitando fuga de información y expansión dimensional.

2. Implementación de *ordered boosting*, que reduce el sesgo y mejora la generalización.
3. Uso de árboles simétricos (*oblivious trees*), que actúan como regularización implícita (Prokhorenkova et al., 2018).

Ajuste de hiperparámetros

La configuración óptima del modelo se determinó mediante búsqueda en malla (*grid search*) con validación cruzada estratificada de 5 pliegues sobre el conjunto de entrenamiento, empleando el puntaje F1 como criterio de optimización.

La Tabla 1 presenta los hiperparámetros considerados, su función dentro del algoritmo y los valores explorados.

Tabla 1.
Hiperparámetros considerados en la búsqueda en malla con validación cruzada de 5 pliegues.

Hiperparámetro	Función	Valores explorados
Profundidad (*depth*)	Controla la complejidad individual de cada estimador	6, 8
Tasa de aprendizaje (*learning_rate*)	Regula la contribución marginal de cada árbol al ensamble	0.03, 0.05
Iteraciones (*iterations*)	Número máximo de iteraciones (árboles) del ensamble	400, 600
Coeficiente de regularización L2 (*L2_leaf_reg*)	Coeficiente de regularización L2 sobre los valores de las hojas	1.0, 3.0

Fuente: elaboración propia.

Evaluación del desempeño

El comportamiento predictivo del modelo se caracterizó a partir de la matriz de confusión, que sintetiza las predicciones correctas e incorrectas en términos de verdaderos positivos (VP), verdaderos negativos (VN), falsos positivos (FP) y falsos negativos (FN).

A partir de esta matriz se calcularon las métricas de exactitud, precisión, sensibilidad (*recall*), especificidad y puntaje F1.

La exactitud corresponde a la proporción de las predicciones correctas; la precisión mide la proporción de VP entre las predicciones positivas; la sensibilidad cuantifica la capacidad de identificar correctamente los casos positivos; la especificidad la de reconocer los casos negativos y el puntaje F1 se calculó como la media armónica entre precisión y sensibilidad.

Adicionalmente, se emplearon métricas independientes: el área bajo la curva ROC (AUC-ROC, por sus siglas en inglés), como medida de capacidad discriminativa (Fawcett, 2006) y el área bajo la curva precisión-sensibilidad (PR-AUC, por sus siglas en inglés), particularmente informativa en contextos con desbalance entre clases (Saito & Rehmsmeier, 2015).

Para estimar la incertidumbre asociada al desempeño en el conjunto de prueba, se construyeron intervalos de confianza (IC) al 95% mediante remuestreo *bootstrap* con 1,000 réplicas con reposición.

Análisis de interpretabilidad

Con el objetivo de identificar e interpretar los factores predictores del modelo se emplearon valores SHAP, un método de interpretabilidad basado en la teoría de juegos que asigna a cada variable una contribución a la predicción del modelo (Lundberg & Lee, 2017).

Los valores SHAP cuantifican, para cada observación y cada variable, la contribución marginal de dicho predictor a la diferencia entre la predicción individual y la predicción promedio del clasificador, permitiendo descomponer cada predicción como la suma de efectos atribuibles a los predictores.

Matemáticamente hablando, esta descomposición puede expresarse como f(x)=E[f(X)+ ∑_(i=1)^M▒ϕ_i], donde E[f(X)] representa la salida esperada del modelo o valor de referencia, M corresponde al número de predictores y φ_i al valor SHAP asociado con el predictor i (Lundberg et al., 2020).

Un valor SHAP positivo indica que la variable contribuye a aumentar la salida del modelo respecto al valor de referencia, mientras que un valor negativo indica una contribución en sentido contrario; su magnitud refleja la intensidad de dicha contribución. En este estudio los valores se calcularon mediante el algoritmo *TreeSHAP*, una implementación para modelos basados en árboles (Lundberg et al., 2020).

El análisis se realizó sobre el conjunto de prueba a dos niveles complementarios. Primero, se estimó la importancia agregada por bloque ecológico mediante la suma de los valores SHAP medios absolutos de las variables que integran cada bloque.

Segundo, se examinó la distribución de los valores SHAP individuales por cada variable mediante diagramas tipo *beeswarm*, que permiten visualizar simultáneamente la magnitud y dirección de los efectos.

Resultados

Selección del umbral operativo

El modelo produce, para cada observación, una probabilidad estimada de pertenencia a la clase positiva, definida como haber presenciado violencia en el hogar antes de los 15 años.

Para convertir estas probabilidades en clasificaciones binarias, fue necesario establecer un umbral de decisión. Este valor influye directamente en el balance entre sensibilidad y especificidad, por lo que su selección tiene implicaciones metodológicas y sustantivas para la interpretación del modelo.

Con el fin de transparentar esta decisión, se evaluaron cuatro criterios de selección del umbral en el conjunto de validación:

1. El umbral por defecto (0.500).
2. La maximización del puntaje F1.
3. La maximización de la exactitud balanceada.
4. La maximización de la sensibilidad sujeta a una especificidad mínima de 0.80.

Posteriormente, cada umbral identificado en validación se aplicó al conjunto de prueba sin reajuste, con el objetivo de estimar su desempeño en datos no utilizados durante la selección del umbral. Los resultados se presentan en la Tabla 2 del apartado de Anexos.

El criterio seleccionado fue C4, correspondiente a un umbral de 0.485. Esta elección responde a dos consideraciones complementarias. En primer lugar, desde una perspectiva de equilibrio entre métricas, este umbral produce un perfil del clasificador en el que la sensibilidad, la especificidad, el puntaje F1 y la exactitud se mantienen en rangos consistentes en el conjunto de prueba, evitando configuraciones extremas como la observada en C2, donde la maximización del F1 reduce la especificidad a 0.581.

En segundo lugar, la selección del umbral fundamenta el propósito analítico del estudio. El modelo no se concibe como una

herramienta de clasificación operativa a nivel individual, sino como un instrumento para identificar y analizar los factores asociados a la exposición a violencia en la infancia mediante técnicas de interpretabilidad.

Desde el punto de vista interpretativo, mantener una especificidad elevada dentro del rango de equilibrio contribuye a que las inferencias derivadas se sustenten en un conjunto de predicciones positivas con menor proporción de FP.

Desempeño predictivo del modelo

La optimización de hiperparámetros mediante búsqueda en malla (en inglés, *grid search*) con validación cruzada estratificada de 5-pliegues, aplicada sobre el conjunto de entrenamiento, permitió identificar la configuración óptima del modelo CatBoost.

Los valores seleccionados fueron: *depth* = 6, *learning rate* = 0.05, *iterations* = 600 y coeficiente de *l2_leaf_reg* = 3.0. Esta configuración alcanzó un puntaje F1 promedio de 0.74 a lo largo de los pliegues de validación cruzada, lo que indica un desempeño consistente del modelo durante el proceso de ajuste.

Posteriormente, el modelo final fue reentrenado utilizando la totalidad del conjunto de entrenamiento con los hiperparámetros óptimos obtenidos.

Los resultados del modelo en los conjuntos de entrenamiento, validación y prueba se presentan en la Tabla 3 (ver Anexos). Con el umbral por defecto, el modelo alcanzó un AUC de 0.880 en entrenamiento, 0.777 en validación y 0.770 en prueba, lo que indica una capacidad discriminativa consistente.

En términos de sensibilidad, se observa una disminución progresiva de 0.691 en entrenamiento a 0.587 en prueba, mientras que la especificidad se mantiene relativamente alta en todos los conjuntos.

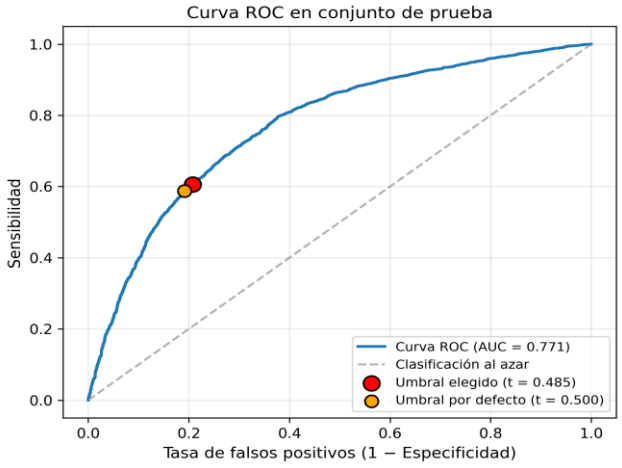
Al emplear el umbral optimizado, se observa un incremento en la sensibilidad en los tres conjuntos (0.710, 0.596 y 0.605, respectivamente), acompañado de una ligera disminución en la precisión y especificidad. Este comportamiento evidencia el ajuste del modelo hacia una mayor capacidad de detección de los casos positivos.

La Figura 1 presenta la curva ROC del modelo evaluado en el conjunto de prueba, con un AUC-ROC de 0.771, lo que indica una adecuada capacidad discriminativa entre las clases a lo largo de distintos umbrales de decisión.

Además, se incluyen dos puntos de operación (1) el correspondiente C1 y (2) el asociado al C4, los cuales permiten ilustrar el comportamiento del modelo en condiciones específicas de clasificación. En conjunto, la curva evidencia que el modelo mantiene un desempeño consistente por encima de la clasificación aleatoria en todo el rango de decisión.

Figura 1.

Curva ROC del modelo CatBoost en el conjunto de prueba, con los puntos de operación correspondientes al umbral por defecto y al umbral elegido.

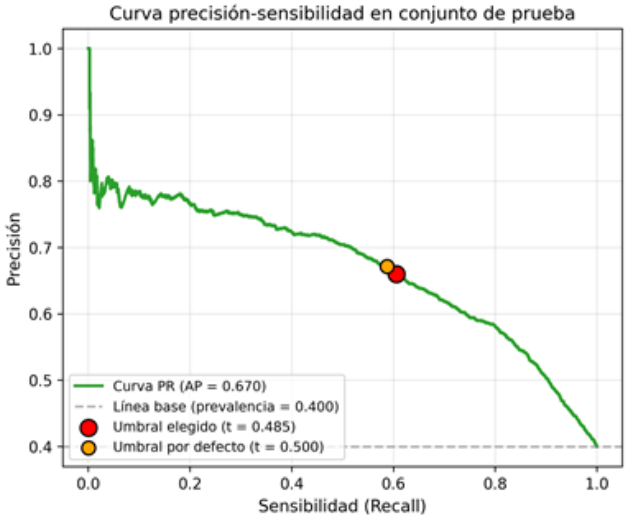


Fuente: elaboración propia.

El análisis precisión-sensibilidad del modelo en el conjunto de prueba muestra un AP de 0.670 (Figura 2), valor superior a la prevalencia de la clase positiva en la muestra (0.400) que constituye la línea base para esta métrica en problemas con desbalance.

Figura 2.

Curva precisión-sensibilidad del modelo CatBoost en el conjunto de prueba, con los puntos de operación correspondientes al umbral por defecto y al umbral elegido.



Fuente: elaboración propia.

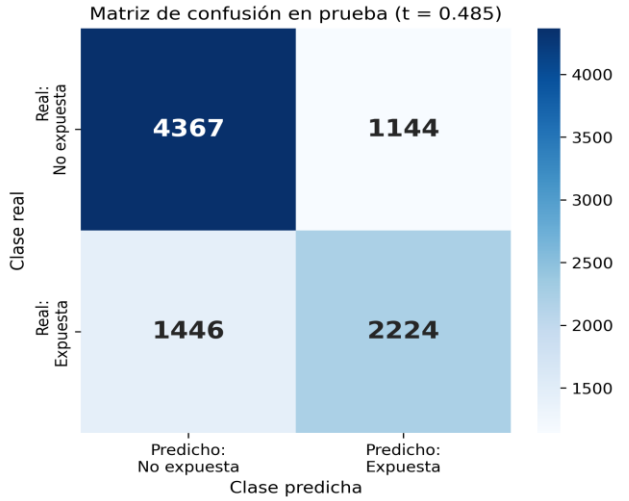
La Figura 3 presenta la matriz de confusión del modelo en el conjunto de prueba bajo el umbral operativo seleccionado. De las 5,511 mujeres que reportaron no haber sido expuestas a violencia en la infancia, el modelo clasificó correctamente a 4,367 (especificidad = 0.792) y clasificó erróneamente a 1,144 como expuestas (proporción de FP = 0.208).

Por su parte, de las 3,670 mujeres que reportaron exposición, el modelo identificó a 2,224 (sensibilidad = 0.606) y no detectó a 1,446 (proporción de FN = 0.394).

Este patrón de errores es consistente con el perfil del clasificador derivado del criterio de selección del umbral, orientado a mantener un equilibrio entre métricas.

Figura 3.

Matriz de confusión del modelo en el conjunto de prueba bajo el umbral operativo seleccionado.



Fuente: elaboración propia.

Para cuantificar la incertidumbre asociada al desempeño del modelo, se construyeron IC al 95% para cada métrica mediante remuestreo bootstrap con 1,000 réplicas con reposición. Los resultados se presentan en la Tabla 4.

Tabla 4.

IC al 95% de las métricas de desempeño del modelo en el conjunto de prueba, estimados mediante remuestreo bootstrap con 1,000 réplicas.

Métrica	Estimación	IC inferior	IC superior
Exactitud	0.718	0.709	0.727
Precisión	0.660	0.644	0676
Sensibilidad	0.606	0.590	0.622
Especificidad	0.793	0.782	0.803
Puntaje F1	0.632	0.617	0.645
AUC	0.771	0.760	0.781

Fuente: elaboración propia.

La amplitud reducida de los intervalos (entre 1.8 y 3.2 puntos porcentuales para todas las métricas), junto con el hecho de que el límite inferior del intervalo del AUC (0.760) se mantiene claramente por encima del valor correspondiente a la clasificación aleatoria (0.500), respalda la estabilidad estadística del clasificador.

Adicionalmente, la sensibilidad observada en el conjunto de validación (0.569) se encuentra dentro del IC estimado para el conjunto de prueba (0.590-0.622), lo que sugiere una adecuada capacidad de generalización del umbral seleccionado hacía datos no utilizados durante la fase de calibración.

Importancia de las variables predictoras

Con el objetivo de identificar e interpretar los factores asociados a la exposición a violencia en la infancia, se calcularon valores SHAP sobre el conjunto de prueba. El análisis se realizó a dos niveles complementarios:

1. Agregación por bloques del modelo ecológico.
2. Análisis de la distribución de los efectos individuales sobre la salida del modelo.

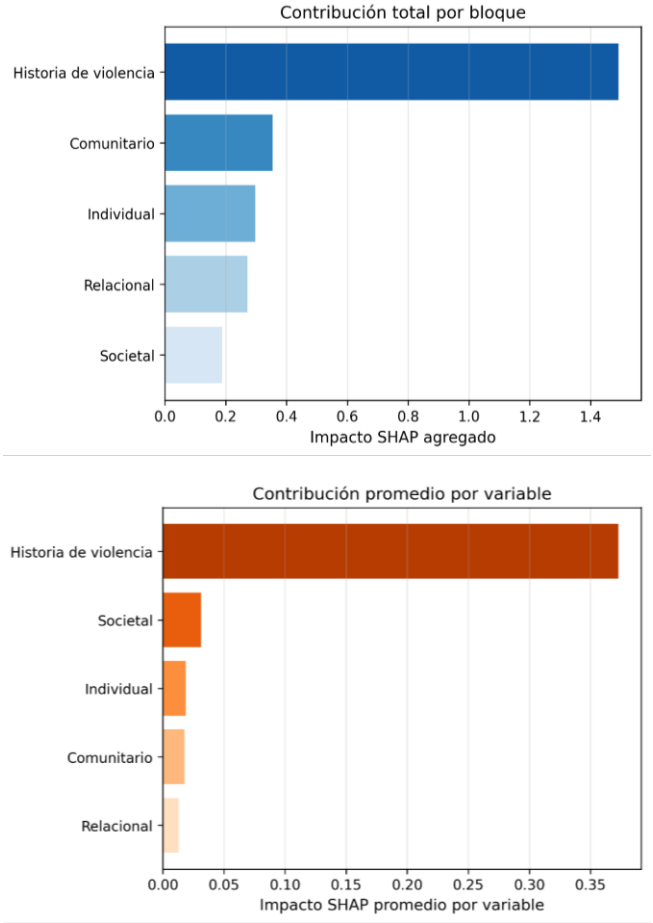
Distribución del impacto por bloque ecológico

La Figura 4 muestra la distribución del impacto predictivo de los cinco bloques del modelo ecológico considerados en el análisis.

El panel azul muestra la contribución total por bloque, calculada como la suma de los valores SHAP medios absolutos de las variables que lo integran. El panel marrón rojizo ajusta esta medida por el número de variables en cada bloque, proporcionando una estimación comparable del peso predictivo por variable.

Los resultados evidencian un patrón claramente diferenciado: el bloque historia de violencia concentra la mayor parte del impacto del modelo. En términos de contribución total, este bloque alcanza un valor aproximado de 1.5, casi cuatro veces superior al del bloque comunitario (≈ 0.35) y más de siete veces al bloque societal (≈ 0.20).

Figura 4.
Importancia de las variables por bloque del modelo ecológico estimada mediante valores SHAP. El panel izquierdo muestra la contribución total por bloque, mientras que el panel derecho presenta la contribución promedio por variable dentro del bloque.



Fuente: elaboración propia.

Esta diferencia se intensifica al considerar la contribución promedio por variable, mientras que el bloque historia de violencia presenta un valor cercano a 0.38, el resto de los bloques se mantiene por debajo de 0.05.

Es importante destacar que, en conjunto, estos resultados sugieren que las variables asociadas a la exposición a violencia en etapas tempranas (tanto de la propia mujer como de su entorno cercano) constituyen los predictores más informativos del modelo.

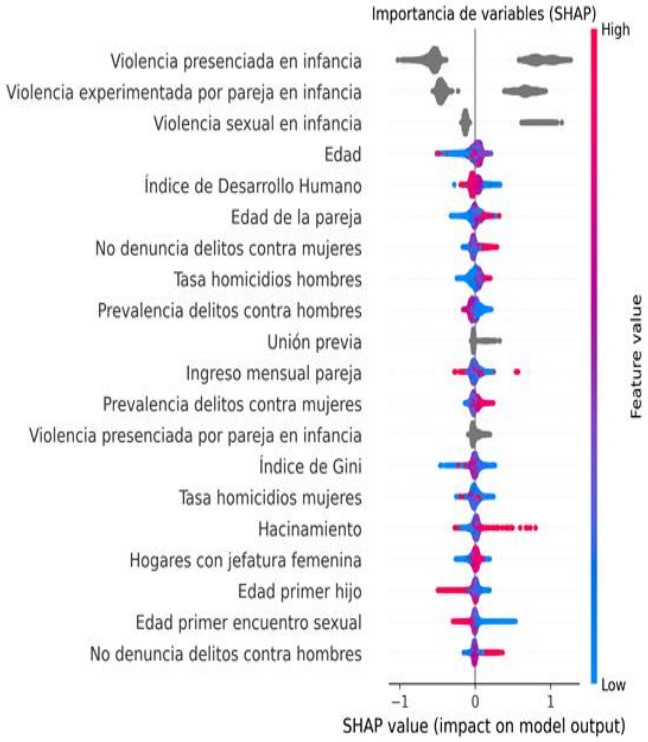
Variables más influyentes y direcciones de efecto

Para profundizar en el contenido del bloque dominante y caracterizar el comportamiento de las variables más relevantes, se examinó la distribución de los valores SHAP individuales en el conjunto de prueba.

La Figura 5 presenta un diagrama tipo *beeswarm*, en el que cada punto representa una observación. La posición horizontal indica la magnitud y dirección del efecto de cada variable (rojo para los valores altos y azul para valores bajos). En el caso de variables binarias, la interpretación se centra en la separación de sus efectos sobre el eje horizontal.

Figura 5.

Diagrama beeswarm de los valores SHAP de las 20 variables más influyentes del modelo en el conjunto de prueba.



Fuente: Lundberg & Lee (2017).

Los tres predictores con mayor impacto pertenecen al bloque historia de violencia, lo que confirma el peso central de las experiencias tempranas de violencia en la predicción del modelo.

La variable violencia presenciada en infancia ocupa la primera posición: las observaciones con respuesta afirmativa tienden a producir valores SHAP positivos cercanos a +1.0, incrementando sustancialmente la probabilidad predicha de exposición.

De forma similar, la violencia experimentada por la pareja en la infancia ocupa la segunda posición y muestra un patrón consistente con la transmisión intergeneracional y relacional de la violencia.

La violencia sexual en infancia se ubica en tercer lugar; aunque su prevalencia en la muestra es menor, su presencia genera impacto positivo considerable sobre la predicción.

A partir de la cuarta posición, la magnitud del efecto disminuye de manera marcada; la edad presenta un comportamiento compatible con efectos de cohorte: los valores SHAP tienden a ser más positivos en edades altas y más negativos en edades bajas, lo que sugiere diferencias generacionales en la exposición, reconocimiento o reporte retrospectivo de experiencias de violencia en el hogar.

Por su parte, el IDH muestra un patrón que debe interpretarse con cautela: su asociación con valores SHAP positivos no implica que el desarrollo humano constituya un factor de riesgo, sino que podría reflejar diferencias territoriales en la visibilización, identificación o reporte de experiencias de violencia.

Otras variables con impacto moderado incluyen la edad de la pareja, la no denuncia de delitos contra mujeres, la prevalencia de delitos contra hombres, la unión previa y el ingreso de la pareja.

En conjunto, estos resultados muestran que el modelo articula factores de historia de violencia, trayectorias individuales, condiciones relacionales y características contextuales, aunque el peso predictivo dominante recae en las experiencias tempranas de violencia.

Discusiones

Los resultados arrojados por el análisis comprueban lo que las teorías feministas han sabido explicar, desde otras ciencias como la antropología, la sociología, o la ciencia política, en torno a los factores que recrudecen las situaciones de vulnerabilidad de las mujeres, y por ende, agravan la violencia.

Antes mencionamos que la violencia por razones de género no se interpreta como actos individuales, sino como el resultado de relaciones jerarquizadas y desiguales, constituidas en el patriarcado. A este, Lagarde (1996) lo define como:

> … un orden social genérico de poder, basado en un modo de dominación donde el paradigma es el hombre. Este orden asegura la supremacía de los hombres y de lo masculino sobre la inferiorización previa de las mujeres y lo femenino. Es asimismo un orden de dominio de unos hombres sobre otros y de enajenación de las mujeres. (p. 52)

De lo anterior entendemos que la subordinación de grupos específicos de la sociedad es posible gracias a que las formas de relacionarnos no son neutrales, sino constituidas bajo paradigmas de segregación (género, raza, clase, entre otros).

En tal sentido, la violencia constituye un mecanismo idóneo para sostener lo anterior, y es la expresión de un lenguaje de poder que comunica, reproduce y reafirma esas jerarquías sociales.

Desde esa perspectiva, los actos de violencia no son meramente instrumentales, sino que poseen una dimensión expresiva, pues transmiten mensajes sobre el lugar que ocupan las mujeres en el orden social y sobre los límites de sus libertades y autonomía. De ahí que Segato (2003) insista en que la violencia opera como una pedagogía social que enseña, disciplina y regula comportamientos, tanto de quienes ejercen la violencia, como de quienes la observan y la viven. Lo afirma especialmente en el caso de las violaciones, a las que define como:

> … un acto violento casi en estado puro, vale decir, despojado de finalidades instrumentales… un acto violento sin sentido atraviesa a un sujeto y sale a la superficie de la vida social como revelación de una latencia, una tensión que late en el sustrato de la ordenación jerárquica de la sociedad. (Segato, 2003, pp. 22-23)

Sin embargo, tales interpretaciones pueden valer para cualquier expresión de violencia por razón de género.

Lo anterior explica que la exposición a dinámicas de violencia en la infancia no sólo implica la vivencia del daño, sino también la internalización de dichos lenguajes de poder. Niñas, niños y adolescentes aprenden a interpretar las relaciones sociales a partir de códigos violentos, naturalizando la desigualdad y la violencia como formas normales de interacción.

De ello se infiere que la exposición en la infancia a la violencia, presenciada o experimentada, no sólo reproduce condiciones de vulnerabilidad, sino que contribuye a la reproducción simbólica del orden de género.

De las encuestas se infiere que la exposición a la violencia en la infancia se produce en los espacios domésticos, por lo que es necesario referir que, contrario a lo que se suele establecer, estos no son un ámbito privado ajeno a lo político, sino que son espacios centrales de producción y reproducción de relaciones de poder.

Develar las dinámicas violentas en los hogares permite romper la invisibilización que por siglos ha permitido su reproducción.

Esas formas, entendidas como menos visibles, no operan de manera aislada, sino que se sostienen en el entramado de la cotidianidad pública y vuelven posible la emergencia de formas más extremas de violencia.

En palabras de Lagarde (1996) "La organización social patriarcal orienta el desarrollo a partir de la violencia –doméstica, privada y pública, personal e institucional– sobre las mujeres y recrea la violencia al convertirla en mecanismo de reproducción de su dominio" (pp. 120-121).

Así, las expresiones de violencia, incluyendo por supuesto las que este análisis revela (físicas y sexuales), suceden en contextos en los que previamente se han establecido relaciones de control, subordinación y desigualdad. Tal continuidad robustece la idea de que las expresiones de violencia se encuentran interconectadas y se refuerzan mutuamente.

La literatura ha documentado que la exposición a la violencia en la infancia se asocia con una mayor probabilidad de victimización y perpetración en etapas posteriores.

De acuerdo con Kim et al. (2019), la violencia aconsejada, presenciada y sufrida en los ámbitos familiar y comunitario influyeron para que los estudiantes en edad adolescente, partícipes en su investigación, aceptaran la violencia tanto de forma interpersonal, así como una estrategia válida para resolver problemas, lo que tiene como consecuencia una mayor posibilidad de vivirla o reproducirla.

Aunado a lo anterior, desde una perspectiva interseccional, el lenguaje de poder al que nos hemos referido y su articulación con otras estructuras de desigualdad, como la clase social, el territorio o la edad, explican que las formas a través de las cuales se expresa la violencia varían en función de los contextos específicos.

Debido a esto, es necesario analizar la violencia no como un fenómeno uniforme, sino como una práctica situada.

Así, la división sexual del trabajo, la feminización de la pobreza, la asignación diferenciada de los trabajos de cuidado y la persistente dependencia económica de las mujeres recrudecen las relaciones desiguales en el ámbito doméstico.

En ese sentido, el IDH, que incorpora dimensiones como el ingreso, la educación, la salud, el acceso a servicios públicos, entre otros, permite observar el impacto que tienen las desigualdades estructurales en la reducción de la capacidad de las mujeres para ejercer su autonomía, y afrontar las violencias que experimentan.

En contextos de niveles inferiores de desarrollo humano se identifican dos aspectos. El primero de ellos es un aumento de la violencia, y el segundo, las restricciones simbólicas y materiales que tienen las mujeres para abandonar las situaciones de violencia.

El análisis SHAP permitió examinar empíricamente el peso relativo de los predictores asociados a la exposición a violencia durante la infancia. Los hallazgos muestran que tres variables del bloque historia de violencia (violencia presenciada en la infancia, violencia experimentada por la pareja durante su propia infancia y violencia sexual sufrida antes de los 15 años) concentran la mayor contribución predictiva del modelo.

En contraste, las variables estructurales, relacionales y territoriales (como el IDH municipal, el hacinamiento del hogar, la edad de la pareja y las tasas estatales de no denuncia) presentan contribuciones más modestas, aunque sistemáticas. Esta jerarquía sugiere que, frente a antecedentes de violencia familiar, los factores contextuales y socioeconómicos modulan el riesgo, pero no constituyen predictores dominantes por sí solos.

Más que señalar características individuales aisladas, este patrón evidencia la centralidad del entorno doméstico de origen en la configuración del riesgo y respalda empíricamente la noción de transmisión intergeneracional de la violencia documentada en la literatura (Herrenkohl et al., 2022).

Un hallazgo que requiere especial cautela es la asociación positiva entre el IDH y la probabilidad estimada de exposición a violencia infantil. Esta dirección, aparentemente contraintuitiva frente a los enfoques que vinculan desigualdad estructural y violencia, no debe interpretarse como evidencia de que el IDH constituye un factor de riesgo.

Una explicación plausible es que refleje diferencias en la disposición o capacidad para identificar y reportar experiencias de violencia infantil entre municipios con distintos niveles de institucionalidad, escolaridad y acceso a información.

En conjunto, los resultados no solo son consistentes con las teorías feministas, sino que la refuerzan al mostrar que la violencia experimentada en la infancia se configura en entramados familiares y estructurales que exceden las trayectorias individuales y demandan respuestas igualmente estructurales.

Las políticas de prevención, atención, sanción y erradicación de la violencia contra las mujeres deben incorporar explícitamente la dimensión intergeneracional documentada en este estudio, reconociendo que los hogares son espacios políticos en los que se reproducen jerarquías de género, y que la prevención de la violencia en la vida adulta requiere intervenir sobre trayectorias de exposición temprana.

La articulación entre aprendizaje automático explicable y marcos interpretativos feministas abre la posibilidad de aproximaciones empíricas capaces de dialogar con la complejidad de los fenómenos sociales sin reducirlos a regularidades estadísticas descontextualizadas.

Limitaciones

El presente estudio presenta algunas limitaciones que deben considerarse al interpretar los resultados. En primer lugar, la naturaleza transversal de la ENDIREH 2021 impide establecer relaciones causales entre

los predictores analizados y la exposición a violencia durante la infancia; por tanto, los patrones identificados por el modelo deben interpretarse en términos de asociación y capacidad predictiva, y no como relaciones de causa y efecto.

En segundo lugar, el análisis se realizó a partir de una base de datos secundaria previamente integrada y procesada, por lo que las decisiones adoptadas durante las etapas originales de vinculación de fuentes, imputación, y depuración de observaciones no pudieron ser controladas directamente en este estudio.

Finalmente, debido a que la variable objetivo se construye a partir del reporte retrospectivo de experiencias ocurridas antes de los 15 años, no puede descartarse posibles diferencias en el recuerdo, reconocimiento o reporte de dichas experiencias.

Referencias

Casique, I., & Castro, R. (2019). Cambios y constantes en los niveles y factores asociados a las violencias de parejas en México. En R. Castro, (Ed.), De parejas, hogares, instituciones y espacios comunitarios: Violencias contra las mujeres en México (ENDIREH 2016) (pp. 161–270). Universidad Nacional Autónoma de México.

Consejo Nacional de Evaluación de la Política de Desarrollo Social, CONEVAL. (2020). *Medición de la pobreza: Índice de Gini 2020*. https://www.coneval.org.mx/Medicion/Paginas/Indice-de-Gini.aspx

Consejo Nacional de Población, CONAPO. (2020). *Índices de marginación 2020.* https://www.gob.mx/conapo/documentos/indices-de-marginacion-2020-284372

Fawcett, T. (2006, june). An introduction to ROC analysis. *Pattern Recognition Letters, 27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Granados-Martínez, A. (2022, septiembre). Interseccionalidad en salud, pobreza y vulnerabilidad. Papeles de Población, 28(111), 137–161. https://doi.org/10.22185/24487147.2022.111.06

Guedes, A., Bott, S., García-Moreno, C., & Colombini, M. (2016, june). Bridging the gaps: A global review of

intersections of violence against women and violence against children. Global Health Action, 9(1), 31516. https://doi.org/10.3402/gha.v9.31516

Herrenkohl, T. I., Fedina, L., & Mason, W. A. (2022). Child maltreatment, youth violence, intimate partner violence, and elder mistreatment: A review and theoretical analysis of research on violence across the life course. *Trauma, Violence, & Abuse, 23*(1), 314–328. https://doi.org/10.1177/1524838020939119

Instituto Nacional de Estadística y Geografía, I.N.E.G.I. (2021c). *Estadísticas de migración 2021*. https://www.inegi.org.mx/temas/migracion/

Instituto Nacional de Estadística y Geografía, I.N.E.G.I. (2019). *Encuesta Nacional de Calidad e Impacto Gubernamental 2019.* https://www.inegi.org.mx/programas/encig/2019/

Instituto Nacional de Estadística y Geografía, I.N.E.G.I. (2020a). *Censo de Población y Vivienda 2020*. https://www.inegi.org.mx/programas/ccpv/2020/

Instituto Nacional de Estadística y Geografía, I.N.E.G.I. (2020b). *Censo Nacional de Gobiernos Municipales y Demarcaciones Territoriales de la Ciudad de México 2020*. https://www.inegi.org.mx/programas/cngmd/2020/

Instituto Nacional de Estadística y Geografía, I.N.E.G.I. (2021a). *Encuesta Nacional de Victimización y Percepción sobre Seguridad Pública 2021*. https://www.inegi.org.mx/programas/envipe/2021/

Instituto Nacional de Estadística y Geografía, I.N.E.G.I. (2021b). *Encuesta Nacional sobre la Dinámica de las Relaciones en los Hogares 2021.* https://www.inegi.org.mx/programas/endireh/2021/

Instituto Nacional de Estadística y Geografía, I.N.E.G.I. (2021d). *Estadísticas de mortalidad: Defunciones por homicidio 2021.* https://www.inegi.org.mx/temas/mortalidad/

Kim, J., Lee, B., & Farber, N. B. (2019, december). Where do they learn violence? The roles of three forms of violent socialization in childhood. *Children and Youth Services Review,*

107. https://doi.org/10.1016/j.childyouth.2019.104494

Krug, E. G., Mercy, J. A., Dahlberg, L. L., & Zwi, A. B. (2002, october). The world report on violence and health. *The Lancet, 360*(9339), 1083–1088. https://doi.org/10.1016/S0140-6736(02)11133-0

Lagarde, M. (1996). *Género y feminismo: Desarrollo humano y democracia*. Instituto de la Mujer.

Lagarde, M. (2008). Antropología, feminismo y política: Violencia feminicida y derechos humanos de las mujeres. En M. L. Bullen & M. C. Díez (Coords.), *Retos teóricos y nuevas prácticas* (pp. 209–240). Ankulegi Antropologia Elkartea.

Ley General de Acceso de las Mujeres a una Vida Libre de Violencia [L.G.A.M.V.L.V.], reformada, Diario Oficial de la Federación [DOF], 1 de febrero de 2007. (México).

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In I. Guyon, U. von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. 30, pp. 4765–4774). Curran Associates.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S. I. (2020, january). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence, 2*(1), 56–67. https://doi.org/10.1038/s42256-019-0138-9

Programa de las Naciones Unidas para el Desarrollo, P.N.U.D. (2023, 15 de febrero). *Informe de Desarrollo Humano Municipal 2010–2020: Una década de transformaciones locales en México*. https://www.undp.org/es/mexico/publicaciones/informe-de-desarrollo-humano-municipal-2010-2020-una-decada-de-transformaciones-locales-en-mexico-0

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: Unbiased boosting with categorical features. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, & R. Garnett (Eds.), *Advances in neural information*

processing systems (Vol. 31, pp. 6638–6648). Curran Associates.

Saito, T., & Rehmsmeier, M. (2015, march). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE, 10*(3), e0118432. https://doi.org/10.1371/journal.pone.0118432

Segato, R. L. (2003). *Las estructuras elementales de la violencia*. Universidad Nacional de Quilmes.

Strasser, C., & Haensch, A. C. (2025, july). Mexico 2021: Psychological intimate partner violence against women and the role of childhood violence exposure: A machine learning approach. *[Preprint]. arXiv*. https://doi.org/10.48550/arXiv.2507.22592

Torres, J. A., & Martínez-Zarzoso, I. (2022, february). Determinants of emotional intimate partner violence against women and girls with children in Mexican households: An ecological framework. *Journal of Interpersonal Violence, 37*(23–24), NP22704–NP22731. https://doi.org/10.1177/08862605211072179

Anexos

Tabla 2.

Comparación del desempeño del modelo CatBoost bajo distintos criterios de selección del umbral.

	Validación					Prueba			
Criterio	Umbral	Exact	Sens	Esp	F1	Exact	Sens	Esp	F1
C1	0.500	0.720	0.578	0.815	0.623	0.720	0.587	0.809	0.627
C2	0.300	0.682	0.816	0.593	0.672	0.678	0.822	0.581	0.671
C3	0.390	0.710	0.723	0.701	0.666	0.702	0.728	0.685	0.661
C4	0.485	0.721	0.596	0.804	0.631	0.718	0.606	0.792	0.632
Nota. C1: criterio por defecto; C2: maximización del puntaje F1; C3: maximización de la exactitud balanceada; C4: maximización sujeta a una especificidad ≥ 0.80; Exact: exactitud; Sens: sensibilidad; Esp: especificidad; F1: puntaje F1.									

Fuente: elaboración propia.

Tabla 3.

Desempeño del modelo en los conjuntos de entrenamiento, validación y prueba bajo dos umbrales de clasificación.

Conjunto	Umbral	Exactitud	Precisión	Sens	Esp	F1	AUC
Entrenamiento	0.500	0.804	0.793	0.691	0.880	0.738	0.880
Validación	0.500	0.719	0.674	0.577	0.814	0.622	0.777
Prueba	0.500	0.720	0.671	0.587	0.808	0.626	0.771
Entrenamiento	0.485	0.806	0.784	0.710	0.870	0.745	0.880
Validación	0.485	0.720	0.669	0.596	0.803	0.630	0.777
Prueba	**0.485**	**0.717**	**0.660**	**0.605**	**0.792**	**0.631**	**0.771**
Nota: Exact: exactitud; Sens: sensibilidad; Esp: especificidad; F1: puntaje F1							

Fuente: elaboración propia.