

## **Predicción de la violencia severa contra las mujeres usando Random Forest y LASSO: un enfoque ecológico del caso peruano**

### **I. Introducción**

De acuerdo con la Organización Mundial de la Salud (OMS, 2021, p. 2), casi 1 de cada 3 mujeres en el mundo ha sufrido de violencia doméstica alguna vez en su vida. En el Perú, alrededor del 60% de mujeres ha experimentado algún tipo de violencia por parte de su esposo o compañero (Instituto Nacional de Estadística e Informática [INEI], 2019a), y alrededor del 30% ha experimentado violencia física en particular (INEI, 2019b). Más recientemente, entre los años 2020 y 2022, se registraron 860 casos de feminicidios y tentativas de feminicidio (Defensoría del Pueblo, 2020, 2021, 2022). Si bien estas situaciones pueden compartir muchas de sus causas, el grado en el que la violencia se presenta en cada una de ellas varía, y también lo hace su frecuencia. Así, mientras la *violencia de pareja* es un fenómeno relativamente común entre las mujeres, lo que se identifica como *feminicidios* y sus tentativas afectan a menos del 1% de la población femenina. Así también, aunque todos los grados de violencia de pareja resultan condenables, no todos poseen la misma dinámica. En otras palabras, el término “violencia de pareja” tiende a aglomerar diferentes escenarios; desde situaciones espontáneas de agresión verbal, hasta potenciales intentos de homicidio. En este sentido, afinar la identificación de las distintas formas de violencia de pareja puede representar una mejora sustancial en su abordaje.

Por lo mismo, diversos autores han indagado acerca de los factores explicativos o predictores de la de violencia de pareja, en sus distintos grados. Mientras algunos estudios han priorizado distinguir entre los tipos de violencia<sup>1</sup>, otros se han centrado en proponer o desarrollar escalas sobre su intensidad (Echebura et al., 2009; Gnisci y Pace, 2016; Vatnar y Bjørkly, 2013). La OMS (2012), por ejemplo, distingue un subtipo de violencia de pareja caracterizado por la iteración y alta intensidad de múltiples formas de abuso (*battering*). También se ha evaluado como violencia severa a aquella que deja consecuencias que van, desde moretones, hasta lesiones que requieren intervención quirúrgica (Vatnar y Bjørkly, 2013)<sup>2</sup>. Al mismo tiempo, la literatura subraya cada vez más la relevancia de análisis que integren los diferentes niveles de factores de esta problemática (Heise, 1998; Gnisci y Pace, 2016; Hernández et al., 2018). La mayoría de estos trabajos emplea un enfoque de inferencia causal, y son bastantes menos los que utilizan un enfoque predictivo con herramientas de *Machine Learning*<sup>3</sup>. Una de las grandes ventajas de este enfoque son sus algoritmos de selección automática de variables, que alcanzan una mayor precisión que la selección *humana* o manual (Kan et al., 2019; Ludwing et al., 2015).

En este sentido, la presente investigación se plantea identificar las variables capaces de predecir la violencia severa de pareja contra las mujeres, utilizando los algoritmos de Machine Learning *Random Forest* y regresión LASSO. En la medida que se busca integrar y distinguir el peso de las distintas dimensiones de factores asociados, en línea con el enfoque socio-ecológico de la violencia de Heise (1998), se emplean variables a nivel distrital e individual. Las bases de datos a usar son el Censo Poblacional (2017), la Encuesta Demográfica y de Salud Familiar y el Registro Nacional de Municipalidades, a cargo del INEI. Se trabajará con data transversal agrupada de los años 2016-2019 para estos dos últimos. Se espera que el resultado sea un insumo para la elaboración de métricas más finas sobre violencia de pareja en el país, que no requieran consultar directamente sobre la incidencia de este fenómeno - el cual, por su nivel de sensibilidad, suele llevar a un problema de sub-reporte de casos (Johnson et al., 2008; McCleary-Sills et al., 2016).

---

<sup>1</sup> Para el caso del Perú, -entre otros- Andía (2019), Díaz y Miranda (2010) y Mendoza (2017).

<sup>2</sup> A este tipo de violencia, que, pese a no ser formalmente identificada como un intento de feminicidio, amenaza la vida o la integridad de las mujeres, es al que se alude con el término de “violencia severa” en este trabajo.

<sup>3</sup> Entre estos, Hossain y colegas (2021) y Amusa y colegas (2022).

## II. Revisión de literatura

Entre la literatura empírica referente a violencia de pareja (VP), es posible identificar diferentes enfoques a partir del tipo de factores que se consideran en el análisis de la problemática. Esta sección se propone dar luces sobre la dinámica de la violencia severa de pareja<sup>4</sup>, reconociendo la importancia de las características de la relación entre sus miembros, así como las del entorno inmediato y la sociedad en las que esta se sitúa.

El estudio de los factores de la relación de pareja indaga en la predicción de la violencia (severa o letal) especialmente a partir de indicadores de violencia previa, por lo general, basados en reportes policiales o judiciales. Las investigaciones en esta línea han logrado validar empíricamente su premisa teórica respecto a que los feminicidios íntimos y los episodios de violencia severa -fenómenos con alto grado de violencia- están vinculados a la historia de violencia de la pareja; y, por ende, es posible estimar su ocurrencia a partir de esta última (Berk et al., 2016; Cavlin, 2020; Echebura et al., 2009; López-Ossorio et al., 2015). Para el caso español, por ejemplo, Echebura y otros (2009) plantearon un instrumento de evaluación del riesgo de violencia severa y feminicidio de pareja, y hallaron como factores (ítems) más empíricamente significativos a -por parte del agresor- la intención de generar daños severos, la amenaza con objetos potencialmente peligrosos, la presencia de celos o comportamiento controlador, la justificación de la violencia, y -por parte de la víctima- la percepción de estar en peligro de muerte. En una línea similar, pero empleando herramientas de *machine learning*, Berk y otros (2016, para Estados Unidos) identificaron que la incidencia de arrestos previos pronostica una mayor probabilidad de nuevos arrestos por violencia doméstica, y alcanzaron una eficacia del 90% en su predicción de infractores no-reincidentes de esta.

Por otro lado, también se ha desarrollado bibliografía que integra explicaciones sociales-estructurales y sociopsicológicas en su planteamiento. En cuanto a literatura internacional, Hornung y otros (1981) evalúan los patrones de logro educativo y ocupacionales en parejas heterosexuales como factores de riesgo de abuso conyugal, en base a su teoría de incompatibilidad de estatus. Entre una muestra representativa de EE. UU. hallaron que las parejas experimentan una mayor incidencia de violencia conyugal cuando la mujer presenta un sobre-logro laboral respecto a su pareja masculina, o un menor nivel educativo que este. Estudios del caso peruano respaldan estos hallazgos. En particular, se ha identificado una relación positiva entre el hecho de que una mujer trabaje y sufra VP, al propiciar situaciones de conflicto entre los miembros (Andía, 2019; Castro y Delgado, Díaz y Miranda, 2010; 2014; Durán, 2019; Mendoza, 2017; Hernández, 2015 - con indicadores regionales como el porcentaje de población femenina económicamente activa). Dentro de este grupo, el desempleo del esposo/compañero y el bajo nivel educativo propio operan como factores de riesgo para las mujeres (Durán, 2019; Mendoza, 2017), mientras que la educación superior en ambos miembros de la pareja es un factor protector (Mendoza, 2017; Valdivia, 2013).

En la misma línea, otros autores han postulado explicaciones macrosociales de la violencia a partir de la Teoría de la Desorganización Social (*Social Disorganization Theory*). Esta sostiene que los contextos con determinadas características son más propensos a tener tasas más altas de criminalidad por su reducida capacidad de ejercer control social formal o informal entre su población (ver, por ejemplo, Sampson et al., 1997). Para explicar la importancia de las redes sociales informales en estos contextos, este enfoque, además, subraya la debilidad institucional que los caracteriza (como dificultad en el acceso a sistemas de protección para las mujeres violentadas (McCleary-Sills et al., 2016)). Pinchevsky y Wright realizan una revisión de literatura al respecto, e identifican que la *desventaja social concentrada* -aterrizada en la presencia conjunta de variables como la proporción de población afroamericana, pobre, con educación básica, que

---

<sup>4</sup> En situaciones en las cuales un varón agrede a una mujer.

conforman hogares monoparentales, beneficiaria de ayuda pública, hacinada y desempleada, entre otros- se relaciona directa y significativamente con la incidencia de VP de tipo letal (2012, p. 124). La tolerancia social a la violencia, asimismo, se vincula con una mayor incidencia de casos de violencia no letal (Browning, 2002; Koenig et al., 2006; Wright & Benson, 2010). En cambio, la eficacia colectiva -cohesión social y capacidad de intervención por el bien común (Sampson et al., 1997)- tiene un efecto protector (aunque limitado) sobre ambas. Benavides y colegas (2019) aplican esta aproximación en el Perú, aterrizada en las variables de normas de género (comportamiento controlador por parte del esposo) y capital social, y encuentran que los contextos con menor incidencia de VP se caracterizan por, pese a su desventaja social, poseer lazos sociales más fuertes entre sus habitantes.

Una aproximación similar y bastante holística es la del enfoque socio-ecológico de la violencia contra las mujeres, que propone la integración de factores en 4 niveles: la historia personal, el microsistema, el exosistema y el macrosistema. Heise (1998) en base a una extensa revisión de literatura propone algunos factores explicativos tentativos (no necesariamente causales, pero sí empíricamente relevantes). Estos son (para la historia personal) haber presenciado violencia marital, haber sufrido abuso de niña, y el rechazo por parte del padre; (microsistema) el dominio o el control económico masculino en la familia, el consumo de alcohol, y el conflicto verbal; (exosistema) un bajo estatus socioeconómico/laboral, el aislamiento de la mujer y la familia, y la interacción con infractores; (macrosistema) masculinidades agresivas, roles de género rígidos, y la aceptación de la violencia interpersonal/castigos físicos (Heise, 1998, p. 265). Hernández y colegas (2018) aplican este enfoque para el caso peruano, e identifican como determinantes de la violencia con riesgo de feminicidio al haber sido alguna vez víctima de violencia sexual y el justificar la violencia contra las mujeres. Además, encuentran que “el efecto de los predictores aumenta cuando la desventaja social en el distrito es mayor” (Hernández et al., 2018, p. 85). Los autores operacionalizan este concepto en una variable compuesta de indicadores del Censo de Población y Vivienda peruano (2007), a nivel distrital, complementando la información a nivel de mujeres/hogares de la ENDES.

En la literatura de Machine Learning, dos investigaciones han considerado predictores en estos niveles. Amusa y otros (2020) identificaron predictores asociados al riesgo de experimentar VP para mujeres en Sudáfrica, en base a 3 técnicas basadas-en-árboles: el miedo al esposo/compañero, las actitudes hacia la violencia, los antecedentes de abuso, y el consumo de alcohol o drogas por parte del esposo/compañero. Hossain y otros (2021) se enfocaron en predecir la violencia doméstica (contra cualquier miembro de la familia) durante la pandemia COVID-19, y sus predictores más importantes fueron el nivel de ingresos familiar durante la pandemia y el nivel educativo de los miembros de la familia.

En cuanto a la distinción del grado de VP, la OMS (2012) distingue un subtipo caracterizado por la iteración y alta intensidad de múltiples formas de abuso (*battering*). Por su parte, autores como Vatnar y Bjørkly (2013) han elaborado una escala de VP considerando ejemplos y consecuencias específicas para el abuso en 4 niveles: moderadamente severo, severo, muy severo y extremadamente severo. Las consecuencias a partir del segundo nivel van desde moretones y depresión, hasta hemorragias internas e intentos de suicidio (Vatnar y Bjørkly, 2013, p. 778). En la presente investigación, la *violencia severa* ha sido operacionalizada como una variable dicotómica que refiere al hecho de haber experimentado, alguna vez, por el esposo o compañero, alguno de 4 escenarios específicos de violencia física que, en línea con la literatura, pueden afectar la integridad o poner en riesgo la vida de las mujeres. Estos son: haber sido pateada o arrastrada; estrangulada o quemada (o un intento de alguno); amenazada con cuchillo, pistola u otra arma; y atacada con un cuchillo, pistola u otra arma (ver Tabla 1).

### **III. Metodología**

Podemos entender al aprendizaje estadístico como un “amplio set de herramientas para interpretar la información” (James et al., 2021: 16, traducción propia). El aprendizaje automático o *Machine Learning* opera bajo un enfoque predictivo, cuyo objetivo es identificar los parámetros que minimicen el error de generalización para nuevas observaciones. Con esto, a diferencia del enfoque de inferencia causal, ninguna de estas herramientas puede determinar la manera en que se relacionan las variables de *entrada* y *salida*. No obstante, una de las grandes ventajas de este enfoque son sus algoritmos de selección automática de variables, que alcanzan una mayor precisión que la selección *humana* o manual (Kan et al., 2019; Ludwing et al., 2015).

Con el objetivo de integrar los factores del hogar y su contexto al modelar la predicción de violencia severa, la presente investigación empleará 3 bases de datos del INEI: la Encuesta Demográfica y de Salud Familiar (ENDES), el Censo de Población y Vivienda, y el Registro Nacional de Municipalidades (RENAMU). Esto implica que, si bien la unidad de análisis *final* es la mujer, se incluirán variables con información del distrito de residencia para cada caso, como se ha hecho en investigaciones previas (por ejemplo, Benavides et al., 2019; y Hernández et al., 2018). Se trabajará, además, con un modelo transversal agrupado (*pooled cross sectional model*) de muestras de los años 2016-2019; con casos nuevos cada año en la ENDES y el RENAMU, y con casos del año 2017 en el Censo. La data será procesada y analizada en el programa R Studio.

La ENDES recoge información a nivel nacional sobre temas en salud reproductiva, materna e infantil, además de violencia familiar. Sus muestras son representativas por área geográfica (urbana o rural), regiones naturales y departamentos del país. En el período estudiado, el número de viviendas encuestadas varió entre 35 910 (2016) y 36 760 (2019) (INEI, 2016a, 2019). Las mujeres en unión conyugal (casadas o en unión consensual), sobre quienes se aplicará la variable de interés, sin embargo, solo conforman una parte de esta muestra. En el 2016, por ejemplo, representaron al 57,5% de encuestadas (INEI, 2016a). Los módulos que se utilizarán para el estudio refieren a las características básicas de la encuestada (módulo 66), la composición del hogar (módulo 64 y 67) y la violencia familiar (módulo 73).

Por otro lado, el Censo de Población y Vivienda, por su propia naturaleza, posee resultados estadísticamente representativos a nivel distrital, además de nacional y departamental. En su edición del año 2017 fueron censadas 29 381 884 personas, en 1 874 distritos (INEI, 2018). Las secciones que se utilizarán para este estudio son las de “Características y servicios de la vivienda”, “Características del hogar” y “Características de la población”. La data será obtenida del *Sistema de Consulta de Base de Datos* del INEI, que facilita el tratamiento de la información por distrito. El RENAMU, por su parte, tiene como unidad de análisis a las municipalidades provinciales y distritales, y su muestra también es su población total. Entre los años estudiados, entre 1 655 (2016) y 1 678 (2019) municipalidades distritales cumplieron con la entrega de su formulario (INEI, 2016b, 2019c). Se consultará, en específico, el módulo de “Servicios Públicos Locales”, que servirá para medir conceptos como la eficacia colectiva y el capital social y el nivel de presencia institucional. En R Studio, la información de las distintas bases de datos será unida a partir del ubigeo distrital, presente como variable en cada una de estas.

En la medida en que la variable de interés clasifica a la muestra en dos grupos muy desiguales - por ejemplo, en 2018, la situación más común de violencia severa se presentó en menos del 10,0% de los casos (INEI, 2019b)-, se aplicará un balanceo muestral. Estudios previos en *machine learning* han demostrado que el balanceo muestral puede incrementar sustancialmente la precisión de los modelos (Batista et al., 2004; Bellinger et al., 2018; Elreedy y Atiya, 2019; Hossain et al., 2021). Para este propósito, se aplicará la Técnica de Sobremuestreo de Minorías Sintéticas (o *SMOTE*, por sus siglas en inglés). Esta consiste en generar casos adicionales de la clase minoritaria a partir de otros casos cercanos entre sí (*K-nearest neighbors*), seleccionados

aleatoriamente. En cuanto a la aplicación en R Studio, se empleará la función “SMOTE” de la librería “DMwR”.

Para estimar la predicción de la variable de interés se utilizarán dos algoritmos no-paramétricos orientados a la estimación de la importancia de variables. El primero consiste en los *bosques aleatorios* o *Random Forests* (RF), una técnica basada en la agregación de múltiples árboles de clasificación (*ntree* en la librería “randomForest” de R Studio). Cada árbol opera con un número de predictores  $m$  (*mtry*, en R Studio) elegidos al azar, que equivale a la raíz del número total de predictores  $p$  (James et al., 2021). Cada árbol impone un voto unitario para *sintetizar* el subconjunto de predictores en la categoría más común de la variable de interés, y el error de este ejercicio puede medirse utilizando el índice de Gini. Los resultados de los RF son interpretables como una selección de las variables más importantes para predecir la categoría de la variable dependiente. Con esto, el algoritmo ayuda a seleccionar los predictores necesarios para un modelo parsimonioso (Genuer et al., 2010). Por lo mismo (y otras razones) diversas investigaciones respaldan su efectividad por sobre otros métodos basados-en-árboles (Amusa et al., 2020; Fuerte-Celis y Sánchez-Castañeda, 2021; Genuer et al., 2010; Hossain et al., 2021).

El rendimiento del RF será comparado con el de una segunda herramienta: la regresión LASSO (por las siglas en inglés para Operador de Selección y Contracción Mínima Absoluta), que predice apropiadamente variables continuas (Ludwing et al., 2015; Kan et al., 2019; Ruck et al., 2022), pero también funciona para variables discretas. Este tipo de regresión -en R Studio, comandos “glmnet” y “predict” de la librería “glmnet” -opera añadiendo un término de penalidad o parámetro de tuneo (*lambda*) a la estimación de mínimos cuadrados ordinarios; de manera que aminora el peso de cada predictor, y puede llevar a que algunos sean iguales a 0 (James et al., 2021). Con esto, al igual que RF, mitiga el riesgo de *overfitting*, al tiempo que incrementa la interpretabilidad de sus resultados (en comparación a regresiones como la Ridge). En la medida en que esta investigación espera observar el mayor número de predictores posible, se escogerá el menor valor de *lambda* que no perjudique el poder predictivo del algoritmo (la serie de opciones se obtiene con la función *seq* en R Studio).

Antes de ser procesada, la muestra balanceada será dividida entre un conjunto de entrenamiento y un conjunto de testeo. Para mejorar el rendimiento predictivo, tanto para las estimaciones de LASSO como de RF, se utilizará además una herramienta de validación cruzada en 10-grupos (para ambos pasos, se usa el comando “trainControl” de la librería “caret” de R Studio).

Es así como, se espera que el algoritmo desarrollado en esta investigación determine las variables más importantes para estimar la predicción de la violencia severa. En detalle, esto implica distinguir los factores de la historia personal, las características del hogar, las características del distrito y las normas culturales que -aunque a través de vías solo presumibles- estarían interviniendo en un tipo de violencia particularmente aguda. Para honrar en su utilidad para las políticas públicas, es preciso primero reconocer sus limitaciones. Y es que, a diferencia de otras herramientas de Machine Learning que resultan directamente alternativas más eficientes para medir un atributo<sup>5</sup>, este algoritmo está entrenado con variables de una fuente que ya ofrece una medición no-sesgada de la variable de interés. En términos más simples, no existe una encuesta nacional que mida solo los predictores de violencia severa, sin medir la violencia severa en sí misma. En este sentido, la presente investigación representaría una contribución a la elaboración de métricas más finas sobre VP en el país (y otros países con un contexto similar). Tales nuevas métricas no tendrían que consultar directamente por temas tan sensibles como la violencia severa, ya que serían capaces de inferirla a partir de los otros factores. De esta manera, se atenuaría el

---

<sup>5</sup> Ver, por ejemplo, para el caso de la pobreza a Jean et al (2016) o Chagalj (2019).

problema de sub-reporte de casos, frecuente en este tipo de estudios (Johnson et al., 2008; McCleary-Sills et al., 2016).

#### IV. Anexos

[Estas tablas son un mapeo o borrador preliminar de las variables que me interesara considerar, planeo terminar de desarrollarlas conforme avance con la investigación]

<b>Tabla 1: Variables dependientes y predictivas en la ENDES (2016-2019)</b>	
Nombre	Descripción
Violencia severa (dependiente)	Identifica (1 = sí, 0 = no) el hecho de que la encuestado haya experimentado alguna vez por su esposo o compañero: un intento de estrangularla o quemarla; una amenaza con cuchillo, pistola u otra arma; o un ataque con cuchillo, pistola u otra arma.
Ser casada	Identifica (1 = sí, 0 = no) si la encuestada
Convivencia	Identifica (1 = sí, 0 = no) si la encuestada
Primaria completa, mujer	Identifica (1 = sí, 0 = no) si la encuestada
Secundaria completa, mujer	Identifica (1 = sí, 0 = no) si la encuestada
Educación superior, mujer	Identifica (1 = sí, 0 = no) si la encuestada
Primaria completa, varón	Identifica (1 = sí, 0 = no) si el esposo o compañero de la encuestada
Secundaria completa, varón	Identifica (1 = sí, 0 = no) si el esposo o compañero de la encuestada
Educación superior, varón	Identifica (1 = sí, 0 = no) si el esposo o compañero de la encuestada
Educación básica, pareja	Identifica (1 = sí, 0 = no) si ambos miembros de la pareja
Educación superior, pareja	Identifica (1 = sí, 0 = no) si ambos miembros de la pareja
Educación superior, solo varón	
Educación superior, solo mujer	
Desempleo femenino	Identifica (1 = sí, 0 = no) si la encuestada
Desempleo masculino	Identifica (1 = sí, 0 = no) si el esposo o compañero de la encuestada
Tipo de trabajo: mano de obra calificada	Identifica (1 = sí, 0 = no) si la encuestada
Tipo de trabajo: mano de obra no calificada	Identifica (1 = sí, 0 = no) si la encuestada
Mujer gana más que varón	
Varón gana más que mujer	
Área geográfica	Identifica (1= urbano, 0= rural) si el hogar se ubica en un área urbana o rural.
Pobreza del hogar	Identifica (1= sí, 0= no) si el hogar es pobre.
Número de hijos	Describe el número de hijos de la encuestada
Padre le pegó alguna vez a madre	Identifica (1 = sí, 0 = no) si la encuestada

Pareja ejerció violencia psicológica (situaciones humillantes o amenazas)	Identifica (1 = sí, 0 = no) si la encuestada
Pareja ejerció violencia física no-severa ¿Desagregar?	Identifica (1 = sí, 0 = no) si la encuestada alguna vez fue empujada, sacudida o se le tiró algo; abofeteada o retorcida del brazo; o golpeada con el puño o algo que pudo dañarla, por parte de su esposo o compañero
Consumo de alcohol por parte del esposo o conviviente	Identifica (1 = sí, 0 = no) si el esposo o compañero de la encuestada
Pareja limitó su contacto con amistades	
Pareja limitó su contacto con familiares	
Pareja insiste en saber a dónde va	
Alguna vez fue víctima de violencia sexual	
Índice de justificación de violencia	Que justifique la violencia contra la mujer en cinco situaciones: salir sin avisarle al esposo; descuidar a los hijos; discutir con el esposo; rehusarse a tener relaciones sexuales con la pareja; y quemar la comida
Violencia contra los niños	Creencias sobre la violencia contra los niños (aceptación del castigo físico)

Tabla 2: Variables predictivas en el Censo de Población y Vivienda (2017)	
Nombre	Descripción
Mujeres con seguro de salud	Porcentaje de
Pobreza distrital	Porcentaje de hogares pobres en el distrito
Población mayor de edad	Porcentaje de
Población con castellano como lengua materna	Porcentaje de
Población residencialmente estable	Porcentaje de
Población desempleada	Porcentaje de
Población con secundaria completa	Porcentaje de
Jefas de hogar	Porcentaje de
Hogares beneficiarios de programas sociales	Porcentaje de
Hacinamiento	Porcentaje de viviendas en el distrito con

Tabla 3: variables explicativas en el RENAMU (2016-2019)	
Nombre	Descripción
Comisaría	Distrito cuenta con comisaría
Policías por habitante	Número de policías por habitante
Centro de Emergencia Mujer	Distrito cuenta con CEM
Establecimiento de Salud	Distrito cuenta con establecimiento de salud

Organización social predominantemente femenina	Presencia de alguna organización social de mujeres (Club de madres, Vaso de Leche, Cocina Comunal, o Wawa Wasi) / plataforma de participación vecinal en el distrito
Otro tipo de organización social	

## V. **Bibliografía**

Amusa, L. B., Bengesai, A. V., & Khan, H. T. (2022). Predicting the vulnerability of women to intimate partner violence in South Africa: Evidence from tree-based machine learning techniques. *Journal of interpersonal violence*, 37(7-8), NP5228-NP5245. <https://doi.org/10.1177/0886260520960110>

Björkly, S., & Vatnar, S. K. (2013). Intimate partner violence: A review of the literature. *Journal of perinatal education*, 22(4), 206-215. <https://doi.org/10.1891/1058-1243.22.4.206>

Batista, G. E., Prati, R. C., & Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD explorations newsletter*, 6(1), pp. 20-29. <https://doi.org/10.1145/1007730.1007735>

Bellinger, C., Drummond, C., & Japkowicz, N. (2018). Manifold-based synthetic oversampling with manifold conformance estimation. *Machine Learning*, 107, 605-637.

Benavides, M., León, J., Etesse, M., Espezúa, L., & Stuart, J. (2019). Exploring the association between segregation and physical intimate partner violence in Lima, Peru: the mediating role of gender norms and social capital. *SSM-population health*, 7, 100338. <https://doi.org/10.1016/j.ssmph.2018.100338>

Berk, R. A., Sorenson, S. B., & Barnes, G. (2016). Forecasting domestic violence: A machine learning approach to help inform arraignment decisions. *Journal of Empirical Legal Studies*, 13(1), pp. 94-115. <https://doi.org/10.1111/jels.12098>

Browning, C. R. (2002). The span of collective efficacy: Extending social disorganization theory to partner violence. *Journal of Marriage and Family*, 64(4), 833-850. <https://doi.org/10.1111/j.1741-3737.2002.00833.x>

Castro, E. & Delgado, M. (2014). Determinantes de la violencia doméstica física, sexual & emocional contra la mujer en el Perú: 2012. En Casas, C. & Grimaldi, D. (Ed.), *Economía aplicada Ensayos de investigación económica 2014* (2da ed., pp. 11-34). Fondo Editorial Universidad del Pacífico.

Çavlin, A. (2020). Women in Danger of Femicide: Predicting Population at Risk of Femicide from Intimate Partner Violence Survivors in Turkey. *Journal for Women's Studies*, 21(1), pp. 1–20. <https://doi.org/10.33831/jws.v21i1.77>

Defensoría del Pueblo. (2021) ¿Qué pasó con ellas? Reporte Igualdad y No Violencia [número 11]. <https://www.defensoria.gob.pe/deunavezportodas/wp-content/uploads/2021/07/Reporte-11-Qu%C3%A9-pas%C3%B3-con-ellas.pdf>

Defensoría del Pueblo. (2022) ¿Qué pasó con ellas? Reporte Igualdad y No Violencia [número 23]. <https://www.defensoria.gob.pe/deunavezportodas/wp-content/uploads/2022/04/Reporte-23.pdf>



Defensoría del Pueblo. (2023) ¿Qué pasó con ellas? Reporte Igualdad y No Violencia número 35.  
<https://www.defensoria.gob.pe/wp-content/uploads/2023/01/Reporte-Igualdad-y-No-Violencia-Anual.pdf>

Díaz, R., & Miranda, J. (2010). Aproximación del costo económico y determinantes de la violencia doméstica en el Perú. Lima: Centro de Investigación Económica & Social (CIES).  
<https://cies.org.pe/sites/default/files/investigaciones/aproximacion-del-coste-economico-y-determinantes-de-la-violencia-domestica-en-el-peru.pdf>

Echeburua, E.; Fernandez-Montalvo, J.; de Corral, P.; Lopez-Goni, J. J. (2009). Assessing Risk Markers in Intimate Partner Femicide and Severe Violence: A New Assessment Instrument. *Journal of Interpersonal Violence*, 24(6), pp. 925–939.

Elreedy, D., & Atiya, A. F. (2019). A comprehensive analysis of synthetic minority oversampling technique (SMOTE) for handling class imbalance. *Information Sciences*, 505, 32-64.

Gnisci, A., & Pace, A. (2016). Lethal domestic violence as a sequential process: Beyond the traditional regression approach to risk factors. *Current Sociology*, 64(7), 1108–1123.  
<https://doi.org/10.1177/0011392116629809>

Hornung, C. A., McCullough, B. C. & Sugimoto, T. (1981). Status Relationships in Marriage: Risk Factors in Spouse Abuse. *National Council on Family Relations*, 43(3), 675-692. Recuperado de <https://psycnet.apa.org/record/1982-07969-001>

Hossain, M. M., Asadullah, M., Rahaman, A., Miah, M. S., Hasan, M. Z., Paul, T., & Hossain, M. A. (2021). Prediction on domestic violence in bangladesh during the covid-19 outbreak using machine learning methods. *Applied System Innovation*, 4(4), 77.

<https://doi.org/10.1016/j.ins.2019.07.070>

Instituto Nacional de Estadística e Informática. (2016a). Características generales de las mujeres.  
[https://www.inei.gob.pe/media/MenuRecursivo/publicaciones\\_digitales/Est/Lib1417/libro.pdf](https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1417/libro.pdf)

Instituto Nacional de Estadística e Informática. (2016b). Estadísticas Municipales.  
[https://www.inei.gob.pe/media/MenuRecursivo/publicaciones\\_digitales/Est/Lib1417/libro.pdf](https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1417/libro.pdf)

Instituto Nacional de Estadística e Informática. (2018). Perú: Perfil Sociodemográfico. Informe Nacional.  
[https://www.inei.gob.pe/media/MenuRecursivo/publicaciones\\_digitales/Est/Lib1539/libro.pdf](https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1539/libro.pdf)

Instituto Nacional de Estadística e Informática. (2019<sup>a</sup>). Violencia contra las mujeres, niñas y niños [https://www.inei.gob.pe/media/MenuRecursivo/publicaciones\\_digitales/Est/Endes2019/](https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Endes2019/)

Instituto Nacional de Estadística e Informática. (2019b). Perú: Indicadores de Violencia Familiar y Sexual, 2012-2019.  
[https://www.inei.gob.pe/media/MenuRecursivo/publicaciones\\_digitales/Est/Lib1686/libro.pdf](https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1686/libro.pdf)

Instituto Nacional de Estadística e Informática. (2019c). Perú: Indicadores de Gestión Municipal 2019. [https://www.inei.gob.pe/media/MenuRecursivo/publicaciones\\_digitales/Est/Lib1720/](https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1720/)

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An introduction to statistical learning: With applications in R (2nd ed.). Springer.

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301), 790-794.  
<https://doi.org/10.1126/science.aaf7894>

Johnson, H., Ollus, N., & Nevala, S. (2008). Violence against women: An international perspective. Springer Science & Business Media.

Koenig, M. A., Stephenson, R., Ahmed, S., Jejeebhoy, S. J., & Campbell, J. (2006). Individual and contextual determinants of domestic violence in North India. *American Journal of Public Health*, 96(1), 132-138. <https://doi.org/10.2105/AJPH.2004.057422>

López-Ossorio, J. J., González-Álvarez, J. L., & Andrés-Pueyo, A. (2016). Eficacia predictiva de la valoración policial del riesgo de la violencia de género. *Psychosocial Intervention*, 25(1), pp. 1-7. <https://doi.org/10.1016/j.psi.2015.10.002>

McCleary-Sills, J., Namy, S., Nyoni, J., Rweyemamu, D., Salvatory, A., & Steven, E. (2016). Stigma, shame and women's limited agency in help-seeking for intimate partner violence. *Global Public Health*, 11(1-2), 224-235. <https://doi.org/10.1080/17441692.2015.1115175>

Mendoza, Á. N. (2017). *El efecto del empleo sobre la violencia doméstica: evidencia para las mujeres peruanas* (Tesis de licenciatura). Recuperado de <https://hdl.handle.net/11042/3222>

Organización Mundial de la Salud. (2021). Violence Against Women Prevalence Estimates, 2018. <https://www.who.int/publications/i/item/9789240022256>

Pinchevsky, G. M.; Wright, E. M. (2012). The Impact of Neighborhoods on Intimate Partner Violence and Victimization. *Trauma, Violence, & Abuse*, 13(2), pp. 112-132. <https://doi.org/10.1177/1524838012445641>

Sampson, R. J., Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science*, 277(5328), 918-924. <https://doi.org/10.1126/science.277.5328.918>

World Health Organization. (2012). Understanding and addressing violence against women. [https://www.who.int/reproductivehealth/publications/violence/VAW\\_Report\\_Full\\_2012.pdf](https://www.who.int/reproductivehealth/publications/violence/VAW_Report_Full_2012.pdf)

Wright, E. M., & Benson, M. L. (2010). The relationship between neighborhood poverty and intimate partner violence: An examination of the social disorganization hypothesis. *Violence and Victims*, 25(5), 695-706. <https://doi.org/10.1891/0886-6708.25.5.695>