

# Predicción de IPV en México

*Propuesta de investigación*

**Camila Riancho**

**Diego Fasan**

**Ronny M. Condor**

Profesora: Noelia Romero

Asistente: Victoria Oubiña

# 1. Introducción

La violencia doméstica es un fenómeno social que afecta a comunidades en todo el mundo, trascendiendo barreras culturales, económicas y geográficas. La Organización Mundial de la Salud (2022, 1) define a la violencia de pareja como todo comportamiento dentro de una relación íntima que causa daño físico, sexual o psicológico, incluidos actos de agresión física, coerción sexual, abuso psicológico y conductas controladoras". A nivel mundial, se calcula que el 27 % de las mujeres de entre 15 y 49 años que tuvieron pareja alguna vez han sido víctimas de violencia física y/o sexual por parte de sus parejas en algún momento de sus vidas, y 13 % lo ha sido en los últimos 12 meses. En el caso de América Latina, estas cifras se reducen ligeramente a 25 % y 8 %, respectivamente (WHO, 2018).

La violencia doméstica es un problema muy relevante porque genera perjuicios de corto y largo plazo en la salud física, mental y sexual de las mujeres que son víctimas, así como también en la salud y el bienestar presente y futuro de sus hijos. Además, la violencia de pareja también genera costos sociales y económicos para la comunidad en general (WHO, 2021). Por lo tanto, estudiar este problema en profundidad es de gran relevancia.

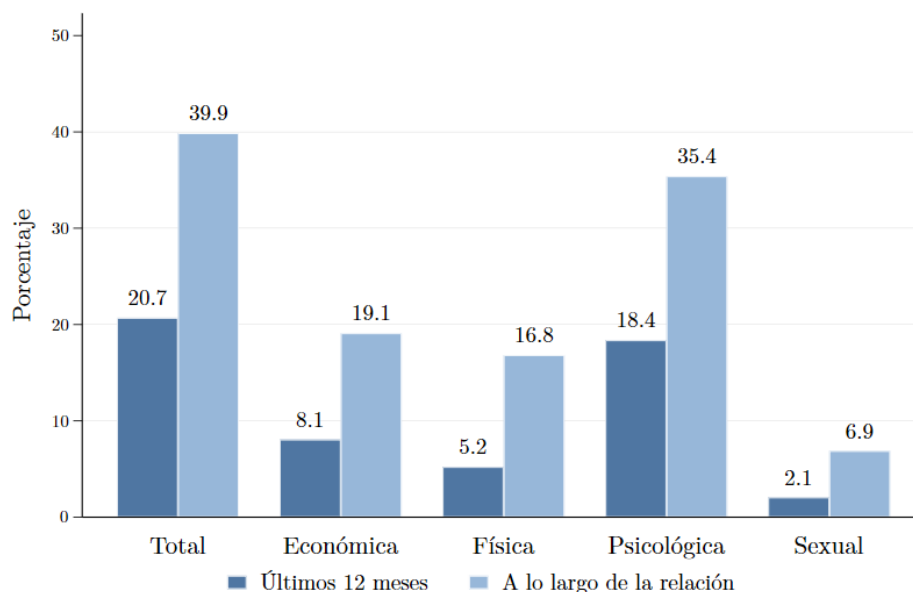
Uno de los grandes inconvenientes que surgen al momento de intentar prevenir y combatir la violencia doméstica es el hecho de que en su mayor parte ocurre en secreto. A modo de ejemplo, puede mencionarse que de acuerdo a una encuesta en Canadá de 2019 solo el 19 % de las víctimas de violencia por parte de sus parejas reportaron el incidente en la policía (Government of Canada, 2022). De manera similar, solo el 21,1 % de las mujeres que sufren violencia doméstica en Argentina hicieron una denuncia (Ministerio de las Mujeres, 2022), y este número se reduce a 13,1 % para el caso de México (INEGI, 2021).

Por lo tanto, identificar a aquellas mujeres que sufren en secreto es una necesidad fundamental para cualquier policy-maker preocupado por combatir la violencia doméstica. Con respecto a esta cuestión, los avances de los últimos años vinculados a Big Data y Machine Learning pueden ser de gran utilidad, ya que brindan la posibilidad de aprovechar los datos cada vez más numerosos con los que contamos para realizar predicciones sobre la probabilidad de que una mujer sea víctima de violencia. No obstante, al evaluar los avances que se han hecho en este sentido, podemos notar que muchos de estos trabajos que utilizan estos métodos se basan en datos de denuncias policiales, los cuales representan una pequeña porción de los casos de violencia totales (Berk et al., 2020; Rodríguez-Rodríguez et al., 2020). Por su parte, los trabajos que utilizan otras fuentes de datos más amplios, como encuestas de hogares, para predecir

el riesgo de que una mujer sea víctima de violencia de parte de su pareja, introducen como predictores variables muy directamente vinculadas a violencia o que indagan sobre cuestiones muy íntimas de la vida de las mujeres (por ejemplo, si la mujer fue víctima de abuso en su niñez, si es víctima de violencia por parte de otras personas que no son su pareja, si ella o su pareja consumen drogas, etc.).

En el presente trabajo, proponemos brindar un aporte a esta literatura desde una perspectiva de política pública. En particular, nos preguntamos si es posible predecir que una mujer está siendo víctima de violencia por parte de su pareja en base a predictores “no sensibles”, que no indaguen directamente en cuestiones de violencia o comportamientos muy íntimos de las mujeres, sino que se centren en características sociodemográficas y de estructura del hogar y la pareja, así como también opiniones de las mujeres sobre roles de género, recursos sociales de las mujeres, y cuestiones vinculadas a la distribución de tareas y decisiones dentro del hogar. Creemos que, en caso de poder desarrollar un modelo que tenga un buen desempeño para predecir violencia en base a este tipo de preguntas, en un segundo paso se podría construir un cuestionario que, al estar basado en preguntas no tan sensibles, podría aplicarse en escuelas, hospitales u otros ámbitos y permitir identificar mujeres en riesgo de estar sufriendo violencia a un nivel más desagregado que el que permiten las encuestas anónimas a nivel nacional que preguntan por violencia directamente.

**Figura 1:** Violencia en el ámbito de pareja

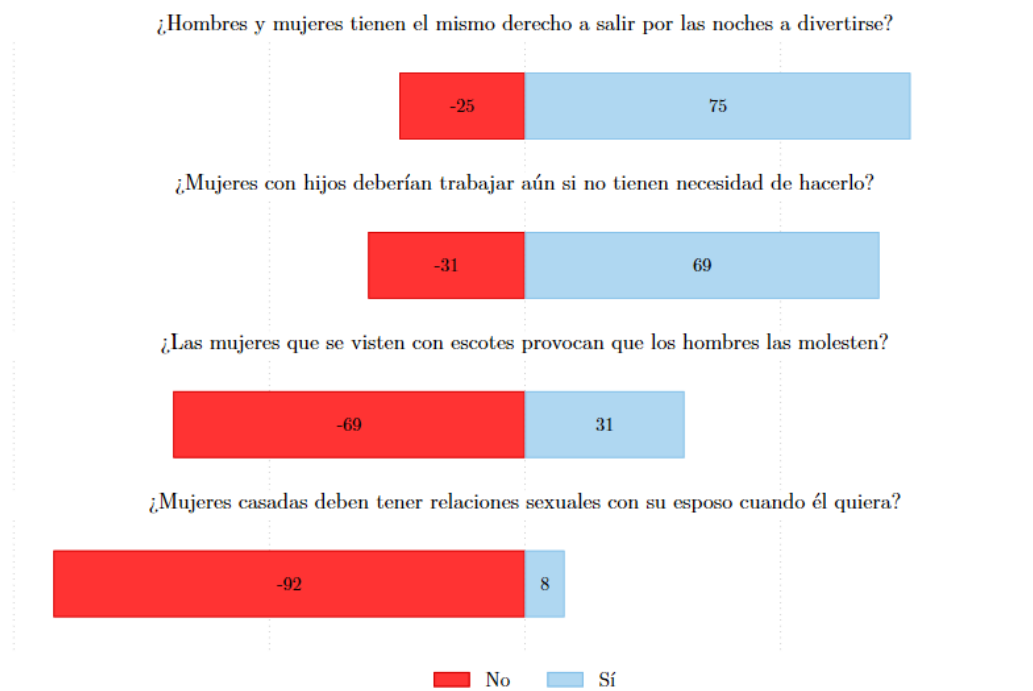


*Fuente: ENDIREH 2021. Elaboración propia.*

Con este objetivo en mente, utilizaremos la Encuesta Nacional sobre la dinámica de las Relaciones

en los Hogares (ENDIREH) de 2021 en México para clasificar las mujeres que sufren violencia de parte de sus parejas. Se trata de una encuesta con un gran número de observaciones (140.784 viviendas), que incluye muchas preguntas referidas a violencia. Respecto a la prevalencia de la violencia de pareja, de acuerdo a la ENDIREH de 2021, el 39,9 % de las mujeres mayores a 15 años reportan haber experimentado algún tipo de violencia (psicológica, física, sexual y/o económica) de parte de su pareja en algún momento de su relación, mientras que 20,7 % reportan haber vivido al menos una situación de violencia de parte de su pareja en el último año. Al momento de evaluar los distintos tipos de violencia por separado, se observa que el 18,4 % de las mujeres sufrieron violencia psicológica de parte de sus parejas en el último año, 5,2 % sufrió violencia física, 2,1 % sufrió violencia sexual y 8,1 % experimentó violencia económica (véase figura 1).

**Figura 2: Creencias**



*Fuente: ENDIREH 2021. Elaboración propia.*

Por otro lado, además de información sobre violencia, la ENDIREH incluye un conjunto amplio de variables que pueden usarse como predictores de la violencia de pareja. Por ejemplo, la encuesta indaga sobre las creencias de las mujeres sobre los roles de género. En particular, una de cada cuatro mujeres encuestadas considera que hombres y mujeres no tienen el mismo derecho a salir de noche a divertirse. Asimismo, el 31 % de encuestadas considera que las mujeres con escote provocan a que los hombres las molesten. (véase figura 2). Esto nos da una idea de las creencias de las mismas mujeres, que podrían correlacionar, a su vez, con experiencias de

violencia en la pareja ([Capaldi et al., 2012](#)).

El trabajo se organiza de la siguiente manera. En la sección 2 presentamos la literatura previa relevante. En el apartado 3, realizamos una breve descripción de la base de datos. En la sección 4, presentamos la metodología que seguiremos, y en el último apartado desarrollamos las conclusiones y posibles limitaciones del estudio.

## 2. Literatura previa

La violencia doméstica es una problemática que ha sido ampliamente estudiada desde una diversidad de disciplinas y puntos de vista. En particular, un gran número de trabajos han estudiado la correlación entre la probabilidad de sufrir violencia de género y distintos factores. Entre las variables que la literatura asocia con la violencia de pareja pueden mencionarse factores sociodemográficos de las mujeres, como la religión, un menor nivel de ingreso, un menor nivel educativo, estar desempleadas, ser jóvenes y ser divorciadas o separadas ([Abramsky et al., 2011](#); [Capaldi et al., 2012](#); [Rahman et al., 2023](#)). También se han identificado como relevantes a variables vinculadas a la estructura de la pareja, que correlacionan positivamente con la violencia doméstica, tales como el hecho de que la mujer cohabite con su pareja sin estar casada formalmente, una mayor diferencia de edad entre el hombre y la mujer, una mayor brecha salarial en favor del hombre, que la pareja sea reciente (menor a 5 años), que se hayan casado jóvenes, que la mujer tenga hijos de relaciones pasadas ([Aizer, 2010](#); [Abramsky et al., 2011](#); [Capaldi et al., 2012](#); [Jensen et al., 2023](#)). Asimismo, se han identificado factores de riesgo vinculados a comportamientos y creencias tanto de la mujer como de su pareja, en particular, que su pareja le haya sido infiel, que su pareja tenga problemas de conducta, comportamientos antisociales o una baja autoestima, que ella y/o su pareja estén sometidos a estrés o depresión, que discutan y tengan conflictos recurrentes, que experimenten celos, que consuman drogas y/o alcohol, que tengan actitudes machistas o crean que la violencia de género está justificada, y que la mujer tenga poco poder de decisión dentro del hogar ([Abramsky et al., 2011](#); [Amusa et al., 2020](#); [Capaldi et al., 2012](#); [McDougal et al., 2021](#); [Rahman et al., 2023](#); [Valdez-Santiago et al., 2013](#)). Más aún, es importante tener en cuenta que pareciera existir cierta transmisión intergeneracional de la violencia, en el sentido de que se han observado asociaciones positivas entre que una mujer sea víctima de violencia por parte de su pareja, que su madre o su suegra hayan sido víctimas de violencia doméstica y que ella o su pareja hayan sido víctimas de violencia en su niñez ([Abramsky et al., 2011](#); [Capaldi et al., 2012](#); [Valdez-Santiago et al., 2013](#)). Del mismo

modo, las mujeres que sufren violencia doméstica de parte de sus parejas también son víctimas de violencia en otros ámbitos de parte de otras personas [Abramsky et al. \(2011\)](#); [McDougal et al. \(2021\)](#). Por último, es interesante mencionar que contar con un entorno social de apoyo (tener interacciones frecuentes con amigos, participar de actividades sociales, contar con apoyo de parte de los padres y vecinos, etc.) está asociada a una menor probabilidad de ser víctima de violencia en el hogar ([Capaldi et al., 2012](#)).

Más allá de estudiar las correlaciones de la violencia doméstica con distintas variables, es interesante y necesario explorar si a partir de estos y otros factores es posible predecir el riesgo de que una mujer sea víctima de violencia por parte de su pareja. En este sentido, hay distintos tipos de estudios vinculados a IPV que han hecho uso de modelos de Machine Learning para hacer clasificaciones y predicciones vinculadas a violencia doméstica.

A modo de ejemplo, una literatura interesante que ha surgido en los últimos años es la que utiliza técnicas de análisis computacional de texto (Natural Language Processing) para identificar contenido vinculado a violencia doméstica en reportes de la policía y posteos en redes sociales ([Castorena et al., 2021](#); [Hui et al., 2023](#); [Neubauer et al., 2021](#)), o incluso que desarrollan modelos en base a imágenes de heridas y golpes para clasificar víctimas de violencia doméstica ([Majumdar et al., 2018](#)).

Asimismo, hay trabajos que utilizan datos administrativos del sistema policial y judicial para predecir el número de denuncias que se presentarán en tribunales en un futuro cercano en una zona determinada ([Rodríguez-Rodríguez et al., 2020](#)), o que se basan en los datos contenidos en las denuncias policiales por violencia de género para predecir que la mujer presentará una segunda denuncia de mayor gravedad en un futuro cercano ([Berk et al., 2020](#)).

La desventaja de estudiar y predecir la IPV en base a datos de denuncias es que, como presentamos previamente, las mujeres que denuncian la violencia doméstica son un bajo porcentaje del total de mujeres que experimentan dicha violencia, y además es muy posible que dicho subgrupo de mujeres no sea representativo de las demás víctimas de IPV. Por ende, otras fuentes de datos en base a las cuales es posible estudiar esta cuestión son las encuestas, como la ENDIREH en México, o la Demographic Health Survey (DHS), que incluyen preguntas sobre violencia doméstica, pero cuyos protocolos de anonimato, privacidad y confidencialidad reducen los problemas de sub-reporte. Hay un importante número de investigaciones que aplicaron modelos de machine learning a este tipo de encuestas para predecir que una mujer sea víctima de violencia de parte de su pareja ([Amusa et al., 2020](#); [McDougal et al., 2021](#); [Rahman et al., 2023](#)).

Nuestra propuesta se apoya sobre estas investigaciones previas, pero busca brindar un aporte orientado a la política pública. En particular, estos trabajos que han buscado predecir la violencia de pareja en base a encuestas suelen utilizar como predictores preguntas sensibles, que es esperable que las mujeres no estén dispuestas a contestar en un contexto no anónimo, tales como el hecho de consumir drogas, haber estado expuesta a otros tipos de violencia, que su madre haya sido víctima de violencia de género, variables muy específicas sobre su vida sexual, etc. No obstante, dado el amplio número de factores que correlacionan con la violencia doméstica, es útil preguntarnos sobre la posibilidad de identificar mujeres que sufren violencia en base a variables no sensibles, vinculadas a cuestiones sociodemográficas y características de la relación de pareja y de la vida social de la mujer. Dicho modelo podría ser una herramienta relevante para policy-makers interesados en identificar mujeres que sufren violencia doméstica, ya que podría permitir la construcción de cuestionarios que puedan ser realizados a mujeres en contextos no anónimos, en los que presumiblemente muchas no estarían dispuestas a contestar preguntas que tratan directamente sobre violencia o que indagan sobre experiencias muy íntimas.

Con este objetivo en mente, el presente trabajo utilizará la encuesta ENDIREH de México de 2021 para intentar predecir la probabilidad de que una mujer haya experimentado violencia por parte de su pareja en el último año, utilizando como predictores variables sociodemográficas, aquellas vinculadas a la estructura del hogar, y a opiniones y actividades de la mujer, que no tratan directamente sobre violencia. La decisión de centrarnos en dicha encuesta se explica por el gran número de observaciones que tiene la muestra (140.784 viviendas), por la amplia variedad de información que ofrece sobre distintos tipos de violencia, y por el hecho de que, además de las típicas variables sociodemográficas, incluye otras preguntas que presumiblemente pueden funcionar como buenos predictores de violencia doméstica, por ejemplo, opinión sobre roles de género, división de tareas y toma de decisiones dentro del hogar, así como también información referida a los recursos sociales de la mujer (actividades sociales de las que participa, si cuenta con el apoyo de familia, amigos y vecinos, etc.). Además, si bien la ENDIREH se ha usado para estudiar el efecto de ciertos factores o eventos, por ejemplo las leyes de divorcio, sobre la violencia ([García-Ramos, 2021](#)), o para predecir la violencia en subgrupos particulares de la población, como pueblos originarios ([Valdez-Santiago et al., 2013](#)), a la fecha no parece haber sido utilizada para predecir violencia de pareja en la población en general a partir de predictores no sensibles, que es el objetivo de nuestro trabajo. A continuación, en la siguiente sección se presenta y describe en mayor detalle esta base de datos.

### 3. Base de datos

Para realizar el presente estudio utilizaremos la Encuesta Nacional sobre la Dinámica de las Relaciones en los Hogares (ENDIREH) de México, la cual es realizada por el Instituto Nacional de Estadística y Geografía (INEGI) a mujeres mayores de 15 años. En sus 5 ediciones (2003, 2006, 2011, 2016, 2021) ofrece información auto reportada sobre situaciones de violencia física, económica, sexual, psicológica y patrimonial a mujeres mayores de 15 años a lo largo de su vida y en los últimos 12 meses. Además, presenta datos sobre lugares donde fueron agredidas y las personas que ejercieron la violencia.

Asimismo, además de preguntas referidas a violencia, la encuesta incluye preguntas vinculadas a características de la vivienda y el hogar, variables sociodemográficas de la mujer y su pareja, opiniones sobre roles de género, división de tareas y decisiones en el hogar, acceso a la salud, y recursos sociales (actividades fuera del hogar y red de apoyo en situaciones de emergencia) . La tabla 1 resume las secciones que consideramos relevantes *ex-ante* para el objetivo de nuestra investigación, ya que incluyen variables que pueden correlacionar con violencia de pareja, pero que no tratan específicamente de violencia en otros ámbitos, ni de comportamientos sexuales de las mujeres, ni de consumo de drogas y sustancias. Muchas de las preguntas contenidas en estas secciones podrían resultar relevantes a la hora de predecir la violencia intra-hogar, y además son no tan invasivas y lo suficientemente sutiles como para que no sea evidente que al preguntarlas se está intentando indagar sobre cuestiones de violencia.

**Cuadro 1:** Principales secciones de la encuesta que usaremos para extraer los predictores

Nombre de Sección	Título
Sección I	Características de la vivienda y hogares de la vivienda
Sección II	Características sociodemográficas de residentes de la vivienda
Sección IV	Ingresos y roles de pareja
Sección VI	Opinión sobre roles de género
Sección VIII	Ámbito laboral
Sección X	Acceso a la salud obstétrica y partos
Sección XVI	Recursos sociales
Sección XVII	División de trabajo intra hogar
Sección XV	Decisiones y libertad personal

La encuesta aborda la violencia de manera específica, explorando diversos ámbitos como el escolar, laboral, comunitario y familiar, junto con aspectos educativos, laborales, lugares frecuentados y dinámicas familiares. En el presente trabajo, buscaremos predecir únicamente si una mujer ha sido víctima de al menos una situación de violencia física y/o sexual de parte de su pareja en



los últimos 12 meses. Asimismo, dado que estamos interesados en estudiar la violencia reciente de parte de la pareja, consideraremos únicamente a aquellas mujeres que están casadas o unidas, no las que están solteras o divorciadas.

Como se mencionó anteriormente, en esta problemática existe mucho subreporte por diversos motivos, desde vergüenza al declarar, amenazas, etc. En ese sentido, el INEGI trata de asegurarse de que las mujeres respondan la encuesta estando solas con la encuestadora, en un ambiente de confianza y privacidad. En la recolección del 2021, el 97 % de las mujeres respondieron la encuesta en total privacidad desde el inicio hasta el final de la entrevista. Para nuestra propuesta de trabajo, nos enfocaríamos en estas respuestas únicamente, para evitar los errores de medición que podrían tener lugar si las mujeres que sufrieron violencia no se animaron a declararlo en la encuesta porque su marido las estaba escuchando.

## 4. Metodología

En el presente trabajo, buscaremos clasificar a las mujeres casadas o unidas en función de si experimentaron o no violencia física o sexual de parte de su pareja en el último año. Para eso, nuestra variable dependiente será una variable binaria que tome valor 1 si la mujer reporta haber sufrido en el último año al menos 1 de las situaciones de abuso físico o sexual por las que se pregunta en la encuesta (golpes, bofetones, disparos, heridas con arma blanca, aventón de objeto, exigirle tener relaciones sexuales, etc.) y 0 en caso contrario.

Como predictores, usaremos todas las variables de las secciones presentadas en el apartado anterior, excluyendo algunas preguntas en particular que son sensibles. Por ejemplo, de la sección de atención obstetricia se incluirá la pregunta sobre el seguro de salud, pero no sobre el número de abortos ni sobre las situaciones de violencia experimentadas en el parto. Del mismo modo, de la sección “Ámbito Laboral” se incluirán solo las preguntas referidas a la ocupación de la mujer, pero no vinculadas a las situaciones de violencia que experimentó en el trabajo. Las variables categóricas las convertiremos a variables binarias. Además, crearemos algunas otras variables relevantes en base a la literatura revisada, tales como la diferencia de la edad entre la mujer y su pareja, una dummy para clasificar si la mujer gana más o no y el valor de esa diferencia, cantidad de personas por habitación del hogar, entre otras.

Los métodos que usaremos para hacer la predicción son métodos basados en árboles (CART, Random Forest, Bagging y Gradient Boosting). La decisión de usar este tipo de métodos es que permiten realizar predicciones en contextos de no-linealidades e interacciones entre variables

independientes, y la literatura revisada (Capaldi, 2012) sugiere que hay un número importante de factores que parecen correlacionar con violencia de pareja, no de manera lineal, sino condicional a la presencia de otros factores. Para estimar estos modelos, tomaremos al 80 % de la muestra como el grupo de entrenamiento y al 20 % restante como grupo de testeo. Sobre el grupo de entrenamiento aplicaremos cross validation para optimizar los hiperparámetros de los distintos modelos (profundidad, número de árboles, número de predictores, etc.) y luego nos quedaremos con el modelo que obtenga una mejor predicción.

Al momento de evaluar el desempeño de los distintos modelos, hay que tener en cuenta que, dada la naturaleza de nuestro problema, contamos con un desbalance en los datos. Como vimos en la figura 1, la proporción de mujeres que experimentaron violencia física o sexual es mucho menor que la proporción de mujeres que no la sufrieron. Este desbalance muestral puede tener un impacto negativo en el rendimiento del modelo, ya que podría sesgar la predicción hacia la clase mayoritaria. Dado este problema, conocido como *class imbalance* en la literatura de machine learning, usar el accuracy para evaluar el desempeño del modelo puede no ser tan informativo (Satpathy, 2020). Es por eso que para evaluar la performance de los modelos que estimemos tendremos en cuenta también el área bajo la curva ROC (AUC), recall (porcentaje de las mujeres que sufren violencia que son clasificadas correctamente) y el F1, que es una medida que combina precision (proporción de todos los clasificados como positivos que efectivamente lo son) y recall.

Asimismo, también aplicaremos una solución propuesta por Chawla et al. (2002) para los casos donde existe class imbalance, que es la de Synthetic Minority Oversample Technique (SMOTE). Esta técnica consiste en generar datos sintéticos de la muestra subrepresentada, en nuestro caso, mujeres que sufrieron violencia en los últimos 12 meses, para poder una muestra más equilibrada<sup>1</sup>.

Finalmente, seleccionaremos el modelo con el mejor desempeño, bajo las distintas métricas que analizaremos, e identificaremos a los predictores más importantes.

## 5. Conclusiones y limitaciones

El objetivo de nuestra propuesta de investigación es predecir la violencia doméstica en México a partir de preguntas no tan invasivas. Para lograr este objetivo, utilizamos datos de la Encuesta

---

<sup>1</sup>Definitivamente, hay más detalles técnicos detrás. Por ejemplo, algo crucial es que solo se use cross-validation únicamente para los datos de train. Además, el grado de oversampling (muestras sintéticas creadas) también es clave y se podría probar con distintas configuraciones para encontrar alguna que sea óptima.

Nacional sobre la Dinámica de las Relaciones en los Hogares (ENDIREH), una encuesta representativa a nivel nacional que contiene datos demográficos y sobre prácticas culturales, roles de género y creencias asociadas con el empoderamiento femenino, facilitando así la predicción de ocurrencias de violencia de pareja. Optamos por métodos basados en árboles debido a su capacidad para capturar relaciones no lineales y su resistencia al sobreajuste. Esta elección se sustenta en la literatura previa que destaca las interacciones entre el IPV y distintos factores.

Nuestra propuesta tiene además serias implicaciones de política. En base a los resultados, podríamos identificar aquellas preguntas que contribuyen mejor a la predicción de hogares donde existe violencia. Esto es relevante para la creación de sistemas de alerta temprana, los cuales podrían prevenir casos de violencia. Si bien reconocemos que la ENDIREH tiene muchos objetivos adicionales, si nos concentráramos exclusivamente en predecir IPV, se podría diseñar un cuestionario menos extenso para identificar los hogares en riesgo, a partir de estas preguntas no tan sensibles que las mujeres estén más dispuestas a responder. Esto es especialmente útil en subgrupos específicos, como personas migrantes, población indígena, o otros contextos sub-representados.

Una posible limitación de nuestro estudio es que es probable que la variable violencia de pareja tenga error de medición, ya que posiblemente hay un porcentaje de mujeres que sufre violencia pero optó por no decirlo. En ese caso, puede ser más difícil que el modelo genere una buena predicción. De todas maneras, si bien difícilmente lo solucione por completo, el hecho de usar solamente aquellas encuestas que se realizaron en privado con la mujer debería reducir este problema.

## Referencias

- Abramsky, T., Watts, C. H., Garcia-Moreno, C., Devries, K., Kiss, L., Ellsberg, M., Jansen, H. A., and Heise, L. (2011). What factors are associated with recent intimate partner violence? findings from the who multi-country study on women's health and domestic violence. *BMC public health*, 11(1):1–17.
- Aizer, A. (2010). The gender wage gap and domestic violence. *American Economic Review*, 100(4):1847–1859.
- Amusa, L., Bengesai, A., and Khan, H. (2020). Predicting the vulnerability of women to intimate

- partner violence in south africa:evidence from tree-based machine learning techniques. *Journal of Interpersonal Violence*, 37:1–18.
- Berk, A., , and Sorenson, S. (2020). Algorithmic approach to forecasting rare violent events. *Criminology and Public Policy*, 19(1):213–233.
- Capaldi, D. M., Knoble, N. B., Shortt, J. W., and Kim, H. K. (2012). A systematic review of risk factors for intimate partner violence. *Partner abuse*, 3(2):231–280.
- Castorena, C., Abundez, I., Alejo, R., Granda-Gutiérrez, E., and Rendón, E.and Villegas, O. (2021). Deep neural network for gender-based violence detection on twitter messages. *Mathematics*, 9:807.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357.
- García-Ramos, A. (2021). Divorce laws and intimate partner violence: Evidence from mexico. *Journal of Development Economics*, 150:102623.
- Government of Canada, S. C. (2022). The Daily — Victims of police-reported family and intimate partner violence in Canada.
- Hui, V., Eby, M., Constantino, R. E., Lee, H., Zelazny, J., Chang, J. C., He, D., and Lee, Y. J. (2023). Examining the supports and advice that women with intimate partner violence experience received in online health communities: text mining approach. *Journal of medical internet research*, 25:e48607.
- INEGI (2021). Encuesta Nacional sobre la Dinámica de las Relaciones en los Hogares: Principales resultados.
- Jensen, S. K., Murray, S. M., Placencio-Castro, M., Kajani, U., Amponsah, D., Sezibera, V., and Betancourt, T. S. (2023). Family violence reduction within a parenting intervention in rwanda: a mixed-methods study. *Pediatrics*, 151(Supplement 2).
- Majumdar, P., Chhabra, S., Singh, R., and Vatsa, M. (2018). On detecting domestic abuse via faces. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 2173–2179.
- McDougal, L., Dehingia, N., Bhan, N., Singh, A., McAuley, J., and Raj, A. (2021). Opening closed doors: using machine learning to explore factors associated with marital sexual violence in a cross-sectional study from india. *BMJ Open*, 11(12):1–9.

- Ministerio de las Mujeres, G. y. D. (2022). Encuesta de prevalencia de violencia contra las mujeres.
- Neubauer, L., Straw, I., Mariconti, E., and Tanczer, L. M. (2021). A systematic literature review of the use of computational text analysis methods in intimate partner violence research. *Mathematics*, 9:807.
- Rahman, M., Jamil, K., Nahar, Q., Chakraborty, N., Haider, M. M., and Khan, S. (2023). Factors that provide protection against intimate partner physical violence among married adolescents in bangladesh. *Frontiers in public health*, 11:1125056.
- Rodríguez-Rodríguez, I., Rodríguez, J.-V., Pardo-Quiles, D.-J., Heras-González, P., and Chatzigiannakis, I. (2020). Modeling and forecasting gender-based violence through machine learning techniques. *Applied Sciences*, 10(22):8244.
- Satpathy, S. (2020). Smote for imbalanced classification with Python.
- Valdez-Santiago, R., Híjar, M., Martínez, R. R., Burgos, L. Á., and Monreal, M. d. l. L. A. (2013). Prevalence and severity of intimate partner violence in women living in eight indigenous regions of mexico. *Social Science & Medicine*, 82:51–57.
- WHO (2018). *Violence against women prevalence estimates*. World Health Organization, Geneva.
- WHO (2021). *Violence against women*.