



El futuro
es de todos

Gobierno
de Colombia

Modelos de probabilidad de vulneración para la prevención de violencia contra niñas, niños y adolescentes en Colombia



Lina Arbeláez
Directora General

Liliana Pulido Villamil
Subdirectora General

**Directora de Planeación y
Control de la Gestión**
Amanda Castellanos Mendoza

Subdirector de Monitoreo y Evaluación
Néstor Fernando Zuluaga Giraldo

**Equipo Técnico de la Subdirección de
Monitoreo y Evaluación**
Edda Patricia Izquierdo López
Olga Lucía Corzo Velásquez

**Equipo Técnico del Grupo de Estadística
y Gestión de la Información (ICBF)**
Leidy Johana Lozano Lavao
Javier Rubio Sáenz
José D. Niño Muriel
Campo E. Pinillos

**Coordinación editorial
Jefe Oficina Asesora de
Comunicaciones**
Ximena Ramírez Ayala

**Corrección de estilo, diagramación
y diseño**
Grupo Imagen Corporativa

Edición
Junio 2022

Contenido

Introducción	6
I. Violencia contra niñas, niños y adolescentes en Colombia	8
1. Caracterización de la violencia contra niñas, niños y adolescentes	8
1.1. Tipos o formas de violencia	9
1.2. Contextos en los que se presenta la violencia	10
1.3. Agresores(as)	10
1.4. Daños que produce	11
2. Cifras de la violencia contra niñas, niños y adolescentes en Colombia	11
2.1 Registro de violencia contra niñas, niños y adolescentes	12
2.2 Ingresos de niñas, niños y adolescentes al Proceso Administrativo de Restablecimiento de Derechos (PARD) por motivo de violencia	15
2.3 Violencia contra niñas, niños y adolescentes durante la pandemia del COVID-19	19
II. Datos y metodología de los modelos de probabilidad de vulneración	22
Datos	23
1.1 Características de las niñas, niños y adolescentes	23
1.2. Características del hogar	25
1.3. Características de la vivienda	28
1.4. Características del municipio	29
2. Metodología de los Modelos de Probabilidad de Vulneración de Derechos de Niñas, Niños y Adolescentes	30

3.	Ética y discriminación	35
III.	Resultados	37
1.	Modelo de probabilidad de vulneración por violencia física, psicológica o negligencia (maltrato)	37
1.1	Desempeño general	37
1.2	Valores SHAP	41
1.3	Caracterización de los casos con probabilidad alta de ser vulnerados	43
1.4	Distribución geográfica de los casos con mayor probabilidad de vulneración	46
2.	Modelo de probabilidad de vulneración por violencia sexual	48
2.1	Desempeño general	48
2.2	Valores SHAP	50
2.3	Caracterización de los casos con probabilidad alta de ser vulnerados	52
2.4	Distribución geográfica de los casos con mayor probabilidad de vulneración	55
IV.	Conclusiones y recomendaciones	57
	Glosario	62
	Referencias	65
	Anexo 1. Detalle variables modelo	67
	Anexo 2. Proporción de casos con riesgo de vulneración por violencia física, psicológica o negligencia (maltrato) estimado alto para el total nacional y respecto a variables seleccionadas	81
	Anexo 3. Proporción de casos con riesgo de vulneración por violencia sexual estimado alto para el total nacional y respecto a variables seleccionadas	82

Lista de siglas y abreviaturas

AATI	Asociaciones de Autoridades Tradicionales Indígenas
ACURIS	Asociación de Comunidades Unidas del Río Isana y Cuyarí
CDC	Centros para el Control y la Prevención de Enfermedades de los Estados Unidos
DANE	Departamento Administrativo Nacional de Estadística
DIJIN	Dirección de Investigación Criminal e INTERPOL
DNP	Departamento Nacional de Planeación
EVCNNA	Encuesta de violencia contra niños, niñas y adolescentes
FGN	Fiscalía General de la Nación
ICBF	Instituto Colombiano de Bienestar Familiar
INMLCF	Instituto Nacional de Medicina Legal y Ciencias Forenses
INS	Instituto Nacional de Salud
MEN	Ministerio de Educación Nacional
ODS	Objetivos de Desarrollo Sostenible de las Naciones Unidas
OIM	Organización Internacional para las Migraciones
OMS	Organización Mundial de la Salud
PARD	Proceso Administrativo de Restablecimiento de Derechos
PPS	Organización Panamericana de la Salud
RLCPCD	Registro para la Localización y Caracterización de las Personas con Discapacidad
RUV	Registro Único de Víctimas
SIEDCO	Sistema de Información Estadístico, Delincuencial, Contravencional y Operativo de la Policía Nacional
SIMAT	Matrícula
SISPRO	Sistema Integrado de Información para la Protección Social
SIVIGILA	Sistema Nacional de Vigilancia en Salud Pública
SNBF	Sistema Nacional de Bienestar Familiar
SPOA	Sistema Penal Oral Acusatorio de la Fiscalía General de la Nación
UNICEF	Fondo de las Naciones Unidas para la Infancia
UNODC	Oficina de las Naciones Unidas contra el Delito

Introducción

De acuerdo con la Organización Mundial de la Salud (OMS) (2020), se estima que alrededor de 1.000 millones de niñas, niños y adolescentes son víctimas de algún tipo de violencia en todo el mundo cada año. En Colombia, de acuerdo con los resultados de la primera Encuesta de violencia contra niños, niñas y adolescentes (EVCNNA) (OIM, 2019), el 41,5 % de las personas entre 18 y 24 años reportaron haber sufrido algún tipo de violencia antes de cumplir 18 años, el 32,1 % física, el 15,2 % violencia psicológica y el 11,4 % violencia sexual. Adicionalmente, en el periodo 2012-2020, 171.157 niñas, niños y adolescentes ingresaron por motivo de violencia al Proceso Administrativo de Restablecimiento de Derechos (PARD) del Instituto Colombiano de Bienestar Familiar (ICBF). De este total, 80.050 niñas, niños y adolescentes ingresaron por motivo de violencia física, psicológica o negligencia (maltrato), y 91.107 por violencia sexual.

Dada la prevalencia de la violencia contra las niñas, niños y adolescentes en Colombia y el mundo, y los daños y afectaciones que causa a la integridad personal y al desarrollo integral de esta población (ICBF, 2017), esta ha sido reconocida como una violación a los derechos humanos en el marco normativo internacional y nacional. Diferentes Estados miembros de la Organización de las Naciones Unidas y organizaciones internacionales como el Fondo de las Naciones Unidas para la Infancia (Unicef), la Organización Panamericana de la Salud (OPS), la OMS, End Violence Against Children, Together for Girls y la Oficina de las Naciones Unidas contra el Delito (UNODC), entre otras, se han articulado para diseñar estrategias con el fin de prevenir la violencia contra las niñas, niños y adolescentes en todo el mundo.

Diferentes gobiernos se han sumado a las acciones que Colombia adelantan. En 2018, la administración de Iván Duque, en el Plan Nacional de Desarrollo 2018-2022 «Pacto por Colombia, pacto por la equidad», adoptó los Objetivos de Desarrollo Sostenible (ODS) de las Naciones Unidas y estableció como propósito «crear las condiciones para el desarrollo integral de las niñas, niños y adolescentes, y prevenir y resolver las situaciones de violencia y vulneración de derechos que los afectan» (p. 270). Para esto, estipuló la construcción de una estrategia nacional contra las violencias bajo el liderazgo del ICBF, con el fin de fortalecer la acción institucional en torno a la prevención de las violencias contra las niñas, niños y adolescentes en el país, al igual que adoptar herramientas tecnológicas para contribuir a la focalización de acciones.

Producto de lo anterior, en 2019 el ICBF creó **la Alianza Nacional contra la Violencia hacia Niñas, Niños y Adolescentes** por medio de la cual se realiza una convocatoria a todos los sectores del país para proteger a la población menor de 18 años contra todas las formas de violencia. Uno de los componentes de la alianza es producir, sistematizar y analizar conocimiento e información estratégica sobre los determinantes, manifestaciones y efectos de las formas de violencia contra niñas, niños y adolescentes, y adoptar herramientas tecnológicas para desarrollar acciones de priorización y focalización basadas en evidencia para proteger a los menores de edad.

Este boletín documenta el primer esfuerzo institucional en Colombia que brinda insumos

para la focalización de acciones de prevención de violencias contra las niñas, niños y adolescentes, mediante la generación de dos modelos de probabilidad de vulneración fundamentados en el uso de analítica de datos y herramientas de aprendizaje de máquinas (machine learning). Modelos similares son utilizados en países como Estados Unidos, Nueva Zelanda y Reino Unido para alinear mejor los recursos disponibles con los requerimientos de las comunidades, y así generar entornos más seguros para las niñas, niños, adolescentes y familias.

Los modelos desarrollados por el ICBF permiten estimar la probabilidad de que una niña, niño o adolescente registrado en la base de datos Sisbén IV pueda ser identificado como víctima de violencia física, psicológica y negligencia (maltrato) o de violencia sexual. El objetivo de los modelos es contribuir al diseño de intervenciones centradas en la prevención de la ocurrencia de violencias desde un enfoque intersectorial, dirigir los esfuerzos institucionales a los grupos poblacionales y territorios con mayor riesgo, y la posible identificación en terreno de casos de vulneración de derechos.

El modelo de probabilidad de vulneración por violencia física, psicológica o negligencia (maltrato) en promedio permite identificar de forma acertada a las niñas, niños y adolescentes con registro de derechos vulnerados y no vulnerados registrados en el Sisbén IV en el 84,2 % de los casos y el de probabilidad de vulneración por violencia sexual en el 83,8¹%. Las variables más importantes en la predicción de la probabilidad de vulneración por violencia física, psicológica y negligencia (maltrato), es decir, las que en su ausencia alterarían más la probabilidad estimada, son la edad, la privación por rezago escolar al interior del hogar (para el mismo u otro miembro del hogar) y estar estudiando. Por su parte, las más importantes para predecir la probabilidad de ocurrencia de violencia sexual son el sexo, la edad, y la privación por rezago escolar al interior del hogar. En términos absolutos y como proporción del total de niñas, niños y adolescentes en la base de Sisbén IV, Bogotá tiene una participación importante en los casos con riesgo estimado alto de vulneración por violencia física, psicológica o negligencia (maltrato) y por violencia sexual.

El boletín se divide en cuatro apartados. En primer lugar, se presenta el marco conceptual y las estadísticas sobre las violencias ejercidas contra la niñez y la adolescencia en el territorio nacional. El siguiente expone los datos y la metodología empleados para el desarrollo de los modelos de probabilidad de vulneración. En el tercero se muestran los resultados obtenidos frente a la probabilidad de que una niña, niño o adolescente registrado en la base de datos del Sisbén IV sea identificado como víctima de violencia física, psicológica o negligencia (maltrato) o violencia sexual. Finalmente, en el cuarto, se presentan las principales conclusiones y recomendaciones dirigidas a los agentes del Sistema Nacional de Bienestar Familiar (SNBF) y al Subsistema de Protección de Derechos para emplear la información generada por los modelos de probabilidad de vulneración con el fin de focalizar esfuerzos y prevenir la violencia contra las niñas, niños y adolescentes.

1. Estos valores corresponden a la medida de Balanced Accuracy en el umbral de riesgo que maximiza su valor. Para mayor información ver la sección 1.1 Desempeño general en el título III. Resultados.



I. Violencia contra niñas, niños y adolescentes en Colombia

De acuerdo con el Lineamiento técnico del ICBF para la atención de niños, niñas y adolescentes con derechos inobservados, amenazados o vulnerados por causa de la violencia, la violencia contra esta población es «toda acción, omisión, abuso, uso de la fuerza o del poder que se expresa a través de la violencia física, psicológica, sexual y la negligencia, así como a través de las amenazas de tales actos, la cual se puede presentar en distintos ámbitos y ser ejercido por parte de sus padres, representantes legales o cualquier otra persona; produce daño y afecta la integridad personal, el desarrollo integral de los niños, niñas y adolescentes» (ICBF, 2017, p. 24). Concretamente, en la EVCNNA se indica que la violencia causa problemas de relacionamiento social, desempeño académico, problemas de salud mental, consumo de sustancias psicoactivas, discapacidad y muerte (OIM, CDC, Ministerio de Salud y Protección Social, 2019).

1. Caracterización de la violencia contra niñas, niños y adolescentes

Con el objetivo de determinar las acciones idóneas para el restablecimiento de derechos de las niñas, niños y adolescentes, el ICBF ha caracterizado la situación de violencia a partir de: 1) los tipos y formas a través de las cuales se expresa la violencia; 2) los ámbitos o contextos en que esta se presenta; 3) los(as) agresores(as); y 4) las afectaciones o daños que produce (ver Diagrama 1).

Diagrama 1. Aspectos caracterización de la violencia contra niñas, niños y adolescentes.



Fuente: elaboración propia con base en el Lineamiento técnico del ICBF para la atención de niños, niñas y adolescentes con sus derechos inobservados, amenazados o vulnerados por causa de la violencia.

1.1 Tipos o formas de violencia

La violencia física ocurre cuando una persona recurre a la fuerza corporal para causar dolor, malestar, lesiones y daños a la salud de las niñas, niños y adolescentes. Se puede ejercer a través de cualquier parte del cuerpo (cachetadas, puñetazos, patadas), objetos (cables, cigarrillos), armas (cuchillos, puñales), sustancias, agentes químicos u objetos que producen quemaduras (líquidos hirviendo, ácido) u otras formas como ahogamiento o ingesta forzosa. El castigo físico o corporal recoge de manera única o combinada las formas de violencia descritas anteriormente.

La violencia psicológica sucede cuando una acción u omisión de una persona tiene como fin degradar, discriminar o controlar las acciones, comportamientos, creencias y decisiones de las niñas, niños y adolescentes, a través de formas como humillar, rechazar, aterrorizar, aislar, ser permisivos, restringir la autonomía (sobreimplicación, infantilización o parentalización), sobreexigencia, instrumentalizar, exposición a violencia al interior del hogar o cualquier otra conducta que implique un perjuicio en la salud mental o el desarrollo personal.

La omisión o negligencia se refiere a la falta de protección y cuidado mínimo de la niña, niño o adolescente por parte de los progenitores o encargados del cuidado. Existe negligencia cuando los responsables del cuidado no protegen

de la exposición al peligro, ni atienden, o satisfacen las necesidades básicas de las niñas, niños y adolescentes, sean estas físicas, psicológicas, educativas o de salud, teniendo los medios, el conocimiento y acceso a la prestación de servicios (ICBF, 2017). La negligencia no se debe confundir con hechos accidentales, por desconocimiento o por condiciones de pobreza.

La violencia sexual, según la Ley 1146 de 2007 corresponde a «todo acto o comportamiento de tipo sexual ejercido sobre una niña, niño o adolescente, utilizando la fuerza o cualquier forma de coerción física, psicológica o emocional, aprovechando las condiciones de indefensión, de desigualdad y las relaciones de poder existentes entre víctima y agresor». Se expresa por medio de actos sexuales (en donde puede no existir penetración por vía anal, vaginal u oral), acceso carnal, explotación sexual comercial, acoso sexual, matrimonio temprano, esclavitud sexual, mutilación genital femenina, embarazo forzado, desnudez forzada, aborto forzado y esterilización o anticoncepción forzada (ICBF, s. f.).

Es importante resaltar que la ruta de atención del proceso de restablecimiento de derechos (PARD) por causa de violencia sexual es diferente de la ruta seguida para la atención de violencia física, psicológica o negligencia (maltrato). Esto debido a que por su complejidad requiere atención especializada y cualificada que se brinda en conjunto con otros sectores corresponsables (ICBF, 2018).

1.2 Contextos en los que se presenta la violencia

En sus diferentes tipologías la violencia puede manifestarse en tres contextos:

El **privado** se refiere al contexto familiar o de hogares de cuidado de las niñas, niños y adolescentes donde se configuran las relaciones particulares e interpersonales.

El **escenario público** está relacionado con los espacios vitales para el desarrollo de las niñas, niños y adolescentes y el ejercicio de sus derechos, los cuales deben ser espacios seguros, protectores y de cuidado. En el escenario público se construye y fortalece el ejercicio de la ciudadanía y la construcción de lo social, lo comunitario y lo político. Pueden ser institucionales (instituciones educativas, de protección, de salud, de justicia o de defensa, ámbito laboral, centros de atención en privación de libertad) y comunitarios (parques, vecindario, calle, espacio público, encuentros comunitarios) en los que las niñas, niños y adolescentes suelen realizar actividades a diario.

1.3 Agresores(as)

Dada la multiplicidad de contextos en los que la violencia puede generarse, los y las agresoras de las niñas, niños y adolescentes también pueden ser múltiples. Pueden ser familiares, cuidadores, actores institucionales o desconocidos.

1.4 Daños que produce

La afectación que causa la violencia es mayor en las niñas, niños y adolescentes que en población mayor de edad, en tanto que no solo se refiere a los daños físicos, sexuales y psicológicos, sino también al daño potencial en el desarrollo integral y la construcción de la identidad. Los tipos o formas de violencia referenciados antes pueden producir diferentes tipos de daño, por ejemplo, la negligencia puede ocasionar daños físicos y psicológicos.

Además del tipo de daño (físico, psicológico o sexual), el efecto de la violencia puede ser alterado por la frecuencia de los hechos violentos, la intensidad, el periodo evolutivo de la vida de la niña, niño o adolescente en que se presenta; el vínculo que existe con el agresor y la existencia de situaciones de vulnerabilidad especiales (por ejemplo, discapacidad, migración o desplazamiento).

2. Cifras de la violencia contra niñas, niños y adolescentes en Colombia

De acuerdo con la OMS (2020), cada año alrededor de 1.000 millones de niñas, niños y adolescentes son víctimas de violencia en sus diferentes formas en todo el mundo. En 2018 el Ministerio de Salud y Protección Social en conjunto con la Organización Internacional para las Migraciones (OIM) y los Centros para el Control y la Prevención de Enfermedades de los Estados Unidos (CDC) realizaron la primera Encuesta de Violencia contra Niños, Niñas y Adolescentes (EVCNNA) con el objetivo de establecer una línea de base de la magnitud de las diferentes formas de violencia que afectan a los menores, factores de riesgo y protección y consecuencias. El 41,5 % de las personas de 18 a 24 años reportaron haber sufrido algún tipo de violencia antes de cumplir 18 años, el 32,1 % violencia física, el 15,2 % violencia psicológica y el 11,4 %, violencia sexual². La prevalencia de violencia física fue mayor entre los hombres (37,5 %) que entre las mujeres (26,5 %) mientras que, para la violencia psicológica (9,5 % para los hombres y 21,2 % para las mujeres) y la violencia sexual (7,8 % y 15,3 %, respectivamente), la prevalencia entre las mujeres fue aproximadamente el doble que entre los hombres.

Adicionalmente, en Colombia, es posible aproximarse al comportamiento de la violencia contra las niñas, niños y adolescentes usando los registros realizados por diferentes entidades públicas que intervienen en la identificación, protección y judicialización de esta situación: la Policía Nacional, la Fiscalía General de la Nación (FGN), el Instituto Nacional de Medicina Legal y Ciencias Forenses (INMLCF) y el ICBF. No obstante, es importante resaltar que existe un subregistro importante de este tipo de violencia. Este se explica porque suele presentarse en ámbitos privados y la denuncia por parte de quienes la conocen puede ser

2. El 3,4 % sufrió violencia sexual solamente, el 19,5 % solo física, el 4,9 % solo psicológica, el 10,1 % dos tipos de violencia y el 3,6 % los tres tipos de violencia.

limitada por factores como normalización de la violencia como práctica de crianza, miedo a las consecuencias que pueda tener la denuncia para el agresor, la víctima o la familia; desconocimiento de los mecanismos para denunciar o la creencia de que el hecho quedará en la impunidad, que se asocia a falta de confianza en las autoridades.

A continuación, se presenta la información referente a la violencia contra niñas, niños y adolescentes desde los registros de la Policía Nacional, la FGN y el INMLCF y los registros de PARD del ICBF.

2.1 Registro de violencia contra niñas, niños y adolescentes

En Colombia, las principales entidades encargadas de recibir las denuncias de cualquier presunta violación al Código Penal son la Policía Nacional y la FGN. Solo en los casos en los que menores de edad se encuentran involucrados es posible presentar la denuncia al ICBF. Los sistemas de información de denuncias de la Policía Nacional (SIEDCO³) y de la FGN (SPOA⁴) se encuentran en proceso de integración desde 2017 con el lanzamiento del portal ADenunciar, por lo que 2018 es el año cero de integración de las estadísticas de denuncias. Ambas instituciones aún se encuentran resolviendo la integración y homologación de sus estadísticas históricas.

Adicionalmente, cuando la violencia tiene manifestaciones físicas evidentes que obligan al afectado a dirigirse a una clínica u hospital, el Instituto Nacional de Salud (INS) y el INMLCF se convierten en otra fuente de información de casos de abuso o maltrato y, sobre todo, de violencia sexual (sistema de gestión de datos SISPRO-SIVIGILA⁵).

En la Tabla 1 se presentan las cifras de violencia al interior del hogar y violencia sexual contra niñas, niños y adolescentes de acuerdo con tres de las fuentes mencionadas antes. Las diferencias en el reporte se deben a que las instituciones conocen el fenómeno en diferentes circunstancias. Por ejemplo, los casos de violencia al interior del hogar fatales registrados por el INMLCF serían solo aquellos que requirieron la realización de un examen médico-legal.

3. Sistema de Información Estadístico, Delincuencial, Contravencional y Operativo de la Policía Nacional (SIEDCO).

4. Sistema Penal Oral Acusatorio de la Fiscalía General de la Nación (SPOA).

5. Sistema Integrado de Información para la Protección Social (SISPRO) y Sistema Nacional de Vigilancia en Salud Pública (SIVIGILA).

Tabla 1. Violencia al interior del hogar y violencia sexual contra niñas, niños y adolescentes, 2016-2020

Violencia al interior del hogar contra NNA					
	2016	2017	2018	2019⁶	2020⁶
Reporte delito – Policía ¹	6199	8825	11623	15354	13098
Conteo de víctimas según entrada de noticias criminales – FGN ²	7641	10272	11629	17499	11890
Casos no fatales registrados – INMLCF ³	11574	11816	12089	9556	5069 ⁴

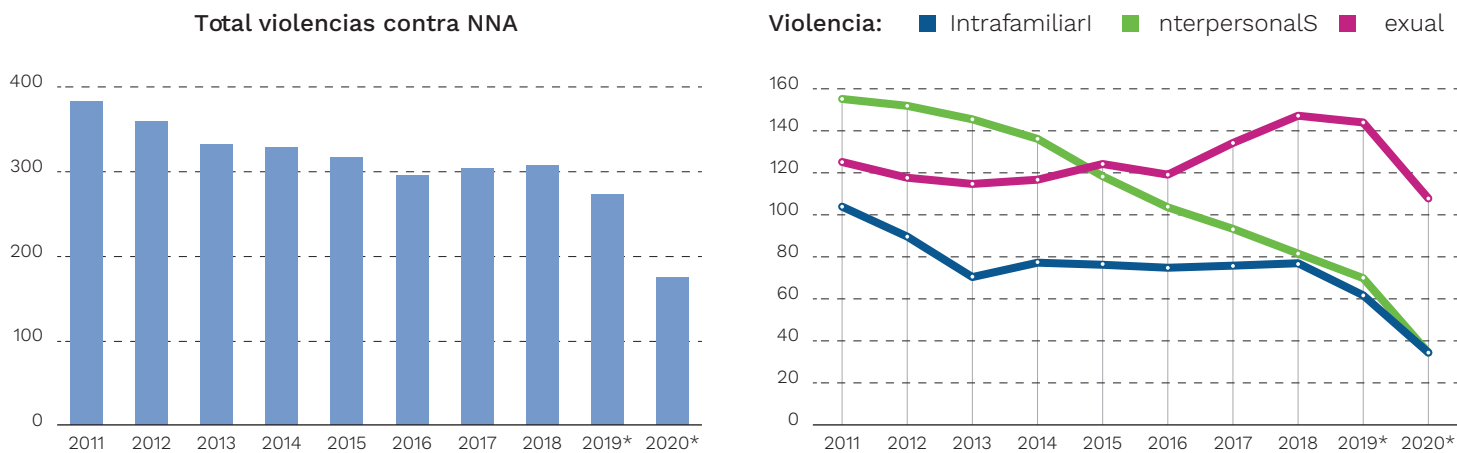
Violencia sexual contra NNA					
	2016	2017	2018	2019⁶	2020⁶
Reporte delito – Policía ⁵	16842	20359	24696	27163	19744
Conteo de víctimas según entrada de noticias criminales – FGN ²	18811	23892	30004	35566	26872
Exámenes médico-legales por presunto delito sexual – INMLCF ³	18416	20663	22794	22211	15359 ⁴

Fuente: Datos abiertos - DIJIN Policía Nacional (11 de mayo de 2021), Reporte delito violencia al interior del hogar Policía Nacional¹. Datos abiertos - Fiscalía General de la Nación (11 de mayo de 2021), Conteo de víctimas según las entradas de noticias criminales por delito al Sistema Penal Oral Acusatorio en la Ley 906 de 2004 y Ley 1098 de 2006². INMLCF. Cálculo violencias INML_2011-2019³. INMLCF. Información preliminar de lesiones no fatales de causa externa en Colombia, enero a diciembre de 2020⁴. Datos abiertos - DIJIN Policía Nacional (11 de mayo de 2021) Reporte delitos sexuales Policía Nacional⁵. Datos preliminares a diciembre⁶.

Es importante recordar que el INMLCF tiene como propósito ofrecer apoyo técnico para satisfacer las necesidades tanto de la FGN como del sistema judicial ordinal en general (jueces, policía judicial, Defensoría del Pueblo), de manera que estos puedan tomar decisiones con el suficiente material probatorio. Por estos motivos, tipifica y calcula casos de violencia al interior del hogar, interpersonal y sexual, lo que lo convierte en fuente general de diversos tipos de violencia fatal y no fatal en todos los contextos (familiar y externo) en tanto conoce de ellos y ofrece material probatorio sobre su realización.

Esta es la principal razón por la que el indicador de violencia contra niñas, niños y adolescentes incluido dentro del Plan Nacional de Desarrollo 2018-2022 toma como numerador la suma de los casos tipificados por el INMLCF, como lo son: violencia interpersonal, violencia de pareja, violencia al interior del hogar y exámenes médico-legales por presunto delito sexual entre población menor de 18 años. Como se aprecia en la siguiente gráfica, dos de los hechos medidos en el indicador disminuyeron durante el período 2011-2020 y solo los casos de presunto delito sexual han aumentado. En los datos preliminares a diciembre de 2020 se observa una disminución en los casos tipificados para todos los hechos.

Gráfica 1. Tasa de violencia sexual, violencia al interior del hogar, violencia interpersonal y violencia total contra niñas, niños y adolescentes por 100.000 habitantes, 2011-2020



Fuente: INMLCF. Eje secundario Violencia Total. *Datos preliminares a diciembre.

El INMLCF entiende la violencia al interior del hogar como cualquier forma de maltrato, ya sea físico, psicológico o sexual, que tiene lugar entre los miembros de una familia y, por lo tanto, incluye violencia hacia niñas, niños y adolescentes, hacia el adulto mayor, entre otros familiares y violencia de pareja. El último componente tiene la mayor participación (en promedio 65 % de los casos reportados entre 2010 y 2018) dentro del conteo total de los casos de violencia en la familia denunciados. La violencia contra niñas, niños y adolescentes es el tercer componente más importante; este ha disminuido 22 % en el período 2010-2019, evidenciando un decrecimiento más acelerado en comparación con el resto de los

componentes. En los últimos 4 años el número de casos reportados ha disminuido en unos 300 casos en promedio por año, ritmo inferior al presentado en el cuatrienio anterior, en el que el número llegó a disminuir hasta en 2.000 casos anuales.

Finalmente, es importante resaltar que la tendencia observada en el reporte no necesariamente refleja la violencia experimentada por los ciudadanos, sino que se puede relacionar también con la tolerancia de la población hacia la violencia y su confianza en las instituciones. Así, por ejemplo, es posible que se observe un aumento en las denuncias mientras la violencia experimentada por los ciudadanos se mantiene igual o incluso disminuye, si hay mayor información sobre las rutas y los canales de denuncia, mayor pedagogía para prevenir la violencia, un fortalecimiento en la institucionalidad o un aumento en la confianza de la ciudadanía hacia las instituciones. Por el contrario, una disminución en las denuncias puede reflejar mayor tolerancia hacia la violencia entre los ciudadanos, la existencia de barreras para denunciar o alta desconfianza en las instituciones.

2.2 Ingresos de niñas, niños y adolescentes al Proceso Administrativo de Restablecimiento de Derechos (PARD) por motivo de violencia

El Proceso Administrativo de Restablecimiento de Derechos (PARD) del ICBF consiste en un conjunto de acciones planificadas y organizadas para la atención integral de las niñas, niños y adolescentes, en el marco de la corresponsabilidad de la familia, la sociedad y el Estado, con el fin de garantizar y promover el ejercicio pleno de sus derechos, prevenir su amenaza o vulneración y restablecer aquellos que les han sido vulnerados. Este proceso toma como referencia el marco de la Ley de Infancia y Adolescencia (Ley 1098 de 2006), la Ley 1878 de 2018, la Convención Internacional sobre los Derechos del Niño y las normas y tratados internacionales ratificados por Colombia en lo referente al reconocimiento, garantía, protección y restablecimiento de los derechos de las niñas, niños y adolescentes, así como la prevención de su amenaza o vulneración.

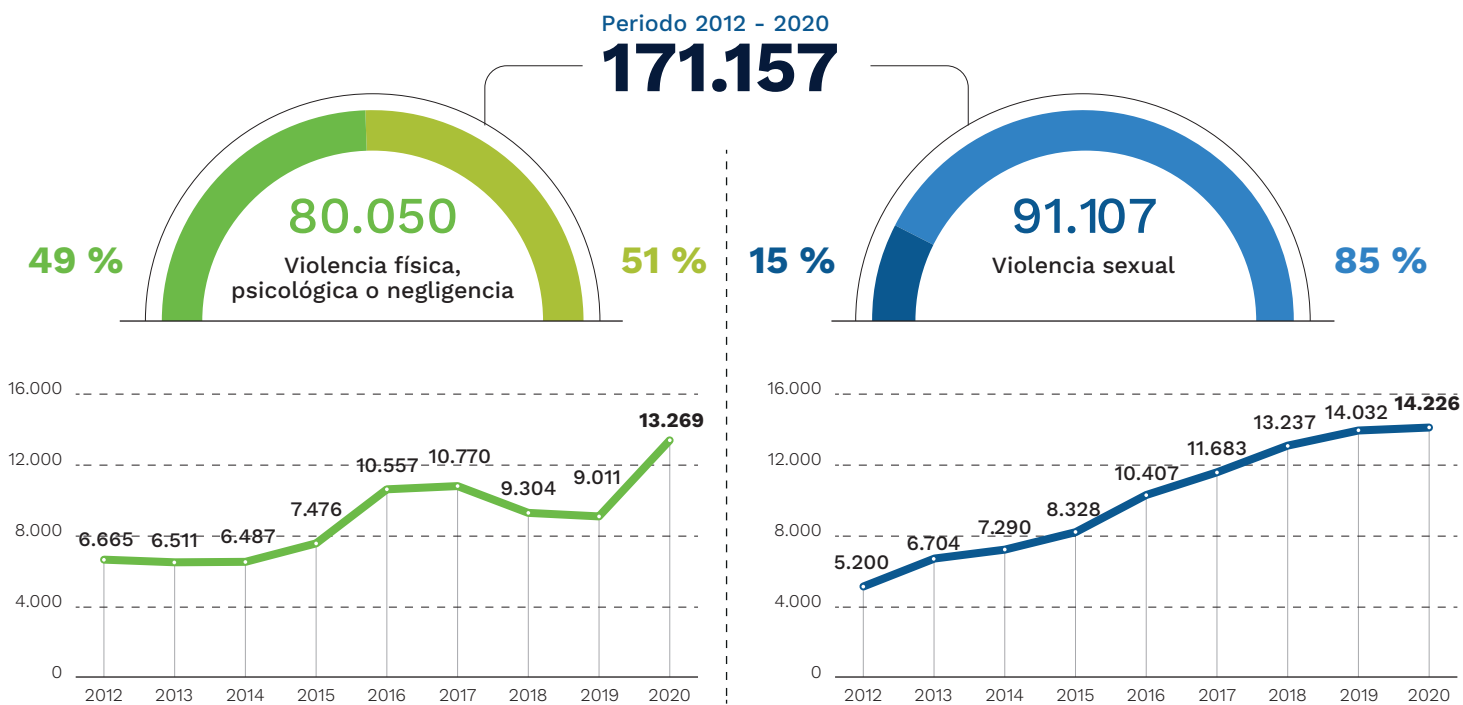
En Colombia, durante el periodo 2012 a 2019, 143.662 niñas, niños y adolescentes ingresaron al PARD del ICBF por algún motivo de violencia⁶. De este total, 66.781 ingresaron por motivo de violencia física, psicológica o negligencia (maltrato), y 76.881 por violencia sexual (Gráfica 2). Entre 2016 y 2019, los ingresos por violencia física, psicológica y negligencia (maltrato) disminuyeron 14,6% y los de violencia sexual aumentaron 34,8 %. En 2019, el 22,5 % de los ingresos a PARD fueron motivados por violencia física, psicológica y negligencia (maltrato) y el 35,1 % por violencia sexual.

6. Entre 2012 y 2020 ingresaron 171.157 niñas, niños y adolescentes al PARD del ICBF, 80.050 por motivo de violencia física, psicológica o negligencia (maltrato), y 91.107 por violencia sexual. El ingreso al PARD por ambos tipos de violencia ha crecido en el periodo, 99 % y 173,6 %, respectivamente. Las comparaciones no incluyen la vigencia 2020 por dos razones: en primer lugar, se observó un incremento atípico en los casos de violencia física, psicológica o negligencia (maltrato) (47,3 %) que puede relacionarse con el aislamiento preventivo obligatorio. Adicionalmente, en la vigencia 2020 se implementaron mejoramientos en el sistema de información que, entre otros cambios, hicieron que al registrar un PARD fuera obligatorio registrar el motivo de ingreso, lo que puede introducir variaciones en la distribución de motivos de ingreso PARD frente a vigencias anteriores.

Entre 2012 y 2019, la violencia física, psicológica y la negligencia (maltrato) se presentó de forma relativamente equitativa entre las niñas y las adolescentes y los niños y los adolescentes (51 % y 49 %, respectivamente). Los casos se distribuyen de forma relativamente homogénea entre los grupos de edad (37 % para la primera infancia, 33,4 % para los niños y niñas entre 6 y 11 años y 28,9 % para los y las adolescentes entre 12 y 17 años). Los departamentos con mayor número de ingresos a PARD por maltrato en relación con la población de 0 a 17 años en 2019 fueron Nariño, Quindío y Caldas (ver Mapa 1).

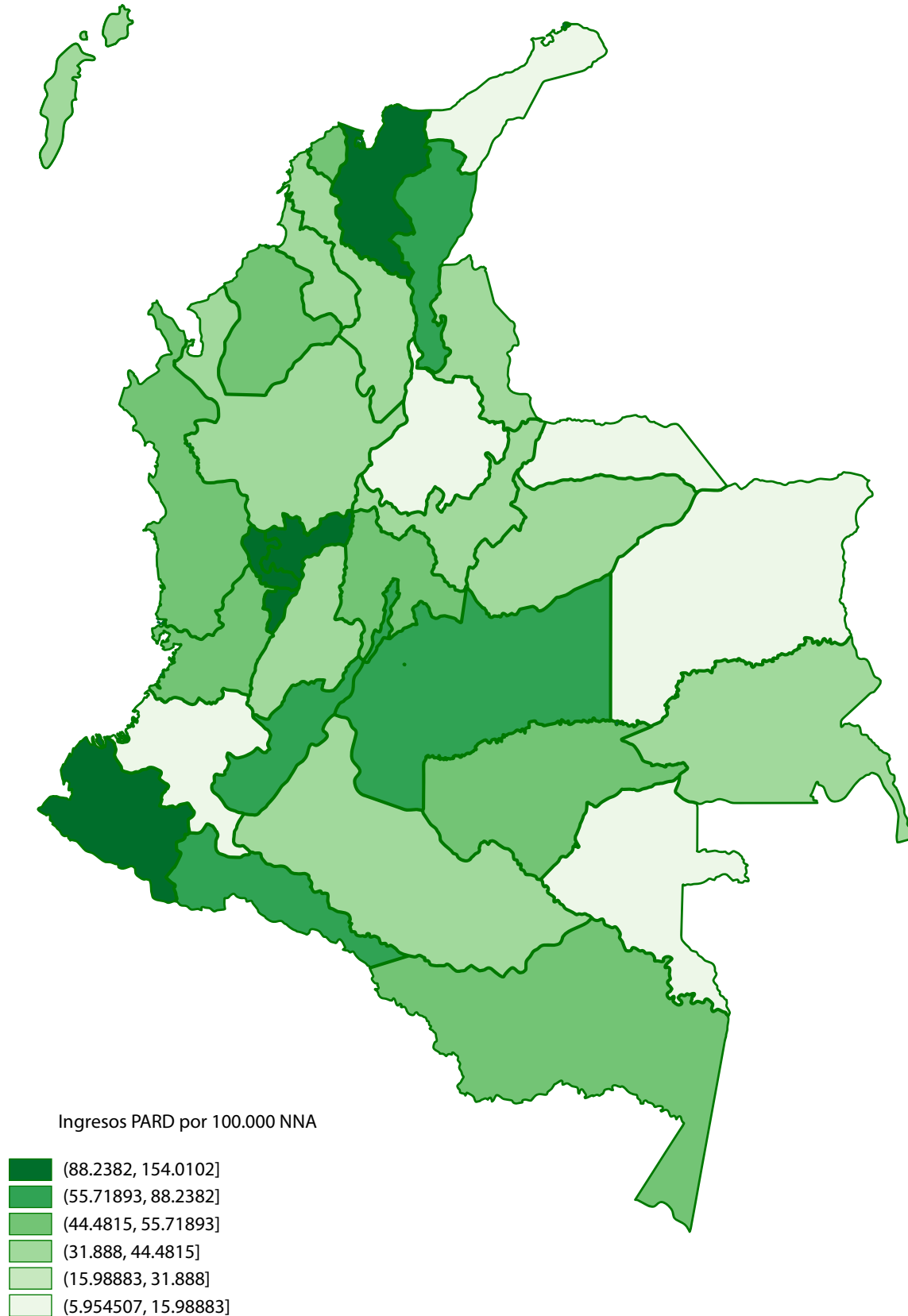
En el mismo periodo, el 84,8 % de los ingresos a PARD por violencia sexual se presentó en niñas y adolescentes mujeres, lo que reflejó la persistencia de relaciones desiguales de poder existentes entre géneros. Por grupos de edad, este tipo de violencia se concentra en las y los adolescentes entre los 12 y 17 años (50,2 %), seguido de las niñas y niños entre los 6 y 11 años (34,2 %), y la primera infancia (14,8 %). En 2019, los departamentos con mayor número de ingresos como proporción de la población de 0 a 17 años por este motivo fueron Amazonas, Guaviare, Quindío, Putumayo y Guainía (ver Mapa 2).

Gráfica 1. Tasa de violencia sexual, violencia al interior del hogar, violencia interpersonal y violencia total contra niñas, niños y adolescentes por 100.000 habitantes, 2011-2020



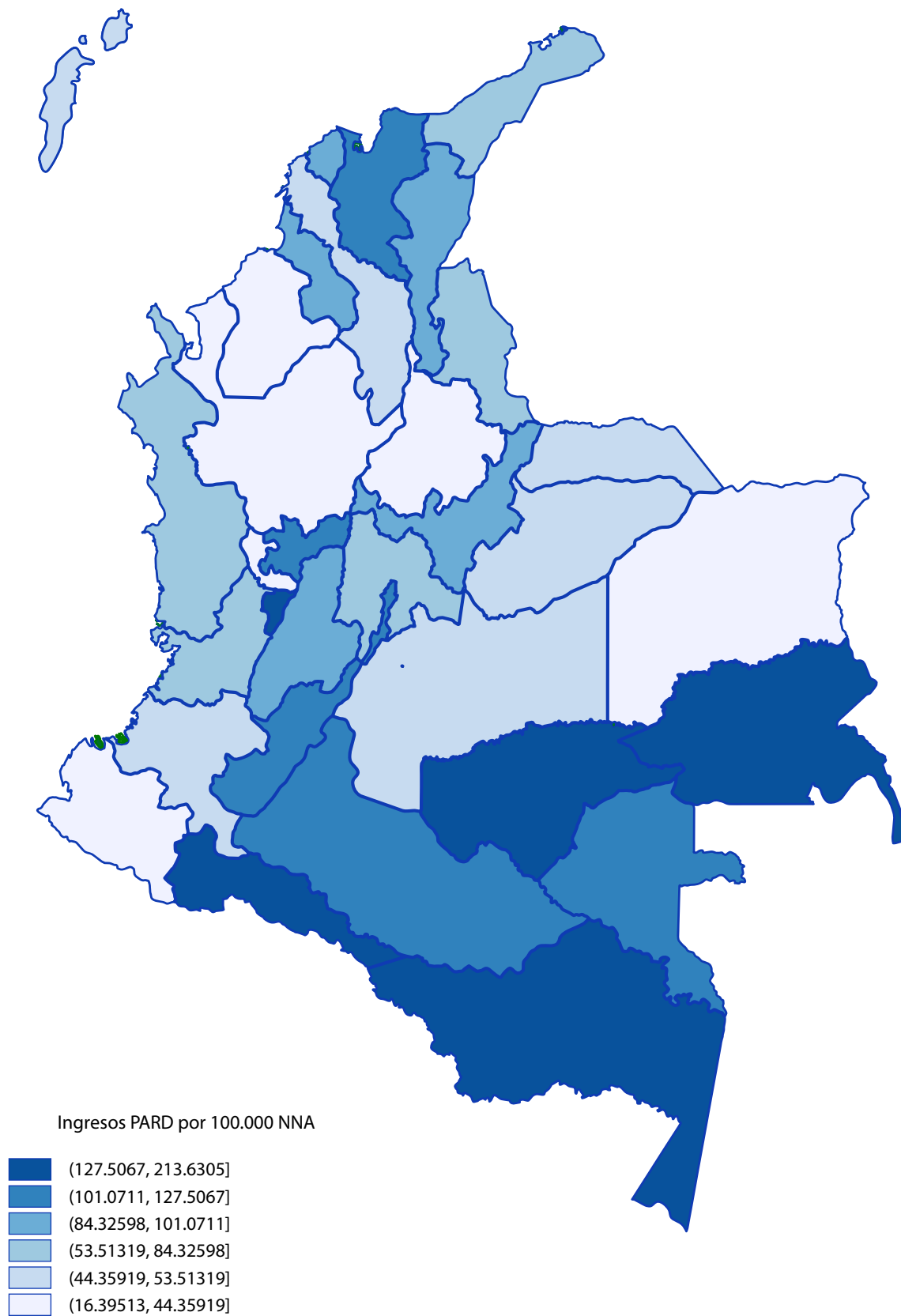
Fuente: Sistema de Información Misional (SIM).
Fecha de corte: 31 de diciembre de 2020.

Mapa 1. Ingresos PARD de NNA por cada 100.000 NNA por motivo de violencia física, psicológica y negligencia (maltrato), 2019.



Fuente: Sistema de Información Misional (SIM).
Fecha de corte: 31 de diciembre de 2019.

Mapa 2. Ingresos PARD de NNA por cada 100.000 NNA por motivo sexual, 2019.



Fuente: Sistema de Información Misional (SIM).
Fecha de corte: 31 de diciembre de 2019.

2.3 Violencia contra niñas, niños y adolescentes durante la pandemia del COVID-19

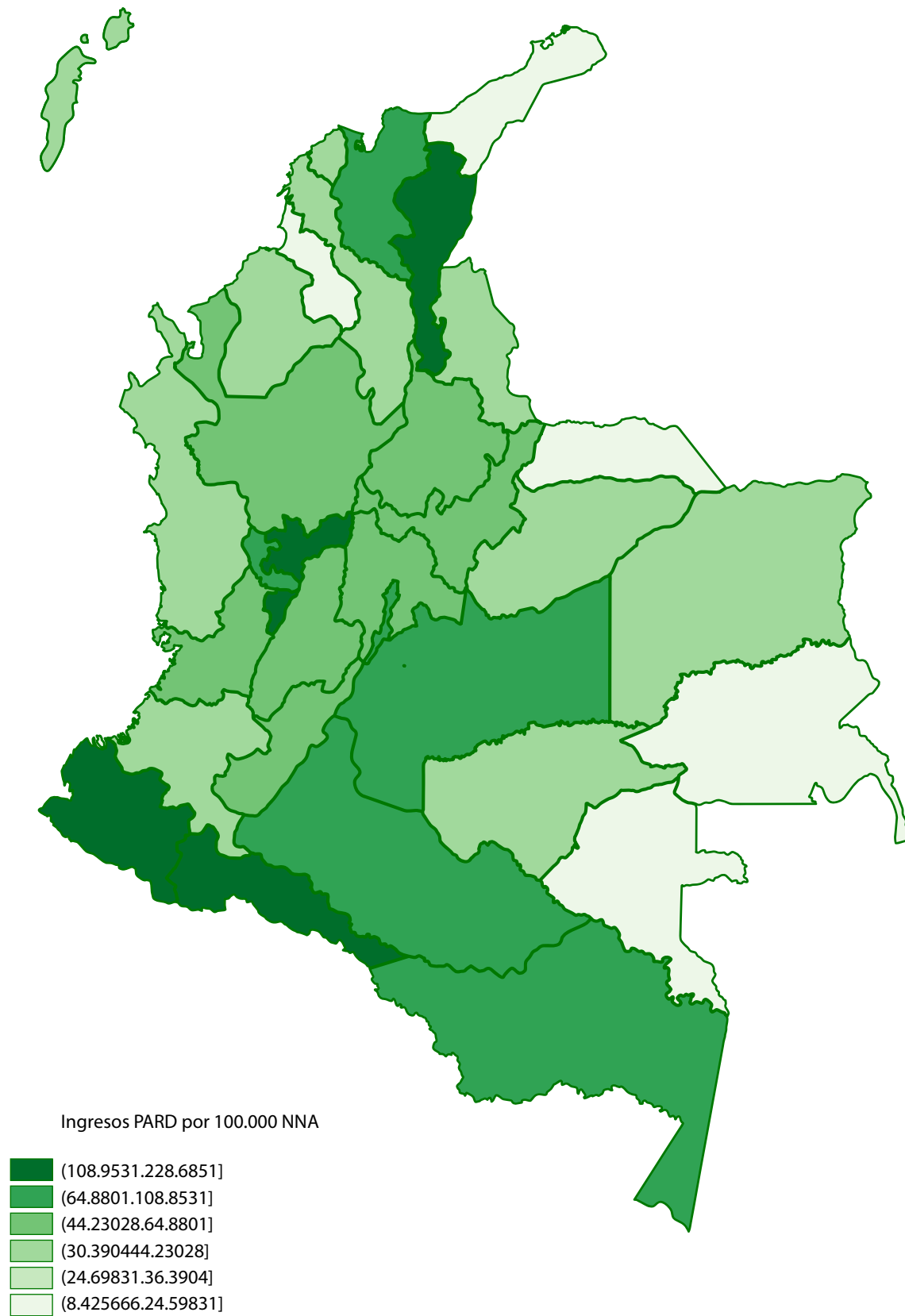
De acuerdo con la OMS (2020), las medidas de distanciamiento físico adoptadas como respuesta a la pandemia de la COVID-19 crearon condiciones de pérdida de ingreso económico, aislamiento y hacinamiento, causando situaciones de estrés y ansiedad que incrementaron de forma importante el riesgo de que las niñas, niños y adolescentes fueran objeto o testigos de violencia entre sus padres y cuidadores. Durante el confinamiento, los llamados a las líneas de ayuda y el reporte de violencias contra niñas, niños y adolescentes aumentaron en países como China, Francia, India, Sur África y Reino Unido (OMS, 2020). En países donde se observó incremento en el reporte, como Estados Unidos y Filipinas, este se ha explicado por el menor contacto de las niñas, niños y adolescentes con los proveedores de servicios que puedan detectar el abuso como profesores, trabajadores sociales y personal de salud (OMS, 2020). Se estima que las tasas de violencia se mantendrán por encima de los niveles anteriores a la pandemia, debido a efectos a largo plazo en el empleo, seguridad financiera del hogar y menores posibilidades de acceder a programas de protección y bienestar social.

En Colombia, entre 2019 y 2020 se observa una caída en el reporte de violencias contra niñas, niños y adolescentes en el registro de la Policía Nacional, la FGN y el INMLCF. Sin embargo, en los registros de ingresos PARD del ICBF se observa un incremento, especialmente en el ingreso por motivo de violencia física, psicológica o negligencia (maltrato) (47,3 %), revirtiendo la tendencia decreciente que se observaba desde 2017.

En 2020, los departamentos con mayor número de ingresos PARD como proporción de la población de 0 a 17 años por motivo de violencia fueron similares a los observados en 2019, aunque con diferencias en el orden y la magnitud de las tasas de violencia observadas. Por motivo de violencia física, psicológica y negligencia (maltrato), los departamentos con el mayor número relativo de ingresos a PARD en 2020 son Caldas, Nariño, Putumayo, Quindío, Cesar y Risaralda (ver Mapa 3). La mayor tasa registrada en Caldas fue de 228,7, mientras que en 2019, la mayor tasa se observó en Nariño con 154. Con relación a la violencia sexual, los departamentos con mayor cantidad de ingresos fueron Guaviare, Guainía, Amazonas, Putumayo, Quindío y Caldas (ver Mapa 4).

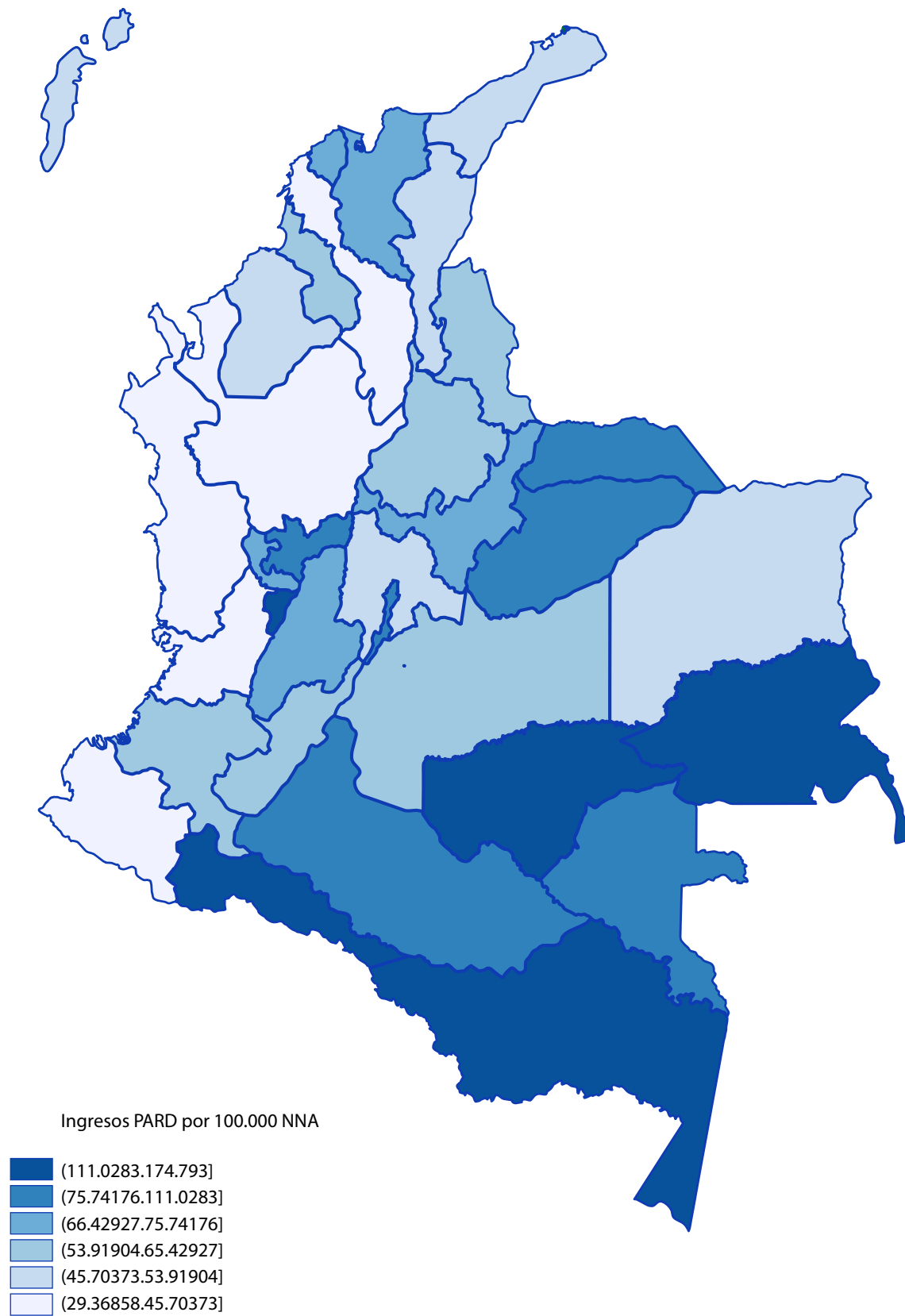


Mapa 3. Ingresos PARD de NNA por cada 100.000 NNA por motivo de violencia física, psicológica y negligencia (maltrato), 2020.



Fuente: Sistema de Información Misional (SIM).
Fecha de corte: 31 de diciembre de 2020.

Mapa 4. Ingresos PARD de NNA por cada 100.000 NNA por motivo sexual, 2020.



Fuente: Sistema de Información Misional (SIM).
Fecha de corte: 31 de diciembre de 2020.

II. Datos y metodología de los modelos de probabilidad de vulneración

A pesar de las graves consecuencias de la violencia contra las niñas, niños y adolescentes, la respuesta institucional suele ser predominantemente reactiva debido a que los gobiernos cuentan con recursos limitados para financiar los esfuerzos de prevención, y la información y herramientas disponibles para tomar decisiones respecto a la prevención son insuficientes (Administrative Data Research Facilities Network, 2019). Esta situación implica que la respuesta institucional puede llegar demasiado tarde con consecuencias trágicas para las niñas, niños y adolescentes y sus familias.

De acuerdo con lo anterior y con el objetivo de mejorar la alineación entre los servicios de prevención y los requerimientos de las comunidades, países como Estados Unidos, Nueva Zelanda y Reino Unido han avanzado en el desarrollo de modelos que predicen la probabilidad de que se produzcan hechos de violencia contra niñas, niños y adolescentes. Los modelos pueden enfocarse en predecir el riesgo asociado a los individuos u hogares, o el riesgo asociado al lugar o ambiente en el que se encuentra la niña, niño o adolescente (Administrative Data Research Facilities Network, 2019). Modelos similares en estos países, estiman la probabilidad de que las niñas, niños y adolescentes sean explotados, de que regresen al cuidado institucional o de que cometan actos de violencia. Los modelos de analítica predictiva compilan grandes cantidades de datos estructurados y no estructurados de diferentes fuentes y mediante herramientas de aprendizaje de máquinas (machine learning)⁷ descubren patrones y correlaciones a través de los cuales se pueden predecir tendencias y patrones de comportamiento. Las predicciones obtenidas por estos modelos no solo permiten mejorar la eficiencia, objetividad y equidad en las decisiones frente a la asignación de los recursos para la prevención, en comparación con decisiones basadas en criterios como la experiencia personal o la intuición, sino que son más acertadas que las basadas en técnicas de análisis multivariable estándar (Pryce et al., 2018).

No obstante, se ha señalado que los modelos de analítica predictiva pueden generar problemas éticos significativos como el uso de información sin consentimiento y la estigmatización de las niñas, niños, adolescentes y familias, especialmente de los más vulnerables. La contención de estos riesgos requiere contar con políticas de seguridad de la información adecuadas, datos suficientemente diversos para evitar la generación de resultados sesgados y asegurar la transparencia del modelo, para lo cual se debe revelar la información relativa a los datos empleados y el desarrollo del algoritmo (Gillingham, 2016).

Este es el objetivo de esta sección, en la que inicialmente se describirán los datos usados en los modelos de probabilidad de vulneración del ICBF y, posteriormente, se detallará el procedimiento seguido para su desarrollo.

7. El aprendizaje de máquinas es una rama de la inteligencia artificial y las ciencias de la computación que se enfoca en el uso de datos y algoritmos para imitar la forma en que los humanos aprenden, aumentando de forma gradual su precisión. El aprendizaje de máquinas es un componente importante del creciente campo de la ciencia de datos. Mediante el uso de métodos estadísticos, los algoritmos se entrenan para hacer clasificaciones o predicciones y descubrir información clave dentro de los proyectos de minería de datos. Esta información, posteriormente, se usa para la toma de decisiones que impactan el desempeño de las organizaciones. IBM (15 de julio de 2020) Machine Learning. (original en inglés).

1. Datos

Los modelos de analítica predictiva desarrollados por el ICBF permiten estimar la probabilidad de que una niña, niño o adolescente registrado en la base de datos Sisbén IV del Departamento Nacional de Planeación (DNP) pueda ser víctima de violencia sexual o de violencia física, psicológica y negligencia (maltrato).

De acuerdo con lo anterior, la principal base de datos utilizada en los modelos es el Sisbén IV que contiene el conjunto de niñas, niños y adolescentes para los cuales se estimará el riesgo de padecer violencia física, psicológica o negligencia (maltrato) o violencia sexual. La base del Sisbén IV usada tiene fecha de corte a junio de 2021, y cuenta con 24.852.809 registros, de los cuales 7.236.230 corresponden a niñas, niños y adolescentes⁸. Estos registros fueron recolectados entre 2017 y 2021.

Otra fuente de información es la base de ingresos a un Proceso Administrativo de Restablecimiento de Derechos (PARD) del ICBF la cual se emplea como fuente para identificar a las niñas, niños y adolescentes que han tenido una vulneración de derechos por causa de violencia y para los cuales se ha iniciado un proceso de restablecimiento de derechos.

Adicionalmente, se integra información del Sistema Integrado de Matrícula (SIMAT) del Ministerio de Educación Nacional (MEN); el Registro Único de Víctimas (RUV) de la Unidad para las Víctimas; el Registro para la Localización y Caracterización de Personas con Discapacidad (RLCPD) del Ministerio de Salud; el sistema de información Cuéntame del ICBF; la base de datos de cultivos ilícitos del Observatorio de Drogas de Colombia; el SPOA de la FGN; el FORENIS del INMLCF; y Estadísticas Vitales y Proyecciones de Población del DANE. En el Anexo 1 se detalla la totalidad de las variables de la base de datos.

A continuación, se presentan estadísticas descriptivas de las variables recolectadas y creadas a partir de otras que permiten caracterizar a las niñas, niños y adolescentes, sus hogares, sus viviendas y los municipios donde residen. Para esto se usa la base de datos construida y depurada de acuerdo con la etapa 2 del procedimiento que se describe en detalle en la sección **II. 2 Metodología de los modelos de probabilidad de vulneración de derechos de niñas, niños y adolescentes** de este documento.

1.1 Características de las niñas, niños y adolescentes

En la base de datos consolidada se observa una distribución equitativa en la participación por grupos de edad y entre niñas y adolescentes mujeres y niños y adolescentes hombres. El 31,21 % de los registros corresponden a hijos únicos. El 18,82 % de las niñas, niños y adolescentes habría sido afectado por el conflicto armado (por desplazamiento u otro motivo) y solo el 2,93 % registra alguna discapacidad.

8. La base completa del Sisbén IV fue procesada para generar variables descriptivas de conformación del hogar de las niñas, niños y adolescentes identificados (p. e. jefatura de hogar, actividad y edad de los padres, tipo de familia, entre otros).

La mayoría de las niñas, niños y adolescentes (69,54 %) son asegurados por el régimen subsidiado de salud. El 68,66 % de las niñas y niños menores de 5 años permanece la mayor parte del tiempo bajo el cuidado de su padre o madre en casa, pero una proporción importante (22,47 %) es cuidado en una institución como un hogar comunitario, jardín, centro de desarrollo infantil o colegio. El 89,57 % de las niñas, niños y adolescentes mayores de 6 años saben leer y escribir y el 56,11 % se encuentran estudiando, de acuerdo con el registro en el SIMAT. De estos últimos, el 13,01 % está matriculado en jornada única. El 4,53 % de las niñas, niños y adolescentes escolarizados se encuentran en extraedad, es decir son mayores que la edad que deberían tener de acuerdo con el grado educativo en el que están matriculados. El 34,4 % de las niñas, niños y adolescentes en la base recibieron algún servicio en las modalidades de prevención del ICBF.

El 1,87 % de las niñas, niños y adolescentes registra haber ingresado a un proceso de restablecimiento de derechos (PARD). De estos, el 22,22 % (0,42 % del total de niñas, niños y adolescentes) ingresó por motivo de violencia física, psicológica o negligencia (maltrato) y el 29,17 % (0,55 %) por motivo de violencia sexual.

Tabla 2. Características de las niñas, niños y adolescentes

General						
Edad (años)	Entre 0 y 5	33,04 %	Entre 6 y 11	32,95 %	Entre 12 y 17	33,83 %
Sexo	Hombre	51,14 %	Mujer	48,86 %		
Tipo de hijo - Lugar	Mayor	29,18 %	Entre 6 y 11	31,21 %	Otro	39,61 %
Discapacidad	Si	2,93 %	No	97,07 %		
Unión temprana	No	98,4 %	Si	1,6 %		
Víctima del conflicto armado	Si	18,82 %	No	81,18 %		
Salud, Educación y Cuidado						
Tipo seguridad social	Ninguna Subsidiado	6,65 % 69,54 %	Contributivo No sabe	22,16 % 0,79 %	Especial	0,86 %
Con quién permanece menor de 5 años	Institución (1) Empleada (4) Solo (7)	22,47 % 0,42 % 0,02 %	Padres en casa (2) Pariente >18 años (5) Otro (8)	68,66 % 6,92 % 0,34 %	Padres en trabajo (3) Pariente <18 años (6)	0,95 % 0,22 %

Salud, Educación y Cuidado						
Beneficiario programas prevención ICBF	Si	34,4 %	No	65,6 %		
Sabe leer y escribir¹	Si	89,57 %	No	10,43 %		
Estudia (MEN)	Si	56,11 %	No	43,89 %	Otro	39,61 %
Se encuentra en extra edad	Si	4,53 %	No	95,47 %		
Jornada estudio²	Si	13,59 % 0,09 %	Mañana Fin de semana	51,96 % 1,09 %	Tarde Única	19,45 % 13,01 %
Ingreso PARD						
Ingreso a PARD	Si	1,87 %	No	98,13 %		
PARD por maltrato	Si	0,42 %	No	99,58 %		
PARD por violencia sexual	Si	0,55 %	No	99,45 %		

1 Excluye menores de 5 años.

2 Excluye los que no cruzan con SIMAT.

Fuente: elaboración propia

1.2 Características del hogar

Casi la mitad de las niñas, niños y adolescentes viven en hogares con su padre y madre. La segunda composición de hogar más común es vivir solo con la madre (37,74 %) y la tercera es vivir sin ninguno de los padres (6,53 %). Más de la mitad de las niñas, niños y adolescentes habitan en hogares con jefatura femenina (57,51 %). En promedio las niñas, niños y adolescentes viven en hogares conformados por 4,17 personas (con una desviación estándar de 1,57) de las cuales 2,15 (desviación estándar de 1,11) son menores de edad.

Solo en el 8,94 % de los hogares tanto la figura materna como la paterna contribuyen con la generación de ingresos y en el 19,4 % la figura materna es la única que contribuye. La mayoría de las niñas, niños y adolescentes tienen madres y padres con nivel educativo media (36,96 % y 30,7 % respectivamente), y primaria (26,84 % y 36,91 %).

El promedio de gasto mensual por persona en los hogares es de \$170.056,9 (con desviación estándar de \$190.980,1). Del total de gasto, el 56,85 % (desviación estándar 27,8 %) se destina a alimentación. El 32,72 % de las niñas, niños y adolescentes de la base viven en hogares clasificados en el grupo A del Sisbén IV correspondiente a hogares en situación de pobreza extrema, el 41,81 % al grupo B que corresponde a pobreza moderada, el 22,64 % al grupo C, hogares en situación de vulnerabilidad y los restantes a hogares en el grupo D que no están en situación de pobreza ni vulnerabilidad socioeconómica.

Las niñas, niños y adolescentes en promedio viven en hogares con 3,85 privaciones⁹ del Índice de Pobreza Multidimensional (desviación estándar de 2,24). El 30,03 % de los hogares son considerados pobres de acuerdo con el indicador de pobreza multidimensional. El 54,17 % de las niñas, niños y adolescentes viven en hogares afectados por bajo logro educativo y el 45,52 % en hogares donde algún miembro presenta rezago escolar. Solo el 20,11 % viven en hogares afectados por desempleo de larga duración, pero el 73,03 % vive en hogares afectados por trabajo informal. Aunque en promedio los niños viven en hogares con 2,22 personas por cuarto usado para dormir (desviación estándar de 1,32), el 38,12 % de los niños vive en hogares con privación por hacinamiento crítico (es decir, hay más de 2,5 personas por habitación usada para dormir).

Tabla 3. Características del hogar

Hogar						
Composición hogar	Sin madre ni padre (1) Solo madre (4)	6,53% 37,74%	Solo madre y su cónyuge (2) Solo padre (5)	3,86% 4,33%	Solo padre y su cónyuge (3) Con madre y padre (6)	1,42% 46,11%
Jefe de hogar mujer	Si	57,51%	No	42,49%		38,81%
Generador de ingresos	Ninguno (1) Ambos (4)	32,85% 8,94%	Solo figura materna (2)	19,40%	Solo figura paterna (3)	38,81%
Grupo Sisbèn	A D	32,72% 2,83%	B	41,81%	C	22,64%
Edad madre	Sin madre de 32 a 35 años	6,64% 16,15%	De 0 a 25 años De 36 a 41 años	15,84% 19,53%	De 26 a 31 años 42 años o mayor	21,37% 20,46%
Edad padre	Sin padre de 36 a 40 años	39,55% 11,5%	De 0 a 29 años De 41 a 48 años	11% 13,21%	De 30 a 35 años 49 años o mayor	12,42% 12,24%
Nivel educativo madre	Ninguno (0) Secundaria (3) Universitaria (6)	3,79% 18,6% 4,07%	Preescolar (1) Media (4) Postgrado (7)	0,09% 36,96% 0,39%	Primaria (2) Técnica/ Tecnológica (5)	26,84% 9,2%

9. Privación respecto a condiciones vitales con relación a educación, situación de la niñez y juventud, salud, trabajo, acceso a servicios públicos domiciliarios y vivienda (DANE, 2014).

Privaciones

Pobreza multidimensional	Si	30,03%	No	69,97%		
Privación por bajo logro educativo	Si	54,17%	No	45,83%		
Privación por analfabetismo	Si	10,47%	No	89,53%		
Privación por inasistencia escolar	Si	9,84%	No	90,16%		
Privación por rezago escolar	Si	45,52%	No	54,48%		
Privación acceso a cuidados para PI	Si	18,68%	No	81,32%		
Privación por desempleo de larga duración	Si	20,11%	No	79,89%		
Privación por trabajo informal	Si	73,07%	No	26,93%		
Privación por acceso a salud	Si	4,73%	No	95,27%		
Privación por hacinamiento	Si	38,12%	No	61,88%		

Fuente: Elaboración propia

1.3 Características de la vivienda

Aunque el 68,62% de las niñas, niños y adolescentes vive en viviendas ubicadas en las cabeceras municipales y el 85,91% vive en una casa o apartamento, el 43,46% vive en hogares que deben compartir la cocina o el sanitario con otros hogares u otras viviendas, solo el 14,5% tiene acceso a internet, el 60,52% a alcantarillado, el 70,87% a recolección de basuras, el 53,14% a gas y el 75,11% a acueducto. En promedio sus hogares reciben agua solo durante 3,99 días a la semana (desviación estándar de 3,32). La mayoría de las niñas, niños y adolescentes viven en viviendas en estrato 1, 2 o sin estrato (94,09%).

Tabla 4. Características de la vivienda

Vivienda						
Ubicación de la vivienda – Clase	Cabecera	68,62 %	Centro poblado	10,54 %	Entre 12 y 17	33,83 %
Tipo de vivienda	Casa (1) Otro (4)	64,24 % 0,56 %	Apartamento (2) Vivienda indígena (5)	21,67 % 0,85 %		
Estrato	No tiene 3 6	8,19 % 5,64 % 0,01 %	1 4	56,77 % 0,23 %	Otro	39,61 %
Comparte cocina o sanitario	Si	43,46 %	No	56,54 %		
Acceso a energía eléctrica	Sí	94,6 %	No	5,4 %		
Acceso a acueducto	Si	75,11 %	No	24,89 %		
Acceso a alcantarillado	Sí	60,52 %	No	39,48 %	Otro	39,61 %
Acceso a gas	Si	53,14 %	No	46,86 %		
Acceso a recolección de basuras	Sí	70,87 %	No	29,13 %		
Acceso a Internet	Si	14,5 %	No	85,5 %		

Fuente: Elaboración propia

1.4 Características del municipio

Las niñas, niños y adolescentes se distribuyen en mayor proporción en los municipios con menor porcentaje de población rural (el promedio ponderado por la cantidad de niñas, niños y adolescentes en cada municipio es 32,33 % y el promedio simple de porcentaje de población rural para los municipios es 55,59%¹⁰), menor proporción de población étnica (promedio ponderado de 12,86 % y promedio simple de 13,07 %), menor proporción de población indígena (promedio ponderado de 4,45 % y promedio simple 5,87 %), mayor tasa de fecundidad por cada 100.000 mujeres de 10 a 14 años (promedio ponderado 299,44 y promedio simple de 284,1), y mayor tasa de fecundidad por cada 100.000 mujeres de 15 a 19 años (promedio ponderado 6.517,78 y promedio simple de 6242,21). Con respecto a la violencia, la tasa de violencia al interior del hogar por cada 100.000 habitantes (139,19 promedio ponderado y 82,45 promedio simple), la tasa de violencia interpersonal por cada 100.000 habitantes (210,49 promedio ponderado y 139,3 promedio simple), y la tasa de víctimas de delitos sexuales por cada 100.000 habitantes (52,09 promedio ponderado y 45,85 promedio simple) son mayores en los municipios donde se encuentra mayor cantidad de niñas, niños y adolescentes.



10. El promedio ponderado por la cantidad de niñas, niños y adolescentes da mayor peso al valor asociado a los municipios con mayor cantidad. Así, por ejemplo, el hecho de que el promedio ponderado de población rural sea menor que el promedio simple, indica que el porcentaje de población rural es menor en los municipios con mayor cantidad de niñas, niños y adolescentes.

2. Metodología de los Modelos de Probabilidad de Vulneración de Derechos de Niñas, Niños y Adolescentes

En el marco de la Alianza Nacional contra la Violencia hacia Niñas, Niños y Adolescentes, el ICBF, en conjunto con el DNP y en el marco del Subsistema de Protección de Derechos, ha trabajado en el desarrollo de modelos predictivos de vulneraciones que afectan a las niñas, niños y adolescentes. Esto en línea con el objetivo de «crear las condiciones para anticipar y resolver las violencias y vulneraciones contra niñas, niños y adolescentes», contenido en el Plan Nacional de Desarrollo 2018-2022.

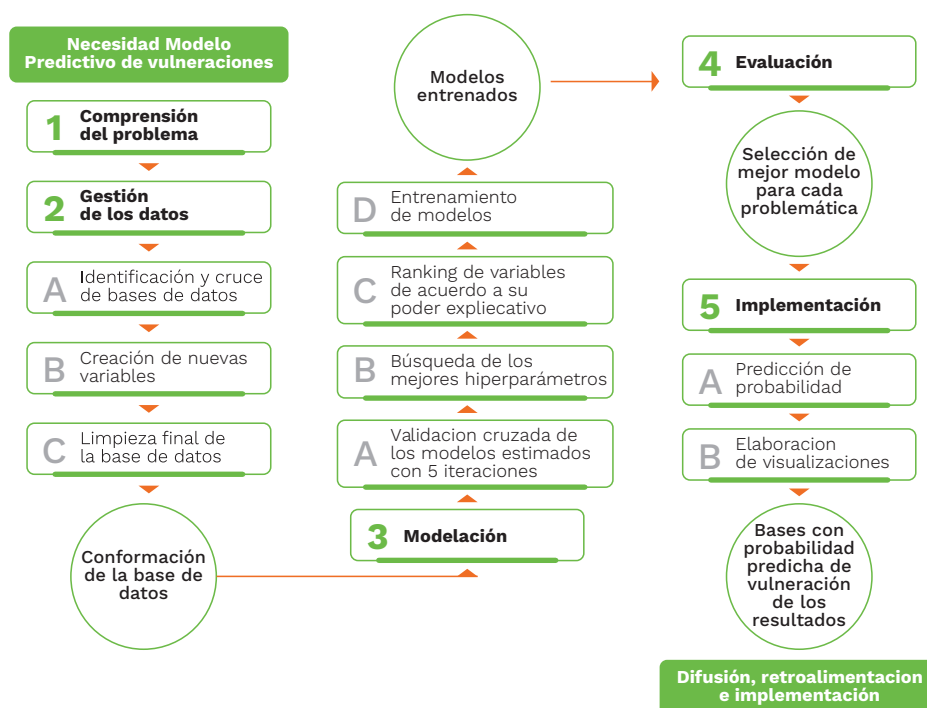
Los modelos de probabilidad de vulneración de derechos de niñas, niños y adolescentes permiten predecir el riesgo asociado a los individuos con base en las características socioeconómicas de las niñas y niños, sus familias y sus entornos. Los modelos identifican patrones complejos en grandes conjuntos de datos mediante un algoritmo de aprendizaje de máquinas (machine learning) que revisa los datos y es capaz de predecir la probabilidad o riesgo de que se presente una vulneración de derechos.

Los dos modelos que se documentan en este boletín permiten estimar el riesgo de que una niña, niño o adolescente sea víctima de violencia física, psicológica o negligencia (maltrato), o de violencia sexual. Estas vulneraciones de derechos fueron priorizadas debido al daño y afectación que causan a la integridad personal y desarrollo integral de esta población, así como a que en conjunto representaron el 57,7 % de los ingresos al PARD en 2019 y el 68,8 % en 2020.

La metodología para el desarrollo de los dos modelos de probabilidad de vulneración se desarrolló en 5 etapas (ver Diagrama 2):

Diagrama 2. Metodología desarrollo modelos de probabilidad de vulneración

Fuente: Elaboración propia



1. Comprensión del problema: Después de haber identificado la necesidad de desarrollar modelos que permitan estimar la probabilidad de ocurrencia de violencias a niñas, niños y adolescentes, se inició la fase de comprensión del problema que incluyó el desarrollo de modelos probabilísticos preliminares que fueron difundidos en instancias internas y externas al ICBF. Esto permitió contar con un conocimiento acumulado de la problemática y con la retroalimentación necesaria para guiar los siguientes pasos.

2. Gestión de los datos: esta etapa tiene como objetivo generar la base de datos que alimenta el entrenamiento de los modelos de probabilidad de vulneración. Su producto es una base de datos con 5.936.805 registros de niñas, niños y adolescentes (equivalente al 82 % del total nacional de niñas, niños y adolescentes registrados en la base Sisbén IV con corte a junio de 2021) y las variables descritas en el Anexo 1, que permiten caracterizar sus condiciones socioeconómicas, las de sus familias y entornos. Los pasos seguidos para obtener la base de datos son:

a. Identificación y cruce de bases de datos: las principales fuentes de información empleadas en este ejercicio son las descritas en el título de esta sección 1. Datos. Estas delimitan el universo de niñas, niños y adolescentes (con la base de Sisbén IV); introducen las variables de resultado (la información de ingreso a PARD por violencia física, psicológica o negligencia (maltrato) o violencia sexual) que actuarán como producto de referencia para el algoritmo de aprendizaje de máquinas (machine learning); el resto de las variables que permitirán caracterizar a las niñas, niños y adolescentes, sus hogares y sus entornos. Este paso incluye también la exploración de los datos con estadísticas descriptivas o regresiones para validar supuestos y probar hipótesis.

b. Creación de nuevas variables a partir de los datos originales (Feature engineering): a partir del conocimiento de la problemática y la información disponible en las bases de datos utilizadas se crearon nuevas variables para caracterizar a las niñas, niños y adolescentes. Algunas variables creadas son: Tipo de familia (identifica si el niño vive con sus dos padres, solo con su padre / madre, solo con su padre / madre y el cónyuge de este o esta, o sin ninguno de sus dos padres); Número de NNA (indica el número de menores de edad que habita en el hogar); Extraedad escolar (indica si la privación por tener una mayor edad a la que deberían tener de acuerdo al grado educativo en el que están matriculados se identifica para algún miembro del hogar); Unión temprana (indica si la niña, niño o adolescente se encuentra en unión temprana debido a que tiene hijos, está embarazada, o si su estado civil es casado o en unión libre).

c. Limpieza final de la base de datos: Dado que la principal fuente, el Sisbén IV, no recolecta información para todas las personas todos los años, no siempre se cuenta con información socioeconómica de las niñas, niños y adolescentes

para el mismo año en que sucedió la vulneración de sus derechos por causa de violencia. Por esta razón, se decidió excluir de la etapa de modelación los registros de aquellas niñas, niños o adolescentes vulnerados para los cuales no se dispone de información socioeconómica reciente. Es decir, en los que la diferencia entre el año de la vulneración y el año en que se realizó la encuesta Sisbén es mayor a tres.

De acuerdo con lo anterior, el universo de estudio se limitó a los registros de niñas, niños y adolescentes en Sisbén IV que cumplen con los siguientes criterios: 1) la fecha de encuesta es igual o mayor al año 2017; y 2) la diferencia entre la fecha de registro en el Sisbén, la fecha de ingreso al PARD y la fecha de las demás variables utilizadas es máximo de tres años. Al aplicar estas restricciones, la base se redujo de 7.236.230 millones registros a 5.936.805.

3. Modelación: el objetivo de esta etapa es contar con modelos confiables (cuyos resultados sean robustos ante distintas especificaciones del modelo y de los datos) y con un alto porcentaje de precisión (que permitan identificar de forma acertada a las niñas, niños y adolescentes con un alto riesgo de ser vulnerados).

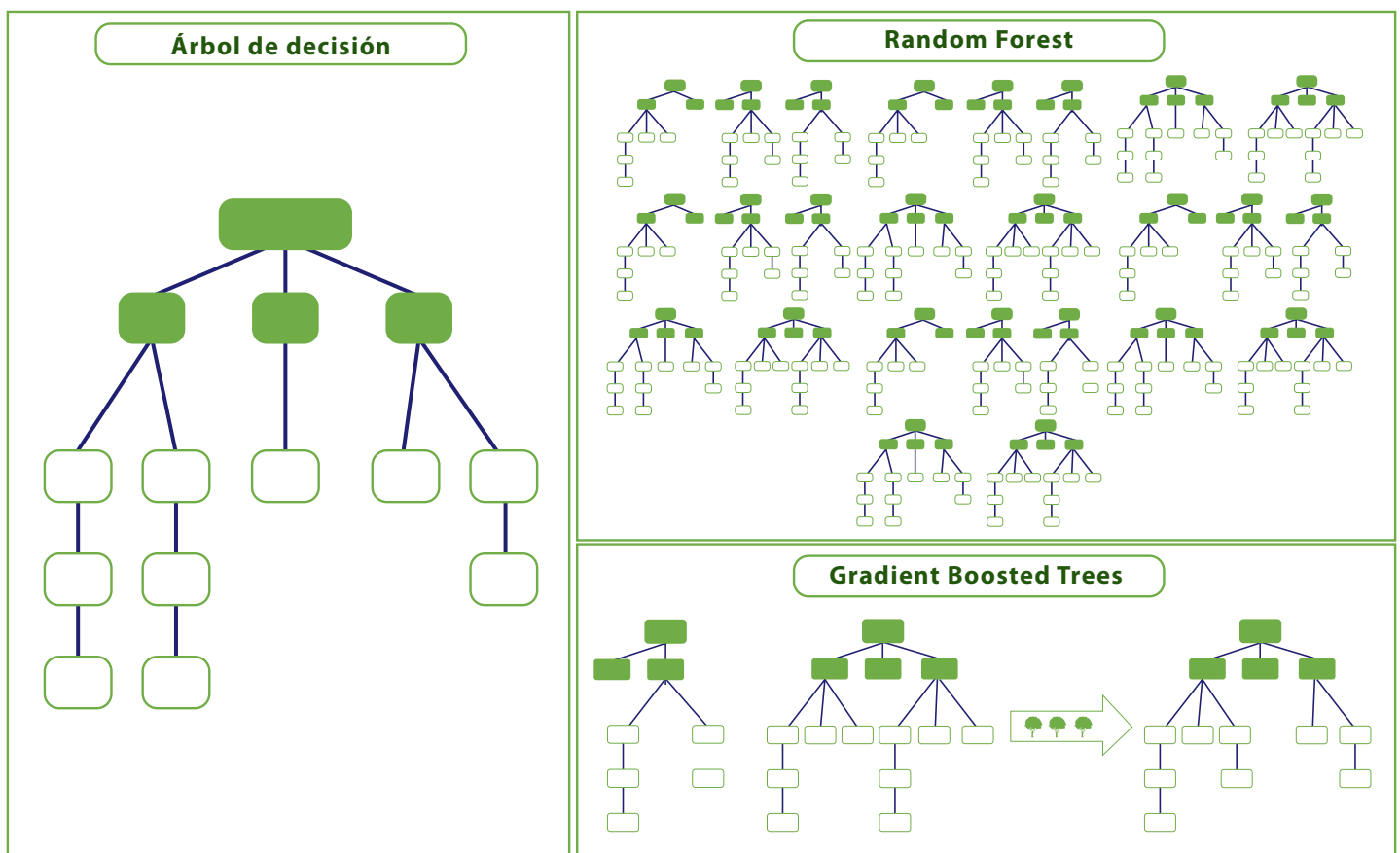
Garantizar la confiabilidad y precisión de los modelos de analítica predictiva es especialmente importante en los casos en los que los datos son desbalanceados, como es el de los modelos de probabilidad de vulneración por violencia contra niñas, niños y adolescentes. Se dice que los datos son desbalanceados porque la proporción de niñas, niños y adolescentes cuyos derechos han sido vulnerados por causa de violencia sexual y han ingresado al PARD es cercana al 0,34% de los registros en el Sisbén IV y de 0,24% en el caso de violencia física, psicológica o negligencia. En términos de probabilidad, esto indica que el registro de vulneración de derechos en niñas, niños y adolescentes por causa de violencia es un «evento raro».

En esta etapa se emplearon dos tipos de modelos que en la literatura se han destacado por su buen desempeño en tareas de clasificación: Random Forest (Bosque aleatorio) y Gradient Boosted Trees (Árbol de potenciación del gradiente). Ambos parten de árboles de decisión (Decision Trees) que permiten predecir el valor de una variable de destino (en este caso la vulneración por violencia), partiendo de variables de entrada (las características de las niñas, niños y adolescentes y sus entornos) a partir de una serie de pasos secuenciales (ver Diagrama 3).

Un árbol de decisión es simple de ejecutar y entender, pero una de sus mayores limitaciones es que cambios pequeños en los datos de entrenamiento, potencialmente, pueden causar grandes cambios en el resultado final. Los modelos de Random Forest y Gradient Boosted Trees corrigen esta limitación creando múltiples árboles de decisión con diferentes parámetros. En el Random Forest los árboles de decisión son creados de forma independiente con submuestras

de datos y de variables escogidas aleatoriamente; al final sus resultados son promediados o agregados por la regla de mayorías para obtener predicciones más cercanas a los valores reales. Por su parte, en el modelo de Gradient Boosted Trees los árboles de decisión se crean de manera secuencial, es decir, para crear un árbol se observa el error de todos los creados anteriormente, de modo que el último lo haga un poco mejor, por lo que los resultados de los diferentes árboles de decisión se van combinando durante el proceso. El proceso se termina cuando el desempeño de todos los árboles de decisión combinados se considera adecuado (Glen, 2019).

Diagrama 3. Esquema de árbol de decisión, Random Forest y Gradient Boosted Trees



Fuente: Elaboración propia

En el entrenamiento y evaluación de los modelos se siguieron estas estrategias:

a. Validación cruzada de los modelos estimados con 5 iteraciones (5-fold cross-validation): la validación cruzada consiste en dividir los datos de forma aleatoria en grupos de, aproximadamente, el mismo tamaño, de modo que en

la versión con 5 iteraciones, en cada una el 20 % de los datos se emplea para la evaluación del modelo, mientras que el 80 % restante se emplea para el entrenamiento. Este proceso genera 5 estimaciones del error cuyo promedio se emplea como estimación final. De esta forma, se busca que todos los datos sirvan tanto para entrenar como para evaluar el modelo y que los resultados sean robustos, sin importar el conjunto de datos elegido.

b. Búsqueda de mejores hiperparámetros (Grid search): para los modelos de Random Forest y Gradient Boosted Trees se probaron y evaluaron distintas combinaciones de parámetros como, por ejemplo, el número de árboles y la profundidad de los árboles (número de brazos desde la raíz del árbol hasta un nodo). Adicionalmente, para el modelo de Gradient Boosted Trees se ajustó la tasa de aprendizaje o regularización (Learning rate) que controla el grado de mejora de un árbol con respecto al anterior, es decir, el ritmo al que aprende el modelo.

c. Ranking de variables de acuerdo con su poder explicativo: para evaluar la relación entre la vulneración (variable objetivo) y cada una de las variables explicativas, se realizaron pruebas chi-cuadrado y ANOVA que permiten probar hipótesis referidas a distribuciones de frecuencias. También se calculó la importancia de cada variable explicativa (Feature importance) con modelos Logit, Random Forest y Gradient Boosted Trees. Tomando en cuenta estos criterios, se realizó un ranking de variables para hacer cuatro grupos de variables de acuerdo con su poder explicativo: un grupo que contenía el 25 % de variables más importantes, otro que contenía el 50 %, otro con el 75 % y, finalmente, un grupo con la totalidad de las variables.

d. Entrenamiento de modelos: Para el entrenamiento tanto de los modelos de Random Forest y Gradient Boosted Trees se utilizó la librería XGBoost (<https://xgboost.readthedocs.io/en/latest/>). Esta consiste en una colección de recursos ampliamente usada para resolver problemas de ciencia de datos de una manera rápida y precisa. La librería implementa algoritmos de aprendizaje de máquinas bajo el marco de Gradient Boosting. El modelo de Random Forest puede ser entrenado en la librería también si se entiende como un caso particular del modelo de Gradient Boosted Trees en el que la tasa de aprendizaje es igual a cero.

4. Evaluación: el desempeño de los modelos entrenados fue evaluado teniendo en cuenta los criterios de predicción y simplicidad. Con respecto al primer criterio se tuvieron en cuenta métricas de desempeño como el acierto en la predicción en los vulnerados (Recall), y el acierto promedio de predicción para los vulnerados y no vulnerados (Balanced accuracy). Bajo el segundo criterio, si se encontraba que dos modelos tenían un desempeño en predicción similar, pero uno empleaba menos variables en su entrenamiento, se prefería el modelo con menos variables. De esta manera, para la etapa de implementación, se eligió el modelo con un mejor balance entre desempeño en predicción y simplicidad.

5. Implementación: los modelos con mejor desempeño para cada una de las problemáticas fueron versiones de Gradient Boosted Trees. A partir del entrenamiento se obtuvieron los productos:

- a. Predicción de la probabilidad que tiene cada niña, niño o adolescente en la base de sufrir una vulneración por violencia física, psicológica o negligencia (maltrato) o violencia sexual. Esta predicción se realiza sobre el total de 7.236.230 niños, niñas y adolescentes que se encuentran en la base del Sisbén IV al corte de junio de 2021.
- b. Visualizaciones de los resultados que identifican las características más relacionadas con la vulneración y las zonas geográficas que concentran niñas, niños y adolescentes con un alto riesgo de vulneración.

3. Ética y discriminación

Los modelos de analítica predictiva han demostrado experimentalmente que pueden predecir tendencias y patrones de comportamiento de forma más acertada que modelos multivariados estándar o juicios expertos basados en la experiencia o la intuición (Pryce et al., 2018). En particular, Chouldechova et al. (2018) indica que los juicios humanos pueden ser afectados por sesgos cognitivos como sobrevalorar casos recientes especialmente violentos, pero no relacionados con el caso actual (heurística de disponibilidad), o evaluar la probabilidad de vulneración con base en características fáciles de observar como el barrio donde se vive, para aproximarse a condiciones de riesgo más pertinentes como el uso de frecuente de sustancias psicoactivas (discriminación estadística o heurística de representatividad). El uso de los algoritmos podría entonces informar y mejorar las decisiones humanas para que sean más eficientes y justas.

Sin embargo, existe la preocupación de que el uso de modelos de analítica predictiva pueda llevar a la estigmatización de grupos vulnerables solo porque los gobiernos usualmente tienen más información de los mismos debido a que son los principales beneficiarios de programas sociales. Así, las familias más vulnerables serán identificadas como de alto riesgo e investigadas con mayor probabilidad, lo que reforzará el sesgo original, si el algoritmo toma en cuenta el hecho de haber sido investigado anteriormente por una situación de riesgo como un predictor de mayor riesgo futuro. Esto generaría un problema de calibración, lo que significa que el algoritmo sobreestimaría el riesgo para algunos grupos con relación a otros.

Siguiendo a Chouldechova et al. (2018), un modelo de analítica predictiva estaría bien calibrado con respecto a una variable asociada con vulnerabilidad (en el ejemplo propuesto en el artículo la raza), si para cada puntaje de riesgo s ($S=s$), la proporción de casos con este puntaje que efectivamente sufren el evento adverso ($Y=1$) es la misma para cada grupo de raza ($G \in \{g_1, g_2, \dots, g_k\}$). En otras palabras, para ningún grupo (y especialmente para ningún

grupo vulnerable) el modelo tiene una mayor proporción de casos en los que de acuerdo con la predicción ocurriría el evento adverso, pero en realidad este no ocurre, lo que se conoce como falso positivo¹¹.

$$P(Y=1|S=s,G=g_j) = P(Y=1|S=s) \quad \forall j \quad (1)$$

Alternativamente,

$$P(\text{evento adverso}=1|S=s,G=\text{Caucásico}) = P(\text{evento adverso}=1|S=s,G=\text{Afroamericano}) = P(\text{evento adverso}=1|S=s,G=\text{Latino}) = P(\text{evento adverso}=1|S=s,G=\text{Asiático})$$

de los grupos de Sisbén IV. Los resultados de esta evaluación que se presentan en la siguiente sección, señalan que no existiría discriminación en las predicciones del modelo asociadas a las condiciones de vulnerabilidad identificadas por los grupos del Sisbén IV. Este resultado es positivo, pero es necesario recordar que el modelo requiere constante revisión y posibles ajustes posteriores para mantener o mejorar su desempeño ante cambios en los datos y procesos que puedan ocurrir.



11. Para mayor detalle sobre falsos positivos, ver el título 1 de la sección III. Resultados.

III. Resultados

Gracias a la complejidad de su algoritmo los modelos de analítica predictiva son mejores prediciendo tendencias y patrones de comportamiento que los modelos multivariantes estándar. Sin embargo, esta misma complejidad hace que la interpretabilidad y comunicación de sus resultados sea más complicada que en modelos más simples como un probit o logit. En esta sección, se van a presentar los resultados de los modelos de probabilidad de vulneración por causa de violencia física, psicológica o negligencia (maltrato) y por causa de violencia sexual desde cuatro diferentes aproximaciones para facilitar su comprensión:

1. Evaluación del desempeño de cada modelo en general y con respecto a cada uno de los grupos de Sisbén IV, usando diferentes métricas que permiten valorar el rendimiento de modelos de clasificación en analítica predictiva.
2. Presentación de las variables con mayor importancia en la probabilidad de vulneración estimada, usando la herramienta denominada valores SHAP.
3. Caracterización de los casos con probabilidad alta de ser vulnerados que corresponden al 20 % de los registros con mayor riesgo de vulneración estimado en cada modelo.
4. Distribución geográfica de los casos con mayor probabilidad de vulneración.

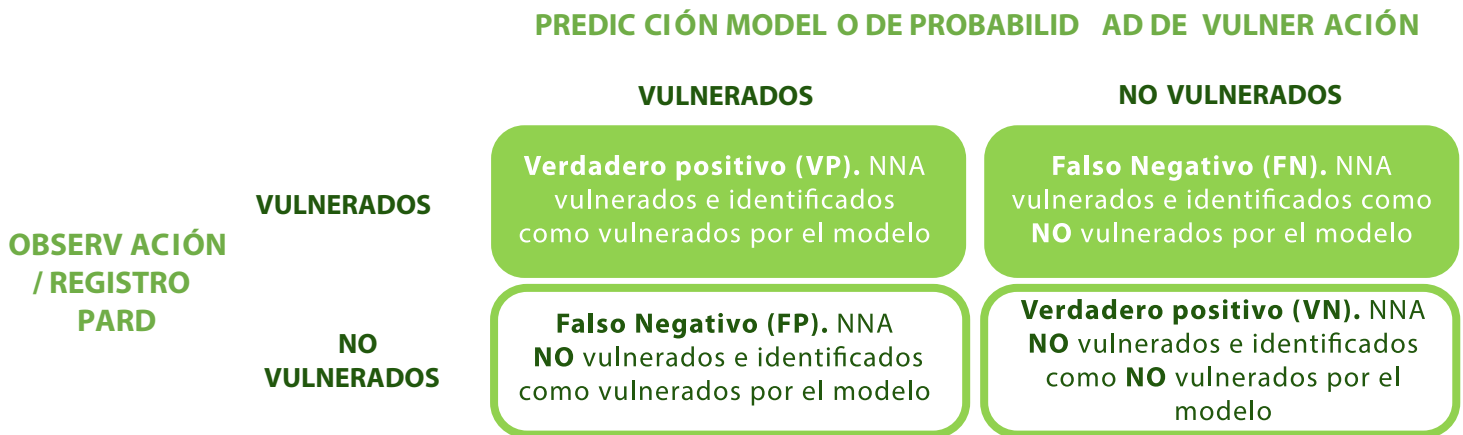
1. Modelo de probabilidad de vulneración por violencia física, psicológica o negligencia (maltrato)

1.1 Desempeño general

En general, el desempeño del modelo de probabilidad de vulneración por violencia física, psicológica o negligencia (maltrato) se evalúa determinando la capacidad de este para identificar correctamente las niñas, niños y adolescentes para los cuales se registrará vulneración de derechos y aquellos para los cuales no se registrará.

Una manera de explicar visualmente cómo se evalúa el desempeño del modelo es emplear una matriz de confusión (ver Diagrama 4). En cada columna de la matriz se muestra el resultado estimado del modelo frente a la vulneración o no de las niñas, niños y adolescentes y en cada fila se muestra el estado observado frente al registro de la vulneración o no de las niñas, niños y adolescentes en un PARD. Así, por ejemplo, en el primer cuadrante de la matriz se encuentran las niñas niños y adolescentes que el modelo predice que serán vulnerados y además se observa en los registros que son vulnerados (verdaderos positivos). Un modelo será mejor si puede maximizar el número de verdaderos positivos y de verdaderos negativos y, consecuentemente, minimizar el número de falsos positivos (niñas, niños y adolescentes para los cuales el modelo predice vulneración, pero para los cuales no se observa vulneración) y falsos negativos (niñas, niños y adolescentes para los cuales el modelo predice no vulneración, pero para los cuales se observa vulneración).

Diagrama 4. Matriz de confusión del modelo de probabilidad de vulneración

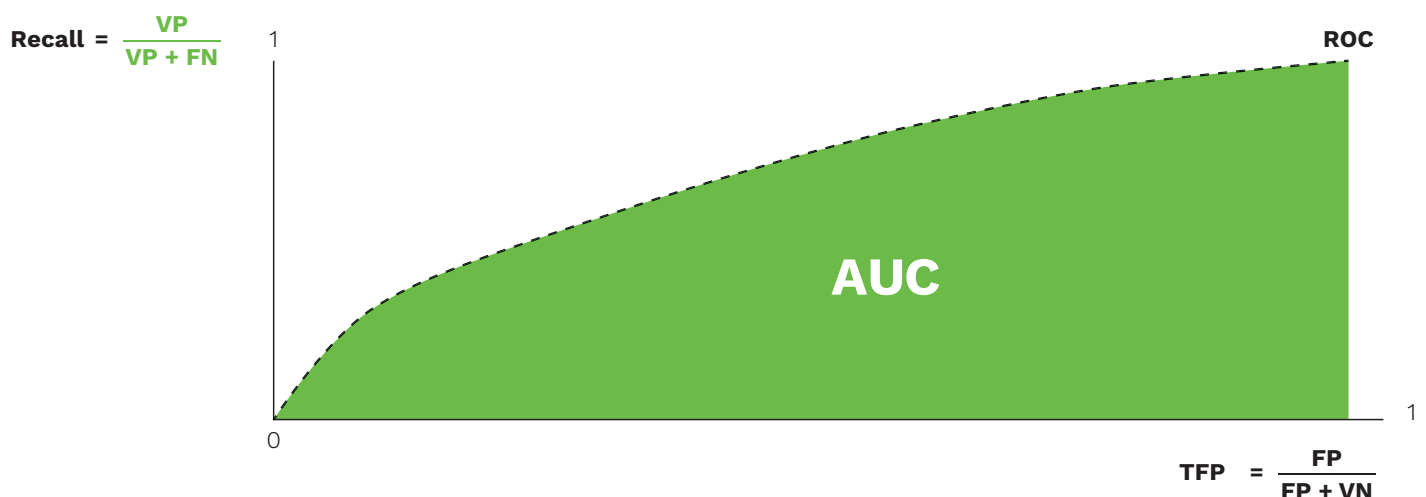


Fuente: Elaboración propia

Existen una serie de métricas para evaluar el desempeño del modelo entre las cuales la más general es la AUC (Area Under the Curve), que es una medida resumen de la habilidad del modelo para distinguir entre casos positivos y negativos en diferentes umbrales de clasificación. Un mayor valor de AUC indica un mejor desempeño del modelo, siendo 1 una clasificación perfectamente precisa (sin falsos positivos ni falsos negativos).

El valor de la AUC corresponde al área por debajo de la ROC (curva del operador del receptor) que relaciona la tasa de verdaderos positivos del modelo (llamada Exhaustividad o Recall) con la tasa de falsos positivos (TFP) en diferentes cortes de riesgo para determinar si una niña, niño o adolescente será registrado como vulnerable o no (Umbral de clasificación). La ROC es siempre creciente debido a que a medida que se reduce el umbral de clasificación, el modelo clasificará más casos como positivos, aumentando tanto los verdaderos positivos como los falsos positivos (Diagrama 5).

Diagrama 5. AUC (área bajo la curva ROC)



VP=Verdadero positivo, FN= Falso negativo, FP=Falso positivo, VN= Verdadero negativo

Fuente: Elaboración propia

Para el modelo de probabilidad de vulneración por violencia física, psicológica o negligencia (maltrato) el valor de la AUC en general es de 92,38 % lo que se considera sobresaliente. Cuando se calcula la AUC del modelo para cada uno de los grupos del Sisbén IV es posible observar que se mantiene un desempeño sobresaliente (por encima del 90 %) y que no se observan variaciones importantes, por lo cual se considera consistente el resultado obtenido de la AUC general.

Otra medida general para evaluar el desempeño del modelo de probabilidad de vulneración de maltrato es la precisión promedio para diferentes umbrales de clasificación. Esta corresponde a la proporción de casos verdaderos positivos sobre el total de casos verdaderos positivos y falsos positivos. La precisión promedio del modelo es de 25,2 %, indicando que en promedio para cada umbral de clasificación 1 de cada 4 niñas, niños y adolescentes identificados como en riesgo de vulneración por maltrato serán efectivamente maltratados. La precisión promedio del modelo para el grupo D de Sisbén IV (33,47 %) es mayor que en los demás grupos, indicando que en promedio hay menos casos de falsos positivos en este grupo. La precisión promedio es menor para los grupos B y C (25,83 % y 25,53 % respectivamente) que para el grupo A (27,79 %), por lo que el modelo la proporción de casos falsos positivos no aumenta de forma monótona con el nivel de vulnerabilidad.

El desempeño del modelo de probabilidad de vulneración por violencia física, psicológica o negligencia (maltrato) posteriormente se evaluó mediante otras métricas empleando el umbral de clasificación en el cual se maximiza el valor del Balanced Accuracy (BACC) para la totalidad de los registros y los registros pertenecientes a cada grupo del Sisbén IV (ver Tabla 5 línea 3).

Tabla 5. Desempeño modelo de predicción de vulneración por violencia física, psicológica y negligencia (maltrato) – Total y por grupo de Sisbén IV

Medidas/Grupos	Total	A	B	C	D
1. AUC	92.380 %	92.435 %	91.714 %	91.057 %	94.842 %
2. Precisión promedio	25.209 %	25.791 %	23.833 %	23.532 %	33.466 %
3. Prevalencia de vulneración verdadera	0.242 %	0.297 %	0.269 %	0.220 %	0.153 %

Medidas/Grupos	Total	A	B	C	D
4. Mejor umbral de clasificación Balanced Accuracy (BACC)	0.271 %	0.356 %	0.349 %	0.281 %	0.306 %
4.1. Prevalencia de Vulneración Estimada en el umbral de BACC	16.917 %	15.800 %	15.536 %	15.606 %	9.237 %
4.2 Recall en el umbral de BACC VP/(VP+FN)	85.180 %	84.027 %	81.835 %	81.173 %	84.241 %
4.4. Especificidad en el umbral de BACC (VN/(VN+FP))	83.248 %	84.403 %	84.643 %	84.538 %	90.878 %
4.4. Balanced Accuracy en el umbral de BACC	84.214 %	84.215 %	83.239 %	82.856 %	87.560 %
4.5. Precisión en el umbral de BACC (VP/(VP+FP))	1.217 %	1.581 %	1.416 %	1.145 %	1.399 %

VP=Verdadero positivo, FN= Falso negativo, FP=Falso positivo, VN= Verdadero negativo

Fuente: Elaboración propia

La métrica Balanced Accuracy promedia el valor de los valores de Recall (tasa de verdaderos positivos) y de la Especificidad (Specificity) que equivale a la tasa de verdaderos negativos. Esta medida es adecuada para valorar el desempeño de modelos de analítica predictiva cuando los datos no están balanceados, es decir, cuando la mayoría de los registros se encuentran en una sola clase. Este es el caso de la vulneración por causa de violencia física, psicológica o negligencia (maltrato), que se observa solo en el 0,24 % de las niñas, niños y adolescentes considerados. La tasa de prevalencia verdadera u observada es mayor entre los pertenecientes al grupo A de Sisbén IV y disminuye a medida que se reduce la vulnerabilidad. La prevalencia de vulneración estimada por el modelo es mucho mayor que la observada (16,92 %) y también disminuye a medida que se reduce el nivel de vulnerabilidad asociado a cada grupo del Sisbén IV.

El valor del Balanced Accuracy observado es superior al 82 % para todos los grupos considerados. El Recall (tasa de verdaderos positivos) es de 85,18 %, indicando que aproximadamente 17 de cada 20 niños que serán maltratados serán identificados por el modelo. El valor de especificidad indica que el modelo puede identificar correctamente el 83,25 % de los casos en los que una niña, niño o adolescente no será registrado como objeto de vulneración por causa de violencia física, psicológica o negligencia (maltrato), o equivalente que solo en el 16,75 % de los casos el modelo dará falsas alarmas sobre la vulneración de derechos. Los valores de Recall y Especificidad son similares entre los grupos de Sisbén IV, excepto por el

valor de Especificidad para el grupo D que indica una proporción aproximadamente 5 % menor de falsas alarmas.

La medida de Precisión es baja (1,22 %), lo que se explica por el hecho de que solo para un porcentaje muy pequeño de las niñas, niños y adolescentes en la base de datos se reportaron situaciones de vulnerabilidad por violencia física, psicológica o negligencia (maltrato) (los datos no están balanceados). La diferencia entre la prevalencia de vulneración verdadera (0.242 % para el total) y la prevalencia de vulneración estimada por el modelo (16.917 %) implica la existencia de una gran proporción de falsos positivos, lo que también afecta negativamente la Precisión del modelo. Este valor indica que el 1,22 % de las niñas, niños y adolescentes que se identifican como en riesgo de vulneración efectivamente entrarán a un proceso de restablecimiento de derechos (PARD), lo que puede reflejar el nivel de subregistro existente.

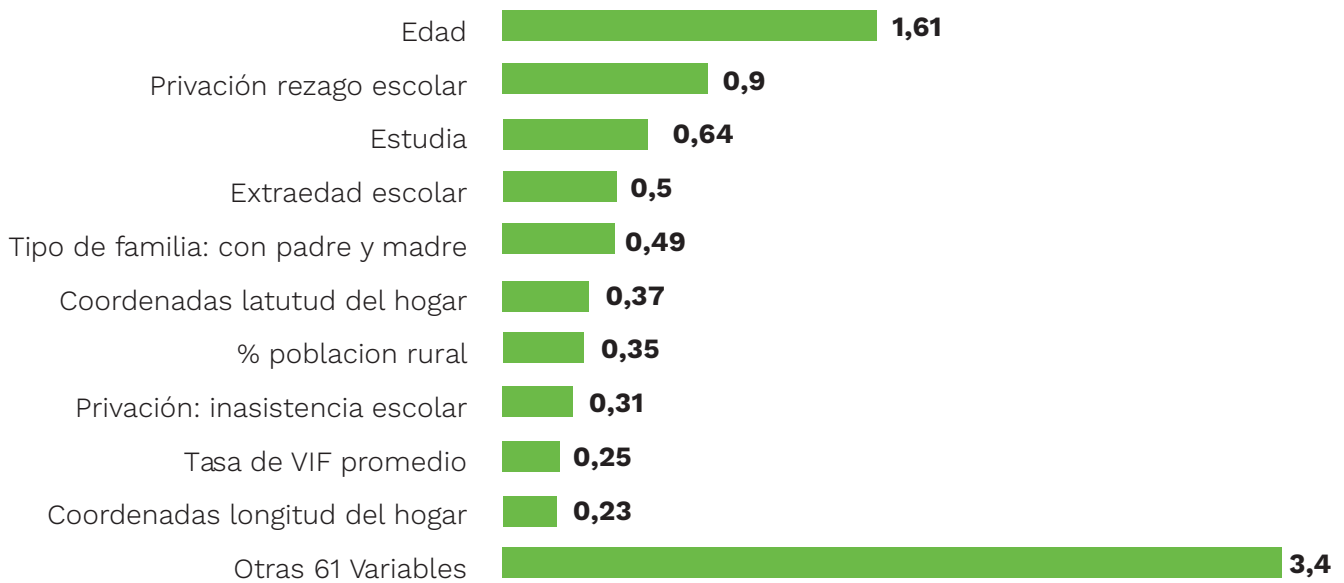
1.2 Valores SHAP

Los valores SHAP (Shapley Additive Explanations) son una herramienta cuyas ideas fundamentales surgen de la teoría de juegos y que permite identificar la contribución de cada variable en la probabilidad de vulneración estimada por el modelo. La intuición detrás de los valores SHAP parte de comparar un escenario inicial, en el que conocemos los resultados del modelo que incluye todas nuestras variables, y compararlo con los resultados de un escenario en el que no incluye una de estas variables. La diferencia entre ambos escenarios permite conocer cuál es el impacto real de dicha variable en la predicción del modelo.

Las magnitudes de los valores SHAP de una misma variable se presentan como logaritmos de la razón de probabilidades (log odds ratio en inglés), que permiten medir la asociación entre dos variables. Valores negativos de los log odds ratio representan reducciones en la probabilidad estimada, mientras que valores positivos significan un efecto positivo en la probabilidad.

En la Gráfica 3 se muestran las 10 variables que tienen más influencia en la predicción de la probabilidad de vulneración por violencia física, psicológica o negligencia (maltrato), dicho de otra forma, las que más importancia tienen para el modelo al momento de realizar su predicción. Es preciso mencionar que los valores no indican una relación directa entre las variables y la probabilidad estimada, sino que señalan la medida en que cada variable afecta los resultados del modelo en comparación con un escenario en el que no se incluyera.

Gráfica 3. Valores SHAP Modelo de probabilidad de vulneración por violencia física, psicológica o negligencia (maltrato)

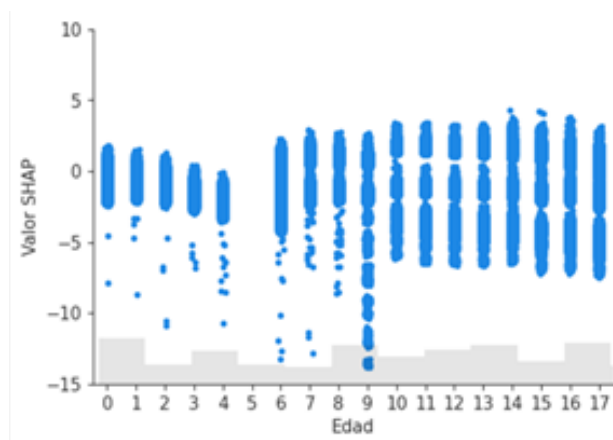


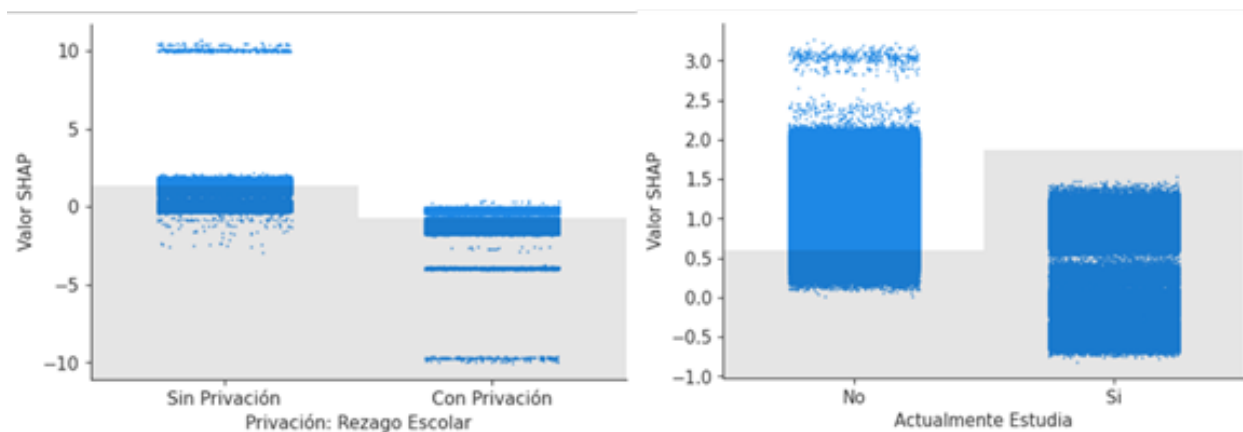
Fuente: Elaboración propia

Las variables más importantes en la predicción de la probabilidad de violencia física, psicológica y negligencia (maltrato), es decir, las que en su ausencia alterarían más la probabilidad estimada son la edad, la privación por rezago escolar al interior del hogar (para el mismo u otro miembro del hogar) y la variable que indica si está estudiando o no. Además de las 10 variables que más contribuyen en representar la probabilidad de vulneración estimada por el modelo, tenemos 61 variables con una participación de 3,4 en la contribución.

En la Gráfica 4 se presenta la relación entre las tres variables más importantes en la predicción de la probabilidad de violencia física, psicológica y negligencia (maltrato) y los valores SHAP asociados a las mismas variables.

Gráfica 3. Valores SHAP Modelo de probabilidad de vulneración por violencia física, psicológica o negligencia (maltrato)





Fuente: Elaboración propia

La información de edad indica que la probabilidad de ser víctimas de vulneración por violencia física, psicológica o negligencia (maltrato) es importante para las niñas y niños menores de dos años y disminuye gradualmente hasta los cuatro años. A partir de los cinco años aumenta la dispersión en los valores SHAP, es decir, hay niñas, niños y adolescentes en los que una determinada edad se asocia con una mayor probabilidad de ser vulnerados; mientras que, para otros, la probabilidad es menor en la misma edad. Esto indica que hay otras variables con una importante influencia al momento de explicar la probabilidad estimada del maltrato. Finalmente, a la edad de 17 años se puede ver que los efectos son en su mayoría negativos, lo que indica una reducción en la probabilidad de ser vulnerados.

Al observar la relación entre la variable de privación por rezago escolar y sus respectivos valores SHAP, se encuentra que las niñas, niños y adolescentes que viven en hogares sin esta privación tienen en promedio valores SHAP positivos. Sucede lo contrario en las niñas, niños y adolescentes que viven en hogares con esta privación, cuyos valores SHAP son menores y en mayor medida negativos. En este sentido, puede decirse que la privación por rezago escolar aumenta la probabilidad de maltrato estimada. Finalmente, el análisis de la variable que indica si la niña, niño o adolescente se encuentra estudiando en una institución educativa indica que, en promedio, la condición de no estudiar aumenta la probabilidad que el modelo estima de ser vulnerado por violencia física, psicológica o negligencia (maltrato), en comparación con las niñas, niños y adolescentes que sí estudian.

1.3 Caracterización de los casos con probabilidad alta de ser vulnerados

Con el objetivo de operacionalizar la probabilidad de vulneración por violencia física, psicológica o negligencia (maltrato) estimada por el modelo, se identificaron las características asociadas a las niñas, niños y adolescentes cuya probabilidad de vulneración estimada hace

parte del grupo con probabilidad de vulneración estimada alta (correspondiente al 20 % más alto).

La distribución entre niñas y adolescentes mujeres y, niños y adolescentes hombres en el grupo con el 20 % con mayor probabilidad de vulneración estimada por violencia física, psicológica o negligencia (maltrato) es similar, y refleja que la vulneración por causa de maltrato no se presenta con mayor frecuencia en ninguno de los sexos. Dado el rango de edad y sexo, la mayor participación en el grupo de probabilidad de vulneración estimada alta se da entre los niños de 6 a 9 años (26,5 %) y la más baja entre los adolescentes hombres entre 14 y 17 años (12,74 %).

Tabla 6. Probabilidad de vulneración por violencia física, psicológica o negligencia (maltrato) por nivel de riesgo

Rango Riesgo			Cantidad		
Probabilidad	Mín.	Max.	Total	Niñas y adolescentes mujeres	Niños y adolescentes hombres
20 % mayor probabilidad estimada	0,2712124 %	100 %	1.438.961	717.700 20,45 %	721.261 19,62 %
Total	0 %	100 %	7.219.935	3.526.565	3.693.370

Fuente: Elaboración propia

No hay diferencias importantes en la participación de las niñas, niños y adolescentes que estudian y los que no, y los que saben leer y escribir y los que no, pero las niñas, niños y adolescentes que se encuentran en extraedad tienen una mayor participación en el grupo con probabilidad de vulneración estimada alta (34,11 % frente a 19,32 % para los que no tienen esta condición). El acceso a internet se asocia con una menor participación (16,68 % frente a 20,57 %).

Estar en una unión temprana por matrimonio, tener hijos o estar en condición de embarazo, al igual que haber sido atendido por servicios de prevención del ICBF, no afecta de forma importante la participación en el grupo con probabilidad estimada alta de sufrir violencia física, psicológica o negligencia (maltrato).

En cuanto a la familia, las niñas, niños y adolescentes con familias compuestas por padre y madre tienen la menor participación en el grupo con probabilidad de vulneración alta (11,5 %). Por el contrario, vivir sin su padre ni su madre (35,39 %), solo con su padre (34,05 %) o con su padre y el cónyuge de él (33,16 %) tienen la mayor participación. El número de menores en el hogar se correlaciona de manera positiva con la participación en el grupo con probabilidad de vulneración alta. Para los hogares con más de 10 menores la participación es de 38,48 %. Las niñas y niños menores de 5 años que reciben otro tipo de cuidado (45,63 %), permanecen solos en casa (44,6 %) o al cuidado de un pariente menor de 18 años (35,97 %) tienen mayor participación en el grupo con probabilidad de vulneración por violencia sexual estimada alta. La edad de la madre no genera variaciones importantes (es de alrededor de 20 % en todos los

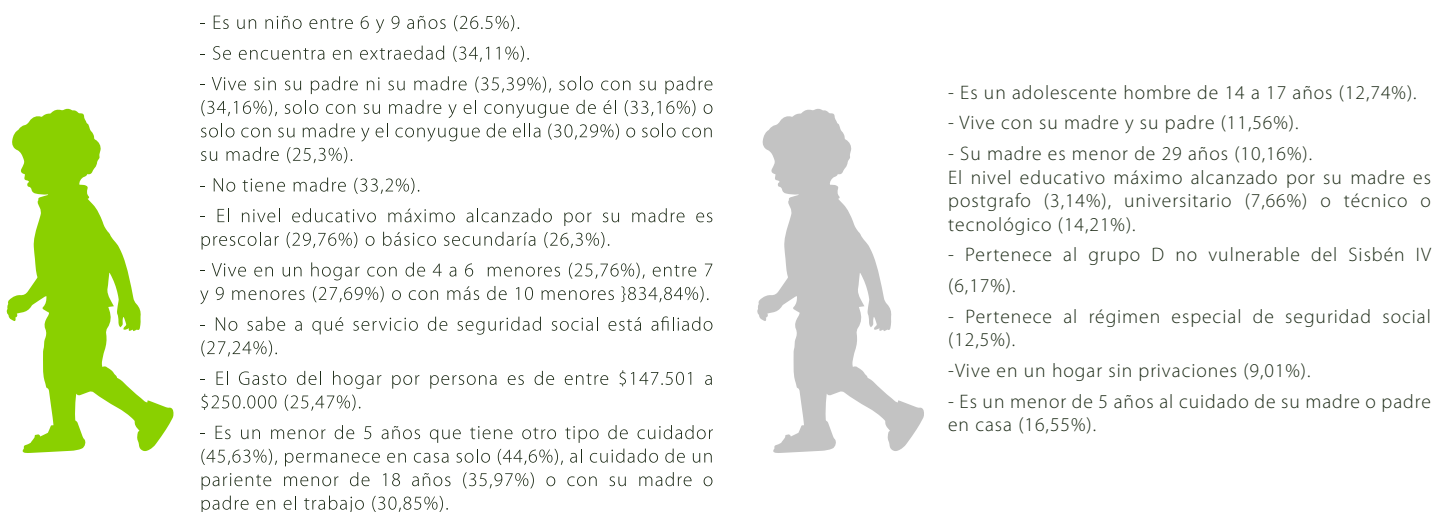
grupos) en la participación en el grupo con probabilidad de vulneración por violencia física, psicológica o negligencia (maltrato) estimada alta. Las niñas, niños y adolescentes que no tienen madre tienen mayor participación (33,2 %). Las niñas, niños y adolescentes cuya madre alcanzó el nivel educativo universitario o posgrado tienen menor participación en el grupo con probabilidad estimada alta (7,66 % y 3,14 %, respectivamente).

Se observa una menor participación en la probabilidad de vulneración alta estimada por violencia sexual en las niñas, niños y adolescentes pertenecientes en el grupo de Sisbén IV No vulnerable (D), y en los hogares sin privaciones. Sin embargo, la participación es mayor en los hogares con gasto mensual por persona mayores a \$147.501.

En la Gráfica 5 se presentan las características de las niñas, niños y adolescentes asociadas con mayor y menor proporción de casos con riesgo alto estimado de violencia física, psicológica o negligencia (maltrato). Para identificar las características asociadas con una mayor participación se tomó como referencia el valor de 25 % y para las que se asociación con una menor participación el valor de 15 %. Excepto para la variable que identifica el cuidador de las niñas y niños menores de 5 años, para la cual el valor de referencia para mayor participación es 30% y para menor participación 20 %.

En el Anexo 2 se puede encontrar la proporción de casos con riesgo alto estimado de vulneración por violencia física, psicológica o negligencia (maltrato) para el total nacional y respecto a variables seleccionadas. El ICBF ha puesto a disposición de la ciudadanía un tablero de visualización de resultados donde se pueden realizar consultas para el total nacional y para cada departamento y municipio¹².

Gráfica 5. Características de las niñas, niños y adolescentes asociadas con mayor y menor proporción de casos con riesgo alto estimado de violencia física, psicológica o negligencia (maltrato)



Fuente: Elaboración propia

12. Link para consulta detallada sobre el Modelo probabilístico de vulneración de derechos de niñas, niños y adolescentes: https://public.tableau.com/app/profile/anal.tica.institucional.icbf/viz/ModeloProbabilisticodeVulneracindeDerechos_16183556026450/Portada

1.4 Distribución geográfica de los casos con mayor probabilidad de vulneración

Los municipios con mayor proporción de casos de alta probabilidad estimada de vulneración por violencia física, psicológica o negligencia (maltrato) se identificaron tomando como referencia la cantidad de niñas, niños y adolescentes en cada municipio que hacen parte del 20 % de niñas, niños y adolescentes con mayor riesgo estimado de probabilidad de vulneración¹³. La cantidad en cada municipio fue dividida por el total de niñas, niños y adolescentes en cada municipio que se encuentran en la base del Sisbén IV.

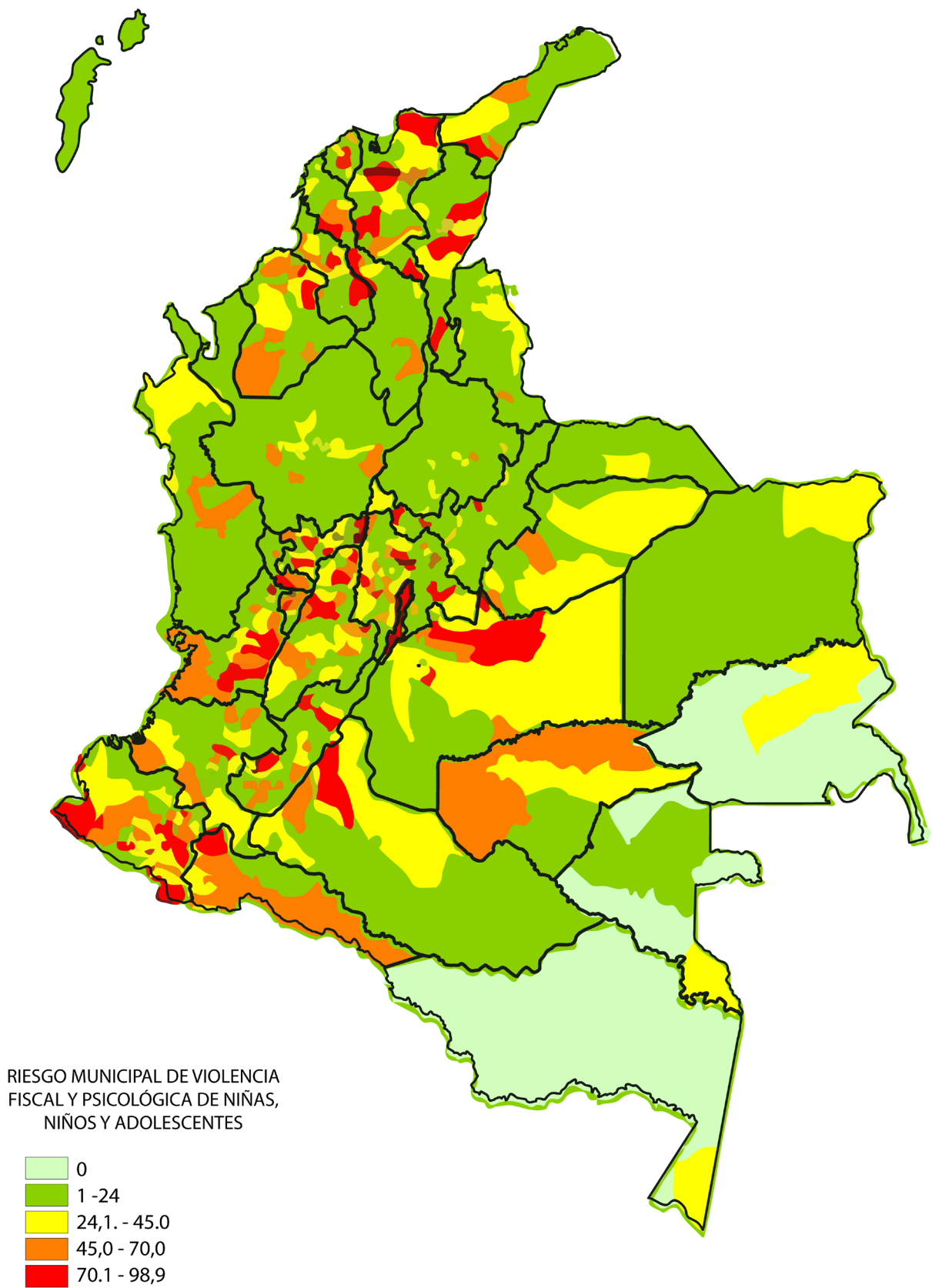
Los municipios con mayor cantidad de niñas, niños y adolescentes con riesgo alto de vulneración por violencia física, psicológica o negligencia (maltrato) son Túquerres-Nariño (97,46 %), La Unión-Nariño (96,82 %) y Salamina-Caldas (92,81 %); y en cuanto departamentos, se encuentra que Bogotá tiene la mayor proporción de niñas, niños y adolescentes con probabilidad alta de vulneración predicha sobre el total de esta población en la base del Sisbén IV (39,4 %), seguida por Nariño (37,83 %), Valle del Cauca (36,39 %) y Meta (35,99 %). De estos, solo Nariño hizo parte de los departamentos con mayor número de ingresos a PARD por maltrato en relación con la población de 0 a 17 años en 2019.

Actualmente el modelo permite focalizar microzonas de alto riesgo de violencias en todos los municipios del país y se proyecta combinar sus resultados con otros datos geográficos o temáticos que posibiliten perfilar y focalizar, por ejemplo, riesgos por Unidades de Servicio de primera infancia, riesgos por Instituciones Educativas, entre otros. Su uso como herramienta para la focalización de acciones de las entidades integrantes de la alianza, permite actuar de forma preventiva y anticipada a la ocurrencia de violencias contra niñas, niños y adolescentes en zonas específicas de cada municipio, y a su vez, orientar acciones e iniciativas coordinadas entre las entidades del ámbito nacional y las entidades territoriales.



13. Las niñas, niños y adolescentes que se encuentran dentro del grupo con el 20 % de mayor riesgo estimado de vulneración por violencia física, psicológica o negligencia (maltrato) tienen probabilidades estimadas de sufrir violencia por maltrato que varían entre 0,27 % y 100 %.

Mapa 5. Proporción de casos de alta probabilidad estimada de vulneración por violencia física, psicológica o negligencia (maltrato) por municipio



Fuente: Elaboración propia

2. Modelo de probabilidad de vulneración por violencia sexual

2.1 Desempeño general

El desempeño del modelo de probabilidad de vulneración por violencia sexual se evalúa teniendo en cuenta su capacidad para identificar correctamente las niñas, niños y adolescentes para los que se registrará vulneración de sus derechos y aquellos para los que no. La medida AUC (Area Under the Curve) del modelo en general es de 92,08 %, lo que indica que, en promedio, en los diferentes umbrales de clasificación el modelo es capaz de clasificar correctamente este porcentaje de casos positivos y casos negativos (para ver más detalles sobre el cálculo de la AUC ver el título 1.1 de esta sección). El valor de la AUC se mantiene por encima del 90 % para los grupos de Sisbén IV A y B, y ligeramente por debajo para los grupos C y D, evidenciando para todos los grupos un desempeño sobresaliente o excelente.

Otra medida general para evaluar el desempeño del modelo de probabilidad de vulneración de maltrato es la precisión promedio para diferentes umbrales de clasificación. Esta corresponde a la proporción de casos verdaderos positivos sobre el total de casos verdaderos positivos y falsos positivos. La precisión promedio del modelo es de 23,08%, indicando que para cada umbral de clasificación, aproximadamente, 1 de cada cuatro niñas, niños y adolescentes identificados como en riesgo de vulneración por violencia sexual serán efectivamente registrados como violentados. La precisión promedio del modelo es menor a medida que se reduce la vulnerabilidad, siendo de 22,35% para el grupo D de Sisbén IV.

El desempeño del modelo de probabilidad de vulneración por violencia sexual posteriormente se evaluó en el umbral de clasificación, en el cual se maximiza el valor del Balanced Accuracy para la totalidad de los registros y los registros pertenecientes a cada grupo del Sisbén IV (ver Tabla 7 línea 3).

La métrica Balanced Accuracy promedia el valor de los valores de Recall (tasa de verdaderos positivos) y de la Especificidad (Specificity) que equivale a la tasa de verdaderos negativos. Esta medida es adecuada para valorar el desempeño de modelos de analítica predictiva con datos desbalanceados como la vulneración por causa de violencia sexual que tienen una prevalencia de 0,337 % entre las niñas, niños y adolescentes considerados. La menor tasa de prevalencia verdadera u observada se encuentra en el grupo A de Sisbén IV, pero la menor tasa de prevalencia estimada por el modelo se encuentra en el grupo D (15,49 %).

El valor del Balanced Accuracy observado es superior al 80 % para todos los grupos considerados. El Recall (tasa de verdaderos positivos) es de 87,03 % para la totalidad de las observaciones, indicando que aproximadamente 87 de cada 100 niños que serán registrados como víctimas de violencia sexual son identificados por el modelo. El Recall se reduce a medida que también lo hace la vulnerabilidad reflejada en los grupos de Sisbén IV, siendo de 76,77 % para el grupo D. En cuanto a la Especificidad que indica la capacidad del modelo para identificar los casos en los que una niña, niño o adolescente no será registrado como objeto de vulneración por causa de violencia sexual, se observa que este es de 80,5 % para la totalidad

de las observaciones, lo que indica que el modelo da falsas alarmas en aproximadamente el 19,5 % de los casos. Para el grupo D del Sisbén IV el valor de la Especificidad es de 84,74 % indicando que este grupo tiene también una menor tasa de falsos positivos. Si bien es cierto que estos resultados implican que el modelo predice menos vulneración por causa de violencia sexual en el grupo D de Sisbén IV, las diferencias son relativamente bajas por lo que no implican la existencia de discriminación importante en el algoritmo contra los grupos más vulnerables.

Tabla 7. Desempeño Modelo de probabilidad de vulneración por violencia sexual – Total y por grupo de Sisbén IV

Medidas/Grupos	Total	A	B	C	D
1. AUC	92.080 %	92.695 %	91.491 %	89.873 %	89.885 %
2. Precisión promedio	23.017 %	24.367 %	22.485 %	21.195 %	22.347 %
3. Prevalencia de vulneración	0.337 %	0.352 %	0.376 %	0.429 %	0.367 %
4. Mejor umbral de clasificación	0.346 %	0.423 %	0.427 %	0.514 %	0.551 %
4.1. Prevalencia Vulneración Estimada en el umbral BACC	19.732 %	16.782 %	18.701 %	19.023 %	15.489 %
4.2 Recall en el umbral de BACC VP/(VP+FN)	87.031 %	85.737 %	84.639 %	81.307 %	76.766 %
4.4. Especificidad en el umbral de BACC	80.496 %	83.462 %	81.548 %	81.245 %	84.737 %
4.4. Balanced Accuracy en el umbral de BACC	83.763 %	84.600 %	83.093 %	81.276 %	80.752 %
4.5. Precisión en el umbral de aBACC	1.488 %	1.800 %	1.701 %	1.832 %	1.821 %

VP=Verdadero positivo, FN= Falso negativo, FP=Falso positivo, VN= Verdadero negativo

Fuente: Elaboración propia

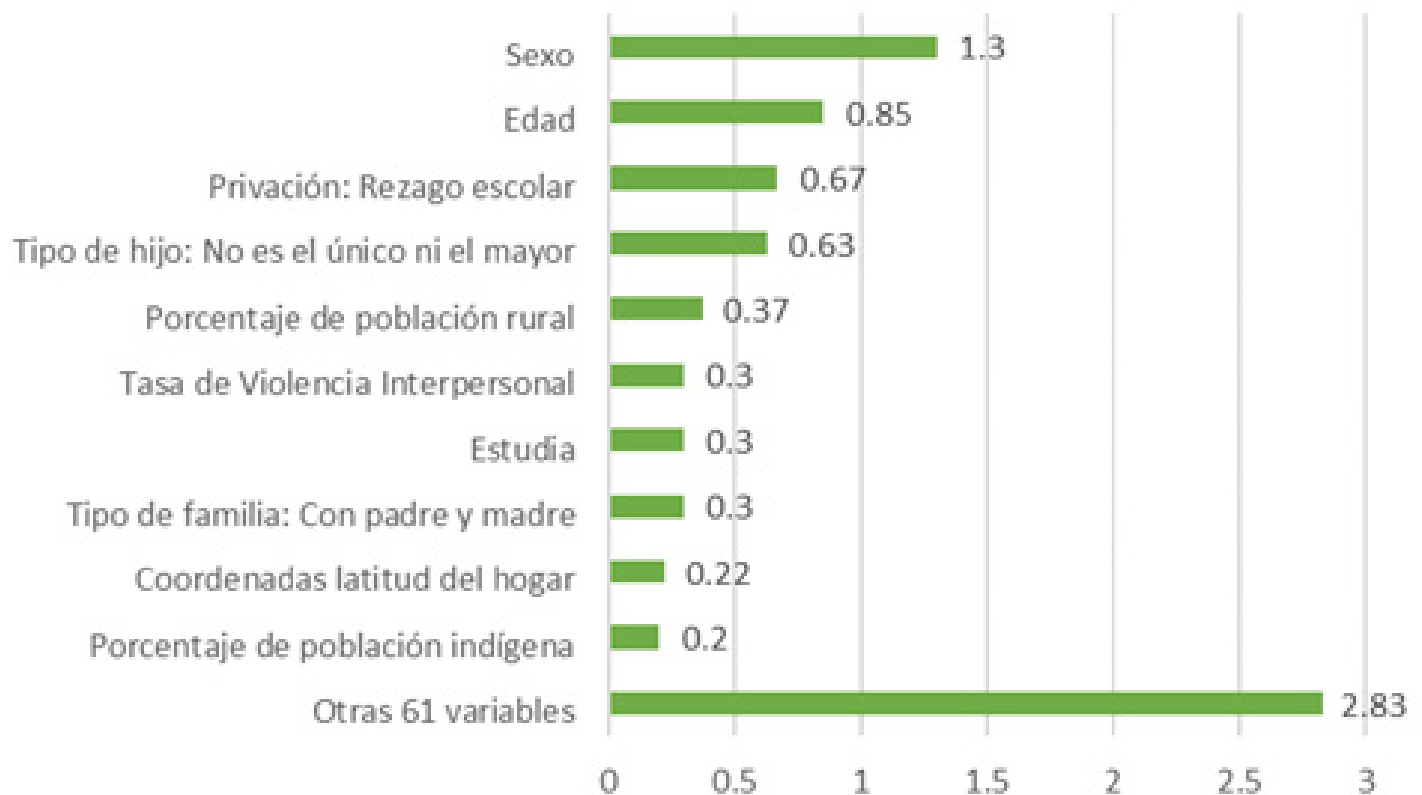
La medida de Precisión en el mejor umbral de clasificación es relativamente baja (1,49 %), lo que se explica por el hecho de que los datos no están balanceados y la diferencia entre la prevalencia observada (0,337 %) y la prevalencia estimada por el modelo (19,73 %), que implica la existencia de una gran proporción de falsos positivos. Este valor indica que el 1,49 % de las niñas, niños y adolescentes que se identifican como en riesgo de vulneración efectivamente entrarán a un proceso de restablecimiento de derechos (PARD), lo que puede reflejar el nivel de subregistro existente.

2.2 Valores SHAP

Los valores SHAP (Shapley Additive Explanations) son una herramienta que permite identificar la contribución de cada variable en la probabilidad de vulneración estimada por el modelo¹⁴. En la Gráfica 6, se muestran las 10 variables que tienen más influencia en la predicción de la probabilidad de vulneración por violencia sexual. Los valores no indican una relación directa entre las variables y la probabilidad estimada, sino que señalan la medida en que cada variable afecta los resultados del modelo en comparación con un escenario en el que no se incluyera.

Así, las variables más importantes en la predicción de la probabilidad de violencia sexual son el sexo, la edad y la privación por rezago escolar al interior del hogar. Es importante resaltar que aparte de las 10 variables que más influyen en la predicción del modelo, se tienen otras 61 variables que presentan una contribución de 2.83 en la probabilidad de vulneración estimada.

Gráfica 6. Valores SHAP Modelo de probabilidad de vulneración por violencia sexual



Fuente: Elaboración propia

14. Para mayor información sobre los valores SHAP ver título III. 1.2.

Las magnitudes de los valores SHAP de una misma variable se presentan como logaritmos de la razón de probabilidades (log odds ratio en inglés), que permiten medir la asociación entre dos variables. Los valores negativos de los log odds ratio representan reducciones en la probabilidad estimada, mientras que los valores positivos significan un efecto positivo en la probabilidad. En la Gráfica 7 se presentan los valores SHAP asociados a los valores de las tres variables más importantes en la predicción de la probabilidad de violencia sexual.

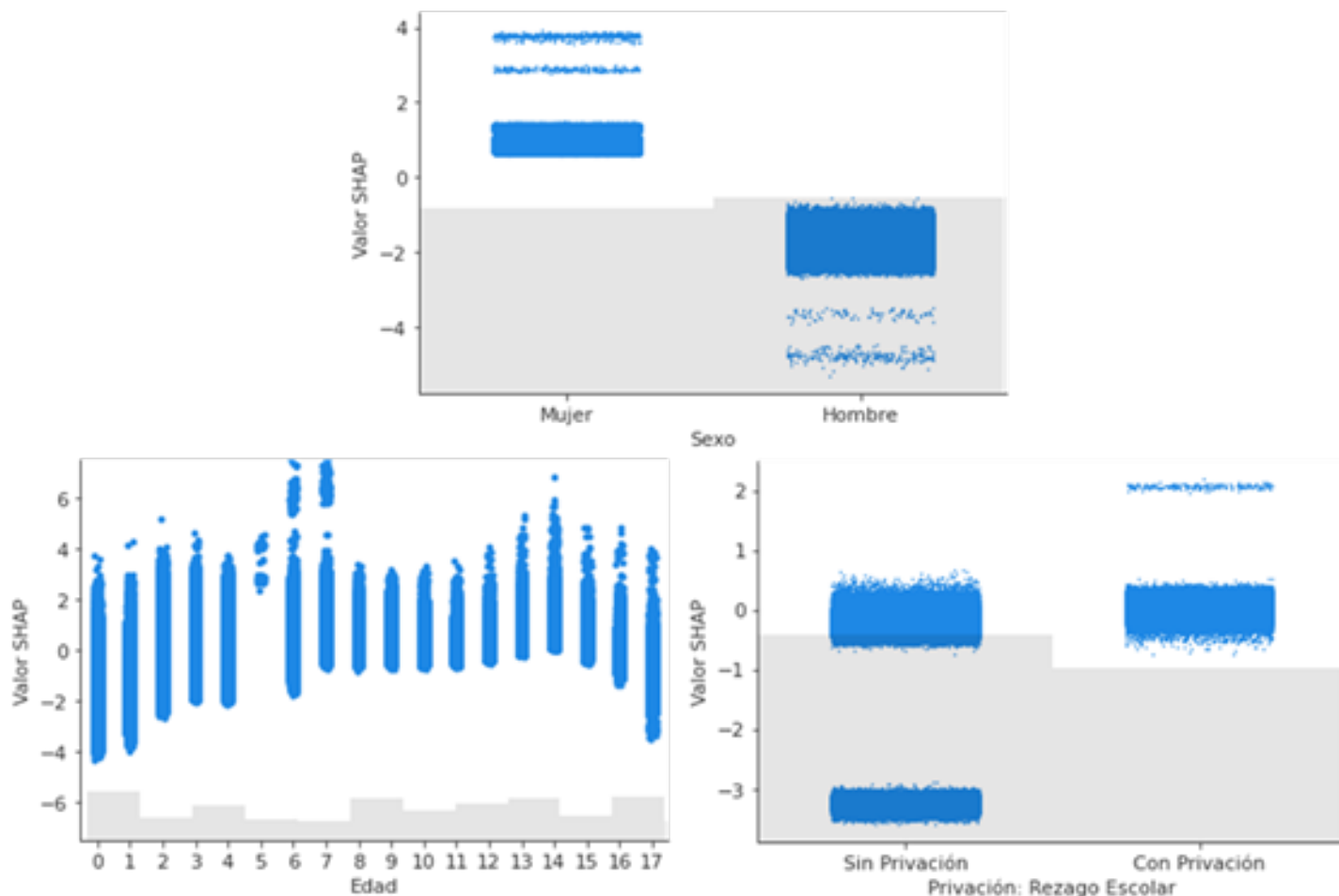
Para todas las niñas y adolescentes mujeres, de forma independiente de otras variables, la probabilidad de vulneración por violencia sexual estimada aumenta de forma considerable (valores SHAP positivos y mayores a 0,5), mientras que para los niños y adolescentes hombres la dispersión en los valores SHAP es mayor y, en general, esta característica disminuye la probabilidad estimada (valores SHAP negativos).

En cuanto a la edad se puede observar que hay una gran dispersión en los valores SHAP para las niñas, niños y adolescentes en los grupos entre 0 y 7 años y entre 16 y 17 años, lo que indica que en estas edades hay otras variables que influyen de forma importante en la probabilidad de vulneración por violencia sexual. Por su parte, los valores SHAP se concentran en valores positivos y con menor dispersión para las edades entre 8 y 14 años. En particular, desde los 10 y hasta su máximo en los 14 años se observa una creciente influencia positiva de la edad en una mayor probabilidad predicha de violencia sexual.

Los valores SHAP para la variable de rezago escolar son similares para la mayoría de las niñas, niños y adolescentes que viven en hogares con y sin esta privación, y tienen valores relativamente bajos (entre -0,5 y 0,5) en comparación con las variables de sexo y edad. Sin embargo, hay un grupo de niñas, niños y adolescentes que viven en hogares sin esta privación cuyo valor SHAP es negativo, lo cual significa que para estos no tener la privación de rezago escolar en su hogar reduce la probabilidad estimada de sufrir violencia sexual. Este grupo corresponde mayoritariamente a niñas y niños menores de 6 años.

Gráfica 7. Valores SHAP de las variables más importantes en la predicción de la probabilidad de violencia sexual





Fuente: Elaboración propia

2.3 Caracterización de los casos con probabilidad alta de ser vulnerados

Con el objetivo de operacionalizar la probabilidad de vulneración por violencia sexual estimada por el modelo, se identificaron las características asociadas a las niñas, niños y adolescentes cuya probabilidad de vulneración estimada hace parte del grupo con probabilidad de vulneración estimada alta (correspondiente al 20 % más alto).

La proporción de niñas y adolescentes mujeres es mucho mayor que la de niños y adolescentes hombres, reflejando el hecho de que la violencia sexual está asociada con el sexo de la víctima. Las niñas y adolescentes representan el 86,1 % del total de los registros con probabilidad alta, y el 35,15 % de las niñas y las adolescentes (frente al 6,01 % de los niños y los adolescentes) tienen alta probabilidad estimada de ser víctimas de violencia sexual.

Dado el rango de edad y sexo, hay una mayor proporción en el grupo de probabilidad de vulneración estimada alta por parte de las niñas entre 10 y 13 años de edad (56,06 %), seguidas por las adolescentes de 14 a 17 años (45,07 %), las niñas entre 6 y 9 años (42,22 %) y los niños entre 6 y 9 años (11,86 %). Los rangos de edad y sexo con menor participación en la probabilidad de vulneración estimada alta son los niños menores de 5 años (2,66 %) y los adolescentes entre 14 y 17 años (5,53 %).

Tabla 8. Rangos de probabilidad de vulneración por violencia sexual

Rango Riesgo			Cantidad		
Probabilidad	Mín.	Max.	Total	Niñas y adolescentes mujeres	Niños y adolescentes hombres
20 % mayor probabilidad estimada	0,3463854 %	100 %	1.438.740 20,98 %	1.238.430 35,15 %	200.310 6.01 %
Total	0 %	100 %	7.219.935	3.526.565	3.693.370

Fuente: Elaboración propia

Las niñas, niños y adolescentes que estudian (26,3 % contra 8,62 % de los que no estudian) y aquellos que saben leer y escribir (27,39 % contra 14,78 % de los que no) tienen una mayor participación en el grupo de probabilidad de vulneración estimada alta. Esto se puede explicar porque las niñas y niños en el rango de edad de 0 a 5 años que son demasiado jóvenes para los grados de educación formal obligatoria y para saber leer y escribir, también tienen la menor participación en la probabilidad de vulneración estimada alta para ambos sexos (11,2 % y 2,66 % respectivamente). La condición de extraedad también implica una mayor proporción de probabilidad de vulneración alta estimada por violencia sexual (29,56% frente a 19,5% para quienes no tienen esta condición).

Acceder a internet es una condición que implica mayor participación en el grupo con probabilidad de vulneración alta estimada (25,2 % frente a 19,15 %). Esta asociación se explica porque en los contextos urbanos es más posible que se denuncien casos de violencia sexual que en los contextos rurales y también se cuenta con mayor acceso a internet.

Estar en una unión temprana por matrimonio, tener hijos o estar en condición de embarazo se asocia con una mayor participación en la probabilidad estimada alta de sufrir violencia sexual (46,24 %), que es más de dos veces mayor que la que experimentan quienes no están en esta condición (19,56 %). Haber participado en una estrategia de protección del ICBF parece ser un factor de protección, dado que quienes tienen esta condición también tienen una menor participación en el grupo con probabilidad de vulneración estimada alta (13,85 % frente a 23,22 %).

En cuanto a la composición de la familia, las niñas, niños y adolescentes que viven con su madre y el cónyuge de ella son los que presentan mayor participación en el grupo con probabilidad estimada alta de vulneración por violencia sexual (32,68 %). La composición de familia que presenta una menor participación son aquellas en las que las niñas, niños y adolescentes viven con su padre y su madre (13,94 %). El número de menores en el hogar no genera variaciones importantes (es de alrededor de 20 % en todos los grupos), excepto para los hogares con más de 10 menores donde el 28,08 % hace parte del grupo con probabilidad de vulneración estimada alta. Las niñas y niños menores de 5 años que asisten a un lugar

comunitario, jardín, centro de desarrollo infantil o colegio (13,55 %), seguidos de aquellos que permanecen con una empleada o niñera en casa (9,65 %) y los que están al cuidado de un pariente mayor de 18 años (9,63 %) son los que tienen mayor participación en el grupo con probabilidad de vulneración por violencia sexual estimada alta.

La edad de la madre no genera variaciones importantes (es de alrededor de 20 % en todos los grupos), excepto para las madres menores de 29 años donde solo el 10,16 % hace parte del grupo con probabilidad de vulneración por violencia sexual estimada alta. Este efecto se puede explicar por la menor participación de los menores de 5 años en el grupo de probabilidad de vulneración alta. Las niñas, niños y adolescentes que no tienen madre tienen la mayor participación (25,56 %). Las niñas, niños y adolescentes cuya madre alcanzó el nivel educativo universitario o posgrado tienen menor participación en el grupo con probabilidad estimada alta (15,83 % y 13,96 %, respectivamente).

Se observa una menor participación en la probabilidad de vulneración alta estimada por violencia sexual en las niñas, niños y adolescentes pertenecientes en el grupo de Sisbén IV No vulnerable (D). Sin embargo, la participación también es menor en los hogares con 7 o más privaciones, y con gasto mensual por persona inferiores a \$147.500.

En la Gráfica 8 se presentan las características de las niñas, niños y adolescentes que se asociaron con una mayor participación en el grupo de probabilidad estimada de vulneración por violencia sexual alta y con una menor participación en el mismo grupo. Para identificar las características asociadas con una mayor participación se tomó como referencia el valor de 25 % y para las que se asociaron con una menor participación el valor de 15 %. Excepto para la variable que identifica el cuidador de las niñas y niños menores de 5 años, para la cual el valor de referencia para mayor participación es 10 % y para menor participación 5 %.

Gráfica 8. Características de las niñas, niños y adolescentes asociadas con mayor y menor proporción de casos con riesgo alto estimado de violencia sexual

Mayor participación



- Es una niña de 10 a 13 años (56,06%), adolescente mujer de 14 a 17 años (45,07%) o niña de 6 a 9 años (42,22%).
- Se encuentra estudiando (26,3%).
- Se encuentra en extraedad (29,56%).
- Tiene acceso a internet (25,2%).
- Sabe leer y escribir (27,39%).
- Vive con su madre y el conyugue de ella (32,68%), sin su padre ni su madre (28,51%), solo con su padre y el conyugue de él (27,74%) o solo con su padre (25,39%).
- No tiene madre (25,56%).
- Se encuentra en una unión temprana (46,24%).
- Vive en un hogar con más de 10 menores (28,08%).
- No sabe a qué servicio de seguridad social está afiliado (25,26%).
- El gasto del hogar por persona es mayor que \$250.000 (28%).
- Es un menor de 5 años que asiste a un lugar comunitario, jardín o centro de desarrollo infantil o colegio (13,55%).

Menor participación



- Es un niño de 0 a 5 años (2,66%), un adolescente hombre de 14 a 17 años (5,52%), un niño entre 10 y 13 años (9,84%), una niña de 0 a 5 años (11,12%) o un niño de 6 a 9 años (11,86%).
- No se encuentra estudiando (8,62%).
- No sabe leer ni escribir (14,78%).
- Vive con su madre y padre (13,94%).
- Su madre es menor de 29 años (10,16%).
- El nivel educativo máximo alcanzado por su madre es posgrado (13,96%).
- Fue atendido por servicios de prevención del ICBF (13,85%).
- Vive en un hogar con más de 10 privaciones (13,42%).
- El gasto del hogar por persona es menor que \$53.000 (14,1%).
- Es un menor de 5 años al cuidado de su madre o padre en el trabajo (1,88%) o en la casa (2,26%), o al cuidado de un pariente menor de 18 años (2,36%).

Fuente: Elaboración propia

Algunas características se asocian con menor participación en el grupo de probabilidad de vulneración alta tanto para violencia física, psicológica y negligencia (maltrato) como para violencia sexual. Entre estas se encuentran ser un adolescente hombre, vivir con su padre y su madre, tener una madre con mayor nivel educativo y ser un menor de 5 años al cuidado de su madre o su padre en casa. Con mayor probabilidad en ambos casos se asocian características como estar en extraedad, vivir sin su padre ni su madre o solo con uno de ellos y su cónyuge o solo con su padre, no tener madre, vivir en un hogar con más de 10 menores, no saber a qué servicio de seguridad se está afiliado o vivir en un hogar con un gasto alto por persona. Mientras que la violencia física, psicológica y negligencia (maltrato) se distribuye de manera relativamente homogénea entre los sexos y grupos de edad, la violencia sexual tiene una menor participación en el grupo de edad de 0 a 5 años y se concentra en las niñas y adolescentes mujeres. Adicionalmente, llama la atención que mientras que en el modelo de violencia física, psicológica y negligencia (maltrato) vivir en un hogar sin privaciones se asocia a una menor proporción de casos con riesgo alto, para la violencia sexual esto sucede en un hogar con más de 10 privaciones. En ambos casos diferentes variables asociadas con vulnerabilidad económica se asocian con menor y con mayor probabilidad de hacer parte del grupo con probabilidad de vulneración alta.

En el Anexo 3, se puede encontrar la proporción de casos con riesgo de vulneración por violencia sexual estimado alto para el total nacional y respecto a variables seleccionadas. El ICBF ha puesto a disposición de la ciudadanía un tablero de visualización de resultados donde se pueden realizar consultas para el total nacional y para cada departamento y municipio¹⁵.

2.4 Distribución geográfica de los casos con mayor probabilidad de vulneración

Los municipios con mayor proporción de casos de alta probabilidad estimada de vulneración por violencia sexual se identificaron tomando como referencia la cantidad de niñas, niños y adolescentes en cada municipio que hacen parte del 20 % de niñas, niños y adolescentes con mayor riesgo estimado de probabilidad de vulneración¹⁶. La cantidad en cada municipio fue dividida por el total de niñas, niños y adolescentes en cada municipio que se encuentran en la base del Sisbén IV.

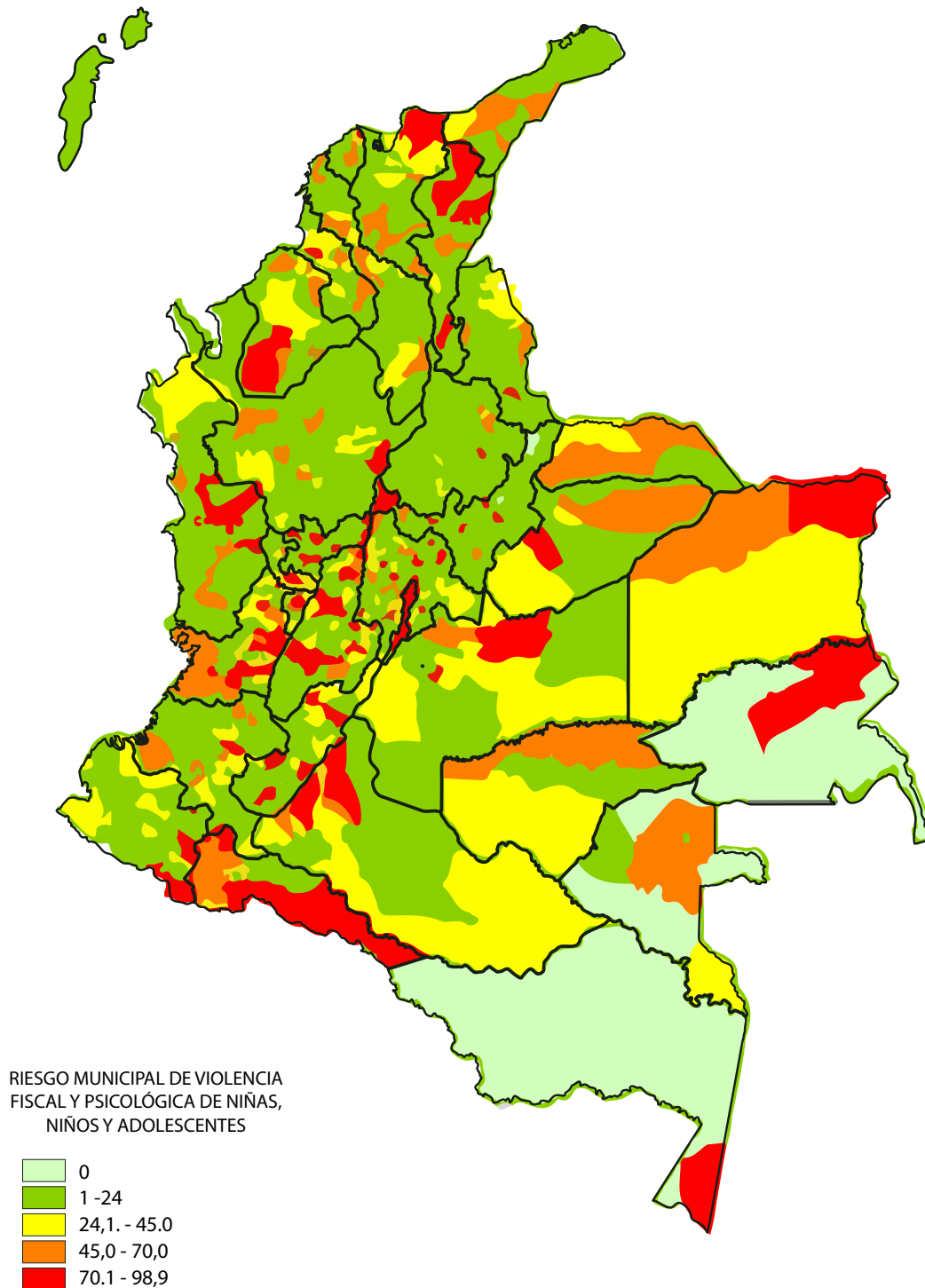
Los municipios con mayor cantidad de niñas, niños y adolescentes con riesgo alto de vulneración por violencia sexual son Manzanares-Caldas (57,25 %), La Dorada-Caldas (55,79 %), Armenia-Quindío (55,5 %) y Neiva-Huila (53,85 %); y en cuanto a departamentos, se encuentra

15. Link para consulta detallada sobre el Modelo probabilístico de vulneración de derechos de niñas, niños y adolescentes: https://public.tableau.com/app/profile/anal.tica.institucional.icbf/viz/ModeloProbabilisticodeVulneracindeDerechos_16183556026450/Portada

16. Las niñas, niños y adolescentes que se encuentran dentro del grupo con el 20 % de mayor riesgo estimado de vulneración por violencia sexual tienen probabilidades estimadas de sufrir violencia sexual que varían entre 0,346 % y 100 %.

que Bogotá tiene la mayor proporción de niñas, niños y adolescentes con probabilidad alta de vulneración sobre el total de esta población en la base del Sisbén IV (47,3 %), seguida por Amazonas (45,1 %), Guainía (40,8 %), Putumayo (30,4 %), Quindío (28,8 %) y Valle del Cauca (28,5 %). Amazonas, Quindío, Putumayo y Guainía se encuentran entre los departamentos con mayor número de ingresos a PARD por violencia sexual como porcentaje de la población de 0 a 17 años en 2019.

Mapa 6. Proporción de casos de alta probabilidadN estimada de vulneración por violencia sexual por municipio



Fuente: Elaboración propia

Actualmente el modelo permite focalizar microzonas de alto riesgo de violencias en todos los municipios del país y se proyecta combinar sus resultados con otros datos geográficos o temáticos que posibiliten perfilar y focalizar, por ejemplo, riesgos por Unidades de Servicio de Primera Infancia, riesgos por Instituciones Educativas, entre otros. Su uso como herramienta para la focalización de acciones de las entidades integrantes de la alianza permite actuar de forma preventiva y anticipada a la ocurrencia de violencias contra niñas, niños y adolescentes en zonas específicas de cada municipio, y a su vez, orientar acciones e iniciativas coordinadas entre las entidades del ámbito nacional y las entidades territoriales.

IV. Conclusiones y recomendaciones

A pesar de que es ampliamente reconocido que la violencia contra las niñas, niños y adolescentes es un fenómeno que causa daños y afectaciones en su integridad y desarrollo, esta se encuentra ampliamente extendida en el mundo donde se estima que alrededor de 1.000 millones de niñas, niños y adolescentes son víctimas de algún tipo de violencia cada año (OMS, 2020) y en Colombia, donde, de acuerdo con los resultados de la EVCNNA, el 41,5 % de las personas de 18 a 24 años reportaron haber sufrido algún tipo de violencia antes de cumplir 18 años, el 32,1 % física, el 15,2 % violencia psicológica y el 11,4% violencia sexual (OIM, CDC y MSPS, 2019).

Existe un consenso desde instancias internacionales (UNICEF, OPS, OMS, UNODC) y nacionales (Alianza Nacional contra las violencias hacia las niñas, niños y adolescentes) frente a la necesidad de adoptar medidas para prevenir y resolver las situaciones de violencia y vulneración de derechos contra las niñas, niños y adolescentes. Sin embargo, la respuesta institucional a la violencia suele ser predominantemente reactiva debido a las limitaciones en la información y herramientas disponibles para tomar decisiones respecto a la distribución de los limitados recursos para la prevención (Administrative Data Research Facilities Network, 2019).

En este contexto, el ICBF desarrolló los primeros modelos de probabilidad de vulneración fundamentados en el uso de analítica de datos y herramientas de aprendizaje de máquinas (machine learning) enfocados en la estimación de la probabilidad de que una niña, niño o adolescente pueda ser víctima de violencia física, psicológica y negligencia (maltrato) o de violencia sexual en Colombia. Estos permitirán identificar y focalizar los esfuerzos institucionales a los grupos poblacionales y territorios con mayor riesgo de sufrir los tipos de violencia que con mayor frecuencia llevan al ingreso de niñas, niños y adolescentes al Proceso Administrativo de Restablecimiento de Derechos (PARD) del ICBF (en 2019 el 22,5 % de los ingresos fueron motivados por violencia física, psicológica y negligencia (maltrato) y el 35,1 % por violencia sexual).

Los modelos de analítica predictiva compilan grandes cantidades de datos estructurados y no estructurados de diferentes fuentes y mediante herramientas de aprendizaje de máquinas (machine learning) descubren patrones y correlaciones a través de los cuales se pueden predecir tendencias y patrones de comportamiento. Modelos de este tipo han sido empleados

en países como Estados Unidos, Nueva Zelanda y Reino Unido para alinear mejor los recursos disponibles con los requerimientos de las comunidades y así mantener a las niñas, niños, adolescentes y familias más seguros.

El ICBF desarrolló modelos de analítica predictiva separados para la estimación de probabilidad de violencia física, psicológica o negligencia (maltrato) y violencia sexual para las niñas, niños y adolescentes registrados en la base del Sisbén IV con corte a junio de 2021. La variable que identifica el registro de una vulneración de derechos es el ingreso a un PARD del ICBF. Estas bases fueron complementadas con información proveniente de diferentes fuentes como el SIMAT del MEN; el RUV de la UARIV; y Estadísticas Vitales y Proyecciones de Población del DANE. La base consolidada cuenta con 7.236.230 registros de niñas, niños y adolescentes en Sisbén IV con fecha de encuesta igual o posterior a 2017, y para los cuales la diferencia entre la fecha de registro en el Sisbén, la fecha de ingreso al PARD y la fecha de las demás variables utilizadas es máximo de tres años.

En la fase de modelación se emplearon diferentes estrategias para contar con modelos confiables y con un alto grado de precisión. El desempeño de los modelos entrenados fue evaluado teniendo en cuenta los criterios de predicción y simplicidad, y considerando que no generara sesgos hacia grupos vulnerables.

El modelo de probabilidad de vulneración por violencia física, psicológica o negligencia (maltrato) tiene una AUC (Area Under the Curve) de 92,38 % que indica un desempeño sobresaliente para distinguir entre casos positivos y negativos en diferentes umbrales de clasificación. El valor de la AUC mantiene un desempeño sobresaliente (por encima del 90 %) y sin variaciones importantes cuando se calcula para cada grupo del Sisbén IV, indicando que no existen sesgos en el algoritmo asociados con la vulnerabilidad de los individuos. El modelo también exhibe valores buenos para las métricas de Balanced Accuracy (84,2 %), Recall (85,18 %) y Especificidad (83,25 %) en el mejor umbral de clasificación, lo que indica que aproximadamente 17 de cada 20 niñas, niños o adolescentes que serán maltratados por violencia física, psicológica o negligencia (maltrato) serán identificados por el modelo. Sin embargo, la medida de Precisión del modelo en el mejor umbral de clasificación es baja (1,22 %), lo que se explica por el hecho de que los datos no están balanceados (solo el 1 % de las niñas, niños y adolescentes en la base del Sisbén IV registran un ingreso a PARD).

Las variables más importantes en la predicción de la probabilidad de violencia física, psicológica y negligencia (maltrato), es decir las que en su ausencia alterarían más la probabilidad estimada son la edad (1,6), la privación por rezago escolar al interior del hogar (para el mismo u otro miembro del hogar) (0,94) y el estar o no estudiando (0,56).

Las niñas, niños y adolescentes que se encuentran dentro del grupo con el 20 % de mayor riesgo estimado de vulneración por violencia física, psicológica o negligencia (maltrato) tienen probabilidades estimadas de sufrir violencia por maltrato que varían entre 0,27 % y 100 %. Estos se distribuyen de forma aproximadamente homogénea entre niñas y adolescentes mujeres y niños y adolescentes hombre, y entre los diferentes grupos de edad. Las características para las cuales se observa una mayor proporción de casos con probabilidad estimada alta

de sufrir una vulneración por causa de violencia física, psicológica o negligencia (maltrato) son: estar en extraedad escolar (34,11 %); vivir sin su padre ni su madre (35,39%), solo con su padre (34,05 %), con su padre y el cónyuge de él (33,16%) o con su madre y el cónyuge de ella (30,29 %); no tener madre (33,2 %), vivir en un hogar con más de 10 menores (34,38 %) y ser un menor de 5 años que permanece al cuidado de otro (45,63%), solo en casa (44,6 %), al cuidado de un pariente menor de 18 años (35,97 %) o con su madre o padre en el trabajo (30,85 %).

Los municipios con mayor cantidad de niñas, niños y adolescentes con riesgo alto de vulneración por violencia física o psicológica son Túquerres-Nariño (97,46 %), La Unión-Nariño (96,82 %) y Salamina-Caldas (92,81 %); y en cuanto departamentos, se encuentra que Bogotá tiene la mayor proporción de niñas, niños y adolescentes con probabilidad alta de vulneración predicha sobre el total de esta población en la base del Sisbén IV (39,4 %), seguida por Nariño (37,83 %), Valle del Cauca (36,39 %) y Meta (35,99 %).

El modelo de probabilidad de vulneración por violencia sexual tiene una AUC de 92,08 % que indica un desempeño sobresaliente para distinguir entre casos positivos y negativos en diferentes umbrales de clasificación. El valor de la AUC mantiene un desempeño sobresaliente (por encima del 90 %) para los grupos A y B del Sisbén IV, y ligeramente por debajo para los grupos C y D, evidenciando para todos los grupos un desempeño sobresaliente. El modelo también exhibe valores buenos para las métricas de Balanced Accuracy (83,8 %), Recall (87,03 %) y Especificidad (80,5 %) en el mejor umbral de clasificación, lo que indica que aproximadamente 87 de cada 100 niñas, niños o adolescentes que serán registrados como víctimas de violencia sexual son identificados en la predicción del modelo. Sin embargo, la medida de Precisión del modelo en el mejor umbral de clasificación es baja (1,49 %) lo que se explica por el hecho de que los datos no están balanceados (solo el 1 % de las niñas, niños y adolescentes en la base del Sisbén IV registran un ingreso a PARD).

Las variables más importantes en la predicción de la probabilidad de violencia sexual son el sexo (1,3), la edad (0,85), y la privación por rezago escolar al interior del hogar (0,67).

Las niñas, niños y adolescentes que se encuentran dentro del grupo con el 20 % de mayor riesgo estimado de vulneración por violencia sexual tienen probabilidades estimadas de sufrir violencia sexual que varían entre 0,346 % y 100 %. El riesgo de vulneración alta se concentra en las niñas y adolescentes mujeres. Los grupos por edad y sexo con mayor participación en el grupo con probabilidad de vulneración estimada alta son niñas de 10 a 13 años (56,06 %), adolescentes mujer de 14 a 17 años (45,07 %) y niñas de 6 a 9 años (42,22 %). Una mayor participación en el grupo de riesgo alto también se asocia con vivir con su madre y el cónyuge de ella (32,68 %) y estar en una unión temprana (46,24 %).

Algunas características se asocian con menor participación en el grupo de probabilidad de vulneración alta tanto para violencia física, psicológica y negligencia (maltrato) como para violencia sexual, entre estas se encuentran ser un adolescente hombre, vivir con su padre y su madre, tener una madre con mayor nivel educativo y ser un menor de 5 años al cuidado de su madre o su padre en casa. Con mayor probabilidad en ambos casos se asocian características como estar en extraedad, vivir sin su padre ni su madre o solo con uno de ellos

y su cónyuge o solo con su padre, no tener madre, vivir en un hogar con más de 10 menores, no saber a qué servicio de seguridad se está afiliado, o vivir en un hogar con un gasto alto por persona. Mientras que la violencia física, psicológica y negligencia (maltrato) se distribuye de manera relativamente homogénea entre los sexos y grupos de edad, la violencia sexual tiene una menor participación en el grupo de edad de 0 a 5 años y se concentra en las niñas y adolescentes mujeres. Adicionalmente, llama la atención que, mientras que en el modelo de violencia física, psicológica y negligencia (maltrato) vivir en un hogar sin privaciones se asocia a una menor proporción de casos con riesgo alto, para la violencia sexual esto sucede en un hogar con más de 10 privaciones. En ambos casos diferentes variables asociadas con vulnerabilidad económica se asocian con menor y con mayor probabilidad de hacer parte del grupo con probabilidad de vulneración alta.

Los municipios con mayor cantidad de niñas, niños y adolescentes con riesgo alto de vulneración por violencia sexual son Manzanares-Caldas (57,25 %), La Dorada-Caldas (55,79 %), Armenia-Quindío (55,5%) y Neiva-Huila (53,85%); y en cuanto a departamentos, se encuentra que Bogotá tiene la mayor proporción de niñas, niños y adolescentes con probabilidad alta de vulneración sobre el total de esta población en la base del Sisbén IV (47,3 %), seguida por Amazonas (45,1%), Guainía (40,8 %), Putumayo (30,4 %), Quindío (28,8 %) y Valle del Cauca (28,5 %).

Actualmente el modelo permite focalizar microzonas de alto riesgo de violencias en todos los municipios del país y se proyecta combinar sus resultados con otros datos geográficos o temáticos que posibiliten perfilar y focalizar, por ejemplo, riesgos por Unidades de Servicio de primera infancia, riesgos por Instituciones Educativas, entre otros. Su uso como herramienta para la focalización de acciones de las entidades integrantes de la alianza, permite actuar de forma preventiva y anticipada a la ocurrencia de violencias contra niñas, niños y adolescentes en zonas específicas de cada municipio, y a su vez, orientar acciones e iniciativas coordinadas entre las entidades del ámbito nacional y las entidades territoriales.

A pesar de que los modelos de analítica predictiva son mejores prediciendo tendencias y patrones de comportamiento que los modelos multivariados estándar, la interpretabilidad y comunicación de sus resultados es compleja. Por lo que en este boletín se recurrió a diferentes enfoques para presentar los resultados de cada uno de los modelos desarrollados de modo que se pudiera evidenciar el buen desempeño de los modelos a partir de métricas adecuadas y algunos de los resultados sobresalientes de una manera clara. No obstante, se debe recordar que los resultados presentados en este boletín son una pequeña selección de los resultados de los modelos, por lo que el uso de los tableros de visualización de resultados y otras herramientas desarrolladas por el ICBF es fundamental para que pueda emplearse en el diseño de intervenciones focalizadas centradas en la prevención de la ocurrencia de violencias desde un enfoque intersectorial y la posible identificación en terreno de casos de vulneración de derechos.

Los modelos de analítica predictiva desarrollados por el ICBF y documentados en este boletín requieren constante revisión y posibles ajustes para mantener o mejorar su desempeño ante cambios en los datos y procesos que puedan ocurrir. Algunos aspectos que podrían explorarse para versiones futuras de los modelos son buscar un mecanismo para integrar niñas, niños

y adolescentes no registrados en el Sisbén IV, y complementar la variable de ingreso a un PARD del ICBF con otra información para identificar a las niñas, niños y adolescentes que han tenido una vulneración de derechos por causa de violencia. Esto último porque la variable de ingreso a PARD solo identifica aquellos casos para los cuales se ha iniciado un proceso de



Glosario

5 iteraciones o 5-fold cross-validation es un caso especial del procedimiento de k-fold cross-validation que es un procedimiento de remuestreo usado para estimar la bondad de modelos de machine learning. En este procedimiento la muestra de datos se divide en k grupos con el objetivo de estimar cómo se comportará el modelo en general cuando es usado para realizar predicciones en datos no usados durante el entrenamiento del modelo.

Aprendizaje de máquinas es una rama de la inteligencia artificial y las ciencias de la computación que se enfoca en el uso de datos y algoritmos para imitar la forma en que los humanos aprenden, aumentando de forma gradual su precisión. El aprendizaje de máquinas es un componente importante del creciente campo de la ciencia de datos. Mediante el uso de métodos estadísticos, los algoritmos se entrenan para hacer clasificaciones o predicciones, descubriendo información clave dentro de los proyectos de minería de datos. Esta información, posteriormente, se usa para la toma de decisiones que impactan el desempeño de las organizaciones. IBM (15 de julio de 2020) Machine Learning. (original en inglés).

Árbol de decisión es un mapa de los posibles resultados de una serie de decisiones relacionadas.

Árbol de potenciación del gradiente o Gradient Boosted Trees es un conjunto de árboles de decisión donde cada uno se crea de manera secuencial por lo que los resultados de los diferentes árboles de decisión se van combinando durante el proceso.

AUC o Area Under the Curve es una medida resumen de la habilidad del modelo para distinguir entre casos positivos y negativos en diferentes umbrales de clasificación. El valor de la AUC corresponde al área por debajo de la ROC. Un mayor valor de AUC indica un mejor desempeño del modelo, siendo 1 una clasificación perfectamente precisa (sin falsos positivos ni falsos negativos).

Balanced Accuracy o Acierto promedio de predicción para los vulnerados y no vulnerados es una métrica de calidad del modelo de machine learning que promedia el valor de los valores de Recall (tasa de verdaderos positivos) y de la Especificidad (Specificity) que equivale a la tasa de verdaderos negativos.

Bosque aleatorio o Random Forest es una combinación de árboles predictores tal que cada árbol depende de los valores de un vector aleatorio probado independientemente y con la misma distribución para cada uno de estos.

Búsqueda de mejores hiperparámetros o Grid search es un proceso empleado para buscar los valores óptimos para los hiperparámetros del modelo que resulten en predicciones más acertadas.

Creación de nuevas variables a partir de los datos originales o Feature engineering es un proceso por el cual se crean nuevas variables a partir de los datos disponibles con el fin de

obtener información importante según el conocimiento de la problemática.

Especificidad o Specificity es una métrica de calidad del modelo de machine learning que informa sobre la proporción de verdaderos negativos sobre el total de casos negativos (verdaderos negativos y falsos positivos).

Exhaustividad o tasa de verdaderos positivos o Recall ($VP/(VP+ FN)$) es una métrica de calidad del modelo de machine learning que informa sobre la proporción de casos verdaderos positivos sobre el total de casos positivos (verdaderos positivos y falsos negativos).

Falso negativo significa caso positivo identificado como negativo por el modelo.

Falso positivo significa caso negativo identificado como positivo por el modelo.

Hiperparámetro es una característica del modelo que es externa al mismo y cuyo valor no puede ser estimado a partir de los datos. Su valor es definido antes de empezar el procesamiento.

Importancia de cada variable explicativa (Feature importance) se refiere a las técnicas que asignan un puntaje a las características de entrada de acuerdo con su utilidad para predecir la variable resultado.

Matriz de confusión es una herramienta que permite visualizar el desempeño de un algoritmo de aprendizaje supervisado en cuanto a qué tipos de aciertos y errores está teniendo el modelo a la hora de pasar por el proceso de aprendizaje con los datos. En las columnas se muestra el número de predicciones de cada clase y en las filas las instancias en la clase real.

Mejor umbral de clasificación Balanced Accuracy (BACC) corresponde al umbral de clasificación en el que se obtiene el mayor valor posible de Balanced Accuracy.

Parámetro es una característica interna del modelo y su valor puede ser estimado a partir de los datos.

Precisión o precisionp ($VP/VP+FP$) es una métrica de calidad del modelo de machine learning que informa sobre la cantidad de casos verdaderos positivos que el modelo logra identificar sobre el total de casos que identifica como positivos (verdaderos positivos y falsos positivos).

Prevalencia de vulneración estimada indica la proporción de casos en los que de acuerdo con la predicción del modelo se identifica la existencia de una vulneración.

Prevalencia de vulneración verdadera u observada indica la proporción de casos en los que es posible identificar la existencia de una vulneración en los datos disponibles.

Prueba ANOVA o análisis de varianza es un método estadístico que permite descubrir si los resultados de una prueba son significativos

Pruebas chi-cuadrado de Pearson es una técnica estadística de libre distribución (no paramétrica) que mide la discrepancia entre una distribución de frecuencias observadas y esperadas.

ROC o curva del operador del receptor relaciona la tasa de verdaderos positivos del modelo (Recall) con la tasa de falsos positivos (TFP) en diferentes umbrales de clasificación.

Umbrales de clasificación en los modelos de clasificación, los umbrales de clasificación son puntos de corte que permiten asignar una observación a una categoría definida dada la puntuación de probabilidad obtenida.

Valores SHAP son una herramienta de visualización que permite identificar de manera intuitiva la contribución de cada variable/característica para obtener una determinada predicción y para el modelo en su conjunto. En la probabilidad de vulneración estimada por el modelo. No permite establecer causalidad, sino identificar la importancia y dirección del impacto relativo de cada característica considerada.

Variable de destino/dependiente/resultado/objetivo representa el producto o resultado cuya variación se está estudiando.

Variable de entrada/independiente/explicativa representa insumos o causas, es decir, razones potenciales de variación.

Verdadero negativo significa caso negativo identificado como negativo por el modelo.

Verdadero positivo significa caso positivo identificado como positivo por el modelo.

Referencias

- Administrative Data Research Facilities Network (2019) New Open Source Machine Learning Tools Help Prevent Child Abuse. Recuperado de <https://www.adrf.upenn.edu/single-post/2019/03/29/new-open-source-machine-learning-tools-help-prevent-child-abuse>
- Chouldechova, Alexandra, Emily Putnam-Hornstein, Diana Benavides-Prado, Oleksandr Fialko y Rhema Vaithianathan (2018) A case study of algorithm-assisted decision making in child maltreatment hotline screening decisions, Conference on Fairness, Accountability, and Transparency, Proceedings of Machine Learning Research 81:1-15, 2018.
- Datos abiertos - DIJIN Policía Nacional (11 de mayo de 2021) Reporte Delito Violencia Intrafamiliar Policía Nacional. Recuperado de <https://www.datos.gov.co/Seguridad-y-Defensa/Reporte-Delito-Violencia-Intrafamiliar-Polic-a-Nac/vuyt-mqpw>
- Datos abiertos - DIJIN Policía Nacional (11 de mayo de 2021) Reporte Delitos sexuales Policía Nacional. Recuperado de <https://www.datos.gov.co/Seguridad-y-Defensa/Reporte-Delitos-sexuales-Polic-a-Nacional/fpe5-ymw>
- Datos abiertos- Fiscalía General de la Nación (11 de mayo de 2021) Conteo de Víctimas según las entradas de noticias criminales por delito al Sistema Penal Oral Acusatorio en la Ley 906 de 2004 y Ley 1098 de 2006. Recuperado de <https://www.datos.gov.co/Justicia-y-Derecho/Conteo-de-V-ctimas/sft7-9im5>
- DNP (2019) Plan Nacional de Desarrollo 2018-2022 “Pacto por Colombia, pacto por la equidad” Recuperado de <https://www.dnp.gov.co/DNPN/Paginas/Plan-Nacional-de-Desarrollo.aspx>
- Gillingham, Philip (2016) Predictive Risk Modelling to Prevent Child Maltreatment and Other Adverse Outcomes for Service Users: Inside the ‘Black Box’ of Machine Learning British Journal of Social Work 2016 Jun; 46(4): 1044–1058. Recuperado de <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4986074/>
- Glen, Stephanie (Julio 2019) Decision Tree vs Random Forest vs Gradient Boosting Machines: Explained Simply. Recuperado de <https://www.datasciencecentral.com/profiles/blogs/decision-tree-vs-random-forest-vs-boosted-trees-explained>
- IBM (15 de julio de 2020) Machine Learning. Recuperado de <https://www.ibm.com/cloud/learn/machine-learning>
- ICBF (2018) Lineamiento Técnico para la Atención a Niños, Niñas y Adolescentes con Derechos Amenazados o Vulnerados, Víctimas de Violencia Sexual. Recuperado de https://www.icbf.gov.co/sites/default/files/procesos/lm18.p_lineamiento_tecnico_programa_especializado_de_atencion_victimas_de_violencia_sexual_v2.pdf

ICBF (2019) ABC ALIANZA NACIONAL CONTRA LA VIOLENCIA HACIA NIÑAS, NIÑOS Y ADOLESCENTES. Recuperado de <https://www.icbf.gov.co/abc-alianza-nacional-contr-la-violencia-hacia-ninas-ninos-y-adolescentes#:~:text=Ni%C3%B1os%20y%20Adolescentes%3F-,La%20Alianza%20Nacional%20contra%20la%20Violencia%20hacia%20Ni%C3%B1as%2C%20Ni%C3%B1os%20y,hacia%20la%20ni%C3%B1ez%20y%20adolescencia.&text=Prevenir%20todas%20las%20formas%20de%20violencia%20contras%20los%20Ni%C3%B1os%2C%20ni%C3%B1as%20y%20adolescentes>

ICBF. (2017). Lineamiento técnico para la atención de niños, niñas y adolescentes con sus derechos inobservados, amenazados o vulnerados por causa de la violencia. Recuperado de https://www.icbf.gov.co/sites/default/files/procesos/Im21.p_lineamiento_tecnico_atencion_nna_con_derechos_inobservados_amenazados_y_vulnerados_por_causa_de_la_violencia_v1.pdf

ICBF. (2020). Ingresos de niñas, niños y adolescentes al Proceso Administrativo de Restablecimiento de Derechos (PARD) en Colombia con fecha de corte a 31 de diciembre de 2020. Sistema de Información Misional (SIM).

INMLCF. Cálculo violencias INML_2011-2019

INMLCF. Información preliminar de lesiones no fatales de causa externa en Colombia. Enero a diciembre de 2020.

OIM, CDC, Ministerio de Salud y Protección Social (2019) Encuesta de Violencia contra Niños, Niñas y Adolescentes EVCNNA – 2018. Recuperado de https://reliefweb.int/sites/reliefweb.int/files/resources/2020-3-17_Colombia-VACS-Final-Report-Spanish.pdf

OMS (2020) Global status report on preventing violence against children 2020. Recuperado de <https://www.who.int/teams/social-determinants-of-health/violence-prevention/global-status-report-on-violence-against-children-2020>

OMS. (2019). Violencia contra los niños. Recuperado de <https://www.who.int/es/news-room/fact-sheets/detail/violence-against-children>

Pryce, Jessica, Anna Telick, Ying Zhang and Kreig (2018) Fields White Paper Using Artificial Intelligence, Machine Learning, and Predictive Analytics in Decision Making FLORIDA INSTITUTE FOR CHILD WELFARE. Recuperado de <https://ficw.fsu.edu/sites/g/files/upcbnu1106/files/Final%20Reports/FICW%20Using%20Artificial%20Intelligence,%20Machine%20Learning,%20and%20Predictive%20Analytics%20in%20Decision-Making.pdf>

The Guardian (19 de noviembre de 2019) Is it right to use AI to identify children at risk of harm? Recuperado de <https://www.theguardian.com/society/2019/nov/18/child-protection-ai-predict-prevent-risks>

Anexo 1. Detalle variables modelo

Variable	Valores	Base de datos	Fuente
Año en que fue realizada la encuesta	2017	Sisbén IV	DNP
	2018		
	2019		
	2020		
	2021		
Fecha de nacimiento			
Edad calculada del niño al momento de la encuesta	Número	Sisbén IV	DNP
Tipo de hijo-lugar	1 – Mayor	Sisbén IV	DNP
	2 - Único		
	3 – Otro		
Discapacidad	0 – No	Sisbén IV	DNP
	1 – Sí		
Discapacidad no puede bañarse	0 – No	Sisbén IV	DNP
	1 – Sí		
Unión temprana (estado civil, tiene hijos, estado de embarazo)	0 – No	Sisbén IV	DNP
	1 – Sí		
Identificador 5		RUV	UARIV
Víctima del conflicto armado	0 – No	RUV	UARIV
	1 – Sí		

Variable	Valores	Base de datos	Fuente
Seguridad social	0- Ninguna	Sisbén IV	DNP
	1- Contributivo		
	2- Especial (Fuerzas Armadas, Ecopetrol, universidades públicas, magisterio)		
	3- Subsidiado (EPS-S)		
	9- No sabe		
Acudió a una institución prestadora de servicios de salud	0 – No	Sisbén IV	DNP
	1 – Sí		
Lo atendieron	0 – No	Sisbén IV	DNP
	1 – Sí		
Seguridad social	1- Asiste a un lugar comunitario, jardín o centro de desarrollo infantil o colegio	Sisbén IV	DNP
	2- Con su padre o madre en la casa		
	3- Con su padre o madre en el trabajo		
	4- Con empleada o niñera en la casa		
	5- Al cuidado de un pariente de 18 años o más		
	6- Al cuidado de un pariente menor de 18 años		
	7- En casa solo		
	8 - Al cuidado de otro		
	9- No aplica por flujo		
Atendido en un servicio de prevención en el año de la encuesta o en un año anterior	0 – No	Cuéntame	ICBF
	1 – Sí		
Sabe leer y escribir anterior	0 – No	Sisbén IV	DNP
	1 – Sí		

Variable	Valores	Base de datos	Fuente
Actualmente estudia (asiste al preescolar, escuela, colegio o universidad)	0 – No	Sisbén IV	DNP
	1 – Sí		
Estudia (MEN)	0 – No	SIMAT	MEN
	1 – Sí		
Estado definitivo matrícula			
Extraedad	0 – No	SIMAT	MEN
	1 – Sí		
Jornada estudio	1 – Completa	SIMAT	MEN
	2 – Mañana		
	3 – Tarde		
	4 – Fin de semana		
	5 – Nocturna		
	6 – Única		
Grado	Número	SIMAT	MEN
Recibe o toma desayuno o almuerzo donde permanece la mayor parte del tiempo entre semana	0 – No	Sisbén IV	DNP
	1 – Sí		
	9 – No aplica		
Motivo ingreso PARD	Alfanumérica	Cuéntame	ICBF
	0 - No vulnerado	Cuéntame	ICBF
Motivo ingreso PARD agrupado	1 - Vulnerado por violencia sexual		
	2 - Vulnerado por violencia física, psicológica o negligencia (maltrato)		
	3 - Otro tipo de vulneración		
Año de vulneración	2011	Cuéntame	ICBF
	2012		
	2013		
	2014		
	2015		

Variable	Valores	Base de datos	Fuente
Año de vulneración	2016	Cuéntame	ICBF
	2017		
	2018		
	2019		
	2020		
	2021		
Código caso		Cuéntame	ICBF
Diferencia entre el año de vulneración y el año de la encuesta	Número	Cuéntame	
¿El niño ha tenido un PARD por violencia sexual? anterior	0 - No vulnerado	Cuéntame	ICBF
	1 - Vulnerado		
Año de violencia sexual	2011	Cuéntame	ICBF
	2012		
	2013		
	2014		
	2015		
	2016		
	2017		
	2018		
	2019		
	2020		
	2021		
Edad en el año en que tuvo vulneración por violencia física, psicológica o negligencia (maltrato)		Cuéntame	ICBF
Diferencia entre el año en que tuvo vulneración por violencia física, psicológica o negligencia (maltrato) y el año de la encuesta		Cuéntame	
Identificador hogar			
Identificador 6			
Identificador 7			

Variable	Valores	Base de datos	Fuente
Identificador 8			
Composición del hogar	1 - Sin padre y sin madre	Sisbén IV	DNP
	2 - Solo madre y cónyuge		
	3 - Solo padre y cónyuge		
	4 - Solo madre		
	5 - Solo padre		
	6 - Con madre y padre		
Total de personas del hogar con documento válido	1 a 25	Sisbén IV	DNP
Número de niños menores de 18 años en el hogar		Sisbén IV	DNP
Jefatura femenina	0 – No	Sisbén IV	DNP
	1 – Sí		
Edad del mayor de los padres	Número	Sisbén IV	DNP
Edad del menor de los padres	Número	Sisbén IV	DNP
Edad del padre	Número	Sisbén IV	DNP
Edad del padre	0 - Sin padre	Sisbén IV	DNP
	1 - 1er quintil (0 a 29)		
	2 - 2do quintil (30 a 35)		
	3 - 3er quintil (36 a 40)		
	4 - 4to quintil (41 a 48)		
	5 - 5to quintil (49 y más)		

Variable	Valores	Base de datos	Fuente
Edad de la madre	Número	Sisbén IV	DNP
Edad de la madre	0 - Sin padre	Sisbén IV	DNP
	1 - 1er quintil (0 a 29)		
	2 - 2do quintil (30 a 35)		
	3 - 3er quintil (36 a 40)		
	4 - 4to quintil (41 a 48)		
	5 - 5to quintil (49 y más)		
Total de personas del hogar con documento válido	1 a 25	Sisbén IV	DNP
Diferencia de edad de los padres	0 - Sin padre	Sisbén IV	DNP
	1 - 1er quintil (0 a 29)		
	2 - 2do quintil (30 a 35)		
	3 - 3er quintil (36 a 40)		
	4 - 4to quintil (41 a 48)		
	5 - 5to quintil (49 y más)		
Nivel educativo del padre	0- Ninguno	Sisbén IV	DNP
	1- Preescolar		
	2- Básica primaria (1o. - 5o)		
	3- Básica secundaria (6o. - 9o.)		
	4- Media (10o. 13o.)		
	5- Técnico o tecnológico		
	6- Universitario		
	7- Postgrado		
	9- No aplica por flujo		

Variable	Valores	Base de datos	Fuente
Generador de ingresos del hogar	0- Ninguno	Sisbén IV	DNP
	2 - Solo figura materna		
	3 - Solo figura paterna		
	4 - Ambos		
Ingresos		Sisbén IV	DNP
Ingresos del hogar		Sisbén IV	DNP
Ingresos per cápita del hogar		Sisbén IV	DNP
Ingresos per cápita del hogar, imputado		Sisbén IV	DNP
ind_ingr_salario		Sisbén IV	DNP
Valor ingresos arriendos		Sisbén IV	DNP
vlr_ingr_col_mayor		Sisbén IV	DNP
Valor ingresos cosecha		Sisbén IV	DNP
Gastos per cápita del hogar, imputado		Sisbén IV	DNP
Gastos en alimentos per cápita del hogar, imputado		Sisbén IV	DNP
Valor gastos en arriendo		Sisbén IV	DNP
Valor gastos en celular		Sisbén IV	DNP

Variable	Valores	Base de datos	Fuente
Valor gastos en educación		Sisbén IV	DNP
Valor gastos en salud		Sisbén IV	DNP
Valor gastos en servicios públicos		Sisbén IV	DNP
vlr_ingr_col_mayor		Sisbén IV	DNP
Valor gastos en transporte		Sisbén IV	DNP
Valor gastos en otros		Sisbén IV	DNP
Porcentaje de gastos del hogar que corresponden a gasto en alimentos, luego de imputación de gasto total y gasto en alimentos		Sisbén IV	DNP
Nivel Sisbén	A01-A05	Sisbén IV	DNP
	B01-B07		
	C01-C18		
	D01-D21		
Clasificación Sisbén		Sisbén IV	DNP
Grupo Sisbén	A	Sisbén IV	DNP
	B		
	C		
	D		
Indicador de pobreza multidimensional	0 - Sin privación	Sisbén IV	DNP
	1 - Con privación		
Número de privaciones		Sisbén IV	DNP

Variable	Valores	Base de datos	Fuente
Bajo logro educativo	0 - Sin privación	Sisbén IV	DNP
	1 - Con privación		
Analfabetismo	0 - Sin privación	Sisbén IV	DNP
	1 - Con privación		
Rezago escolar	0 - Sin privación	Sisbén IV	DNP
	1 - Con privación		
Barreras a servicios para cuidado de la primera infancia	0 - Sin privación	Sisbén IV	DNP
	1 - Con privación		
Desempleo de larga duración	0 - Sin privación	Sisbén IV	DNP
	1 - Con privación		
Trabajo informal	0 - Sin privación	Sisbén IV	DNP
	1 - Con privación		
Sin aseguramiento en salud	0 - Sin privación	Sisbén IV	DNP
	1 - Con privación		
Barreras de acceso a servicios de salud	0 - Sin privación	Sisbén IV	DNP
	1 - Con privación		
Sin acceso a fuente de agua mejorada	0 - Sin privación	Sisbén IV	DNP
	1 - Con privación		
Inadecuada eliminación de excretas	0 - Sin privación	Sisbén IV	DNP
	1 - Con privación		
Material inadecuado de pisos	0 - Sin privación	Sisbén IV	DNP
	1 - Con privación		
Material inadecuado de paredes exteriores	0 - Sin privación	Sisbén IV	DNP
	1 - Con privación		
Valor línea de pobreza		Sisbén IV	DNP
Tipo de vivienda	1- Casa	Cuéntame	ICBF
	2- Apartamento		
	3- Cuarto		
	4- Otro tipo de vivienda		
	5- Vivienda indígena		

Variable	Valores	Base de datos	Fuente
La vivienda ocupada por este hogar es	1- En arriendo o subarriendo	Sisbén IV	DNP
	2- Propia, la están pagando		
	3- Propia, totalmente pagada		
	4- Con permiso del propietario		
	5- Posesión Sin título, ocupante de hecho		
Cuantos grupos de personas que manejan su propio presupuesto (hogares) hay en la vivienda	Número	Sisbén IV	DNP
El hogar comparte sanitario o cocina con otros hogares de la vivienda u otras viviendas	0 – No	Sisbén IV	DNP
	1 – Sí		
Número de personas por cuarto usado para dormir, imputado		Sisbén IV	DNP
Número de personas por cuarto usado exclusivamente para dormir, imputado		Sisbén IV	DNP
Vivienda afectada por un evento natural	0 – No	Sisbén IV	DNP
	1 – Sí		
Número total de veces que ha sido afectada la vivienda por algún evento natural (inundación, avalancha, terremoto, incendio, vendaval, hundimiento)	Número	Sisbén IV	DNP
¿El hogar tiene cocina?	0 – No	Sisbén IV	DNP
	1 – Sí		
¿El hogar posee nevera?	0 – No	Sisbén IV	DNP
	1 – Sí		
¿El hogar posee lavadora?	0 – No	Sisbén IV	DNP
	1 – Sí		
¿El hogar posee computador para uso del hogar?	0 – No	Sisbén IV	DNP
	1 – Sí		

Variable	Valores	Base de datos	Fuente
¿El hogar posee conexión a internet?	0 – No	Sisbén IV	DNP
	1 – Sí		
Estrato (recibo de energía o de acueducto)	0. No tiene1	Sisbén IV	DNP
	1. 1		
	2. 2		
	3. 3		
	4. 4		
	5. 5		
	6. 6		
Acceso a energía eléctrica	0 – No	Sisbén IV	DNP
	1 – Sí		
Acceso a acueducto	0 – No	Sisbén IV	DNP
	1 – Sí		
Número de días que llega el		Sisbén IV	DNP
Acceso a alcantarillado	0 – No	Sisbén IV	DNP
	1 – Sí		
Acceso a gas natural domiciliario	0 – No	Sisbén IV	DNP
	1 – Sí		
Acceso a recolección de basuras	0 – No	Sisbén IV	DNP
	1 – Sí		
Código DANE departamento			DANE
Código DANE municipio			DANE
Capital	0 – No		DANE
	1 – Sí		
Ubicación de la vivienda-Clase	1- Cabecera	Sisbén IV	DNP
	2- Centro poblado		
	3- Rural Disperso		
Coordenada 11			

Variable	Valores	Base de datos	Fuente
Coordenada 12			
Coordenada 12			
Coordenada 21			
Coordenada 22			
Coordenada 31			
Coordenada 32			
Coordenada 41			
Coordenada 42			
Ruralidad	1 - Ciudades y aglomeraciones	Sistema de ciudades	DNP
	2 - Intermedio		
	3 - Rural		
	4 - Rural disperso		
Porcentaje de población rural en el municipio		CENSO	DANE
Hectáreas sembradas de coca que tiene el municipio, en el año anterior de la encuesta		Base de datos de cultivos ilícitos	Observatorio de Drogas de Colombia
Porcentaje de población con alguna pertenencia étnica		CENSO	DANE
Porcentaje de población con alguna pertenencia indígena		CENSO	DANE
Tasa de fecundidad específica en mujeres entre 10 y 14 años, en el año anterior de la encuesta		CENSO	DANE
Tasa de fecundidad específica en mujeres entre 15 y 19 años, en el año anterior de la encuesta		CENSO	DANE
Tasa de homicidios por cada 100.000 niños entre 0 y 17 años, en el año anterior de la encuesta		FORENSIS	INMCLCF

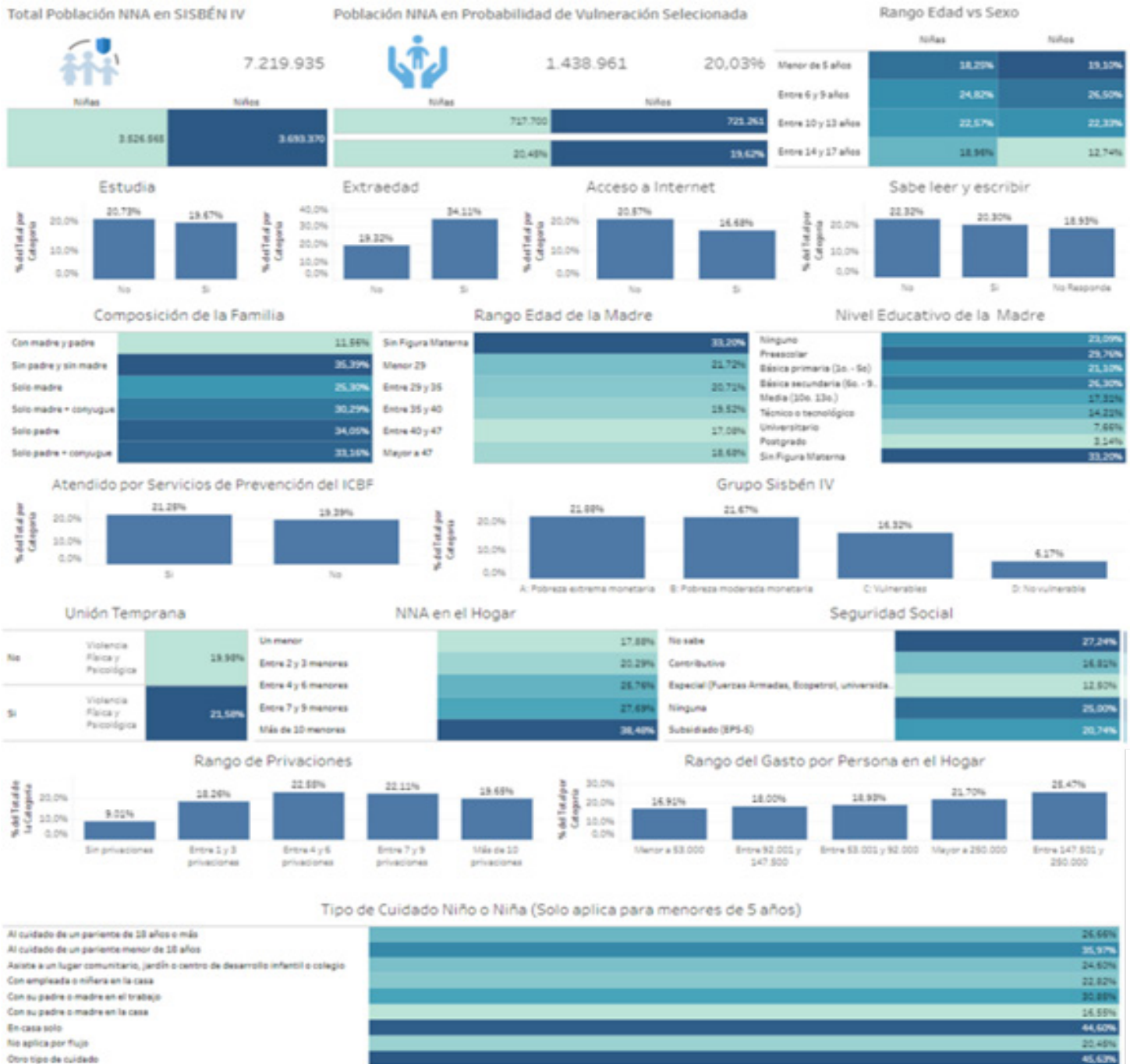
Variable	Valores	Base de datos	Fuente
Tasa de homicidios por cada 100.000 habitantes, en el año anterior de la encuesta		FORENSIS	INMCLCF
Tasa de suicidios por cada 100.000 niños entre 0 y 17 años, en el año anterior de la encuesta		FORENSIS	INMCLCF
Tasa de suicidios por cada 100.000 habitantes, en el año anterior de la encuesta		FORENSIS	INMCLCF
Tasa de violencia al interior del hogar por cada 100.000 niños entre 0 y 17 años, en el año anterior de la encuesta		FORENSIS	INMCLCF
Tasa de violencia al interior del hogar por cada 100.000 habitantes, en el año anterior de la encuesta		FORENSIS	INMCLCF
Tasa de violencia interpersonal por cada 100.000 niños entre 0 y 17 años, en el año anterior de la encuesta		FORENSIS	INMCLCF
Tasa de suicidios por cada 100.000 niños entre 0 y 17 años, en el año anterior de la encuesta		FORENSIS	INMCLCF
Tasa de violencia interpersonal por cada 100.000 habitantes, en el año anterior de la encuesta		FORENSIS	INMCLCF
Tasa de exámenes médico-legales por presunto abuso sexual por cada 100.000 niños entre 0 y 17 años, en el año anterior de la encuesta		FORENSIS	INMCLCF
Tasa de exámenes médico-legales por presunto abuso sexual por cada 100.000		FORENSIS	INMCLCF

Variable	Valores	Base de datos	Fuente
habitantes, en el año anterior de la encuesta			
Tasa de víctimas de delitos sexuales por presunto abuso sexual por cada 100.000 niños entre 0 y 17 años, en el año anterior de la encuesta		SPOA	Fiscalía General de la Nación
Tasa de víctimas de delitos sexuales por cada 100.000 habitantes, en el año anterior de la encuesta		SPOA	Fiscalía General de la Nación
Error			Calculada
Marca			Calculada
Estado			Calculada
Predicción			Calculada
Replicación			Calculada

Fuente: Elaboración propia

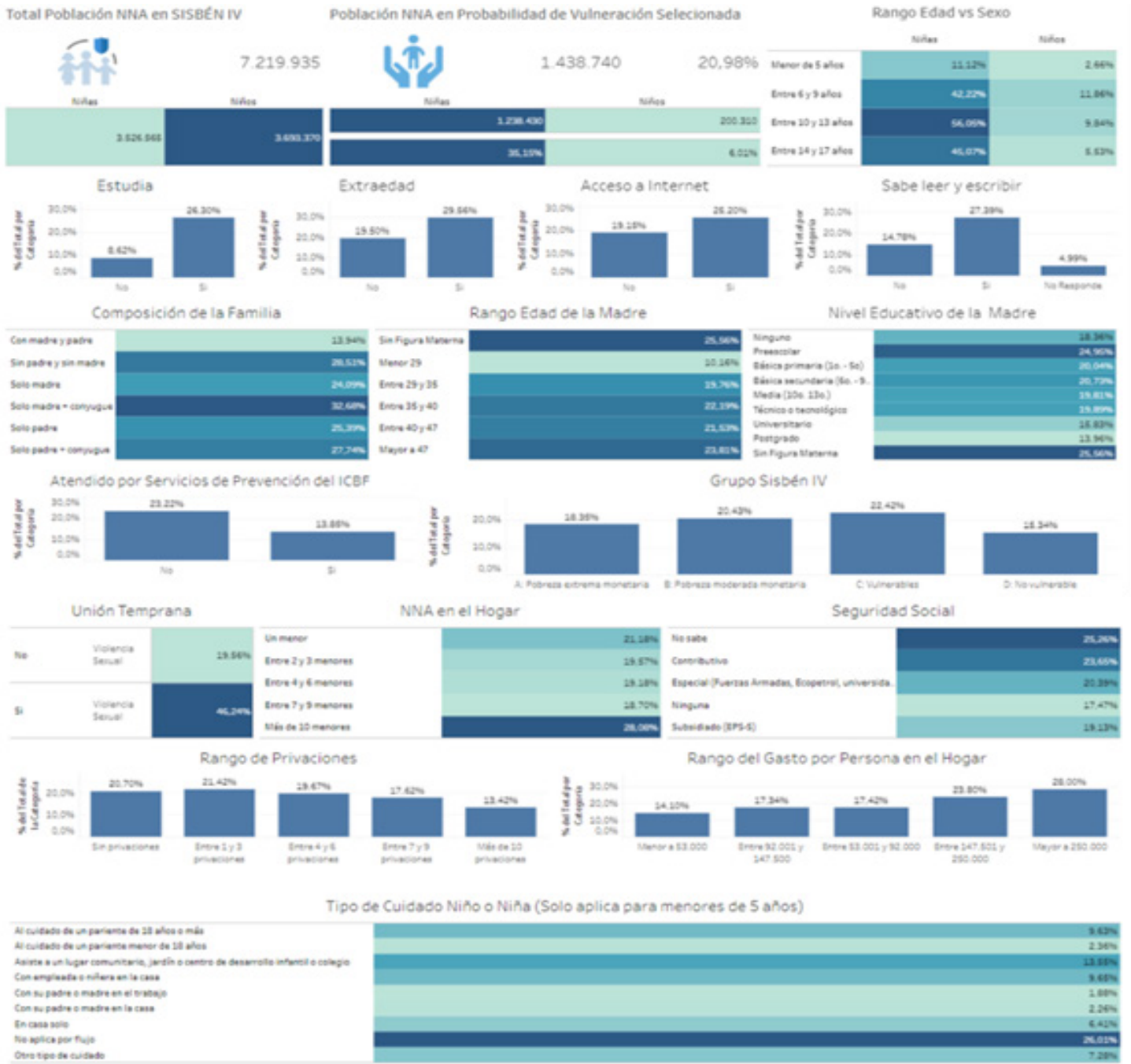


Anexo 2. Proporción de casos con riesgo de vulneración por violencia física, psicológica o negligencia (maltrato) estimado alto para el total nacional y respecto a variables seleccionadas



Fuente: Grupo de Analítica Institucional Dirección de Planeación y Control de Gestión ICBF

Anexo 3. Proporción de casos con riesgo de vulneración por violencia sexual estimado alto para el total nacional y respecto a variables seleccionadas



Fuente: Grupo de Analítica Institucional Dirección de Planeación y Control de Gestión ICBF



**BIENESTAR
FAMILIAR**



LÍNEA DE
ATENCIÓN A
NIÑOS, NIÑAS
Y ADOLESCENTES.
PROTECCIÓN - EMERGENCIA - ORIENTACIÓN

Línea gratuita nacional ICBF:
01 8000 91 80 80
www.icbf.gov.co

 ICBFColombia

 @ICBFColombia

 @icbfcolombiaoficial